Identification and characterisation of novel phages of *Pectobacterium* and *Erwinia*

A dissertation presented to Cork Institute of Technology for the degree of

Doctor of Philosophy

by

Colin Buttimer, B.Sc.

Department of Biological Sciences, Cork Institute of Technology,
Bishopstown, Cork, Ireland

Research supervisors:
Professor Aidan Coffey
Dr Jim O'Mahony

Submitted to Cork Institute of Technology,
November 2018

CORK INSTITUTE OF TECHNOLOGY
INSTITIÚID TEICNEOLAÍOCHTA CHORCAÍ

Declaration

I declare that this thesis, which I submit to Cork Institute of Technology, is my own personal effort. Where any of the content presented is the result of input or data from a related collaborative research programme, this is duly acknowledged, such that it is possible to ascertain how much of the work is my own. I have not already obtained a degree in Cork Institute of Technology or elsewhere on the basis of this work. Furthermore, I took reasonable care to ensure that the work is original, and, to the best of my knowledge, does not breach copyright law, and has not been taken from other sources except where such work has been cited and acknowledged within the text.

Signed _____

Student Number _____

Date _____

Table of Contents

for the laughs but not so much the drama! Caoimhe, for the late evening chats. Lisa, no better travel buddy at a scientific conference. Jennifer, sharing your passion for Mario Kart and Pokémon. Kate for your help with statistical analysis, a godsend!

My project students Linda and Frank, your hard work made my own work that much better. Other CREATE Erasmus students Matthias, Jane, Elisa and Minh for sharing German culture and food!

To Dad (Francis) and Mum (Shirley), enabling me to explore my major passion in life. Yes, I have finally finished writing that book! My sisters (Avril and Lorna), thank you for your support and the numerous times that you allowed me to use your place as a pit stop for my conference related adventures.

Finally, to my friends (previously mentioned and not mentioned), thank you so much for your advice, distractions, support and care over the duration of my PhD, it has been indispensable.

List of abbreviations

aa	amino acid
BLAST	basic local alignment search tool
bp	base pair
ca.	circa
CFU	colony forming units
cm	centimetre
CMP	cytidine monophosphate
Co.	county
cs	clear spot
CsCl	caesium chloride
CVP	crystal violet pectate
dCMP	deoxycytidine monophosphate
DNA	deoxyribonucleic acid
dNDP	nucleoside diphosphate
dNMP	nucleoside monophosphate
dNTP	nucleoside triphosphate
dTMP	deoxythymidine monophosphate
dUMP	deoxyuridine monophosphate
DTR	direct terminal repeats
dUMP	deoxyuridine monophosphate
EOP	efficiency of plaquing
ESI-MS/MS	electrospray ionization-mass spectrometry
Freq.	frequency
g	relative centrifugal force
GBDP	genome BLAST distance phylogeny
h	hour
HCl	hydrochloric acid
ICTV	International Committee on Taxonomy of Viruses
IPTG	isopropyl β-D-1-thiogalactopyranoside
kDa	kilodalton

L	litre
LB	Lysogeny broth
M	molar
MALDI-TOF	spectrometry matrix-assisted laser desorption ionization-time of flight mass
mL	millilitre
ML	maximum likelihood
mg	milligram
n=	number
NAD^+	nicotinamide adenine dinucleotide
NaCl	sodium chloride
NaOH	sodium hydroxide
min	minute
no.	number
nm	nanometre
OD_{600}	optical density at 600 nm
ORF	open reading frame
PCR	polymerase chain reaction
PFU	plaque forming units
PFGE	pulsed-field gel electrophoresis
pH	measure of the hydrogen ion concentration of a solution
pVOG	prokaryotic virus orthologous group
RNA	ribonucleic acid
RNAP	RNA polymerase
rpm	rotations per minute
RT-PCR	reverse transcriptase polymerase chain reaction
SAR	signal-arrest-release
SDS-PAGE	sodium dodecyl polyacrylamide gel electrophoresis
SRE	soft rot *Enterobacteriaceae*
TEM	transmission electron microscopy
TMP	thymidine monophosphate
tRNA	transfer RNA

UV	ultraviolet
v	volume
V	Volt
vRNA	polymerase virion-associated polymerase
w	weight
°C	Celsius
µg	microgram
µL	microliter
µM	micromolar
µm	micrometre
%	per cent

List of posters, presentations, publications and taxonomy proposals

Poster presentations

Buttimer, C., Hendrix, H., Neve, H., Noben, J-P., Franz, C., O'Mahony, J., Lavigne, R., Coffey, A. (2018). Novel N4-like phages of *Pectobacterium atrosepticum*. EMBO conference, Viruses of Microbes. Wroclaw, Poland. Jul 9-13,2018.

Buttimer, C., Lucid, A., McAuliffe, O., Ross, R, P., Hill, C., Coffey, A. (2017). Characterization of bacteriophages of *Pectobacterium atrosepticum* and assessing their application as a biocontrol. 46th Annual Food Research Conference. Cork, Ireland. Dec 6-7, 2017.

Buttimer, C., Born, Y., Hendrix, H., Oliveria, H., McAuliffe, O., Neve, H., Ross, R. P., Hill, C., Noben, J-P., O'Mahony, J., Loessner, M, J., Fieseler, L., Lavigne, R., Coffey, A. (2017). Hairy jumbo phages of *Enterobacteriaceae*. 22nd Evergreen Conference. Aug 6-11, 2017.

Buttimer, C., Oliveira, H., Hendrix H., Lavigne R., McAuliffe, O., Ross R.P., Hill, C., Coffey, A. (2016). Analysis of the atypical 'hairy' *Enterobacteriaceae* phage vB_PcaM_CBB. EMBO conference, Viruses of Microbes. Liverpool, U.K. July 18-22, 2016.

Buttimer, C., Lucid, A., McAuliffe, O., Ross, R.P., Hill, C., Coffey, A. (2015). Characterization of bacteriophages that are inhibitory to *Pectobacterium atrosepticum*, a common potato pathogen. 44th Annual Food Research Conference. Cork, Ireland. Dec 14, 2015.

Buttimer, C., McAuliffe, O., Ross, R.P., Hill, C., Coffey, A. (2015). Genomes of bacteriophages infecting *Pectobacterium*. Phage in interactions. Leuven, Belgium. Sep 29, 2015.

Buttimer, C., McAuliffe, O., Ross, R.P., Hill, C., Coffey, A. (2015). Isolation and characterisation of soil-borne bacteriophages of *Pectobacterium atrosepticum*. EuroSciCon conference, Bacteriophage. London, U.K., Jan 27-29, 2015.

Oral presentations

Buttimer, C., Coffey, A. (2018). Novel N4-like phages of *Pectobacterium atrosepticum*. Oxford bacteriophage conference, Oxford, UK. Sep. 11-12, 2018

Buttimer, C., Coffey, A. (2014) Isolation and characterisation of bacteriophages for the elimination of the potato pathogen *Pectobacterium*. IPSAM meeting. Cork, Ireland. Apr 28-29, 2014

Peer-reviewed publications

Buttimer, C., Lucid, A., Neve, H., Franz, C., O'Mahony, J., Turner, D., Lavigne, R., Coffey, A. (2018). *Pectobacterium atrosepticum* Phage vB_PatP_CB5: A member of the proposed genus "*Phimunavirus*." *Viruses* 10, 394. https://doi.org/10.3390/v10080394

Buttimer, C., Hendrix, H., Lucid, A., Neve, H., Noben, J.-P., Franz, C., O'Mahony, J., Lavigne, R., Coffey, A. (2018). Novel N4-like bacteriophages of *Pectobacterium atrosepticum*. *Pharmaceuticals* 11. https://doi.org/10.3390/ph11020045

Ajuebor, J., **Buttimer, C.,** Arroyo-Moreno, S., Chanishvili, N., Gabriel, E.M., O'Mahony, J., McAuliffe, O., Neve, H., Franz, C., Coffey, A. (2018). Comparison of *Staphylococcus* phage K with close phage relatives commonly employed in phage therapeutics. *Antibiotics* 7. doi.org/10.3390/antibiotics7020037

Buttimer, C., Born, Y., Lucid, A., Loessner, M.J., Fieseler, L., Coffey, A. (2018). *Erwinia amylovora* phage vB_EamM_Y3 represents another lineage of hairy *Myoviridae*. *Research in Microbiology*. https://doi.org/10.1016/j.resmic.2018.04.006

Buttimer, C., Hendrix, H., Oliveira, H., Casey, A., Neve, H., McAuliffe, O., Ross, R.P., Hill, C., Noben, J.-P., O'Mahony, J., Lavigne, R., Coffey, A. (2017). Things are getting hairy: Enterobacteria bacteriophage vB_PcaM_CBB. *Frontiers in Microbiology*. 8, 44. https://doi.org/10.3389/fmicb.2017.00044

Buttimer, C., McAuliffe, O., Ross, R.P.P., Hill, C., O'Mahony, J., Coffey, A., O'Mahony, J., Coffey, A. (2017). Bacteriophages and bacterial plant diseases. *Frontier in Microbiology* 8, 34. https://doi.org/10.3389/fmicb.2017.00034

Endersen, L., **Buttimer, C.,** Nevin, E., Coffey, A., Neve, H., Oliveira, H., Lavigne, R., O'Mahony, J., (2017). Investigating the biocontrol and anti-biofilm potential of a three phage cocktail against *Cronobacter sakazakii* in different brands of infant formula. *International Journal of Food Microbiology* 253. https://doi.org/10.1016/j.ijfoodmicro.2017.04.009

O'Sullivan, L., **Buttimer, C.,** McAuliffe, O., Bolton, D., Coffey, A. (2016). Bacteriophage-based tools: recent advances and novel applications. *F1000Research* 5, 2782. https://doi.org/10.12688/f1000research.9705.1

Buttimer, C., O'Sullivan, L., Elbreki, M., Neve, H., McAuliffe, O., Ross, R.P., Hill, C., O'Mahony, J., Coffey, A. (2016). Genome sequence of jumbo phage vB_AbaM_ME3 of *Acinetobacter baumanni*. *Genome Announcements* 4. https://doi.org/10.1128/genomeA.00431-16

Taxonomy proposals to the International Committee on Taxonomy of Viruses

Kropinski, A.M., Adriaenssens, E.M., Withmann, J., **Buttimer, C.,** Coffey, A. (2018). To create one new genus*, Cbunavirus*, containing two species in the family *Podoviridae*.

Buttimer, C., Turner, D., Coffey, A., Neve, H., Kropinski, A.M., Adriaenssens, E.M. (2018). To create one new genus, *Phimunavirus*, containing four species in the subfamily *Autographivirinae*, family *Podoviridea*. 2018.110B

Turner, D., **Buttimer, C.**, X, Coffey, A., Neve, H., Kropinski, A.M., Adriaenssens, E.M. (2018). To create one new genus *Metrivirus*, containing one species in the family *Myoviridae*. 2018.108B

Thesis introduction

Losses in crop yields due to disease need to be reduced to meet increasing global food demands associated with growth in the human population. There is a well-recognised need to develop new environmentally-friendly control strategies to combat bacterial crop diseases. There are several crop diseases for which no effective bactericidal agents are currently available, such as potato blackleg and soft rot disease caused by *Pectobacterium atrosepticum* and other members of soft rot *Enterobacteriaceae* (Czajkowski et al., 2011). Furthermore, current control measures involving the use of traditional chemicals or antibiotics are losing their efficacy due to the natural development of bacterial resistance to these agents, as seen for fire blight of the pear and apple tree caused by *Erwinia amylovora* (de León Door et al., 2013; Mayerhofer et al., 2009; Ordax et al., 2006; Russo et al., 2008). Bacteriophages (phage), the viruses of bacteria, have received increased research interest in recent years as an environmentally friendly means of controlling bacterial diseases. However, not all phages possess the features that would enable them to be effective bactericidal agents. To this end, this thesis provides a detailed study of phages that infect *Pectobacterium atrosepticum* and *Erwinia amylovora.* The knowledge gained in the execution of this PhD thesis contributes to the pool knowledge about the lifestyles of the phages examined thus enabling a more informed choice with regard to the selection of suitable phages for biocontrol applications for the relevant phytopathogens.

Chapter 1 is a review of phages and their application as a biocontrol against bacterial plant diseases in broad terms. The literature shows that phages for biocontrol applications should have a number of features such as strictly lytic lifestyles, suitable host range and genomes that possess no toxins and other virulence factors. Additionally, it indicates that phage biocontrol possesses several advantages over chemical controls in that tailor-made phage cocktails can be adapted to target specific disease-causing bacteria. Unlike chemical control measures, phage mixtures can be easily adapted for bacterial resistances that may develop over time. Phages have also been deemed suitable for organic growers for the control of crop disease. However, there exist some disadvantages to their use as bactericidal agents such a limited host

range, biostability and physical access to their host bacterial strain. But the use of UV protective agents and the timing of phage application can overcome some of these shortcomings. Nevertheless, this chapter shows that phage-based products against phytopathogens are starting to become available on the market.

In **Chapter 3**, stem samples of potato crops exhibiting blackleg were taken from three farms in Co. Cork, Ireland, and were found to be infected with *P. atrosepticum*. Three closely related phages specific to this phytopathogen were isolated and characterised, namely vB_PatP_CB1, vB_PatP_CB3 and vB_PatP_CB4 (abbreviated as CB1, CB3 and CB4). Both CB1 and CB3 were determined to infect 12 strains and CB4 10 strains of the 19 strains of *P. atrosepticum* tested. Morphology, latent periods, burst sizes and their stability at various temperatures and pHs were also examined. Genome sequencing of the three phages revealed that they shared a minimum nucleotide identity of 93% with each other. Their genomes exhibited an *Enquartavirinae* genome organization, possessing several conserved proteins associated with phages of this group, like the type species *Escherichia* virus N4. Tandem electrospray ionization-mass spectrometry (ESI-MS/MS) allowed the identification of ten structural proteins that form the virion of CB1, six which are conserved in phage N4. *In vitro* experiments demonstrated that the phages suppress soft rot formation upon co-inoculation with *P. atrosepticum* on whole tubers. Results of this study indicate that CB1 related phages could be good candidates for phage-based control.

Chapter 4, describes *Pectobacterium* phage vB_PatP_CB5 (abbreviated as CB5) which specifically infects the bacterium. The phage was characterized in detail and transmission electron microscopy (TEM) micrographs indicated that it belongs to the *Podoviridae* family. CB5 has significant pairwise nucleotide identity (≥80%) to *P. atrosepticum* phages φM1, Peat1 & PP90 and shares common genome organization. Phylograms constructed using conserved proteins and whole-genome comparison based amino acid sequences show that these phages form a distinct clade within the *Autographivirinae*. They also possess conserved RNA polymerase recognition and specificity loop sequences. Their lysis cassette resembles that of *KP34virus*, containing in sequential order, a U-spanin, a holin and a SAR endolysin. However, they possess low pairwise nucleotide identity to the type phage of the *KP34virus* genus, *Klebsiella*

phage KP34. In addition, phage KP34 does not possess several conserved proteins associated with these *P. atrosepticum* phages. As such, we propose the allocation of phages CB5, Peat1, ɸM1 and PP90 to a separate new genus designated '*Phimunavirus*'.

Chapter 5, provides a detailed description of *Pectobacterium* phage vB_PatM_CB7 (abbreviated as CB7), which specifically infects *P. atrosepticum*. This phage can be placed with the genus *Cr3virus*, the only genus of phage currently defined by the International Committee on Taxonomy of Viruses (ICTV) with phages infecting *P. atrosepticum*. Host range, morphology, latent periods, burst sizes, and stability at different conditions of temperature and pH were examined. Analysis of its genome shows that it shares features with other Cr3-like phages, highlighting conservation within the genus. Conserved elements include a putative early promoter possessing elements like that of the *Escherichia coli* sigma70 promoter, which was shared with other genus members. A number of dissimilarities also exist regarding DNA methylation and nucleotide metabolism with some members not possessing homologs of a putative cytosine methylase and anaerobic nucleotide reductase subunits NrdD and NrdG respectively. Furthermore, the genome of CB7 was identified to have possibly the largest number of homing endonucleases in the literature to date, having 23 from both the HNH and LAGLIDADG families. Examination of the HNH homing endonuclease residing within introns of genes of the large terminase, DNA polymerase, ribonucleotide reductase subunits NrdA and NrdB showed that they are splicing-competent but dissimilar to the nicotinamide phosphoribosyltransferase intron. ESI-MS/M was also performed on the virion of CB7, allowing the identification of 26 structural proteins, 20 of which were found to be shared with the type phages of the genera of *V5virus* and *SE1virus*. Phylogenetic analysis conducted on *Vequintavirinae* in this study indicated the existence of a novel clade represented by *Klebsiella* phages vB_KpnM_KB57. The results of this study provide greater insights into the phages of the *Cr3virus* genus as well as the subfamily *Vequintavirinae*.

Chapter 6, examines *Pectobacterium* phage vB_PcaM_CBB (abbreviated as CBB) is a 'jumbo' phage belonging to the family *Myoviridae.* It possesses highly atypical whisker-like structures along the length of its contractile tail. It has a broad host range with the

capability of infecting species of the genera *Erwinia*, *Pectobacterium* and *Cronobacter*. With a genome of 355,922 bp, excluding a predicted terminal repeat of 22,456 bp, phage CBB has one of the largest genomes sequenced to date. Its genome was predicted to encode 554 open reading frames (ORFs) with 33 tRNAs. Using a combination of BLASTP, Interproscan, HHpred and virion proteome analysis, 29% of its predicted ORFs could be assigned functions, involved in DNA replication, nucleotide metabolism, virion structure and peptidoglycan degradation. Putative endolysin CBB_187 was demonstrated to possess signal-arrest-release (SAR) activity and putative virion-associated lysin CBB_239 shown to be capable of peptidoglycan degradation. Protein comparison shows that CBB shares between 33-38% of its proteins with *Cronobacter* phage GAP32, coliphages PBECO4 and 121Q as well as *Klebsiella* phage vB_KleM_Rak2. This work presents a detailed and comparative analysis of phage CBB, a highly atypical jumbo *Myoviridae* phage, contributing to a better understanding of phage diversity and biology.

Chapter 7, examines *Erwinia amylovora* phage vB_EamM_Y3 (abbreviated as Y3), another example of a jumbo myoviruses with whisker-like structures along the surface of its contractile tail. It possesses a genome of 261,365 kbp with 333 predicted ORFs. Using a combination of BLASTP, Interproscan and HHpred, about 21% of its putative proteins could be assigned functions involved in nucleotide metabolism, DNA replication, virion structure and cell wall degradation. The phage was found to have a SAR endolysin (Y3_301) possessing a soluble lytic transglycosylase domain. Like other SAR endolysins, inducible expression of Y3_301 caused *Escherichia coli* lysis, which is dependent on the presence of an N-terminal signal sequence. Phylogenetic analysis showed that its closest relatives are other jumbo phages including *Pseudomonas aeruginosa* phage PaBG and *P. putida* phage Lu11, sharing 105 and 87 homologous proteins, respectively. Like these phages, Y3 also shares a distant relationship to *Ralstonia solanacearum* phage ΦRSL1 (sharing 55 homologous proteins). As these phages are unrelated to the Rak2-like group of hairy phages, Y3 along with Lu11 represent a second lineage of hairy myoviruses.

Chapter 1. Literature review

A manuscript based upon this chapter has been published in Frontiers in Microbiology.

1.1. Importance of crop diseases

The human population is expected to reach 9.6 billion by 2050 and this will result in increased demands for food. It has been estimated that the global food supply may need to grow by as much as 70% in order to meet these demands (UN, 2013). For such growth, it has been predicted that crop supply may have to increase as much as 80-110% (Ray et al., 2013). To achieve these yields, the impact of crop disease has to be reduced. It has been estimated that at least 10% of global food production is lost to plant diseases (Strange and Scott, 2005). The major pathogens of plants are parasitic plants, oomycetes, nematodes, viruses, fungi and bacteria. Among the latter, there are over 200 plant pathogenic bacterial species (Considine and Considine, 1995). Those considered to be the most important belonging to the genera of *Pseudomonas, Ralstonia, Agrobacterium, Xanthomonas*, *Erwinia*, *Xylella*, *Pectobacterium* and *Dickeya* (Mansfield et al., 2012)*.*

1.2. Bacteriophages, their life cycles and their morphology

Bacteriophages (phages) are the most abundant biological entity in the biosphere with an estimated number of 10^{31}, as total prokaryotic cell numbers are understood to be around 10^{30} in the biosphere and phage numbers are believed to be at least 10 times greater than this value (Whitman et al., 1998; Wommack and Colwell, 2000). Phages are specific viruses of bacteria that subvert the metabolism of their bacterial hosts in order to replicate. Of the phages that have been identified, the majority belong to the tailed phages; and these form the Taxonomic Order: *Caudovirales* (Ackermann, 2007). These phages possess icosahedral heads containing genomes comprised of double-stranded DNA. The order *Caudovirales* is made up of three phage families; *Myoviridae* which have rigid contractile tails, *Podoviridae* with short, non-contractile tails and *Siphoviridae* with long flexible tails. Phages belonging to other families have highly variable morphologies with genomes of varying nucleic acid composition.

1.3. History of bacteriophages and their use as antibacterial agents towards plant diseases

The discovery of bacteriophages is credited to Frederick Twort (Twort, 1915) and Felix d'Herelle (d'Herelle 1917). Similar findings of antibacterial agents that hinted on the existence of phage had been made prior to that of Twort and d'Herelle (Abedon et al.,

2011). However, they were the first to suggest this phenomenon as being viral in origin. The potential of phages as antibacterial agents was quickly recognised, with d'Herelle in 1919 demonstrating the capability of his phage preparations to treat dysentery patients in the Hôpital des Enfants-Malades in Paris (Wilkinson, 2001). Following this work, many early studies and attempts were made to use phages to treat staphylococcal infections, cholera and bubonic plaque of humans (Sulakvelidze et al., 2001). This pre-antibiotic era approach became known as bacteriophage therapy. Studies were also initiated with the aim of using phages to control plant diseases. Mallmann and Hemstreet in 1924 showed that the filtrate of decomposing cabbage could be used to inhibit the "cabbage-rot organism" *Xanthomonas campestris pv. campestris*. In 1925, Kotila and Coons demonstrated with bioassays that they could use phage to prevent soft rot by *Pectobacterium atrosepticum* and *Pectobacterium carotovorum* subsp *carotovorum* on slices of potato tuber and carrot respectively (Coons and Kotila, 1925; Kotila and Coons, 1925). The first field trials were also done by Thomas (1935), who showed that he could reduce the incidence of Stewart's wilt disease by treating seeds with phage against the phytopathogen *Pantoea stewartii* from 18% (untreated) to 1.5% (treated). However this type of research became neglected as an understanding of the nature of phage was poor at the time, and data on their efficacy was limited (Okabe and Goto, 1963).

1.4. Bacteriophage types used for therapy/biocontrol

From a terminology perspective, the term bacteriophage therapy is usually reserved for human and animal applications. For plants, the term bacteriophage biocontrol is more often used. In recent years, several studies have been published on phage biocontrol on a number of important bacterial plant pathogens, with many very promising results (Table 1.1). The main deciding factor whether a phage is applicable for biocontrol (and also therapy in humans or animals) is whether a phage is exclusively lytic (virulent) or instead temperate in nature. Virulent phages are those which conduct infections that ultimately result in lysis of their host bacterium with the release of progeny phage particles. Temperate phages can follow the lytic route of infection but also follow the route of lysogeny, where the phage genome integrates into the bacterial chromosome or persists as a plasmid. In this form, the phage is

known as a prophage (Łobocka et al., 2004). With this strategy, the phage genome replicates as part of the bacterial genome of its host until a trigger switches it into the lytic cycle. These triggers can be chemical or physical (UV light or heat) in nature (Brunner and Pootjes, 1969; Müller et al., 2012). It is interesting to note that certain plant extracts can also trigger these events (Sato, 1983). Often, prophage DNA can increase the fitness of the bacterial host due to genes present on prophage genome. For example, in the case of plant pathogens, the *P. atrosepticum* prophages ECA41 and ECA29 both improve the motility of the bacterial host (Evans et al., 2010). Prophages may also harbour genes for toxins, e.g. shiga, cholera and diphtheria toxins (Abedon and Lejeune, 2005). Another concern with these phages is the spread of virulence genes by transduction (specialised), where these phages can excise themselves from their host genomes incorporating host DNA into their own genomes facilitating the horizontal transfer of genetic material among bacteria (Griffiths et al., 2000). Also, some lytic bacteriophages are capable of transduction (generalised), where they accidentally pack bacterial DNA into their own capsid heads during the later stages of the lytic cycle (Klumpp et al., 2008). There is also a third mechanism of phage-host interaction identified in filamentous phages (*Inovirus family*). Here, phages form a non-lethal chronic infection with continuous production of progeny phages. However, suitability of these phages for biocontrol is questionable as their infection can have varying effects on host virulence, as shown with phytopathogen *Ralstonia solanacearum* with its phage φRSS1 causing increased virulence (Yamada, 2013), although it has been shown possible to isolate virulent filamentous phage (Kuo et al., 1994). Another undesirable property in a phage intended for biocontrol is the ability to bring about superinfection exclusion to its host during infection. This prevents secondary infection of the host by another phage (Lu and Henning, 1994).

Ideally, a phage for biocontrol applications should be exclusively lytic and possess a host range which allows productive infection on all strains of the pathogen genus/species being targeted. Also, current opinion is that phages should be able to lyse the host quickly while producing high numbers of progeny phage and diffuse easily through the environment to which they are being applied. However, there was a report of a phage (φRSL1) of the phytopathogen *R. solanacearum* which was described

as not to be highly lytic but still exhibited great biocontrol effect. The current standing theory of this phage's disease prevention ability is that it is capable of co-existing without complete removal of its host from the soil surrounding crop roots, forming an equilibrium of infection that maintains the phage's population but yet suppresses bacteria pathogenicity (Fujiwara et al., 2011).

While a given phage's infection properties may appear to have great potential with *in vitro* studies, this does not necessarily translate into biocontrol potential in the field. Balogh showed in a study of three phages of *X.citri* pv *citri* exhibiting lytic activity in overlay plate assays that two of these phages were unable to lyse their host bacterium on grapefruit leaves, and indeed were later shown to be ineffective for the suppression of citrus canker in greenhouse trials (Balogh, 2006). Attention should also be paid to the receptors that a given phage recognises on a bacterial target. This can aid in the creation of phage mixtures with a reduced likelihood of host resistance (Frampton et al., 2014), and as such can lead to the development of phage combinations where individual members work in synergy to eliminate the target bacterium (Born et al., 2011).

1.5. Advantages of phage biocontrol over other strategies

Unlike chemical biocides, phages occur naturally in the environment and humans are thus exposed to them on a daily basis without any harm. After application, their numbers increase if their target bacterial host species are accessible to them. However, they tend to persist in high numbers in any environment only as long as the host is present (Iriarte et al., 2012). Thus, phages are unlike copper-based pesticides which can potentially accumulate in the soil (Hirst et al., 1961; Pietrzak and McPhail, 2004). Phages generally have a narrow host range, typically being limited to strains within a particular species of bacteria. This can allow the creation of phage mixtures which can target bacterial species within a given genus of bacteria only. This could be a specific bacterial phytopathogen or it could be a particular bacterium in a microbial community whose suppression could help improve crop growth. Basit et al (1992) for example, isolated a phage which was unable to infect the desired strain of *Bradyrhizobium japonicum* which could aid soybean crop growth due to its nitrogen

fixation properties but could inhibit competing bacteria which did not possess this feature, thus allowing enhanced nitrogen fixation to occur.

Biofilm formation is an important factor in the virulence of phytopathogens such as *E. amylovora* (Koczan et al., 2011; Li and Wang, 2014). It is an attribute which has been shown to be involved in bacterial phytopathogen resistance to copper bactericides (Rodrigues et al., 2008). Phages have evolved to overcome this biofilm barrier through the use of depolymerase enzymes on their virions but can also be released on host lysis, which allows them to degrade biofilm material, allowing the phage anti-receptor to gain access to the receptors on the surface of their host bacterium (Born et al., 2013). There is a growing demand by consumers for food produce that is free from chemicals biocides and preservatives. This has resulted in the restricted use of chemicals to produce "Organic label" crops. The requirements of such food require the absence of chemical residues in crop production and processing (Lohr, 2001). Since phages are naturally occurring in the environment, they can be registered as biopesticides, making them suitable for more consumer-friendly organic farming (OmniLytics, 2006).

1.6. Potential issues concerning the use of phage in biocontrol

The main limitation for the application of phages in biocontrol in most settings is bacterial host-range. While this can be an advantage in certain circumstances, developing a phage-biocide that eliminates every member of a particular bacterial genus or species can be a challenge. Frequently the development of phage mixtures (cocktails) overcomes this disadvantage. Occasionally (but nevertheless, rarely) a phage is isolated which has an unexpectedly broad-host-range. One example of this is a phage isolated from sewage and shown to target *Pectobacterium* and also enteric bacteria associated with humans (Pirhonen and Palva, 1988). Thus, careful attention should be given to ensure full understanding of likely host-range of a phage to avoid inefficacy or indeed to avoid the elimination of non-target potentially beneficial bacteria. In the latter context, instances of phage infecting beneficial bacteria resulting in reduced crop yield have been reported (Ahmad and Morgan, 1994; Basit et al., 1992).

It is believed that phages do not directly interact with plants. However, a number of phage-like genes have been identified in wheat, corn and *Arabidopsis* cress (Chang et al., 1999; Hedtke et al., 1997; Ikeda and Gray, 1999) which would suggest incorporation of phage DNA into the genomes of these crops and thus a possible a role in their evolution.

1.7. Advantages of bacteriophages in the context of host resistance

Like antibiotics and copper sprays, for which resistance has been reported, there is also the possibility of bacteria becoming resistant to phage infection following constant exposure. However, unlike chemicals, phages are biological entities which can evolve and overcome these biological alterations in their hosts. There has always been a constant race between phage and bacteria in nature. This is indicated by the fact that up 10-20% of bacterial populations in certain habitats are lysed daily because of phage infection (Suttle, 1994). In the context of phage resistance, Qiao *et al* (2010) found that *Pseudomonas syringae* phage phi2954 was dependent on a host protein glutaredoxin 3 for successful infection. Mutant host strains without this protein were shown to be resistant to the phage. Nevertheless, these authors showed it was possible to isolate mutants of the phage that had become independent of this host protein for infection and this observation has been developed and employed in certain phages aimed at biocontrol. Flaherty et al (2001) also showed that phages could evolve to overcome phage resistance in target bacteria and these were referred to as H-mutants. This allowed the development of phages with broader host ranges.

In addition to simple mutation-based phage resistance, bacterial phytopathogens can also possess other more complex resistance mechanisms such as the altruistic abortive infection (Abi) systems which give a bacterial host population immunity against a phage by causing phage-infected cells to commit suicide in order to prevent phage reproduction (Parma et al., 1992). While a number of these systems have been identified in *Lactococcus* starter culture strains found in dairy fermentations (Chopin et al., 2005; Coffey and Ross, 2002), recently such a system was identified in the phytopathogen *P. atrosepticum* and was termed ToxIN. This was characterised as a plasmid-encoded Type III protein-RNA toxin-antitoxin system. The toxic protein ToxN is bound to RNA antitoxin ToxI in its inactive form. However, when phage infection

occurred, ToxI RNA antitoxin became unbound from ToxN causing the death of the bacterial host cell (Fineran et al., 2009). Indeed, Blower et al (2012) also showed using phage phiTE, that the phage was capable of creating mutants that could overcome this system by producing a pseudo ToxI RNA antitoxin preventing ToxN toxic activity

Another mode of phage resistance is CRISPR/Cas systems, which are used by bacteria as well as archaea to form an immunity to protect from infection by foreign DNA such as phage. These systems are comprised of clustered regularly interspaced short palindromic repeat (CRISPR) arrays and CRISPR associated (Cas) proteins. In a recent study of 1,724 bacterial and archaeal genomes, it was found that these systems were present in 10% of studied genomes. Previous studies had estimated CRISPR/Cas prevalence values of 40% and 80% of studied bacteria and archaeal genomes, respectively (Burstein et al., 2016). These have been detected in phytopathogens such as *P.atrosepticum* (Przybilski et al., 2011), *E.amylovora* (Rezzonico et al., 2011) and *Xanthomonas oryzae* (Semenova et al., 2009). CRISPR arrays are comprised of short stretches of DNA (termed spacers), which are transcribed into short RNAs which interact with Cas proteins to detect and cut foreign DNA that match the sequence of the spacer (protospacer). Spacer sequences are acquired during exposure to foreign DNA in phage or plasmids, and thus they provide a genetic immunity from invasion by foreign DNA due to previous encounters (Marraffini and Sontheimer, 2008). However, it is also possible for phages to evolve to overcome these systems. Indeed, Semenova et al (2009) detected a spacer in *X.oryzae* which matched a protospacer of phage Xop411. However, the phage was still able to infect this bacterium, due to a mutation having occurred in the protospacer sequence.

Bacteria developing resistance against phage infection is not necessarily a negative development in the context of phage biocontrol. Phage-resistance mutations in bacteria frequently are accompanied by a fitness cost, one example being a reduction in virulence, resulting in reduced disease severity. This results from the fact that molecules involved in phage attachment are frequently also involved in the virulence process. Examples include lipopolysaccharide (LPS) (Evans et al., 2010a), extracellular polysaccharide (EPS) (Ayers et al., 1979), flagella (Addy et al., 2012; Evans et al., 2010b) and pili (Ahern et al., 2014). Thus, mutations leading to resistance frequently

compromise virulence. There are, however, a few examples where these mutations in bacteria surface structures did not lead to reduced virulence as seen with LPS production mutants of *Pectobacterium* and *Dickeya* (Pirhonen et al., 1988; Schoonejans et al., 1987).

1.8. Bacteriophage and chemicals

Phages have been shown to be stable in certain agrichemicals (Ravensdale et al., 2010). However, precautions need to be taken with some chemicals being combined with phage. Chemical biocides typically contain a range of phage inactivating substances such as surfactants and chelators (Chattopadhyay et al., 2002; Yamamoto et al., 1968). Also, copper-based bacteriocides have been shown to inactivate phage, but this inactivation can be avoided with the delayed application of phage (4-7 days) after initial application of copper-based bactericide (Iriarte et al., 2007).

1.9. The complexity of bacteriophage interaction with soil

The rhizosphere is the area of soil which is in close proximity to the roots of a plant. There are a number of factors which can affect phage activity in this environment such as pH, moisture levels, the presence of organic matter and soil type. A number of these factors either individually or in combination can cause phage inactivation. Different soil types affect the survival of phage. For example, clay loam soils appear better at maintaining phage at low soil moisture levels and high soil temperatures than that of sandy loam soils (Straub et al., 1992) As well, low soil pH can also negativity affect phage survivability (Sykes et al., 1981).

