UNIVERSITY OF LEICESTER

Isolation and characterisation of bacteriophages infecting *Legionella* spp.

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

by

Neda Nezam-Abadi, BSc, MSc

Department of Respiratory Sciences

University of Leicester

Leicester

United Kingdom

Supervised by

Dr. Edouard E. Galyov and Prof. Martha R. J. Clokie

March 2019

Abstract

Isolation and characterisation of bacteriophages infecting *Legionella* spp.

Neda Nezam-Abadi

Legionella spp. are waterborne pathogens that can cause Legionnaires' disease, a potentially fatal acute pneumonia. These pathogens can be resistant to many disinfectants and thermal treatments. Furthermore, detection of *Legionella* spp. is often difficult. Thus, infections with these organisms can be difficult to both diagnose and treat. Bacteriophages are a potential tool to be exploited to aid in the elimination and detection of *Legionella* spp. in both the environment and in patients, with bacteriophage based diagnostics and therapeutics already successfully developed for many organisms. However, little is known about the bacteriophages capable of infecting *Legionella* spp. Thus, with little knowledge available in the literature, the aim of this study was to obtain more information on *Legionella* spp. bacteriophages using both experimental and bioinformatic approaches. Here, a diverse range of *Legionella* spp. were collected either through isolation from 262 water and soil samples from natural reservoirs and man-made systems, or obtained from Public Health England. These strains were then characterised, used to isolate bacteriophages from the same 262 samples, and were exposed to prophage-inducing agents to isolate bacteriophages directly from the bacteria themselves. While no bacteriophages were isolated from the samples, virus-like particles were observed following induction of *Legionella* spp.. These particles were not like any currently identified bacteriophages, and warrant further investigation to confirm they are in fact bacteriophages. A bioinformatic approach was conducted in parallel to these investigations. Using this approach, a total of 717 *Legionella* spp. genomes in the GenBank database were screened for the presence of prophage-like regions. Although putative complete prophage elements were identified in the *Legionella* spp. genomes, further investigation concluded that these were due to ϕX-174 contamination that had not been removed by the submitting group before adding their genomes to GenBank. Other elements were discovered, but with little core recognisable bacteriophage genes. These results show that much more fundamental knowledge remains to be understood about *Legionella* spp. bacteriophages before extensive work is carried with the aim of developing bacteriophage diagnostics and therapeutics for *Legionella* spp. infections.

Acknowledgements

Finally, it is time to finish this chapter of my life, which took a total of 6.5 years (15 months on my first project and 5.3 years on the second, current, one). I cannot say that it was not fun, but it did take a lot of hard work and perseverance, as working on multiple jobs alongside a PhD is not the easiest way to do one. However, all the ups and downs during this journey have most definitely made me stronger.

I would not have been able to come this far without the support of my family and friends. My mum, who I have not seen for the last eight years, but whose unconditional love and encouragement has always helped me to look positively at any challenges/hardships. I cannot find the words to thank her and I am just delighted and proud of having her as my mum. To my brothers Mohammad, Babak, and especially Siamak, without whom I could not have been here today. Malakout who taught me how to believe in myself, embrace my values and become an independent person; I owe you so much! Radhika and Niloo for tolerating my craziness and being there for me whenever I got stuck in a challenging situation. Sara for being an amazing friend who I bothered a lot on a daily basis through the thesis writing process. I particularly want to thank Chris for constantly supporting me all the way through this journey and for helping me to survive my PhD; Don't worry, you won't hear "Is this a plaque?" every 5 minutes anymore! I would also like to thank the escapegoats gang (Bandar, Ananthi, Faizal, Mohammed, and Chris) who made me laugh and with whom I have shared a lot of fun times, Mercy and Sally for our morning pep talks. Thanks to all the members of lab 105-107/230, past and present.

I would also like to thank my supervisors Dr. Ed Galyov and Prof. Martha Clokie for their guidance. Dr. David Jenkins, Dr. Julian Clokie, Dr. Simon kilvington, Dr. Srwa Rashid, and Dr. Janet Nale for providing samples for use in isolating *Legionella* spp. and screening for bacteriophages. I would also like to thank Dr. Samuel Collins and other lab members at Public Health England for providing 37 *Legionella* spp. isolates. Prof. Andrea Cooper and Dr. John Pearl for their support and for putting up with me while working for them, and the members of their lab, Ganisha, Mrinal, and Ben, for distracting me by their large ordering lists that helped me to switch my mind off from my PhD every once in a while. I want to say thank you to all of you, along with those who have come into my life and left.

I would like to dedicate this thesis to my father. You may not be here with me in body, but you will always in my life in spirit, and will never forget you. You have always been one of the main reasons that kept me carrying on with my education. Thank you.

Contents

List of Tables

List of Figures

Abbreviations

ANI	Average nucleotide identify
BYE-α	Buffered yeast extract-α
BCYE-α	Buffered charcoal yeast extract-α
CDC	Centers for disease control
CDM	Chemically modified medium
CFU	Colony forming unit
dH$_2$O	Distilled water
DNA	Deoxyribonucleic acid
ds	Double-stranded
EPS	Extracellular polymeric substances
ESCMID	European Society of Clinical Microbiology and Infectious Diseases'
ESGLI	ESCMID Study Group for Legionella Infections
EWGLI	European Working Group for Legionella Infections
LCV	*Legionella*-containing vacuoles
LPS	Lipopolysaccharides
LTM	*Legionella* transparent medium
MICs	Minimal inhibitory concentrations
mip	Macrophage infectivity potentiator
MitC	Mitomycin C
MWY	Wadowsky and Yee
NCTC	National Collection of Type Cultures
NFX	Norfloxacin
NS	NanoSight
NTA	Nanoparticle tracking analysis
OD	Optical density

PCR	Polymerase Chain Reaction
RNA	Ribonucleic acid
PHAST	PHAge Search Tool
PHASTER	PHAge Search Tool Enhanced Release
PHE	Public Health England
pVOGs	Prokaryotic Virus Orthologous Groups
qPCR	Quantitative polymerase chain reaction
SBT	Sequence-based typing
ss	Single-stranded
TAE	Tris-acetate-EDTA
TEM	Transmission electron microscopy
UPH$_2$O	Ultra-pure water
UV	Ultra-violet
VBNC	Viable but nonculturable
WHO	World Health Organisation

1. Introduction

1.1 *Legionella* species

The genus *Legionella* is a group of facultative, intracellular, Gram-negative, pleomorphic bacilli with polar flagella that belong to the class Gammaproteobacteria (Taylor et al., 2009). These organisms are recognised as important nosocomial and community-acquired human pathogens, with infection by these organisms often associated with a considerable level of morbidity and mortality (Borges A., Simões M., Martínez-Murcia A., 2012). These bacteria can act as opportunistic pathogens in humans, where they can cause severe diseases, namely legionellosis.

Legionellosis is the collective name for two individual *Legionella*-associated conditions, Legionnaires' disease and Pontiac fever (Fields et al., 2002; Bartram et al., 2007). Legionnaires' disease is a severe, acute form of pneumonia, and can often have a high mortality rate, with death seen in approximately $10 - 30$ % of individuals (Phin et al., 2014). Pontiac fever on the other hand is a non-fatal, mild flu-like disease that exhibits symptoms after incubation time of ~ 36 hours (Edelstein, 2007). Legionnaires' disease is named after its first diagnosis, during a 1976 outbreak in Philadelphia, USA, among attendees at the 58th annual convention of the American Legion, celebrating the bicentennial of the Declaration of Independence (Fraser et al., 1977; Winn, 1988). Pontiac fever is similarly named after an outbreak, namely an outbreak of acute fever at a health department facility in Pontiac, Michigan, USA in 1968 (Glick et al., 1978).

Legionella spp. are aquatic bacteria, found in natural water systems such as lakes and rivers (Carvalho et al., 2008; Parthuisot et al., 2010). However, they are frequently found in man-made water systems, including both industrial systems, such as cooling towers, and domestic systems, such as showers, baths, and air conditioning units (Castilla et al., 2008; Goutziana et al., 2008; Kuroki et al., 2017). These man-made water systems create ideal environments to support the growth of *Legionella* spp. as temperatures of $\sim$ 32 - 42°C favour their growth (Yee and Wadowsky, 1982). Infections by *Legionella* spp. most commonly occur following inhalation of *Legionella*-containing water-borne aerosols from these man-made systems (Hines et al., 2014), which was the case in both the 1976 Philadelphia and 1968 Pontiac outbreaks, where spread of the bacteria was found to be caused from contaminated air conditioning

systems in both cases (Glick et al., 1978; Abu Kwaik et al., 1998).

Several factors impact the probability of developing Legionnaires' disease and Pontiac fever. The chances of developing Legionnaires' disease are increased by factors such as the individuals smoking status, whether they have other pre-existing conditions such as chronic obstructive pulmonary disease, diabetes, immunodeficiency, and their age, with older age (> 50 years) more associated with the development of Legionnaires' disease (Farnham et al., 2014; Peabody et al., 2017). However, these factors do not increase the chance of developing Pontiac fever in human, instead is witnessed primarily in younger individuals, with a median age of ~ 30 (Tossa et al., 2006).

Within *Legionella* spp., cases of legionellosis are primarily associated with one particular species, *L. pneumophila*, which causes ~ 90 % of all *Legionella*-associated disease (Newton et al., 2010). This species itself can be sub-classified into 14 serogroups, by using polyvalent or monoclonal antisera targeting distinct lipopolysaccharide structures (Ciesielski et al., 1986; Yong et al., 2010; Lück et al., 2013). Of these, *L. pneumophila* serogroup 1 is the most well-known and studied species of *Legionella*, as it is responsible for the vast majority of *L. pneumophila* disease cases. For example, an analysis of *Legionella* spp. isolated from legionellosis patients in Japan from 2008 - 2016 observed that 98.1 % of all isolates were *L. pneumophila*, while 87.1 % of all isolates belonged to *L. pneumophila* serogroup 1 (Amemura-Maekawa et al., 2018). *L. pneumophila* serogroup 1 is also notable as members of this serogroup were the causative serogroup of the 1976 American Legion Legionnaires' disease outbreak (McKinney et al., 1979).

There are ~ 60 other species within the genus *Legionella*, some of which are also important clinically, such as *Legionella longbeachae* and *Legionella micdadei* (also known as *Tatlockia micdadei*) (Donlan et al., 2002; Brenner, 2018). For example, in 1981, *L. longbeachae* was first isolated from the respiratory tracts of the patients who were hospitalised in Long Beach, California, USA suffering with pneumonia, with *L. longbeachae* identified as the causative organism (McKinney et al., 1981). In some parts of the world, *L. longbeachae* can even be as important of a pathogen as *L. pneumophila*, such as in New Zealand where it can cause as much as 30 - 50 % of all *Legionella*-associated disease (Graham et al., 2012). *L. micdadei* on the other hand is often seen in sporadic outbreaks. For example, in 1988, *L. micdadei* was associated with an outbreak of Pontiac fever in Lochgoilhead, UK, with the source found to be a leisure complex whirlpool spa (Goldberg et al., 1989). In a more severe case, an outbreak of pneumonia among transplant patients was found to be due to *L. micdadei* isolated from hot water supplies within a hospital in Manhattan, New York, USA (Knirsch et al., 2000). While also able to cause disease for most examples, the remainder of the *Legionella* spp. have more frequently been found in the environmental samples, and are rarely implicated in human disease (Fields et al., 2002; Yu et al., 2002; Joseph et al., 2016).

1.1.1 Epidemiology

After the identification of *L. pneumophila* in the 1970s, surveillance schemes for Legionnaires' disease were formed to monitor the disease in numerous countries (Phin et al., 2014). However, regardless of all the attempts to provide better surveillance scheme, the global incidence of infections by *Legionella* spp. is still poorly documented because of factors such as the differences in awareness levels, diagnostic methods and reporting among countries, as well as difficulties in confirming microbiological diagnoses, that likely leads to underestimations in its incidence globally (Burillo et al., 2017). In England, monitoring is conducted by Public Health England (PHE), and the numbers of reported cases of Legionnaires' disease, in England and Wales, and Scotland (from Health Protection Scotland), from 2000 - 2016 can be found in figure 1.1.
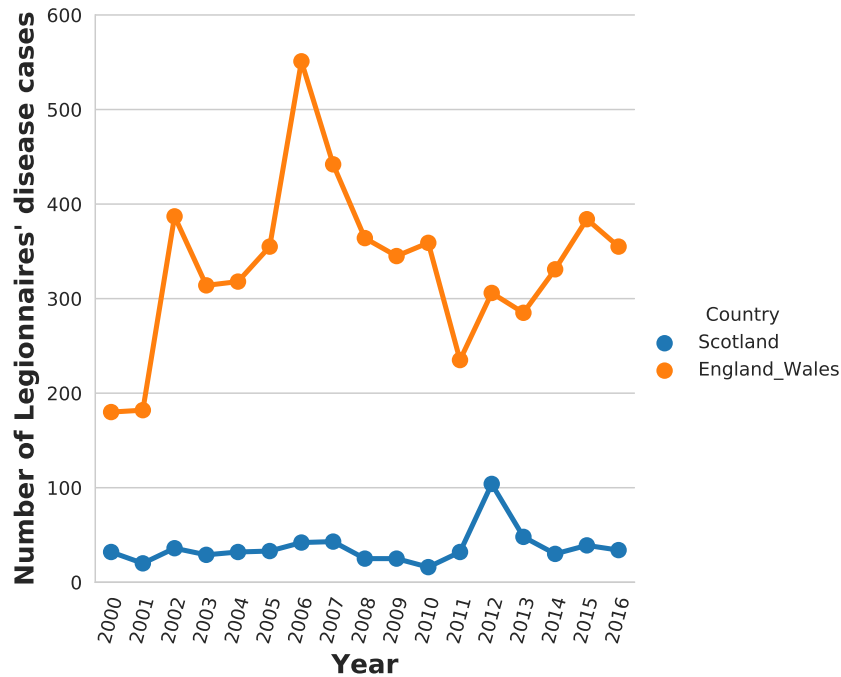
Figure 1.1. Number of confirmed cases of Legionnaires' disease by year of onset of symptoms in Scotland, and England and Wales between 2000 - 2016. Total number of confirmed cases of Legionnaires' disease in Scotland (blue line), and England and Wales (orange line) between 2000 and 2016, where each dot represents a value for the corresponding year. All values for Scotland were obtained from data published by Health Protection Scotland, while for England and Wales values come from data published by Health Protection Agency (2000 - 2011), Public Health England (2012 - 2016).

Among all the confirmed Legionnaires' disease cases reported by PHE for England and Wales from 2000 - 2016, between ∼ 20 - 50 of these cases resulted in death of the patient (Figure 1.2). A report by the World Health Organisation (WHO) estimates that the mortality rate of this disease can be 5 - 10 % (World Health England, 2018), which correlates well with the figures observed by PHE (Figures 1.1

and 1.2). However, misdiagnosis of this disease and any delay in initiating the treatment can increase this rate up to 20 % (Saleri and Ryan, 2018). Death rates are the worst in immunosuppressed patients, where as many as 40 – 80 % who contract Legionnaires' disease die from the condition (World Health England, 2018).
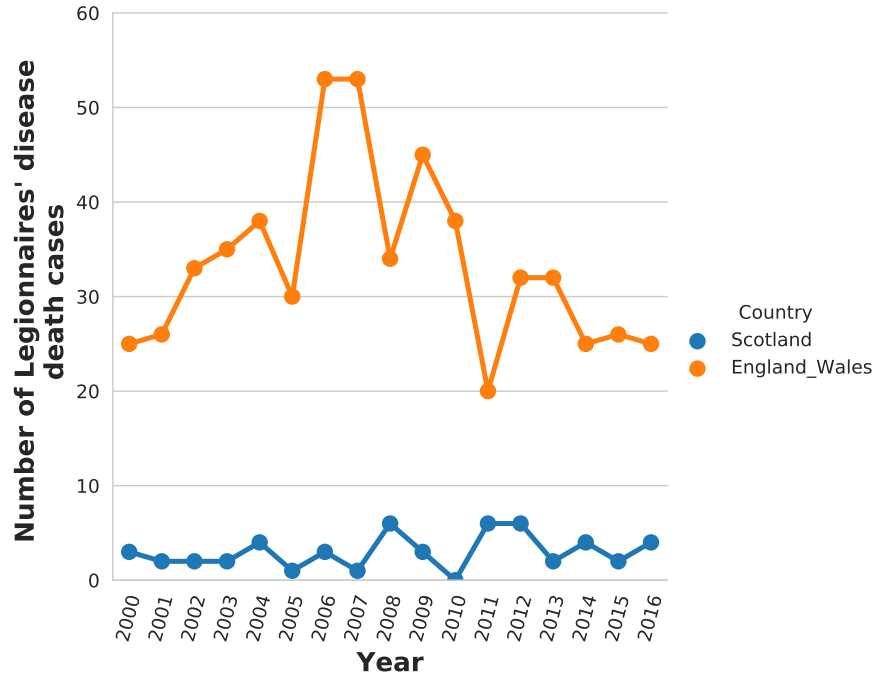
Figure 1.2. Number of fatalities from Legionnaires' disease by year in Scotland, and England and Wales between 2000 - 2016. The blue line shows the number of death cases in Scotland, while the orange line represents the data for England and Wales. All data for Scotland were obtained from data published by Health Protection Scotland, while for England and Wales values come from data published by Health Protection Agency (2000 - 2011), Public Health England (2012 - 2016).

Incidence of Legionnaires' has been correlated with season, with reported cases increasing during the warmer months of the year and decreasing when it is colder (Beauté et al., 2013). In England and Wales, this trend has been stable over the last $\sim$ 15 years (figure 1.3), with the peak number of cases presenting between July and September. Similar trends have been observed for legionellosis in other countries, such as New Zealand, with reports beginning to increasing during spring, which is has been speculatively linked to increased outdoor activity, such as gardening as *Legionella* spp. can often be found in gardening materials like potting soils and compost (Casati et al., 2009, 2010; Graham et al., 2012). In other cases increased occurrence of *Legionella*-related disease during these months could possibly be due to greater survival and replication of *Legionella* spp. in the higher temperatures, allowing increased contact with the man-made systems like spas, ventilation and air conditioning units, and decorative water fountains that then spread the organisms.
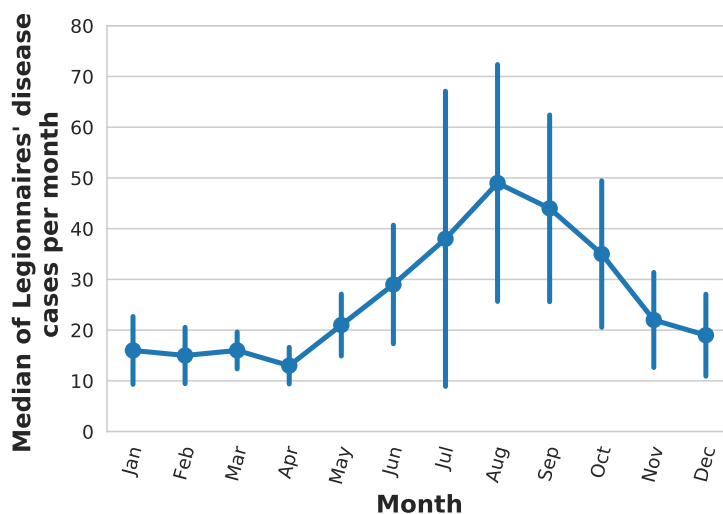
Figure 1.3. Number of confirmed cases of Legionnaires' disease by month in England and Wales from 2000 - 2016. Each point shows the median number of confirmed cases of Legionnaires' disease for a given month between 2000 and 2016 in England and Wales. Error bars indicate ± standard deviation (SD).

1.1.2 Life cycle of *Legionella* spp.

Legionella spp. can replicate and survive in an array of environments through use of a number of replication strategies, best characterised in *L. pneumophila*. In nature, for *L. pneumophila*, these replication strategies can consist of either through planktonic replication, as inhabitants of a biofilm, or by replicating intracellularly inside eukaryotic cells such as protozoa or alveolar macrophages (Figure 1.4) (Byrne and Swanson, 1998; Molofsky and Swanson, 2004; Taylor et al., 2009; Borges A., Simões M., Martínez-Murcia A., 2012; Gomez-Valero and Buchrieser, 2013).

When growing extracellularly, *L. pneumophila* follows a biphasic life cycle (Figure 1.4 B), allowing them to survive under different environmental conditions. When nutrients are sufficient, *L. pneumophila* cells are in the replicative (non-virulent) phase, where they are non-motile, non-flagellated cells that are more sensitive to stresses, and freely able to replicate (Steinert et al., 2002). However, in nutrient poor environments, cells enters the transmissive (virulent) phase, where they become motile, flagellated cells, that are more resistant to stress, where they are unable to replicate until finding a nutrient-rich environment again (Steinert et al., 2002; Newton et al., 2010).

Legionella spp. are also able to grow intracellularly inside cells such as protozoa in environmental niches, or even within macrophage in a human host (Figure 1.4 D and E) (Rowbotham, 1980; Nash et al., 1984). In the environment *Legionella* spp. can replicate in a number of protozoan hosts, such as *Hartmannella*, *Acanthamoeba*, *Naegleria*, and *Tetrahymena*, that are commonly found in aquatic systems (Rowbotham, 1980; Tyndall and Domingue, 1982; Barbaree et al., 1986) (Figure 1.4 D).

5

Inhalation of aerolised droplets containing *Legionella* spp. on the other hand allows *Legionella* spp. to encounter human alveolar macrophages, which can act as another growth niche for the bacteria (Baskerville et al., 1981). Intracellular replication allows *Legionella* spp. to survive under harsh conditions, such as nutrient deprivation, disinfectants (chlorine), and when water temperatures are elevated (Rowbotham, 1980; Wadowsky et al., 1988; Storey et al., 2004; Dupuy et al., 2011; Price et al., 2011).

Most of the steps for intracellular growth in both amoebae and lung alveolar macrophages are very similar. In both cases, the intracellular life cycle begins with the binding of the bacterium to the cell surface, where it then enters the cell through in a process called 'coiling phagocytosis'. Here, pseudopods, projections from the phagocyte cell membrane, coil around the *Legionella* spp. cells multiple times leading to internalisation of the bacteria into the phagocytic cell, leading to the establishment of *Legionella*-containing vacuoles (LCV) (Horwitz, 1984; Bozue and Johnson, 1996). LCV are able to act as a replication niche in both amoebae and macrophage through the bacteria's ability to evade bactericidal actions against it by remodelling the LCV surface (Swanson and Isberg, 1995; Kagan and Roy, 2002). *Legionella* spp. achieve this through use of the Dot/Icm type IVB secretion system (defective organelle trafficking/intracellular multiplication) that allows secretion of around 300 *Legionella* spp. effector proteins into the host cell cytosol (Rêgo et al., 2010; Segal, 2013). These effectors are then believed to interrupt cellular processes, including the phagocyte-lysosome pathway, while providing further safety by recruiting host cellular components such as endoplasmic reticulum to the vacuole surface, blocking attachment of bactericidal agents e.g. lysozyme (Hoffmann et al., 2014; Hilbi et al., 2017). Bacteria are then able to proliferate in these vacuoles, continuing until nutrient levels decreases. At this point a transition happens, where they become flagellated and virulent, before being released from the host eukaryote (Alli et al., 2000; Molmeret et al., 2010). *Legionella* spp. leave their host cells by lysing them through either necrosis for amoeba or pyroptosis for alveolar macrophage cells, both of which lead to pore formation on the eukaryotic membrane, leading to release of the bacterial cells (Gao and Abu Kwaik, 2000; Silveira and Zamboni, 2010).

Replication involving biofilm formation is another approach that can be taken by *Legionella* spp. (Figure 1.4 C). Biofilms are aggregates of bacterial cells within a matrix of extracellular polymeric substances (EPS), which consists of polysaccharides, DNA, and numerous glycomolecules produced by the bacteria (Flemming et al., 2007). Biofilm formation starts by the attachment of a bacterium to a surface, followed by bacterial replication to form microcolonies as well as production of EPS molecules like exopolysaccharides to form a three-dimensional biofilm (Donlan, 2001; Hall-Stoodley et al., 2004). However, when environmental and nutritional conditions change and become unfavourable for maintenance of the biofilm, some of the bacteria may detach from the biofilm by producing matrix-degrading enzymes, allowing them to disperse into the environment to find a new surface to establish a new biofilm (Watnick and Kolter, 2000; Petrova and Sauer, 2016). In the case of *L. pneumophila*, it
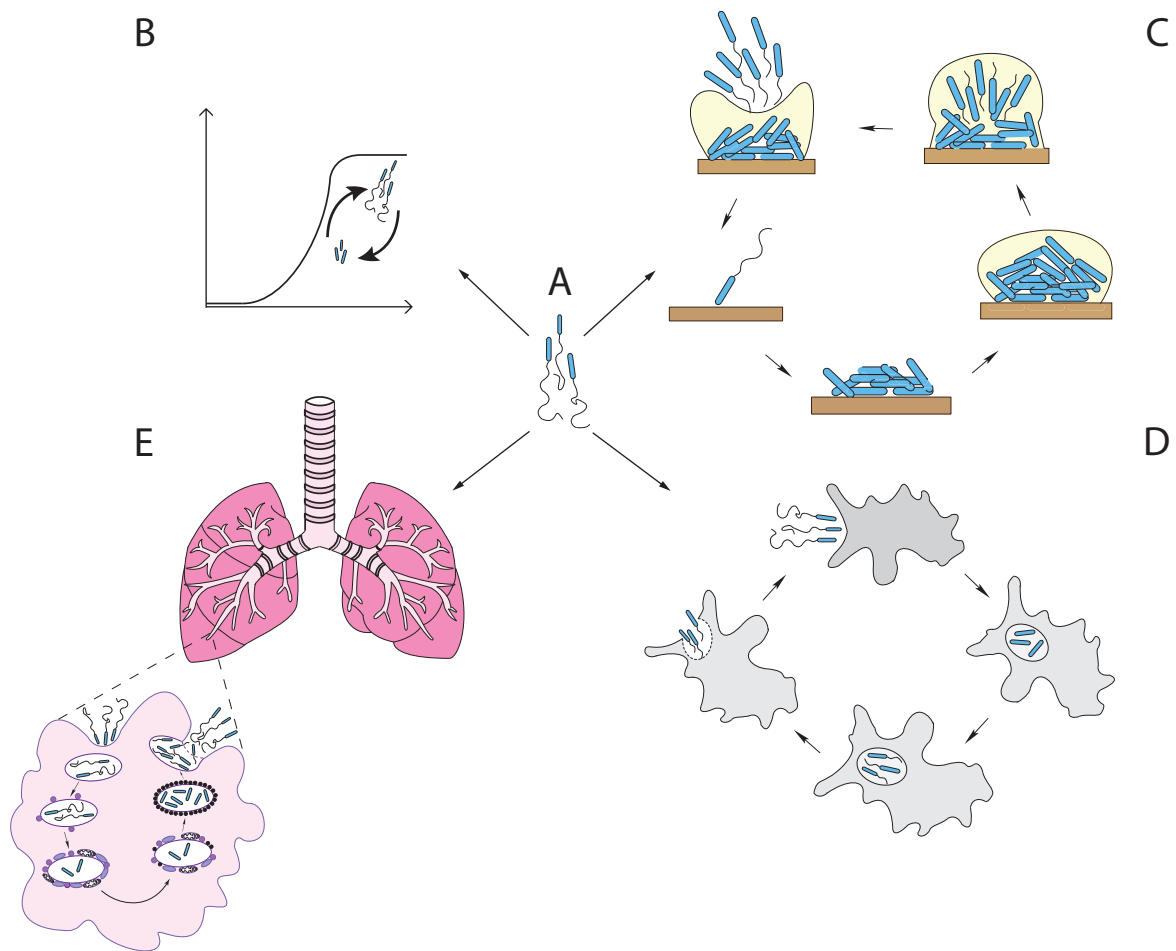
Figure 1.4. Overview of the possible replication strategies used by _L. pneumophila._
(A) The transmissive _Legionella_ spp. can exist extracellular environment planktonic cells, with the possibility of entering distinct replicative pathways depending on the environment. (B) In aqueous environments, _L. pneumophila_ can follow a biphasic life cycle, where it switches between the replicative and transmissive forms depending on factors such as nutrient availability. (C) _L. pneumophila_ may form/join biofilms, either in natural or artificial aquatic habitats. (D) Transmissive _L. pneumophila_ can be engulfed by eukaryotic cells such as protozoa when in nature, where the bacteria then replicates inside the eukaryote. (E) Internalisation within alveolar macrophages can likewise lead to intracellular replication.

has been shown that biofilms count as one of the major reservoir for _Legionella_ spp. within man-made water systems (Carvalho et al., 2007; Serrano-Suárez et al., 2013).

Another state that _Legionella_ spp. can use is the viable but non-culturable (VBNC) state. In aquatic environments, stress can also induce _Legionella_ spp. to enter the VBNC state as a survival strategy (Hussong et al., 1987; Oliver, 2010). During this state cells are at their lowest level of metabolic activity and do not divide, halting bacterial growth (Shleeva et al., 2002). While these cells remain viable, they have lost their culturability, meaning they cannot be recovered through culturing (Oliver, 2000). Switching to this state is often triggered by conditions such as low nutrient availability, chemical

stresses e.g. disinfectants, or exposure to heat stress (Li et al., 2014). Cells that have undergone this switch are described as VBNC, while the cells are viable, they can no longer be cultured under standard laboratory conditions, leading to difficulties in their detection (Oliver, 2000; Tronel and Hartemann, 2009). However, VBNC *Legionella* spp. can be brought back to the culturable state, as has been shown experimentally following passage of the VBNC *Legionella* spp. through eukaryotic cells such as amoebae (Steinert et al., 1997; García et al., 2007).

1.1.3 Treatment and prevention of *Legionella* spp. infections

As mentioned earlier, man-made water systems, e.g. cooling towers and hot-water supplies, that are able to produce aerosols are considered high risk sources of *Legionella* spp. infection (Bartram et al., 2007; Burillo et al., 2017). To prevent potential outbreaks from these sources various treatments are used to reduce/remove contaminating *Legionella* spp. Treatments include intermittent temperature increase to 60°C, continuous chemical treatment e.g. chlorine and hydrogen peroxide, regular flushing of the water systems, and cleaning of sediment from water storage tanks (Fields et al., 2002; Kim et al., 2002; Marchesi et al., 2016). However, no single treatment will normally lead to total elimination of the bacterium. For example, chemical agents such as chlorine are a common treatment used in disinfection protocols, and have been shown to be effective at removing planktonic *L. pneumophila* (Cooper and Hanlon, 2010). However, when these organisms exist within biofilms they often show decreased sensitivity to chemical treatments (Cargill et al., 1992). Similarly, *Legionella* spp. have a reduced sensitivity to chemical treatment when growing intracellularly within protozoa. For instance, García et al. (2007) observed that when co-cultured with *Acanthamoeba polyphaga*, *L. pneumophila* were ∼ 5x more resistant to sodium hypochlorite. Similar results have been observed by other groups, including for interactions between *Legionella* spp. and other protozoa (Kilvington and Price, 1990; Donlan et al., 2005). Along with biofilm formation and intracellular replication in protozoa, other factors make the removal of *Legionella* spp. especially difficult when the organisms are under nutrient-deprived conditions. Under these conditions *Legionella* spp. are also resistant to heat treatments, and also become highly motile, meaning they can disseminate throughout the water system (Barker et al., 1992; Abdel-Nour et al., 2013; Falkinham, 2015) (see Section 1.1.2). Therefore, alternative means of water disinfection are required.

Although inhalation of *Legionella*-containing droplets has the potential to lead to development of either Pontiac fever or Legionnaires' disease in an individual, intervention is normally only required for one of these conditions. There is no real need for any specific clinical interventions for patients dealing with Pontiac fever, as this tends not to be a life-threatening condition (Lim et al., 2009; Phin et al., 2014). However, in the case of Legionnaires' disease, urgent action is required to combat this type of pneumonia (Wever et al., 2000). Treatment for Legionnaires' disease involves treatment with

antibiotics, including macrolides, fluoroquinolones, and rifamycins Wilson et al. (2018). Antibiotic resistance in *Legionella* spp. is uncommon, and has not been observed clinically, although as with all bacteria this still remains a possibility, and development of resistance has selected for *in vitro* (Bruin et al., 2014; Sikora et al., 2017; Massip et al., 2017). In the UK, upon diagnosis of this disease, mild cases of community-acquired pneumonia are treated with oral fluoroquinolones or macrolides, and in severe cases fluoroquinolones are recommended, possibly in combination with either macrolides or rifampicin (Lim et al., 2009). However, before *Legionella* spp. infection is confirmed, treatment of broad-spectrum antibiotics such as β-lactam antibiotics is recommended (Lim et al., 2009). Yet β-lactam are often ineffective in treating *Legionella*-associated infections, as *Legionella* spp. encode β-lactamase enzymes that inhibit their effect (Fu and Neu, 1979; Cunha et al., 2016). Thus, rapid diagnoses of *Legionella* infections are important to ensure correct treatments are applied.

However, Legionnaires' disease can often be difficult to diagnose, as symptoms can easily be misdiagnosed as other types of pneumonia (Swartz, 1979; Murdoch, 2003). When a patient is suffering from severe pneumonia, it is recommended that the patient is screened to determine if *Legionella* is cause of the pneumonia. For this, the patients urine is screened using the *Legionella* urine antigen test, which works by detecting *Legionella* lipopolysaccharide, providing the result on the same day (Berdal et al., 1979; Tilton, 1979; Lim et al., 2009). However, this test only detects infections caused by *L. pneumophila* serogroup 1. Thus, if this test is negative then culture will be relied upon, which can take a number of days for the *Legionella* spp. to grow to be identified. Therefore, new broader and rapid diagnostic assays would be necessary to detect new and emerging *Legionella* spp. and assess their potential to cause disease.

1.2 Bacteriophages

Bacteriophages are one possibility that may prove useful in the detection and elimination of *Legionella* spp. in both environmental and clinical settings. Bacteriophages are viruses that exclusively infect bacteria and propagate by utilising their host's metabolic machinery. They are the most abundant and diverse organisms on the planet, with an estimated 10^{31} bacteriophages in the whole biosphere, existing in all habitats where bacteria are present (Clokie et al., 2011; Hatfull and Hendrix, 2011). This means that in these environments bacteriophages often outnumber their bacterial hosts, with approximately 10 to 100 bacteriophage particles per bacterial cell (Clokie et al., 2011; Hatfull and Hendrix, 2011; Díaz-Muñoz and Koskella, 2014). By outnumbering bacteria, bacteriophages can have important roles in the ecology of bacteria through their interactions as predators, parasites, or as mutualists (Braga et al., 2018).

As predators, bacteriophages are able to lyse bacterial cells, altering the density of bacterial popula-

tions. For example, in nature, bacteriophage lysis of bacteria plays an important role in shaping the total prokaryote biomass in the oceans, where they lyse $\sim$ 15 - 40 % of the ocean's bacteria everyday (Suttle, 1994; Danovaro et al., 2011). This lysis results in the release of organic carbon from the bacteria into the ocean, a significant proportion of all biological carbon recycling in the ocean (Wilhelm and Suttle, 1999; Clokie and Mann, 2006). The impact of predatory role of bacteriophages can also be seen in industries involving the use of bacterial fermentation, e.g. dairy production, where contamination of these with bacteriophages can lead to significant levels of lysis of the fermenting bacteria, reducing their productivity (Brüssow, 2001). Bacterial lysis by bacteriophages is also thought to participate in altering genetic diversity within and between bacterial population, causing a large selection pressure on the bacterial population to adapt to avoid infection and death (Koskella, 2013). Adaptation will result in forming new resistant genotypes that are distinct from the original susceptible bacterial cells (Lenski, R. E. and Levin, 1985; Scanlan et al., 2015).

The parasitic relation between bacteriophages and bacteria is when bacteriophage DNA integrate into the bacterial genome, staying initially dormant as elements called prophages (explained in more detail in Section 1.2.2). Here, the bacteriophage replicates via the hosts own replicative processes, being replicated as part of the bacterial chromosome, but at the same time still has the ability to begin producing new viral cells that will eventually lead to lysis and death of the host bacteria (Mills et al., 2013). The impact of this can be very high, with prophage-associated bacterial lysis believed to contribute greatly to altering *Pseudomonas aeruginosa* population densities in cystic fibrosis (James et al., 2015).

Bacteriophages can also have a mutualistic relationship with their bacterial hosts when integrated into the host chromosomes as prophages. For example, bacteriophages can contain genes that provide beneficial characteristics to their hosts fitness that, following infection and integration, allows the bacteria to gain new characteristics that allow it to have a better chance of survival in certain environments, therefore also boosting the chances of the bacteriophage survival (Casjens, 2003). In the case of *Bacillus anthracis* for instance, prophages have been observed to provide long-term colonisation benefits to the bacterial population when colonising soil and the earthworm gut, by promoting sporulation, producing exopolysaccharides, and by aiding in the formation of biofilms (Schuch and Fischetti, 2009).

1.2.1 Bacteriophage classification

Bacteriophages are known as the largest and most diverse viral group and since the first applications of electron microscopy and negative staining to bacteriophages in 1959, over 6000 bacterial viruses have been described morphologically (Clokie et al., 2011; Ackermann and Prangishvili, 2012). Established in 1966, the International Committee on Taxonomy of Viruses (ICTV) is currently the only international body dealing with virus taxonomy (Fenner, 1995; Adriaenssens and Brister, 2017). Over the years

their work has led to important standardisation of viral taxonomy, including providing guidelines for classifying newly discovered viruses based on criteria such as the viruses' morphology and type of nucleic acid, among others (Ackermann, 2009; Salmond and Fineran, 2015). Bacteriophages cover a range of morphologies including tailed, polyhedral, filamentous, and pleomorphic virions, to even viruses with lipid or lipoprotein envelopes (Fauquet, 2008; Ackermann, 2009). Genetically, the nucleic acid composition of their genomes can either take the form of double-stranded (ds) or single-stranded (ss), DNA or RNA, with genome sizes ranging from $\sim$ 2.5 Kb (*Leuconostoc* bacteriophage L5) to almost 500 Kb (*Bacillus megaterium* bacteriophage G) (Hatfull, 2008; Salmond and Fineran, 2015; Mavrich and Hatfull, 2017). The makeup of these genomes is similarly diverse, with no ubiquitously conserved genes across all bacteriophages genomes, unlike the genomes of bacteria (i.e. 16S rRNA) (Shapiro and Putonti, 2018).

Of the bacteriophages described by the ICTV classification schemes, the order of dsDNA tailed bacteriophages, the *Caudovirales*, are the most commonly observed (covering over 95 % of all identified bacteriophages) (Abedon, 2009; Ackermann, 2011). Originally consisting of 3 families, this order has recently expanded to contain 4 families following the 2017 ICTV update, encompassing 22 subfamilies and 166 genera (Lavigne et al., 2012; Adriaenssens et al., 2018). The 3 original families within the *Caudovirales* are the *Myoviridae*, *Siphoviridae*, and *Podoviridae*, with the recent addition being the *Ackermannviridae* which were originally within the *Myoviridae* family (Lavigne et al., 2012; Adriaenssens et al., 2018). Morphologically the members of this order are distinguished based on their tail structures, with *Myoviridae* (e.g. T4) and *Ackermannviridae* (e.g. ViI) possessing contractile tails, while *Siphoviridae* (e.g. λ) and *Podoviridae* (e.g. T7) both have non-contractile tails. Tail length is used to distinguish *Siphoviridae* and *Podoviridae* with long tails classified as *Siphoviridae* and short tails *Podoviridae* (Lavigne et al., 2012). *Myoviridae* and *Ackermannviridae* are distinguishable structurally by their tail tip complex, as *Ackermannviridae* have a unique tail tip complex that is not found in other bacteriophages, composed of "prongs" and "stars" (Adriaenssens et al., 2012).

The genome of tailed bacteriophages can range from $\sim$ 11.5 Kb in *Mycoplasma* bacteriophage P1 to $\sim$ 500 Kb in *B. megaterium* bacteriophage G (Seaman and Day, 2007; Hatfull and Hendrix, 2011). While showing vast sequence diversity, a common gene architecture is generally maintained, with variations of genes corresponding to bacteriophage infection, replication, assembly, and propagation all generally present in a bacteriophage genome (Brüssow and Hendrix, 2002; Mavrich and Hatfull, 2017). For example, Grose and Casjens (2014) observed that while comparison of the genomes of 337 tailed bacteriophages could produce 56 distinct clusters of similar bacteriophages, they observed that homologues for head assembly proteins, terminases, portal proteins and major capsid proteins existed between all clusters, while bacteriophage components involved in DNA replication and lysis (e.g. DNA polymerases and holins respectively) were found in some but not all clusters. An example of the genome components and arrangements for bacteriophage λ can be found in Figure 1.5.
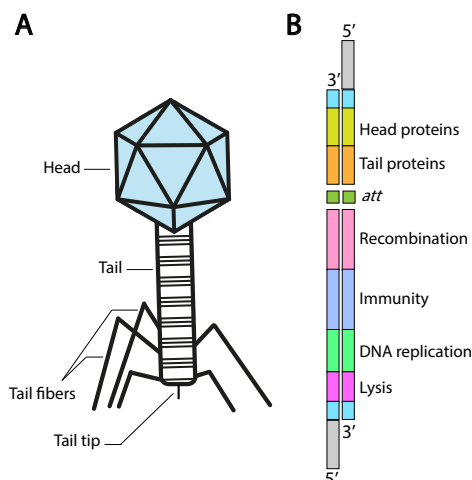
Figure 1.5. Schematic structure of bacteriophage λ genome arrangement. (A) illustrates the physical shape and some examples of structural components for bacteriophage λ and (B) shows the organisation of the bacteriophage λ genome. Figure is adapted from (Tropp, 2012).

Bacteriophage groups beyond the *Caudovirales* include other dsDNA bacteriophages such as the *Corticoviridae*, *Tectiviridae*, and *Plasmaviridae*, ssDNA bacteriophages such as the *Inoviridae* and *Microviridae*, ssRNA viruses such as the *Leviviridae*, and dsRNA viruses like the *Cystoviridae* (Oksanen and Bamford, 2012b,a; Maniloff, 2012; Day, 2012; Cherwa and Fane, 2012; van Duin and Olsthoorn, 2012; Poranen and Bamford, 2012). Similar to the *Caudovirales* members of these families can have important implication on bacterial ecology. For example, members of the *Inoviridae*, the filamentous bacteriophages, are known to be important in biofilm formation in many bacteria, e.g. *P. aeruginosa* (Rice et al., 2009). However, as many of the bacteriophages outside the *Caudovirales* can be difficult to isolate and culture, their dynamics are considerably understudied in comparison to the *Caudovirales*. Despite this, some of these bacteriophages taxa have contributed to important scientific advances, e.g. the *Leviviridae* MS2 and *Microviridae* φX174, which were the first RNA and DNA genomes sequenced in 1976 and 1977, respectively, with MS2 actually being the first genome ever sequenced (Fiers et al., 1976; Sanger et al., 1978).

1.2.2 Bacteriophage life cycles

Bacteriophages can replicate through a number of distinct life cycles including the lytic, lysogenic, and chronic cycles (Abedon, 2008; Drulis-Kawa et al., 2012; Lorenz et al., 2016). In all bacteriophages life cycles, infection begins by the adsorption of bacteriophages to a target bacterial cell, then followed by injection of their viral DNA into the adsorbed host. From here, the processes of the bacteriophage life cycles then diverge.

In the lytic cycle, injection is proceeded by utilising the hosts internal machinery to generate new

viral particles, with these particles then released from the host by lysis of the bacterial host cell (Figure 1.6 A). Bacteriophages that replicate exclusively through the lytic cycle are called virulent bacteriophages, and while most obligate lytic bacteriophages studied are members of the *Caudovirales*, members of other bacteriophage families can also replicate exclusively through the lytic cycle, with the entire RNA-based *Cystoviridae* and *Leviviridae* believed to do so (Hobbs and Abedon, 2016).

In the lysogenic cycle, used by temperate bacteriophages, rather than immediately creating new virions within the cell bacteriophages remain within the bacteria by integrating their DNA with the host chromosome, existing as elements called 'prophage' (Knoll and Mylonakis, 2014). In λ for example, the expression of bacteriophage genes required for the lytic pathway are prevented by the action of the cI repressor. The cI repressor protein is responsible for maintaining the lysogenic state and inhibiting the lytic development of the prophages that are already integrated into the chromosome by repressing the two lytic promoters, *pR* and *pL* (Lee et al., 2018; Sarkar-Banerjee et al., 2018). Integration into the chromosome by bacteriophages can either be the result of site-specific recombination between two specific attachment sites, one on the bacterial chromosome (*attB*) and the other on the bacteriophage genome (*attP*) (as seen in bacteriophage λ) (Fogg et al., 2014), or through random transposition (as seen in bacteriophage Mu) (Harshey, 2014). Integrated prophages allow the bacteriophages to replicate alongside the bacteria, during the hosts normal DNA replication process (Figure 1.6 B). Prophages can play a vital role in their hosts' fitness, enhancing characteristics such as their nutrient utilisation, and even altering their virulence (Fortier and Sekulovic, 2013). For instance, Wang et al. (2010) observed in *Escherichia coli*, that prophages increase the hosts resistance to oxidative stress, acidic environments, osmotic stress, and even increased resistance to β-lactam antibiotics and biofilm formation. Some of these genes can also alter virulence in the bacteria, as is the case in *E. coli* and *Vibrio cholerae*, where prophages can encoded Shiga toxin genes (Stx) (Brüssow et al., 2004), Cholera toxin (Kaper et al., 1995), respectively.

Prophages can exist integrated into the bacterial for long periods, but can be triggered to undergo lytic replication when the host cell is exposed to stress stimuli or this can also occur spontaneously (Figure 1.6 B) (Howard-Varona et al., 2017). Stress stimuli can include stresses such as UV light or chemical like mitomycin C (MitC) and norfloxacin (NFX) (Cruz Martín et al., 2006; Łoś et al., 2009; Casjens and Hendrix, 2015). Upon exposure to damaging stress stimuli many bacteria attempt to repair any damage to their DNA by the action of the SOS regulatory system, best characterised in *E. coli*. In *E. coli*, this SOS mechanism is comprised of two key proteins, LexA (a repressor) and RecA (an inducer) (Žgur-Bertok, 2013). In the absence of DNA damage, a LexA dimer binds to SOS boxes, sets of 20 base pair consensus palindromic DNA sequences, repressing the transcription of SOS genes required for DNA repair (Žgur-Bertok, 2013). The SOS genes are a group of over 50 genes that in presence of DNA-damaging agents repair damage to the cells DNA (Simmons et al., 2008). Examples of some of these SOS genes include *ligA*, which encodes a DNA ligase that seals nicks in the DNA (Condra

and Pauling, 1982), *recN*, which allows recombinational repair of damaged DNA (Picksley et al., 1984; Wang and Smith, 1988), and *polA* which encodes a DNA polymerase allowing DNA synthesis (Wechsler and Gross, 1971).
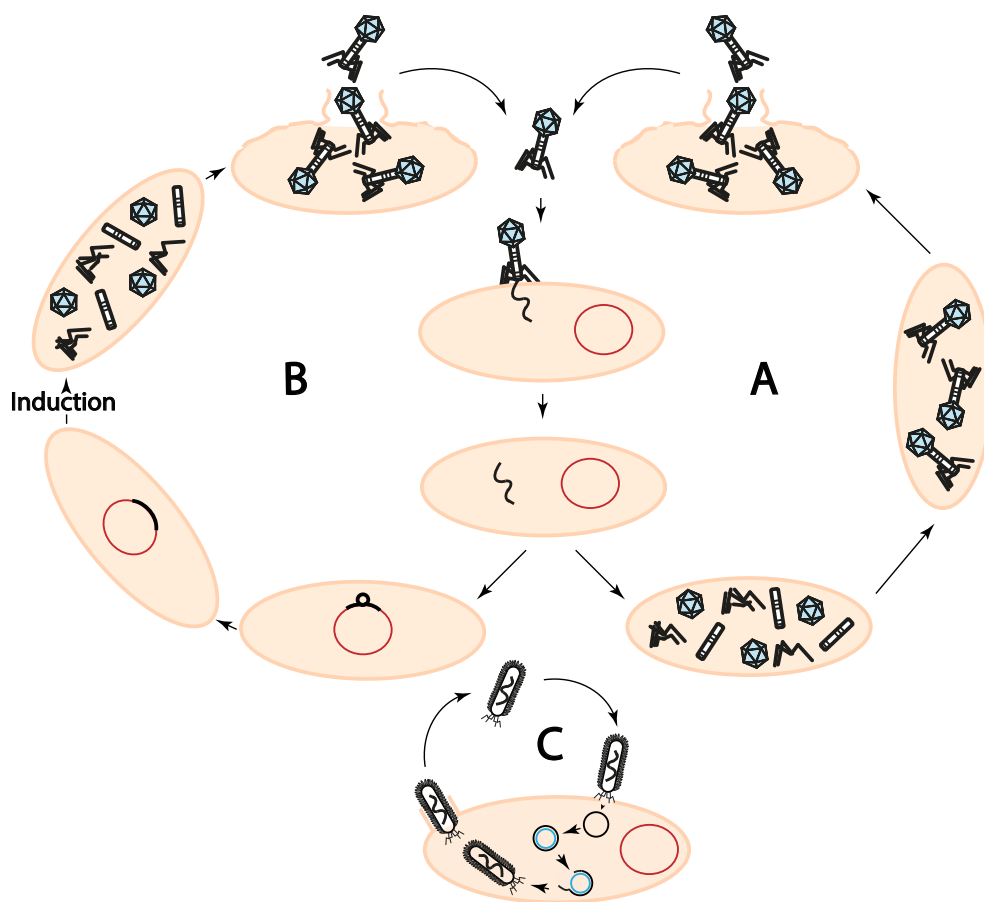
Figure 1.6. Schematic representation of the bacteriophage life cycles. The diagram represent the three most common bacteriophage life cycles: lytic (A), lysogenic (B), and chronic (C). (A) Following injection of bacteriophage DNA into the host cell, the bacteriophage then uses the host cell machinery to synthesise the required proteins to build new bacteriophage particles. The heads and sheaths are assembled separately, the new genetic material packed into the head and new bacteriophage particles assembled. Bacteriophage enzymes then attack the host cell wall, causing release of the new bacteriophage progeny into the surrounding environment through lysis of the host cell. (B) Synthesis of new bacteriophage is repressed, and the bacteriophage DNA instead integrates the host genome, remaining integrated until the bacterial cell is exposed environmental stress e.g. DNA-damaging agents. At this point the prophage excises and begins to initiation of the lytic cycle. (C) The chronic cycle begins with injection of the ssDNA bacteriophage genome into the host cell (black circle). The DNA is rapidly converted to dsDNA within the host (black and blue circles show template and nascent strand respectively). The double-stranded replicative form of DNA then undergoes transcription, producing mRNAs directing the synthesis of bacteriophage-encoded proteins. The ssDNA bacetriophage genomes are continuously replicated, one strand at a time, through rolling circle replication, as are bacteriophage proteins. An assembled bacteriophages are then continually secreted out from the host cell without the hosts death.

When DNA damage occurs the replication of DNA is blocked, resulting in the creation of ssDNA.

This ssDNA then binds to RecA, forming ssDNA-RecA nucleoproteins. The activity of ssDNA-RecA nucleoprotein catalyses self-cleavage of LexA, inactivating its repressor function, leading to high level expression of the SOS genes (Prudhomme et al., 2006; Butala et al., 2009; Baharoglu and Mazel, 2014; Ghodke et al., 2019). In bacteriophage λ, following exposure to stress stimuli the activated RecA attached with ssDNA binds to the prophage cI repressor, which mimics the structure of LexA, promotes the cleavage of DNA-binding domain in cI repressor (Lemire et al., 2011; Kreuzer, 2013). Without the dimerisation domain, the cI repressor can no longer form dimers and bind to the operators, causing the repression of the prophage. As the repressors drop off the operators, transcription of genes that are required for excision of prophage from the chromosome initiates, leading to the lytic cycle (Atsumi and Little, 2006). Excision requires site-specific recombination between *attP-attB* sequences that exists at the junctions between the prophage DNA and the chromosomal DNA. The recombination between *attP* and *attB* sequences happen in the presence of both integrase (Int) and excisase (Xis) proteins, resulting in excision of bacteriophage DNA as a circle from the chromosome and subsequent lytic development. Temperate *V. cholerae* bacteriophages are another example of bacteriophages that can be induced following DNA damage and stimulation of the bacterial SOS system.

The decision process between the lytic and lysogenic cycle in temperate bacteriophage remains poorly understood, but is known to be influenced by various factors. In the case of bacteriophage λ for instance, the decision is known to be influenced by factors including nutritional state, density of other bacteriophages in the surrounding environment, and the compatibility between the bacteriophage *attP* site and the host's *attB* site (Oppenheim et al., 2005; Zeng et al., 2010). In fact even molecular signalling between bacteriophages has recently been shown to influence the lytic-lysogenic decision in temperate bacteriophage (Erez et al., 2017).

There are some bacterial viruses that take an alternative strategy, beyond the lytic and lysogenic cycles, called the chronic replication cycle (Figure 1.6 C). In this cycle bacteriophages neither lyse nor kill their host cell, instead release involves secretion from the host cells through extrusion in the case of *Inoviridae* or via budding for *Plasmaviridae* (Weinbauer, 2004; Hobbs and Abedon, 2016). This process has been best characterised *Inoviridae*, and where replication begins again with adsorption to the host cell, and injection of the bacteriophage nucleic acid. For *Inoviridae* this is ssDNA, which after injection, is then converted into dsDNA using the bacterial hosts machinery. The newly formed dsDNA is then transcribed, and the resulting mRNA is translated leading to production of protein A. Protein A then nicks the parental strand of dsDNA, and rolling circle replication takes place to generate more copies of the DNA. Viral particles are then assembled and packaged with copies of the original injected strand, while the strand synthesised within the host now acts as a template to form more copies of the original host ssDNA. Each completed bacteriophage particle then leaves the host cell by forming through extrusion from the cell without damage the cell (Tropp, 2012).

1.2.3 Discovery of bacteriophage

Bacteriophage research began in 1915 with the observations of Frederick William Twort, an English bacteriologist, when studying vaccinia virus, the primary component of the smallpox vaccine (Twort, 1915). Twort (1915) noticed that contaminating micrococci colonies in his preparations were unable to be subcultured, and over time became 'glassy and transparent'. Crucially, he observed that touching these glassy areas and introducing it to a pure culture of healthy micrococci led to spread of the glassy area through the healthy microccoci culture, sometimes leading to the death of all microccoci. Twort suggested that his observations could either be explained as an unusual manifestation of the bacterial life cycle, an enzyme with power of growth, or potentially an "ultra-microscopic virus" (Twort, 1915; Keen, 2015). However, Twort was not certain about the possibility of an ultra-microscopic virus, due to the lack of knowledge on such a virus at the time. He was unable to conduct any further research into finding a definitive conclusion to his work due to financial difficulties and disruption by World War I, in which he participated. He instead focused on propagating vertebrate viruses, such as the cowpox virus (Twort, 1915; Ackermann, 2011).

In 1917, during an outbreak of dysentery among French soldiers during World War I, a French-Canadian microbiologist called Félix d'Hérelle independently observed the same phenomenon. D'Hérelle was examining the biological activity of stool filtrates from the dysentry patients and discovered that these filtrates were antagonistic to the dysentery bacteria, causing lysis in liquid culture and creating zones of lysis on the surface of agar seeded with the bacteria (Service, 2007). D'Hérelle concluded this activity to be caused by an "invisible microbe" infecting the dysentry bacteria, coining the term bacteriophage to describe them. D'Hérelle was the first person who recognised the nature of the agent discovered, and would go on to study their role in the microbial world and investigate their potential in treating bacterial infections (Summers, 2011; Keen, 2015).

1.2.4 Bacteriophage therapy

The therapeutic potential of bacteriophages, termed bacteriophage therapy, was a major focus of bacteriophage research after their discovery. D'Hérelle himself initiated its use, successfully treating a 12-year old boy for dysentery at the Hôpital des Enfants-Malades, Paris, in 1919 (Summers, 1999). However, with the advent of antibiotics in the 1940's, the interest and acceptance of bacteriophage therapy was limited in Western countries, due to the simplicity and wider spectrum of action of antibiotics (Sulakvelidze et al., 2001; Kutateladze and Adamia, 2010; Kutter et al., 2010; Knoll and Mylonakis, 2014). Although bacteriophage therapy was overshadowed by antibiotics in the United Kingdom, the United States and most of Western Europe, bacteriophages were widely used to treat a variety of diseases across the former Soviet Union (Fernández et al., 2018). Numerous clinical trials were

conducted there, covering a range of conditions including lung infections, gastrointestinal diseases, and skin injuries such as wounds and burns (Abedon et al., 2011). One major site of Soviet bacteriophage research, known today as the "Eliava Institute of bacteriophage, Microbiology, and Virology" (which was co-founded by Félix d'Hérelle and George Eliava, a Georgian microbiologist, in 1923) remains one of the most active centres for bacteriophage therapy research (Sulakvelidze et al., 2001).

Although the use of antibiotics led to the abandonment of bacteriophage-therapy in the West, the rise of antibiotic resistance has recently revived their interest in bacteriophage therapy (Meaden and Koskella, 2013; Laxminarayan et al., 2016). For example, in the UK the issue of antibiotic resistance reached new prominence in 2011, after being the focus of the annual report of the Chief Medical Officer for England, Prof. Dame Sally Davies, who described it as a national emergency (Davies et al., 2013). The global impact of antibiotic resistance has also led to similar attention in other countries (Carlet, 2015; O'Neill, 2016; Bassetti et al., 2017). This has produced a great number of discussions into potential ways to combat antibiotic resistance, including a 2016 Department of Health (England)/Wellcome Trust jointly commissioned review into possible strategies for tackling antibiotic resistance, which found bacteriophage therapy to be one of the most promising options (Czaplewski et al., 2016).

In bacteriophage therapy, primarily bacteriophages belonging to the *Caudovirales* are used, as these bacteriophages are more readily isolated in the laboratory and are also often easier to culture *in vitro*. Of these, only obligately lytic members are used to treat bacterial infections. This is because the use of temperate bacteriophage risks possible lysogenic replication, which does not always lead to death of bacterial cell and may even influence fitness and virulence traits of the infected organism (or even provide resistance to the bacterium through superinfection immunity) (Gill and Hyman, 2010; Karumidze et al., 2012). Treatment with obligately lytic bacteriophages has been suggested to potentially be more effective in treating infections than other therapies. For instance, bacteriophages often infect only a small number of susceptible bacteria within a population meaning there will be minimal disruption to non-target populations, their ability to 'autodose' where each viral replication increases the concentration of bacteriophages at the site of infection, and their ability to degrade biofilm through the action of depolymerising enzymes (Flores et al., 2011; Loc-Carrillo and Abedon, 2011). However, there are also disadvantages to their use. For example, their narrow host range can mean their use may be limited against an unidentified infection, as unless the infection is caused by the species or even strain targeted by the bacteriophage then no infection would take place (Loc-Carrillo and Abedon, 2011). For reasons like this, bacteriophage therapeutics are mainly used as mixtures of multiple bacteriophages in one preparation, known as a bacteriophage 'cocktail' (Gill and Hyman, 2010). Cocktails improve the effectiveness of bacteriophage therapy as by using more bacteriophages within a preparation a wider range of susceptible hosts can be targeted, improving the chances that a bacteriophage able to infect the target organism is within the perpetration (Chan et al., 2013). For

example, the 'Intesti' bacteriophage cocktail produced by the Eliava institute which is used to treat urinary tract, oral, and gut disorders contains bacteriophages active against multiple bacterial genera, including *E. coli*, *Salmonella* spp., *Shigella* spp., *Proteus* spp., *Stapylococcus aureus* and *P. aeruginosa* (Zschach et al., 2015).

Our knowledge on the clinical effectiveness of bacteriophage preparations remains mainly from trials carried out in Eastern Europe during Soviet times (Abedon et al., 2011). This is because the development of modern large scale clinical trials has been impeded by regulatory hurdles, as bacteriophages do not fall into many of the traditional definitions of medicinal products (Furfaro et al., 2018). A few modern trials have been conducted through, and have shown promise. In the phase II clinical trial of Wright et al. (2009) into the use of bacteriophages to treat antibiotic-resistant *P. aeruginosa* in 24 patients with chronic otitis media, bacteriophages were shown to be inherently safe to use and improved patient outcome. They saw that the application of one dose of bacteriophage, although not sufficient to remove the targeted bacteria completely, was able to significantly reduce bacterial densities (Wright et al., 2009). In a recent clinical study, bacteriophage therapy was given to a 15 year old patient with cystic fibrosis, who was suffering from a disseminated *Mycobacterium abscessus* infection following a lung transplantation. A three-bacteriophage cocktail was given intravenously and successfully eliminated the systemic infection without any adverse side-effects (Dedrick et al., 2019). Case studies into the use of bacteriophage therapy have also shown success, including the recent case where bacteriophage therapy was used to successfully treat a patient with a multi-drug resistant *Acinetobacter baumannii* infection, despite multiple previous antibiotic treatments producing no improvements (Schooley et al., 2017).

1.2.5   Other applications of bacteriophages

While trials are ongoing into bacteriophages as treatments for human infections, bacteriophage cocktails are also being used to remove bacteria for other purposes. One example is in food manufacturing, where they can be used to prevent food contamination with potential pathogens or prolong shelf life by reducing growth of spoilage bacteria (Pérez Pulido et al., 2016). Commercial products for these purposes already exist and are used at nearly all stages of the food preparation process, such as Agriphage™(Omnilytics) which is used in the control of *Xanthomonas campestris* or *Pseudomonas syringae* in food crops (Buttimer et al., 2017), or ListShield™(Intralytix) and ShigaShield™(Intralytix) which provide protection against *Listeria monocytogenes* and *Shigella* spp., respectively, in final food products (de Melo et al., 2018). Bacteriophages are also used in biocontrol in other industries, including as a means of controlling bacterial densities in aqueous environments as a form of water treatment (Satyanarayana et al., 2012). For example, Zhang and Hu (2013) investigated the impact of bacteriophage and chlorine treatments, separately and in combination, in preventing and removing established

P. aeruginosa biofilms. They saw that treatment with chlorine only removed around $\sim$ 40 % of *P. aeruginosa* biofilm, while the bacteriophage only treatment removed $\sim$ 89 %. However, they observed that the use of chlorine and bacteriophage in combination removed biofilm by $\sim$ 96 %. Therefore, the combination of bacteriophage and chlorine treatment can potentially be a promising approach for removing biofilm from water systems, even possibly for *Legionella* spp. biofilms.

Bacteriophages are not only useful in their ability to control bacterial densities, but can also be used for other purposes such as detecting bacteria. There are numerous ways that bacteriophages can be used in detecting bacteria (Schmelcher and Loessner, 2014). In one example, bacteriophage genomes can be modified to carry genes encoding bioluminescence or fluorescence proteins. In such a system, when the bacteriophages are free in suspension no luminescence is observed from the bacteriophage, but upon injection of the bacteriophage DNA into its host the bacteriophage encoded bioluminescent or fluorescent proteins are synthesised, facilitating visual detection (O'Sullivan et al., 2016). Jain et al. (2012) demonstrated that bacteriophage modified to express green fluorescent protein (GFP) upon infection of *Mycobacterium tuberculosis* were a useful diagnostic tool to identify *M. tuberculosis* in clinical samples, as culture identification of *M. tuberculosis* can take between $\sim$ 4 - 8 weeks while the GFP-modified bacteriophage allowed identification within 12 hours. They also proposed that this may have a further use in determining drug sensitivity of the *M. tuberculosis*, by comparing fluorescence between antibiotic untreated and treated samples, with absence of fluorescence only in the treated samples indicating drug sensitivity. This could be a highly useful development as this process will again take $\sim$ 12 hours from collection of samples, while by culture this could take weeks if not months to determine (Jain et al., 2012). Such as system would also be highly valuable in the detection of *Legionella* spp. as these organisms can often take between 3 and 5 days to isolate by culture, meaning identification and antibiotic susceptibility testing could be conducted in hours rather than days.

1.2.6 Bacteriophages and *Legionella* species

Despite their wide distribution and diversity in the biosphere, attempts to isolate bacteriophages able to infect *Legionella* spp. have not either been widely carried out, or have proven unsuccessful, with only one report claiming to have isolated bacteriophages infecting *L. pneumophila*. In that study, Lammertyn et al. (2008) described the enrichment of environmental water samples that led to the isolation of four *Myoviridae* able to infect a range of *Legionella* spp.. However, personal communications with the authors about these bacteriophages revealed that the isolated bacteriophages have since been lost (Elke Lammertyn, personal communications) and no new bacteriophages have since been obtained, raising questions about the validity of their initial observations.

The only other work discussing *Legionella* spp. bacteriophages beyond this is a bioinformatics study by Gomez-Valero et al. (2014), which showed the presence of a putative prophage within one strain

of *L. micdadei.* However, the comparative analysis conducted on only 11 *Legionella* spp. genomes, covering five species (*L. hackeliae* (n = 1), *L. micdadei* (n = 1), *L. fallonii* (n = 1), *L. pneumophila* (n = 7), and *L. longbeachae* (n = 1)), with no physical bacteriophage particles identified. Thus to date, no bacteriophages with the ability to infect any of the *Legionella* spp. have been observed.

1.3   Project Aims

This project is based on the hypothesis that bacteriophages could have a great number of uses in the identification, prevention, and treatment of infection and colonisation by *Legionella* spp. However, as there is currently limited information available about the existence of bacteriophages infecting *Legionella* spp., the first step in this process is the identification of viruses infecting *Legionella* spp. Therefore, the aim of this project was to isolate and characterise bacteriophages infecting *Legionella* spp.

To achieve this, the following specic objectives were outlined:

1. Establish a collection of environmental *Legionella* spp. isolates to use as hosts for isolating free bacteriophages and as a possible source of temperate bacteriophages by:

    (a) Optimise liquid culture for supporting the growth of *Legionella* spp. and our future bacteriophage work

    (b) Collecting samples from both natural and man-made aquatic and soil environments

    (c) Detection of *Legionella* spp. within each sample using qPCR-based screening

    (d) Screen samples for culturable *Legionella* spp. isolates

    (e) Characterise *Legionella* spp. isolates using both serogroup and sequenced-based typing

    (f) Determine susceptibilities of these isolates to disinfectant and antibiotic treatments that are commonly used to eliminate them

2. Isolate free bacteriophages using the *Legionella* spp. collection as hosts by:

    (a) Optimise methods involved in the bacteriophage isolation process for *Legionella* spp.

    (b) Use filtrate from natural and man-made aquatic and solid sources used for *Legionella* spp. to attempt to isolate bacteriophages by enrichment and direct plating

    (c) Use a well-based assay to avoid the need to culture *Legionella* spp. on a solid medium and potentially isolate bacteriophages unable to form plaques

3. Identify and isolate temperate bacteriophages existing as prophages within the *Legionella* spp. collection by:

(a) Inducing cultures of *Legionella* spp., using DNA-damaging agents

(b) Examine the presence of prophages within publicly available (GenBank) *Legionella* spp. genomes sequences bioinformatically

2. Methods

2.1 General methodologies

2.1.1 Bacterial strains

All *Legionella* spp. strains used in this study are detailed in supplementary table S4. *L. pneumophila* NCTC 11192 and *L. micdadei* NCTC 11371 were purchased from the National Collection of Type Cultures (NCTC) of PHE. Thirty-seven *Legionella* spp. strains were obtained from Dr. Samuel Collins (PHE), sourced originally from the a Culture Collection of PHE. The other eighteen strains of *L. pneumophila* were obtained during the current work, and were isolated from environmental water samples from around the UK.

2.1.2 Media and growth conditions

The compositions of all media can be found in supplementary table S1. Unless otherwise stated, all *Legionella* spp. isolates were cultured on BCYE-α 1 % agar plates, supplemented with 4 % cysteine and 2.5 % iron (III) pyrophosphate. For liquid culture, *Legionella* spp. strains were routinely cultured in BYE-α broth, supplemented with 4 % cysteine and 2.5 % III, shaken at 100 revolutions per minute (rpm). Both cultures on agar plates and in broth were incubated at 37°C.

2.1.3 General DNA analysis methods

2.1.3.1 Polymerase Chain Reaction (PCR)

For PCR, unless otherwise stated, a single bacterial colony was resuspended in 40μl of ultra-pure H_2O and heated at 100°C for 10 min to release the cells DNA. Cells debris was then collected by centrifugation at 11000 xg for 1 min and 1μl of the supernatant then used as template in the PCR reactions.

All PCR reactions were conducted using, Biotaq DNA polymerase (Bioline). All reaction conditions and compositions were depend on the purpose of the assay being carried out. Further information on the PCR reactions can be found in the appropriate section.

2.1.3.2 Agarose gel electrophoresis

To visualise the DNA fragments produced by PCR, gel electrophoresis was done on 1 % (w/v) agarose gels in TAE (Tris-acetate-EDTA) buffer (composition can be found in supplementary table S1) containing 0.05 μl ml^{-1} Gel Red (biotium). 5μl of each sample unless otherwise stated was then loaded onto the gel using DNA loading dye (New England Biolabs). Electrophoresis was conducted at 100 V for 60 minutes and the gels were then imaged using a ChemiDoc Imaging System (Biorad). The resulting DNA product bands were visualised and compared to a GeneRuler 1 kb (Thermo-Scientific).

2.1.3.3 Purification of DNA fragments

After agarose gel electrophoresis (section 2.1.3.2), if only a single band was expected and present, the E.Z.N.A Cycle Pure Kit (Omega biotek) was used to purify the DNA directly from the remainder of the PCR reaction, according to the manufacturers' instructions.

2.1.3.4 DNA sequencing

DNA sequencing of PCR products was performed by GATC Biotech (Germany). For sequencing, following purification of the PCR products (Section 2.1.3.3), 30 μl of $\sim$ 30 ng μl^{-1} of the DNA to be sequenced was sent together with 30 μl of both forward and reverse primers (10 μM). The sequencing reaction was conducted by GATC Biotech using an ABI 3730x1 sequencer.

2.2 Comparison of *L. pneumophila* growth in liquid media

To define the best available liquid medium for growing *Legionella* spp., the growth of *L. pneumophila* NCTC 11192 was compared in five variants of BYE-α liquid media; filter-sterilised BYE-α, autoclaved BYE-α with no charcoal added at any stages, BYE-α with added charcoal, BYE-α with charcoal removed by filtration following autoclaving, and BYE-α with charcoal removed by centrifugation following autoclaving. Each of these media were prepared according to the recipes in supplementary table S1.

The inoculum for these assays was prepared by inoculating a portion of colony growth on a BCYE-α plate into 5 ml of sterile SM buffer to a turbidity similar to McFarland 0.5 (OD$_{600}$: $\sim$0.1). 300 μl of

the OD_{600} adjusted inoculum was then added into 30 ml of each of the five liquid media being tested. From these, 2 ml aliquots were then removed and transferred into 13 x 7 ml bijous, and incubated at 37°C, 100 rpm for 96 hours. Both the optical density (at 600 nm) and colony numbers (CFU ml^{-1}) was measured at each time-points; 0, 4, 6, 8, 10, 12, 14, 24, 48, 72, and 96 hours (11 in total), by removing and measuring the contents of one of these tubes. Optical density was measured using a spectrophotometer, while colony numbers were measured by spotting 10μl of a set of 10-fold serial dilutions of each time point onto BCYE-α plates, followed by incubation overnight at 37°C.

2.3 Isolation of *Legionella* spp. from environmental samples

2.3.1 Collection and processing of environmental water and soil samples

For isolation of *Legionella* spp. from environmental samples, a total of 108 water samples originating from different locations in the UK (Bath(4), Hampshire(14), Essex (7), Margate (2), Leicester (34)), as well as other countries; i.e., Kenya (3), Nigeria (31), Iran (2), and Iraq (11) were obtained (Figure 2.1). These samples were collected from a number of different sources in these locations (e.g. wells, rivers, spas, thermal springs, cooling towers, seas, and canals) with a full list of the location and sample source given in Supplementary table S2. The samples from Nigeria and Iraq were graciously provided by Dr. Janet Nale and Dr. Srwa Rashid respectively, while the samples from Hayling Island, and Leicester General Hospital were kindly provided by Dr. Julian Clokie and Dr. David Jenkins, respectively. Spa and thermal spring water samples were obtained from Bath, UK, from Dr. Simon Kilvington.

A second set of 88 water samples were collected from five cities around the Midlands, UK (Leicester, Nottingham, Derby, Birmingham, and Coventry) (Supplementary table S3). During sampling, 1 L water samples were collected from each site with additional 50 ml samples collected from nearby (within 500 m) the 1 L source. A total of 66 soil samples were also collected from these locations by collecting soil from the surface to the depth of 10 cm into a sterile 50 ml falcon tube. The co-ordinates of all of these samples can be found in Supplementary table S3. All samples were stored at 4°C until required.

2.3.2 Processing and culturing of the collected samples

To isolate bacteria from each water sample, bacteria within the water samples were first concentrated. Here, water samples were first passed through a 0.22 μm filter. The filter paper was then collected and placed into a sterile 50 ml centrifuge tube containing 10 ml of 1:40 Ringer's solution. This tube was then vortexed for 1 minutes to free bacteria from the filter. This homogenised sample would then
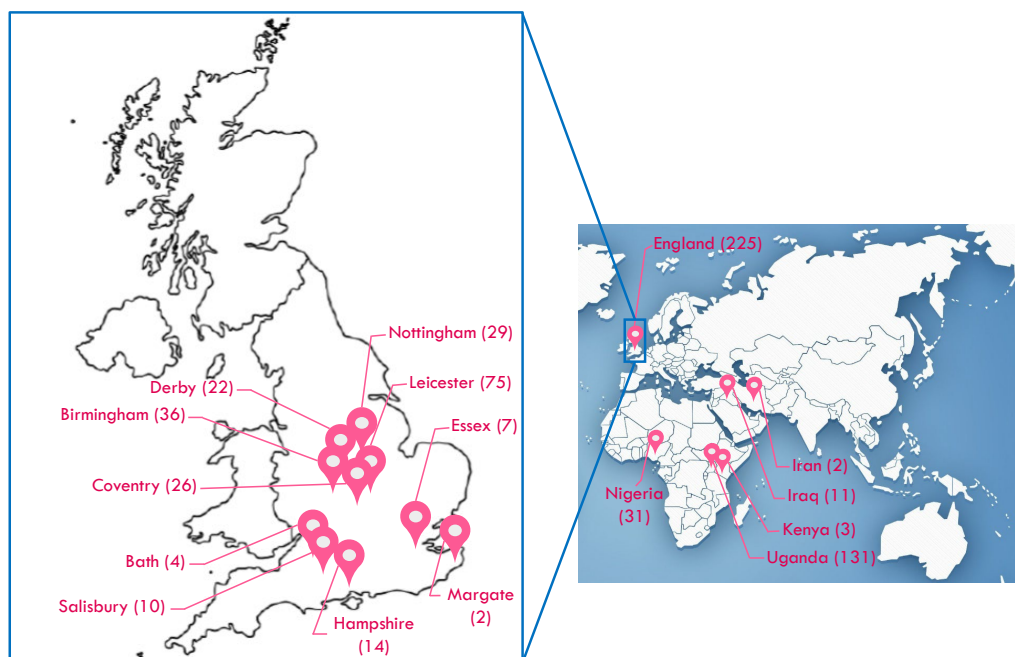
Figure 2.1. Maps of locations where samples were collected for use in this study. Water and soil samples were collected from each of the countries marked on the world map (pink points), with country name indicated and the the number of samples collected from that in parentheses.

be used in the downstream culturing of *Legionella* spp.. The filtrate from the 0.22 μm was stored at 4°C to be used in the screening for bacteriophage against *Legionella* spp.. (Figure 2.2).

To isolate bacteria from the soil samples, 2 g of every soil sample were first homogenised in 15 ml 1:40 Ringer's solution plus 5 - 6 3 mm glass beads. Samples were then vortexed for 1 minutes to release and separate the cells from larger soil particles. As the supernatants contained noncellular material e.g. stones and clay particles, the homogenised soil were first centrifuged for 5 minutes at 500 xg, whereby only the the large particles would sediment. The resulting supernatant was then processed similar to the water samples above by filter-concentration using a 0.22 μm filter (Figure 2.2).

To try and isolate *Legionella* spp. from the collected samples each filter-concentrated sample first underwent heat-treatment followed by culturing on selective media plates. Heat-treatment was used to reduce the growth of non-*Legionella* organisms and enhance the recovery of *Legionella* spp.. Heat-treatment was conducted by submerging a 5 ml aliquot of each sample at 55°C in a water bath for 15 minutes. The samples were then allowed to cool at room temperature for 10 min, and then 100 μl of a series of 10-fold dilutions were then spread over MWY agar plates. Plates where then incubated at 37°C for 72 hours.

After this, any colonies that appeared like *Legionella* spp. were then re-streaked from a single colony

onto both BCYE-α agar plates with cysteine and BCYE-α agar plates without cysteine. Colonies that grew on BCYE-α agar plates with cysteine but not BCYE-α agar plates without cysteine were designated as *Legionella* spp. colonies.

Polymerase chain reaction (PCR) using primers (mipLesnsens and mipLensrev) amplifying the *mip* gene (Supplementary table S6) was then conducted on each *Legionella* spp. isolate (Figure 2.2). The amplification reaction was carried out in a total volume 23 μl standard reaction mixture (1 μl of each of the 10 μM forward and reverse primers, 1.5 μl 50 μM MgCl$_2$, 2.5 μl 10x reaction buffer, 0.25 μl Biotaq DNA polymerase, 1 μl 10 μM dNTPs, and made to 23 μl with upH$_2$O) and 2 μl DNA template. Amplification program included an initial denaturation at 95$\circ$C for 5 minutes followed by 30 cycles of denaturation at 95$\circ$C for 30 seconds, primer annealing at 50$\circ$C for 30 seconds and extension at 72$\circ$C for 45 seconds, with final extension step at 72$\circ$C for 5 minutes

The PCR products were purified and sent to GATC Biotech (Germany) for sequencing using Sanger sequencing.
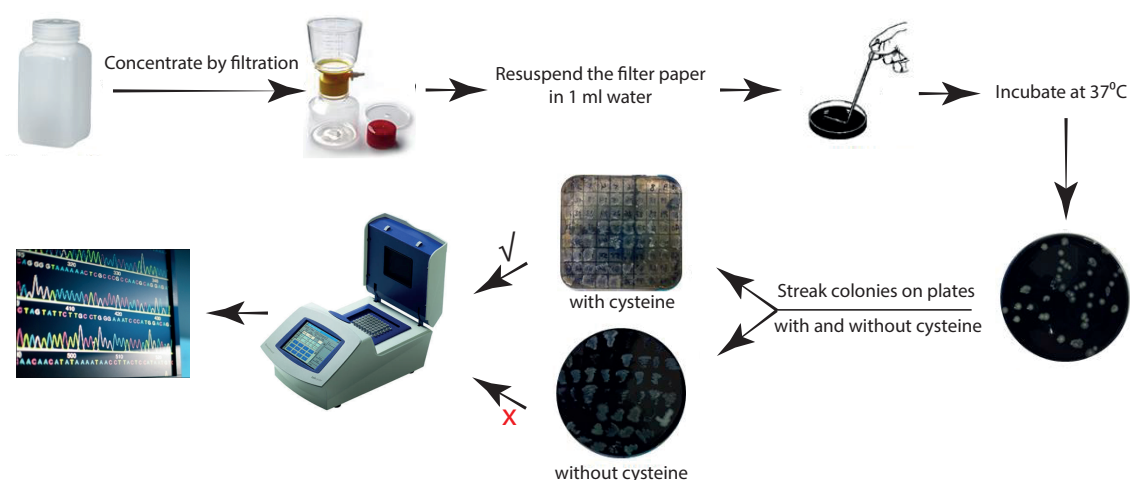
Figure 2.2. Flow diagram of the process used in culturing the collected samples. The figure shows each step in the sample processing method, from collecting the sample, filter-concentrating the samples, culturing the organisms, and identifying any potential *Legionella* spp. using cysteine selection and PCR.

2.4 Serogroup and sequence-based typing of *L. pneumophila* isolates

2.4.1 Serogroup typing

Serogroup typing of all *L. pneumophila* isolates was conducted using the *Legionella* Latex Test Kit (DR0800M; Oxoid, Basingstoke, UK) following the manufacturer's instructions.

2.4.2 Sequence-based typing

For sequence-based typing a total of 45 clinical and environmental isolates of *L. pneumophila* were analysed, including 20 *L. pneumophila* serogroup 1, 24 *L. pneumophila* serogroups 2 - 14, and 1 *L. pneumophila* strain that presented a cross-reaction with the reagents with the *Legionella* Latex kit.

For sequence-based typing, PCR was performed using primers targeting seven specific genes; *asd* encodes the *L. pneumophila* aspartate-beta-semialdehyde dehydrogenase, *mip* encodes macrophage infectivity potentiator protein, *pilE* encodes type IV pilin, *flaA* encodes flagellum subunit protein, *proA* encodes a zinc metalloprotease, *momps* encodes an outer membrane protein, and either *neuA* or *neuAH* which encode an enzyme involved in the biosynthesis of the lipopolysaccharide (LPS) (Gaia et al., 2003, 2005; Ratzow et al., 2007; Farhat et al., 2011). *neuAH*, was only used in PCRs if PCR with *neuA* failed.

PCR using each of these primers was conducted as follows:

First bacterial DNA was extracted by resuspending one colony in 100 μl ultra-pure water and heated at 100°C for 10 min. Then 5 μl of the supernatant was used as a template along with 2 μl of 2.5 mM MgCl$_2$, 4.4 μl of 10 mM Reaction buffer, 1.6 μl of 10mM dNTPs, 0.8 μl of the appropriate 10 mM forward and reverse primer (Supplementary table S5; Integrated DNA Technologies, UK), 4 μl template DNA, and 0.4 μl of BioTaq DNA polymerase (New England Biolabs, UK) in a PCR reaction.

Amplification was performed using the following conditions: initial denaturation at 95°C for 5 minutes followed by 35 cycles of denaturation at 95°C for 30 seconds, primer annealing at 55°C for 30 seconds and extension at 72°C for 45 seconds, with final extension step at 72°C for 10 minutes.

PCR products were then analysed by gel electrophoresis with the PCR products then purified and sent to GATC Biotech (Germany) for sequencing.

2.4.3 Sequence quality tool and SBT Database

To identify the allelic profile and a sequence type (ST) of the isolates, the sequences from both forward and reverse sequencing reactions for each primer set (.abi) were submitted to the "Sequence Quality Tool" through `http://www.hpa-bioinformatics.org.uk/cgi-bin/legionella/sbt/seq_assemble_legionella1.cgi`. This tool assembles the contigs to produce a similar sequence. The consensus sequences were then trimmed to the correct length by alignment with the available reference sequences in the website. The trimmed sequences were then matched against those in the SBT database. For each isolate, allele profile (7 digits) were obtained for each allele gene by the combination of alleles at each of the loci using the following order: *flaA*, *pilE*, *asd*, *mip*, *momps*, *proA*, and *neuA/neuAH*. Finally, genotypes (sequence type or ST) of each isolate were determined based on the allelic profiles of each gene (`http://www.hpa-bioinformatics.org.uk/legionella/legionella_sbt/php/sbt_homepage.php`). This is an online tools to perform these analysis are available on the European study Group on *Legionella* Infections (ESGLI) of the European Society of Clinical Microbiology and Infectious Diseases (ESCMID) website.

2.5 Susceptibility testing of *Legionella* spp.

2.5.1 Antibiotic susceptibility of *Legionella* spp.

The Antibiotic Susceptibility of *Legionella* spp. strains were tested against four antimicrobial agents using E-test strips on an agar plate. The tested antimicrobials were ciprofloxacin, rifampicin, clarithromycin, and erythromycin. The range / MIC scale printed of above antibiotics on the strips that were used in this study were 0.002 - 32 μg ml^{-1} for rifampicin, 0.16 - 256 μg ml^{-1} for erythromycin, 0.002 - 32 μg ml^{-1} for ciprofloxacin, and 0.16 - 256 μg ml^{-1} for clarithromycin. To perform this test, an overnight culture was first prepared by inoculating a single colony of each isolate that was grown on BCYE-α agar plate and incubated at 37°C for 48 hours, into BYE-α broth. 100 μl of an overnight culture was then subcultured into 10 ml of fresh broth and incubated at 37°C until it reached to an optical density of 0.09 - 0.1 at 600 nm. After that, 150 μl of each inoculum was spread on a BCYE-α agar plate and an E-test strip was applied to the surface according to the manufacturer's instructions (Biomerieux). The plates were then incubated at 37°C for 48 h before reading the the MIC values. The MIC value was determined the lowest concentration of antibiotic that completely inhibited the lawn of bacterial growth and were compared between *L. pneumophila* and other species of *Legionella*. All the tests were as three biological repeats.

2.5.2 Chlorine susceptibility of *Legionella* spp.

To check the sensitivity of all the *Legionella* spp. in our collection to the common disinfectant that are used to remove *Legionella* contamination from water systems, the lawn of 44 strains of *Legionella* spp. were exposed to sodium hypochlorine (2 %). To do this, a single colony of the targeted isolate of *Legionella* spp. was inoculated into 5 ml BYE-α broth and incubated overnight at 37°C, 100 rpm. On the following day, 100 μl of the overnight culture were subcultured into 10 ml of fresh BYE-α broth and incubated at 37°C, 100 rpm for 6 - 7 hours until the optical density at 600 nm reached to 0.1, corresponding to 10^8 CFU ml^{-1}. 150 μl of the bacterial suspension was then spread on the entire surface of the 90 mm BCYE-α agar plate using a spreader. Blank diffusion disks (Oxoid, Hampshire, UK) were then applied using a sterile pair of forceps within 3 - 4 minutes of inoculation when the plates were dry. After that, 10 μl of the above pre-made disinfectant stock (2 %) were spotted on top of the disk and the plates were incubated at 37°C for 48 hours. Inhibition zone diameters (mm) were then measured by a ruler at the point of complete inhibition. Results obtained were used to classify isolates as being resistant or susceptible to sodium hypochlorite that have been used. The bigger the diameter of the inhibition zone, the more susceptible is the strain to the disinfectant.

2.5.3 Environmental samples DNA extraction

To allow qPCR to be performed on soil and water samples collected from the Midlands, UK, DNA was extracted from the sample concentrates following filtration process as explained in sections 2.3.1. Instagene matrix (BioRad, Watford,UK) was used to extract DNA from the both water and homogenised soil samples. To do this, 1 ml of the filter-concentrated samples were first spun down at 10,000 xg for 15 minutes, followed by adding 180 μl of InstaGene matrix to the pellet and incubated at 56°C for 30 minutes in a dry shaker-heating block at 1400 rpm. The samples were then vortexed at high speed for 10 seconds, were placed in a 100°C heat block for 10 minutes, and were then re-vortexed for extra 10 seconds. Following the centrifugation of the suspension at 10,000 ×g for 3 minutes, the supernatant was collected and stored at -20°C for using in qPCR assay.

