Physiological and genomic characterization of six virulent bacteriophages of Shiga toxin-producing *Escherichia coli* O157:H7 for biocontrol and detection applications

by

Beatriz Iara Cabral e Pacheco

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Chemical Engineering

Department of Chemical and Materials Engineering

University of Alberta

Abstract

Despite multiple control strategies in all steps of the food production chain, outbreaks of foodborne pathogens are common worldwide. More effective ways to prevent, eliminate or at least detect their presence before products reach consumers are of great interest to the general public, the food industry, and public health agencies. A promising approach is the use of bacteriophages (phages), viruses that specifically infect bacteria, in the biocontrol and detection of these pathogens. *Escherichia coli* O157:H7 is an important bacterial pathogen commonly associated with the contamination of vegetables, meat and other animal products. Healthy cattle are the primary reservoir of *E. coli* O157:H7 and consequently of its predators, phages. In the present work, six *E. coli* O157:H7 phages previously isolated from commercial feedlots in Southern Alberta were characterized in terms of morphology, host range and lytic capability, adsorption kinetics, and virulence dynamics. The six phages belong to the *Siphoviridae* family. The screening of 30 different *E. coli* O157:H7 strains against all phages revealed that at least 22 were susceptible to all six phages, with 14 being extremely sensitive. Adsorption kinetics results combined with two different reaction models (single-step and sequential adsorption) indicated adsorption occurred in two steps: an initial reversible binding followed by an irreversible one. Moreover, the phage genomes were sequenced and analyzed. The six phages shared high genomic similarity between themselves and with other viruses isolated from the same trial but at different geographical locations and collected at different sampling times. We thus hypothesize these phages represent different variants of the dominant *E. coli* O157:H7 phage in their ecological niche. Furthermore, although the genomes of five of the six phages shared 99.9% pairwise similarity, their host range and lytic capability, adsorption

kinetics, and infection dynamics differed. Twelve nonsynonymous point mutations differentiated the genetic codes of these phages and, notably, six of these mutations were located in genes encoding putative tail fibers, responsible for host recognition and binding. Thus, these point mutations are likely causes of differences in the phage phenotypes observed, and would be represent interesting locations for phage genetic engineering to tailor host specificity. The present work sets a framework for the identification and selection of phages with potential for biocontrol and/or detection of *E. coli* O157:H7, provides valuable insights on phage host relationship mechanisms, and suggests genetic basis for traits of interest.

Acknowledgements

Someone wise once said it takes a village to raise a child. A wise friend told me it also takes a village to get someone through grad school. This is my simple thank you that does not do my village justice.

First, I'd like to thank my supervisors Dr. Dominic Sauvageau, Dr. Anastasia Elias, and Dr. Yan Niu, for their generosity and patience with me, and for being true role models by showing me where hard work and passion can lead you.

I am grateful to Dr. Tim McAllister, Dr. Kim Stanford and their groups for welcoming me in their labs, where most of the experiments of this thesis were conducted.

I'd like to thank Arlene Oatway from the Biological Sciences Microscopy Unit for her amazing help with imaging, and Matthew Walker and the Public Health Agency of Canada for the phages sequencing.

Many thanks to my friends from the lab and office, Nadia, Mark, Kieran, Ana, Diana, and Fabini, for all the fruitful discussions, long lunches, coffee breaks, beers, and bets that made my routine brighter and lighter. Thank you, Melissa, for always having lab tips, words of encouragement, and warm hugs to keep me going. Thank you, Miranda, for all the anxiety interventions and the instant friendship.

To my dancer friends from Swing-Out Edmonton and Miss Behavin', especially Anna, Jillian, and Chelsea, for welcoming me into their hobbies, groups, and hearts. I learned so much from you.

Thank you, my Canadian family, Camila, Zé Roberto, Lara, Vitor, and Cadoo for being my home away from home; and my friends around the world for always being just one call or message away, reminding me I am never alone.

À minha incrível família, por sempre acreditar em mim mais do que eu acredito em mim mesma. Aos meus exemplares pais Joice e Orlando por sempre estarem ao meu lado, me apoiando em todas as minhas decisões, ou discordando com e por amor. Ao meu irmão Hugo por ser meu amigo mais incondicional e sincero. Esse título é de vocês.

It indeed takes a village and I am deeply grateful for having found such a supportive one. I could not have done it without you.

Table of contents

List of Tables

List of Figures

1. Introduction

As a major cause of acute illnesses and mortality worldwide, foodborne pathogens are a constant risk for consumers and a challenge to the food industry and public health agencies. Even though multiple hygiene and control measures are undertaken throughout the production chain, outbreaks associated with contaminated food products remain common.

Shiga toxin-producing *Escherichia coli* O157:H7 (STEC O157:H7) is an important class of foodborne bacterial pathogens. Outbreaks of STEC O157:H7 are most commonly associated with animal products such as meat, or vegetable products that had contact with contaminated water, such as leafy greens and fruits. Infections with this bacteria can be asymptomatic, cause mild to severe gastrointestinal illness, or evolve to extraintestinal complications such as renal failure (Besser, Griffin and Slutsker, 1999; Briones, 2013). STEC O157:H7 is estimated to cause 73,000 illnesses, 2,100 hospitalizations and 60 deaths yearly in the U.S. alone (Rangel et al., 2005). In the first half of 2018, 210 infections by STEC O157:H7 were reported across the U.S. due to the consumption of contaminated romaine lettuce, with 96 patients requiring hospitalization and five deaths (CDC, 2018a). In Alberta, 119 cases of STEC O157:H7 linked to pork contamination were reported in 2014 (Honish *et al.*, 2017) and, in 2012, an outbreak at the XL Foods beef processing plant led to 18 cases of illness but over $27 million in losses to the industry. This indicates the urgent need of more efficient tools to control or at least detect this pathogen prior to the consumption of contaminated products.

STEC O157:H7 is a common part of the microbiota of several animals. In fact, healthy cattle are the primary reservoir of STEC O157:H7 in nature, spreading the bacteria through their feces (Besser, Griffin and Slutsker, 1999) – animals in commercial feedlots in Southern Alberta have been found to shed up to 10^8 CFU (colony forming units) of this bacteria per g of manure (Munns *et al.*, 2016). Consequently, cattle are also the reservoir the predator of this bacterium: bacteriophages (phages), viruses that infect and kill STEC O157:H7. Their capacity to lyse bacteria, along with their specificity towards their host, make phages an interesting platform for the control and detection of bacterial pathogens.

For food safety purposes, phage mixtures (often referred to as cocktails) are commercially available for use in ready-to-eat products; with approval for use in Canada, the USA, and the European Union. These products are so far mainly targeting common pathogens such as Listeria and Salmonella. In addition, many technologies have harnessed the highly specialized and specific recognition mechanisms of phages to detect bacterial pathogens (reviewed in Mahony *et al.*, 2011; Janjarasskul and Suppakul, 2018).

Technological advances in genetic sequencing propelled genetic characterization of phages and have made obvious we have just tackled the surface of their diversity – in fact, phages have even been described as the "dark matter of the biosphere" (Hatfull, 2015). For example, it is estimated that 50% of the phage genes available in public databases have no assigned function (Klumpp, Fouts and Sozhamannan, 2013). The better understanding of phage structures and genomes, combined with progress in synthetic biology and genetic engineering in the past decade made tailoring phages to specific applications a rapidly growing field of investigation.

The present study is based on the hypothesis that genetic comparison of phages sampled from a single environmental niche can provide a basis for the understanding of factors involved in virulence and host recognition, and can serve as the basis for the selection of phages for specific applications.

In order to achieve this goal, six STEC O157:H7 phages previously isolated from commercial feedlots in Southern Alberta were characterized with respect to their physiology, genetic composition, host range, infectivity, and adsorption. This approach assesses traits important to phage applications and provides insights into their interactions with host cells. Sequencing of the phage genomes revealed interesting evolutionary relationships with other phages, both isolated from feedlot trials in Alberta and from distant geographical locations. Bioinformatics analysis highlighted mutations that could be responsible for the different performance of the phage candidates. This work provides a framework for the identification and selection of STEC O157:H7 phages for further applications and engineering of these viruses.

2. Literature Review

In this section, relevant background information for the current work is discussed. The bacterial pathogen *Escherichia coli* O157:H7 and the relevance of using bacteriophages to target it are presented.

2.1 *Escherichia coli* O157:H7

2.1.1 Physiology and ecology

Escherichia coli O157:H7 is a gram-negative bacillus first identified as a human pathogen in 1982, due to an outbreak of hemorrhagic colitis in Oregon and Michigan linked to contaminated and undercooked beef patties. *E. coli* O157:H7 can cause a range of human infections, which can be asymptomatic, cause mild to severe bloody diarrhea, or evolve to extraintestinal problems such as hemolytic uremic syndrome (HUS). HUS is characterized by acute renal failure, low platelet counts, and hemolytic anemia (Besser, Griffin and Slutsker, 1999; Briones, 2013).

The main virulence factor of some strains of *E. coli* O157:H7 is the production of Shiga toxins (stx), a family of proteins toxic to eukaryotic cells (Besser, Griffin and Slutsker, 1999). Hence, *E. coli* O157:H7 producing this toxin are referred to as a STEC – Shiga toxin-producing *E. coli.* Shiga toxins were so named for one of its types, stx1, being virtually identical to the toxin produced by *Shigella dysenteriae* type 1. The genes that encode stx were most likely transferred from *Shigella* to *E. coli* by phage-mediated horizontal gene transfer (Herold, Karch and Schmidt, 2004). Even though not all stx genes are phage-encoded, most of the characterized ones are present in bacterial genomes encoded in sequences of functional or defective prophages - most commonly of lambdoid morphology (Besser, Griffin and Slutsker, 1999; Herold, Karch and Schmidt, 2004).

The primary transmission mode of *E. coli* O157:H7 in humans is through the consumption of contaminated food and water, although direct transmission from person to person and through occupational exposure have also been reported (Besser, Griffin and Slutsker, 1999). Meat products and vegetables are most commonly associated with outbreaks, with ground beef, romaine lettuce, sprouts, roast beef, raw milk, salami, and

unpasteurized apple juice being the most common vectors (Besser, Griffin and Slutsker, 1999).

STEC O157:H7 is a normal component of the rumen microbiota, being spread to the environment through the animal's feces, and, in fact, healthy cattle are the primary reservoir of the bacteria (Besser, Griffin and Slutsker, 1999). The abundance of STEC O157:H7 varies significantly from animal to animal: cattle that shed over 10^4 colony forming units (CFU) of the organism per gram of feces are considered super-shedders (Matthews *et al.*, 2006), and concentrations up to 10^9 CFU/g of feces were reported in animals from feedlots in Southern Alberta (Stephens, McAllister and Stanford, 2009; Munns *et al.*, 2016). Other factors affecting STEC O157:H7 are geographical location, season, and method and scope of sampling (Bach *et al.*, 2002).

2.1.2 Epidemiology

The largest *Escherichia coli* outbreak ever registered occurred in Germany from May to July 2011, when 3,816 cases were reported, with 845 involving HUS and 54 deaths (Frank *et al.*, 2011). The strain associated with this outbreak was a highly pathogenic enteroaggregative *E. coli* O154:H4, which additionally carried Shiga toxin genes. Moreover, the strain was resistant to some classes of antibiotics, such as beta-lactams and cephalosporins (Frank *et al.*, 2011). The origin of the outbreak was never defined, but sprouts were identified as the most likely contamination vehicle (Frank *et al.*, 2011). Merabishvili *et al*. (2012) isolated, in only three days, three candidate phages that could have been used as prophylactic therapy agents for the bacteria associated with this outbreak (Merabishvili *et al.*, 2012).

In Alberta, Canada, 119 cases of STEC O157:H7 infections were reported between July and October 2014. Twenty-three patients were hospitalized, and six developed HUS. No deaths were registered. This outbreak was linked to the consumption of pork products, and generated significant costs with hospitalizations, product recalls, destruction of suspected products, temporary food facilities closures, and loss of consumer trust in food processors, among others (Honish *et al.*, 2017).

From March to June 2018, 42 cases of *E. coli* O157:H7 infections were registered in the Edmonton area, Alberta. Of these, 14 patients needed hospitalization

and one died. Twenty-one of these cases were linked to a single restaurant, and pork sausages were identified as the primary vehicle of infection (Alberta Health Services, 2018).

Recently, in the United States, 210 peoples were infected over 36 states with *E. coli* O157:H7 after consuming romaine lettuce from April to June 2018 (CDC, 2018a). Ninety-six patients were hospitalized, 27 developed HUS, and five deaths were reported. The U.S. Centers for Diseases Control and Prevention (CDC) identified the source of contamination as canal water in a growing region (CDC, 2018a). Eight cases were reported across Canada, associated with consumption of romaine lettuce contaminated with a bacterium showing high genetic identity, indicating a common source that was not identified (PHAC, 2018b).

An unrelated outbreak of *E. coli* O157:H7 in romaine lettuce was declared in November 2018, with 43 reported cases across 12 U.S. states, leading to 16 hospitalizations (CDC, 2018b). The U.S. Food and Drug Administration (FDA) demanded all romaine lettuce be labelled with its harvest location so consumers could avoid those grown in northern and central California, which were pointed as the contamination source (CDC, 2018b). Twenty-four related cases were reported across 4 Canadian provinces. The Canadian Food Inspection Agency (CFIA) recommended residents of Ontario, Quebec, and New Brunswick to avoid eating romaine lettuce and salad mixes containing romaine lettuce until further notice (PHAC, 2018a).

2.1.3 Mitigation strategies

Different pre-harvest strategies have been tested to minimize STEC O157:H7 contamination on cattle before entering the food chain. General animal husbandry and basic good production practices such as clean water and feeds (free from fecal contamination), properly drained environments, and reduced animal density decrease the bacterial load to which animals are expose to and, consequently, the risk of STEC O157:H7 contamination (Beef Industry Food Safety Council, 2013).

The effects of different diet components and feeding strategies – including fasting (Buchko *et al.*, 2000), hay and corn silage (Berry *et al.*, 2006), forage and grain diets (Baale *et al.*, 2004), and barley (Buchko *et al.*, 2000) – on the prevalence and fecal

shedding of *E. coli* O157:H7 have been studied. These studies on dietary shifts found conflicting results on the impact of feed on *E. coli* O157:H7 populations (Callaway *et al.*, 2009).

Vaccines based on different mechanisms were developed against STEC O157:H7. Econiche (Bioniche Life Sciences, Inc., Belleville, Ontario, Canada) is licensed for use in Canada and targets *E. coli* O157:H7 type III secreted proteins, reducing its ability to colonize animals (Allen *et al.*, 2011). This vaccine was shown to reduce (but not eliminate) rectal colonization, fecal shedding, and hide contamination of treated animals, although the results were affected by seasonality and the proportion of the herd vaccinated (Smith *et al.*, 2009).

Probiotic preparations have also been suggested as an alternative agent to reduce *E. coli* O157:H7 colonization in cattle by competing for the same ecological niche. Bovamine (Chr. Hansen, Hoersholm, Denmark), a *Lactobacillus*-based product used mainly for growth promotion in the U.S., was assessed as a control agent for pathogenic bacteria. However, a feedlot trial found no effect on the supplementation of the direct-fed microbial on *E. coli* O157:H7 shedding (Cull *et al.*, 2012).

2.2 Phages

Bacteriophages – or phages – are viruses that specifically infect bacteria. Like all other viruses, phages depend on their host for survival and replication, hindering the targeted host cell machinery for the production and release of new viral particles. Viruses are considered a gray area between living and non-living matter, as they are composed of genetic material coated with proteins lacking its metabolic functions (Delbrück, 1970; Villarreal, 2004).

Phages were first described by two independent studies just over a century ago (Salmond and Fineran, 2015). In 1915, British pathologist Frederick Twort described a “condition or disease” observed in micrococci. Twort extensively tried to grow the “self-destructive material” with no success, hypothesizing it was a virus of unknown nature and of smaller dimensions than the ones known at that time. He also attempted to isolate the “material” from diseased patients and environmental samples (Twort, 1915).

In 1917, French-Canadian microbiologist Félix D'Hérelle isolated what he then called an antagonistic microbe against *Shigella spp.* from patients with dysentery. He observed cultures of the pathogen crashing consistently, but when different bacteria were inoculated with the same "anti-microbe," they were not affected. D'Hérelle then coined the term bacteriophage (bacteria eater in Latin) and immediately recognized the therapeutic potential of his discovery, treating dysentery with his phage isolates (D'Hérelle, 2007). Phage therapy research gathered interest in the scientific world until the late 1930s when regulatory concerns and the emergence of antibiotics slowed it down (Salmond and Fineran, 2015).

Since their discovery, the study of phages has contributed to science immensely. Compared to traditional virology, which mostly focused on extracellular particles, early phage research focused mostly on host-virus interactions, thus enabling the clarification of core biological processes (Norrby, 2008; Salmond and Fineran, 2015). In 1952, Alfred Hershey and Martha Chase demonstrated that DNA was the bearer of genetic information by infecting bacteria with radiolabeled phages (Hershey and Chase, 1952). Phage research was also essential to the development of molecular and synthetic biology, providing tools such as restriction modifications and enzymes, early cloning vectors such as M13 and phage λ, DNA ligases and polymerases, and, more recently, the CRISPR-Cas system (Barrangou *et al.*, 2007; Salmond and Fineran, 2015).

2.2.1 Life cycles

Figure 2.1 shows the two predominant life cycles a phage can undergo. The first steps for both cycles involve the phage encountering, attaching and injecting its DNA into the host bacterium. During the lytic cycle, the host DNA gets disrupted immediately upon infection, and its cell machinery gets hindered to replicate phage DNA and express phage proteins and structures. New virions then get assembled and the host cell is lysed, liberating new phage particles to the environment. These new progeny can then meet new bacteria and reinitiate the cycle (Salmond and Fineran, 2015; Howard-Varona *et al.*, 2017).

On the other hand, during the lysogenic cycle, the phage DNA remains dormant inside the host, integrating itself to the host genome or remaining in the form of a

plasmid, until a trigger starts phage replication and the lytic cycle. During this lysogenic phase, stable phages inside the host are termed prophages or prophage regions. These regions are replicated with the host genome and are passed along as the bacterial cell divides (Salmond and Fineran, 2015; Howard-Varona *et al.*, 2017).

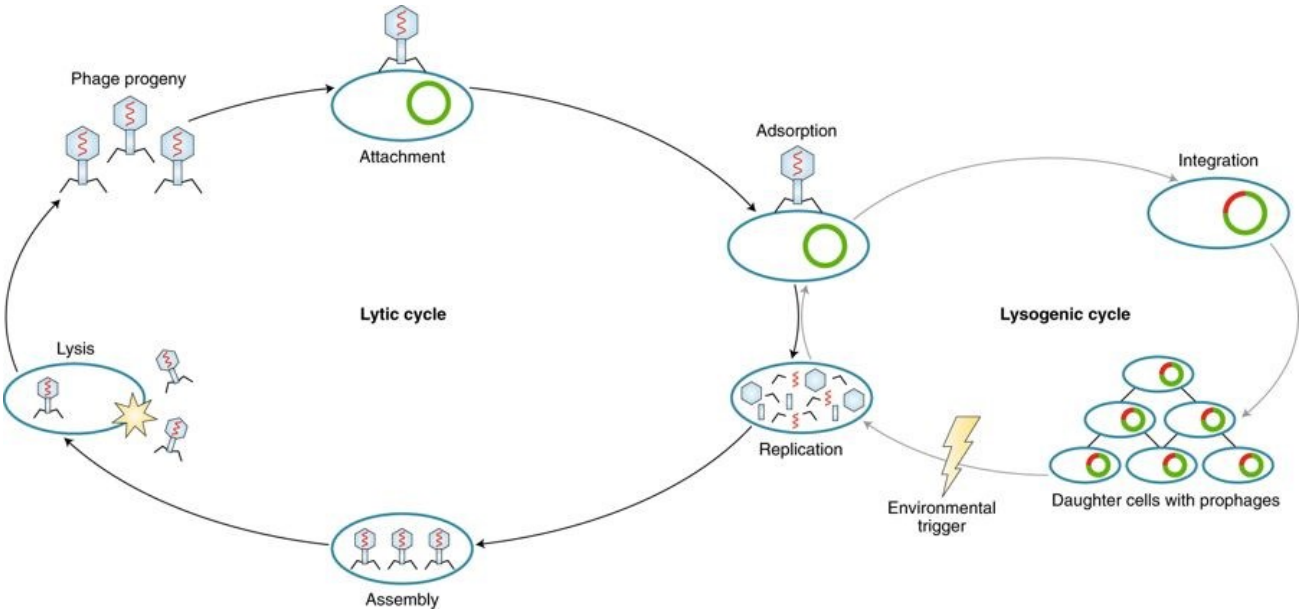


Figure 2.1 - Phages can generally undergo two different life cycles. The lytic cycle involves attachment of the phage to the bacteria, transfer of its genetic material, disruption of host cell genome, replication and assembly of new phages, disruption of host and release of new viral particles. During the lysogenic cycle, temperate phages remain in the host either integrated to their DNA or as a plasmid, multiplying with the cell until an environmental factor triggers the lytic cycle. Image from Breitbart et al. 2018.

The encounters of phages with bacterial hosts occur randomly, as phages do not have host-sensing machinery; hence the higher the number of phage particles per bacteria present in an environment, the higher the probability of phage attachment and, eventually, of infection. Experimentally, the ratio of phage added to bacteria is called multiplicity of infection (MOI). Thus, a higher MOI means more phage are present per bacteria, which increases the likelihood of a phage meeting its host (Salmond and Fineran, 2015; Nobrega *et al.*, 2018). MOI has had different definitions in literature – but most commonly it is the amount of phage added to bacterial cultures at high densities. However, for more complex systems at lower bacterial densities, MOI should not be used as an indicator of infection probability and, for phage therapy applications, the phage progeny should also be considered (Abedon, 2016).

Phages use a plethora of mechanisms and structures to target and attach to potentially sensitive hosts, and this interaction ultimately defines a phage's host range. Tailed phages commonly have receptor-binding proteins (RBP) at the end of their tails to recognize and attach to the receptors on the host cell surface. The diversity of RBPs and receptors pose a challenge to the study of phages, and many adsorption mechanisms remain poorly or not defined. Tail fibers, spikes and tips often attach to lipolysaccharides (LPS), teichoic acids, and porins (Bertozzi Silva, Storms and Sauvageau, 2016; Nobrega *et al.*, 2018).

Thus, bacteria susceptibility to phages is defined by specific structures on the cell wall, and this can be used to classify bacterial strains. A common and quick diagnostic method for pathogenic bacteria such as *E. coli* is phage typing – by infecting strains with a standardized set of phages, one can quickly assess a simple measurement of their similarities through the infection susceptibility patterns. The current phage typing set for *E. coli* O157:H7 is composed of 16 phages that have identified 62 bacteria phage types so far (Khakhria, Duck and Lior, 1990).

Initial phage interactions with the host, in which the viral particle can adsorb to and desorb from the host, are thought to be often reversible and followed by an irreversible binding. Once irreversible binding occurs, conformational changes take place to inject the phage DNA into the host. Some phages have membrane-penetrating proteins specifically for this function (Ackermann, 1998; Nobrega *et al.*, 2018).

The mechanism of DNA ejection of the phage capsid and injection into the host is poorly understood. Due to the condensed state of DNA packaged into the capsid, it was thought the resulting high osmotic pressure was sufficient to eject the DNA upon infection; but this model does not explain all modes of ejection. Presently, different models are hypothesized to explain the driving force of phage DNA ejection, including water flow and ion exchange (Molineux and Panja, 2013).

Strictly lytic phages will immediately start gene expression, replication and formation of phage genomes and structures. Transcription of phage genes is generally thought to be divided into three categories: early, middle, and late genes. Early genes code for the preparation of the host for the phage, the degradation of bacterial DNA and the overtake of its metabolic functions. Middle genes initiate phage DNA replication, and

late genes induce the synthesis of regulatory, structural, and lysis proteins, besides DNA packaging (Ackermann, 1998a).

Temperate phages, on the other hand, can go into the lysogenic cycle after DNA injection. In this case, their genetic material is integrated to the host genome to produce lysogens through the mediation of integrase enzymes, or stay in the host cytoplasm to enter a latency period. This vegetative state can be reverted spontaneously or caused by inducing agents. DNA damaging agents such as ultraviolet light, mytomicin C and high temperatures have been used to induce prophages (Ackermann, 1998). Upon infection, temperate phages have to make lysis-lysogeny decisions. The mechanisms behind it were extensively studied for phage lambda, where the decision is ultimately probabilistic, with influence from the number of prophages co-infecting the cell and the host nutritional state (Oppenheim *et al.*, 2005). It was recently discovered that phages of the SPbeta group infecting *Bacillus* bacteria release a small peptide to communicate with other phages. Upon further infections, the phage progeny assesses the concentration of the communication peptide and lysogenizes if the concentration is above a threshold, thus avoiding to kill all the host cells in the environment (Erez *et al.*, 2017). This quorum sensing strategy might be present in other phages and illustrates how new phage mechanisms remain to be discovered (Callaway, 2017).

Assembled phages can be released by lysing the host cell. This process can be enzyme-mediated by peptidoglycan hydrolases (endolysins) and holins. Endolysins attack the murein session of the cell wall, while holins damage the plasma membrane (Ackermann, 1998).

Evolution pushes both phage and bacteria to adapt and develop a multitude of mechanisms to either guarantee or avoid infection. Bacteria antiphage systems is a research area in rapid expansion – groundbreaking mechanisms have been discovered in the last decade, and many undoubtedly remain to be elucidated (Samson *et al.*, 2013).

Bacteria can prevent phage adsorption by blocking or altering their cell wall and membrane components; by producing extracellular polymeric matrixes, physically blocking phage receptors; or by producing competitive inhibitors specifically to phage receptors. Phages evolve to evade these strategies by changing their RBPs or

producing polymer-degrading enzymes, for example (Labrie, Samson and Moineau, 2010; Samson *et al.*, 2013).

Many bacteria also encode restriction-modification (R-M) systems whose primary function is to protect the cells from foreign DNA, including phages. Phage DNA entering bacterial cells can get recognized and promptly degraded by restriction enzymes. To circumvent this, phages have evolved antirestriction strategies such as accumulating point mutations on endonucleases recognition sites or using methylated nucleotides variants and unusual bases in their genome. Phage T4, for example, uses hydroxymethylcytosine (HMC) instead of cytosine in its genome to avoid R-M systems (Labrie, Samson and Moineau, 2010).

Another bacterial mode of defence against phages is the CRISPR-Cas system (clustered regularly interspaced short palindromic repeats and CRISPR-associated proteins) (Barrangou *et al.*, 2007). These proteins target and cleave invading DNA in a specific manner, by recognizing sequences that previously invaded cells and whose information is stored in the form of DNA spacers in the CRISPR loci of the bacterial genome. CRISPR-Cas is often referred to as the adaptive immune system of bacteria (Barrangou and Marraffini, 2014), and this highly specific yet adaptable endonuclease ability has become a powerful tool for molecular biology and genetic engineering (Jinek *et al.*, 2012; Hsu, Lander and Zhang, 2014). Phages can evade this system by mutating or deleting these recognizable sequences in their genomes; through anti-CRISPR genes that produce proteins able to neutralize CRISPR-Cas surveillances complexes; or even by encoding their own CRISPR-Cas systems targeting the bacterial antiviral proteins (Barrangou *et al.*, 2007; Labrie, Samson and Moineau, 2010; Samson *et al.*, 2013).