Levels of adsorption of phage are affected differently in differing soil types, with levels of hindrance varying from one phage type to another (Goyal and Gerba, 1979). Phage can become bound to soil components such as clays (kaolinite and montmorillonite) as these minerals possess positively and negatively charged surfaces to which phage can adsorb (Schiffenbauer and Stotzky, 1982). Such adsorption can be influenced by pH (Goyal and Gerba, 1979; Loveland et al., 1996) as well the presence of organic materials (Zhuang and Jin, 2003). Under favourable conditions, phages have been identified that persist at relatively stable concentrations for several weeks in soil (Fujiwara et al., 2011).

1.10. Bacteriophage in the phyllosphere

The phyllosphere is the portion of the plant which is above the ground and phages can readily be isolated from this location. How phages get there naturally has not been defined precisely, although it is possible that they originate in the soil from which the plant germinated or alternatively get deposited by insect vectors. Indeed, phages for the phytopathogens *Pantoea stewartii* and *Erwinia herbicola* var. *herbicola* have been isolated from corn flea beetles (Woods et al., 1981). Another route is the translocation of phage from the roots to leaves of plants through the plant vascular system. And it has been shown that phages of *R. solanacearum*, *Xanthomonas perforans* and *Xanthomonas euvesicatoria* can translocate though tomatoes plants, phage of *Xanthomonas oryzae* through the rice seedlings and phages of *E. amylovora* though apple seedlings and firethorn (Iriarte et al., 2012; Kolozsváriné Nagy et al., 2015; Rao and Srivastava, 1973). However, this translocation may be influenced by the phage type, plant age, plant size, plant species, plant health and possibly soil type in which the plant is growing (Iriarte et al., 2012). It has also been reported that *E. amylovora* phages could pass from the leaves to the roots of apple seedlings (Kolozsváriné Nagy et al., 2015). The phyllosphere is nevertheless a harsh environment for phages to survive and it has been reported that their numbers can rapidly decline during daylight hours (Balogh et al., 2003; Iriarte et al., 2007). The destructive influence of UV light from the sun has been reported to be a limiting factor for the application of phages for successful biocontrol. The radiation causes the formation of lesions in DNA which can block DNA replication and transcription. An *in vivo* study with phage phiXV3-16, Iriarte *et al* demonstrated a direct relationship between phage reduction on tomato leaves and increasing UVA+B dose. They also showed in an *in vitro* study that UV was capable of inactivating phage used against *Xanthomonas campestris pv. vesicatoria,* preventing it from exerting a biocontrol effect (Iriarte et al., 2007). Phage sensitivity against UV light has been shown to occur also with phage of *Dickeya solani and E. amylovora* phages (Born et al., 2015; Czajkowski et al., 2014). However, there have been phages isolated against the phytopathogen *Pseudomonas syringae pv. actinidiae* which can tolerate extended UV-B doses (Yu et al., 2015). Other potential factors that could cause phage decline on the phyllosphere are desiccation,

temperature, pH as well as certain chemicals produced by plants (Delitheos et al., 1997; Erskine, 1973; Iriarte et al., 2007).

1.11. Bacteriophage application methods for optimal biocontrol performance on plants

One of the limitations to effective phage biocontrol on crops is the possibility of poor persistence on the phyllosphere due to the factors discussed in the previous section. However, a number of methods have been found to reduce this problem. Survival of phage can be improved in the phyllosphere and rhizosphere if they are accompanied by a viable host. This can be an avirulent strain of the pathogen being targeted or indeed another species of bacteria which occurs naturally in that environment (Bae et al., 2012; Iriarte et al., 2012; Svircev et al., 2006). It has also been found that avoiding daylight during application can improve phage-based biocontrol. Indeed it has been demonstrated that applying phage to tomato leaves in the evening resulted in longer phage persistence in the phyllosphere, giving phage more time to infect and kill their bacterial targets (Balogh et al., 2003; Iriarte et al., 2007).

Born *et al* conducted studies with a number of substances to investigate if they gave phage protection against UV and reported that natural extracts from carrot, red pepper, and beetroot all gave protection as did casein, soy peptone and also purified aromatic amino acids, astaxanthin and Tween 80. None of these substances had a compromising effect on phage infection and stability (Born et al., 2015). Thus, it appears that a wide range of substances can enhance phage performance in the phyllosphere with the main requirement being that they need to absorb UV thus limiting phage exposure. Biodegradable polymers have also been shown to give these protective effects (Khalil et al., 2016). In addition, Balogh et al also showed an enhanced phage activity by combining the following preparations with phage, namely (i) 0.5% pregelatinized corn (PCF) and 0.5% sucrose, (ii) 0.5% Casecrete NH400, 0.5% sucrose and 0.25% PCF and (iii) 0.75% skim milk and 0.5% sucrose. These tests were performed in greenhouse trials and in field trials on tomato plants with phages against *Xanthomonas campestris* pv. *vesicatoria.* All formulations were used under a variety of different conditions, but generally demonstrated enhanced disease protection (Balogh et al., 2003).

Soil-based phage delivery is another approach that has been looked at to improve phage persistence in the phyllosphere. Iriarte et al showed that a proprietary mixture of phage (OmniLytics Inc.) active against *X. perforans* strain 97-2 could translocate to the upper leaves of a tomato plant from its roots. They demonstrated that these phages which were applied to soil at levels of 10^8PFU/mL could be detected at titres of 10^4 PFU/g in leaf tissue for 7 days, whereas with a direct foliar application of the same phage mix, phages were undetectable 1 to 2 days after application (Iriarte et al., 2012). This work would suggest that the phage control of foliar plant diseases could be controlled by applying the phages to surrounding soil of a plant rather than by foliar spraying.

1.12. Combination of protective methods appear to be the best direction for bacteriophage biocontrol

There is evidence to support that combining phage with a number of methods used to control crop disease results in better control. The bacterium *Pantoea agglomerans* has been used as a biocontrol agent to suppress the growth of the agent of fire blight *E. amylovora* and is being sold under the band name Bloomtime® (Mikiciński et al., 2016). However, it has been reported that combining this bacterium with phage biocontrol can give enhanced protection that is comparable to that achieved with the antibiotic streptomycin (Boulé et al., 2011; Svircev et al., 2006). A similar observation of enhanced control was seen using a bacteriocin-producing strain of *R. solanacearum* with a phage to combat tobacco bacterial wilt (Tanaka et al., 1990). In another study, combining phage with Acibenzolar-S-methyl (ASM) was shown to have improved protection against bacterial spot of tomato in the field (Obradovic et al., 2004). However, combinations of phage with copper-based pesticides do not appear to produce synergistic effects. Treatment with copper-mancozeb as seen with citrus canker and bacterial spot of citrus fruits did not produce synergy against *Xanthomonas axonopodis* pv.*citri* or *Xanthomonas axonopodis* pv*. citrumelo,* respectively (Balogh et al., 2008). As mentioned previously, this could be due to phage sensitivity to the components of these copper-based sprays.

1.13. Improved understanding of bacterial host diversity should aid phage biocontrol and improve its success in the future

Recent years have seen recognition of the increasing diversity and complexity of bacterial phytopathogens mainly due to advances in molecular techniques (16S rRNA sequencing). For example, *X. campestris* pv*. vesicatoria,* which was previously a single species has since been divided into four (Jones et al., 2004). Another example is of the soft rot *Erwinia* group which has undergone a significant taxonomic reshuffle with the creation of novel species and genera (Gardan, 2003; Hauben et al., 1998; Samson et al., 2005). These developments are very important, as while the afflictions caused by these bacteria may appear identical on their respective crop targets, the phage sensitivities of the pathogens are likely to differ significantly, but nevertheless are likely to have some correlation with their taxonomic groupings. For example, the soft rot *Erwinia* group, which affects potato crops, has more recently been reclassified into two new bacterial genera (*Pectobacterium* and *Dickeya*), and these are relatively distinct from the point of view of phage susceptibilities (Czajkowski, 2016).

1.14. Phytopathogens targeted for phage biocontrol and how they are currently managed

There are a number of important bacterial plant pathogens that have received attention for phage biocontrol in recent years (Table 1.1) as existing approaches are having limited efficacy or their use is restricted in certain regions of the world. The following section discusses selected crop pathogens where phage biocontrol has been evaluated and is showing promise.

1.14.1. *Dickeya* and *Pectobacterium*

Both *Dickeya* and *Pectobacterium* belong to the family of *Enterobacteriaceae*, which collectively can be referred to as the soft rot *Enterobacteriaceae* (SRE). Both genera characteristically produce several cell-wall-degrading enzymes that allow them to infiltrate and macerate the plant tissue on which they feed (Pérombelon, 2002). The plant host range of both bacterial genera is very broad: species belonging to *Dickeya* have been reported to infect ten monocot and eleven dicot families, while those of *Pectobacterium* are reported to infect eleven monocot and sixteen dicot families (Ma et al., 2007).

P. carotovorum subsp. *carotovorum* has a wide host range and global distribution, while *P. atrosepticum* is primarily found in temperate climates with a host range mainly limited to the potato (Pérombelon, 2002). *P. wasabie* and *P. carotovurum* subsp. *brasilensis* are also found to infect potato in a number of regions worldwide (Lee et al., 2014; Waleron et al., 2013). In Europe, *Dickeya dianthicola* is reported to be very important in potato disease, although more recently, a new *Dickeya* species called *D. solani* is being more frequently identified. Both also cause disease in other regions of the world (Toth et al., 2011). The economic impact of these potato infections can be severe. In the Netherlands, they cause annual losses in the seed potato sector of as much €30 million per year and in Israel, potato yield losses due to *Dickeya* have been as much as 20-25% (Prins and Breukers, 2008; Tsror (Lahkim) et al., 2008). In Ireland, blackleg has been reported to be one of the top causes for ware potato losses along with blight (MacConnell, 2002).

With regard to the potato, there are no effective bactericides to protect against SRE and the most effective approach has been through careful culturing practices, involving avoidance of contamination and the removal of diseased plants and/or diseased tissue. Certification systems are also employed. These involve the propagation of seed plants using healthy tissue culture plantlets followed by propagation in greenhouses, and then open field grow-out production. It is accompanied by careful monitoring and removal of diseased plants before releasing for general production. The generation number of these crops is also kept low to limit bacterial build up. However, the success of these certification schemes has been variable and heavily weather dependent (Czajkowski et al., 2011; De Boer, 2004).

1.14.2. *Erwinia amylovora*

E. amylovora, a member of the family of *Enterobacteriaceae,* is the causative agent of fire blight which is a destructive disease that occurs to species of the plant family *Rosaceae.* The disease has been reported in 40 countries across North America, Europe, the Pacific Rim and the Middle East (Bonn and van der Zwet, 2000)*.* It heavily affects apple and pear production in a number of regions, with costs estimated as much as $100 million per year in the USA due to production losses and control measures (Norelli et al., 2003). It is considered to be a quarantine concern in countries

belonging to plant protection agencies of APPPC (Asia and Pacific Plant Protection Commission), COSAVE (Comite Regional de Sanidad Vegetal para el Cono Sur), EPPO (Europe and Mediterranean Plant Protection Organisation) and IAPC (Inter-African Phytosanitary Council) (CABI, 2016).

Pathogenesis typically involves the bacterium entering a susceptible plant host through the nectarthodes of its flowers, but it may also enter the plant through other openings such as wounds (Bubán and Orosz-Kovács, 2003). Once in the plant, it is capable of moving through the intracellular space of parenchyma, where at the latter stages it may reach the xylem vessels. Under favourable conditions, the disease can present itself as wilting, necrosis of tissue and dieback of the plant (Vanneste and Eden-Green, 2000). The bacterium does not produce cell-wall-degrading enzymes but the exopolysaccharide amylovoran, biofilm formation capacity, motility, a type III secretion system and quorum sensing are all understood to be features in its virulence (Piqué et al., 2015).

Traditionally, control of fire blight relies on cultural practices involving the removal of diseased tissue as well as preventative sprays containing copper or antibiotics (Norelli et al., 2003). However, issues with these chemical controls is copper tolerance of the pathogen and also the long term of use antibiotics (such as streptomycin) as a control strategy may be limited in the future, with growing concern of antibiotic resistance and the resulting restricted used of antibiotics for agriculture in certain regions of the world such as EU countries (de León Door et al., 2013; Mayerhofer et al., 2009; Ordax et al., 2006; Russo et al., 2008). As mentioned, biological controls using antagonistic bacteria have shown a capacity for controlling the disease (Mikiciński et al., 2016)

1.14.3. *Ralstonia solanacearum*
R. solanacearum is a Gram-negative soil-borne bacterium. It is considered to be one of most destructive phytopathogens with a host range of up to 200 plant species from over 50 families (Denny, 2007). The bacterium is highly heterogeneous, historically being divided into five races (based on plant host range) and five biovars (based on carbon utilisation) (Denny, 2007). It causes diseases of economically important crops, such as bacterial wilt of tobacco, banana and tomato as well as brown rot of the potato (Sanchez Perez et al., 2008). The bacterium has global distribution (Sanchez

Perez et al., 2008), and with regard to tomato and potato production, has quarantine status in the EU (Anonymous, 2000). The species has a considerable economic impact: for example, brown rot of the potato has been estimated to exceed more than €950 million in losses per year worldwide (Scherf et al., 2010). Infection begins by the bacterium entering the host plant through its roots where it will then colonise the xylem. Infection typically leads to the development of yellowing of the plant, stunted growth, wilting and death, although the bacterium is also capable of asymptomatic infections (Sanchez Perez et al., 2008). Typical methods of control include the use of cultural practices such as selection of planting time, crop rotation, using clean seedlings and the use of resistant cultivars (Mariano et al., 1998). However, the use of such cultivars has shown a negative correlation between resistance and yields (Yuliar et al., 2015). Also, resistance possessed by these cultivars tends to be strain specific (Wang et al., 2000).

1.14.4. *Pseudomonas syringe*

The bacterial phytopathogen *P. syringea* belongs to the class of *Gammaproteobacteria* (Hirano and Upper, 2000). The species is currently subdivided into more than 50 pathovars, with different pathovars representing different strains with differing plant host ranges (Hirano and Upper, 1990; Parkinson et al., 2011). Strains of most pathovars typically exhibit narrow host ranges, with pathovar P. *syringea* pv. *syringea* being an exception, having been reported to infect more than 80 plant species (Hirano and Upper, 2000).

P. syringea pv. *tomato* causes necrotic lesions surrounded by a yellow chlorotic halo on tomato, a disease known as bacterial speck (Cruz et al., 2010). The pathovar can also infect members of genera of *Arabidopsis* and *Brassica* in a laboratory setting (Elizabeth and Bender, 2007). The disease reduces yields while also affecting fruit quality (Fatmi, 2003). Pathogenesis by the bacterium involves the invasion of plant tissue from natural openings, such as stomata, where a type III secretion system plays a major role in its virulence with the release of effectors to overcome the plant immune system (Xin and He, 2013). It is spread by contaminated tomato seeds but can also survive as an epiphyte for extended periods on tomato plant surfaces and is dispersed in windblown rain (McCarter, 1983; Preston, 2000; Smitley and McCarter, 1982). Control of the

organism typically involves the use of uncontaminated seeds and the use of bactericides (copper and streptomycin) to limit its spread (Fatmi, 2003; Preston, 2000). However, copper tolerant strains of the bacterium have been reported (Alexander et al., 1999).

1.14.5. *Xanthomonas* species

Xanthomonas is a large genus, which belongs to the class of *Gammaproteobacteria*, containing at least 27 official species, many of which also possess a number of pathovars. Collectively, the genus host range is broad: infecting around 400 plant hosts, a number of which are important crops such as rice, banana, tomato and citrus fruits. Species and pathovars of this genus typically exhibit a high degree of host- as well as tissue-specificity, invading either the xylem or intercellular spaces of the mesophyll parenchyma tissue (Ryan et al., 2011).

X. campestris pv. *vesicatoria* is the causative agent of bacterial spot disease of tomato and pepper, with the disease having been identified in many countries worldwide (Jones et al., 2005). This tomato disease can be very severe with yield losses of up to 50% reported for tomatoes, grown both in greenhouses and fields in the USA and Caribbean (Camesano, 2015). The disease is caused by the bacterium entering the plant through stomata or wounds. The bacteria then colonise the intercellular space of the plant, inducing water-soaked lesions that later become necrotic, which can result in defoliation and severely spotted fruit (Thieme et al., 2005). Control of the disease has involved preventative cultural practices such as avoiding unnecessary crop damage and using uncontaminated seed, but also includes the use of resistant cultivars as well as chemical controls with copper or streptomycin (Goode and Sasser, 1980). However, the use of resistant cultivars has not always been successful and there have been reports of bacteria developing resistance to the above two agents (Goode and Sasser, 1980; McDonald and Linde, 2002; Ritchie and Dittapongpitch, 1991).

1.14.6. *Xylella fastidiosa*

Xylella fastidiosa belongs to the class of *Gammaproteobacteria.* It is a xylem-limited phytopathogen that requires insect vectors (such as sharpshooters) for its distribution and infection of its host plants (Chatterjee et al., 2008). It causes disease on a number of crops such as the grape, citrus, almond, peach and coffee (Hopkins and Purcell,

2002). While it has primarily been contained in the Americas, it has been identified in Europe in recent years causing disease on olive trees (Hopkins and Purcell, 2002; Loconsole et al., 2014). Disease caused by the bacterium is believed to be induced by the formation of biofilm aggregates in the vascular system, which restricts the movement of nutrients and water throughout the plant (Chatterjee et al., 2008). It causes Pierce disease of the grapevine, a highly destructive infection, which heavily affects grape production in the USA, and has been estimated to cost as much as $104.4 million annually to the state of California (Tumber et al., 2014). Existing control methods have been limited in their management of the disease and include removal of infected plants and control of the infected insect vector populations with neonicotinoid-based insecticides (Janse and Obradovic, 2010). However, the use of these insecticides has seen restrictions in recent years due to their possible effects on honey bee populations (Anonymous, 2013; Lu et al., 2014).

1.15. A critical summary of recent phage biocontrol studies on crops

There is growing evidence showing that phages have promising biocontrol applications for a number of plant diseases in different crops. The following section describes recent studies that have been conducted since the year 2000 and the findings from these are summarised in Table 1.1.

The most common crops that appear to benefit from the application of phages for biocontrol in the recent scientific literature are the potato and the tomato, as both have been the focus of numerous recent studies. The bacterial pathogens in the case of the potato are predominantly the SRE. As mentioned above, one of the most important SREs in Europe is *D. solani;* and the potential of phage to control this phytopathogen have been assessed indicating strong potential for disease control. For example, Adriaenssens *et al* (2012) conducted a bioassay and a field trial using phage (LIMEstone1). The bioassay involved the incubation of seed tubers (cultivar Bintje), which had either been inoculated with the bacteria or co-inoculated with the bacteria and the phage (multiplicity of infection [MOI] of 100). They showed that tubers inoculated with the bacteria alone would experience 40% maceration of tuber tissue, while those co-inoculated with the phage and bacteria exhibited no more than 10% maceration of tuber tissue. Similar results were observed with the seed tuber cultivar

Kondor. The field trial using the same phage against the same pathogen also suggested it was capable of exerting this biocontrol effect *in-planta,* as phage treated infected seed potatoes resulted in higher crop yields than those without phage treatment. Similar findings were reported by Czajkowski et al (2014) who also isolated phages specific for *D. solani*. These workers conducted bioassays with tuber slices incubated with the bacterial pathogen with or without phages (MOI of 0.01) and showed that the application of phages could prevent potato tuber tissue maceration by up to 70%. SREs other than *D. solani* were also studied for their susceptibility to phages by the same group. They found that the application of phages (MOI of 0.01) to control *P. carotovorum subsp carotovorum* and *P. wasabie* destruction could prevent damage of up to 80% on tuber slices and up to 95% on whole tubers against tissue maceration from a mixed bacterial infection (Czajkowski et al., 2015). Such data is highly encouraging as many SRE infections tend to result for a mixture of genera/species. Aside from potato, SRE infections have also been controlled by phage in lettuce, with high levels of disease prevention being reported (Lim et al., 2013). Aside from the SRE problem, potato infections from the Gram-positive bacterium *Streptomyces scabies* result in the formation of a corky lesion (known as common scab) on the tuber and indeed other root vegetables also, as well as causing the reduced growth of seedlings (Lerat et al., 2009). This pathogen has also been successfully treated in potato by phage biocontrol and thus has implications for other crops also as demonstrated by Goyer (2005). In conclusion, the above studies indicate strong potential for phage-based control of these diseases.

Another crop which the focus of several studies in the context of phage therapy has been the tomato, which is commonly infected by *R. solanacearum* (also causes brown rot in the potato) and *X. campestris* pathovars. Again, phage biocontrol approaches have been demonstrated to give a significant reduction in bacterial wilt (*Ralstonia*) and leaf spot caused by *Xanthomonas*. Indeed, the successful trials against *R. solanacearum* reported by Mansfield et al (2012) are significant considering the wide host range of the bacterium. Similarly, in the case of *Xanthomonas*, the observed beneficial effect of the application of phages can also be extrapolated to other plants affected by pathogens belonging to the same genus. Indeed studies on the elimination

of *Xanthomonas* using phage have been conducted with successful outcomes on both grapefruit and orange (Balogh et al. 2008) as well as onion (Lang et al. 2007). A variety of other crop infections have also been reduced in severity in other phage biocontrol studies. These include *Pseudomonas* infections of mushrooms (brown blotch) and leeks (bacterial blight) and infection of the grapevine by *Xylella* (Das et al., 2015).

1.16. The commercialisation of phage for biocontrol in crop disease

In recent years a number of phage biocontrol products have reached the market. A USA based company Omnilytics was the first company to receive registration (from the US Environmental protection agency) for their phage-based biopesticide product Agriphage. The product is designed for the control of bacterial spot or speck of tomatoes and peppers (specific for *X. campestris* pv. *vesicatoria*, or *P. syringae* pv. *tomato*). This product has also received an OMRI listing making it suitable for use by commercial organic growers (OmniLytics, 2006). A Hungarian company Enviroinvest was the second company to receive registration for their biopesticide named Erwiphage for the control of fire blight of apple trees (specific for *Erwinia amylovora*) (Enviroinvest, n.d.). There is also a Scottish company, APS biocontrol, which has developed a bacteriophage-based wash solution (Biolyse) for potatoes tubers, which is to be used for prevention of soft rot disease (specific against soft rot *Enterobacteriaceae*) during storage (APS Biocontrol Ltd, n.d.). Interestingly this product has been reported to be used by the Tesco supermarket chain (Branston, 2012).

However, there are delays in some regions of the world with phage biocontrol due to legislation that hinders the use of phage biocontrol approaches for the of control of bacterial plant diseases. A problem with phage-mediated biocontrol is that phage mixtures/cocktails need to be updated constantly in order to lyse as many newly emerging strains of the target bacterium as possible. This approach is used by Omnilytics (OmniLytics, 2004). This allows a phage cocktail to be adapted to the relevant disease-causing bacterial strains in a given situation, also facilitating counteraction of any phage resistance development during the phage application. However, EU regulations (1107/2009 EC) require that any change to one of the components of a phage cocktail would require reregistering which requires time and expense, making the US approach currently unfeasible in the EU (Doffkay et al., 2015).

Legislation governing phage biocontrol may need to become more malleable in the EU for the best application and performance of phage products as biopesticides.

1.17. Other phage applications of the past and possible future with regards to phytopathogens

Phage typing schemes have been employed for a number of phytopathogens for epidemiology studies (Ahmad et al., 2014; Toth et al., 1999). These systems allow the identification of a particular strain of a species based on their susceptibility to series of phages. The downfall of this method, however, is that it depends on the isolation of pure cultures for identification as well as the maintenance of stocks of typing phage as well as host strains for which to propagate them. Nowadays, studies of phytopathogens have moved away from phage typing due to its tendency to generate false positives and false negatives results as well as its low resolution and the development of new and improved molecular techniques (Czajkowski, 2016)

A number of phage-based detection systems have been developed for human and animal pathogens (van der Merwe et al., 2014). Recently, however, work has been published on the promising application of these methods for the detection of plant pathogens. Such a detection system has been developed for *R. solanacearum,* which is based on detection of the bacterium by phage propagation followed with quantitative PCR (qPCR). Samples that contain the bacterium will cause added phage titres to increase, these titre increases can then be detected using qPCR. This method was found to be faster than conventional methods with greater sensitivity allowing detection of 10^2 CFU/g of soil, 10^3 CFU/ml from drainage water from potted plants and 10^2 CFU/g in 0.1 g of leaf tissue. The method also doesn't require the destruction of a plant for the detection of bacterium unlike those currently used to detect *R. solanacearum* (Kutin et al., 2009). It is possible to engineer phage of phytobacteria into reporter systems that can emit a detectable bioluminescent signal during infection. A "*luxAB*-tagged" reporter phage was developed for *Pseudomonas cannabina pv. alisalensis* (agent of bacterial blight of crucifers) which was shown capable of detecting the bacteria within minutes. This phage was also capable of emitting a detectable signal during infection of both cultures and diseased plant samples (Schofield et al., 2013). Both mentioned systems have advantages over other

molecular detection methods, in that phage propagation requires active metabolism, conveniently limiting it to viable bacterial cells.

1.18. Discussion

Effective control of plant disease typically calls for a disease control strategy that involves a number of integrated approaches. Currently, the use of phage biocontrol is emerging, but as yet uncommon practice. However, phages do possess several properties which can add to the arsenal of controls for crop diseases. They are natural, making them suitable for organic farming. They can be used to create phage cocktails with tailored host ranges. Also, phages naturally have the potential to evolve to adapt to overcome phage-resistance or overcome new strains of bacteria. They can be combined with other chemical or biocontrol agents. A possible limitation to their use is their sensitivity to UV light and to certain soil conditions. However, approaches have been found to overcome some of these limitations with the use of UV protectant formulas and timing of the application of phage to crops to avoid interaction with chemical pesticides and exposure to UV light. In addition to biocontrol applications, there is also good potential for phage-based diagnostics for plant pathogenic bacteria with a high sensitivity aimed specifically at viable bacteria.

Many pesticide companies are moving away from investment in chemical pesticides and increasingly directing their attention to biopesticides. The pesticide market is worth $56 billion with the biopesticides forming only $2-3 billion of this. However, the growth of the biopesticide sector is expected to outpace chemical pesticides in the future (Marrone, 2014). This change is believed to be due to increasing customer demand for chemical residue free foods and increasing legalisation on the use of synthetic pesticides in certain regions in the word. In addition, many biopesticide products are potentially cheaper to develop and quicker to bring to the market (Marrone, 2014). With this economic environment, one can expect to see increased activity in the development of phage biocontrol as a viable approach for crop disease control in the future.

Table 1.1. Summary of bacteriophage biocontrol experiments which have been conducted since the year 2000 to the present.

Pathogen	Host	Disease	Information	Reference
Streptomyces scabies	Potato	Common scab	Seed tubers treated with phage ΦAS1 resulted in producing tuber progeny with reduced levels of surface lesion of scab (1.2%) compared with tubers harvested from non -treated seed tubers (23%).	McKenna et al., 2001
Xanthomonas campestris pv. vesicatoria	Tomato	Bacterial spot	Greenhouse experiments with formulated phage cocktails could reduce disease severity with formulated phage cocktails providing better protection in comparison to unformulated. A similar effect was found in three consecutive field trials.	Balogh et al. 2003
Xanthomonas campestris pv. vesicatoria	Tomato	Bacterial spot	In field experiments, phage treatment was comparable to disease control with copper-mancozeb. Combination of phage and plant activator (ASM) resulted in enhanced control.	Obradovic et al., 2004
Streptomyces scabies	Radish	Common scab	Phages Stsc1 and Stsc3 could prevent disease development by treating radish seedlings. Non-treated radishes had 30% less weight than negative control, with phage treated radishes having masses similar to the negative control.	Goyer 2005
Xanthomonas axonopodis pv. allii	Onion	Xanthomonas leaf blight of onion	Field trial showed that weekly and biweekly applications of phage could reduce disease severity, a result which was comparable to treatments of weekly applications of copper-mancozeb.	Lang et al., 2007
Xanthomonas axonopodis pv. *citri*	Grapefruit	Asiatic citrus canker	Five greenhouse experiments utilising phage treatment could reduce disease severity by 59%. However, using a skim milk formulation of phage did not have increased disease control. Phage treatment was also capable of reducing	Balogh et al. 2008

			disease occurrence in a citrus nursery. Control was less effective than copper-mancozeb. The combination did not give increased disease control.	
Xanthomonas axonopodis pv. citrumelo	Orange	Citrus bacterial spot	Phage treatments reduced citrus spot occurrence by 35% and 48% in two trials in a commercial citrus nursery. Control was equal or less effective than copper-mancozeb. The combination did not give increased disease control	Balogh et al. 2008
Ralstonia solanacearum	Tomato	Bacterial wilt	Tomato plants treated with phage ΦRSL1 showed no symptoms of bacterial wilt during the experimental period; whereas all untreated plants showed wilting 18 days post infection.	Fujiwara et al., 2011
Pseudomonas tolaasii	Mushrooms	Brown blotch Disease	The formation of blotches was completely blocked by co-incubation of phages with the pathogen.	Kim et al. 2011
Erwinia amylovora	Pear, apple trees	Fire blight	Phages ΦEa1337-26 and ΦEa 2345 reduced infection of detached pear tree blossoms by 84% and 96%, respectively with *Pantoea agglomerans* as a carrier. Also, infection of potted apple tree blossoms could be reduced by 54% with phage ΦEa1337-26 and *P. agglomerans*. Control achieved was comparable to streptomycin.	Boulé et al. 2011
Ralstonia solanacearum	Tomato	Bacterial wilt	Simultaneous treatment of phage PE204 with R. solanacearum of the rhizosphere of tomato completely inhibited bacterial wilt. However, pre-treatment with phage before the inoculation of the pathogen was not effective with control of bacterial wilt, whereas post-treatment of PE204 delayed disease development.	Bae et al., 2012

Dickeya solani	Potato	Soft rot/ Blackleg	Phage vB_DsoM_LIMEstone1 and vB_DsoM_LIMEstone2 reduced soft rot of inoculated tubers in bioassays and in field trials which produced a potato crop with higher yields.	Adriaenssens et al., 2012b
Pectobacterium carotovorum subsp. carotovorum	Lettuce	Soft rot	Greenhouse trials showed that phage PP1 could significantly reduce disease development on lettuce plants.	Lim et al., 2013
Dickeya solani	Potato	Soft rot	Bioassays with phage ΦD1, ΦD2, ΦD3, ΦD4, ΦD5, ΦD7, ΦD9, ΦD10, ΦD11 could reduce the incidence of soft rot by up to 30–70% on co-inoculated potato slices with pathogen and phage.	Czajkowski et al., 2014
Xylella fastidiosa	Grapevines	Pierce's Disease	X. fastidiosa levels in grapevines were significantly reduced on pre and post inoculation of a four phage (Sano, Salvo, Prado and Paz) cocktail. Pierce disease symptoms could be stopped using phage treatment post infection as well as applying phage prophylactically to grapevines.	Das et al., 2015
Pectobacterium carotovorum subsp. carotovorum, *Pectobacterium wasabiae,* *Dickeya solani*	Potato	Soft rot	Bioassays with phage ΦPD10.3 and ΦPD23.1 could reduce the severity of soft rot of tubers by 80% on potato slices and 95% with whole tubers from a mixed pathogen infection.	Czajkowski et al., 2015
Pseudomonas syringae pv. porri	Leek	Bacterial blight	Specific-bio-assays demonstrated the in-*planta* efficacy of phages KIL1, KIL2, KIL3, and KIL3b. However, phage cocktail of six phages (KIL1, KIL2, KIL3, KIL4, and KIL5 and KIL3b), were tested with two parallel field trial experiments in three locations which showed variable results. In one trial, symptom development was attenuated.	Rombouts et al., 2016

Chapter 2. Materials and methods

2.1. Bacterium isolation, bacteriophage isolation and cultivation conditions

2.1.1. Cultivation conditions of bacteria and phages

For chapters 3, 4, 5 and 6, all bacteria cultures were grown at 25°C unless stated otherwise, using Lysogeny broth (LB) medium (Sigma- Aldrich, St. Louis, MO, USA). For chapter 7, all *Erwinia* and *Pantoea* strains were grown in LB medium at 28°C.

Phages CB1, CB3, CB4, CB5 and CB7 were routinely cultivated at 25°C with LB agar (1.5% w/v agar) and LB overlays (0.4% w/v agar) using *P. atrosepticum* strains DSM 18077 (for phage CB1) and DSM 30186 (for phages CB3, CB4, CB5, CB7). Phage CBB was routinely cultivated at 25°C with LB overlays using 0.2% w/v agarose as described by Serwer *et al* (2007), on LB agar (1.5% w/v agar) and using *P. carotovorum* subsp. *carotovorum* CB BL19-1-37 as the host strain. Stocks of phage were produced according to the plate lysis method as previously described (Sambrook and Russell, 2001a). Additionally, phages were also propagated by lysis in broth, a volume of LB broth was inoculated with 1% overnight culture, this was grown with agitation to an OD_{600} of ca. 0.2. To this, phage was added to a final concentration of ca. 1×10^9 plaque forming units (PFU)/mL and further grown with agitation until complete clearing of broth was observed. The lysate was then filter sterilised (0.45 μm pore-size filter, Sarstedt, Nümbrecht, Germany) or underwent CsCl gradient purification (as described in section 2.2.4.). To determine phage titre, a 10-fold dilution series of phage stock was tested using the double overlay method as previously described (Sambrook and Russell, 2001b).

Phage Y3 was routinely cultivated at 28°C with the soft agar overlay method (Adams, 1959) using LB plates and LC soft agar (LB supplemented with 10 mM $CaCl_2$ and 2 mM $MgSO_4$; 0.4% agar), and *E. amylovora* 1/79 as the host strain.

2.1.2. Isolation of *Pectobacterium*

Crystal violet pectate (CVP) agar was used to isolate *Pectobacterium* strains from stems of potato plants presenting symptoms of blackleg (Hélias et al., 2012). Bacterial identification was achieved by using biochemical and physiological tests (Perombelon and Van Der Wolf, 2002), genus and species-specific PCRs (Darrasse et al., 1994; De

Boer and Ward, 1995; Kang et al., 2003) and MALDI-TOF mass spectrometry (Bruker Daltonics Biotyper, Billerica, MA, USA).

2.1.3. Bacteriophage isolation

Five grams of soil (for the isolation of CB1, CB3, CB4, CB5 and CB7) or 5ml of activated sludge (for the isolation of CBB) was enriched with 30 mL of LB broth, followed by the addition of 300 µL of an overnight culture of *P. atrosepticum* (DSM 18077, DSM 30184, DSM 30185 and DSM 30186) or *P. carotovorum* subsp. *carotovorum* (CB BL19-1-37). After incubation for 18 h at 25°C, this was centrifuged to pellet solid matter with supernatant then being filtered (0.45 µm pore-size filter, Sarstedt, Nümbrecht, Germany). The supernatant was spotted (10 µL) onto LB overlays seeded with different strains of *Pectobacterium*. Phages were isolated by picking individual plaques and then re-plating and re-isolating to ensure purity(Sambrook and Russell, 2001c).