2.5.4 Quantitative Polymerase Chain Reaction (qPCR) assay

To detect and quantify the presence of *Legionella* spp. in environmental samples, qPCR was conducted using simplex assay for *L. pneumophila* serogroup 1, as well as duplex assay for identification of non-serogroup 1 *L. pneumophila* and other species of *Legionella* based on Collins et al. (2015) with few modifications. The sensitivity and specificity of the primers were tested using extracted bacterial genomic DNA of *L. pneumophila* NCTC 11192, *L. pneumophila* serogroup 2 - 14, and *L. longbeachae*

as positive controls, plus *P. aeruginosa* and *Haemophilus influenzae* as negative controls. All qPCR assays were performed using the Applied Biosystems 7500 Fast Real-Time PCR System (Thermo-Scientific). Two separate assays were conducted, a simplex assay for *L. pneumophila* serogroup 1 and a duplex assay for identifying non-serogroup 1 *L. pneumophila* and other *Legionella* app..
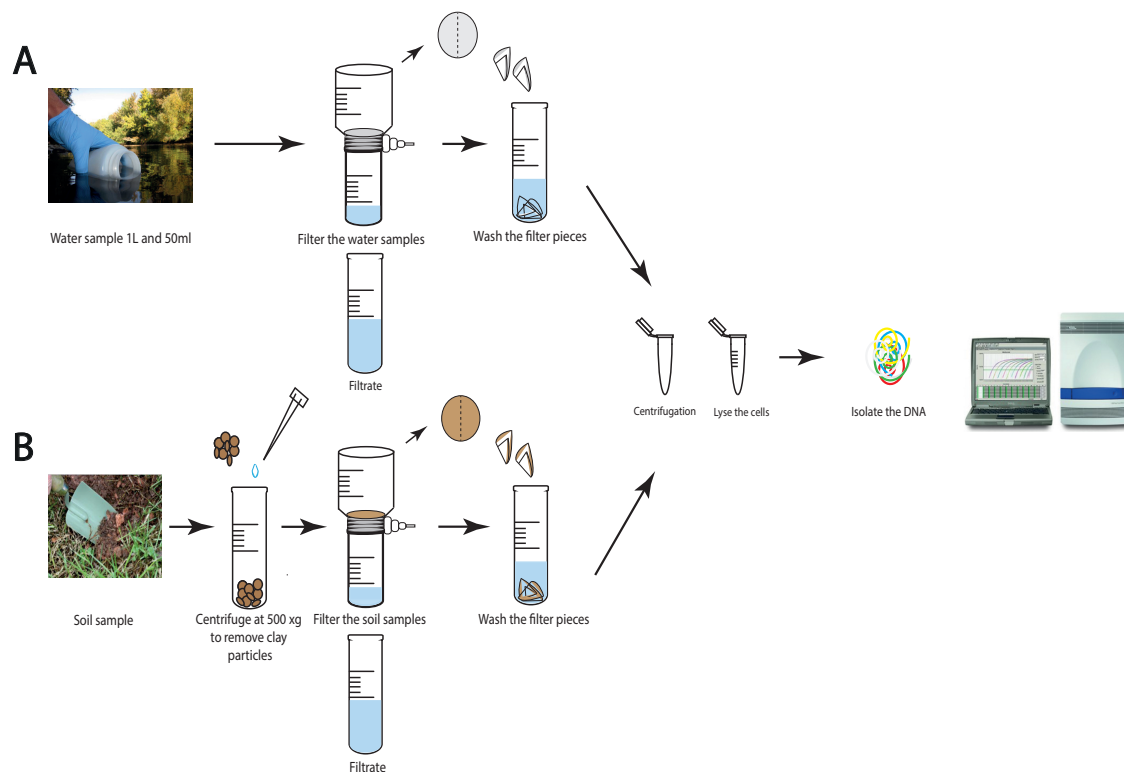
Figure 2.3. Schematic diagram of the work flow from sample collection to qPCR analysis.
The process starts with collecting samples; water (A) and soil (B) from different locations, followed by filter-concentration of them to concentrate the bacterial content and generating bacteria-free filtrate. This was completed using vacuum-filtration units. In case of soil samples, they went through an extra centrifugation step compared to the water samples to remove clay particles from them. The collected bacterial cells on filters were then resuspended in Ringer solution before their DNA was extracted and were used for qPCR or being streaked on agar plate for culturing purpose. The retained potentially bacteria-free filtrates were stored at 4°C for screening with the purpose of finding bacteriophage infecting *Legionella* spp..

qPCR reaction mixtures were performed in a total volume of 30 μl consisting of 15 μl TaqMan™ Environmental Master Mix 2.0 (Thermo-Scientific), 10 μl template DNA, 1.2 μl of 10 μM of both forward and reverse primers for *wzm* and *mip* and 1.8 μl of 10 μM of both forward and reverse primers for *ssrA*, plus 0.45 μl probe. The thermal cycling conditions were: 95°C for 10 minutes followed by 45 cycles of 95°C for 15 seconds and 60°C for 1 min. Calibration standards derived from three strains; *L. pneumophila* NCTC 11192, *L. pneumophila* serogroup 2 - 14, and *L. longbeachae*, ranging from 10^0 to 10^8 per reaction were used to generate standard curves for each qPCR run. For this purpose, *L. pneumophila* NCTC 11192 was used for simplex assay and both *L. pneumophila* serogroup 2 - 14 and *L. longbeachae* were used for duplex assay. A negative control (UltraPure™ DNase/RNase-Free Distilled

Water, ThermoFisher Scientific, UK) were included in all assays. Flouresence data was authomatically collected after each amplification cycle. This approach was conducted based on amplification of three genes; *ssrA*, *mip* and *wzm*, which exist as single copies in the *Legionella* spp. genomes. The primers and probes used in this technique are listed in supplementary table S6.

In order to estimate the number of *Legionella* spp. in each qPCR assay, aliquots of the above control strains corresponding to 10^0 and 10^8 copy number were used. Three replicates of each concentration of these strains were included on each run of qPCR to minimise the intra-assay variations. The tested samples were also 10-fold serially diluted for these times to decrease the inhibitory effect of the substances in the samples.

All data were analysed using AB 7500 Software and python programming scripts. The python scripts used AB 7500 Software raw data output as input data to visualise results using boxplots and tests for significance using Analysis of Variance (ANOVA) test. The python script was conducted to facilitate easy manipulation for each dataset. Results from all assays were compared to determine any significant differences in number of identified *Legionella* spp. between different locations that samples were collected from.

2.6 Isolation of free bacteriophages infecting *Legionella* spp. from environmental samples

2.6.1 Enrichment

To isolate bacteriophages infecting *Legionella* spp., environmental samples were first collected and the presence of *Legionella* bacteriophages were investigated using enrichment assay. The samples that were used in enrichment were the first set of samples collected (Supplementary table S2). The enrichment process was conducted by inoculating 40 ml of water samples individually with 1 ml of each of *Legionella* spp. strains in our laboratory collection and incubation at 37°C for 72 hours. The presence of bacteriophage capable of infecting above strains was then investigated using spot test and plaque assay as sections 2.9.1 and 2.9.2.

2.6.2 Sample preparation for well-assay

To allow the visual detection of bacteriophage by monitoring reductions in bacterial density in the presence of the sample and speed up the process of sample screening for the presence of *Legionella* bacteriophages, well assay was conducted. For this, all 403 samples (Supplementary tables S8) from environmental and man-made sources from various regions around the world (Figure 2.1) were pilled

up together based on the diverse categories that they were belong to (Supplementary table S7) and concentrated using either tangential flow filtration or Amicon Ultra-15 Centrifugal Filter Units into 14 different groups. Sample 1 contained all the Midlands' qPCR-positive samples for *wzm* genes; sample 2: all the environmental samples that were collected in first round of sampling; sample 3: the Midlands' qPCR-positive samples for *mip* genes; sample 4: the Midlands' qPCR-positive samples for *ssrA* genes; sample 5: tap water sample from University of Leicester, which were culture-positive for *L. pneumophila* serogroup 1, sample 6; tap water sample from Leicester General Hospital, Leicester, culture-positive for *L. pneumophila* serogroup 1; sample 7: environmental water samples from Nigeria; sample 8: water samples from cooling tower; sample 9: the only Midlands water sample that was negative for *Legionella* spp.; sample 10: environmental water samples from Bath, UK, culture-positive for *L. pneumophila* serogroup 2 - 14; sample 11: environmental water sample from Public Health England; sample 12: environmental water sample from Uganda, East Africa, treated with chloroform; sample 13: environmental water sample, from Uganda, East Africa, with no treatment; and sample 14: sewage sample from London, UK. Amicon Filter Unit (Merck Millipore) was used only for samples from Uganda (sample numbers 12 and 13). To do this, following taking a 300 μl aliquot of all the samples, the ones that were treated with chloroform were mixed together and the ones without any treatment were added together. The 50 kDa amicon centrifugal filter units were then used to concentrate each group of above samples individually. For this purpose, the samples were centrifuged for 5 minutes initially at 4000 xg and centrifugation was then carried on until the final volume dropped to 1 ml. The samples were then stored at 4°C before being used for screening for *Legionella* spp. bacteriophage. For the rest of the samples, following adding them together for each group, tangential flow filtration a 50 kDa filter unit was used to concentrate them into a final volume of 20 ml.

2.6.3   Well-assay

To avoid the need of plaque assay to identify and purify the potential *Legionella* bacteriophages on a solid medium, as well as speeding up the process of screening samples, well assay was conducted. To do this, 160 μl of BYE-α broth, supplemented with and without 5 mM $MgCl_2$ and 5 mM $CaCl_2$, was added separately to each well of a 96-well microtitre plate, followed by adding 40 μl of the target *Legionella* spp. cells from an exponentially growing culture (OD 0.1 at 600nm). 10 μl of the concentrated environmental samples to be screened as preparaed in section 2.6.2 were then added to an individual well. Control wells with sterile ultra-pure water and no environmental sample added served as negative controls. The plates were then incubated in Multiskan™ GO Microplate Spectrophotometer (ThermoFisher Scientific, UK) at 37°C for 72 hours. The wells were monitored for the lysis of *Legionella* cells by recording a drop in optical density at 600 nm compared to the negative control. In case of observing any reduction in growth, the lysate from a well which showed a reduction in growth was

removed and centrifuged at 13,000 xg for 5 minutes in a benchtop centrifuge to pellet any cell debris. The supernatant was then serially diluted and used for another round of well assays to ensure a clonal isolates. The dilution before the one that does not lyse would potentially contain a single bacteriophage.

2.7 Optimisation of approaches to isolate bacteriophages against *Legionella* spp.

2.7.1 Influence of inoculum conditions of *L. pneumophila*

To test whether different stages of the growth of *L. pneumophila* could affect the plaque formation, plaque assay was performed on two diverse inoculum conditions; exponential and post-exponential phases. For this purpose, the relationship between nutrient concentrations and the initiation of potential bacteriophage production was tested by collecting bacteria from cultures at different growth phases, and resuspended with supernatant of the opposite growth stage. For instance, to test the effect of post-exponential phase *L. pneumophila* (host strain of respective bacteriophage) on infectivity of bacteriophage, stationary phase *L. pneumophila* E1 cells ($OD_{600} = 1$) were collected by centrifugation at 3000 $\times$g for 5 minutes and then inoculated into fresh BYE-α broth. While, in second condition, an exponential-phase growth of the same host stains ($OD_{600} = 0.5$) were centrifuged at 3000 $\times$g for 5 min, and resuspended with supernatants obtained by centrifugation of post-exponential-phase culture. Both spot test and plaque assay were performed as sections 2.9.1 and 2.9.2 using the cultures prepared in the manner just described and potential bacteriophage lysate that was collected from the plate that produced single plaques once.

2.7.2 Effect of replacement of charcoal with catalase in the growth medium on plaque formation

To test whether the presence of charcoal could have a negative effect on infectivity of bacteriophage or masking the plaque formation, charcoal was replaced with catalase in bottom and overlay agar for both spot test and plaque assay. To do this, both the top and bottom agars were prepared as explained in supplementary table S1, except replacing the charcoal with 250 U ml^{-1} catalase (Sigma) following autoclaving the agars. The prepared agars then were used for both spot test and plaque assay as sections 2.9.1 and 2.9.2.

2.7.3 *Legionella* transparent media

To remove the challenge of having charcoal in the medium, *Legionella* spp. Transparent Medium (LTM), was used, instead of a charcoal-based medium. LTM was first prepared as explained in supplementary table S1. The plaque assay was then performed as explained in section 2.9.2 and LTM was used instead of BCYE-α medium.

2.8 Isolation of *Legionella* spp. bacteriophage via induction of lysogens

2.8.1 Sensitivity of all *Legionella* spp. strains to induction agents using a disk diffusion assay

In order to increase the possibility of inducing prophages within the genomes of *Legionella* spp., it was necessary to choose strains that were more sensitive to DNA-damaging agents, indicating their potential to be more inducible. Thus, disk diffusion assay was performed to test the susceptibility of 44 strains of *Legionella* spp. isolates to prophage inducing agents; MitC and NFX, under identical conditions. All strains were first cultured on BCYE-α agar plates for 48 hours at 37°C. A single colony was then suspended in 5 ml BYE-α broth and incubated overnight at 37°C. On following day, 100 μl of an overnight culture was resuspended in 10 ml of fresh BYE-α broth and incubated between 6 - 10 hours to get to OD$_{600}$ = 0.1 (the incubation time was varied from one strain to another). After that, 150 μl of bacterial culture were drawn out of the suspension and spread to the entire surface of the 90 mm plate using a spreader. Blank diffusion disks (Oxoid, Hampshire, UK) were used for each of DNA-damaging agents; MitC (10 mg ml^{-1}) (Sigma) and NFX (10 mg ml^{-1}) (Sigma). Disks were applied using a sterile pair of forceps within 3 - 4 minutes of inoculation when the plates were dry. After that, 10 μl of the above pre-made antibiotic stocks (10 mg ml^{-1}) were spotted on top of the disk and the plates were incubated at 37°C for 48 hours. Inhibition zone diameters (mm) were then measured by a ruler at the point of complete inhibition. Results obtained were used to classify isolates as being resistant or susceptible to the both inducing agents that have been used. The bigger the diameter of the inhibition zone, the more susceptible is the strain to the DNA-damaging agent. Therefore, the outliers (the strains suspected of exhibiting higher sensitivity to either MitC or NFX) could potentially be used as lysogenic strains for prophage induction from their genomes.

2.8.2 Prophage induction using broth micro-dilution method

Following performing the susceptibility test using inducing agents to choose the potential lysogenic strains, the strains that showed more sensitivity to inducing agents, were chosen to check whether exposure to the inducing agents would excising the prophage from these strains or not. Therefore, prophage induction was conducted using broth micro-dilution method. To do this, bacterial growth curve was performed in a 96-well plate. A two-fold serial-dilution of the appropriate inducing agent was conducted in a 96-well plate and they were then inoculated with a known volume of targeted bacterial isolate. For this, an aqueous stock solution of 400 μg ml^{-1} of MitC was prepared and sterilised through a membrane filter of 0.22 μm of pore size (Sarstedt). Serial dilution was achieved through the addition of 40 μl of working stock (400 μg ml^{-1}) of either MitC or NFX to the first well, corresponding to each outlier, plus 160 μl of BYE-α broth (final concentration of 80 μg ml^{-1}). Two-fold serial dilutions were then made of this in subsequent wells, which was initially filled with 100 μl of fresh BYE-α broth until the concentration of the inducing agent gets to 0.00025 μg ml^{-1}. Then, few colonies of the target *Legionella* spp. were transferred to a falcon tube with 10 ml of BYE-α broth to produce optical density of 1. *Legionella* spp. without adding any inducing agents and BYE-α broth were included in each batch of broth microdilution tests as a control. Following preparing the dilutions, 100 μl of the bacterial suspension was added to each well, and the microtitre plates were sealed with a thin layer of parafilm and incubated for 52 h at 37°C in a Multiskan GO Microplate Spectrophotometer (Thermo-Scientific) for measurement of bacterial density (OD$_{600}$) every 15 minutes over the 52 hour time course. Results were obtained from the Thermo-Scientific SkanIt Software, and exported to Microsoft Excel for further analysis. The reduction in optical density of the cell density were taken as an indicator of bacterial cell lysis and prophage release.

2.8.3 Larger volume prophage induction with MitC or NFX

After estimating the required concentration of prophage inducing agents for observing the reduction in optical density using broth micro-dilution test, to achieve a high concentration of potential viral particles for further investigation and checking by TEM, the induction process was conducted in a larger volume (200 ml). A larger volume was chosen to increase the yield of released potential viral particles, after being exposed to MitC or NFX. For this purpose, 4 ml of an overnight culture of the tested strain of *Legionella* was transferred into a glass flask containing 600 ml of fresh BYE-α broth and grown at 37°C with shaking (100 rpm) for 13 - 16 hours until the optical density at 600 nm was 0.5. Thereafter, the culture was split into 3 x 200 ml cultures and induction with MitC or NFX was then conducted at the concentration that gave the largest reduction in optical density in section 2.8.2. No DNA-damaging agents was added to the third 200 ml aliquot and was kept as a control. The cultures

were then incubated at 37°C for 24 hours. The resulting lysates were then centrifuged at 3500 ×g for 10 minutes at 4°C to remove the bacteria and cell fragments from the medium, and the supernatant was then filtered through 0.22 μm filter unit and stored at 4°C. The lysate was then concentrated using ultracentrifuge and the bacteriophage DNA extraction method was conducted on the concentrated lysate. Moreover, transmission electron microscopy (TEM) was performed to investigate the presence of potential viral particle following induction process. The experiments were done with three biological replicates.

2.8.4 Prophage induction using ultraviolet light

The prophage induction using ultraviolet light was conducted on *L. micdadei* NCTC 11371. For this, 1.2 ml of an overnight bacterial culture of *L. micdadei* NCTC 11371 to be tested for lysogeny was transferred to 120 ml fresh BYE-α broth and was incubated at 37°C for $\sim$ 16 hours or until it gets to OD_{600}= 0.6 before being centrifuged at 6000 ×g for 10 minutes at room temperature. Pelleted cells were then resuspended in 120 ml sterile 0.1 M $MgSO_4$ and splitted into 2 x 60 ml, one was used for ultraviolet light (UV) induction and the other as a negative control (with out induction). The first 60 ml aliquot of cell culture suspension was then transferred to 120 mm square petri dishes and irradiated in a "Stratalinker® UV crosslinker" for 0.5 μJ cm^{-2}. The cells were then added to double strength BYE-α broth in a 1:1 ratio. The same procedure, except UV irradiation was conducted for negative control. Additionally, a 60 ml aliquot of fresh double strength BYE-α broth, plus the same volume of sterile 0.1 M $MgSO_4$ was include as a control for liquid media. At this point the 3 x 60 ml suspensions were then separated into 20 x 3 ml aliquots in 7 ml bijou, one bijou per time point. These bijous were then incubated at 37°C, shaken at 100 rpm for 72 hours. Samples were taken at 1 hour intervals from 1-6 hours and then for 8, 20, 22, 24, 26, 28, 30, 43, 48, 54, 67, and 72 hours. At each time point 2 ml was removed from the corresponding bijou and from this 1 ml was used to measure the absorbance at 600 nm (BYE-α broth with 0.1 M $MgSO_4$ was used as control zero of absorbance) and the other 1 ml was filtered through 0.22 μm filter for further analysis of potential viral-like particles sizes and concentration.

2.8.5 Testing the relationship between DNA concentration in induced lysate and inducer concentration in *L. micdadei*

To determine the required concentration of inducing agents to achieve higher yield of DNA from viral particles, the culture of *L. micdadei* NCTC 11371 was first exposed to a range of DNA-damaging agents concentrations using broth dilution method as explained in section 2.8.2. The bacteriophage DNA extraction was then conducted on each lysate and the concentration of obtained DNA was

measured using Qubit. To do this, a 2 ml working stock solution of 2000 ng ml^{-1} of MitC and NFX was prepared separately in 7 ml bijou tubes by adding 10 μl of working stock of either above inducing agents (400 μg ml^{-1}) into 1990 ml BYE-α broth. Two-fold serial dilutions were then made for both MitC and NFX, achieving the final concentration of 1000 ng ml^{-1} to 1.95 ng ml^{-1}. Following preparing the dilutions, 1 ml of *L. micdadei* NCTC 11371 cells from an exponentially growing culture (OD$_{600}$=0.2) was transferred to each bijou tube producing the final optical density of 0.1. The tubes were then incubated for 12 h at 37°C. The culture of *L. micdadei* NCTC 11371 without adding any inducing agents into the BYE-α broth was included in each batch of the assay tests as a control. The lysates were filtered through 0.22 μm syringe filters and stored at 4°C. On the same day, 750 μl of the filtered lysates were used for bacteriophage DNA extraction technique (Section 2.10).

2.9 Characterisation of bacteriophages

2.9.1 Spot test

To screen samples for presence of bacteriophages infecting *Legionella* spp., spot tests were carried out as illustrated in figure 2.4. For this purpose, charcoal was removed from Buffered Charcoal Yeast Extract (BCYE-α) agar by centrifugation after autoclaving to facilitate the screening for bacetriophage plaques. Then, 150 μl of overnight *Legionella* spp. culture (OD600 0.1) was mixed with a 3 ml molten 0.4 % BCYE-α semi solid agar and immediately poured onto a 1.2 % BCYE-α agar plate (Supplementary table S1). The plate was allowed to solidify, before 10 μl drops of bacteriophage suspension was spotted on top of the o.4 % agar. Plates were then incubated overnight at 37°C. After that, plates were checked for zones of lysis at the site of sample addition. Zones of lysis were then picked and transferred into 500 μl SM buffer (10 mM NaCl, 8 mM MgSO4.7H2O and 50 mM Tris-Cl) (Suppleentary table S1) using scalpel and kept overnight at 4°C. On the following day, the suspension was mixed by inverting the tube, before centrifugation at 10000 ×g for 10 min. The supernatant was then filtered and used to re-challenge the host in plaque assay as section 2.9.2.

2.9.2 Plaque assay

To carry out a plaque assay as shown in figure 2.4, the tested strain of *Legionella* spp. host was grown in BYE-α broth to OD$_{600}$ 0.1. 150 μl of the culture and 100 μl of sample caused lysis during spot test was dispensed into a Bijou, followed by 3 ml of 0.4 % agar. This was then mixed by gentle inversion and immediately poured onto 1.2 % BCYE-α plates (charcoal was removed by centrifugation). Before incubating the plates at 37°C, plates were allowed to set for 5 minutes at room temperature. Plates were checked for plaque formation on the following day.
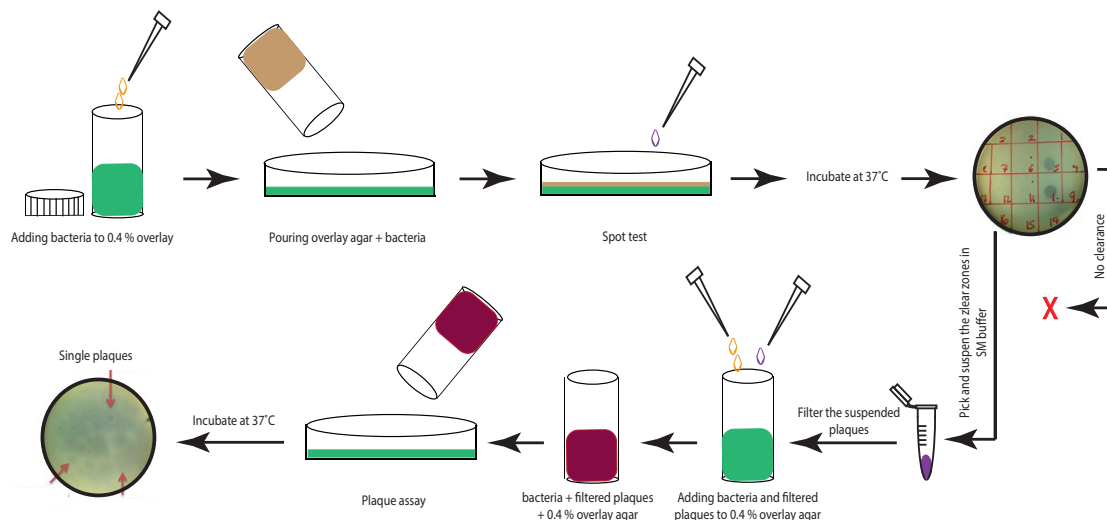
Figure 2.4. Schematic diagram of the work flow for detection of bacteriophages using spot test and plaque assay. Diagram shows the process of sample processing for bacteriophage detection using agar plates. First, spot test are conducted to determine which sample are able to produce zones of lysis on a bacterial lawn. This started by mixing the culture of individual *Legionella* spp. (orange) with overlay agar (Green), followed by spotting the enriched environmental sample (purple) on top of the overlay agar which was then incubated at 37° for 1-2 days. If a zone of clearance is produced, the zone of lysis was picked, suspended in SM buffer (purple), and then stored at 4°C overnight. The following day, the suspension was filtered to remove any agar particles and bacteria from the potential bacteriophage stock, and an aliquot of this stock was then mixed with a culture of the *Legionella* spp. that the zone of lysis was observed on plus overlay agar (claret) before being poured onto a petri-dish, and again incubated at 37° for 1-2 days. Following incubation the plates were then examined for the presence of single plaques to confirm the presence of bacteriophage in the sample.

2.9.3 Enumeration of prophage-like elements using Nanoparticle Tracking Analysis

In order to measure/estimate the particle size and concentration of viral-like particles within the lysate that achieved through prophage induction process, Nanoparticle Tracking Analysis (NTA) was used. The assay was performed according to the recommended protocol of the machine's manufacturer (Nanosight LM10). Briefly, 300 to 400 μl aliquots of a suspension of particles of induced lysate of *L. micdadei* NCTC 11371 culture at 19 different time-points were injected into the NanoSight LM10 (NanoSight Ltd., Amesbury, UK) machine's specimen chamber until the liquid reached the nozzle's tip. The Nanosight uses a finely focused laser beam that is introduced to the sample through a glass prism. The number of particles were measured at room temperature for 90 seconds. The data were then captured and analysed with NTA 2.0 Analytical Software.

2.9.4 Bacteriophage particle concentration using ultracentrifuge

For further concentration of potential viral particles, the induced bacteriophage lysates were ultra-centrifuged (Sorvall Discovery 100SE) at 151,000 xg for 3 hours at 4°C and then the supernatants were discarded. The pellets were then resuspended in 500 μl of BYE-α and stored at 4°C until later analysis.

2.9.5 Analysis of temperate virus-like particles using transmission electron microscopy

For further characterisation of potential viral particle, TEM was used to identify the morphologies of the particles, following conentarting of the lysates. To prepare the samples for observing under TEM, 200-mesh copper grids were coated with thin film of pioloform (a base for the carbon to be applied to). Before stabilising the pioloform coated copper grids by applying a thin coating of carbon, they were allowed to air dry for 1 h. The grids were made hydrophilic side up by high voltage glow discharge with argon gas for 30 s. 5 μl of bacteriophage preparation was deposited onto a 200-mesh Formvar/carbon-coated copper grid (Pelco International, Redding, CA) and was incubated for 2 min. The excess liquid was then removed from the grid by blotting with a piece of Whatman paper. The sample on the grid was washed twice in 5 μl of distilled-H_2O prior to staining. Finally, $2 \times 5\mu$l of 1% uranyl acetate was deposited onto 200-mesh Formvar/carbon-coated copper grid and the dye was removed immediately by a piece of Whitman paper. To achieve good contrast, care was taken to leave a thin film of the uranyl acetate stain on the grid and to avoid complete blotting. The grid was then allowed to air dry for 3 - 5 minutes and stored in a Petri dish lined with filter paper for TEM examination. The grids was observed at 80 kV with a JEOL-1400 transmission electron microscope (JEOL, Tokyo, Japan) equipped with a Megaview III digital camera with iTEM software (Gatan, Pleasanton, CA). Micrographs were recorded using the iTEM software (version 1.3), after the exposure time was set to 10 s.

2.10 Bacteriophage DNA extraction

Phenol/chloroform/isopropanol method was used to extract the bacteriophage DNA. Prior to viral DNA extraction, the lysate was concentrated using ultracentrifuge (Section 2.9.4). Then, 10 μl of each of RNase A (30 mg ml^{-1}) and DNase 1(30mg ml^{-1}) was added to 500 μl ultracentrifuged lysate and the mixture was then incubated overnight at 37°C. After testing the absence of genomic DNA using PCR targeting universal 16S rRNA gene, DNA extraction followed by the addition of equal volume of phenol to the previously generated resuspended bacetriophage pellet. The mixture was

then centrifuged at 20,000 ×g for 10 minutes at 4°C. The aqueous layer was extracted with an equal volume of Phenol:chloroform:isoamyl alcohol (25:24:1, v/v/v) solution, mixed well and was allowed to stand at room temperature for 2 minutes before being centrifuged at 20,000 ×g for 10 minutes. The aqueous layer was transferred to a new eppendorf tube and was extracted by adding an equal volume of chloroform: isoamyl alcohol (24:1, v/v) to the aqueous, mixed well and left to stand at room temperature for 2 minutes and then centrifuged at 20,000 ×g for 10 minutes. The resulting top layer was then transferred to a new eppendorf tube and mixed well with 2 volumes of isopropanol (propanol-2-ol) and 0.4 volumes of sodium acetate (7.5 M), and incubated overnight at 4°C. The following day, the sample were centrifuged at 21,000 ×g for 20 minutes, before removing the supernatant and the resulting DNA pellet was then left to be dried for 5 minutes. The pellet was washed with 500 μl of 75 % ethanol followed by a final centrifugation for 10 minutes at 21,000 ×g to recover the DNA pellet. After removing the residual liquid using a pipette, the DNA pellet was air-dried for 20 minutes in flow hood before being dissolved in in 70 μl of Elution Buffer (EB, 10 mM Tris-HCl, pH 8.5). The Qubit dsDNA HS Assay Kit from Life Technologies was used to quantify the bacteriophage DNA concentration. Afterwards, 150 ng of the eluted bacteriophage DNA sample was mixed with gel loading dye before being run on a 1 % agarose gel for 60 minutes at 100 V to detect a visible band of DNA.

2.11 Bioinformatics work

All bioinformatic analysis was processed using the CLIMB (the Cloud Infrastructure for Microbial Bioinformatics), a shared computing infrastructure (Connor et al., 2016).

2.11.1 Data collection of *Legionella* spp. genomes and identification of prophages across genomes

For identification of prophage elements within the genomes of *Legionella* spp., the 717 genomes of 73 *Legionella* spp. (these consisted of both complete and incomplete bacterial genomes) of the NCBI (the National Center for Biotechnology Information) database (https://www.ncbi.nlm.nih.gov/) were analysed. Three genomes of *L. pneumophila* lab-strains that were isolated from environmental samples were also included in this analysis. Prophage sequences were then identified using PHAge Search Tool Enhanced Release (PHASTER) (Arndt et al., 2016) and Prokaryotic Virus Orthologous Groups (pVOGs) (Grazziotin et al., 2017). PHASTER assigns a completeness score to each putative prophage based on factors such as the presence of genes homologous to a known bacteriophage, and the proportion of bacteriophage genes in the identified region denoted them either as complete (intact), questionable, or incomplete. Prophages that assigned as complete have a score of 100 or above, while questionable prophages score between 60-90. Those with a score < 60 are designated incomplete holding only < 50%

of the genes in their regions that are related to a known bacteriophage (Zhou et al., 2011). pVOGs is a website providing access to the most updated database of 9,518 orthologous groups shared among nearly 3,000 complete genomes of viruses that infect bacteria and archaea, including the functional annotation of viral proteins and identification of genes and viruses in uncharacterised DNA samples. Here, pVOGs database was used to reannotate the genes of *Legionella* spp. available in NCBI database, plus the available lab-strains, in order to classify the prophage elements by analysis of bacteriophage-related genes within them. To do this, the Viral clusters of Orthologous Groups (VOG) hidden Markov models (hmm) profiles were first downloaded from `http://dmk-brain.ecn.uiowa.edu/VOG/downloads.html` and the genome annotation was then conducted using prokka. Prokka is a software tool that uses a variety of database to annotate bacterial, archaeal and viral genomes quickly (Seemann, 2014). Here, prokka was used on downloaded VOG hmm profiles and the GenBank sequence of *Legionella* spp. genomes to assign function to their predicted CDS features for more precise prophage classification.

2.11.2 Phylogenetic analysis and genome comparisons of *Legionella* spp. complete prophages

To compare the similarity between *L. pneumophila* complete prophages and different members of *Microviridae* family, prophage phylogenetic trees were generated using selected bacteriophage genes related to *Microviridae* family (i.e., minor and major spike, DNA packaging, DNA initiation replication, DNA maturation, major capsid, and external scaffolding). The *Microviridae* bacteriophage that was used for the phylogeny were *Enterobacteria* phage G4 (NC_001420.2), *Enterobacteria* phage alpha3 (NC_001330.1), *Chlamydia* virus Chp1 (D00624.1), *Spiroplasma* virus SpV4 (M17988.1), *Enterobacteria* phage ID1 (DQ079880.1), *Enterobacteria* phage MED1 (KJ997912.1), *Enterobacteria* phage NC51 (DQ079891.1), *Coliphage* ϕ-X174 (NC_001422.1), *Enterobacteria* phage WA11 (DQ079895.1), and *Bdellovibrio* phage ϕ-MH2K (AF306496_1). For this, a stringent block alignment for each viral structural genes was created by simple amino acid comparison using Multiple Alignment using Fast Fourier Transform (MAFFT), a multiple sequence alignment program (Katoh and Standley, 2013). Then, maximum likelihood phylogenetic trees were constructed using RAxML (Randomized Axelerated Maximum Likelihood) program and these were visualised using the program FigTree 1.4.0. Furthermore, genome comparison regions were generated using the Blastn setting (Altschul et al., 1990) and synteny comparison figure created with EasyFig 2.2.2 (Sullivan et al., 2011), which is a Python application to show regions of sequence identity of multiple genomic loci with an easy-to-use graphical user interface (GUI).

2.11.3 Identification of integron elements in *Legionella* spp.

The sequences and annotations of 717 *Legionella* spp. genomes were downloaded from NCBI Gen-Bank (last accessed in December 2018, `https://www.ncbi.nlm.nih.gov/assembly`). All the above downloaded genomes were then screened for the presence of integron elements using IntegronFinder program (Cury et al., 2016), applying default settings (`https://github.com/gem-pasteur/Integron_Finder`). IntegronFinder is an online available program that can be used to identify integrons and their distinctive components; the *attC* sites with the use of covariance models, integron-integrases using hmm profiles, and other features (i.e, promoters, *attI* site) using pattern matching. In this study, a total of 32,822 sequence contigs of *Legionella* spp., both the downloaded genomes from NCBI GenBank and the lab-strains, were screened through IntegronFinder for ensuring an automatic and accurate identification of integrons, cassette arrays, and *attC* sites when localised less than 4kb apart. The threshold of 4kb was an default option, which was twice the size of the largest known cassettes ($\sim$ 2kb) to compromise between sensitivity (large values decrease the probability of missing cassettes) and specificity (small values are less likely to put together two independent integrons) of the identified profiles. The searches for the elements of the integron can be interpreted in three classes; complete integrons, *In0* elements, and CALIN. The identified elements with *intI* and at least one *attC* site were called as complete, while the elements with only *intI* and no *attC* sites were classified as *In0*. The cluster of attC site lacking inntegron-integrase (CALIN) elements contained at least two *attC* sites and none *intI*.

2.11.4 Data analysing and visualisation

All the data was processed and analysed using a Linux based operating system. All graphs were generated using the programs Pandas, Scipy, Numpy, Matplotlib, Seaborn, and Python3.

2.11.5 Statistical analysis

All the data was statistically analysed using either one-way ANOVA or two-sample t-test.

3. Isolation and characterisation of *Legionella* spp. from environmental samples

3.1 Abstract

Legionella spp. are the causative agents of legionellosis, a set of complex conditions including Legionnaires' disease and Pontiac fever. Legionellosis outbreaks are an emerging public health problem, contributing to increased mortality and morbidity, especially in cases that lack early diagnosis and treatment. Bacteriophages may provide an alternative means of treatment, diagnosis, and environmental cleaning and monitoring. However, no bacteriophages are currently known that infect *Legionella* spp.. To isolate any new bacteriophages, a collection of host strains is needed. Therefore, this chapter describes the collection of a diverse range of *Legionella* spp. isolated either from a total of 262 water and soil samples collected from natural reservoirs and man-made systems and analysed for the presence of *Legionella* spp. along with other strains obtained from PHE. Each isolate within the collection was characterised into subgroups by serotyping and sequence typing, corresponding to 20 serogroup 1 and 24 serogroup 2 - 14 isolates that covered 12 recognisable STs as well as 6 new sequence types (STs). These were then examined for their sensitivity to antibiotics used in the treatment of *Legionella* spp. infections and a common chlorine-based water treatment agent.

3.2 Introduction

Legionella spp. are waterborne pathogens that are common in hot and warm aquatic environments (Okafor, 2011). They replicate either directly within these niches, or within other organisms, such as amoebae (Albert-Weissenberger et al., 2007). In man-made systems, *Legionella* spp. can colonise and proliferate to high numbers, and are spread by aerosols produced by the devices, such as in the cases of cooling towers, evaporative condensers, spa pools, humidifiers, garden hoses, and even domestic hot water systems, if operating between 20° to 45°C (Farrell et al., 1990). Any water systems able to produce aerosols are considered high risk sources of *Legionella* spp. infections, due to the possibility of inhaling airborne water droplets produced that contain viable *Legionella* spp.. Inhalation of *Legionella* spp. containing droplets is a health risk due to the possibly of developing legionellosis upon the droplet reaching the lungs (Bartram et al., 2007).

To combat this risk, chlorine and its derivatives are commonly used to remove or reduce the density of *Legionella* spp. within water systems (WHO, 2015). However, this process can be challenging, due to the ability of this bacterium to escape treatment through biofilm formation and residing inside protozoa (Cervero-Aragó et al., 2015; Buse et al., 2019). This can also make detection of *Legionella* spp. within these water systems challenging (Montagna et al., 2017).

Detection is also problematic in clinical cases. For instance, Legionnaires' disease, the serious pneumonia form of legionellosis, can easily be initially misdiagnosed as other types of pneumonia, delaying diagnosis (Swartz, 1979; Murdoch, 2003). Diagnostic tests are available though, such as the *Legionella* urine antigen test where the patient's urine is screened for the presence of *Legionella* lipopolysaccharide (Berdal et al., 1979; Tilton, 1979; Lim et al., 2009). However, this test only detects infections caused by *L. pneumophila* serogroup 1 (Burillo et al., 2017; Miller et al., 2018). Therefore, culturing can often be required for diagnosis which can take time, anywhere from 3 to 14 days, due to the slow growing nature of the organisms, if it even grows at all. Lack of culturability in *Legionella* spp. is common, as *Legionella* spp. can exist in the VBNC state where, although present in the niche, the organisms are unable to be cultured in the laboratory (Oliver, 2000; Pierre et al., 2017). Thus, in both the cases of environmental and clinical detection a broader, and faster, diagnostic assay would allow treatment to begin sooner for infected individuals, and reduce the potential for outbreaks to occur in the first place by detecting them in environmental samples. Therefore, alternative approaches for rapid detection and elimination of *Legionella* spp. from environmental samples, and from patients, are crucial.

Bacteriophages are one potential alternative tool that could be used in both the detection and elimination of *Legionella* spp., as bacteriophage-based diagnostics have already been successfully developed for other bacteria such as, *B. anthracis* (Schofield and Westwater, 2009) and *M. tuberculosis* (Piuri et al., 2009), and bacteriophage-based therapeutics are also showing promise (Schooley et al., 2017).

Yet, so far in the literature only one group have described the isolation of bacteriophages infecting *Legionella* spp. (Lammertyn et al., 2008). However, these isolated bacteriophages no longer exist within the collection of that group (Elke Lammertyn, personal communication), meaning no known bacteriophages infecting *Legionella* spp. are currently available. Therefore, if bacteriophage-based diagnostics and therapeutics are to be developed, then new bacteriophages against these organisms need to be isolated. To isolate bacteriophages, it is vital to have a broad range of host strains due to the often narrow infective range of bacteriophages (Flores et al., 2011; Khan Mirzaei and Nilsson, 2015).

The first step of this study then aimed to collect of a diverse set of *Legionella* spp. by isolating these organisms from different natural and man-made water systems, for the purpose of eventually using them as hosts for bacteriophage screening. Isolates would then be tested to see if they were *L. pneumophila* or one of the other *Legionella* spp., followed by subgrouping the collection into serogroups and sequence types (ST) to measure the diversity of the isolates. Finally, the sensitivity of all the *L. pneumophila* and non-*pneumophila Legionella* spp. to a chlorine-based water treatment, sodium hypochlorite, and antibiotics frequently used in the treatment of *Legionella* spp. infections was examined to determine any characteristics that may be relevant to clinical treatment/environmental cleaning.

3.3 Results

3.3.1 Comparison of the growth of *L. pneumophila* in liquid media

In order to isolate *Legionella* spp. from environmental samples for use as hosts for bacteriophage isolation, it was a prerequisite to have an efficient liquid medium to support the rapid growth of this fastidious and slow-growing bacterium to a high cell density. Currently, there is no standard liquid medium used for the growth of *Legionella* spp. (Samuel Collins from Public Health England, personal communications). In the literature, different media for growing *Legionella* spp. in liquid medium are used, including BYE-α (Ristroph et al., 1980; Chatfield and Cianciotto, 2013), chemically modified medium (CDM) (Pine et al., 1979), and *Legionella* transparent medium (LTM) (Armon and Payment, 1990). BYE-α was selected here as it is known to support the growth of *L. pneumophila* (Ristroph et al., 1980; Chatfield and Cianciotto, 2013), as well as other species of *Legionella*; such as *L. longbeachae*, *L. micdadei*, *L. anisa*, *L. bozemanae*, and *L. dumoffii* (O'Connell et al., 1995; Edelstein et al., 2005; Stewart et al., 2009).

BYE-α is commonly used for culturing *Legionella* spp. on solid media, termed BCYE-α, due to a difference between its use in solid media and as a broth base. The solid media BCYE-α is different due to the use of activated charcoal. Normally for BCYE-α solid cultures, activated charcoal is added

to improve the growth of strains by removing superoxide radicals that form during the autoclaving of the media (Hoffman et al., 1983; Barker et al., 1986). However no charcoal is normally used in BYE-α as a liquid media. Therefore, the growth of *L. pneumophila* NCTC 11192 was compared across five modifications of BYE-α liquid growth media to determine which media would produce the most optimal growth of *Legionella* spp..

The liquid growth conditions used were based on the standard preparation of BYE-α media for liquid growth of *Legionella* spp., with modifications. Five variants of BYE-α liquid media were explored; filter-sterilised BYE-α, autoclaved BYE-α with no charcoal added at any stage, BYE-α with added charcoal (not removed), BYE-α with charcoal removed by filtration following autoclaving, and BYE-α with charcoal removed by centrifugation following autoclaving. Each of these media were inoculated with a suspension of *L. pneumophila* NCTC 11192 in SM buffer that was adjusted to a turbidity equivalent to McFarland 0.5, with growth monitored by optical density and colony production on BCYE-α agar for each culture over 96 hours (Fig. 3.1 A and B, respectively). Optical density for the BYE-α-charcoal culture was not measured as this media is dark black containing charcoal particles that gives inaccurate readings.

Growth was observed to be similar across all tested liquid media. One-way analysis of variance (ANOVA) demonstrated no significant difference in optical densities produced in each liquid media ($F = 0.03$, $P = 0.99$; Fig. 3.1 A). Trends were also similar. In each case exponential growth was observed after 4 hours, plateauing into stationary phase after around 24 hours at $\sim$ OD$_{600}$ 1.5, where it remained until the end of the experiment. This was also the case for colony formation, with no significant difference again observed by One-way ANOVA analysis between each of the liquid media ($F = 0.04$, $P = 0.99$; Fig. 3.1 B). In all cultures an initial increase in number of observed colony forming units was observed from 0 - 20 hours, followed by a plateau at $\sim$ 5 x 10^9 CFU ml^{-1}, until around 55 hours followed by a slight reduction in colony numbers. As no significant differences were observed between the tested liquid media, autoclaved BYE-α with no charcoal added at any stage was chosen for use in further experiments due to its simple preparation.

3.4 Isolation of *Legionella* spp. from environmental samples

In order to increase the probability of isolating bacteriophages infecting *Legionella* spp., it is essential to have a diverse panel of *Legionella* spp. isolates to screen potential bacteriophage-containing samples on. This is because bacteriophages can often be highly specific in which hosts they infect (Flores et al., 2011). Having a broad range of *Legionella* spp. isolates would increase the probability of having an organism sensitive to bacteriophage infection. Therefore, the next step in the bacteriophage isolation process was obtaining a culturable collection of *Legionella* spp. strains from solid and aquatic natural
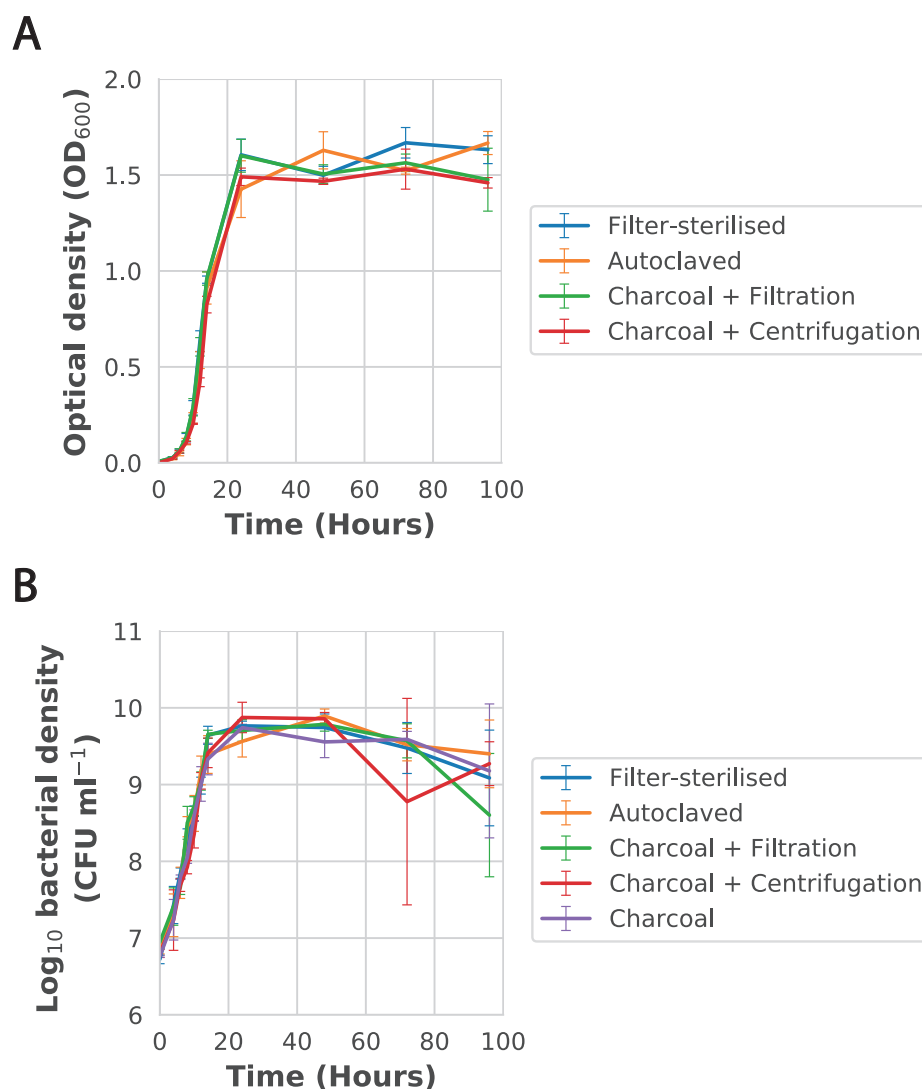
Figure 3.1. Comparison of the growth of *L. pneumophila* NCTC 11192 in five liquid media. The figure shows the growth profiles observed for *L. pneumophila* NCTC 11192 over 96 hours in five different culture media; Filter-sterilised BYE-α, autoclaved BYE-α, BYE-α with added charcoal, BYE-α with charcoal removed by filtration, and BYE-α with charcoal removed by centrifugation. (A) shows the monitored bacterial density in four broth. However, BYE-α with charcoal was omitted, due to the darkness of the broth. (B) represents the measurements of bacterial viable cells in all five liquid media over the same time-period. In each figure lines indicate the mean value at each time point (N = 3), while error bars indicate $\pm$ SD

and man-made environments.

To isolate a collection of *Legionella* spp., two rounds of sampling was conducted. In the first round a set of 108 samples were collected from environmental and man-made sources from different locations in the United Kingdom (Margate (2), Hampshire (14), Bath (4), Essex (7), Leicester (34)), as well as from other countries around the world, namely Nigeria (31), Iraq (11), Iran (2), and Kenya (3) (Supplementary table S2). In the second set a total of 154 liquid and solid samples were collected

from rivers and the surrounding areas in 5 cities within the Midlands, United Kingdom. These were samples from Leicester (41), Birmingham (36), Derby (22), Coventry (26), and Nottingham (29) (Supplementary table S3).

For the first set of samples (108 samples from various countries, collected from cooling tower, rivers, wells etc.), samples were tested for the presence of *Legionella* spp. using selective agar plates, medium of Wadowsky and Yee (MWY) (Descours et al., 2014). This agar is described as being selective for the growth of *Legionella* spp. as while containing antibiotics such as polymyxin B, anisomycin, and vancomycin (Wadowsky and Yee, 1981), it also contains two dyes; bromothymol blue and bromocresol purple, that aid in discriminating *Legionella* spp. from other organisms by acting as pH indicators, which show *L. micdadei* as blue-gray colonies, and *L. pneumophila* and some other *Legionella* spp. as green colonies (Vickers et al., 1981; Corry et al., 2011).

Before plating, a heat-treatment step was also used on each sample, as the presence of other microbes within the environmental samples can lead to difficulties in isolating *Legionella* spp., due to their slow growth and potential growth inhibition by other organisms (Bopp et al., 1981; Héchard et al., 2005). *Legionella* spp. are often more resistant to brief exposure to higher temperatures (e.g. 55°C for 15 minutes used here) than many other bacteria, therefore this step can minimise the concentration of viable non-*Legionella* spp., allowing for better isolation of *Legionella* spp. (Edelstein et al., 1982; Percival and Williams, 2013).

Any resulting colonies were then picked and each streaked on standard (BCYE-α) agar plates supplemented with or without cysteine. Since *Legionella* spp. require cysteine for their growth, only the isolates that grew on BCYE-α agar plates containing cysteine were selected. A total of 18 isolates passed this criteria from the 108 collected water samples, despite screening $\sim$ 3,120 colonies. Each of these colonies were then examined using PCR targeting the *L. pneumophila mip* gene (macrophage infectivity potentiator), a gene that can be used to discriminate between *L. pneumophila* and other *Legionella* spp. (Ratcliff et al., 1998). This indicated that all 18 isolates were positive for *mip* and were therefore *L. pneumophila*.

With the first set of samples collected only producing a small number of *Legionella* spp. isolates a second sample set (154 samples collected from 5 cities) was screened using the same method. However, this time no *Legionella* spp. colonies were detected. Therefore, quantitative polymerase chain reaction (qPCR) was used to determine if any *Legionella* spp. were present in any of these samples, that were unable to grow due to being in the VBNC state or due to inhibition from other organisms.

The qPCR assay used is based on the assay used by Collins et al. (2015), which describes the detection of *Legionella* spp. using two qPCR assays. One assay is a multiplex reaction targeting the *ssrA* and *mip* genes, while the other targets *wzm* (described in more detail in section 2.5.4). By quantifying these targets, these assays can be used to discriminate between the densities of specifically *L.*
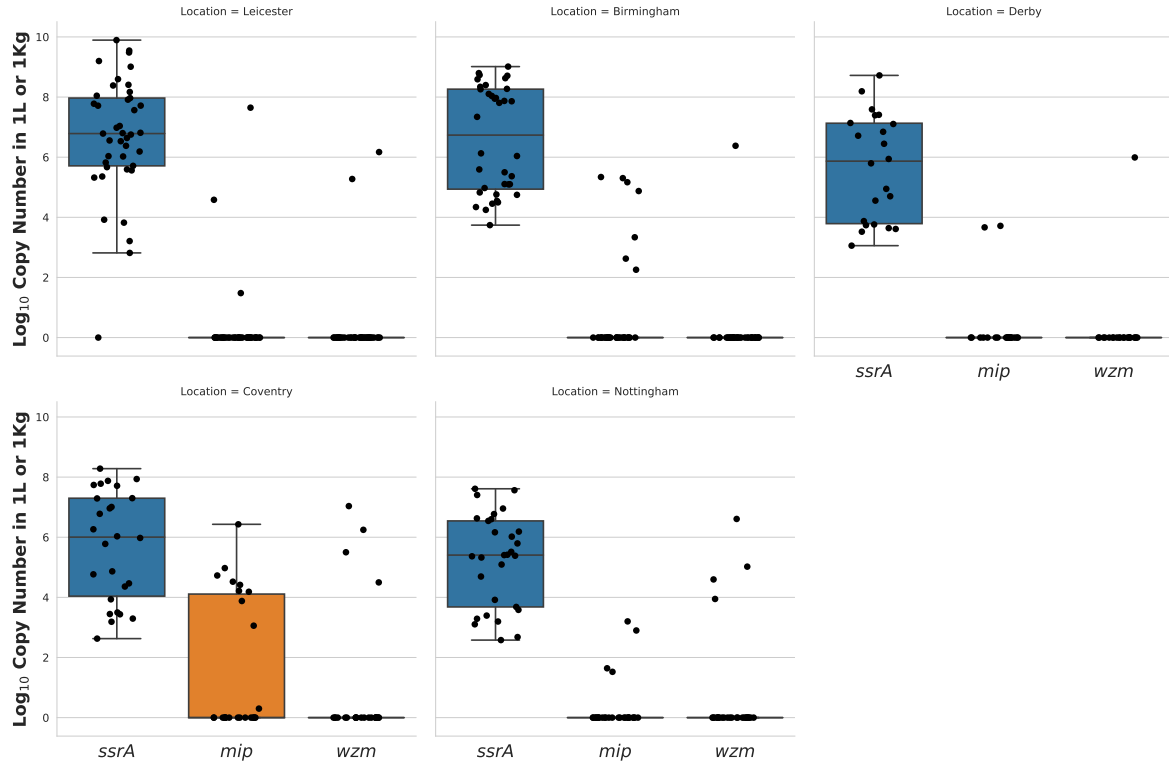
Figure 3.2. *Legionella* **spp. within samples from five cities in the Midlands, UK using qPCR.** Boxplots show the quantities each of the 3 target genes detected in each of the samples from the five cities. Targeted genes are *wzm* (for *L. pneumophila* serogroup 1, *mip* (for other *L. pneumophila* srogroups), and *ssrA* (for other *Legionella* spp.). The boxplots indicate the number of copies of each gene detected in each sample. The black horizontal line in each box indicates the median value, while the upper and lower regions of the box indicates the 25 % and 75 % quartiles, while whisker extending from the boxes correspond to 1.5 × the interquartile range (max/min values). Each black circle indicates a single sample.

pneumophila serogroup 1, the total density of *L. pneumophila*, and the collective abundance of all *Legionella* spp.. This is because the *ssrA* gene can be found in all *Legionella* spp. (Thurman et al., 2011), while *mip* is specific for *L. pneumophila* serogroups 1 - 14 (Engleberg et al., 1989), and *wzm* is only present in serogroup 1 strains of *L. pneumophila* (Mérault et al., 2011). An example of this can be seen in figure 3.3.

From the 154 samples, only 1 sample (0.64 %) was negative for any *Legionella* spp. by qPCR, while *L. pneumophila* were detected in 25 samples (16.23 %), with 12 (7.79 %) samples being positive for *L. pneumophila* serogroup 1 (Figure 3.3). The distribution of the positive samples with each of these genes in each of the 5 cities can be seen in figure 3.2. Comparing all 5 cities; Leicester, Birmingham, Derby, Coventry, and Nottingham (Figure 3.2 A-E respectively), significant differences were observed with the levels of detectable *ssrA* (one-way ANOVA, $F = 4.04$, $P = 0.003$) and *mip* (one-way ANOVA, $F = 3.37$, $P = 0.01$) in the samples. No significant differences were found in the quantity of *wzm* found
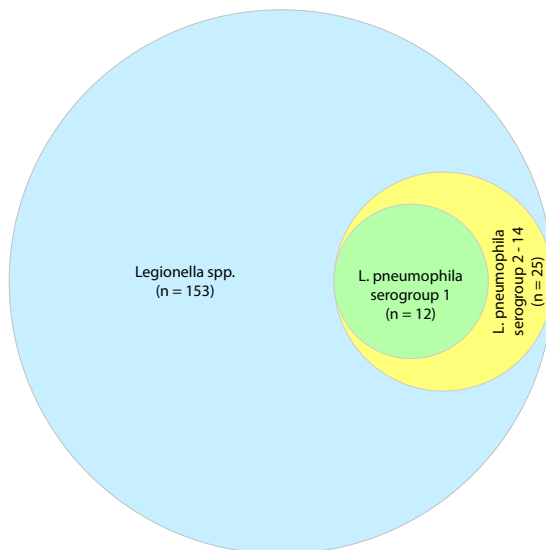
Figure 3.3. The proportion of *Legionella*-positive samples detected by qPCR. Venn diagram showing the proportions of screened environmental samples from rivers around the Midlands, UK, that were positive for either *Legionella* spp. (*ssrA* positive only; Blue circle; n = 153), *L. pneumophila* 1 - 14 (positive for *ssrA* and *mip*; Yellow circle; n = 25), and *L. pneumophila* serogroup 1 (positive for *ssrA*, *mip* and *wzm*; Green circle; n = 12) using the Collins et al. (2015) qPCR assay.

among the tested locations in the Midlands (one-way ANOVA, $F = 1.23$, $P = 0.29$).

Although the isolation of *Legionella* spp. from the first set of samples led to only 18 isolates being obtained, and the second sample led to no new *Legionella* spp. being isolated, a further 37 *Legionella* spp. isolates were graciously provided by Dr. Samuel Collins at Public Health England. Of these 37 strains of *Legionella* spp., 27 were *L. pneumophila*, 2 were *L. longbeachae*, and 1 isolate for each of the following *Legionella* spp., *L. sainthelensi*, *L. bozemanae*, *L. anisa*, *L. hackeliae*, *L. oakridgensis*, *L. gormanii*, *L. feeleii*, *L. dumoffii*.

3.5   Type analysis of *L. pneumophila* isolates

Currently, the genus *Legionella* comprises over 50 species (Lück et al., 2013), with *L. pneumophila* itself able to be sub-classified into 14 serogroups by using a latex agglutination test to detect antigens targeting distinct lipopolysaccharide structures (Yong et al., 2010; Lück et al., 2013). To determine the diversity within the collection of *L. pneumophila* hosts in the current study a two step strain typing approach was used to further classify these isolates. In the first approach serogroup typing using a latex agglutination test was used to determine if the *L. pneumophila* isolates belonged to the more pathogenic serogroup 1 or were part of the less pathogenic serogroups 2 - 14. This was then followed by sequence based typing to subclassify each serotyped *L. pneumophila* isolate at the genotype level.

Table 3.1. Serogroup Typing of *Legionella* spp. using Oxoid *Legionella* Latex Test

Strain	Sample ID	Source	Reference	*sg 1	*sg 2 - 14	*other	Serogroup
CLP	NCTC 11192	Human lung	National Collection of Type Cultures (NCTC)	+	-	-	1
E1	N/A	Roman Baths (Bath, UK)	This study	-	+	-	2 - 14
E2	N/A	Roman Baths (Bath, UK)	This study	-	+	-	2 - 14
GBIn1	N/A	Roman Baths inlet (Bath, Uk)	This study	-	+	-	2 - 14
GBIn3	N/A	Roman Baths inlet (Bath, Uk)	This study	-	+	-	2 - 14
GBIn6	N/A	Roman Baths inlet (Bath, Uk)	This study	-	+	-	2 - 14
GBO1	N/A	Roman Baths outlet (Bath, Uk)	This study	-	+	-	2 - 14
GBO3	N/A	Roman Baths outlet (Bath, Uk)	This study	-	+	-	2 - 14
GBO6	N/A	Roman Baths outlet (Bath, Uk)	This study	-	+	-	2 - 14
GBO4	N/A	Roman Baths outlet (Bath, Uk)	This study	+	+	+	Cross reaction
SS3	N/A	Roman Sacred Spring (Bath, Uk)	PHE	-	+	-	2 - 14
SS4	N/A	Roman Sacred Spring (Bath, Uk)	PHE	-	+	-	2 - 14
SS6	N/A	Roman Sacred Spring (Bath, Uk)	PHE	-	+	-	2 - 14
GH1	N/A	Hot-tap water(Leicester, UK)	This study	+	-	-	1
GH11	N/A	Hot-tap water(Leicester, UK)	This study	+	-	-	1
GH18	N/A	Hot-tap water(Leicester, UK)	This study	+	-	-	1
NL1	N/A	Hot-tap water (Leicester, UK)	This study	+	-	-	1
NL4	N/A	Hot-tap water (Leicester, UK)	This study	+	-	-	1

Table 3.1. – Continued from overleaf

Strain	Sample ID	Source	Reference	*sg 1	*sg 2 - 14	*other	Serogroup
NL7	N/A	Hot-tap water (Leicester, UK)	This study	+	-	-	1
Phe1	89957	Hospital calorifier	PHE	+	-	-	1
Phe11	PO1318902-07	ND	PHE	-	+	-	2 - 14
Phe12	PO1323950-05	ND	PHE	-	+	-	2 - 14
Phe13	PO1324253-01	Drain point	PHE	-	+	-	2 - 14
Phe14	PO11406064-03	Ship tap	PHE	-	+	-	2 - 14
Phe15	6820-01	ND	PHE	-	+	-	2 - 14
Phe16	2075-01	ND	PHE	-	+	-	2 - 14
Phe17	1467-01	ND	PHE	-	+	-	2 - 14
Phe18	1753-03	ND	PHE	-	+	-	2 - 14
Phe19	7839-03	ND	PHE	-	+	-	2 - 14
Phe2	PO1311089-03	Ship shower	PHE	+	-	-	1
Phe20	9072-01	ND	PHE	-	+	-	2 - 14
Phe21	4111-02	ND	PHE	-	+	-	2 - 14
Phe22	56476	ND	PHE	-	+	-	2 - 14
Phe23	2851-01	ND	PHE	-	+	-	2 - 14
Phe24	1646-02	ND	PHE	+	-	-	1
Phe25	91157	ND	PHE	+	-	-	1
Phe26	0394-08	ND	PHE	+	-	-	1
Phe27	0394-06	ND	PHE	+	-	-	1

Table 3.1. – Continued from overleaf

Strain	Sample ID	Source	Reference	*sg 1	*sg 2 - 14	*other	Serogroup
Phe28	2579-03	ND	PHE	+	-	-	1
Phe30	7784-03	ND	PHE	+	-	-	1
Phe4	PO1312868-06	Domestic tap	PHE	+	-	-	1
Phe5	PO1315682-08	Hospital mains	PHE	+	-	-	1
Phe6	PO1323309-06	Hospital tap	PHE	+	-	-	1
Phe7	PO1404336-01	Hospital tap	PHE	+	-	-	1
Phe8	PO1403754-04	Ship shower	PHE	+	-	-	1

3.5.1 Serogroup Typing

All *L. pneumophila* strains isolated as part of this study (n = 18), plus those that were received from PHE (n = 27), were serogroup-typed using the *Legionella* latex test (Oxoid, Basingstoke, UK). This test is able to classify the *L. pneumophila* isolates into either serogroup 1 or as belonging to one of serogroups 2 - 14 (i.e. it does not discriminate between the serogroups within this classification e.g. serogroup 3 vs serogroup 11). The results from the serogroup typing are shown in table 3.1.

Of all 45 examined isolates, 20 were found to belong to *L. pneumophila* serogroup 1 (44.5 %) and 24 to *L. pneumophila* seroroup 2 - 14 (53.5 %). One isolate, *L. pneumophila* GBO4, was found to react with all three test reagents, antibodies specific for *L. pneumophila* serogroup 1, those for *L. pneumophila* serogroup 2 - 14, and even the set that reacted with other *Legionella* spp.. Even after numerous replicates, and subculturing single colonies further in case the stock contained mixed *L. pneumophila* serogroups, this result was consistently observed. No explanation was found for the cross reaction observed for *L. pneumophila* GBO4, but as it was confirmed by PCR earlier that this strain belongs to *L. pneumophila* it was still included in the collection of *L. pneumophila* strains for further subclassification by sequence-based typing.

3.5.2 Sequence-based typing

To further characterise the *L. pneumophila* isolates, typing at a molecular level was conducted using sequence-based typing. Sequence-based typing has been recommended by the European Working Group for *Legionella* Infections (EWGLI) for the sub-classification and genotyping of *L. pneumophila* (Mentasti et al., 2012; Mentasti, M. and Fry, 2012; Lück et al., 2013). Sequence-based typing is performed by sequencing amplicons from seven gene targets; *flaA*, *pilE*, *asd*, *mip*, *momps*, *proA*, and either *neuA* or *neuAH*, with *neuAH* used when *neuA* amplification was negative (Mentasti, M. and Fry, 2012; Mentasti et al., 2014). The sequences are then submitted to an online database available at the PHE website (http://www.hpa-bioinformatics.org.uk/legionella/legionella_sbt/php/sbt_homepage.php), where the sequences are then compared to a set of known sequences to determine sequence homology. Base on matches an allele number is then assigned to each individual gene input, and depending on the combination of allele numbers found within a strain, an overall sequence type (ST) for that strain is determined using the European Society of Clinical Microbiology and Infectious Diseases' (ESCMID) ESGLI online tool (ESCMID Study Group for Legionella Infections; `http://bioinformatics.phe.org.uk/legionella/legionella_sbt/php/sbt_homepage.php`) (e.g. allelic profile 1,4,3,1,1,1,1 = ST1) (Gaia et al., 2005; Mentasti et al., 2014).

Targets of either *neuA* or *neuAh* and the 6 other target genes were successfully amplified by PCR from all 45 *L. pneumophila* isolates. *NeuA* was amplified for 44/45 (97.8 %). However, *L. pneu-*

Table **3.2.** Sequence-based typing of *L. pneumophila*

Strain	Sample ID	Source	Reference	*flaA*	*pilE*	*asd*	*mip*	*mompS*	*proA*	*neuA*	*neuAH*	ST
CLP	NCTC 11192	Human lung	NCTC	3	4	1	1	14	9	1		ST36
E1	N/A	Roman Baths (Bath, UK)	This study	10	22	7	28	16	18	6		ST337
E2	N/A	Roman Baths (Bath, UK)	This study	10	22	7	28	16	18	6		ST337
GBIn1	N/A	Roman baths inlet (Bath, Uk)	This study	10	22	7	28	16	18	6		ST337
GBIn3	N/A	Roman baths inlet (Bath, Uk)	This study	10	22	7	28	16	18	6		ST337
GBIn6	N/A	Roman baths inlet (Bath, Uk)	This study	10	22	7	28	16	18	6		ST337
GBO1	N/A	Roman baths outlet (Bath, Uk)	This study	10	22	7	28	16	18	6		ST337
GBO3	N/A	Roman baths outlet (Bath, Uk)	This study	10	22	7	28	16	18	6		ST337
GBO6	N/A	Roman baths outlet (Bath, Uk)	This study	10	22	7	28	16	18	6		ST337
GBO4	N/A	Roman baths outlet (Bath, Uk)	This study	17	23	13	20	32	22	-	205	ST707
GH1	N/A	Hot-tap water(Leicester, UK)	This study	1	4	3	1	1	1	1		ST1
GH11	N/A	Hot-tap water(Leicester, UK)	This study	1	4	3	1	1	1	1		ST1
GH18	N/A	Hot-tap water(Leicester, UK)	This study	1	4	3	1	1	1	1		ST1
SS3	N/A	Roman Sacred Spring (Bath, Uk)	PHE	10	22	7	28	16	18	6		ST337
SS4	N/A	Roman Sacred Spring (Bath, Uk)	PHE	10	22	7	28	16	18	6		ST337
SS6	N/A	Roman Sacred Spring (Bath, Uk)	PHE	10	22	7	28	16	18	6		ST337
NL1	N/A	Hot-tap water (Leicester, UK)	This study	1	4	3	1	1	1	9		ST8
NL4	N/A	Hot-tap water (Leicester, UK)	This study	1	4	3	1	1	1	9		ST8

Table 3.2. – Continued from overleaf

Strain	Sample ID	Source	Reference	flaA	pilE	asd	mip	mompS	proA	neuA	neuAH	ST
NL7	**N/A**	Hot-tap water (Leicester, UK)	This study	1	4	3	1	1	1	9		ST8
Phe1	**89957**	Hospital calorifier	PHE	1	4	3	1	1	1	1		ST1
Phe11	**PO1318902-07**	PHE	ND	7	6	17	28	13	11	1		ST2412
Phe12	**PO1323950-05**	ND	PHE	2	6	17	3	13	11	7		ST2413
Phe13	**PO1324253-01**	Drain point	PHE	6	10	19	28	19	4	6		ST191
Phe14	**PO11406064-03**	Ship tap	PHE	3	10	1	28	14	9	13		ST93
Phe15	**6820-01**	ND	PHE	12	7	11	25	11	12	1		ST2410
Phe16	**2075-01**	ND	PHE	6	10	19	28	19	4	6		ST191
Phe17	**1467-01**	ND	PHE	2	10	3	28	9	14	1		ST851
Phe18	**1753-03**	ND	PHE	11	14	16	31	15	13	1		ST2418
Phe19	**7839-03**	ND	PHE	1	4	3	16	2	1	1		ST1180
Phe2	**PO1311089-03**	Ship shower	PHE	1	4	3	1	1	1	1		ST1
Phe20	**9072-01**	ND	PHE	6	10	17	28	4	14	3		ST2414
Phe21	**4111-02**	ND	PHE	12	7	11	25	11	12	1		ST2410
Phe22	**56476**	ND	PHE	5	1	22	30	6	10	1		ST856
Phe23	**2851-01**	ND	PHE	11	14	16	31	15	13	1		ST2418
Phe24	**1646-02**	ND	PHE	7	6	14	3	13	11	9		ST2411
Phe25	**91157**	ND	PHE	7	6	17	3	13	11	11		ST59
Phe26	**0394-08**	ND	PHE	7	6	17	3	13	11	11		ST59
Phe27	**0394-06**	ND	PHE	7	6	17	3	13	11	11		ST59

Table 3.2. – Continued from overleaf

Strain	Sample ID	Source	Reference	*flaA*	*pilE*	*asd*	*mip*	*mompS*	*proA*	*neuA*	*neuAH*	ST
Phe28	2579-03	ND	PHE	7	6	17	3	13	11	11		ST59
Phe30	7784-03	ND	PHE	7	6	17	3	13	11	11		ST59
Phe4	PO1312868-06	Domestic tap	PHE	7	6	14	3	13	11	9		ST2411
Phe5	PO1315682-08	Hospital mains	PHE	7	6	17	3	13	11	11		ST59
Phe6	PO1323309-06	Hospital tap	PHE	7	6	17	3	13	11	11		ST59
Phe7	PO1404336-01	Hospital tap	PHE	1	4	3	4	1	1	15		ST79
Phe8	PO1403754-04	Ship shower	PHE	1	4	3	1	1	1	1		ST1

mophila GBO4, the isolate that also reacted with all 3 serotyping reagents in Section 3.5.1, did not amplify with *neuA* gene, and therefore, *neuAh* was amplified and sequenced instead. From the 45 isolates, 18 distinct STs were identified (Figure 3.4). Of these 18 STs, 12 were assigned to a known STs, while 6 STs were identified as novel STs and were uploaded to the EWGLI database and given the ST numbers ST2410, ST2411, ST2412, ST2413, ST2414, and ST2418. The most prominent STs in this study were ST337 and ST59 with 11 (26 %) and 7 (16 %) strains of the *L. pneumophila* collection identified as these STs respectively (Figure 3.4).
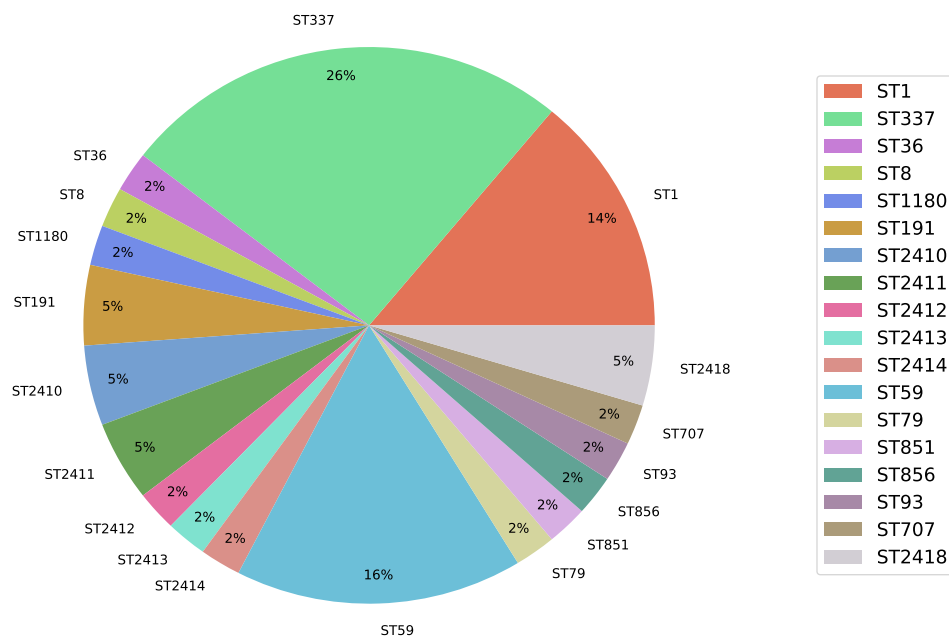
Figure 3.4. Proportion of the sequence types identified within the collection of *L. pneu-mophila* isolates. Pie chart shows the percentage of strains identified as a particular sequence type from the collection of *L. pneumophila* isolates that was used in this study (n = 45). Each colour represents a different sequence type, with sequence type indicate both on the outer rim of the corresponding slice, and within the key on the right of the pie chart.

3.6 Susceptibility testing of *Legionella* spp.

For further characterisation of *Legionella* spp. strains in the collection, the susceptibility of these isolates were tested against antimicrobials and a water treatment agent. In the first test, the sensitivity of *Legionella* spp. isolates to four antibiotics was tested, from the macrolides, fluoroquinolones, and rifamycins families that are commonly used for treating infections caused by this bacterium (Bartlett et al., 2000; Sabrià et al., 2005). This allowed the examination of whether any of these strains showed resistance to the common antibiotics used in the treatment of *Legionella* spp. infections. The sensitivity of the same isolates were then also tested against sodium hypochlorite, which is commonly used for

chemical disinfection of *Legionella* spp. in many water systems (Kim et al., 2002; Joseph et al., 2011; Health and Safety Executive, 2013). In total 44/57 strains from the collection were examined. This was because any isolates that were isolated from the same sample and were identified as the same sequence type were noted as duplicates and only one strain from any non-unique set was included here.

3.6.1 Antibiotic susceptibility of *Legionella* spp. isolates

The susceptibility of these isolates to different antimicrobial agents was first checked by assessing the *in vitro* activities of two macrolides (erythromycin and clarithromycin), a flouroquinolone (ciprofloxacin), and a rifamycin (rifampcin) against 44 isolates within the *Legionella spp.* collection (1 clinical and 43 environmental isolates). Antibiotic screening was conducted according to guidelines adapted from European Committee on Antimicrobial Susceptibility Testing (EUCAST) for antimicrobial susceptibility testing of *L. pneumophila* (Bruin et al., 2012). Here, the minimal inhibitory concentrations (MICs) of the 4 antimicrobials were obtained using an agar diffusion method with commercially available E-test strips (BioMérieux) which contain an exponential gradient of antibiotic concentrations along the strip. Isolates on plates were incubated with the strips for 48 hours, and MICs were then recorded based on the first concentration noted on the the strip where bacterial growth was inhibited (Wiegand et al., 2008). The *Legionella* spp. that were tested with the above antimicrobial agents, included 17 *L. pneumophila* serogroup 1, 17 *L. pneumophila* serogroup 2 - 14, 1 *L. pneumophila* cross-reacted between the reagents for identifying *pneumophila* and 9 non-*pneumophila* strains, one strain each of *L. longbeachae*, *L. micdadei*, *L. bozemanae*, *L. anisa*, *L. hackeliae*, *L. oakridegensis*, *L. gormanii*, *L. feelei*, and *L. dumoffii* strains. The distribution of the MIC values of 4 antibiotics tested are shown in figure 3.5.

After exposing of *L. pneumophila* and other *Legionella* spp. to all four antibiotics, variable sensitivities were observed, as can be noted in Figure 3.5. For rifampicin, the *L. longbeachae* strain Phe_37 and *L. hackeliae* strain Phe_34 showed the lowest level of sensitivity to rifampicin (0.0476 and 0.047 μg ml^{-1}, respectively). The *L. pneumophila* strains Phe_19, Phe_2, and GBO1 showed the lowest level of sensitivity to erythromycin, ciprofloxacin, and clarithromycin at 0.833, 0.543, and 0.231 μg ml^{-1}, respectively. Interestingly, for the other *Legionella* spp. the sensitivity of Phe_34 to all antibiotics was low compared to the other *Legionella* spp., and showed the lowest sensitivity for erythromycin, ciprofloxacin, and clarithromycin among the *Legionella* spp. isolates.

All the isolates were more sensitive to rifampicin in comparison to the other antibiotics, with growth of *Legionella* spp. inhibited at concentrations of between 0.002 - 0.006 μg ml^{-1} for *L. pneumophila* and 0.002 - 0.047 μg ml^{-1} for the non-*pneumophila Legionella* spp. (Figure 3.5 A). The sensitivity of *L. pneumophila* to rifampicin was significantly greater than the other non-*pneumophila Legionella* spp.

(Two-sample *t-test*, $t = 3.33$, $P = 0.0017$). However, no significant difference in observed MICs was noted between MICs of the *L. pneumophila* strains and the other non-*pneumophila Legionella* spp. when they exposed to erythromycin (Figure 3.5 B; *t-test* = -0.102, $P = 0.918$), ciprofloxacin (Figure 3.5 C; *t-test* = 1.51, $P = 0.136$), and clarithromycin (Figure 3.5 D; *t-test* = 0.21, $P = 0.82$).

None of the strains isolated showed resistance to the antibiotics tested.
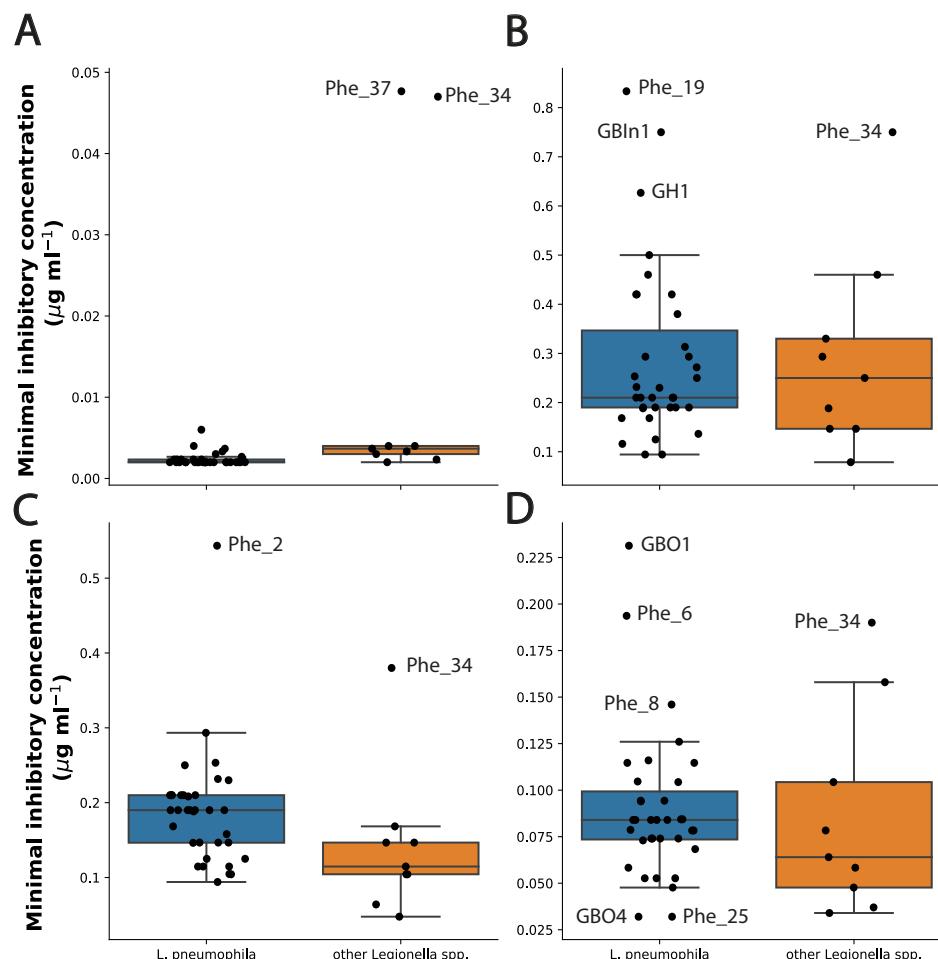
Figure 3.5. Antibiotic susceptibility of the *L. pneumophila* versus non-*pneumophila Legionella* spp.. Boxplots show the distribution of the minimal inhibitory concentration (MIC) of antibiotics against the *L. pneumophila* and non-*pneumophila Legionella* spp. used in this study. Blue boxes represent values obtained for *L. pneumophila* while orange boxes show non-*pneumophila Legionella* spp.. (A) shows the MIC values for rifampicin, (B) for erythromycin , (C) for ciprofloxacin , and (D) shows those for clarithromycin. Boxes show the 25 % to 75 % quartiles, the solid horizontal line within the box indicates the median of each distribution, and whiskers extend to 1.5 × the interquartile range, with outliers beyond represented marked as black circles and are labelled. Each black circle shows the mean MIC value obtained from 3 biological replicates for a particular isolate.

3.6.2 Sodium hypochlorite susceptibility

Strains were further characterised to determine their sensitivity to a disinfectant that is used in the water treatment process to remove bacteria such as *Legionella* spp., sodium hypochlorite. Sodium hypochlorite was selected as according to the EWGLI guidelines for control of Legionnaires' disease and national water regulations, continuous addition of chlorine in the form of sodium hypochlorite is often employed for the disinfection of water (Joseph et al., 2011). To do this, a blank sensitivity disk was impregnated with 10 μl of 2 % sodium hypochlorite and placed on an individual agar plate for each of 44 strains, which were spread over the surface. Following this, the zone of inhibition caused by this bactericide on the lawn of *Legionella* was measured after 48 hours incubation at 37°C as described in section 2.5.2. The larger the zones of lysis, the greater the sensitivity to the bactericidal agent.

The susceptibility of the *Legionella* spp. collection to sodium hypochlorite were compared again between *L. pneumophila* and the non-*pneumophila Legionella* spp., and the distributions of the both of these groups can be found in figure 3.6. *L. pneumophila* isolates were found to be more sensitive to sodium hypochlorite compared with non-*pneumophila* isolates, as larger zones of inhibition were observed around the sodium hypochlorite containing disk. *L. pneumophila* showed zones of inhibition ranging from 28.5 - 36.8 mm with a median diammeter of 32.16 mm, while for the non-*pneumophila Legionella* spp zones produced were between 17.33 - 31.83 mm, with a median diammeter of 25.66 mm (Figure 3.6). The higher sensitivity of *L. pneumophila* isolates was statistically significant to bactericidal activity of sodium hypochlorite (Two-sample *t-test* = 5.4, *P* = < 0.05).

Among *L. pneumophila* isolates, two serogroup 1 isolates, *L. pneumophila* Phe_24 and NCTC 11192 showed the lowest and highest sensitivity to sodium hypochlorite respectively (28.5 mm and 36.83 mm respectively). As for the other *Legionella* spp., *L. hackeliae* showed the smallest zone of inhibition (17.33 mm diammeter), while the largest zone of inhibition was observed by *L. micdadei* (31.83 mm diammeter).

3.7 Discussion

To date, very little has been published on bacteriophages infecting *Legionella* spp., and as a result, there is a lack of knowledge on the abundance and diversity of bacteriophages that infect this bacterium. Because of this, if the current study aimed to investigate the potential for bacteriophages to be used in the treatment and detection of *Legionella* spp. the isolation of new bacteriophages was required. To isolate bacteriophages against *Legionella* spp. we would first require a broad range of isolates of *Legionella* spp. to increase the chances of finding a bacteriophage. Interestingly, Lammertyn et al. (2008) obtained all four of their bacteriophages using only a single host strain, *L. pneumophila* NCTC
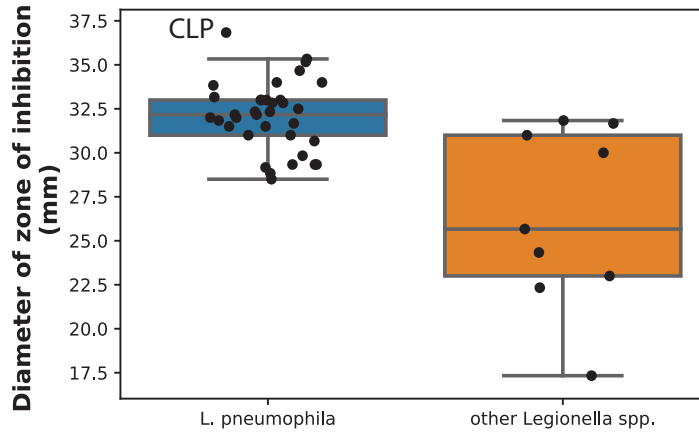
Figure 3.6. Susceptibility of *L. pneumophila* versus non-*pneumophila Legionella* spp. to sodium hypochlorite using a disk diffusion assay. Boxplots represent the distribution of the diameter of zones of inhibition produced following exposure of each of the 44 *L. pneumophila* and non-*pneumophila Legionella* spp. strains collected to 2 % sodium hypochlorite. Blue boxes represent values obtained for *L. pneumophila* while orange boxes show non-*pneumophila Legionella* spp.. Boxes show the 25 % to 75 % quartiles, the solid horizontal line within the box indicates the median of each distribution, and whiskers extend to 1.5 × the interquartile range. Each black circle shows the mean diameter obtained from 3 biological replicates for a particular isolate.

11192. Although isolated on a serogroup 1 *L. pneumophila* strain, the four bacteriophages were found to be able to infect every other *L. pneumophila* serogroup, except serogroup 13. Two of the four bacteriophages were even able to infect other *Legionella* spp., those being *L. bozemanae*, *L. micdadei*, *L. longbeachae*, and *L. dumoffii* (Lammertyn et al., 2008). Therefore, as the previously observed bacteriophages were able to infect multiple serogroups and species of *Legionella*, by having a wide collection of host serogroups and species in the present work the probability of having one strain able to be infected by such a bacteriophage will increase.