2.2.2 Classification and taxonomy

Phages are typically classified according to the characteristics of their free virions, their genome compositions, and their host ranges. Electron microscopy is widely used in phage research and constitutes the fastest diagnostic technique for virology; hence the field's historical reliance on phage morphology for classification. Phage micrograph collections date to as early as 1943, accounting for different morphotypes

and groups: Hans W. Ackermann, phage researcher and microscopy extraordinaire, documented over 5500 phage micrographs over his life (H. W. Ackermann 2007, H. W. Ackermann 2012).

The development of staining methods for isolated, purified viruses in 1955 (Hall, 1955) allowed researchers to collect extensively, image, and group phage samples. The principle of staining is to expose phages to an electron-dense solution of a salt of small molecular size and high molecular weight. Phages are said to be negatively stained when the surroundings uptake the stain, and viral particles appear white or gray in contrast (H.W. Ackermann 2003). The main stains used in phage research are phosphotungstates and uranyl acetate (H.W. Ackermann 2012; H.W. Ackermann 2003).

Bacterial and archaeal viruses are currently classified in 22 families, according to their capsid and tail morphology and the nature of their genetic material, as indicated in Figure 2.2. Guidelines for this classification are under the responsibility of the International Committee on Taxonomy of Viruses (ICTV) Bacterial and Archaeal Viruses Subcommittee (BAVS) (Adriaenssens and Brister, 2017).

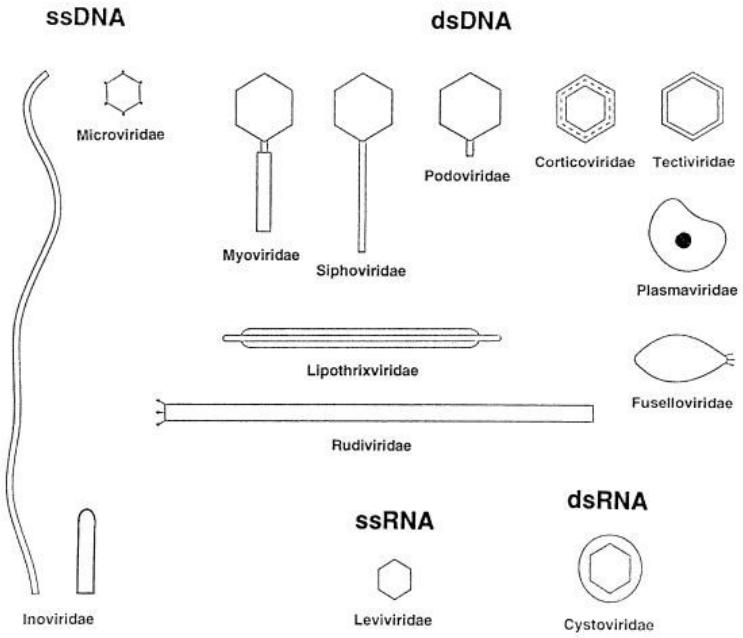


Figure 2.2 - Non-comprehensive phage families diagram. Image from (Ackermann 2003)

Tailed phages constitute the Caudovirales order, which is divided among four families (ICTV 2017a) and constitutes the largest of all virus groups. Its members are

easily recognizable through microscopy due to their tubular tails. Tailed phages able to infect a multitude of hosts of medical and industrial relevance have been isolated – enterobacteria, clostridia, lactobacillus, pseudomonas and even archaea and cyanobacteria (Ackermann, 1998; Salmond and Fineran, 2015). The Siphoviridae family are the most frequent group of the Caudovirales order (H. W. Ackermann 1998). The latest viral species list released by the ICTV shows 602 out of the 1023 tailed phages registered belong to this family, as represented in Figure 2.3 (ICTV 2017b).

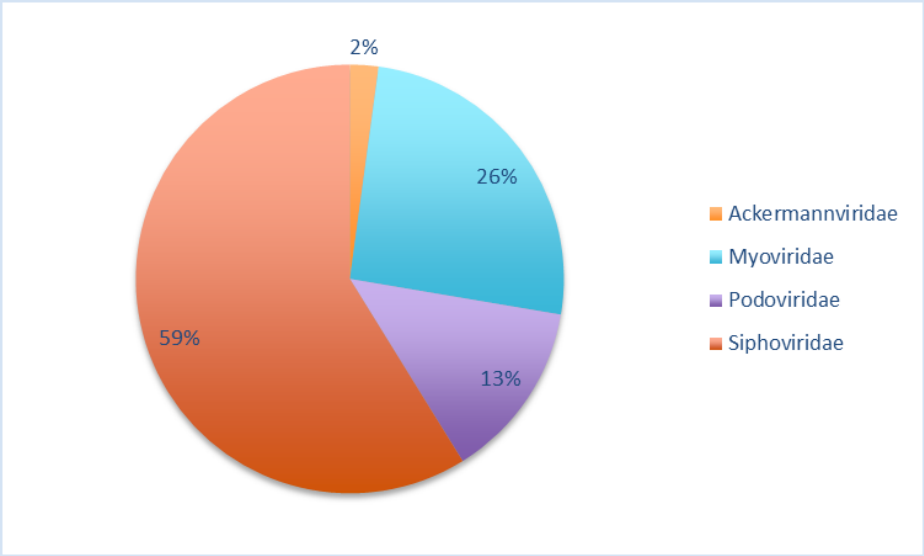


Figure 2.3 - Tailed phage families' distribution. Data from the International Committee on Taxonomy of Viruses Master Species List (ICTV 2017b)

Classification at lower taxa level - at the subfamily and genus rankings - is becoming more common as sequencing technologies evolve and phage genomic information is more widely available (Adriaenssens and Brister, 2017). This requires viral taxonomy to be continuously reevaluated and updated; for example, the BAVS votes yearly on proposals for new taxonomy groups (Adriaenssens *et al.*, 2018).

The lack of markers and conserved genes – such as 16S rRNA for bacteria and 18S rRNA for Protozoa –, the high genetic diversity - both in terms of genetic composition, with phages varying between circular and linear genomes, single or double-stranded genetic material, DNA or RNA, and genome variability within similar phages due to horizontal gene transfer and high mutation rates –, and the recent and

ongoing development of molecular techniques enabling phage sequencing and identification all make phage classification challenging. The hierarchical phage classification scheme established by the ICTV predates the omics era and thus heavily relies on morphology and host range (Nelson, 2004; Aiewsakun *et al.*, 2018).

Proposals to incorporate genomic and proteomic information to phage taxonomy are frequent. Currently, these data provide guidelines for lower taxa classification, at the order, genus and species level (Adriaenssens and Brister, 2017). For example, Rohwer and Edwards proposed classifying phages according to a proteome tree through BLASTP comparisons input in a distance tree (Rohwer and Edwards, 2002). This would categorize phages relative to both their close relatives and all other phages on the database. Proux and colleagues from the Nestle Research Center proposed a taxonomy system based on comparative genomics of structural genes, as these are relatively conserved within groups, and often provide interesting insight on phage evolution (Proux *et al.*, 2002; Chibani-Chennoufi *et al.*, 2004).

More recently, an international group of phage experts developed a pipeline to classify phages based on comparative genomics (Aiewsakun *et al.*, 2018). This GRAViTy platform (Genome Relationship Applied to Virus Taxonomy) computes similarity matrixes between higher groups and then within family-level taxa based on gene homology. The authors found several inconsistencies in the current ICTV classification, such as phages clustering closer to other groups than within their assigned family (Aiewsakun *et al.*, 2018).

Although problematic, phage hierarchical classification based on morphology can include phages that are not yet sequenced and relies on relatively quick and straightforward diagnostic methods. However, it does not allow for the classification of virus-like particles, such as lysogenic phage regions in bacteria not able to produce virions nor non-cultivated phage sequences obtained from metagenomics studies. The analysis of omics data also enables a more detailed classification of phages and predict biological features of phages (Nelson 2004; Rohwer and Edwards 2002; Aiewsakun et al. 2018).

2.2.3 Ecology and genetics

Viruses are the most abundant biological entities in the world, endemic to all environments where life is found and are estimated to outnumber bacteria 10-fold. Oceans have an average concentration of 10 million viral particles per millilitre of surface water (Wommack and Colwell, 2000), and thus there are more viruses in one litre of seawater than humans on the planet (10^{10} versus 7.6×10^{9}) (Breitbart *et al.*, 2018).

Before the development of more accurate and versatile techniques in the last two decades, most ecological studies of viral communities focused on aquatic environments, as water samples are simple to collect and quantification of viral particles can be done through the use of a microscope (Hatfull, 2015). The main barrier to assessing viral diversity in different environments was limited by the purifying methods available for viral particles, usually requiring enrichment with specific hosts (Reyes *et al.*, 2012). The advance of next-generation sequencing, computational methods, and metagenomics uncoupled genetic characterization from culture-based assays, allowing direct sequencing of environmental samples. Metagenomic surveys have shown substantial, and yet mostly uncharacterized, viral communities – called viromes – in microbial communities of ecosystems as diverse as ocean water (Coutinho *et al.*, 2017), plants (Brown *et al.*, 2018), Antarctic soil (Adriaenssens *et al.*, 2017), the human skin (Foulongne *et al.*, 2012) and gut (Reyes *et al.*, 2012), animal microbiota such as cow rumen (Berg Miller *et al.*, 2012), pig digestive tract (Shan *et al.*, 2011), and bat guano (Li *et al.*, 2010), among many others.

The main challenges to characterizing and identifying specific viruses from viral metagenomes include the lack of phylogenetic markers, laborious assembly due to highly permuted viral genomes, difficulties in assigning host specificity from genetic information, high number of viral genes of unknown function, and poorly characterized prophage regions in bacterial genomes (Reyes *et al.*, 2012). However, despite these difficulties, metagenomic studies have provided an impressive overview of the genetic diversity of viruses that is still vastly unknown – phages have been referred to as the dark matter of the biological world (Hatfull, 2015).

The mostly unexplored diversity and overwhelming prevalence of phages in all microbial ecosystems studied highlight their importance in, for example, animal health. However, the complexity of animal microbiota systems renders the study and modelling of the dynamic impact of viruses not feasible at this point. It remains difficult to understand how a single phage strain, let alone a group of strains, can impact the dynamics of complex bacterial communities. This being said, various mechanisms that lay the ground to investigate the roles of phages in microbiota dynamics have been proposed (Figure 2.4) (De Paepe *et al.*, 2014).

The “kill the winner” model is the most easily observed in laboratory settings and consists in the increased chance of infection by lytic phages and acceleration of their multiplication with increasing host density. Phages would thus control the number of the fastest growing members of the community. In the “kill the relative” model, spontaneous prophage induction of a lysogenized population can cause the death of strains that are close relatives of a bacterium. As prophages provide immunity to its carrier bacteria, the lysogenic bacterial community would have the advantage to dominate its niche. On the other hand, prophages can be detrimental to their carriers, as induction of the phage can lyse the host and cause dysbiosis, as per the “community shuffling model”. The “invade the relative” model refers to prophages invading close bacterial relatives of its original hosts, establishing new strains without lysing them (Mills *et al.*, 2013; De Paepe *et al.*, 2014).

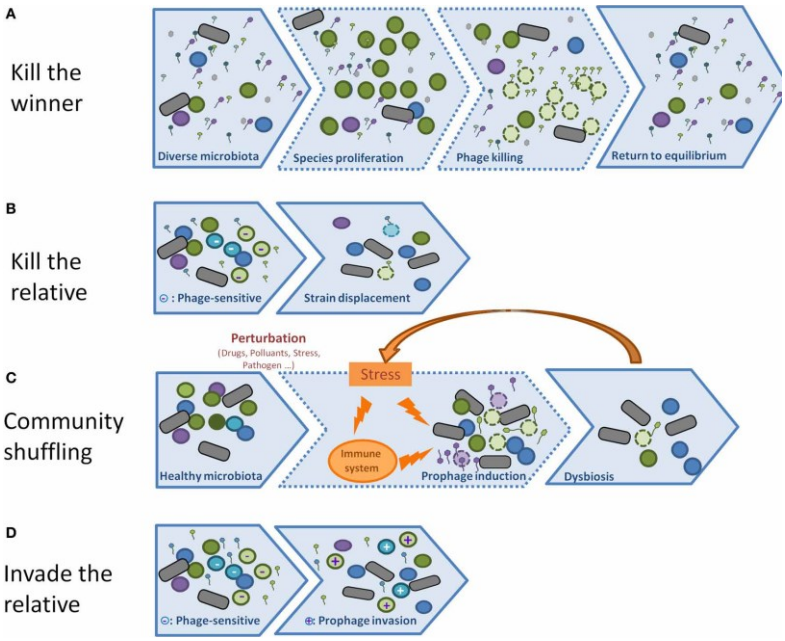


Figure 2.4 - Models of phage influence in microbial communities through predation and lysogenic induction. Figure from De Paepe et al. 2014.

Phages also play an essential role in bacterial evolution and adaptation to new environments through horizontal gene transfer (HGT), a process mediated by phages in which the genes from a bacterium are transferred to another bacterium. HGT is an important process for pathogenicity and strain fitness in microbial communities. Recently, metagenomics, single-cell, and genomics studies renewed interest in phage-mediated HGT events (Shterzer and Mizrahi, 2015; Touchon, Moura de Sousa and Rocha, 2017). The three main HGT mechanisms are lysogenic conversion, specialized transduction, and generalized transduction.

Lysogenic conversion refers to the phenotypic changes in the host due to the expression of traits from temperate phages. Observed influences of prophage accessory genes include motility, virulence, biofilm formation, and antibiotic tolerance. Acquiring these traits can be risky to the host, as induction of the phage can cause cell death. However, mutations can inactivate the prophage and reduce or eliminate its ability to kill the host cell (defective prophages). Defective phages are commonly observed in bacterial genomes and can correspond to significant portions of its make up – up to 13.5% of the genome of *E. coli* O157:H7 strain EC4115, and 16% of strain Sakai

(Hayashi *et al.*, 2001; Bobay, Touchon and Rocha, 2014; Touchon, Moura de Sousa and Rocha, 2017) originate from defective phages.

In generalized transduction, host DNA gets erroneously packed into the phage virion during prophage induction and, upon infection of other cells, this chromosomal DNA from the lysed cell is integrated to the genome of the new host by homologous recombination. The transferred genetic material can thus be from anywhere in the genome, as this process results from errors in DNA packaging. Specialized transduction, on the other hand, results from inaccurate excision of the induced prophage from the host's genome, hence packaging into the capsid and transferring specific regions (Touchon, Moura de Sousa and Rocha, 2017).

The amount of genetic material that can be packaged for transduction depends on the volume of the phage capsid, and can surpass the size of the original phage genome. Transduction has been historically considered a rare event (Touchon, Moura de Sousa and Rocha, 2017; Davison, 2018); however, a recent study combining genomics and transcriptomics techniques proposed a mechanism they called lateral transduction that transferred material between *Staphylococcus aureus* cells 1000 times more frequently than previously described transduction mechanisms (Chen *et al.*, 2018). This recent discovery illustrates how modern techniques can help elucidate phage-mediated gene transfer mechanisms that are most likely a regular part of phages' life cycle (Davison, 2018).

2.2.4 Phage genomics

Prior to rapid and affordable sequencing methods, DNA fingerprinting techniques were used to characterize phages. Most commonly, DNA samples are subjected to pulsed field gel electrophoresis (PFGE) to create a genetic fingerprint of a population. Unlike conventional electrophoresis methods, the electric field to which samples are subjected in PFGE is abruptly perturbed in order to resolve larger DNA sizes (Herschleb, Ananiev and Schwartz, 2007). Another common approach consists in digesting phage DNA with restriction enzymes before gel electrophoresis, allowing comparison of the patterns produced by different genomes. This is called restriction fragment length polymorphism (RFLP). Fingerprinting techniques are still used for quick

screening of phages prior to sequencing or to roughly assess the genetic similarity of phages (Hallewell *et al.*, 2014). Fingerprinting methods are also used for quick analyses and diagnoses of genotypes of bacteria – since 1996, the US CDC maintains the PulseNet database which contains different PFGE patterns of the foodborne pathogens *E. coli* O157:H7, *Salmonella*, *Shigella,* and *Listeria* (Swaminathan *et al.*, 2001).

The development of the Sanger sequencing method 40 years ago (Sanger, Nicklen and Coulson, 1977) revolutionized biology and was the main sequencing technique for the following 30 years of its publication (often referred to as first-generation sequencing) mainly in its automated form. However, shotgun sequencing principles – based on the sequencing of random DNA fragments followed by assembly of a consensus sequence based on sequence overlaps – and massive parallel sequencing allowed for the democratization of genomics. Second-generation sequencing (or next-generation sequencing – NGS) methods require less work force, space, and capital investment than Sanger sequencing (Shendure *et al.*, 2017). The main NGS platforms are Roche's 454 (no longer developed); Illumina's Solexa and NovaSeq; the Polonator by Dover; Applied Biosystem's SOLiD (no longer developed) and single molecule analysis by VisiGen and Helicos. Nowadays the Illumina platform is dominant (Mardis, 2008; Shendure *et al.*, 2017). It is important to note that these methods require the preparation of libraries through the amplification of sequences, leading to potential errors linked to the copy process.

More recently, third-generation sequencing technologies – PacBio and nanopore sequencing – have become commercially available. These do not rely on template amplification, hence are not subjected to copy errors and information loss (caused by, for example, methylation patterns). Third-generation methods can sequence single molecules in real-time but are still under development because of high error rates and lower throughputs than most NGS methods (Shendure *et al.*, 2017). These methods are promising for phage sequencing as they allow, at least theoretically, the analysis of a single phage instead of the average population as with other NGS platforms.

The first relatively large genome sequenced was that of phage λ, a model organism in molecular biology. It consisted of 48,502 bases and was sequenced through automated Sanger sequencing (Sanger *et al.*, 1982). Since then, the availability

of sequences has grown exponentially and, since 1982, the number of bases publicly available in GenBank doubles approximately every 18 months (Shendure et al. 2017; NCBI 2018). Following this trend, and propelled by the regained interested in phage research, the number of phage genomes compiled in the NCBI Genome database (National Center for Biotechnology Information, U.S. National Library of Medicine) more than doubled in the last three years (Philipson *et al.*, 2018).

Common challenges to phage sequencing include: the large number of short reads with context-specific errors that, due to their small genomes, become hard to correct; the presence of repetitive regions render phage genomes difficult to assemble with reliability. Summed to these are the everyday challenges of NGS, such as data storage, processing, and analysis. An attractive approach is to correct errors through deep sequencing – i.e. by achieving higher sequence coverages and consequently data reliability with sequences obtained through other NGS methods or circular consensus (Klumpp, Fouts and Sozhamannan, 2013; Philipson *et al.*, 2018). Another critical limitation to phage genomics is the large pool of phage genes of unknown function. Phages have high diversity and mutation rates, and it is estimated that more than 50% of the currently predicted phage gene products in databases are hypothetical (Klumpp, Fouts and Sozhamannan, 2013). The development of bioinformatics tools to predict protein functions and efforts in comparative genomics are used to tackle this problem. However, as there is no unified phage annotation and comparison pipeline, this process is labour intensive (Klumpp, Fouts and Sozhamannan, 2013).

Often, the quality of genomes found in databases is tied to the intended application. For example, high-quality genome standards are found for phages considered “safe” or “promising” for phage therapy or vaccine development. This means that all open reading frames (ORFs) have been identified, 100% of the genome is found in a consensus sequence with deep coverage, the phage lifestyle has been elucidated, and purity indicators such as lack of population diversity have been defined (Koonin and Galperin, 2003; Philipson *et al.*, 2018).

2.2.5 Applications

The use of phage preparations to treat bacterial infections (phage therapy) was exploited since phages were discovered – D'Hérelle's laboratory in Paris sold at least five different preparations, marketed by a company that would become L'Oréal (Kutateladze and Adamia, 2008). However, clinical use and applied research on phages were virtually abandoned in Western countries due to the discovery of antibiotics and mixed results of phage therapy – used even to treat non-bacterial infections. After World War II, countries behind the Iron Curtain continued to rely on phage therapy. Nowadays, phage mixtures are commercialized over the counter in Georgia, Poland, and Russia. The Eliava Institute in Tbilisi, Georgia, has an extensive collection of phages that can be used to treat different infections - this costs at least €3,900 and up to €7,000 for severe cases (Kutateladze and Adamia, 2008; Kingwell, 2015; Eliava Phage Therapy Center, 2017).

With the emergence and spread of multi-drug resistant bacteria due to the abusive use of antibiotics, there is regained interest in phage therapy. One of the main challenges to the progress of phage therapy is the lack of robust clinical trials – the main recent ones were the treatment of children with acute diarrhea in Bangladesh (Sarker *et al.*, 2016), funded by Nestlé, and the PhagoBurn project that treated burn patients in France and Belgium, funded by the European Union (EU) (Kingwell, 2015; Jault *et al.*, 2018). Both projects showed limited efficiency of phage therapy and were discontinued due to the lack of observed efficacy in patients. This failure is mainly attributed to mistakes in assessing the bacterial pathogens and, consequently, the phages for the trials and dosages.

Another challenge to the development of phage therapy is regulation: phage preparations for human use are classified as medicinal products in the USA and as drugs in the EU. Thus, each phage candidate is subjected to the same requirements as other drugs in terms of aseptic manufacturing, clinical trials, and quality assurance. This fails to comprise particularities of phage applications such as the need to mix phages in so-called phage cocktails to circumvent bacterial resistance and broaden host range – according to the EU, and US regulation, each phage of the cocktail has to be assessed individually as an active compound. Thus, the current regulation is pointed out by

researchers as too rigid and unnecessarily costly for phage therapy (Furfaro, Payne and Chang, 2018; Pirnay *et al.*, 2018).

Other antimicrobial applications of phages are shown in Figure 2.5.

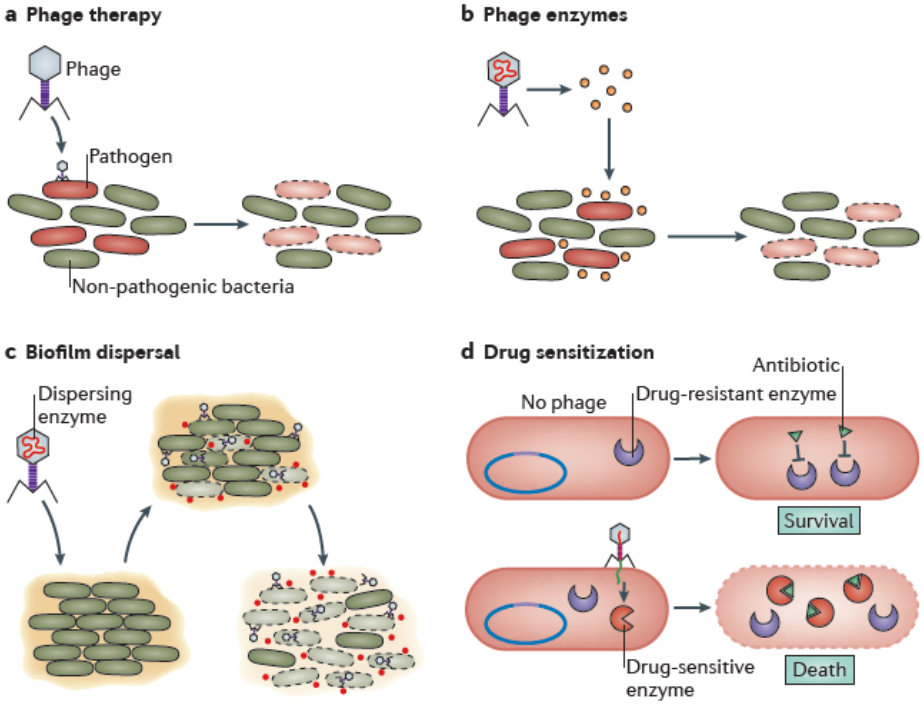


Figure 2.5 - Phage antimicrobial applications. Phage therapy (a) and phage-encoded enzymes (b) are used to target and lyse specific bacterial infections. Phages can be used to disperse biofilms (c) or to introduce antibiotic-sensitive genes into resistant bacteria (d). Figure from (Salmond and Fineran, 2015).

Phages have been used to target and sensitize resistant bacteria. Researchers have, for example, used phages as drug delivery platforms by conjugating antibiotics to phages to improve drug delivery (Edgar *et al.*, 2012), or used phages as gene delivery vehicles to inject specific dominant genes to confer antibiotic sensitivity (Edgar *et al.*, 2012), to suppress bacterial stress responses (Lu and Collins, 2009), and to inject CRISPR-Cas systems targeting resistance-conferring plasmid (Yosef *et al.*, 2015).

Phage-encoded endolysins and virion-associated peptidoglycan hydrolases involved in the lysis and adsorption steps of the phage life cycle, respectively, degrade the peptidoglycan layer of bacterial cell walls and can cause lysis due to osmotic imbalance. These enzymes can be used as alternatives to integral phages as antimicrobials. Advantages of phage enzymes over phage therapy include broader host spectrum, no potential transmission of virulence factors, and less likely development of

bacterial resistance (Roach and Donovan, 2015; Rodríguez-Rubio *et al.*, 2015). These cell-wall-degrading enzymes were successfully expressed in common chassis and demonstrated biocontrol activity against several bacteria relevant to public health. Examples of this include prevention of nasal and skin colonization by Staphylococcal biofilms (Fenton *et al.*, 2013), and treatment of bacteremia due to *Streptococcus pyogenes* in mice (Lood *et al.*, 2014).

The antimicrobial capacity of phage enzymes was also demonstrated with food pathogens, both for detection and sanitization of abiotic surfaces and food matrixes. A ΦH5 endolysin was able to control and lyse bovine and human isolates of *S. aureus* in pasteurized milk (Lood *et al.*, 2014). Solanki and colleagues immobilized an antilisterial phage enzyme in silica nanoparticles incorporated into a thin polymer film. The endolysin was able to effectively kill the surrogate strain *Listeria innocua* in buffer and lettuce (Solanki *et al.*, 2013).

Detection systems based on domains from phage endolysins can consist on binding the enzymes to a separable substrate to collect the pathogen and further detect it with molecular methods. This approach was used to separate *Listeria* cells from different food samples using functionalized magnetic beads with the cell wall binding domains from phage endolysins (Kretzer *et al.*, 2007). After separation, the beads were enriched and plated. This method was further developed to eliminate the enrichment step and detect *Listeria* cells in raw milk via real-time PCR (Walcher *et al.*, 2010). The merging of cell binding domains from phage endolysins to fluorescent proteins was also used to detect and classify different strains of *Listeria* in mixed cultures, milk, and cheese (Schmelcher *et al.*, 2010). However, biochemical incompatibilities (pH, the presence of proteolytic agents, ionic strength, and so forth), inefficiency against Gram-negative bacteria, and mass transfer limitations are all challenges to the use of these enzymes in biocontrol of pathogen in food matrixes (Lood *et al.*, 2014).