Erwinia phage Y3 was originally isolated from soil as described previously (Born et al., 2011). Briefly, 10 g of soil was enriched in 200 mL SM-buffered (50 mM Tris, 100 mM NaCl, 8 mM $MgSO_4$, pH 7.4) LB broth, which had been inoculated 1:20 with an overnight culture of *E. amylovora* 1/79. After overnight incubation at 28°C, filtered supernatants were spotted onto a freshly growing lawn of *E. amylovora* 1/79 and plaques were picked with a Pasteur pipette. After isolation, Y3 was propagated to high titres with the soft agar overlay method.

2.2. Bacteriophage manipulations

2.2.1. Host range

Host range of phages was tested by spotting (µL) a serial dilution (neat to 10^{-9}) of a phage suspension onto LB overlays seeded with bacteria of interest. The efficiency of plaquing (EOP) values was determined for sensitive strains by dividing phage titre on target bacterium by phage titre on host bacterium. In the case of the CB1-like phages, the determination of EOP values employed the use of a phage lysate with a titre of $1x10^7$ PFU/mL. Bacterial strains used in host range study of *Pectobacterium* phages CB1, CB3, CB4, CB5, CB7 and CBB are listed in Appendix, Table S2.1.

2.2.2. Single step growth curve assay

The single step growth curve assay was conducted in a similar manner as described previously (Park et al., 2012; Yang et al., 2010). The host bacteria (strain DSM 30186) were grown to an OD_{600} of 0.20-0.23 (c. $1x10^8$ colony forming units (CFU)/mL), followed by centrifugation of 2 mL in a microfuge to pellet bacteria. The pellet was resuspended in 1 mL of LB broth phage suspension to yield an approximate multiplicity of infection (MOI) of $5x10^{-4}$ following incubation at 25 °C for one min. This was then centrifuged to pellet bacteria, and the supernatant was removed, thus separating bound from unbound phages. The bacterial pellet with bound phage was then resuspended in 10 mL of LB and incubated aerobically in a water bath at 25 °C with agitation at 60 rpm. At five-min intervals, aliquots were removed to measure phage titre by the overlay method. Based on the number of PFU/mL of each replicate, the latent period and the burst size were determined, by dividing the average PFU/mL of the latent period by the average PFU/mL of the last four time points of the experiment.

2.2.3. Biophysical stability

Phage stability was tested by incubating phage suspension of 10^6 PFU/mL in SM buffer (50 mM Tris.HCl pH 7.5, 100 mM NaCl, 8 mM $MgSO_4$) at different temperatures for one hour and incubating phage suspension in pH buffer ranging from 2 to 12 (10 mM trisodium citrate, 10 mM boric acid, and 150 mM KCl, adjusted with NaOH or HCl) for 24 h (Adriaenssens et al., 2012b).

2.2.4. CsCl gradient purification

Isopycnic centrifugation through CsCl gradients was performed on phage Y3, as previously described (Sambrook and Russell, 2001d). However, for phages CB1, CB3, CB4, CB5, CB7 and CBB, this was conducted with a number of modifications. A high titre phage lysate ($>1 \times 10^9$ PFU/mL), was precipitated using polyethylene glycol (15% w/v PEG8000, 1 M NaCl) at 4 °C overnight and centrifuged, after which the pellet was resuspended in TMN buffer (10 mM Tris-HCl pH 7.4, 10 mM $MgSO_4 \cdot 7H2O$, 0.5 M NaCl), and where necessary a chloroform phase separation step (1:1) was conducted to remove debris. The resulting phage preparation was placed onto a CsCl step gradient composed of 1.3, 1.5, and 1.7 g/mL layers and spun in a 100 Ti rotor (Beckman Coulter,

Brea, CA, USA) at 200,480 g for 3 h at 4 °C. Resulting phage bands were collected and subjected to dialysis with two changes of Tris-HCl buffer (10 mM, pH 7.5) at 4 °C.

2.2.5. Transmission electron microscopy

CsCl purified phages were absorbed onto freshly prepared ultra-thin carbon films. Negatively staining of phages CB1, CB3, CB4 involved the use of 2% (w/v) uranyl acetate and 1% phosphotungstic acid, for phages CB5 and CB7 the use of 1% (w/v) uranyl acetate and for phage CBB the use of 2% (w/v) uranyl acetate. Micrographs were taken using a Tecnai 10 transmission electron microscope (FEI Thermo Fisher, Eindhoven, the Netherlands) at an acceleration voltage of 80 kV with a MegaView G2 CCD-camera (EMSIS, Muenster, Germany).

For transmission electron microscopy of phage Y3, CsCl-purified phages were negatively stained with 2% ammonium molybdate and then analysed as previously as described (Klumpp et al., 2008).

2.3. Bacteriophage genome and proteome analysis

2.3.1. Genomic DNA isolation and restriction digestion

Genomic DNA was extracted from phages CB1, CB3, CB5, CB7 as previously described (Pickard, 2009). CsCl purified phage particles were treated with DNase and RNase, followed by treatment with 10% SDS and proteinase K followed by DNA extraction with phenol: chloroform: isoamyl alcohol (25:24:1 v/v) and chloroform: isoamyl alcohol (24:1 v/v).

For the extraction of genomic DNA from phage Y3, CsCl-purified phage particles had free nucleic acids degraded with treatment of 15 min at 37°C by DNase I (final concentration: 1 U/ml) and RNase A (50 µg/ml). Nucleases were inactivated and phage capsids degraded 1 h at 56°C with EDTA (20 mM, pH 8.0), proteinase K (50 µg/ml), and SDS (0.5% [w/v]). DNA was then purified using a phenol-chloroform extraction followed by ethanol precipitation (Sambrook and Russell, 2001e)

DNA samples were digested with BamHI, Ssp1 and ClaI according to manufacturer's protocols (New England BioLabs, Ipswich, MA, USA). The digested DNA was analysed by agarose gel electrophoresis.

2.3.2. Genomic DNA sequencing

Prior to sequencing, DNA quality and quantity were estimated using both a Nanodrop (ND-1000, Thermo Fisher, Waltham, MA, USA) and by visualisation after agarose gel electrophoresis.

For phages CB1, CB3, CB4 and CB7, genomic sequencing was outsourced to the Centre for Genomic Research at the University of Liverpool, UK. Illumina MiSeq system, with a TruSeq DNA Nano LT library sample preparation kit for library preparation. Library quality was assessed using the Agilent Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and Qubit measurements prior to being sequenced with paired-end reads of 2 ×250 bp. *De novo* assembly was conducted using Spades genome assembler v.3.10 (St. Petersburg, Russia).

The genome of phage CBB was sequenced with a high throughput Illumina MiSeq System outsourced at Nucleomics Core (VIB, Belgium). Libraries were processed with a custom NEBNext® Ultra™ DNA Kit to generate 500-bp fragments with individual barcodes. The quality of each library preparation was controlled using an Agilent Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and Qubit measurements, before being pooled together with a non-homologous genome and sequenced with 2x150 bp paired-end reads. Demultiplex, quality controlled (above Q30) and trimmed reads were *de novo* assembled using CLC Bio Genomics Workbench v7.0 (Aarhus, Denmark) into a single contig.

Genome sequencing of phages CB5 and Y3 was outsourced to GATC Biotech (Konstanz, Germany). To conduct sequencing DNA libraries were first created by DNA fragmentation, adapter ligation followed by a size selection and amplification. DNA libraries were then measured and quantified on a fragment analyser before sequencing with 2 x 300 bp paired-end reads using the Illumina Hiseq system (Illumina, San Diego, CA, USA). The *de novo* assembly was performed using default parameters with CLC Genomics Workbench v8.0 (Aarhus, Denmark).

2.3.3. BAL-31 nuclease treatment of genomic DNA

A total of 40 µg of phage genomic DNA was digested with BAL-31 (0.5 units per µg) (New England BioLabs, Ipswich, MA, USA) at 30°C. Over a series of time intervals, aliquots were taken, treated with EGTA for BAL-31 deactivation and then subjected to digestion with restriction enzyme BglII at 37°C (New England BioLabs, Ipswich, MA, USA) (Klumpp et al., 2008). Resulting genomic DNA was visualised by agarose gel electrophoresis.

2.3.4. Pulsed-field gel electrophoresis

Pulsed-field gel electrophoresis (PFGE) was conducted as described by Lingohr et al. (2009). A CsCl purified phage suspension was mixed with an equal volume of low-melting-point agarose (Bio-Rad Laboratories, Hercules, CA, USA). Prepared plugs were placed in lysis buffer (50mM Tris, 50mM EDTA, 1% SDS) and digested with proteinase K for 2 h at 54 ˚C. Plugs were then washed twice with TE buffer (10mM Tris, 1mM EDTA [pH8.0]) and subjected to electrophoresis on a 1% agarose gel using Bio-Rad CHEF-DR® II PFGE apparatus (Bio-Rad Laboratories, Hercules, CA, USA) at 6V/cm (200V) with 60-120 switch time ramp for 24 h. Yeast chromosome PFGE markers (Bio-Rad Laboratories, Hercules, CA, USA) were used to allow estimation of phage genome size.

2.3.5. Bioinformatic analysis

2.3.5.1. Identification of ORFs, tRNA genes, putative protein function, promoters and terminators and the creation of circular genomic maps.

Open reading frames (ORFs) of CB1, CB3, CB4, CB5, CB7 and CBB were predicted with GLIMMER (Delcher et al., 1999) and GenmarkS (Besemer et al., 2001). The RAST server (http://rast.nmpdr.org/; (Becker et al., 2005)) was employed with GLIMMER (Delcher et al., 1999) to predict ORFs of Y3.

Further analysis of predicted ORF gene products was conducted with BLASTP (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins),   Pfam (http://pfam.xfam.org/search#tabview=tab1; (Finn et al., 2015)), InterProScan (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3998142/;(Mitchell et al., 2014)) and HHpred (https://toolkit.tuebingen.mpg.de/#/tools/hhpred; (Söding et al., 2005)). With the detection of ORFs with transmembrane domains and lipoprotein cleavage signal

being identified with the use of TMHMM v.2 (http://www.cbs.dtu.dk/services/TMHMM/; (Krogh et al., 2001)) and LipoP v.1 (http://www.cbs.dtu.dk/services/LipoP/; (Juncker et al., 2003)), respectively. The molecular weights of the predicted ORFs were estimated using the batch protein molecular weight determination of the sequence manipulation suite (http://www.bioinformatics.org/sms2/protein_mw.html). The presence of transfer RNA genes was investigated with the use of tRNAscan-SE (http://lowelab.ucsc.edu/tRNAscan-SE/; (Lowe and Eddy, 1997)) and ARAGORN (http://130.235.46.10/ARAGORN/; (Laslett and Canback, 2004)). Potential Rho-independent terminators in were identified using ARNold (http://rna.igmors.u-psud.fr/toolbox/arnold; (Naville et al., 2011)) with Mfold QuikFold (http://unafold.rna.albany.edu/?q=DINAMelt/Quickfold; (Zuker, 2003)) using RNA energy rules 3.0 to verify predictions, with putative single-stranded hairpin promoters identification being assisted with Mfold QuikFold using DNA energy rules.

Putative promotors of phages CB7, CBB and Y3, were determined by the submission of 100bp of DNA sequence upstream of each gene to MEME (Multiple Em for Motif Elicitation) (http://meme-suite.org/tools/meme) (Bailey et al., 2009). Codon usage of CBB was analysed using the University of Georgia's amino acid and codon usage statistics services (http://www.cmbl.uga.edu/software/codon_usage.html). Circular genome maps of CB7 and Y3 were drawn using GCview (http://stothard.afns.ualberta.ca/cgview_server/ ; (Grant and Stothard, 2008)).

2.3.5.2. Comparative genomics of *Pectobacterium* CB1-like phages

The linear genomic comparison maps of CB1 and other N4-like phages was created with the use of either BLASTN or TBLASTX to determine genome homology and was visualised with Easyfig (Sullivan et al., 2011). Artemis Comparison Tool (ACT) was used for the identification of feature variations between the genomes of the CB1-like phages, with homology being assessed with BLASTN (Carver et al., 2005). Phylogenetic analysis employed the use of the DNA polymerase and vRNA polymerase proteins of 38 N4-like phages as well as those of CB1-like phages (Appendix, Table S2.2) using MEGA7 (Kumar et al., 2016), involving the use of MUSCLE for sequence alignment

(Edgar, 2004), with the construction of phylograms using the maximum likelihood (ML) method based on the Jones–Thornton–Taylor model (Jones et al., 1992), with the robustness of the trees being assessed with bootstrapping (1000). The heat map comparing the genomes of 38 N4-like phages and CB1-like phages was generated using Gegenees, using accurate parameters (fragment length: 200 bp; step size: 100 bp; threshold: 0%) (Ågren et al., 2012).

2.3.5.3. Comparative genomics of *Pectobacterium* phage CB5

To determine shared proteins among phiM1-like phage proteomes Coregenes 3.5 (http://gateway.binf.gmu.edu:8080/CoreGenes3.5/custdata.html; (Turner et al., 2013)) was used. Translated ORFs from phage φM1 were searched against hidden Markov model profiles downloaded from the prokaryotic Virus Orthologous Groups (pVOGs) database (PMID: 27789703, (Grazziotin et al., 2017)) using hmmscan (PMID: 22039361, (Eddy, 2011)) with an E-value cutoff of 1×10^{-3}. Matches to pVOG profiles were considered significant at an E-value of $\leq 1 \times 10^{-15}$ and $\geq 35\%$ coverage of the profile HMM. The linear genomic comparison maps were created with the use of either BLASTN or TBLASTX, to determine similarity, and then visualised with Easyfig (Sullivan et al., 2011). Phylograms were generated based on the amino acid sequence of the major capsid protein of phage CB5 and 52 members of *Autographivirinae* (Appendix, Table S2.3) using MEGA7 (Kumar et al., 2016), applying MUSCLE for sequence alignment (Edgar, 2004) with the construction of phylograms using the maximum likelihood (ML) method based on the Whelan and Goldman substitution model (Whelan and Goldman, 2001), with the robustness of the trees was assessed by bootstrapping (1000). VICTOR was employed using all pairwise comparisons of the amino acid sequences (same phages as described previously) which employs the Genome-BLAST Distance Phylogeny (GBDP) method (Meier-Kolthoff et al., 2013) under settings recommended for prokaryotic viruses (Meier-Kolthoff and Göker, 2017). The resulting intergenomic distances (including 100 replicates each) were used to infer a balanced minimum evolution tree with branch support via FASTME including SPR postprocessing (Lefort et al., 2015) for each of the formulas D0, D4 and D6, respectively. The trees were rooted at the midpoint (Farris, 1972) and visualised with

FigTree (Rambaut, 2006). Taxon boundaries at the species, genus, subfamily and family level were estimated with the OPTSIL program (Göker et al., 2009), the recommended clustering thresholds (Meier-Kolthoff and Göker, 2017) and an F value (fraction of links required for cluster fusion) of 0.5 (Meier-Kolthoff et al., 2014). The heat map comparing the genomes of phage CB5 and 31 phages of *Autographivirinae* was generated using Gegenees utilising TBLASTX, using accurate parameters (fragment length: 200 bp; step size: 100 bp, threshold set to 5%) (Ågren et al., 2012). Alignment of the RNAP proteins for the examination of catalytic active residues and residues of the recognition and specificity loop was conducted using MUSCLE on MEGA7.

2.3.5.4. Comparative genomics of *Pectobacterium* phage CB7

To determine shared proteins among phage proteomes of *Cr3virus* Coregenes 3.5 (http://gateway.binf.gmu.edu:8080/CoreGenes3.5/custdata.html; (Turner et al., 2013)) was used. The linear genomic comparison maps were created with the use of either BLASTN or TBLASTX, to determine homology, and then visualised with Easyfig (Sullivan et al., 2011). Artemis Comparison Tool (ACT) was used for identification of feature variations between phage genomes of *Vequintavirinae*, with homology being assessed with TBLASTX (Carver et al., 2005). Phylograms were generated based on the amino acid sequence of the major capsid protein, larger terminase and DNA polymerase of phage CB7 and 17 members of *Vequintavirinae* were created using MEGA7 (Kumar et al., 2016), applying MUSCLE for sequence alignment (Edgar, 2004) with the construction of phylograms using the maximum likelihood (ML) method based on the Whelan and Goldman substitution model (Whelan and Goldman, 2001), with the robustness of the trees was assessed by bootstrapping (1000). VICTOR was employed using all pairwise comparisons of the amino acid sequences (same phages as described previously) which employs the Genome-BLAST Distance Phylogeny (GBDP) method (Meier-Kolthoff et al., 2013) under settings recommended for prokaryotic viruses (Meier-Kolthoff and Göker, 2017). The resulting intergenomic distances (including 100 replicates each) were used to infer a balanced minimum evolution tree with branch support via FASTME including SPR postprocessing (Lefort et al., 2015) for each of the formulas D0, D4 and D6, respectively. The trees were rooted at the midpoint (Farris, 1972) and visualised with FigTree (Rambaut, 2006). Taxon boundaries

at the species, genus, subfamily and family level were estimated with the OPTSIL program (Göker et al., 2009), the recommended clustering thresholds (Meier-Kolthoff and Göker, 2017) and an F value (fraction of links required for cluster fusion) of 0.5 (Meier-Kolthoff et al., 2014). The heat map comparing the genomes of phage CB7 and 17 members of *Vequintavirinae* was generated using Gegenees utilising TBLASTX, using accurate parameters (fragment length: 200 bp; step size: 100 bp, threshold set to 0%) Ågren et al., 2012).

2.3.5.5. Comparative genomics of *Pectobacterium* phage CBB

Coregenes (Turner et al., 2013) was used for total proteome comparisons between phages, with the BLASTP threshold set at 75% (excluding phage CBB terminal repeat ORFs CBB_555 to CBB_605). Genome comparison of the Rak2-like phages was visualised using Easyfig (Sullivan et al., 2011) with comparison of genome sequences facilitated by TBLASTX (excluding phage CBB terminal repeat ORFs CBB_555 to CBB_605 of CBB) and reorientation of genomes such that the largest ORF was set to the first position to improve visual understanding (as starting point among Genbank files was not uniform). Phylograms were generated using MEGA7 (Kumar et al., 2016) involving the use of MUSCLE for sequence alignment (Edgar, 2004), with the construction of phylograms using the maximum likelihood (ML) method based on the Jones–Thornton–Taylor model (Jones et al., 1992). Analysis using the portal vertex protein of phage CBB was conducted as described previously (Brewer et al., 2014). The full-length protein was used in the BLASTP search to find phages with homologous portal proteins. A conserved internal region from 100 different portal vertex proteins from different phages was then extracted and aligned with MUSCLE. The resulting alignment was then used to create a phylogram based on the Maximum likelihood method (Jones et al., 1992), with the robustness of the trees being assessed by bootstrapping (100)

2.3.5.6. Comparative genomics of *Erwinia* phage Y3

Genome comparisons between phages were performed using TBLASTX and visualised using Easyfig (Sullivan et al., 2011). Multiple sequence alignments were created with MEGA7 (Kumar et al., 2016) using MUSCLE. Phylograms were constructed using the portal vertex (Y3_003), large terminase (Y3_004), DNA polymerase (Y3_173) and ATP

dependent DNA helicase (Y3_176) among phages possessing homology to ΦRSL1, with the large terminase and portal vertex protein being used among jumbo myoviruses of *E. amylovora*. To estimate the robustness of the phylograms, the maximum-likelihood algorithm (Jones et al., 1992) based on the Whelan and Goldman substitution model (Whelan and Goldman, 2001), was conducted with bootstrap support (1000). A total proteome comparison between phages was conducted using Coregenes 3.5 with the BLASTP threshold set at 75% (http://gateway.binf.gmu.edu:8080/CoreGenes3.5/custdata.html; (Turner et al., 2013)).

2.3.6. Bacteriophage virion ESI-MS/MS proteome analysis

Phage capsid proteins were extracted from high titre CsCl purified phage (>1x10^9 PFU/mL) using chloroform:methanol extraction (1:1:0.75, v/v/v). The resulting protein pellet was resuspended in loading buffer (1% SDS, 6% sucrose, 100 mM dithiothreitol, 10 mM Tris pH 6.8, 0.0625% w/v bromophenol blue) and heated to 95˚C for 5 min to resuspend the pellet. This was subsequently loaded onto a 12% SDS-PAGE gel after which gel electrophoresis was conducted. The resulting gel was then stained using Gelcode™ Blue Safe Protein Stain (Thermo Scientific, Waltham, MA, USA) to visualise virion proteins. Gel fragments were extracted and subjected to trypsinisation which were analysed using tandem electrospray ionisation-mass spectrometry (ESI-MS/MS) exactly as described previously (Van den Bossche et al., 2014).

2.3.7. Determination of HNH endonuclease splicing

A similar approach was taken as previously described (Ceyssens et al., 2008; Landthaler et al., 2002). Freshly grown host bacteria (strain DSM 30186) at an OD$_{600}$ of 0.20-0.23 was infected with phage at an MOI of 5x10^{-4}, in the same manner as described for single-step growth curve assay as previously described (as described in section 2.2.2.). At 15 min, 30 min and 45 min, 1 mL of infected cells were pelleted and then resuspended in phosphate buffer saline (PBS). Total RNA was then extracted using the High Pure RNA Roche extraction kit (Roche, Basel, Switzerland) and subsequently reverse transcription PCR (RT-PCR) was performed using the cDNA synthesis kit (Bioline, London, UK) according to manufactures protocol. PCRs were performed to amplify products using primers specific to regions situated at 5' and 3' ends of HNH

endonucleases being examined (Table 2.1), using RedTaq Readymix (Sigma-Aldrich, St. Louis, MO, USA) with cDNA and a control using CB7 genomic DNA. These products were then inspected by agarose gel electrophoresis and Sanger sequencing for the presence of splicing.

Table 2.1. Primers used for the creation of PCR products from cDNA and genome DNA of *Pectobacterium* phage CB7 for the examination of HNH endonuclease splicing.

HNHs being examined	Primer name	Primer sequence 5'-3'
CB7_2	CB7_F	CGTTGCTTCAGTTCTTCC
	CB7_R	ATCGTAGACGATCCACG
CB7_50 and CB7_52	CB7_50-52F	ACGACTACGTATTGACG
	CB7_50-52R	CCTTCCATGACCATACC
CB7_54	CB7_54F	CGACACAGCGACTTATCC
	CB7_54R	GACATTGACAGAGTGTCG
CB7_196	CB7_196F	CAACATCAACGCACTGG
	CB7_196R	TTCAGCCGACATAGAGG
CB7_202	CB7_202F	CAAGGTAAGTATGCTGC
	CB7_202R	TTCTTCGCCACAGAAGG
CB7_245	CB7_245F	CGCAGATGATGTCTTCC
	CB7_245R	AACATTGTGAGCATTACC

2.3.8. Accession number of phage genomes

All phage genomes examined in this thesis were submitted to Genbank (Table 2.2).

Table 2.2. Genbank accession number of phages genomes

Phage	Genbank accession no.
Pectobacterium phage vB_PatP_CB1	KY514264
Pectobacterium phage vB_PatP_CB3	KY514265
Pectobacterium phage vB_PatP_CB4	KY549659
Pectobacterium phage vB_PatP_CB5	KY953156
Pectobacterium phage vB_PatM_CB7	KY514263
Pectobacterium phage vB_PccM_CBB	KU574722
Erwinia phage vB_EaM_Y3	KY984068

2.4. Cloning, recombinant protein expression and purification

2.4.1. Cloning of putative peptidoglycan degrading proteins of *Pectobacterium* phage CBB and *Erwinia* phage Y3

For the cloning of protein CBB_187, its gene was amplified using forward primer CBB_187-PciI and reverse primer CBB_187-XhoI. The truncated version of CBB_187, namely rTM-CBB_187 without the transmembrane domain, was amplified using forward primer rTM-CBB_187 and the reverse primer CBB_187-XhoI (Table 2.3). For the cloning of protein CBB_239, its gene was amplified using forward primer CBB_239-PciI and reverse primer CBB_239-XhoI (Table 2.3). All PCR reactions were conducted using Velocity DNA polymerase (Bioline, London, UK), using CBB genomic DNA as a template. Resulting PCR products were digested (NcoI and XhoI) and ligated in-frame into vector pET28a, in a manner that would allow Induced proteins to possess a 6 x Histidine tag at the N-terminus. The pET28a vector with the insert of CBB_187 and the transmembrane tunicated version of this protein (rTM-CBB_187) were chemically transformed into *E.coli* Lemo21 (DE3) (New England Biolabs, Ipswich, MA, USA), while the insert of the CBB_239, in pET28a, was transformed into *E.coli* BL21 (DE3) (New England Biolabs, Ipswich, Massachusetts), following the manufacturer's standard protocol.

For the cloning of protein Y3_301, its gene was amplified using forward primer Y3_301-PciI and reverse primer Y3_301-Xho1. The truncated version of Y3_301, namely rTM-Y3_301 without the transmembrane domain was amplified using forward primer rTM-Y3_301-PciI and the above reverse primer Y3_301-XhoI (Table 2.3). All amplifications were conducted with High Fidelity PCR Master mix (Roche, Basel, Switzerland) using genomic DNA as a template. The resulting PCR product was digested and ligated in-frame into vector pET28a, with the resulting construct transformed in *E. coli* Lemo21 (DE3) (Schlegel et al., 2012), following the manufacturer's standard protocol (New England Biolabs, Ipswich, MA, USA).

DNA sequencing was performed on purified plasmid to ensure the integrity of all constructs.

Table 2.3. List of primers used for the cloning of CBB_187/ rTM-CBB_187, CBB239 and Y3_301/rTM-Y3_301. Primer overhangs, restriction sites and coding sequences are indicated by uppercase, underlined and lowercase letters, respectively.

Product	Primer name	Primer sequence 5'-3'	Restriction enyzme
CBB_187	CBB_187-PciI	CATAC<u>Atgt</u>ataaagcgatacaagaactc	PciI
	CBB_187-XhoI	AAGCTT<u>CTCGAG</u>aagactactcaataataagtcaacacg	XhoI
rTM-CBB_187	rTM-CBB_187-Pci1	CAT<u>ACATG</u>tcacatgatcgtactacatcattagt	PciI
CBB_239	CBB_239-BspH1	CA<u>TC</u>atgagcaattgggttttaac	BspHI
	CBB_239-Xho1	AAGCTT<u>CTCGAG</u>tagagattctacaatacgtgc	XhoI
Y3_301	Y3_301-PciI	<u>CATAC</u>atgttagttatcatcaacaatcg	PciI
	Y3_301-XhoI	<u>AAGCTT</u>ctcgagggaagaaggagtatgct	XhoI
rTM-Y3_301	rTM-Y3_301-PciI	<u>CATACAT</u>gttacgcatttcgccgtc	PciI

2.4.2. Overexpression of protein CBB_239

Recombinant CBB_239 was overexpressed in the pET28a/*E.coli* BL21 (DE3) expression system by growing in LB broth containing kanamycin (40 µg/mL) with a 1% inoculum of an overnight culture. This culture was incubated at 30 °C (150 rpm.) to an optical density (600nm) of 0.2-0.4, after which the culture was induced with IPTG (400 µM) and incubation at 30°C for a further 4 h. Cells were harvested by centrifugation (14,534 g for 20 min) and the resulting cell pellet was resuspended in a buffer of 20mM sodium phosphate (pH7.4), 0.5 M NaCl, 40 mM imidazole. At this point in the procedure cells could be stored at -80°C or lysed using Bugbuster lysis mastermix reagent (Merck Millipore, Billerica, MA, USA) with agitation at 120 rpm. at room temperature for 20 min. Soluble and insoluble protein fractions were separated by centrifugation at 7,085

g for 20 min. Protein fractions were examined by using SDS-PAGE (Sambrook et al., 2001).

2.4.3. Purification and quantification of protein CBB_239

The soluble fraction of lysate was applied to an ÄKTA pure chromatography system (GE Healthcare, Little Chalfont, UK) using a sample pump S9-S (GE Healthcare, Little Chalfont, UK). A Histrap FastFlow crude 1 mL column (GE Healthcare, Little Chalfont, UK) was used for selective protein binding, using a running buffer of 10mM sodium phosphate (pH 7.4), 0.5M NaCl, 20 mM imidazole and an elution buffer of 10mM sodium phosphate (pH 7.4), 0.5M NaCl, 500 mM imidazole. Factions that contained the purified protein of interest were pooled and desalted by loading onto a 5 mL HiTrap desalting column (GE Healthcare, Little Chalfont, UK) using a desalting buffer of 10mM sodium phosphate (pH 7.4), 150 mM NaCl. The resulted purified protein was quantified using a Pierce BCA protein assay kit (Thermo Scientific, Waltham, MA, USA) following the manufacturer's standard protocols

2.5. Lytic activity tests

2.5.1. Demonstrate signal-arrest-release (SAR) activity

SAR endolysin activity of CBB187/rTM-CBB_187 or Y3_301/rTM-Y3_301 in pET28a/*E. coli* Lemo21 (DE3) expression system was determined by using 40mL liquid cultures (triplicate) in LB both using selective markers kanamycin (40 μg/mL) and chloramphenicol (40 μg/mL) with 1% inoculum of an overnight culture. Cultures were agitated (150 rpm) at 30 °C with the optical density (600nm) being measured at 1 h intervals. At hour 4 (OD_{600}~0.4) half the cultures were induced with IPTG (400 μM) and readings were continued up to hour 8. Averages and standard deviations were determined using Microsoft Excel.

2.5.2. Zymogram analysis

The assay involved the preparation of a 12% SDS-PAGE gel but with the addition of autoclaved bacterial cells to the separation gel. A purified putative peptidoglycan degrading protein sample was mixed with an equal volume of protein gel loading dye (1% [w/v]SDS, 6% [w/v] sucrose, 100 mM 1,4 dithiothreitol, 10 mM Tris-HCl pH6.8, 0.0625% [w/v] bromophenol blue), this was then heated to 95°C for 5 min, after which

the sample was applied to the gel and put running at 200 V for 40 – 60 min. The gel was then soaked in distilled water at room temperature with gentle agitation for 30-60 min before being transferred into renaturation buffer (25mM Tris-HCl, pH 8.0, 1% Triton X100). This was then again incubated at room temperature with gentle agitation, after 15 min the gel was checked every five min for the presence of clearing.

2.5.3. Turbidity reduction assay

Turbidity reduction assay was conducted as previously described by Lavigne et al (2004). Briefly, *P. carotovorum* subsp. *carotovorum* BL19-1-37 was grown to an OD_{600nm} of 0.6 which was pelleted by centrifugation at 7,085 g for 10 min. The resulting pellet was resuspended in chloroform-saturated 50 mM Tris-HCl (pH 8), incubated at room temperature and agitated at 200 rpm for 45 min. Cells were then again pelleted and washed twice with and resuspended in 50 mM Tris-HCl (pH 8) to a final OD_{600nm} of 0.1. Fifty microliters of purified protein were added to 150 µL of treated cells in triplicate. For a negative control 50 uL of a buffer of 10mM sodium phosphate (pH 7.4), 150 mM NaCl was added to 150 µL of treated cells. The OD_{600nm} was then read every min for 10 min.

2.6. Assessing biocontrol potential

2.6.1. Tuber rot assay to determine biocontrol potential of CB1, CB3 and CB4 mixture

The potato whole tuber rot assay was used to assess the potential of a phage mixture comprised of phages CB1, CB3 and CB4 to prevent infection of potato tissue by *P. atrosepticum* and was conducted in a similar manner as previously described (Czajkowski et al., 2015). The phage mixture was made by adding an equal number of phages to a total titre approx. $1x10^7$ PFU/mL to SM buffer. *P. atrosepticum* strains, DSM 18077 and DSM 30186, were resuspended in deionised water with cell numbers of approx. $1x10^7$ CFU/mL and mixed in equal quantities. Ware potato tubers were obtained at a local supermarket, these were first washed with tap water, then surface sterilised for 10 min in 70% isopropanol, then again washed with tap water and allowed to dry on tissue paper at room temperature.

Whole tuber assay involved tubers being incised at the rose end (opposite the stolon end) to remove a 0.5 cm transverse slice. To which 100 µL of either bacteria or water was added, left to absorb and then followed with the addition of 100 µL of SM buffer

or phage mixture. These were then allowed 30 min to sit at room temperature before the detached potato slices were reattached using sterile toothpicks to tubers. Tubers were then incubated or 72 hrs at 25°C in a humid box. To assess the protective effect of the phage mixture the weight of rotten tissue was determined for each tuber.

2.6.2. Statistical analysis of data from tuber rot assay

Figure 3.7 was generated with Excel. Statistical analysis was performed with IBM SPSS Software (version 24). Normality of data was assessed with the Shapiro-Wilks test at a significance level of 0.05. Non-parametric tests were chosen for data not normally distributed. Comparison of the weight of soft rot tissue of infected potatoes, phage treated and untreated, were performed with the Mann-Withney U test.

Chapter 3. Novel N4-like bacteriophage of *Pectobacterium atrosepticum*

Transmission electron microscopy was conducted by Horst Neve.

Mass spectrometry analysis of phage virion proteins was conducted by Hanne Hendrix.

Genome sequencing of CB1-like phages was outsourced to the University of Liverpool, UK.

Statistical analysis of the biocontrol experiment was assisted by Kate Hayes.

MALDI-TOF analysis was assisted by Monika Koziel.

3.1. Introduction

The genera of *Pectobacterium* and *Dickeya*, often referred to collectively as the soft rot *Enterobacteriaceae* (SRE), are phytopathogens that cause economically important losses in a wide range of arable and ornamental crops. These bacteria are all Gram-negative, non-spore forming, facultative anaerobic rods that are typified by the production of extracellular pectinolytic enzymes during plant infection (Pérombelon, 2002; Toth et al., 2011). Both genera are considered to be among the top ten most important plant pathogens (Mansfield et al., 2012). To date, the predominant SRE species and subspecies which cause blackleg and soft rot of the potato crop in Europe are *Pectobacterium atrosepticum*, *Pectobacterium carotovorum* subsp. *carotovorum*, *Pectobacterium parmentieri (*formally *Pectobacterium wasabiae*), *Dickeya dianthicola* and *Dickeya solani* (Khayi et al., 2016; Toth et al., 2011; van der Wolf et al., 2017; Waleron et al., 2013)*. P. atrosepticum* is traditionally believed to be one of the more important causative agents for blackleg in temperate climates (Toth et al., 2003), and the dominant blackleg causative agent in Scotland (van der Wolf et al., 2017).