To begin this process, the composition of the liquid media that would be used to culture *Legionella* spp. in all further sections was examined to try and obtain maximal growth of *Legionella* spp.. By obtaining maximum bacterial densities the probability a bacteriophage interacting with a bacterial cell will be higher than at lower densities (Abedon and Katsaounis, 2018). This is because the number of bacteriophages lost from an environment due to interacting and adsorbing to a host bacterial cell increases as a function of bacterial densities, which can be seen by the equation:

$$P_t = P_0 e^{-kNt}$$

where P_t, is the number of bacteriophages at time t, P_0 is the number of bacteriophages initially in the environment, k is the bacteriophage adsorption constant, and N is the density of bacteria in that

environment (Abedon, 2012). Host density is also an important factor in determining eventual bacteriophage densities, as for each cell that is adsorbed to by a bacteriophage, new viruses will be released from that cell, a concept that is applied in bioreactor cultures during production of bacteriophages to try and produce maximum bacteriophage titres (Agboluaje and Sauvageau, 2018). While no significant differences were found between the growth of *L. pneumophila* NCTC 11192 between media containing activated charcoal (either only during the autoclaving process or at all times) and charcoal free media, the maximum density observed in each culture was high, ~ 9.89 x 10^9 CFU ml^{-1}. This concentration will be high enough to have a sufficient probability of leading to bacteriophage adsorption should bacteriophages infecting *Legionella* spp. be present, and enough to support the production of higher densities of bacteriophages from adsorbed bacteriophages.

The liquid growth optimisation experiments were conducted using *L. pneumophila* NCTC 11192, a serogroup 1 isolate that had been purchased from the National Collection of Type Cultures (NCTC), and although this strain is the same that was used by Lammertyn et al. (2008) to isolate their bacteriophages, a larger collection would improve the chances of bacteriophage isolation. To obtain these, a set of 108 samples from environmental and man-made water systems were obtained from various regions around the world. Several different approaches; such as acid, buffer or heat pre-treatment, as well as heat enrichment, to perform environmental monitoring for *Legionella* spp. have been proposed to improve chances of isolating *Legionella* spp. (Reinthaler et al., 1993; Centers for Disease Control (CDC), 2005). Here, heat treatment prior to culturing was selected because unlike many other bacteria, *Legionella* spp. are relatively resistant to higher temperature (55°C - 60°C) (Leoni et al., 2001; Chang et al., 2007). Therefore, heat-treatment combined with antibiotic selective media was used to increase the chance of isolating *Legionella* spp. by eliminating/reducing the density of competing bacteria. Screening these samples resulted in isolation of 18 strains of *L. pneumophila*, confirmed by amplification of the *mip* gene, which is not present in other *Legionella* spp. (Hayden et al., 2001). These numbers are consistent with other investigations. For example, Carvalho et al. (2007) screened 67 water samples from various man-made water systems in São Paulo, Brazil, and found only 9 samples were positive for *L. pneumophila* (13.4 % of samples). However, this can vary drastically though across sample geographical locations and the source being sampled (Whiley and Taylor, 2016).

Collection of a further 154 environmental samples from five cities in the Midlands, UK, to try and expand the number of *Legionella* spp. did not give any more *Legionella* spp. isolates. It was possible then that no *Legionella* spp. were present in any of these samples. However, other non-*Legionella* colonies were isolated from these samples, and it has been suggested by Toze et al. (1990) that a lack of *Legionella* spp. when other organisms are present could be due to an anti-*Legionella* effect, inhibiting *Legionella* spp. growth. In the environment, *Legionella* spp. are in constant contact with non-*Legionella* bacteria that share the environment (Toze et al., 1990). Although there is only limited evidence on inhibition of the growth of *Legionella* spp. by non-*Legionella* bacteria, studies have shown

that some organisms e.g. *Aeromonas* spp., *Acinetobacter* spp., *Bacillus* spp., *Flavobacterium* spp., and *Pseudomonas* spp., *Staphylococcus* spp., and *Streptococcus* spp. (Flesher et al., 1980; Héchard et al., 2005; Guerrieri et al., 2008; Marchand et al., 2011; Loiseau et al., 2015) are capable of inhibiting or limiting the growth of *Legionella* spp.. Corre et al. (2018) isolated 273 non-*Legionella* spp. (covering 4 phyla) from 5 freshwater environments and found that 178 (65.2 %) of these isolates were able to inhibit the growth of *L. pneumophilla in vitro*, indicating that organisms able to inhibit the growth of *Legionella* spp. are present in water samples such as those sampled here. Another possibility is the presence of VBNC *Legionella* spp. that are known to be present in some samples but are unable to cultured (Whiley and Taylor, 2016; Kirschner, 2016).

Culture-independent analysis of these samples by qPCR was used to examine these possibilities. qPCR indicated that 153 of the 154 ($\sim$ 99.3 %) were positive for *Legionella* spp., indicating that *Legionella* spp. are present in the samples collected. This would suggest then that growth is either being inhibited by other organisms or because these cells are in the VBNC state. Differences in the detection of *Legionella* spp. between culture and qPCR are common. Diederen et al. (2007) were only able to identify *Legionella* spp. in 8/357 potable water water samples from around the Netherlands, while qPCR analysis of these same samples indicated the presence of *Legionella* spp. DNA in 311/357 samples. In river samples, Parthuisot et al. (2010) were able to culture *Legionella* spp. from 15/72 samples while qPCR indicated their presence in all 72 samples. However, qPCR amplification can also occur for both living and dead *Legionella* spp. within a sample, so in some cases values could include some false positive results due to amplification of dead cells. Despite this, as the focus of the current study was to eventually isolate bacteriophages infecting *Legionella* spp., the presence of *Legionella* spp. DNA in the samples could be a good indication of bacteriophages presence, as bacteriophages are believed to be in every environment where bacteria are found, as they rely on a host for replication.

With only 18 *Legionella* spp. isolates collected from the two rounds of sampling, 37 additional isolates of *Legionella* spp. were graciously provided from the collection of the Biosafety, Air and Water Microbiology Group, National Infection Service, PHE, 27 being *L. pneumophila*. Other species in the set provided by PHE consisted of *L. longbeachae* (2 isolates), *L. sainthelensi* (1 isolate), *L. bozemanii* (1 isolate), *L. anisa* (1 isolate), *L. hackeliae* (1 isolate), *L. oakridgensis* (1 isolate), *L. gormanii* (1 isolate), *L. feeleii* (1 isolate), and *L. dumoffii* (1 isolate). Additionally, 1 isolate of each of *L. pneumophila* and *L. micdadei* were purchased from NCTC. This made the total number of *Legionella* spp. in the collection of this study 57 isolates.

All 45 isolates of *L. pneumophila* in the collection were then further sub-classified to determine the diversity of these strains using serogroup typing and sequence-based typing. This would allow the further characterisation of the *L. pneumophila* and also limit the use of repetitive strains in the

eventual bacteriophage samples screening. Serogroup typing of the 45 strains of *L. pneumophila* showed that 20 of these isolates belonged to serogroup 1 and 24 belonged to serogroup 2 - 14. One isolate, *L. pneumophila* GBO4, was found to react with all three test reagents, antibodies specific for *L. pneumophila* serogroup 1, those for *L. pneumophila* serogroup 2 - 14, and for other *Legionella* spp.. Sequence-based typing of the *L. pneumophila* isolates showed that the collection covered 18 STs, 12 currently known STs and 6 novel STs, the majority belonging to either ST337 (26 %) or ST59 (16 %). A study by Quero et al. (2018) found that ST337 can be identified within hospital environments, although was not directly associated with disease. Although this may not be the same for all members of a sequence-type, ST59 is a known disease causing agent. From analysis of the STs deposited within the CDC *Legionella* reference collection during the last 30 years, Kozak-Muiznieks et al. (2014) noted that ST59 was the causative agent of an outbreak in Washington, USA, in 2011 (3 cases). This was a recurrent outbreak, with the strain detected in environmental samples leading to the disease for at least one year despite the treatment. Other outbreak-associated STs from the study of Kozak-Muiznieks et al. (2014) were found within the the current *L. pneumophila* collection. Of particular note is ST36, which was associated with the most outbreaks covered in their study in the USA, including 21 cases in Deleware in 1994, 4 cases in North Carolina in 2005, 6 cases in Texas in 2006, 5 cases in Georgia in 2009, and 11 cases in Illinois in 2012. ST1 which covers 14 % of the current *L. pneumophila* collection was associated with one outbreak in Nevada in 2012, but has also been identified in clinical cases in the UK, as has ST8, ST59, and ST191 (Harrison et al., 2009; Kozak-Muiznieks et al., 2014). With a collection of *L. pneumophila* isolates that span both environmental and potentially clinically relevant STs the isolates were then further characterised for their sensitivity to antibiotics used in the treatment of disease and a chlorine-based chemical used in water treatment.

The most commonly used antibiotics to treat Legionnaires' disease are macrolides, fluoroquinolones, and rifamycin (Bartlett et al., 2000; Sabrià et al., 2005), which are able to accumulate in phagocytes meaning they can tackle intracellular *Legionella* spp. (Edelstein and Edelstein, 1989). Only one case of antibiotic resistant *Legionella* spp. has been identified in a clinical setting, with a *L. pneumophila* serogroup 1 isolate obtained from a patient with severe pneumonia showing resistance to the fluoroquinolone, ciprofloxacin (Bruin et al., 2014). Resistance can also be readily developed *in vitro*, and while no treatment failures due to resistance have been reported, these are signs that this still could be a possibility in the future (Dowling et al., 1984; Nielsen et al., 2000; Almahmoud et al., 2009).

In this study, the antimicrobial susceptibility of 44 of the *Legionella* spp. isolates was assessed against four antibiotics to evaluate their sensitivity to common treatments. Although these data would not directly aid in isolating bacteriophages, it would allow further characterisation of diversity within the *Legionella* spp. collection. No isolates of *L. pnuemophila* were found to be resistant to any of the antibiotics based on the "epidemiological cut-off values" determined by Bruin et al. (2012). Epidemiological cut-off values are defined by determining the distribution of MIC's across the wild-

type population (organisms with no known resistance mechanisms), and identifying its upper limit of MIC (Kahlmeter et al., 2003). Bruin et al. (2012) determined epidemiological cut-off values for 10 antibiotics used in the treatment of *L. pnuemophila*, including the four examined in this study using only isolates from *L. pnuemophila* serogroup 1. They determined that the epidemiological cut-off values for rifampicin was 0.032 mg L^{-1}, erythromycin was 1.0 mg L^{-1}, ciprofloxacin was 1.0 mg L^{-1}, and clarithromycin was 0.5 mg L^{-1}. No *L. pnuemophila* isolates in this current study had an MIC greater than this threshold. However, no epidemiological cut-off values have been set for either *L. pnuemophila* serogroup 2 - 14 or non-*pnuemophila Legionella* spp., therefore while Phe_34 and Phe_37 (*L. hackeliae* and *L. longbeachae*, respectively) are above the epidemiological cut-off values for rifampicin it is unknown if these actually contain resistance mutations or just naturally have higher MIC's than *L. pnuemophila* serogroup 1 isolates. The susceptibility of *Legionella* spp. strains in the gathered collection were also tested against a common water treatment agent, sodium hypochlorite, using a disk diffusion assay. Our observations showed that *L. pneumophila* strains were more sensitive to 2 % sodium hypochlorite, compared to non-*pneumophila Legionella* isolates.

To conclude, a broad and diverse strain collection of environmental isolates of *Legionella* spp. was collected, including *L. pneumophila* serogroup 1 belonging to ST's that have been previously associated with disease outbreaks. This panel can then be used as a collection of host isolates to attempt to isolate bacteriophages in the next chapter.

4. Isolation and characterisation of bacteriophages infecting *Legionella* spp.

4.1 Abstract

Legionella spp. are pathogens with a number of key characteristics that make the detection and treatment of environmental contamination and associated human infections challenging. There is therefore a need to identify and develop new diagnostic and treatment methods targeting these organisms. Bacteriophages have been suggested as a possible alternative, and have shown promise as a potential tool in the detection and treatment of many bacteria. However, currently there are no known bacteriophage infecting any of the *Legionella* spp. available in public database. This chapter then aimed to isolate and characterise bacteriophages infecting *Legionella* spp. using two common approaches: enrichment of environmental samples (for free bacteriophages from the environment) and induction of *Legionella* spp. strains from the collection of *Legionella* spp. isolated in the previous chapter (for prophages present within bacterial host). For isolation by enrichment, a first set of 108 samples were enriched and analysed for the presence of free bacteriophages by attempting to detect plaque formation. However, only zones of confluent lysis were observed in samples showing clearance, with plaque formation only observed for 1 sample, which was not reproducible. Concentration of a total of 403 water and soil samples from natural reservoirs and man-made systems also showed non-reproducible decreases to bacterial density when bacterial growth was measured in a well assay. For isolation by induction, *Legionella* spp. treated with the prophage-inducing agents: MitC and NFX, showed some changes in growth profile possibly indicating prophage induction. Following TEM analysis no particles similar to the common *Caudovirales* were observed, however other virus-like particles were observed, which may be currently undescribed bacteriophage.

4.2 Introduction

Water systems with the ability to produce aerosols are considered high risk sources of *Legionella* spp. infection, due to the risk of inhalation of aerosolised droplets containing the organisms (Bartram et al., 2007). If droplets containing *Legionella* spp. become inhaled, the individual is at risk of developing legionellosis, the collective name covering Pontiac fever, a mild-flu like disease, and Legionnaires' disease, a severe and potentially deadly form of pneumonia (Fields et al., 2002; Bartram et al., 2007). In 2018, a total of 480 confirmed cases of Legionnaires' disease were reported in England and Wales (Public Health England, 2018). Across Europe, The European Surveillance System (TESSy) reported $\sim$ 8,500 confirmed case of Legionnaires' disease across 30 European countries in 2017, an increase from $\sim$ 6,500 confirmed cases the previous year (European Centre for Disease Prevention and Control, 2018, 2019). In both 2017 and 2016 TESSy reported that the fatality rate from all cases was $\sim$ 8 % (European Centre for Disease Prevention and Control, 2018, 2019). Thus, these pathogens pose an important public health threat across European countries.

To combat the threat caused by these organisms, organisations such as the Health and Safety Executive (HSE) in the UK regulate the *Legionella* spp. levels in environmental and man-made systems by routine environmental monitoring and epidemiological investigations following outbreaks of Legionnaires' disease (Health and Safety Executive, 2013). One way to prevent disease occurrence is by the elimination of *Legionella* spp. from water and aerosol-producing systems (Mérault et al., 2011). This is achieved by the treatment of potentially contaminated water with chemical disinfectants e.g. chlorine-based bleaches, or by using high temperatures to reduce/remove *Legionella* spp. (Kim et al., 2002; Marchesi et al., 2016). However, *Legionella* spp. have been shown to resist the action of many disinfectants and heat treatments, due to their ability to form biofilms and obtain refuge within amoebae (Barker et al., 1992; Falkinham, 2015; Abu Khweek et al., 2013; Abdel-Nour et al., 2013).

In cases where infection does happen, clinical detection of *Legionella* spp. can also be problematic. For example, Legionnaires' disease cases can easily be misdiagnosed as other types of pneumonia, causing a delay in diagnosis and administering appropriate treatment (Swartz, 1979; Murdoch, 2003). Currently the *Legionella* urine antigen test is used for diagnosis of Legionnaires' disease (Lim et al., 2009), yet this test can only be used accurately to detect infections caused by *L. pneumophila* serogroup 1 (Burillo et al., 2017; Miller et al., 2018). A study by Helbig et al. (2003) tested the ability of the urine antigen tests to detect *Legionella* spp. infections in known culture-positive patients and found that the test was accurate for travel and community acquired infections where *L. pneumophila* serogroup 1 were more frequently found ($\sim$ 95 % and 80 % of examined cases respectively), but was far less accurate for nosocomial infections where significantly less *L. pneumophila* serogroup 1 isolates were observed ($\sim$ 45 % of cases). Therefore, culturing can often be required for diagnosis, which can take anywhere from

3 to 14 days to confirm, due to the slower growing nature of these organisms (Bartram et al., 2007; Pierre et al., 2017). These delays in effective diagnosis could then possibly lead to serious implications for the patient if infection was due to an alternative *L. pneumophila* serogroup or *Legionella* spp.

In both the cases of environmental and clinical detection a broader and faster alternative approach to detecting *Legionella* spp. is required for the rapid detection and elimination of *Legionella* spp.. One approach that has been suggested for many other bacteria for similar purposes is the application of bacterial viruses, bacteriophages, to overcome these issues. Bacteriophages have already been shown to be effective when used as diagnostics and antimicrobials against various bacterial infections (Karumidze et al., 2012; Satyanarayana et al., 2012; Abedon, 2015). For example, in a clinical trial into the use of bacteriophages to treat chronic otitis caused by antibiotic-resistant *P. aeruginosa*, Wright et al. (2009) observed that a cocktail of bacteriophages was able to significantly reduce *P. aeruginosa* numbers compared to placebo, and also showed signs of improved clinical outcome. Sergueev et al. (2010) developed a bacteriophage-based qPCR diagnostic test for the identification of *Yersinia pestis* in clinical/environmental samples. In their assay, *Y. pestis* bacteriophages are added to the sample and primers targeting bacteriophage DNA are used to detect and quantify bacteriophage replication. Detection of bacteriophage replication can confirm the presence of *Y. pestis* within a sample in ~ 4 hours, while conventional clinical tests can take around $48 - 72$ hours. Quantification of bacteriophage replication also has the additional benefit in this assay of allowing quantification of the number of bacteria in the sample within this time frame (Sergueev et al., 2010).

There is a significant issue though in the use of bacteriophages in detecting and treating *Legionella* spp.. Currently, there is only one mention of the isolation of bacteriophages infecting *Legionella* spp. within the literature, an article by Lammertyn et al. (2008). They describe the isolation of 4 bacteriophages infecting *Legionella* spp. that had been isolated from environmental water systems around Belgium. However, these bacteriophages were all lost during laboratory purification following publication (Elke Lammertyn, personal communication). Thus, there are currently no know bacteriophages able to infect *Legionella* spp. that can be used in the development of new methods of detecting and controlling *Legionella* spp..

The aim of this chapter then was to isolate new bacteriophages infecting *Legionella* spp., to allow the development of bacteriophage therapeutics and diagnostics. This would be attempted through two distinct approaches. First, the water and soil samples from environmental and man-made sources would be screened for the presence of free *Legionella* spp. bacteriophages using the *Legionella* spp. isolates from chapter 3 as potential hosts. In the second method, as prophages are found throughout bacterial genomes (Casjens, 2003), the *Legionella* spp. isolates would also be screened for the presence of induceable prophages within their genomes using common bacteriophage induction agents followed by TEM to observe any induced bacteriophage particles.

4.3 Results

4.3.1 Screening of samples from environmental and man-made sources for bacteriophages infecting *Legionella* spp. using spot tests and plaque assays

Bacteriophages are common throughout nature. It has been estimated that as many as 10^{31} bacteriophages exist within the biosphere, more than 10x the estimated number of bacteria (Hatfull and Hendrix, 2011; Clokie et al., 2011). As bacteriophages obligately require a host to replicate, they are found in and around all locations that bacteria are found, both in environmental and host-associated systems (Koskella et al., 2011; Brown-Jaque et al., 2016). Therefore, to isolate bacteriophages infecting *Legionella* spp. here, samples from environmental and man-made sources where *Legionella* spp. are commonly identified were screened for the presence of bacteriophages using sample "enrichment". In this process the target bacterial species is mixed with the sample to be screened, together with liquid culture media that allows for high levels of bacterial replication within the reaction (Twest and Kropinski, 2009). By allowing bacterial densities to increase, the probability of an interaction occurring between a bacterium and a bacteriophage increases and also allow bacteriophage to undergo many round of replication. This will result in an increase in bacteriophage titre within the reaction mixture, potentially leading to sufficient numbers to produce plaques on a lawn of the host organism for visual detection.

The 108 water or soil samples collected in chapter 3 from environmental and man-made systems from various geographical locations (Figure 2.1) were first used for enrichment. The 44 strains of *Legionella* spp. that were collected in chapter 3 were used as the host cells in these enrichments, and cover 17 *L. pneumophila* serogroup 1, 17 *L. pneumophila* serogroup 2 - 14, 1 *L. pneumophila* that cross-reacted with the serogrouping antibodies, and one strain each of *L. longbeachae*, *L. micdadei*, *L. bozemanae*, *L. anisa*, *L. hackeliae*, *L. oakridegensis*, *L. gormanii*, *L. feelei*, and *L. dumoffii*. This included *L. pneumophila* NCTC 11192, which is the same strain of *L. pneumophila* used by Lammertyn et al. (2008) in their bacteriophage isolation study. After enrichment, visual screening for bacteriophages was conducted by spotting lysate that had been passed through a 0.22 μm filter on a lawn of the host to determine if any inhibition of bacterial growth would occur for the tested sample.

Spotting filtrates produced zones of lysis on several host lawns, such as the the clearances caused by two samples collected, shown in figure 4.1 A, one from Kurdistan, Iraq (sample 5) and another from Margate sea (sample 10) on the lawn of *L. pneumophila* E1. A total of 10 isolates showed at least one zone of clearance after spotting of the enrichment filtrate. Strain *L. pneumophila* E1 was found to produce zones of clearance for 23 samples when the enrichment filtrate was spotted onto its lawn,

the highest number among the tested collection. Meanwhile, sample 5 from Kurdistan, Iraq produced zones of clearance on the most bacterial lawns, producing clearance on *L. pneumophila* strains E1, NCTC 11192, GBO4, GH1, GBIn1 and SS1.

For samples where zones of clearance were observed, each zones were then transferred into 500 μl fresh BYE-α broth, and incubated at 4°C to allow bacteriophages to diffuse from the agar. After incubation, aliquots were then filtered through a 0.22 μm filter and used in a plaque assay, where 100 μl of the aliquot was mixed directly with the host strain with this mix then used to make the lawn of bacteria. This process would allow the visualisation of individual bacteriophage plaques if the zone of lysis was caused by a bacteriophage and not another form of an antagonist. This would also allow individual plaques to be selected, leading to the isolation of a pure stock of any bacteriophage. Only one sample produced single plaques on a lawn of *L. pneumophila* E1 (Figure 4.1 B). A total of 35 single plaques were then picked from this plate and transferred into BYE-α broth and kept overnight at 4°C, filtered through a 0.22 μm filter, and then stored at 4°C until needed for further analysis.
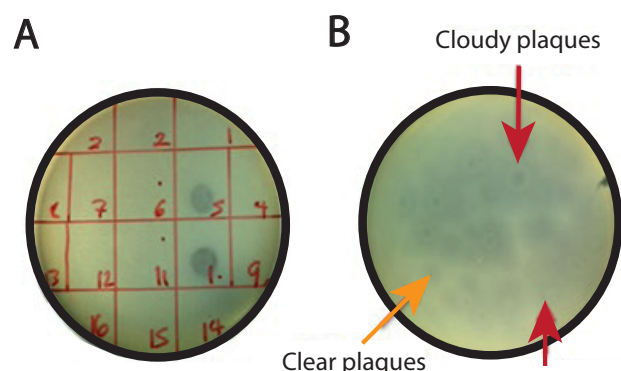
Figure 4.1. Examples of results obtained from the spot test and plaque assay tests. (A) shows the result of spotting enriched filtrates onto the lawn of *L. pneumophila* E1. Each section indicates the region where the filtrate spot of a single sample was added. The clear zoned produced by spotting a sample from Kurdistan, Iraq, and by a sample from Margate Sea, UK, can be found seen in sections numbered 5 and 10 respectively. (B) shows the result of the plaques assay produced from mixing the sample from Iraq, with *L. pneumophila* E1 strain. This produced single plaques, including those with different morphologies e.g. the plaques with a cloudy centre, and plaques that produced clearer plaques.

To further examine the single plaques formed on the lawn of *L. pneumophila* E1 by sample 5, the 35 picked single plaques were examined by TEM to determine if they did contain bacteriophages. From the TEM analysis it was found that bacteriophages were present in each plaque, and that only members of the *Caudovirales* were noted. In particular, members of the *Myoviridae* and *Siphoviridae* families were observed (Figure 4.2). The lysates were then going to be used for further propagation and plaque-purification. This would be particularly useful in some TEM images, where multiple bacteriophage morphologies could be observed within one of the single plaques meaning a mixed population (e.g. Figure 4.2 A, where both *Siphoviridae* and *Myoviridae* can be seen). However, all the attempts to

obtain single plaques again from each of the isolated plaques failed, and even repeating the plaque assays with the same enriched sample failed to produce any single plaques again, even after numerous attempts.
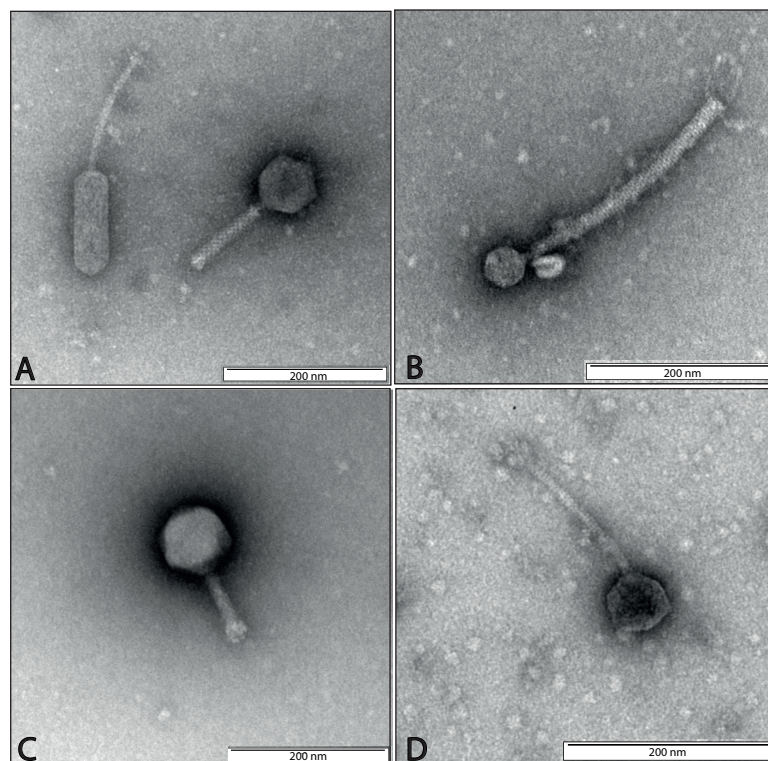
Figure 4.2. Electron micrograph of single plaque lysates obtained from the sample from Kurdistan, Iraq, following enrichment using *L. pneumophila* E1. Each section contains a TEM images showing the bacteriophage observed from an extracted single plaques produced by the sample from Kurdistan, Iraq, sample on the lawn of *L. pneumophila* E1. (A) show a sample containing more than one family of bacteriophage, a *Siphoviridae* (left) and a *Myoviridae* (right). In (B) and (C), *Myoviridae* can be seen, and (D) shows a *Siphoviridae*. Scale bars represent 200 nm in each case.

Due to the failed attempts to reproduce the observation of single plaques again using BCYE-α plates, additional forms of *Legionella* spp. media was used to try and observe single plaques again. Distinct media were investigated to determine if this was a factor in controlling whether single plaque formation occurred. For instance modifying BCYE-α by replacing charcoal with catalase or a media with no charcoal, *Legionella* transparent media (LTM), as activated carbon can cause virus inactivation (Cookson and North, 1967). Previous works by Byrne and Swanson (1998) and Al-Bana et al. (2014) demonstrated a correlation between the growth phase of *Legionella* spp. and two morphologically distinct forms; a replicating form in exponential growth and a transmissive form in stationary phase. As *Legionella* spp. can switch between two morphologies based on growth phase, plaque assays were also conducted using target bacteria in either of these growth phases. However, none of these modifica-

tions allowed single plaques to be formed on the test strains for any of the samples that showed zones of clearance. Therefore an alternative approach to isolating bacteriophage from from environmental samples was required

4.3.2 Screening of samples from environmental and man-made sources for bacteriophages infecting *Legionella* spp. using well assay

Following the failed attempts to obtain reproducible single plaques and as most samples where zone of clearance were observed did not produce single plaques, an alternative method of detecting bacteriophages was conducted. In this method two BYE-α broths; one supplemented with divalent cations known to aid in the bacteriophage adsorption process (MgCl$_2$ and CaCl$_2$) (Rountree, 1955; Rohan and Kropinski, 2009), and one without any divalent cations, were used to culture the host organism together with the samples from environmental and man-made sources in wells of a 96-well microtitre plate (Supplementary table S1). The well assay was chosen as it occurs in liquid media and may allow visual detection of bacteriophages by monitoring reductions in bacterial density in the presence of the sample. Upon inoculation of each well with the target exponential-growing *Legionella* spp., the growth pattern of *Legionella* spp. were then monitored every 15 minutes for 72 hours. For any cultures that showed a reduction in growth compared to the control (bacterial culture without adding any samples), the resulting lysates were then centrifuged and used in the assay for another round to confirm the lysate would be capable of causing consistent growth reduction, which would be expected if the sample did contain bacteriophages.

This time, samples included the second sample set described in chapter 3 which were obtained from rivers around the Midlands, UK. To improve throughput, and also the chance of finding bacteriophages in the environmental samples, all 403 of the samples to be screened were separated into 14 sample groups based on factors such as, location, and being culture- or qPCR-positive for *Legionella* spp. (Supplementary table S7), with each sample group concentrated together using either tangential flow filtration and Amicon Ultra-15 centrifugal filter units. Sample 1 contained all the Midlands' qPCR-positive samples for *wzm* genes; sample 2: all the environmental samples that were collected in the first round of sampling; sample 3: the Midlands' qPCR-positive samples for *mip* genes; sample 4: the Midlands' qPCR-positive samples for *ssrA* genes; sample 5: tap water sample from University of Leicester, which were culture-positive for *L. pneumophila* serogroup 1, Sample 6; tap water sample from Leicester General Hospital, Leicester, that were culture-positive for *L. pneumophila* serogroup 1; sample 7: environmental water samples from Nigeria; sample 8: water samples from cooling tower; sample 9: the only Midlands water sample that was negative for *Legionella* spp. by qPCR; sample 10: environmental water samples from Bath, UK, culture-positive for *L. pneumophila* serogroup 2 - 14; sample 11: environmental water sample from PHE; sample 12: environmental water sample from
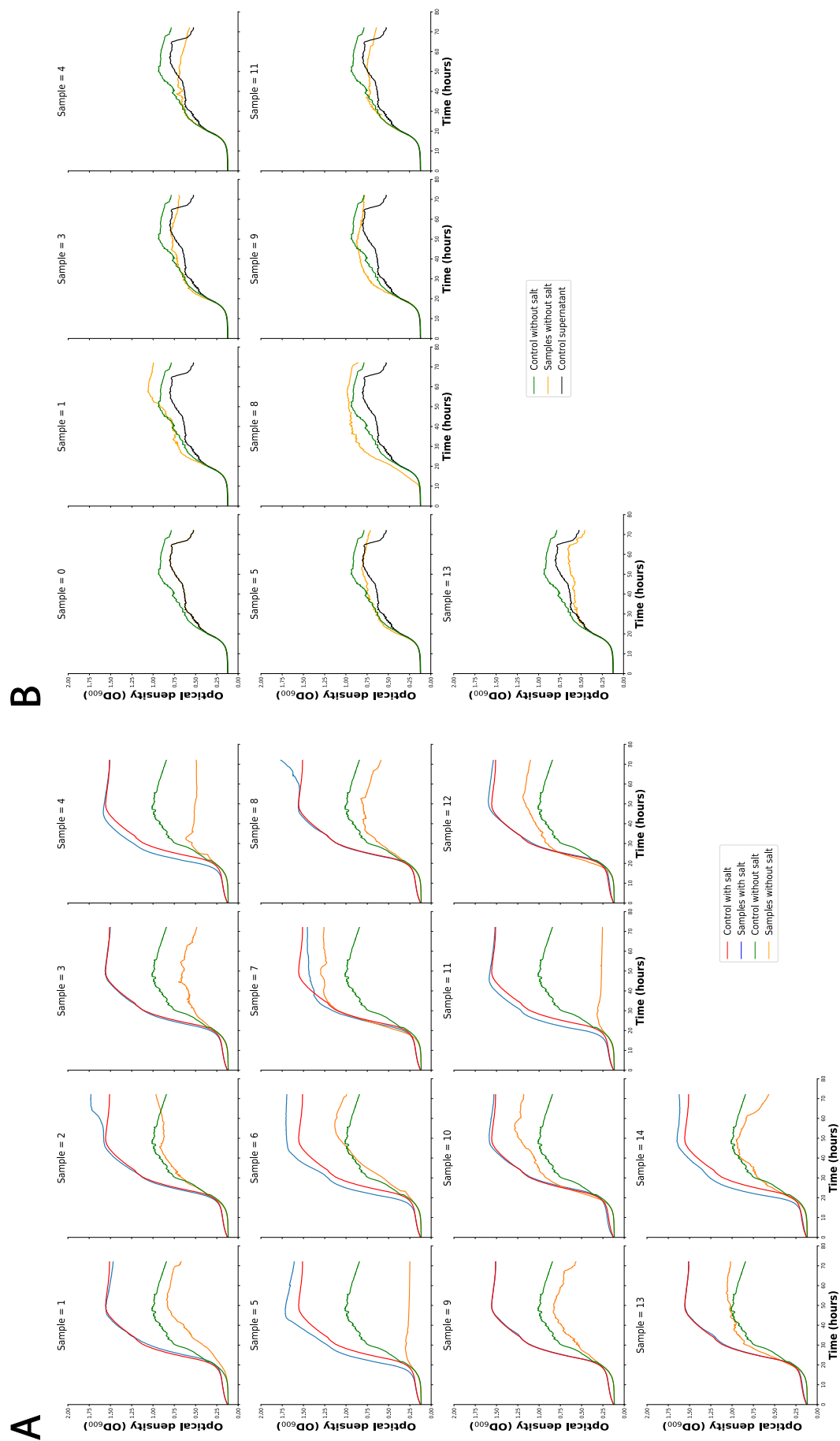
Figure 4.3. Growth of *L. pneumophila* NCTC 11192 in the presence of concentrated samples from environmental and man-made sources. This figure shows growth curves for *L. pneumophila* NCTC 11192 when co-cultured with different concentrated environmental and man-made samples. Each sub-plot shows the result of incubation with a different samples, while each line shows growth of NCTC 11192 with one concentrated sample either in the presence of divalent cation salts (blue line) or without salts (orange line), NCTC 11192 growth with or without the presence of salt (red and green lines respectively). (A) shows the result of the first round of co-culture, while (B) shows the growth of NCTC 11192 when the filtrate of samples from the first round of incubations where a reduction was noted, were re-incubated for a second round with fresh NCTC 11192. The additional black line indicates growth of NCTC 11192 when incubated with the supernatant from the control (no sample) from the initial round of incubations.

74
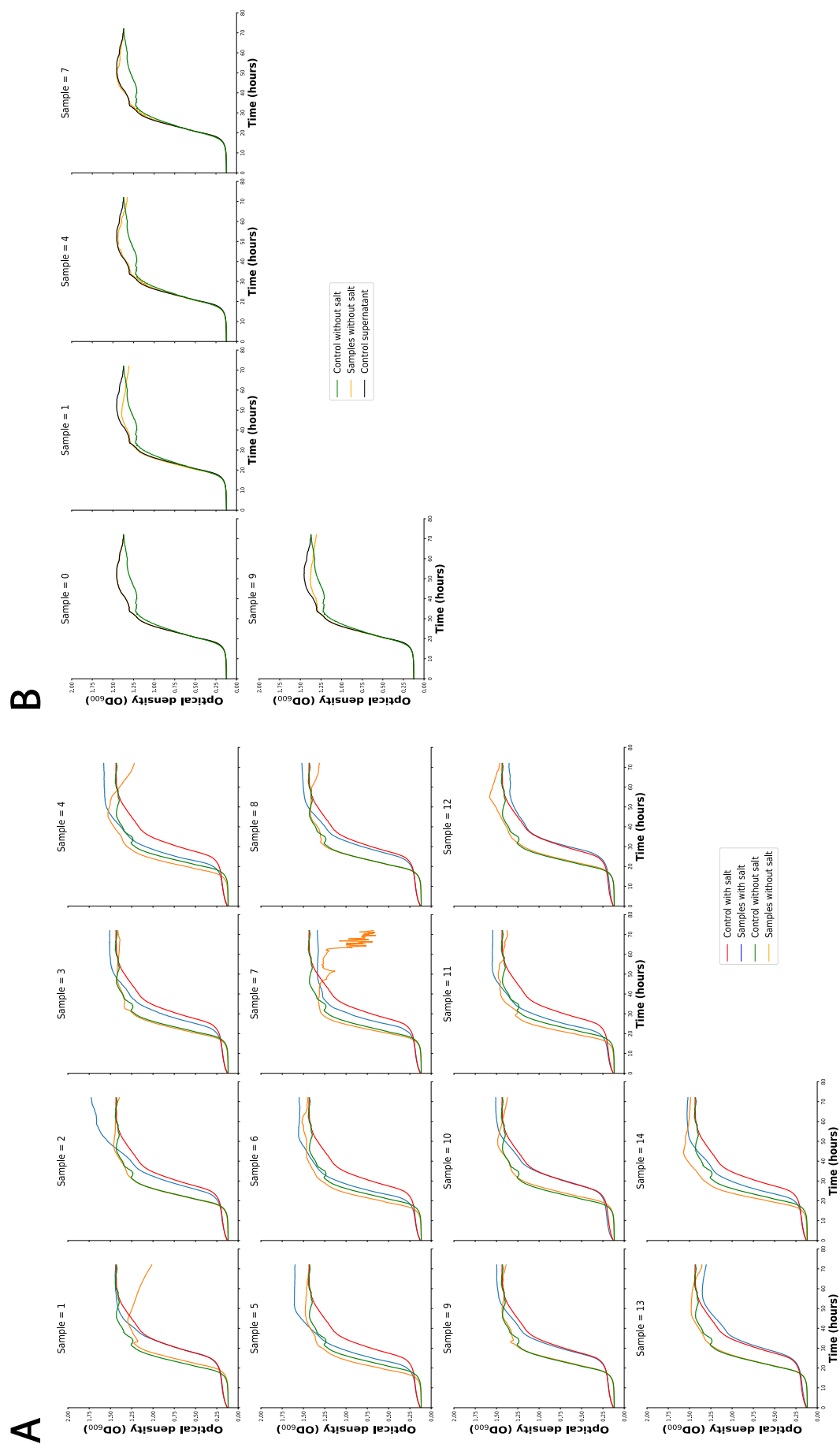
Figure 4.4. Growth of _L. pneumophila_ E1 in the presence of concentrated samples from environmental and man-made sources. This figure shows growth curves for _L. pneumophila_ E1 when co-cultured with different concentrated environmental and man-made samples. Each sub-plot shows the result of incubation with a different samples, while each line shows growth of E1 with one concentrated sample either in the presence of divalent cation salts (blue line) or without salts (orange line), E1 growth with no sample with or without the presence of salt (red and green lines respectively). (A) shows the result of the first round of co-culture, while (B) shows the growth of E1 when the filtrate of samples from the first round of incubations where a reduction was noted, were re-incubated for a second round with fresh E1. The additional black line indicates growth of E1 when incubated with the supernatant from the control (no sample) from the initial round of incubations.

75
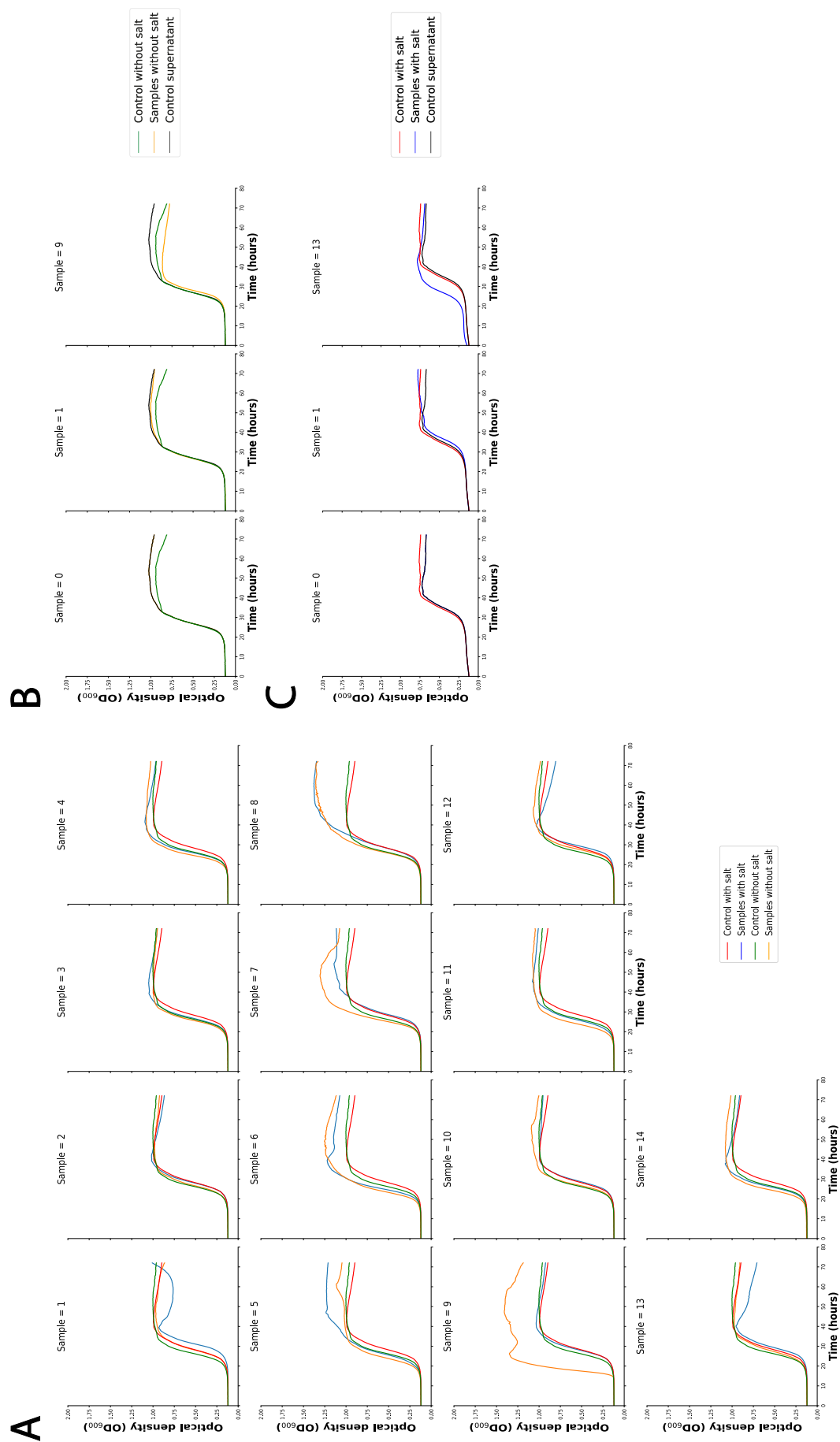
Figure 4.5. Growth of _L. pneumophila_ NCTC 11371 in the presence of concentrated samples from environmental and man-made sources. This figure shows growth curves for _L. pneumophila_ NCTC 11371 when co-cultured with different concentrated environmental and man-made samples. Each sub-plot shows the result of incubation with a different samples, while each line shows growth of NCTC 11371 with one concentrated sample either in the presence of divalent cation salts (blue line) or without salts (orange line), NCTC 11371 growth with or without the presence of salt (red and green lines respectively). (A) shows the result of the first round of co-culture, while (B) shows the growth of NCTC 11371 when the filtrate of samples from the first round of incubations where a reduction was noted, were re-incubated for a second round with fresh NCTC 11371. The additional black line indicates growth of NCTC 11371 when incubated with the supernatant from the control (no sample) from the initial round of incubations.

Uganda, East Africa, treated with chloroform; sample 13: environmental water sample, from Uganda, East Africa, with no treatment; and sample 14: sewage sample from London, UK.

The concentrated samples were then tested on 21 of the *Legionella* spp. isolates, covering one of each *L. pneumophila* sequence type, and one of each of the non-*pneumophila Legionella* spp. Although this assay was conducted on 21 isolates from the *Legionella* spp. collection, here only the result of *L. pneumophila* NCTC 11192, *L. pneumophila* E1, and *L. micdadei* NCTC 11371 are presented as an example of the results observed for *L. pneumophila* serogroup 1, *L. pneumophila* serogroup 2 - 14, and other *Legionella* spp. isolates, respectively. The complete results for the 21 isolates can be seen in supplementary figure S1 - S19. At least one sample led to reduced growth for all *legionella* spp. strains except *L. pneumophila* NL1 where no reduction was observed. For *L. pneumophila* NCTC 11192 for instance, reductions in growth were observed with samples 1, 3, 4, 5, 8, 9, 11, and 13 that were enriched in BYE-α supplemented with salts (figure 4.3 A). However, these reductions were not observed for any of the samples for this strain in conditions where the divalent cation salts were added. When the lysates of the samples that did show reductions were reused for an additional round to determine if this would occur again, which should be the case if viral replication was the cause of the reduction. However, similar growth to the controls were observed in every case where a second round of growth was examined. This would indicate then that the reductions may not be due to the action of replicating bacteriophages, with a possible alternative explanation for the initial but non-recurring reduction being an inhibitor that is being diluted with reuse in an additional passage. A similar outcome was observed with *L. pneumophila* E1 and *L. micdadei* NCTC 11371 (figure 4.4 and 4.5, respectively). For *L. pneumophila* E1 samples 1, 4, 7, and 9 produced reduced bacterial growth compared to their controls, in the absence of divalent cation salts (Figure 4.4 A). However, again no difference in growth between the control and the sample was observed for the a subsequent passage of the lysate (Figure 4.4 B). For *L. micdadei* NCTC 11371 samples 1 and 9 without salts and samples 1 and 13 in the presence of salts reductions in growth were observed (Figure 4.5 A). However, again a second passage of the lysate from the samples were reductions were observed showed no differences compared to their controls (Figure 4.5 B).

As all attempts from both plating and liquid culture had failed to lead to the isolation of bacteriophages infecting *Legionella* spp. an alternative approach was required. With a collection already assembled of bacterial isolates, and bacteriophages able to persist directly within bacterial cells as prophages, the isolation of integrated bacteriophages would be a good alternative isolation route to attempt.

4.4 Isolation of *Legionella* spp. bacteriophages via induction of lysogens

Prophages are widespread among bacteria, and in some cases can makeup about 20 % of the whole bacterial genome (Casjens, 2003). Regularly more than one prophage element can be found in a single bacterial genome. In *E. coli* O157:H7 EDL933 for instance, there are a total 18 bacteriophage-like regions, ranging from 7.5 kb to 61.6 kb in size (Perna et al., 2001). These prophages can often be induced from their integrated state, producing viable bacteriophages. In a study by Engelhardt et al. (2011), ∼ 46 % of bacteria isolated from seafloor sediments were able to produce bacteriophages following treatment with the inducing agent MitC. In *Legionella* spp. a study has previously claimed to have identified a prophage within the genome of one , *L. micdadei*. Gomez-Valero et al. (2014) believed they had identified a prophage in the genome of *L. micdadei* NCTC 11371 based on gene annotations following bioinformatic analysis. They identified a region encoding 73 proteins that was annotated as containing a few bacteriophage-associated proteins, while the majority of genes within the region encoded hypothetical proteins. As no success was achieved in isolating bacteriophages using either spot test and plaque assays, or the well assay, in the previous sections the focus of the current study turned to isolating temperate bacteriophages that may be within the genomes of the *Legionella* spp. in the strain collection; which included *L. micdadei* NCTC 11371.

4.4.1 Sensitivity of all *Legionella* spp. strains to induction agents using a disk diffusion assay

Given the evidence of the presence of prophage-like elements in the genome of *L. micdadei* (Gomez-Valero et al., 2014), it stands to reason that other *Legionella* spp. could similarly contain prophage elements. To examine this, the 44 *Legionella* spp. strains available within the collection from chapter 3 were to be examined for the presence of prophage. To begin, the sensitivity of all 44 available strains to two DNA-damaging agents; MitC and NFX, both known to facilitate prophage induction (Raya and Elvira, 2009; McDonald et al., 2010; Nale et al., 2016; Oliveira et al., 2017), was examined. In some bacteria the presence of a prophage increases the sensitivity of bacteria to inducing agents like MitC (Muschel and Schmoker, 1966). Therefore, strains showing a higher overall sensitivity to the inducing agents compared to other members of their species may be an indication of the presence of prophages in their genomes. To examine this, a blank disk was impregnated with 10 μl of either 10 mg ml^{-1} of MitC or NFX, and placed on an agar plate where the strain to be tested had been spread over the surface. Following incubation, the zone of inhibition was then measured to determine the sensitivity of each of the strains, with larger zones of inhibition meaning greater sensitivity. Figure 4.6 shows the

sizes of the zones of inhibition observed for all isolates.
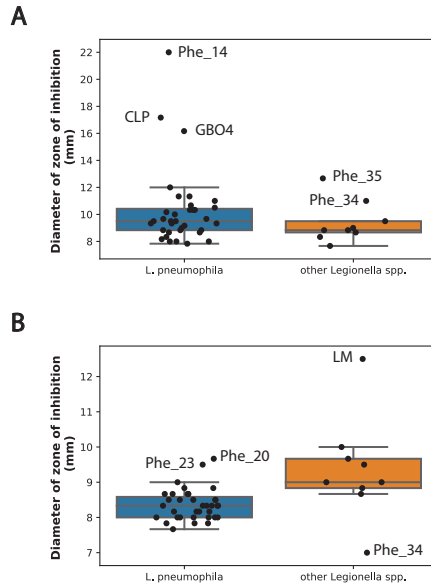
Figure 4.6. Distribution of the diameters of inhibition zones for *L. pneumophila* and other *Legionella* spp. in the presence of MitC and NFX in a disk diffusion assay. A) shows the result for MitC while B) shows the results for NFX Boxplots show the distribution of the diameter of zones of inhibition produced following exposure of each of the 44 *Legionella* spp. strains within the strain collection. Each box shows median (black horizontal line in box) and 25 % and 75 % quartiles. Whisker extend to 1.5 × the interquartile range. Outliers beyond the whiskers are represented as black circles labelled with the name of their associated strain. Blue boxes show the results for the 35 *L. pneumophila* isolates tested, and the orange boxes show the results for the 9 *Legionella* spp. isolates.

Exposure of *Legionella* spp. to MitC showed that the median zone of inhibition for *L. pneumophila* isolates was 9.5 mm, while for non-*pneumophila Legionella* spp., the median zone of inhibition observed was 8.83 mm for MitC (Figure 4.6 A). For NFX the median zone of inhibition observed was 9 mm for non-*pneumophila Legionella* spp., compared with 8.33 mm for the *L. pneumophila* isolates (Figure 4.6 B). There was no statistically significant difference between the sensitivity of *L. pneumophila* strains and non-*pneumophila Legionella* spp. isolates to MitC (two sample *t*-test: $t = 0.89$, $P = 0.37$). However, the sensitivity of *L. pneumophila* and non-*pneumophila Legionella* spp. to NFX were found to be significantly different, with non-*pneumophila Legionella* spp. being more sensitive to NFX than the *L. pneumophila* isolates (two sample *t*-test: $t = -5.29$, $P = 0.001$).

Among all the tested *Legionella* spp. isolates, *L. pneumophila* strains Phe_14, NCTC 11192 (CLP), and GBO4 showed higher sensitivities to MitC compared to the distribution of the other *L. pneumophila* isolated (Figure 4.6 A). For NFX, *L. pneumophila* strains Phe_20 and Phe_23 showed sensitivity beyond the distribution of the other *L. pneumophila* strains (Figure 4.6 B). As for non-*pneumophila Legionella* spp., *L. oakridgensis* Phe_35 and *L. hackeliae* Phe_34 showed higher sensitivity to MitC than the other non-*pneumophila Legionella* spp. (Figure 4.6 A). *L. micdadei* strain NCTC 11371 (LM;

the same strain where a putative prophage was identified by Gomez-Valero et al. (2014)) was the only non-*pneumophila Legionella* spp. found to produce a larger zone of inhibition with NFX 4.6 B). No outliers showed higher sensitivity to both MitC and NFX for either the *L. pneumophila* strains or non-*pneumophila Legionella* spp..

4.4.2 Growth of strains with increased sensitivity to prophage inducers in a broth micro-dilution method

To determine if the larger zones of inhibition noted for the 8 strains of *Legionella* spp. (1 *L. pneumophila* serogroup 1, 3 *L. pneumophila* serogroup 2 - 14, 1 *L. pneumophila* cross-reacted with the reagents of the *Legionella* latex kit, 1 *L. micdadei*, 1 *L. oakridgensis*, and 1 *L. anisa*) in the previous section were associated with the presence of induceable prophages, growth of these bacteria in a broth microdilution method was conducted. This would allow potential induction of prophages to be detected based on changes in the bacterial growth trends with increasing inducer concentration, an idea which has been shown previously to successfully allow identification of strains containing prophage in *Lactococcus lactis* (Oliveira et al., 2017). In this, an exponentially growing culture of each bacteria was exposed to a series of dilutions of either MitC or NFX and incubated for 52 hours with culture density examined every 15 minutes. Again, although 8 strains were used only the results for the *L. pneumophila* serogroup 1 isolate NCTC 11192, the *L. pneumophila* serogroup 2 - 14 strain Phe_20, and *L. micdadei* NCTC 11371 are shown as representative examples (figures 4.7, 4.8, and 4.9, respectively). The results for the other isolates tested can be found in Supplementary figures S20 - S24.

Across tested strains it was observed that, as expected, bacterial growth inhibition increased with increasing antibiotic concentration. For some of the cultures there appeared to be declines in bacterial density that may indicate the lysis of bacteria due to release of bacteriophages. For instance in *L. pneumophila* NCTC 11192 (CLP) population growth showed signs of density decline when exposed to 0.078 μg ml^{-1} MitC (Figure 4.7 A: J) and higher concentrations until around 0.625 μg ml^{-1} MitC where complete inhibition began to be observed. No similar reductions in density were observed for treatment of this strain with NFX (Figure 4.7 B). This is interesting as NCTC 11192 (CLP) showed a higher sensitivity to MitC, but not NFX during the disk diffusion assays. *L. pneumophila* Phe_20 showed an inhibition in growth when exposed to either MitC or NFX. For MitC reductions were observed for a number of concentrations of MitC beginning at $\sim$ 0.004 μg ml^{-1} MitC until $\sim$ 0.156 (Figure 4.8 A:F-K). For NFX, reductions were observed for 2 of the 19 concentrations tested, 0.078 and 0.156 μg ml^{-1} (Figure 4.8 B:J and K). Phe_20 showed a higher sensitivity to MitC, but not NFX during the disk diffusion assays. No noticeable declines in NCTC 11371 (LM) density were observed for *L. micdadei* NCTC 11371, which is interesting as this is the organism that Gomez-Valero et al. (2014) identified a potential prophage within.
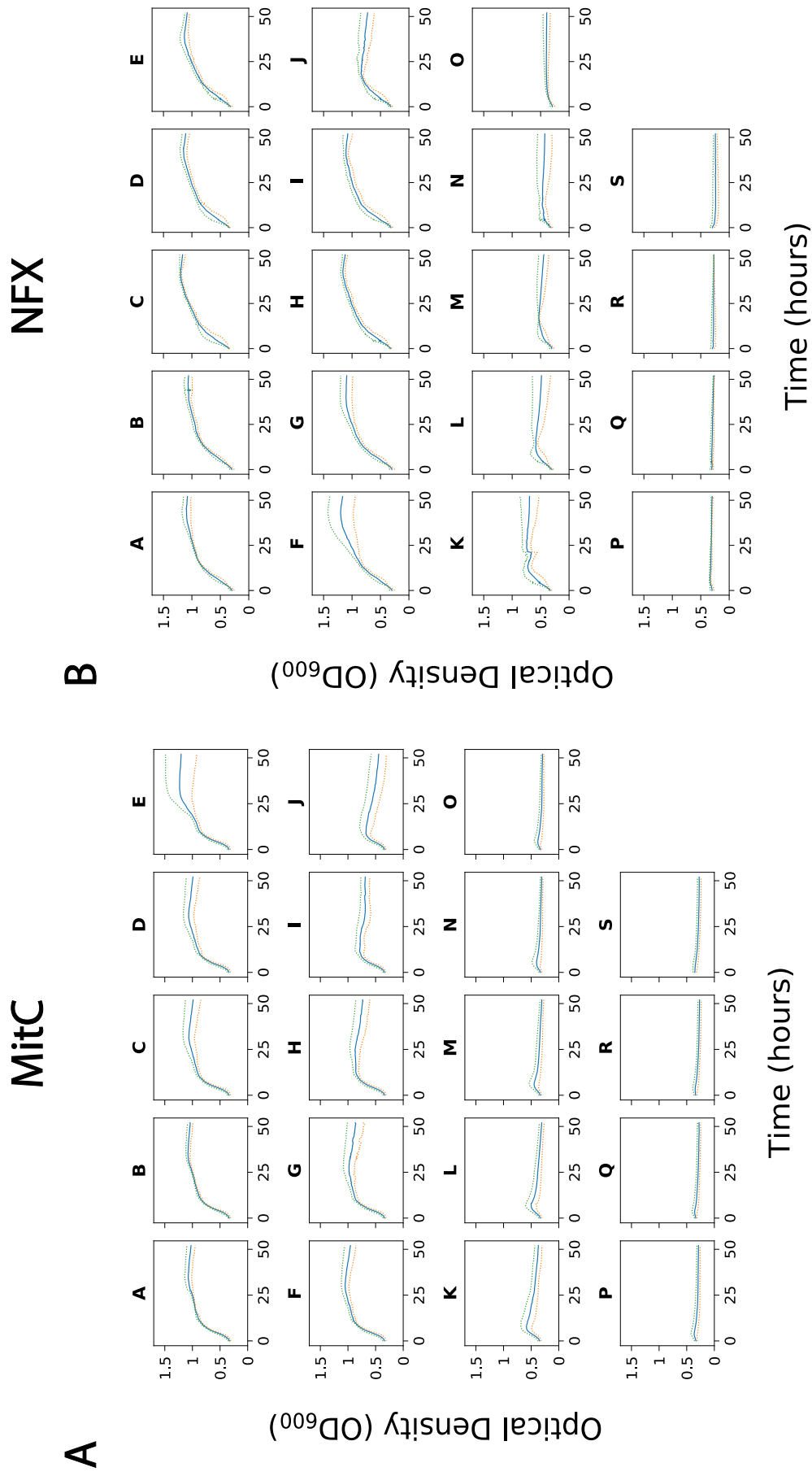
Figure 4.7. Growth curves of *L. pneumophila* NCTC 11192 (CLP) following treatment with the inducing agents MitC and NFX. The figure shows the growth profiles of *L. pneumophila* NCTC 11192 following treatment with a series of 2-fold serial dilutions of MitC ranging from 40 μg ml^{-1} to 0.00025 μg ml^{-1}. Each subplot shows the growth profile of *L. pneumophila* NCTC 11192 with one concentration of MitC: (A) untreated (0 μg ml^{-1}; (B) 0.00025 μg ml^{-1}; (C) 0.0005 μg ml^{-1}; (D) 0.001 μg ml^{-1}; (E) 0.002 μg ml^{-1}; (F) 0.004 μg ml^{-1}; (G) 0.009 μg ml^{-1}; (H) 0.019 μg ml^{-1}; (I) 0.039 μg ml^{-1}; (J) 0.078 μg ml^{-1}; (K) 0.156 μg ml^{-1}; (L) 0.312 μg ml^{-1}; (M) 0.625 μg ml^{-1}; (N) 1.25 μg ml^{-1}; (O) 2.5 μg ml^{-1}; (P) 5 μg ml^{-1}; (Q) 10 μg ml^{-1}; (R) 20 μg ml^{-1}; (S) 40 μg ml^{-1}. Blue line indicate mean bacterial density (n = 3). The green and orange dashed lines show growth $\pm$ sd.

81

Figure 4.8. Growth curves of *L. pneumophila* Phe_20 following treatment with the inducing agents MitC and NFX. The figure shows the growth profiles of *L. pneumophila* Phe_20 following treatment with a series of 2-fold serial dilutions of MitC ranging from 40 μg ml^{-1} to 0.00025 μg ml^{-1}. Each subplot shows the growth profile of *L. pneumophila* Phe_20 with one concentration of MitC: (A) untreated (0 μg ml^{-1}; (B) 0.00025 μg ml^{-1}; (C) 0.0005 μg ml^{-1}; (D) 0.001 μg ml^{-1}; (E) 0.002 μg ml^{-1}; (F) 0.004 μg ml^{-1}; (G) 0.009 μg ml^{-1}; (H) 0.019 μg ml^{-1}; (I) 0.039 μg ml^{-1}; (J) 0.078 μg ml^{-1}; (K) 0.156 μg ml^{-1}; (L) 0.312 μg ml^{-1}; (M) 0.625 μg ml^{-1}; (N) 1.25 μg ml^{-1}; (O) 2.5 μg ml^{-1}; (P) 5 μg ml^{-1}; (Q) 10 μg ml^{-1}; (R) 20 μg ml^{-1}; (S) 40 μg ml^{-1}. Blue line indicate mean bacterial density (n = 3). The green and orange dashed lines show growth $\pm$ sd.

Figure 4.9. Growth curves of _L. micdadei_ NCTC 11371 (LM) following treatment with the inducing agents MitC and NFX. The figure shows the growth profiles of _L. pneumophila_ NCTC 11371 (LM) following treatment with a series of 2-fold serial dilutions of MitC ranging from 40 μg ml^{-1} to 0.00025 μg ml^{-1}. Each subplot shows the growth profile of _L. pneumophila_ NCTC 11371 (LM) with one concentration of MitC: (A) untreated (0 μg ml^{-1}; (B) 0.00025 μg ml^{-1}; (C) 0.0005 μg ml^{-1}; (D) 0.001 μg ml^{-1}; (E) 0.002 μg ml^{-1}; (F) 0.004 μg ml^{-1}; (G) 0.009 μg ml^{-1}; (H) 0.019 μg ml^{-1}; (I) 0.039 μg ml^{-1}; (J) 0.078 μg ml^{-1}; (K) 0.156 μg ml^{-1}; (L) 0.312 μg ml^{-1}; (M) 0.625 μg ml^{-1}; (N) 1.25 μg ml^{-1}; (O) 2.5 μg ml^{-1}; (P) 5 μg ml^{-1}; (Q) 10 μg ml^{-1}; (R) 20 μg ml^{-1}; (S) 40 μg ml^{-1}. Blue line indicate mean bacterial density (n = 3). The green and orange dashed lines show growth $\pm$ sd.

4.4.3 Larger volume prophage induction and TEM analysis of temperate *Legionella* spp. virus-like particles

The growth reductions observed following treatment of bacteria with MitC or NFX found for some of the isolates in the well assay may indicate lytic induction of bacteriophages from the bacterial cells. To investigate this further, treatment with the inducers was repeated in larger volume cultures (200 ml) using the concentration of MitC or NFX observed in the previous assay to produce a notable reduction in bacterial density where applicable. Following incubation with the inducer, each cell lysate was then concentrated 100-fold by ultracentrifugation and examined using TEM to attempt detection of bacteriophage particles. These larger-scale inductions were conducted on all 8 isolates that showed higher sensitivity to the inducing agents in Section 4.4.1.
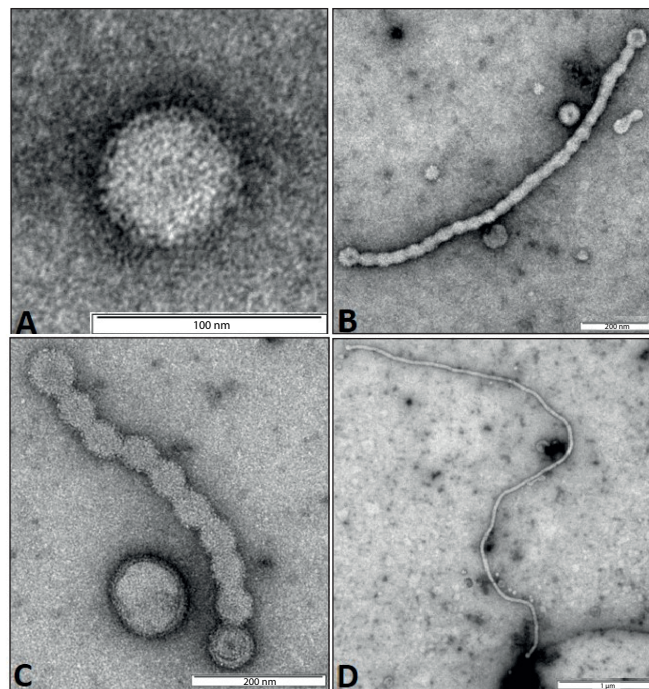
Figure 4.10. Electron micrographs of potential virus-like particles identified in the lysate of *L. pneumophila* GBO4. Representative transmission electron microscopy images of potential virus-like particles from induction of *L. pneumophila* GBO4. A) shows a virus-like particle with a spherical morphology. In B-D filamentous particles with regularly repeating bumps resembling beads on a string can be seen, but with 'strings' of very different lengths. Scale bars are indicated at the bottom right hand corner of each of the micrographs.

TEM analysis did not show the presence of any of the commonly observed *Caudovirales* in the induced lysates, nor the presence of any other currently described bacteriophages. However, in all strains examined other virus-like particles were observed. Two distinct morphologies for these virus-like particles were noted. The first morphology were spherical particles with a diameter of roughly ~ 58 - 70 nm (e.g. Figure 4.10). These observed particles looked to be enveloped, possibly showing spike

complexes throughout their surface. The second morphology were filamentous particles which seem to be made up of repeating units in a 'beads on a string'-like arrangement. Each of these particles had a width of roughly $\sim$ 50 - 70 nm with slight variations in the widths of each of the inner units, with the widest units those at the ends of the filaments (Figure 4.10 B-D). These filaments varied widely in length, for instance the particle in figure 4.10 B is $\sim$ 1.24 μm, 4.10 C is $\sim$ 615 nm, and 4.10 D is $\sim$ 5.79 μm. The units of the filamentous virus-like particles look very similar and are roughly consistent with the sizes of the individual spherical particles. However, as these are not similar to currently known bacteriophages, further analysis was required.

4.4.3.1 Relationship between DNA concentration in induced lysate and inducer concentration in *L. micdadei* NCTC 11371

After identifying viral-like particles under TEM, bacteriophage DNA extraction was then attempted for all of the lysates containing these particles. However, following loading the extracted DNA samples on an agarose gel, none of the strains showed any DNA bands. To confirm that DNA was not able to be extracted from these particles a range of concentrations of each of the induction agents was tested to determine if the concentrations of MitC and NFX being used was limiting induction resulting in absence of DNA bands. For this, cultures of *L. micdadei* NCTC 11371 were exposed to a range of concentrations of the induction agents (0 - 1000 ng ml^{-1}) using a broth dilution method. Viral DNA extraction was then performed on the lysates, and changes in viral DNA concentration measured using Qubit (see Section 2.8.5 for more detailed methods). Should an increase in the concentration of DNA detected be found following exposure to the DNA damaging agents, this could indicate the release of bacteriophage particles containing DNA and an appropriate MitC or NFX could be used to induce enough cells to allow DNA visualisation. This method would also have the additional benefit of allowing estimations of the quantity of viral particles released at each concentration as increased DNA should correlate with particle release.

It was found that across the concentrations tested no obvious changes in concentration of the extracted DNA occurred upon treatments with either MitC or NFX (Figure 4.11). Linear regression analysis showed that there was no linear relationship between the DNA concentration observed and the concentration of MitC or NFX that the isolate was incubated with (r^2 = 0.046 and 0.002, respectively). Upon comparison of the highest DNA concentrations obtained for each of the induction agents, 3.18 ng μl^{-1} and 4.44 ng μl^{-1} of MitC and NFX respectively, to the DNA concentration of the non-exposed sample, no significant differences were observed (Two-sample *t-test*, t = -0.309, p = 0.77 for MitC and t = -0.003, p = 0.99 for NFX). As no increases in DNA concentrations were observed between any inducer concentration and the non-exposed lysate this would indicate that either the virus-like particles are possibly not bacteriophage particles, or that these viruses may be RNA viruses instead. However, due
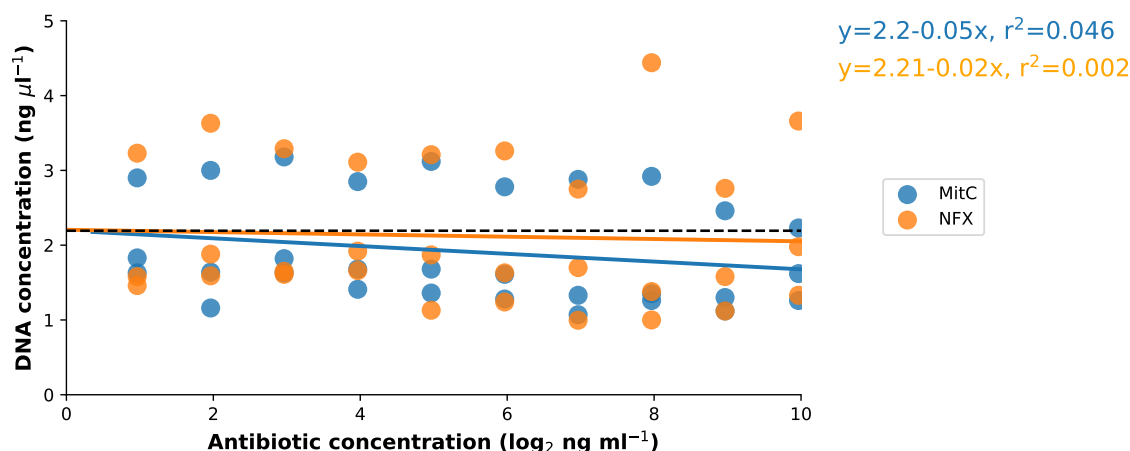
Figure 4.11. Quantification of lysate DNA concentrations following exposure of *L. micdadei* NCTC 11371 to the common bacteriophage inducing agents MitC and NFX. This figure shows the linear regression analysis of lysate DNA concentrations observed after exposure of *L. micdadei* NCTC 11371 to MitC (blue line and points) and NFX (orange line and points). Each point for both inducers indicate the mean lysate DNA concentration produced from 3 biological replicates for each inducer concentration. The black dashed line represents the mean lysate DNA concentration found for *L. micdadei* NCTC 11371 in the absence of inducers (N = 3).

to time constraints in this current work this latter point could not be investigated further.

4.4.4 Enumeration of Bacteriophage particles using Nanoparticle Tracking Analysis

As virus-like particles DNA extraction did not detect viral DNA, an alternative means of quantifying particles was attempted. Therefore, direct quantification of virus-like particles was attempted with a nanoparticle tracking analysis (NTA) -based approach using NanoSight (NS) technology. The NTA-based approach is based on utilising laser-illuminated optical microscopy for direct and real-time visualisation of nanoparticles in liquid suspensions. The nanoparticles are detected and counted in a few seconds or minutes as light-scattering centres moving under Brownian motion. This approach has previously been used successfully to estimate concentrations of both eukaryotic viruses and bacteriophages (Anderson et al., 2011; Kramberger et al., 2012). *L. micdadei* NCTC 11371 was chosen to be measured as it showed both the virus-like particles and is also suggested to harbour a putative complete prophage.

Samples were induced using UV light as detailed in Section 2.8.4. UV light is another commonly used means to induce prophages from bacterial genomes (Raya and Elvira, 2009). When tested on *L. micdadei* NCTC 11371 was qualitatively found to produce more of the spherical and filamentous viral-like particles (results not shown), although no DNA bands were still able to be observed following viral

DNA extraction. After UV exposure cultures were then aliquoted in bijou tubes containing broth and incubated at 37°C for 72 hours with NS used to measure particle concentration at regular intervals. The results of this can be found in figure 4.12.
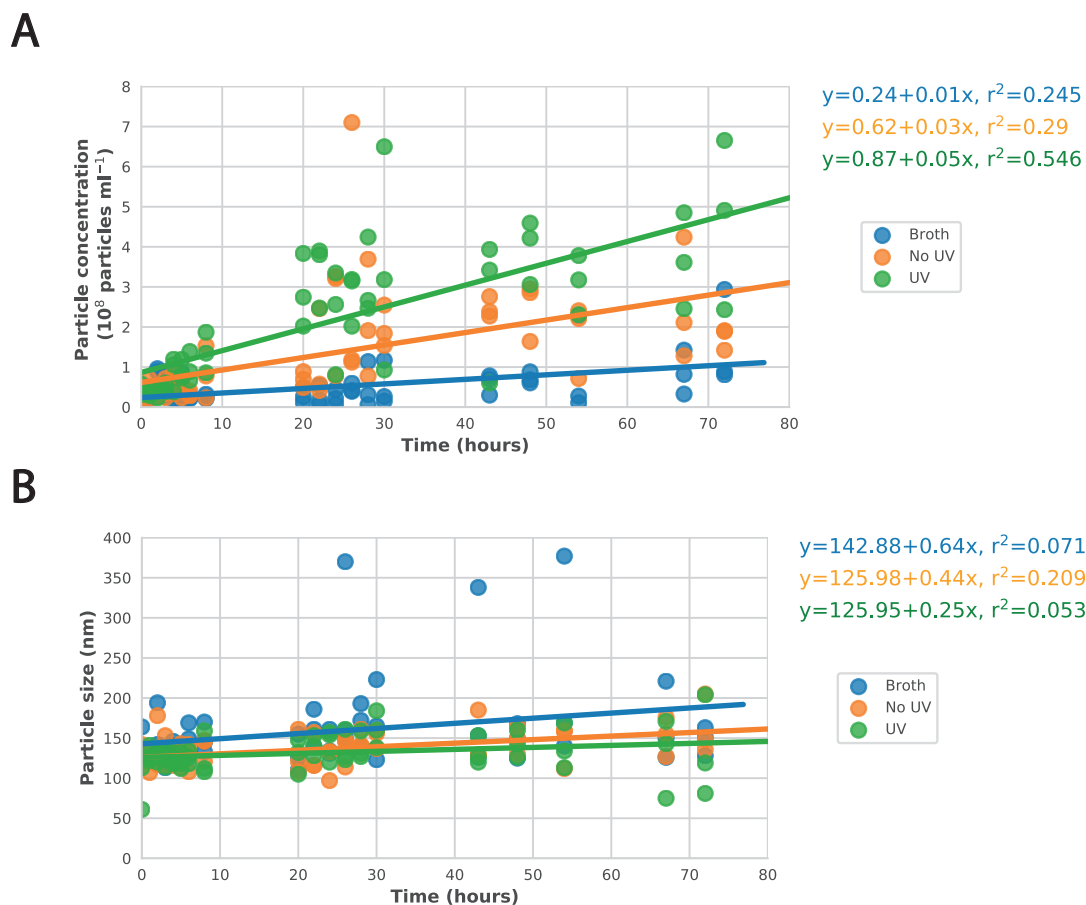
Figure 4.12.   Linear regression of particle concentration and size in the lysate of *L. micdadei* NCTC 11371 following exposure to UV induction. (A) shows estimated particle concentrations within the lysates using Nanosight measurement over time, while in (B) the mean size of these particles is shown. Each line represents the different test conditions; induced *L. micdadei* NCTC 11371 lysate (green line), *L. micdadei* NCTC 11371 without induction (orange line), and the liquid medium with no bacteria (blue line) (N = 3).

A greater positive correlation between number of particles and time was observed for cells that were exposed to UV, increasing from $\sim 1 \times 10^8$ particles ml^{-1} to $\sim 5 \times 10^8$ particles ml^{-1} (Figure 4.12 A; $r^2 = 0.546$). In the absence of UV, particle concentrations also increased over time to a lesser extent, from $\sim 1 \times 10^8$ particles ml^{-1} to $\sim 3 \times 10^8$ particles ml^{-1} (Figure 4.12 A). However, this was less predictable than the increases seen in the UV treated cells ($r^2 = 0.29$). BYE-α (Broth) remained roughly the same as the original time point, with an increase of $\sim 0.8 \times 10^8$ particles ml^{-1} between the first and last time points.

NS also allows estimation of particle size. However, particle size remained roughly consistent for each of

the bacterial conditions tested. Mean particle size was found to maintain around 130 - 150 nm for both, the non-treated and treated samples. Mean particle size was found not to be significantly different between any of the examined conditions and the broth only sample at the end of the incubation (two sample t-test, $P = > 0.05$; Figure 4.12 A).

4.5 Discussion

Despite the wide distribution and diversity of bacteriophages in the biosphere, some organisms still have no known associated bacteriophages. This is the case for *Legionella* spp.. The present study aimed to rectify this by isolating bacteriophages against *Legionella* spp., with the goal of characterising them for use in future development of bacteriophage therapeutics and diagnostics. To begin this process, the strains obtained in the previous chapter were used as hosts to screen the 108 samples from environmental and man-made sourced described in section 2.3.1 for bacteriophages able to infect *Legionella* spp. using sample enrichment. Sample enrichment is a commonly used methodology in the isolation of bacteriophages, and has been applied successfully to isolate bacteriophages from a range of sample sources including environmental samples such as sea and river water, soil, and even air samples, as well as host-associated samples, like saliva and faeces (Hitch et al., 2004; Li et al., 2010; Viazis et al., 2011; Yordpratum et al., 2011; Mattila et al., 2015; Magare et al., 2017). In their paper Lammertyn et al. (2008) used only a single host strain *L. pneumophila*, NCTC 11192, for their enrichments of 34 water samples from around Belgium. Using this method they isolated a total of 4 bacteriophages, with 2 of these bacteriophages later found to infect a wide range of *Legionella* spp. isolates. Therefore, it was believed that by using a wider strain collection (which included *L. pneumophila* NCTC 11192) and a larger number of samples, the current study would have a greater chance of isolating *Legionella* spp. bacteriophages.

This began promisingly, with reproducible zones of clearance observed for some but not all enriched samples following spot testing. However, when plaque assays were used to try and isolate single plaques, only one sample was found able to produce single plaques. This was also not reproducible, with no single plaques observed even after multiple attempts under the same conditions and when using other media. This is consistent with the observations of Lammertyn et al. (2008), as they observed from their screening that 4/34 enrichments where able to produce zones of clearance by spot testing, but only occasionally could single plaques be observed using plaque assays.

The reason for the inconsistency in plaque formation in the study by Lammertyn et al. (2008) and our current study is unclear. Plaque formation can occur spontaneously though in some bacterial strains. For instance, filamentous bacteriophage Pf of *P. aeruginosa* is capable of lytic superinfection of the Pf lysogen, leading to plaque formation (Webb et al., 2003). This is not constantly the case

though as lytic superinfection actually only occurs from a superinfection variant of Pf (SI Pf) that develops within some mature cells, with SI Pf development also being linked to mutations in the $mutS$ mismatch repair gene (Rice et al., 2009; Hui et al., 2014). Martínez and Campos-Gómez (2016) observed that *P. aeruginosa* PAO1 lysogens with defective $mutS$ genes were capable of producing spontaneous plaques when plated, while WT strains did not produce any spontaneous plaques. Therefore, plating of wild-type *P. aeruginosa* PAO1 could produce plaques infrequently, occurring when a spontaneous mutants of with defective $mutS$ is used to form the bacterial lawn. A similar mechanism could exist in *Legionella* spp., which would explain the infrequency of observed plaques.

Bacteriophages belonging to the *Caudovirales* were identified following TEM analysis of extracted single plaques, meaning it is possible that the plaques were due to the action of bacteriophages. To examine this further, bacterial growth in liquid media was examined, to attempt to observe common bacteriophage-associated lysis dynamics. For this, a total of 403 samples were concentrated into 14 sample groups by either tangential flow filtration using a 50 kDa filter unit or a 50 kDa Amicon filter. Concentrated samples were then incubated with a culture of the target host in a microtitre plate, with target bacterial density monitored in the wells every 15 minutes for 52 hours. It was expected that any bacteriophages in the samples would be concentrated allowing bacteriophage replication to more easily reach the 'inundation threshold', the concentration of bacteriophages required to begin causing reductions in bacterial population density (Cairns et al., 2009).

Some reductions in bacterial density were observed for the initial round of incubation. However, further passage of any sample that gave a reduction did not give any reduction in a second passage. This would not make sense for a reduction caused by bacteriophages, as bacterial lysis in the first cycle would indicate bacteriophage replication has occurred meaning increase bacteriophage titres that should then also cause reductions in a second transfer. The presence of inhibitors or bacteriocins in the sample may explain seeing only reductions in the first round but not the second, as the inhibitors/bacteriocins would have been diluted out in the second transfer. This would also explain why only zones of clearance were found during spot testing (Hockett and Baltrus, 2017). Bacteriocin-like substances found in water samples have been shown to inhibit bacterial growth. For example, Means and Olson (1981) found that bacteria isolated from potable water samples, such as *Flavobacterium* spp. and *Moraxella* spp. could produce bacteriocins that were able to inhibit the growth of Coliforms. Many other water-associated bacteria are known already to produce substances that can inhibit *Legionella* spp. growth. Corre et al. (2018) found that of 273 bacteria isolated from freshwater environments 178 of them were able to inhibit *Legionella* spp. growth, including by producing diffusible molecules and/or volatile organic compounds, which might explain the observations in the present study.

As the isolation of free bacteriophages within the samples had not been successful, isolation of bacteriophages directly from bacteria was then attempted. This was believed to be a promising way to

isolate bacteriophages against *Legionella* spp. as a putative prophage has previously been identified within the genome of *L. micdadei* NCTC 11371 (Gomez-Valero et al., 2014). Isolating bacteriophages directly from lysogenised target organisms can be extremely useful means of establishing a bacteriophage collection (Raya and Elvira, 2009). For example, Hargreaves et al. (2013) examined 27 strains of *Clostridium difficile* isolated from estuarine systems in England and were able to isolate bacteriophages from 20/27 (74 %) of these strains through induction, covering bacteriophages with distinct morphologies and genetic diversity. In the current study, the sensitivity of the 44 strains of *Legionella* spp. collected in the previous chapter to the common inducing agents MitC and NFX was first examined using a disk diffusion assay. Strains containing prophages can have a higher sensitivity to inducing agents (Muschel and Schmoker, 1966; Selva et al., 2009), and therefore if a strain presented a larger zone of inhibition around the disk then this may be a sign of prophage carriage. From this, eight strains were identified with greater sensitivity than their counterparts to either MitC or NFX.

Growth curves with serial dilutions of the inducing agents were then conducted for these strains to determine if the greater sensitivity also gave reductions in bacterial density. Prophage induction leads to a reduction in bacterial density due to lysis of the host cell, and is commonly seen following induction of bacteriophages in other organisms (Gillis and Mahillon, 2014), and reductions in density were found for some of the eight strains tested. However, to examine if these reductions were due to the action of induced prophages, inductions were repeated in larger volumes, and the resulting lysate then concentrated and examined by TEM. No particle resembling common *Caudovirales* morphologies were noted, but other potential virus-like particles with either a spherical or a 'beads on a string'-like morphology were observed. These particles appeared to increase in concentration over time in induced samples, as measured by NS, possibly indicating that they are bacteriophage from within the cells. Small spherical bacteriophages have been identified, such as the *Cystoviridae*, e.g. *Pseudomonas* bacteriophage phi6 (Vidaver et al., 1973; Poranen et al., 2017). *Cystoviridae* are dsRNA viruses (Poranen et al., 2017), and to date no integrating RNA bacteriophages have been identified, with these viruses being described as 'obligately lytic' (Hobbs and Abedon, 2016). However, there have been cases of a pseudolysogenic-like carrier state described for *Cystoviridae* (Cuppels et al., 1979; Onodera et al., 1992).

Romantschuk and Bamford (1981) for example, examined cultures of *Pseudomonas phaseolicola* (now known as *Pseudomonas syringae* pathovar *phaseolicola*) that had been exposed to phi6 in previous passages and noted that these cultures were still able to produce viral particles. They found that the level of viral release could be variable, but that virion release correlated with changes in bacterial growth, i.e. low levels of phi6 release would allow higher growth, while higher phi6 release would give lower growth rates. Romantschuk and Bamford (1981) also observed that passage of phage-carrier colonies could maintain the carrier state for multiple passages but, unlike traditional *Caudovirales* lysogeny, passage for multiple rounds could lead to loss of bacteriophages, meaning strains were no

longer carriers. Similar dynamics have also been observed for another family of RNA viruses the *Leviviridae* (Pourcel et al., 2017). Attempts to isolate DNA from the virus-like particles in the current study were unsuccessful, even when a range of induction conditions were tested. Therefore, it is possible that the particles observed in the current study may be novel RNA viruses. However, due to time constraints this was not able to be investigated further and remains to be explored.

The 'beads on a string'-like morphology of the other virus-like particles is not similar to any bacteriophages currently known, nor does it appear similar to any described cell-associated filamentous particles that have been described for *Legionella* spp.. Currently described filamentous bacteriophage families, such as the *Inoviridae*, or the archeal viruses belonging to the order *Ligamenvirales*, have thinner filaments. The widths of the particles in these families are $\sim$ 10 nm for *Inoviridae* and $\sim$ 25 - 30 nm for *Ligamenvirales* (Day, 2012; Prangishvili, 2012a,b), while those identified here have a diameter of $\sim$ 50 nm. The length of the virus-like particles identified here, $\sim$ 6 μm, are also larger than that observed for both the *Ligamenvirales* ($\sim$ 0.75 - 2.5 μm) and even the largest *Inoviridae* ($\sim$ 4 μm) (Day, 2012; Prangishvili, 2012a,b). The presence of filamentous bacteriophages within bacteria is believed to be severely underestimated, with only a small proportion of such viruses having been isolated. Roux et al. (2019) recently examined the prevalence of one filamentous virus family, the *Inoviridae*, in the genomes of 56,868 bacteria and archaea, and 6,412 metagenomes and found that roughly 6 % of bacterial and archaeal genomes while 35 % of metagenomes contained *Inoviridae* (Roux et al., 2019). This can vary widely among bacteria though. Castillo et al. (2018) identified bioinformatically that as many as 45 % of *Vibrio* spp. may contain *Inoviridae*-like prophages. Therefore, many other currently unrecognised families of filamentous bacteriophages could also exist.

Interestingly, the dimensions of the 'beads on a string'-like particles show similarity to the individual spherical particles, as each 'bead' present on the longer filament is similar to those of the individual spherical particles. It is possible then that the spherical virus-like particles could be an alternative morphotype of the longer filaments. This again is something that has not been observed for bacteriophages, but is found in eukaryotic viruses. For instance, *Influenza* viruses are most commonly observed as spherical virions, but are pleomorphic (Lamb and Choppin, 1983). They can exist with either spherical, bacilliform, or filamentous morphologies, with the filamentous form able to be over 50 μm in length (Dadonaite et al., 2016). These viruses, like the *Cystoviridae*, are RNA viruses and therefore, as mentioned further exploration of these particles as potential RNA viruses may prove more successful in the future.

To conclude, attempts to isolate bacteriophages in this current study were unsuccessful. However, induction of *Legionella* spp. strains did show possible evidence of pleomorphic RNA virus-like particles. However, due to the time constraints in the current work this could not be examined further. While physical isolation was not successful, bioinformatic screening of *Legionella* spp. for putative prophages

was also conducted in parallel to this work, and will be discussed in the next chapter.

5. Bioinformatic screening for putative prophage elements within the genomes of *Legionella* spp.

5.1 Abstract

Bacteriophages are present throughout nature, in environmental samples, samples from man-made systems, and in host-associated systems. They can exist either free in the environment, or within bacteria, where they replicate passively along with the host as 'prophages'. Prophages have been found in over half of all sequenced bacteria, where they can make up as much as 20 % of a bacteria's genome, often by the presence of more than one prophage in a single bacterial genome. However, despite their abundance throughout nature and within the genomes of other bacteria, no bacteriophages infecting *Legionella* spp. are currently available for study in the literature. Due to this, and as previous attempts to isolate bacteriophages infecting *Legionella* spp. from environmental samples and directly from bacterial strains experimentally in the previous chapters using different approaches had proved unsuccessful, the aim of this study was to obtain more information on *Legionella* spp. bacteriophages using a bioinformatic approach. To do this, the genomes of 717 *Legionella* spp. belonging to 73 species were screened for putative prophages using a prophage program, PHASTER (PHAge Search Tool Enhanced Release). From this analysis, 631 putative prophage-like elements were identified, with 11 potentially intact prophages. However, further examination of these 11 elements revealed that although these regions did encode *Microviridae* proteins, these were likely ϕX-174 contaminants that were present in the genomes due to poorly curated genomes being submitted to the public databases. Of the other 620 elements identified, only a small number contained genes with potential functions as bacteriophage structure-associated proteins, with these regions either only showing a single or very few other structure-associated genes, meaning that they may not be bacteriophages, or that they are elements that are so distinct from currently known bacteriophages that they cannot be confidently identified yet.