Phage-based biosanitation products to combat food pathogens are commercially available and approved for use in Canada, the USA, and the European Union. Advantages of using whole phage preparations include their specificity, self-replicating and self-limiting (as long as there is a host threshold present, phages will multiply), relatively low cost to isolate and propagate, and prolonged shelf life (Sillankorva,

Oliveira and Azeredo, 2012). The first product approved for this use was Listex (Micreos, Wageningen, The Netherlands), against *Listeria*. Nowadays the same company also offers products and cocktails for *Salmonella* and *E. coli* biocontrol for food safety purposes. Intralytix (Baltimore, USA) markets phage-based products targeting *Listeria monocytogenes, E. coli* O157:H7, *Salmonella,* and *Shigella* strains for use in ready-to-eat products, processing facilities, pet food, and animal hides.

Phage-based biosensors for food safety relying on all steps of the infection cycle have been developed, such as detection of phage-induced lysis (Blasco *et al.*, 1998), of phage amplification (Stewart *et al.*, 1998), and expression of reporter genes(Loessner *et al.*, 1996). Advantages of this last strategy are the fast and sensitive detection of only viable pathogen cells. Besides, phages can act as signal amplifiers as infection proceeds. Sample6 Detect (Sample6, Boston, U.S.A.) is a commercially available pathogen detection system based on the expression of bioluminescent proteins upon phage infection, allowing single cell in-plant detection (Schmelcher and Loessner, 2014).

2.2.6 Bovine and *E. coli* O157:H7 phages

As phages are found in all environments where their host bacteria are present, the rumen with its dense and diverse microbial populations is clearly a rich environment for viruses. Early studies of bacterial and archaeal viruses in ruminants date from the mid-1960s, when phages were first suggested to be ordinary members of the rumen population, rather than transient ones (Hoogenraad *et al.*, 1967; Gilbert and Klieve, 2015). Until recent years, ruminal phage studies were limited to morphological and general characterization. Several groups of phages were isolated from rumen contents, indicating their diversity and richness (Klieve and Bauchop, 1988).

In the last decade, technological advances have made genetic characterization and metagenomics studies of the rumen virome possible. Besides the general challenges to virome studies, ruminal phage populations also undergo great variation with feeding time and type, and several phage genes are annotated as of bacterial origin (Gilbert and Klieve, 2015). To avoid this and assess phage diversity in the rumen, Miller and colleagues looked into CRISPR loci in 14 rumen bacteria reference genomes

and 15 metagenomic sets of rumen data. The authors identified up to 28,000 different viral genotypes, of which 78% did not match any phage genes in databases, indicating the unknown range of ruminal phages. The authors also identified a predominance of prophages over lytic phages, and of phages that predate the most abundant phyla in the rumen (Berg Miller *et al.*, 2012).

Phages of *E. coli* O157:H7 have been isolated from feedlot cattle, water feeds and manure slurs (Niu *et al.*, 2009b), swine samples (Morita *et al.*, 2002; Lee *et al.*, 2016), and industrial settings such as cucumber fermentation (Lu and Breidt, 2015), among others. These phages have been assessed for biocontrol and as detection agents of *E. coli* O157:H7, and their genomes provide important insights into phage ecology and applications.

Morita and collaborators isolated and characterized the lytic coliphage PP01, which can selectively lyse O157:H7 strains but not other O-serogroups (Morita *et al.*, 2002). The authors were able to change the host range of the standard coliphage T2 to infect *E. coli* O157:H7 by cloning the distal proteins of PP01's tail fibre into its genome (Yoichi *et al.*, 2005). The same group also developed a biosensor with the same phage, by expressing green fluorescent protein attached to the phage capsid, allowing visualization of *E. coli* O157:H7 through fluorescence microscopy (Oda *et al.*, 2004).

Niu and collaborators analyzed the prevalence of *E. coli* O157:H7 and its phages in feedlots of Southern Alberta (Niu *et al.*, 2009b). Phages were recovered from all types of samples (pooled fecal pats, water from troughs, manure slurry from pens floors, and rectal fecal samples) on ten pens monitored over seven months. The authors demonstrated a higher prevalence of phages in slurry samples, and higher prevalence of *E. coli* O157:H7 in rectal fecal samples, indicating the relevance of the sampling strategy for phage and host assessments. Moreover, this study produced a library of 255 *E. coli* O157:H7 phage isolates. Further characterization of some of these isolates revealed a great diversity of highly lytic candidate phages for biocontrol (Niu *et al.*, 2012), and even a new subfamily (Tunavirinae) for T1-like viruses based on the clustering of genetic data of newly discovered phages (Niu *et al.*, 2014).

While conducting a phage therapy experiment against *E. coli* O157:H7 in feedlot animals in Canada, Kropinski et al. observed that animals would consistently shed a

different phage than those inoculated, which indicates the importance of evaluating endemic phages and their ecological niches in phage applications (Kropinski *et al.*, 2012). The characterization of this phage, named Rogue1, revealed a striking genetic resemblance to JK06, isolated independently in Israel. Since then, similar phages were also isolated in Mexico (Amarillas, Cháidez-Quiroz, *et al.*, 2013), suggesting an evolutionary relationship between these widely distributed phages (Niu *et al.*, 2014).

2.2.7 Phage engineering

The growing number of phage and bacterial genomes available in public databases, a better understanding of phage structures and interactions with their hosts, and the advance of molecular and synthetic biology techniques have opened the door to the genetic engineering of phages. Tailoring phages to specific applications is a young but fast-growing field of study. Genetic engineered phages can broaden and overcome limitations existing in many phage applications. Some examples of potential applications of phage engineering include the reduction of toxicity of phage cocktails to human hosts for phage therapy, the inhibition of resistance mechanisms in targeted bacteria, and the introduction of reporter genes for pathogen detection. Deleting genes or inserting heterologous genes of interest is also important to elucidate their functions. Hence, genetic engineering of phages can be extremely valuable to both applied and fundamental sciences (Pires et al. 2016; Salmond and Fineran 2015; Costa et al. 2018).

Homologous recombination is the most common and well-established phage genetic engineering method. It is based on the natural ability of prokaryotic and cells to recombine heterologous genetic material into their own DNA by recognizing similar regions in two sequences. To alter a phage's genome via homologous recombination, the desired modification must be flanked with standard sequences of the phage genome. It is then inserted into a plasmid that is loaded into the bacterial host. Upon phage infection and replication, the modification is recombined into a small fraction of the phage progeny (Pires *et al.*, 2016; Costa *et al.*, 2018).

Recombination methods have been successfully utilized to add genes of interest into phages for over 20 years – the expression of reporter genes such as the bioluminescent protein luciferase upon phage infection to specific hosts, for example,

allows the detection of pathogenic bacteria in complex food matrixes (Loessner *et al.*, 1996; Loessner, Rudolf and Scherer, 1997). However, low recombination frequencies and lack of efficient screening methods to recover successful recombinants make this technique time and labour-intensive (Pires *et al.*, 2016).

To overcome the lower efficiencies of methods based on homologous recombination, Lu and colleagues developed a technique using yeast as an intermediate vector for genetic engineering of phages (Lu *et al.*, 2012). By using *Saccharomyces cerevisiae*, one can also take advantage of its efficient recombination machinery and clone the artificial yeast chromosome replication vector (YAC) backbones and phage DNA directly into yeast cells. The phage DNA is replicated in the intermediate vector, then isolated and transformed into the bacterial host, starting the phage lytic cycle. This technique was used, for example, to express fully synthetic phage øX174, both wild-type and engineered genome versions (Jaschke *et al.*, 2012). The YAC method was also used to alter phages' host range by swapping tail fibres genes – Ando and colleagues were able to modify coliphages T7 and T3 to target *Yersinia* and *Klebsiella* bacteria, and conversely alter a *Klebsiella* phage to infect *E. coli* (Ando *et al.*, 2015). While promising and enabling, the efficiency of the YAC method is limited by the bacterial host transformation capacity, which renders the method difficult for phages with genomes larger than 70 kbp (Pires *et al.*, 2016).

A recently described method uses a similar approach but circumvents bacterial host transformation capacities by assembling phage synthetic DNA *in vitro* using standard molecular biology methods, then introducing this synthetic DNA to cell wall deficient bacteria (L-form bacteria). The proof of concept experiments using *Listeria monocytogenes* recovered virulent virions not only from native and rebooted naked phage DNA, but also virions of *Bacillus* and *Staphylococcus* phages. As the bacterial host does not have a cell wall, and only pure DNA is inserted to reboot the phage, there is no adsorption step involved in the phage replication process (Kilcher *et al.*, 2018). This breaks the host specificity barrier to engineer phages and allows broader applications to phages of bacterial families with limited molecular and synthetic biology tools available, for example.

The development and democratization of DNA sequencing and genomics techniques significantly increased our knowledge of phage basic mechanisms and diversity. With further development of molecular and synthetic biology methods, engineered phages can increase the range of applications of these viruses in medicine, industry, and agriculture.

3. Methods

3.1 Phages, bacteria and media

The six phages used in this study were previously isolated from commercial feedlots in southern Alberta (Niu *et al.*, 2009b). Phages AKS26, AKS41, AKS46, AKSC2, and AKS66 originated from different pens collected at different sampling times at the same feedlot. Phage AHS184 was isolated from a different feedlot.

Bacterial cultures of *Escherichia coli* O157:H7 strain R508 previously isolated from cattle (Niu *et al.*, 2009b) were grown overnight at 37°C and used as the standard host for bacteriophage amplification and titration. Strains used for host range and virulence assays were previously isolated and characterized from cattle and human clinical isolates (Niu et al. 2009b). Table 3.1 lists all strains used in this project.

The media used in this study were Tryptic Soy Broth (TSB) for experiments in liquid phase and Tryptic Soy Agar (TSA) for plate assays. TSB medium was prepared by adding Bacto TSB preparation (Becton Dickinson, New Jersey, USA) to deionized water to a concentration of 30 g/L, as per the supplier's indications. TSB medium was then sterilized by autoclave (121 ºC, 20 psig) for a minimum of 15 min. TSA plates were either purchased premade (Dalynn Biologicals, Calgary, Canada) or prepared by adding 15 g/L of Agar A (Becton Dickinson, New Jersey, USA) to the TSB broth preparation. Phage stocks were plated in Modified Nutrient Agar (MNA) (Dalynn Biologicals, Calgary, Canada) composed of: 20 g/L Nutrient agar; NaCl 8.5 g/L; Agar 10.0 g/L; $CaCl_2$ 8.3 mg/L; $FeCl_3$ 1.1mg/L; $MgSO_4$. $7H_20$ 0.5g/L.

Table 3.1 - Host strains and phage types

E. coli O157:H7 strain	Phage Type
R508	14
EC19990299	1
EC19990300	2
EC19990295	4
EC19990298	14
EC19990293	21
EC19990296	23
EC19990294	24
EC19990297	32
EC19990301	31
EC19990302	33
EC20110005	8
EC20110006	10
EC20110007	14a
EC20110008	28
EC20110009	34
EC20110010	38
EC20110011	45
EC20110012	46
EC20110013	47
EC20110014	48
EC20110015	49
EC20110016	50
EC20110017	51
EC20110018	54
EC20110019	63
EC20110020	67
EC20110021	68
EC20110022	80
EC20110023	74
EC20110024	88

3.2 Phage propagations

Phage master stocks were obtained from phage eluates previously obtained from cow fecal matter and feedlot puddles by picking and diluting single clear plaques from *E. coli* O157:H7 strain R508 plates three sequential times (Niu *et al.*, 2009b). These isolates were diluted ten times in lambda diluent (8 mM $MgSO_4 \cdot 7H_2O$ 10 mM Tris-HCl, pH 7.5), centrifuge at 11,000 xg for 10 min in an Eppendorf 5424R microcentrifuge (Eppendorf, Mississauga, ON, Canada), and then filtrating the supernatant through 0.2 µm syringe filters (Corning Inc, New York, USA) to remove bacteria and/or agar residues.

These phage filtrates were tittered according to the method described in the section below and used to infect mid-log liquid bacterial cultures of *E. coli* O157:H7 strain R508 at an OD (optical density – absorbance of samples at 600 nm) of 0.45-0.55, corresponding to approximately 10^8 CFU/mL. Mid-log cultures were obtained from incubating single colonies of host stocks TSA plates no longer than 3 days old overnight, and then adding 1 ml of overnight culture to 9 ml TSB and incubate for 1-1.5 h at 37°C and 170 rpm, until the desired OD was reached.

For initial infections, the ratio between addition of inoculates (i.e. host-phage mixture) and final propagating volume was kept at 0.01~0.1. For master stocks, phage initial infections were carried out at a total volume of 1.6 mL mTSB (TSB with 10 mM $MgSO_4$) at a MOI of 0.001 (1 mL of host mid-log culture, 100 µL diluted phage filtrate at ~ 10^5 PFU (plaque forming units) per mL, and 500 µL mTSB) for 15 min at 37 °C. After that, the host-phage mixture was transferred to 18.4 mL mTSB in 50 mL Falcon tubes and incubated for 5 h at 37 °C and 190 rpm. Cultures were then centrifuged at 6,000 xg for 10 min at 4°C in a Sorvall Legend X1 centrifuge (Thermo Fischer Scientific, Hampton, NH, USA) and filtrated through 0.22 µm Steriflip centrifuge tubes filtration systems (MilliporeSigma,Burlington, MA, USA). Samples were titrated as indicated on session 1.3 below and stored at 4°C.

Phage working stocks were propagated similarly, using master stocks to infect host cells. For phage laboratory stocks – used on all the experiments listed below - a higher final volume was generated. To do so, initial infections were carried out in 50 mL Falcon tubes containing 6 mL mid-log host prepared as before, 0.1 mL phage working

stock ~ 10^7 PFU/mL, and 0.5 mL mTSB. After 10 min at 37 °C and 130 rpm, inoculates were transferred to sterile 500 mL Erlenmeyer flasks containing 192.5 mL mTSB. Propagation mixture was incubated for 5 h at 37 °C and 190 rpm, harvested, and centrifuged at 5250 xg for 30 min at 4°C in a Sorvall Legend X1 centrifuge (Thermo Fischer Scientific, Hampton, NH, USA). Supernatants were filtrated in 0.20 µm Nalgene® vacuum filtration systems (MilliporeSigma,Burlington, MA, USA) and stored at 4°C.

3.3 Determination of phage titer

Titers of phage filtrate lysates were determined in duplicates via the soft agar overlay technique (Sambrook and Russell, 2001), where a bacterial host lawn is grown in a top half-strength agar layer and infected with phage dilutions. After overnight growth, individual clearing lawn spots (plaques) are counted for phage enumeration.

Phage samples of unknown concentration were serially diluted in lambda diluent or TSB medium. A lawn of stardard host was obtained by mixing 100 µL of *E. coli* O157:H7 grown overnight with 3 mL of molten overlay agar – 3 g/L UltraPure Agarose (Thermo Fischer Scientific, Hampton, NH, USA), 1.23 g/L $MgSO_4 \cdot 7H_2O$; or soft TSA (30 g/L TSB preparation, 7.5 g/L Agar A) – and pouring this mixture over MNA or TSA plates.

If a single phage dilution was plated, 100 µL of the diluted phage were mixed with the molten agar and host mixture. If multiple dilutions were casted over the same plate, 10 µL spots of each dilution were pipetted over the overlay and let to air dry close to a flame. Plates were incubated at 37°C for 18 – 24 h. Phage individual plaques were counted and the titers were calculated as the average of plaque count number times the reciprocal of the dilution factor. Titers are reported in plaque forming units per mL (PFU/mL).

3.4 Phage Morphology

The following steps were taken to prepare samples of phage lysates for Transmission Electron Microscopy (TEM). To reduce the presence of bacterial debris, 1 mL of filtered (0.2 µm syringe-filter, Corning, New York, USA) phage lysates (~10^{11}

PFU/mL) were centrifuged at 20,000 × g at 4 °C for one hour in an Eppendorf 5424R microcentrifuge (Eppendorf, Mississauga, ON, Canada). The supernatant was discarded and pellets were resuspended in 1 mL of lambda diluent overnight at 4 °C. Samples were then washed a second time by centrifuging at the same conditions and resuspended in 100 µL of lambda diluent.

10 µL of washed samples were deposited on copper grids carbon-coated with Formvar films, stained with 4% uranyl acetate (pH 4.5) for 4 min. Samples were imaged using a Morgagni 268 (FEI, Hillsboro, Oregon, USA) transmission electron microscope (TEM) with a beam intensity of 80 kV and 110,000 times magnification. Phage capsid and tail sizes were determined by measuring at least five bacteriophages per sample using the microscope image processing software. Statistical analysis of morphology results was conducted with Analysis of Variance (ANOVA), with Tukey test set at 95% confidence level by GraphPad Prism 5.04 software (La Jolla, CA, USA).

3.5 Genome size and DNA restriction pattern

Bacterial nucleic acids were removed from 30 mL of filtered phage lysates (~10^{11} PF/mL) via treatment with 10 µg/mL DNase1 (Sigma-Aldrich, Oakville, ON, Canada) and 20 µg/mL RNaseA (Sigma-Aldrich, Oakville, ON, Canada) for 1 h at room temperature and slow mixing. Samples were precipitated by adding NaCl (Sigma-Aldrich, Oakville, ON, Canada) and polyethylene glycol (PEG) 8000 (Sigma-Aldrich, Oakville, ON, Canada) to a final concentration of 1 M and 10% w/v, respectively, and slow mixing overnight at 4 °C. They were then centrifuged at 14 000 × g at 4 °C for 1 h in a Sorvall Legend X1 centrifuge (Thermo Fischer Scientific, Hampton, NH, USA). Pellets containing the concentrated phage DNA were resuspended in 3 mL of lambda diluent.

The genome size of the phages investigated was determined by pulsed-field gel electrophoresis (PFGE) as described in (Lingohr, Frost and Johnson, 2009). Briefly, 200 µL of concentrated phage DNA were treated with Proteinase K (Invitrogen, Burlington, ON, Canada), and embedded in 1% SeaKem Gold agarose (Lonza, Shawinigan, QC, Canada): 1% sodium dodecyl sulphate (SDS) (Sigma-Aldrich, Oakville, ON, Canada) plugs. Samples were subjected to electrophoresis in 0.5 × Tris-borate EDTA (Sigma-

Aldrich, Oakville, ON, Canada) buffer at 12 °C for 18 h using a Chef DR-III Mapper electrophoresis system (Bio-Rad, Mississauga, ON, Canada), with pulse times of 2.2–54.2 s, 6 V/cm. Molecular weight markers were prepared similarly using *Salmonella braenderup* H9812 digested by *Xba*I (Invitrogen, Burlington, ON, Canada). Results were analyzed using BioNumerics software (Applied Maths, Austin, TX, USA).

For the restriction fragment length polymorphism analysis (RFLP), plugs were prepared as previously described, then digested overnight with 50 units of restriction enzymes EcoRI and HindIII (NEB Biolabs, Beverly, MA, USA) overnight at 37 °C. Samples were submitted to electrophoresis for 5 h at 12°C, with pulse times of 1.0–42.2 s, 6 V/cm. 1 kb plus DNA Ladder (Thermo Fischer Scientific, Hampton, NH, USA) was used as molecular weight marker.

3.6 Phage adsorption kinetics

Phage adsorption experiments were performed as in (Storms *et al.*, 2010), with modifications. Mid-log bacterial host cultures (OD_{600} = 0.1) of *E. coli* O157:H7 R508 grown in TSB broth were diluted to about 10^8 CFU/mL in the same medium and infected with phage stocks (titer of 10^9 – 10^{11} PFU/mL diluted in TSB broth) at an MOI of 0.01 at room temperature. The amount of free phage remaining in the infection mixture was measured by filtering 1 mL of the mixture through a 0.2 µm SFCA Corning syringe filter (Corning Inc, Corning, NY) at different times over 10-min experiments. A fresh syringe filter was used for each sample. Concentrations of free phages were determined by the soft layer overlay technique (Section 1.3) and the fraction of remaining free phages was reported as a fraction of the initial phages added to the infection mixture.

3.6.1 Adsorption models

The phage adsorption experiments results were fitted to two different models using Matlab R2017b (Mathworks, Natick, Massachusetts, USA), using an ordinary differential equations solver coupled with a least-square non-linear curve fit to estimate rate constant values.

3.6.1.1 Model 1

The first mathematical model is known as sequential model and it describes phage adsorption to the host as a two-step process as shown in equation (1), where P is the free phage particle, H is the bacterial host, R is the reversible phage-host complex, and I the irreversible phage-host complex.

$$P + H \underset{k_2}{\overset{k_1}{\rightleftharpoons}} R \xrightarrow{k_3} I \quad (1)$$

The concentrations of each species over time can be determined through first-order reaction rate laws (equations 2-5):

$$\frac{dC_p}{dt} = -k_1 C_p C_H + k_2 C_r \quad (2)$$

$$\frac{dC_r}{dt} = k_1 C_p C_H - (k_2 + k_3) C_r \quad (3)$$

$$\frac{dC_i}{dt} = k_3 C_r \quad (4)$$

$$\frac{dC_h}{dt} = 0 \quad (5)$$

$$C_0 - C_p = C_r + C_i \quad (6)$$

Where C_p is the free phage concentration, C_h is the host concentration, C_r is reversible phage–host complex concentration, C_i is the irreversible phage–host complex concentration, C_0 is initial phage concentration, and k_1,k_2, and k_3 are rate constants.

3.6.1.2 Model 2

The second mathematical model is a simplification of the first and considers phage adsorption as a single step process. Equation 7 illustrates this model, where P is the free phage particle, H is the bacterial host, and I the irreversible phage-host complex.

$$P + H \xrightarrow{k} I \quad (7)$$

First order reaction rate laws for this process are:

$$\frac{dC_p}{dt} = -k_1 C_p C_h \quad (8)$$

$$\frac{dC_h}{dt} = 0 \quad (9)$$

Where C_p is the free phage concentration, C_h is the host concentration, and k is the adsorption rate constant.

3.7 Dynamics of infection and virulence

Virulence dynamics were evaluated through bacterial reduction curves for five *E. coli* O157:H7 host strains of different phage types (PT): EC19990293 (PT 21); EC19990295 (PT 4); EC19990296 (PT 23); EC19990298 (PT 14); EC19990299 (PT 1); EC19990300 (PT 2).

Overnight cultures of each host (1 mL) were used to inoculate 10 mL of fresh medium in a 50 mL Falcon tube, agitated at 250 rpm and incubated at 37 °C until the optical density at 600 nm reached 0.5 (corresponding to approximately 10^8 CFU/mL). Reduction curves were generated using 96-well plates in a total volume of 250 µL per well. Host cells were added to reach a cell concentration of 10^8 CFU/ml in each well. High-titer phage stocks (~10^{10} PFU/mL) were serially diluted in each row, from a concentration of 10^9 PFU/mL to 10^2 PFU/mL, yielding initial MOIs ranging from 1 to 10^{-7}. At least three columns were used per phage tested to serve as replicates. Wells containing medium-only (blank), standard host (*E. coli* O157:H7 R508) and phage (positive lysis control), and uninfected bacteria (negative lysis control) were included on every plate. Figure 3.1 shows a microplate layout used for this experiment. Optical density was measured every 30 min for a total of 5 h at 600 nm with a Bio-Tek Synergy HTX Gen 5110 (BioTek Instruments, Vermont, USA). Experiments were performed at 37 °C.

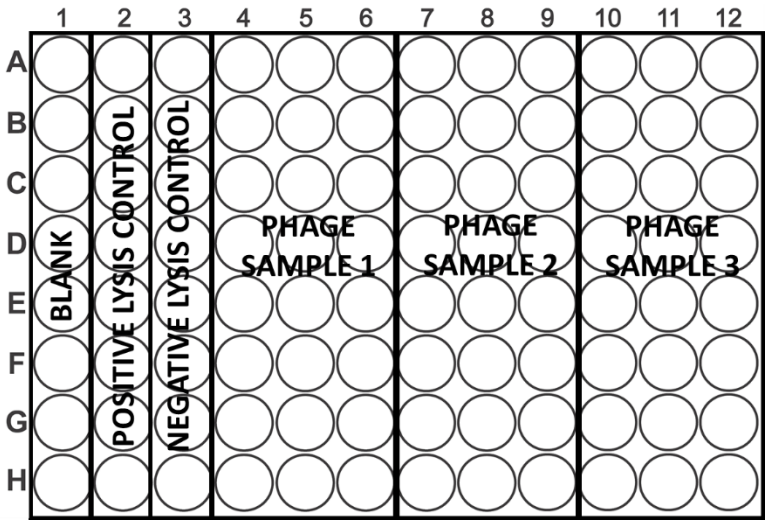


Figure 3.1 - 96-well plate layout for infection dynamics and virulence assays

Reduction curves were integrated to generate local and general Virulence Indexes as described by Storms *et al.* (2019).

3.8 Host range and lytic capability

The host range and lytic capability of the investigated phages against *E. coli* O157:H7 strains were assessed via a different microplate virulence assay (Niu *et al.*, 2009a). 30 previously phage typed STEC O157:H7 isolates from human and animal samples were used for this assay, as listed in Table 3.1.

In a 96-well plate, high-titer phage stocks (~10^{11} PFU/mL) were serially diluted in each row, from a concentration of 10^{9} PFU/mL to 10^{2} PFU/mL, yielding initial MOIs ranging from 1 to 10^{-7}. Overnight cultures of STEC O157:H7 were diluted to reach a cell concentration of 10^{8} CFU/ml in each well. At least 2 columns were used per phage to serve as replicates, as shown in Figure 3.2 below. Wells containing medium-only (blank), standard host (*E. coli* O157:H7 R508) and phage (positive lysis control), and uninfected bacteria (negative lysis control) were included on every plate.

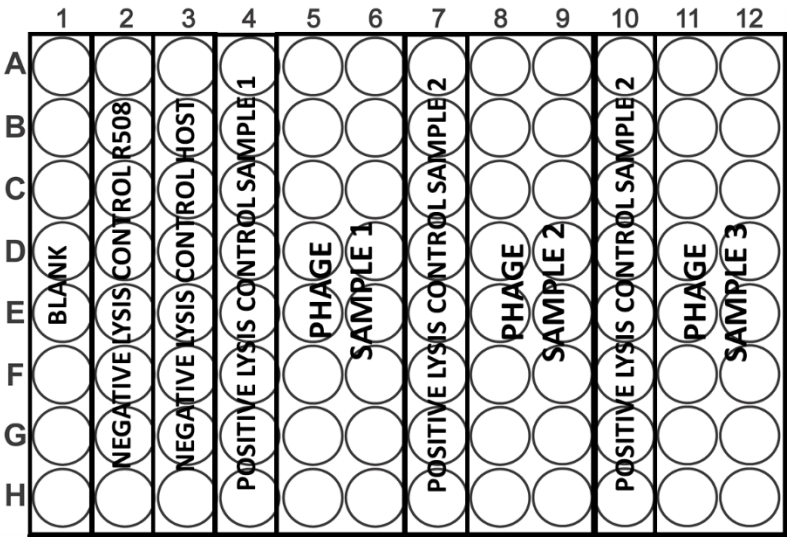


Figure 3.2 - 96-well plate layout for host range and lytic capability assays

Plates were incubated at 37 °C for 5 h. After the incubation period, wells were visually examined for turbidity and the highest phage dilution - i.e. lowest phage concentration - to cause full bacterial lysis (no discernable turbidity) was recorded.

The initial number of bacterial cells added to each well was calculated by plate counts of the serially diluted hosts. Results are expressed as the lowest initial MOI to cause full lysis.