Strategies for the control of potato soft rot and blackleg are limited, whether they be physical, chemical or biological. In addition, no blackleg resistant potato cultivars are available. The current approach to disease control is the use of cultivation practices to minimise levels of infection, including contamination avoidance and the removal of diseased plants and/or tissue. Seed certification schemes are also employed. However, the success of these schemes vary and are highly weather dependent (Czajkowski et al., 2011; De Boer, 2004).

Phages, the viruses of bacteria, are being investigated as a potential biocontrol strategy for many problematic bacteria, including phytopathogens. Indeed, phages of *Erwinia amylovora*, *Xanthomonas campestris* pv. *vesicatoria*, *Agrobacterium tumefaciens* and many others have been investigated for their disease control potential (Buttimer et al., 2017b; Frampton et al., 2012). In addition, a number of SRE phages have also been recently isolated and characterized (Czajkowski, 2016), some of these having also been assessed for their potential as biocontrol agents of their respective phytopathogenic hosts with promising results.

Proof of concept experiments using SRE phages on potato whole tubers has demonstrated that phage biocontrol has the capability to inhibit soft rot caused by *P.*

carotovorum subsp. *carotovorum*, *P. parmentieri* and *D. solani* (Adriaenssens et al., 2012b; Czajkowski et al., 2015; Smolarska et al., 2018). A field trial was conducted by Adriaenssens *et al.* (2012) using phage vB_DsoM_LIMEstone1 against *D. solani* and showed that phage treatment against blackleg resulted in decreased disease severity and improved yields. However, SRE species causing disease that were not sensitive to the phage limited the success of the overall outcome. Investigation of phage biocontrol of SRE also has not been limited to the potato. Phage biocontrol was shown, for example, to reduce the incidence of soft rot by 50% on tuber plugs of Calla lily due to *P. carotovorum* subsp. *carotovorum* infection in greenhouse trials (Ravensdale et al., 2010). Promising results were also demonstrated using phage against this species infecting Chinese cabbage (Lim et al., 2013).

We here report the isolation and characterisation of *P. atrosepticum* phages vB_PatP_CB1, vB_PatP_CB3 and vB_PatP_CB4. To the author's knowledge, this is the first report of N4-like phages to be identified infecting bacteria belonging to SRE. *Escherichia* phage N4 is a lytic podovirus, which was isolated in the 1960s from sewers in Genoa, Italy (Molina et al., 1965). The phage is typified by the possession of a virion-associated RNA (vRNA) polymerase, which it injects along with its genome into its host *Escherichia coli* at the beginning of infection to initiate transcription of its DNA and an overall conserved transcriptional scheme (Glucksmann et al., 1992). To date, Genbank contains at least 56 genome sequences of N4-like phages infecting hosts belonging to the classes of *Alpha-*, *Beta-* and *Gamma-proteobacteria*, with each phage genome encoding the hallmark feature of a vRNA polymerase. Six phage genera have been defined among these so far, these being *G7cvirus*, *Lit1virus*, *Ea92virus*, *Luz7virus* and *N4virus* (Adriaenssens et al., 2017; Wittmann et al., 2016b, 2016a, 2015) with another two genera (*Sp58virus* and *Dss3virus*) having been proposed (Wittmann et al., 2015). Furthermore, using these newly isolated N4-like phages, we demonstrated their potential for biocontrol with the suppression of soft rot formation with their co-inoculation with *P. atrosepticum* on potato tubers.

3.2. Results

3.2.1. Isolation of SRE from potato crops symptomatic of blackleg

In 2013, nineteen potato stem samples symptomatic for blackleg, each representing a crop, were collected from three distinct farms in Co. Cork, Ireland. Isolate identification was based on cavity formation on crystal violet pectate (CVP) medium, production of reducing substances from sucrose, acid production from α-methyl-glucoside, *Pectobacterium* genus-specific along with *P. atrosepticum* and *P. carotovorum* subsp. *carotovorum* species-specific PCRs and MALDI-TOF mass spectrometry. Based on these results, fourteen plants were found to be infected with *P. atrosepticum* with a remaining plant found to be infected with both *P. atrosepticum* and *P. carotovorum* subsp. *carotovorum* (Appendix, Table S3.1)*.

3.2.2. Isolation of bacteriophages, host range and general characteristics

Thirteen phage isolates were obtained from soil samples that were collected from potato grading machinery and potato fields from two of the three farms mentioned above. These were subjected to genomic DNA comparison. Restriction digestion analysis, employing BamH1, allowed the identification of ten isolates that produced three similar band patterns (Figure 3.1). Phages producing these patterns were found in both the grading machinery and the field soils. An example of each was taken for further study, namely CB1, CB3 and CB4. These isolates produced clear plaques with an approximate diameter of 2-3 mm (overlay concentration 0.4% w/v agar in LB) on their respective host strains of *P. atrosepticum,* with narrow halos occasionally being observed to be surrounding plaques.

The host ranges of these phages were examined using 31 bacterial strains (local Irish isolates and reference strains) from five different species belonging to SRE, namely *P. atrosepticum* (19 strains), *P. carotovorum* subsp. *carotovorum* (4 strains), *D. chrysanthemi* bv. *chrysanthemi* (1 strain), *D. dianthicola* (3 strains) and *D. solani* (4 strains). These phages were able to form plaques on 15 strains of their host species, *P. atrosepticum*, with no plaque formation or inhibition of growth being observed on the other species. Slight variations in lytic ability were found among the three phages, with

CB1 and CB3 determined to form plaques on 12 strains, while CB4 was only found to infect 10 of the 19 strains tested (Table 3.1).
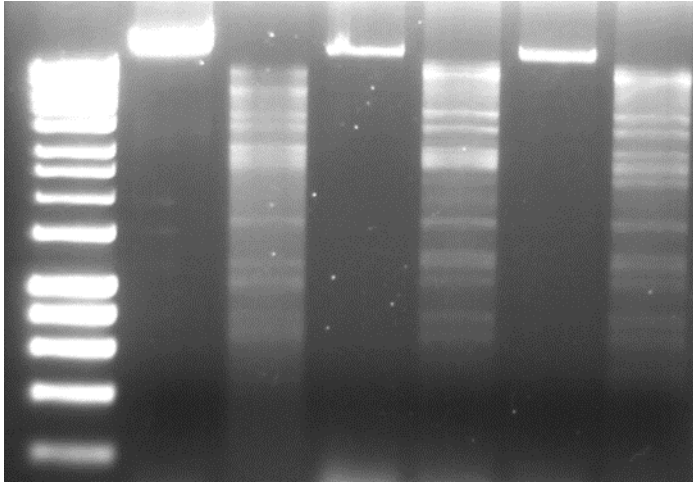
Figure 3.1. Genomic DNA of *Pectobacterium* phages CB1, CB3 and CB4, BamHI-digested (lane 3, 5 and 7, respectively) and undigested (lane 2, 4 and 6, respectively). Lane 1; DNA marker (Hyperladder 1kb, Bioline). Gel concentration 1 % w/v agarose.

Table 3.1. Host range of *Pectobacterium* phages CB1, CB3 and CB4 on 31 strains of various members of the soft rot *Enterobacteriaceae* determined by spot testing with serial dilutions of phage suspensions. The efficiency of plating (EOP) values were determined for sensitive strains. Results recorded as –, no infection; CS, the presence of clear spot with no plaque formation; number, EOP value; *, host strain of phage. EOP values determined by spot testing in triplicate.

Bacteria		Bacteriophage infection		
Species	**Strain**	**CB1**	**CB3**	**CB4**
Pectobacterium atrosepticum	DSMZ 18077 (type strain)	1.000*	2.263	0.909
	DSMZ 30184	2.00E-05	CS	CS
	DSMZ 30185	0.371	0.126	0.136
	DSMZ 30186	0.003	1.000*	1.000*
	CB BL1-1	CS	CS	CS
	CB BL2-1	0.571	0.737	0.336
	CB BL3-1	0.529	0.605	0.218
	CB BL4-1	0.571	0.789	0.245
	CB BL5-1	CS	0.158	CS
	CB BL7-1	CS	CS	CS
	CB BL9-1	0.027	0.007	0.164
	CB BL11-1	CS	0.279	0.127
	CB BL12-2	2.86E-06	0.079	0.255
	CB BL13-1	1.71E-05	0.063	0.436
	CB BL14-1	CS	CS	CS
	CB BL15-1	-	-	-
	CB BL16-1	0.005	CS	CS
	CB BL18-1	CS	0.037	CS
	CB BL19-1	0.005	CS	CS
Pectobacterium carotovorum subsp. carotovorum	DSMZ 30168 (type strain)	-	-	-
	DSMZ 30169	-	-	-
	DSMZ 30170	-	-	-
	CB BL19-1-37	-	-	-
Dickeya chrysanthemi bv chrysanthemi	LMG 2804	-	-	-
Dickeya dianthicola	PD 482	-	-	-
	PD 2174	-	-	-
	GBBC 1538	-	-	-
Dickeya solani	sp. PRI 2222	-	-	-
	LMG 25865	-	-	-
	GBBC 1502	-	-	-
	GBBC 1586	-	-	-

Examination of the morphology of the three phages by transmission electron microscopy showed that they can be classified as members of the *Podoviridae* family (Ackermann, 2001). They featured a C1 morphotype with isometric capsids (ca. 70 nm) and short non-contractile tails (length: ca. 25 nm) (Figure 3.2, Table 3.2). Head and tail

measurements are consistent with previously reported N4-like phages (Wittmann et al., 2015). Additionally, a set of (putatively six) short whiskers (length: ca. 25 nm) attached to a collar structure (width: ca. 19 nm) were observed. At their distal ends, the whiskers terminate with elongated globular appendices (ca. 12 nm x 7 nm). (Wittmann et al., 2015). The three phages were named in accordance with the nomenclature set out by Kropinski *et al.* (2009).
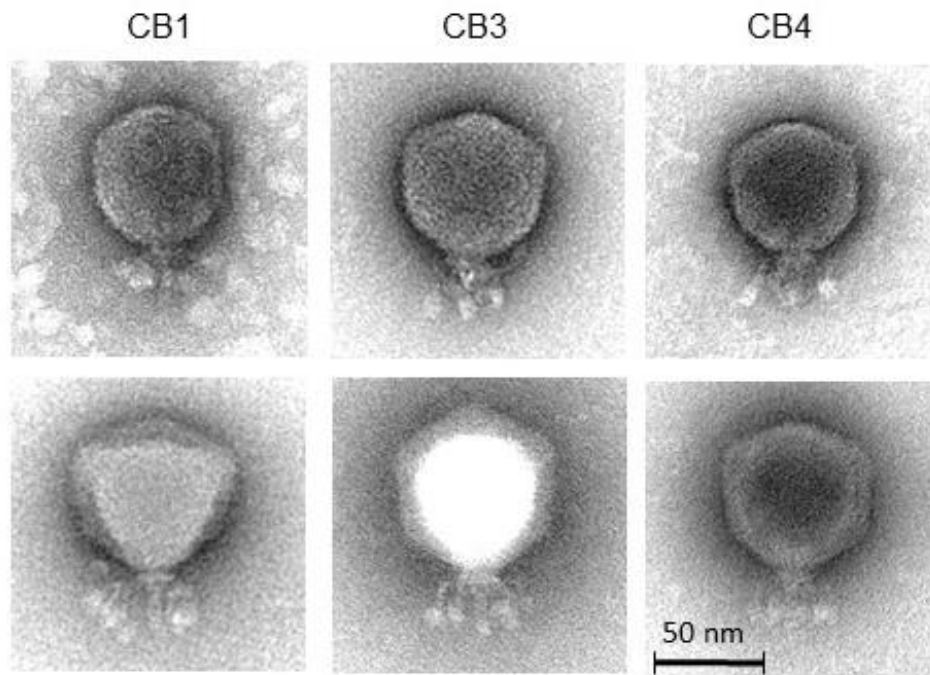
Figure 3.2. Transmission electron micrographs of negatively stained *Pectobacterium* phages CB1, CB3 and CB4 using 2% (w/v) uranyl acetate (top) and 1% (w/v) phosphotungstic acid (bottom).

Table 3.2. Dimensions of *Pectobacterium* phage CB1, CB3 and CB4 negatively stained with 2% (w/v) uranyl acetate (head sizes) or with 1% phosphotungstic acid (all other measurements). *Whisker length from collar to distal end of the whisker ball.

	Head (nm)	Tail length (nm)	Collar width (nm)	Whisker length* (nm)	Whisker ball length (nm)	Whisker ball width (nm)
CB1	67.8±4.4 (n=19)	23.9±1.7 (n=8)	18.5±1.3 (n=8)	23.4±3.0 (n=14)	12.3±1.4 (n=16)	6.3±0.7 (n=17)
CB3	71.8±1.7 (n=6)	25.8±2.3 (n=8)	19.6±0.8 (n=7)	25.6±1.7 (n=7)	12.0±1.5 (n=9)	7.4±0.9 (n=9)
CB4	70.4±2.5 (n=12)	25.5±1.0 (n=5)	19.3±0.8 (n=6)	23.5±1.6 (n=3)	11.8±0.9 (n=5)	6.6±0.6 (n=5)

The one-step-growth curve assay, under standard conditions using LB medium, showed that the latent period of CB1 was 60 min with an approximate burst size of

207 plaque forming units (PFU)/cell. For CB3, a latent period of 65 min with an approximate burst size of 246 PFU/cell observed. For CB4, a latent period of 65 min with an approximate burst size of 158 PFU/cell observed (Appendix, Figure S3.1). Phage viability under different environmental conditions was also examined. Over a duration of one hour all phages were found to be stable between -18°C and 50 °C. They were also found to be stable between pH 5 and 11 for 24 hrs (Appendix, Figure S3.2, S3.3).

3.2.3. Genome and proteome analysis

3.2.3.1. Genomes of phages CB1, CB3 and CB4 show an N4virus organisation

The genome sequences obtained for phages CB1, CB3 and CB4 were 75,394 bp, 75,522 bp and 75,973 bp with coverage of 2057x, 1882x and 1765x, respectively. The genomes of these phages are likely linear with direct terminal repeats (DTRs) with estimated sizes of 647 bp (for CB1 and CB3) and 648 bp (for CB4). This estimation is based on the identification of a localised region with roughly double the read depth in comparison to average read depth across the whole genome of each phage, a similar finding to previously reported N4-like phages (Fouts et al., 2013; Li et al., 2014). However, it is noteworthy that termini at the ends of the DTRs for these phages can be asymmetric, as seen in the case of *Escherichia* phage N4 itself (Ohmori et al., 1988). It is not known whether DTR asymmetry exists in phages CB1, CB3 and CB4. As well as this, the average GC content of these genomes was found to be 49%, just below the GC content associated with the host bacterium *P. atrosepticum* at 50-51% (Bell et al., 2004; Nikolaichik et al., 2014).

The number of predicted open reading frames (ORFs) determined on the non-redundant genome of these phages were 97, 102 and 100 for CB1, CB3 and CB4, respectively (Appendix, Table S3.2, S3.3, S3.4). Proteins predicted to play roles in transcription, DNA replication, virion morphogenesis and host lysis were identified in all three phages as well as putative genes for homing endonucleases (Figure 3.3). No integrase, excisionase or repressor genes were detected suggesting that all three phages have an exclusively lytic lifecycle. Moreover, no ORFs were identified for pathogenicity or known toxins.
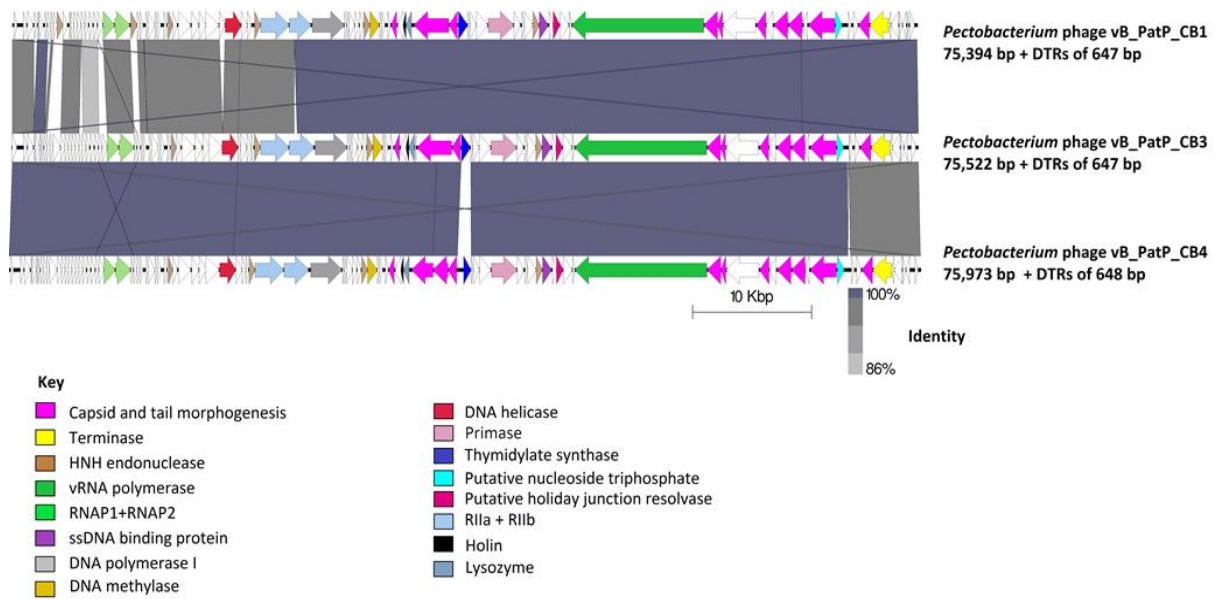
Figure 3.3. Comparison of the genomes of *Pectobacterium* phages CB1, CB3 and CB4 employing BLASTN and visualised with Easyfig. Genome maps comprise of arrows indicating locations of ORFs among the different phage genomes. Arrows have been colour-coded describing their predicted roles (see key), and lines between genome maps indicate levels of homology.

Phages CB3 and CB4 exhibited a 97% genome-wide nucleotide identity (BLASTN algorithm) and can thus be considered different isolates of the same phage species (Adriaenssens and Brister, 2017). The major feature variations found between them being their ORFs for thymidylate synthase and N4 gp32-like protein (Appendix, Table S3.5). The only major difference identified with host range between these phages was the ability of CB3 to infect two extra *P. atrosepticum* strains (CB BL5-1 and CB BL18-1) (Table 3.1). In addition, phage CB4 also possessed two genes for tRNAs (for asparagine & glutamine). In contrast, phage CB1 was only 95% similar to CB3 and 93% similar to CB4. This places CB1 on the boundary of speciation with the other two phages. CB1 possesses the same predicted thymidylate synthase gene and has no tRNA genes like CB3. However, the predicted early gene region possesses variations in its ORF content not shared with CB3 and CB4. These differences include the presence of different hypothetical proteins and the presence of two additional HNH endonucleases. There is also a notable sequence variation of the rIIB protein of CB1 to that of the other two phages. Interestingly, CB1 shares its N4 gp32-like gene with CB4 (Appendix, Table S3.5).

The overall genomic architecture of the CB1-like phages resembles those belonging to the N4-like group (Figure 3.4), possessing all 18 core proteins of these phages identified by Li et al (Li et al., 2016) (Appendix, Table S3.6), thus allowing the N4-like designation. When comparing these phages to *Escherichia* phage N4, the gene order of most structural proteins is well conserved with major variations associated with the genes involved in host lysis and genes for structural proteins believed to play roles in tail morphogenesis. This observation has been noted among other N4-like phages (Ceyssens et al., 2010; Kulikov et al., 2012). These phages do not fall within any genera currently defined in the N4-like group to date. Phylograms based on the vRNA and DNA polymerase proteins show that they form their own distinct clade among these phages (Figure 3.5), with their closest evolutionary relationship appearing to be with the phages of the genera of *Lit1virus* and *Luz7virus*. Additionally, when protein homology was calculated with Gegenees analysis (TBLASTX), they show limited identity to other N4-like members (Figure 3.6).

Figure 3.4. Comparison of the genomes of *Pectobacterium* phage CB1 with those phages representing the genera of *N4virus*, *G7cvirus*, *Ea92virus*, *Lit1virus* and *Luz7virus* employing TBLASTX and visualised with Easyfig. Genome maps comprise of arrows indicating locations of genes on the different phage genomes, and lines between genome maps indicate the level of homology.

(A)

74 — Pseudomonas phage RWG (AIZ94788.1)	
100 — Pseudomonas phage LIT1 (YP 003358435.1)	*Litvirus*
— Pseudomonas phage vB Pae575P-3 (ANT44317.1)	
— Pseudomonas phage LUZ7 (YP 003358324.1)	*Luz7virus*
100 — Pseudomonas phage KPP21 (YP 009219031.1)	
— Acinetobacter phage Presley (YP 009007605.1)	
— Pectobacterium phage vB PatP CB1	
100 — Pectobacterium phage vB PatP CB3	**CB1-like phages**
62 — Pectobacterium phage vB PatP CB4	
— Vibrio phage pVco-5 (ARM71084.1)	
100 — Vibrio phage phi 1 (YP 009198578.1)	
100 — Vibrio phage JSF3 (APD18062.1+APD18064.1)	
96 — Vibrio phage VCO139 (AGI61870.1)	
87 — Vibrio phage JA-1 (YP 008126803.1)	
— Vibrio phage pVa5 (APC46020.1)	
100 — Pseudoalteromonas phage pYD6-A (YP 007674265.1)	
100 — Vibrio phage VBP32 (YP 007676595.1)	
100 — Vibrio phage VBP47 (YP 007674119.1)	
— Enterobacter phage EcP1 (YP 007003163.1)	
— Escherichia phage Pollock (YP 009152145.1)	
100 — Salmonella phage FSL SP-058 (YP 008239451.1)	*'Sp58virus'*
98 — Salmonella phage FSL SP-076 (YP 008240179.1)	
100 — Pseudomonas phage ZC03 (AMD43470.1)	
— Pseudomonas phage ZC08 (AMD43475.1)	
— Erwinia phage vB EamP-S6 (YP 007005799.1)	
99 — Sulfitobacter phage EE36phi1 (YP 002898979.1)	*'Dss3virus'*
97 — Silicibacter phage DSS3phi2 (YP 002899062.1)	
— Dinoroseobacter phage vBDshPR2C (AID16870.1)	
100 — Dinoroseobacter phage DFL12phi1 (YP 009043709.1)	
— Sulfitobacter phage phiCB2047-B (YP 007675812.1)	
78 — Escherichia phage vB EcoP PhAPEC7 (YP 009056171.1)	*G7cvirus*
100 — Escherichia phage vB EcoP G7C (YP 004782168.1)	
— Escherichia phage ECBP1 (YP 006908814.1)	
— Escherichia phage N4 (YP 950517.1)	*N4virus*
82 — Erwinia phage vB EamP Rexella (ANJ65282.1)	*Ea92virus*
100 — Erwinia phage vB EamP Gutmeister (ANJ65360.1)	
— Erwinia phage Ea9-2 (YP 009007430.1)	
— Pseudomonas phage inbricus (ATW58103.1)	
100 — Achromobacter phage phiAxp-3 (YP 009208696.1)	
95 — Achromobacter phage JWDelta (AHC56569.1)	
100 — Achromobacter phage JWAlpha (YP 009004757.1)	

70

Figure 3.5. Phylogenetic analysis using the (A) DNA polymerase (log likelihood = -16535.57) and (B) vRNA polymerase (log likelihood = -103761.67) protein sequences of *Pectobacterium* phages CB1, CB3 and CB4 and 38 other N4-like phages. Phages belonging to the genera of *G7cvirus*, *Lit1virus*, *Ea92virus*, *Luz7virus* and *N4virus* and proposed genera *Sp58virus* and *Dss3virus* are highlighted. The amino acid sequences were compared using MUSCLE. The tree was constructed using the maximum likelihood algorithm. The percentages of replicate trees were assessed with the bootstrap test (1000).

Organism	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41
1: Ea9-2	100	90	86	31	31	31	32	32	31	31	31	25	20	22	22	22	19	19	19	19	18	18	18	18	18	18	23	23	21	22	21	22	22	20	20	20	20	20	21	21	21
2: Rexella	90	100	91	31	31	31	32	32	31	31	31	25	20	22	23	22	19	19	19	19	18	18	18	18	18	18	23	23	21	22	21	22	22	20	20	20	20	20	21	21	21
3: vB_EamP_Gutmeister	91	96	100	32	32	32	33	33	32	32	32	25	20	22	23	23	19	19	19	19	19	18	18	18	18	18	23	23	22	22	22	22	22	20	20	20	20	20	21	21	21
4: JWAlpha	31	31	31	100	94	55	45	36	34	34	34	24	20	21	21	21	19	19	19	19	19	18	18	18	18	18	23	23	21	21	21	22	22	20	20	21	21	21	21	21	21
5: JWDelta	31	31	30	92	100	54	44	35	34	34	34	23	20	20	21	21	19	19	19	19	18	18	18	18	18	18	23	23	21	21	21	22	22	20	20	21	21	21	20	20	20
6: phiAxp-3	31	31	31	54	54	100	49	39	34	34	34	23	20	21	21	21	19	19	19	19	18	18	18	18	18	18	23	23	21	22	22	22	22	20	20	21	21	21	20	20	20
7: inbricus	33	33	33	46	46	51	100	42	37	37	36	25	20	21	22	22	20	20	20	19	18	18	18	18	18	18	24	24	22	22	22	22	22	21	21	21	21	21	21	21	21
8: N4	34	34	33	38	38	41	43	100	48	48	47	25	21	22	22	23	20	20	20	20	19	19	19	19	19	19	25	24	22	22	22	23	23	20	20	21	21	21	22	22	22
9: ECBP1	33	33	33	36	37	37	38	48	100	79	74	24	22	23	23	23	20	20	20	20	20	19	19	19	19	19	24	24	22	22	22	22	22	21	20	21	21	21	21	21	21
10: vB_EcoP_PhAPEC7	32	32	32	36	36	36	37	47	77	100	76	24	21	23	23	23	20	20	20	20	19	19	19	19	19	19	24	24	22	22	22	22	22	20	20	20	20	21	21	21	21
11: vB_EcoP_G7C	32	32	32	35	35	36	37	46	72	76	100	24	21	23	23	23	20	20	20	20	19	19	19	19	19	19	24	24	22	22	22	22	22	20	20	20	20	21	21	21	21
12: vB_EamP-S6	24	25	24	24	24	24	24	24	23	23	23	100	22	24	24	24	20	20	20	20	19	18	18	18	18	18	21	21	20	21	21	21	21	20	20	20	20	20	20	20	20
13: EcP1	23	23	22	23	23	23	23	23	23	24	24	25	100	31	31	31	22	22	22	22	21	21	21	21	21	21	22	22	21	21	21	21	21	21	21	21	21	21	22	21	21
14: Pollock	24	24	23	22	22	22	23	23	24	24	24	26	29	100	67	66	21	22	23	23	21	22	21	21	21	21	22	21	21	22	22	22	22	22	22	21	21	21	21	21	21
15: FSL SP-058	23	23	23	22	22	22	22	23	23	23	23	25	28	63	100	87	22	22	22	22	21	21	21	21	21	21	21	20	20	20	20	21	21	21	21	21	21	21	21	21	21
16: FSL SP-076	23	23	23	22	22	22	22	23	23	24	23	25	28	62	87	100	22	22	22	22	21	21	21	21	21	21	21	20	20	20	20	21	21	21	21	21	21	21	21	21	21
17: pVa5	19	19	19	20	20	20	20	20	20	20	20	20	20	21	21	21	100	34	34	33	20	19	19	19	19	19	19	19	18	19	19	19	19	20	20	20	20	20	20	20	20
18: VBP47	19	19	19	20	20	20	20	19	20	20	20	20	20	22	22	22	34	100	99	56	20	20	20	20	20	20	19	19	18	18	18	19	19	21	21	20	21	21	21	21	21
19: VBP32	19	19	19	20	20	20	20	20	20	20	20	20	20	22	22	22	34	99	100	57	20	20	20	20	20	20	19	19	18	18	18	19	19	21	21	21	21	21	21	21	21
20: pYD6-A	19	19	19	19	20	19	20	20	20	20	20	20	20	22	22	22	33	57	57	100	20	20	20	20	20	20	19	19	18	18	18	18	18	20	20	20	20	20	20	20	21
21: Presley	19	19	18	19	19	19	19	19	19	19	19	19	19	20	20	20	20	20	20	19	100	18	18	18	18	18	18	18	20	20	20	20	20	20	20	20	20	20	22	22	22
22: pVco-5	19	19	19	19	19	19	19	19	19	19	19	19	19	21	21	21	20	20	20	20	19	100	37	36	36	36	19	19	18	18	18	18	18	18	18	19	19	19	19	19	19
23: phi 1	19	19	19	20	20	19	19	20	20	20	20	20	20	22	21	21	20	21	21	21	19	40	100	78	78	78	20	20	19	18	18	18	18	20	20	20	20	20	20	20	20
24: JSF3	19	19	19	19	19	19	19	19	19	19	19	20	20	22	21	22	20	21	21	21	19	38	76	100	97	98	19	19	18	18	18	18	18	20	20	20	20	20	20	19	20
25: VCO139	19	19	19	19	19	19	19	19	19	19	19	20	20	22	21	21	20	21	20	21	19	38	76	98	100	99	19	19	18	18	18	18	18	19	19	19	19	19	20	19	19
26: JA-1	19	19	19	19	19	19	19	19	19	19	19	20	20	22	21	21	20	20	20	20	19	38	76	98	98	100	19	19	18	18	18	18	18	19	19	19	19	19	19	19	19
27: ZC08	24	24	24	24	24	24	24	24	24	24	24	22	21	21	21	21	19	19	19	19	19	19	19	19	19	19	100	91	21	21	21	21	21	19	19	19	19	19	19	19	19
28: ZC03	24	24	24	24	24	24	25	25	24	24	24	22	19	19	19	19	19	19	19	19	19	19	19	19	19	19	92	100	21	21	21	21	21	19	19	19	19	20	19	20	20
29: phiCB2047-B	22	22	22	22	22	22	22	22	21	22	21	21	19	20	20	20	18	18	18	18	18	18	18	18	18	18	20	20	100	39	39	40	40	19	19	19	19	19	19	19	19
30: EE36P1	22	22	22	22	22	22	22	22	21	21	22	21	19	19	20	20	18	18	18	18	18	17	17	17	17	17	20	20	39	100	79	55	55	18	19	19	19	19	19	19	19
31: DSS3P2	21	22	21	21	21	22	22	21	21	21	21	21	19	19	19	19	18	18	18	18	18	17	17	17	17	17	20	20	38	78	100	55	54	18	18	19	19	19	19	19	19
32: DFL12phi1	21	21	21	22	22	22	22	21	21	21	21	21	19	19	19	19	18	18	18	18	18	17	17	17	17	17	20	20	39	53	54	100	93	18	18	19	19	19	18	18	18
33: vBDshPR2C	21	21	21	22	22	22	22	21	21	21	21	21	19	19	19	19	18	18	18	18	18	17	17	17	17	17	20	20	39	53	53	94	100	18	18	19	19	19	19	18	18
34: LUZ7	20	20	19	20	20	20	20	19	19	20	20	20	19	20	20	20	20	20	20	20	20	18	18	18	18	18	18	18	18	18	18	18	18	100	80	40	40	41	22	22	22
35: KPP21	20	20	19	20	20	20	20	19	19	19	19	19	20	20	20	20	20	20	20	20	20	18	18	18	18	18	18	18	18	19	18	18	18	80	100	40	40	41	22	22	22
36: LIT1	20	20	20	21	21	21	21	20	20	20	20	20	19	20	20	20	20	20	20	20	20	18	18	18	18	18	19	19	19	19	19	19	19	40	40	100	88	88	22	22	22
37: RWG	20	20	20	21	21	21	21	20	20	20	20	19	19	20	20	20	20	20	20	20	20	18	18	18	18	18	19	19	19	19	19	19	19	41	41	88	100	90	22	22	22
38: vB_Pae575P-3	20	20	20	21	21	21	21	20	20	20	20	20	19	20	20	20	20	20	20	20	20	18	19	19	19	19	19	19	19	19	19	19	19	42	42	88	90	100	23	22	22
39: vB_Pat_CB4	21	21	20	21	21	20	21	21	21	21	21	20	19	20	20	20	20	20	20	20	22	19	18	18	18	18	19	19	19	19	19	19	19	22	22	22	22	22	100	96	92
40: vB_Pat_CB3	21	21	20	21	21	21	21	21	21	21	21	20	19	20	20	20	20	20	20	20	22	19	18	18	18	18	19	19	19	19	19	19	19	22	22	22	22	22	97	100	94
41: vB_Pat_CB1	21	21	20	20	20	20	21	21	21	21	21	20	19	20	20	20	20	20	20	20	22	19	18	18	18	18	19	19	19	19	19	19	19	22	22	22	22	22	92	94	100

Figure 3.6 TBLASTX heat map generated using Gegenees with accurate parameters – fragment length: 200 bp; and step size: 100 bp; threshold: 0%. The map includes the genomes of 38 N4-like phages with phages representing the genera *G7cvirus* (red), *Lit1virus* (brown), *Ea92virus* (green), *Luz7virus* (pink) and *N4virus* (yellow) and proposed genera *Sp58virus* (orange) and *Dss3virus* (grey), with the *Pectobacterium* phages CB1, CB3 and CB4 coded in blue.

3.2.3.2. Transcription

Transcription of the CB1-like phage genomes likely happens in a similar manner to that of *Escherichia* phage N4. All three phages possess a vRNA (CB1_77, CB3_82, CB4_81), which conducts transcription of single-stranded DNA that is initiated at hairpin promotors located within the early gene region. These promoters are composed of a 5 nucleotide hairpin with a 3 nucleotide loop possessing a central purine (Glucksmann et al., 1992). Inspection of the early gene region of phages CB1, CB3 and CB4 reveal the presence of three potential hairpin loops for all three, which appear analogous to hairpin loops in phage N4 (Appendix, Table S3.7). Middle gene transcription in phage N4 is conducted by a heterodimeric RNA polymerase related to the T7-like RNA polymerase family (Carter et al., 2003), with the CB1-like phages also possessing homologs (CB1_22, 23, CB3_24, 25 and CB4_24, 25). Finally, late gene transcription of phage N4 involves its ssDNA-binding protein activating the *E. coli* sigma 70 RNA polymerase directing it towards the N4 late promotors (Choi et al., 1995), with the ssDNA-binding protein also having homologs among the CB1-like phages (CB1_72, CB3_77 and CB4_76). Within the non-redundant parts of the genomes of the CB1-like phages, 18 putative rho-independent terminators were identified for CB1 and CB3 with 17 identified for CB4 (Appendix, Table S3.8, S3.9, S3.10). Comparison between these phages showed that 16 of these terminators were located at similar locations among their respective genomes. (Appendix, Table S3.11).