5.2 Introduction

Bacteriophages are obligate parasites of bacteria, with bacteriophage replication only possible following infection of a target bacterial cell (Kutter, 2001). Their replication occurs mainly through either of two cycles, the lytic and lysogenic cycles (Doss et al., 2017). In the lytic cycle, virus replication leads to the eventual lysis of the host cell and the release of new progeny virions (Weynberg, 2018). In the lysogenic cycle, bacteriophages remain within the cell following injection of their DNA, either by integrating into the bacterial chromosome or by existing as episomes within the cytoplasm. These bacteriophage elements are called prophages, and allow the virus to replicate passively during the normal bacterial DNA replication processes (Doss et al., 2017).

Prophages are commonly found in bacterial genomes. In a study by Touchon et al. (2016), 2110 complete bacterial genomes covering 1196 bacterial species were screened for the presence of prophages. They found that 46 % of these genomes contained at least 1 prophage element. One possible explanation of why $\sim$ 50 % of bacteria contain prophages is that these viruses act as a major sources of genetic diversity between bacterial strains, and can contribute new characteristics to their host bacteria, including fitness advantages; such as, antibiotic resistance and toxin production (Howard-Varona et al., 2017). Prophages are also able to re-enter the lytic cycle following induction and excision, which although leading to the death of the infected cell, can have benefits at the population level. For instance, prophage release can modify the structure and diversity of bacterial communities surrounding the lysogenised strains by lysing competitor bacterial strains also sensitive to the virus, eliminating them from the environment and liberating nutrients from within the cells, which could aid in the lysogenised population's survival (Duerkop et al., 2012).

With isolation of bacteriophages infecting *Legionella* spp. from both environmental samples and directly from bacterial isolates proving unsuccessful in the previous chapters, a bioinformatic approach to detect prophages of *Legionella* spp. within the bacterial genomes was next used to try and identify these viruses. Such studies have been used successfully previously to detect and genetically characterise bacteriophages against a range of bacterial genera; such as, *Helicobacter* spp., *Mycobactreium* spp., and *Streptococcus* spp. (Tang et al., 2013; Fan et al., 2014, 2016). This approach has also been used to identify potential prophages in the genomes of organisms where no bacteriophages have been isolated experimentally. For instance, no bacteriophages have been isolated for *Moraxella catarrhalis*, yet through the bioinformatic screening of 95 *M. catarrhalis* genomes Ariff et al. (2015) were able to identify 32 novel *M. catarrhalis* prophages likely belonging to the *Siphoviridae* family. Their attempts to isolate *M. catarrhalis* bacteriophages were unsuccessful, and thus with no success in isolating prophages in the current study, a bioinformatic process may elucidate potential prophage elements within *Legionella* spp. genomes. Gomez-Valero et al. (2014) have previously used a bioinformatic approach to identify a

putative prophage within one *Legionella* spp., *L. micdadei* NCTC 11371. They suggested the existence of a putative full length prophage within the genome of *L. micdadei* NCTC 11371 based on finding a region that encoded 73 proteins, from which 16 were bacteriophage-related proteins (e.g., bacteriophage capsid, tail, and replication proteins) (Gomez-Valero et al., 2014). Therefore, in this chapter the aim was to bioinformatically screen all available genome sequences for the *Legionella* spp. within the GenBank database for the presence of prophage elements, to finally identify bacteriophages infecting these organisms.

5.3 Bioinformatic identification of putative prophage-like elements across all *Legionella* spp..

With no bacteriophages able to be isolated by either the screening of samples, nor from the inductions of individual *Legionella* spp. isolates, an alternative approach to identify *Legionella* spp. bacteriophages was required. A bioinformatic approach was selected as there are numerous *Legionella* spp. genomes available, covering a range of species. At the time of this work, a total of 717 genomes from 73 *Legionella* spp. were available within the GenBank database. The majority of these genomes, 581 (81.03 %) belonged to *L. pneumophila*, but as many other species had at least one genome available. Characterisation of the complete set of 717 genomes would allow an increased chance of bacteriophage identification, and also allow inter-species comparisons of prophage diversity to be examined.

To analyse the genomes, all of the *Legionella* spp. genomes were examined using the prophage finding software PHASTER (Arndt et al., 2016). This software was the same used by Ariff et al. (2015) in their approach for detecting prophages in *M. catarrhalis*. PHASTER is a web-based tool that predicts prophage sequences within bacterial genomes and plasmids. It does this by comparing genes within a bacterial genome to a database of bacteriophage genes from known free bacteriophages and prophages, leading to predictions of prophage genes within a screened bacterial genome (Arndt et al., 2016). Each identified putative prophage region is then given a completeness score, based on the number of known bacteriophage genes/proteins identified within that genomic region. Putative prophages with a score > 90 are considered 'intact prophages', those with a score of 60 - 90 are 'questionable prophages', and those with scores < 60 are classified as 'incomplete'.

From the 717 *Legionella* spp. genomes, PHASTER identified a total of 631 putative prophage elements, with prophage identified within 40 of the 73 species (54.8 %) of *Legionella* (Supplementary table S9). Of the identified regions, PHASTER classified 67.29 % as incomplete elements (605/631), 1.66 % as questionable elements (15/631), and 1.22 % as potentially intact prophage (11/631) (Figure 5.1 A). Although the majority of the identified prophage-like elements were classified as incomplete, their overall median size, 13,185 bp, was still larger than the size of the smallest known dsDNA tailed
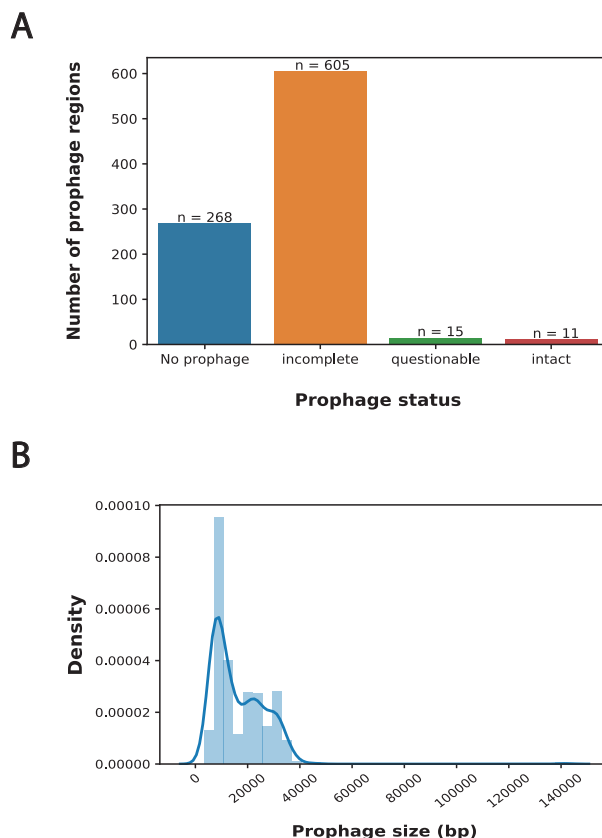
Figure 5.1. Number, classification and size distribution of putative prophages within *Legionella* spp.. Figure (A) shows the number of putative prophage identified by PHASTER within 717 *Legionella* spp. genomes. Coloured bars indicate the the number of these putative prophage elements based on their PHASTER classifications. These are designated as intact (red), questionable (green), incomplete (orange). The blue bar indicates the total number of *Legionella* spp. genomes where no prophage were identified. Figure (B) shows the distribution of lengths of the putative prophage identified within *Legionella* spp.

bacteriophage (*Mycoplasma* phage P1, $\sim$ 11.5 kbp). The putative prophage elements ranged in size from 3,313 bp (in *Legionellales bacterium* UBA6786) to 141,545 bp (in *L. pneumophila* HO92620872), (Figure 5.1 B). The GC contents of the identified elements were between 38.66 % and 44.75 %, with a median GC content of 43.15 %, which is slightly higher than the average GC content of *Legionella* spp. of $\sim$ 38.3%. Putative attachment sites were also identified for 353 of the 631 (54.22 %) prophage-like elements (Supplementary table S9).

The number of putative prophage elements per species of *Legionella* was variable (Figure 5.2). Up to 6 putative prophage elements were found in a single genome, with two genomes containing this number, *L. pneumophila* strain F-4198 and *Legionella tunisiensis* strain LegM. Although *L. pneumophila* F-4198 and *L. tunisiensis* LegM contained the highest number of identified putative prophage elements, all 6 of the elements found were classified as incomplete in *L. pneumophila* F-4198 genome, while 5
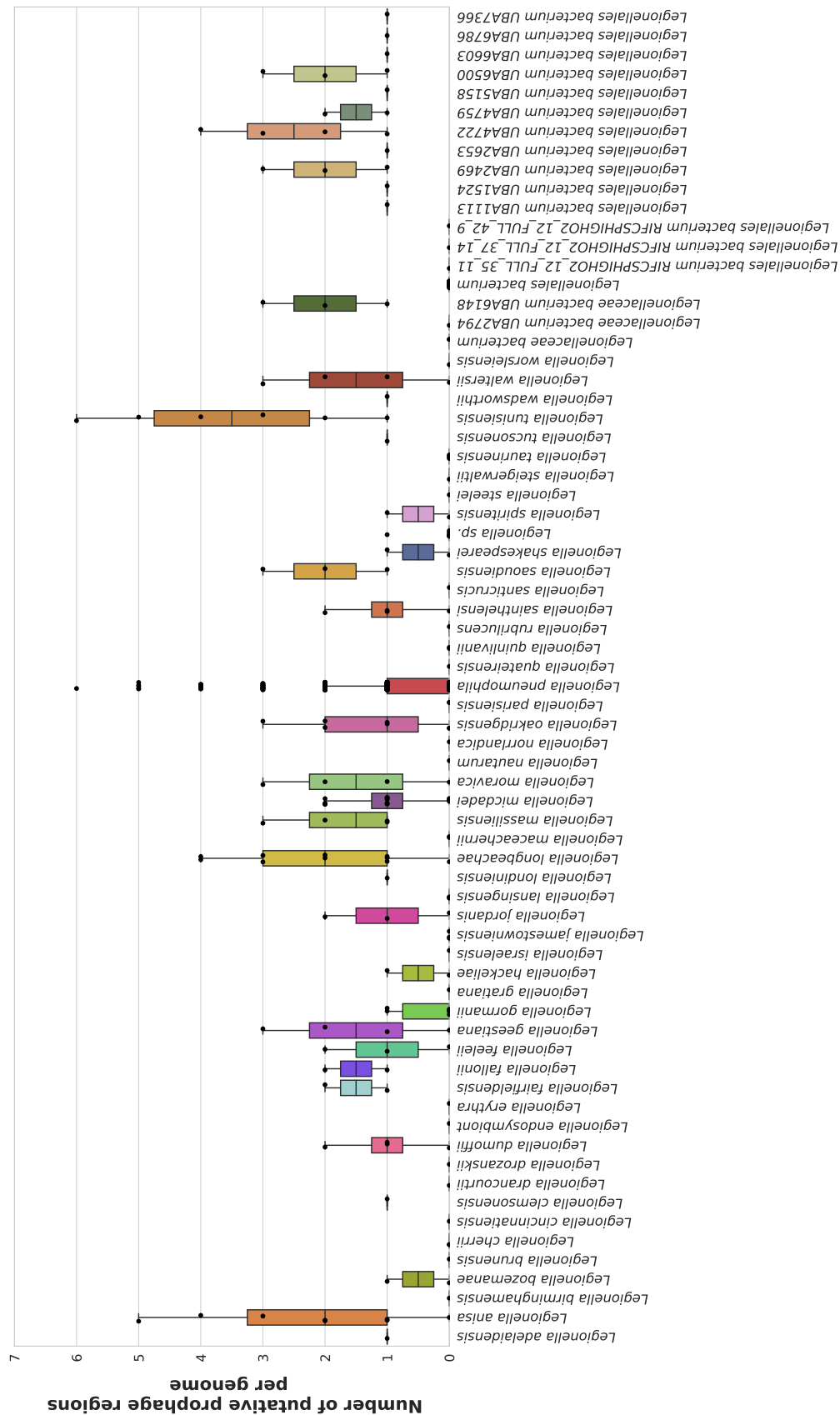
Figure 5.2. Distribution of numbers of putative prophage elements identified per strain among *Legionella* spp. with genomes in GenBank. Figure shows the distribution of the number of putative prophage elements identified per strain within each of the 717 genomes for all 73 *Legionella* spp. analysed in this study using PHASTER. The black horizontal line in each box indicates the median value, while the upper and lower regions of the box indicates the 25 % and 75 % quartiles, while whisker extending from the boxes correspond to 1.5 × the interquartile range. Each black circle indicates a single examined genome using PHASTER.

of the 6 in *L. tunisiensis* LegM were classified as incomplete and the other element was classified as questionable. The median number of putative prophage elements identified per genome was found to be one. No strains contained more than one element that was classified as complete.

5.4 Examination of 'intact' putative prophage regions identified by PHASTER

PHASTER identified a number of potential prophage regions, with the majority classified as 'incomplete' elements. However, PHASTER did identify 11 putative 'intact' prophage elements among the 717 genomes of *Legionella* spp.. Therefore, to confirm this 'intact' classification, manual examination of the gene content within these 11 regions was conducted. Manual examination would allow the determination of if these regions likely did encode full bacteriophage proteins, if they instead encoded partial bacteriophage that are missing key genes (i.e. are cryptic prophage), or if they were actually not bacteriophages, instead being other recombinational elements or other false positive detections.

To do this, the elements classified as complete were first re-annotated using prokka and the Prokaryotic Virus Orthologous Groups database 'pVOGs' (Grazziotin et al., 2017). The pVOGs database contains groups of orthologous genes identified from known bacterial and archeal virus genomes across multiple taxa. This database would act as a highly accurate database of bacteriophage genes, possibly allowing identification of bacteriophage genes within the prophage region that was not annotated as bacteriophages through comparison to the PHASTER database. The re-annotated regions were then manually examined for common bacteriophage genes responsible for structural-associated functions; such as, capsid, portal, and tail fibre proteins.

Table 5.1. *L. pneumophila* strains containing prophage elements identified as 'intact' by PHASTER

Strain	Length of region	Isolation source	Country	Reference
L. pneumophila **12_5223**	5,397	Patient	UK	McAdam et al. 2014
L. pneumophila **12_5251**	5,233	Patient	UK	McAdam et al. 2014
L. pneumophila **12_5329**	5,670	Patient	UK	McAdam et al. 2014
L. pneumophila **12_5383**	5,547	Patient	UK	McAdam et al. 2014
L. pneumophila **12_5415**	3,955	Patient	UK	McAdam et al. 2014
L. pneumophila **HO4020049**	5,946	Patient	UK	McAdam et al. 2014
L. pneumophila **HO80160261**	3,3761	Environmental	UK	McAdam et al. 2014
L. pneumophila **HO80160262**	28,916	Environmental	UK	McAdam et al. 2014

L. pneumophila **HO80160263**	8,077	Environmental	UK	McAdam et al. 2014
L. pneumophila **HO92620872**	141,546	Environmental	UK	McAdam et al. 2014
L. pneumophila **Twr292**	5,386	Ashiyu foot spa	Japan	Watanabe et al. 2015

Each of these 11 putatively intact prophages were found only in genomes of *L. pneumophila*, with a list of these genomes given in Table 5.1. Putatively, intact prophage regions ranged in size from between 3,955 bp to 141,545 bp in length, with a median length of 5,670 bp. Manual examination of these regions revealed that each of these putative prophage regions contained at least one annotation to a bacteriophage 'capsid' gene. In particular, a gene annotated as virus orthologous group (VOG) 0173 was identified in all 11 of the putative intact prophage. VOG0173 corresponds to gpD, an external scaffolding protein involved in capsid formation in members of the *Microviridae* family (Cherwa and Fane, 2012; Grazziotin et al., 2017). Another capsid-associated VOG, VOG0176, was found in 10/11 of the putative complete prophage, and corresponds to gpF which is another protein encoded within *Microviridae* that acts as the major capsid protein. Further examination of the genes within each of these regions showed that they contain a number of genes associated with the *Microviridae* family, including major and minor spike proteins, a DNA packaging protein, a DNA maturation protein, and a DNA replication initiation protein.

As each putative intact prophage region contains numerous genes associated with *Microviridae*, it is possible that these elements encode viruses within the *Microviridae* family. However, members of the *Microviridae* family have genomes between $\sim$ 4.4 - 6.1 kb in length (Cherwa and Fane, 2012), and while the median length of the identified regions was 5,670 bp, three of the putative prophage regions, HO80160263 (8,077 bp), HO80160262 (28,916 bp), and HO92620872 (141,546 bp) were larger than any known *Microviridae*. Re-examination of these three putative prophage regions showed that they likely did encode smaller viruses. For the complete region in HO92620872, positions 1,724 - 122,253 contain a number of genes (at irregular intervals) that encode various DNA binding/recombination genes but no virus structural genes, possibly indicating that these are just recombinant DNA elements rather than bacteriophages. Positions 126,790 - 132,044 though, contain all the *Microviridae* genes, meaning this 5,254 bp region is likely the real 'intact' prophage region. The same was also true for HO80160262 and HO80160263 with a 5,490 bp and 5,092 region encoding all of the *Microviridae*-associated VOGs in these genomes respectively.

In addition to this, by screening each of the 'questionable' and 'incomplete' elements for the VOG's found in the complete elements similar bacteriophage genes were also observed in *L. pneumophila* 12_5414. This was assigned as a questionable prophage element by PHASTER with a length of 28,108 bp, with the *Microviridae*-associated VOGs contained within a 6,563 bp region. No other *Microviridae*-associated VOGs were found in any of the other PHASTER predicted prophage elements.

5.4.1 Similarity between *Microviridae* identified in *L. pneumophila* and other known *Microviridae*

With potential *Microviridae* identified in some of the *L. pneumophila* genomes in GenBank, the similarity of these potential bacteriophage sequences to known *Microviridae* were then examined. To do this, a set of phylogenetic trees were constructed for the potential *L. pneumophila* prophages and other sequenced members of the *Microviridae* family. The *Microviridae* family consists of two subfamilies, the *Bullavirinae* and the *Gokushovirinae*, each containing three genera. The *Bullavirinae* contain the *Alphatreviruses*, e.g. *Escherichia* virus alpha3, the *Gequatroviruses*; such as, *Escherichia* virus G4, and the *Sinsheimerviruses*, which include *Escherichia* virus ϕX174. The *Gokushovirinae* contain the *Bdellomicroviruses*, including *Bdellovibrio* virus MH2K, the *Chlamydiamicroviruses*; such as, *Chlamydia* virus Chp1, and the *Spiromicroviruses*, e.g. *Spiroplasma* virus SpV4 (ICTV, 2018). Comparison of the identified putative prophages to other known, classified, *Microviridae* would then also allow prediction/identification of which genera within the *Microviridae* family the putative prophage belonged to. Therefore, for phylogenetic tree construction the sequences of the 7 bacteriophage structural genes examined were compared between the 12 putative *Microviridae* prophages within the *L. pneumophila* genomes and known *Microviridae* belonging to each genera; the accessions of the *Microviridae* genomes used in these comparisons can be found table 5.2.

Table 5.2. Table of known *Microviridae* genomes used to construct the phylogenetic trees.

Strain	Genus	NCBI accession number
Enterobacteria phage alpha3	*Alphatrevirus*	NC_001330.1
Bdellovibrio phage phiMH2K	*Bdellomicroviruses*	AF306496_1
Chlamydia virus Chp1	*Chlamydiamicroviruses*	D00624.1
Enterobacteria phage G4	*Gequatroviruses*	NC_001420.2
Spiroplasma virus SpV4	*Spiromicroviruses*	M17988.1
Enterobacteria phage ID1	*Sinsheimerviruses*	DQ079880.1
Enterobacteria phage MED1	*Sinsheimerviruses*	KJ997912.1
Enterobacteria phage NC51	*Sinsheimerviruses*	DQ079891.1
Coliphage ϕ-X174	*Sinsheimerviruses*	NC_001422.1
Enterobacteria phage WA11	*Sinsheimerviruses*	DQ079895.1

Phylogenetic analysis revealed that for each examined gene, all of the putative *L. pneumophila* prophage elements were highly similar to each other. In terms of their groupings to the example genera, they
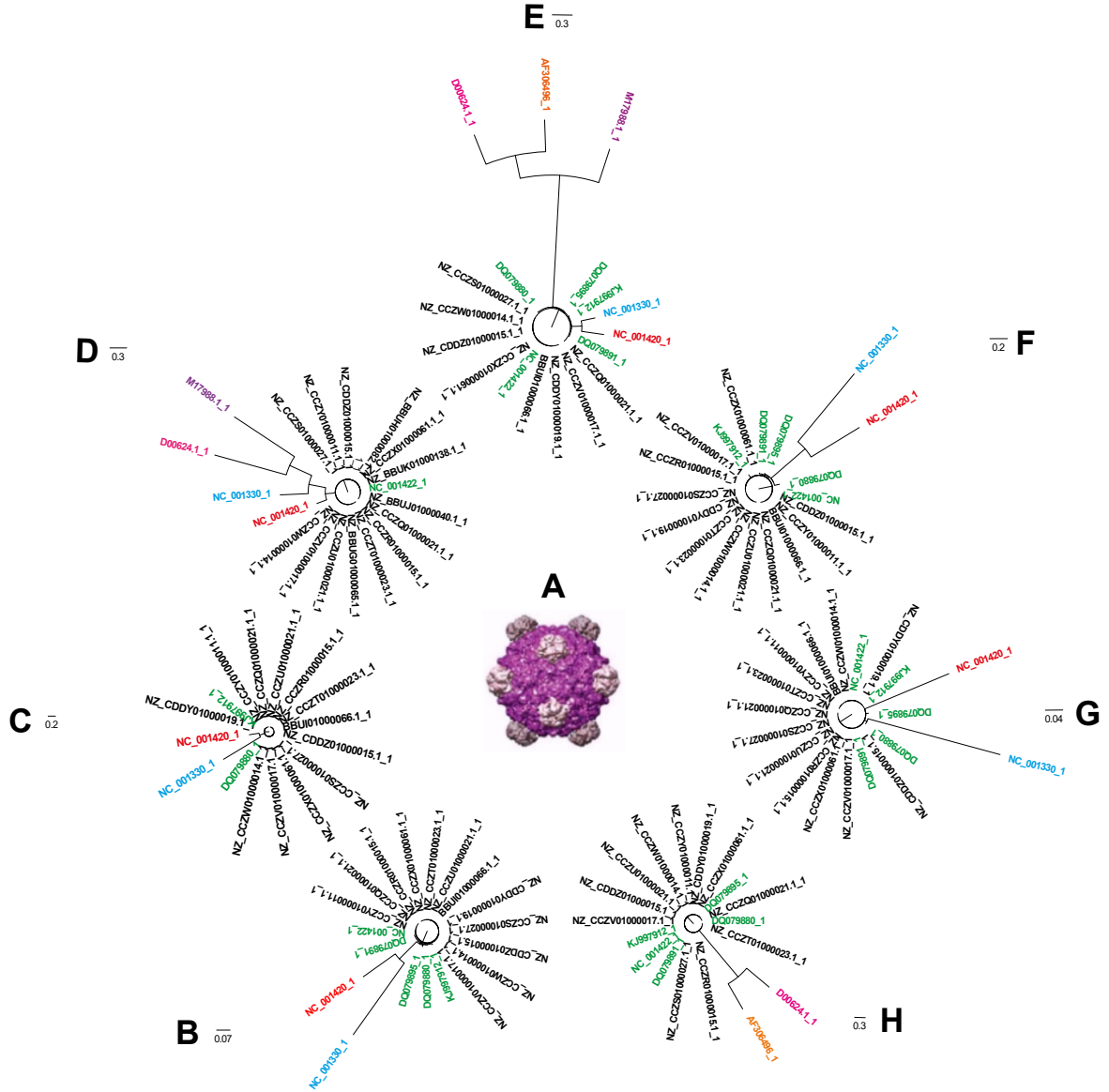
Figure 5.3. Phylogenetic trees showing similarity between putative *Microviridae* identified in *L. pneumophila* genomes from GenBank and other known *Microviridae*. Figure shows the phylogenetic relationships between the 12 *Microviridae* identified in *L. pneumophila* genomes from GenBank and other known *Microviridae*. (A) shows a graphical representation of a *Microviridae* particle. (B) shows a tree of similarities between minor spike proteins, (C) of DNA maturation proteins, (D) of DNA replication initiation proteins, (E) of major capsid proteins, (F) of major spike proteins, (G) of external scaffolding proteins, and (H) of DNA packaging proteins. The *Microviridae* identified in *L. pneumophila* genomes are shown in black, while other *Microviridae* genera are shown as follows: *Alphatrevirus* are in blue, *Bdellomicroviruses* are in orange, *Chlamydiamicrovirus* are in pink, *Gequatroviruses* are in red, *Spiromicroviruses* are in purple, and *Sinsheimerviruses* are in green.

all grouped more closely to the *Sinsheimerviruses*, i.e. *Escherichia* virus φ-X174 (Figure 5.3). Further comparison of gene organisations of the *Sinsheimervirus* φ-X174 to all 12 putative *L. pneumophila Microviridae* showed identical gene organisation (Figure 5.4). At the DNA sequence level, overall av-

erage nucleotide identify (ANI) across the entire prophage regions compared to ϕ-X174 revealed that all of the identified prophages had > 99.7 % ANI to ϕ-X174.

This would indicate then that the 12 prophages identified in the *L. pneumophila* are likely contamination of ϕ-X174 present within assemblies of the *L. pneumophila* genomes, not true *L. pneumophila* bacteriophages. To examine this possibility further, the genome sequence of the *L. pneumophila* strains found to contain complete prophages were re-downloaded from the JGI Integrated Microbial Genomes and Microbiome (IMG/M) public database instead of GenBank. The JGI IMG/M database identifies and removes ϕ-X174 contamination from sequence files following data submission, and therefore should remove contaminants from files. Re-examination of the 12 *L. pneumophila* genomes obtained from the JGI IMG/M using PHASTER this time, led to identification of no intact prophages. Therefore, the 'complete' prophage (and one 'questionable' prophage) initially identified from the GenBank genomes were not from true bacteriophages but ϕ-X174 contamination within the submitted assemblies.
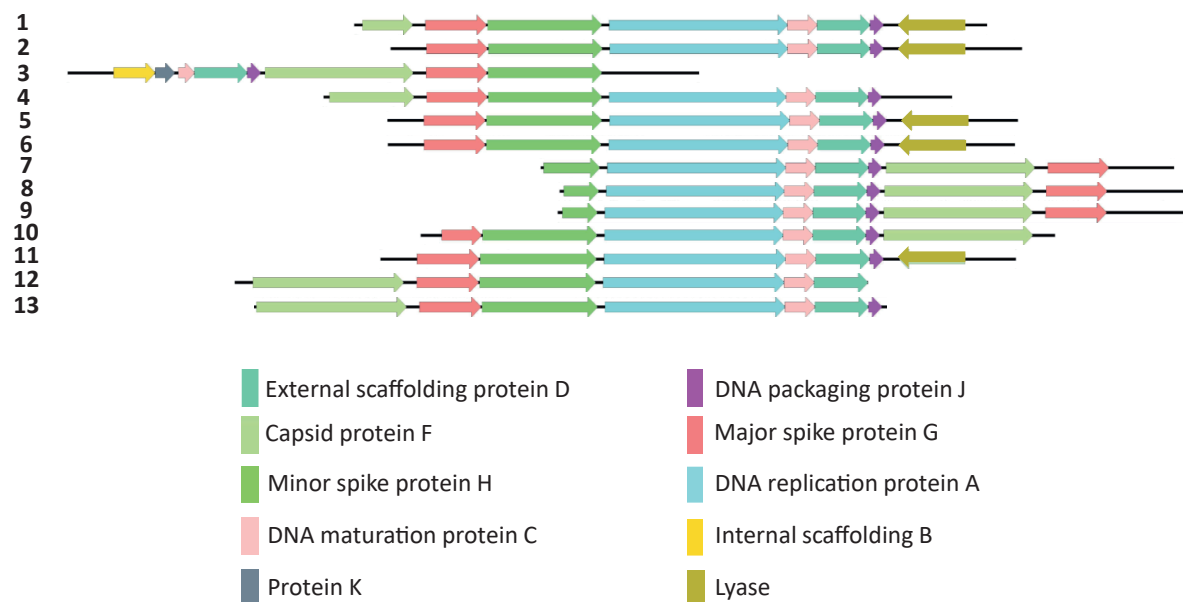
Figure 5.4. Genome organisation of the *Microviridae* identified in *L. pneumophila* genomes from GenBank. Figure shows the comparison of genome organisations for each of the *Microviridae* identified by PHASTER and ϕ-X174. Each line shows the organisation of one *Microviridae* from each of the following *L. pneumophila* genomes, (1) 12_5223, (2) 12_5251, (3) 12_5329, (4) 12_5383, (5) 12_5414, (6) 12_5415, (7) HO4020049, (8) HO801602161, (9) HO80160262, (10) HO80160263, (11) HO92620872, (12) Twr292, and the genome of the reference *Sinsheimervirus Escherichia* virus ϕ-X174 (13). Each arrow is coloured by gene function, with colour and function indicated under the alignments.

102

5.5 Examination of bacteriophage-related genes within 'questionable' and 'incomplete' elements detected by PHASTER

Since the *Microviridae* initially identified in the putative 'complete' prophages by PHASTER for the 717 *Legionella* spp. were all found to be ϕ-X174 contamination, both the questionable and incomplete putative prophage-like elements that were detected by PHASTER were then re-annotated using pVOGs and examined for the presence of bacteriophage-related structural genes. To do this, the pVOG annotated regions were screened for common bacteriophage genes; such as, capsid, tail, terminase, and integrase genes. Although some hits were identified, no complete set of these genes was observed in any single prophage regions. Capsid-associated genes were found in 12 prophage elements (1.90 %), plate-associated genes in 6 elements (1.11 %), tail-associated genes in 11 elements (1.74 %), portal-associated genes in 1 elements (0.16 %), terminase in 7 elements (1.11 %), while integrases were found in 372 of the elements (58.95 %).

The putative capsid-associated genes belonged to four VOGs; VOG0038, VOG0173, VOG0176, and VOG4573. VOGs VOG0173 and VOG0176, are two of the *Microviridae*-associated VOGs observed in the 'questionable' *L. pneumophila* 12_5414 element, and were thus ignored. VOG0038 corresponds to a group of proteins found in *Podoviridae* that may encode a protein of unknown function within the viral capsid. VOG4573 corresponds to a Clp protease which may be involved in the capsid maturation process in members of the *Caudovirales*. VOG0038 was found in 6 *Legionella* spp. genomes all of which were *L. pneumophila*, while VOG4573 was found in 5 genomes, 3 *L. pneumophila*, 1 *L. tunisiensis* and an uncharacterised *Legionellaceae*. No putative prophage regions identified by PHASTER contained more than one of these two capsid-associated gene clusters. Those prophage regions containing VOG0038 also contained all of the putative bacteriophage plate-associated genes that were found, each of which belonged to VOG0352 (Table 5.3).

Manual examination of the putative prophage regions containing hits to the capsid-associated genes revealed hits to other bacteriophage VOGs. For example, the 14,401 bp putative prophage region in *L. pneumophila* 2531STDY5467390 (classified as incomplete) which contained VOG0038 was found to contain six other VOGs, including VOGs corresponding to bacteriophage structural genes. This included VOG1086 which contains genes that may be associated with bacteriophage tail assembly, and VOG0352 which corresponds to possible plate proteins (Table 5.3). The regions containing VOG4573 also showed hits to other VOGs but were mainly to genes involved in DNA interactions/replication.

The putative tail-associated genes belonged to three VOGs; VOG1086, VOG4588, and VOG4783. VOG1086 corresponds to a set of hypothetical genes that possibly act as tail assembly-like protein, VOG4588 contains genes that encode a minor tail protein in *Siphoviridae* and *Myoviridae*, while

Table 5.3. Table of *Legionella* spp. with the most diverse bacteriophage structural genes (Ca = Capsid; Ta = Tail; Pl = Plate; Po = Portal; Te = Terminase; and In = Integrase) within *Legionella* spp.

Strain	GenBank No.	Size	pVOGs					
			Ca	Ta	Pl	Po	Te	In
L. pneumophila 2531STDY5467390	GCA_900052095	14,401 bp	1	1	1	0	0	0
L. pneumophila 2531STDY5467369	GCA_900052275	14,401 bp	1	1	1	0	0	0
L. pneumophila 2531STDY5467395	GCA_900058565	14,401 bp	1	1	1	0	0	0
L. pneumophila 2531STDY5467396	GCA_900058575	14,401 bp	1	1	1	0	0	0
L. pneumophila 2531STDY5467366	GCA_900060715	14,401 bp	1	1	1	0	0	0
L. pneumophila 2531STDY5467374	GCA_900061585	14,401 bp	1	1	1	0	0	0
L. pneumophila 2531STDY5467287	GCA_900053365	6,952 bp	1	0	0	0	0	0
L. pneumophila 2531STDY5467356	GCA_900057245	6,952 bp	1	0	0	0	0	0
L. pneumophila 2531STDY5467345	GCA_900062505	6,952 bp	1	0	0	0	0	0
Legionellaceae bacterium UBA6148	GCA_002423125	8,148 bp	1	0	0	0	0	0
L. tunisiensis LegM	GCA_000308315	7,675 bp	1	0	0	0	0	0
L. pneumophila TUM 13948	GCA_000695015	23,194 bp	0	1	0	0	0	4
L. pneumophila ATCC 33823	GCA_001582305	26,723 bp	0	1	0	0	0	2
L. pneumophila NY29	GCA_001601155	21,362 bp	0	1	0	0	0	1
L. spiritensis NCTC11990	GCA_900186965	15,550 bp	0	1	0	0	0	3
Legionellales bacterium UBA2469	GCA_002342175	31,258 bp	0	1	0	0	0	0
L. micdadei ATCC 33218	GCA_001648625	9,793 bp	0	0	0	1	2	1
Legionellales bacterium UBA6500	GCA_002436445	28,685 bp	0	0	0	0	2	2
L.fairfieldensis ATCC 49588	GCA_000621525	22,279 bp	0	0	0	0	1	1

Table **5.3.** – Continued from overleaf

Strain	GenBank No.	Size	pVOGs					
			Ca	Ta	Pl	Po	Te	In
L. micdadei LMI	GCA_000953635	9,793 bp	0	0	0	1	2	1
L. pneumophila Lorraine	GCA_000306865	16,130 bp	0	0	0	0	1	1

VOG4783 encodes a series of hypothetical proteins that may be tail accessory fibers in *Siphoviridae*. Again no element contained more than one of these tail genes. VOG1086 was found in 6 *Legionella* spp genomes, all of which were *L. pneumophila*. In fact all the isolates containing VOG1086 were also those six *L. pneumophila* that contained the capsid-associated VOG0038. VOG4588 was found in three *L. pneumophila* putative prophage regions and one *L. spiritensis* region. Most of the putative prophage elements containing this VOG did not contain any other potentially structural-associated genes, but did contain numerous VOGs with potential roles in DNA recombination, including multiple annotated as being potential integrase genes (Table 5.3). Only one putative prophage region contained VOG4783, that being a region in *Legionellales* bacterium UBA2469.

The sole portal-associated gene (VOG0254) was found in *L. micdadei* strain ATCC 33218, the strain of *L. micdadei* Gomez-Valero et al. (2014) believed they had identified a prophage in. *L. micdadei* strain ATCC 33218 also contained VOGs corresponding to an integrase gene (VOG0221), and 2 terminase subunits (VOG0612 and VOG4544). The terminase small subunit ortholog group VOG0612 was found in three other putative prophage regions other than *L. micdadei* strain ATCC 33218, another *L. micdadei* strain, a *L. fairfieldensis* genome, and *Legionellales* bacterium UBA6500. The two *L. micdadei* strains, ATCC 33218 and LMI were also the strains that contained a hit to VOG4544, which contains correlates to a group of genes encoding a terminase large subunit. The other VOG found in the putative prophage regions that was associated with terminases was VOG1517 which encodes a putative terminase small subunit. This was found in *L. pneumophila* Lorraine alongside an integrase gene.

This was the full extent of common key bacteriophage genes identified within the prophage elements predicted by PHASTER. While structural-associated genes were identified in only a handful of prophage predictions, the remaining regions encoded a range of other VOG genes that are less clearly bacteriophages in origin. These include genes encoding products such as glycosyltransferases, methyltransferases, and restriction endonucleases that, although have been found within bacteriophages, are more bacterial-associated. This could possibly indicated that PHASTER has detected a number of other transposable elements.

5.6   Investigation of the number of integrons within *Legionella* spp.

Apart from the described structural genes, a large number of integrases were also identified within the screened genomes of *Legionella* spp. using PHASTER. Further examination of the putative prophage elements containing these integrases revealed that, while also containing a number of other VOGs, most did not contain any bacteriophage structural genes. Instead genes involved in other more bacteria-associated functions were present in these elements. One possibility for this could be that these

elements are not prophage elements but are instead other mobile genetic elements. One common integrase containing mobile element found in bacterial genomes are integrons (Gillings, 2014), and are found in as many as $\sim 10\%$ of sequenced genomes (Boucher et al., 2007). No study to date has examined the presence of integrons within *Legionella* spp., and therefore, to investigate if these elements do exist in the genomes of *Legionella* spp., and whether the integrase-containing regions identified by PHASTER may be integrons, the 717 *Legionella* spp. genomes in GenBank were screened for presence of integrons within them using an integron identification tool, IntegronFinder (Cury et al., 2016).
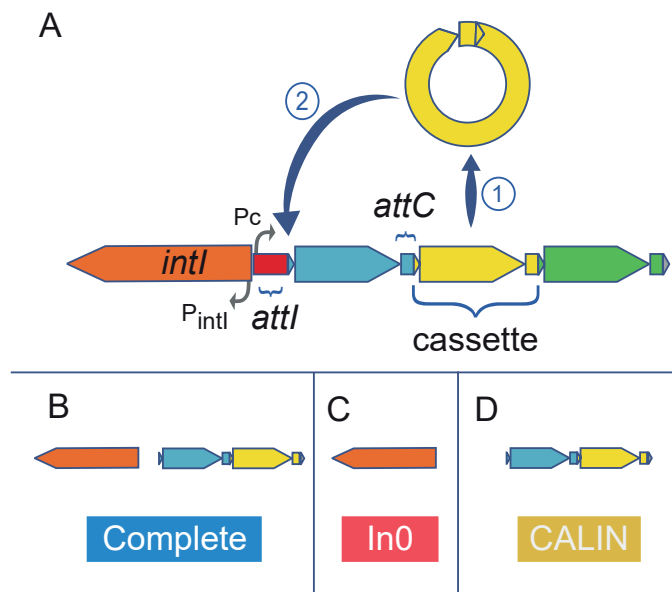
Figure 5.5. Schematic representation of the three categories of integrons identified by IntegronFinder. (A) shows the elements that compose an integron, a specific integron integrase gene (*intI*, orange), a promoter (*PintI*), an *attI* recombination site (red), and an array of gene cassettes (blue, yellow and green) including a promoter (*Pc*) for the cluster of cassettes. (B-D) shows the three classification of integron that can be identified using integron finder, (B) are complete integrons, which include an integrase and at least one attC site, (C) are In0 elements that are composed of an integron integrase but no attC sites, and (D) are the clusters of attC sites lacking integron-integrases (CALIN) elements, and are composed of at least two attC sites. Figure was adapted from (Cury et al., 2016)

IntegronFinder is a program for the identification of integrons, genetic elements containing a site-specific recombination system to integrate and transmit genes (Hall and Collis, 1995). This program identifies integrons into three categories of elements; (i) the "complete integrons", which contain *intI* and at least one *attC* site (Figure 5.5 B), (ii) The "In0" elements, which includes only an *intI* with no *attC* sites (Figure 5.5 C), and (iii) the "Cluster of *attC* site Lacking Integron-integrase (CALIN)" elements, which include at least two *attC* sites with no *intI* (Figure 5.5 D).

IntegronFinder identified a total of 583 integron elements. However, no complete integrons were identified, nor any In0 elements. Instead, each of the integron elements identified by IntegronFinder

were from occurrences of isolated *attC* sites without nearby integrases, i.e. only CALIN elements were found (Figure 5.6). Lengths of the CALIN integrons identified varied among *Legionella* spp. with distances between *attC* sites also varying in length from 41 bp to 6,113 bp, with a median size of 79 bp. The largest number of these elements were found in *L. pneumophila* (50.6 %), which also contained the largest number of genomes screened. Although the majority of CALIN integrons were in *L. pneumophila*, the largest number of CALIN regions in a single genome were identified in some of the uncharacterised *Legionellales* species, with *Legionellales bacterium* UBA7366 containing 19 elements, and *Legionellales bacterium* UBA1524 containing 18 elements. In contrast, where integrons were found other *Legionella* spp.; such as, *L. londiniensis*, *L. rubrilucens*, and *L. saoudiensis* among others showed only one integron element in their genomes. No integron elements were identified in 27/73 species of *Legionella*; such as, for *L. gormanii*, *L. birminghamensis*, and *L. cherrii*. The complete diversity in number of CALIN elements identified in the *Legionella* spp. genomes can be found in figure 5.7 and Supplementary table S10.
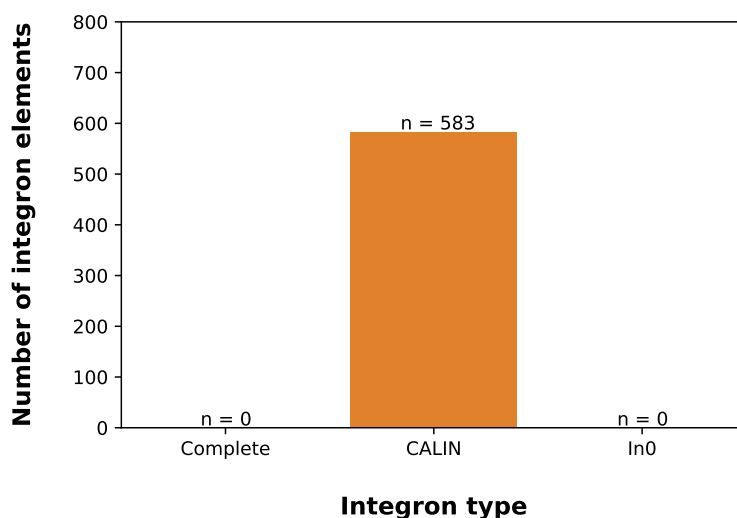
Figure 5.6. The number of integron elements identified within all *Legionella* spp. genomes in GenBank using IntegronFinder. Figure shows the number of the three types of integron elements detected by IntegronFinder within *Legionella* spp., based on presence of an integrase gene (*intI*), an *attI* recombination site, and *attC* sites.

5.7 Discussion

Attempts in the current work to isolate bacteriophage infection *Legionella* spp. was unsuccessful; even after screening large numbers of samples from environmental and man-made sources known to contain *Legionella* spp.. This was also despite attempts to induce viruses from inside a range of *Legionella* spp. isolates. This is strange as bacteriophages can frequently inhabit bacterial cells, either as elements
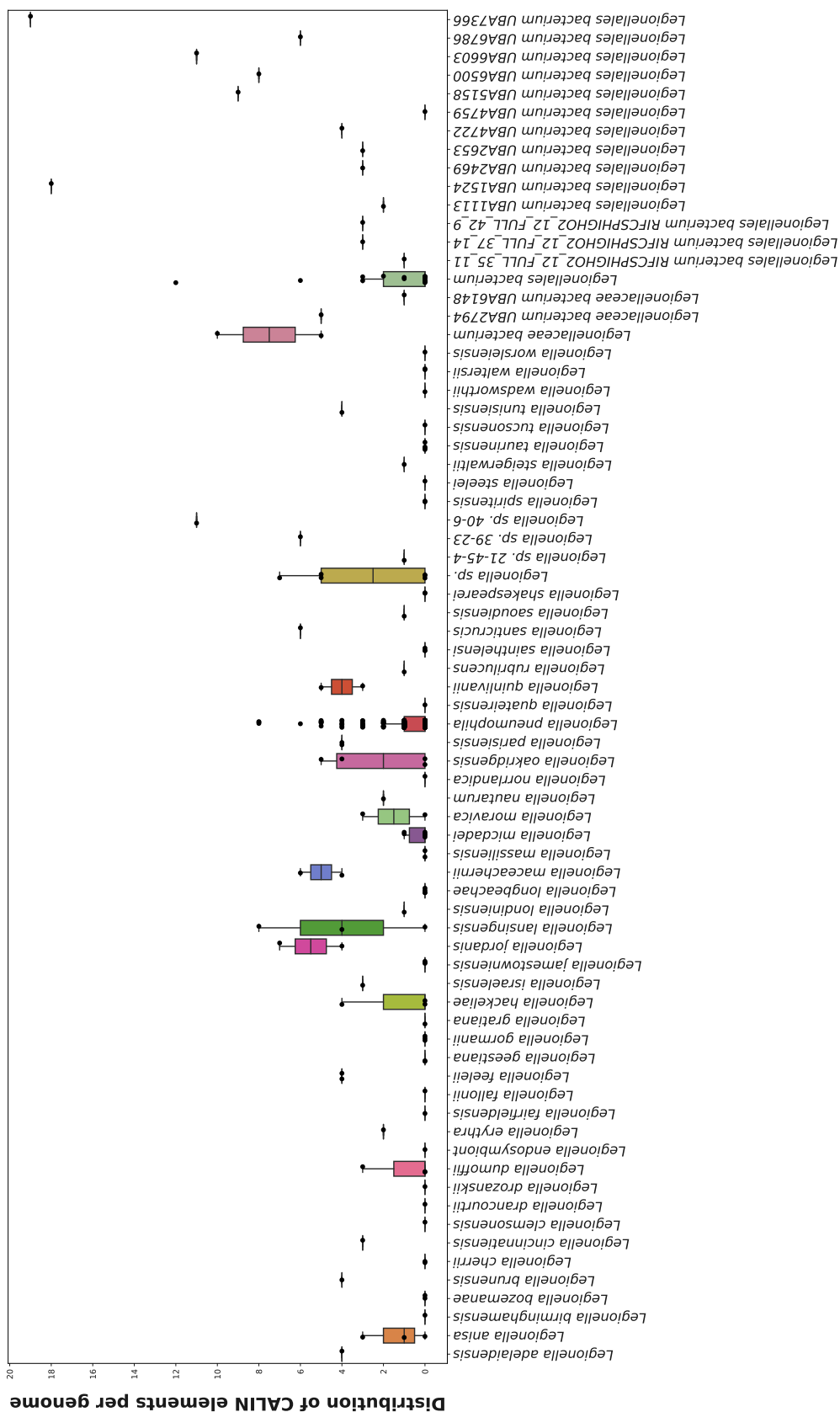
Figure 5.7. Distribution of the number of CALIN elements identified within each genome by IntegronFinder for each of the 73 *Legionella* spp. available in GenBank. Figure shows the distribution of the number of CALIN elements identified per strain within each of the 717 genomes for all 73 *Legionella* spp. analysed in this study using IntegronFinder. The black horizontal line in each box indicates the median value, while the upper and lower regions of the box indicates the 25 % and 75 % quartiles, while whisker extending from the boxes correspond to 1.5 × the interquartile range. Each black circle indicates a single examined genome.

integrated into the bacterial genomes, or by existing in the cytoplasm as episomes. These are called 'prophage' elements and are highly prevalent across bacterial genomes, in some cases making up $\sim$ 20 % of a bacteria's genome (Casjens, 2003). Examination of *Legionella* spp. genomes for prophages bioinformatically has also shown their potential presence, when Gomez-Valero et al. (2014) showed that a putative prophage may exist in the genome of one strain of *L. micdadei*. As bioinformatic analysis had shown success previously and experimental had so far failed to give any bacteriophages, in this chapter a bioinformatic approach was attempted to try and identify prophages within all of the publicly available *Legionella* spp. genomes.

Bioinformatics approaches have been previously used to identify putative prophage elements within the genomes of other bacterial species such as for, *M. catarrhalis* (Ariff et al., 2015), *Mycobacterium abscessus* (Sassi et al., 2013), and *Mannheimia haemolytica* (Niu et al., 2015). These computer-based studies can be highly accurate in predicting the existence of experimentally isolatable bacteriophages, rather than just acting as theoretical estimation of a prophages existence. For instance, in the study by Lorenz et al. (2016) a bioinformatic approach was initially used to identify prophage within the genome of *Vibrio campbellii* strain ATCC BAA-1116. Bioinformatic analysis revealed the presence of two putative *Myoviridae* within the bacterial genome, which was then confirmed experimentally in their study after exposure of the bacteria to the inducing agent MitC let to the observation of two sets of morphologically distinct *Myoviridae* particles along with enrichment of the predicted prophage DNA within the induced lysate (Lorenz et al., 2016).

In the bioinformatic study by Gomez-Valero et al. (2014) analysis was conducted on only 11 *Legionella* spp. genomes, covering five species, *L. hackeliae* (1), *L. micdadei* (1), *L. fallonii* (1), *L. pneumophila* (7), and *L. longbeachae* (1). Therefore, to increase the chances of detecting *Legionella* spp. bacteriophages in the current work, a larger *Legionella* spp. genome collection was used by exploring prophage presence in all *Legionella* spp. genomes available within the GenBank database. This collection contained a total of 717 bacterial genomes, covering a total of 73 *Legionella* spp.. The presence of prophages within the bacterial genomes was examined using PHASTER, a prophage prediction tool that has been used successfully for identifying prophages in a number of bacterial species (Crispim et al., 2018). For instance, the earlier described Lorenz et al. (2016) study actually used the previous release of PHASTER, PHAST, to bioinformatically predict the prophage regions in *V. campbellii*. PHASTER identifies potential prophage regions by comparison of coding sequences within the bacterial genome to a database of known bacteriophages and their genes (Arndt et al., 2016), allowing it to also make predictions on how complete an identified element is.

PHASTER identified 631 putative prophage elements within the 717 *Legionella* spp. genomes. Only 11 of the putative prophage elements were designated as potentially 'intact' regions, meaning they contain a large number of genes corresponding to known bacteriophages. The median size of these predicted

'intact' elements was 5,670 kb, ruling out the possibility that these elements were intact *Caudovirales*, as the smallest known *Caudovirales* bacteriophage genome is 11.5 Kb belonging to *Mycoplasma pulmonis* (Tu et al., 2001). Instead, this is similar to the complete genome size of members of the *Microviridae* family, which have genomes ranging between $\sim$ 4 - 7 kb in size (Roux et al., 2012). Re-annotation of the putative prophage identified by PHASTER using the pVOGs database, a well curate database of bacteriophage genes, confirmed that these regions were likely *Microviridae* as numerous genes with annotations to *Microviridae* VOGs were present in the 'intact' putative prophage elements. Alignment of the 11 putative intact elements and an additional element classified as 'questionable' that contained *Microviridae* VOGs to known reference genera of *Microviridae* showed that all of these were highly similar to members of the *Sinsheimervirus* genus, particularly ϕX-174. These regions were eventually confirmed to be likely ϕX-174 due to each element having an ANI of at least 99.7 % to ϕX-174 and identical genome arrangements.

ϕX-174 DNA is commonly used as a quality and calibration control for Illumina sequencing runs due to its relatively small genome size (Brandt and Albertsen, 2018). As it is commonly used as a control in Illumina sequencing, one of the key good practice steps in the processing of reads from Illumina sequencing runs is the removal of reads corresponding to ϕX-174 DNA (Wright et al., 2017). However, despite this ϕX-174 is still often found within bacterial genomes in the public databases. For instance, Mukherjee et al. (2015) examined 18,000 publicly available bacterial genomes for the presence of ϕX-174 contamination. They found that > 5.5 % of these genomes were contaminated with ϕX-174, and that $\sim$ 10 % of the contaminated genomes had actually been published in the literature. The bioinformatic pipelines used in the studies found to contain ϕX-174 contamination in the current work make no mention of the removal of ϕ-X174 in processing of their sequence data (McAdam et al., 2014; Watanabe et al., 2015), and therefore it would not be unrealistic for the *Microviridae* to have actually been introduced during the sequencing process and were not removed bioinformatically. This would mean that the prophage elements found to be highly similar to ϕX-174 were just contaminants that were not removed before read assembly and submission to the GenBank database. However, this does indicate that the prophage methodology used in the current work is highly capable of identifying prophages with similarities to known bacteriophages.

The majority of the putative prophage elements identified by PHASTER were classified as 'incomplete' elements. Further examination of these elements led to identification of some bacteriophage genes after re-annotation by pVOGs, including those associated with bacteriophage structural proteins in a small number of putative prophage regions. Although these elements were designated as incomplete or questionable they could therefore actually be complete elements. This is because PHASTER's score designations are based on the similarity of each gene/protein to a known bacteriophage gene, and thus obtaining a high score will depend on the presence of well-annotated viral genomes in its database. With no characterised bacteriophages infecting *Legionella* spp. available in the literature, and thus

PHASTER's database, PHASTER would be unable to recognise these genes meaning a low score would be given to these regions if the *Legionella* spp. bacteriophages were distinct from those in its database. Thus, although poorly scored for completeness, these incomplete or questionable regions may still be complete prophage elements.

This database issue may also give an explanation of the overall high number of 'incomplete' elements predicted by PHASTER. This is a limitation faced by all other widely used prophage identification software e.g. PhiSpy (Akhter et al., 2012), Prophinder (Lima-Mendez et al., 2008), and PhageWeb (de Sousa et al., 2018), as each rely on homology comparison to known viruses. Alternatively, these elements could have been 'intact' elements at one point, but over time have become degraded within the bacterial cell (Canchaya et al., 2003), and are now cryptic bacteriophages. Cryptic prophages are common in bacteria (Wang et al., 2010), and are prophages that have become defective in some parts of their genome meaning they are unable to form virulent particles (Bobay et al., 2014). However, this scenario is unlikely to explain the high number of incomplete elements found in the current study though due to the total abundance of incomplete vs intact elements. For example, in the study of Ariff et al. (2015) investigating the presence of prophage in 95 *M. catarrhalis* genomes, they found that there were ~ 3 times more incomplete prophages than the either questionable or intact prophages. In the current study, there were ~ 23 times as many incomplete and questionable as intact prophages. This would mean that either there is a ~ 8 times greater selection for cryptic elements in *Legionella* spp. compared to *M. catarrhalis*, or that *Legionella* spp. rarely undergo lysogeny with the intact and questionable elements being signs of infections from many years ago, both of which are highly unlikely scenarios.

Despite only a small number of prophage structural genes being identified within the questionable and incomplete elements, a large number of integrases (372) were found. None of the element's that contained integrases contained bacteriophage structural genes, instead containing more bacterial-associated genes. Therefore, to examine if these integrase containing elements were prophage-associated, or from other transposable DNA elements, the presence of integrons, recombinational elements that contain an integrase, was examined. Following analysis of all the genomes of *Legionella* spp., no complete integrons, elements with both an integrase and at least one *attC* site, were found. Instead, a high number of CALIN elements, integron elements that do not contain an integrase were found throughout the *Legionella* spp. examined. The absence of integrase region identified by IntegronFinder but not PHASTER could again be due to the databases being used, as IntegronFinder and PHASTER are using different integrase-containing databases. If these are true CALIN elements, that do not overlap with the prophage sequences, then the reason for the integron elements missing integrases could be because integrons are continuously under selective pressure (Engelstädter et al., 2016), meaning these integron elements were complete at some point but lost their integrases over the time.

In conclusion, although all available *Legionella* spp. genomes within the public databases were screened for prophage elements, no confident detections could be found. However, a number of potential prophage-like elements containing common key bacteriophage genes were observed which may prove useful in future studies attempting to isolate bacteriophages against *Legionella* spp..

6. Concluding discussion

Bacteriophages are the viral predators of bacteria, and are important parts of the global microbiome, where they can have important roles in regulating the numbers of bacterial populations (Koskella, 2013). Because of this ability, and their high level of specificity to only infect a small number of bacterial hosts (Flores et al., 2011), bacteriophages are now becoming recognised as useful tools in the clearance or detection of bacteria that are otherwise difficult to remove or identify (Salmond and Fineran, 2015). For example, therapeutic use of bacteriophages to treat antimicrobial resistant organisms has already been shown to be successful in cases, such as *P. aeruginosa*-associated graft infections (Chan et al., 2018), *Staphylococcus*-associated diabetic ulcers (Fish et al., 2016), and chronic lung infections caused by *Achromobacter xylosoxidans* (Hoyle et al., 2018). For bacteria that are challenging to detect, bacteriophages have also been shown to be useful. For instance, Swift et al. (2013) developed a highly accurate bacteriophage-based assay to detect *Mycobacterium avium* subsp. *paratuberculosis* in cattle within 48 hours, a diagnosis that would normally have taken $\sim$ 16 weeks by culture of the organism.

Legionella spp. are aquatic bacteria that are normally found within environmental and man-made water systems (Parthuisot et al., 2010; Kuroki et al., 2017). *Legionella* spp. are a significant healthcare risk, as inhalation of water particles containing the bacteria can lead to legionellosis, a set of conditions that can range from having mild flu-like symptoms, Pontiac fever, to a form of potentially fatal pneumonia, Legionnaires' disease (Fields et al., 2002; Bartram et al., 2007). These organisms can be both difficult to remove and difficult to detect within environmental/man-made systems, as well as in patients infected with the organism. For instance, in natural water and man-made systems containing *Legionella* spp., treatments, such as exposure to high temperatures or chlorine-based bleaches are used to try and eliminate the bacteria (Kim et al., 2002; Marchesi et al., 2016). However, through the bacteria's ability to grow within eukaryotic cells e.g. amoebae, and by forming biofilms, treatments may often not be fully effective (Barker et al., 1992; Abdel-Nour et al., 2013; Falkinham, 2015). In patient diagnosis, the *Legionella* urine antigen test is currently used to detect cases of Legionnaires' disease (Lim et al., 2009). However, this test has been shown to only accurately detect infections caused by *L. pneumophila* serogroup 1 (Burillo et al., 2017; Miller et al., 2018), while with infections caused by all other non-*pneumophilla Legionella* species, and *L. pneumophila* serogroup 2 - 14 shown

to produce inconsistent results using this test (Helbig et al., 2003). The use of bacteriophages in detection and elimination of *Legionella* spp. may be able to overcome these issues.

Although bacteriophages are widespread in nature, very little has been published on bacteriophages infecting *Legionella* spp.. Only two studies in the literatures have mentioned the identification or isolation of bacteriophages infecting *Legionella* spp.. In one of these studies, Lammertyn et al. (2008) described the isolation of four bacteriophages from river water samples that were able to infect *L. pneumophila* and in some cases other *Legionella* spp.. However, these isolated bacteriophages are no longer able to be examined, as they were lost during further work with no stocks of these viruses still in existence (personal communications, Elke Lammertyn). In the other study, Gomez-Valero et al. (2014) found bioinformatically that there may be a prophage element within the genome of *L. micdadei* NCTC 11371 after finding 16 bacteriophage-related proteins (e.g., bacteriophage capsid, tail, and replication proteins) within one ~ 70 open reading frame long region. Therefore, there are currently no confirmed bacteriophages in existence that are able to infect *Legionella* spp., meaning that if bacteriophage-based treatments and diagnostics were to be developed, bacteriophages against these organisms would need to be isolated. The present study aimed to rectify this by attempting to isolate bacteriophages against *Legionella* spp..

However, before any attempts could be made to isolate bacteriophages infecting *Legionella* spp., it was important to have a broad range of *Legionella* spp. isolates for use in the isolation process. Two sets of samples, one consisting of 108 samples from various environmental and man-made sources around the UK and other countries and another set of 154 sample from water and soil systems around the Midlands, UK, were screened to try and isolate such a collection. A total of 18 strains of *Legionella* spp. were isolated, all of which were *L. pneumophila*. Although this is only a small number, these results were consistent with other studies. Dimitriadi and Velonakis (2014) collected only 6 *Legionella* spp. isolates despite screening 100 water samples, while numerous other studies across a range of sampling locations have also found similar results (Whiley and Taylor, 2016).

An additional 37 strains were able to be sourced from the collection of PHE, giving a total of 57 isolates in the *Legionella* spp. collection, 45 being *L. pneumophila*. These *L. pneumophila* strains were found to be made up of 20 serogroup 1 isolates, 24 serogroup 2 - 14, and 1 isolate that had a cross-reaction with all three antibodies used to distinguish between serogroup 1, serogroup 2 - 14, and the other *Legionella* spp.. Sequence-based typing further separated the collection of *L. pneumophila* isolated to 18 STs, from which 6 were novel STs that were uploaded to the EWGLI public database. The majority of these strains belonged to ST337, which can be identified within hospital environments, but with strains in this ST not frequently associated with causing disease. A number of ST59 isolates were also in the collection, and previously members in this ST have directly associated with disease outbreaks (Kozak-Muiznieks et al., 2014; Quero et al., 2018). Other STs found that have had members associated

with disease outbreaks include ST36, ST1, ST8, and ST191 which have been associated with outbreaks of legionellosis in the USA and UK (Harrison et al., 2009; Kozak-Muiznieks et al., 2014). By having 27 distinct *Legionella* spp. isolates (18 distinct STs of *L. pneumophila* plus an additional 9 different species of *Legionella*), the current study should be 27x more likely to isolate bacteriophage than the study of Lammertyn et al. (2008), who were able to isolate four *Legionella* bacteriophages using only a single strain of *L. pneumophila*.

The low recovery levels from the screened samples could be due to factors, such as the presence of anti-*Legionella* molecules produced by other organisms in the samples (Toze et al., 1990; Corre et al., 2018) that are known to limit *Legionella* spp. growth. Alternatively *Legionella* spp. in some of the samples could be in the VBNC state, where they are still viable but in a dormant state, unable to be grown under laboratory conditions (Whiley and Taylor, 2016; Kirschner, 2016). Either of these scenarios can be supported by the observations from qPCR of the second sample set to detect *Legionella* spp. where, although no *Legionella* spp. were isolated from these samples, amplification of *Legionella* spp. genes was found in 153/154 examined samples. This again is common, with the chance of identifying *Legionella* spp. by qPCR often found to be likely than identifying these organisms by culturing (Diederen et al., 2007; Parthuisot et al., 2010). Although identification of *Legionella* spp. by qPCR does not aid in the isolation of new *Legionella* spp. strains, it could aid in the isolation of bacteriophages. Bacteriophages are found in nearly all environments where their bacterial host exist (Koskella et al., 2011; Brown-Jaque et al., 2016), and by having positive detection of *Legionella* spp. in the samples by qPCR, this could be an indication that they also contain bacteriophages.

Sample enrichment is one of the common approaches that are used in the isolation of bacteriophages against bacterial isolates (Li et al., 2010; Yordpratum et al., 2011). However, when the collected samples were incubated with the assembled strain collection and examined for plaque formation on solid media, a small number of samples produced only a zone of lysis, not single plaques. Only on one occasion one sample, K2 from Kurdistan, Iraq, produced single plaques. This is consistent with the observations of Lammertyn et al. (2008), who were infrequently and unreliably able to observed single plaques formation, while zones of clearance were often found. As an alternative approach, the screening of these samples was re-done to try and observe bacteriophage infection in liquid rather than solid media using a well-assay (Millard, 2009). Although some reductions in bacterial density were observed, further passage of the same samples did not give any reductions in a second round, which is one of the distinct features of bacteriophage replication compared to other antimicrobials (Abedon, 2014), removing the possibility of finding potential bacteriophages within the samples. Therefore, the initial reductions in bacterial density could be due to the presence of inhibitors or bacteriocins in the samples that become diluted following passage to the next cycle, rather than bacteriophages, which would also explain the presence of only zones of clearance and no plaque formation found with the tests on solid media (Abedon, 2018). Therefore, although the qPCR indication of *Legionella* spp. in the

collected samples might have been a promising indication that bacteriophages may be present within the samples, enrichment of all samples for bacteriophages using the assembled strain collection proved unsuccessful.

The assembled strain collection was then itself the focus of bacteriophage isolation efforts. Prophage elements, bacteriophages that either integrated their DNA into the host chromosome or persist in the cell cytoplasm following DNA injection, are common in bacteria with roughly half of sequenced bacteria containing at least one prophage element (Touchon et al., 2016). These elements can often be induced into the lytic cycle, creating and releasing new viruses into the environments after being exposed to stress stimuli, such as antibiotics or reactive oxygen species (Canchaya et al., 2003). In other bacteria where no bacteriophages have been isolated from sample enrichments, temperate bacteriophages have been isolated from induction of bacterial strains. Hammerl et al. (2016) used common bacteriophage inducers (MitC, UV, and heat) to examine a strain of *Brucella inopinata*, where they had identified a prophage bioinformatically within its chromosome. This led to the isolation of temperate bacteriophage BiPBO1 (Hammerl et al., 2016). This approach was thought to be promising since the presence of a putative prophage within the genome of one isolate of *L. micdadei* had already been reported by Gomez-Valero et al. (2014). Therefore, this method was then attempted to isolate bacteriophages directly from the genomes of the diverse *Legionella* spp. strain collection assembled.

To begin, the sensitivity of the strains in the *Legionella* spp. collection to two commonly used inducing agents, MitC and NFX, was examined to see if any of the isolates would potentially be prophage carrier. As prophage carriage can lead to increased sensitivity to these inducing agents (Muschel and Schmoker, 1966), any strains showing increased sensitivity to the inducing agents may be carrying prophage. Eight strains showed greater sensitivity to these agents compared to all other isolates. Retesting their sensitivity in a microdilution assay also showed that for some of these strains bacterial density declined in the presence of certain concentrations of target inducer, potentially indicating the lysis of bacteria following prophage release.

Following this, larger-scale induced lysates were then analysed using TEM, leading to the observation of the presence of two morphologically distinct possible virus-like particles in the lysates. One particle had a spherical shape between 58 - 70 nm in diameter, while the other had a filamentous morphology of similar diameter, but with vastly varied length up to $\sim 6~\mu$m. Other spherical bacteriophages have been discovered, such as the *Cystoviridae* e.g. *Pseudomonas* bacteriophage phi6 (Vidaver et al., 1973). Although the virus-like particles here are similar shapes to the *Cystoviridae*, the sperical virus-like particles are smaller than the *Cystoviridae*, which are normally ~ 85 nm in diameter (Poranen et al., 2017). Also *Cystoviridae* are dsRNA viruses, that are normally believed to be obligately lytic, although carrier states have been observed (Cuppels et al., 1979; Onodera et al., 1992). The filamentous particles had a very unusual shape, appearing to be long filaments composed of individual repeating

units in a 'bead on a string'-like morphology, with the repeating units roughly the same size as the spherical particles. No bacteriophage particle with this morphology has been reported, not even the currently known filamentous bacteriophages. The 'bead on a string'-like virus-like particles had a diameter of $\sim$ 50 nm by between $\sim$ 500 nm - 6 μm in length, larger than any other known filamentous bacteriophages/archeal viruses (e.g. the *Inoviridae* family of bacteriophage are $\sim$ 10 nm in diameter and up to $\sim$ 4 μm in length, while the archeal *Ligamenvirales* are $\sim$ 25 - 30 nm in diameter and $\sim$ 0.75 - 2.5 μm in length) (Day, 2012; Prangishvili, 2012a,b).

Although not similar to any known bacteriophages, if the spherical and filamentous particles are actually different morphotypes of the same virus then this is similar to some eukaryotic viruses, such as the *Influenza* virus. *Influenza* viruses are pleomorphic viruses that can be seen as either spherical, bacilliform, or filamentous particles (Lamb and Choppin, 1983), with the dominant morphotype produced related to environmental selection (Seladi-Schulman et al., 2013). The filaments of these viruses can also get to extremely long lengths, > 50 μm (Dadonaite et al., 2016), far beyond that of the longest filaments found in the *Legionella* spp. filamentous virus-like particles. When DNA was attempted to be extracted from the *Legionella* spp. virus-like particles, even over a gradient of induction conditions, no DNA was found within the viral fraction. Therefore, as the closest viruses in appearance to the *Legionella* spp. virus-like particles are RNA viruses, the next logical step would be to examine these particles to see if they encoded viral RNA. However, further characterisation of these particles could not be conducted due to time restrictions.

In parallel to the experimental inductions to isolate prophages from the collection of *Legionella* spp. strains, a bioinformatic approach was also used to try and detect viruses within publicly available *Legionella* spp. genomes. A bioinformatic approach was believed to be a good alternative analysis method process, due to it previously being used successfully to identify bacteriophages in other, difficult, organisms. For instance, Dyson et al. (2016) had a similar lack of success in isolating bacteriophages for *Gordonia alkanivorans*, *Gordonia alkanivorans malaquae*, and *Nocardia brevicatena* experimentally, so they examined the whole genome sequences of these organisms using a prophage identification tool, PHAST. Using this tool they identified genes of bacteriophage origin in 83 % of the bacterial genomes. In 26 % of the genomes PHAST had identified the presence of at least one potentially complete prophage (Dyson et al., 2016). Thus, if bacteriophages were not going to be able to be isolated experimentally, a bioinformatic approach may at least allow identification of potential future targets. Gomez-Valero et al. (2014) had already utilised a bioinformatic approach to screen a total of 11 *Legionella* spp. genomes to try and detect any putative prophages in their bacterial genomes. This led to the identification of one potential prophage element in the genome of *L. micdadei* NCTC 11371. Therefore, to expand on their study, all available genomes within the GenBank database, 717 genomes in total, were analysed using an updated version of PHAST, PHASTER.

Upon screening of the 717 *Legionella* spp. genomes using PHASTER, 11 *Legionella* spp. genomes were predicted to contain a prophage region that had been classified as potentially intact, while 15 regions in bacterial genomes had been identified as questionable, and 605 as incomplete. Further analysis of the 11 potentially complete prophage regions identified by PHASTER revealed that all of these regions harboured genes encoding proteins belonging to the *Microviridae* family. Comparison of these to each other, a questionable element found to contain *Microviridae* genes, and other known *Microviridae*, revealed that all of the 12 *Microviridae* in the *Legionella* spp. genomes were identical to each other, and ϕX-174. ϕX-174 is commonly used as a quality and calibration control for Illumina sequencing. Removal of ϕX-174 reads from sequencing data is one of the first steps that should be conducted in the quality control process for all genome sequencing data (Wright et al., 2017). However, a large number of microbial genome sequences within the public databases are still contaminated with ϕX-174 from them (Mukherjee et al., 2015). This is likely what has happened here, with PHASTER detecting this contaminating sequence and noting it as a potential complete *Legionella* spp. prophages. However, this does show that PHASTER would be able to find complete prophage elements within the analysed genomes if they were present. With only questionable and incomplete elements identified other than the ϕX-174 contamination, this could mean either of two possibilities. Either these elements have now been 'domesticated' by their bacterial hosts to become cryptic bacteriophages that are missing a number of key genes (Canchaya et al., 2003; Bobay et al., 2014) or, if any prophages are present in the *Legionella* spp. genomes, they are then distinct from those bacteriophage sequences currently available in the prophage databases.

In summary, a broad and diverse collection of *Legionella* spp. was obtained in this study and used as both host strains for trying to isolate bacteriophages infecting *Legionella* spp. from environmental samples, and as targets for prophage induction to release any prophages from their chromosomes. Although attempts in isolating bacteriophages from environmental samples were unsuccessful, induction of *Legionella* spp. strains did possibly give evidence of virus-like particles that may be RNA-based. However, due to time restrictions these particles could not be investigated in further detail. Bioinformatic analysis of all available *Legionella* spp. genomes in the GenBank database to find potential prophage also proved unsuccessful. However, although no prophage elements were detected confidently, a number of genomes harboured genes with potential functions as bacteriophages, making these regions a potential target for further investigation in the future.

7. Future work

Although the current study was not successful in definitively isolating bacteriophages infecting *Legionella* spp., a number of interesting observations and questions were noted that warrant future study. Of most importance is the nature of the virus-like particles identified following induction of the *Legionella* spp. strains. In particular, since attempts to isolate DNA from these particles were unsuccessful in this study, do these particles contain RNA instead? This could be investigate by extracting and measuring the concentration of the viral RNA using gradients of induction agents just as conducted in the current study for DNA extractions. Alternatively, epifluorescence microscopy could be used (Ortmann and Suttle, 2009). This would allow visual detection of the presence of viral DNA or RNA within the particles, confirming that they are not just empty particles. Also if so and these particles do contain RNA, do both types contain the same RNA genomes? Finally, if RNA genomes are detected, then these genomes could be sequenced and could act as the beginning of a sequence database to investigate publicly available *Legionella* spp. RNA transcriptomes/total meta-transcriptomes for potential new variants of these RNA viruses either in *Legionella* spp. or in other organisms.

For experimental isolation of both *Legionella* spp. and *Legionella* spp. bacteriophages, could amoeba or similar organisms assist in bacterial recovery? Do these amoeba act as reservoirs for bacteriophage? (Magnet et al., 2015) group showed that the isolation of *Legionella* spp. can be improved in presence of *Acanthamoeba* in screened samples, and could amoeba act as an environment that allows bacteriophages to replicate in *Legionella* spp.? For this reason, does co-culturing with amoebae increase the chance of identifying live *Legionella*-positive samples? and if so can this assist in the isolation of *Legionella* spp. bacteriophages?

Also related to bacteriophage isolation, is a possible reason for the lack of bacteriophages being detected during sample screening due to the presence of iron within the culturing media? Iron is one of the factors that could have a detrimental effect on activity of bacteriophages (Sagripanti et al., 1993; Penner et al., 2016). Therefore, could removing iron from the liquid and solid media in this study aid in the isolation bacteriophages against *Legionella* spp.? Although Centers for Disease Control (CDC) (2005) guidelines recommended that *Legionella* spp. growth considerably depends on the presence of

L-cysteine-HCl and iron salts, a study by Barker et al. (1986) demonstrated that supplemented iron had no considerable effect on growth of *L. pneumophila*. Therefore, this should not interfere with bacterial growth and could be easily examined.

Finally, could shotgun metagenomic sequencing of induced lysates from *Legionella* spp. aid in the detection of *Legionella* spp. prophages? As the PHASTER database, and those of all other homology-based prophage finding tools, do not contain any sequences related to *Legionella* spp. they will continually struggle to detect 'complete' bacteriophages. However, shotgun metagenomic sequencing of induced lysates would require no knowledge of the bacteriophages, with any induced prophages being sequenced. These sequences would then begin to add to databases such as PHASTERs that will then aid in prophage predictions.

8. Supplementary Information

Table S1. Table of Media/Buffer/Reagents used in this study

Media/Buffer/Reagent name	Composition
Cysteine (4 %)	0.4 g L-Cysteine hydrochloride monohydrate (Sigma)
	Distilled water to 10 ml
	Sterilise by filtration
Iron (2.5 %)	0.25 g Iron(III) pyrophosphate (Sigma)
	Distilled water to 10 ml
	Sterilise by filtration
MgCl$_2$ (1 M)	40.66 g MgCl$_2$ (Fisher)
	Distilled water to 200 ml
	Sterilise by filtration
CaCl$_2$ (1 M)	29.40 g CaCl$_2$ (Sigma)
	Distilled water to 200 ml
	Sterilise by filtration
BYE-α (filter-sterilised) Broth	4 g N-(2-Acetamido)-2-aminoethanesulfonic acid (ACES) Buffer (Sigma)
	1.12 g Potassium Hydroxide (Fisher Scientific)
	4.8 g Bacto™ Yeast Extract (BD Biosciences)
	0.4 g α-Ketoglutaric acid (Sigma)
	Distilled water to 400 ml
	Adjust pH to 6.9 with HCl
	Sterilise by filtration
	4 ml Cysteine (4 %)*

Media/Buffer/Reagent name	Composition
BYE-α (autoclaved) Broth	4 ml Iron (2.5 %)*
	* Added after filter-sterilisation
	4 g N-(2-Acetamido)-2-aminoethanesulfonic acid (ACES) Buffer (Sigma)
	1.12 g Potassium Hydroxide (Fisher Scientific)
	4.8 g Bacto™ Yeast Extract (BD Biosciences)
	0.4 g α-Ketoglutaric acid (Sigma)
	Distilled water to 400 ml
	Adjust pH to 6.9 with HCl
	Sterilise by autoclave
	4 ml Cysteine (4 %)*
	4 ml Iron (2.5 %)*
	* Added after autoclaving
BYE-α (charcoal + filtration) Broth	4 g N-(2-Acetamido)-2-aminoethanesulfonic acid (ACES) Buffer (Sigma)
	1.12 g Potassium Hydroxide (Fisher Scientific)
	4.8 g Bacto™ Yeast Extract (BD Biosciences)
	0.4 g α-Ketoglutaric acid (Sigma)
	0.8 g Activated Charcoal (Sigma)
	Distilled water to 400 ml
	Adjust pH to 6.9 with HCl
	Sterilise and remove charcoal by filtration
	4 ml Cysteine (4 %)*
	4 ml Iron (2.5 %)*
	* Added after filter-sterilisation
BYE-α (charcoal + centrifugation) Broth	4 g N-(2-Acetamido)-2-aminoethanesulfonic acid (ACES) Buffer (Sigma)
	1.12 g Potassium Hydroxide (Fisher Scientific)
	4.8 g Bacto™ Yeast Extract (BD Biosciences)
	0.4 g α-Ketoglutaric acid (Sigma)
	0.8 g Activated Charcoal (Sigma)

Media/Buffer/Reagent name	Composition
BYE-α (with charcoal) Broth	Distilled water to 400 ml
	Adjust pH to 6.9 with HCl
	Sterilise by autoclave
	remove charcoal by centrifugation
	4 ml Cysteine (4 %)*
	4 ml Iron (2.5 %)*
	* Added after centrifugation
	4 g N-(2-Acetamido)-2-aminoethanesulfonic acid (ACES) Buffer (Sigma)
	1.12 g Potassium Hydroxide (Fisher Scientific)
	4.8 g Bacto™ Yeast Extract (BD Biosciences)
	0.4 g α-Ketoglutaric acid (Sigma)
	0.8 g Activated Charcoal (Sigma)
	Distilled water to 400 ml
	Adjust pH to 6.9 with HCl
	Sterilise by autoclave
	4 ml Cysteine (4 %)*
	4 ml Iron (2.5 %)*
	* Added after autoclaving
BYE-α Broth	4 g N-(2-Acetamido)-2-aminoethanesulfonic acid (ACES) Buffer (Sigma)
	1.12 g Potassium Hydroxide (Fisher Scientific)
	4.8 g Bacto™ Yeast Extract (BD Biosciences)
	0.4 g α-Ketoglutaric acid (Sigma)
	Distilled water to 400 ml
	Adjust pH to 6.9 with HCl
	Sterilise by autoclave
	4 ml Cysteine (4 %)*
	4 ml Iron (2.5 %)*
	* Added after autoclaving

Media/Buffer/Reagent name	Composition
BYE-α Broth with divalent cations	4 g N-(2-Acetamido)-2-aminoethanesulfonic acid (ACES) Buffer (Sigma)
	1.12 g Potassium Hydroxide (Fisher Scientific)
	4.8 g Bacto™ Yeast Extract (BD Biosciences)
	0.4 g α-Ketoglutaric acid (Sigma)
	Distilled water to 400 ml
	Adjust pH to 6.9 with HCl
	Sterilise by autoclave
	4 ml Cysteine (4 %)*
	2 ml $MgCl_2$ (1 M)*
	2 ml $CaCl_2$ (1 M)*
	* Added after autoclaving
BCYE-α 1.2% Agar	4 g ACES Buffer (Sigma)
	1.12 g Potassium Hydroxide (Fisher Scientific)
	4.8 g Bacto™ Yeast Extract (BD Biosciences)
	0.4 g α-Ketoglutaric acid (Sigma)
	4.8 g Agar bacteriological (VWR)
	0.8 g Activated Charcoal (Sigma)
	Distilled water to 400 ml
	Adjust pH to 6.9 with HCl
	Sterilise by autoclave
	4 ml Cysteine (4 %)*
	4 ml Iron (2.5 %)*
	* Added after autoclaving
BCYE-α 1.2% Agar (without Charcoal)	4 g ACES Buffer (Sigma)
	1.12 g Potassium Hydroxide (Fisher Scientific)
	4.8 g Bacto™ Yeast Extract (BD Biosciences)
	0.4 g α-Ketoglutaric acid (Sigma)
	4.8 g Agar bacteriological (VWR)
	0.8 g Activated Charcoal (Sigma)

Media/Buffer/Reagent name	Composition
BCYE-α 0.4% Overlay agar	Distilled water to 400 ml
	Adjust pH to 6.9 with HCl
	Sterilise by autoclave
	Remove the Charcoal by centrifugation
	4 ml Cysteine (4 %)*
	4 ml Iron (2.5 %)*
	* Added after autoclaving and centrifugation
	0.625 g ACES Buffer (Sigma)
	0.175 g Potassium Hydroxide (Fisher Scientific)
	0.75 g Bacto™ Yeast Extract (BD Biosciences)
	0.0625 g α-Ketoglutaric acid (Sigma)
	0.25 g Agar bacteriological (VWR)
	0.125 g Activated Charcoal (Sigma)
	Distilled water to 60 ml
	Adjust pH to 6.9 with HCl
	Sterilise by autoclave
	Remove the Charcoal by centrifugation
	0.625 ml Cysteine (4 %)*
	0.625 ml Iron (2.5 %)*
	* Added after autoclaving and centrifugation
***Legionella* transparent medium (LTM)**	6.66 g Proteose peptone 3 (BD Biosciences)
	0.44 g Bacto™ Yeast Extract (BD Biosciences)
	4.44 g ACES buffer (Sigma)
	0.11 g Iron (Sigma)
	0.44 g α-Ketoglutarate (Sigma)
	7.55 g Agar bacteriological (VWR)
	Distilled water to 400 ml
	Adjust pH to 6.9 with HCl
	Sterilise by autoclave

Media/Buffer/Reagent name	Composition
	1.77 ml 10% Cysteine* (Sigma)
	1.4 ml 1M Potassium Hydroxide* (Fisher Scientific)
	44.44 ml 20% Bovine serum albumin (BSA) fraction-V* (Sigma)
	* Added after autoclaving
BCYE-α 1.2% Agar	4 g ACES Buffer (Sigma)
with catalase	1.12 g Potassium Hydroxide (Fisher Scientific)
	4.8 g Bacto™ Yeast Extract (BD Biosciences)
	0.4 g α-Ketoglutaric acid (Sigma)
	4.8 g Agar bacteriological (VWR)
	Distilled water to 400 ml
	Adjust pH to 6.9 with HCl
	Sterilise by autoclave
	4 ml Cysteine (4 %)*
	4 ml Iron (2.5 %)*
	250 U ml^{-1} catalase (Sigma)*
	* Added after autoclaving
SM Buffer	5.8 g NaCl (Fisher)
	2 g MgSO4·7H2O (Fisher)
	50ml 1M Tris HCl (pH 7.5)
	Distilled water to 1 L
	Sterilise by autoclave
50 X TAE buffer	484 g Tris-base (Fisher)
	Dissolve in 1.6 L distilled water
	200 ml 0.5 M Ethylenediaminetetraacetic acid (EDTA) (Sigma)
	114.2 ml Glacial Acetic Acid (Fisher)
	Distilled water to 2 L
	Dilute 50 fold in distilled water before use

Table S2. Table of Samples (set 1) used in this study for isolation of *Legionella* spp.