3.9 Whole genome sequencing

3.9.1 DNA extraction

Phage genomic DNA was purified using a Phage DNA Isolation kit (Norgen Biotek, Thorold, ON, Canada), with modifications to the manufacturer's protocol. 2mL of filtered phage lysates (10^8 – 10^9 PFU/mL) were centrifuged at 8000 xg for 10 min at room temperature in an Eppendorf 5424R microcentrifuge (Eppendorf, Mississauga, ON, Canada). 1300 µl of the phage supernatant was treated with 1mg/mL of DNase 1 and 2 mg/mL of RNase A for 15 min at room temperature. Manufacturer's instructions were followed for loading and washing of samples, except on the elution step, when 50 µL of buffer was used over a single step.

The integrity and purity of the extracted DNA were analyzed by submitting undigested samples to gel electrophoresis (80 V, 2h running time) against a DNA ladder

λ DNA/HindIII marker (NEB Biolabs, Beverly, MA, USA); and by verifying samples' absorbance at 260 and 280 nm using a Nanodrop 2000 (Thermo Fischer Scientific, Hampton, NH, USA).

A PicoGreen Assay for double-stranded DNA kit (Thermo Fischer Scientific, Hampton, NH, USA) was used to determine the concentration of samples, following the manufacturer's instructions. A DNA standard curve was built using Lambda DNA (Invitrogen, Burlington, ON, Canada) diluted in TE buffer 1X (Thermo Fischer Scientific, Hampton, NH, USA) to concentrations ranging from 1 ng/mL to 1000 ng/mL. Samples were diluted in TE buffer 1X to a final volume of 125 µL. 125 µL of the PicoGreen dye diluted 1/200 in TE buffer 1X were pippeted into clear bottom black 96-well plates (Corning Inc, New York, USA) and mixed with 125 µL of DNA solution (sample or standard). Fluorescence was measured in a BioTek Synergy HTX Gen 5110 (BioTek Instruments, Vermont, USA) at an excitement wavelength of 485 nm and an emission wavelength of 580 nm.

3.9.2 Sequencing, annotation and comparative genomics

The genomes of the phages of interest were sequenced at the Public Health Agency of Canada (PHAC) National Microbiology Laboratory (Winnipeg, MB, Canada). Sample libraries were prepared using a MiSeq Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA, USA). Whole genome sequencing was performed by 150 bp paired-end read sequencing on an Illumina MiSeq sequencer using the MiSeq Micro Reagent Kit V2 and 300 cycles on the MiSeq platform (Illumina, San Diego, CA, USA) to obtain an average phage genome coverage of 70–300x. Data was assembled using SPAdes v3.10 (Bankevich *et al.*, 2012) followed by visual inspection.

Genomes were initially annotated automatically using the RAST server (Bankevich *et al.*, 2012). Predicted proteins were confirmed, when possible, and scanned for homologues via BLASTP (Altschul *et al.*, 1990, 1997). Transfer-RNA (tRNA) encoding sequences were predicted using ARAGORN (Laslett and Canback, 2004) and tRNAscan-SE (Lowe and Chan, 2016). Rho-independent terminators were screened with TransTermHP (Kingsford, Ayanbule and Salzberg, 2007). The Neural Network Promoter Prediction Tool (Reese, 2001) followed by visual inspection was

used to identify promoter regions. The TMHMM Server v.2.0 (Krogh *et al.*, 2001), Phobius (Kall, Krogh and Sonnhammer, 2007), and SPLIT 4.0 (Juretić, Zoranić and Zucić, 2002) were used to identify transmembrane protein domains.

Genomes were analyzed using Geneious R11 (Biomatters, Auckland, New Zealand) and aligned using the progressiveMauve feature (Darling, Mau and Perna, 2010). Specific protein-coding sequences were aligned using the Geneious alignment tool. Phylogenetic trees were build with representative siphophages as described by Niu *et al.* (2014), using VICTOR (Meier-Kolthoff and Göker, 2017) and visualized using FigTree 1.4.3 (Rambaut, 2006).

4. Results

This chapter describes the experimental results obtained in the present study, characterizing the six phages of interest.

4.1 Morphology

The six phages of interest produced plaques of similar morphology on the STEC O157:H7 host strain R508. Figure 4.1 displays the phages' plaques with clear centers and surrounding haloes, and ranging diameters of 0.2-0.8 mm.

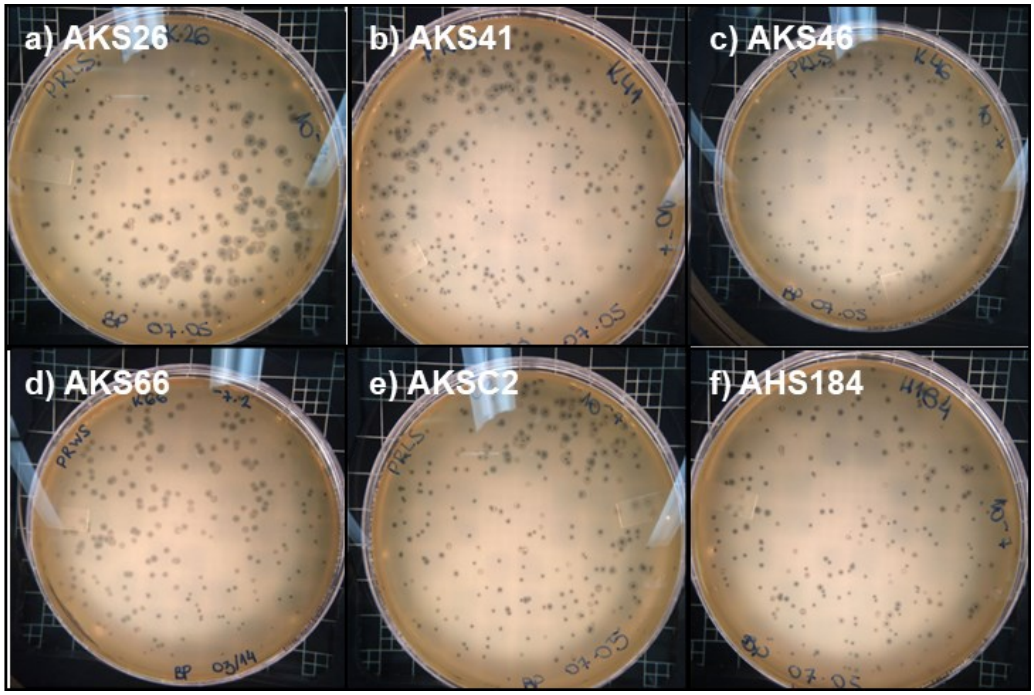


Figure 4.1 - Plaque morphology of the six phages of interest. a) AKS26; b) AKS41; c) AKS46; d) AKS66; e) ACS2; and f) AHS184 plated on host *E. coli* O157:H7 R508

In order to assess their morphology, the six phages of interest were imaged by TEM, as shown in Figure 4.2 below.

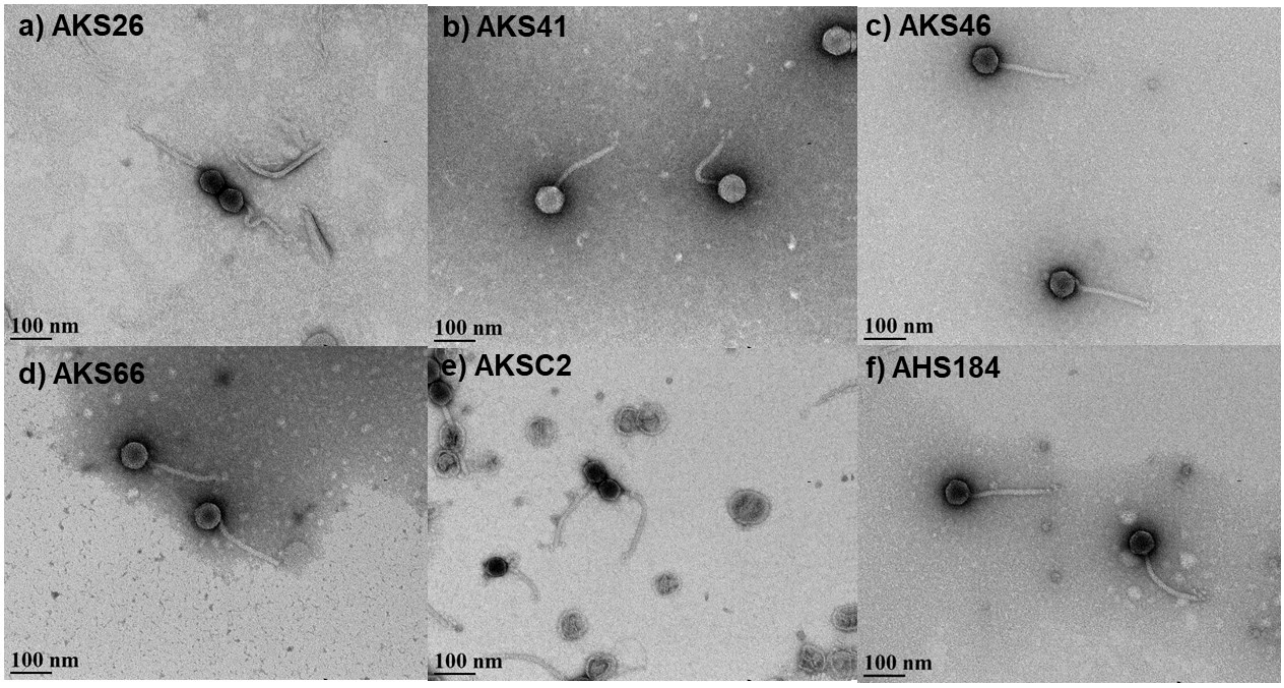


Figure 4.2 - Transmission electron micrographs of phages of interest. a) AKS26; b) AKS41; c) AKS46; d) AKS66; e) ACS2; and f) AHS184 negatively stained with uranyl acetate.

Table 4.1 below shows the phage sizes as determined by TEM. All six phages displayed icosahedral heads ranging from 51 to 61 nm in diameter, and long, flexible, non-contractile tails of 169 to 181 nm in length. Their structures suggest all the phage isolates are members of the *Siphoviridae* family. Phage capsid diameters did not differ significantly (α=0.05) except for phages AKS66, that significantly differed from phages and AKS41, AKS46, and ACS2. Only the tail lengths of isolates AKS46 and AKS66 statistically differ at a significance level of 0.05.

Table 4.1 Phage sizes determined by TEM

Phage	Capsid diameter (nm)	Tail length (nm)
AKS26	51±4	175±17
AKS41	59±10 *	163±15
AKS46	56±5 *	181±8***
AKS66	47±3 *,**	144±33***
ACS2	60±8 **	161±21
AHS184	53±4	179±11

*,**,*** indicates measurements that are significantly different (α=0.05).

4.2 Genome size and characterization

Figure 4.3 shows the electrophoresis gel for genome size estimation of the six phage isolates. The image was analyzed using the Bionumerics software to determine genome sizes, which are reported in Table 4.2.

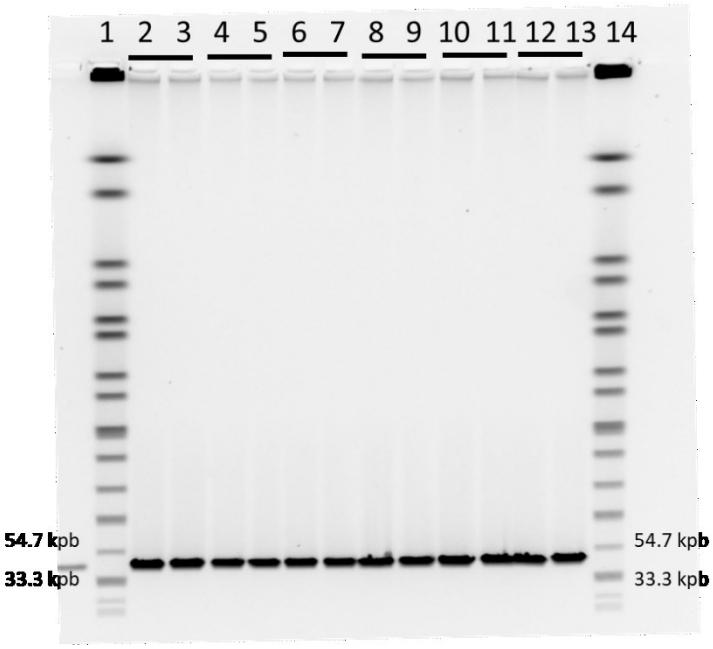


Figure 4.3 - Genome size estimation through PFGE. Lanes: 1 and 14) Salmonella branderup standard; 2 and 3) AKS26; 4 and 5) AKS41; 6 and 7) AKS46; 8 and 9) ACS2; 10 and 11) AKS66; 12 and 13) AHS184.

Table 4.2 - Genome sizes as determined by gel electrophoresis

Phage	Genome size (kbp)
AKS26	40.6±0.1
AKS41	40.8±0.0
AKS46	40.6±0.1
ACS2	40.8±0.3
AKS66	40.1±0.4
AHS184	41.8±0.3

The error reported is the standard deviation based on n = 2.

Samples AKS26, AKS41, AKS46, ACS2, and AKS66 had genomes ranging between 40.1 and 40.8 kbp – it is also important to note that sequencing results showed these genomes to be of equal size, at 45.7 kbp (see Table 4.3 below) –, whereas phage

AHS184 had a slightly larger genome at approximately 41.8 kpb – confirmed to be 46.7 kbp by sequencing (Table 4.3 below). These values are smaller than phage T1 genome size of, consisting of about 49 kbp of non-redundant nucleotides (Roberts, Martin and Kropinski, 2004), and smaller than phage vB_EcoS_Rogue1, comprised of 44.3 kb (Kropinski *et al.*, 2012). However, these genome sizes are consistent to the 45.7 kbp of phage AKS96 (Niu et al. 2014), isolated at the same feedlot as samples AKS26, AKS41, AKS46, ACS2, and AKS66; whereas phage AHS184 has a genome size similar to phages AHP24, AHS24, and AHP42, from the same geographical location, at 46.7, 46.4, and 46.8 kpb respectively (Niu et al. 2014).

Figure 4.4 displays the results for RFLP for the six phages of interest, providing the genome digestion patterns using restriction enzymes EcoRI and HindIII. RFLP was used to provide a fingerprint of the phages' genomes and quickly assess their similarity prior to sequencing. Phages AKS26, AKS41, AKS46, ACS2, and AKS66 displayed the same RFLP patterns regardless of the enzyme used, suggesting they are closely related genetically. Phage AHS184 exhibited different digestion patterns than the other five phages (lane 7 compared to 2-6, and lane 14 compared to 9-13).

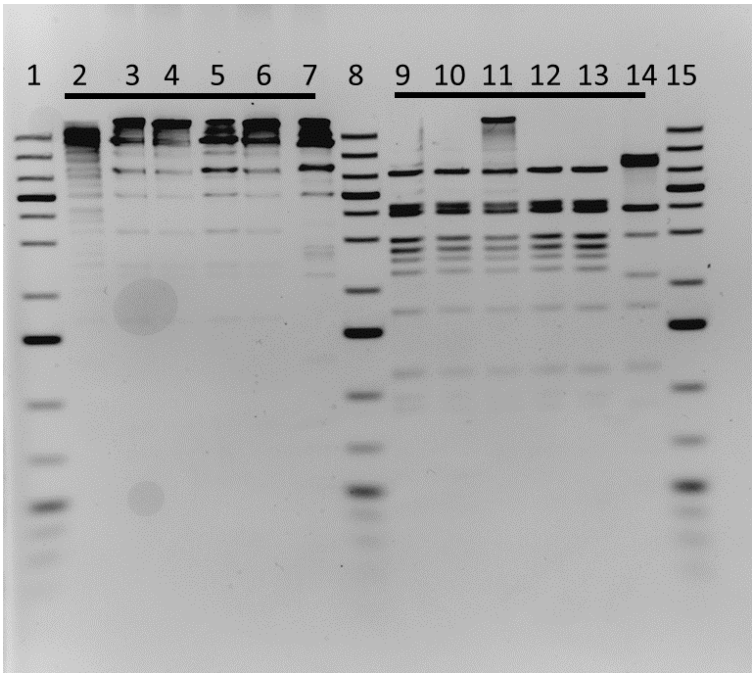


Figure 4.4 - RFLP results for the size phage samples. Lanes: 2-7: EcoRI digested samples. 9-14: HindIII digested samples. 1, 8, and 15: 1kb plus DNA ladder. 2 and 9: AKS26. 3 and 10: AKS41. 4 and 11: ACS2. 5 and 12: AKS46. 6 and 13: AKS66. 7 and 14: AHS184.

4.3 Genomics

To further determine their genotype, 6 phages were subjected to DNA extraction and whole genome sequencing. Table 4.3 shows general genomic features of the six phages of interest.

Table 4.3 - General genomic features of the six phages of interest

Feature	Phage					
	AKS26	**AKS41**	**AKS46**	**AKS66**	**AKSC2**	**AHS184**
Size (bp)	45,745	45,745	45,741	45,746	45,746	46,720
G+C content (%)	43.9	43.9	43.9	43.9	43.9	43.8
Total ORFs	121	121	120	119	119	129
Average ORF size (bp)	417	415	420	423	423	398
% of genome coding for proteins	90.9	90.9	91.03	91.05	91.05	91.17
Number of hypothetical proteins	38	38	38	37	38	45
Number of tRNAs	1	1	1	1	1	1
Number of promoters	96	96	96	95	95	98
Number of rho-independent terminators	22	22	22	21	21	27

ORF: open reading frame

All six phages consist of 45.7-46.7 kbp of double-stranded DNA, collinear with the Rogue1 virus of the *Tunavirinae* subfamily. Phage AHS184 was the only one with a genome size different than 45.74 kbp, with a genome of 46.72 kbp. The G+C content of the phages was lower than that of T1 phage (45.6%) (Roberts, Martin and Kropinski, 2004), and slightly lower than that of Rogue1 (44.2%) (Kropinski *et al.*, 2012) .

Figure 4.5 shows the alignment and comparison of the six phages. Green regions on the figure mean an area of high genomic similarity among all phages, whereas yellow areas mean one or more phages differ from the consensus sequence. The gray lines correspond to each phage and the black regions indicate where these differences from the consensus sequence are located. The overall pairwise similarity of the six phages' genomes was 96.3% using the Geneious pairwise identity algorithm. However, five out of the six phages – namely AKS26, AKS41, AKS46, AKS66, and

AKSC2 – shared 99.9% similarity. Phage AHS184 show differences from the other five phages, including a bigger genome size.

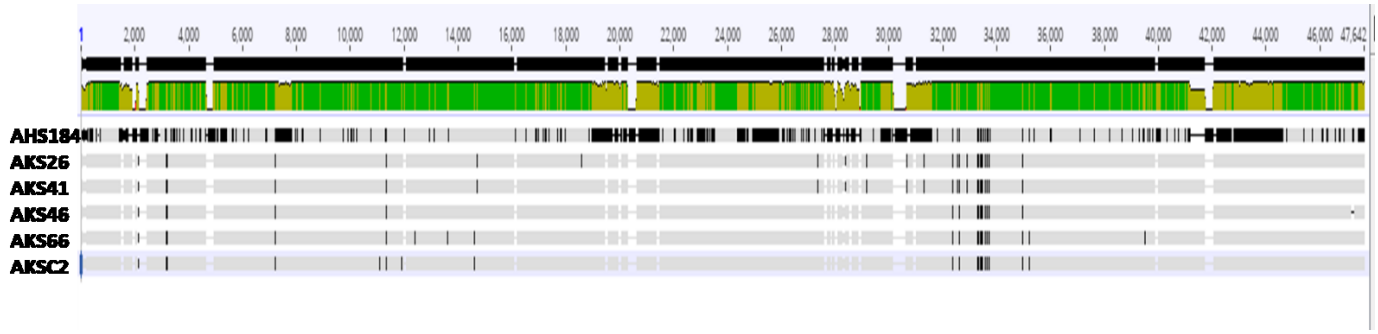


Figure 4.5 - Alignment and comparison of the six phages of interest using the Geneious pairwise identity algorithm

Figure 4.6 shows a BLAST map of the six phages compared to phage vB_EcoS_Rogue1, a STEC O157:H7 phage isolated in a previous feedlot trial. Differently than the pairwise nucleotide alignment in Figure 4.5, the BLAST map compares protein coding regions on the genomes using the BLAST algorithm and highlights significant alignments, indicated by darker colors on the circles, with the reference sequences. Phage AHS184 shares a higher similarity with Rogue1 than with the other five phages.

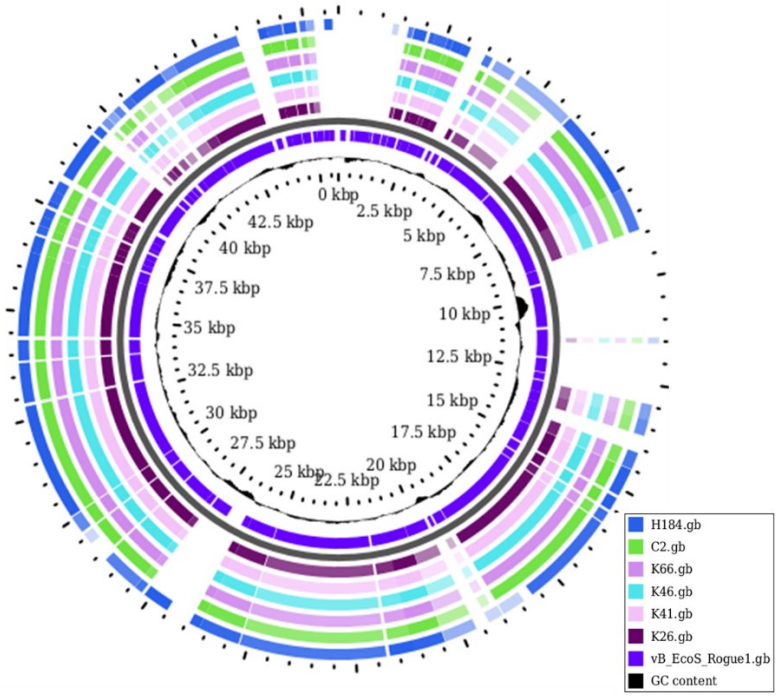


Figure 4.6 - BLAST Map of the six phages of interest against vB_EcoS_Rogue1 as a reference sequence

The five phages that shared high genome similarity had 15 point nonsynonymous mutations that differentiated them. Table 4.4 lists these mutations compared to the consensus sequence.

Table 4.4 – Nonsynonymous point mutations of phages AKS26, AKS41, AKS46, AKS66, and AKSC2

Phages	Position	ORF number	Gene product	Amino acid mutation	Mutation type
AKSC2	10,400	32	Hypothetical protein	Met → Val	Point mutation
AKS41, AKS26	10,662	32	Hypothetical protein	Asn → Thr	Point mutation
AKSC2	11,240	58	Tail Fiber	Pro → Gln	Point mutation
AKS66	11,718	58	Tail Fiber	Arg → Ser	Point mutation
AKS66	12,939	34	Hypothetical protein	Ala → Thr	Point mutation
AKS41, AKS26	14,046	60	Hypothetical protein	Asn → Asp	Point mutation
AKS26	17,903	8	Phage tail length tape	Val → Ala	Point mutation
AKS41, AKS26	26,350	58	Hypothetical protein	Pro → Ser	Point mutation
AKS41, AKS26	27,291	40	Anti-repressor protein	Frameshift	Deletion
AKS41, AKS26	30,741	87	Tail Fiber	Ile → Met	Point mutation
AKS41, AKS26	31,282	87	Tail Fiber	Val → Ala	Point mutation
AKS41, AKS26	31,664	87	Tail Fiber	Thr → Ala	Point mutation
AKS41, AKS26	31,809	87	Tail Fiber	Phe → Asp	Point mutation
AKS66	37,865	18	Hypothetical protein	Asp → Tyr	Point mutation
AKS 66	45,292	-	Non-coding region		Deletion

ORF: open reading frame

Out of the 12 nonsynonymous point mutations, 6 occurred in hypothetical proteins of unknown function and 6 occurred in genes encoding for putative tail fibers. Figure 4.7 is a graphical depiction of the comparison of the two putative genes for tail

fibers of phages AKS26, AKS41, AKS46, AKS66, and AKSC2, and the nonsynonymous mutations between them – the notation P11240Q indicates a mutation from proline to glycine at the 11240 position of the genome.

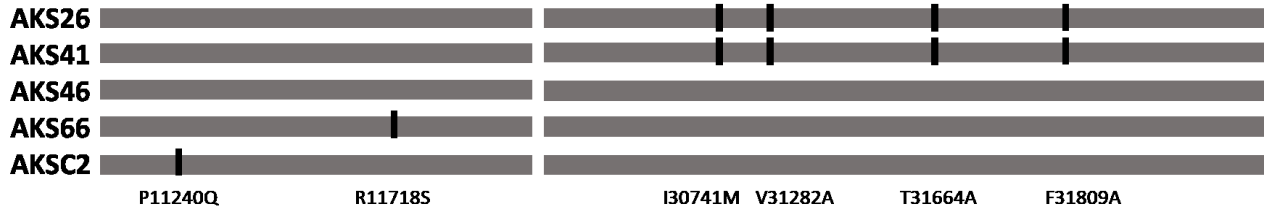


Figure 4.7 – Nonsynonymous point mutations analysis of two putative tail fiber genes of phages AKS26, AKS41, AKS46, AKS66, and AKSC2.

A phylogenetic tree of representative phages of the Siphoviridae family (Niu et al. 2014) is presented in Figure 4.8. Phylogenetic trees picture evolutionary relationships between species, and branching points on it represent species' common ancestors. Species' relatedness can be inferred from phylogenetic trees based on their common ancestors – the more recent the common ancestor (i.e. distant from the tree root), the more related the species are. The six phages characterized on this work are clustered together due to their high genome similarity. They are also clustered close to other phages isolated from the same feedlot trials in Southern Alberta (vB_EcoS_AHP24, AHS24, AHP42 and AKS96) (Niu et al. 2014; Niu et al. 2009a). Furthermore, the phages are placed in the Tunavirinae subfamily, and the Jk06likevirus genera.