3.2.3.3. DNA replication, metabolism and methylation

The CB1-like phages have several components of a DNA replication system including a DNA polymerase I (CB1_46, CB3_50, CB4_49), a helicase (CB1_40, CB3_42, CB4_42) and a primase (CB1_67, CB3_72, CB4_71). Their ability to alter the nucleotide pool of their host exploiting a thymidylate synthase (CB1_63, CB3_68, CB4_67) allows the conversion of dUMP to dTMP and a putative nucleoside triphosphate pyrophosphohydrolase (CB1_87, CB3_92, CB4_91), which allows conversion of nucleoside triphosphates to their monomer form. These phages also possess a DNA adenine methylase (CB1_53, CB3_57, CB4_56) (IPR012327). Indeed, restriction digestion patterns of the genomic DNA of CB1, CB3 and CB4 using ClaI indicate that

their DNA is likely deoxyadenosine-methylated (Appendix, Figure S3.4, S3.5, S3.6). Such methylation is known to occur on the DNA of a number of other phages such as P1 and T4 of *E. coli*. It can have regulatory functions involved in phage DNA packaging and transcription but additionally can provide resistance against host restriction endonucleases (Murphy et al., 2013).

3.2.3.4. Cell lysis

Escherichia phage N4 possesses a signal-arrest-release (SAR) endolysin (*N*-acetylmuramidase, pfam05838) (Stojković and Rothman-Denes, 2007). Such endolysins are transported to the inner membrane by the host *sec* system and depend on pin holins to cause the collapse of membrane potential to induce their cell wall degrading activity (Savva et al., 2014). A typical feature of such endolysins is the possession of an N-terminal transmembrane domain. However, the CB1-like phage endolysin (Endolysin lambda type, IPR034691) (CB1_60, CB3_64, CB4_63) lacks this feature, thus indicating that their endolysin likely depends on the predicted class II holin (CB1_59, CB3_63, CB4_62) for release into the host cell periplasm to reach cell wall peptidoglycan (Savva et al., 2014). Situated next to these predicted ORFs for the endolysin and holin of the CB1-like phages are two overlapping ORFs (CB1_58, 58a, CB3_62, 62a, CB4_61, 61a), which possess one of the typical gene arrangements of a spanin rz and rz1 pair (Young, 2014). In this spanin pair arrangement, the gene for the rz protein is the larger of the two, encoding a protein with a transmembrane domain. The smaller gene encodes a lipoprotein. However, in the CB1-like example, the ORF predicted to encode the lipoprotein is the larger of the two with the smaller protein lacking the predicted transmembrane domain.

3.2.3.5. Structural proteome of phage CB1

There is a minimum of ten proteins that have been identified to form the virion of *Escherichia* phage N4 (Choi et al., 2008), with *in silico* analysis showing six to be shared with phage CB1. These six are the vRNA polymerase (CB1_77), the major capsid (CB1_83), the portal protein (CB1_86) and structural proteins of unknown function resembling those of gp52 (CB1_79), gp54 (CB1_81) and gp67 (CB1_90) of N4. Those

not found to be shared with CB1 were the tail sheath (gp65), the tail appendage (gp66), the head decorating protein (gp17) and one other structural protein (gp51) that has previously been suggested to be an internal virion protein (Hardies et al., 2016). Using *in silico* analysis, three proteins were identified to play potential roles in the morphogenesis of the tail structure of CB1 not shared with N4, namely CB1_57, 61 and 62. The CB1_61 ORF is likely to encode a tail spike given that it possesses an SGNH hydrolase-type esterase domain (IPR013830), suggesting enzymatic activity like the tail spike (gp63.1) of N4-like *Escherichia* phage G7C, which deacetylates host surface polysaccharides (Prokhorov et al., 2017). Mass spectrometry (ESI-MS/MS) conducted on the structural proteome of CB1 verified the presence of the shared N4 structural proteins along with the three identified putative tail proteins and one hypothetical protein (CB1_78) (Table 3.3). This latter protein of the CB1 virion mirrors the position of gp51 of phage N4 but shares no homology. All CB1 structural proteins identified possess homologs in phages CB3 and CB4, with the putative tail spike protein CB1_61 being split into two ORFs on the genome of CB4 (CB4_64, 65). Furthermore, the large terminase responsible for virion DNA packaging was identified for all three phages (CB1_91, CB3_97, CB4_95).

Table 3.3. Results of tandem mass spectrometry of proteins of the *Pectobacterium* phage CB1 virion.

ORF	Predicted function	Molecular mass (kDa)	No. of unique peptides	Sequence coverage %
CB1_57	Putative tail protein	15.59	6	71
CB1_61	Putative tail spike protein	104.98	19	29
CB1_62	Putative tail protein	26.01	10	55
CB1_77	Virion associated RNA polymerase (N4 gp50-like)	399.92	37	13
CB1_78	Unknown structural protein	41.64	15	52
CB1_79	Structural protein (N4 gp52-like)	14.2	3	19
CB1_81	Structural protein (N4 gp54-like)	28.82	9	53
CB1_83	Major capsid protein (N4 gp56-like)	43.33	33	75
CB1_86	Portal protein (N4 gp59-like)	83.31	27	45
CB1_90	Structural protein (N4 gp67-like)	33.72	9	40

3.2.3.6. Selfish genetic elements

Homing endonucleases are mobile genetic elements with endonuclease activity that only promote the spread of their own encoding gene (Stoddard, 2005). Several homing endonucleases of the HNH family (IPR003615) have been identified on the genomes of the three CB1-like phages. Phage CB1 itself was found to possess six homing endonucleases (CB1_24, 33, 39, 43, 52, 71), whereas phages CB3 (CB3_36, 47, 56, 76) and CB4 (CB4_36, 46, 55, 75) both possess four. Such genes have also been identified among other N4-like phages (Fouts et al., 2013; Kulikov et al., 2012).

3.2.4. Bacteriophage biocontrol on whole tubers

The phage mixture consisting of phages CB1, CB3 and CB4 was assessed for its ability to suppress soft rot caused by a mixed infection of by their host *P. atrosepticum* strains, DSM 18077 and DSM 30186, using whole tuber rot assays. The potato cultivar Rooster was selected for these experiments, given that it is the predominant variety grown in the Republic of Ireland (it comprised 60% of the Irish crop in 2014) (Bord Bia the Irish Food Board, 2015).

The whole tuber rot assay was carried out independently in triplicated and results were averaged. The assay involved two sets of ten tubers being treated with bacteria (designated sets (a) and (b), 100 µL at approx.$1.0x10^7$ CFU/mL), which were allowed to absorb into the tuber tissue. A third set (c) was treated with water. Sets (a) and (c) were treated with SM while (b) was treated with the phage mixture (100 µL at $1x10^7$ PFU/mL). Following incubation, the average weight of rotten tissue from each set of tubers was recorded. Set (a) (bacteria + SM buffer) was 5.39 ± 3.121 g; Set (b) (bacteria + phage) was 0.311 ± 0.498 g, thus the average weight of infected tubers treated with phage was less than that treated without phage, with this result being statistically significant ($p<0.0005$). Indicating phage treatment limited soft rot formation. No rot was observed for set (c) (Figure 3.7). Look to Appendix, Figure S3.7 for the visual outcome of treated tuber sets (a), (b) and (c).
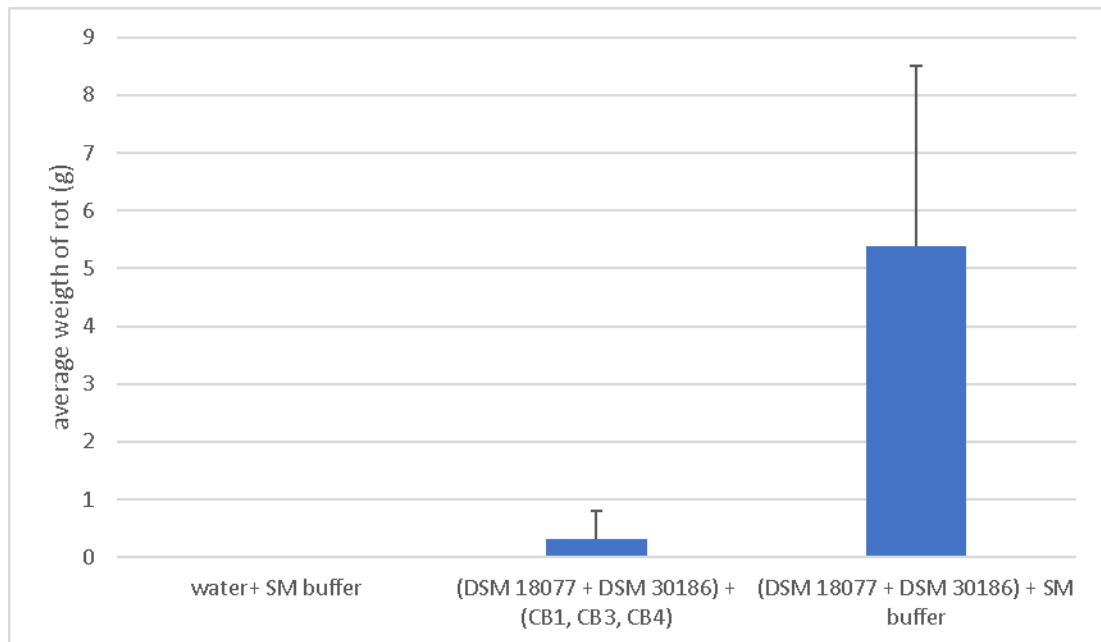
Figure 3.7. Protective effective effect of a phage mixture containing *Pectobacterium* phages CB1, CB3 and CB4 on whole tubers against a mixed infection of *Pectobacterium atrosepticum* strains DSM 18077 and DSM 30186. The assay [*n*=10] was carried out independently in triplicate and results were averaged.

3.3. Discussion

Of the sixteen samples of potato crops symptomatic of blackleg obtained from three potato farms in Co. Cork from which pectolytic isolates were found, fifteen were identified to be infected with *P. atrosepticum* with the remaining infection identified to be caused by a mixture of *P. atrosepticum* with *P. carotovorum* subsp. *carotovorum.* This study focused on the identification and isolation of exclusively lytic phages of *P. atrosepticum* from the potato crop environment, followed by an assessment of their potential for biocontrol applications against soft rot and/or blackleg disease. *P. atrosepticum* is highly relevant to Irish potato horticulture, and to date, relatively few reports focusing on phage biocontrol of this pathogen have appeared in the scientific literature.

During an enrichment screening for *P. atrosepticum* phages, three closely related phage isolates were obtained, namely CB1, CB3 and CB4. Genome sequencing revealed them to be N4-like, sharing a similar gene order and possessing all 18 core genes found among all N4-like members. For phage therapy applications involving the treatment of *Pseudomonas aeruginosa* murine infections, N4-like phages have been deemed to be

safe and effective (Shigehisa et al., 2016). This suggested that the three identified phages could be good candidates for biocontrol applications.

Phages CB1, CB3 and CB4 appear to possess relatively broad host ranges within their host species, collectively spanning 15 of the 19 tested *P. atrosepticum* strains (Table 3.1). By comparison, among the other members of the N4-like phages (infecting *Escherichia*, *Pseudomonas*, *Vibrio* and *Roseovarius*), many have been found to propagate only on the strain used in the original isolation (Chan et al., 2014; Katharios et al., 2017; Kulikov et al., 2012). The relative broad host range observed is desirable for their use in biocontrol applications. Nevertheless, the application of these phages would have limited biocontrol potential where other SRE species may be involved in disease. Additionally, *P. atrosepticum* strains tested consisted predominantly of those from an Irish environment. It is unknown if strains from a different geographical location would have a similar susceptibility to these phages. However, resistance could be overcome by supplementing the phage mixture with additional phages targeting resistant *P. atrosepticum* strains as well as other SRE species. The generation of phage cocktails is an important approach being adopted for phage preparations for use in the food industry and veterinary medicine as it ensures the widest possible host range against the targeted bacteria (Chan et al., 2013; García et al., 2008). Phages with host ranges limited within their respective host species have been observed for a number of other studies focusing on SRE phage biocontrol (Adriaenssens et al., 2012b; Lim et al., 2013; Smolarska et al., 2018). Nevertheless, a few notable exceptions have been described such as *Dickeya* phages D4 and D5 being able to infect several *Dickeya* species and *Dickeya* phages φPD10.3 and φPD23.1 being able to infect *D. solani*, *P. carotovorum* subsp*. carotovorum* and *P. parmentieri* (Czajkowski, 2016; Smolarska et al., 2018). To date, no broad host range phage infecting multiple SRE species including *P. atrosepticum* has been described.

The tuber rot assay in this study indicated that phages CB1, CB3 and CB4 appeared to possess some potential for the inhibition of soft rot formation of potato tubers (Figure 3.7). A similar finding was previously demonstrated using phages against *P. carotovorum* subsp. *carotovorum*, *D. solani* and *P. parmentieri* (Adriaenssens et al., 2012b; Czajkowski et al., 2015; Smolarska et al., 2018). An approximate multiplicity of infection (MOI) of 1 was used in our experiments to achieve this effect, lower than that

reported with *Dickeya* phage vB_DsoM_LIMEstone1 at an MOI of 10 and 100, (Adriaenssens et al., 2012b) but higher than that with *Dickeya* phages ΦPD10.3 and ΦPD23.1 against *P. carotovorum subsp. carotovorum, D. solani* and *P. parmentieri* at an MOI of 0.01 (Czajkowski et al., 2015). However, further work is necessary to examine the inhibitory effect of soft rot on potato tubers by the CB1-like phages in more detail as well as additional studies to explore if the phages could inhibit blackleg formation on whole plants. Obviously, other environmental conditions that could potentially affect the viability of these phages in field applications such as UV light, salinity, soil composition, agrochemical factors (including fertilisers, copper based compounds, etc.) would also require investigation.

Of the ten structural proteins identified to form the CB1 virion (Table 3.4), four were not shared with that of *Escherichia* phage N4 (CB1_57, 61, 62 and 78). These proteins appear to be associated with tail components of the CB1 virion. Their lack of homology is likely due to their adaption to allow recognition of its host bacterium and enable subsequent DNA injection. For N4-like *Escherichia* phage G7C to recognise its host, its tail spike esterase domain must first deacetylate the O-antigen of its host lipopolysaccharide (Prokhorov et al., 2017). This may also be the case for the CB1-like phages with their putative tail spike (CB1_61, CB3_66 and CB4_75, 76) possessing an SGNH hydrolase-type esterase domain (IPR013830).

Phylogenetic analysis (based on phylograms and Gegenees (TBLASTX) of CB1-like phages show that they are distinct members within the N4-like group, with their closest evolutionary relationship being found with phages of the genera *Lit1virus* and *Luz7virus*. This analysis also indicates the existence of higher-order taxonomic relationships between genera of the N4-like phages. Such relationships are becoming apparent due to the increasing number of N4-like genomes being added to public databases. Such a relationship was identified between the phages belonging to the genera of *N4virus*, *G7cvirus* and *Ea92virus* with *Achromobacter* phage JWAlpha, which has been proposed to form the subfamily *Enquartavirinae* (Wittmann et al., 2015). A similar connection was also found with the genera of *Luz7virus* and *Lit1virus* (Figure 3.5, Table 3.3). These relationships are indicated with clades that are closely situated within phylograms and the sharing of protein sequence identity of ~30-40% in Gegenees analysis.

A taxonomy proposal describing the CB1-like phages as a new genus has been submitted to the International Committee for the Taxonomy of Viruses (ICTV), proposing the genus of *Cbunavirus*.

Chapter 4. *Pectobacterium atrosepticum* phage vB_PatP_CB5; a member of the proposed genus '*Phimunavirus*'

A manuscript based upon this chapter has been published in Viruses

Transmission electron microscopy was conducted by Horst Neve.

Genome sequence of phage CB5 was outsourced to GATC Biotech, Germany.

4.1. Introduction

In the post-genomic era, the number of bacteriophage (phage) genomes being deposited into public databases, such as NCBI GenBank, has substantially increased due in part to the ever-decreasing cost of DNA sequencing. This growing quantity of genomic data has led to increasing insights into the evolutionary relationships between phages. Originally, phage taxonomic classification was based on morphology, nucleic acid composition and physicochemical characteristics (Calendar, 2006). More recently, classification has developed to the point where nucleotide and protein homology can be usefully employed to tease out phylogenetic relationships. This has led to the creation of subfamilies within of the *Myoviridae*, *Siphoviridae* and *Podoviridae*, such as the subfamily *Autographivirinae*. This subfamily encompassed what was previously known as the T7 supergroup (Lavigne et al., 2008). Key defining features of the subfamily *Autographivirinae* include the presence of a single RNA polymerase (RNAP) gene and a typical genomic organisation with genes positioned on the Watson strand (King et al., 2012). To date, this subfamily currently encompasses seven genera: *T7virus*, *SP6virus*, *Phikmvvirus*, *Fri1virus*, *KP32virus*, *Pradovirus* and *KP34virus* (Adriaenssens et al., 2017; Lavigne et al., 2008).

The bacterial genera *Pectobacterium* and *Dickeya*, often referred to collectively as the soft rot *Enterobacteriaceae*, are phytopathogens that cause economically important losses in a wide range of arable crops, thus potentially impacting food biosecurity. They are Gram-negative, facultative anaerobic rod-shaped cells that are typified by the production of extracellular pectinolytic enzymes during infection of plants (Pérombelon, 2002; Toth et al., 2011). Within the last two years, a limited number of *Pectobacterium* and *Dickeya* phages have been reported whose genome sequences have been described to resemble phages of *KP34virus* and *Phikmvvirus*. The first of these to have their genome described was *Pectobacterium atrosepticum* phage Peat1 (accession no. KR604693) by Kalischuk *et al.* (Kalischuk et al., 2015). Phage ϕM1 (accession no. JX290549) was subsequently described by Blower *et al.* (Blower et al., 2017), after isolation and characterisation by Toth *et al.* (Toth et al., 1997). Related phages have also been described for *Pectobacterium carotovorum* subsp. *carotovorum* (namely phage PPWS1, accession no. LC063634.2) and *Dickeya* (phage BF25/12, accession no. KT240186.1) (Alič et al., 2017; Hirata et al., 2016). In addition, a *P.*

atrosepticum phage, PP90 (accession no. KX278419.1), has been deposited to the public databases as well as *P. carotovorum* subsp*. carotovorum* phage PP16 (accession no. KX278418). The latter two display high level of amino acid sequence similarity and possess a similar genomic organisation of genes to phages of *KP34virus*.

In this study, we describe the newly isolated *P. atrosepticum* phage vB_PatP_CB5. Phylogenetic analysis of its genome shows a close evolutionary relationship with *P. atrosepticum* phages φM1, Peat1 and PP90 (termed the PhiM1-like phages from here onwards in this article). Based on these findings, we propose the formation of the bacteriophage genus '*Phimunavirus*' to formally classify these phages, with *Pectobacterium* phage φM1 designated as the type phage.

4.2. Results

4.2.1. Isolation, host range, growth characteristics and morphology

Phage CB5 was isolated from soil samples collected from potato grading machinery on a farm in Co. Cork, Ireland, during the year 2013, as mentioned previously (Buttimer et al., 2018a). Host range was determined on 31 bacterial strains from five different species belonging to soft rot *Enterobacteriaceae*, namely *P. atrosepticum* (19 strains), *P. carotovorum* subsp. *carotovorum* (4 strains), *D. chrysanthemi* bv. *chrysanthemi* (1 strain), *D. dianthicola* (3 strains) and *D. solani* (4 strains). The phage possesses a narrow host range, in that it is only capable of forming plaques on the phage's host strain (DSM 30186) and two other strains of *P. atrosepticum* (Table 4.1)*.* Additionally, spot tests showed that the phage had an inhibitory effect on 15 of the other 16 strains of *P. atrosepticum* tested with the observation of zones of clearing at high phage titres despite the absence of distinct plaques. No plaque formation or inhibition was detected for any other bacterial species tested. A similar narrow host range has also been reported for PhiM1-like *Pectobacterium* phage φM1, with an infectivity range confined to a small number of *P. atrosepticum* strains (Toth et al., 1997). This limited host range has also been observed among phages of *KP34virus* and *Phikmvvirus* (Ceyssens et al., 2006; Eriksson et al., 2015). On the propagating host strain DSM 30186, CB5 was found to produce clear plaques with an approximate diameter of 3 mm (Appendix, Figure S4.1). One-step-growth curve assay, under standard conditions

using LB medium, demonstrated that phage CB5 possessed a latent period of 45 mins with an approximate burst size of 43 PFU/cell (Figure 4.1).

Table 4.1. Host range of *Pectobacterium* phage vB_PatP_CB5 (CB5) on 31 strains of the soft rot *Enterobacteriaceae* determined by spot testing with serial dilutions of phage.

Species	Strain	Sensitivity
Pectobacterium atrosepticum	DSM 18077 (type strain)	++
	DSM 30184	+
	DSM 30185	+
	DSM 30186	++*
	CB BL1-1	+
	CB BL2-1	+
	CB BL3-1	+
	CB BL4-1	+
	CB BL5-1	+
	CB BL7-1	+
	CB BL9-1	+
	CB BL11-1	+
	CB BL12-2	++
	CB BL13-1	+
	CB BL14-1	+
	CB BL15-1	-
	CB BL16-1	+
	CB BL18-1	+
	CB BL19-1	+
Pectobacterium carotovorum subsp. *carotovorum*	DSM 30168 (type strain)	-
	DSM 30169	-
	DSM 30170	-
	CB BL19-1-37	-
Dickeya chrysanthemi bv *chrysanthemi*	LMG 2804	-
Dickeya dianthicola	PD 482	-
	PD 2174	-
	GBBC 1538	-
Dickeya solani	sp. PRI 2222 (D36)	-
	LMG 25865 (D10)	-
	GBBC 1502	-
	GBBC 1586	-

Results recorded as ++, sensitive; +, the presence of clear spot with no plaque formation; –, no infection; *host strain of phage.
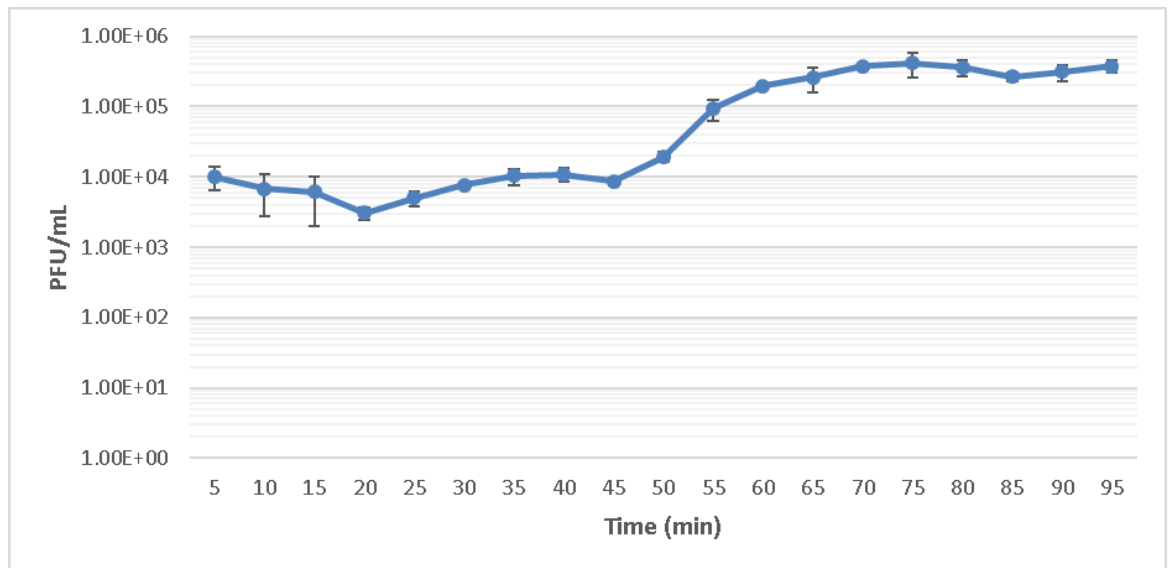
Figure 4.1. Single-step growth curve of *Pectobacterium* phage vB_PatP_CB5 infection of host bacterium DSM 30186. The assay was independently repeated in triplicate and the results were averaged.

Examination of the morphology of the phage by transmission electron microscopy (Figure 4.2), shows it can be classified as a member of the *Podoviridae* family, featuring a C1 morphotype (Ackermann, 2001) with an icosahedral head (63.1 ± 3.6 nm in diameter, *n*=25) with clearly distinguishable hexagonal outlines and a short non-contractile tail (13.1 ± 1.8 nm, *n*=11), and short appendices (length: ca. 10.1 ± 1.7 nm, *n*=10) visible at the head/tail connection site. These head and tail dimensions are consistent with previously reported phages of *KP34virus* and *Phikmvvirus* (Eriksson et al., 2015; Lavigne et al., 2003). The phage was formally named in accordance with the nomenclature set out by Kropinski *et al.* (2009).
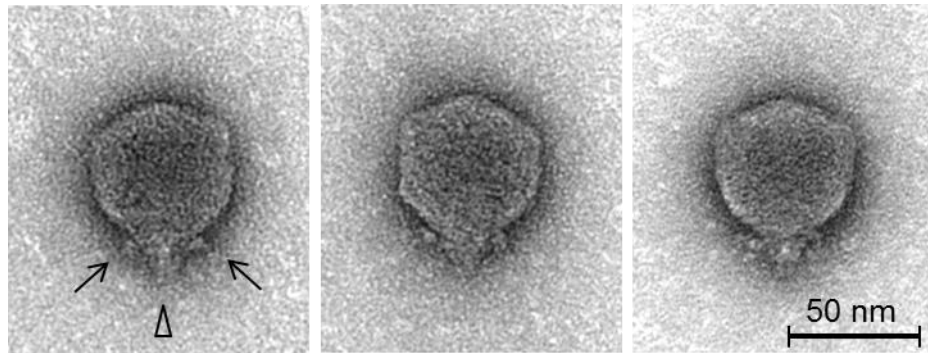
Figure 4.2. Transmission electron microscopy of negatively stained *Pectobacterium* phage CB5 using 1% (w/v) uranyl acetate. Triangle indicates the short conical tail structure, and arrows indicate short appendages (whiskers) beneath the capsid. Scale bars represent 50 nm.

4.2.2. General genome information of CB5

The genome sequence size obtained for phage CB5 is 44,262 bp (coverage >1000x) and examination of sequence reads indicated the presence of direct terminal repeats (DTRs) with an estimated size of 287 bp. This estimation is based on the identification of a localized region with more than double the read depth compared to average read depth across the whole genome (Appendix, Figure S4.2). This approach to detecting DTRs has also been applied to a number of other phages (Buttimer et al., 2017a; Fouts et al., 2013; Li et al., 2014). Additionally, the average G+C% content of its genome was found to be 49%: close to the range that is typically expected for its host bacterium, *P. atrosepticum* at 50-51% (Bell et al., 2004; Nikolaichik et al., 2014).

The genome of CB5 was predicted to contain 60 ORFs. These, apart from one, were found to read in the 5' to 3' direction, with GC skew correlating well with transcription (Marín and Xia, 2008). Based on analysis using a combination of BLASTP, InterProScan and HHpred, putative functions to 33 of the 60 predicted ORFs (55%) were assigned. These gene products can be categorized into DNA replication, virion structure and host lysis functions (Appendix, Table S4.1). Of the assigned ORFs, five are predicted to encode homing endonucleases of the HNH family (CB5_17, 23, 29, 41, 48) (IPR003615). No integrase, excisionase nor repressor genes were detected suggesting the phage has an exclusively lytic lifecycle. Furthermore, no tRNA genes were identified.

4.2.3. Comparative genomics of PhiM1-like phages

The four phages within the proposed genus '*Phimunavirus*' possess genomes of similar size that share a high degree of sequence similarity and share a large number of conserved proteins (Table 4.2). Genome sizes (excluding DTRs) range from 43,534 bp (φM1) to 45,633 bp (Peat1), with nucleotide pairwise identity between the four phages ranging from 82 to 86% (BLASTN). Total ORF numbers range from 52 (φM1) to 61 (Peat1), with Coregenes analysis showing that they collectively share a minimum of 39 proteins, including 32 which were affiliated to a known pVOG (Table 4.3). These coding sequences are spread across the entire genome and are not associated/limited to particular genomic modules. Of the four phages, only φM1 possesses a tRNA gene (for isoleucine). G+C content among the four phages is highly similar, ranging from 48.7 to 49.2%. Additionally, these phages show limited similarity to *Klebsiella* phage KP34 (7 to 9% identity). Indeed, Coregenes shows that phage KP34 shares 29 proteins with the PhiM1-like phages (Table 4.3). The major variations of conserved proteins of phage KP34 with these phages are five hypothetical proteins located in the DNA replication and nucleotide metabolism region. Also, additional hypothetical proteins were identified immediately downstream of the genes encoding the RNAP and large terminase proteins, as well as the ORFs for their predicted holin and tail spike.

Table 4.2. Properties of the seven phages belonging to the proposed genus of '*Phimunavirus*'.

Phage	Genome size (bp)	DTRs (bp)	G+C content, %	ORFs	tRNA	DNA identity, %*	Homologous proteins, %**
φM1	43,534	293	49.18	52	1	100	100
CB5	44,262	287	48.98	60	0	84	73
Peat1	45,633	NA	48.86	61	0	86	87
PP90	44,570	NA	48.89	56	0	86	80

* DNA identity in comparison to φM1 using BLASTN.
**Number of homologous proteins in comparison to φM1 using Coregenes.
NA, not available.

Table 4.3. Thirty-nine conserved genes among phages (φM1, Peat1, CB5, PP90) of the proposed genus 'Phimunavirus', as determined by Coregenes, and their details in regard to the type phage φM1. Conserved proteins of these phages shared with *Klebsiella* phage KP34 are highlighted in bold with italics. Additionally, pVOGs determined from translated ORFs of φM1 of these conserved proteins are presented.

no.	Product	φM1 Accession no.	φM1 locus tag	pVOG
1	*hypothetical protein*	*AFQ22488.1*	*PhiM1_03*	*VOG6006*
2	*hypothetical protein*	*AFQ22489.1*	*PhiM1_04*	*VOG1073*
3	*hypothetical protein*	*AFQ22493.1*	*PhiM1_08*	*VOG5528*
4	*hypothetical protein*	*AFQ22494.1*	*PhiM1_09*	*-*
5	*putative peptidase*	*AFQ22495.1*	*PhiM1_10*	*VOG5332*
6	*hypothetical protein*	*AFQ22496.1*	*PhiM1_11*	*VOG5704*
7	*putative DNA primase*	*AFQ22497.1*	*PhiM1_12*	*VOG4551*
8	*putative DNA helicase*	*AFQ22499.1*	*PhiM1_14*	*VOG0025*
9	hypothetical protein	AFQ22501.1	PhiM1_16	-
10	hypothetical protein	AFQ22503.1	PhiM1_18	-
11	*DNA polymerase*	*AFQ22505.1*	*PhiM1_20*	*VOG0026*
12	*hypothetical protein*	*AFQ22506.1*	*PhiM1_21*	*VOG1076*
13	*DNA exonuclease*	*AFQ22507.1*	*PhiM1_22*	*VOG0028*
14	hypothetical protein	AFQ22508.1	PhiM1_23	-
15	*DNA endonuclease VII*	*AFQ22510.1*	*PhiM1_25*	*VOG8238*
16	*putative metallophosphoesterase*	*AFQ22512.1*	*PhiM1_27*	*VOG1606*
17	hypothetical protein	AFQ22514.1	PhiM1_29	VOG1254
18	hypothetical protein	AFQ22515.1	PhiM1_30	VOG9679
19	*putative RNA polymerase*	*AFQ22516.1*	*PhiM1_31*	*VOG0019*
20	hypothetical protein	AFQ22517.1	PhiM1_32	VOG1406
21	hypothetical protein	AFQ22518.1	PhiM1_33	VOG9202
22	*putative structural protein*	*AFQ22519.1*	*PhiM1_34*	*VOG8332*
23	*putative head-tail connector protein*	*AFQ22520.1*	*PhiM1_35*	*VOG0030*
24	*putative scaffolding protein*	*AFQ22521.1*	*PhiM1_36*	*VOG0031*
25	*putative endonuclease*	*AFQ22522.1*	*PhiM1_37*	*-*
26	*putative capsid protein*	*AFQ22523.1*	*PhiM1_38*	*VOG4572*
27	*putative tail tubular protein A*	*AFQ22524.1*	*PhiM1_39*	*VOG4592*
28	*putative tail tubular protein B*	*AFQ22525.1*	*PhiM1_40*	*VOG0034*
29	*putative internal core protein A*	*AFQ22526.1*	*PhiM1_41*	*VOG1080*
30	*putative internal core protein B*	*AFQ22527.1*	*PhiM1_42*	*VOG3794*
31	*putative internal core protein C*	*AFQ22528.1*	*PhiM1_43*	*VOG0038*
32	*putative tail fiber protein*	*AFQ22529.1*	*PhiM1_44*	*-*
33	*putative DNA maturase A*	*AFQ22530.1*	*PhiM1_45*	*VOG0041*
34	*putative DNA maturase B*	*AFQ22531.1*	*PhiM1_46*	*VOG4544*
35	hypothetical protein	AFQ22532.1	PhiM1_47	-

36	*putative Rz1A protein*	*AFQ22534.1*	*PhiM1_49*	*VOG1082*
37	putative holin	AFQ22535.1	PhiM1_50	VOG0765
38	*endolysin*	*AFQ22536.1*	*PhiM1_51*	*VOG4565*
39	phage tail spike protein	AFQ22537.1	PhiM1_52	VOG4640