Sample ID	Set No.	Year	City, Country	Location	Coordinates	Water	Soil
Lake-S1	1	2014	Chafford Hundred, Essex, United kingdom	Chafford Gorges Nature Park	51.489976, 0.301790	Y	N
Lake-S2	1	2014	Chafford Hundred, Essex, United kingdom	Chafford Gorges Nature Park	51.488543, 0302853	Y	N
Lake-S9	1	2014	Chafford Hundred, Essex, United kingdom	Chafford Gorges Nature Park	ND	Y	N
Lake-S10	1	2014	Chafford Hundred, Essex, United kingdom	Chafford Gorges Nature Park	ND	N	Y
Lake-S11	1	2014	Chafford Hundred, Essex, United kingdom	Chafford Gorges Nature Park	ND	Y	N
Lake-S12	1	2014	Chafford Hundred, Essex, United kingdom	Chafford Gorges Nature Park	ND	N	Y
Pond-S13	1	2014	Chafford Hundred, Essex, United kingdom	Chafford Gorges Nature Park	ND	Y	N
Stream U-S3	1	2014	Leicester, United kingdom	Loddington Road, Tilton on the Hill	ND	Y	N
Stream Lower-S4	1	2014	Leicester, United kingdom	Loddington Road, Tilton on the Hill	ND	Y	N
Puddle 1-S5	1	2014	Leicester, United kingdom	Loddington Road, Tilton on the Hill	ND	Y	N

Table S2. – Continued from overleaf

Sample ID	Set No.	Year	City, Country	Location	Coordinates	Water	Soil
Puddle 2-S6	1	2014	Leicester, United kingdom	Loddington Road, Tilton on the Hill	ND	Y	N
Drain-S7	1	2014	Leicester, United kingdom	Loddington Road, Tilton on the Hill	ND	Y	N
Muldy pud-S8	1	2014	Leicester, United kingdom	Loddington Road, Tilton on the Hill	ND	Y	N
Shower head-S16	1	2014	Leicester, United kingdom	Queens Road	ND	Y	N
Wanlip-1	1	2014	Leicestershire, United kingdom	Wanlip	ND	Y	N
Wanlip-2	1	2014	Leicestershire, United kingdom	Wanlip	ND	N	Y
Wanlip-3	1	2014	Leicestershire, United kingdom	Wanlip	ND	Y	N
Wanlip-4	1	2014	Leicestershire, United kingdom	Wanlip	ND	N	Y
Wanlip-5	1	2014	Leicestershire, United kingdom	Wanlip	ND	Y	N
Wanlip-6	1	2014	Leicestershire, United kingdom	Wanlip	ND	N	Y
Wanlip-7	1	2014	Leicestershire, United kingdom	Wanlip	ND	Y	N
Wanlip-8	1	2014	Leicestershire, United kingdom	Wanlip	ND	Y	N
Wanlip-9	1	2014	Leicestershire, United kingdom	Wanlip	ND	N	Y
Wanlip-10	1	2014	Leicestershire, United kingdom	Wanlip	ND	Y	N
Wanlip-11	1	2014	Leicestershire, United kingdom	Wanlip	ND	Y	N
H-A	1	2014	Hayling Island, Hampshire, United kingdom	Hayling Seafront	ND	Y	N
H-B	1	2014	Hayling Island, Hampshire, United kingdom	Shore mud surface	ND	Y	N

129

Table S2. – Continued from overleaf

Sample ID	Set No.	Year	City, Country	Location	Coordinates	Water	Soil
H-C	1	2014	Hayling Island, Hampshire, United kingdom	Small ditch, 2nd ditch along Hayling Billy trail	ND	Y	N
H-D	1	2014	Hayling Island, Hampshire, United kingdom	Fish tank, 21-23°C	ND	Y	N
H-E	1	2014	Hayling Island, Hampshire, United kingdom	Water tank no 1 from East end in Seafront	ND	Y	N
H-F	1	2014	Hayling Island, Hampshire, United kingdom	Water tank no 2 in Seafront	ND	Y	N
H-G	1	2014	Hayling Island, Hampshire, United kingdom	Water tank no 3 in Seafront	ND	Y	N
H-H	1	2014	Hayling Island, Hampshire, United kingdom	Water tank no 4 in Seafront	ND	Y	N
H-I	1	2014	Hayling Island, Hampshire, United kingdom	Top of porch in Seafront 51 (in stainless steel container)	ND	Y	N
H-J	1	2014	Hayling Island, Hampshire, United kingdom	Meridian centre, Havant, On top of air con machine labelled Robt Dyas	ND	Y	N
H-K	1	2014	Hayling Island, Hampshire, United kingdom	Housing Assn flats Havant air con unit tray nearest road	ND	Y	N
H-L	1	2014	Hayling Island, Hampshire, United kingdom	Meridian centre Havant drip from machine labelled O2	ND	Y	N
H-M	1	2014	Hayling Island, Hampshire, United kingdom	Meridian centre Havant machine labelled Genix. drip from base	ND	Y	N

Table S2. – Continued from overleaf

Sample ID	Set No.	Year	City, Country	Location	Coordinates	Water	Soil
H-N	1	2014	Hayling Island, Hampshire, United kingdom	Havant spring (old spring in centre of town)	ND	Y	N
S10	1	2014	Iraq	ND	ND	Y	N
IK2	1	2014	Iraq	ND	ND	Y	N
SK9	1	2014	Iraq	ND	ND	Y	N
K2	1	2014	Iraq	ND	ND	Y	N
SK4	1	2014	Iraq	ND	ND	Y	N
SK3	1	2014	Iraq	ND	ND	Y	N
SK6	1	2014	Iraq	ND	ND	Y	N
Ski	1	2014	Iraq	ND	ND	Y	N
SK5	1	2014	Iraq	ND	ND	Y	N
SK7	1	2014	Iraq	ND	ND	Y	N
SK8	1	2014	Iraq	ND	ND	Y	N
Kenya1	1	2013	Kenya	ND	ND	Y	N
Kenya2	1	2013	Kenya	ND	ND	Y	N
Kenya3	1	2013	Kenya	ND	ND	Y	N
M-UK	1	2014	Margate, United kingdom	Margate Sea	ND	Y	N
M-UK2	1	2014	Margate, United kingdom	Margate Sea	ND	N	Y
KD1	1	2015	Nigeria	ND	ND	Y	N
KD2	1	2015	Nigeria	ND	ND	Y	N
KD3	1	2015	Nigeria	ND	ND	Y	N

131

Table S2. – Continued from overleaf

Sample ID	Set No.	Year	City, Country	Location	Coordinates	Water	Soil
KD4	1	2015	Nigeria	ND	ND	Y	N
KD5	1	2015	Nigeria	ND	ND	Y	N
KD6	1	2015	Nigeria	ND	ND	Y	N
KD7	1	2015	Nigeria	ND	ND	Y	N
KD8	1	2015	Nigeria	ND	ND	Y	N
KD10	1	2015	Nigeria	ND	ND	Y	N
KD11	1	2015	Nigeria	ND	ND	Y	N
KD12	1	2015	Nigeria	ND	ND	Y	N
KD13	1	2015	Nigeria	ND	ND	Y	N
KD14	1	2015	Nigeria	ND	ND	Y	N
KD15	1	2015	Nigeria	ND	ND	Y	N
KD16	1	2015	Nigeria	ND	ND	Y	N
KD17	1	2015	Nigeria	ND	ND	Y	N
KD18	1	2015	Nigeria	ND	ND	Y	N
KD19	1	2015	Nigeria	ND	ND	Y	N
KD20	1	2015	Nigeria	ND	ND	Y	N
KD21	1	2015	Nigeria	ND	ND	Y	N
KD22	1	2015	Nigeria	ND	ND	Y	N
KD23	1	2015	Nigeria	ND	ND	Y	N
KD24	1	2015	Nigeria	ND	ND	Y	N

Table S2. – Continued from overleaf

Sample ID	Set No.	Year	City, Country	Location	Coordinates	Water	Soil
KD25	1	2015	Nigeria	ND	ND	Y	N
KD26	1	2015	Nigeria	ND	ND	Y	N
KD27	1	2015	Nigeria	ND	ND	Y	N
KD28	1	2015	Nigeria	ND	ND	Y	N
KD29	1	2015	Nigeria	ND	ND	Y	N
KD30	1	2015	Nigeria	ND	ND	Y	N
KD31	1	2015	Nigeria	ND	ND	Y	N
KD32	1	2015	Nigeria	ND	ND	Y	N
IR1	1	2015	Esfahan, Iran	ND	ND	Y	N
IR2	1	2015	Esfahan, Iran	ND	ND	N	Y
Cooling Tower A	1	2015	Leicester, United kingdom	ND	ND	Y	N
Cooling Tower B	1	2015	Leicester, United kingdom	ND	ND	Y	N
Spa water	1	2014	Bath, United kingdom	ND	ND	Y	N
Sacred Spring	1	2014	Bath, United kingdom	ND	ND	Y	N
GB Outlet	1	2014	Bath, United kingdom	ND	ND	Y	N
GB Intlet	1	2014	Bath, United kingdom	ND	ND	Y	N
GH1	1	2015	Leicester, United kingdom	General hospital	ND	Y	N
GH2	1	2015	Leicester, United kingdom	General hospital	ND	Y	N

133

Table S2. – Continued from overleaf

Sample ID	Set No.	Year	City, Country	Location	Coordinates	Water	Soil
GH3	1	2015	Leicester, United kingdom	General hospital	ND	Y	N
GH4	1	2015	Leicester, United kingdom	General hospital	ND	Y	N
GH5	1	2015	Leicester, United kingdom	General hospital	ND	Y	N
GH6	1	2015	Leicester, United kingdom	General hospital	ND	Y	N
GH7	1	2015	Leicester, United kingdom	General hospital	ND	Y	N
GH8	1	2015	Leicester, United kingdom	General hospital	ND	Y	N
GH9	1	2015	Leicester, United kingdom	General hospital	ND	Y	N
GH10	1	2015	Leicester, United kingdom	General hospital	ND	Y	N
GH11	1	2015	Leicester, United kingdom	General hospital	ND	Y	N
NatLei	1	2015	Leicester, United kingdom	University of Leicester	52.623594, -1.123939	Y	N
MSBLei	1	2015	Leicester, United kingdom	University of Leicester	52.624031, -1.124974	Y	N
AdrianLei	1	2015	Leicester, United kingdom	University of Leicester	52.623594, -1.123939	Y	N

Table S3. Table of Samples (set 2) used in this study for isolation of *Legionella* spp.

Sample ID	Set No.	Year	City, Country	Location	Coordinates	Water	Soil
1	2	2016	Leicester, United kingdom	Bradgate Park	52.687463, -1.205110	Y	N
2	2	2016	Leicester, United kingdom	Bradgate Park	52.687463, -1.205110	Y	N
3	2	2016	Leicester, United kingdom	Bradgate Park	52.687463, -1.205110	N	Y
4	2	2016	Leicester, United kingdom	Bradgate Park	52.696947, -1.200714	Y	N
5	2	2016	Leicester, United kingdom	Bradgate Park	52.696947, -1.200714	Y	N
6	2	2016	Leicester, United kingdom	Bradgate Park	52.696947, -1.200714	Y	N
7	2	2016	Leicester, United kingdom	Bradgate Park	52.696872, -1.200039	N	Y
B1	2	2016	Leicester, United kingdom	Bradgate Park	52.696938, -1.199721	Y	N
8	2	2016	Leicester, United kingdom	Watermeade County Park	52.673832, -1.108674	Y	N
9	2	2016	Leicester, United kingdom	Watermeade County Park	52.673832, -1.108674	N	Y
10	2	2016	Leicester, United kingdom	Watermeade County Park	52.673832, -1.108674	N	Y
B2	2	2016	Leicester, United kingdom	Watermeade County Park	52.673871, -1.108998	Y	N
11	2	2016	Leicester, United kingdom	Watermeade County Park	52.673871, -1.108998	Y	N
12	2	2016	Leicester, United kingdom	Watermeade County Park	52.673871, -1.108998	Y	N
B3	2	2016	Leicester, United kingdom	Watermeade County Park	52.672939, -1.113342	Y	N
B4	2	2016	Leicester, United kingdom	Watermeade County Park	52.672773, -1.114555	Y	N
13	2	2016	Leicester, United kingdom	Watermeade County Park	52.672773, -1.114555	Y	N
14	2	2016	Leicester, United kingdom	Watermeade County Park	52.672939, -1.113342	Y	N
15	2	2016	Leicester, United kingdom	Watermeade County Park	52.670502, -1.112030	N	Y
16	2	2016	Leicester, United kingdom	Watermeade County Park	52.670502, -1.112030	N	Y

Table S3. – Continued from overleaf

Sample ID	Set No.	Year	City, Country	Location	Coordinates	Water	Soil
17	2	2016	Leicester, United kingdom	Watermeade County Park	52.670710, -1.111469	N	Y
B5	2	2016	Leicester, United kingdom	Abbey Park	52.646948, -1.135588	Y	N
18	2	2016	Leicester, United kingdom	Abbey Park	52.646948, -1.135588	Y	N
19	2	2016	Leicester, United kingdom	Abbey Park	52.646948, -1.135588	N	Y
B6	2	2016	Leicester, United kingdom	Abbey Park	52.646828, -1.136359	Y	N
20	2	2016	Leicester, United kingdom	Abbey Park	52.646828, -1.136359	Y	N
21	2	2016	Leicester, United kingdom	Abbey Park	52.645759, -1.135759	N	Y
22	2	2016	Leicester, United kingdom	Abbey Park	52.645513, -1.135678	N	Y
23	2	2016	Leicester, United kingdom	Abbey Park	52.645646, -1.135104	Y	N
24	2	2016	Leicester, United kingdom	Abbey Park	52.645646, -1.135104	Y	N
25	2	2016	Leicester, United kingdom	Abbey Park	52.645646, -1.135104	Y	N
26	2	2016	Leicester, United kingdom	Abbey Park	52.645646, -1.135104	Y	N
27	2	2016	Leicester, United kingdom	Abbey Park	52.645646, -1.135104	Y	N
B7	2	2016	Leicester, United kingdom	Aylestone meadows	52.614014, -1.150801	Y	N
29	2	2016	Leicester, United kingdom	Aylestone meadows	52.614014, -1.150801	N	Y
30	2	2016	Leicester, United kingdom	Aylestone meadows	52.613298, -1.151139	Y	N
31	2	2016	Leicester, United kingdom	Aylestone meadows	52.613298, -1.151139	Y	N
32	2	2016	Leicester, United kingdom	Aylestone meadows	52.613298, -1.151139	N	Y
33	2	2016	Leicester, United kingdom	Aylestone meadows	52.612952, -1.151509	Y	N
34	2	2016	Leicester, United kingdom	Aylestone meadows	52.612952, -1.151509	N	Y

Table S3. – Continued from overleaf

Sample ID	Set No.	Year	City, Country	Location	Coordinates	Water	Soil
B8	2	2016	Nottingham, United kingdom	Marina Coldwick County Park, River Trent	52.944146, -1.0966265	Y	N
35	2	2016	Nottingham, United kingdom	Marina Coldwick County Park, River Trent	52.9489524, -1.0867406	Y	N
36	2	2016	Nottingham, United kingdom	Marina Coldwick County Park, River Trent	52.9489524, -1.0867406	Y	N
37	2	2016	Nottingham, United kingdom	Marina Coldwick County Park, River Trent	52.9489524, -1.0867406	N	Y
38	2	2016	Nottingham, United kingdom	Marina Coldwick County Park, River Trent	52.9489524, -1.0867406	N	Y
39	2	2016	Nottingham, United kingdom	Marina Coldwick County Park, River Trent	52.944146, -1.0966265	N	Y
40	2	2016	Nottingham, United kingdom	Marina Coldwick County Park, River Trent	52.9477258, -1.0947691	Y	N
41	2	2016	Nottingham, United kingdom	Marina Coldwick County Park, River Trent	52.9477258, -1.0947691	N	Y
B9	2	2016	Nottingham, United kingdom	Nottingham War Memorial path, River Trent	52.9326942, -1.1407918	Y	N
42	2	2016	Nottingham, United kingdom	Nottingham War Memorial path, River Trent	52.9326942, -1.1407918	Y	N
43	2	2016	Nottingham, United kingdom	Nottingham War Memorial path, River Trent	52.9326942, -1.1407918	N	Y

Table S3. – Continued from overleaf

Sample ID	Set No.	Year	City, Country	Location	Coordinates	Water	Soil
44	2	2016	Nottingham, United kingdom	Nottingham War Memorial path, River Trent	52.9322854, -1.1433526	Y	N
45	2	2016	Nottingham, United kingdom	Nottingham War Memorial path, River Trent	52.9322854, -1.1433526	Y	N
46	2	2016	Nottingham, United kingdom	Nottingham War Memorial path, River Trent	52.9322854, -1.1433526	N	Y
47	2	2016	Nottingham, United kingdom	Nottingham War Memorial path, River Trent	52.9324739, -1.1458628	Y	N
48	2	2016	Nottingham, United kingdom	Nottingham War Memorial path, River Trent	52.9324739, -1.1458628	Y	N
49	2	2016	Nottingham, United kingdom	Nottingham War Memorial path, River Trent	52.9324739, -1.1458628	N	Y
50	2	2016	Nottingham, United kingdom	Nottingham War Memorial path, River Trent	52.9337192, -1.1481735	Y	N
51	2	2016	Nottingham, United kingdom	Nottingham War Memorial path, River Trent	52.9337192, -1.1481735	Y	N
52	2	2016	Nottingham, United kingdom	Nottingham War Memorial path, River Trent	52.9337192, -1.1481735	N	Y
53	2	2016	Nottingham, United kingdom	Beeston Marina, River Trent	52.9152832, -1.1978652	Y	N
54	2	2016	Nottingham, United kingdom	Beeston Marina, River Trent	52.9152832, -1.1978652	N	Y
B10	2	2016	Nottingham, United kingdom	Beeston Marina, River Trent	52.9152832, -1.1978652	Y	N

Table S3. – Continued from overleaf

Sample ID	Set No.	Year	City, Country	Location	Coordinates	Water	Soil
55	2	2016	Nottingham, United kingdom	Beeston Marina, River Trent	52.9120615, -1.2036644	N	Y
56	2	2016	Nottingham, United kingdom	Beeston Marina, River Trent	52.9120615, -1.2036644	N	Y
57	2	2016	Nottingham, United kingdom	Beeston Marina, River Trent	52.9128907, -1.2064073	N	Y
58	2	2016	Nottingham, United kingdom	Beeston Marina, River Trent	52.9128907, -1.2064073	Y	N
59	2	2016	Nottingham, United kingdom	Beeston Marina, River Trent	52.9129733, -1.2077444	Y	N
60	2	2016	Nottingham, United kingdom	Beeston Marina, River Trent	52.9129733, -1.2077444	Y	N
B11	2	2016	Derby, United kingdom	Darley park, River Derwent	52.9349881, -1.4774818	Y	N
61	2	2016	Derby, United kingdom	Darley park, River Derwent	52.9339100, -1.4779371	Y	N
62	2	2016	Derby, United kingdom	Darley park, River Derwent	52.9339100, -1.4779371	N	Y
63	2	2016	Derby, United kingdom	Darley park, River Derwent	52.9349881, -1.4774818	N	Y
64	2	2016	Derby, United kingdom	Darley park, River Derwent	52.9356105, -1.4773299	Y	N
65	2	2016	Derby, United kingdom	Darley park, River Derwent	52.9356105, -1.4773299	N	Y
66	2	2016	Derby, United kingdom	Darley park, River Derwent	52.9389139, -1.4742524	Y	N
67	2	2016	Derby, United kingdom	Darley park, River Derwent	52.9389139, -1.4742524	N	Y
B12	2	2016	Derby, United kingdom	Derby River gardens, River Derwent	52.9230669, -1.4724647	Y	N
68	2	2016	Derby, United kingdom	Derby River gardens, River Derwent	52.9230669, -1.4724647	N	Y
69	2	2016	Derby, United kingdom	Derby River gardens, River Derwent	52.9233177, -1.4699280	Y	N

Table S3. – Continued from overleaf

Sample ID	Set No.	Year	City, Country	Location	Coordinates	Water	Soil
70	2	2016	Derby, United kingdom	Derby River gardens, River Derwent	52.9233177, -1.4699280	N	Y
71	2	2016	Derby, United kingdom	Derby River gardens, River Derwent	52.9236015, -1.4733475	Y	N
72	2	2016	Derby, United kingdom	Derby River gardens, River Derwent	52.9236015, -1.4733475	N	Y
73	2	2016	Derby, United kingdom	River Derwent	52.9358584, -1.4778211	Y	N
74	2	2016	Derby, United kingdom	River Derwent	52.9358584, -1.4778211	N	Y
B12	2	2016	Derby, United kingdom	River Derwent	52.9369435, -1.4778854	Y	N
75	2	2016	Derby, United kingdom	River Derwent	52.9369435, -1.4778854	N	Y
76	2	2016	Derby, United kingdom	River Derwent	52.9338556, -1.4784782	Y	N
77	2	2016	Derby, United kingdom	River Derwent	52.9338556, -1.4784782	N	Y
78	2	2016	Derby, United kingdom	River Derwent	52.9331247, -1.4788495	Y	N
79	2	2016	Derby, United kingdom	River Derwent	52.9331247, -1.4788495	N	Y
B13	2	2016	Coventry, United kingdom	Charterhouse field, River Sherbourne	52.4011107, -1.4960789	Y	N
80	2	2016	Coventry, United kingdom	Charterhouse field, River Sherbourne	52.4011107, -1.4960789	Y	N
81	2	2016	Coventry, United kingdom	Charterhouse field, River Sherbourne	52.4011107, -1.4960789	N	Y

Table S3. – Continued from overleaf

Sample ID	Set No.	Year	City, Country	Location	Coordinates	Water	Soil
82	2	2016	Coventry, United kingdom	Charterhouse field, River Sherbourne	52.3988136, -1.4933323	Y	N
83	2	2016	Coventry, United kingdom	Charterhouse field, River Sherbourne	52.3988136, -1.4933323	N	Y
84	2	2016	Coventry, United kingdom	Charterhouse field, River Sherbourne	52.3977406, -1.4922403	Y	N
85	2	2016	Coventry, United kingdom	Charterhouse field, River Sherbourne	52.3977406, -1.4922403	N	Y
86	2	2016	Coventry, United kingdom	Charterhouse field, River Sherbourne	52.3998494, -1.4947694	Y	N
87	2	2016	Coventry, United kingdom	Charterhouse field, River Sherbourne	52.3998494, -1.4947694	N	Y
B14	2	2016	Coventry, United kingdom	River Sherbourne	52.3873389, -1.4906839	Y	N
88	2	2016	Coventry, United kingdom	River Sherbourne	52.3873389, -1.4906839	N	Y
89	2	2016	Coventry, United kingdom	River Sherbourne	52.3866282, -1.4921739	Y	N
90	2	2016	Coventry, United kingdom	River Sherbourne	52.3866282, -1.4921739	N	Y
91	2	2016	Coventry, United kingdom	River Sherbourne	52.3868887, -1.4917464	Y	N
92	2	2016	Coventry, United kingdom	River Sherbourne	52.3868887, -1.4917464	N	Y
93	2	2016	Coventry, United kingdom	River Sherbourne	52.3871981, -1.4912200	Y	N
94	2	2016	Coventry, United kingdom	River Sherbourne	52.3871981, -1.4912200	Y	N
95	2	2016	Coventry, United kingdom	River Sherbourne	52.3871981, -1.4912200	N	Y

Table S3. – Continued from overleaf

Sample ID	Set No.	Year	City, Country	Location	Coordinates	Water	Soil
B15	2	2016	Coventry, United kingdom	River Sowe	52.374792, -1.5045912	Y	N
96	2	2016	Coventry, United kingdom	River Sowe	52.374792, -1.5045912	N	Y
97	2	2016	Coventry, United kingdom	River Sowe	52.3749010, -1.5040745	Y	N
98	2	2016	Coventry, United kingdom	River Sowe	52.3749010, -1.5040745	N	Y
99	2	2016	Coventry, United kingdom	River Sowe	52.374779, -1.504394	Y	N
100	2	2016	Coventry, United kingdom	River Sowe	52.374780, -1.504474	N	Y
101	2	2016	Coventry, United kingdom	River Sowe	52.3748488, -1.5034512	Y	N
102	2	2016	Coventry, United kingdom	River Sowe	52.3748488, -1.5034512	N	Y
B16	2	2016	Birmingham, United kingdom	River Tame	52.5251064, -1.9125102	Y	N
103	2	2016	Birmingham, United kingdom	River Tame	52.5251064, -1.9125102	N	Y
104	2	2016	Birmingham, United kingdom	River Tame	52.5261459, -1.9155284	Y	N
105	2	2016	Birmingham, United kingdom	River Tame	52.5261459, -1.9155284	N	Y
106	2	2016	Birmingham, United kingdom	River Tame	52.5256319, -1.9176071	Y	N
107	2	2016	Birmingham, United kingdom	River Tame	52.5256319, -1.9176071	N	Y
108	2	2016	Birmingham, United kingdom	River Tame	52.5255452, -1.9218443	Y	N
109	2	2016	Birmingham, United kingdom	River Tame	52.5255452, -1.9218443	N	Y
B17	2	2016	Birmingham, United kingdom	Hazelwell Park, River Rea	52.4298006, -1.9167880	Y	N
110	2	2016	Birmingham, United kingdom	Hazelwell Park, River Rea	52.4298006, -1.9167880	N	Y
111	2	2016	Birmingham, United kingdom	Hazelwell Park, River Rea	52.4309448, -1.9167237	Y	N
112	2	2016	Birmingham, United kingdom	Hazelwell Park, River Rea	52.4309448, -1.9167237	N	Y

Table S3. – Continued from overleaf

Sample ID	Set No.	Year	City, Country	Location	Coordinates	Water	Soil
113	2	2016	Birmingham, United kingdom	Hazelwell Park, River Rea	52.4290150, -1.9171300	Y	N
114	2	2016	Birmingham, United kingdom	Hazelwell Park, River Rea	52.4290150, -1.9171300	N	Y
115	2	2016	Birmingham, United kingdom	River Rea	52.4266312, -1.9182884	Y	N
116	2	2016	Birmingham, United kingdom	River Rea	52.4266312, -1.9182884	Y	N
B18	2	2016	Birmingham, United kingdom	Holders Lane and Pebble Mill fieldds, River Rea	52.4397609, -1.9116027	Y	N
117	2	2016	Birmingham, United kingdom	Holders Lane and Pebble Mill fieldds, River Rea	52.4397609, -1.9116027	N	Y
118	2	2016	Birmingham, United kingdom	Holders Lane and Pebble Mill fieldds, River Rea	52.4404010, -1.9116754	Y	N
119	2	2016	Birmingham, United kingdom	Holders Lane and Pebble Mill fieldds, River Rea	52.4404010, -1.9116754	N	Y
120	2	2016	Birmingham, United kingdom	Holders Lane and Pebble Mill fieldds, River Rea	52.4411325, -1.9117485	Y	N
121	2	2016	Birmingham, United kingdom	Holders Lane and Pebble Mill fieldds, River Rea	52.4411325, -1.9117485	N	Y
122	2	2016	Birmingham, United kingdom	Holders Lane and Pebble Mill fieldds, River Rea	52.4429334, -1.9118863	Y	N
123	2	2016	Birmingham, United kingdom	Holders Lane and Pebble Mill fieldds, River Rea	52.4429334, -1.9118863	N	Y

Table S3. – Continued from overleaf

Sample ID	Set No.	Year	City, Country	Location	Coordinates	Water	Soil
124	2	2016	Birmingham, United kingdom	Holders Lane and Pebble Mill fieldds, River Rea	52.4446139, -1.9121914	Y	N
125	2	2016	Birmingham, United kingdom	Holders Lane and Pebble Mill fieldds, River Rea	52.4446139, -1.9121914	N	Y
B19	2	2016	Birmingham, United kingdom	Amphitheatre, Canon Hill Park, Canoe pool	52.4541230, -1.9014770	Y	N
126	2	2016	Birmingham, United kingdom	Amphitheatre, Canon Hill Park, Canoe pool	52.4541230, -1.9014770	N	Y
127	2	2016	Birmingham, United kingdom	Amphitheatre, Canon Hill Park, Canoe pool	52.4542280, -1.9008705	Y	N
128	2	2016	Birmingham, United kingdom	Amphitheatre, Canon Hill Park, Canoe pool	52.4542280, -1.9008705	N	Y
129	2	2016	Birmingham, United kingdom	Amphitheatre, Canon Hill Park, Canoe pool	52.4536827, -1.9009767	Y	N
130	2	2016	Birmingham, United kingdom	Amphitheatre, Canon Hill Park, Canoe pool	52.4536827, -1.9009767	N	Y
131	2	2016	Birmingham, United kingdom	Amphitheatre, Canon Hill Park, Canoe pool	52.4531723, -1.9018917	Y	N
132	2	2016	Birmingham, United kingdom	Amphitheatre, Canon Hill Park, Canoe pool	52.4531723, -1.9018917	N	Y
133	2	2016	Birmingham, United kingdom	Amphitheatre, Canon Hill Park, Canoe pool	52.4534908, -1.9025113	Y	N

144

Table S3. – Continued from overleaf

Sample ID	Set No.	Year	City, Country	Location	Coordinates	Water	Soil
134	2	2016	Birmingham, United kingdom	Amphitheatre, Canon Hill Park, Canoe pool	52.4534908, -1.9025113	N	Y
28	2	2016	Leicester, United kingdom	Aylestone meadows	52.614014, -1.150801	N	Y

Table S4. The list of the *Legionella* spp. isolates present in our laboratory collection

Organism	Strain	Sample ID	Source	Provided by
L. pneumophila	**CLP**	NCTC 11192	Human lung	Purchased
L. pneumophila	**E1**	N/A	Roman Baths (Bath & UK)	In this study
L. pneumophila	**E2**	N/A	Roman Baths (Bath & UK)	In this study
L. pneumophila	**GBIn1**	N/A	Roman Baths inlet (Bath & Uk)	In this study
L. pneumophila	**GBIn3**	N/A	Roman Baths inlet (Bath & Uk)	In this study
L. pneumophila	**GBIn6**	N/A	Roman Baths inlet (Bath & Uk)	In this study
L. pneumophila	**GBO1**	N/A	Roman Baths outlet (Bath & Uk)	In this study
L. pneumophila	**GBO3**	N/A	Roman Baths outlet (Bath & Uk)	In this study
L. pneumophila	**GBO4**	N/A	Roman Baths outlet (Bath & Uk)	In this study
L. pneumophila	**GBO6**	N/A	Roman Baths outlet (Bath & Uk)	In this study
L. pneumophila	**SS4**	N/A	Roman Sacred Spring (Bath & Uk)	In this study
L. pneumophila	**SS3**	N/A	Roman Sacred Spring (Bath & Uk)	In this study
L. pneumophila	**SS6**	N/A	Roman Sacred Spring (Bath & Uk)	In this study
L. pneumophila	**GH1**	N/A	Hot-tap water (Leicester & UK)	In this study
L. pneumophila	**GH11**	N/A	Hot-tap water(Leicester & UK)	In this study
L. pneumophila	**GH18**	N/A	Hot-tap water(Leicester & UK)	In this study
Legionella micdadei	**LM**	NCTC 11371	ND	Purchased
L. pneumophila	**NL1**	N/A	Hot-tap water (Leicester & UK)	In this study
L. pneumophila	**NL4**	N/A	Hot-tap water (Leicester & UK)	In this study
L. pneumophila	**NL7**	N/A	Hot-tap water (Leicester & UK)	In this study

Table S4. – Continued from overleaf

Organism	Strain	Sample ID	Source	Provided by
L. pneumophila	**Phe1**	89957	Hospital calorifier	PHE
L. longbeachae	**Phe10**	B & Q top soil	Commercial top soil	PHE
L. pneumophila	**Phe11**	PO1318902-07	ND	PHE
L. pneumophila	**Phe12**	PO1323950-05	ND	PHE
L. pneumophila	**Phe13**	PO1324253-01	Drain point	PHE
L. pneumophila	**Phe14**	PO11406064-03	Ship tap	PHE
L. pneumophila	**Phe15**	6820-01	ND	PHE
L. pneumophila	**Phe16**	2075-01	ND	PHE
L. pneumophila	**Phe17**	1467-01	ND	PHE
L. pneumophila	**Phe18**	1753-03	ND	PHE
L. pneumophila	**Phe19**	7839-03	ND	PHE
L. pneumophila	**Phe2**	PO1311089-03	Ship shower	PHE
L. pneumophila	**Phe20**	9072-01	ND	PHE
L. pneumophila	**Phe21**	4111-02	ND	PHE
L. pneumophila	**Phe22**	56476	ND	PHE
L. pneumophila	**Phe23**	2851-01	ND	PHE
L. pneumophila	**Phe24**	1646-02	ND	PHE
L. pneumophila	**Phe25**	91157	ND	PHE
L. pneumophila	**Phe26**	0394-08	ND	PHE
L. pneumophila	**Phe27**	0394-06	ND	PHE

147

Table S4. – Continued from overleaf

Organism	Strain	Sample ID	Source	Provided by
L. pneumophila	**Phe28**	2579-03	ND	PHE
L. pneumophila	**Phe30**	7784-03	ND	PHE
Fluoribacter bozemanae (formerly *Legionella bozemanii*)	**Phe31**	NCTC 11368	ND	PHE
Legionella anisa	**Phe32**	NCTC 11974	ND	PHE
Legionella hackeliae	**Phe34**	NCTC 11979	ND	PHE
Legionella oakridgensis	**Phe35**	NCTC 11531	Industrial cooling tower in 1981	PHE
Legionella gormanii	**Phe36**	NCTC 11401	wet soil from a creek bank	PHE
Legionella longbeachae	**Phe37**	NCTC 11477	ND	PHE
Legionella feeleii	**Phe38**	ND	ND	PHE
Fluoribacter dumoffii (formerly *L. dumoffii*)	**Phe39**	NCTC 11370	ND	PHE
L. pneumophila	**Phe4**	PO1312868-06	Domestic tap	PHE
L. pneumophila	**Phe5**	PO1315682-08	Hospital mains	PHE
L. pneumophila	**Phe6**	PO1323309-06	Hospital tap	PHE
L. pneumophila	**Phe7**	PO1404336-01	Hospital tap	PHE
L. pneumophila	**Phe8**	PO1403754-04	Ship shower	PHE
Legionella sainthelensi	**Phe9**	PO1418133-37	Compost	PHE
L. pneumophila	**SS3**	N/A	Roman Sacred Spring (Bath & Uk)	PHE

Table S5. The list of the primers with respect to the Genebank accession number of the reference sequence used for allele assignment and SBT

Gene	Primer Name	Primer Sequence	Fragment Size(bp)	GenBank accession no. of reference sequence	Reference
flaA	flaA-587F	5'-GCGTATTGCTCAAAATACTG-3'	414	X83232	(Mentasti et al., 2012)
	flaA-960R	5'-CCATTAATCGTTAAGTTGTAGG-3'			
pilE	pilE-35F	5'-CACAATCGGATGGAACACAAACTA-3'	460	AF048690	(Gaia et al., 2005)
	pilE-453R	5'-GCTGGCGCCACTCGGTATCT-3'			
asd	asd-511F	5'-CCCTAATTGCTCTACCATTCAGATG-3'	576	AF034213	Gaia2005
	asd-1039R	5'-CGAATGTTATCTGCGACTATCCAC-3'			
mip	mip-74F	5'-GCTGCAACCGATGCCAC-3'	559	AJ496269	Gaia2005
	mip-595R	5'-CATATGCAAGACCTGAGGGAAC-3'			
mompS	mompS-450F	5'-TTGACCATGAGTGGGATTGG-3'	711	AF078136	MentastiM.;Frya
	mompS-1116R	5'-TGGATAAATTATCCAGCCGGACTTC-3'			Gaia2005
mompS	mompS-1015R**	5'-CAGAAGCTGCGAAATCAG-3'		AF078136	Gaia2003
proA	proA-1107F	5'-GATCGCCAATGCAATTAG-3'	481	M32884	Gaia2003
	proA-1553R	5'-ACCATAACATCAAAAGCC-3'			
neuA	neuA-196F	5'-CCGTTCAATATGGGGCTTCAG-3'	459	AJ007311	(Ratzow et al., 2007)
	neuA-634R	5'-CGATGTCGATGGATTCACTAATAC-3'			
neuAH	neuAH-L	5'-ATCCAGCAGTTTTAMAAATTTAGG-3'	791-794*	FN256429 & FR750546	(Mentasti et al., 2014)
	neuAH-R	5'-TGGCTGCATAAAYTAATTCTTTAGCCA-3'			

* Due to the INDEL of a GAA triplet

** mompS-1015R instead of mompS-1116R is used for the reverse sequencing reaction of the mompS target

Table S6. The list of the primers for PCR and qPCR assays

Gene	Primer Name	Primer Sequence	Fragment Size(bp)	Reference
mip	mipLesnsens	5′-ATGAAGATGAAAATTGGTGACTGCAG-3′	607	(Chaabna et al., 2013)
	mipLensrev	5′-CAACGCTACGTGGGCCATA-3′		
ssrA	LsppF	5′-GGCGACCTGGCTTC-3′	101	(Collins et al., 2015)
	LsppR	5′-GGTCATCGTTTGCATTTATATTTA-3′		
	LsppP	5′-FAM-ACGTGGGTTGCAA-MGBNFQ*-3′		
mip	LpnF	5′-TTGTCTTATAGCATTGGTGCCG-3′	115	(Collins et al., 2015)
	LpnR	5′-CCAATTGAGCGCCACTCATAG-3′		
	LpnP	5′-Cy3-CGGAAGCAATGGCTAAAGGCATGCA-BHQ2**-3′		
wzm	Lsg1F	5′-TGCCTCTGGCTTTGCAGTTA-3′	70	(Collins et al., 2015)
	Lsg1R	5′-CACACAGGCACAGCAGAAACA-3′		
	Lsg1P	5′-VIC-TTTATTACTCCACTCCAGCCGAT-MGBNFQ*-3′		

* Minor grove bender nonfluorescent quencher

** Black Hole Quencher-2

150

Table S7. Table of samples were used in well assay

Samples Group No.	Source
Sample 1	All the Midlands' positive samples for *wzm* genes
Sample 2	All the environmental miscellaneous samples
Sample 3	All the Midlands' positive samples for *mip* genes
Sample 4	All the Midlands' positive samples for *ssrA* genes
Sample 5	Tap water sample from University of Leicester, positive for *wzm* genes
Sample 6	Tap water sample from Leicester General Hospital, Leicester
Sample 7	Environmental water samples from Nigeria
Sample 8	Water samples from cooling tower
Sample 9	The Midlands water sample negative for *Legionella* spp.
Sample 10	Environmental water sample from Bath, UK, positive for *mip* genes
Sample 11	Environmental water sample from Public Health England, positive for *Legionella* spp.
Sample 12	Environmental water sample from Uganda, East Africa, treated with chloroform
Sample 13	Environmental water sample, from Uganda, East Africa, no treatment
Sample 14	Swage sample from London, UK

Table S8. Table of Samples used in this study for isolation of bacteriophages against *Legionella* spp.

Sample ID	City, Country	Location	Coordinates	Water	Soil	Samples Group No.
Lake-S1	Chafford Hundred, Essex, United kingdom	Chafford Gorges Nature Park	51.489976, 0.301790	Y	N	2
LakeS2	Chafford Hundred, Essex, United kingdom	Chafford Gorges Nature Park	51.488543, 0302853	Y	N	2
Stream U-S3	Leicester, United kingdom	Loddington Road, Tilton on the Hill	ND	Y	N	2
Stream Lower-S4	Leicester, United kingdom	Loddington Road, Tilton on the Hill	ND	Y	N	2
Puddle 1-S5	Leicester, United kingdom	Loddington Road, Tilton on the Hill	ND	Y	N	2
Puddle 2-S6	Leicester, United kingdom	Loddington Road, Tilton on the Hill	ND	Y	N	2
Drain-S7	Leicester, United kingdom	Loddington Road, Tilton on the Hill	ND	Y	N	2
Muldy pud-S8	Leicester, United kingdom	Loddington Road, Tilton on the Hill	ND	Y	N	2
S10	Iraq	ND	ND	Y	N	2
Shower head-S16	Leicester, United kingdom	Queens Road	ND	Y	N	2
Biofilm-WAN-1	Leicestershire, United kingdom	Wanlip	ND	Y	N	2
Biofilm-WAN-2	Leicestershire, United kingdom	Wanlip	ND	Y	N	2
River-WAN-3	Leicestershire, United kingdom	Wanlip	ND	Y	N	2

Table S8. – Continued from overleaf

Sample ID	City, Country	Location	Coordinates	Water	Soil	Samples Group No.
Biofilm-WAN-5	Leicestershire, United kingdom	Wanlip	ND	Y	N	2
River-WAN-6	Leicestershire, United kingdom	Wanlip	ND	Y	N	2
River-WAN-7	Leicestershire, United kingdom	Wanlip	ND	Y	N	2
River-WAN-8	Leicestershire, United kingdom	Wanlip	ND	Y	N	2
Biofilm-WAN-9	Leicestershire, United kingdom	Wanlip	ND	Y	N	2
River-WAN-10	Leicestershire, United kingdom	Wanlip	ND	Y	N	2
River-WAN-11	Leicestershire, United kingdom	Wanlip	ND	Y	N	2
H-A	Hayling Island, Hampshire, United kingdom	Hayling Seafront	ND	Y	N	2
H-B	Hayling Island, Hampshire, United kingdom	Shore mud surface	ND	Y	N	2
H-C	Hayling Island, Hampshire, United kingdom	Small ditch, 2nd ditch along Hayling Billy trail	ND	Y	N	2
H-D	Hayling Island, Hampshire, United kingdom	Fish tank, 21-23ºC	ND	Y	N	2
H-E	Hayling Island, Hampshire, United kingdom	Water tank no 1 from East end in Seafront	ND	Y	N	2
H-F	Hayling Island, Hampshire, United kingdom	Water tank no 2 in Seafront	ND	Y	N	2
H-G	Hayling Island, Hampshire, United kingdom	Water tank no 3 in Seafront	ND	Y	N	2

153

Table S8. – Continued from overleaf

Sample ID	City, Country	Location	Coordinates	Water	Soil	Samples Group No.
H-H	Hayling Island, Hampshire, United kingdom	Water tank no 4 in Seafront	ND	Y	N	2
H-I	Hayling Island, Hampshire, United kingdom	Top of porch in Seafront 51 (in stainless steel container)	ND	Y	N	2
H-J	Hayling Island, Hampshire, United kingdom	Meridian centre, Havant, On top of air con machine labelled Robt Dyas	ND	Y	N	2
H-K	Hayling Island, Hampshire, United kingdom	Housing Assn flats Havant air con unit tray nearest road	ND	Y	N	2
H-L	Hayling Island, Hampshire, United kingdom	Meridian centre Havant drip from machine labelled O2	ND	Y	N	2
H-M	Hayling Island, Hampshire, United kingdom	Meridian centre Havant machine labelled Genix. drip from base	ND	Y	N	2
H-N	Hayling Island, Hampshire, United kingdom	Havant spring (old spring in centre of town)	ND	Y	N	2
IK2	Iraq	ND	ND	Y	N	2
SK9	Iraq	ND	ND	Y	N	2
K2	Iraq	ND	ND	Y	N	2
SK4	Iraq	ND	ND	Y	N	2
SK3	Iraq	ND	ND	Y	N	2

154

Table S8. – Continued from overleaf

Sample ID	City, Country	Location	Coordinates	Water	Soil	Samples Group No.
SK6	Iraq	ND	ND	Y	N	2
Ski	Iraq	ND	ND	Y	N	2
SK5	Iraq	ND	ND	Y	N	2
SK7	Iraq	ND	ND	Y	N	2
SK8	Iraq	ND	ND	Y	N	2
Kenya1	Kenya	ND	ND	Y	N	2
Kenya2	Kenya	ND	ND	Y	N	2
Kenya3	Kenya	ND	ND	Y	N	2
M-UK	Margate, United kingdom	Margate Sea	ND	Y	N	2
KD1	Nigeria	ND	ND	Y	N	7
KD2	Nigeria	ND	ND	Y	N	7
KD3	Nigeria	ND	ND	Y	N	7
KD4	Nigeria	ND	ND	Y	N	7
KD5	Nigeria	ND	ND	Y	N	7
KD6	Nigeria	ND	ND	Y	N	7
KD7	Nigeria	ND	ND	Y	N	7
KD8	Nigeria	ND	ND	Y	N	7
KD10	Nigeria	ND	ND	Y	N	7
KD11	Nigeria	ND	ND	Y	N	7
KD12	Nigeria	ND	ND	Y	N	7

155

Table S8. – Continued from overleaf

Sample ID	City, Country	Location	Coordinates	Water	Soil	Samples Group No.
KD13	Nigeria	ND	ND	Y	N	7
KD14	Nigeria	ND	ND	Y	N	7
KD15	Nigeria	ND	ND	Y	N	7
KD16	Nigeria	ND	ND	Y	N	7
KD17	Nigeria	ND	ND	Y	N	7
KD18	Nigeria	ND	ND	Y	N	7
KD19	Nigeria	ND	ND	Y	N	7
KD20	Nigeria	ND	ND	Y	N	7
KD21	Nigeria	ND	ND	Y	N	7
KD22	Nigeria	ND	ND	Y	N	7
KD23	Nigeria	ND	ND	Y	N	7
KD24	Nigeria	ND	ND	Y	N	7
KD25	Nigeria	ND	ND	Y	N	7
KD26	Nigeria	ND	ND	Y	N	7
KD27	Nigeria	ND	ND	Y	N	7
KD28	Nigeria	ND	ND	Y	N	7
KD29	Nigeria	ND	ND	Y	N	7
KD30	Nigeria	ND	ND	Y	N	7
KD31	Nigeria	ND	ND	Y	N	7
KD32	Nigeria	ND	ND	Y	N	7

156

Table S8. – Continued from overleaf

Sample ID	City, Country	Location	Coordinates	Water	Soil	Samples Group No.
IR1	Esfahan, Iran	ND	ND	Y	N	2
IR2	Esfahan, Iran	ND	ND	Y	N	2
Cooling Tower A	Leicester, United kingdom	ND	ND	Y	N	8
Cooling Tower B	Leicester, United kingdom	ND	ND	Y	N	8
Spa water	Bath, United kingdom	ND	ND	Y	N	10
Sacred Spring	Bath, United kingdom	ND	ND	Y	N	10
GB Outlet	Bath, United kingdom	ND	ND	Y	N	10
GB Intlet	Bath, United kingdom	ND	ND	Y	N	10
GH1	Leicester, United kingdom	General hospital	ND	Y	N	6
GH2	Leicester, United kingdom	General hospital	ND	Y	N	6
GH3	Leicester, United kingdom	General hospital	ND	Y	N	6
GH4	Leicester, United kingdom	General hospital	ND	Y	N	6
GH5	Leicester, United kingdom	General hospital	ND	Y	N	6
GH6	Leicester, United kingdom	General hospital	ND	Y	N	6
GH7	Leicester, United kingdom	General hospital	ND	Y	N	6
GH8	Leicester, United kingdom	General hospital	ND	Y	N	6
GH9	Leicester, United kingdom	General hospital	ND	Y	N	6
GH10	Leicester, United kingdom	General hospital	ND	Y	N	6
GH11	Leicester, United kingdom	General hospital	ND	Y	N	6
1	Leicester, United kingdom	Bradgate Park	52.687463, -1.205110	Y	N	4

157

Table S8. – Continued from overleaf

Sample ID	City, Country	Location	Coordinates	Water	Soil	Samples Group No.
2	Leicester, United kingdom	Bradgate Park	52.687463, -1.205110	Y	N	4
3	Leicester, United kingdom	Bradgate Park	52.687463, -1.205110	N	Y	4
4	Leicester, United kingdom	Bradgate Park	52.696947, -1.200714	Y	N	4
5	Leicester, United kingdom	Bradgate Park	52.696947, -1.200714	Y	N	4
6	Leicester, United kingdom	Bradgate Park	52.696947, -1.200714	Y	N	4
7	Leicester, United kingdom	Bradgate Park	52.696872, -1.200039	N	Y	4
B1	Leicester, United kingdom	Bradgate Park	52.696938, -1.199721	Y	N	4
8	Leicester, United kingdom	Watermeade County Park	52.673832, -1.108674	Y	N	4
9	Leicester, United kingdom	Watermeade County Park	52.673832, -1.108674	N	Y	4
10	Leicester, United kingdom	Watermeade County Park	52.673832, -1.108674	N	Y	4
B2	Leicester, United kingdom	Watermeade County Park	52.673871, -1.108998	Y	N	4
11	Leicester, United kingdom	Watermeade County Park	52.673871, -1.108998	Y	N	4
12	Leicester, United kingdom	Watermeade County Park	52.673871, -1.108998	Y	N	4
B3	Leicester, United kingdom	Watermeade County Park	52.672939, -1.113342	Y	N	4
B4	Leicester, United kingdom	Watermeade County Park	52.672773, -1.114555	Y	N	4
13	Leicester, United kingdom	Watermeade County Park	52.672773, -1.114555	Y	N	4
14	Leicester, United kingdom	Watermeade County Park	52.672939, -1.113342	Y	N	4
15	Leicester, United kingdom	Watermeade County Park	52.670502, -1.112030	N	Y	4
16	Leicester, United kingdom	Watermeade County Park	52.670502, -1.112030	N	Y	4
17	Leicester, United kingdom	Watermeade County Park	52.670710, -1.111469	N	Y	4

158

Table S8. – Continued from overleaf

Sample ID	City, Country	Location	Coordinates	Water	Soil	Samples Group No.
B5	Leicester, United kingdom	Abbey Park	52.646948, -1.135588	Y	N	4
18	Leicester, United kingdom	Abbey Park	52.646948, -1.135588	Y	N	1
19	Leicester, United kingdom	Abbey Park	52.646948, -1.135588	N	Y	4
B6	Leicester, United kingdom	Abbey Park	52.646828, -1.136359	Y	N	4
20	Leicester, United kingdom	Abbey Park	52.646828, -1.136359	Y	N	3
21	Leicester, United kingdom	Abbey Park	52.645759, -1.135759	N	Y	4
22	Leicester, United kingdom	Abbey Park	52.645513, -1.135678	N	Y	4
23	Leicester, United kingdom	Abbey Park	52.645646, -1.135104	Y	N	4
24	Leicester, United kingdom	Abbey Park	52.645646, -1.135104	Y	N	4
25	Leicester, United kingdom	Abbey Park	52.645646, -1.135104	Y	N	4
26	Leicester, United kingdom	Abbey Park	52.645646, -1.135104	Y	N	4
27	Leicester, United kingdom	Abbey Park	52.645646, -1.135104	Y	N	4
B7	Leicester, United kingdom	Aylestone meadows	52.614014, -1.150801	Y	N	4
29	Leicester, United kingdom	Aylestone meadows	52.614014, -1.150801	N	Y	4
30	Leicester, United kingdom	Aylestone meadows	52.613298, -1.151139	Y	N	4
31	Leicester, United kingdom	Aylestone meadows	52.613298, -1.151139	Y	N	3
32	Leicester, United kingdom	Aylestone meadows	52.613298, -1.151139	N	Y	4
33	Leicester, United kingdom	Aylestone meadows	52.612952, -1.151509	Y	N	4
34	Leicester, United kingdom	Aylestone meadows	52.612952, -1.151509	N	Y	1
B8	Nottingham, United kingdom	Marina Coldwick County Park, River Trent	52.944146, -1.0966265	Y	N	4

Table S8. – Continued from overleaf

Sample ID	City, Country	Location	Coordinates	Water	Soil	Samples Group No.
35	Nottingham, United kingdom	Marina Coldwick County Park, River Trent	52.9489524, -1.0867406	Y	N	4
36	Nottingham, United kingdom	Marina Coldwick County Park, River Trent	52.9489524, -1.0867406	Y	N	4
37	Nottingham, United kingdom	Marina Coldwick County Park, River Trent	52.9489524, -1.0867406	N	Y	1
38	Nottingham, United kingdom	Marina Coldwick County Park, River Trent	52.9489524, -1.0867406	N	Y	4
39	Nottingham, United kingdom	Marina Coldwick County Park, River Trent	52.944146, -1.0966265	N	Y	4
40	Nottingham, United kingdom	Marina Coldwick County Park, River Trent	52.9477258, -1.0947691	Y	N	4
41	Nottingham, United kingdom	Marina Coldwick County Park, River Trent	52.9477258, -1.0947691	N	Y	4
B9	Nottingham, United kingdom	Nottingham War Memorial path, River Trent	52.9326942, -1.1407918	Y	N	4
42	Nottingham, United kingdom	Nottingham War Memorial path, River Trent	52.9326942, -1.1407918	Y	N	3
43	Nottingham, United kingdom	Nottingham War Memorial path, River Trent	52.9326942, -1.1407918	N	Y	4
44	Nottingham, United kingdom	Nottingham War Memorial path, River Trent	52.9322854, -1.1433526	Y	N	4

Table S8. – Continued from overleaf

Sample ID	City, Country	Location	Coordinates	Water	Soil	Samples Group No.
45	Nottingham, United kingdom	Nottingham War Memorial path, River Trent	52.9322854, -1.1433526	Y	N	1
46	Nottingham, United kingdom	Nottingham War Memorial path, River Trent	52.9322854, -1.1433526	N	Y	4
47	Nottingham, United kingdom	Nottingham War Memorial path, River Trent	52.9324739, -1.1458628	Y	N	4
48	Nottingham, United kingdom	Nottingham War Memorial path, River Trent	52.9324739, -1.1458628	Y	N	1
49	Nottingham, United kingdom	Nottingham War Memorial path, River Trent	52.9324739, -1.1458628	N	Y	4
50	Nottingham, United kingdom	Nottingham War Memorial path, River Trent	52.9337192, -1.1481735	Y	N	4
51	Nottingham, United kingdom	Nottingham War Memorial path, River Trent	52.9337192, -1.1481735	Y	N	1
52	Nottingham, United kingdom	Nottingham War Memorial path, River Trent	52.9337192, -1.1481735	N	Y	4
53	Nottingham, United kingdom	Beeston Marina, River Trent	52.9152832, -1.1978652	Y	N	4
54	Nottingham, United kingdom	Beeston Marina, River Trent	52.9152832, -1.1978652	N	Y	4
B10	Nottingham, United kingdom	Beeston Marina, River Trent	52.9152832, -1.1978652	Y	N	4
55	Nottingham, United kingdom	Beeston Marina, River Trent	52.9120615, -1.2036644	N	Y	3
56	Nottingham, United kingdom	Beeston Marina, River Trent	52.9120615, -1.2036644	N	Y	4

Table S8. – Continued from overleaf

Sample ID	City, Country	Location	Coordinates	Water	Soil	Samples Group No.
57	Nottingham, United kingdom	Beeston Marina, River Trent	52.9128907, -1.2064073	N	Y	4
58	Nottingham, United kingdom	Beeston Marina, River Trent	52.9128907, -1.2064073	Y	N	4
59	Nottingham, United kingdom	Beeston Marina, River Trent	52.9129733, -1.2077444	Y	N	4
60	Nottingham, United kingdom	Beeston Marina, River Trent	52.9129733, -1.2077444	Y	N	4
B11	Derby, United kingdom	Darley park, River Derwent	52.9349881, -1.4774818	Y	N	4
61	Derby, United kingdom	Darley park, River Derwent	52.9339100, -1.4779371	Y	N	4
62	Derby, United kingdom	Darley park, River Derwent	52.9339100, -1.4779371	N	Y	4
63	Derby, United kingdom	Darley park, River Derwent	52.9349881, -1.4774818	N	Y	4
64	Derby, United kingdom	Darley park, River Derwent	52.9356105, -1.4773299	Y	N	4
65	Derby, United kingdom	Darley park, River Derwent	52.9356105, -1.4773299	N	Y	4
66	Derby, United kingdom	Darley park, River Derwent	52.9389139, -1.4742524	Y	N	3
67	Derby, United kingdom	Darley park, River Derwent	52.9389139, -1.4742524	N	Y	4
B12	Derby, United kingdom	Derby River gardens, River Derwent	52.9230669, -1.4724647	Y	N	4
68	Derby, United kingdom	Derby River gardens, River Derwent	52.9230669, -1.4724647	N	Y	4
69	Derby, United kingdom	Derby River gardens, River Derwent	52.9233177, -1.4699280	Y	N	4
70	Derby, United kingdom	Derby River gardens, River Derwent	52.9233177, -1.4699280	N	Y	4

Table S8. – Continued from overleaf

Sample ID	City, Country	Location	Coordinates	Water	Soil	Samples Group No.
71	Derby, United kingdom	Derby River gardens, River Derwent	52.9236015, -1.4733475	Y	N	4
72	Derby, United kingdom	Derby River gardens, River Derwent	52.9236015, -1.4733475	N	Y	4
73	Derby, United kingdom	River Derwent	52.9358584, -1.4778211	Y	N	4
74	Derby, United kingdom	River Derwent	52.9358584, -1.4778211	N	Y	4
B12	Derby, United kingdom	River Derwent	52.9369435, -1.4778854	Y	N	4
75	Derby, United kingdom	River Derwent	52.9369435, -1.4778854	N	Y	4
763	Derby, United kingdom	River Derwent	52.9338556, -1.4784782	Y	N	1
77	Derby, United kingdom	River Derwent	52.9338556, -1.4784782	N	Y	4
78	Derby, United kingdom	River Derwent	52.9331247, -1.4788495	Y	N	4
79	Derby, United kingdom	River Derwent	52.9331247, -1.4788495	N	Y	4
B13	Coventry, United kingdom	Charterhouse field, River Sherbourne	52.4011107, -1.4960789	Y	N	3
80	Coventry, United kingdom	Charterhouse field, River Sherbourne	52.4011107, -1.4960789	Y	N	3
81	Coventry, United kingdom	Charterhouse field, River Sherbourne	52.4011107, -1.4960789	N	Y	4
82	Coventry, United kingdom	Charterhouse field, River Sherbourne	52.3988136, -1.4933323	Y	N	4

Table S8. – Continued from overleaf

Sample ID	City, Country	Location	Coordinates	Water	Soil	Samples Group No.
83	Coventry, United kingdom	Charterhouse field, River Sherbourne	52.3988136, -1.4933323	N	Y	4
84	Coventry, United kingdom	Charterhouse field, River Sherbourne	52.3977406, -1.4922403	Y	N	1
85	Coventry, United kingdom	Charterhouse field, River Sherbourne	52.3977406, -1.4922403	N	Y	4
86	Coventry, United kingdom	Charterhouse field, River Sherbourne	52.3998494, -1.4947694	Y	N	4
87	Coventry, United kingdom	Charterhouse field, River Sherbourne	52.3998494, -1.4947694	N	Y	1
B14	Coventry, United kingdom	River Sherbourne	52.3873389, -1.4906839	Y	N	4
88	Coventry, United kingdom	River Sherbourne	52.3873389, -1.4906839	N	Y	4
89	Coventry, United kingdom	River Sherbourne	52.3866282, -1.4921739	Y	N	3
90	Coventry, United kingdom	River Sherbourne	52.3866282, -1.4921739	N	Y	4
91	Coventry, United kingdom	River Sherbourne	52.3868887, -1.4917464	Y	N	4
92	Coventry, United kingdom	River Sherbourne	52.3868887, -1.4917464	N	Y	4
93	Coventry, United kingdom	River Sherbourne	52.3871981, -1.4912200	Y	N	4
94	Coventry, United kingdom	River Sherbourne	52.3871981, -1.4912200	Y	N	1
95	Coventry, United kingdom	River Sherbourne	52.3871981, -1.4912200	N	Y	4
B15	Coventry, United kingdom	River Sowe	52.374792, -1.5045912	Y	N	1
96	Coventry, United kingdom	River Sowe	52.374792, -1.5045912	N	Y	4

Table S8. – Continued from overleaf

Sample ID	City, Country	Location	Coordinates	Water	Soil	Samples Group No.
97	Coventry, United kingdom	River Sowe	52.3749010, -1.5040745	Y	N	3
98	Coventry, United kingdom	River Sowe	52.3749010, -1.5040745	N	Y	4
99	Coventry, United kingdom	River Sowe	52.374779, -1.504394	Y	N	4
100	Coventry, United kingdom	River Sowe	52.374780, -1.504474	N	Y	3
101	Coventry, United kingdom	River Sowe	52.3748488, -1.5034512	Y	N	4
102	Coventry, United kingdom	River Sowe	52.3748488, -1.5034512	N	Y	3
B16	Birmingham, United kingdom	River Tame	52.5251064, -1.9125102	Y	N	4
103	Birmingham, United kingdom	River Tame	52.5251064, -1.9125102	N	Y	4
104	Birmingham, United kingdom	River Tame	52.5261459, -1.9155284	Y	N	4
105	Birmingham, United kingdom	River Tame	52.5261459, -1.9155284	N	Y	4
106	Birmingham, United kingdom	River Tame	52.5256319, -1.9176071	Y	N	4
107	Birmingham, United kingdom	River Tame	52.5256319, -1.9176071	N	Y	3
108	Birmingham, United kingdom	River Tame	52.5255452, -1.9218443	Y	N	4
109	Birmingham, United kingdom	River Tame	52.5255452, -1.9218443	N	Y	4
B17	Birmingham, United kingdom	Hazelwell Park, River Rea	52.4298006, -1.9167880	Y	N	3
110	Birmingham, United kingdom	Hazelwell Park, River Rea	52.4298006, -1.9167880	N	Y	4
111	Birmingham, United kingdom	Hazelwell Park, River Rea	52.4309448, -1.9167237	Y	N	4
112	Birmingham, United kingdom	Hazelwell Park, River Rea	52.4309448, -1.9167237	N	Y	4
113	Birmingham, United kingdom	Hazelwell Park, River Rea	52.4290150, -1.9171300	Y	N	4
114	Birmingham, United kingdom	Hazelwell Park, River Rea	52.4290150, -1.9171300	N	Y	4

Table S8. – Continued from overleaf

Sample ID	City, Country	Location	Coordinates	Water	Soil	Samples Group No.
115	Birmingham, United kingdom	River Rea	52.4266312, -1.9182884	Y	N	3
116	Birmingham, United kingdom	River Rea	52.4266312, -1.9182884	Y	N	4
B18	Birmingham, United kingdom	Holders Lane and Pebble Mill fieldds, River Rea	52.4397609, -1.9116027	Y	N	3
117	Birmingham, United kingdom	Holders Lane and Pebble Mill fieldds, River Rea	52.4397609, -1.9116027	N	Y	4
118	Birmingham, United kingdom	Holders Lane and Pebble Mill fieldds, River Rea	52.4404010, -1.9116754	Y	N	4
119	Birmingham, United kingdom	Holders Lane and Pebble Mill fieldds, River Rea	52.4404010, -1.9116754	N	Y	4
120	Birmingham, United kingdom	Holders Lane and Pebble Mill fieldds, River Rea	52.4411325, -1.9117485	Y	N	4
121	Birmingham, United kingdom	Holders Lane and Pebble Mill fieldds, River Rea	52.4411325, -1.9117485	N	Y	4
122	Birmingham, United kingdom	Holders Lane and Pebble Mill fieldds, River Rea	52.4429334, -1.9118863	Y	N	4
123	Birmingham, United kingdom	Holders Lane and Pebble Mill fieldds, River Rea	52.4429334, -1.9118863	N	Y	3
124	Birmingham, United kingdom	Holders Lane and Pebble Mill fieldds, River Rea	52.4446139, -1.9121914	Y	N	4
125	Birmingham, United kingdom	Holders Lane and Pebble Mill fieldds, River Rea	52.4446139, -1.9121914	N	Y	4

Table S8. – Continued from overleaf

Sample ID	City, Country	Location	Coordinates	Water	Soil	Samples Group No.
B19	Birmingham, United kingdom	Amphitheatre, Canon Hill Park, Canoe pool	52.4541230, -1.9014770	Y	N	4
126	Birmingham, United kingdom	Amphitheatre, Canon Hill Park, Canoe pool	52.4541230, -1.9014770	N	Y	4
127	Birmingham, United kingdom	Amphitheatre, Canon Hill Park, Canoe pool	52.4542280, -1.9008705	Y	N	3
128	Birmingham, United kingdom	Amphitheatre, Canon Hill Park, Canoe pool	52.4542280, -1.9008705	N	Y	4
129	Birmingham, United kingdom	Amphitheatre, Canon Hill Park, Canoe pool	52.4536827, -1.9009767	Y	N	4
130	Birmingham, United kingdom	Amphitheatre, Canon Hill Park, Canoe pool	52.4536827, -1.9009767	N	Y	4
131	Birmingham, United kingdom	Amphitheatre, Canon Hill Park, Canoe pool	52.4531723, -1.9018917	Y	N	4
132	Birmingham, United kingdom	Amphitheatre, Canon Hill Park, Canoe pool	52.4531723, -1.9018917	N	Y	4
133	Birmingham, United kingdom	Amphitheatre, Canon Hill Park, Canoe pool	52.4534908, -1.9025113	Y	N	4
134	Birmingham, United kingdom	Amphitheatre, Canon Hill Park, Canoe pool	52.4534908, -1.9025113	N	Y	1
Birm	Birmingham, United kingdom	Hazelwell Park, River Rea	52.4298006, -1.9167880	Y	N	4
NatLei	Leicester, United kingdom	University of Leicester	52.623594, -1.123939	Y	N	5

Table S8. – Continued from overleaf

Sample ID	City, Country	Location	Coordinates	Water	Soil	Samples Group No.
MSBLei	Leicester, United kingdom	University of Leicester	52.624031, -1.124974	Y	N	5
AdrianLei	Leicester, United kingdom	University of Leicester	52.623594, -1.123939	Y	N	5
PHE_01/30_1	United kingdom	Hot tap, commercial building water distribution system	ND	Y	N	11
PHE_01/30_2	United kingdom	Cold tap, commercial building water distribution system	ND	Y	N	11
PHE_01/21_H	United kingdom	Hot tap, commercial building water distribution system	ND	Y	N	11
PHE_01/21_C	United kingdom	Cold tap, commercial building water distribution system	ND	Y	N	11
PHE_01/210_H	United kingdom	Hot tap, commercial building water distribution system	ND	Y	N	11
PHE_01/210_C	United kingdom	Cold tap, commercial building water distribution system	ND	Y	N	11
PHE_01/210_H2	United kingdom	Hot tap, commercial building water distribution system	ND	Y	N	11
PHE_ENV_R	United kingdom	Blended tap (with TMV), commercial building water distribution system	ND	Y	N	11

Table S8. – Continued from overleaf

Sample ID	City, Country	Location	Coordinates	Water	Soil	Samples Group No.
PHE_ENV_L	United kingdom	Blended tap (with TMV), commercial building water distribution system	ND	Y	N	11
PHE_01/24a	United kingdom	Blended tap (with TMV), commercial building water distribution system	ND	Y	N	11
U2	Uganda	Sewage 1	ND	Y	N	12
U4	Uganda	Sedimentation Tank	ND	Y	N	12
U9	Uganda	EN 1	ND	Y	N	12
U10	Uganda	EN 2	ND	Y	N	12
U11	Uganda	EN 3	ND	Y	N	12
U12A	Uganda	EN 4	ND	Y	N	12
U12B	Uganda	Sewage	ND	Y	N	12
U13	Uganda	IN 5	ND	Y	N	12
U14	Uganda	IN 4	ND	Y	N	12
U15	Uganda	Sewage	ND	Y	N	12
U25	Uganda	IN 3	ND	Y	N	12
U26	Uganda	IN 1	ND	Y	N	12
U32	Uganda	Water Hierson	ND	Y	N	12
U33	Uganda	Lab	ND	Y	N	12
U35	Uganda	Sa	ND	Y	N	12

Table S8. – Continued from overleaf

Sample ID	City, Country	Location	Coordinates	Water	Soil	Samples Group No.
U36	Uganda	F1	ND	Y	N	12
U37	Uganda	F2	ND	Y	N	12
U38	Uganda	F3	ND	Y	N	12
U39	F4	ND	Y	N	12	
U40	Uganda	Gd 1	ND	Y	N	12
U41	Uganda	Fine Screen	ND	Y	N	12
U42	Uganda	Sewage 1	ND	Y	N	12
U43	Uganda	Rx 1	ND	Y	N	12
U44	Uganda	Rx 2	ND	Y	N	12
U45	Uganda	Gd 2	ND	Y	N	12
U46	Uganda	Sample	ND	Y	N	12
U47	Uganda	Sample	ND	Y	N	12
U48	Uganda	Sample	ND	Y	N	12
U49	Uganda	Sample	ND	Y	N	12
U50	Uganda	Gd 3	ND	Y	N	12
U51	Uganda	S1	ND	Y	N	12
U52	Uganda	S2	ND	Y	N	12
U53	Uganda	S3	ND	Y	N	12
U54	Uganda	rx	ND	Y	N	12
U55	Uganda	Gd 4	ND	Y	N	12

Table S8. – Continued from overleaf

Sample ID	City, Country	Location	Coordinates	Water	Soil	Samples Group No.
U56	Uganda	Sample	ND	Y	N	12
U57	Uganda	Min	ND	Y	N	12
U58	Uganda	Sample	ND	Y	N	12
U59	Uganda	Sample	ND	Y	N	12
U60	Uganda	Fine Screen	ND	Y	N	12
U61	Uganda	N3	ND	Y	N	12
U62	Uganda	N3	ND	Y	N	12
U75	Uganda	WH	ND	Y	N	12
U76	Uganda	Trickle Filter	ND	Y	N	12
U34	Uganda	Sample	ND	Y	N	13
U63	Uganda	IN 5	ND	Y	N	13
U64	Uganda	IN 4	ND	Y	N	13
U65	Uganda	IN 3	ND	Y	N	13
U66	Uganda	IN 1	ND	Y	N	13
U67	Uganda	GAK	ND	Y	N	13
U68	Uganda	GAK	ND	Y	N	13
U69	Uganda	GAK	ND	Y	N	13
U70	Uganda	GAK	ND	Y	N	13
U71	Uganda	Fine Screen	ND	Y	N	13
U72	Uganda	Sample	ND	Y	N	13

171

Table S8. – Continued from overleaf

Sample ID	City, Country	Location	Coordinates	Water	Soil	Samples Group No.
U73	Uganda	Sewage	ND	Y	N	13
U74	Uganda	Sewage	ND	Y	N	13
U77	Uganda	Sample	ND	Y	N	13
U78	Uganda	FSI	ND	Y	N	13
U79	Uganda	Sample	ND	Y	N	13
U80	Uganda	Sample	ND	Y	N	13
U81	Uganda	Sample	ND	Y	N	13
U82	Uganda	Sample	ND	Y	N	13
U83	Uganda	Sample	ND	Y	N	13
U84	Uganda	Sample	ND	Y	N	13
U85	Uganda	Primary Waste	ND	Y	N	13
U86	Uganda	Fine Screen	ND	Y	N	13
U87	Uganda	F4	ND	Y	N	13
U88	Uganda	Sewage	ND	Y	N	13
U89	Uganda	Sedimentation	ND	Y	N	13
U90	Uganda	Tickle Filter	ND	Y	N	13
U91	Uganda	F3	ND	Y	N	13
U92	Uganda	Water	ND	Y	N	13
U93	Uganda	Sewage	ND	Y	N	13
U94	Uganda	F2	ND	Y	N	13

172

Table S8. – Continued from overleaf

Sample ID	City, Country	Location	Coordinates	Water	Soil	Samples Group No.
U95	Uganda	F1	ND	Y	N	13
U96	Uganda	Sewage	ND	Y	N	13
Lon	London	Sewage	ND	Y	N	14
28	Leicester, United kingdom	Aylestone meadows	52.614014, -1.150801	N	Y	9

Figure S1. Growth of *L. pneumophila* GBO4 in the presence of concentrated samples from environmental and man-made sources. This figure shows growth curves for *L. pneumophila* GBO4 when co-cultured with different concentrated environmental and man-made samples. Each sub-plot shows the result of incubation with a different samples, while each line shows growth of GBO4 with one concentrated sample either in the presence of divalent cation salts (blue line) or without salts (orange line), GBO4 growth with no sample with or without the presence of salt (red and green lines respectively). (A) shows the result of the first round of co-culture, while (B) shows the growth of GBO4 when the filtrate of samples from the first round of incubations where a reduction was noted, were re-incubated for a second round with fresh GBO4. The additional black line indicates growth of GBO4 when incubated with the supernatant from the control (no sample) from the initial round of incubations.

Figure S2. Growth of *L. pneumophila* GH1 in the presence of concentrated samples from environmental and man-made sources. This figure shows growth curves for *L. pneumophila* GH1 when co-cultured with different concentrated environmental and man-made samples. Each sub-plot shows the result of incubation with a different samples, while each line shows growth of GH1 with one concentrated sample either in the presence of divalent cation salts (blue line) or without salts (orange line), GH1 growth with no sample with or without the presence of salt (red and green lines respectively). (A) shows the result of the first round of co-culture.
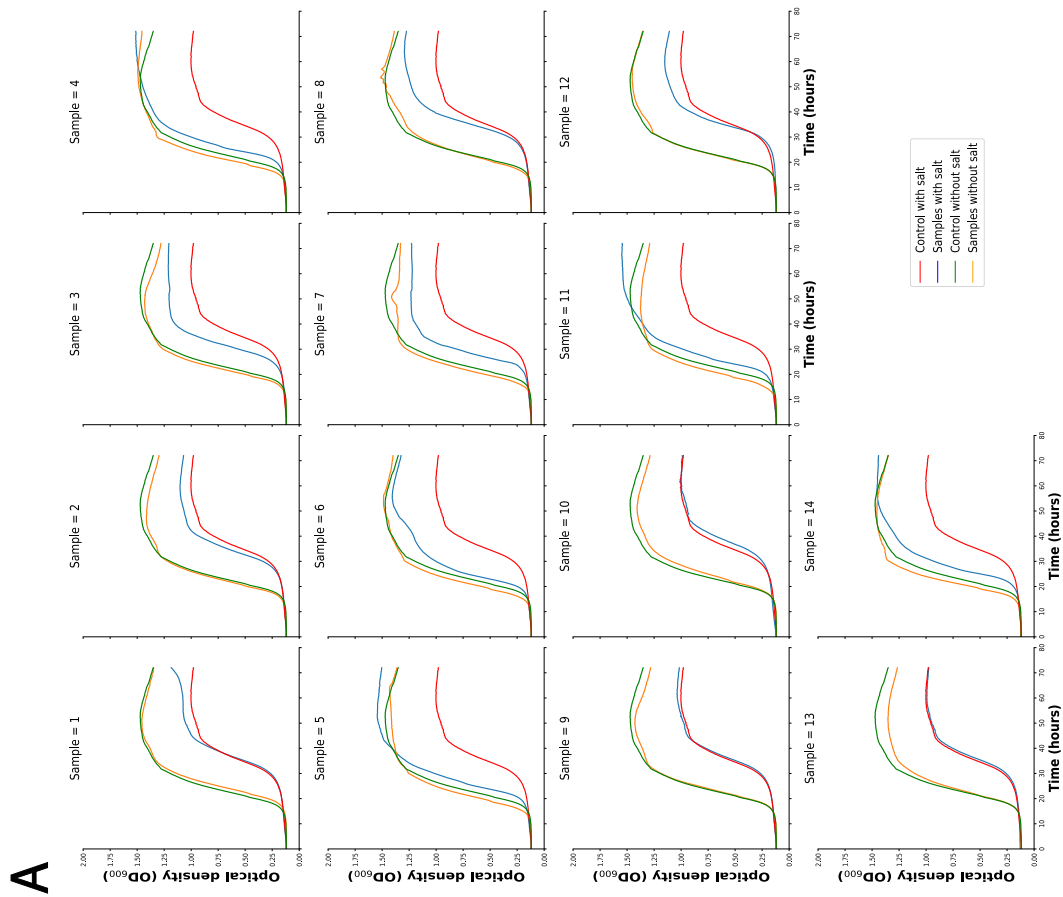
Figure S3. Growth of *L. pneumophila* NL1 in the presence of concentrated samples from environmental and man-made sources. This figure shows growth curves for *L. pneumophila* NL1 when co-cultured with different concentrated environmental and man-made samples. Each sub-plot shows the result of incubation with a different samples, while each line shows growth of NL1 with one concentrated sample either in the presence of divalent cation salts (blue line) or without salts (orange line), NL1 growth with no sample with or without the presence of salt (red and green lines respectively). (A) shows the result of the first round of co-culture.
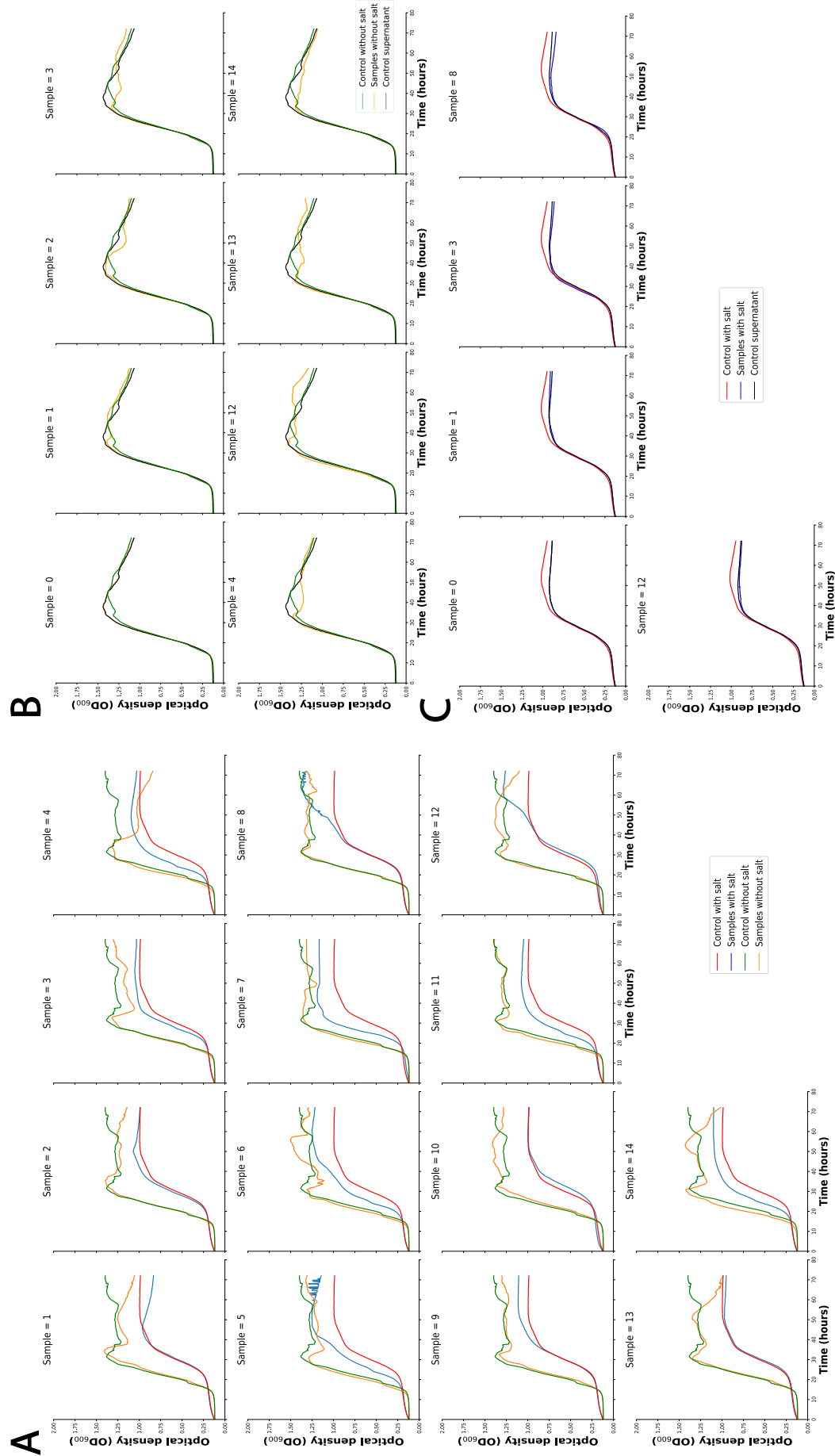
Figure S4. Growth of *L. pneumophila* Phe_11 in the presence of concentrated samples from environmental and man-made sources.
This figure shows growth curves for *L. pneumophila* Phe_11 when co-cultured with different concentrated environmental and man-made samples. Each sub-plot shows the result of incubation with a different samples, while each line shows growth of Phe_11 with one concentrated sample either in the presence of divalent cation salts (blue line) or without salts (orange line), Phe_11 growth with no sample with or without the presence of salt (red and green lines respectively). (A) shows the result of the first round of co-culture, while (B) and (C) show the growth of Phe_11 when the filtrate of samples from the first round of incubations where a reduction was noted, were re-incubated for a second round with fresh Phe_11. The additional black line indicates growth of Phe_11 when incubated with the supernatant from the control (no sample) from the initial round of incubations.

Figure S5. Growth of *L. pneumophila* Phe_12 in the presence of concentrated samples from environmental and man-made sources.
This figure shows growth curves for *L. pneumophila* Phe_12 when co-cultured with different concentrated environmental and man-made samples. Each sub-plot shows the result of incubation with a different samples, while each line shows growth of Phe_12 with one concentrated sample either in the presence of divalent cation salts (blue line) or without salts (orange line), Phe_12 growth with no sample with or without the presence of salt (red and green lines respectively). (A) shows the result of the first round of co-culture, while (B) shows the growth of Phe_12 when the filtrate of samples from the first round of incubations where a reduction was noted, were re-incubated for a second round with fresh Phe_12. The additional black line indicates growth of Phe_12 when incubated with the supernatant from the control (no sample) from the initial round of incubations.

Figure S6. Growth of *L. pneumophila* Phe_13 in the presence of concentrated samples from environmental and man-made sources.
This figure shows growth curves for *L. pneumophila* Phe_13 when co-cultured with different concentrated environmental and man-made samples. Each sub-plot shows the result of incubation with a different samples, while each line shows growth of Phe_13 with one concentrated sample either in the presence of divalent cation salts (blue line) or without salts (orange line), Phe_13 growth with no sample with or without the presence of salt (red and green lines respectively). (A) shows the result of the first round of co-culture, while (B) and (C) show the growth of Phe_13 when the filtrate of samples from the first round of incubations where a reduction was noted, were re-incubated for a second round with fresh Phe_13. The additional black line indicates growth of Phe_13 when incubated with the supernatant from the control (no sample) from the initial round of incubations.

179

Figure S7. Growth of *L. pneumophila* Phe_14 in the presence of concentrated samples from environmental and man-made sources. This figure shows growth curves for *L. pneumophila* Phe_14 when co-cultured with different concentrated environmental and man-made samples. Each sub-plot shows the result of incubation with a different samples, while each line shows growth of Phe_14 with one concentrated sample either in the presence of divalent cation salts (blue line) or without salts (orange line), Phe_14 growth with no sample with or without the presence of salt (red and green lines respectively). (A) shows the result of the first round of co-culture, while (B) and (C) show the growth of Phe_14 when the filtrate of samples from the first round of incubations where a reduction was noted, were re-incubated for a second round with fresh Phe_14. The additional black line indicates growth of Phe_14 when incubated with the supernatant from the control (no sample) from the initial round of incubations.
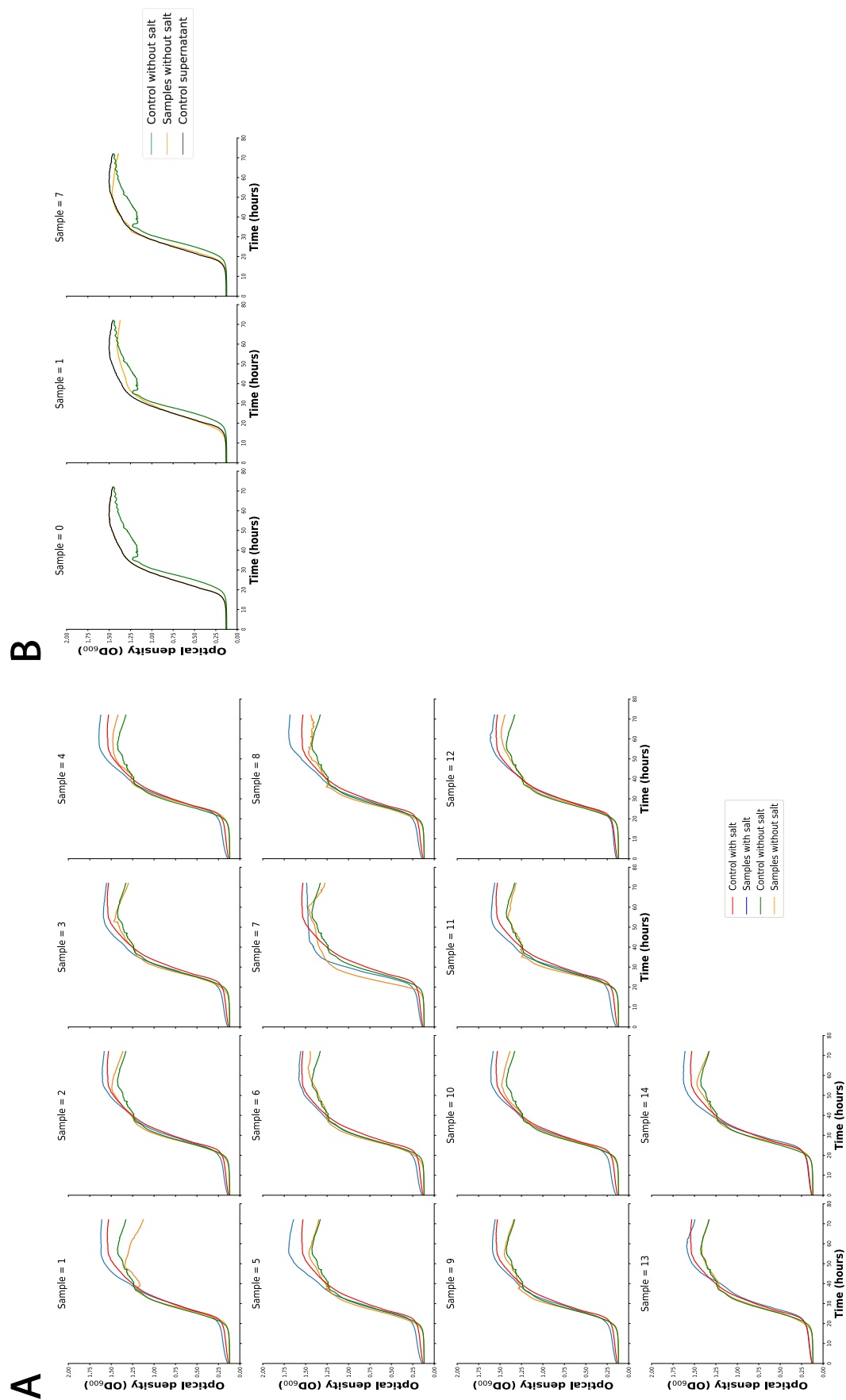
180

Figure S8. Growth of *L. pneumophila* Phe_17 in the presence of concentrated samples from environmental and man-made sources. This figure shows growth curves for *L. pneumophila* Phe_17 when co-cultured with different concentrated environmental and man-made samples. Each sub-plot shows the result of incubation with a different samples, while each line shows growth of Phe_17 with one concentrated sample either in the presence of divalent cation salts (blue line) or without salts (orange line), Phe_17 growth with no sample with or without the presence of salt (red and green lines respectively). (A) shows the result of the first round of co-culture, while (B) and (C) show the growth of Phe_17 when the filtrate of samples from the first round of incubations where a reduction was noted, were re-incubated for a second round with fresh Phe_17. The additional black line indicates growth of Phe_17 when incubated with the supernatant from the control (no sample) from the initial round of incubations.