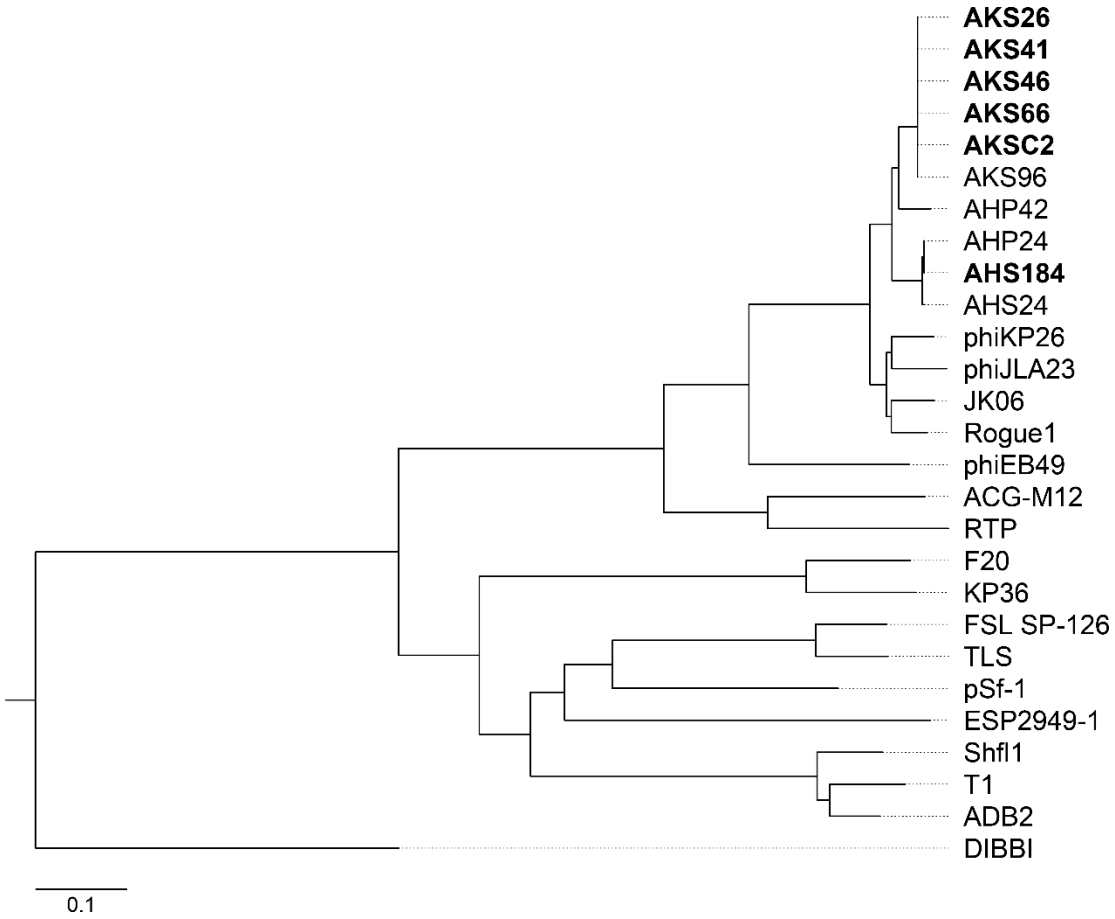
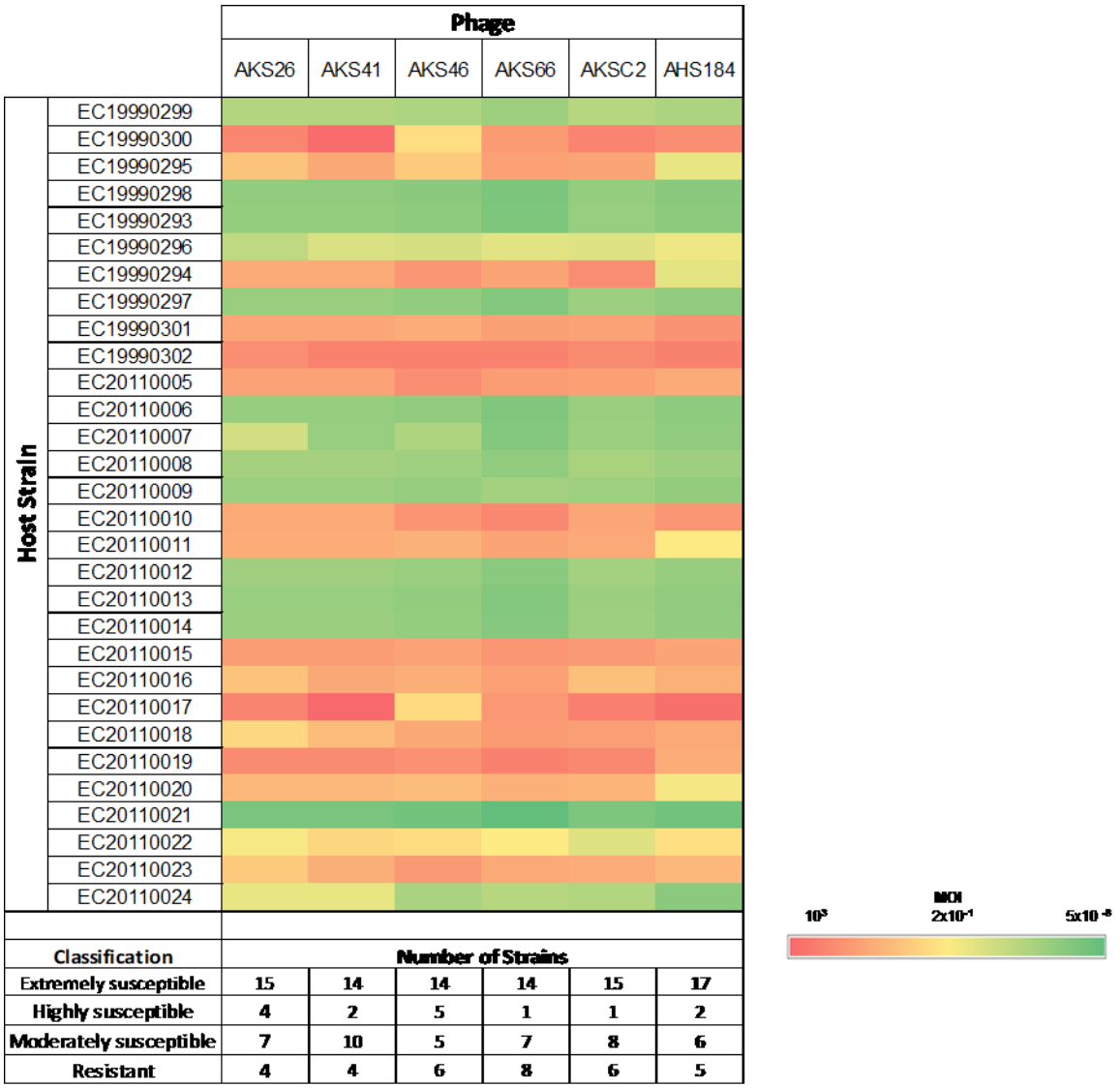


Figure 4.8 – Phylogenetic analysis of the whole genomes of the six phages of interest (in bold) and 25 T1-like reference phages. Scale bar represents 0.1 substitutions.

4.4 Host range and lytic capability

The phages of interest were then tested to determine their host range, an important feature for phage applications. Ideally, phages should recognize and lyse as many STEC O157:H7 as possible, but not other *E. coli* strains. The heat map in Figure 4.9 below shows the comparison of the lowest MOI to cause full lysis of the host population after 5 h of infection for the six phages of interest against 30 different STEC O157:H7 strains. These strains were then classified as extremely (MOI<0.01), highly (0.01≤MOI<1) or moderately susceptible (1≤MOI<10), or resistant (MOI>10) to the phages tested, as proposed by Niu et al. (2009a). In Figure 4.9, darker greens mean a that full visual lysis was caused by a lower MOI (less phage per bacterial cell), while

darker reds mean lysis was not observed or only observed at higher MOIs (10 phages per bacterial cell).



Classification	Number of Strains					
Extremely susceptible	15	14	14	14	15	17
Highly susceptible	4	2	5	1	1	2
Moderately susceptible	7	10	5	7	8	6
Resistant	4	4	6	8	6	5

Figure 4.9 - Heat map and strain classification for host range and lytic capability of the six phage samples against 30 different E. coli O157:H7 strains

The host range assay showed at least 14 out of the 30 STEC O157:H7 strains tested were extremely susceptible to the six phages (MOI required for complete lysis <0.01), with 17 host strains extremely susceptible to phage AHS184. Amongst these are STEC O157:H7 phage types 14 and 14a (strains EC19990298 and, EC20110007, respectively) those most commonly associated with outbreaks in North America.

The number of resistant strains (MOI required for complete lysis >10) varied for the six phages – 4 strains were resistant to phages AKS26 and AKS41; 5 to AHS184; 6 to AKS46 and ACS2; and 8 to AKS66.

4.5 Dynamics of infection and virulence

The infection of five different STEC O157:H7 strains of varying Phage Types by the six phages of interest was evaluated at different MOIs. The death curves generated for each host infected by each of the six phages can be found in Figures 4.10 to 4.13. The comparisons of the dynamics of infection for each combination of phage and host provide information on the virulence and infectivity of the phages.

Phages AKS26 and AKS41 (Figure 4.10a and 4.10b, respectively) were not able to lyse the EHEC strain EC19990295, even at the highest MOI tested, whereas phages AKS66 and AHS184 (Figure 4.10d and 4.10f) managed to control the bacterial growth after 5h at the highest MOIs tested. Against EHEC strain EC19990296 (Figure 4.11), all phages were able to prevent bacterial growth for at least 3h at an MOI of 1, but not at MOIs lower than 10^{-2}.

On host EHEC EC19902298, phage AKS26 showed the fastest population-wide lysis at low MOIs (e.g. at 2h at an MOI of 10^{-7} in Figure 4.12a.), while phage AHS184 showed only partial lysis even at an MOI of 10^{-6} (Figure 4.12f.) Against host EHEC EC19902299 (Figure 4.13), all phages were able to control bacterial growth at least until an MOI of 10^{-4}, with page AKS26 (Figure 4.13a) showing preventing host growth after 5h for all MOIs tested.

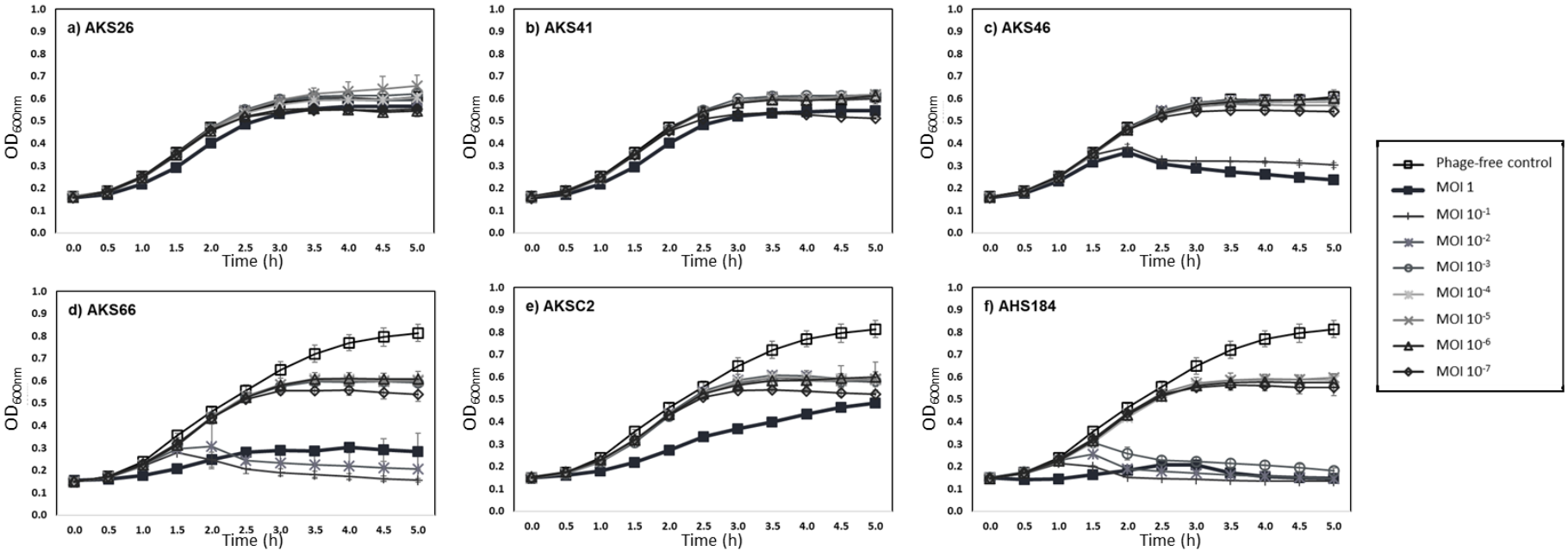


Figure 4.10 - Death curves of host EHEC EC19990295 (phage type 4) infected by each of the six phages of interest (a) AKS26, b) AKS41, c) AKS66, d) AKS66, e) AKSC2, f) AHS184) at MOIs ranging from 10-7 to 1. For each comparison, a control consisting of the host grown.

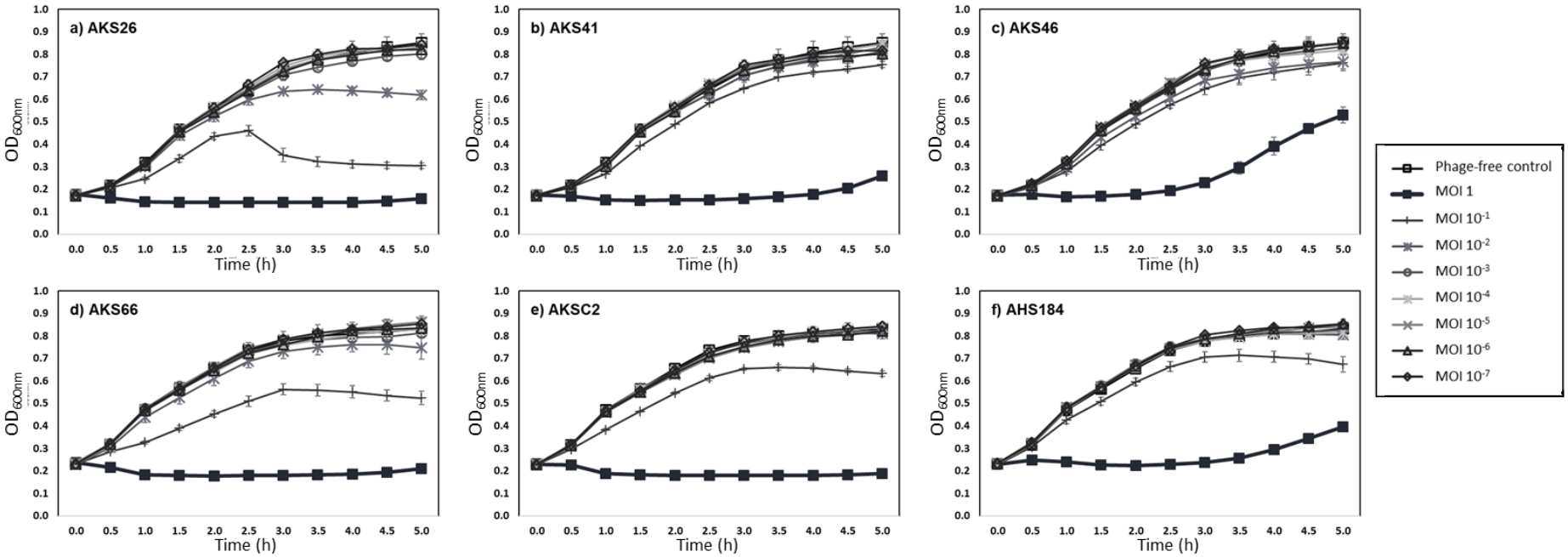


Figure 4.11 - Death curves of host EHEC EC19990296 (phage type 23) infected by each of the six phages of interest (a) AKS26, b) AKS41, c) AKS66, d) AKS66, e) AKSC2, f) AHS184) at MOIs ranging from 10^{-7} to 1. For each comparison, a control consisting of the host grown in the absence of phages is shown.

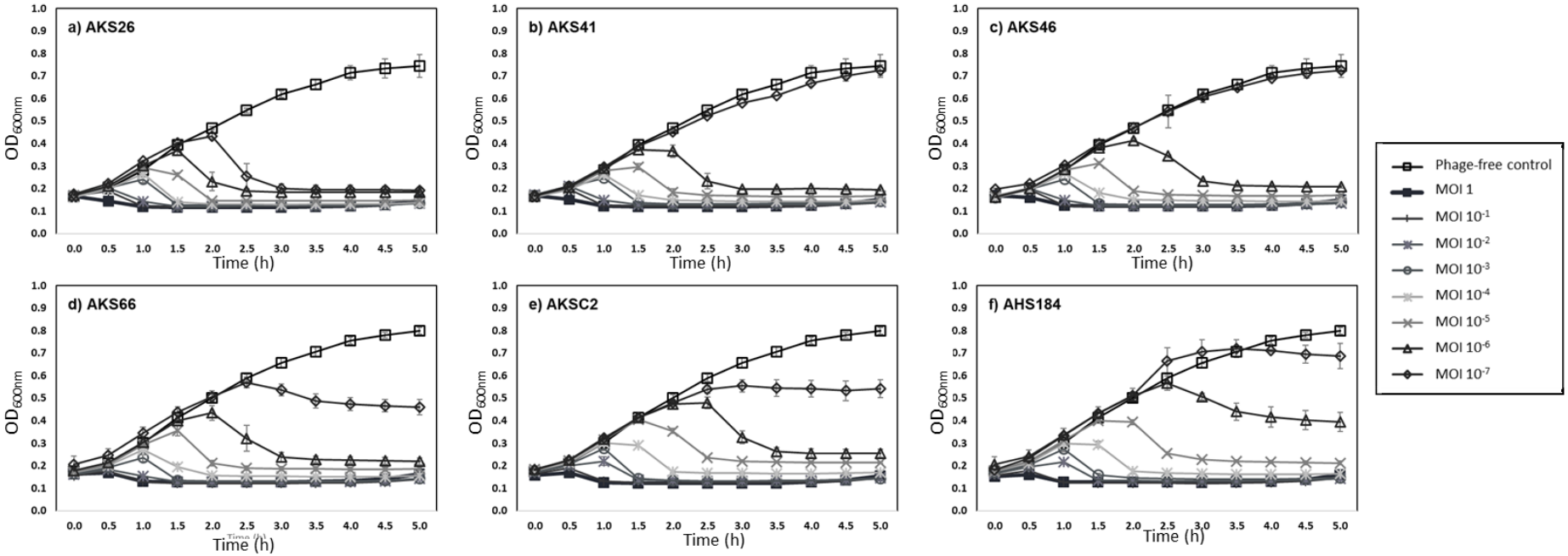


Figure 4.12 - Death curves of host EHEC EC19990298 (phage type 14) infected by each of the six phages of interest (a) AKS26, b) AKS41, c) AKS66, d) AKS66, e) AKSC2, f) AHS184) at MOIs ranging from 10^{-7} to 1. For each comparison, a control consisting of the host grown in the absence of phages is shown.

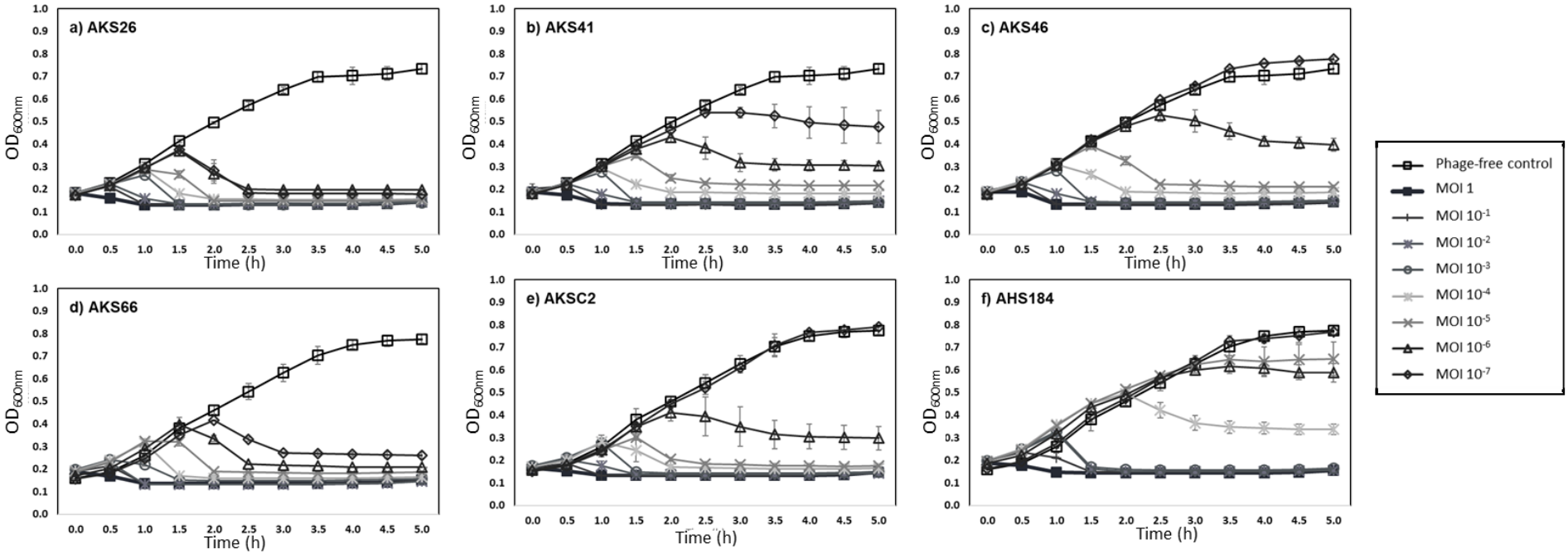


Figure 4.13 - Death curves of host EHEC EC19990299 (phage type 1) infected by each of the six phages of interest (a) AKS26, b) AKS41, c) AKS66, d) AKS66, e) AKSC2, f) AHS184) at MOIs ranging from 10^{-7} to 1. For each comparison, a control consisting of the host grown in the absence of phages is shown.

Virulence Indexes were established for each phage against each of the 5 hosts tested - Virulence Indexes vary from zero to one, with indexes close to one indicating higher virulence of a phage against a specific host strain at the conditions tested. These were calculated through the integration of the death curves shown in Figures 4.11 to 4.14 as described by Storms *et al.* (2019), and the resulting values are shown in Table 4.5 below. Overall, the phage isolates were more virulent against strain EC1992298, which is of the same phage type as the standard host. Overall, host strains EC19902295 and EC19902296 showed the least susceptibility to the phages tested; with virulence indexes ranging from 0.02 for phages AKS26 and AKS41 to 0.27 for AHS 184 (the only phage showing a Virulence Index above 0.20) for EC19990295, and all equal to or below 0.10 for EC19990296. In contrast, the most susceptible hosts were EC19990293 (all virulence indexes above 0.28), EC19990299 (all values above 0.25) and EC19990298 (all values above 0.41). Phage AKS26 displayed the overall most virulence among the six isolates, with the strongest virulence against EC19990293, EC19990298 and EC19990299.

Table 4.5 - Virulence Indexes of the six phage isolates against five EHEC host strains

Host strain	**Phage type**	**Phage**					
		AKS26	**AKS41**	**AKS46**	**AKSC2**	**AKS66**	**AHS184**
EC19902293	21	0.44	0.38	0.32	0.28	0.33	0.29
EC19902295	4	0.02	0.02	0.05	0.09	0.18	0.27
EC19902296	23	0.10	0.07	0.06	0.09	0.10	0.04
EC19902298	14	0.58	0.54	0.52	0.43	0.49	0.41
EC1990229	1	0.53	0.45	0.41	0.44	0.45	0.25

4.6 Adsorption kinetics

The adsorption kinetics of the six phages was determined against their isolated host (STEC R508) by measuring the concentration of free phages over time. To assess kinetic parameters, the data was normalized according to the initial phage concentration. As shown in Figure 4.14, adsorption dynamics were similar for each phage, except for AKS66 that displayed slower adsorption to the standard host.

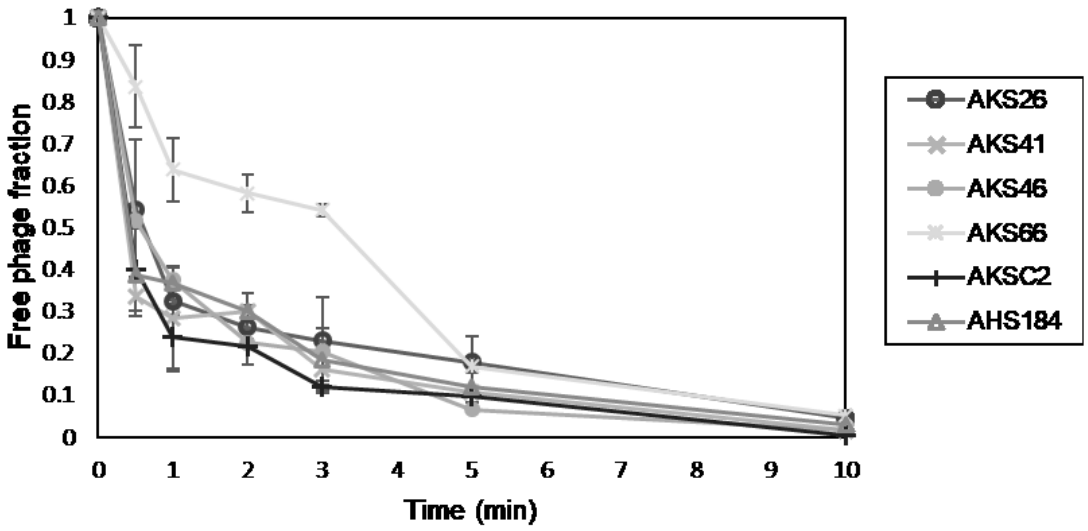


Figure 4.14 - Adsorption kinetics of the six phages of interest. Each data point corresponds to the average value of experimental triplicates with respective standard deviations.

Two different models were fitted to the data, and rate constants were estimated for each phage using each model. Model 1 describes the adsorption mechanism as a series of two reactions: a first reversible one, that can result on the phage desorbing of the host or not. The second reaction in Model 1 refers to an irreversible adsorption of the phage to the host leading to infection. In Model 2, on the other hand, the adsorption mechanism is lumped into a single irreversible reaction, meaning that all encounters of phage and host would lead to a successful infection.

Figure 4.15 shows the experimental results and models for each phage tested. As can be seen, Model 1 consistently describes the data better than Model 2, except for phage AKS66, to which both models had similar fits. The models residuals (difference between observed and predicted values) are shown in Figure 4.16 and the sum of the squared residuals for each phage and model is listed on Table 4.6.