The gene order is highly conserved among the PhiM1-like phages (Figure 4.3). Their genome architecture is arranged so that the predicted early and middle gene regions end with an RNAP gene (CB5_39), with ORFs within these regions involved in DNA replication and nucleotide metabolism, but also expected to be involved in host take-over. The position of ORFs for RNAP in the PhiM1-like phages is shared with phages of the genera *KP34virus*, *Fri1virus* and *KMVvirus*, but not with those of the genera *T7virus*, *SP6virus* and *KP32virus* where the RNAP is situated at the early gene region (Figure 4.4). The late gene region of the PhiM1-like phages is associated with virion morphogenesis and host lysis roles. Gene order between PhiM1-like phages and KP34-like phages is highly conserved apart from the position of an ORF encoding a conserved protein (CB5_36, PhiM1_27, AXI77_gp27, PP90_28) possessing a calcineurin-like phosphoesterase domain (IPR004843) (Figure 4.4).
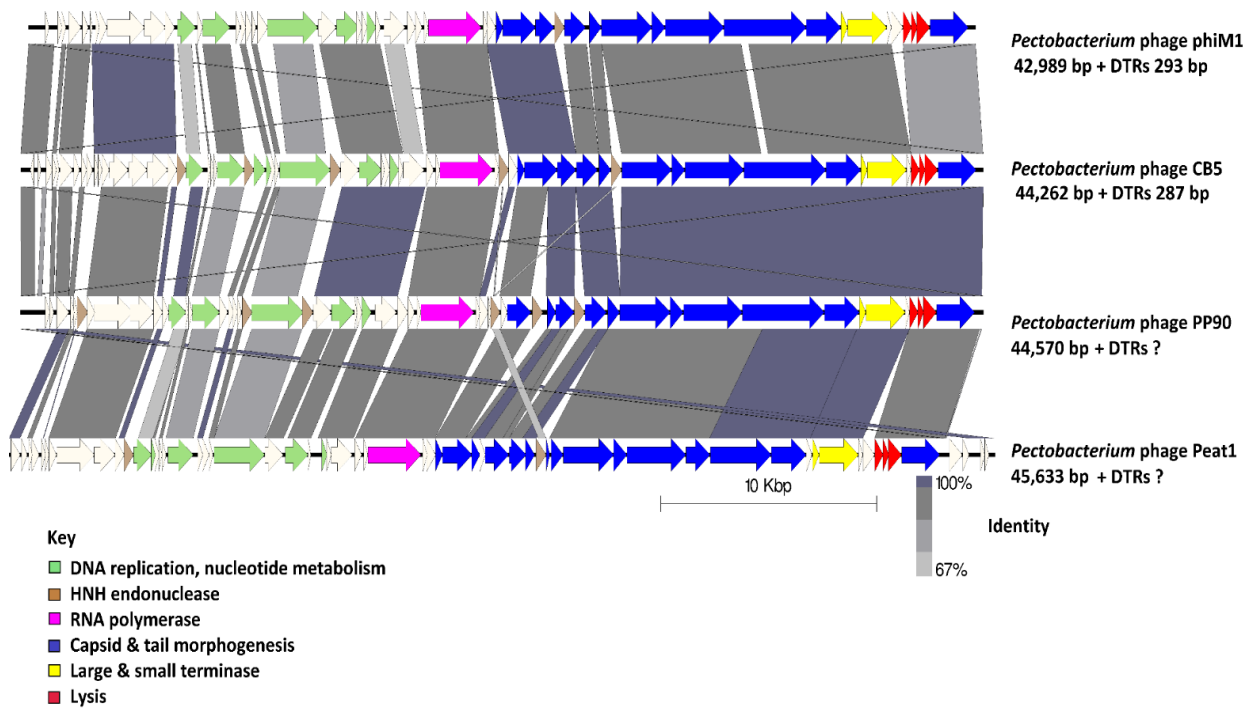
Figure 4.3. Comparison of the genomes of the phages that form the proposed genus of '*Phimunavirus*'; *Pectobacterium* phages CB5 and *Pectobacterium* phages φM1, Peat1 and PP90 using currently available annotation from Genbank, employing BLASTN and visualised with Easyfig. The genome maps comprise of arrows indicating locations and orientation of ORFs among different phage genomes. Arrows have been colour-coded describing their predicted roles (see key), and shading between the genome maps indicate the level of identity.
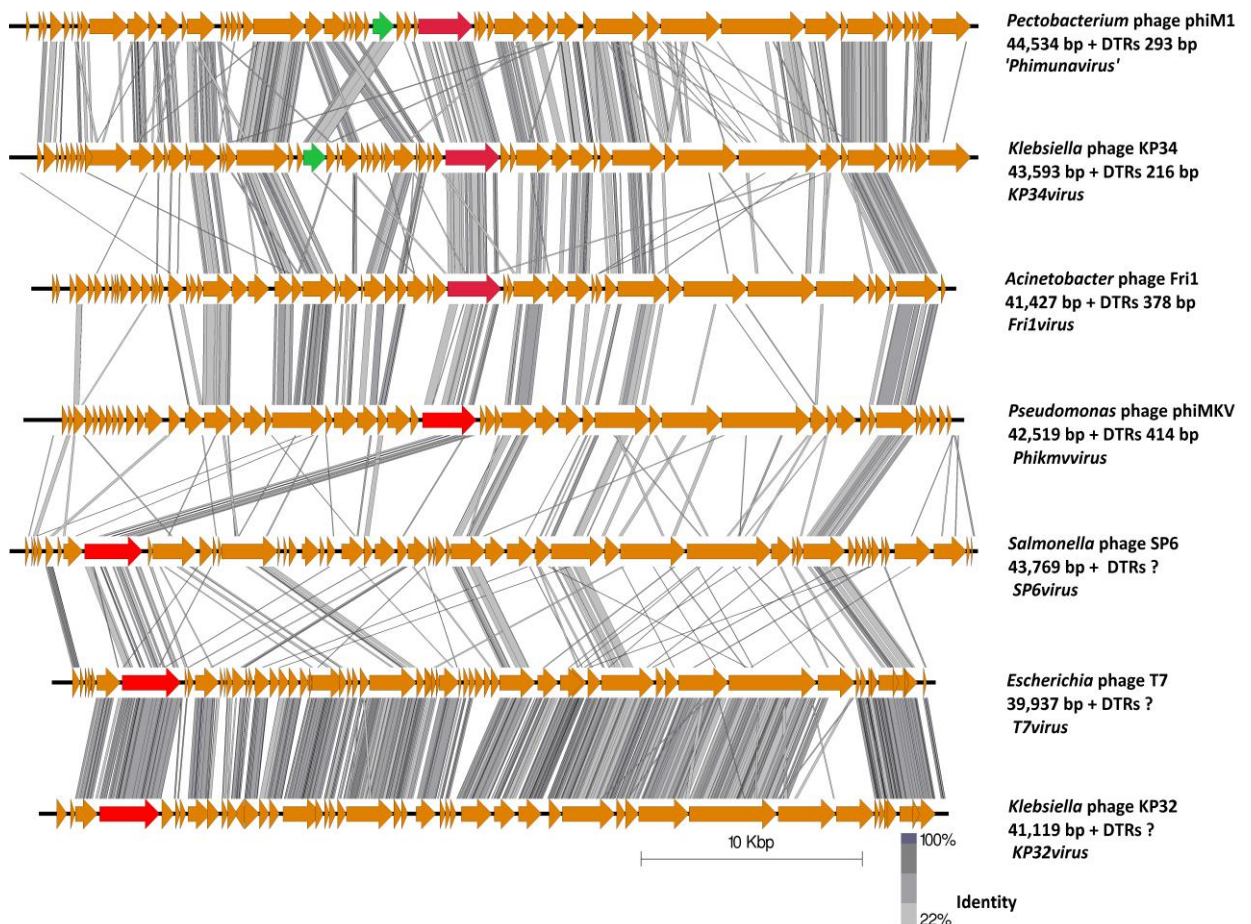
Figure 4.4. Pairwise comparison of the type phages representing six genera of the subfamily *Autographivirinae*; *KP34virus*, *Phikmvvirus*, *Frivirus*, *SP6virus*, *T7virus* and against *Pectobacterium* phage φM1. Maps were created using currently available annotation from Genbank with comparisons employing TBLASTX and visualised with Easyfig. The genome maps comprise of arrows indicating locations and orientation of ORFs. Lines between genome maps indicate the level of identity. ORF of a conserved protein shared between φM1 and KP34 that does not mirror genomic position is colour-coded green. The ORF encoding the RNAP shared between genera of *Autographivirinae* has been colour-coded red.

4.2.4. *Phimunavirus* evolutionary position the *Autographivirinae*

To determine the PhiM1-like phages' evolutionary relationship to other phages within the *Autographivirinae*, a phylogenetic analysis was conducted using the amino acid sequence of the major capsid protein (Figure 4.5). The resulting phylogram showed that the PhiM1-like phages form their own clade on a branch containing an additional clade representing the *P. carotovorum* subsp. *carotovorum* phages PP16 and PPSW1 and *Dickeya* phage BF25/12. The branch that these phages form, was found to be positioned close to phages of the *KP34virus* genus along with sister groups consisting

of *Vibrio* phage VP93 and the *Pantoea* phage LIMElight, which have previously been described to possess a close evolutionary relationship to the *KP34virus* members (Eriksson et al., 2015). This analysis was performed with the head-tail connector protein, large terminase subunit, and tail tubular proteins A and B. These analyses produced tree exhibiting similar relationships, albeit with weaker bootstrap support values (Appendix, Figure S4.3). Whole-genome comparison based on amino acid sequences was performed using VICTOR and the resulting phylogram (formula D4, yielding average support of 71 %) presented a similar conclusion like that based on major head protein sequence (Figure 4.6). Additionally, analysis using VICTOR could cluster these 53 phage genomes into 13 genera and four subfamilies, with PhiM1-like phages being placed in their own genera with *P. carotovorum* subsp. *carotovorum* phages PP16 and PPSW1 and *Dickeya* phage BF25/12 (Appendix, Table S4.2). Further analysis using Gegenees (TBLASTX) based on protein similarity, indicate the PhiM1-like phages form a clade with high identity values of ≥80%. Additionally, as seen in the phylograms, it was observed that *P. carotovorum* subsp. *carotovorum* phages PP16 and PPSW1 and *Dickeya* phage BF25/12 share an evolutionary relationship with PhiM1-like phages (identity values ≥54%) (Figure 4.7).
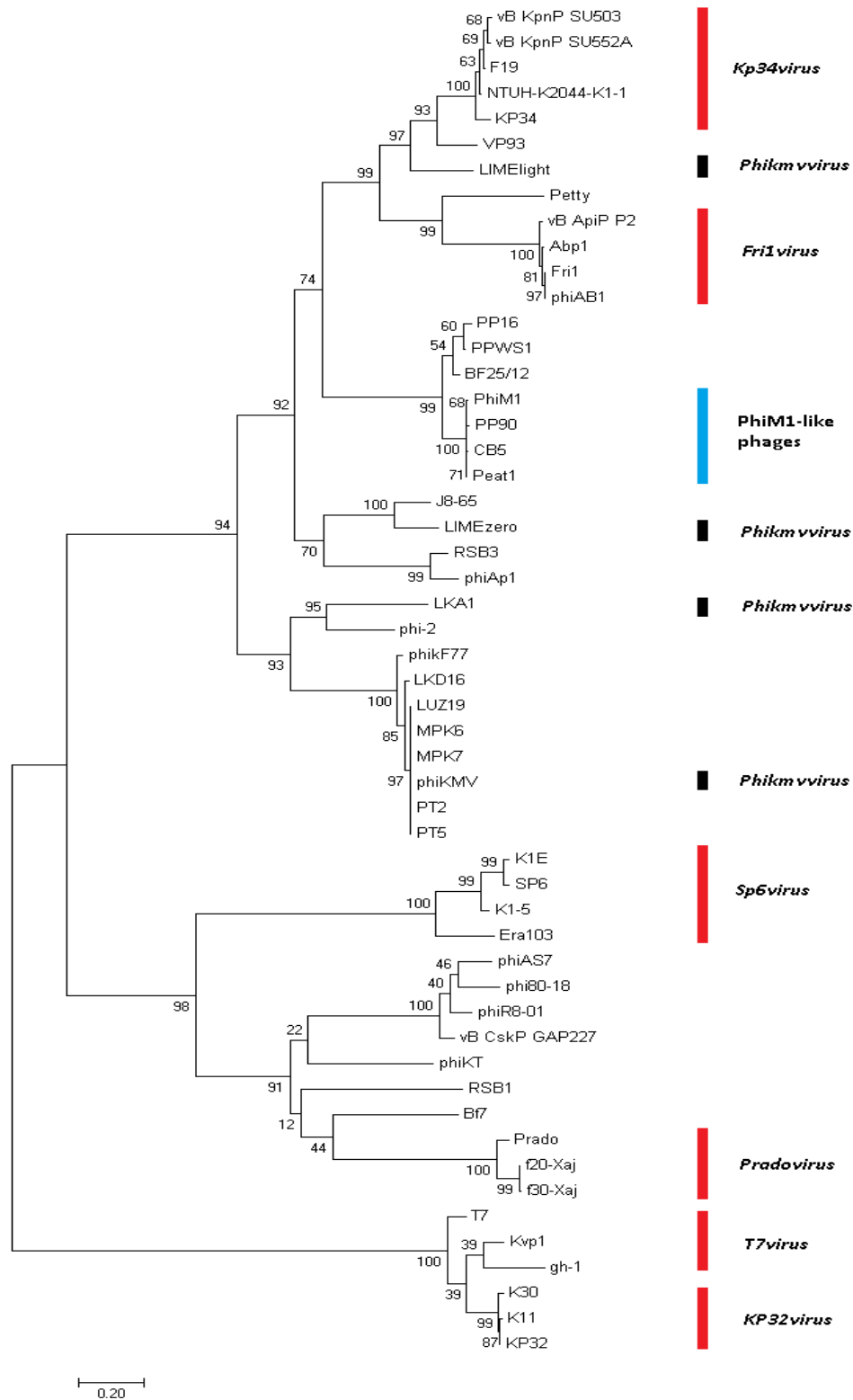
Figure 4.5. Phylogenetic analyses of amino sequences of the major capsid (log likelihood = 13809.83) of *Pectobacterium* phage CB5 and 52 members of the *Autographivirinae* subfamily using maximum likelihood (Whelan and Goldman substitution model), with 1000 bootstrap replicates. Members of the *T7virus*, *SP6virus*, *KP34virus, Frivirus, Pradovirus, KP32virus* are illustrated.
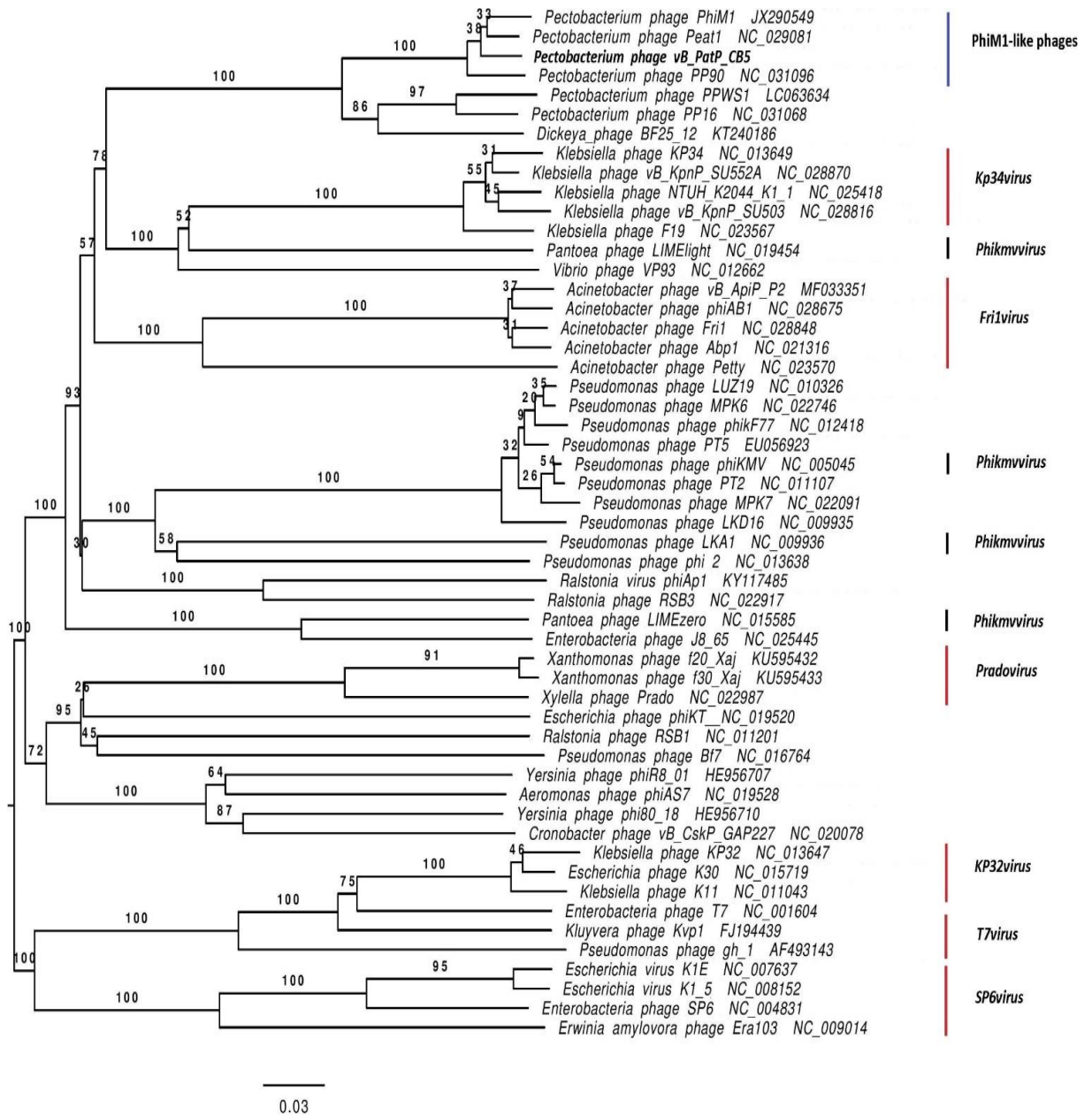
Figure 4.6. VICTOR generated phylogenomic GBDP trees of *Pectobacterium* phage CB5 and 52 members of the *Autographivirinae* subfamily inferred using the formula D4 and yielding average support of 71 %. The numbers above branches are GBDP pseudo-bootstrap support values from 100 replications. Members of the *T7virus*, *SP6virus*, *KP34virus*, *Frivirus*, *Pradovirus*, *KP32virus* are illustrated.

Phage	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
1: *Acinetobacter* phage Abp1	100	86	80	80	34	20	19	19	19	18	18	20	20	20	20	20	20	21	21	21	22	21	21	21	18	18	18	18	18	18	18	18
2: *Acinetobacter* phage phiAB1	87	100	81	80	34	20	20	19	19	18	18	20	20	20	20	20	20	21	21	21	21	21	21	20	18	18	18	18	18	18	18	18
3: *Acinetobacter* phage vB_ApiP_P2	81	81	100	80	33	20	20	19	19	18	18	20	20	20	20	20	20	20	21	20	21	21	21	20	18	18	18	18	18	18	18	18
4: *Acinetobacter* phage Fri1	81	80	80	100	33	19	20	19	19	18	18	20	20	20	20	20	20	21	21	21	21	21	21	21	18	18	18	18	18	18	18	18
5: *Acinetobacter* phage Petty	34	35	34	33	100	20	20	19	19	18	18	20	20	20	20	20	20	21	21	21	21	21	21	21	18	18	19	18	18	18	18	18
6: *Ralstonia* virus phiAp1	18	18	18	18	18	100	42	22	22	20	20	20	20	20	20	20	20	21	20	20	20	20	20	22	22	22	22	22	22	22	22	22
7: *Ralstonia* phage RSB3	19	19	19	19	19	42	100	23	23	21	21	20	20	20	20	20	20	21	21	21	21	21	21	21	23	23	22	22	22	22	23	23
8: *Pseudomonas* phage LKA1	18	18	18	18	18	23	23	100	31	20	20	20	20	20	20	20	20	21	21	21	21	21	21	21	28	28	28	28	28	28	28	28
9: *Pseudomonas* phage phi-2	18	18	18	18	18	22	23	30	100	20	19	20	20	20	19	20	20	21	21	20	20	21	21	21	29	29	29	29	29	29	29	29
10: Enterobacteria phage J8-65	17	17	17	17	17	21	21	21	20	100	47	20	19	18	18	19	19	21	21	21	20	20	20	21	20	20	20	20	20	20	20	20
11: *Pantoea* phage LIMEzero	17	17	17	17	17	20	20	20	19	46	100	19	18	18	18	18	18	20	20	20	19	19	20	20	19	19	19	19	19	19	19	19
12: *Pantoea* phage LIMElight	19	19	19	19	19	20	20	20	20	19	19	100	30	30	31	31	30	24	24	24	23	23	23	23	20	20	20	20	20	20	20	20
13: *Klebsiella* phage NTUH-K2044-K1-1	19	19	19	19	19	20	20	20	20	18	18	31	100	76	75	76	73	25	25	25	25	25	25	25	20	20	20	20	20	20	20	20
14: *Klebsiella* phage vB_KpnP_SU503	19	19	19	19	19	20	20	20	20	18	18	31	77	100	77	76	73	25	25	25	25	25	25	25	20	20	20	20	20	20	20	20
15: *Klebsiella* phage KP34	19	19	19	19	19	20	20	20	20	18	18	31	76	76	100	78	74	25	25	25	25	25	25	25	20	20	20	20	20	20	20	20
16: *Klebsiella* phage vB_KpnP_SU552A	19	19	19	19	19	20	20	20	20	19	18	32	76	75	77	100	75	26	26	25	25	25	25	26	20	20	20	20	20	20	20	20
17: *Klebsiella* phage F19	19	19	19	19	19	20	20	20	20	18	18	31	74	73	74	76	100	25	25	25	25	25	25	25	20	20	20	20	20	20	20	20
18: *Dickeya* phage BF2512	19	20	19	19	19	21	21	21	21	21	20	24	25	25	25	25	25	100	61	60	57	56	55	56	21	21	21	21	21	21	21	21
19: *Pectobacterium* phage PP16	20	20	20	20	20	21	21	21	21	21	20	24	25	25	25	26	25	61	100	72	55	55	54	55	21	21	21	21	21	21	21	21
20: *Pectobacterium* phage PPWS1	20	20	19	19	20	21	21	21	21	20	20	24	25	25	25	25	24	59	71	100	55	55	54	54	20	21	21	21	21	21	21	21
21: *Pectobacterium* phage PP90	21	21	20	20	21	21	21	21	21	20	20	24	25	25	25	25	25	56	55	55	100	85	82	80	20	20	20	20	20	20	20	20
22: *Pectobacterium* phage vB_PatP_CB5	20	20	20	20	20	21	21	21	21	20	19	23	25	25	25	25	25	56	56	56	84	100	85	83	21	21	20	20	20	20	20	21
23: *Pectobacterium* phage Peat1	20	20	20	20	20	21	21	21	21	20	20	23	25	25	25	25	25	55	54	54	81	86	100	86	20	20	20	20	20	20	20	20
24: *Pectobacterium* phage PhiM1	20	20	20	20	20	21	21	21	21	21	20	24	25	25	25	26	25	57	55	55	82	83	87	100	21	21	21	21	21	21	21	21
25: *Pseudomonas* phage LKD16	17	17	17	17	17	22	22	27	29	19	19	19	19	19	19	19	20	20	20	20	20	20	20	20	100	79	79	79	77	78	78	80
26: *Pseudomonas* phage phikF77	17	17	17	17	17	22	22	27	29	19	19	19	20	20	19	20	20	21	20	20	20	20	20	20	79	100	85	84	81	81	82	84
27: *Pseudomonas* phage LUZ19	17	17	17	17	17	22	22	27	29	19	19	19	20	20	19	19	19	20	20	20	20	20	20	20	78	85	100	90	85	86	84	86
28: *Pseudomonas* phage MPK6	17	17	17	17	17	22	22	27	29	19	19	19	20	20	19	19	20	21	20	20	20	20	20	20	79	85	91	100	88	86	84	87
29: *Pseudomonas* phage MPK7	17	17	17	17	17	22	22	28	29	19	19	19	20	19	19	20	20	21	20	20	20	20	20	20	77	81	86	88	100	87	86	84
30: *Pseudomonas* phage PT2	17	17	17	17	18	22	22	27	29	19	19	19	19	20	19	20	20	21	20	20	20	20	20	20	78	82	86	85	87	100	95	91
31: *Pseudomonas* phage phiKMV	17	17	17	18	18	22	22	27	29	19	19	20	20	20	20	20	20	21	21	20	20	20	20	20	79	83	85	85	86	96	100	94
32: *Pseudomonas* phage PT5	17	17	17	17	17	22	22	27	29	19	19	19	19	19	19	19	19	20	20	20	20	20	20	20	80	84	87	87	84	92	93	100

Figure 4.7. TBLASTX heat map generated using Gegenees with accurate parameters – fragment length: 200 bp; and step size: 100 bp with the threshold set to 5%. The map includes the genomes of 32 phages of Autographivirinae with phages representing the genera *Fri1virus* (yellow), *KP34virus* (brown), *Phikmvvirus* (green) and the proposed genus of *'Phimunavirus'* (blue).

4.2.5. RNAP of the PhiM1-like phages

The RNAP protein has been used as a feature to establish genera within the subfamily of *Autographivirinae*. Within the amino acid sequence of the RNAP of phage T7, the key catalytic residues Asp537, Lys631, Tyr639 and Asp812; and the regions of functional importance, namely the recognition loop (93-101 aa) and the specificity loop (739-770 aa) are generally well conserved among different clades within *Autographivirinae* (Ceyssens et al., 2006; Eriksson et al., 2015). Analysis of PhiM1-like phages shows that they all have the catalytic residues Asp537, Lys631 and Asp812. Comparisons of the recognition loop and specificity loop of these phages show that they are vastly different to that of φKMV, with the recognition loop of KP34 and Fri1 possessing a small resemblance to that of these phages (Table 4.4). Furthermore, sequence variation is evident between the PhiM1-like phages and the closely related *P. carotovorum* subsp. *carotovorum* phages PP16 and PPSW1 and *Dickeya* phage BF25/12.

Table 4.4. Alignment of the recognition and specificity loops of the RNAP of PhiM1-like phages to phages φKMV and KP34. Underlined amino acids/residues show sites of substitutions in comparison to φM1.

Phage	Recognition loop	Specificity loop
φKMV	HQEAKAAGPAAKL	EEVRVRLRAEAVEYVTLYEAK-DE
KP34	MRNVKAPGIGGKY	EEVRVRIDCMNLSAVLVHNRDFKT
Fri1	VKKQKIRGVGGKY	VTKTVAIRSMGINNIAYRYPD-NQ
φM1	ICSKGTRGVGGKY	SITRVSLKALGVALNMRVFDD-HS
CB5	ICSKGTRGVGGKY	SITRVSLKALGVALNMRVFDD-HS
Peat1	ICSKGTRGVGGKY	SITRVSLKALGVALNMRVFDD-HS
PP90	ICSKGTRGVGGKY	SITRVSLKALGVALNMRVFDD-HS
BF25/12	MCSTGSRGLGGKY	DSTRINLNALGTQLVMRTFND-HL
PP16	ICTTGNRGLNGKY	DSTRIELRSLGIKLVMRTFDD-TQ
PPWS1	ICTTGNRGLNGKY	DSTRIELRSLGIKLVMRTFDD-TQ

4.2.6. Early gene region

Early region ORFs are those that are expected to be transcribed immediately upon internalisation of phage DNA within the host. *In silico* analysis of ORFs for hypothetical proteins within this region is typically limited (in the context of determining their functions) for the majority of phages due to their vast diversity. However, it can be expected that these ORFs are involved in the redirection of host proteins to a role in

the phage infection cycle through stimulation or inhibition of protein-protein interactions (Roucourt and Lavigne, 2009). Here, we define the early genes among the PhiM1-like phages as those positioned before the DNA primase ORF (Figure 4.3). Six ORFs were found to be conserved within this region using Coregenes (Appendix, Table S4.3). Of these six ORFs, only one could be assigned the putative function of a peptidase (IPR007484).

4.2.7. DNA replication, repair and related metabolism

PhiM1-like phages encode ORFs for proteins involved in the replication and repair of DNA (Appendix, Table S4.4), including a primase, helicase, DNA polymerase, endonuclease VII and a putative 5' exonuclease. The order of these ORFs is conserved among the genomes of these phages with variation existing in the context of the presence of ORFs for hypothetical proteins and homing endonucleases among the primase, helicase, DNA polymerase and endonuclease VII. Additional variations identified were that CB5 possesses an ORF encoding a putative nucleatidyl transferase, which is absent among other PhiM1-like phages; and that CB5 and Peat1 lack an ORF encoding a putative polynucleotide 5' kinase / 3'phosphatase that is shared between PP90 and φM1.

4.2.8. Structure related genes

Discussion of these genes also includes those specifying large and small terminase (maturase) proteins which play a role in the packaging of DNA into the phage capsid. A total of 12 ORFs predicted to be involved in virion morphogenesis are shared among the PhiM1-like phages (Appendix, Table S4.5). These include a head-tail connector protein, a virion scaffolding protein, major capsid protein, tail tubular proteins A and B, internal virion proteins A, B and C, a tail fibre and a tail spike. The order in which the ORFs for these proteins occur in the genomes of these phages is highly conserved. Minor differences within this syntenic region were due to the presence of homing endonucleases among the ORFs or in some cases splitting ORFs, for example, the head-tail connector protein of PP90 (PP90_39, 40). Additionally, it was noted that some structural proteins were encoded by split ORFs without the presence of homing endonucleases, as seen for the major capsid protein (AX177_gp38, 39), tubular protein A (AX177_gp41, 42) and internal virion protein C of Peat1 (AX177_gp46, 47).

The predicted tail spike protein of these phages possesses the P22 tail spike domain (IPR015331). The P22 tail spike is characterised by the right-handed beta-helix architecture first observed for pectate lyase (Jenkins et al., 1998). This protein is capable of breaking down saccharides upon binding to host cell surface (Andres et al., 2010), which may be the case also for the PhiM1-like tail spike. The internal virion protein B of these phages may also possess enzymatic activity, with HHpred analysis indicating homology to phage proteins with lysozyme activity (best hit against PhiM1_42; *Escherichia* phage P1 endolysin Lyz, PDB accession no. 1XJU_A). This suggests that this protein may play a role in the breakdown of cell wall peptidoglycan during injection of phage genomic DNA into its host cell, like Gp16 of phage T7 (Moak and Molineux, 2000).

4.2.9. Lysis cassette of PhiM1-like phages that of *KP34virus*

The PhiM1-like phages possess three proteins predicted to cause host lysis. These are arranged in a conserved order: a spanin, a holin and an endolysin (Appendix, Table S4.6). The endolysins of these phages are predicted to possess an N-terminal transmembrane domain with lysozyme domain (IPR023347), indicating a likely function as a signal-arrest-release (SAR) endolysin, similar to that described for *Pseudomonas* phage φKMV (Briers et al., 2011). SAR endolysins use the host *sec* translocon system to enable their transport to the cell's inner membrane. The most likely holin of these phages is the pin-hole holin variety. These can provide narrow channels for ion movement causing membrane depolarization and activation of the SAR endolysin resulting in the degradation of cell wall peptidoglycan (Briers et al., 2011; Pang et al., 2009). Spanins are proteins responsible for the destruction of the outer membrane of Gram-negative hosts allowing phage progeny release. The predicted spanin of these phages is comprised of a single protein with an N-terminal outer-membrane lipoprotein signal and a C-terminal transmembrane domain, classifying them of the u-spanin variety (Young, 2014). The lysis cassette configuration of the PhiM1-like phages resembles that of *Klebsiella* phage KP34, sharing the same gene variations to the type phage φKMV of *Phikmvvirus*. The spanin of φKMV is composed of a two protein component system, an i-spanin integral cytoplasmic membrane protein (Rz) and an o-spanin outer membrane lipoprotein (Rz1) (Berry et

al., 2013; Young, 2014). Variation also exists with the order of occurrence of the genes for these proteins (holin, SAR endolysin, Rz and Rz1) in the lysis cassette of φKMV in comparison to the PhiM1-like phages and phage KP34 (Figure 4.8).

(A)

Phage	Spanin	Holin	Endolsyin
KP34	40,338 - 40,735	40,735 - 40,986	40,970 - 41,578
CB5	41,208 - 41,597	41,587 - 41,856	41,840 - 42,445
Peat1	40,069 - 40,458	40,448 - 40,717	40,701 - 41,717
φM1	40,481 - 40,867	40,860 - 41,126	41,110 - 41,715
PP90	41,149 - 41,538	41, 528 - 41,979	41,781 - 42,386

(B)

Phage	Holin	Endolysin	Rz	Rz1
φKMV	40,351 - 40,548	40,545 - 41,027	40985 - 41,314	41,202 - 41, 401
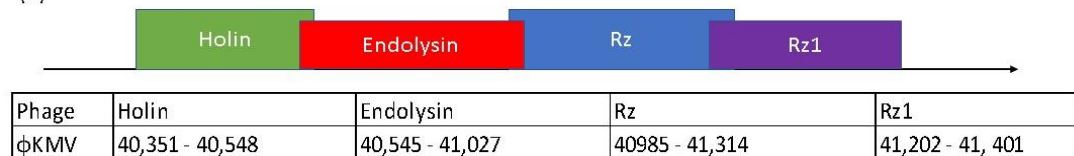
Figure 4.8. Lysis cassette scheme of (A) *Klebsiella* phage KP34 and PhiM1-like phages compared to (B) *Pseudomonas* phage φKMV.

4.3. Discussion

In silico analysis shows that the *Pectobacterium* phage CB5 belongs to a distinct group of phages (φM1, Peat1 and PP90) that infect *P. atrosepticum* and that can be classified as members of the subfamily *Autographivirinae*. In this article, we propose the creation of the genus 'Phimunavirus' to encompass these phages, the genus name is derived from the allocated type phage φM1, the first representative of these phages to be described. This proposal is supported due to these phages sharing a high DNA pairwise identity of ≥80% (BLASTN), with a highly conserved gene order (Figure 4.3) and a shared protein content of ≥60% (Coregenes). They also form a distinct clade when compared to other phages of *Autographivirinae* on phylograms based on their major capsid protein sequence (Figure 4. 5) and whole-genome comparison based on amino acid sequences employing VICTOR (Figure 4.6). Furthermore, Gegenees analysis (TBLASTX) based on amino acid sequence also indicates the existence of this clade (Figure 4.7). Additionally, the recognition and specificity loop of the RNAP between

these phages is highly conserved (Table 4.4). These phages share a number of characteristics with members of the *KP34virus* genus. Such as a similar arrangement of genes (Figure 4.4) like that of the lysis cassette, arranged sequentially in the order of a u-spanin, holin and SAR endolysin (Figure 4.8). However, DNA pairwise identity of these phages with the type phage, *Klebsiella* phage KP34, is low at 7-9%. In addition, Coregenes analysis showed that the PhiM1-like phages also possess a number of conserved proteins not shared with KP34 (39 vs 29 proteins, respectively) (Table 4.3). Furthermore, the amino acid sequence of the recognition and specificity loop of the RNAP of KP34 differs with that of PhiM1-like phages (Table 4.4). Interestingly, phylograms of the conserved proteins (Figure 5 and Appendix, Figure S4.3) and whole-genome comparison based on amino acid sequence (Figure 4.6) show a more closely placed evolutionary relationship of *Vibrio* phage V93 and *Pantoea* phage LIMElight with the *KP34virus* genus than with the PhiM1-like phages. Features that exclude these phages from being incorporated into *KP34virus* genus were that phage V93 possesses a distinct lysis cassette and encodes two tail proteins, while phage LIMElight encodes a tail fibre protein with greater similarity to phages of *SP6virus* than to that of KP34 (Eriksson et al., 2015).