Figure S9. Growth of *L. pneumophila* Phe_19 in the presence of concentrated samples from environmental and man-made sources.
This figure shows growth curves for *L. pneumophila* Phe_19 when co-cultured with different concentrated environmental and man-made samples. Each sub-plot shows the result of incubation with a different samples, while each line shows growth of Phe_19 with one concentrated sample either in the presence of divalent cation salts (blue line) or without salts (orange line), Phe_19 growth with no sample with or without the presence of salt (red and green lines respectively). (A) shows the result of the first round of co-culture, while (B) and (C) show the growth of Phe_19 when the filtrate of samples from the first round of incubations where a reduction was noted, were re-incubated for a second round with fresh Phe_19. The additional black line indicates growth of Phe_19 when incubated with the supernatant from the control (no sample) from the initial round of incubations.
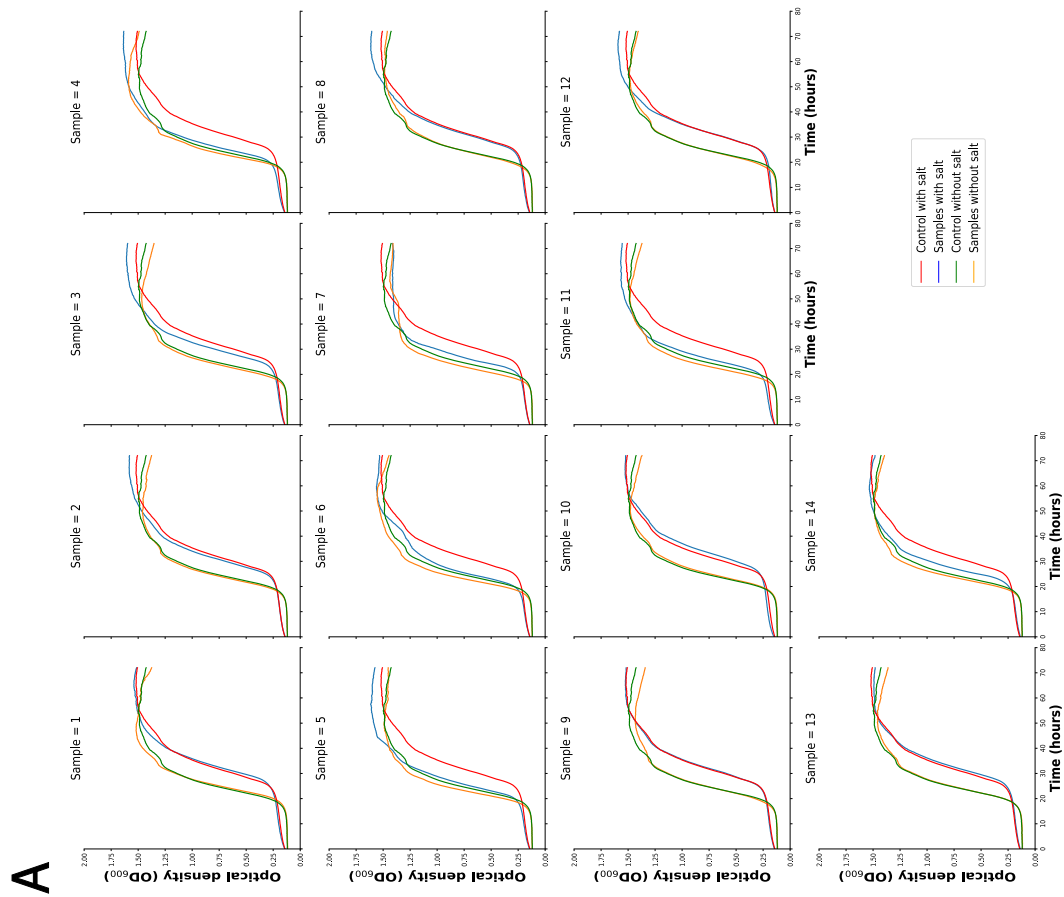
182

Figure S10. Growth of *L. pneumophila* Phe_20 in the presence of concentrated samples from environmental and man-made sources. This figure shows growth curves for *L. pneumophila* Phe_20 when co-cultured with different concentrated environmental and man-made samples. Each sub-plot shows the result of incubation with a different samples, while each line shows growth of Phe_20 with one concentrated sample either in the presence of divalent cation salts (blue line) or without salts (orange line), Phe_20 growth with no sample with or without the presence of salt (red and green lines respectively). (A) shows the result of the first round of co-culture, while (B) shows the growth of Phe_20 when the filtrate of samples from the first round of incubations where a reduction was noted, were re-incubated for a second round with fresh Phe_20. The additional black line indicates growth of Phe_20 when incubated with the supernatant from the control (no sample) from the initial round of incubations.
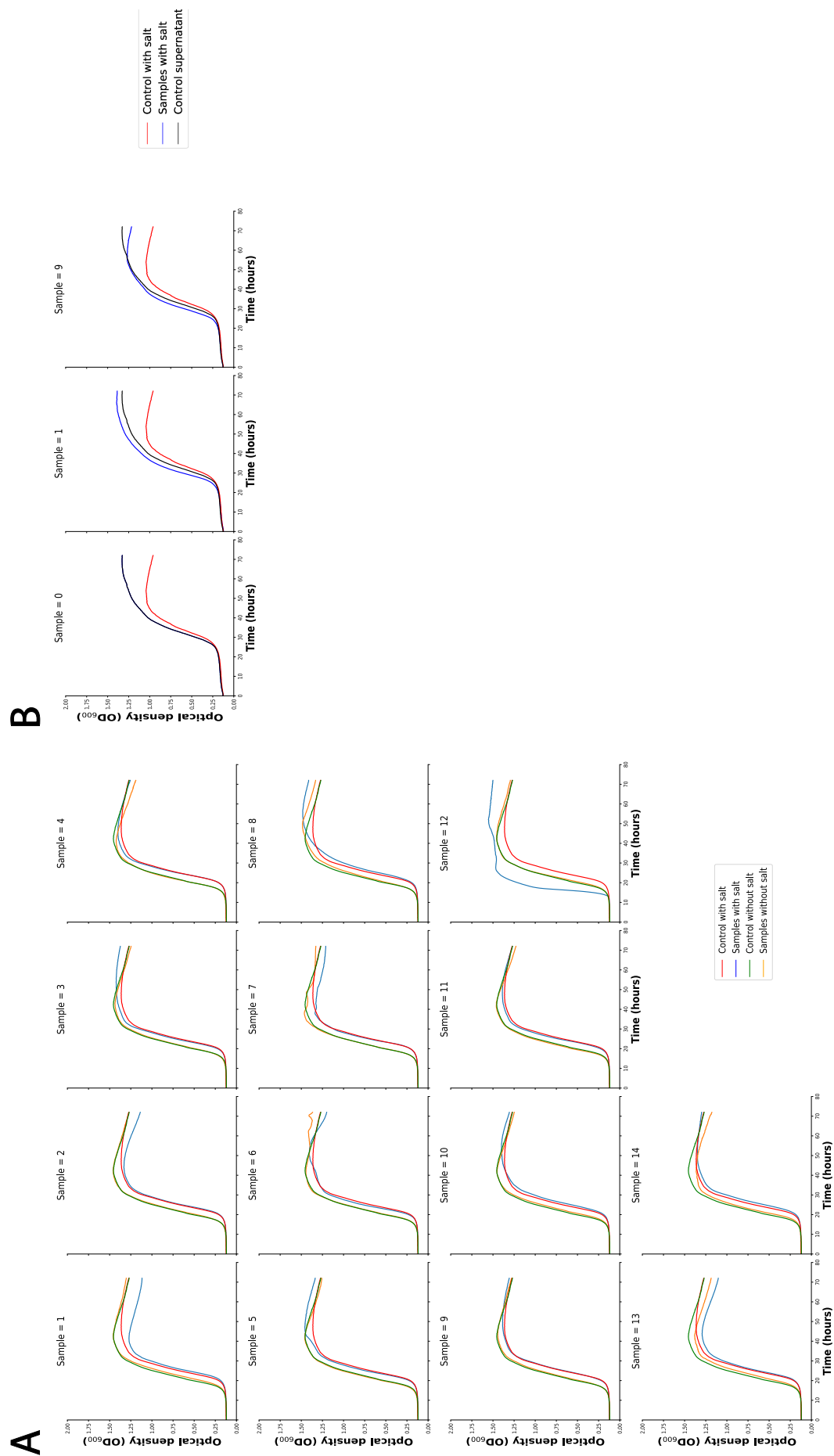
Figure S11. Growth of *L. pneumophila* Phe_21 in the presence of concentrated samples from environmental and man-made sources. This figure shows growth curves for *L. pneumophila* Phe_21 when co-cultured with different concentrated environmental and man-made samples. Each sub-plot shows the result of incubation with a different samples, while each line shows growth of Phe_21 with one concentrated sample either in the presence of divalent cation salts (blue line) or without salts (orange line), Phe_21 growth with no sample with or without the presence of salt (red and green lines respectively). (A) shows the result of the first round of co-culture, while (B) shows the growth of Phe_21 when the filtrate of samples from the first round of incubations where a reduction was noted, were re-incubated for a second round with fresh Phe_21. The additional black line indicates growth of Phe_21 when incubated with the supernatant from the control (no sample) from the initial round of incubations.
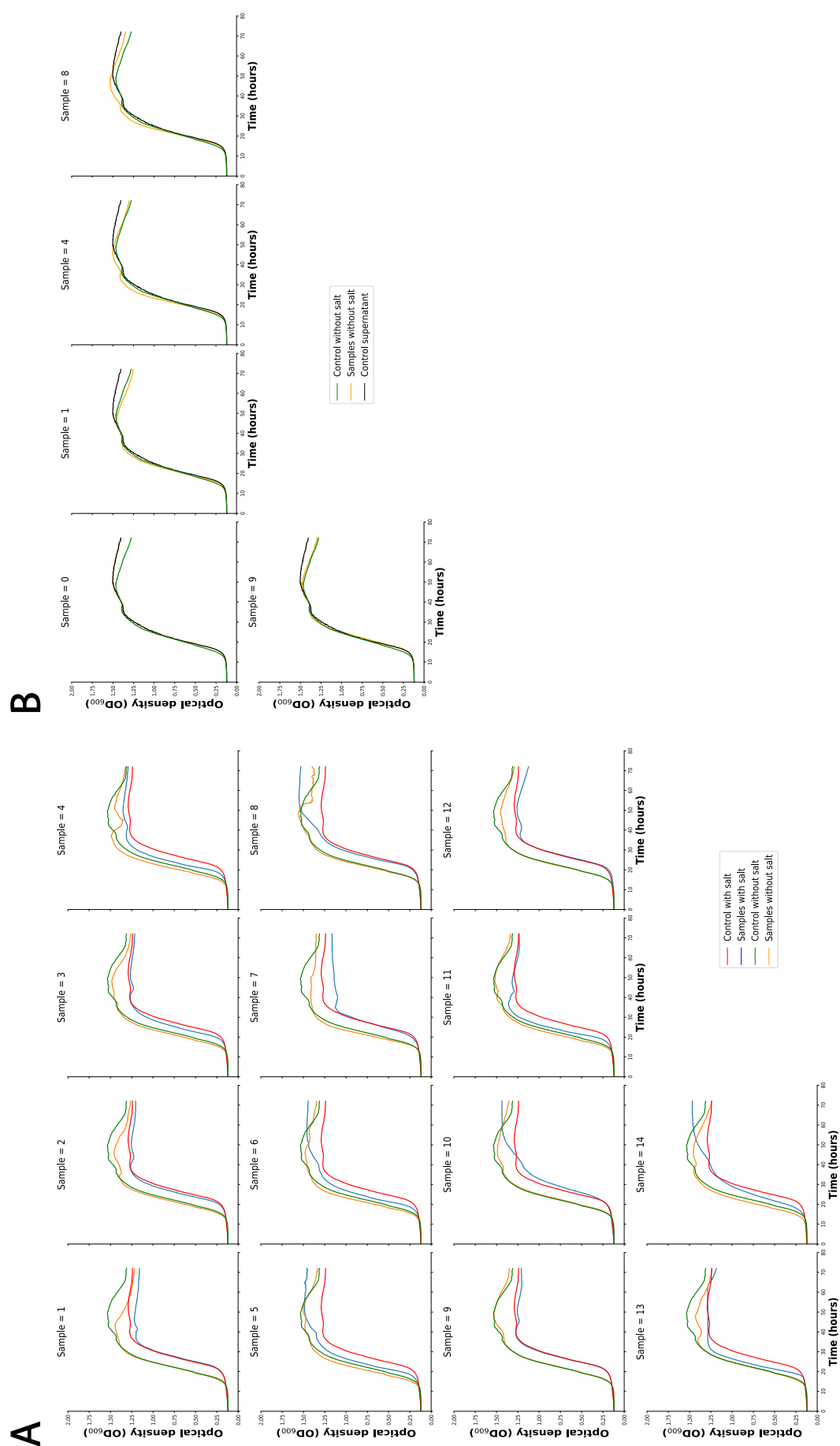
184
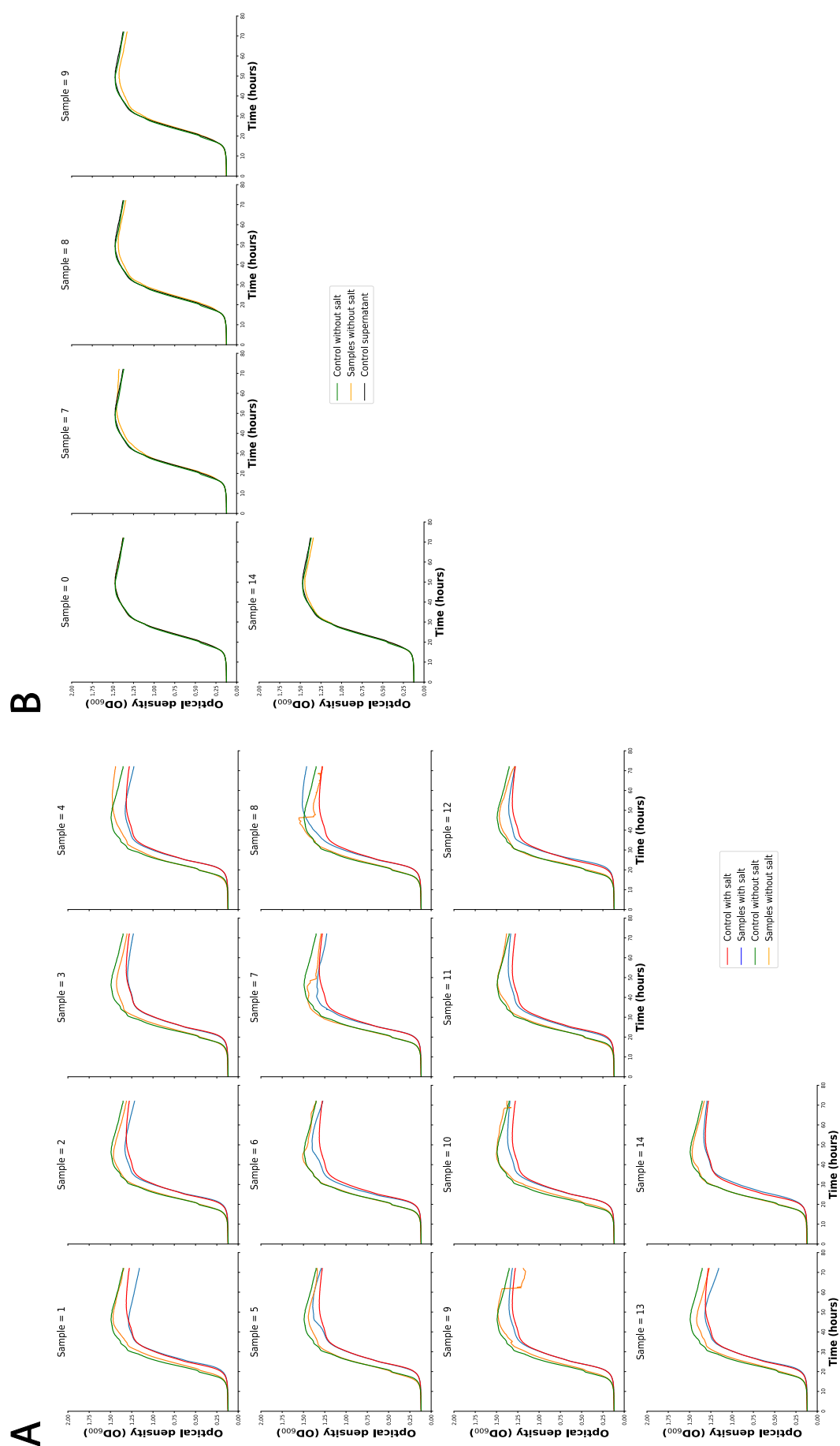
Figure S12. Growth of *L. pneumophila* Phe_22 in the presence of concentrated samples from environmental and man-made sources.
This figure shows growth curves for *L. pneumophila* Phe_22 when co-cultured with different concentrated environmental and man-made samples. Each sub-plot shows the result of incubation with a different samples, while each line shows growth of Phe_22 with one concentrated sample either in the presence of divalent cation salts (blue line) or without salts (orange line), Phe_22 growth with no sample with or without the presence of salt (red and green lines respectively). (A) shows the result of the first round of co-culture, while (B) shows the growth of Phe_22 when the filtrate of samples from the first round of incubations where a reduction was noted, were re-incubated for a second round with fresh Phe_22. The additional black line indicates growth of Phe_22 when incubated with the supernatant from the control (no sample) from the initial round of incubations.
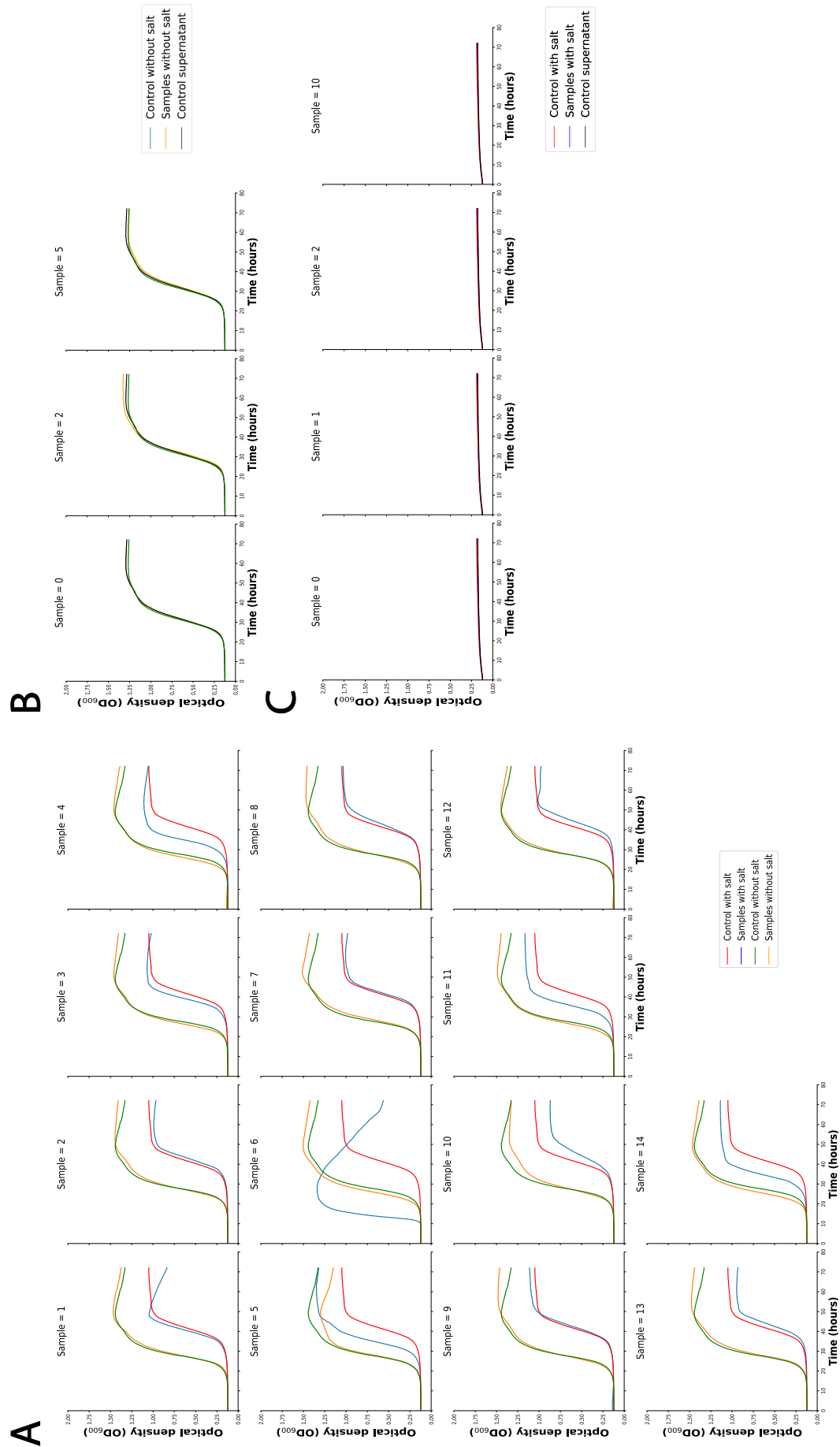
Figure S13. Growth of *L. Bozemanii* (Phe_31) in the presence of concentrated samples from environmental and man-made sources.
This figure shows growth curves for *L. Bozemanii* (Phe_31) when co-cultured with different concentrated environmental and man-made samples. Each sub-plot shows the result of incubation with a different samples, while each line shows growth of Phe_31 with one concentrated sample either in the presence of divalent cation salts (blue line) or without salts (orange line), Phe_31 growth with no sample with or without the presence of salt (red and green lines respectively). (A) shows the result of the first round of co-culture, while (B) and (C) show the growth of Phe_31 when the filtrate of samples from the first round of incubations where a reduction was noted, were re-incubated for a second round with fresh Phe_31. The additional black line indicates growth of Phe_31 when incubated with the supernatant from the control (no sample) from the initial round of incubations.
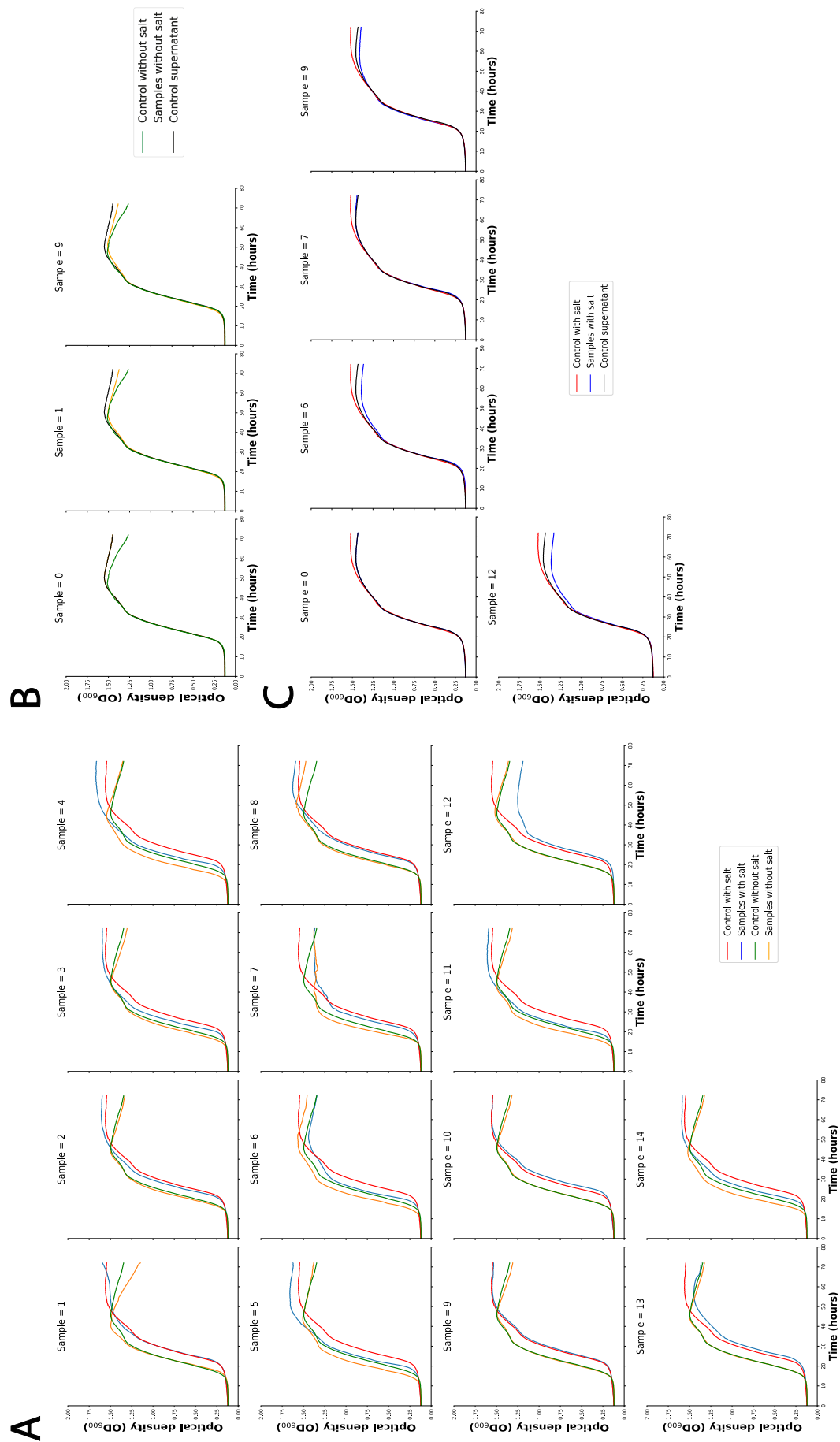
186
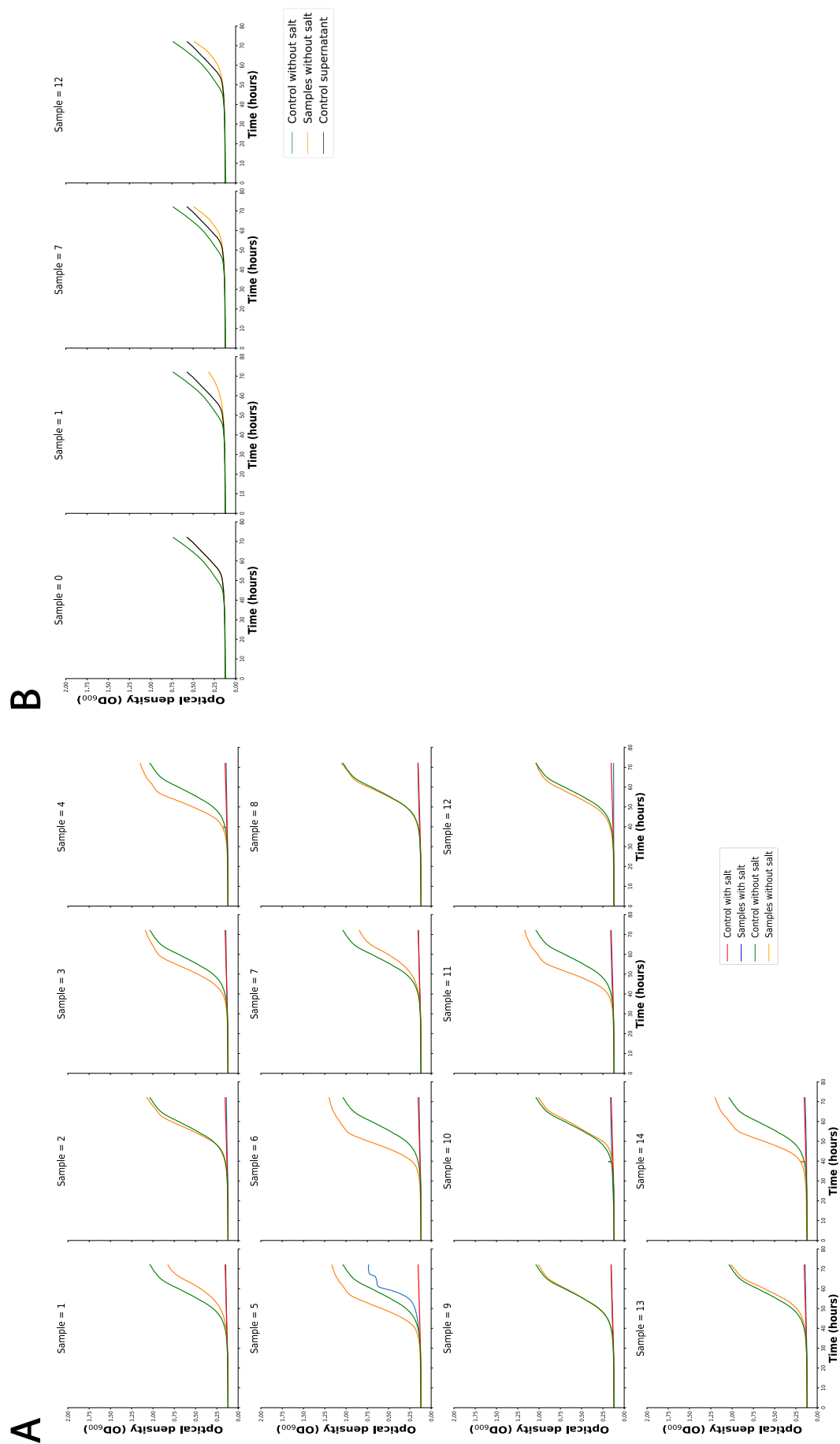
Figure S14. Growth of *L. gormanii* (Phe_36) in the presence of concentrated samples from environmental and man-made sources. This figure shows growth curves for *L. gormanii* (Phe_36) when co-cultured with different concentrated environmental and man-made samples. Each sub-plot shows the result of incubation with a different samples, while each line shows growth of Phe_36 with one concentrated sample either in the presence of divalent cation salts (blue line) or without salts (orange line), Phe_36 growth with no sample with or without the presence of salt (red and green lines respectively). (A) shows the result of the first round of co-culture, while (B) shows the growth of Phe_36 when the filtrate of samples from the first round of incubations where a reduction was noted, were re-incubated for a second round with fresh Phe_36. The additional black line indicates growth of Phe_36 when incubated with the supernatant from the control (no sample) from the initial round of incubations.
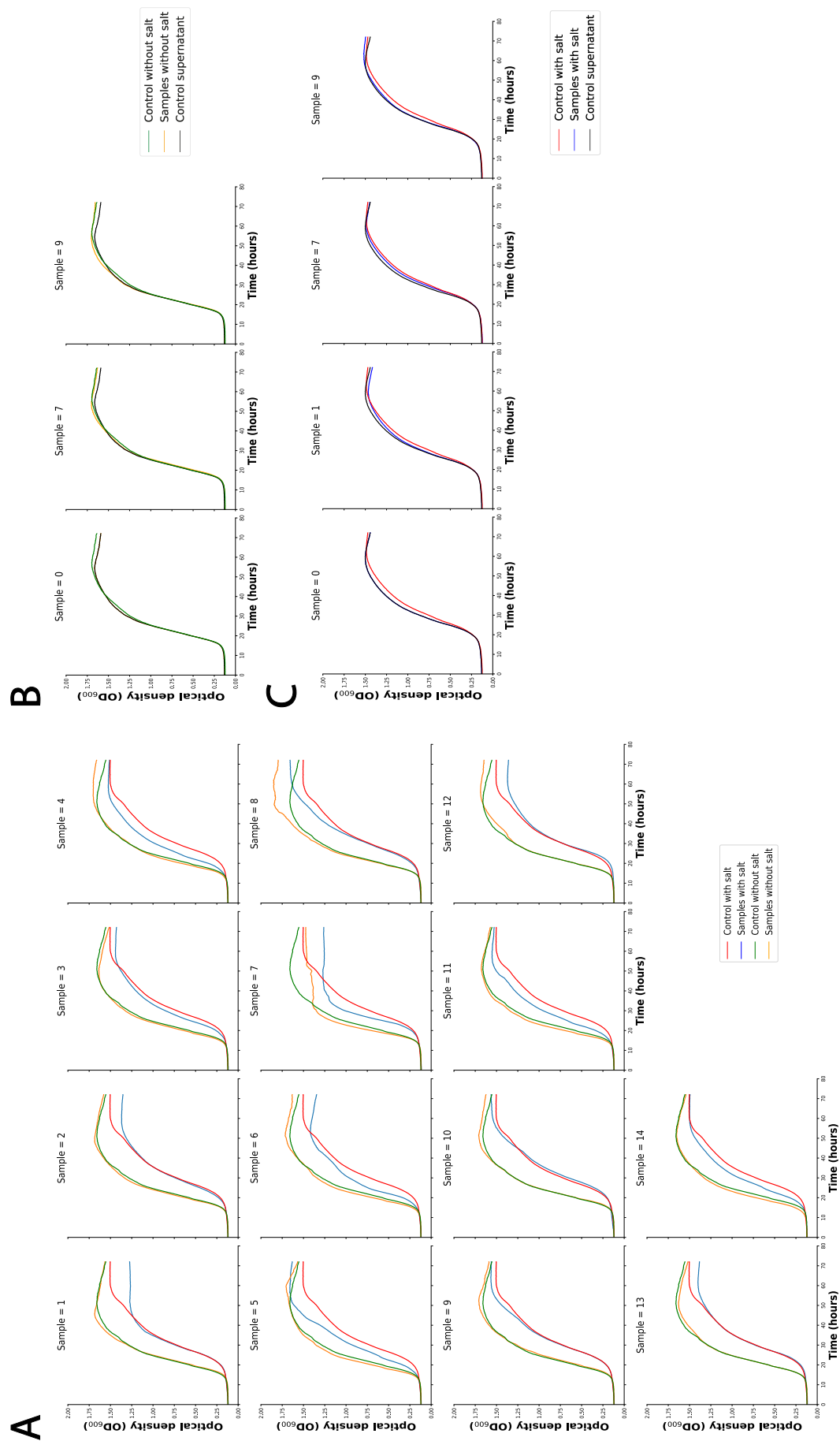
187
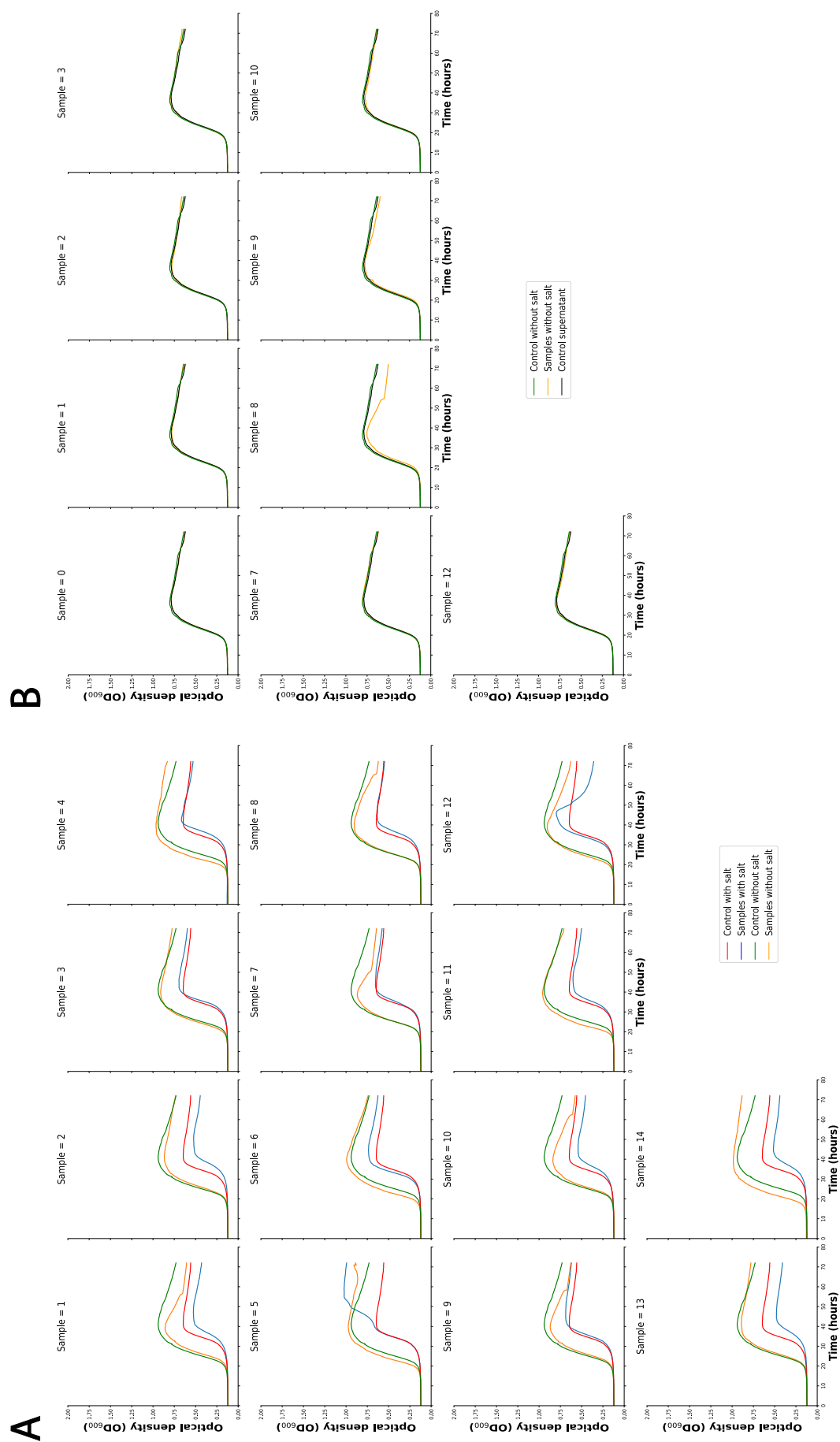
Figure S15. Growth of *L. dumoffii* (Phe_39) in the presence of concentrated samples from environmental and man-made sources. This figure shows growth curves for *L. dumoffii* (Phe_39) when co-cultured with different concentrated environmental and man-made samples. Each sub-plot shows the result of incubation with a different sample, while each line shows growth of Phe_39 with one concentrated sample either in the presence of divalent cation salts (blue line) or without salts (orange line), Phe_39 growth with no sample with or without the presence of salt (red and green lines respectively). (A) shows the result of the first round of co-culture, while (B) shows the growth of Phe_39 when the filtrate of samples from the first round of incubations where a reduction was noted, were re-incubated for a second round with fresh Phe_39. The additional black line indicates growth of Phe_39 when incubated with the supernatant from the control (no sample) from the initial round of incubations.
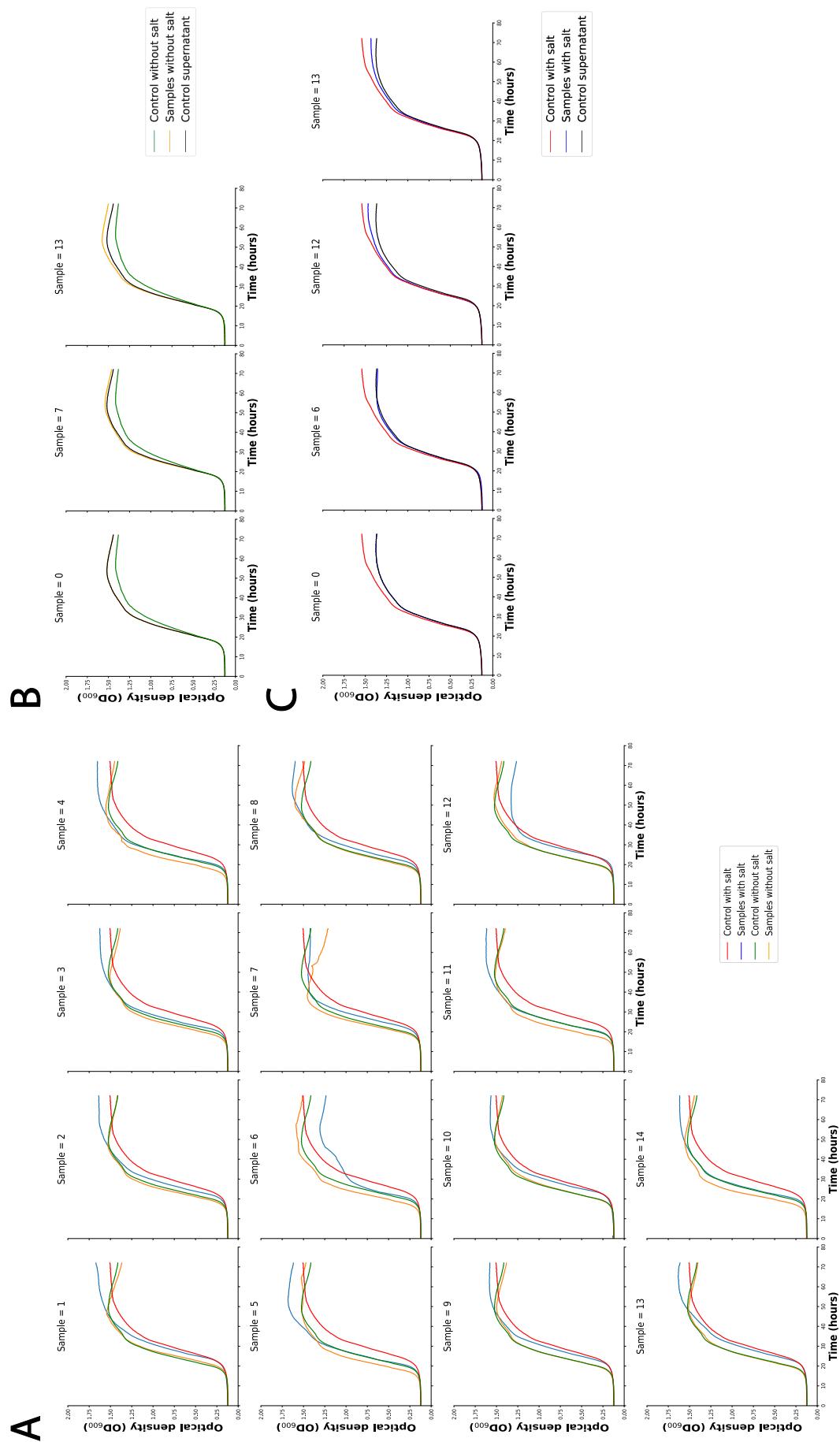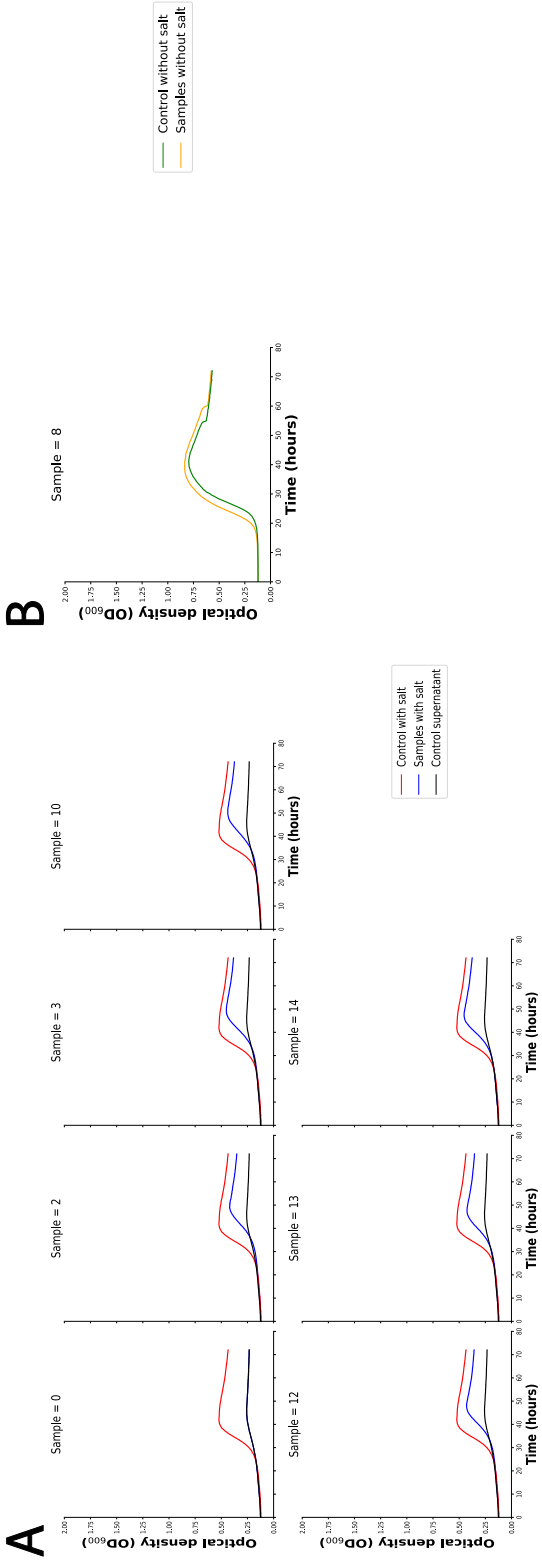
Figure S16. Growth of *L. dumoffii* (Phe_39) in the presence of concentrated samples from environmental and man-made sources. This figure; (A) and (B), show the growth of Phe_39 when the filtrate of samples from the first round of incubations (S15) where a reduction was noted, were re-incubated for a second round with fresh Phe_39. The additional black line indicates growth of Phe_39 when incubated with the supernatant from the control (no sample) from the initial round of incubations.
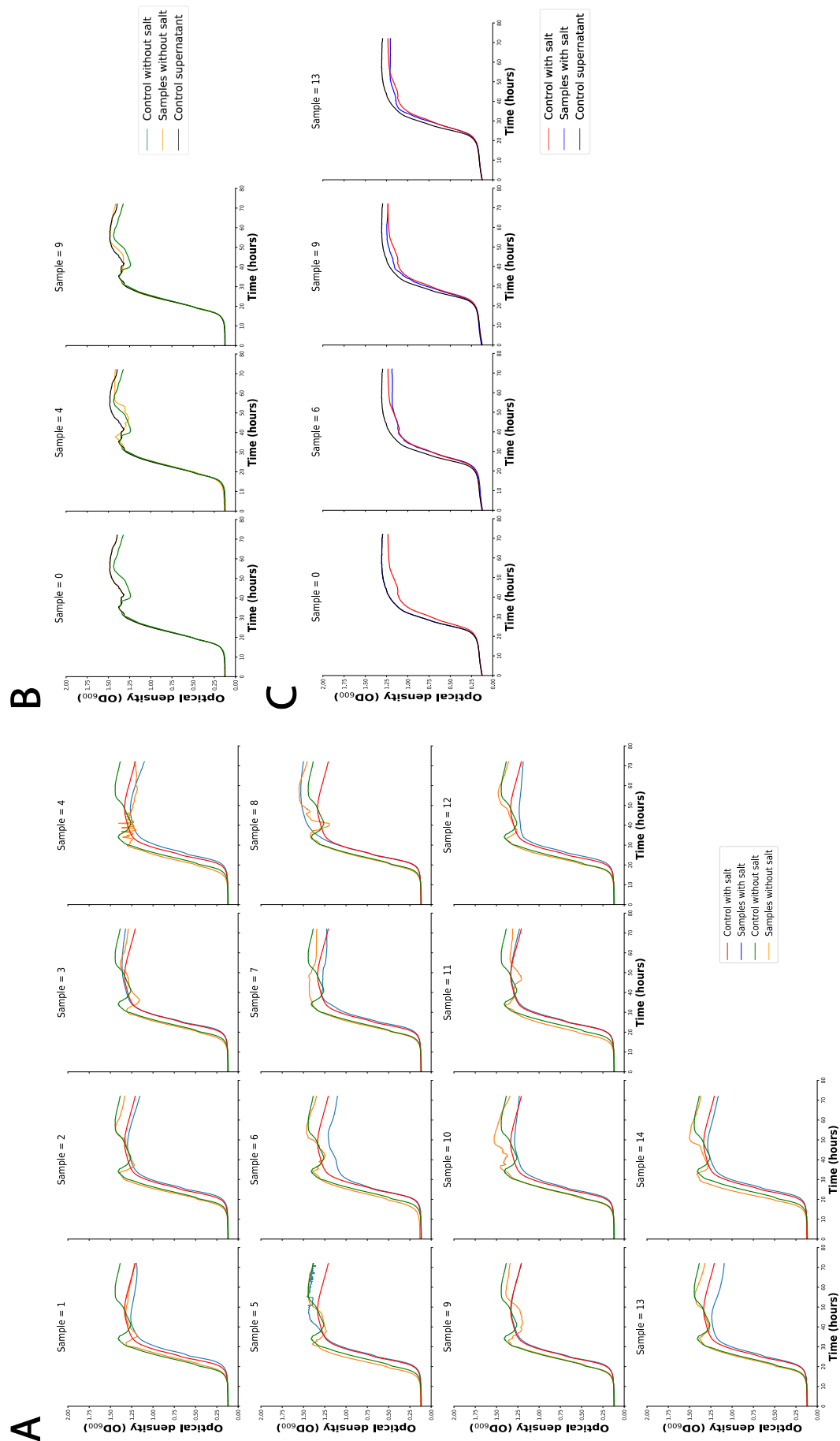
Figure S17. Growth of *L. pneumophila* Phe_4 in the presence of concentrated samples from environmental and man-made sources.
This figure shows growth curves for *L. pneumophila* Phe_4 when co-cultured with different concentrated environmental and man-made samples. Each sub-plot shows the result of incubation with a different samples, while each line shows growth of Phe_4 with one concentrated sample either in the presence of divalent cation salts (blue line) or without salts (orange line), Phe_4 growth with no sample with or without the presence of salt (red and green lines respectively). (A) shows the result of the first round of co-culture, while (B) and (C) show the growth of Phe_4 when the filtrate of samples from the first round of incubations where a reduction was noted, were re-incubated for a second round with fresh Phe_4. The additional black line indicates growth of Phe_4 when incubated with the supernatant from the control (no sample) from the initial round of incubations.
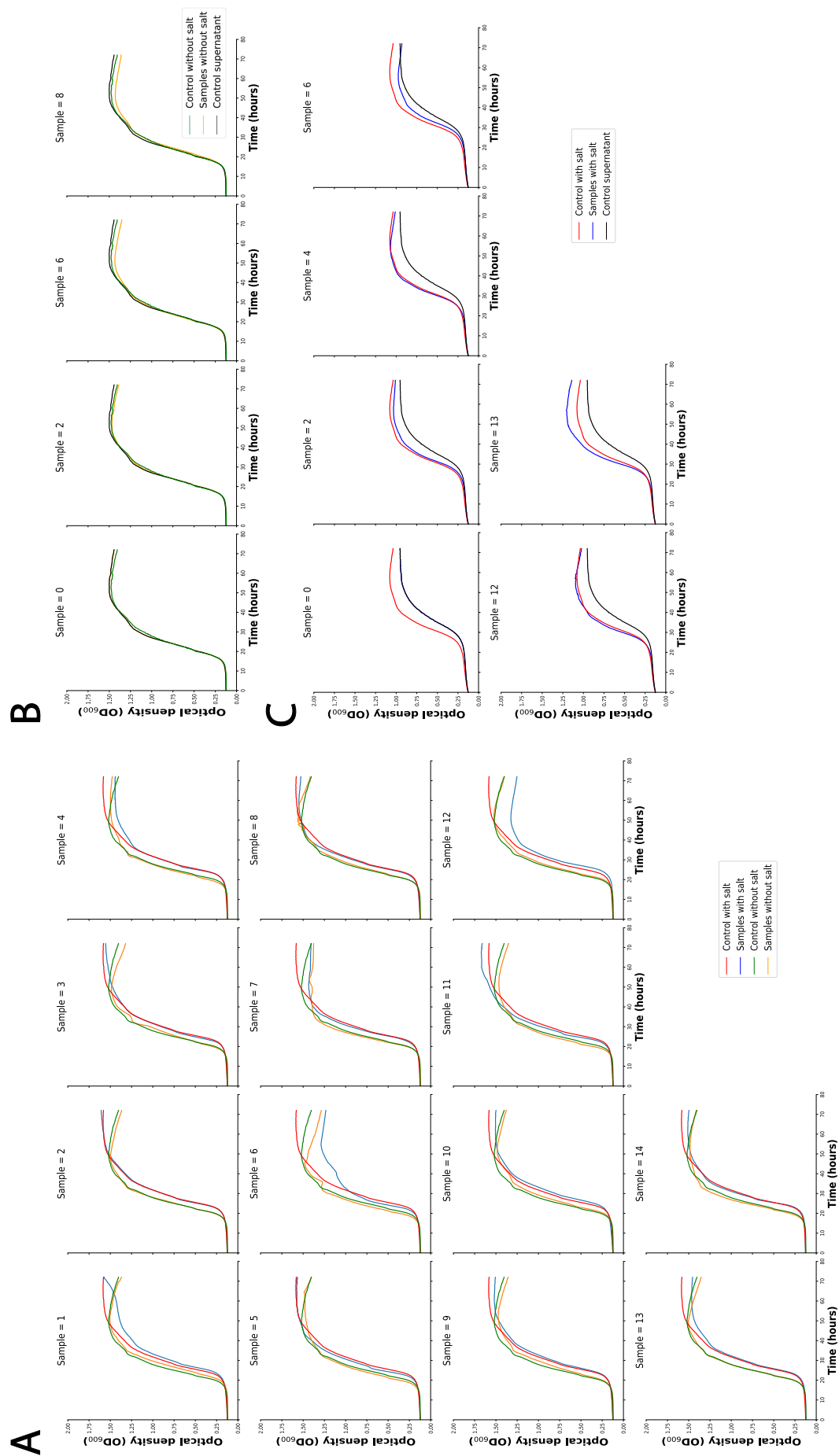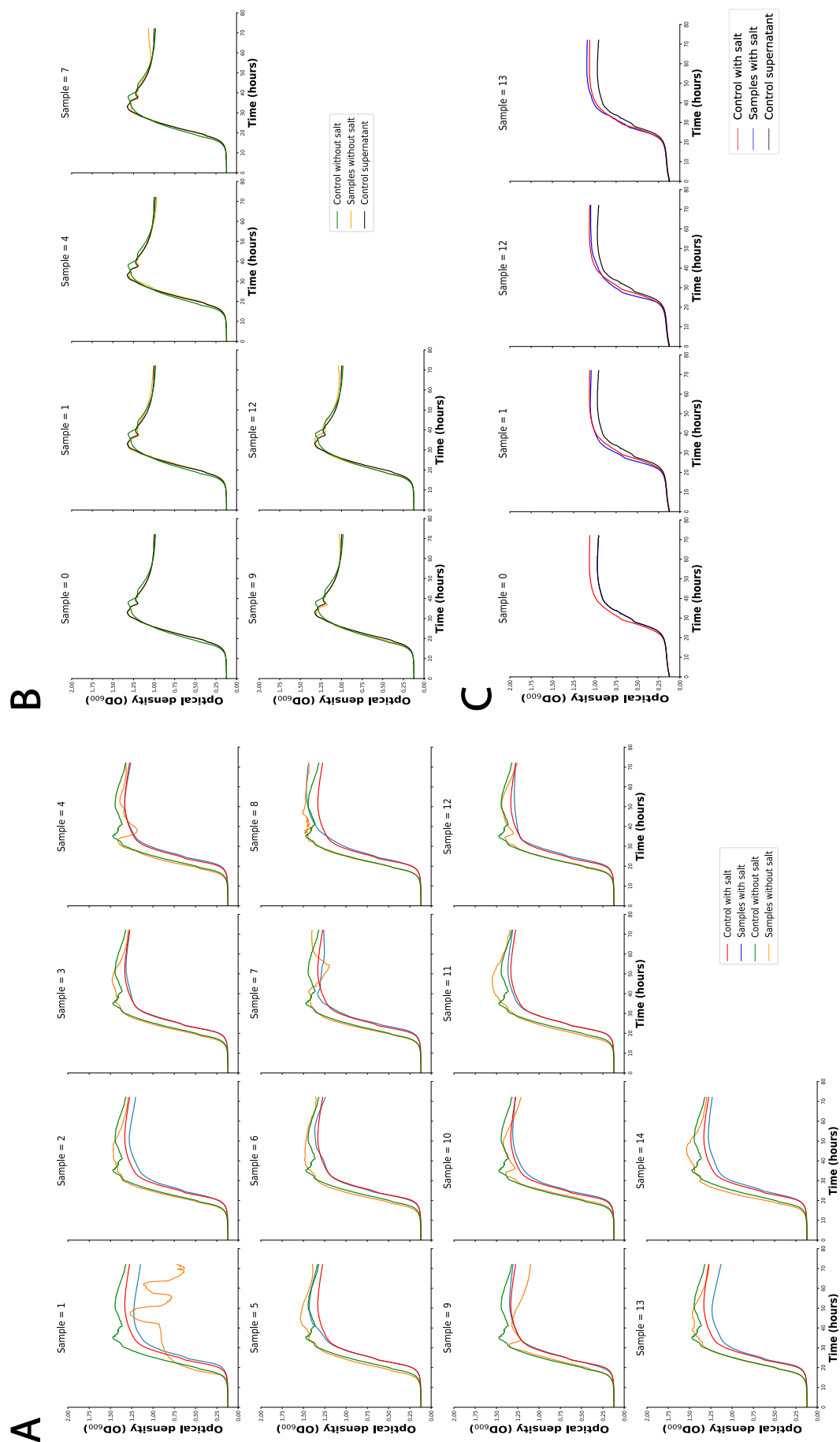
Figure S18. Growth of *L. pneumophila* Phe_5 in the presence of concentrated samples from environmental and man-made sources.
This figure shows growth curves for *L. pneumophila* Phe_5 when co-cultured with different concentrated environmental and man-made samples. Each sub-plot shows the result of incubation with a different samples, while each line shows growth of Phe_5 with one concentrated sample either in the presence of divalent cation salts (blue line) or without salts (orange line), Phe_5 growth with no sample with or without the presence of salt (red and green lines respectively). (A) shows the result of the first round of co-culture, while (B) and (C) show the growth of Phe_5 when the filtrate of samples from the first round of incubations where a reduction was noted, were re-incubated for a second round with fresh Phe_5. The additional black line indicates growth of Phe_5 when incubated with the supernatant from the control (no sample) from the initial round of incubations.
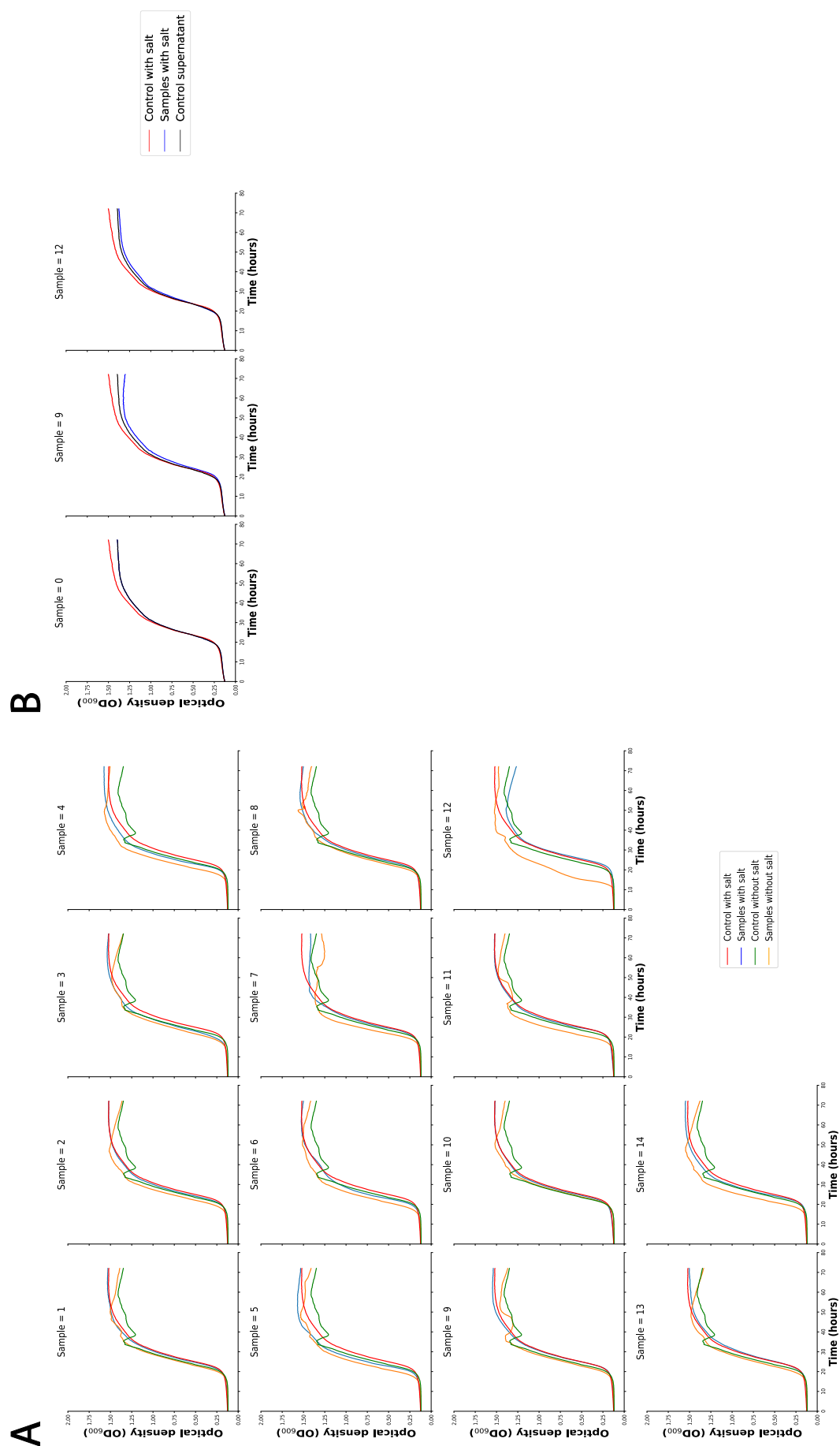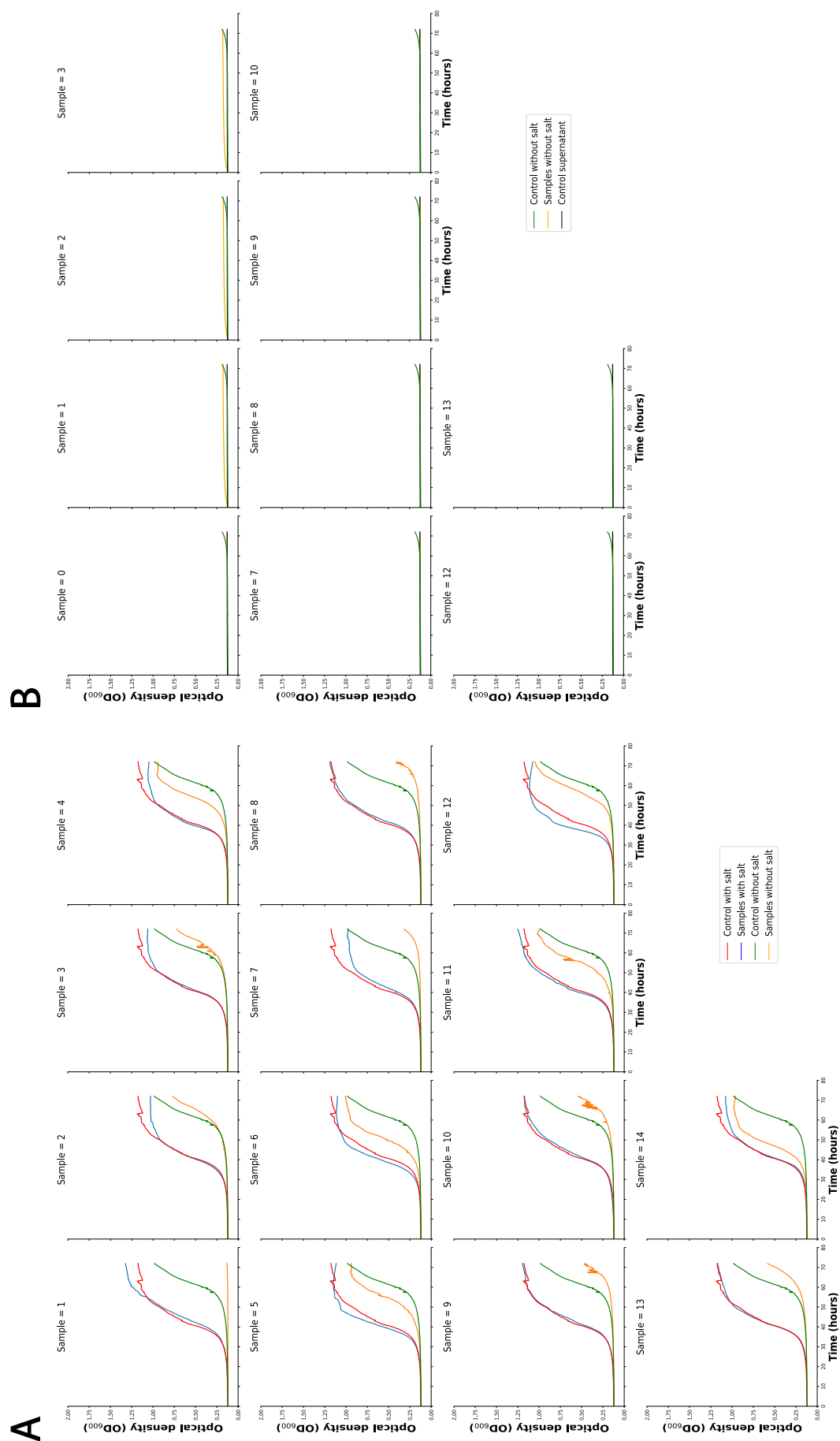
Figure S19. Growth of *L. pneumophila* Phe_7 in the presence of concentrated samples from environmental and man-made sources. This figure shows growth curves for *L. pneumophila* Phe_7 when co-cultured with different concentrated environmental and man-made samples. Each sub-plot shows the result of incubation with a different samples, while each line shows growth of Phe_7 with one concentrated sample either in the presence of divalent cation salts (blue line) or without salts (orange line), Phe_7 growth with no sample with or without the presence of salt (red and green lines respectively). (A) shows the result of the first round of co-culture, while (B) shows the growth of Phe_7 when the filtrate of samples from the first round of incubations where a reduction was noted, were re-incubated for a second round with fresh Phe_7. The additional black line indicates growth of Phe_7 when incubated with the supernatant from the control (no sample) from the initial round of incubations.

192
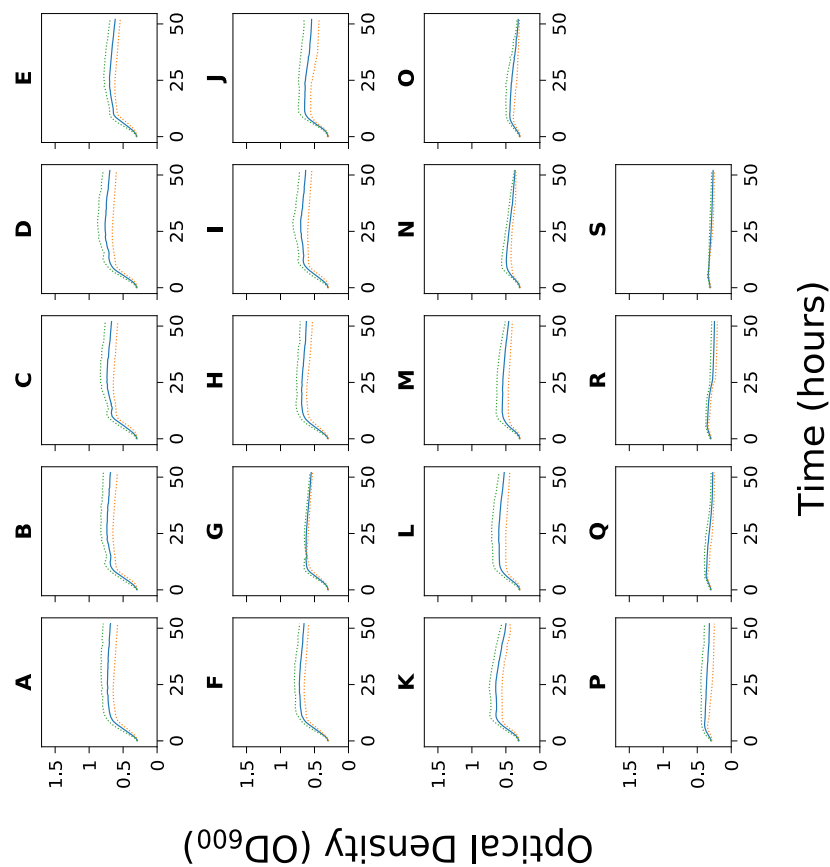
Figure S20. Growth curves of *L. hackeliae* Phe_34 following treatment with NFX. The figure shows representative growth profiles observed for *L. hackeliae* Phe_34 following treatment with 2-fold serial dilutions of NFX concentrations (40 μg ml^{-1} to 0.00025 μg ml^{-1}). Each subplot shows the growth profile of *L. hackeliae* Phe_34 exposed to a certain concentration of NFX: (**A**) untreated (0 μg ml^{-1}; (**B**) 0.00025 μg ml^{-1}; (**C**) 0.0005 μg ml^{-1}; (**D**) 0.001 μg ml^{-1}; (**E**) 0.002 μg ml^{-1}; (**F**) 0.004 μg ml^{-1}; (**G**) 0.009 μg ml^{-1}; (**H**) 0.019 μg ml^{-1}; (**I**) 0.039 μg ml^{-1}; (**J**) 0.078 μg ml^{-1}; (**K**) 0.156 μg ml^{-1}; (**L**) 0.312 μg ml^{-1}; (**M**) 0.625 μg ml^{-1}; (**N**) 1.25 μg ml^{-1}; (**O**) 2.5 μg ml^{-1}; (**P**) **5 μg ml^{-1}**; (**Q**) 10 μg ml^{-1}; (**R**) 20 μg ml^{-1}; (**S**) 40 μg ml^{-1}. Blue line indicate mean bacterial growth (n = 3). Each green and orange dashed lines show growth $\pm$ sd.

Figure S21. Growth curves of *L. pneumophila* Phe_14 following treatment with MitC and NFX. The figure shows representative growth profiles observed for *L. pneumophila* Phe_14 following treatment with 2-fold serial dilutions of NFX concentrations (40 μg ml^{-1} to 0.00025 μg ml^{-1}). Each subplot shows the growth profile of *L. pneumophila* Phe_14 exposed to a certain concentration of NFX: (**A**) untreated (0 μg ml^{-1}; (**B**) 0.00025 μg ml^{-1}; (**C**) 0.0005 μg ml^{-1}; (**D**) 0.001 μg ml^{-1}; (**E**) 0.002 μg ml^{-1}; (**F**) 0.004 μg ml^{-1}; (**G**) 0.009 μg ml^{-1}; (**H**) 0.019 μg ml^{-1}; (**I**) 0.039 μg ml^{-1}; (**J**) 0.078 μg ml^{-1}; (**K**) 0.156 μg ml^{-1}; (**L**) 0.312 μg ml^{-1}; (**M**) 0.625 μg ml^{-1}; (**N**) 1.25 μg ml^{-1}; (**O**) 2.5 μg ml^{-1}; (**P**) **5 μg ml^{-1}**; (**Q**) 10 μg ml^{-1}; (**R**) 20 μg ml^{-1}; (**S**) 40 μg ml^{-1}. Blue line indicate mean bacterial growth (n = 3). Each green and orange dashed lines show growth $\pm$ sd.

Figure S22. Growth curves of *L. pneumophila* GBO4 following treatment with MitC and NFX. The figure shows representative growth profiles observed for *L. pneumophila* GBO4 following treatment with 2-fold serial dilutions of MitC concentrations (40 μg ml^{-1} to 0.00025 μg ml^{-1}). Each subplot shows the growth profile of *L. pneumophila* GBO4 exposed to a certain concentration of MitC: (**A**) untreated (0 μg ml^{-1}; (**B**) 0.00025 μg ml^{-1}; (**C**) 0.0005 μg ml^{-1}; (**D**) 0.001 μg ml^{-1}; (**E**) 0.002 μg ml^{-1}; (**F**) 0.004 μg ml^{-1}; (**G**) 0.009 μg ml^{-1}; (**H**) 0.019 μg ml^{-1}; (**I**) 0.039 μg ml^{-1}; (**J**) 0.078 μg ml^{-1}; (**K**) 0.156 μg ml^{-1}; (**L**) 0.312 μg ml^{-1}; (**M**) 0.625 μg ml^{-1}; (**N**) 1.25 μg ml^{-1}; (**O**) 2.5 μg ml^{-1}; (**P**) **5 μg ml^{-1}**; (**Q**) 10 μg ml^{-1}; (**R**) 20 μg ml^{-1}; (**S**) 40 μg ml^{-1}. Blue line indicate mean bacterial growth (n = 3). Each green and orange dashed lines show growth $\pm$ sd.

Figure S23. Growth curves of _L. pneumophila_ Phe_23 following treatment with MitC. The figure shows representative growth profiles observed for _L. pneumophila_ Phe_23 following treatment with 2-fold serial dilutions of MitC concentrations (40 μg ml⁻¹ to 0.00025 μg ml⁻¹). Each subplot shows the growth profile of _L. pneumophila_ Phe_23 exposed to a certain concentration of MitC: **(A)** untreated (0 μg ml⁻¹; **(B)** 0.00025 μg ml⁻¹; **(C)** 0.0005 μg ml⁻¹; **(D)** 0.001 μg ml⁻¹; **(E)** 0.002 μg ml⁻¹; **(F)** 0.004 μg ml⁻¹; **(G)** 0.009 μg ml⁻¹; **(H)** 0.019 μg ml⁻¹; **(I)** 0.039 μg ml⁻¹; **(J)** 0.078 μg ml⁻¹; **(K)** 0.156 μg ml⁻¹; **(L)** 0.312 μg ml⁻¹; **(M)** 0.625 μg ml⁻¹; **(N)** 1.25 μg ml⁻¹; **(O)** 2.5 μg ml⁻¹; **(P) 5 μg ml⁻¹**; **(Q)** 10 μg ml⁻¹; **(R)** 20 μg ml⁻¹; **(S)** 40 μg ml⁻¹. Blue line indicate mean bacterial growth (n = 3). Each green and orange dashed lines show growth ± sd.

Figure S24. Growth curves of *L. oakridgensis* Phe_35 following treatment with MitC and NFX. The figure shows representative growth profiles observed for *L. oakridgensis* Phe_35 following treatment with 2-fold serial dilutions of NFX concentrations (40 μg ml^{-1} to 0.00025 μg ml^{-1}). Each subplot shows the growth profile of *L. oakridgensis* Phe_35 exposed to a certain concentration of NFX: (**A**) untreated (0 μg ml^{-1}); (**B**) 0.00025 μg ml^{-1}; (**C**) 0.0005 μg ml^{-1}; (**D**) 0.001 μg ml^{-1}; (**E**) 0.002 μg ml^{-1}; (**F**) 0.004 μg ml^{-1}; (**G**) 0.009 μg ml^{-1}; (**H**) 0.019 μg ml^{-1}; (**I**) 0.039 μg ml^{-1}; (**J**) 0.078 μg ml^{-1}; (**K**) 0.156 μg ml^{-1}; (**L**) 0.312 μg ml^{-1}; (**M**) 0.625 μg ml^{-1}; (**N**) 1.25 μg ml^{-1}; (**O**) 2.5 μg ml^{-1}; (**P**) 5 μg ml^{-1}; (**Q**) 10 μg ml^{-1}; (**R**) 20 μg ml^{-1}; (**S**) 40 μg ml^{-1}. Blue line indicate mean bacterial growth (n = 3). Each green and orange dashed lines show growth $\pm$ sd.

Table S9. Table of the list of prophage elements identified within all *Legionella* spp. genomes in GenBank using PHASTER

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	*attL* site	Prophage GC content (%)
Legionella bozemanae	WIGA	GCA_001467045	0	N/A	N/A	N/A	N/A
Legionella bozemanae	ATCC 33217.1	GCA_001648595	1	incomplete	28615	yes	37.8
Legionella dumoffii	Tex-KL	GCA_000236145	1	incomplete	14513	yes	40.15
Legionella dumoffii	NY23	GCA_000236165	1	incomplete	31470	yes	40.59
Legionella dumoffii	NY23	GCA_000236165	2	incomplete	4885	no	36.47
Legionella dumoffii	NY23	GCA_001467605	0	N/A	N/A	N/A	N/A
Legionella gormanii	LS-13	GCA_001467685	0	N/A	N/A	N/A	N/A
Legionella gormanii	ATCC 33297	GCA_001648685	1	incomplete	12716	yes	40.97
Legionella gormanii	ATCC 33342	GCA_900156395	0	N/A	N/A	N/A	N/A
Legionella gormanii	LS-13	GCA_001467685	0	N/A	N/A	N/A	N/A
Legionella gormanii	ATCC 33297	GCA_001648685	1	incomplete	12716	yes	40.97
Legionella gormanii	ATCC 33342	GCA_900156395	0	N/A	N/A	N/A	N/A
Legionella pneumophila	GH1	N/A	0	N/A	N/A	N/A	N/A
Legionella adelaidensis	1762-AUS-ELade	GCA_001467055	1	incomplete	13697	yes	35.06
Legionella anisa	Linanisette	GCA_000333755	1	incomplete	10351	no	38.05
Legionella anisa	Linanisette	GCA_000333755	2	incomplete	11535	no	36.97
Legionella anisa	WA-316-C3	GCA_001467525	0	N/A	N/A	N/A	N/A
Legionella anisa	FDAARGOS_200	GCA_002082905	1	incomplete	11539	no	36.98
Legionella anisa	FDAARGOS_200	GCA_002082905	2	incomplete	14921	yes	34.67

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	attL site	Prophage GC content (%)
Legionella anisa	FDAARGOS_200	GCA_002082905	3	questionable	21533	yes	37.72
Legionella anisa	FDAARGOS_200	GCA_002082905	4	incomplete	24929	yes	40.24
Legionella anisa	FDAARGOS_200	GCA_002082905	5	incomplete	35490	yes	39.14
Legionella birminghamensis	CDC1407-AL-14	GCA_001467505	0	N/A	N/A	N/A	N/A
Legionella brunensis	ATCC 43878	GCA_001467025	0	N/A	N/A	N/A	N/A
Legionella cherrii	DSM 19213Q775	GCA_000621385	0	N/A	N/A	N/A	N/A
Legionella cherrii	ORW	GCA_001467035	0	N/A	N/A	N/A	N/A
Legionella cincinnatiensis	CDC72-OH-14	GCA_001467545	0	N/A	N/A	N/A	N/A
Legionella clemsonensis	CDC-D5610	GCA_002240035	1	incomplete	21717	yes	36.68
Legionella drancourtii	LLAP12	GCA_000162755	0	N/A	N/A	N/A	N/A
Legionella drozanskii	ATCC 700990	GCA_001467585	0	N/A	N/A	N/A	N/A
Legionellaceae bacterium	SAT2595	GCA_002707695	0	N/A	N/A	N/A	N/A
Legionellaceae bacterium	SP3112	GCA_002719855	0	N/A	N/A	N/A	N/A
Legionellaceae bacterium	UBA2794	GCA_002352055	0	N/A	N/A	N/A	N/A
Legionellaceae bacterium	UBA6148	GCA_002423125	1	incomplete	7278	no	41.05
Legionellaceae bacterium	UBA6148	GCA_002423125	2	incomplete	6350	no	40.37
Legionellaceae bacterium	UBA6148	GCA_002423125	3	incomplete	8147	no	42.97
Legionellales bacterium	NORP184	GCA_002400645	0	N/A	N/A	N/A	N/A
Legionellales bacterium	MED753	GCA_002690495	0	N/A	N/A	N/A	N/A
Legionellales bacterium	MED607	GCA_002691585	0	N/A	N/A	N/A	N/A

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	attL site	Prophage GC content (%)
Legionellales bacterium	MED126	GCA_002692205	0	N/A	N/A	N/A	N/A
Legionellales bacterium	NAT247	GCA_002698645	0	N/A	N/A	N/A	N/A
Legionellales bacterium	NAT241	GCA_002698825	0	N/A	N/A	N/A	N/A
Legionellales bacterium	NAT232	GCA_002698845	0	N/A	N/A	N/A	N/A
Legionellales bacterium	MED841	GCA_002701115	0	N/A	N/A	N/A	N/A
Legionellales bacterium	SAT3214	GCA_002707205	0	N/A	N/A	N/A	N/A
Legionellales bacterium	SAT2734	GCA_002707565	0	N/A	N/A	N/A	N/A
Legionellales bacterium	SAT2787	GCA_002708265	0	N/A	N/A	N/A	N/A
Legionellales bacterium	SP1517	GCA_002709565	0	N/A	N/A	N/A	N/A
Legionellales bacterium	SAT1342	GCA_002714285	0	N/A	N/A	N/A	N/A
Legionellales bacterium	SP3029	GCA_002719415	0	N/A	N/A	N/A	N/A
Legionellales bacterium	SP289	GCA_002721165	0	N/A	N/A	N/A	N/A
Legionellales bacterium	SP228	GCA_002722435	0	N/A	N/A	N/A	N/A
Legionellales bacterium	SP195	GCA_002722995	0	N/A	N/A	N/A	N/A
Legionellales bacterium	RIFCSPHIGHO2_12_FULL_35_11	GCA_001803185	0	N/A	N/A	N/A	N/A
Legionellales bacterium	RIFCSPHIGHO2_12_FULL_37_14	GCA_001804735	0	N/A	N/A	N/A	N/A
Legionellales bacterium	RIFCSPHIGHO2_12_FULL_42_9	GCA_001804755	0	N/A	N/A	N/A	N/A

200

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	attL site	Prophage GC content (%)
Legionellales bacterium	UBA1113	GCA_002313395	1	incomplete	24096	yes	44.85
Legionellales bacterium	UBA1524	GCA_002323365	1	incomplete	6417	no	36.97
Legionellales bacterium	UBA2469	GCA_002342175	1	incomplete	11585	no	38.45
Legionellales bacterium	UBA2469	GCA_002342175	2	incomplete	31257	yes	37
Legionellales bacterium	UBA2469	GCA_002342175	3	incomplete	8746	no	36.55
Legionellales bacterium	UBA2653	GCA_002359535	1	incomplete	11539	no	36.9
Legionellales bacterium	UBA4722	GCA_002404295	1	incomplete	5885	no	40.81
Legionellales bacterium	UBA4722	GCA_002404295	2	incomplete	8218	no	42.34
Legionellales bacterium	UBA4722	GCA_002404295	3	incomplete	8567	no	40.01
Legionellales bacterium	UBA4722	GCA_002404295	4	incomplete	9571	no	40.59
Legionellales bacterium	UBA4759	GCA_002403515	1	incomplete	7223	no	36.89
Legionellales bacterium	UBA4759	GCA_002403515	2	incomplete	14652	no	34.23
Legionellales bacterium	UBA5158	GCA_002413665	1	incomplete	11284	no	36.58
Legionellales bacterium	UBA6500	GCA_002436445	1	incomplete	28684	yes	39.01
Legionellales bacterium	UBA6500	GCA_002436445	2	incomplete	11624	no	43.11
Legionellales bacterium	UBA6500	GCA_002436445	3	incomplete	7773	no	42.42
Legionellales bacterium	UBA6603	GCA_002433445	1	incomplete	6428	no	42.81
Legionellales bacterium	UBA6786	GCA_002452935	1	incomplete	3313	no	40.01
Legionellales bacterium	UBA7366	GCA_002469885	1	incomplete	7217	no	43.96
Legionella endosymbiont	PsAG	GCA_002776555	0	N/A	N/A	N/A	N/A

Table S9. – Continued from overleaf

of Polyplaxserrata

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	*attL* site	Prophage GC content (%)
Legionella erythra	SE-32A-C8	GCA_001467615	0	N/A	N/A	N/A	N/A
Legionella fairfieldensis	ATCC 49588	GCA_000621525	1	incomplete	22278	yes	34.62
Legionella fairfieldensis	ATCC 4958	GCA_000621525	2	incomplete	23262	yes	37.27
Legionella fallonii	LLAP-10	GCA_000953135	1	incomplete	9569	yes	39.3
Legionella fallonii	LLAP-10	GCA_000953135	2	incomplete	9826	no	37.26
Legionella feeleii	WO-44C	GCA_001467625	0	N/A	N/A	N/A	N/A
Legionella feeleii	ATCC 35072	GCA_001648615	1	incomplete	15315	yes	37.45
Legionella feeleii	ATCC 35072	GCA_001648615	2	incomplete	8662	no	44.92
Legionella geestiana	DSM 21217	GCA_000621365	1	incomplete	7663	no	47.79
Legionella geestiana	DSM 21217	GCA_000621365	2	incomplete	10917	yes	51.54
Legionella geestiana	DSM 21217	GCA_000621365	3	incomplete	7070	no	47.63
Legionella geestiana	ATCC 49504	GCA_001467645	0	N/A	N/A	N/A	N/A
Legionella gratiana	Lyon8420412	GCA_001467695	0	N/A	N/A	N/A	N/A
Legionella hackeliae	LHA	GCA_000953655	1	questionable	45590	yes	36.89
Legionella hackeliae	798-PA-HL	GCA_001467705	0	N/A	N/A	N/A	N/A
Legionella israelensis	Bercovier4	GCA_001467785	0	N/A	N/A	N/A	N/A
Legionella jamestowniensis	JA-26-G1-E2	GCA_001467745	0	N/A	N/A	N/A	N/A
Legionella jamestowniensis	974010_12	GCA_001691475	0	N/A	N/A	N/A	N/A
Legionella jamestowniensis	DSM 19215	GCA_900114725	0	N/A	N/A	N/A	N/A

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	*attL* site	Prophage GC content (%)
Legionella jordanis	BL-540	GCA_001467765	0	N/A	N/A	N/A	N/A
Legionella jordanis	ATCC 33623	GCA_001648675	1	incomplete	30955	yes	42.26
Legionella jordanis	ATCC 33623	GCA_001648675	2	incomplete	19729	yes	36.27
Legionella lansingensis	ATCC 49751	GCA_000622185	0	N/A	N/A	N/A	N/A
Legionella lansingensis	ATCC 49751	GCA_001467795	0	N/A	N/A	N/A	N/A
Legionella lansingensis	NCTC 12830	GCA_900187355	0	N/A	N/A	N/A	N/A
Legionella londiniensis	ATCC 49505	GCA_001467825	1	incomplete	18408	yes	41.98
Legionella longbeachae	NSW150	GCA_000091785	1	incomplete	8601	no	37.55
Legionella longbeachae	NSW150	GCA_000091785	2	incomplete	9624	no	36.6
Legionella longbeachae	NSW150	GCA_000091785	3	incomplete	20846	yes	38.22
Legionella longbeachae	NSW150	GCA_000091785	4	incomplete	9927	no	38.35
Legionella longbeachae	D-4968	GCA_000176095	0	N/A	N/A	N/A	N/A
Legionella longbeachae	FDAARGOS_201	GCA_002073455	1	questionable	31298	yes	36.46
Legionella longbeachae	F1157CHC	GCA_002113845	1	questionable	34451	yes	37.01
Legionella longbeachae	F1157CHC	GCA_002113845	2	incomplete	9618	no	36.58
Legionella longbeachae	F1157CHC	GCA_002113845	3	incomplete	20846	yes	38.22
Legionella longbeachae	F1157CHC	GCA_002113845	4	incomplete	8979	no	39.01
Legionella maceachernii	PX-1-G2-E2	GCA_001467845	0	N/A	N/A	N/A	N/A
Legionella maceachernii	ATCC 35300	GCA_900167045	0	N/A	N/A	N/A	N/A
Legionella massiliensis	PRJEB110	GCA_000756695	1	incomplete	24961	yes	39.85

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	*attL* site	Prophage GC content (%)
Legionella massiliensis	PRJEB110	GCA_000756695	2	incomplete	11332	yes	43.25
Legionella massiliensis	PRJEB110	GCA_000756695	3	incomplete	10741	yes	38.19
Legionella massiliensis	PRJEB6598	GCA_000756815	1	incomplete	40575	yes	40.44
Legionella moravica	DSM 19234	GCA_000423305	1	incomplete	7604	no	35.21
Legionella moravica	DSM 19234	GCA_000423305	2	incomplete	14495	yes	38.78
Legionella moravica	DSM 19234	GCA_000423305	3	incomplete	25258	yes	41.74
Legionella moravica	ATCC 43877	GCA_001467865	0	N/A	N/A	N/A	N/A
Legionella nautarum	ATCC 49506	GCA_001467895	0	N/A	N/A	N/A	N/A
Legionella norrlandica	LEGN	GCA_000770585	0	N/A	N/A	N/A	N/A
Legionella oakridgensis	ATCC 33761	GCA_000512355	1	incomplete	7188	no	35.61
Legionella oakridgensis	ATCC 33761	GCA_000512355	2	incomplete	35702	yes	40.19
Legionella oakridgensis	ATCC 33761	GCA_000512355	3	incomplete	10537	no	38.61
Legionella oakridgensis	RV-2-2007	GCA_000512715	0	N/A	N/A	N/A	N/A
Legionella oakridgensis	OakRidge-10	GCA_001467925	0	N/A	N/A	N/A	N/A
Legionella oakridgensis	ATCC 33761	GCA_001648605	1	incomplete	7332	no	35.57
Legionella oakridgensis	ATCC 33761	GCA_001648605	2	incomplete	11059	no	37.45
Legionella parisiensis	PF-209-C-C2	GCA_001467945	0	N/A	N/A	N/A	N/A
Legionella parisiensis	DSM 19216	GCA_001736145	0	N/A	N/A	N/A	N/A
Legionella pneumophila	Philadelphia1	GCA_000008485	1	questionable	30754	yes	37.22
Legionella pneumophila	Paris.	GCA_000048645	1	incomplete	18768	yes	38.52

204

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	attL site	Prophage GC content (%)
Legionella pneumophila	Lens.	GCA_000048665	1	incomplete	13185	yes	38.25
Legionella pneumophila	Corby	GCA_000092545	1	incomplete	19033	yes	38.24
Legionella pneumophila	2300/99Alcoy	GCA_000092625	1	incomplete	19762	yes	36.99
Legionella pneumophila	130b	GCA_000211115	1	incomplete	7746	no	37.95
Legionella pneumophila	130b	GCA_000211115	2	incomplete	7624	no	36.01
Legionella pneumophila	ATCC 43290	GCA_000239175	1	incomplete	18804	yes	38.51
Legionella pneumophila	ATCC 43290	GCA_000239175	2	questionable	13415	yes	37.11
Legionella pneumophila	Hextuple.2q	GCA_000277025	1	incomplete	30754	yes	37.22
Legionella pneumophila	Hextuple.3a	GCA_000277065	1	incomplete	30754	yes	37.22
Legionella pneumophila	HL06041035	GCA_000306845	1	incomplete	31586	yes	37.52
Legionella pneumophila	HL06041035	GCA_000306845	2	incomplete	17415	yes	37.27
Legionella pneumophila	Lorraine	GCA_000306865	1	incomplete	16129	yes	35.67
Legionella pneumophila	Lorraine	GCA_000306865	2	incomplete	7624	no	36.14
Legionella pneumophila	Lorraine	GCA_000306865	3	incomplete	16331	yes	37.47
Legionella pneumophila	Lorraine	GCA_000306865	4	incomplete	22769	yes	40.05
Legionella pneumophila	LPE509	GCA_000347615	1	incomplete	26063	yes	38.18
Legionella pneumophila	Thunder Bay	GCA_000404245	1	questionable	30751	yes	37.15
Legionella pneumophila	121004_c1	GCA_000455845	0	N/A	N/A	N/A	N/A
Legionella pneumophila	Leg01_11	GCA_000465675	0	N/A	N/A	N/A	N/A
Legionella pneumophila	Leg01_53	GCA_000465695	0	N/A	N/A	N/A	N/A