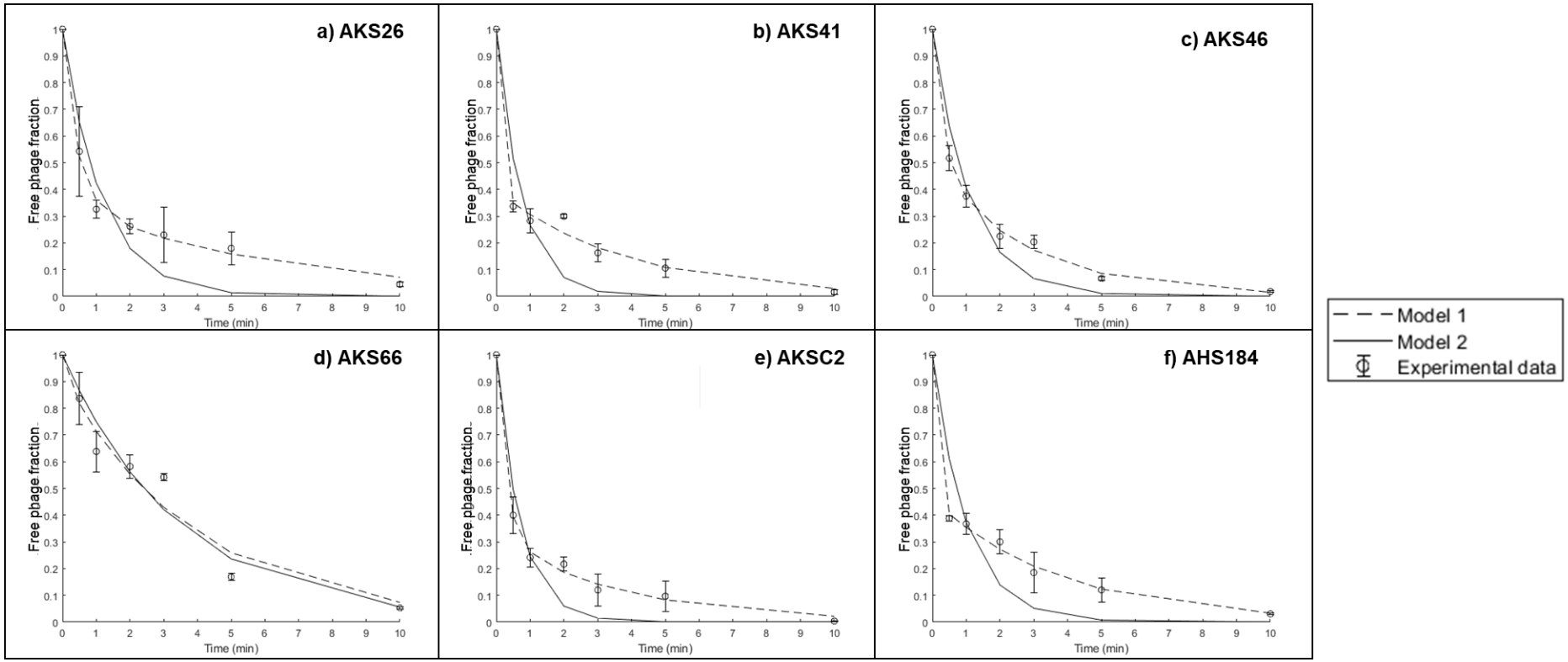


Figure 4.15 - Mathematical modeling of the adsorption kinetics of the six phages of interest (a) AKS26, b) AKS41, c) AKS66, d) AKS66, e) AKSC2, f) AHS184). For each phage, two different models are presented in comparison with the experimental data. Each data point corresponds to the average value of experimental triplicates with its respective standard deviations.

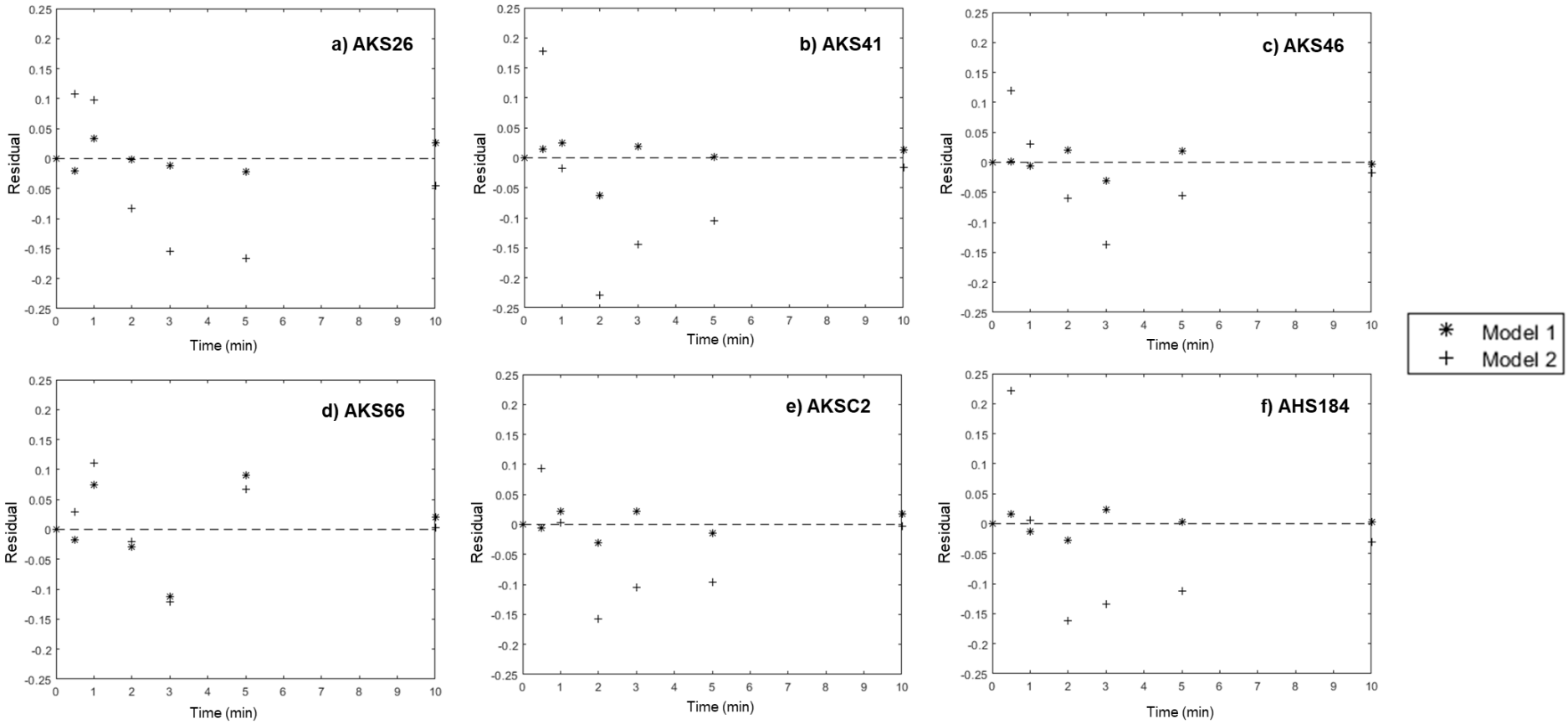


Figure 4.16 - Residual values for each of the two adsorption mathematical models of the six phages of interest (a) AKS26, b) AKS41, c) AKS66, d) AKS66, e) AKSC2, f) AHS184)

Table 4.6 - Sum of squared residuals values for each phage and model

	Sum of Squared Residuals	
Phage	**Model 1**	**Model 2**
AKS26	0.0029	0.0813
AKS41	0.0052	0.1167
AKS46	0.0018	0.0412
AKS66	0.0280	0.0329
AKSC2	0.0025	0.0539
AHS184	0.0018	0.1067

Model 1 presents randomly dispersed and smaller residuals than Model 2, except for phage AKS66. The average sum of squared residuals of Model 1 is one order of magnitude smaller than that of Model 2 – 0.0070 and 0.0721, respectively.

The second model generally underestimates the free-phage concentration earlier in the infection time (under 2 minutes) and then overestimates it, getting to a final free-phage concentration close to zero after 10 minutes, which is not observed experimentally.

Table 4.7 lists the rate constants for each phage obtained from Model 1. k_1 and k_2 are the rate constants of the forward and reverse reactions, respectively, of the initial adsorption step, and k_3 is the rate constant of the irreversible second step. The k_1 values ranged from 5.8×10^{-13} for phage AKS66 to 2.0×10^{-11} for AHS184; k_2 from 6.9×10^{-6} for phage AKS26 to 1.7×10^{-4} for AHS184; and k_3 from 2.4×10^{-6} for phage AKS26 to 1.8×10^{-5} for phages AKS66 and AKSC2. Generally, from this model, phage AHS184 showed the fastest adsorption rate on the tested host and AKS66 had the slowest. The highest calculated desorption rate constant also belongs to AHS184, meaning more reversible binding to the host are predicted to occur with that phage, whereas phage AKS26 has the lowest desorption rate constant.

Table 4.7 - Rate constants for each phage calculated using Model 1 (two-step process)

	Rate constant (k) (mL·min^{-1})		
Phage	k_1	k_2	k_3
AKS26	$1.6 \cdot 10^{-11} \pm 5.3 \cdot 10^{-13}$	$6.9 \cdot 10^{-6} \pm 5.8 \cdot 10^{-6}$	$2.4 \cdot 10^{-6} \pm 2.1 \cdot 10^{-6}$
AKS41	$1.0 \cdot 10^{-11} \pm 1.7 \cdot 10^{-10}$	$6.3 \cdot 10^{-5} \pm 2.1 \cdot 10^{-3}$	$4.3 \cdot 10^{-6} \pm 3.2 \cdot 10^{-6}$
AKS46	$1.8 \cdot 10^{-12} \pm 8.4 \cdot 10^{-13}$	$1.1 \cdot 10^{-5} \pm 1.4 \cdot 10^{-5}$	$4.3 \cdot 10^{-6} \pm 3.2 \cdot 10^{-6}$
AKS66	$5.8 \cdot 10^{-13} \pm 5.1 \cdot 10^{-12}$	$2.0 \cdot 10^{-5} \pm 5.3 \cdot 10^{-4}$	$1.8 \cdot 10^{-5} \pm 2.0 \cdot 10^{-4}$
AKSC2	$2.5 \cdot 10^{-12} \pm 1.1 \cdot 10^{-12}$	$2.0 \cdot 10^{-5} \pm 5.3 \cdot 10^{-5}$	$1.8 \cdot 10^{-5} \pm 2.0 \cdot 10^{-5}$
AHS184	$2.0 \cdot 10^{-11} \pm 5.8 \cdot 10^{-11}$	$1.7 \cdot 10^{-4} \pm 5.1 \cdot 10^{-4}$	$4.9 \cdot 10^{-6} \pm 2.0 \cdot 10^{-6}$

Results are expressed with 95% confidence intervals

The same set of results was fitted to Model 2, a simplified one-step model. Table 4.8 below shows the resulting adsorption rate constants for Model 2. Using this model, phage AKSC2 showed the fastest adsorption rate constant and AKS66, the slowest.

Table 4.8 - Adsorption rate constants for each phage calculated using Model 2 (single-step process)

Phage	**Adsorption rate constant (k) (mL·min^{-1})**
AKS26	$8.6 \cdot 10^{-12} \pm 4.2 \cdot 10^{-12}$
AKS41	$1.3 \cdot 10^{-11} \pm 2.4 \cdot 10^{-11}$
AKS46	$9.0 \cdot 10^{-12} \pm 3.1 \cdot 10^{-12}$
AKS66	$2.9 \cdot 10^{-12} \pm 7.9 \cdot 10^{-13}$
AKSC2	$1.4 \cdot 10^{-11} \pm 6.3 \cdot 10^{-12}$
AHS184	$9.9 \cdot 10^{-12} \pm 5.7 \cdot 10^{-12}$

Results are expressed with 95% confidence intervals

5. Discussion

This section will discuss the findings previously described and the implications of such in the characterization of the studied phages and their use in different applications.

5.1 Identification of phages

As seen in Figures 4.1 and 4.2, the six phages described in this work have similar plaque and virion morphologies. The observed icosahedral heads and long, flexible, non-contractile tails classify the six phages as members of the Siphoviridae family. This family was first described in 1960 (Bradley and Kay, 1960) and is represented by phages T1 and T5. Tailed phages of this morphology are the most commonly isolated and described in literature (Figure 2.3).

The micrographs were obtained from staining with uranyl acetate. Known artifacts of this type of staining are: production of both positive and negative staining; swelling of protein structures such as tails; and shrinkage of positively stained capsids (Ackermann, 2012). In Figure 4.2 most phage capsids appear darker on the images (notably for ACS2) and thus were positively stained. Therefore, the measurement of phage head structures showed in Table 4.1 might be underestimating the actual capsid size. The dimensions of the six phages are similar – only the capsid diameter of AKS66 is significantly smaller than those of AKS41, AKS46, and AKSC2. The length of the tail fibers of AKS66 is also significantly smaller than AKS46.

The genome sizes and restriction patterns assessed through gel electrophoresis (Figures 4.3 and 4.4) reveal that five of the six phages (AKS26, AKS41, AKS46, ACS2, and AKS66) showed high genome similarity, which was further confirmed by genome sequencing – including similar GC contents, number of predicted ORFs, and overall identity (Table 4.3 and Figure 4.5). The morphology, genome size, digestion and sequencing results demonstrate that these five phages are likely variants of a parent phage. In contrast, phage

AHS184 was confirmed by sequencing to have a larger genome size than the other five phages (46.7 kbp instead of 45.7 kpb) and lower pairwise similarity (96.3%).

A high degree of similarity between phages isolated from the same feedlot pen from different sample sources (e.g. fecal pats and manure slurry) and even different feedlots was observed before (Niu et al. 2009a; Niu et al. 2014). The five phages described in the present work that share high genome similarity were isolated from different pens of the same commercial feedlot at different sampling times, whereas the sixth phage, AHS184, was isolated at a different facility in Southern Alberta. The six phages described in this study are closely related to other T1-like phages isolated in the same feedlot trial, namely AHP24, AHS24, AHP42, and AKS96, as seen in Figure 4.8. We hypothesize this family of phages is widely spread throughout feedlots in Southern Alberta and represent the dominant variants in their ecological niches, which could explain the isolation of related phages at different geographical locations over time.

The genome of phage AHS184 has numerous substitutions to the consensus sequence of the six phages. Notably, genes encoding putative tail fibers and tail length tape-measure proteins differ in AHS184. Tail length proteins assist the regulation of phages' tail fibers assembly and thus determines its sizes (Katsura and Hendrix, 1984; Katsura, 1990), whereas tail fibers are receptor binding proteins that recognize the host cell. These differences in AHS184 could at least partially explain the differences in its host range and why it is generally broader than for the other phages characterized. Most interestingly, the genome of phage AHS184 encodes four hypothetical proteins with unknown function that are not present in the other five phages. However, all of these proteins are conserved in other closely related phages found in the NCBI database: namely, Rogue1 (1 gene product found), AHS24 (2 gene products found), AHP42 (1 gene product found), KP26 (1 gene product found), and phiJLA23 (1 gene product found). The fact that these proteins are present in phages of different origins – these phages were isolated from feedlot cattle feces in southern Alberta (Rogue1) (Kropinski *et al.*, 2012), a commercial cattle feedlot in Southern Alberta

(AHS24 and AHP42) (Niu et al. 2014), and farm animal feces in Mexico (phiJLA23) (Amarillas, Cristobal Chaidez, *et al.*, 2013) - suggests they likely play roles in phage fitness.

An interesting finding from Figure 4.8 is the evolutionary relationship between the six phages described in this work and other phages isolated from feedlot trials, namely Rogue1 and JK06 (Kropinski *et al.*, 2012) isolated in Canada and Israel, respectively, and phages phiKP26 and phiJLA23, isolated from animal feces in Mexico (Amarillas, Cristobal Chaidez, *et al.*, 2013; Amarillas, Cristóbal Chaidez, *et al.*, 2013). Notably, phage phiKP26 also infects *Salmonella* and has tail fibers that differ significantly from other coliphages (Amarillas, Cristóbal Chaidez, *et al.*, 2013). The fact that phages isolated from such distant regions share a recent common ancestor and high genome similarity is outstanding. Even though the genomic data cannot explain how this occurred, it is remarkable that such phenomena can be detected and, albeit phages have such rich diversity, we can still find evolutionary relationships on samples isolated independently.

The identification of phages with close genetic relatedness from samples isolated at distant geographical location was reported before. An example is the discovery of the highly abundant crAssphage (cross-assembly phage), identified in 2014 through computational analysis of publicly available human fecal metagenomics data and comprising up to 90% of viral-like particle reads in some datasets (Dutilh *et al.*, 2014). This phage is predicted to infect members of the gut commensal phylum Bacteroidetes, the most abundant in the human gut (Dutilh *et al.*, 2014; Yutin *et al.*, 2018). Since then, crAssphage-like viruses have been found to be nearly ubiquitous in healthy human gut studies worldwide and have been identified in metagenomics datasets from different continents (Manrique *et al.*, 2016; Yutin *et al.*, 2018).

Furthermore, phage Rogue1 was isolated after failed attempts to use STEC O157:H7 phages to control this bacterial pathogen in sheep. After animals were inoculated with chosen therapeutic phages, Kropinski and colleagues (2012) would consistently detect a different phage being shed, indicating the

“rogue” phage, instead of the administrated phages, was dominating the ecological niche in the treated animals. Upon further characterization, the Rogue1 phage was showed to be of *Siphoviridae* morphology, whereas the phages initially assessed for therapeutic potential were myoviruses. The same phage emerged again four years later in a different trial at the same facility (Kropinski *et al.*, 2012). This illustrates the importance of characterizing and evaluating endemic phages in phage therapy, as these viruses are often the most suited to dominate their ecological niches, and can be interesting candidates for a variety of applications.

The genomic similarity of the phages studied with the vB_EcoS_Rogue1 virus and vB_EcoS_AKS96 previously described in literature (Kropinski et al. 2012; Niu et al. 2014) supports their classification as members of the Tunavirinae subfamily and the Rogue01likevirus genera. This is also supported by their close clustering in the phylogenetic tree shown in Figure 4.8.

Currently, the names of phage species are composed of three elements: the bacterial host genus followed by the word “phage,” and a unique identifier. This unique identifier contains information on the bacterial host genus, phage morphology, and an exclusive short name or code (Adriaenssens and Brister, 2017). Following the guidelines for assigning phage names to new isolates (E. Adriaenssens and Brister 2017), the phages characterized in this work were officially designated as *E. coli* phages vB_EcoS_AKS26, vB_EcoS_AKS41, vB_EcoS_AKS46, vB_EcoS_AKSC2, vB_EcoS_AKS66, and vB_EcoS_AHS184, indicating they are **v**iruses of **b**acteria, infecting ***Escherichia coli***, of **S**iphoviridae morphology. The unique names indicate their location of isolation (**A**lberta feedlots **K** or **H**), sample type (**S**lurry), and number of the collection sample.

5.2 Host range

Even though phages AKS26, AKS41, AKS46, ACS2, and AKS66 share 99.9% genome similarity (Figure 4.5), their host range and lytic capability, infection dynamics, virulence indexes, and adsorption kinetics are different (Figure 4.9, Table 4.5, and Figure 4.15, respectively). For example, phage

AKS26 and AKS41 show a broader host range than the other four phages (more STEC O157 strains were extremely susceptible to it, and fewer were resistant) (Figure 4.9) and its virulence index was generally higher over the STEC O157:H7 strains tested (Table 4.5).

There are no standardized definition for the term “broad host range”; it has been used in literature to describe phages that infect multiple strains of the same species, bacteria of different species, and bacteria for different genera (Ross, Ward and Hyman, 2016). Moreover, the methods used for the determination of host range are also subjected to limitations. For example, the spot testing method, although simple and rapid, often overestimates host range by placing a large amount of phages on a growing bacteria lawn. This can generate false positives due to cell lysis without infection, either by lysis from without (cell burst due to a large number of phages adsorbing to it) or lysis from residual endolysins in the phage stock (Abedon, 2011; Ross, Ward and Hyman, 2016), thus testing dilutions of phages is desirable. On the other hand, testing for plaque formation to assess host range can generate false negatives as not all phages are able to form visible plaques as this is highly dependent on the viral progeny generated (Ross, Ward and Hyman, 2016).

In the present work, host range was assessed by visually inspecting the turbidity of liquid cultures of different STEC O157:H7 strains after 5h of infection at different MOIs, which also allows to identify and compare the lytic capabilities of the phages (Figure 4.9). This circumvents the limitations mentioned above, but could generate false negatives in the case of, for example, a host strain mutates to develop resistance early on the infection process and, even though the phage would be able to generate progeny, the final turbidity of the culture would indicate otherwise.

The nonsynonymous mutations that differentiate the genomes of phages AKS26, AKS41, AKS46, ACS2, and AKS66 are centered on genes encoding RBPs, as indicated in Table 4.4 – out of the 12 nonsynonymous mutations, six are single point mutations in putative tail fibers-encoding genes. Recognition of the host cell by the phage often determines its host range and it has been

extensively shown that mutations in RBPs can alter phage host range (de Jonge *et al.*, 2019).

Modulation of host range can be achieved by altering RBP-encoding genes: Yoichi et. al. (2005) demonstrated that by altering the tail fibers of T2 by cloning genes 37 and 38 of phage PP01 encoding its tail fibers made T2 able to infect an *E. coli* O157:H7 strain. The same research group later showed that the host range of phage T2 can be broadened by switching its tail fibers to those of myophage IP008 (Mahichi *et al.*, 2009).

Even single point mutations in RBP-encoding genes can significantly impact phage host specificity. Le et. al. (2013) isolated two *Pseudomonas aeruginosa* phages, namely PaP1 and JG004, with different host specificities but with a single nucleotide mutation on RBP-encoding genes. By cloning the putative tail fiber and baseplate genes into chimeric phages, the authors were able to switch their host specificity. Moreover, they also obtained natural mutants from challenge experiments with the same single-point mutation on the C-terminal of the putative RBP gene (Le *et al.*, 2013), suggesting the end of that protein was determinant of host susceptibility. However, the authors only sequenced these putative tail fiber and baseplate genes instead of the whole genome of the phages, hence the contribution of other genes cannot be discarded. Additionally, only two host strains were tested for this work instead of a range of different strains.

Mutations on the C-terminal portion of the tail fiber proteins of phage λ were also shown to alter its host range (Wang, Hofnung and Charbit, 2000). In the 5 phages that share high genome similarity described in this study, the point mutations found in putative RBP-encoding genes are not located towards the C-terminal end of the protein (Figure 4.7). However, as their proteins structures are unknown, it cannot be inferred that the mutations do not occur towards the tip of the RBPs. An example of that is T4 tail fibers – due to the tail fiber tip protein folding, the receptor binding domain is actually expressed by the middle of gene 37 and not towards the C-terminal of the protein primary structure (Bartual *et al.*, 2010).

Moreover, phages with broader host range have been found to bind to more abundant and common receptors on bacterial cell surfaces or to possess multiple RBPs. Polyvalent phages (i.e. those that bind to different receptors molecules) have been found to carry multiple RBPs – phi92, a myovirus of the rV5-like genus, is able to infect multiple laboratorial *E. coli* and *Salmonella* strains and expresses five types of tail fibers and tail spikes pointing at different directions from the baseplate on the same virion (Schwarzer *et al.*, 2012).

Phage AKS26 and AKS41 presented the lowest number of strains resistant to them (Figure 4.9). These phages do not carry different genes from AKSC2, AKS46, and AKS66, but rather show the same point mutations on three hypothetical proteins, four point mutations in tail fiber encoding genes, and one in an anti-repressor protein encoding gene (Table 4.4). Additionally, phage AKS26 carries a point mutation in the gene encoding for its tail length tape. Mutations in tail fibers could alter their structure and thus their specificity by, for example, generating an RBP with easier spatial access to the receptors on the host cell membrane. Phage anti-repressor proteins allow a bacteria that already possesses a lysogenic phage region in its genome to get superinfected (Botstein *et al.*, 1975). Mutations leading to more efficient anti-repressor proteins could lead to broader host ranges by allowing more bacterial strains to be susceptible to the phage, whereas mutations in tape length proteins could alter the structure and folding of the RBPs. Presumably, it is possible that all these mutations might play a synergetic role in affecting the host range of the phage, which merits further investigation.

5.3 Adsorption kinetics

The first notable observation when comparing adsorption kinetics was that the sequential model (two-step process) described the experimental data better than the one step process (lumped model) for all phages, except AKS66 for which both models were equivalent (Table 4.6 and Figure 4.15). This indicates reversible interactions with the bacterial hosts are relevant for the phages of interest and that the desorption step is important for the description of adsorption

kinetics. This is consistent with the findings of Storms et al. (2012), in which the authors successfully used a one-step model to depict the adsorption kinetics of T-series phages (namely, T2, T4, T5, T6, and T7), but not a phage λ variant without side tail fibers. Phage T1, a representative of *Siphoviridae* morphology, is historically described with a two-step adsorption process (Garen and Puck, 1951). Since the phages characterized in the present study are siphoviruses and do not have long side tail fibers, the results presented in this study further support the hypothesis that the reversible step is significant for this type of phages. This also highlights how mathematical modeling can aid in understanding mechanisms involved in biological processes.

Secondly, phage AKS66 displayed notably slower adsorption kinetics than the other four phages that share high genome similarity. Its calculated adsorption rate constant in the sequential model was one order of magnitude smaller than all others (Table 4.7). This indicates the first step of the adsorption process is more strongly limiting to AKS66 than to the other studied phages.

As phage AKS66 only presents one substitution from the consensus sequence on its tail fibers encoding genes (Figure 4.7), it can be hypothesized the change from arginine to serine at position 11,718-11,721 of the phage genome could be responsible for the slower adsorption and different host range. Arginine has a larger side chain than serine, and arginine is negatively charged, whereas serine is uncharged. Hence this amino acid mutation could affect protein folding or conformation and thus, its specificity. Other nucleotide substitutions in AKS66 (Table 4.4) include two nonsynonymous point mutations, one at an intergenic region and another within the coding region of a protein of unknown function. The latter is similar to sequences of other phages in databases (*Escherichia* phage vB_EcoS_AHP42 and *Escherichia* virus KP26).

Despite its slow adsorption, phage AKS66 has a higher virulence index against strain EC19990295 than its four closely related phages (AKS26, AKS41, AKS46, AKSC2) (Table 4.5). This means that other factors influencing virulence and infection dynamics – such as large burst size and short lysis time – overcome the slow adsorption rate.

The fastest adsorbing phage was AHS184, followed by AKS26 and AKS46 (k_1). AKS26 also presents the lowest desorption rate (k_2) of the six phages. AHS184, on the other hand, has a desorption rate two orders of magnitude larger than AKS26, reiterating the adsorption dynamics differs for these phages. This also indicates that, under the same conditions, AKS26 phages desorb from the host 100 times slower than AHS184.

Under the one-step model, the adsorption rates of the phage described in this study (Table 4.8) are 3 orders of magnitude lower than those described in literature for the reference phage T1 against its standard host *E. coli B* (Garen and Puck, 1951), and AKFV33, a T5-like phage isolated at the same feedlot trial against STEC O157:H7 (Niu et al. 2012) – both phages presented *k* values of the order of 10^{-9} mL/min, whereas the phages characterized in this study have adsorption rate values of the order of 10^{-12} mL/min. A note of caution is due here since these adsorption rates were determine under different conditions and with different bacterial hosts, and factors such as medium composition, ionic strength and temperature significantly affect the adsorption kinetics of phages (Garen and Puck, 1951; Storms *et al.*, 2012).

5.4 Virulence

All the phages studied showed highest virulence indexes against STEC O157:H7 strain EC19990298, which is the same phage type as the host strain R508 used for isolation. This is somewhat expected as both bacterial strains have the same phage receptors on their surfaces and affinity to the phages RBPs, and hence show similar phage susceptibility. These results may indicate some isolation bias – as the phages were isolated using strain R508, they could infect and lyse this bacteria and their relatives more efficiently, while performing poorly on other distantly related STEC O157:H7 strains such as EC19990295. Prior studies have noted that using a single host strain for phage isolation has a tendency to select for phages with narrow host ranges. Variations of standard isolation procedures using multiple hosts have been suggested to overcome this issue (Ross, Ward and Hyman, 2016). However, the six phages of interest were

isolated in a strategically selected animal isolate of phage type 14, the same commonly associated with outbreaks in North America (phage types 14 and 14a). This being said, the isolation procedure used for the phages of interest did not put a strict selection pressure on the virulence towards this phage type. Essentially, higher virulence could still have been observed against another strain or even bacterial phage type. An example of this is the high virulence of phages AKS26 and AKS66 against STEC strain EC19990299 of phage type 1.