It is clear that *P. carotovorum* subsp. *carotovorum* phages PP16 and PPSW1 and *Dickeya* phage BF25/12 share a close evolutionary relationship with PhiM1-like phages, sharing DNA pairwise identity between 47% and 55%. They also form a clade that sits on the same branch with that of PhiM1-like phages on phylograms comparing the major capsid protein and whole-genome comparison based on amino acid sequences of phages of *Autographivirinae* (Figure 4.5, 4.6). Furthermore, Gegenees analysis (TBLASTX) shows an identity of between 54% and 60% (Figure 4.7). However, the relationship is more distant than that between phages of the suggested '*Phimunavirus*' genus. Thus casting doubt if they should be placed in the genus, therefore we have chosen to exclude phages PP16, PPSW1 and BF25/12 from the genus at this point in time until the availability of more data on related phages.

Not all *Pectobacterium* phages reported to date that have been classified as belonging to *Autographivirinae* resemble phage φM1. For example, *Pectobacterium carotovorum* subsp. *carotovorum* phage PP2 possesses homology to *Cronobacter* phage

vB_CsaP_Gap277, a phage that has been identified to represent a distinct genus within *Autographivirinae* (Abbasifar et al., 2013; Lim et al., 2017).

Finally, It has been commented that the split of the *Phikmvvirus* genus into genera better reflecting evolutionary relationships is to date incomplete (Nowicki et al., 2017). This concern was highlighted in phylograms constructed in this study with the observation of phage LIMElight, which is currently classified as a member of the *Phikmvvirus*, is in fact placed between the genera *Fri1virus* and *KP34virus* (Figure 4.5, 4.6)*.* Since the creation of the *Autographivirinae* subfamily, many more phages related to it have been sequenced. It is clear from the phylograms constructed in this study that a taxonomic reassessment of these phages is required to adequately reflect their genomic diversity.

Chapter 5. Isolation and characterisation of *Pectobacterium* phage vB_PaM_CB7; further insights into the genus *Cr3virus*

Genome sequence of phage CB7 was outsourced to the University of Liverpool, UK.

Transmission electron microscopy was conducted by Horst Neve.

Mass Spectrometry analysis of phage proteins was conducted by Hanne Hendrix.

5.1. Introduction

Soft Rot *Enterobacteriaceae* (SRE) is a group of economically important phytopathogenic bacteria that consist of the genera of *Pectobacterium* and *Dickeya*, both typified by the production of extracellular pectinolytic enzymes upon plant infection (Pérombelon, 2002; Toth et al., 2011, 2003). Of these, *Pectobacterium atrosepticum* has traditionally been the most dominant SRE affecting the potato crop in temperate climates, causing potato blackleg and soft rot disease (Pérombelon, 2002; Toth et al., 2003), infections for which there are no effective bactericides. Control strategies used to date are mainly culturing approaches such as the removal of diseased tissues and/or plants and the implementation of seed certification schemes (Czajkowski et al., 2011).

Examination of the phages of several well studied SRE members have shown a good potential of phages to be exploited for biocontrol (Adriaenssens et al., 2012b; Czajkowski et al., 2015; Smolarska et al., 2018). The phages of *P. atrosepticum* have also received this attention (Buttimer et al., 2018a). Currently, it is known that there are at least three kinds of phages from the order *Caudovirales* that infect this bacterium (Blower et al., 2012; Buttimer et al., 2018a, 2018b). Of these groups, one has been recently formally established as a genus by the ICTV, namely *Cr3virus* (Krupovic et al., 2016). This genus and the genera of *V5virus* and *Se1virus* form the subfamily *Vequintavirinae,* which falls within the *Myoviridae* family (Krupovic et al., 2016). Before the establishment of the *Cr3virus* genus, members of this group had often been described as being rV5-like, referring to the closely related *Escherichia* phage rV5, the type phage of *V5virus* (Blower et al., 2012). The type phage of *Cr3virus* is *Cronobacter* phage CR3 (accession no. JQ691612), the first representative of the *Cr3virus* to have its genome described (Shin et al., 2012), with information on *Pectobacterium* phage ΦTE (accession no. JQ015307) and *Cronobacter* phage PBES 02 (accession no. KT353109) becoming available later (Blower et al., 2012; Lee et al., 2016). These three phages and *Cronobacter* phage CR8 (accession no. KC954774) and CR9 (accession no. JQ691611), for which genome sequences are available on public databases (Genbank), officially form the *Cr3virus* genus. Furthermore, genomes of *Pectobacterium* phages DU_PP_I (accession no. MF979560) and DU_PP_IV (accession

no. MF979563) have come available on public databases within the last year and both of these share high pairwise nucleotide identity with phage CR3 (>70%), with a similar gene arrangement and so they can, therefore, be described as additional members within this genus.

This study reports the analysis of the newly-isolated *Pectobacterium* phage CB7. Phylogenetic analysis of its genome shows that it can be placed within the genus *Cr3virus*. Analysis of this phage provides greater insights into phages of *Cr3virus*, regarding transcription, HNH endonuclease gene splicing, DNA related metabolism and virion structure.

5.2. Results

5.2.1. Isolation of CB7, host range and general characteristics

Phage CB7 was isolated from a soil sample collected from potato grading machinery on a farm in Co. Cork, Ireland, during the year 2013. The phage was differentiated from other phage isolates based upon band patterns observed after restriction digestion analysis of its genomic DNA using BgII and SspI (Figure 5.1). Its host range was examined using 31 bacterial strains from five different SRE species, namely *P. atrosepticum* (19 strains), *P. carotovorum* subsp. *carotovorum* (four strains), *D. chrysanthemi* bv. *chrysanthemi* (one strain), *D. dianthicola* (three strains), and *D. solani* (four strains). The phage was found to possess a narrow host only exhibiting activity against its host strain *P. atrosepticum* (DSM 30186) and four other strains of the same species (Table 5.1). It produced clear plaques with an approximate diameter of 1-2 mm (0.4% LB agar overlay) on the host strain (Appendix, Figure S5.1).
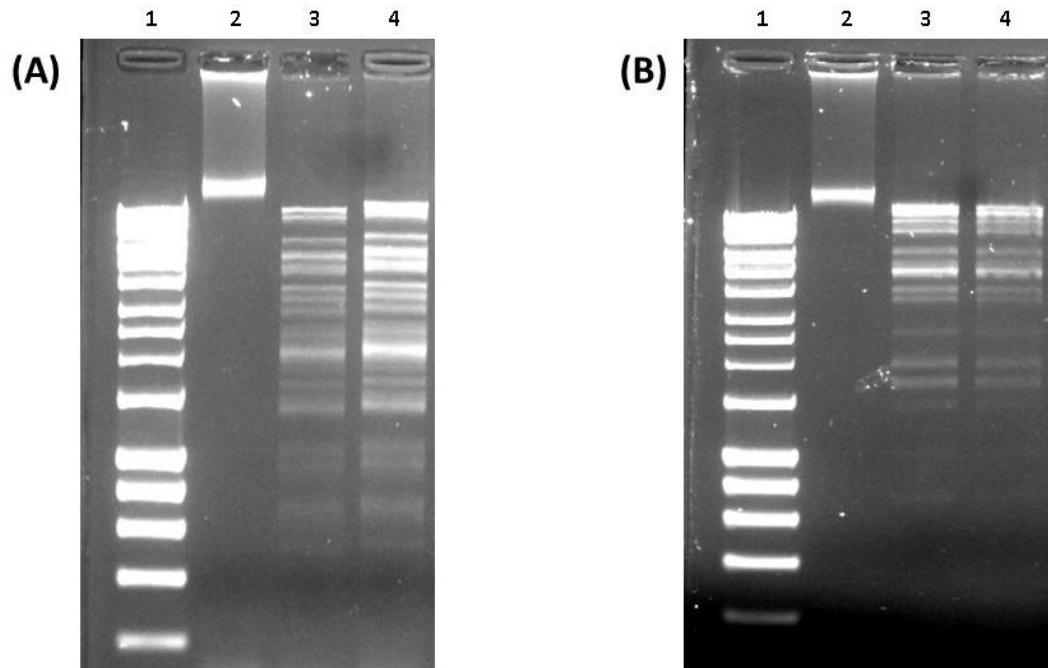
Figure 5.1. Genomic DNA of *Pectobacterium* phage CB7 digested with restriction enzyme BglII (A) and SspI (B). Lane 1; DNA marker (Hyperladder 1kb, Bioline), Lane 2; undigested CB7 genomic DNA, Lanes 3 and 4; digestion of CB7 genomic DNA with BglII and SspI respectively. The gel concentration was 0.8 % (w/v).

Table 5.1. Host range of *Pectobacterium* phage vB_PatM_CB7 (CB7) on 31 strains of soft rot *Enterobacteriaceae* as determined by spot testing with a serial dilution of phage.

Species	Strain	Sensitivity
Pectobacterium atrosepticum	DSMZ 18077 (type strain)	-
	DSMZ 30184	-
	DSMZ 30185	+
	DSMZ 30186	+*
	CB BL1-1	-
	CB BL2-1	+
	CB BL3-1	+
	CB BL4-1	+
	CB BL5-1	-
	CB BL7-1	-
	CB BL9-1	-
	CB BL11-1	-
	CB BL12-2	-
	CB BL13-1	-
	CB BL14-1	-
	CB BL15-1	-
	CB BL16-1	-
	CB BL18-1	-
	CB BL19-1	-
Pectobacterium carotovorum subsp. *carotovorum*	DSMZ 30168 (type strain)	-
	DSMZ 30169	-
	DSMZ 30170	-
	CB BL19-1-37	-
Dickeya chrysanthemi bv. *chrysanthemi*	LMG 2804	-
Dickeya dianthicola	PD 482	-
	PD 2174	-
	GBBC 1538	-
Dickeya solani	sp. PRI 2222 (D36)	-
	LMG 25865 (D10)	-
	GBBC 1502	-
	GBBC 1586	-
Erwinia amylovora	LMG 2024	-
	GBBC 403	-
Erwinia mallotivora	LMG 1271	-

Results recorded as +, sensitive; –, no infection; * host strain of phage

The one-step growth curve assay, using standard growing conditions with LB medium, showed that CB7 possesses a latent period of 55 min with an approximate burst size of 154 PFU/cell (Figure 5.2). This differs to the CR3-like phage *Cronobacter* phage PBES 02 which has a latent period of 30 min and a burst size of 250 PFU/cell, the only

representative for which such data is available (Lee et al., 2016). The stability of the phage was also examined under different conditions of temperature and pH. Over one hour, the phage was found to be stable between the temperatures of -18 ºC and 55 ºC (Figure 5.3), and over a period of 24 h was stable between a pH of 5 and 11 (Figure 5.4). Temperature stability of CB7 is similar to that of phage PBES 02, but its pH stability range appears to be slightly wider (Lee et al., 2016).
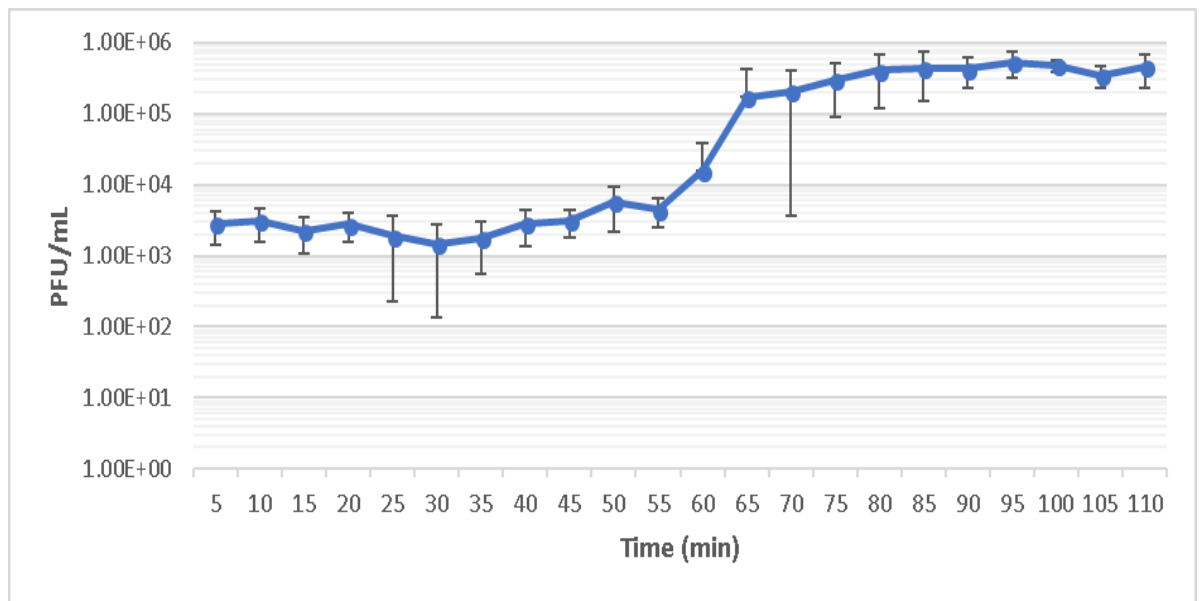
Figure 5.2. Single step growth curve analysis of *Pectobacterium* phage CB7 on *P. atrosepticum* strain DSM 30186. The assay was performed in triplicate and the results were averaged.
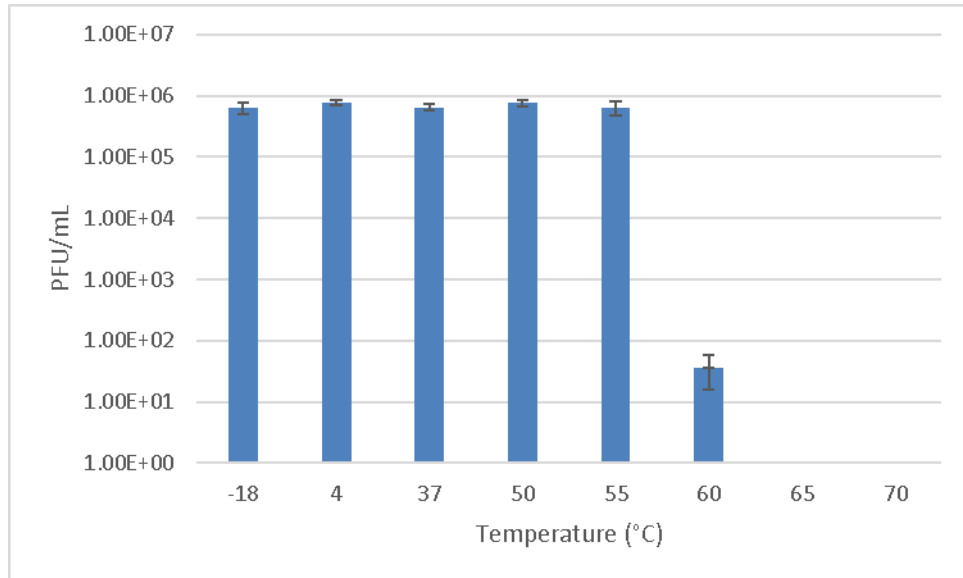
Figure 5.3. Stability of *Pectobacterium* phage CB7 to various temperatures upon one-hour exposures. Error bars represent the standard deviation of biological repeats (*n*=3).
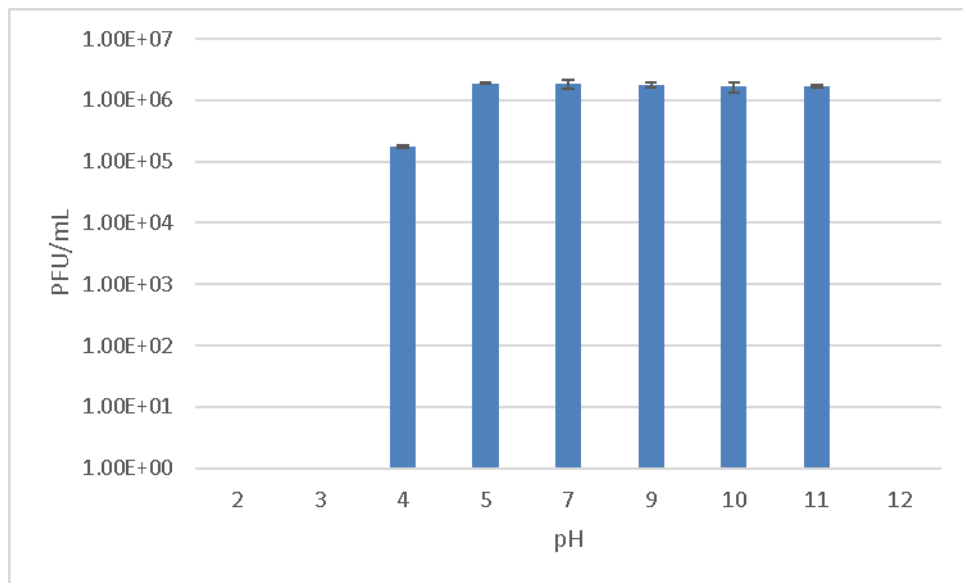
Figure 5.4. Stability of *Pectobacterium* phage CB7 to various pH values upon 24 hours of exposure. Error bars represent the standard deviation of biological repeats (*n*=3).

Transmission electron microscopy showed that CB7 possesses an A1 morphotype (Figure 5.5), allowing it to be classified as a member of the family of *Myoviridae* (Ackermann, 2001). The phage possesses an icosahedral head (84.05 ± 3.88 nm in diameter, *n*=11), with clearly distinguishable hexagonal outlines, and a contractile tail (123.39 ± 2.63 nm x 20.18 ± 1.05 nm, *n*=8) possessing tail fibres (length: 42.40 ± 3.68

nm, *n*=6). The phage was named in accordance with the nomenclature set out by Kropinski *et al.* (2009).
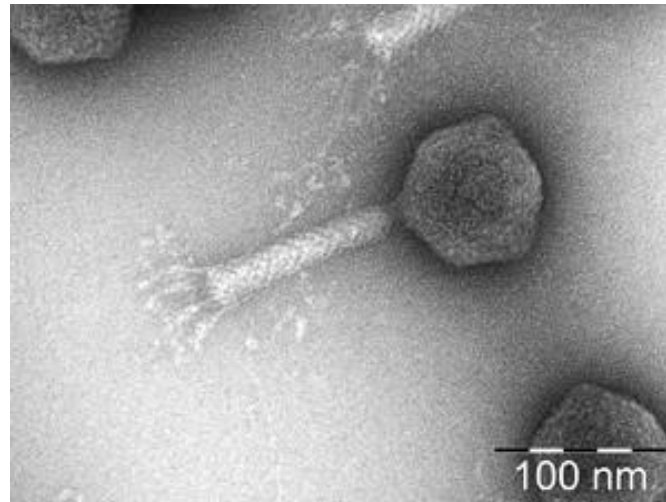
Figure 5.5. Transmission electron micrographs of negatively stained *Pectobacterium* phage CB7 with 1% (w/v) uranyl acetate. Scale bar represents 100 nm.

5.2.2. General genome characteristics of CB7

Genome sequencing of CB7 revealed a size of 142,778 bp (coverage 2554x) and examination of sequence reads indicated that its genome is likely circular permuted like that of Coliphage T4 (Casjens and Gilcrease, 2009) due to there being no region possessing significantly higher read depths than that of the average read depth across the entire genome. BAL-31 exonuclease time course treatment of the genome followed by restriction digestion with BglII agrees with the above finding with no bands found to be preferentially degraded following examination by agarose gel electrophoresis (Figure 5.6). A similar result was obtained with the closely related *Pectobacterium* phage ΦTE (Blower et al., 2012). Therefore, since the genome is believed to be circularly permuted the start position of its Genbank file was matched to that of *Cronobacter* phage CR3. Additionally, the G+C content of its genome was found to be 50.1%, a value similar to what would be typically expected for its host species, *P. atrosepticum* (50-51%) (Bell et al., 2004; Nikolaichik et al., 2014).
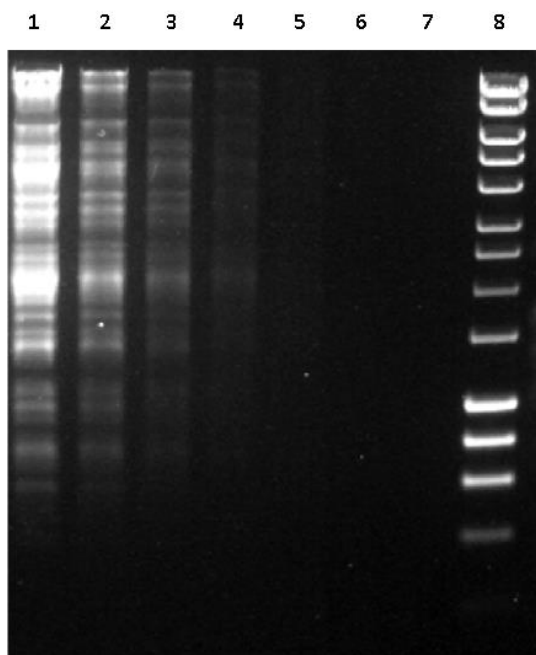
Figure 5.6. Treatment of *Pectobacterium* phage CB7 genomic DNA with BAL-31 for 60 min (Lane 1), 80 min (Lane 2), 100 min (lane 3), 120 min (Lane 4), 140 min (Lane 4), 160 min (Lane 5), 180 min (Lane 6) and (Lane 7) 200 min followed with restriction digestion with BglII. Molecular weight marker (Hyperladder 1kb, Bioline) (lane 8). Gel concentration 1% (w/v) agarose.

The genome of CB7 was determined to contain 253 ORFs (Figure 5.7 and Appendix, Table S5.1). Based upon the orientation of these ORFs the genome can be divided into four regions, namely region 1 (CB7_241 - 37), region 2 (CB7_38 - 103), region 3 (CB7_104 - 143) and region 4 (CB7_144 - 240). ORF orientation correlates well with GC skew (Marín and Xia, 2008). Using *in silico* analysis based on protein sequence homology, protein structure homology, lipoprotein and transmembrane analysis, and as well as ESI-MS/MS spectrometry, it was possible to identify a role for 81 (32%) of these putative proteins. The remaining proteins could be categorised as either hypothetical proteins (9), conserved hypothetical proteins (143), putative lipoproteins (3), conserved (4) and hypothetical (13) transmembrane proteins. Additionally, one tRNA gene for tyrosine was identified. No ORFs were identified to encode integrase, excisionase nor repressor proteins, indicating that the phage likely follows an exclusively lytic lifestyle.

Figure 5.7. Summary of the genomic organisation of the 142,778 bp genome of *Pectobacterium* phage CB7. On the outer ring, putative ORFs are represented by purple arrowheads labelled with predicted function where possible. On the middle ring is shown GC content relative to the mean GC content of the genome. On the inner ring, GC shew is illustrated where green represents a positive skew and purple a negative skew. Created using GCview.

111

5.2.3. Phylogenetic analysis; CB7 is a member of *Cr3virus* and the status of *Vequintavirinae*

Examination of the genome of CB7 with the BLASTN algorithm showed that the phage has significant homology with phages of the genus *Cr3virus* (≥60% identity), sharing a similar arrangement of genes (Table 5.2, Figure 5.8). Coregenes analysis also showed a large number of shared proteins (Table 5.2), confirming CB7 to be a member of the genus *Cr3virus*. Furthermore, construction of phylograms (with confident bootstrap values) using the large terminase, major head and DNA polymerase proteins of phages within the subfamily of *Vequintavirinae* placed CB7 within a clade that represents the genus of *Cr3virus* (Figure 5.9). This was further supported by the creation of a GBDP phylogram using VICTOR (Figure 5.10) and Gegenees analysis (TBLASTX) where CB7 was shown to possess 55% to 86% identity within a clade representing *Cr3virus* (Figure 5.11).

Table 5.2. Properties of the five phages of *Cr3virus* and *Pectobacterium* phages DU_PP_I,

Phage	Accession no.	Genome size (bp)	G+C content, %	ORFs	tRNA	Identity (%)*	Shared proteins (%)**
Cronobacter phage CR3	JQ691612	149,273	50.9	265	18	100	100
Cronobacter phage CR8	KC954774	149,162	50.8	269	17	94	80
Cronobacter phage CR9	JQ691611	151,924	50.6	281	17	70	89
Cronobacter phage PBES 02	KT353109	149,732	50.7	270	14	64	90
Pectobacterium phage DU_PP_I	MF979560	144,959	50.1	267	8	73	80
Pectobacterium phage DU_PP_IV	MF979563	145,233	50.3	268	8	73	80
Pectobacterium phage vB_PatM_CB7	KY514263	142,778	50.1	253	1	64	70
Pectobacterium phage ФTE	JQ015307	142,349	50.1	242	2	62	63

DU_PP_V and CB7 and comparison of all seven with the type phage CR3

*DNA identity in comparison to CR3 using BLASTN., **Number of homologous proteins in comparison to CR3 using Coregenes.
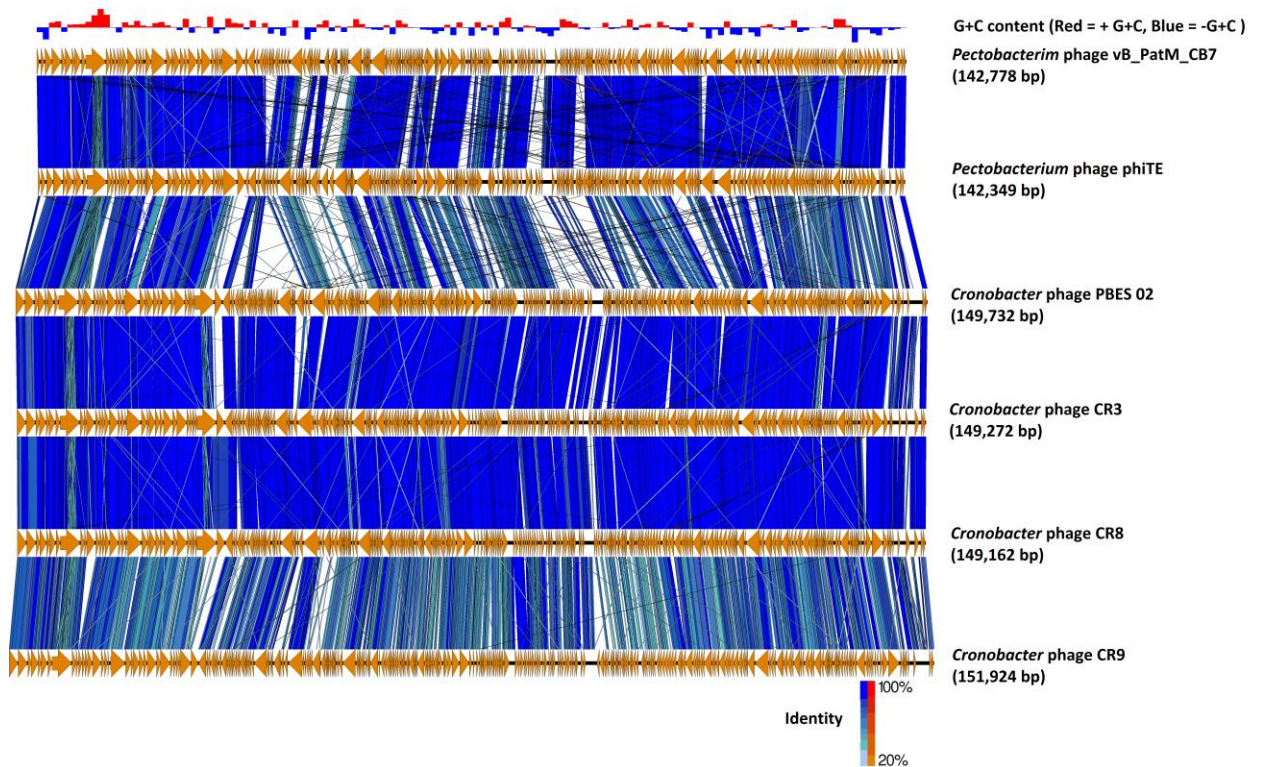
Figure 5.8. Comparison of the genome of *Pectobacterium* phage CB7 to other CR3-like phages (*Pectobacterium* phage ΦTE and *Cronobacter* phages PBES 02, CR3, CR8 and CR9) using currently available annotations employing TBLASTX and visualised with Easyfig. A bar chart shows the G+C skew of the CB7 genome, genome maps comprise of orange arrows indicating locations of genes among the different phage genomes; and lines between genome maps indicate the level of identity (blue/turquoise; genes sharing orientation, red/orange; genes with inverted orientation). The large terminase was set as the first gene among all genomes.
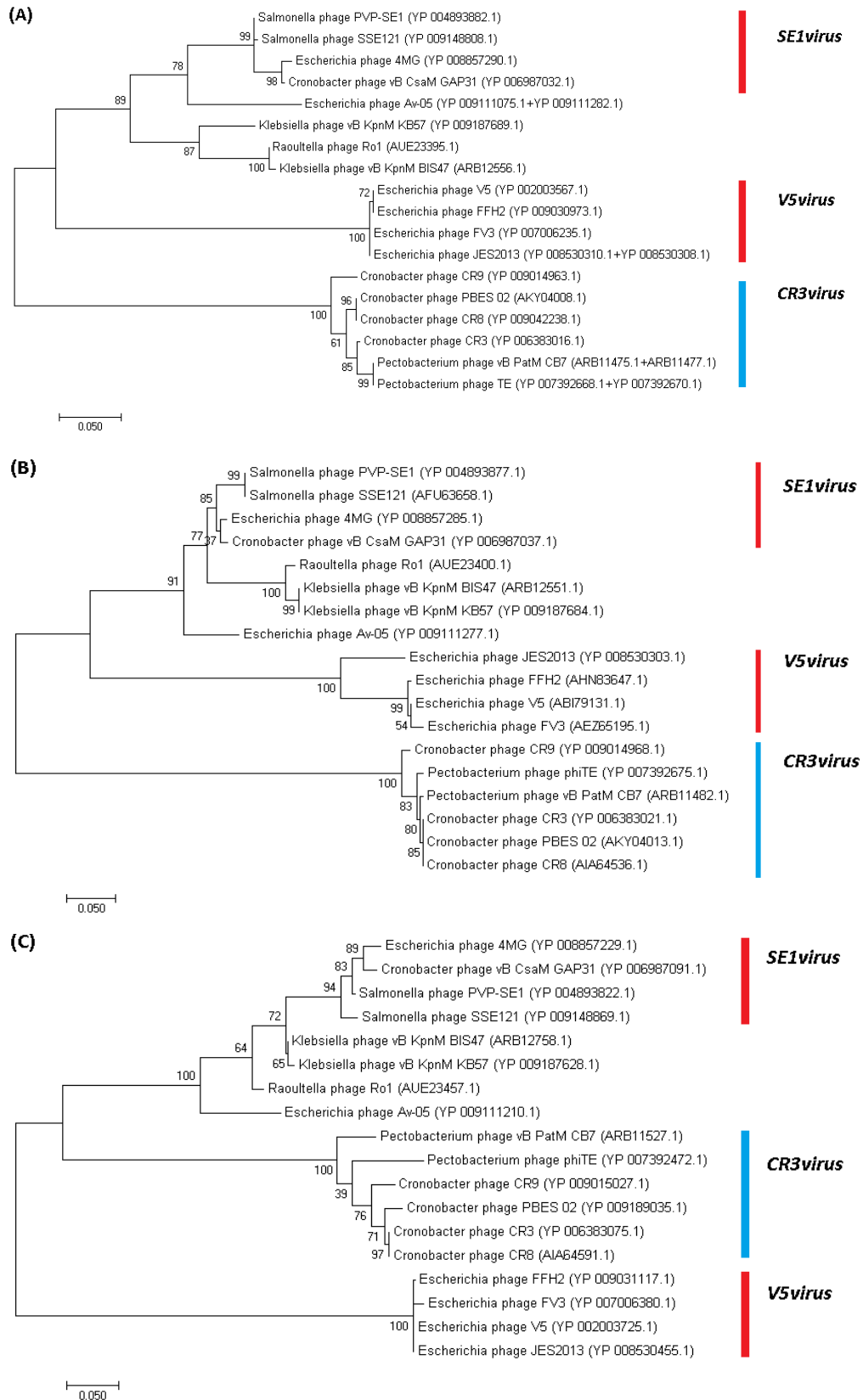
Figure 5.9. Phylograms constructed using the large terminase (A), major head (B) and DNA polymerase (C) of *Pectobacterium* phage CB7 and homologs of 17 phages of the subfamily of *Vequintavirinae* analysed using maximum likelihood (Whelan and Goldman substitution model), with 1000 bootstrap replicates using MEGA7. Members of the genera of *Se1virus*, *V5virus* and *Cr3virus* are illustrated.

114
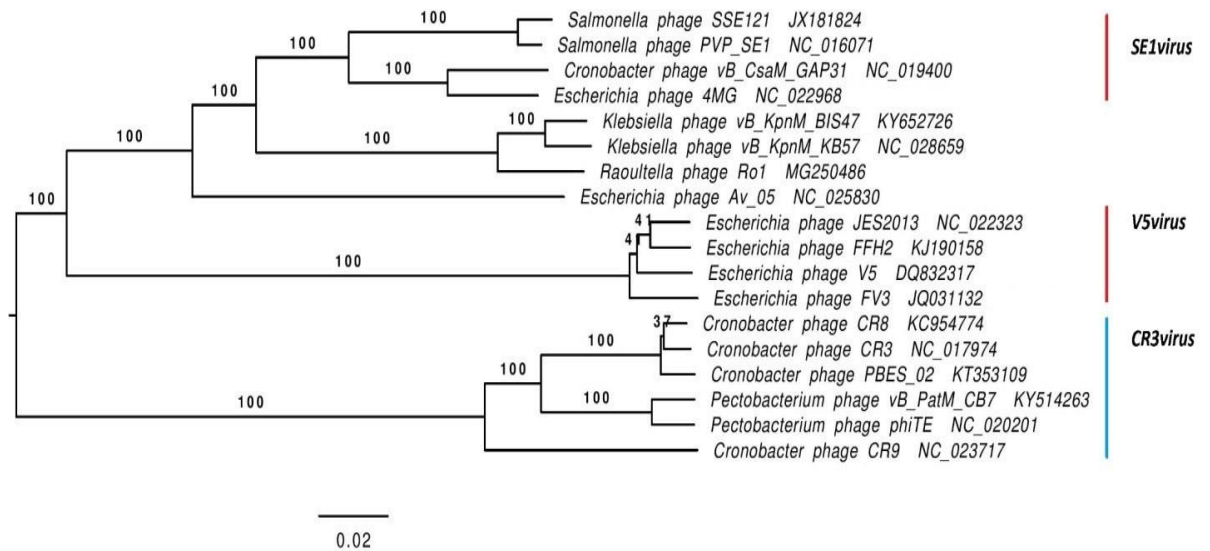
Figure 5.10. VICTOR generated phylogenomic GBDP tree of *Pectobacterium* phage CB7 and 17 members of *Vequintavirinae* subfamily inferred using the formula D4 and yielding average support of 87%. The numbers above branches are GBDP pseudo-bootstrap support values from 100 replications. Members of the genera of *Se1virus*, *V5virus* and *Cr3virus* are illustrated.