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	attL site	Prophage GC content (%)
Legionella pneumophila	Leg01/20_c1	GCA_000465915	0	N/A	N/A	N/A	N/A
Legionella pneumophila	Leg01_16	GCA_000468235	1	incomplete	6807	no	38.32
Legionella pneumophila	Leg01_16	GCA_000468235	2	incomplete	7137	no	35.74
Legionella pneumophila	Leg01_16	GCA_000468235	3	incomplete	18841	yes	38.43
Legionella pneumophila	Leg01_16	GCA_000468235	4	incomplete	19120	yes	40.28
Legionella pneumophila	Leg01_16	GCA_000468235	5	incomplete	4839	no	41.45
Legionella pneumophila	W1046	GCA_000500125	1	incomplete	8620	no	38.55
Legionella pneumophila	ATCC 43736	GCA_000586075	1	incomplete	19812	yes	37.27
Legionella pneumophila	ATCC 43703	GCA_000586095	1	incomplete	11774	no	38.96
Legionella pneumophila	ATCC 43290	GCA_000586115	1	incomplete	18059	yes	36.71
Legionella pneumophila	ATCC 43283	GCA_000586135	1	incomplete	30304	yes	36.82
Legionella pneumophila	ATCC 43130	GCA_000586155	1	incomplete	21539	yes	40.23
Legionella pneumophila	ATCC 43130	GCA_000586155	2	incomplete	11620	no	37.15
Legionella pneumophila	ATCC 35289	GCA_000586175	1	incomplete	13773	yes	36.78
Legionella pneumophila	ATCC 35251	GCA_000586195	1	incomplete	8740	yes	39.87
Legionella pneumophila	ATCC 35096	GCA_000586215	1	incomplete	30225	yes	37.04
Legionella pneumophila	ATCC 33823	GCA_000586235	1	incomplete	33519	yes	37.77
Legionella pneumophila	ATCC 33823	GCA_000586235	2	incomplete	9526	no	35.72
Legionella pneumophila	ATCC 33737	GCA_000586255	1	incomplete	26590	yes	37.36
Legionella pneumophila	ATCC 33216	GCA_000586275	1	incomplete	21912	yes	38.56

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	attL site	Prophage GC content (%)
Legionella pneumophila	ATCC 33216	GCA_000586275	2	incomplete	9853	no	41.09
Legionella pneumophila	ATCC 33215	GCA_000586295	1	incomplete	16442	yes	37.26
Legionella pneumophila	ATCC 33156	GCA_000586315	1	incomplete	9853	no	41.08
Legionella pneumophila	ATCC 33156	GCA_000586315	2	questionable	8805	no	39.59
Legionella pneumophila	ATCC 33155	GCA_000586335	1	incomplete	33362	yes	38.97
Legionella pneumophila	ATCC 33154	GCA_000586355	1	incomplete	31916	yes	35.39
Legionella pneumophila	ATCC 33152	GCA_000586375	1	incomplete	14074	yes	38.51
Legionella pneumophila	JCM7571	GCA_000614785	1	incomplete	21445	yes	36.37
Legionella pneumophila	TUM 13947	GCA_000694995	1	incomplete	13773	yes	36.78
Legionella pneumophila	TUM 13948	GCA_000695015	1	incomplete	23193	yes	36.07
Legionella pneumophila	Nagoya-1	GCA_000699225	1	incomplete	6908	no	40.43
Legionella pneumophila	Quebec 2012	GCA_000724005	1	incomplete	22610	yes	36.31
Legionella pneumophila	D-5864	GCA_000785905	1	incomplete	14072	yes	38.4
Legionella pneumophila	12_3965	GCA_000823305	1	incomplete	14072	yes	38.4
Legionella pneumophila	12_4030	GCA_000823325	1	incomplete	13185	yes	38.25
Legionella pneumophila	12_4042	GCA_000823345	1	incomplete	13185	yes	38.25
Legionella pneumophila	12_4053	GCA_000823365	1	incomplete	13185	yes	38.25
Legionella pneumophila	12_4054	GCA_000823385	1	incomplete	23056	yes	40.25
Legionella pneumophila	12_4058	GCA_000823405	1	incomplete	13185	yes	38.25
Legionella pneumophila	12_4117	GCA_000823425	1	incomplete	9717	yes	37.72

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	*attL* site	Prophage GC content (%)
Legionella pneumophila	12_4117	GCA_000823425	2	incomplete	24241	yes	39.18
Legionella pneumophila	12_4169	GCA_000823445	1	incomplete	13185	yes	38.25
Legionella pneumophila	12_4240	GCA_000823465	1	incomplete	23056	yes	40.25
Legionella pneumophila	12_4251	GCA_000823485	1	incomplete	30354	yes	36.81
Legionella pneumophila	12_4437	GCA_000823505	1	incomplete	13185	yes	38.25
Legionella pneumophila	12_4480	GCA_000823525	1	incomplete	23056	yes	40.25
Legionella pneumophila	12_4499	GCA_000823545	1	incomplete	13185	yes	38.25
Legionella pneumophila	12_4555	GCA_000823565	1	incomplete	23056	yes	40.25
Legionella pneumophila	12_4561	GCA_000823585	1	incomplete	13185	yes	38.25
Legionella pneumophila	12_4563	GCA_000823605	1	incomplete	23056	yes	40.25
Legionella pneumophila	12_4903	GCA_000823625	1	incomplete	15961	yes	36.12
Legionella pneumophila	12_4903	GCA_000823625	2	incomplete	13185	yes	38.25
Legionella pneumophila	12_4904	GCA_000823645	1	incomplete	15396	yes	37.11
Legionella pneumophila	12_5064	GCA_000823665	1	incomplete	19845	yes	37.87
Legionella pneumophila	12_5064	GCA_000823665	2	incomplete	7624	no	35.99
Legionella pneumophila	12_5223	GCA_000823685	1	incomplete	13185	yes	38.25
Legionella pneumophila	12_5223	GCA_000823685	2	intact	5397	no	44.66
Legionella pneumophila	12_5251	GCA_000823705	1	incomplete	23056	yes	40.25
Legionella pneumophila	12_5251	GCA_000823705	2	intact	5233	no	44.75
Legionella pneumophila	12_5329	GCA_000823725	1	incomplete	23056	yes	40.25

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	attL site	Prophage GC content (%)
Legionella pneumophila	12_5329	GCA_000823725	2	intact	5670	no	43.59
Legionella pneumophila	12_5383	GCA_000823745	1	incomplete	23056	yes	40.25
Legionella pneumophila	12_5383	GCA_000823745	2	intact	5547	no	44.54
Legionella pneumophila	12_5414	GCA_000823765	1	incomplete	23056	yes	40.25
Legionella pneumophila	12_5414	GCA_000823765	2	questionable	28108	yes	37.95
Legionella pneumophila	12_5415	GCA_000823785	1	incomplete	17377	yes	37.2
Legionella pneumophila	12_5415	GCA_000823785	2	intact	3955	no	44.64
Legionella pneumophila	12_5415	GCA_000823785	3	incomplete	23183	yes	38.52
Legionella pneumophila	HO4020049	GCA_000823805	1	intact	5946	no	43.48
Legionella pneumophila	HO80160261	GCA_000823825	1	incomplete	23056	yes	40.25
Legionella pneumophila	HO80160261	GCA_000823825	2	incomplete	15714	yes	37.14
Legionella pneumophila	HO80160261	GCA_000823825	3	intact	33761	yes	38.66
Legionella pneumophila	HO80160262	GCA_000823845	1	incomplete	23056	yes	40.25
Legionella pneumophila	HO80160262	GCA_000823845	2	intact	28916	yes	38.98
Legionella pneumophila	HO80160263	GCA_000823865	1	intact	8077	no	42.67
Legionella pneumophila	HO80160263	GCA_000823865	2	incomplete	13185	yes	38.25
Legionella pneumophila	HO92620872	GCA_000823885	1	incomplete	23056	yes	40.25
Legionella pneumophila	HO92620872	GCA_000823885	2	intact	141545	yes	38.86
Legionella pneumophila	Bnt314	GCA_000950675	0	N/A	N/A	N/A	N/A
Legionella pneumophila	Ofk308	GCA_000950695	0	N/A	N/A	N/A	N/A

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	attL site	Prophage GC content (%)
Legionella pneumophila	Twr292	GCA_000950745	1	intact	5386	no	44.7
Legionella pneumophila	Ymg289	GCA_000953915	0	N/A	N/A	N/A	N/A
Legionella pneumophila	Ymt294	GCA_000953935	0	N/A	N/A	N/A	N/A
Legionella pneumophila	PtVF89/2014	GCA_001549915	0	N/A	N/A	N/A	N/A
Legionella pneumophila	PtVF66/2014	GCA_001549925	0	N/A	N/A	N/A	N/A
Legionella pneumophila	ATCC 33153	GCA_001582135	1	incomplete	18768	yes	38.52
Legionella pneumophila	JNLH86	GCA_001582145	1	incomplete	19531	yes	37.24
Legionella pneumophila	ATCC 35289	GCA_001582155	1	incomplete	18768	yes	38.52
Legionella pneumophila	ATCC 43283	GCA_001582165	1	incomplete	30304	yes	36.82
Legionella pneumophila	ATCC 43130	GCA_001582215	1	incomplete	21539	yes	40.23
Legionella pneumophila	ATCC 43130	GCA_001582215	2	incomplete	11620	no	37.15
Legionella pneumophila	ATCC 35096	GCA_001582225	1	incomplete	32106	yes	37.14
Legionella pneumophila	ATCC 35096	GCA_001582225	2	incomplete	14947	yes	38.43
Legionella pneumophila	BJ-23	GCA_001582235	1	incomplete	18768	yes	38.52
Legionella pneumophila	JX1	GCA_001582245	1	incomplete	7624	no	36.12
Legionella pneumophila	SZ099	GCA_001582295	1	incomplete	32044	yes	37.84
Legionella pneumophila	ATCC 33823	GCA_001582305	1	incomplete	26722	yes	36.58
Legionella pneumophila	ATCC 33823	GCA_001582305	2	incomplete	9526	no	35.72
Legionella pneumophila	ICDC-LP002	GCA_001582315	1	incomplete	24864	yes	37.56
Legionella pneumophila	Qin1	GCA_001582325	1	incomplete	7624	no	36

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	attL site	Prophage GC content (%)
Legionella pneumophila	Sctan	GCA_001582375	1	incomplete	14072	yes	38.4
Legionella pneumophila	SH135	GCA_001582385	1	incomplete	21402	yes	37.11
Legionella pneumophila	ZJ050052	GCA_001582395	1	incomplete	18768	yes	38.52
Legionella pneumophila	FS_10_1101a-3	GCA_001582405	1	incomplete	13186	yes	38.25
Legionella pneumophila	FS_10_1101a-3	GCA_001582405	2	incomplete	23609	yes	39.25
Legionella pneumophila	WD_9_1102a	GCA_001582455	1	incomplete	12321	yes	36.31
Legionella pneumophila	WD_9_1102a	GCA_001582455	2	incomplete	23607	no	34.84
Legionella pneumophila	SH095	GCA_001582465	1	incomplete	7624	no	36.1
Legionella pneumophila	ATCC 33152	GCA_001582475	1	incomplete	16581	yes	38.23
Legionella pneumophila	WD_4_1102a	GCA_001582485	1	incomplete	8048	no	40.24
Legionella pneumophila	WD_4_1102a	GCA_001582485	2	incomplete	7612	no	36.1
Legionella pneumophila	SZ069	GCA_001582535	1	incomplete	19408	yes	36.97
Legionella pneumophila	SZ069	GCA_001582535	2	incomplete	9741	no	39.95
Legionella pneumophila	SZ069	GCA_001582535	3	incomplete	7624	no	36.14
Legionella pneumophila	ATCC 33155	GCA_001582545	1	incomplete	7700	no	37.42
Legionella pneumophila	TL-12	GCA_001582555	1	incomplete	13186	yes	38.25
Legionella pneumophila	ATCC 33154	GCA_001582565	1	incomplete	29333	yes	37.7
Legionella pneumophila	AH104	GCA_001582615	1	incomplete	16254	yes	37.85
Legionella pneumophila	FS_4_1103	GCA_001582625	1	incomplete	9853	no	41.09
Legionella pneumophila	WX2011046	GCA_001582635	1	incomplete	14072	yes	38.4

211

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	attL site	Prophage GC content (%)
Legionella pneumophila	WD_4_1102b-36	GCA_001582645	0	N/A	N/A	N/A	N/A
Legionella pneumophila	ATCC 33216	GCA_001582695	1	incomplete	7624	no	36.04
Legionella pneumophila	ATCC 33216	GCA_001582695	2	incomplete	22608	yes	38.98
Legionella pneumophila	ATCC 33216	GCA_001582695	3	incomplete	10063	no	41.36
Legionella pneumophila	SZ2012007	GCA_001582705	1	incomplete	7624	no	36.04
Legionella pneumophila	SZ2012007	GCA_001582705	2	incomplete	22608	yes	38.98
Legionella pneumophila	SZ2012007	GCA_001582705	3	incomplete	10063	no	41.36
Legionella pneumophila	NX0702	GCA_001582715	1	incomplete	16254	yes	37.85
Legionella pneumophila	BJ7	GCA_001582735	1	incomplete	26999	yes	38.75
Legionella pneumophila	WX2011036	GCA_001582775	1	incomplete	7624	no	36.14
Legionella pneumophila	WX2011036	GCA_001582775	2	incomplete	9853	no	41.09
Legionella pneumophila	BJ-9	GCA_001582785	1	incomplete	13773	yes	36.78
Legionella pneumophila	WX2012012	GCA_001582795	1	incomplete	13185	yes	38.26
Legionella pneumophila	ATCC 33156	GCA_001582825	1	incomplete	8647	no	40.89
Legionella pneumophila	ATCC 33156	GCA_001582825	2	incomplete	6910	no	40.08
Legionella pneumophila	SZ059	GCA_001582855	1	incomplete	7624	no	36
Legionella pneumophila	SZ059	GCA_001582855	2	incomplete	7924	no	41.21
Legionella pneumophila	SH202	GCA_001582865	1	incomplete	32443	yes	37.32
Legionella pneumophila	WX2011029	GCA_001583565	1	incomplete	10923	yes	38.38
Legionella pneumophila	WX2011029	GCA_001583565	2	incomplete	34921	yes	37.82

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	attL site	Prophage GC content (%)
Legionella pneumophila	WX2011029	GCA_001583565	3	questionable	8974	no	38.09
Legionella pneumophila	SZ2012006	GCA_001583575	1	incomplete	28817	yes	37.22
Legionella pneumophila	SZ2012006	GCA_001583575	2	incomplete	9741	no	39.95
Legionella pneumophila	SZ2012006	GCA_001583575	3	incomplete	7624	no	36.04
Legionella pneumophila	Yu237	GCA_001583585	1	incomplete	18768	yes	38.52
Legionella pneumophila	Hu6	GCA_001583595	1	incomplete	18768	yes	38.52
Legionella pneumophila	SZ026	GCA_001583645	1	incomplete	20051	yes	38.97
Legionella pneumophila	SH0030_1	GCA_001583655	1	incomplete	34036	yes	38.81
Legionella pneumophila	SH0030_1	GCA_001583655	2	incomplete	5655	no	41.23
Legionella pneumophila	F-4185	GCA_001590615	1	incomplete	26590	yes	37.36
Legionella pneumophila	D-7158	GCA_001590645	1	incomplete	26590	yes	37.36
Legionella pneumophila	D-7119	GCA_001590695	1	incomplete	26590	yes	37.36
Legionella pneumophila	Toronto-2005	GCA_001592705	1	questionable	35824	yes	35.48
Legionella pneumophila	NY16	GCA_001600895	1	incomplete	27505	yes	37.29
Legionella pneumophila	NY24	GCA_001600905	1	incomplete	24340	yes	35.23
Legionella pneumophila	NY19	GCA_001600915	1	incomplete	21461	yes	36.86
Legionella pneumophila	NY18	GCA_001600925	1	incomplete	25327	yes	39.26
Legionella pneumophila	NY18	GCA_001600925	2	questionable	10325	no	36.66
Legionella pneumophila	NY23	GCA_001600975	1	incomplete	16552	yes	38.32
Legionella pneumophila	NY30	GCA_001600985	1	incomplete	11873	yes	37.99

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	*attL* site	Prophage GC content (%)
Legionella pneumophila	NY26	GCA_001600995	1	incomplete	8154	no	38.31
Legionella pneumophila	NY25	GCA_001601005	1	incomplete	6811	no	36.82
Legionella pneumophila	NY25	GCA_001601005	2	incomplete	7169	yes	41.16
Legionella pneumophila	NY17	GCA_001601055	1	incomplete	26091	yes	38.43
Legionella pneumophila	NY27	GCA_001601075	1	incomplete	15170	yes	38.84
Legionella pneumophila	NY20	GCA_001601085	1	incomplete	14074	yes	38.51
Legionella pneumophila	NY21	GCA_001601115	1	incomplete	21731	yes	39.26
Legionella pneumophila	NY22	GCA_001601135	1	incomplete	7624	no	36.01
Legionella pneumophila	NY22	GCA_001601135	2	incomplete	7236	no	40.74
Legionella pneumophila	NY29	GCA_001601155	1	incomplete	21361	yes	39.73
Legionella pneumophila	NY1	GCA_001601165	1	incomplete	8936	no	41.47
Legionella pneumophila	NY1	GCA_001601165	2	incomplete	7624	no	36.01
Legionella pneumophila	NY28	GCA_001601175	1	incomplete	17111	yes	38.2
Legionella pneumophila	NY28	GCA_001601175	2	incomplete	6227	no	38.2
Legionella pneumophila	NY3	GCA_001601215	1	incomplete	7624	no	36.01
Legionella pneumophila	NY4	GCA_001601235	1	incomplete	10523	no	39.98
Legionella pneumophila	NY2	GCA_001601245	1	incomplete	6909	no	40.65
Legionella pneumophila	NY2	GCA_001601245	2	incomplete	7624	no	36.01
Legionella pneumophila	NY5	GCA_001601325	1	incomplete	13185	yes	38.25
Legionella pneumophila	NY6	GCA_001601355	1	incomplete	13185	yes	38.25

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	attL site	Prophage GC content (%)
Legionella pneumophila	NY9	GCA_001601375	1	incomplete	7624	no	36.01
Legionella pneumophila	NY13	GCA_001601395	1	incomplete	7624	no	36
Legionella pneumophila	NY14	GCA_001601415	1	incomplete	7624	no	36
Legionella pneumophila	NY14	GCA_001601415	2	incomplete	9745	no	43.38
Legionella pneumophila	NY12	GCA_001601425	1	incomplete	11663	no	39.63
Legionella pneumophila	NY12	GCA_001601425	2	incomplete	7624	no	36.01
Legionella pneumophila	NY15	GCA_001601455	1	incomplete	23396	yes	35.22
Legionella pneumophila	NY11	GCA_001601475	1	incomplete	7624	no	36.01
Legionella pneumophila	NY10	GCA_001601485	1	incomplete	7703	no	40.03
Legionella pneumophila	NY7	GCA_001601505	1	incomplete	18447	yes	38.67
Legionella pneumophila	NY7	GCA_001601505	2	incomplete	7624	no	36.16
Legionella pneumophila	NY8	GCA_001601535	1	incomplete	13185	yes	38.25
Legionella pneumophila	L10-023	GCA_001610735	1	incomplete	23824	yes	37.84
Legionella pneumophila	PtVFX/2014	GCA_001639045	0	N/A	N/A	N/A	N/A
Legionella pneumophila	D-7632	GCA_001652645	1	incomplete	26417	yes	36.69
Legionella pneumophila	D-7632	GCA_001652645	2	incomplete	24275	yes	37.32
Legionella pneumophila	D-7631	GCA_001652665	1	incomplete	26417	yes	36.69
Legionella pneumophila	D-7631	GCA_001652665	2	incomplete	24275	yes	37.32
Legionella pneumophila	D-7630	GCA_001652685	1	incomplete	26417	yes	36.69
Legionella pneumophila	D-7630	GCA_001652685	2	incomplete	24275	yes	37.32

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	*attL* site	Prophage GC content (%)
Legionella pneumophila	OLDA	GCA_001677075	1	incomplete	18768	yes	38.52
Legionella pneumophila	Pontiac	GCA_001677115	1	incomplete	30682	yes	36.66
Legionella pneumophila	Lp01	GCA_001685545	1	incomplete	36109	yes	36.58
Legionella pneumophila	Lp01	GCA_001685545	2	incomplete	7236	no	40.74
Legionella pneumophila	Lp01	GCA_001685545	3	incomplete	7624	no	36.01
Legionella pneumophila	Lp01	GCA_001685545	4	incomplete	7694	no	37.69
Legionella pneumophila	JR32	GCA_001685575	1	incomplete	7236	no	40.74
Legionella pneumophila	JR32	GCA_001685575	2	incomplete	36109	yes	36.58
Legionella pneumophila	JR32	GCA_001685575	3	incomplete	7624	no	36.01
Legionella pneumophila	JR32	GCA_001685575	4	incomplete	7694	no	37.69
Legionella pneumophila	E10_P	GCA_001752705	1	incomplete	30680	yes	37.21
Legionella pneumophila	E11_U	GCA_001752725	1	incomplete	30754	yes	37.22
Legionella pneumophila	Philadelphia_1_ATCC	GCA_001752745	1	incomplete	30754	yes	37.22
Legionella pneumophila	Philadelphia_1_CDC	GCA_001752765	1	incomplete	30754	yes	37.22
Legionella pneumophila	Philadelphia_2	GCA_001752785	1	incomplete	30754	yes	37.22
Legionella pneumophila	Philadelphia_3	GCA_001752805	1	incomplete	30754	yes	37.22
Legionella pneumophila	Philadelphia_4	GCA_001752825	1	incomplete	30754	yes	37.22
Legionella pneumophila	C1_S	GCA_001752845	1	incomplete	30754	yes	37.22
Legionella pneumophila	C2_S	GCA_001752865	1	incomplete	30754	yes	37.22
Legionella pneumophila	C3_O	GCA_001752885	1	incomplete	30754	yes	37.22

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	attL site	Prophage GC content (%)
Legionella pneumophila	C4_S	GCA_001752905	1	incomplete	30754	yes	37.22
Legionella pneumophila	C5_P	GCA_001752925	1	incomplete	30681	yes	37.21
Legionella pneumophila	C6_S	GCA_001752945	1	incomplete	30754	yes	37.22
Legionella pneumophila	C7_O	GCA_001752965	1	incomplete	30752	yes	37.22
Legionella pneumophila	C8_S	GCA_001753065	1	incomplete	30754	yes	37.22
Legionella pneumophila	C9_S	GCA_001753085	1	incomplete	30754	yes	37.22
Legionella pneumophila	C10_S	GCA_001753105	1	incomplete	30754	yes	37.22
Legionella pneumophila	C11_O	GCA_001753125	1	incomplete	30754	yes	37.22
Legionella pneumophila	E1_P	GCA_001753145	1	incomplete	30754	yes	37.22
Legionella pneumophila	E2_N	GCA_001753265	1	incomplete	30754	yes	37.22
Legionella pneumophila	E3_N	GCA_001753285	1	incomplete	26063	yes	38.18
Legionella pneumophila	E4_N	GCA_001753305	1	incomplete	30754	yes	37.22
Legionella pneumophila	E5_N	GCA_001753325	1	questionable	32755	yes	37.16
Legionella pneumophila	E6_N	GCA_001753345	1	incomplete	30754	yes	37.22
Legionella pneumophila	E7_O	GCA_001753365	1	incomplete	30754	yes	37.22
Legionella pneumophila	E8_O	GCA_001753385	1	incomplete	30754	yes	37.22
Legionella pneumophila	E9_O	GCA_001753405	1	incomplete	30754	yes	37.22
Legionella pneumophila	FFI102	GCA_001766275	1	incomplete	22871	yes	37.06
Legionella pneumophila	FFI102	GCA_001766275	2	incomplete	10860	no	37.4
Legionella pneumophila	FFI103	GCA_001766295	1	incomplete	10860	no	37.4

217

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	attL site	Prophage GC content (%)
Legionella pneumophila	FFI104	GCA_001766315	1	incomplete	13098	yes	38.22
Legionella pneumophila	FFI104	GCA_001766315	2	incomplete	9651	no	36.31
Legionella pneumophila	FFI105	GCA_001766335	1	incomplete	9651	no	36.31
Legionella pneumophila	FFI105	GCA_001766335	2	incomplete	13185	yes	38.25
Legionella pneumophila	FFI329	GCA_001766355	1	incomplete	10860	no	37.4
Legionella pneumophila	FFI337	GCA_001766375	1	incomplete	9651	no	36.31
Legionella pneumophila	FFI337	GCA_001766375	2	incomplete	13185	yes	38.25
Legionella pneumophila	Detroit-1	GCA_001886795	1	incomplete	7624	no	36.13
Legionella pneumophila	Detroit-1	GCA_001886795	2	incomplete	36662	yes	36.73
Legionella pneumophila	Detroit-1	GCA_001886795	3	incomplete	31292	yes	39.72
Legionella pneumophila	Detroit-1	GCA_001886795	4	incomplete	11995	no	38.6
Legionella pneumophila	Detroit-1	GCA_001886795	5	incomplete	35153	yes	40.32
Legionella pneumophila	Dallas1E	GCA_001886835	1	incomplete	5742	no	39.02
Legionella pneumophila	Dallas1E	GCA_001886835	2	incomplete	32044	yes	37.84
Legionella pneumophila	Dallas1E	GCA_001886835	3	incomplete	11995	no	38.6
Legionella pneumophila	Dallas1E	GCA_001886835	4	incomplete	35153	yes	40.32
Legionella pneumophila	Philadelphia-1	GCA_001941585	1	incomplete	30754	yes	37.22
Legionella pneumophila	F4469	GCA_001969405	1	incomplete	26417	yes	36.69
Legionella pneumophila	F4469	GCA_001969405	2	incomplete	26005	yes	37.62
Legionella pneumophila	F4468	GCA_001989475	1	incomplete	26417	yes	36.69

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	attL site	Prophage GC content (%)
Legionella pneumophila	F4468	GCA_001989475	2	incomplete	26005	yes	37.62
Legionella pneumophila	ATCC 43290	GCA_001997245	0	N/A	N/A	N/A	N/A
Legionella pneumophila	MississaugaCont14	GCA_002002625	0	N/A	N/A	N/A	N/A
Legionella pneumophila	SudburyCont18	GCA_002002645	0	N/A	N/A	N/A	N/A
Legionella pneumophila	FDAARGOS_202	GCA_002082955	1	incomplete	30765	yes	37.2
Legionella pneumophila	D6026	GCA_002813715	1	incomplete	7625	no	36.14
Legionella pneumophila	D6026	GCA_002813715	2	incomplete	32044	yes	37.84
Legionella pneumophila	D6026	GCA_002813715	3	incomplete	11995	no	38.6
Legionella pneumophila	D6026	GCA_002813715	4	incomplete	35153	yes	40.32
Legionella pneumophila	D5945	GCA_002813735	1	incomplete	7625	no	36.14
Legionella pneumophila	D5945	GCA_002813735	2	incomplete	32044	yes	37.84
Legionella pneumophila	D5945	GCA_002813735	3	incomplete	11995	no	38.6
Legionella pneumophila	D5945	GCA_002813735	4	incomplete	35153	yes	40.32
Legionella pneumophila	LG63	GCA_002934165	0	N/A	N/A	N/A	N/A
Legionella pneumophila	LG59	GCA_002934185	0	N/A	N/A	N/A	N/A
Legionella pneumophila	LG61	GCA_002934205	0	N/A	N/A	N/A	N/A
Legionella pneumophila	LG57	GCA_002967055	0	N/A	N/A	N/A	N/A
Legionella pneumophila	D-4954	GCA_003003515	1	incomplete	23511	yes	39
Legionella pneumophila	D-4954	GCA_003003515	2	incomplete	33704	yes	36.32
Legionella pneumophila	NY24(D-7706)	GCA_003003535	1	incomplete	35422	yes	34.82

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	attL site	Prophage GC content (%)
Legionella pneumophila	NY23(D-7705)	GCA_003003555	1	incomplete	35422	yes	34.82
Legionella pneumophila	D-3137	GCA_003003595	1	incomplete	32044	yes	37.84
Legionella pneumophila	D-5387	GCA_003003675	1	incomplete	7624	no	36.2
Legionella pneumophila	D-5387	GCA_003003675	2	incomplete	6965	no	37.74
Legionella pneumophila	D-5387	GCA_003003675	3	incomplete	32044	yes	37.84
Legionella pneumophila	D-5387	GCA_003003675	4	incomplete	11995	no	38.6
Legionella pneumophila	D-5387	GCA_003003675	5	incomplete	20076	yes	38.56
Legionella pneumophila	LosAngeles1(D-7696)	GCA_003003755	1	incomplete	6759	no	42.54
Legionella pneumophila	LosAngeles1(D-7696)	GCA_003003755	2	incomplete	9853	no	41.08
Legionella pneumophila	LosAngeles1(D-7696)	GCA_003003755	3	incomplete	22611	yes	39.37
Legionella pneumophila	Bellingham1	GCA_003003815	1	incomplete	27921	yes	38.58
Legionella pneumophila	Lansing3	GCA_003003865	1	incomplete	35231	yes	39.86
Legionella pneumophila	D-5744	GCA_003003955	1	incomplete	32044	yes	37.84
Legionella pneumophila	D-7708	GCA_003004065	1	incomplete	19472	yes	35.51
Legionella pneumophila	D-7708	GCA_003004065	2	incomplete	8274	no	40.29
Legionella pneumophila	U8W(D-7160)	GCA_003004115	1	incomplete	26590	yes	37.36
Legionella pneumophila	Knoxville1	GCA_003004135	1	incomplete	27939	yes	38.62
Legionella pneumophila	Burlington1	GCA_003004155	1	incomplete	30754	yes	37.22
Legionella pneumophila	Birmingham1	GCA_003004175	1	incomplete	13185	yes	38.25
Legionella pneumophila	D-5265	GCA_003004195	1	incomplete	10259	yes	36.91

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	*attL* site	Prophage GC content (%)
Legionella pneumophila	D-5265	GCA_003004195	2	incomplete	20285	yes	36.08
Legionella pneumophila	D-4058	GCA_003004215	1	incomplete	32044	yes	37.84
Legionella pneumophila	D-4040	GCA_003004235	1	incomplete	19760	yes	37.3
Legionella pneumophila	D-4040	GCA_003004235	2	incomplete	20288	yes	36.07
Legionella pneumophila	F-4198	GCA_003004255	1	incomplete	7624	no	36.13
Legionella pneumophila	F-4198	GCA_003004255	2	incomplete	6965	no	37.73
Legionella pneumophila	F-4198	GCA_003004255	3	incomplete	32044	yes	37.84
Legionella pneumophila	F-4198	GCA_003004255	4	incomplete	11995	no	38.6
Legionella pneumophila	F-4198	GCA_003004255	5	incomplete	20076	yes	38.56
Legionella pneumophila	F-4198	GCA_003004255	6	incomplete	12129	no	39.81
Legionella pneumophila	Flint2 (D-7477)	GCA_003004275	1	incomplete	7624	no	36.01
Legionella pneumophila	Flint2 (D-7477)	GCA_003004275	2	incomplete	20861	yes	39.2
Legionella pneumophila	Allentown1	GCA_003004295	1	incomplete	7767	no	38.9
Legionella pneumophila	Allentown1	GCA_003004295	2	incomplete	7624	no	36.12
Legionella pneumophila	Allentown1	GCA_003004295	3	incomplete	10417	yes	34.94
Legionella pneumophila	Allentown1	GCA_003004295	4	incomplete	17179	yes	40.31
Legionella pneumophila	Allentown1	GCA_003004295	5	incomplete	22547	yes	39.01
Legionella pneumophila	Albuquerque1	GCA_003004315	1	incomplete	7624	no	36.01
Legionella pneumophila	Albuquerque1	GCA_003004315	2	incomplete	7301	no	40.73
Legionella pneumophila	D-7787	GCA_003004335	1	incomplete	26613	yes	35.48

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	*attL* site	Prophage GC content (%)
Legionella pneumophila	2531STDY5467311	GCA_900048905	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467317	GCA_900048915	1	incomplete	12252	yes	38.22
Legionella pneumophila	2531STDY5467382	GCA_900048925	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467383	GCA_900048935	1	incomplete	23766	yes	39.8
Legionella pneumophila	2532STDY5467498	GCA_900048945	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467593	GCA_900048955	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467557	GCA_900048965	1	incomplete	7624	no	36.14
Legionella pneumophila	2531STDY5467360	GCA_900049245	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467402	GCA_900049255	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467403	GCA_900049265	1	incomplete	6951	no	39.37
Legionella pneumophila	2531STDY5467467	GCA_900049275	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467623	GCA_900049285	1	incomplete	7624	no	36.13
Legionella pneumophila	2532STDY5467622	GCA_900049295	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467621	GCA_900049305	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467550	GCA_900049375	1	incomplete	7624	no	36.14
Legionella pneumophila	2531STDY5467302	GCA_900050175	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467306	GCA_900050185	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467335	GCA_900050195	1	incomplete	23824	yes	37.84
Legionella pneumophila	2531STDY5467339	GCA_900050205	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467340	GCA_900050215	1	incomplete	23056	yes	40.25

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	attL site	Prophage GC content (%)
Legionella pneumophila	2531STDY5467437	GCA_900050225	1	incomplete	14037	yes	35.46
Legionella pneumophila	2531STDY5467437	GCA_900050225	2	incomplete	30117	yes	39.58
Legionella pneumophila	2531STDY5467318	GCA_900050235	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467555	GCA_900050245	1	incomplete	7624	no	36.14
Legionella pneumophila	2531STDY5467468	GCA_900050555	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467477	GCA_900050565	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467456	GCA_900050975	1	incomplete	13185	yes	38.25
Legionella pneumophila	2531STDY5467363	GCA_900050985	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467446	GCA_900050995	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467663	GCA_900051005	1	incomplete	7624	no	36.14
Legionella pneumophila	2531STDY5467451	GCA_900051015	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467634	GCA_900051025	1	incomplete	11727	yes	38.73
Legionella pneumophila	2531STDY5467282	GCA_900051655	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467321	GCA_900051665	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467338	GCA_900051675	1	incomplete	11727	yes	38.73
Legionella pneumophila	2531STDY5467385	GCA_900051685	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467348	GCA_900051695	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467330	GCA_900051705	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467409	GCA_900051715	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467463	GCA_900051725	1	incomplete	13185	yes	38.25

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	attL site	Prophage GC content (%)
Legionella pneumophila	2532STDY5467499	GCA_900051735	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467554	GCA_900051745	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467648	GCA_900051755	1	incomplete	11727	yes	38.73
Legionella pneumophila	2531STDY5467305	GCA_900052055	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467324	GCA_900052065	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467290	GCA_900052075	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467299	GCA_900052085	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467390	GCA_900052095	1	incomplete	14400	no	39.85
Legionella pneumophila	2531STDY5467390	GCA_900052095	2	incomplete	16254	yes	37.85
Legionella pneumophila	2532STDY5467476	GCA_900052105	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467304	GCA_900052255	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467292	GCA_900052265	1	incomplete	7624	no	36.14
Legionella pneumophila	2531STDY5467369	GCA_900052275	1	incomplete	14400	no	39.85
Legionella pneumophila	2532STDY5467531	GCA_900052285	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467529	GCA_900052295	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467474	GCA_900052305	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467478	GCA_900052315	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467604	GCA_900052325	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467588	GCA_900052335	1	incomplete	7624	no	36.14
Legionella pneumophila	2531STDY5467293	GCA_900052905	0	N/A	N/A	N/A	N/A

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	attL site	Prophage GC content (%)
Legionella pneumophila	2531STDY5467284	GCA_900052915	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467322	GCA_900052925	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467392	GCA_900052935	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467560	GCA_900052945	1	incomplete	7624	no	36.14
Legionella pneumophila	2531STDY5467283	GCA_900053335	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467301	GCA_900053345	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467329	GCA_900053355	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467287	GCA_900053365	1	incomplete	6951	no	39.37
Legionella pneumophila	2531STDY5467434	GCA_900053375	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467481	GCA_900053385	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467482	GCA_900053395	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467485	GCA_900053405	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467337	GCA_900053415	1	incomplete	12094	yes	37.13
Legionella pneumophila	2532STDY5467573	GCA_900053425	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467568	GCA_900053435	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467595	GCA_900053445	1	incomplete	7624	no	36.14
Legionella pneumophila	2531STDY5467288	GCA_900053665	1	incomplete	27748	yes	38.15
Legionella pneumophila	2531STDY5467380	GCA_900053675	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467326	GCA_900053685	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467297	GCA_900053695	0	N/A	N/A	N/A	N/A

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	attL site	Prophage GC content (%)
Legionella pneumophila	2531STDY5467376	GCA_900053705	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467484	GCA_900053715	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467519	GCA_900053725	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467520	GCA_900053735	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467640	GCA_900053745	1	incomplete	19880	yes	37
Legionella pneumophila	2531STDY5467469	GCA_900054155	1	incomplete	20447	yes	34.72
Legionella pneumophila	2532STDY5467626	GCA_900054165	1	incomplete	7624	no	36.14
Legionella pneumophila	2531STDY5467312	GCA_900054385	1	incomplete	7624	no	36.14
Legionella pneumophila	2531STDY5467310	GCA_900054395	1	incomplete	22610	yes	36.31
Legionella pneumophila	2531STDY5467331	GCA_900054405	1	incomplete	12094	yes	37.13
Legionella pneumophila	2532STDY5467532	GCA_900054415	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467475	GCA_900054425	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467653	GCA_900054435	1	incomplete	27433	yes	37.42
Legionella pneumophila	2532STDY5467583	GCA_900054645	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467613	GCA_900054655	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467610	GCA_900054665	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467611	GCA_900054675	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467619	GCA_900054685	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467647	GCA_900054695	1	incomplete	11727	yes	38.73
Legionella pneumophila	2532STDY5467500	GCA_900055195	0	N/A	N/A	N/A	N/A

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	*attL* site	Prophage GC content (%)
Legionella pneumophila	2532STDY5467502	GCA_900055205	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467582	GCA_900055215	1	incomplete	7624	no	36.14
Legionella pneumophila	2531STDY5467350	GCA_900055575	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467460	GCA_900055585	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467491	GCA_900055595	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467580	GCA_900055605	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467553	GCA_900055615	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467571	GCA_900055625	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467607	GCA_900055635	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467516	GCA_900056185	1	incomplete	19049	yes	40.26
Legionella pneumophila	2532STDY5467516	GCA_900056185	2	incomplete	25002	yes	35.62
Legionella pneumophila	2532STDY5467551	GCA_900056195	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467617	GCA_900056205	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467616	GCA_900056215	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467633	GCA_900056225	1	incomplete	19880	yes	37
Legionella pneumophila	2532STDY5467639	GCA_900056235	1	incomplete	23824	yes	37.84
Legionella pneumophila	2532STDY5467635	GCA_900056245	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467332	GCA_900056645	1	incomplete	13185	yes	38.25
Legionella pneumophila	2531STDY5467334	GCA_900056655	1	incomplete	21473	yes	36.7
Legionella pneumophila	2531STDY5467334	GCA_900056655	2	incomplete	7721	no	37.66

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	attL site	Prophage GC content (%)
Legionella pneumophila	2531STDY5467315	GCA_900056915	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467401	GCA_900056925	1	incomplete	7624	no	36.01
Legionella pneumophila	2532STDY5467541	GCA_900056935	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467540	GCA_900056945	1	incomplete	7624	no	36.14
Legionella pneumophila	2531STDY5467316	GCA_900057235	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467356	GCA_900057245	1	incomplete	6951	no	39.37
Legionella pneumophila	2531STDY5467440	GCA_900057255	1	incomplete	7004	yes	36.8
Legionella pneumophila	2531STDY5467440	GCA_900057255	2	incomplete	7721	no	37.66
Legionella pneumophila	2532STDY5467592	GCA_900057265	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467645	GCA_900057275	1	incomplete	11727	yes	38.73
Legionella pneumophila	2531STDY5467319	GCA_900057545	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467336	GCA_900057555	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467435	GCA_900057565	1	incomplete	7236	no	40.74
Legionella pneumophila	2531STDY5467435	GCA_900057565	2	incomplete	25002	yes	35.62
Legionella pneumophila	2532STDY5467547	GCA_900057575	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467543	GCA_900057585	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467574	GCA_900057595	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467625	GCA_900057605	1	incomplete	7624	no	36.14
Legionella pneumophila	2531STDY5467449	GCA_900057615	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467631	GCA_900057625	1	incomplete	23824	yes	37.84

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	*attL* site	Prophage GC content (%)
Legionella pneumophila	2531STDY5467327	GCA_900057735	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467358	GCA_900057745	1	incomplete	19473	yes	36.87
Legionella pneumophila	2531STDY5467358	GCA_900057745	2	incomplete	8579	no	38.24
Legionella pneumophila	2531STDY5467399	GCA_900057755	1	incomplete	20748	yes	38.67
Legionella pneumophila	2532STDY5467492	GCA_900057765	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467493	GCA_900057775	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467576	GCA_900057785	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467514	GCA_900058335	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467509	GCA_900058345	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467652	GCA_900058355	1	incomplete	11727	yes	38.73
Legionella pneumophila	2531STDY5467379	GCA_900058555	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467395	GCA_900058565	1	incomplete	14400	no	39.85
Legionella pneumophila	2531STDY5467396	GCA_900058575	1	incomplete	14400	no	39.85
Legionella pneumophila	2532STDY5467522	GCA_900058585	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467527	GCA_900058595	1	incomplete	10959	yes	34.53
Legionella pneumophila	2532STDY5467527	GCA_900058595	2	incomplete	7236	no	40.74
Legionella pneumophila	2532STDY5467527	GCA_900058595	3	incomplete	7624	no	36.13
Legionella pneumophila	2532STDY5467563	GCA_900058605	1	incomplete	7624	no	36.14
Legionella pneumophila	2531STDY5467291	GCA_900058805	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467364	GCA_900058815	0	N/A	N/A	N/A	N/A

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	attL site	Prophage GC content (%)
Legionella pneumophila	2532STDY5467600	GCA_900058825	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467548	GCA_900058835	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467637	GCA_900058845	1	incomplete	23824	yes	37.84
Legionella pneumophila	2532STDY5467642	GCA_900058855	1	incomplete	23824	yes	37.84
Legionella pneumophila	2532STDY5467641	GCA_900058865	1	incomplete	26067	yes	38.21
Legionella pneumophila	2531STDY5467406	GCA_900059575	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467362	GCA_900059585	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467603	GCA_900059595	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467614	GCA_900059605	1	incomplete	7624	no	36.14
Legionella pneumophila	2531STDY5467391	GCA_900059935	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467367	GCA_900059945	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467397	GCA_900059955	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467285	GCA_900059965	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467365	GCA_900059975	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467512	GCA_900059985	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467447	GCA_900059995	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467535	GCA_900060005	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467596	GCA_900060015	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467661	GCA_900060025	1	incomplete	23824	yes	37.84
Legionella pneumophila	2531STDY5467307	GCA_900060195	0	N/A	N/A	N/A	N/A

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	attL site	Prophage GC content (%)
Legionella pneumophila	2531STDY5467295	GCA_900060205	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467373	GCA_900060365	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467394	GCA_900060375	1	incomplete	7624	no	36.12
Legionella pneumophila	2531STDY5467404	GCA_900060385	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467490	GCA_900060395	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467494	GCA_900060405	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467542	GCA_900060415	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467549	GCA_900060425	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467544	GCA_900060435	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467537	GCA_900060445	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467584	GCA_900060455	1	incomplete	7624	no	36.14
Legionella pneumophila	2531STDY5467387	GCA_900060695	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467388	GCA_900060705	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467366	GCA_900060715	1	incomplete	14400	no	39.86
Legionella pneumophila	2531STDY5467368	GCA_900060725	1	incomplete	30304	yes	36.82
Legionella pneumophila	2531STDY5467370	GCA_900060735	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467400	GCA_900060745	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467495	GCA_900060755	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467517	GCA_900060765	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467627	GCA_900060775	1	incomplete	7624	no	36.14

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	attL site	Prophage GC content (%)
Legionella pneumophila	2532STDY5467556	GCA_900060785	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467651	GCA_900060795	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467314	GCA_900061455	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467320	GCA_900061465	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467381	GCA_900061475	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467289	GCA_900061485	1	incomplete	20245	yes	36.02
Legionella pneumophila	2531STDY5467303	GCA_900061495	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467298	GCA_900061505	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467344	GCA_900061515	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467349	GCA_900061525	1	incomplete	7624	no	36.18
Legionella pneumophila	2531STDY5467351	GCA_900061535	1	incomplete	7624	no	36.14
Legionella pneumophila	2531STDY5467352	GCA_900061545	1	incomplete	23824	yes	37.84
Legionella pneumophila	2531STDY5467353	GCA_900061555	1	incomplete	13185	yes	38.25
Legionella pneumophila	2531STDY5467357	GCA_900061565	1	incomplete	23824	yes	37.84
Legionella pneumophila	2531STDY5467389	GCA_900061575	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467374	GCA_900061585	1	incomplete	14400	no	39.85
Legionella pneumophila	2531STDY5467398	GCA_900061595	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467412	GCA_900061605	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467439	GCA_900061615	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467436	GCA_900061625	1	incomplete	10959	yes	34.53

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	*attL* site	Prophage GC content (%)
Legionella pneumophila	2531STDY5467436	GCA_900061625	2	incomplete	7236	no	40.74
Legionella pneumophila	2531STDY5467426	GCA_900061635	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467453	GCA_900061645	1	incomplete	13185	yes	38.25
Legionella pneumophila	2531STDY5467455	GCA_900061655	1	incomplete	13185	yes	38.25
Legionella pneumophila	2531STDY5467458	GCA_900061665	1	incomplete	13185	yes	38.25
Legionella pneumophila	2531STDY5467459	GCA_900061675	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467461	GCA_900061685	1	incomplete	29647	yes	38.51
Legionella pneumophila	2531STDY5467462	GCA_900061695	1	incomplete	29046	yes	38.63
Legionella pneumophila	2531STDY5467457	GCA_900061705	1	incomplete	17159	yes	37.51
Legionella pneumophila	2531STDY5467464	GCA_900061715	1	incomplete	13185	yes	38.25
Legionella pneumophila	2531STDY5467465	GCA_900061725	1	incomplete	29647	yes	38.51
Legionella pneumophila	2532STDY5467525	GCA_900061935	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467496	GCA_900061945	1	incomplete	7694	no	37.69
Legionella pneumophila	2531STDY5467378	GCA_900061975	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467386	GCA_900062315	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467511	GCA_900062325	0	N/A	N/A	N/A	N/A
Legionella pneumophila	OLDA_1	GCA_900062335	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467480	GCA_900062345	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467513	GCA_900062355	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467515	GCA_900062365	0	N/A	N/A	N/A	N/A

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	attL site	Prophage GC content (%)
Legionella pneumophila	2532STDY5467523	GCA_900062375	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467524	GCA_900062385	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467530	GCA_900062395	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467483	GCA_900062405	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467624	GCA_900062415	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467503	GCA_900062425	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467504	GCA_900062435	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467505	GCA_900062445	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467507	GCA_900062455	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467570	GCA_900062465	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467510	GCA_900062475	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467518	GCA_900062485	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467416	GCA_900062495	1	incomplete	12094	yes	37.13
Legionella pneumophila	2531STDY5467345	GCA_900062505	1	incomplete	6951	no	39.37
Legionella pneumophila	2531STDY5467417	GCA_900062515	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467450	GCA_900062525	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467575	GCA_900062535	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467591	GCA_900062545	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467578	GCA_900062555	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467581	GCA_900062565	1	incomplete	7624	no	36.14

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	*attL* site	Prophage GC content (%)
Legionella pneumophila	2532STDY5467577	GCA_900062575	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467546	GCA_900062585	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467565	GCA_900062595	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467561	GCA_900062605	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467559	GCA_900062615	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467628	GCA_900062625	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467538	GCA_900062635	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467545	GCA_900062645	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467602	GCA_900062655	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467562	GCA_900062665	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467539	GCA_900062675	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467566	GCA_900062685	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467572	GCA_900062695	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467608	GCA_900062705	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467564	GCA_900062715	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467586	GCA_900062725	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467567	GCA_900062735	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467598	GCA_900062745	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467585	GCA_900062755	1	incomplete	7624	no	36.14
Legionella pneumophila	2531STDY5467414	GCA_900062855	1	incomplete	13773	yes	36.78

235

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	*attL* site	Prophage GC content (%)
Legionella pneumophila	2532STDY5467533	GCA_900062865	1	incomplete	7236	no	40.74
Legionella pneumophila	2532STDY5467533	GCA_900062865	2	incomplete	25002	yes	35.62
Legionella pneumophila	2532STDY5467662	GCA_900062875	1	incomplete	23824	yes	37.84
Legionella pneumophila	2531STDY5467328	GCA_900063035	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467300	GCA_900063045	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467308	GCA_900063055	1	incomplete	16901	yes	39.69
Legionella pneumophila	2531STDY5467325	GCA_900063065	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467479	GCA_900063075	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467501	GCA_900063085	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467506	GCA_900063095	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467486	GCA_900063105	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467558	GCA_900063115	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467615	GCA_900063125	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467587	GCA_900063135	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467612	GCA_900063145	1	incomplete	7624	no	36.14
Legionella pneumophila	2531STDY5467323	GCA_900063785	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467313	GCA_900063795	1	incomplete	30026	yes	36.56
Legionella pneumophila	2531STDY5467466	GCA_900063805	1	incomplete	13185	yes	38.25
Legionella pneumophila	2531STDY5467471	GCA_900063815	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467489	GCA_900063825	0	N/A	N/A	N/A	N/A

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	attL site	Prophage GC content (%)
Legionella pneumophila	2532STDY5467552	GCA_900063835	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467605	GCA_900063845	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467597	GCA_900063855	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467594	GCA_900063865	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467589	GCA_900063875	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467601	GCA_900063885	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467620	GCA_900063895	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467629	GCA_900063905	1	incomplete	23824	yes	37.84
Legionella pneumophila	2532STDY5467630	GCA_900063915	1	incomplete	23824	yes	37.84
Legionella pneumophila	2532STDY5467660	GCA_900063925	1	incomplete	21690	yes	36.54
Legionella pneumophila	2532STDY5467654	GCA_900063935	1	incomplete	19880	yes	37
Legionella pneumophila	2531STDY5467354	GCA_900064175	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467488	GCA_900064185	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467521	GCA_900064195	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467618	GCA_900064205	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467599	GCA_900064315	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467606	GCA_900064325	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467644	GCA_900064335	1	incomplete	23824	yes	37.84
Legionella pneumophila	2532STDY5467636	GCA_900064345	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467638	GCA_900064355	0	N/A	N/A	N/A	N/A

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	attL site	Prophage GC content (%)
Legionella pneumophila	2532STDY5467650	GCA_900064365	1	incomplete	19880	yes	37
Legionella pneumophila	2532STDY5467649	GCA_900064375	1	incomplete	23824	yes	37.84
Legionella pneumophila	2532STDY5467657	GCA_900064385	1	incomplete	23824	yes	37.84
Legionella pneumophila	2532STDY5467508	GCA_900064485	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467534	GCA_900064495	1	incomplete	7624	no	36.14
Legionella pneumophila	2531STDY5467341	GCA_900064715	1	incomplete	13773	yes	36.78
Legionella pneumophila	2531STDY5467411	GCA_900064725	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467528	GCA_900064735	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467487	GCA_900064745	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467579	GCA_900064755	1	incomplete	7624	no	36.14
Legionella pneumophila	2532STDY5467590	GCA_900064765	1	incomplete	7624	no	36.14
Legionella pneumophila	2531STDY5467413	GCA_900065305	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467415	GCA_900065315	1	incomplete	21473	yes	36.7
Legionella pneumophila	2531STDY5467415	GCA_900065315	2	incomplete	7721	no	37.66
Legionella pneumophila	2531STDY5467333	GCA_900065435	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467646	GCA_900065445	1	incomplete	21054	yes	39.83
Legionella pneumophila	2532STDY5467497	GCA_900065895	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2532STDY5467632	GCA_900065905	1	incomplete	38143	yes	38.57
Legionella pneumophila	2532STDY5467643	GCA_900065915	1	incomplete	19473	yes	36.87
Legionella pneumophila	2532STDY5467643	GCA_900065915	2	incomplete	8579	no	38.23

238

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	attL site	Prophage GC content (%)
Legionella pneumophila	2531STDY5467406	GCA_900070125	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467356	GCA_900070135	1	incomplete	21473	yes	36.69
Legionella pneumophila	2531STDY5467356	GCA_900070135	2	incomplete	7721	no	37.66
Legionella pneumophila	2531STDY5467314	GCA_900070145	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467350	GCA_900070155	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467373	GCA_900070165	0	N/A	N/A	N/A	N/A
Legionella pneumophila	2531STDY5467308	GCA_900070185	1	incomplete	21473	yes	36.7
Legionella pneumophila	2531STDY5467308	GCA_900070185	2	incomplete	7721	no	37.66
Legionella pneumophila	LP_423	GCA_900072995	0	N/A	N/A	N/A	N/A
Legionella pneumophila	LP_617	GCA_900073005	0	N/A	N/A	N/A	N/A
Legionella pneumophila	H034980467	GCA_900073025	0	N/A	N/A	N/A	N/A
Legionella pneumophila	Pontiac-1	GCA_900073035	0	N/A	N/A	N/A	N/A
Legionella pneumophila	H034680033	GCA_900073045	0	N/A	N/A	N/A	N/A
Legionella pneumophila	H034800427	GCA_900073055	1	incomplete	20921	yes	37.48
Legionella pneumophila	H034800427	GCA_900073055	2	incomplete	25002	yes	35.62
Legionella pneumophila	Lpm7613	GCA_900092465	1	incomplete	8901	no	37.05
Legionella pneumophila	ST23	GCA_900119755	1	incomplete	20233	yes	36.03
Legionella pneumophila	ST62	GCA_900119765	1	incomplete	23824	yes	37.84
Legionella pneumophila	ST37	GCA_900119775	1	incomplete	30754	yes	37.22
Legionella pneumophila	ST42	GCA_900119785	1	incomplete	12080	yes	37.17

239

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	attL site	Prophage GC content (%)
Legionella pneumophila	NCTC 11286	GCA_900186855	1	incomplete	27939	yes	38.62
Legionella pneumophila	NCTC 11985	GCA_900186925	1	incomplete	31655	yes	36.66
Legionella pneumophila	NCTC 11985	GCA_900186925	2	incomplete	32106	yes	37.14
Legionella quateirensis	ATCC 49507	GCA_001467955	0	N/A	N/A	N/A	N/A
Legionella quinlivanii	CDC1442-AUS-ELqui	GCA_001467975	0	N/A	N/A	N/A	N/A
Legionella quinlivanii	DSM 21216	GCA_900108145	0	N/A	N/A	N/A	N/A
Legionella rubrilucens	WA-270A-C2	GCA_001468125	0	N/A	N/A	N/A	N/A
Legionella sainthelensi	ATCC 35248	GCA_000621685	1	incomplete	22216	yes	34.94
Legionella sainthelensi	Mt.St.Helens-4Lsai	GCA_001468105	0	N/A	N/A	N/A	N/A
Legionella sainthelensi	LA01-117	GCA_002848365	1	incomplete	28864	yes	36.59
Legionella sainthelensi	LA01-117	GCA_002848365	2	incomplete	17326	yes	35.38
Legionella santicrucis	SC-63-C7Lsan	GCA_001468135	0	N/A	N/A	N/A	N/A
Legionella sp.	LH-SWC	GCA_001465875	1	incomplete	8007	yes	41.62
Legionella sp.	LH-SWC	GCA_001465875	2	incomplete	8328	no	37.39
Legionella sp.	LH-SWC	GCA_001465875	3	incomplete	10151	no	39.88
Legionella shakespearei	DSM 23087	GCA_000373765	1	incomplete	31184	yes	40.44
Legionella shakespearei	ATCC 49655	GCA_001468025	0	N/A	N/A	N/A	N/A
Legionella sp.	DS3.009	GCA_002786095	0	N/A	N/A	N/A	N/A
Legionella sp.	BC.3.72	GCA_002786165	0	N/A	N/A	N/A	N/A
Legionella sp.	FW.3.37	GCA_002786205	0	N/A	N/A	N/A	N/A

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	attL site	Prophage GC content (%)
Legionella sp.	BC.3.64	GCA_002786245	0	N/A	N/A	N/A	N/A
Legionella sp.	Gammapro2	GCA_002863045	1	incomplete	17730	yes	39.12
Legionella sp.	21-45-4	GCA_002255255	0	N/A	N/A	N/A	N/A
Legionella sp.	39-23	GCA_001897665	0	N/A	N/A	N/A	N/A
Legionella sp.	40-6	GCA_001899505	0	N/A	N/A	N/A	N/A
Legionella spiritensis	Mt.St.Helens-9	GCA_001468165	0	N/A	N/A	N/A	N/A
Legionella spiritensis	NCTC 11990	GCA_900186965	1	incomplete	15549	yes	38.86
Legionella steelei	IMVS3376	GCA_001468005	0	N/A	N/A	N/A	N/A
Legionella steigerwaltii	SC-18-C9	GCA_001468065	0	N/A	N/A	N/A	N/A
Legionella taurinensis	Genessee01	GCA_003070625	0	N/A	N/A	N/A	N/A
Legionella taurinensis	Genessee03	GCA_003070645	0	N/A	N/A	N/A	N/A
Legionella taurinensis	Genessee04	GCA_003070665	0	N/A	N/A	N/A	N/A
Legionella taurinensis	Genessee02	GCA_003070675	0	N/A	N/A	N/A	N/A
Legionella tucsonensis	ATCC 49180	GCA_001468035	1	incomplete	20274	yes	34.21
Legionella sp.	LegM	GCA_000308315	1	questionable	32624	yes	39.19
Legionella sp.	LegM	GCA_000308315	2	incomplete	9450	no	42.86
Legionella sp.	LegM	GCA_000308315	3	incomplete	7835	no	39.56
Legionella sp.	LegM	GCA_000308315	4	incomplete	7674	no	39.9
Legionella sp.	LegM	GCA_000308315	5	incomplete	17508	yes	38.53
Legionella sp.	LegM	GCA_000308315	6	incomplete	7880	no	40.91

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	attL site	Prophage GC content (%)
Legionella wadsworthii	ATCC 33877	GCA_000701265	1	incomplete	9004	no	35.69
Legionella waltersii	ATCC 51914	GCA_001468085	0	N/A	N/A	N/A	N/A
Legionella waltersii	NCTC 13017	GCA_900187095	1	incomplete	29192	yes	37.19
Legionella waltersii	NCTC 13017	GCA_900187095	2	questionable	36013	yes	37.37
Legionella waltersii	NCTC 13017	GCA_900187095	3	incomplete	6002	no	36.17
Legionella worsleiensis	ATCC 49508	GCA_001467535	0	N/A	N/A	N/A	N/A
Legionella pneumophila	Phe_14	N/A	0	N/A	N/A	N/A	N/A
Legionella pneumophila	E1	N/A	0	N/A	N/A	N/A	N/A
Legionella micdadei	LMI	GCA_000953635	1	incomplete	18861	no	40.24
Legionella micdadei	LMI	GCA_000953635	2	incomplete	9792	no	35.72
Legionella micdadei	TATLOCKLmic	GCA_001467875	0	N/A	N/A	N/A	N/A
Legionella micdadei	ATCC 33218	GCA_001648625	1	incomplete	9792	no	35.72
Legionella micdadei	ATCC 33218	GCA_001648625	2	incomplete	20318	yes	40.51
Legionella micdadei	NZ2015	GCA_002085715	1	incomplete	19597	yes	40.34
Legionella micdadei	NZ2016	GCA_002085735	1	incomplete	25078	yes	38.45
Legionella micdadei	ATCC 33218	GCA_900102325	0	N/A	N/A	N/A	N/A
Legionella micdadei	LMI	GCA_000953635	1	incomplete	18861	no	40.24
Legionella micdadei	LMI	GCA_000953635	2	incomplete	9792	no	35.72
Legionella micdadei	TATLOCKLmic	GCA_001467875	0	N/A	N/A	N/A	N/A
Legionella micdadei	ATCC 33218	GCA_001648625	1	incomplete	9792	no	35.72

Table S9. – Continued from overleaf

Bacteria	Strain	Identifier	Prophage number	Prophage Status	Prophage size (bp)	*attL* site	Prophage GC content (%)
Legionella micdadei	ATCC 33218	GCA_001648625	2	incomplete	20318	yes	40.51
Legionella micdadei	NZ2015	GCA_002085715	1	incomplete	19597	yes	40.34
Legionella micdadei	NZ2016	GCA_002085735	1	incomplete	25078	yes	38.45
Legionella micdadei	ATCC 33218	GCA_900102325	0	N/A	N/A	N/A	N/A

243

Table S10. The table of the number of CALIN elements identified within all *Legionella* spp. genomes in GenBank using IntegronFinder

Bacteria	Strain	Identifier	Number of CALIN elements
Legionellaceae bacterium	UBA6148	GCA_002423125.1	1
Legionella saoudiensis	LH-SWC	GCA_001465875.1	1
Legionellales bacterium	NORP184	GCA_002400645.1	3
Legionellales bacterium	MED753	GCA_002690495.1	2
Legionellales bacterium	MED1261	GCA_002692205.1	3
Legionellales bacterium	NAT247	GCA_002698645.1	6
Legionellales bacterium	SAT3	GCA_002707205.1	12
Legionellales bacterium	SAT1342	GCA_002714285.1	1
Legionellales bacterium	SP195	GCA_002722995.1	1
Legionella micdadei	ATCC 33218	GCA_900102325.1	1
Legionella micdadei	ATCC 33218	GCA_001648625.1	1
Legionella sp.	39-23	GCA_001897665.1	6
Legionella quinlivanii	DSM 21216	GCA_900108145.1	5
Legionella quinlivanii	CDC1442-AUS-E	GCA_001467975.1	3
Legionella hackeliae	798-PA-H	GCA_001467705.1	4
Legionellales bacterium	RIFCSPHIGHO2_12_FULL_42_9	GCA_001804755.1	3
Legionella brunensis	ATCC 43878	GCA_001467025.1	4
Legionellales bacterium	UBA6500	GCA_002436445.1	8
Legionella nautarum	ATCC 49506	GCA_001467895.1	2
Legionella maceachernii	ATCC 35300	GCA_900167045.1	6

244

Table S10. – Continued from overleaf

Bacteria	Strain	Identifier	Number of CALIN elements
Legionella maceachernii	PX-1-G2-E2	GCA_001467845.1	4
Legionella feeleii	ATCC 35072	GCA_001648615.1	4
Legionella feeleii	WO-44C	GCA_001467625.1	4
Legionellales bacterium	UBA6786	GCA_002452935.1	6
Legionella pneumophila	Mississauga	GCA_002002625.1	1
Legionella pneumophila	Sudbury	GCA_002002645.1	1
Legionella pneumophila	JCM_7571	GCA_000614785.1	1
Legionella pneumophila	Nagoya-1	GCA_000699225.1	8
Legionella pneumophila	Ymt294	GCA_000953935.1	2
Legionella pneumophila	130b	GCA_000211115.2	8
Legionella pneumophila	12.4437	GCA_000823505.1	1
Legionella pneumophila	12.4904	GCA_000823645.1	1
Legionella pneumophila	12.5415	GCA_000823785.1	1
Legionella pneumophila	HO80160263	GCA_000823865.1	2
Legionella pneumophila	HO4020049	GCA_000823805.1	1
Legionella pneumophila	2531STDY5467307	GCA_900060195.1	1
Legionella pneumophila	2531STDY5467317	GCA_900048915.1	1
Legionella pneumophila	2531STDY5467312	GCA_900054385.1	1
Legionella pneumophila	2531STDY5467381	GCA_900061475.1	1
Legionella pneumophila	2531STDY5467310	GCA_900054395.1	1

Table S10. – Continued from overleaf

Bacteria	Strain	Identifier	Number of CALIN elements
Legionella pneumophila	2531STDY5467329	GCA_900053355.1	1
Legionella pneumophila	2531STDY5467383	GCA_900048935.1	1
Legionella pneumophila	2531STDY5467349	GCA_900061525.1	1
Legionella pneumophila	2531STDY5467351	GCA_900061535.1	1
Legionella pneumophila	2531STDY5467356	GCA_900070135.1	3
Legionella pneumophila	2531STDY5467367	GCA_900059945.1	1
Legionella pneumophila	2531STDY5467406	GCA_900070125.1	1
Legionella pneumophila	2531STDY5467415	GCA_900119785.1	3
Legionella pneumophila	2531STDY5467334	GCA_900056655.1	3
Legionella pneumophila	2531STDY5467453	GCA_900061645.1	1
Legionella pneumophila	2531STDY5467426	GCA_900061635.1	1
Legionella pneumophila	2531STDY5467455	GCA_900061655.1	1
Legionella pneumophila	2531STDY5467459	GCA_900061675.1	1
Legionella pneumophila	2531STDY5467458	GCA_900061665.1	1
Legionella pneumophila	2531STDY5467461	GCA_900061685.1	1
Legionella pneumophila	2531STDY5467462	GCA_900061695.1	1
Legionella pneumophila	2531STDY5467460	GCA_900055585.1	1
Legionella pneumophila	2531STDY5467463	GCA_900051725.1	1
Legionella pneumophila	2531STDY5467467	GCA_900049275.1	1
Legionella pneumophila	2531STDY5467468	GCA_900050555.1	1

246

Table S10. – Continued from overleaf

Bacteria	Strain	Identifier	Number of CALIN elements
Legionella pneumophila	2531STDY5467466	GCA_900063805.1	1
Legionella pneumophila	2531STDY5467469	GCA_900054155.1	1
Legionella pneumophila	2531STDY5467471	GCA_900063815.1	1
Legionella pneumophila	2531STDY5467456	GCA_900050975.1	1
Legionella pneumophila	2531STDY5467457	GCA_900061705.1	1
Legionella pneumophila	2531STDY5467464	GCA_900061715.1	1
Legionella pneumophila	2531STDY5467465	GCA_900061725.1	1
Legionella pneumophila	2531STDY5467345	GCA_900062505.1	3
Legionella pneumophila	2531STDY5467450	GCA_900062525.1	1
Legionella pneumophila	2532STDY5467575	GCA_900062535.1	1
Legionella pneumophila	2532STDY5467626	GCA_900054165.1	1
Legionella pneumophila	2532STDY5467591	GCA_900062545.1	1
Legionella pneumophila	2532STDY5467576	GCA_900057785.1	1
Legionella pneumophila	2532STDY5467578	GCA_900062555.1	1
Legionella pneumophila	2532STDY5467581	GCA_900062565.1	1
Legionella pneumophila	2532STDY5467593	GCA_900048955.1	1
Legionella pneumophila	2532STDY5467577	GCA_900062575.1	1
Legionella pneumophila	2532STDY5467580	GCA_900055605.1	1
Legionella pneumophila	2532STDY5467535	GCA_900060005.1	1
Legionella pneumophila	2532STDY5467542	GCA_900060415.1	1

Table S10. – Continued from overleaf

Bacteria	Strain	Identifier	Number of CALIN elements
Legionella pneumophila	2532STDY5467549	GCA_900060425.1	1
Legionella pneumophila	2532STDY5467544	GCA_900060435.1	1
Legionella pneumophila	2532STDY5467560	GCA_900052945.1	1
Legionella pneumophila	2532STDY5467537	GCA_900060445.1	1
Legionella pneumophila	2532STDY5467546	GCA_900062585.1	1
Legionella pneumophila	2532STDY5467565	GCA_900062595.1	1
Legionella pneumophila	2532STDY5467547	GCA_900057575.1	1
Legionella pneumophila	2532STDY5467559	GCA_900062615.1	1
Legionella pneumophila	2532STDY5467561	GCA_900062605.1	1
Legionella pneumophila	2532STDY5467627	GCA_900060775.1	1
Legionella pneumophila	2532STDY5467600	GCA_900058825.1	1
Legionella pneumophila	2532STDY5467628	GCA_900062625.1	1
Legionella pneumophila	2532STDY5467579	GCA_900064755.1	1
Legionella pneumophila	2532STDY5467534	GCA_900064495.1	1
Legionella pneumophila	2532STDY5467545	GCA_900062645.1	1
Legionella pneumophila	2532STDY5467538	GCA_900062635.1	1
Legionella pneumophila	2532STDY5467548	GCA_900058835.1	1
Legionella pneumophila	2532STDY5467550	GCA_900049375.1	1
Legionella pneumophila	2532STDY5467555	GCA_900050245.1	1
Legionella pneumophila	2532STDY5467557	GCA_900048965.1	1

Table S10. – Continued from overleaf

Bacteria	Strain	Identifier	Number of CALIN elements
Legionella pneumophila	2532STDY5467556	GCA_900060785.1	1
Legionella pneumophila	2532STDY5467596	GCA_900060015.1	1
Legionella pneumophila	2532STDY5467602	GCA_900062655.1	2
Legionella pneumophila	2532STDY5467566	GCA_900062685.1	1
Legionella pneumophila	2532STDY5467563	GCA_900058605.1	1
Legionella pneumophila	2532STDY5467572	GCA_900062695.1	1
Legionella pneumophila	2532STDY5467564	GCA_900062715.1	1
Legionella pneumophila	2532STDY5467551	GCA_900056195.1	1
Legionella pneumophila	2532STDY5467586	GCA_900062725.1	1
Legionella pneumophila	2532STDY5467541	GCA_900056935.1	1
Legionella pneumophila	2532STDY5467558	GCA_900063115.1	1
Legionella pneumophila	2532STDY5467543	GCA_900057585.1	1
Legionella pneumophila	2532STDY5467573	GCA_900053425.1	1
Legionella pneumophila	2532STDY5467567	GCA_900062735.1	1
Legionella pneumophila	2532STDY5467568	GCA_900053435.1	1
Legionella pneumophila	2532STDY5467571	GCA_900055625.1	1
Legionella pneumophila	2532STDY5467595	GCA_900053445.1	1
Legionella pneumophila	2532STDY5467598	GCA_900062745.1	1
Legionella pneumophila	2532STDY5467585	GCA_900062755.1	1
Legionella pneumophila	2532STDY5467574	GCA_900057595.1	1

Table S10. – Continued from overleaf

Bacteria	Strain	Identifier	Number of CALIN elements
Legionella pneumophila	2532STDY5467605	GCA_900063845.1	1
Legionella pneumophila	2532STDY5467603	GCA_900059595.1	1
Legionella pneumophila	2532STDY5467604	GCA_900052325.1	1
Legionella pneumophila	2532STDY5467597	GCA_900063855.1	1
Legionella pneumophila	2532STDY5467594	GCA_900063865.1	1
Legionella pneumophila	2532STDY5467590	GCA_900064765.1	1
Legionella pneumophila	2532STDY5467599	GCA_900064315.1	1
Legionella pneumophila	2532STDY5467582	GCA_900055215.1	1
Legionella pneumophila	2532STDY5467606	GCA_900064325.1	1
Legionella pneumophila	2532STDY5467588	GCA_900052335.1	1
Legionella pneumophila	2532STDY5467589	GCA_900063875.1	1
Legionella pneumophila	2532STDY5467601	GCA_900063885.1	1
Legionella pneumophila	2532STDY5467583	GCA_900054645.1	1
Legionella pneumophila	2532STDY5467584	GCA_900060455.1	1
Legionella pneumophila	2532STDY5467592	GCA_900057265.1	1
Legionella pneumophila	2532STDY5467617	GCA_900056205.1	1
Legionella pneumophila	2532STDY5467616	GCA_900056215.1	1
Legionella pneumophila	2532STDY5467625	GCA_900057605.1	1
Legionella pneumophila	2532STDY5467614	GCA_900059605.1	1
Legionella pneumophila	2532STDY5467615	GCA_900063125.1	1

Table S10. – Continued from overleaf

Bacteria	Strain	Identifier	Number of CALIN elements
Legionella pneumophila	2532STDY5467612	GCA_900063145.1	1
Legionella pneumophila	2532STDY5467613	GCA_900054655.1	1
Legionella pneumophila	2532STDY5467587	GCA_900063135.1	1
Legionella pneumophila	2532STDY5467611	GCA_900054675.1	1
Legionella pneumophila	2532STDY5467610	GCA_900054665.1	1
Legionella pneumophila	2532STDY5467607	GCA_900055635.1	1
Legionella pneumophila	2532STDY5467623	GCA_900049285.1	1
Legionella pneumophila	2532STDY5467620	GCA_900063895.1	1
Legionella pneumophila	2532STDY5467622	GCA_900049295.1	1
Legionella pneumophila	2532STDY5467619	GCA_900054685.1	1
Legionella pneumophila	2532STDY5467621	GCA_900049305.1	1
Legionella pneumophila	2532STDY5467618	GCA_900064205.1	1
Legionella pneumophila	2532STDY5467663	GCA_900051005.1	1
Legionella pneumophila	2531STDY5467449	GCA_900057615.1	1
Legionella pneumophila	2532STDY5467629	GCA_900063905.1	1
Legionella pneumophila	2531STDY5467406	GCA_900059575.1	1
Legionella pneumophila	2531STDY5467356	GCA_900057245.1	3
Legionella pneumophila	2531STDY5467373	GCA_900060365.1	1
Legionella pneumophila	LP_617	GCA_900073005.1	1
Legionella pneumophila	ATCC 35251	GCA_000586195.1	5

Table S10. – Continued from overleaf

Bacteria	Strain	Identifier	Number of CALIN elements
Legionella pneumophila	ATCC 35096	GCA_000586215.1	4
Legionella pneumophila	ATCC 33216	GCA_000586275.1	3
Legionella pneumophila	ATCC 33156	GCA_000586315.1	3
Legionella pneumophila	D-5864	GCA_000785905.1	1
Legionella pneumophila	strain_ATCC 33823	GCA_000586235.1	1
Legionella pneumophila	ATCC 33215	GCA_000586295.1	1
Legionella pneumophila	FS_10_1101a-3	GCA_001582405.1	5
Legionella pneumophila	ATCC 33156	GCA_001582825.1	4
Legionella pneumophila	SH003	GCA_001583655.1	1
Legionella pneumophila	ATCC 35096	GCA_001582225.1	4
Legionella pneumophila	ATCC 33823	GCA_001582305.1	1
Legionella pneumophila	SH135	GCA_001582385.1	3
Legionella pneumophila	FS_10_1101a-3	GCA_001582405.1	1
Legionella pneumophila	SH095	GCA_001582465.1	1
Legionella pneumophila	SZ069	GCA_001582535.1	3
Legionella pneumophila	WX2011029	GCA_001583565.1	4
Legionella pneumophila	FS_4_1103abu	GCA_001582625.1	3
Legionella pneumophila	ATCC 33216	GCA_001582695.1	2
Legionella pneumophila	SZ2012007	GCA_001582705.1	2
Legionella pneumophila	WX2011036	GCA_001582775.1	3

Table S10. – Continued from overleaf

Bacteria	Strain	Identifier	Number of CALIN elements
Legionella pneumophila	Hu6 Scaffold7_1	GCA_001583595.1	1
Legionella pneumophila	SH202	GCA_001582865.1	1
Legionella pneumophila	PtVF66_2014	GCA_001549925.1	4
Legionella pneumophila	PtVF89_2014	GCA_001549915.1	5
Legionella pneumophila	NY3	GCA_001601215.1	2
Legionella pneumophila	NY5	GCA_001601325.1	5
Legionella pneumophila	NY6	GCA_001601355.1	1
Legionella pneumophila	NY7	GCA_001601505.1	4
Legionella pneumophila	NY8	GCA_001601535.1	1
Legionella pneumophila	NY11	GCA_001601475.1	2
Legionella pneumophila	NY15	GCA_001601455.1	1
Legionella pneumophila	NY17	GCA_001601055.1	2
Legionella pneumophila	NY18	GCA_001600925.1	2
Legionella pneumophila	NY19	GCA_001600915.1	2
Legionella pneumophila	NY20	GCA_001601085.1	2
Legionella pneumophila	NY21	GCA_001601115.1	2
Legionella pneumophila	NY22	GCA_001601135.1	2
Legionella pneumophila	NY23	GCA_001600975.1	8
Legionella pneumophila	NY24	GCA_001600905.1	3
Legionella pneumophila	NY25	GCA_001601005.1	6

Table S10. – Continued from overleaf

Bacteria	Strain	Identifier	Number of CALIN elements
Legionella pneumophila	NY26	GCA_001600995.1	5
Legionella pneumophila	NY27	GCA_001601075.1	5
Legionella pneumophila	NY28	GCA_001601175.1	3
Legionella pneumophila	NY29	GCA_001601155.1	5
Legionella pneumophila	NY30	GCA_001600985.1	4
Legionella adelaidensis	1762-AUS-E	GCA_001467055.1	4
Legionella moravica	DSM 19234	GCA_000423305.1	3
Legionella lansingensis	ATCC 49751	GCA_000622185.1	4
Legionella lansingensis	ATCC 49751	GCA_001467795.1	8
Legionellales bacterium	UBA2653	GCA_002359535.1	3
Legionellales bacterium	UBA4722	GCA_002404295.1	4
Legionellaceae bacterium	SAT2595	GCA_002707695.1	10
Legionellaceae bacterium	SP3112	GCA_002719855.1	5
Legionellales bacterium	RIFCSPHIGHO2_12_FULL_35_11	GCA_001803185.1	1
Legionella sp.	BC.3.64	GCA_002786245.1	5
Legionella sp.	FW.3.37	GCA_002786205.1	5
Legionella sp.	Gammapro2	GCA_002863045.1	7
Legionella anisa	Linanisette	GCA_000333755.1	1
Legionella anisa	WA-316-C3	GCA_001467525.1	3
Legionella jordanis	ATCC 33623	GCA_001648675.1	4

254

Table S10. – Continued from overleaf

Bacteria	Strain	Identifier	Number of CALIN elements
Legionella jordanis	BL-540	GCA_001467765.1	7
Legionellales bacterium	UBA7366	GCA_002469885.1	19
Legionella erythra	SE-32A-C8	GCA_001467615.1	2
Legionellales bacterium	UBA2469	GCA_002342175.1	3
Legionellales bacterium	UBA1113	GCA_002313395.1	2
Legionella steigerwaltii	SC-18-C9	GCA_001468065.1	1
Legionellaceae bacterium	UBA2794	GCA_002352055.1	5
Legionella santicrucis	SC-63-C7	GCA_001468135.1	6
Legionella oakridgensis	RV-2-2007	GCA_000512715.1	5
Legionella oakridgensis	ATCC 33761	GCA_001648605.1	4
Legionellales bacterium	UBA5158	GCA_002413665.1	9
Legionella cincinnatiensis	CDC72-OH-14	GCA_001467545.1	3
Legionella dumoffii	NY_23	GCA_001467605.1	3
Legionellales bacterium	RIFCSPHIGHO2_12_FULL_37_14	GCA_001804735.1	3
Legionella parisiensis	PF-209-C-C2	GCA_001467945.1	4
Legionella parisiensis	DSM 19216	GCA_001736145.1	4
Legionella tunisiensis	GCA_000308315.1	4	
Legionella sp.	21-45-4	GCA_002255255.1	1
Legionellales bacterium	UBA6603	GCA_002433445.1	11
Legionellales bacterium	UBA1524	GCA_002323365.1	18

255

Table S10. – Continued from overleaf

Bacteria	Strain	Identifier	Number of CALIN elements
Legionella londiniensis	ATCC 49505	GCA_001467825.1	1
Legionella israelensis	Bercovier_4	GCA_001467785.1	3
Legionella rubrilucens	WA-270A-C2	GCA_001468125.1	1
Legionella sp.	40-6	GCA_001899505.1	11

Bibliography

Abdel-Nour, M., Duncan, C., Low, D. E., and Guyard, C. (2013). Biofilms: the stronghold of Legionella pneumophila. *International journal of molecular sciences*, 14(11):21660–75.

Abedon, S. (2015). Ecology of Anti-Biofilm Agents I: Antibiotics versus Bacteriophages. *Pharmaceuticals*, 8(3):525–558.

Abedon, S. T. (2008). Phages, ecology, evolution. In Abedon, S. T., editor, *Bacteriophage Ecology: Population Growth, Evolution, and Impact of Bacterial Viruses*, chapter 1, pages 1–28. Cambridge University Press.

Abedon, S. T. (2009). Phage evolution and ecology. *Advances in applied microbiology*, 67(12):1–45.

Abedon, S. T. (2012). Spatial vulnerability: bacterial arrangements, microcolonies, and biofilms as responses to low rather than high phage densities. *Viruses*, 4(5):663–87.

Abedon, S. T. (2014). Phage therapy: eco-physiological pharmacology. *Scientifica*, 2014:581639.

Abedon, S. T. (2018). *Detection of Bacteriophages: Phage Plaques*.

Abedon, S. T. and Katsaounis, T. I. (2018). Basic Phage Mathematics. *Methods in molecular biology (Clifton, N.J.)*, 1681:3–30.

Abedon, S. T., Kuhl, S. J., Blasdel, B. G., and Kutter, E. M. (2011). Phage treatment of human infections. *Bacteriophage*, 1(2):66–85.

Abu Khweek, A., Fernández Dávila, N. S., Caution, K., Akhter, A., Abdulrahman, B. a., Tazi, M., Hassan, H., Novotny, L. a., Bakaletz, L. O., and Amer, A. O. (2013). Biofilm-derived Legionella pneumophila evades the innate immune response in macrophages. *Frontiers in Cellular and Infection Microbiology*, 3(May):1–8.

Abu Kwaik, Y., Gao, L. Y., Stone, B. J., Venkataraman, C., and Harb, O. S. (1998). Invasion of protozoa by Legionella pneumophila and its role in bacterial ecology and pathogenesis. *Applied and environmental microbiology*, 64(9):3127–33.

Ackermann, H.-w. (2009). Phage classification and characterization. *Methods in molecular biology (Clifton, N.J.)*, 501:127–140.

Ackermann, H.-W. (2011). Bacteriophage Taxonomy. *Microbiology Australia*, pages 90–94.

Ackermann, H.-W. and Prangishvili, D. (2012). Prokaryote viruses studied by electron microscopy. *Archives of Virology*, 157(10):1843–1849.

Adriaenssens, E. and Brister, J. R. (2017). How to Name and Classify Your Phage: An Informal Guide. *Viruses*, 9(4):70.

Adriaenssens, E. M., Ackermann, H.-W., Anany, H., Blasdel, B., Connerton, I. F., Goulding, D., Griffiths, M. W., Hooton, S. P., Kutter, E. M., Kropinski, A. M., Lee, J.-H., Maes, M., Pickard, D., Ryu, S., Sepehrizadeh, Z., Shahrbabak, S. S., Toribio, A. L., and Lavigne, R. (2012). A suggested new bacteriophage genus: "Viunalikevirus". *Archives of Virology*, 157(10):2035–2046.

Adriaenssens, E. M., Wittmann, J., Kuhn, J. H., Turner, D., Sullivan, M. B., Dutilh, B. E., Jang, H. B., van Zyl, L. J., Klumpp, J., Lobocka, M., Moreno Switt, A. I., Rumnieks, J., Edwards, R. A., Uchiyama, J., Alfenas-Zerbini, P., Petty, N. K., Kropinski, A. M., Barylski, J., Gillis, A., Clokie, M. R. C., Prangishvili, D., Lavigne, R., Aziz, R. K., Duffy, S., Krupovic, M., Poranen, M. M., Knezevic, P., Enault, F., Tong, Y., Oksanen, H. M., and Rodney Brister, J. (2018). Taxonomy of prokaryotic viruses: 2017 update from the ICTV Bacterial and Archaeal Viruses Subcommittee. *Archives of Virology*, 163(4):1125–1129.

Agboluaje, M. and Sauvageau, D. (2018). Bacteriophage Production in Bioreactors. *Methods in molecular biology (Clifton, N.J.)*, 1693:173–193.

Akhter, S., Aziz, R. K., and Edwards, R. A. (2012). PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. *Nucleic acids research*, 40(16):e126.

Al-Bana, B. H., Haddad, M. T., and Garduño, R. A. (2014). Stationary phase and mature infectious forms of Legionella pneumophila produce distinct viable but non-culturable cells. *Environmental microbiology*, 16(2):382–95.

Albert-Weissenberger, C., Cazalet, C., and Buchrieser, C. (2007). Legionella pneumophila - a human pathogen that co-evolved with fresh water protozoa. *Cellular and molecular life sciences : CMLS*, 64(4):432–48.

Alli, O. A., Gao, L. Y., Pedersen, L. L., Zink, S., Radulic, M., Doric, M., and Abu Kwaik, Y. (2000). Temporal pore formation-mediated egress from macrophages and alveolar epithelial cells by Legionella pneumophila. *Infection and immunity*, 68(11):6431–40.

Almahmoud, I., Kay, E., Schneider, D., and Maurin, M. (2009). Mutational paths towards increased fluoroquinolone resistance in Legionella pneumophila. *The Journal of antimicrobial chemotherapy*, 64(2):284–93.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, 215(3):403–10.

Amemura-Maekawa, J., Kura, F., Chida, K., Ohya, H., Kanatani, J.-i., Isobe, J., Tanaka, S., Nakajima, H., Hiratsuka, T., Yoshino, S., Sakata, M., Murai, M., Ohnishi, M., Kodama, F., Iwabuchi, K., Fujii, S., Yamaguchi, Y., Numata, N., Konno, T., Kaneko, A., Seto, J., Suzuki, Y., Kikuchi, R., Tomita, N., Hakuta, T., Nagashima, F., Tokoi, Y., Goto, K., Kurosawa, H., Tamura, Y., Kanazawa, S., Kobori, S., Kikuchi, K., Tomita, T., Nakamura, M., Kitahashi, T., Ogura, E., Okuno, R., Nakamura, I., Kurai, D., Watanabe, Y., Kojima, Y., Miyashita, Y., Matsuo, C., Yuzawa, E., Shindo, M., Hosoya, M., Kimura, Y., Yamamoto, K., Kawakami, K., Kitagawa, E., Yanagimoto, K., Igawa, Y., Kasahara, H., Ueda, H., Koyama, T., Sahara, K., Tomita, A., Kanazawa, Y., Hikida, M., Kadokura, Y., Noda, M., Shiraki, Y., Suzuki, M., Agata, N., Matsuo, C., Shimizu, M., Fujita, N., Kimura, Y., Katsukawa, C., Ogasawara, J., Fushiwaki, T., Tsuji, H., Kanazawa, Y., Ekawa, H., Nishiyama, T., Hirooka, M., Hanabara, Y., Uchida, J., Karasudani, T., Murakami, K., Shimizu, Y., Miyamoto, Y., Yoshida, H., Matsunaga, N., Kawano, M., Taguri, T., Harada, Y., Fukushiyama, I., Toda, J., Yasaka, T., Sugitani, W., Sasaki, M., Ogata, K., and Kawano, K. (2018). Legionella pneumophila and Other Legionella Species Isolated from Legionellosis Patients in Japan between 2008 and 2016. *Applied and Environmental Microbiology*, 84(18).