Besides better and broader host recognition and adsorptions mechanisms linked to RBPs discussed above, other several other gene products can affect the virulence of phages. After adsorbing onto the host, lytic phages start to replicate their genomes and, at this stage, phages better adapted to the molecular mechanisms of the host (such as codon usage) or that can evade the bacterial cell defenses such as restriction modification (Rifat *et al.*, 2008) or CRISPR systems (Bondy-Denomy *et al.*, 2015), will cause more successful infections and hence, greater virulence. After genome replication and assembly, the exit of the cell is also an important phase to the virulence of a phage – phages with shorter latent periods and/or producing more efficient endolysins (e.g. capable to lyse diverse cell wall components) will release its progeny more efficiently and thus be more virulent. Another important factor is burst size, as a phage able to produce more progeny per infected cell will display greater against a bacterial culture.

In the present study, phage virulence was analyzed by integrating bacterial reduction curves of different strains of STEC O157:H7 infected with the phages of interest at different MOIs. The reduction curves were obtained by directly measuring the turbidity of the bacterial cultures in a spectrometer, instead of visual inspection as in the method used for host range previously described (Niu *et al.*, 2009a). Capturing the dynamics of infection also provides interesting insights – for example, phage AKS46 can prevent the growth of host strain EC19990296 during the first 2.5 h of infection at the highest MOI tested (Figure 4.11c). However, as the host starts growing after that, an endpoint analysis of turbidity would fail to detect this. For biocontrol applications on phage products

targeted to food manufacturers, for example, progeny is not necessarily a requirement for suitable phages, but rather their ability to quickly disrupt the pathogenic host's genome and so stopping its replication. Determining infection dynamics would be the best way to assess this feature.

Phage AKS26 showed the highest virulence indexes against the strains tested (Table 4.5), except against host EC19990295. This could be due the mutations in genes encoding relevant proteins such as RBPs, as previously discussed. As most of the point mutations found in this phage are located in regions encoding hypothetical proteins, it is hard to hypothesize the factors that confer this phage a higher virulence.

5.5 Implications for potential applications

Based on the results of host range, virulence and adsorption kinetics obtained for the phages investigated in this study, it is possible to determine which phage would be the best candidate(s) for different potential applications.

In order to be considered suitable for biocontrol or detection of food pathogens, phages cannot encode toxins in their genomes nor be able to undergo the lysogenic cycle – i.e. must be strictly lytic (Hagens and Loessner, 2010). None of the six phages of interest presented genes encoding integrases, toxins or similar proteins from information available in databases, which indicates they are likely safe for use in applications. However, a significant number of putative proteins encoded by these phages are of unknown function. These results therefore need to be interpreted with caution.

First, out of the 30 common phage types of human and animal STEC O157:H7 isolates tested, a maximum of 8 were resistant to the six phages (Figure 4.9). Moreover, for diagnostic applications such as biosensors based on genetically modified reporter phages, fast adsorbing phages with shorter latent periods are preferred (Schmelcher and Loessner, 2014) and thus AKS26 could be an interesting candidate. Interestingly, AKS26 also presented the lowest number of resistant STEC O157:H7 strains (Figure 4.9).

For phage therapy applications, phage host-specificity is considered a double-edged sword: phages with narrower host range can preserve the natural microbiota of the ecological niche treated, but pathogens can readily evolve resistance. As a consequence, these phages have to be used in combinations with others in phage cocktails. In contrast, phages with broad host range can have unstable phage-host interactions as their progeny might evolve to better infect non-pathogenic host strains instead of the targeted ones (de Jonge *et al.*, 2019). Thus, testing the effect of phage combinations to surpass the host range limitations of each virus should be of interest. For example, STECO157:H7 strains of phage type 8, such as EC20110008, are a common source of outbreaks in England and Spain (Mora *et al.*, 2004; Launders *et al.*, 2016). This strain is only moderately susceptible to phages AKS26, AKS41, AKSC2, and AHS184 but resistant to phages AKS46 and AKS66 (Figure 4.9), and a combination of these phages might be useful for applications directed biosanitation of food industry surfaces in those countries, for example. In terms of broader host range and lytic capability, phages AHS184, AKS26, and AKS41 could be suitable for biocontrol and therapeutic applications.

6. Conclusions and Future Works

The characterization of six phages infecting strains of STEC O157:H7 has shown how close phage variants can display variations at both the genetic and infective levels. The main findings of the study and recommendations for future research avenues are described below.

6.1 Conclusions

Six phages previously isolated from commercial feedlot cattle were characterized in terms of morphology, host range and lytic capability, adsorption kinetics, and virulence dynamics. Moreover, their genomes were characterized and the main genetic differences between the phages were determined.

Five of the six phages were isolated from the same feedlot at different sampling times and in different pens. The sixth phage originated from a different feedlot. The six phages share a high genome similarity and are closely related to other phages isolated in the course of the same feedlot trial from different pens, samples and times. This suggests these phages are persistent in the feedlots and represent variants of the dominant STEC O157:H7 viruses in their ecological niches.

Another major finding of this project was that, although five of the six phages share 99.9% pairwise genome similarity, they display variations in their host range, adsorption kinetics and virulence (as assessed by virulence index) against different STEC O157:H7 strains. Bioinformatics analysis of their genomes identified twelve nonsynonymous point mutations that differentiate these five phages. Notably, six out of these twelve mutations were located in putative tail fiber genes, responsible for host recognition and attachment. These mutations could thus potentially explain the different traits of the phages.

The reported findings of this thesis shed light on interesting aspects of phage host interactions, such as the relevance of a reversible binding step on the phage adsorption kinetics, and suggests relationships between point mutation in the phage genomes and their possible effects on host recognition and attachment – for example, slower adsorption of AKS66 and its single point

mutation on the gene encoding its tail fiber. By evaluating desirable characteristics relevant to different applications, phages AKS26 and AHS184 were suggested as good candidates for detection and biocontrol of STEC O157:H7, respectively.

6.2 Future works

The present study lays the groundwork for future research into the characterization of STEC O157:H7 phages and assessment of these viruses for different food safety applications.

Further characterization of the six phages presented in this study could be conducted prior to their use in food products. The analysis of host range could be expanded to other STEC serotypes and even other *Enterobacteria*, such as *Shigella* and *Salmonella*. This could identify phages with broader host range, potentially of interest for biocontrol applications, but demeaning to biosensors, for example. The identification of the phage receptors in the host membrane through the use of host knock-outs or cell binding assays could also help to elucidate mechanisms involved in phage-host interactions.

Functional genomics techniques such as structure prediction and identification of conserved domains could aid in the improvement of the phage genome annotation and identification of proteins of interest, such as RBPs. Proteomic studies aimed at characterizing the structural proteins and other hypothetical proteins could help assign functions to genes with currently undefined or putative functions. The elucidation of encoded protein functions also plays a role in guaranteeing the safety of the usage of the phages of interest in food safety applications, as it must be ensured phages do not encode for virulence factors, lysogenation or transduction.

The combination of the six phages of interest in phage cocktails with others of different host ranges can also be of interest in biocontrol applications, targeting relevant foodborne pathogens. It might also be of interest to adapt and amplify the phages in a non-pathogenic host strain to facilitate the scale-up of the

phage production process, facilitate their purification, and ensure the safety of phage cocktails.

A natural progression of this work is to introduce the identified point mutations into other phages to attempt to change the host specificity and adsorption kinetics of a given phage. In this way, it may be possible to control the host range or increase the adsorption rate of the phage through design and engineering. This would allow the tailoring of a phage towards specific applications against STEC O157:H7. Phage engineering techniques can also assist in the definition of protein functions by cloning of candidate genes into other phages or producing phage knock-outs to evaluate the effects of the addition or lack of specific proteins in the progeny phages.

Bibliography

Abedon, S. T. (2011) 'Lysis from without', *Bacteriophage*. Taylor & Francis, 1(1), pp. 46–49. doi: 10.4161/bact.1.1.13980.

Abedon, S. T. (2016) 'Phage therapy dosing: The problem(s) with multiplicity of infection (MOI).', *Bacteriophage*. Taylor & Francis, 6(3), p. e1220348. doi: 10.1080/21597081.2016.1220348.

Ackermann, H.-W. (1998a) 'Tailed Bacteriophages: The Order Caudovirales', *Advances in Virus Research*. Academic Press, 51, pp. 135–201. doi: 10.1016/S0065-3527(08)60785-X.

Ackermann, H.-W. (1998b) 'Tailed Bacteriophages: The Order Caudovirales', in, pp. 135–201. doi: 10.1016/S0065-3527(08)60785-X.

Ackermann, H.-W. (2003) 'Bacteriophage observations and evolution', *Research in Microbiology*. Elsevier Masson, 154(4), pp. 245–251. doi: 10.1016/S0923-2508(03)00067-6.

Ackermann, H.-W. (2007) '5500 Phages examined in the electron microscope', *Archives of Virology*. Springer-Verlag, 152(2), pp. 227–243. doi: 10.1007/s00705-006-0849-1.

Ackermann, H.-W. (2012) 'Bacteriophage Electron Microscopy', *Advances in Virus Research*. Academic Press, 82, pp. 1–32. doi: 10.1016/B978-0-12-394621-8.00017-0.

Adriaenssens, E. and Brister, J. R. (2017) 'How to Name and Classify Your Phage: An Informal Guide', *Viruses*. Multidisciplinary Digital Publishing Institute, 9(12), p. 70. doi: 10.3390/v9040070.

Adriaenssens, E. M. *et al.* (2017) 'Environmental drivers of viral community composition in Antarctic soils identified by viromics', *Microbiome*. BioMed Central, 5(1), p. 83. doi: 10.1186/s40168-017-0301-7.

Adriaenssens, E. M. *et al.* (2018) 'Taxonomy of prokaryotic viruses: 2017 update from the ICTV Bacterial and Archaeal Viruses Subcommittee', *Archives of Virology*. doi: 10.1007/s00705-018-3723-z.

Aiewsakun, P. *et al.* (2018) 'Evaluation of the genomic diversity of viruses infecting bacteria, archaea and eukaryotes using a common bioinformatic platform: steps towards a unified taxonomy', *Journal of General Virology*. Microbiology Society, 99(9), pp. 1331–1343. doi: 10.1099/jgv.0.001110.

Alberta Health Services (2018) *E. coli outbreak investigation expanded in Edmonton*. Available at: https://www.albertahealthservices.ca/news/Page14382.aspx (Accessed: 3 October 2018).

Allen, K. J. *et al.* (2011) 'Vaccination with type III secreted proteins leads to decreased shedding in calves after experimental infection with Escherichia coli O157.', *Canadian journal of veterinary research = Revue canadienne de recherche veterinaire*, 75(2), pp. 98–105. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21731179 (Accessed: 4 December 2018).

Altschul, S. *et al.* (1997) 'Gapped BLAST and PSI-BLAST: a new generation of protein database search programs', *Nucleic Acids Research*. Oxford University Press, 25(17), pp. 3389–3402. doi: 10.1093/nar/25.17.3389.

Altschul, S. F. *et al.* (1990) 'Basic local alignment search tool', *Journal of Molecular Biology*. Academic Press, 215(3), pp. 403–410. doi: 10.1016/S0022-2836(05)80360-2.

Amarillas, L., Cháidez-Quiroz, C., *et al.* (2013) 'Complete genome sequence of a polyvalent bacteriophage, phiKP26, active on Salmonella and Escherichia coli', *Archives of Virology*, 158(11), pp. 2395–2398. doi: 10.1007/s00705-013-1725-4.

Amarillas, L., Chaidez, C., *et al.* (2013) 'Complete genome sequence of a polyvalent bacteriophage, phiKP26, active on Salmonella and Escherichia coli', *Archives of Virology*. Springer Vienna, 158(11), pp. 2395–2398. doi: 10.1007/s00705-013-1725-4.

Amarillas, L., Chaidez, C., *et al.* (2013) 'Complete Genome Sequence of Escherichia coli O157:H7 Bacteriophage phiJLA23 Isolated in Mexico.', *Genome announcements*. American Society for Microbiology (ASM), 1(1). doi: 10.1128/genomeA.00219-12.

Ando, H. *et al.* (2015) 'Engineering Modular Viral Scaffolds for Targeted Bacterial Population Editing', *Cell Systems*. Cell Press, 1(3), pp. 187–196. doi: 10.1016/J.CELS.2015.08.013.

Baale, M. J. Van *et al.* (2004) 'Effect of Forage or Grain Diets with or without Monensin on Ruminal Persistence and Fecal Escherichia coli O157:H7 in Cattle', *Appl. Environ. Microbiol.* American Society for Microbiology, 70(9), pp. 5336–5342. doi: 10.1128/AEM.70.9.5336-5342.2004.

Bach, S. J. *et al.* (2002) 'Transmission and control of Escherichia coli O157:H7 — A review', *Canadian Journal of Animal Science*. NRC Research Press Ottawa, Canada, 82(4), pp. 475–490. doi: 10.4141/A02-021.

Bankevich, A. *et al.* (2012) 'SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing.', *Journal of computational biology : a journal of computational molecular cell biology*. Mary Ann Liebert, Inc., 19(5), pp. 455–77. doi: 10.1089/cmb.2012.0021.

Barrangou, R. *et al.* (2007) 'CRISPR provides acquired resistance against viruses in prokaryotes.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 315(5819), pp. 1709–12. doi: 10.1126/science.1138140.

Barrangou, R. and Marraffini, L. A. (2014) 'CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity', *Mol Cell*, 54(2), pp. 234–244. doi: 10.1016/j.molcel.2014.03.011.

Bartual, S. G. *et al.* (2010) 'Structure of the bacteriophage T4 long tail fiber receptor-binding tip.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 107(47), pp. 20287–92. doi: 10.1073/pnas.1011218107.

Berg Miller, M. E. *et al.* (2012) 'Phage-bacteria relationships and CRISPR elements revealed by a metagenomic survey of the rumen microbiome', *Environmental Microbiology*. Blackwell Publishing Ltd, 14(1), pp. 207–227. doi: 10.1111/j.1462-2920.2011.02593.x.

Berry, E. D. *et al.* (2006) 'Influence of genotype and diet on steer performance, manure odor, and carriage of pathogenic and other fecal bacteria. II. Pathogenic and other fecal bacteria1,2', *Journal of Animal Science*, 84(9), pp. 2523–2532. doi: 10.2527/jas.2005-747.

Bertozzi Silva, J., Storms, Z. and Sauvageau, D. (2016) 'Host receptors for bacteriophage adsorption', *FEMS Microbiology Letters*. doi: 10.1093/femsle/fnw002.

Besser, R. E., Griffin, P. M. and Slutsker, L. (1999) 'ESCHERICHIA COLI O157:H7 GASTROENTERITIS AND THE HEMOLYTIC UREMIC SYNDROME: An Emerging Infectious Disease', *Annual Review of Medicine*, 50(1), pp. 355–367. doi: 10.1146/annurev.med.50.1.355.

BIFSCO Beef Industry Food Safety Council (2013) 'Production Best Practices (PBP) to Aid in the Control of Foodborne Pathogens in Groups of Cattle'. Available at: https://www.bifsco.org/CMDocs/BIFSCO/Best Practices/Production Best Practices.pdf.

Blasco, R. *et al.* (1998) 'Specific assays for bacteria using phage mediated release of adenylate kinase.', *Journal of applied microbiology*, 84(4), pp. 661–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9633663 (Accessed: 11 December 2018).

Bobay, L., Touchon, M. and Rocha, E. P. (2014) 'Pervasive domestication of defective prophages by bacteria', *Proceedings of the National Academy of Sciences of the United States of America*, 111(33), pp. 12127–12132. doi: https://doi.org/10.1073/pnas.1405336111.

Bondy-Denomy, J. *et al.* (2015) 'Multiple mechanisms for CRISPR–Cas inhibition by anti-CRISPR proteins', *Nature*. Nature Publishing Group, 526(7571), pp. 136–139. doi: 10.1038/nature15254.

Botstein, D. *et al.* (1975) 'Role of antirepressor in the bipartite control of repression and immunity by bacteriophage P22', *Journal of Molecular Biology*, 91(4), pp. 439–462. doi: 10.1016/0022-2836(75)90271-5.

Bradley, D. E. and Kay, D. (1960) 'The Fine Structure of Bacteriophages', *Journal of General Microbiology*. Microbiology Society, 23(3), pp. 553–563. doi: 10.1099/00221287-23-3-553.

Breitbart, M. *et al.* (2018) 'Phage puppet masters of the marine microbial realm', *Nature Microbiology*. Nature Publishing Group, 3(7), pp. 754–766. doi: 10.1038/s41564-018-0166-y.

Briones, M. A. (2013) 'Hemolytic Uremic Syndrome', *Transfusion Medicine and Hemostasis*. Elsevier, pp. 669–674. doi: 10.1016/B978-0-12-397164-7.00102-6.

Brown, P. *et al.* (2018) 'Larger than life: Isolation and genomic characterization of a jumbo phage that infects the bacterial plant pathogen, Agrobacterium tumefaciens', *Frontiers in Microbiology*. Frontiers, 9, p. 1861. doi: 10.3389/FMICB.2018.01861.

Buchko, S. J. *et al.* (2000) 'The effect of different grain diets on fecal shedding of Escherichia coli O157:H7 by steers.', *Journal of food protection*, 63(11), pp. 1467–74. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11079685 (Accessed: 4 December 2018).

Buchko, S. J. *et al.* (2000) 'The effect of fasting and diet on fecal shedding of Escherichia coli O157:H7 by cattle', *Canadian Journal of Animal Science*. NRC Research Press Ottawa, Canada, 80(4), pp. 741–744. doi: 10.4141/A00-025.

Callaway, E. (2017) 'Do you speak virus? Phages caught sending chemical messages', *Nature*. doi: 10.1038/nature.2017.21313.

Callaway, T. R. *et al.* (2009) 'Diet, Escherichia coli O157:H7, and cattle: a review

after 10 years.', *Current issues in molecular biology*, 11(2), pp. 67–79. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19351974 (Accessed: 5 December 2018).

CDC, Centers for Disease Control and Prevention (2018) *Multistate Outbreak of E. coli O157:H7 Infections Linked to Romaine Lettuce (Final Update) | Investigation Notice: Multistate Outbreak of E. coli O157:H7 Infections April 2018*. Available at: https://www.cdc.gov/ecoli/2018/o157h7-04-18/index.html (Accessed: 3 August 2018).

CDC, Centers for Disease Control and Prevention (2018) *Outbreak of E. coli Infections Linked to Romaine Lettuce | E. coli Infections Linked to Romaine Lettuce November 2018*. Available at: https://www.cdc.gov/ecoli/2018/o157h7-11-18/index.html (Accessed: 3 December 2018).

Chen, J. *et al.* (2018) 'Genome hypermobility by lateral transduction.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 362(6411), pp. 207–212. doi: 10.1126/science.aat5867.

Chibani-Chennoufi, S. *et al.* (2004) 'Lactobacillus plantarum bacteriophage LP65: a new member of the SPO1-like genus of the family Myoviridae.', *Journal of bacteriology*. American Society for Microbiology Journals, 186(21), pp. 7069–83. doi: 10.1128/JB.186.21.7069-7083.2004.

Costa, A. R. *et al.* (2018) 'Synthetic Biology to Engineer Bacteriophage Genomes', in *Bacteriophage Therapy: From Lab to Clinical Practice*. Springer Science+Business Media, pp. 285–300. doi: 10.1007/978-1-4939-7395-8_21.

Coutinho, F. H. *et al.* (2017) 'Marine viruses discovered via metagenomics shed light on viral strategies throughout the oceans', *Nature Communications*. Nature Publishing Group, 8, p. 15955. doi: 10.1038/ncomms15955.

Cull, C. A. *et al.* (2012) 'Efficacy of a vaccine and a direct-fed microbial against fecal shedding of Escherichia coli O157:H7 in a randomized pen-level field trial of commercial feedlot cattle', *Vaccine*. Elsevier, 30(43), pp. 6210–6215. doi: 10.1016/J.VACCINE.2012.05.080.

D'Herelle, F. (2007) 'On an invisible microbe antagonistic toward dysenteric bacilli: brief note by Mr. F. D'Herelle, presented by Mr. Roux', *Research in Microbiology*. Elsevier Masson, 158(7), pp. 553–554. doi: 10.1016/J.RESMIC.2007.07.005.

Darling, A. E., Mau, B. and Perna, N. T. (2010) 'progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement', *PLoS ONE*. Edited by J. E. Stajich. Public Library of Science, 5(6), p. e11147. doi: 10.1371/journal.pone.0011147.

Davison, A. R. (2018) 'A common trick for transferring bacterial DNA', *Science*, 362(6411), pp. 152–153. doi: 10.1126/science.aav1723.

Delbrück, M. (1970) 'A Physicist's Renewed Look at Biology: Twenty Years Later', *Science*. American Association for the Advancement of Science, 168(3937), pp. 1312–1315. doi: DOI: 10.1126/science.168.3937.1312.

Dutilh, B. E. *et al.* (2014) 'A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes', *Nature Communications*. Nature Publishing Group, 5(1), p. 4498. doi: 10.1038/ncomms5498.

Edgar, R. *et al.* (2012) 'Reversing bacterial resistance to antibiotics by phage-mediated delivery of dominant sensitive genes', *Applied and environmental microbiology*. American Society for Microbiology, 78(3), pp. 744–51. doi:

10.1128/AEM.05741-11.

Eliava Phage Therapy Center (2017) *Price List - Rates for all phage therapy treatments*. Available at: https://eliavaphagetherapy.com/price-list/ (Accessed: 20 October 2018).

Erez, Z. *et al.* (2017) 'Communication between viruses guides lysis–lysogeny decisions', *Nature*. Nature Publishing Group, 541(7638), pp. 488–493. doi: 10.1038/nature21049.

Fenton, M. *et al.* (2013) 'Bacteriophage-Derived Peptidase CHAP K Eliminates and Prevents Staphylococcal Biofilms', *International Journal of Microbiology*, 2013, pp. 1–8. doi: 10.1155/2013/625341.

Foulongne, V. *et al.* (2012) 'Human Skin Microbiota: High Diversity of DNA Viruses Identified on the Human Skin by High Throughput Sequencing', *PLoS ONE*. Edited by A. E. Toland. Public Library of Science, 7(6), p. e38499. doi: 10.1371/journal.pone.0038499.

Frank, C. *et al.* (2011) 'Epidemic Profile of Shiga-Toxin–Producing Escherichia coli O104:H4 Outbreak in Germany', *New England Journal of Medicine*. Massachusetts Medical Society, 365(19), pp. 1771–1780. doi: 10.1056/NEJMoa1106483.

Furfaro, L. L., Payne, M. S. and Chang, B. J. (2018) 'Bacteriophage Therapy: Clinical Trials and Regulatory Hurdles', *Frontiers in Cellular and Infection Microbiology*. Frontiers, 8, p. 376. doi: 10.3389/fcimb.2018.00376.

Garen, A. and Puck, T. T. (1951) 'The first two steps of the invasion of host cells by bacterial viruses', *The Journal of experimental medicine*. Rockefeller University Press, 94(3), pp. 177–89. doi: 10.1084/JEM.94.3.177.

Gilbert, R. A. and Klieve, A. V. (2015) 'Ruminal Viruses (Bacteriophages, Archaeaphages)', in *Rumen Microbiology: From Evolution to Revolution*. New Delhi: Springer India, pp. 121–141. doi: 10.1007/978-81-322-2401-3_9.

Hagens, S. and Loessner, M. (2010) 'Bacteriophage for Biocontrol of Foodborne Pathogens: Calculations and Considerations', *Current Pharmaceutical Biotechnology*, 11(1), pp. 58–68. doi: 10.2174/138920110790725429.

Hall, C. E. (1955) 'Electron densitometry of stained virus particles.', *The Journal of biophysical and biochemical cytology*. Rockefeller University Press, 1(1), pp. 1–12. doi: 10.1083/JCB.1.1.1.

Hallewell, J. *et al.* (2014) 'Differing populations of endemic bacteriophages in cattle shedding high and low numbers of Escherichia coli O157:H7 bacteria in feces.', *Applied and environmental microbiology*. American Society for Microbiology (ASM), 80(13), pp. 3819–25. doi: 10.1128/AEM.00708-14.

Hatfull, G. F. (2015) 'Dark Matter of the Biosphere: the Amazing World of Bacteriophage Diversity.', *Journal of virology*. American Society for Microbiology, 89(16), pp. 8107–10. doi: 10.1128/JVI.01340-15.

Hayashi, T. *et al.* (2001) 'Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12.', *DNA research : an international journal for rapid publication of reports on genes and genomes*, 8(1), pp. 11–22. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11258796 (Accessed: 15 November 2018).

Herold, S., Karch, H. and Schmidt, H. (2004) 'Shiga toxin-encoding bacteriophages – genomes in motion', *International Journal of Medical Microbiology*. Urban & Fischer, 294(2–3), pp. 115–121. doi: 10.1016/J.IJMM.2004.06.023.

Herschleb, J., Ananiev, G. and Schwartz, D. C. (2007) 'Pulsed-field gel electrophoresis'. doi: 10.1038/nprot.2007.94.

Hershey, A. D. and Chase, M. (1952) 'Independent functions of viral protein and nucleic acid in growth of bacteriophage.', *The Journal of general physiology*. Rockefeller University Press, 36(1), pp. 39–56. doi: 10.1085/JGP.36.1.39.

Honish, L. *et al.* (2017) 'Escherichia coli O157:H7 infections associated with contaminated pork products Alberta, Canada, July–October 2014', *Canada Communicable Disease Report (CCDR)*, 43. Available at: http://www.phac-aspc.gc.ca/publicat/ccdr-rmtc/17vol43/dr-rm43-1/assets/pdf/17vol43_1-ar-04-eng.pdf (Accessed: 30 June 2017).

Hoogenraad, N. J. *et al.* (1967) 'Bacteriophages in Rumen Contents of Sheep', *Journal of General Virology*. Microbiology Society, 1(4), pp. 575–576. doi: 10.1099/0022-1317-1-4-575.

Howard-Varona, C. *et al.* (2017) 'Lysogeny in nature: mechanisms, impact and ecology of temperate phages', *The ISME Journal*. Nature Publishing Group, 11(7), pp. 1511–1520. doi: 10.1038/ismej.2017.16.

Hsu, P. D., Lander, E. S. and Zhang, F. (2014) 'Development and applications of CRISPR-Cas9 for genome engineering', *Cell*, 157(6), pp. 1262–1278.

International Committee on Taxonomy of Viruses (ICTV) (2017a) *10th (Online) ICTV Report - Virus Taxonomy: The Classification and Nomenclature of Viruses*. Available at: https://talk.ictvonline.org/ictv-reports/ictv_online_report/ (Accessed: 8 September 2018).

International Committee on Taxonomy of Viruses (ICTV) (2017b) *Master Species List 2017 v1.0*. Available at: https://talk.ictvonline.org/files/master-species-lists/m/msl/7185 (Accessed: 8 September 2018).

Janjarasskul, T. and Suppakul, P. (2018) 'Active and intelligent packaging: The indication of quality and safety', *Critical Reviews in Food Science and Nutrition*. Taylor & Francis, 58(5), pp. 808–831. doi: 10.1080/10408398.2016.1225278.