Anderson, B., Rashid, M. H., Carter, C., Pasternack, G., Rajanna, C., Revazishvili, T., Dean, T., Senecal, A., and Sulakvelidze, A. (2011). Enumeration of bacteriophage particles: Comparative analysis of the traditional plaque assay and real-time QPCR- and nanosight-based assays. *Bacteriophage*, 1(2):86–93.

Ariff, A., Wise, M. J., Kahler, C. M., Tay, C. Y., Peters, F., Perkins, T. T., and Chang, B. J. (2015). Novel Moraxella catarrhalis prophages display hyperconserved non-structural genes despite their genomic diversity. *BMC Genomics*, 16(1):860.

Armon, R. and Payment, P. (1990). A transparent medium for isolation of Legionella pneumophia from environmental water sources. *Journal of Microbiological Methods*, 11(1):65–71.

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., and Wishart, D. S. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic acids research*, 44(W1):W16–21.

Atsumi, S. and Little, J. W. (2006). Role of the lytic repressor in prophage induction of phage as

analyzed by a module-replacement approach. *Proceedings of the National Academy of Sciences*, 103(12):4558–4563.

Baharoglu, Z. and Mazel, D. (2014). SOS, the formidable strategy of bacteria against aggressions. *FEMS Microbiology Reviews*, 38(6):1126–1145.

Barbaree, J. M., Fields, B. S., Feeley, J. C., Gorman, G. W., and Martin, W. T. (1986). Isolation of protozoa from water associated with a legionellosis outbreak and demonstration of intracellular multiplication of Legionella pneumophila. *Applied and environmental microbiology*, 51(2):422–4.

Barker, J., Brown, M. R. W., Collier, P. J., Farrell, I. A. N., and Gilbert, P. (1992). Relationship between Legionella pneumophila and Acanthamoeba polyphaga : Physiological Status and Susceptibility to Chemical Inactivation. *Applied and Environmental Microbiology*, 58(8):2420–2425.

Barker, J., Farrell, I. D., and Hutchison, J. G. (1986). Factors affecting growth of Legionella pneumophila in liquid media. *Journal of medical microbiology*, 22(2):97–100.

Bartlett, J. G., Dowell, S. F., Mandell, L. A., File, T. M., Musher, D. M., and Fine, M. J. (2000). Practice Guidelines for the Management of Community-Acquired Pneumonia in Adults. *Clinical Infectious Diseases*, 31(2):347–382.

Bartram, J., Chartier, Y., Lee, J. V., Bond, K., and Surman-Lee, S. (2007). *Legionella and the prevention of legionellosis*. Geneva, Switzerland.

Baskerville, A., Fitzgeorge, R. B., Broster, M., Hambleton, P., and Dennis, P. J. (1981). Experimental transmission of legionnaires' disease by exposure to aerosols of Legionella pneumophila. *Lancet (London, England)*, 2(8260-61):1389–90.

Bassetti, M., Poulakou, G., Ruppe, E., Bouza, E., Van Hal, S. J., and Brink, A. (2017). Antimicrobial resistance in the next 30 years, humankind, bugs and drugs: a visionary approach. *Intensive care medicine*, 43(10):1464–1475.

Beauté, J., Zucs, P., de Jong, B., and European Legionnaires' Disease Surveillance Network (2013). Legionnaires disease in Europe, 2009-2010. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin*, 18(10):20417.

Berdal, B. P., Farshy, C. E., and Feeley, J. C. (1979). Detection of Legionella pneumonophila antigen in urine by enzyme-linked immunospecific assay. *Journal of clinical microbiology*, 9(5):575–8.

Bobay, L.-m., Touchon, M., and Rocha, E. P. C. (2014). Pervasive domestication of defective prophages by bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 111(33):12127–32.

Bopp, C. a., Sumner, J. W., Morris, G. K., and Wells, J. G. (1981). Isolation of Legionella spp. from

environmental water samples by low-pH treatment and use of a selective medium. *Journal of clinical microbiology*, 13(4):714–9.

Borges A., Simões M., Martínez-Murcia A., S. M. J. (2012). Detection of Legionella spp. in Natural and Man-made Water Systems Using Standard Guidelines. *Journal of Microbiology Research*, 2(4):95–102.

Boucher, Y., Labbate, M., Koenig, J. E., and Stokes, H. W. (2007). Integrons: mobilizable platforms that promote genetic diversity in bacteria. *Trends in microbiology*, 15(7):301–9.

Bozue, J. A. and Johnson, W. (1996). Interaction of Legionella pneumophila with Acanthamoeba castellanii: uptake by coiling phagocytosis and inhibition of phagosome-lysosome fusion. *Infection and immunity*, 64(2):668–73.

Braga, L. P. P., Soucy, S. M., Amgarten, D. E., da Silva, A. M., and Setubal, J. C. (2018). Bacterial Diversification in the Light of the Interactions with Phages: The Genetic Symbionts and Their Role in Ecological Speciation. *Frontiers in Ecology and Evolution*, 6(6):1–12.

Brandt, J. and Albertsen, M. (2018). Investigation of Detection Limits and the Influence of DNA Extraction and Primer Choice on the Observed Microbial Communities in Drinking Water Samples Using 16S rRNA Gene Amplicon Sequencing. *Frontiers in microbiology*, 9(2140).

Brenner, D. J. (2018). Taxonomy of The Family Legionellaceae. In Katz, editor, *Legionellosis*, number v. 1, chapter 3. CRC Press.

Brown-Jaque, M., Muniesa, M., and Navarro, F. (2016). Bacteriophages in clinical samples can interfere with microbiological diagnostic tools. *Scientific reports*, 6(33000).

Bruin, J. P., Ijzerman, E. P. F., den Boer, J. W., Mouton, J. W., and Diederen, B. M. W. (2012). Wild-type MIC distribution and epidemiological cut-off values in clinical Legionella pneumophila serogroup 1 isolates. *Diagnostic microbiology and infectious disease*, 72(1):103–8.

Bruin, J. P., Koshkolda, T., IJzerman, E. P. F., Lück, C., Diederen, B. M. W., Den Boer, J. W., and Mouton, J. W. (2014). Isolation of ciprofloxacin-resistant Legionella pneumophila in a patient with severe pneumonia. *The Journal of antimicrobial chemotherapy*, 69(10):2869–71.

Brüssow, H. (2001). Phages of Dairy Bacteria. *Annual Review of Microbiology*, 55(1):283–303.

Brüssow, H., Canchaya, C., and Hardt, W.-D. (2004). Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiology and molecular biology reviews : MMBR*, 68(3):560–602, table of contents.

Brüssow, H. and Hendrix, R. W. (2002). Phage genomics: small is beautiful. *Cell*, 108(1):13–6.

Burillo, A., Pedro-Botet, M. L., and Bouza, E. (2017). Microbiology and Epidemiology of Legionnaire's Disease. *Infectious Disease Clinics of North America*, 31(1):7–27.

Buse, H. Y., Morris, B., Struewing, I. T., and Szabo, J. G. (2019). Chlorine and monochloramine disinfection of Legionella pneumophila colonizing copper and PVC drinking water biofilms. *Applied and environmental microbiology*, 85(7).

Butala, M., Žgur-Bertok, D., and Busby, S. J. (2009). The bacterial LexA transcriptional repressor. *Cellular and Molecular Life Sciences*, 66(1):82–93.

Buttimer, C., McAuliffe, O., Ross, R. P., Hill, C., O'Mahony, J., and Coffey, A. (2017). Bacteriophages and Bacterial Plant Diseases. *Frontiers in microbiology*, 8(34).

Byrne, B. and Swanson, M. S. (1998). Expression of Legionella pneumophila virulence traits in response to growth conditions. *Infection and immunity*, 66(7):3029–34.

Cairns, B. J., Timms, A. R., Jansen, V. A. A., Connerton, I. F., and Payne, R. J. H. (2009). Quantitative models of in vitro bacteriophage-host dynamics and their application to phage therapy. *PLoS pathogens*, 5(1):e1000253.

Canchaya, C., Fournous, G., Chibani-Chennoufi, S., Dillmann, M.-L., and Brüssow, H. (2003). Phage as agents of lateral gene transfer. *Current Opinion in Microbiology*, 6(4):417–424.

Cargill, K. L., Pyle, B. H., Sauer, R. L., and McFeters, G. A. (1992). Effects of culture conditions and biofilm formation on the iodine susceptibility of Legionella pneumophila. *Canadian journal of microbiology*, 38(5):423–9.

Carlet, J. (2015). Introduction: Antimicrobial resistance - a serious threat for humanity. Technical report.

Carvalho, F. R., Foronda, A. S., and Pellizari, V. H. (2007). Detection of Legionella pneumophila in water and biofilm samples by culture and molecular methods from man-made systems in São Paulo - Brazil. *Brazilian Journal of Microbiology*, 38(4):743–751.

Carvalho, F. R. S., Nastasi, F. R., Gamba, R. C., Foronda, A. S., and Pellizari, V. H. (2008). Occurrence and Diversity of Legionellaceae in Polar Lakes of the Antarctic Peninsula. *Current Microbiology*, 57(4):294–300.

Casati, S., Conza, L., Bruin, J., and Gaia, V. (2010). Compost facilities as a reservoir of Legionella pneumophila and other Legionella species. *Clinical Microbiology and Infection*, 16(7):945–947.

Casati, S., Gioria-Martinoni, A., and Gaia, V. (2009). Commercial potting soils as an alternative infection source of Legionella pneumophila and other Legionella species in Switzerland. *Clinical Microbiology and Infection*, 15(6):571–575.

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

Casjens, S. R. and Hendrix, R. W. (2015). Bacteriophage lambda: Early pioneer and still relevant. *Virology*, 479-480:310–30.

Castilla, J., Barricarte, A., Aldaz, J., García Cenoz, M., Ferrer, T., Pelaz, C., Pineda, S., Baladrón, B., Martín, I., Goñi, B., Aratajo, P., Chamorro, J., Lameiro, F., Torroba, L., Dorronsoro, I., Martínez-Artola, V., Esparza, M. J., Gastaminza, M. A., Fraile, P., and Aldaz, P. (2008). A large Legionnaires' disease outbreak in Pamplona, Spain: early detection, rapid control and no case fatality. *Epidemiology and infection*, 136(6):823–32.

Castillo, D., Kauffman, K., Hussain, F., Kalatzis, P., Rørbo, N., Polz, M. F., and Middelboe, M. (2018). Widespread distribution of prophage-encoded virulence factors in marine Vibrio communities. *Scientific reports*, 8(1):9973.

Centers for Disease Control (CDC) (2005). Procedures for the Recovery of Legionella from the Environment. *U.S Department of Health and Human Services*, pages 1–15.

Cervero-Aragó, S., Rodríguez-Martínez, S., Puertas-Bennasar, A., and Araujo, R. M. (2015). Effect of Common Drinking Water Disinfectants, Chlorine and Heat, on Free Legionella and Amoebae-Associated Legionella. *PloS one*, 10(8):e0134726.

Chaabna, Z., Forey, F., Reyrolle, M., Jarraud, S., Atlan, D., Fontvieille, D., and Gilbert, C. (2013). Molecular diversity and high virulence of Legionella pneumophila strains isolated from biofilms developed within a warm spring of a thermal spa. *BMC microbiology*, 13:17.

Chan, B. K., Abedon, S. T., and Loc-Carrillo, C. (2013). Phage cocktails and the future of phage therapy. *Future microbiology*, 8(6):769–83.

Chan, B. K., Turner, P. E., Kim, S., Mojibian, H. R., Eleferiades, J. A., and Narayan, D. (2018). Phage treatment of an aortic graft infected with Pseudomonas aeruginosa. *Evolution, medicine, and public health*, 2018(1):60–66.

Chang, C.-W., Hwang, Y.-H., Cheng, W.-Y., and Chang, C.-P. (2007). Effects of chlorination and heat disinfection on long-term starved Legionella pneumophila in warm water. *Journal of applied microbiology*, 102(6):1636–44.

Chatfield, C. H. and Cianciotto, N. P. (2013). Culturing, media, and handling of legionella. *Methods in molecular biology (Clifton, N.J.)*, 954:151–62.

Cherwa, J. J. and Fane, B. (2012). Microviridae. In King, A., Adams, M., Carstens, E., and Lefkowitz, E., editors, *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*, number 1997, pages 385–393. Elsevier Science, San Diego.

Ciesielski, C. A., Blaser, M. J., and Wang, W. L. (1986). Serogroup specificity of Legionella pneumophila is related to lipopolysaccharide characteristics. *Infection and immunity*, 51(2):397–404.

Clokie, M. R., Millard, A. D., Letarov, A. V., and Heaphy, S. (2011). Phages in nature. *Bacteriophage*, 1(1):31–45.

Clokie, M. R. J. and Mann, N. H. (2006). Marine cyanophages and light. *Environmental microbiology*, 8(12):2074–82.

Collins, S., Jorgensen, F., Willis, C., and Walker, J. (2015). Real-time PCR to supplement gold-standard culture-based detection of Legionella in environmental samples. *Journal of applied microbiology*, 119(4):1158–69.

Condra, J. H. and Pauling, C. (1982). Induction of the SOS system by DNA ligase-deficient growth of Escherichia coli. *Journal of general microbiology*, 128(3):613–21.

Connor, T. R., Loman, N. J., Thompson, S., Smith, A., Southgate, J., Poplawski, R., Bull, M. J., Richardson, E., Ismail, M., Thompson, S. E., Kitchen, C., Guest, M., Bakke, M., Sheppard, S. K., and Pallen, M. J. (2016). CLIMB (the Cloud Infrastructure for Microbial Bioinformatics): an online resource for the medical microbiology community. *Microbial genomics*, 2(9):e000086.

Cookson, J. T. and North, W. J. (1967). Adsorption of viruses on activated carbon. Equilibriums and kinetics of the attachment of Escherichia coli bacteriophage T4 on activated carbon. *Environmental science & technology*, 1(1):46–52.

Cooper, I. R. and Hanlon, G. W. (2010). Resistance of Legionella pneumophila serotype 1 biofilms to chlorine-based disinfection. *The Journal of hospital infection*, 74(2):152–9.

Corre, M.-H., Delafont, V., Legrand, A., Berjeaud, J.-M., and Verdon, J. (2018). Exploiting the Richness of Environmental Waterborne Bacterial Species to Find Natural Legionella pneumophila Competitors. *Frontiers in microbiology*, 9(3360).

Corry, J. E. L., Curtis, G. D. W., and Baird, R. M. (2011). *Handbook of Culture Media for Food and Water Microbiology.* Royal Society of Chemistry.

Crispim, J. S., Dias, R. S., Vidigal, P. M. P., de Sousa, M. P., da Silva, C. C., Santana, M. F., and de Paula, S. O. (2018). Screening and characterization of prophages in Desulfovibrio genomes. *Scientific reports*, 8(1):9273.

Cruz Martín, M., Ladero, V., and Alvarez, M. A. (2006). PCR identification of lysogenic Lactococcus lactis strains. *Journal fur Verbraucherschutz und Lebensmittelsicherheit*, 1(2):121–124.

Cunha, B. A., Burillo, A., and Bouza, E. (2016). Legionnaires' disease. *Lancet (London, England)*, 387(10016):376–385.

Cuppels, D. A., Vidaver, A. K., and Van Etten, J. L. (1979). Resistance to Bacteriophage 6 by Pseudomonas phaseolicola. *Journal of General Virology*, 44(2):493–504.

Cury, J., Jové, T., Touchon, M., Néron, B., and Rocha, E. P. (2016). Identification and analysis of integrons and cassette arrays in bacterial genomes. *Nucleic acids research*, 44(10):4539–50.

Czaplewski, L., Bax, R., Clokie, M., Dawson, M., Fairhead, H., Fischetti, V. A., Foster, S., Gilmore, B. F., Hancock, R. E. W., Harper, D., Henderson, I. R., Hilpert, K., Jones, B. V., Kadioglu, A., Knowles, D., Ólafsdóttir, S., Payne, D., Projan, S., Shaunak, S., Silverman, J., Thomas, C. M., Trust, T. J., Warn, P., and Rex, J. H. (2016). Alternatives to antibiotics-a pipeline portfolio review. *The Lancet. Infectious diseases*, 16(2):239–51.

Dadonaite, B., Vijayakrishnan, S., Fodor, E., Bhella, D., and Hutchinson, E. C. (2016). Filamentous influenza viruses. *The Journal of general virology*, 97(8):1755–64.

Danovaro, R., Corinaldesi, C., Dell'Anno, A., Fuhrman, J. A., Middelburg, J. J., Noble, R. T., and Suttle, C. A. (2011). Marine viruses and global climate change. *FEMS Microbiology Reviews*, 35(6):993–1034.

Davies, S. C., Fowler, T., Watson, J., Livermore, D. M., and Walker, D. (2013). Annual Report of the Chief Medical Officer: infection and the rise of antimicrobial resistance. *The Lancet*, 381(9878):1606–1609.

Day, L. (2012). Inoviridae. In King, A., Adams, M., Carstens, E., and Lefkowitz, E., editors, *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*, pages 375–383. Elsevier Science, San Diego.

de Melo, A. G., Levesque, S., and Moineau, S. (2018). Phages as friends and enemies in food processing. *Current Opinion in Biotechnology*, 49:185–190.

de Sousa, A. L., Maués, D., Lobato, A., Franco, E. F., Pinheiro, K., Araújo, F., Pantoja, Y., da Costa da Silva, A. L., Morais, J., and Ramos, R. T. J. (2018). PhageWeb - Web Interface for Rapid Identification and Characterization of Prophages in Bacterial Genomes. *Frontiers in genetics*, 9(644).

Dedrick, R. M., Guerrero-Bustamante, C. A., Garlena, R. A., Russell, D. A., Ford, K., Harris, K., Gilmour, K. C., Soothill, J., Jacobs-Sera, D., Schooley, R. T., Hatfull, G. F., and Spencer, H. (2019). Engineered bacteriophages for treatment of a patient with a disseminated drug-resistant Mycobacterium abscessus. *Nature medicine*, 25(5):730–733.

Descours, G., Cassier, P., Forey, F., Ginevra, C., Etienne, J., Lina, G., and Jarraud, S. (2014). Evaluation of BMPA, MWY, GVPC and BCYE media for the isolation of Legionella species from respiratory samples. *Journal of microbiological methods*, 98:119–21.

Díaz-Muñoz, S. L. and Koskella, B. (2014). Bacteria-phage interactions in natural environments. *Advances in applied microbiology*, 89:135–83.

Diederen, B. M. W., de Jong, C. M. A., Aarts, I., Peeters, M. F., and van der Zee, A. (2007). Molecular evidence for the ubiquitous presence of Legionella species in Dutch tap water installations. *Journal of water and health*, 5(3):375–83.

Dimitriadi, D. and Velonakis, E. (2014). Detection of Legionella spp. from Domestic Water in the Prefecture of Arta, Greece. *Journal of pathogens*, 2014(407385).

Donlan, R. (2001). Biofilm Formation: A Clinically Relevant Microbiological Process. *Clinical Infectious Diseases*, 33(8):1387–1392.

Donlan, R. M., Carpenter, J., Murga, R., Brown, E., Besser, R., and Fields, B. (2002). Monochloramine Disinfection of Biofilm-Associated Legionella pneumophila in a Potable Water Model System. In *Legionella*, pages 406–410. American Society of Microbiology.

Donlan, R. M., Forster, T., Murga, R., Brown, E., Lucas, C., Carpenter, J., and Fields, B. (2005). Legionella pneumophila associated with the protozoan Hartmannella vermiformis in a model multi-species biofilm has reduced susceptibility to disinfectants. *Biofouling*, 21(1):1–7.

Doss, J., Culbertson, K., Hahn, D., Camacho, J., and Barekzi, N. (2017). A Review of Phage Therapy against Bacterial Pathogens of Aquatic and Terrestrial Organisms. *Viruses*, 9(3):93–103.

Dowling, J. N., McDevitt, D. A., and Pasculle, A. W. (1984). Disk diffusion antimicrobial susceptibility testing of members of the family Legionellaceae including erythromycin-resistant variants of Legionella micdadei. *Journal of clinical microbiology*, 19(6):723–9.

Drulis-Kawa, Z., Majkowska-Skrobek, G., Maciejewska, B., Delattre, A.-S., and Lavigne, R. (2012). Learning from bacteriophages - advantages and limitations of phage and phage-encoded protein applications. *Current protein & peptide science*, 13(8):699–722.

Duerkop, B. A., Clements, C. V., Rollins, D., Rodrigues, J. L. M., and Hooper, L. V. (2012). A composite bacteriophage alters colonization by an intestinal commensal bacterium. *Proceedings of the National Academy of Sciences*, 109(43):17621–17626.

Dupuy, M., Mazoua, S., Berne, F., Bodet, C., Garrec, N., Herbelin, P., Ménard-Szczebara, F., Oberti, S., Rodier, M.-H., Soreau, S., Wallet, F., and Héchard, Y. (2011). Efficiency of water disinfectants against Legionella pneumophila and Acanthamoeba. *Water research*, 45(3):1087–94.

Dyson, Z. A., Brown, T. L., Farrar, B., Doyle, S. R., Tucci, J., Seviour, R. J., and Petrovski, S. (2016). Locating and Activating Molecular 'Time Bombs': Induction of Mycolata Prophages. *PLOS ONE*, 11(8):e0159957.

Edelstein, P. H. (2007). Urine antigen tests positive for Pontiac fever: implications for diagnosis and pathogenesis. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 44(2):229–31.

Edelstein, P. H. and Edelstein, M. a. (1989). WIN 57273 is bactericidal for Legionella pneumophila grown in alveolar macrophages. *Antimicrobial Agents and Chemotherapy*, 33(12):2132–2136.

Edelstein, P. H., Hu, B., and Edelstein, M. A. C. (2005). In vitro and intracellular activities of LBM415 (NVP PDF-713) against Legionella pneumophila. *Antimicrobial agents and chemotherapy*, 49(6):2533–5.

Edelstein, P. H., Snitzer, J. B., and Bridge, J. A. (1982). Enhancement of recovery of Legionella pneumophila from contaminated respiratory tract specimens by heat. *Journal of clinical microbiology*, 16(6):1061–5.

Engelhardt, T., Sahlberg, M., Cypionka, H., and Engelen, B. (2011). Induction of prophages from deep-subseafloor bacteria. *Environmental microbiology reports*, 3(4):459–65.

Engelstädter, J., Harms, K., and Johnsen, P. J. (2016). The evolutionary dynamics of integrons in changing environments. *The ISME journal*, 10(6):1296–307.

Engleberg, N. C., Carter, C., Weber, D. R., Cianciotto, N. P., and Eisenstein, B. I. (1989). DNA sequence of mip, a Legionella pneumophila gene associated with macrophage infectivity. *Infection and immunity*, 57(4):1263–70.

Erez, Z., Steinberger-Levy, I., Shamir, M., Doron, S., Stokar-Avihail, A., Peleg, Y., Melamed, S., Leavitt, A., Savidor, A., Albeck, S., Amitai, G., and Sorek, R. (2017). Communication between viruses guides lysis-lysogeny decisions. *Nature*, 541(7638):488–493.

European Centre for Disease Prevention and Control (2018). Legionnaires' disease. In: ECDC. Annual epidemiological report for 2016. Technical report, Stockholm: ECDC.

European Centre for Disease Prevention and Control (2019). Legionnaires' disease. In: ECDC. Annual epidemiological report for 2017. Technical Report January, Stockholm: ECDC.

Falkinham, J. O. (2015). Common features of opportunistic premise plumbing pathogens. *International journal of environmental research and public health*, 12(5):4533–4545.

Fan, X., Li, Y., He, R., Li, Q., and He, W. (2016). Comparative analysis of prophage-like elements in Helicobacter sp. genomes. *PeerJ*, 4:e2012.

Fan, X., Xie, L., Li, W. W., Xie, J., Pdf, T. P., Genomics, B. M. C., Fan, X., Xie, L., Li, W. W., Xie, J., Article, I., Url, A., Central, P., and Central, B. (2014). Prophage-like elements present in Mycobacterium genomes. *BMC Genomics*, 15(1):243.

Farhat, C., Mentasti, M., Jacobs, E., Fry, N. K., and Lück, C. (2011). The N-acylneuraminate cytidyltransferase gene, neuA, is heterogenous in Legionella pneumophila strains but can be used as a marker for epidemiological typing in the consensus sequence-based typing scheme. *Journal of clinical microbiology*, 49(12):4052–8.

Farnham, A., Alleyne, L., Cimini, D., and Balter, S. (2014). Legionnaires' disease incidence and risk factors, New York, New York, USA, 2002-2011. *Emerging infectious diseases*, 20(11):1795–1802.

Farrell, I. D., Barker, J. E., Miles, E. P., and Hutchison, J. G. (1990). A field study of the survival of Legionella pneumophila in a hospital hot-water system. *Epidemiology and infection*, 104(3):381–7.

Fauquet, C. (2008). Taxonomy, Classification and Nomenclature of Viruses. In *Encyclopedia of Virology*, pages 9–23. Elsevier.

Fenner, F. (1995). ICNV 1966 to ICTV 1994: the contribution of veterinary virology. *Veterinary microbiology*, 46(1-3):3–13.

Fernández, L., Rodríguez, A., and García, P. (2018). Phage or foe: An insight into the impact of viral predation on microbial communities. *ISME Journal*, 12(5):1171–1179.

Fields, B. S., Benson, R. F., and Besser, R. E. (2002). Legionella and Legionnaires' Disease: 25 Years of Investigation. *Clinical Microbiology Reviews*, 15(3):506–526.

Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Iserentant, D., Merregaert, J., Min Jou, W., Molemans, F., Raeymaekers, A., Van den Berghe, A., Volckaert, G., and Ysebaert, M. (1976). Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. *Nature*, 260(5551):500–7.

Fish, R., Kutter, E., Wheat, G., Blasdel, B., Kutateladze, M., and Kuhl, S. (2016). Bacteriophage treatment of intransigent diabetic toe ulcers: a case series. *Journal of Wound Care*, 25(Sup7):S27–S33.

Flemming, H.-c., Neu, T. R., and Wozniak, D. J. (2007). The EPS matrix: the "house of biofilm cells". *Journal of bacteriology*, 189(22):7945–7.

Flesher, a. R., Kasper, D. L., Modern, P. a., and Mason, E. O. (1980). Legionella pneumophila: growth inhibition by human pharyngeal flora. *The Journal of infectious diseases*, 142(3):313–7.

Flores, C. O., Meyer, J. R., Valverde, S., Farr, L., and Weitz, J. S. (2011). Statistical structure of host-phage interactions. *Proceedings of the National Academy of Sciences of the United States of America*, 108(28):E288–97.

Fogg, P. C. M., Colloms, S., Rosser, S., Stark, M., and Smith, M. C. M. (2014). New applications for phage integrases. *Journal of molecular biology*, 426(15):2703–16.

Fortier, L.-C. and Sekulovic, O. (2013). Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence*, 4(5):354–65.

Fraser, D. W., Tsai, T. R., Orenstein, W., Parkin, W. E., Beecham, H. J., Sharrar, R. G., Harris, J., Mallison, G. F., Martin, S. M., McDade, J. E., Shepard, C. C., and Brachman, P. S. (1977). Legionnaires' Disease. *New England Journal of Medicine*, 297(22):1189–1197.

Fu, K. P. and Neu, H. C. (1979). Inactivation of beta-lactam antibiotics by Legionella pneumophila. *Antimicrobial agents and chemotherapy*, 16(5):561–4.

Furfaro, L. L., Payne, M. S., and Chang, B. J. (2018). Bacteriophage Therapy: Clinical Trials and Regulatory Hurdles. *Frontiers in cellular and infection microbiology*, 8(October):376.

Gaia, V., Fry, N. K., Afshar, B., Lück, P. C., Meugnier, H., Etienne, J., Peduzzi, R., and Harrison, T. G. (2005). Consensus sequence-based scheme for epidemiological typing of clinical and environmental isolates of Legionella pneumophila. *Journal of Clinical Microbiology*, 43(5):2047–2052.

Gaia, V., Fry, N. K., Harrison, T. G., and Peduzzi, R. (2003). Sequence-based typing of Legionella pneumophila serogroup 1 offers the potential for true portability in legionellosis outbreak investigation. *Journal of clinical microbiology*, 41(7):2932–9.

Gao, L.-Y. and Abu Kwaik, Y. (2000). The mechanism of killing and exiting the protozoan host Acanthamoeba polyphaga by Legionella pneumophila. *Environmental Microbiology*, 2(1):79–90.

García, M. T., Jones, S., Pelaz, C., Millar, R. D., and Abu Kwaik, Y. (2007). Acanthamoeba polyphaga resuscitates viable non-culturable Legionella pneumophila after disinfection. *Environmental microbiology*, 9(5):1267–77.

Ghodke, H., Paudel, B. P., Lewis, J. S., Jergic, S., Gopal, K., Romero, Z. J., Wood, E. A., Woodgate, R., Cox, M. M., and van Oijen, A. M. (2019). Spatial and temporal organization of RecA in the Escherichia coli DNA-damage response. *eLife*, 8.

Gill, J. and Hyman, P. (2010). Phage Choice, Isolation, and Preparation for Phage Therapy. *Current Pharmaceutical Biotechnology*, 11(1):2–14.

Gillings, M. R. (2014). Integrons: Past, Present, and Future. *Microbiology and Molecular Biology Reviews*, 78(2):257–277.

Gillis, A. and Mahillon, J. (2014). Influence of lysogeny of Tectiviruses GIL01 and GIL16 on Bacillus thuringiensis growth, biofilm formation, and swarming motility. *Applied and environmental microbiology*, 80(24):7620–30.

Glick, T. H., Gregg, M. B., Berman, B., Mallison, G., Rhodes, W. W., and Kassanoff, I. (1978). Pontiac

fever. An epidemic of unknown etiology in a health department: I. Clinical and epidemiologic aspects. *American journal of epidemiology*, 107(2):149–60.

Goldberg, D. J., Wrench, J. G., Collier, P. W., Emslie, J. A., Fallon, R. J., Forbes, G. I., McKay, T. M., Macpherson, A. C., Markwick, T. A., and Reid, D. (1989). Lochgoilhead fever: outbreak of non-pneumonic legionellosis due to Legionella micdadei. *Lancet (London, England)*, 1(8633):316–8.

Gomez-Valero, L. and Buchrieser, C. (2013). Genome dynamics in Legionella: the basis of versatility and adaptation to intracellular replication. *Cold Spring Harbor perspectives in medicine*, 3(6):a009993–a009993.

Gomez-Valero, L., Rusniok, C., Rolando, M., Neou, M., Dervins-Ravault, D., Demirtas, J., Rouy, Z., Moore, R. J., Chen, H., Petty, N. K., Jarraud, S., Etienne, J., Steinert, M., Heuner, K., Gribaldo, S., Médigue, C., Glöckner, G., Hartland, E. L., and Buchrieser, C. (2014). Comparative analyses of Legionella species identifies genetic features of strains causing Legionnaires' disease. *Genome Biology*, 15(11):505.

Goutziana, G., Mouchtouri, V. A., Karanika, M., Kavagias, A., Stathakis, N. E., Gourgoulianis, K., Kremastinou, J., and Hadjichristodoulou, C. (2008). Legionella species colonization of water distribution systems, pools and air conditioning systems in cruise ships and ferries. *BMC public health*, 8:390.

Graham, F. F., White, P. S., Harte, D. J. G., and Kingham, S. P. (2012). Changing epidemiological trends of legionellosis in New Zealand, 1979-2009. *Epidemiology and infection*, 140(8):1481–96.

Grazziotin, A. L., Koonin, E. V., and Kristensen, D. M. (2017). Prokaryotic Virus Orthologous Groups (pVOGs): a resource for comparative genomics and protein family annotation. *Nucleic acids research*, 45(D1):D491–D498.

Grose, J. H. and Casjens, S. R. (2014). Understanding the enormous diversity of bacteriophages: the tailed phages that infect the bacterial family Enterobacteriaceae. *Virology*, 468-470:421–443.

Guerrieri, E., Bondi, M., Sabia, C., de Niederhäusern, S., Borella, P., and Messi, P. (2008). Effect of bacterial interference on biofilm development by Legionella pneumophila. *Current microbiology*, 57(6):532–6.

Hall, R. M. and Collis, C. M. (1995). Mobile gene cassettes and integrons: capture and spread of genes by site-specific recombination. *Molecular microbiology*, 15(4):593–600.

Hall-Stoodley, L., Costerton, J. W., and Stoodley, P. (2004). Bacterial biofilms: from the Natural environment to infectious diseases. *Nature Reviews Microbiology*, 2(2):95–108.

Hammerl, J. A., Göllner, C., Al Dahouk, S., Nöckler, K., Reetz, J., and Hertwig, S. (2016). Analysis

of the First Temperate Broad Host Range Brucellaphage (BiPBO1) Isolated from B. inopinata. *Frontiers in microbiology*, 7(24).

Hargreaves, K. R., Colvin, H. V., Patel, K. V., Clokie, J. J. P., and Clokie, M. R. J. (2013). Genetically diverse Clostridium difficile strains harboring abundant prophages in an estuarine environment. *Applied and environmental microbiology*, 79(20):6236–43.

Harrison, T. G., Afshar, B., Doshi, N., Fry, N. K., and Lee, J. V. (2009). Distribution of Legionella pneumophila serogroups, monoclonal antibody subgroups and DNA sequence types in recent clinical and environmental isolates from England and Wales (2000-2008). *European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology*, 28(7):781–91.

Harshey, R. M. (2014). Transposable Phage Mu. *Microbiology Spectrum*, 2(5):1–22.

Hatfull, G. F. (2008). Bacteriophage genomics. *Current Opinion in Microbiology*, 11(5):447–453.

Hatfull, G. F. and Hendrix, R. W. (2011). Bacteriophages and their genomes. *Current opinion in virology*, 1(4):298–303.

Hayden, R. T., Uhl, J. R., Qian, X., Hopkins, M. K., Aubry, M. C., Limper, A. H., Lloyd, R. V., and Cockerill, F. R. (2001). Direct detection of Legionella species from bronchoalveolar lavage and open lung biopsy specimens: comparison of LightCycler PCR, in situ hybridization, direct fluorescence antigen detection, and culture. *Journal of clinical microbiology*, 39(7):2618–26.

Health and Safety Executive (2013). Legionnaires' disease: Technical guidance. Part 1: The control of legionella bacteria in evaporative cooling systems. Technical report.

Héchard, Y., Ferraz, S., Bruneteau, E., Steinert, M., and Berjeaud, J.-M. (2005). Isolation and characterization of a Staphylococcus warneri strain producing an anti-Legionella peptide. *FEMS microbiology letters*, 252(1):19–23.

Helbig, J. H., Uldum, S. A., Bernander, S., Lück, P. C., Wewalka, G., Abraham, B., Gaia, V., and Harrison, T. G. (2003). Clinical utility of urinary antigen detection for diagnosis of community-acquired, travel-associated, and nosocomial legionnaires' disease. *Journal of clinical microbiology*, 41(2):838–40.

Hilbi, H., Nagai, H., Kubori, T., and Roy, C. R. (2017). Subversion of Host Membrane Dynamics by the Legionella Dot/Icm Type IV Secretion System. In Steffen Backert, E. G., editor, *Type IV Secretion in Gram-Negative and Gram-Positive Bacteria*, pages 221–242. Springer International Publishing.

Hines, S. A., Chappie, D. J., Lordo, R. A., Miller, B. D., Janke, R. J., Lindquist, H. A., Fox, K. R., Ernst, H. S., and Taft, S. C. (2014). Assessment of relative potential for Legionella species or surrogates inhalation exposure from common water uses. *Water Research*, 56:203–213.

Hitch, G., Pratten, J., and Taylor, P. W. (2004). Isolation of bacteriophages from the oral cavity. *Letters in applied microbiology*, 39(2):215–9.

Hobbs, Z. and Abedon, S. T. (2016). Diversity of phage infection types and associated terminology: the problem with 'Lytic or lysogenic'. *FEMS microbiology letters*, 363(7):1–8.

Hockett, K. L. and Baltrus, D. A. (2017). Use of the Soft-agar Overlay Technique to Screen for Bacterially Produced Inhibitory Compounds. *Journal of visualized experiments : JoVE*, (119):1–5.

Hoffman, P. S., Pine, L., and Bell, S. (1983). Production of superoxide and hydrogen peroxide in medium used to culture Legionella pneumophila: catalytic decomposition by charcoal. *Applied and environmental microbiology*, 45(3):784–91.

Hoffmann, C., Harrison, C. F., and Hilbi, H. (2014). The natural alternative: protozoa as cellular models for Legionella infection. *Cellular microbiology*, 16(1):15–26.

Horwitz, M. A. (1984). Phagocytosis of the Legionnaires' disease bacterium (Legionella pneumophila) occurs by a novel mechanism: engulfment within a pseudopod coil. *Cell*, 36(1):27–33.

Howard-Varona, C., Hargreaves, K. R., Abedon, S. T., and Sullivan, M. B. (2017). Lysogeny in nature: mechanisms, impact and ecology of temperate phages. *The ISME journal*, 11(7):1511–1520.

Hoyle, N., Zhvaniya, P., Balarjishvili, N., Bolkvadze, D., Nadareishvili, L., Nizharadze, D., Wittmann, J., Rohde, C., and Kutateladze, M. (2018). Phage therapy against Achromobacter xylosoxidans lung infection in a patient with cystic fibrosis: a case report. *Research in Microbiology*, 169(9):540–542.

Hui, J. G. K., Mai-Prochnow, A., Kjelleberg, S., McDougald, D., and Rice, S. A. (2014). Environmental cues and genes involved in establishment of the superinfective Pf4 phage of Pseudomonas aeruginosa. *Frontiers in microbiology*, 5(654).

Hussong, D., Colwell, R. R., O'Brien, M., Weiss, E., Pearson, A. D., Weiner, R. M., and Burge, W. D. (1987). Viable Legionella Pneumophila Not Detectable by Culture on Agar Media. *Nature biotechnology*, 5:709–712.

ICTV (2018). International Committee on Taxonomy of Viruses (ICTV) Master Species List 2018b.v1.

Jain, P., Hartman, T. E., Eisenberg, N., O'Donnell, M. R., Kriakov, J., Govender, K., Makume, M., Thaler, D. S., Hatfull, G. F., Sturm, A. W., Larsen, M. H., Moodley, P., and Jacobs, W. R. (2012). $\phi(2)$GFP10, a high-intensity fluorophage, enables detection and rapid drug susceptibility testing of Mycobacterium tuberculosis directly from sputum samples. *Journal of clinical microbiology*, 50(4):1362–9.

James, C. E., Davies, E. V., Fothergill, J. L., Walshaw, M. J., Beale, C. M., Brockhurst, M. A., and

Winstanley, C. (2015). Lytic activity by temperate phages of Pseudomonas aeruginosa in long-term cystic fibrosis chronic lung infections. *The ISME journal*, 9(6):1391–8.

Joseph, C., Lee, J., Surman-Lee, S., Drasar, V., Cresp, S., and Briand, E. (2011). EWGLI Technical Guidelines for the Investigation , Control and Prevention of Travel Associated Legionnaires ' Disease. Technical Report September.

Joseph, S. J., Cox, D., Wolff, B., Morrison, S. S., Kozak-Muiznieks, N. A., Frace, M., Didelot, X., Castillo-Ramirez, S., Winchell, J., Read, T. D., and Dean, D. (2016). Dynamics of genome change among Legionella species. *Scientific reports*, 6(33442).

Kagan, J. C. and Roy, C. R. (2002). Legionella phagosomes intercept vesicular traffic from endoplasmic reticulum exit sites. *Nature cell biology*, 4(12):945–54.

Kahlmeter, G., Brown, D. F. J., Goldstein, F. W., MacGowan, A. P., Mouton, J. W., Osterlund, A., Rodloff, A., Steinbakk, M., Urbaskova, P., and Vatopoulos, A. (2003). European harmonization of MIC breakpoints for antimicrobial susceptibility testing of bacteria. *The Journal of antimicrobial chemotherapy*, 52(2):145–8.

Kaper, J. B., Morris, J. G., and Levine, M. M. (1995). Cholera. *Clinical microbiology reviews*, 8(1):48–86.

Karumidze, N., Thomas, J. A., Kvatadze, N., Goderdzishvili, M., Hakala, K. W., Weintraub, S. T., Alavidze, Z., and Hardies, S. C. (2012). Characterization of lytic Pseudomonas aeruginosa bacteriophages via biological properties and genomic sequences. *Applied microbiology and biotechnology*, 94(6):1609–1617.

Katoh, K. and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, 30(4):772–80.

Keen, E. C. (2015). A century of phage research: bacteriophages and the shaping of modern biology. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 37(1):6–9.

Khan Mirzaei, M. and Nilsson, A. S. (2015). Isolation of phages for phage therapy: a comparison of spot tests and efficiency of plating analyses for determination of host range and efficacy. *PloS one*, 10(3):e0118557.

Kilvington, S. and Price, J. (1990). Survival of Legionella pneumophila within cysts of Acanthamoeba polyphaga following chlorine exposure. *The Journal of applied bacteriology*, 68(5):519–25.

Kim, B. R., Anderson, J. E., Mueller, S. a., Gaines, W. a., and Kendall, a. M. (2002). Literature review–efficacy of various disinfectants against Legionella in water systems. *Water research*, 36(18):4433–44.

Kirschner, A. K. T. (2016). Determination of viable legionellae in engineered water systems: Do we find what we are looking for? *Water research*, 93:276–288.

Knirsch, C. A., Jakob, K., Schoonmaker, D., Kiehlbauch, J. A., Wong, S. J., Della-Latta, P., Whittier, S., Layton, M., and Scully, B. (2000). An outbreak of Legionella micdadei pneumonia in transplant patients: evaluation, molecular epidemiology, and control. *The American journal of medicine*, 108(4):290–5.

Knoll, B. M. and Mylonakis, E. (2014). Antibacterial bioagents based on principles of bacteriophage biology: an overview. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 58(4):528–534.

Koskella, B. (2013). Phage-mediated selection on microbiota of a long-lived host. *Current biology : CB*, 23(13):1256–60.

Koskella, B., Thompson, J. N., Preston, G. M., and Buckling, A. (2011). Local biotic environment shapes the spatial scale of bacteriophage adaptation to bacteria. *The American naturalist*, 177(4):440–51.

Kozak-Muiznieks, N. A., Lucas, C. E., Brown, E., Pondo, T., Taylor, T. H., Frace, M., Miskowski, D., and Winchell, J. M. (2014). Prevalence of sequence types among clinical and environmental isolates of Legionella pneumophila serogroup 1 in the United States from 1982 to 2012. *Journal of clinical microbiology*, 52(1):201–11.

Kramberger, P., Ciringer, M., Štrancar, A., and Peterka, M. (2012). Evaluation of nanoparticle tracking analysis for total virus particle determination. *Virology journal*, 9:265.

Kreuzer, K. N. (2013). DNA Damage Responses in Prokaryotes: Regulating Gene Expression, Modulating Growth Patterns, and Manipulating Replication Forks. *Cold Spring Harbor Perspectives in Biology*, 5(11):a012674–a012674.

Kuroki, T., Watanabe, Y., Teranishi, H., Izumiyama, S., Amemura-Maekawa, J., and Kura, F. (2017). Legionella prevalence and risk of legionellosis in Japanese households. *Epidemiology and infection*, 145(7):1398–1408.

Kutateladze, M. and Adamia, R. (2010). Bacteriophages as potential new therapeutics to replace or supplement antibiotics. *Trends in biotechnology*, 28(12):591–5.

Kutter, E. (2001). Bacteriophages. In *Encyclopedia of Genetics*, pages 179–186. Elsevier.

Kutter, E., De Vos, D., Gvasalia, G., Alavidze, Z., Gogokhia, L., Kuhl, S., and Abedon, S. (2010). Phage Therapy in Clinical Practice: Treatment of Human Infections. *Current Pharmaceutical Biotechnology*, 11(1):69–86.

Lamb, R. A. and Choppin, P. W. (1983). The gene structure and replication of influenza virus. *Annual review of biochemistry*, 52:467–506.

Lammertyn, E., Vande Voorde, J., Meyen, E., Maes, L., Mast, J., and Anné, J. (2008). Evidence for the presence of Legionella bacteriophages in environmental water samples. *Microbial ecology*, 56(1):191–7.

Lavigne, R., Molineux, I., and Kropinski, A. (2012). Caudovirales. In King, A., Adams, M., Carstens, E., and Lefkowitz, E., editors, *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*, pages 39–45. Elsevier Science, San Diego.

Laxminarayan, R., Matsoso, P., Pant, S., Brower, C., Røttingen, J.-A., Klugman, K., and Davies, S. (2016). Access to effective antimicrobials: a worldwide challenge. *The Lancet*, 387(10014):168–175.

Lee, S., Lewis, D. E., and Adhya, S. (2018). The Developmental Switch in Bacteriophage λ: A Critical Role of the Cro Protein. *Journal of Molecular Biology*, 430(1):58–68.

Lemire, S., Figueroa-Bossi, N., and Bossi, L. (2011). Bacteriophage Crosstalk: Coordination of Prophage Induction by Trans-Acting Antirepressors. *PLoS Genetics*, 7(6):e1002149.

Lenski, R. E. and Levin, B. R. (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.

Leoni, E., Legnani, P. P., Bucci Sabattini, M. A., and Righi, F. (2001). Prevalence of legionella spp. in swimming pool environment. *Water Research*, 35(15):3749–3753.

Li, L., Mendis, N., Trigui, H., Oliver, J. D., and Faucher, S. P. (2014). The importance of the viable but non-culturable state in human bacterial pathogens. *Frontiers in microbiology*, 5(258).

Li, L., Yang, H., Lin, S., and Jia, S. (2010). Classification of 17 newly isolated virulent bacteriophages of Pseudomonas aeruginosa. *Canadian journal of microbiology*, 56(11):925–33.

Lim, W. S., Baudouin, S. V., George, R. C., Hill, A. T., Jamieson, C., Le Jeune, I., Macfarlane, J. T., Read, R. C., Roberts, H. J., Levy, M. L., Wani, M., and Woodhead, M. A. (2009). BTS guidelines for the management of community acquired pneumonia in adults: update 2009. *Thorax*, 64(Suppl 3):iii1–iii55.

Lima-Mendez, G., Van Helden, J., Toussaint, A., and Leplae, R. (2008). Prophinder: a computational tool for prophage prediction in prokaryotic genomes. *Bioinformatics (Oxford, England)*, 24(6):863–5.

Loc-Carrillo, C. and Abedon, S. T. (2011). Pros and cons of phage therapy. *Bacteriophage*, 1(2):111–114.

Loiseau, C., Schlusselhuber, M., Bigot, R., Bertaux, J., Berjeaud, J.-M., and Verdon, J. (2015).

Surfactin from Bacillus subtilis displays an unexpected anti-Legionella activity. *Applied microbiology and biotechnology*, 99(12):5083–93.

Lorenz, N., Reiger, M., Toro-Nahuelpan, M., Brachmann, A., Poettinger, L., Plener, L., Lassak, J., and Jung, K. (2016). Identification and Initial Characterization of Prophages in Vibrio campbellii. *PloS one*, 11(5):e0156010.

Łoś, J. M., Łoś, M., Wegrzyn, G., and Wegrzyn, A. (2009). Differential efficiency of induction of various lambdoid prophages responsible for production of Shiga toxins in response to different induction agents. *Microbial Pathogenesis*, 47(6):289–298.

Lück, C., Fry, N. K., Helbig, J. H., Jarraud, S., and Harrison, T. G. (2013). *Typing Methods for Legionella Christian*, volume 954 of *Methods in Molecular Biology*. Humana Press, Totowa, NJ.

Magare, B., Nair, A., and Khairnar, K. (2017). Isolation of bacteriophages from air using vacuum filtration technique: an improved and novel method. *Journal of applied microbiology*, 123(4):896–902.

Magnet, A., Peralta, R. H. S., Gomes, T. S., Izquierdo, F., Fernandez-Vadillo, C., Galvan, A. L., Pozuelo, M. J., Pelaz, C., Fenoy, S., and Del Águila, C. (2015). Vectorial role of Acanthamoeba in Legionella propagation in water for human use. *The Science of the total environment*, 505:889–95.

Maniloff, J. (2012). Plasmaviridae. In King, A., Adams, M., Carstens, E., and Lefkowitz, E., editors, *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*, volume 2, pages 263–266. Elsevier Science, San Diego.

Marchand, a., Verdon, J., Lacombe, C., Crapart, S., Héchard, Y., and Berjeaud, J. M. (2011). Anti-Legionella activity of staphylococcal hemolytic peptides. *Peptides*, 32(5):845–851.

Marchesi, I., Ferranti, G., Mansi, A., Marcelloni, A. M., Proietto, A. R., Saini, N., Borella, P., and Bargellini, A. (2016). Control of Legionella Contamination and Risk of Corrosion in Hospital Water Networks following Various Disinfection Procedures. *Applied and environmental microbiology*, 82(10):2959–65.

Martínez, E. and Campos-Gómez, J. (2016). Pf Filamentous Phage Requires UvrD for Replication in Pseudomonas aeruginosa. *mSphere*, 1(1):1–14.

Massip, C., Descours, G., Ginevra, C., Doublet, P., Jarraud, S., and Gilbert, C. (2017). Macrolide resistance in Legionella pneumophila: the role of LpeAB efflux pump. *The Journal of antimicrobial chemotherapy*, 72(5):1327–1333.

Mattila, S., Ruotsalainen, P., and Jalasvuori, M. (2015). On-demand isolation of bacteriophages against drug-resistant bacteria for personalized phage therapy. *Frontiers in Microbiology*, 6(NOV):1–7.

Mavrich, T. N. and Hatfull, G. F. (2017). Bacteriophage evolution differs by host, lifestyle and genome. *Nature microbiology*, 2(17112).

McAdam, P. R., Vander Broek, C. W., Lindsay, D. S. J., Ward, M. J., Hanson, M. F., Gillies, M., Watson, M., Stevens, J. M., Edwards, G. F., and Fitzgerald, J. R. (2014). Gene flow in environmental Legionella pneumophila leads to genetic and pathogenic heterogeneity within a Legionnaires' disease outbreak. *Genome biology*, 15(11):504.

McDonald, J. E., Smith, D. L., Fogg, P. C. M., McCarthy, A. J., and Allison, H. E. (2010). High-throughput method for rapid induction of prophages from lysogens and its application in the study of Shiga Toxin-encoding Escherichia coli strains. *Applied and environmental microbiology*, 76(7):2360–2365.

McKinney, R. M., Porschen, R. K., Edelstein, P. H., Bissett, M. L., Harris, P. P., Bondell, S. P., Steigerwalt, A. G., Weaver, R. E., Ein, M. E., Lindquist, D. S., Kops, R. S., and Brenner, D. J. (1981). Legionella longbeachae species nova, another etiologic agent of human pneumonia. *Annals of Internal Medicine*, 94(6):739–743.

McKinney, R. M., Thacker, L., Harris, P. P., Lewallen, K. R., Hebert, G. A., Edelstein, P. H., and Thomason, B. M. (1979). Four serogroups of Legionnaires' disease bacteria defined by direct immunofluorescence. *Annals of internal medicine*, 90(4):621–4.

Meaden, S. and Koskella, B. (2013). Exploring the risks of phage application in the environment. *Frontiers in microbiology*, 4(358).

Means, E. G. and Olson, B. H. (1981). Coliform inhibition by bacteriocin-like substances in drinking water distribution systems. *Applied and environmental microbiology*, 42(3):506–12.

Mentasti, M., Fry, N. K., Afshar, B., Palepou-Foxley, C., Naik, F. C., and Harrison, T. G. (2012). Application of Legionella pneumophila-specific quantitative real-time PCR combined with direct amplification and sequence-based typing in the diagnosis and epidemiological investigation of Legionnaires' disease. *European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology*, 31(8):2017–28.

Mentasti, M., Underwood, A., Lück, C., Kozak-Muiznieks, N. A., Harrison, T. G., and Fry, N. K. (2014). Extension of the Legionella pneumophila sequence-based typing scheme to include strains carrying a variant of the N-acylneuraminate cytidylyltransferase gene. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*, 20(7):O435–41.

Mentasti, M. and Fry, N. (2012). Analysis of the N -Acylneuraminate Cytidyltransferase homologue (neuAh) found in some non-serogroup 1 strains of Legionella pneumophila. Technical report.

Mérault, N., Rusniok, C., Jarraud, S., Gomez-Valero, L., Cazalet, C., Marin, M., Brachet, E., Aegerter, P., Gaillard, J. L., Etienne, J., Herrmann, J. L., DELPH-I Study Group, Lawrence, C., and Buchrieser, C. (2011). Specific real-time PCR for simultaneous detection and identification of Legionella pneumophila serogroup 1 in water and clinical samples. *Applied and environmental microbiology*, 77(5):1708–17.

Millard, A. D. (2009). Isolation of cyanophages from aquatic environments. In Clokie, M. and Kropinski, A., editors, *Bacteriophages Methods and Protocols, Volume 1: Isolation, Characterization, and Interactions*, volume 501, chapter 4, pages 33–42. Humana Press.

Miller, J. M., Binnicker, M. J., Campbell, S., Carroll, K. C., Chapin, K. C., Gilligan, P. H., Gonzalez, M. D., Jerris, R. C., Kehl, S. C., Patel, R., Pritt, B. S., Richter, S. S., Robinson-Dunn, B., Schwartzman, J. D., Snyder, J. W., Telford, S., Theel, E. S., Thomson, R. B., Weinstein, M. P., and Yao, J. D. (2018). A Guide to Utilization of the Microbiology Laboratory for Diagnosis of Infectious Diseases: 2018 Update by the Infectious Diseases Society of America and the American Society for Microbiology. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 67(6):e1–e94.

Mills, S., Shanahan, F., Stanton, C., Hill, C., Coffey, A., and Ross, R. P. (2013). Movers and shakers: influence of bacteriophages in shaping the mammalian gut microbiota. *Gut Microbes*, 4(1):4–16.

Molmeret, M., Jones, S., Santic, M., Habyarimana, F., Esteban, M. T. G., and Kwaik, Y. A. (2010). Temporal and spatial trigger of post-exponential virulence-associated regulatory cascades by Legionella pneumophila after bacterial escape into the host cell cytosol. *Environmental microbiology*, 12(3):704–15.

Molofsky, A. B. and Swanson, M. S. (2004). Differentiate to thrive: lessons from the Legionella pneumophila life cycle. *Molecular microbiology*, 53(1):29–40.

Montagna, M. T., De Giglio, O., Cristina, M. L., Napoli, C., Pacifico, C., Agodi, A., Baldovin, T., Casini, B., Coniglio, M. A., D'Errico, M. M., Delia, S. A., Deriu, M. G., Guida, M., Laganà, P., Liguori, G., Moro, M., Mura, I., Pennino, F., Privitera, G., Romano Spica, V., Sembeni, S., Spagnolo, A. M., Tardivo, S., Torre, I., Valeriani, F., Albertini, R., and Pasquarella, C. (2017). Evaluation of Legionella Air Contamination in Healthcare Facilities by Different Sampling Methods: An Italian Multicenter Study. *International journal of environmental research and public health*, 14(7):670.

Mukherjee, S., Huntemann, M., Ivanova, N., Kyrpides, N. C., and Pati, A. (2015). Large-scale contamination of microbial isolate genomes by Illumina PhiX control. *Standards in genomic sciences*, 10(18).

Murdoch, D. R. (2003). Diagnosis of Legionella Infection. *Medical Microbiology*, 36:64–69.

Muschel, L. H. and Schmoker, K. (1966). Activity of mitomycin C, other antibiotics, and serum against lysogenic bacteria. *Journal of bacteriology*, 92(4):967–71.

Nale, J. Y., Spencer, J., Hargreaves, K. R., Buckley, A. M., Trzepiński, P., Douce, G. R., and Clokie, M. R. J. (2016). Bacteriophage Combinations Significantly Reduce Clostridium difficile Growth In Vitro and Proliferation In Vivo. *Antimicrobial Agents and Chemotherapy*, 60(2):968–981.

Nash, T. W., Libby, D. M., and Horwitz, M. A. (1984). Interaction between the legionnaires' disease bacterium (Legionella pneumophila) and human alveolar macrophages. Influence of antibody, lymphokines, and hydrocortisone. *The Journal of clinical investigation*, 74(3):771–82.

Newton, H. J., Ang, D. K. Y., van Driel, I. R., and Hartland, E. L. (2010). Molecular pathogenesis of infections caused by Legionella pneumophila. *Clinical microbiology reviews*, 23(2):274–298.

Nielsen, K., Bangsborg, J. M., and Høiby, N. (2000). Susceptibility of Legionella species to five antibiotics and development of resistance by exposure to erythromycin, ciprofloxacin, and rifampicin. *Diagnostic microbiology and infectious disease*, 36(1):43–8.

Niu, Y. D., Cook, S. R., Wang, J., Klima, C. L., Hsu, Y.-h., Kropinski, A. M., Turner, D., and McAllister, T. A. (2015). Comparative analysis of multiple inducible phages from Mannheimia haemolytica. *BMC microbiology*, 15:175.

O'Connell, W. A., Bangsborg, J. M., and Cianciotto, N. P. (1995). Characterization of a Legionella micdadei mip mutant. *Infection and immunity*, 63(8):2840–5.

Okafor, N. (2011). *Environmental Microbiology of Aquatic and Waste Systems*. Springer Netherlands.

Oksanen, H. and Bamford, D. (2012a). Tectiviridae. In King, A., Adams, M., Carstens, E., and Lefkowitz, E., editors, *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*, volume 1, pages 317–321. Elsevier Science, San Diego.

Oksanen, H. and Bamford, J. (2012b). Corticoviridae. In King, A., Adams, M., Carstens, E., and Lefkowitz, E., editors, *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*, pages 179–182. Elsevier Science, San Diego.

Oliveira, J., Mahony, J., Hanemaaijer, L., Kouwen, T. R., Neve, H., MacSharry, J., and van Sinderen, D. (2017). Detecting Lactococcus lactis prophages by mitomycin C-mediated induction coupled to flow cytometry analysis. *Frontiers in Microbiology*, 8(1343).

Oliver, J. D. (2000). The Public Health Significance of Viable but Nonculturable Bacteria. In Colwell R.R., G. D., editor, *Nonculturable Microorganisms in the Environment*, chapter 16, pages 277–300. Springer US, Boston, MA.

Oliver, J. D. (2010). Recent findings on the viable but nonculturable state in pathogenic bacteria. *FEMS microbiology reviews*, 34(4):415–25.

O'Neill, J. (2016). Tackling drug-resistant infections globally: Final report and recommendations. Technical report.

Onodera, S., Olkkonen, V. M., Gottlieb, P., Strassman, J., Qiao, X. Y., Bamford, D. H., and Mindich, L. (1992). Construction of a transducing virus from double-stranded RNA bacteriophage phi6: establishment of carrier states in host cells. *Journal of virology*, 66(1):190–6.

Oppenheim, A. B., Kobiler, O., Stavans, J., Court, D. L., and Adhya, S. (2005). Switches in Bacteriophage Lambda Development. *Annual Review of Genetics*, 39(1):409–429.

Ortmann, A. C. and Suttle, C. A. (2009). Determination of virus abundance by epifluorescence microscopy. In Clokie, M. and Kropinski, A., editors, *Methods in molecular biology (Clifton, N.J.)*, volume 501, chapter 10, pages 87–95. Humana Press.

O'Sullivan, L., Buttimer, C., McAuliffe, O., Bolton, D., and Coffey, A. (2016). Bacteriophage-based tools: recent advances and novel applications. *F1000Research*, 5(0):2782.

Parthuisot, N., West, N. J., Lebaron, P., and Baudart, J. (2010). High diversity and abundance of Legionella spp. in a pristine river and impact of seasonal and anthropogenic effects. *Applied and environmental microbiology*, 76(24):8201–10.

Peabody, M. A., Caravas, J. A., Morrison, S. S., Mercante, J. W., Prystajecky, N. A., Raphael, B. H., and Brinkman, F. S. L. (2017). Characterization of Legionella Species from Watersheds in British Columbia, Canada. *mSphere*, 2(4):1–11.

Penner, J. C., Ferreira, J. A. G., Secor, P. R., Sweere, J. M., Birukova, M. K., Joubert, L.-M., Haagensen, J. A. J., Garcia, O., Malkovskiy, A. V., Kaber, G., Nazik, H., Manasherob, R., Spormann, A. M., Clemons, K. V., Stevens, D. A., and Bollyky, P. L. (2016). Pf4 bacteriophage produced by Pseudomonas aeruginosa inhibits Aspergillus fumigatus metabolism via iron sequestration. *Microbiology*, 162(9):1583–1594.

Percival, S. L. and Williams, D. W. (2013). *Legionella*. Elsevier, second edi edition.

Pérez Pulido, R., Grande Burgos, M. J., Gálvez, A., and Lucas López, R. (2016). Application of bacteriophages in post-harvest control of human pathogenic and food spoiling bacteria. *Critical reviews in biotechnology*, 36(5):851–61.

Perna, N. T., Plunkett, G., Burland, V., Mau, B., Glasner, J. D., Rose, D. J., Mayhew, G. F., Evans, P. S., Gregor, J., Kirkpatrick, H. A., Pósfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grotbeck, E. J., Davis, N. W., Lim, A., Dimalanta, E. T., Potamousis, K. D., Apodaca, J.,

Anantharaman, T. S., Lin, J., Yen, G., Schwartz, D. C., Welch, R. A., and Blattner, F. R. (2001). Genome sequence of enterohaemorrhagic Escherichia coli O157:H7. *Nature*, 409(6819):529–33.

Petrova, O. E. and Sauer, K. (2016). Escaping the biofilm in more than one way: desorption, detachment or dispersion. *Current Opinion in Microbiology*, 30(607):67–78.

Phin, N., Parry-Ford, F., Harrison, T., Stagg, H. R., Zhang, N., Kumar, K., Lortholary, O., Zumla, A., and Abubakar, I. (2014). Epidemiology and clinical management of Legionnaires' disease. *The Lancet. Infectious diseases*, 14(10):1011–21.

Picksley, S. M., Attfield, P. V., and Lloyd, R. G. (1984). Repair of DNA double-strand breaks in Escherichia coli K12 requires a functional recN product. *Molecular & general genetics : MGG*, 195(1-2):267–74.

Pierre, D. M., Baron, J., Yu, V. L., and Stout, J. E. (2017). Diagnostic testing for Legionnaires' disease. *Annals of clinical microbiology and antimicrobials*, 16(1):59.

Pine, L., George, J. R., Reeves, M. W., and Harrell, W. K. (1979). Development of a chemically defined liquid medium for growth of Legionella pneumophila. *Journal of clinical microbiology*, 9(5):615–26.

Piuri, M., Jacobs, W. R., and Hatfull, G. F. (2009). Fluoromycobacteriophages for rapid, specific, and sensitive antibiotic susceptibility testing of Mycobacterium tuberculosis. *PloS one*, 4(3):e4870.

Poranen, M. and Bamford, D. (2012). Cystoviridae. In King, A., Adams, M., Carstens, E., and Lefkowitz, E., editors, *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*, pages 515–518. Elsevier Science, San Diego.

Poranen, M. M., Mäntynen, S., and Ictv Report Consortium (2017). ICTV Virus Taxonomy Profile: Cystoviridae. *The Journal of general virology*, 98(10):2423–2424.

Pourcel, C., Midoux, C., Vergnaud, G., and Latino, L. (2017). A carrier state is established in Pseudomonas aeruginosa by phage LeviOr01, a newly isolated ssRNA levivirus. *The Journal of general virology*, 98(8):2181–2189.

Prangishvili, D. (2012a). Lipothrixviridae. In King, A., Adams, M., Carstens, E., and Lefkowitz, E., editors, *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*, pages 211–221. Elsevier Science, San Diego.

Prangishvili, D. (2012b). Rudiviridae. In King, A., Adams, M., Carstens, E., and Lefkowitz, E., editors, *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*, number 2008, pages 311–315. Elsevier Science, San Diego.

Price, C. T. D., Al-Quadan, T., Santic, M., Rosenshine, I., and Abu Kwaik, Y. (2011). Host protea-

somal degradation generates amino acids essential for intracellular bacterial growth. *Science (New York, N.Y.)*, 334(6062):1553–7.

Prudhomme, M., Attaiech, L., Sanchez, G., Martin, B., and Claverys, J.-P. (2006). Antibiotic Stress Induces Genetic Transformability in the Human Pathogen Streptococcus pneumoniae. *Science*, 313(5783):89–92.

Public Health England (2018). Monthly Legionella Report November 2018. Technical Report November.

Quero, S., Párraga-Niño, N., Barrabeig, I., Sala, M. R., Pedro-Botet, M. L., Monsó, E., Jané, M., Sabrià, M., and Garcia-Núñez, M. (2018). Population structure of Environmental and Clinical Legionella pneumophila isolates in Catalonia. *Scientific reports*, 8(1):6241.

Ratcliff, R. M., Lanser, J. A., Manning, P. A., and Heuzenroeder, M. W. (1998). Sequence-based classification scheme for the genus Legionella targeting the mip gene. *Journal of clinical microbiology*, 36(6):1560–7.

Ratzow, S., Gaia, V., Helbig, J. H., Fry, N. K., and Lück, P. C. (2007). Addition of neuA, the gene encoding N-acylneuraminate cytidylyl transferase, increases the discriminatory ability of the consensus sequence-based scheme for typing Legionella pneumophila serogroup 1 strains. *Journal of Clinical Microbiology*, 45(6):1965–1968.

Raya, R. and Elvira, M. H. (2009). Isolation of Phage via Induction of Lysogens. In Martha R. J. Clokie, A. M. K., editor, *Bacteriophages: Methods and Protocols, Volume 1: Isolation, Characterization, and Interactions*, volume 501, chapter 3, pages 23–32. Humana Press, illustrate edition.

Rêgo, A., Chandran, V., and Waksman, G. (2010). Two-step and one-step secretion mechanisms in Gram-negative bacteria: contrasting the type IV secretion system and the chaperone-usher pathway of pilus biogenesis. *Biochemical Journal*, 425(3):475–488.

Reinthaler, F. F., Sattler, J., Schaffler-Dullnig, K., Weinmayr, B., and Marth, E. (1993). Comparative study of procedures for isolation and cultivation of Legionella pneumophila from tap water in hospitals. *Journal of clinical microbiology*, 31(5):1213–6.

Rice, S. A., Tan, C. H., Mikkelsen, P. J., Kung, V., Woo, J., Tay, M., Hauser, A., McDougald, D., Webb, J. S., and Kjelleberg, S. (2009). The biofilm life cycle and virulence of Pseudomonas aeruginosa are dependent on a filamentous prophage. *The ISME journal*, 3(3):271–82.

Ristroph, J. D., Hedlund, K. W., and Allen, R. G. (1980). Liquid medium for growth of Legionella pneumophila. *Journal of clinical microbiology*, 11(1):19–21.

Rohan, V. T. and Kropinski, A. M. (2009). Bacteriophage Enrichment from Water and Soil. In Clokie, M. R. J. and Kropinski, A. M., editors, *Bacteriophages: Methods and Protocols, Volume 1:*

Isolation, Characterization, and Interactions, Bacteriophages: Methods and Protocols, chapter 2. Humana Press.

Romantschuk, M. and Bamford, D. H. (1981). phi 6-resistant phage-producing mutants of Pseudomonas phaseolicola. *The Journal of general virology*, 56(Pt 2):287–95.

Rountree, P. M. (1955). The role of divalent cations in the multiplication of staphylococcal bacteriophages. *Journal of general microbiology*, 12(2):275–87.

Roux, S., Krupovic, M., Daly, R. A., Borges, A. L., Nayfach, S., Schulz, F., Cheng, J.-f., Ivanova, N. N., Bondy-denomy, J., Wrighton, K. C., Woyke, T., Visel, A., Kyrpides, N. C., and Eloe-fadrosh, E. A. (2019). Cryptic inoviruses are pervasive in bacteria and archaea across Earth's biomes.

Roux, S., Krupovic, M., Poulet, A., Debroas, D., and Enault, F. (2012). Evolution and diversity of the Microviridae viral family through a collection of 81 new complete genomes assembled from virome reads. *PloS one*, 7(7):e40418.

Rowbotham, T. J. (1980). Preliminary report on the pathogenicity of Legionella pneumophila for freshwater and soil amoebae. *Journal of clinical pathology*, 33(12):1179–1183.

Sabrià, M., Pedro-Botet, M. L., Gómez, J., Roig, J., Vilaseca, B., Sopena, N., Baños, V., and Legionnaires Disease Therapy Group (2005). Fluoroquinolones vs macrolides in the treatment of Legionnaires disease. *Chest*, 128(3):1401–5.

Sagripanti, J. L., Routson, L. B., and Lytle, C. D. (1993). Virus inactivation by copper or iron ions alone and in the presence of peroxide. *Applied and environmental microbiology*, 59(12):4374–6.

Saleri, N. and Ryan, E. T. (2018). Respiratory infections. In Keystone, J. S., Kozarsky, P. E., Connor, B. A., Nothdurft, H. D., Mendelson, M., and Leder, K., editors, *Travel Medicine*, chapter 59, pages 527–538. Elsevier Health Sciences.

Salmond, G. P. and Fineran, P. C. (2015). A century of the phage: Past, present and future. *Nature Reviews Microbiology*, 13(12):777–786.

Sanger, F., Coulson, A. R., Friedmann, T., Air, G. M., Barrell, B. G., Brown, N. L., Fiddes, J. C., Hutchison, C. A., Slocombe, P. M., and Smith, M. (1978). The nucleotide sequence of bacteriophage ϕX174. *Journal of Molecular Biology*, 125(2) : 225 − −246.

Sarkar-Banerjee, S., Goyal, S., Gao, N., Mack, J., Thompson, B., Dunlap, D., Chattopadhyay, K., and Finzi, L. (2018). Specifically bound lambda repressor dimers promote adjacent non-specific binding. *PLoS ONE*, 13(4):1–19.

Sassi, M., Bebeacua, C., Drancourt, M., and Cambillau, C. (2013). The first structure of a my-

cobacteriophage, the Mycobacterium abscessus subsp. bolletii phage Araucaria. *Journal of virology*, 87(14):8099–109.

Satyanarayana, T., Prakash, A., and Johri, B. N. (2012). *Microorganisms in Environmental Management*, volume 9789400722. Springer Netherlands, Dordrecht.

Scanlan, P. D., Hall, A. R., Blackshields, G., Friman, V.-P., Davis, M. R., Goldberg, J. B., and Buckling, A. (2015). Coevolution with bacteriophages drives genome-wide host evolution and constrains the acquisition of abiotic-beneficial mutations. *Molecular biology and evolution*, 32(6):1425–35.

Schmelcher, M. and Loessner, M. J. (2014). Application of bacteriophages for detection of foodborne pathogens. *Bacteriophage*, 4(2):e28137.

Schofield, D. A. and Westwater, C. (2009). Phage-mediated bioluminescent detection of Bacillus anthracis. *Journal of applied microbiology*, 107(5):1468–78.

Schooley, R. T., Biswas, B., Gill, J. J., Hernandez-Morales, A., Lancaster, J., Lessor, L., Barr, J. J., Reed, S. L., Rohwer, F., Benler, S., Segall, A. M., Taplitz, R., Smith, D. M., Kerr, K., Kumaraswamy, M., Nizet, V., Lin, L., McCauley, M. D., Strathdee, S. A., Benson, C. A., Pope, R. K., Leroux, B. M., Picel, A. C., Mateczun, A. J., Cilwa, K. E., Regeimbal, J. M., Estrella, L. A., Wolfe, D. M., Henry, M. S., Quinones, J., Salka, S., Bishop-Lilly, K. A., Young, R., and Hamilton, T. (2017). Development and Use of Personalized Bacteriophage-Based Therapeutic Cocktails To Treat a Patient with a Disseminated Resistant Acinetobacter baumannii Infection. *Antimicrobial agents and chemotherapy*, 61(10).

Schuch, R. and Fischetti, V. A. (2009). The secret life of the anthrax agent Bacillus anthracis: bacteriophage-mediated ecological adaptations. *PloS one*, 4(8):e6532.

Seaman, P. F. and Day, M. J. (2007). Isolation and characterization of a bacteriophage with an unusually large genome from the Great Salt Plains National Wildlife Refuge, Oklahoma, USA. *FEMS Microbiology Ecology*, 60(1):1–13.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics (Oxford, England)*, 30(14):2068–9.

Segal, G. (2013). Identification of Legionella Effectors Using Bioinformatic Approaches. In C., B. and H., H., editors, *Legionella: Methods in Molecular Biology (Methods and Protocols)*, volume 954, pages 595–602.

Seladi-Schulman, J., Steel, J., and Lowen, A. C. (2013). Spherical influenza viruses have a fitness advantage in embryonated eggs, while filament-producing strains are selected in vivo. *Journal of virology*, 87(24):13343–53.

Selva, L., Viana, D., Regev-Yochay, G., Trzcinski, K., Corpa, J. M., Lasa, I., Novick, R. P., and Penadés, J. R. (2009). Killing niche competitors by remote-control bacteriophage induction. *Proceedings of the National Academy of Sciences of the United States of America*, 106(4):1234–1238.

Sergueev, K. V., He, Y., Borschel, R. H., Nikolich, M. P., and Filippov, A. A. (2010). Rapid and sensitive detection of Yersinia pestis using amplification of plague diagnostic bacteriophages monitored by real-time PCR. *PloS one*, 5(6):e11337.

Serrano-Suárez, A., Dellundé, J., Salvadó, H., Cervero-Aragó, S., Méndez, J., Canals, O., Blanco, S., Arcas, A., and Araujo, R. (2013). Microbial and physicochemical parameters associated with Legionella contamination in hot water recirculation systems. *Environmental science and pollution research international*, 20(8):5534–44.

Service, P. (2007). On an invisible microbe antagonistic toward dysenteric bacilli: brief note by Mr. F. D'Herelle, presented by Mr. Roux. *Research in Microbiology*, 158(7):553–554.

Shapiro, J. W. and Putonti, C. (2018). Gene Co-occurrence Networks Reflect Bacteriophage Ecology and Evolution. *mBio*, 9(2):1–14.

Shleeva, M. O., Bagramyan, K., Telkov, M. V., Mukamolova, G. V., Young, M., Kell, D. B., and Kaprelyants, A. S. (2002). Formation and resuscitation of "non-culturable" cells of Rhodococcus rhodochrous and Mycobacterium tuberculosis in prolonged stationary phase. *Microbiology (Reading, England)*, 148(Pt 5):1581–91.

Sikora, A., Gładysz, I., Kozioł-Montewka, M., Wójtowicz-Bobin, M., Stańczak, T., Matuszewska, R., and Krogulska, B. (2017). Assessment of antibiotic susceptibility of Legionella pneumophila isolated from water systems in Poland. *Annals of agricultural and environmental medicine : AAEM*, 24(1):66–69.

Silveira, T. N. and Zamboni, D. S. (2010). Pore formation triggered by Legionella spp. is an Nlrc4 inflammasome-dependent host cell response that precedes pyroptosis. *Infection and immunity*, 78(3):1403–13.

Simmons, L. A., Foti, J. J., Cohen, S. E., and Walker, G. C. (2008). The SOS Regulatory Network. *EcoSal Plus*, 2008(9):1385–1395.

Steinert, M., Emödy, L., Amann, R., and Hacker, J. (1997). Resuscitation of viable but nonculturable Legionella pneumophila Philadelphia JR32 by Acanthamoeba castellanii. *Applied and environmental microbiology*, 63(5):2047–53.

Steinert, M., Hentschel, U., and Hacker, J. (2002). Legionella pneumophila: an aquatic microbe goes astray. *FEMS microbiology reviews*, 26(2):149–162.

Stewart, C. R., Rossier, O., and Cianciotto, N. P. (2009). Surface translocation by Legionella pneumophila: a form of sliding motility that is dependent upon type II protein secretion. *Journal of bacteriology*, 191(5):1537–46.

Storey, M. V., Winiecka-Krusnell, J., Ashbolt, N. J., and Stenström, T.-a. (2004). The efficacy of heat and chlorine treatment against thermotolerant Acanthamoebae and Legionellae. *Scandinavian journal of infectious diseases*, 36(9):656–62.

Sulakvelidze, A., Alavidze, Z., and Morris, J. G. (2001). Bacteriophage Therapy. *Antimicrobial Agents and Chemotherapy*, 45(3):649–659.

Sullivan, M. J., Petty, N. K., and Beatson, S. A. (2011). Easyfig: a genome comparison visualizer. *Bioinformatics (Oxford, England)*, 27(7):1009–10.

Summers, W. C. (1999). *Félix D'Herelle and the Origins of Molecular Biology*. Yale University Press.

Summers, W. C. (2011). In the beginning... *Bacteriophage*, 1(1):50–51.

Suttle, C. A. (1994). The significance of viruses to mortality in aquatic microbial communities. *Microbial ecology*, 28(2):237–43.

Swanson, M. S. and Isberg, R. R. (1995). Association of Legionella pneumophila with the macrophage endoplasmic reticulum. *Infection and immunity*, 63(9):3609–20.

Swartz, M. N. (1979). Clinical aspects of Legionnaires' disease. *Annals of internal medicine*, 90(4):492–5.

Swift, B. M. C., Denton, E. J., Mahendran, S. A., Huxley, J. N., and Rees, C. E. D. (2013). Development of a rapid phage-based method for the detection of viable Mycobacterium avium subsp. paratuberculosis in blood within 48 h. *Journal of microbiological methods*, 94(3):175–9.

Tang, F., Bossers, A., Harders, F., Lu, C., and Smith, H. (2013). Comparative genomic analysis of twelve Streptococcus suis (pro)phages. *Genomics*, 101(6):336–344.

Taylor, M., Ross, K., and Bentham, R. (2009). Legionella, Protozoa, and Biofilms: Interactions Within Complex Microbial Systems. *Microbial Ecology*, 58(3):538–547.

Thurman, K. A., Warner, A. K., Cowart, K. C., Benitez, A. J., and Winchell, J. M. (2011). Detection of Mycoplasma pneumoniae, Chlamydia pneumoniae, and Legionella spp. in clinical specimens using a single-tube multiplex real-time PCR assay. *Diagnostic microbiology and infectious disease*, 70(1):1–9.

Tilton, R. C. (1979). Legionnaires' disease antigen detected by enzyme-linked immunosorbent assay. *Annals of internal medicine*, 90(4):697–8.

Tossa, P., Deloge-Abarkan, M., Zmirou-Navier, D., Hartemann, P., and Mathieu, L. (2006). Pontiac fever: An operational definition for epidemiological studies. *BMC Public Health*, 6:1–10.

Touchon, M., Bernheim, A., and Rocha, E. P. (2016). Genetic and life-history traits associated with the distribution of prophages in bacteria. *The ISME journal*, 10(11):2744–2754.

Toze, S., Sly, L. I., MacRae, I. C., and Fuerst, J. A. (1990). Inhibition of growth ofLegionella species by heterotrophic plate count bacteria isolated from chlorinated drinking water. *Current Microbiology*, 21(2):139–143.

Tronel, H. and Hartemann, P. (2009). Overview of diagnostic and detection methods for legionellosis and Legionella spp. *Letters in applied microbiology*, 48(6):653–6.

Tropp, B. (2012). Viruses in Molecular Biology. In *Principles of Molecular Biology*, chapter 7, pages 223–263. Jones & Bartlett Learning.

Tu, A. H., Voelker, L. L., Shen, X., and Dybvig, K. (2001). Complete nucleotide sequence of the mycoplasma virus P1 genome. *Plasmid*, 45(2):122–6.

Twest, R. V. and Kropinski, A. M. (2009). Bacteriophage Enrichment from Water and Soil. 501:15–21.

Twort, F. (1915). AN INVESTIGATION ON THE NATURE OF ULTRA-MICROSCOPIC VIRUSES. *The Lancet*, 186(4814):1241–1243.

Tyndall, R. L. and Domingue, E. L. (1982). Cocultivation of Legionella pneumophila and free-living amoebae. *Applied and environmental microbiology*, 44(4):954–9.

van Duin, J. and Olsthoorn, R. (2012). Leviviridae. In King, A., Adams, M., Carstens, E., and Lefkowitz, E., editors, *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*, pages 1035–1043. Elsevier Science, San Diego.

Viazis, S., Akhtar, M., Feirtag, J., Brabban, A. D., and Diez-Gonzalez, F. (2011). Isolation and characterization of lytic bacteriophages against enterohaemorrhagic Escherichia coli. *Journal of applied microbiology*, 110(5):1323–31.

Vickers, R. M., Brown, A., and Garrity, G. M. (1981). Dye-containing buffered charcoal-yeast extract medium for differentiation of members of the family Legionellaceae. *Journal of clinical microbiology*, 13(2):380–2.

Vidaver, A. K., Koski, R. K., and Van Etten, J. L. (1973). Bacteriophage phi6: a Lipid-Containing Virus of Pseudomonas phaseolicola. *Journal of virology*, 11(5):799–805.

Wadowsky, R. M., Butler, L. J., Cook, M. K., Verma, S. M., Paul, M. A., Fields, B. S., Keleti, G., Sykora, J. L., and Yee, R. B. (1988). Growth-supporting activity for Legionella pneumophila in tap water cultures and implication of hartmannellid amoebae as growth factors. *Applied and environmental microbiology*, 54(11):2677–82.

Wadowsky, R. M. and Yee, R. B. (1981). Glycine-containing selective medium for isolation of Legionel-laceae from environmental specimens. *Applied and environmental microbiology*, 42(5):768–72.

Wang, T. C. and Smith, K. C. (1988). Different effects of recJ and recN mutations on the postreplication repair of UV-damaged DNA in Escherichia coli K-12. *Journal of bacteriology*, 170(6):2555–9.

Wang, X., Kim, Y., Ma, Q., Hong, S. H., Pokusaeva, K., Sturino, J. M., and Wood, T. K. (2010). Cryptic prophages help bacteria cope with adverse environments. *Nature communications*, 1(9):147.

Watanabe, K., Suzuki, H., Nakao, R., Shimizu, T., and Watarai, M. (2015). Draft Genome Sequences of Five Legionella pneumophila Strains Isolated from Environmental Water Samples. *Genome announcements*, 3(3):1–2.

Watnick, P. and Kolter, R. (2000). Biofilm, City of Microbes. *Journal of Bacteriology*, 182(10):2675–2679.

Webb, J. S., Thompson, L. S., James, S., Charlton, T., Tolker-Nielsen, T., Koch, B., Givskov, M., and Kjelleberg, S. (2003). Cell death in Pseudomonas aeruginosa biofilm development. *Journal of bacteriology*, 185(15):4585–92.

Wechsler, J. A. and Gross, J. D. (1971). Escherichia coli mutants temperature-sensitive for DNA synthesis. *MGG Molecular & General Genetics*, 113(3):273–284.

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

Wever, P. C., Yzerman, E. P., Kuijper, E. J., Speelman, P., and Dankert, J. (2000). Rapid diagnosis of Legionnaires' disease using an immunochromatographic assay for Legionella pneumophila serogroup 1 antigen in urine during an outbreak in the Netherlands. *Journal of clinical microbiology*, 38(7):2738–9.

Weynberg, K. D. (2018). Viruses in Marine Ecosystems: From Open Waters to Coral Reefs. In *Advances in Virus Research*, volume 101, pages 1–38. Elsevier Inc., 1 edition.

Whiley, H. and Taylor, M. (2016). Legionella detection by culture and qPCR: Comparing apples and oranges. *Critical reviews in microbiology*, 42(1):65–74.

WHO (2015). *Water Safety in Distribution Systems*. World Health Organization.

Wiegand, I., Hilpert, K., and Hancock, R. E. W. (2008). Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nature Protocols*, 3(2):163–175.

Wilhelm, S. W. and Suttle, C. a. (1999). Viruses and Nutrient Cycles in the Sea. *BioScience*, 49(10):781–788.

Wilson, R. E., Hill, R. L. R., Chalker, V. J., Mentasti, M., and Ready, D. (2018). Antibiotic susceptibility of Legionella pneumophila strains isolated in England and Wales 2007-17. *The Journal of antimicrobial chemotherapy*, 73(10):2757–2761.

Winn, W. C. (1988). Legionnaires disease: historical perspective. *Clinical microbiology reviews*, 1(1):60–81.

World Health England (2018). Legionellosis.

Wright, A., Hawkins, C. H., Anggård, E. E., and Harper, D. R. (2009). A controlled clinical trial of a therapeutic bacteriophage preparation in chronic otitis due to antibiotic-resistant Pseudomonas aeruginosa; a preliminary report of efficacy. *Clinical otolaryngology : official journal of ENT-UK ; official journal of Netherlands Society for Oto-Rhino-Laryngology & Cervico-Facial Surgery*, 34(4):349–57.

Wright, M. N., Gola, D., and Ziegler, A. (2017). Preprocessing and Quality Control for Whole-Genome Sequences from the Illumina HiSeq X Platform. In Elston, R. C., editor, *Statistical Human Genetics: Methods and Protocols Methods in Molecular Biology*, volume 1666, chapter 30, pages 629–647. Springer New York, 2, illustr edition.

Yee, R. B. and Wadowsky, R. M. (1982). Multiplication of Legionella pneumophila in unsterilized tap water. *Applied and environmental microbiology*, 43(6):1330–4.

Yong, S. F. Y., Goh, F.-N., and Ngeow, Y. F. (2010). Legionella species and serogroups in Malaysian water cooling towers: identification by latex agglutination and PCR-DNA sequencing of isolates. *Journal of water and health*, 8(1):92–100.

Yordpratum, U., Tattawasart, U., Wongratanacheewin, S., and Sermswan, R. W. (2011). Novel lytic bacteriophages from soil that lyse Burkholderia pseudomallei. *FEMS microbiology letters*, 314(1):81–8.

Yu, V. L., Plouffe, J. F., Pastoris, M. C., Stout, J. E., Schousboe, M., Widmer, A., Summersgill, J., File, T., Heath, C. M., Paterson, D. L., and Chereshsky, A. (2002). Distribution of Legionella species and serogroups isolated by culture in patients with sporadic community-acquired legionellosis: an international collaborative survey. *The Journal of infectious diseases*, 186(1):127–8.

Zeng, L., Skinner, S. O., Zong, C., Sippy, J., Feiss, M., and Golding, I. (2010). Decision Making at a Subcellular Level Determines the Outcome of Bacteriophage Infection. *Cell*, 141(4):682–691.

Žgur-Bertok, D. (2013). DNA damage repair and bacterial pathogens. *PLoS pathogens*, 9(11):e1003711.

Zhang, Y. and Hu, Z. (2013). Combined treatment of Pseudomonas aeruginosa biofilms with bacteriophages and chlorine. *Biotechnology and bioengineering*, 110(1):286–95.

Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J., and Wishart, D. S. (2011). PHAST: a fast phage search tool. *Nucleic acids research*, 39(Web Server issue):W347–52.

Zschach, H., Joensen, K. G., Lindhard, B., Lund, O., Goderdzishvili, M., Chkonia, I., Jgenti, G., Kvatadze, N., Alavidze, Z., Kutter, E. M., Hasman, H., and Larsen, M. V. (2015). What can we learn from a metagenomic analysis of a georgian bacteriophage cocktail? *Viruses*, 7(12):6570–6589.