Jaschke, P. R. *et al.* (2012) 'A fully decompressed synthetic bacteriophage øX174 genome assembled and archived in yeast', *Virology*. Academic Press, 434(2), pp. 278–284. doi: 10.1016/J.VIROL.2012.09.020.

Jault, P. *et al.* (2018) 'Efficacy and tolerability of a cocktail of bacteriophages to treat burn wounds infected by Pseudomonas aeruginosa (PhagoBurn): a randomised, controlled, double-blind phase 1/2 trial.', *The Lancet. Infectious diseases*. Elsevier, 0(0). doi: 10.1016/S1473-3099(18)30482-1.

Jinek, M. *et al.* (2012) 'A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity', *Science*, 337(6096), pp. 816–821. doi: 10.1126/science.1225829.

de Jonge, P. A. *et al.* (2019) 'Molecular and Evolutionary Determinants of Bacteriophage Host Range', *Trends in Microbiology*, 27(1), pp. 51–63. doi: 10.1016/j.tim.2018.08.006.

Juretić, D., Zoranić, L. and Zucić, D. (2002) 'Basic charge clusters and predictions of membrane protein topology.', *Journal of chemical information and computer sciences*, 42(3), pp. 620–32. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12086524 (Accessed: 17 May 2018).

Kall, L., Krogh, A. and Sonnhammer, E. L. L. (2007) 'Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server', *Nucleic Acids Research*. Oxford University Press, 35(Web Server), pp. W429–W432. doi: 10.1093/nar/gkm256.

Katsura, I. (1990) 'Mechanism of length determination in bacteriophage lambda tails.', *Advances in biophysics*, 26, pp. 1–18. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2150582 (Accessed: 6 January 2019).

Katsura, I. and Hendrix, R. W. (1984) 'Length determination in bacteriophage lambda tails.', *Cell*. Elsevier, 39(3 Pt 2), pp. 691–8. doi: 10.1016/0092-8674(84)90476-8.

Khakhria, R., Duck, D. and Lior, H. (1990) 'Extended phage-typing scheme for Escherichia coli O157:H7.', *Epidemiology and infection*, 105(3), pp. 511–20. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2249715 (Accessed: 7 December 2018).

Kilcher, S. *et al.* (2018) 'Cross-genus rebooting of custom-made, synthetic bacteriophage genomes in L-form bacteria.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 115(3), pp. 567–572. doi: 10.1073/pnas.1714658115.

Kingsford, C. L., Ayanbule, K. and Salzberg, S. L. (2007) 'Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake', *Genome Biology*. BioMed Central, 8(2), p. R22. doi: 10.1186/gb-2007-8-2-r22.

Kingwell, K. (2015) 'Bacteriophage therapies re-enter clinical trials', *Nature Reviews Drug Discovery*, 14(8), pp. 515–516. doi: 10.1038/nrd4695.

Klieve, A. V and Bauchop, T. (1988) 'Morphological diversity of ruminal bacteriophages from sheep and cattle.', *Applied and environmental microbiology*. American Society for Microbiology (ASM), 54(6), pp. 1637–41. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3415230 (Accessed: 6 December 2018).

Klumpp, J., Fouts, D. E. and Sozhamannan, S. (2013) 'Bacteriophage functional genomics and its role in bacterial pathogen detection', *Briefings in Functional Genomics*. Oxford University Press, 12(4), pp. 354–365. doi: 10.1093/bfgp/elt009.

Koonin, E. and Galperin, M. (2003) 'Genomics: From Phage to Human', in *Sequence - Evolution - Function: Computational Approaches in Comparative Genomics*. Boston, MA, USA: Kluwer Academic. Available at: https://www.ncbi.nlm.nih.gov/books/NBK20263/.

Kretzer, J. W. *et al.* (2007) 'Use of high-affinity cell wall-binding domains of bacteriophage endolysins for immobilization and separation of bacterial cells.', *Applied and environmental microbiology*. American Society for Microbiology (ASM), 73(6), pp. 1992–2000. doi: 10.1128/AEM.02402-06.

Krogh, A. *et al.* (2001) 'Predicting transmembrane protein topology with a hidden markov model: application to complete genomes', *Journal of Molecular Biology*. Academic Press, 305(3), pp. 567–580. doi: 10.1006/JMBI.2000.4315.

Kropinski, A. M. *et al.* (2012) 'Endemic bacteriophages: a cautionary tale for

evaluation of bacteriophage therapy and other interventions for infection control in animals', *Virology Journal*. BioMed Central, 9(1), p. 207. doi: 10.1186/1743-422X-9-207.

Kutateladze, M. and Adamia, R. (2008) 'Phage therapy experience at the Eliava Institute', *Médecine et Maladies Infectieuses*. Elsevier Masson, 38(8), pp. 426–430. doi: 10.1016/J.MEDMAL.2008.06.023.

Labrie, S. J., Samson, J. E. and Moineau, S. (2010) 'Bacteriophage resistance mechanisms', *Nature Reviews Microbiology*. Nature Publishing Group, 8(5), pp. 317–327. doi: 10.1038/nrmicro2315.

Laslett, D. and Canback, B. (2004) 'ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences', *Nucleic Acids Research*. Oxford University Press, 32(1), pp. 11–16. doi: 10.1093/nar/gkh152.

Launders, N. *et al.* (2016) 'Disease severity of Shiga toxin-producing E. coli O157 and factors influencing the development of typical haemolytic uraemic syndrome: a retrospective cohort study, 2009-2012.', *BMJ open*. British Medical Journal Publishing Group, 6(1), p. e009933. doi: 10.1136/bmjopen-2015-009933.

Le, S. *et al.* (2013) 'Mapping the Tail Fiber as the Receptor Binding Protein Responsible for Differential Host Specificity of Pseudomonas aeruginosa Bacteriophages PaP1 and JG004', *PLoS ONE*. Edited by M. J. van Raaij. Public Library of Science, 8(7), p. e68562. doi: 10.1371/journal.pone.0068562.

Lee, H. *et al.* (2016) 'Characterization and Genomic Study of the Novel Bacteriophage HY01 Infecting Both Escherichia coli O157:H7 and Shigella flexneri: Potential as a Biocontrol Agent in Food', *PLOS ONE*. Edited by M. J. van Raaij, 11(12), p. e0168985. doi: 10.1371/journal.pone.0168985.

Li, L. *et al.* (2010) 'Bat guano virome: predominance of dietary viruses from insects and plants plus novel mammalian viruses.', *Journal of virology*. American Society for Microbiology Journals, 84(14), pp. 6955–65. doi: 10.1128/JVI.00501-10.

Lingohr, E., Frost, S. and Johnson, R. P. (2009) 'Determination of Bacteriophage Genome Size by Pulsed-Field Gel Electrophoresis', in *Bacteriophages. Methods in Molecular Biology*. Humana Press, pp. 19–25. doi: 10.1007/978-1-60327-565-1_3.

Loessner, M. J. *et al.* (1996) 'Construction of luciferase reporter bacteriophage A511::luxAB for rapid and sensitive detection of viable Listeria cells.', *Applied and environmental microbiology*. American Society for Microbiology, 62(4), pp. 1133–40. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8919773 (Accessed: 15 November 2018).

Loessner, M. J., Rudolf, M. and Scherer, S. (1997) 'Evaluation of luciferase reporter bacteriophage A511::luxAB for detection of Listeria monocytogenes in contaminated foods.', *Applied and environmental microbiology*. American Society for Microbiology (ASM), 63(8), pp. 2961–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9251182 (Accessed: 15 November 2018).

Lood, R. *et al.* (2014) 'A highly active and negatively charged Streptococcus pyogenes lysin with a rare D-alanyl-L-alanine endopeptidase activity protects mice against streptococcal bacteremia.', *Antimicrobial agents and chemotherapy*. American Society for Microbiology Journals, 58(6), pp. 3073–84. doi: 10.1128/AAC.00115-14.

Lowe, T. M. and Chan, P. P. (2016) 'tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes', *Nucleic Acids Research*, 44(W1), pp.

W54–W57. doi: 10.1093/nar/gkw413.

Lu, T. K. *et al.* (2012) 'Recombinant phage and methods'. USPTO. Available at: https://patents.google.com/patent/US20130122549A1/en (Accessed: 15 November 2018).

Lu, T. K. and Collins, J. J. (2009) 'Engineered bacteriophage targeting gene networks as adjuvants for antibiotic therapy.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 106(12), pp. 4629–34. doi: 10.1073/pnas.0800442106.

Lu, Z. and Breidt, F. (2015) 'Escherichia coli O157:H7 bacteriophage Φ241 isolated from an industrial cucumber fermentation at high acidity and salinity.', *Frontiers in microbiology*. Frontiers Media SA, 6, p. 67. doi: 10.3389/fmicb.2015.00067.

Mahichi, F. *et al.* (2009) 'Site-specific recombination of T2 phage using IP008 long tail fiber genes provides a targeted method for expanding host range while retaining lytic activity', *FEMS Microbiology Letters*. Oxford University Press, 295(2), pp. 211–217. doi: 10.1111/j.1574-6968.2009.01588.x.

Mahony, J. *et al.* (2011) 'Bacteriophages as biocontrol agents of food pathogens.', *Current opinion in biotechnology*, 22(2), pp. 157–63. doi: 10.1016/j.copbio.2010.10.008.

Manrique, P. *et al.* (2016) 'Healthy human gut phageome.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 113(37), pp. 10400–5. doi: 10.1073/pnas.1601060113.

Mardis, E. R. (2008) 'Next-generation DNA sequencing methods', in *Annual Review of Genomics and Human Genetics*. Palo Alto: Annual Reviews, pp. 387–402. doi: 10.1146/annurev.genom.9.081307.164359.

Matthews, L. *et al.* (2006) 'Super-shedding cattle and the transmission dynamics of Escherichia coli O157.', *Epidemiology and infection*. Cambridge University Press, 134(1), pp. 131–42. doi: 10.1017/S0950268805004590.

Meier-Kolthoff, J. P. and Göker, M. (2017) 'VICTOR: genome-based phylogeny and classification of prokaryotic viruses', *Bioinformatics*. Oxford University Press, 33(21), pp. 3396–3404. doi: 10.1093/bioinformatics/btx440.

Merabishvili, M. *et al.* (2012) 'Selection and Characterization of a Candidate Therapeutic Bacteriophage That Lyses the Escherichia coli O104:H4 Strain from the 2011 Outbreak in Germany', *PLoS ONE*. Edited by D. M. Ojcius. Public Library of Science, 7(12), p. e52709. doi: 10.1371/journal.pone.0052709.

Mills, S. *et al.* (2013) 'Movers and shakers: influence of bacteriophages in shaping the mammalian gut microbiota.', *Gut microbes*. Taylor & Francis, 4(1), pp. 4–16. doi: 10.4161/gmic.22371.

Molineux, I. J. and Panja, D. (2013) 'Popping the cork: mechanisms of phage genome ejection', *Nature Reviews Microbiology*. Nature Publishing Group, 11(3), pp. 194–204. doi: 10.1038/nrmicro2988.

Mora, A. *et al.* (2004) 'Phage types and genotypes of shiga toxin-producing Escherichia coli O157:H7 isolates from humans and animals in spain: identification and characterization of two predominating phage types (PT2 and PT8).', *Journal of clinical microbiology*. American Society for Microbiology Journals, 42(9), pp. 4007–15. doi: 10.1128/JCM.42.9.4007-4015.2004.

Morita, M. *et al.* (2002) 'Characterization of a virulent bacteriophage specific for *Escherichia coli* O157:H7 and analysis of its cellular receptor and two tail fiber genes', *FEMS Microbiology Letters*. Oxford University Press, 211(1), pp. 77–83. doi: 10.1111/j.1574-6968.2002.tb11206.x.

Munns, K. D. *et al.* (2016) 'Comparative Genomic Analysis of Escherichia coli O157:H7 Isolated from Super-Shedder and Low-Shedder Cattle.', *PloS one*, 11(3), p. e0151673. doi: 10.1371/journal.pone.0151673.

NCBI - National Center for Biotechnology Information, U. S. N. L. of M. (no date) *GenBank and WGS Statistics*. Available at: https://www.ncbi.nlm.nih.gov/genbank/statistics/ (Accessed: 13 October 2018).

Nelson, D. (2004) 'Phage taxonomy: we agree to disagree.', *Journal of bacteriology*. American Society for Microbiology Journals, 186(21), pp. 7029–31. doi: 10.1128/JB.186.21.7029-7031.2004.

Niu, Y. D. *et al.* (2009a) 'Host range and lytic capability of four bacteriophages against bovine and clinical human isolates of Shiga toxin-producing Escherichia coli O157:H7', *Journal of Applied Microbiology*. Wiley/Blackwell (10.1111), 107(2), pp. 646–656. doi: 10.1111/j.1365-2672.2009.04231.x.

Niu, Y. D. *et al.* (2009b) 'Prevalence and impact of bacteriophages on the presence of Escherichia coli O157:H7 in feedlot cattle and their environment.', *Applied and environmental microbiology*. American Society for Microbiology, 75(5), pp. 1271–8. doi: 10.1128/AEM.02100-08.

Niu, Y. D. *et al.* (2012) 'Genomic, Proteomic and Physiological Characterization of a T5-like Bacteriophage for Control of Shiga Toxin-Producing Escherichia coli O157:H7', *PLoS ONE*. Edited by A. R. Poteete. Public Library of Science, 7(4), p. e34585. doi: 10.1371/journal.pone.0034585.

Niu, Y. D. *et al.* (2014) 'Four Escherichia coli O157:H7 Phages: A New Bacteriophage Genus and Taxonomic Classification of T1-Like Phages', *PLoS ONE*. Edited by M. J. van Raaij. Public Library of Science, 9(6), p. e100426. doi: 10.1371/journal.pone.0100426.

Nobrega, F. L. *et al.* (2018) 'Targeting mechanisms of tailed bacteriophages', *Nature Reviews Microbiology*. Nature Publishing Group, p. 1. doi: 10.1038/s41579-018-0070-8.

Norrby, E. (2008) 'Nobel Prizes and the emerging virus concept', *Archives of Virology*. Springer Vienna, 153(6), pp. 1109–1123. doi: 10.1007/s00705-008-0088-8.

Oda, M. *et al.* (2004) 'Rapid detection of Escherichia coli O157:H7 by using green fluorescent protein-labeled PP01 bacteriophage.', *Applied and environmental microbiology*. American Society for Microbiology (ASM), 70(1), pp. 527–34. doi: 10.1128/AEM.70.1.527-534.2004.

Oppenheim, A. B. *et al.* (2005) 'Switches in Bacteriophage Lambda Development', *Annual Review of Genetics*, 39(1), pp. 409–429. doi: 10.1146/annurev.genet.39.073003.113656.

De Paepe, M. *et al.* (2014) 'Bacteriophages: an underestimated role in human and animal health?', *Frontiers in Cellular and Infection Microbiology*. Frontiers, 4, p. 39. doi: 10.3389/fcimb.2014.00039.

PHAC, Public Health Agency of Canada (2018a) *Public Health Notice - Outbreak*

of E. coli infections linked to romaine lettuce. Available at: https://www.canada.ca/en/public-health/services/public-health-notices/2018/outbreak-ecoli-infections-linked-romaine-lettuce.html (Accessed: 3 December 2018).

PHAC, Public Health Agency of Canada (2018b) *Public Health Notice – Outbreak of E. coli infections linked to romaine lettuce*. Available at: https://www.canada.ca/en/public-health/services/public-health-notices/2018/public-health-notice-outbreak-e-coli-infections-linked-romaine-lettuce.html (Accessed: 3 August 2018).

Philipson, C. W. *et al.* (2018) 'Characterizing Phage Genomes for Therapeutic Applications.', *Viruses*. Multidisciplinary Digital Publishing Institute (MDPI), 10(4). doi: 10.3390/v10040188.

Pires, D. P. *et al.* (2016a) 'Genetically Engineered Phages: a Review of Advances over the Last Decade.', *Microbiology and molecular biology reviews : MMBR*. American Society for Microbiology, 80(3), pp. 523–43. doi: 10.1128/MMBR.00069-15.

Pires, D. P. *et al.* (2016b) 'Genetically Engineered Phages: a Review of Advances over the Last Decade.', *Microbiology and molecular biology reviews : MMBR*. American Society for Microbiology, 80(3), pp. 523–43. doi: 10.1128/MMBR.00069-15.

Pirnay, J.-P. *et al.* (2018) 'Bacteriophage Production in Compliance with Regulatory Requirements', in *Bacteriophage Therapy*. New York: Humana Press, pp. 233–252. doi: 10.1007/978-1-4939-7395-8_18.

Proux, C. *et al.* (2002) 'The dilemma of phage taxonomy illustrated by comparative genomics of Sfi21-like Siphoviridae in lactic acid bacteria.', *Journal of bacteriology*. American Society for Microbiology (ASM), 184(21), pp. 6026–36. doi: 10.1128/JB.184.21.6026-6036.2002.

Rambaut, A. (2006) *FigTree 1.4.3 - a graphical viewer of phylogenetic trees and a program for producing publication-ready figures*. Available at: http://tree.bio.ed.ac.uk/software/figtree/.

Rangel, J. M. *et al.* (2005) 'Epidemiology of Escherichia coli O157:H7 Outbreaks, United States, 1982–2002', *Emerging Infectious Diseases*, 11(4), pp. 603–609. doi: 10.3201/eid1104.040739.

Reese, M. G. (2001) 'Application of a time-delay neural network to promoter annotation in the Drosophila melanogaster genome', *Computers & Chemistry*. Pergamon, 26(1), pp. 51–56. doi: 10.1016/S0097-8485(01)00099-7.

Reyes, A. *et al.* (2012) 'Going viral: next-generation sequencing applied to phage populations in the human gut', *Nature Reviews Microbiology*. Nature Publishing Group, 10(9), pp. 607–617. doi: 10.1038/nrmicro2853.

Rifat, D. *et al.* (2008) 'Restriction Endonuclease Inhibitor IPI* of Bacteriophage T4: A Novel Structure for a Dedicated Target', *Journal of Molecular Biology*, 375(3), pp. 720–734. doi: 10.1016/j.jmb.2007.10.064.

Roach, D. R. and Donovan, D. M. (2015) 'Antimicrobial bacteriophage-derived proteins and therapeutic applications', *Bacteriophage*. Taylor & Francis, 5(3), p. e1062590. doi: 10.1080/21597081.2015.1062590.

Roberts, M. D., Martin, N. L. and Kropinski, A. M. (2004) 'The genome and proteome of coliphage T1', *Virology*. Academic Press, 318(1), pp. 245–266. doi: 10.1016/j.virol.2003.09.020.

Rodríguez-Rubio, L. *et al.* (2015) 'Phage lytic proteins: biotechnological applications beyond clinical antimicrobials', *Critical Reviews in Biotechnology*. Taylor & Francis, pp. 1–11. doi: 10.3109/07388551.2014.993587.

Rohwer, F. and Edwards, R. (2002) 'The Phage Proteomic Tree: a genome-based taxonomy for phage.', *Journal of bacteriology*. American Society for Microbiology Journals, 184(16), pp. 4529–35. doi: 10.1128/JB.184.16.4529-4535.2002.

Ross, A., Ward, S. and Hyman, P. (2016) 'More Is Better: Selecting for Broad Host Range Bacteriophages', *Frontiers in Microbiology*. Frontiers, 7, p. 1352. doi: 10.3389/fmicb.2016.01352.

Salmond, G. P. C. and Fineran, P. C. (2015a) 'A century of the phage: past, present and future', *Nature Reviews Microbiology*. Nature Publishing Group, 13(12), pp. 777–786. doi: 10.1038/nrmicro3564.

Salmond, G. P. C. and Fineran, P. C. (2015b) 'A century of the phage: past, present and future', *Nature Reviews Microbiology*, 13(12), pp. 777–786. doi: 10.1038/nrmicro3564.

Sambrook, J. and Russell, D. W. (2001) *Molecular cloning: a laboratory manual*. 3rd edn. UK: Coldspring-Harbour Laboratory Press.

Samson, J. E. *et al.* (2013) 'Revenge of the phages: defeating bacterial defences', *Nature Reviews Microbiology*. Nature Publishing Group, 11(10), pp. 675–687. doi: 10.1038/nrmicro3096.

Sanger, F. *et al.* (1982) 'Nucleotide sequence of bacteriophage lambda DNA.', *Journal of molecular biology*, 162(4), pp. 729–73. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6221115 (Accessed: 13 November 2018).

Sanger, F., Nicklen, S. and Coulson, A. R. (1977) 'DNA SEQUENCING WITH CHAIN-TERMINATING INHIBITORS', *Proceedings of the National Academy of Sciences of the United States of America*, 74(12), pp. 5463–5467. doi: 10.1073/pnas.74.12.5463.

Sarker, S. A. *et al.* (2016) 'Oral Phage Therapy of Acute Bacterial Diarrhea With Two Coliphage Preparations: A Randomized Trial in Children From Bangladesh', *EBioMedicine*. Elsevier, 4, pp. 124–137. doi: 10.1016/J.EBIOM.2015.12.023.

Schmelcher, M. *et al.* (2010) 'Rapid multiplex detection and differentiation of Listeria cells by use of fluorescent phage endolysin cell wall binding domains.', *Applied and environmental microbiology*. American Society for Microbiology (ASM), 76(17), pp. 5745–56. doi: 10.1128/AEM.00801-10.

Schmelcher, M. and Loessner, M. J. (2014) 'Application of bacteriophages for detection of foodborne pathogens.', *Bacteriophage*. Taylor & Francis, 4(1), p. e28137. doi: 10.4161/bact.28137.

Schwarzer, D. *et al.* (2012) 'A multivalent adsorption apparatus explains the broad host range of phage phi92: a comprehensive genomic and structural analysis.', *Journal of virology*. American Society for Microbiology Journals, 86(19), pp. 10384–98. doi: 10.1128/JVI.00801-12.

Shan, T. *et al.* (2011) 'The fecal virome of pigs on a high-density farm.', *Journal of virology*. American Society for Microbiology Journals, 85(22), pp. 11697–708. doi: 10.1128/JVI.05217-11.

Shendure, J. *et al.* (2017) 'DNA sequencing at 40: past, present and future', *Nature*. Nature Publishing Group, 550(7676), pp. 345–353. doi: 10.1038/nature24286.

Shterzer, N. and Mizrahi, I. (2015) 'The animal gut as a melting pot for horizontal gene transfer', *Canadian Journal of Microbiology*. NRC Research Press, 61(9), pp. 603–605. doi: 10.1139/cjm-2015-0049.

Sillankorva, S. M., Oliveira, H. and Azeredo, J. (2012) 'Bacteriophages and their role in food safety.', *International journal of microbiology*. Hindawi, 2012, p. 863945. doi: 10.1155/2012/863945.

Smith, D. R. *et al.* (2009) 'A Randomized Longitudinal Trial to Test the Effect of Regional Vaccination Within a Cattle Feedyard on *Escherichia coli* O157:H7 Rectal Colonization, Fecal Shedding, and Hide Contamination', *Foodborne Pathogens and Disease*, 6(7), pp. 885–892. doi: 10.1089/fpd.2009.0299.

Solanki, K. *et al.* (2013) 'Enzyme-based listericidal nanocomposites.', *Scientific reports*. Nature Publishing Group, 3, p. 1584. doi: 10.1038/srep01584.

Stephens, T. P., McAllister, T. A. and Stanford, K. (2009) 'Perineal swabs reveal effect of super shedders on the transmission of Escherichia coli O157:H7 in commercial feedlots1', *Journal of Animal Science*, 87(12), pp. 4151–4160. doi: 10.2527/jas.2009-1967.

Stewart, G. S. *et al.* (1998) 'The specific and sensitive detection of bacterial pathogens within 4 h using bacteriophage amplification.', *Journal of applied microbiology*, 84(5), pp. 777–83. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9674131 (Accessed: 11 December 2018).

Storms, Z. J. *et al.* (2010) 'Bacteriophage adsorption efficiency and its effect on amplification', *Bioprocess and Biosystems Engineering*. Springer-Verlag, 33(7), pp. 823–831. doi: 10.1007/s00449-009-0405-y.

Storms, Z. J. *et al.* (2012) 'Modeling bacteriophage attachment using adsorption efficiency', *Biochemical Engineering Journal*, 64, pp. 22–29. doi: 10.1016/j.bej.2012.02.007.

Storms, Z. J. *et al.* (2019) 'The virulence index: A quantitative measurement of bacteriophage virulence', *In preparation for Frontiers in Microbiology*.

Swaminathan, B. *et al.* (2001) 'PulseNet: the molecular subtyping network for foodborne bacterial disease surveillance, United States.', *Emerging infectious diseases*. Centers for Disease Control and Prevention, 7(3), pp. 382–9. doi: 10.3201/eid0703.010303.

Touchon, M., Moura de Sousa, J. A. and Rocha, E. P. (2017) 'Embracing the enemy: the diversification of microbial gene repertoires by phage-mediated horizontal gene transfer', *Current Opinion in Microbiology*. Elsevier Current Trends, 38, pp. 66–73. doi: 10.1016/J.MIB.2017.04.010.

Twort, F. W. (1915) 'AN INVESTIGATION ON THE NATURE OF ULTRA-MICROSCOPIC VIRUSES.', *The Lancet*. Elsevier, 186(4814), pp. 1241–1243. doi: 10.1016/S0140-6736(01)20383-3.

Villarreal, L. P. (2004) 'Are Viruses Alive?', *Scientific American*, December, pp. 100–105. Available at: https://www.jstor.org/stable/26060805?seq=1#metadata_info_tab_contents.

Walcher, G. *et al.* (2010) 'Evaluation of Paramagnetic Beads Coated with Recombinant Listeria Phage Endolysin–Derived Cell-Wall-Binding Domain Proteins for Separation of Listeria monocytogenes from Raw Milk in Combination with Culture-Based and Real-Time Polymerase Chain Reaction–Ba', *Foodborne Pathogens and Disease*. Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA, 7(9), pp. 1019–1024. doi: 10.1089/fpd.2009.0475.

Wang, J., Hofnung, M. and Charbit, A. (2000) 'The C-terminal portion of the tail fiber protein of bacteriophage lambda is responsible for binding to LamB, its receptor at the surface of Escherichia coli K-12.', *Journal of bacteriology*. American Society for Microbiology Journals, 182(2), pp. 508–12. doi: 10.1128/JB.182.2.508-512.2000.

Wommack, K. E. and Colwell, R. R. (2000) 'Virioplankton: viruses in aquatic ecosystems.', *Microbiology and molecular biology reviews*, 64(1), pp. 69–114. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10704475 (Accessed: 17 October 2018).

Yoichi, M. *et al.* (2005) 'Alteration of tail fiber protein gp38 enables T2 phage to infect Escherichia coli O157:H7', *Journal of Biotechnology*. Elsevier, 115(1), pp. 101–107. doi: 10.1016/J.JBIOTEC.2004.08.003.

Yosef, I. *et al.* (2015) 'Temperate and lytic bacteriophages programmed to sensitize and kill antibiotic-resistant bacteria', *Proceedings of the National Academy of Sciences*, 112(23), pp. 7267–7272. doi: 10.1073/pnas.1500107112.

Yutin, N. *et al.* (2018) 'Discovery of an expansive bacteriophage family that includes the most abundant viruses from the human gut', *Nature Microbiology*. Nature Publishing Group, 3(1), pp. 38–46. doi: 10.1038/s41564-017-0053-y.