Phage	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1: Cronobacter phage CR3	100	92	88	66	60	59	22	24	24	24	25	25	24	24	20	20	20	20
2: Cronobacter phage CR8	92	100	90	65	60	59	22	24	24	24	24	25	24	24	20	20	20	20
3: Cronobacter phage PBES 02	88	89	100	65	60	59	22	24	24	24	25	25	25	25	20	20	20	20
4: Cronobacter phage CR9	65	65	65	100	55	53	21	24	24	24	24	25	24	25	20	20	20	20
5: Pectobacterium phage vB_PatM_CB7	62	62	62	57	100	86	22	24	23	23	24	24	26	26	21	21	21	21
6: Pectobacterium phage phiTE	61	61	61	56	86	100	22	24	23	23	24	24	25	26	21	21	21	21
7: Escherichia phage Av-05	24	24	24	24	24	23	100	37	36	36	40	39	41	41	26	26	25	26
8: Raoultella phage Ro1	25	24	25	25	24	23	33	100	75	73	49	49	48	48	26	26	26	26
9: Klebsiella phage vB_KpnM_BIS47	25	24	25	25	23	23	32	74	100	82	50	50	48	48	26	26	26	26
10: Klebsiella phage vB_KpnM_KB57	25	25	25	25	23	23	33	74	84	100	52	50	49	48	27	27	27	27
11: Enterobacteria phage 4MG	25	25	25	25	24	24	35	49	50	50	100	79	63	62	27	27	27	27
12: Cronobacter phage vB_CsaM_GAP31	25	25	25	25	24	24	34	48	50	49	79	100	63	62	27	27	27	27
13: Salmonella phage PVP-SE1	25	25	25	25	25	25	36	48	48	48	64	63	100	94	28	28	28	28
14: Salmonella phage SSE-121	24	24	25	25	25	25	36	47	47	47	62	62	93	100	28	28	28	28
15: Enterobacteria phage vB_EcoM-FV3	21	21	21	21	21	21	24	27	28	28	28	28	29	29	100	86	86	88
16: Escherichia coli bacteriophage rv5	21	21	21	21	21	21	24	27	27	27	28	28	29	29	86	100	90	90
17: Escherichia phage 2 JES-2013	21	21	21	21	21	21	24	27	27	27	28	28	29	29	86	91	100	91
18: Escherichia phage vB_EcoM_FFH2	21	21	21	21	21	21	24	27	27	27	28	28	29	29	87	89	90	100

Figure 5.11. TBLASTX heat map generated using Gegenees with accurate parameters - fragment length: 200 bp; and step size: 100 bp; threshold: 0%. The map includes the genomes of *Pectobacterium* phage CB7 and 17 members of *Vequintavirinae* subfamily. Members of the genera of the *Se1virus* (brown), *Cr3virus* (blue), and *V5virus* (pink) are illustrated.

Analysis performed in this work showed that there are currently eight phage genomes in the public databases that can be classified as *Cr3virus.* New additions to the genus other than CB7, are *Pectobacterium* phages DU_PP_I and DU_PP_IV (Table 5.2). Generally, phages within the *Cr3virus* genus possess genome sizes that range between

142,349 bp (φTE) and 151,924 bp (CR9) and possess an average G+C content of 50±0.34%. Total ORF number range from 242 (φTE) to 281 (CR9), with Coregenes showing that they share a minimum of 166 proteins in comparison to phage CR3. However, there is a large variation in tRNA gene content among these phages. Those infecting *Cronobacter* possess between fourteen (PBES 02) and eighteen (CR3) tRNA genes, while those infecting *Pectobacterium* is much lower: one (CB7) to eight (DU_PP_I and DU_PP_IV) tRNA genes.

5.2.4. Transcription, promoters and terminators

Phage CB7 does not contain an RNA polymerase suggesting that it is totally dependent on host-encoded RNA polymerase for transcription. However, no proteins that could potentially play a role in the takeover of host RNA polymerase causing its redirection to host promoters were identified, as seen in *Escherichia* phage T4 (Hinton, 2010).

A putative promoter with the consensus of AAAA(N3)TGTTGAC(N19)TATAAT was identified at 13 sites on the genome of CB7 (Appendix, Table S5.2). The sequence of this putative promoter resembles the consensus sequence of the sigma70 promoters of *E. coli*, possessing -10, -35 and the UP elements (Hinton, 2010). The locations of this promoter are heavily concentrated within a region containing a large number of ORFs encoding short hypothetical proteins (likely early gene region), with some located within a region possessing genes for a DNA helicase, DNA methylase and DNA polymerase (Figure 5.12). Based on the gene products downstream of these promoter sites, it would suggest the promoter may play a role in early and possibly middle phase transcription. Moreover, analysis of genomes of *Cronobacter* phages CR3, CR8 and CR9 showed that this putative promoter is also present among their genomes, with 21, 21 and 13 sites being identified respectively (Appendix, Table S5.3, S5.4 and S5.5, Figure S5.2, S5.3 and S5.4). Like CB7, the promoter is concentrated within a localised region containing a large number of ORFs encoding short hypothetical proteins (likely early gene region), as well as also being located near a small number of ORFs with gene products involved in DNA replication and nucleotide metabolism (likely middle gene region). Identification of this promoter among these additional phages suggests it may represent part of a conserved transcription strategy being employed among members

of *Cr3virus*. Furthermore, 27 putative rho-independent terminators were also identified on the genome of CB7 (Appendix, Table S5.6).
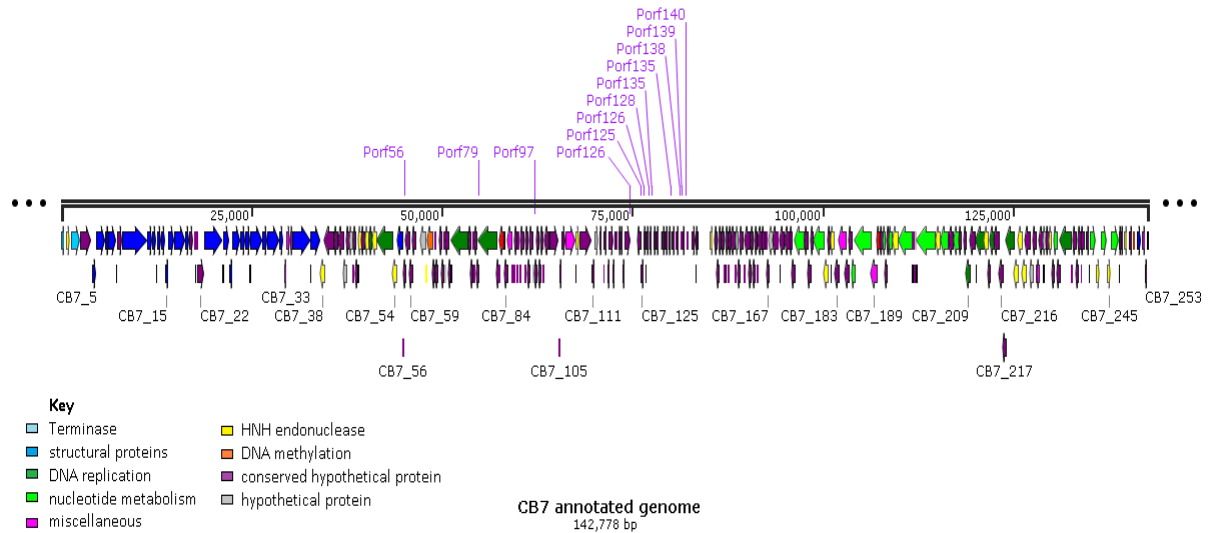
Figure 5.12. Genome map of *Pectobacterium* phage CB7 showing locations of a putative sigma70-like promoter with the consensus sequence AAAA(N4)TGTTGAC(N17)TATAAT. The map comprises of arrows indicating the location of ORFs. Arrows have been colour-coded describing their predicted role (see key). The genome map was created with Snapgene.

5.2.5. DNA replication, methylation and nucleotide metabolism

Within the genome of CB7, ORFs for proteins associated with functions related to DNA replication were mostly found within gene region 2 while those for nucleotide metabolism were mostly found within gene region 4 (Figure 5.7). The phage was identified to have several proteins that would be involved in DNA replication such as a DNA polymerase (CB7_49, 51, 53), DNA ligase (CB7_234), helicase (CB7_80) and a putative primase (CB7_73). There is also a homolog of recombination endonuclease VII of *Escherichia* phage T4 (CB7_214). In phage T4, this protein is involved in DNA mismatch repair and assists in DNA packaging by removing branched replicative DNA (Shcherbakov et al., 2011).

The phage genome also encodes proteins involved in nucleotide metabolism, such as aerobic class I ribonucleotide reductase (RNR) subunits NrdA (CB7_195, 197) and NrdB (CB7_201, 203) with the role of deoxynucleoside triphosphates (dNTP) synthesis during

aerobic conditions. In addition, it also encodes anaerobic class III ribonucleotide reductase subunits NrdG (CB7_187) and NrdD (CB7_188), which enable dNTP production during anaerobic conditions. These classes of RNRs have been identified among a number of phage genomes on Genbank (Lundin et al., 2009). Additionally, the RNR class I function of CB7 can potentially be supported by glutaredoxin (CB7_189) allowing RNR reduction through acting as an electron carrier (Feeney et al., 2012). The phage is also capable of influencing the deoxythymidine monophosphate (dTMP) pool of the host by using deoxyuridine monophosphate (dUMP) with its thymidylate synthase (CB7_206) (IPR003669). It also has a ribose-phosphate pyrophosphokinase (CB7_242) (IPR005946) which forms the precursor phosphoribosyl pyrophosphate (PRPP) which can be used in the biosynthesis of purine and pyrimidine nucleotides. PRPP can also be used by CB7 with its nicotinamide phosphoribosyltransferase (CB7_244, 246) to produce nicotinamide mononucleotide, an intermediate in the production of coenzyme nicotinamide adenine dinucleotide (NAD).

The phage is also capable of influencing the tRNA pool of the host with a tRNA nucleotidyltransferase (CB7_179) (IPR012006) which synthesises or repairs the 3' terminal CCA sequence of tRNA molecules.

The identification of two kinds of DNA methylase enzymes on phage CB7 (CB7_63, 183) indicates that its genome is likely to be adenine and cytosine methylated, given that it has a putative N-6 adenine-specific DNA methylase (DAM) (IPR002052) and DNA cytosine methylase (HHpred analysis: best hit; 5-cytosine DNA methyltransferase of *Entamoeba histolytica*, PDB accession no. 3QV2_A). Indeed, restriction digestion patterns of the genomic DNA of CB7 using ClaI does indicate that its DNA is likely to be DAM methylated (Figure 5.13).
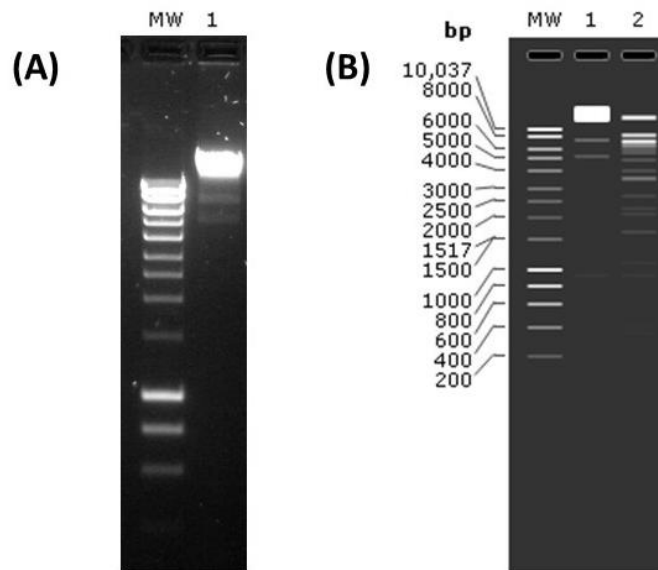
Figure 5.13. (A) Genomic DNA of *Pectobacterium* phage vB_PatM_CB7, which had been digested with restriction enzyme ClaI (Lane 1), with DNA marker (Hyperladder 1kb, Bioline) (Lane MW). (B) *In silico* digest of CB7 genomic DNA with ClaI with DAM methylation (Lane 2); without DAM methylation (Lane 3); DNA marker (Hyperladder 1kb, Bioline) (Lane MW). Gel concentration 1% (w/v) agarose. Image B was generated using Snapgene.

Comparison of CB7 to other members in *Cr3virus* genus using ACT (TBLASTX) showed that the majority of the previously discussed proteins involved in DNA replication, methylation and nucleotide metabolism are shared across the genus (Appendix, Table S5.7). The exceptions to this are a homolog of the putative DNA cytosine methylase not shared with *Pectobacterium* phages ΦTE, DU_PP_I and DU_PP_IV. Another variation identified was that NrdD and NrdG of CB7 were only shared with ΦTE among the genus. These gene product variations suggest that there may be small differences between members of *Cr3virus* regarding DNA methylation and nucleotide metabolism; unless proteins for which a function could not be defined among these phages perform a similar role. Additional comparison (using ACT with TBLASTX) between CB7 and type phages rV5 and PVP-SE1 revealed differences in gene product content. (Appendix, Table S5.8). Both type phages were found not to share a homolog of CB7 DNA ligase. Furthermore, phage rV5 was identified to not have homologs of the CB7 tRNA nucleotidyltransferase and nicotinamide phosphoribosyltransferase.

5.2.6. Selfish genetic elements within the genome of CB7

Homing endonucleases are mobile genetic elements consisting of genes that encode a protein with endonuclease activity that promotes the lateral transfer of their own

encoding gene. These endonucleases can recognise specific DNA sequences at which they initiate catalysis causing DNA strand breakage resulting in the insertion of a homing endonuclease encoding gene, due to DNA cellular mechanisms that rely on homologous recombination (the process being termed homing). Introns are segments of DNA that are removed from a mature mRNA post-transcription, while inteins are self-splicing protein elements that self-excise from a protein precursor with the concomitant post-translation ligation of C- and N-terminal segments called exteins. Homing endonucleases can be found associated with these elements but can simply exist as a free-standing gene (Chevalier and Stoddard, 2001). A significant number of homing endonucleases have been characterised among phages and shown to have recognition sites that lie within genes related to DNA replication and metabolism (Edgell et al., 2000), but they have also been identified to target genes related to virion structure (Kala et al., 2014; Pope et al., 2013).

On the genome of CB7, twenty-one homing endonucleases of the HNH family (IPR003615, IPR029471) were identified (twenty HNHs confirmed with Interproscan, and one HNH identified with HHpred). Five of these HNH homing endonucleases were associated with an intron, with the remaining identified as free-standing genes (Table 5.3). ORFs for these homing endonucleases were found to cover approximately 7 % of the genome of CB7.

A single intron with an HNH gene was found to be interrupting ORFs for the large terminase (CB7_1,3), ribonucleotide-diphosphate reductase (CB7_195, 197) and nicotinamide phosphoribosyltransferase (CB7_144, 246), with two introns with an HNH gene, found to be interrupting ORFs of the DNA polymerase (CB7_49, 51, 53), with an additional HNH gene at the 5' end of this gene product. These can be categorised as being group 1 introns as they possess homing endonucleases of the HNH family (Chevalier and Stoddard, 2001).

RT-PCR was performed to investigate splicing of the previously described introns at the mRNA level. Total RNA was extracted from cells of *P. atrosepticum* strain DSM 30186 infected with phage CB7 at different time points (15 min, 30 min and 45 min). The resulting cDNA was then investigated for splicing using PCR with primers complementary to the 5' and 3' ends of the HNH endonuclease genes being examined

and using CB7 genomic DNA as a control. The size of the resulting PCR product was then compared to that obtained from CB7 genomic DNA (Figure 5.14). Splicing was shown to occur for introns containing an HNH for gene products of the large terminase (*CB7_3*), ribonucleotide reductase NrdB (*CB7_196*), nicotinamide phosphoribosyltransferase (*CB7_202*) and DNA polymerase (*CB7_50,51*). This was supported by the appearance of PCR products derived from cDNA that were smaller than that those obtained from CB7 genomic DNA, indicating that splicing had occurred at the mRNA level causing the removal of HNH ORF sequences, thus resulting in smaller PCR products.

No splicing occurred for DNA polymerase HNH *CB7_54*, situated at the 5' end of ORF *CB7_53*. This is likely to be due to the likely free-standing nature of this HNH and not being part of an intron. However, HNH *CB7_245* is associated with an intron that interrupts the ORFs of nicotinamide phosphoribosyltransferase and was not found to be removed by splicing at the mRNA level. The existence of introns with an HNH that do not splice post-transcription (like that of the intron with HNH *CB7_245*) has been previously described. *Aeromonas* phage Aeh1 anaerobic ribonucleotide reductase subunit NrdA is split by an intron that contains an HNH endonuclease (*mobE*) and does not splice post-transcription. This intron causes NrdA to be translated as two separate peptides that associate with each other post-translation along with ribonucleotide reductase NrdB subunit while still retaining activity (Friedrich et al., 2007). This may also be the case for the nicotinamide phosphoribosyltransferase of CB7. The above HNH splicing results of CB7 were further confirmed by performing Sanger sequencing on the PCR products derived from cDNA.

Three inteins were identified among the ORFs of CB7, namely the putative helicase (CB7_80), the ribonucleotide reductase NrdB (CB_197) and NrdA (CB7_201) (Table 5.3). Furthermore, the inteins of ORFs CB7_80 and CB7_201 were found to contain the homing endonuclease of the LAGLIDADG family (IPR004860). Experiments were not conducted to determine the splicing nature of these elements within the time-frame of this study.

In total, twenty-three homing endonucleases were identified to be present on the genome of CB7. This is one of the largest quantities of homing endonucleases

identified on a phage genome examined in literature to date, greater than that described for *Escherichia* phage T4, which possesses fifteen homing endonucleases - one of the highest known among T4-like phages (Edgell et al., 2010). Most of the homing endonucleases of CB7 are shared with *Pectobacterium* phage ΦTE, and less so with the *Cronobacter* phages of the *Cr3virus* genus (Table 5.3). *Cronobacter* phage CR3 is only predicted to possess two HNH homing endonucleases. The presence of a high number of homing endonucleases appears to be a common trend among many phage types infecting bacteria of SRE group (Adriaenssens et al., 2012b; Buttimer et al., 2018b, 2018a).

Table 5.3. Inteins and homing endonucleases (free-standing or intron-associated) identified in the genome of *Pectobacterium* phage CB7, using ACT (TBLASTX) for comparison with *Pectobacterium* phage ΦTE and *Cronobacter* phage Cr3.

ORF/associated ORF	Selfish genetic element	Homing endonuclease family	Gene product function of the targeted gene	Does splicing occur (mRNA level)?	Shared with ΦTE	Shared with Cr3
CB7_2	intron associated with homing endonuclease	HNH	large terminase (CB7_1,3)	Yes	Shared	No
CB7_38	free-standing homing endonuclease	HNH	_	_	Shared	No
CB7_48	free-standing homing endonuclease	HNH	_	_	No	No
CB7_50	intron associated with homing endonuclease	HNH	DNA polymerase (CB7_49, 51)	Yes	Shared	No
CB7_52	intron associated with homing endonuclease	HNH	DNA polymerase (CB7_51, 53)	Yes	Shared	No
CB7_54	free-standing homing endonuclease	HNH	_	No	shared	No
CB7_62	free-standing homing endonuclease	HNH	_	_	No	No
CB7_80	intein associated with homing endonuclease	LAGLIDADG	putative helicase (CB7_80)	_	Shared	_
CB7_109	free-standing homing endonuclease	HNH	_	_	Shared	No
CB7_143	free-standing homing endonuclease	HNH	_	_	Shared	No
CB7_180	free-standing homing endonuclease	HNH	_	_	Shared	No
CB7_182	free-standing homing endonuclease	HNH	_	_	Shared	No
CB7_196	intron associated with mobile genetic element	HNH	ribonucleotide reductase NrdB, part 1 & 2 (CB7_195, 197)	Yes	No	No
CB7_197	intein with no homing	_	ribonucleotide reductase NrdB,	_	Shared	_

	endonuclease		part 2 (CB_197)			
CB7_201	intein associated with homing endonuclease	LAGLIDADG	ribonucleotide reductase NrdA, part 1 (CB7_201)	_	Shared	_
CB7_202	intron associated with homing endonuclease	HNH	ribonucleotide reductase NrdA (CB7_201, 203)	Yes	Shared	No
CB7_212	free-standing homing endonuclease	HNH	_	_	Shared	No
CB7_219	free-standing homing endonuclease	HNH	_	_	Shared	No
CB7_220	free-standing homing endonuclease	HNH	_	_	Shared	Shared
CB7_221	free-standing homing endonuclease	HNH	_	_	Shared	No
CB7_230	free-standing homing endonuclease	HNH	_	_	Shared	No
CB7_243	free-standing homing endonuclease	HNH	_	_	Shared	No
CB7_245	intron associated with homing endonuclease	HNH	nicotinamide phosphoribosyltransferase (CB7_245)	No	Shared	No
CB7_249	free-standing homing endonuclease	HNH	_	_	No	No
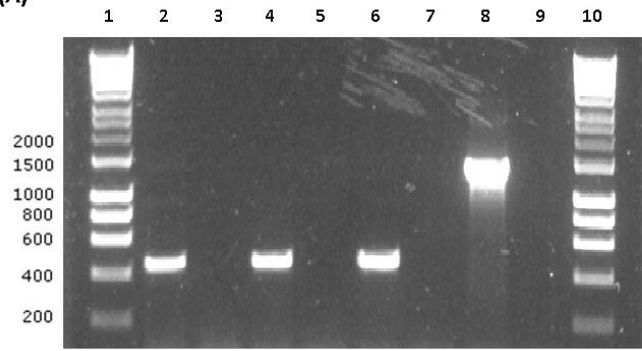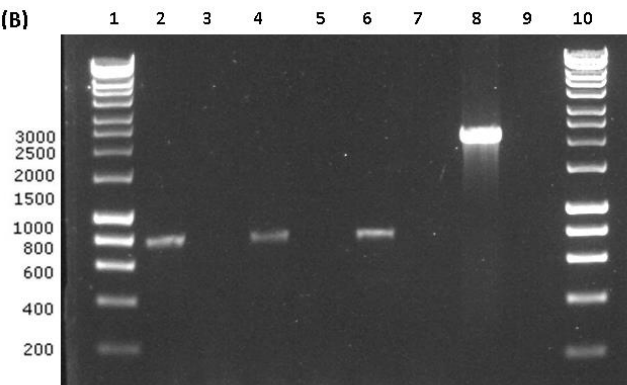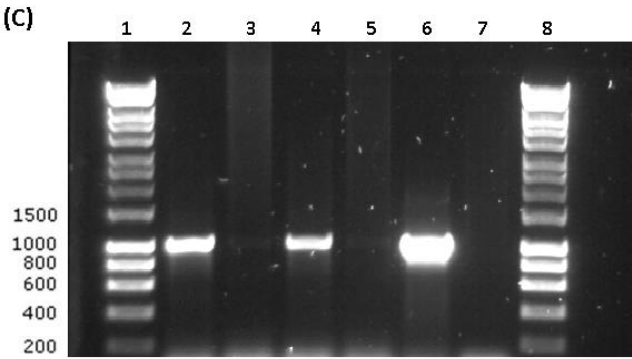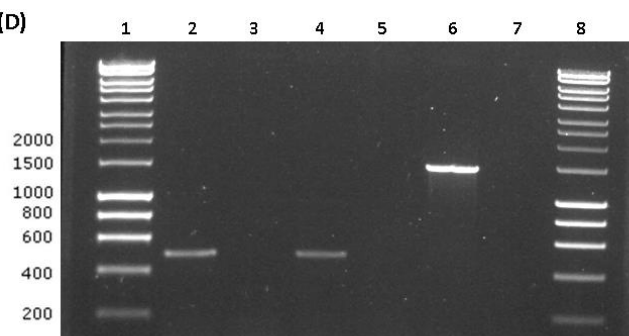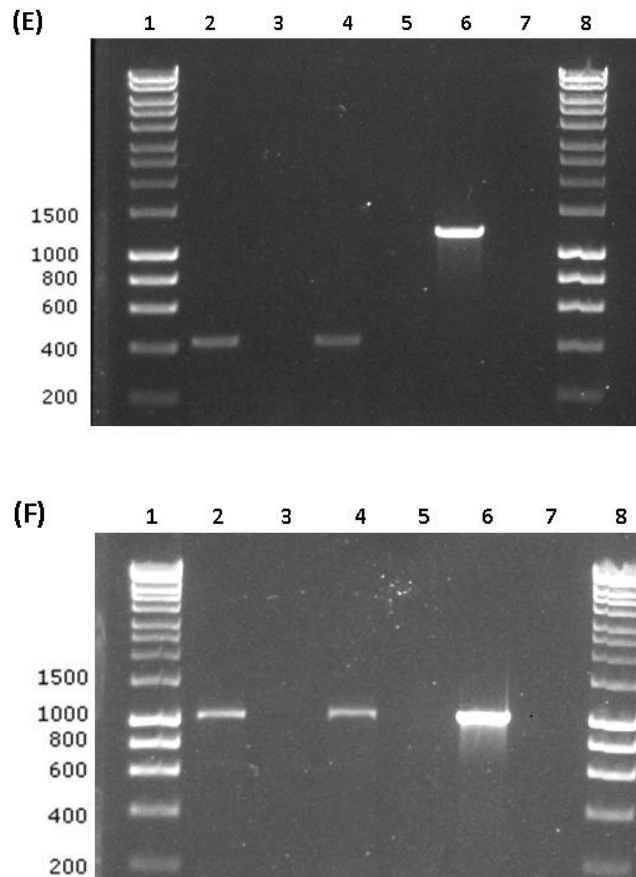
Figure 5.14. Gel images showing PCR products demonstrating splicing of intron-associated homing endonucleases associated with *Pectobacterium* phage CB7 using primers complementary to the 5' and 3' ends of the ORFs. (A) CB7_2; HNH endonuclease situated between ORFs encoding the large terminase (CB7_1,3), (B) CB7_50 and CB7_52; two HNH endonucleases situated between the ORFs encoding the DNA polymerase (CB7_49, 51, 53), (C) CB7_54; HNH endonuclease situated between the ORFs encoding the DNA polymerase (CB7_53) and a structural protein (CB7_55), (D) CB7_196; HNH endonuclease situated between the ORFs encoding the ribonucleotide-diphosphate reductase (CB7_195, 197), (E) CB7_202; HNH endonuclease situated between the ORFs encoding the ribonucleoside triphosphate reductase (CB7_201, 203), (F) CB7_245; HNH endonuclease situated between the ORFs encoding the nicotinamide phosphoribosyltransferase (CB7_244, 246). First and last lanes in each gel contain DNA marker (Hyperladder 1kb, Bioline), the molecular mass of bands is illustrated. Images A & B: PCR products using cDNA (Lanes 2, 4 & 6) and genomic DNA of *Pectobacterium* phage CB7 (Lane 8) and negative controls with water (Lanes 3, 5, 7 & 9). Lanes 2, 4 & 6 used cDNA from phage host infections times points of 15 min and 30 min and 45 min, respectively. Images C, D, E & F: PCR products using cDNA (Lanes 2 & 4) and genomic DNA of *Pectobacterium* phage CB7 (Lane 6) and negative controls with water (Lanes 3, 5 &7). Lanes 2 & 4 used cDNA derived from RNA from phage host infections times points of 15 and 30 min, respectively.

5.2.7. Structural proteome of CB7

The majority of ORFs identified to encode proteins involved in the morphogenesis of the virion of CB7, including the large terminase (CB7_1, 3) which plays a role in the capsid packaging of genomic DNA, are located within gene region 1 of the CB7 genome. One exception was the ORF of structural protein CB7_55, which is situated in gene region 2 (Figure 5.7). ESI-MS/MS analysis was conducted on the virion of CB7, and this is the first phage of the *Cr3virus* genus to have such analysis described. A total of twenty-six proteins were identified to form its virion (Table 5.4). Those for which a function could be inferred were putative tail proteins (CB7_17, 18, 23, 26, 31), tail fibre proteins (CB7_10, 36), tail baseplate proteins (CB7_27, 28, 29, 30) and capsid proteins (CB7_4, 5, 7, 8). *In silico* analysis of the other eleven structural proteins identified failed to find a putative role.

Putative structural protein CB7_32 was identified to encode a tail assembly protein due to it possessing a phage T4 gp38 tail assembly domain (IPR003458). This was not detected by mass spectrometry (MS) analysis. Protein gp38 of phage T4 acts as a chaperone protein in the assembly of the long tail fibre and is not present in the mature phage particle. This may also be the case for the gene product of CB7_32.

Proteins that may be involved in host recognition are CB7_10 and CB7_36 as they may play a role in the structure of phage tail fibres. Gene product CB7_10 possesses a collagen domain (IPR008160); and such domains are commonly associated with phage tail fibre proteins (Ghosh et al., 2012). Additionally, the putative tail collar protein CB7_31 may also play a role in host cell attachment given that it possesses an Ig-domain (IPR003343). Such domains commonly occur on several phage virion proteins such as the tail fibre, the baseplate wedge initiator, the major tail and major capsid proteins. It is believed that this domain may interact weakly with carbohydrates present on the host cell surface (Fraser et al., 2006).

Comparison between phages of the *Cr3virus* genus shows that the majority of the identified structural proteins of phage CB7 are shared among members. Exceptions to this are CB7_55, with homologs of this protein only shared with phages outside the *Cr3virus* genus, such as *Erwinia* phages PhiEaH1 (accession no. YP_009010139) and

vB_Eam_Stratton (accession no. ANZ50590). In addition, the structural protein CB7_251 only has a homolog on phage ΦTE (Appendix, Table S5.9).

When type phages rV5 and PVP-SE1 were first described, the number of proteins that were identified to form the virion of these phages were sixteen and thirty-six proteins respectively. MS analysis of both of these virions allowed the confirmation of six (for rV5) and twenty-five (for PVP-SE1) of these proteins (Kropinski et al., 2013; Santos et al., 2011). Comparisons of these two phages with CB7 show that they share twenty structural proteins, with highly limited homology occurring between several proteins predicted to play roles in tail fibre structure (Appendix, Table S5.10).

Table 5.4. Results of tandem mass spectrometry of proteins of the *Pectobacterium* phage CB7 virion.

ORF	Predicted function	Protein molecular weight (kDa)	No. of unique peptides	Sequence coverage %
CB7_4	putative portal protein	55.37	36	76
CB7_5	putative prohead core protein protease	21.35	2	18
CB7_6	unknown structural protein	40.57	3	15
CB7_7	putative head stabilization/decoration protein	15.6	16	97
CB7_8	putative major head protein	37.6	30	93
CB7_10	putative tail fibre protein	111.95	38	46
CB7_11	unknown structural protein	24.54	7	37
CB7_12	unknown structural protein	17.19	9	48
CB7_13	unknown structural protein	20.14	4	32
CB7_15	unknown structural protein	16.58	5	44
CB7_16	unknown structural protein	26.47	3	16
CB7_17	putative tail sheath protein	50.69	26	89
CB7_18	putative tail tube protein	17.2	6	45
CB7_23	putative tape measure protein	87.72	30	40
CB7_24	unknown structural protein	31.67	5	25
CB7_25	unknown structural protein	14.38	1	15
CB7_26	putative tail protein	36.5	1	5
CB7_27	putative baseplate protein	26.42	9	65
CB7_28	putative tail lysozyme	20.45	1	11
CB7_29	putative baseplate wedge protein	54.37	9	23
CB7_30	putative baseplate protein	24.2	11	57
CB7_31	putative tail collar protein	59.21	19	47
CB7_36	putative tail fibre protein	81.89	19	34
CB7_37	unknown structural protein	43.48	1	21
CB7_55	unknown structural protein	26.53	4	24
CB7_251	unknown structural protein	19.68	1	10

5.2.8. Cell wall degrading enzymes and cell lysis proteins

Peptidoglycan degrading enzymes are used during the initial steps of phage infection to penetrate the host cell wall during injection of phage DNA (virion-associated lysins). They are also employed during host cell lysis at the end of the phage lytic cycle (endolysins) (Oliveira et al., 2018). Three potential peptidoglycan-degrading enzymes were identified in the genome of CB7 (Appendix, Table S5.11). CB7_28 is a putative virion-associated lysin and HHpred analysis of this protein and its homologs among *Cr3virus* and phages rV5 and PVP-SE1 showed that they are homologous to gp25 of

Escherichia phage T4 (using CB7_28; best hit gp25-like lysozyme of *Geobacter sulfurreducens,* PDB accession no. 2IA7_A). This protein of phage T4 forms part of the phage's baseplate and possesses acidic lysozyme activity (Szewczyk et al., 1986). CB7_83 is predicted to be a putative cell wall hydrolase (N-acetylmuramyl-L-alanine amidase) resembling SleB, a protein in *Bacillus subtilis* responsible for hydrolysis of the spore cortex during germination (IPR011105). Homologs of CB7_83 were identified among members of all genera within the *Vequintavirinae* subfamily. A homolog of this protein in *Cronobacter* phage CR3 (namely, CR3_087) has been implicated in host lysis (Shin et al., 2012). Additionally, analysis of this protein with SignalP predicts the presence of a putative N-terminal peptide signal. This feature is recognised by the cell secretary system allowing transport of the protein across the inner membrane into the periplasm of the cell or to the extracellular space, with proteolytic cleavage of the signal peptide sequence (Ivankov et al., 2013). CB7_190 is a potential peptidase (IPR009045) with HHpred analysis showing homology to the endolysin of *Escherichi*a phage T5 (L-alanyl-D-glutamate peptidase, PDB accession no. 2MXZ). This putative peptidoglycan degrading enzyme has only been identified in *Pectobacterium* phages CB7 and ΦTE (phiTE_147) within the *CR3virus* genus with the other genus members having a putative peptidoglycan degrading enzyme of different origin and possibly enzymatic activity (IPR023346).

A putative Rz/Rz1 spanin pair (CB7_252/253) was also identified: these proteins are shared among members of the *Cr3virus* genus as well as phages rV5 and PVP-SE1 (Appendix, Table S5.11). They play a role in the destruction of the outer membrane of Gram-negative cells during host cell lysis allowing progeny phage release at the end of infection. Like other spanin proteins, the CB7_252/25 products are typical examples, one possessing an N-terminal transmembrane domain and the other possessing a lipoprotein signal sequence (Summer et al., 2007). The ORFs of these consist of separate coding sequences where the stop codon of the Rz gene overlaps with the start codon of the Rz1 gene. This gene arrangement is in common with *Escherichia* phage T4 (Summer et al., 2007). For CB7, rV5 and PVP-SE1, this spanin protein pair is not associated with a classic lysis cassette as their genes are located next to those of the large terminase. A similar gene arrangement has also been seen in a number of

Podoviruses such as *Vibrio* phage VP4 and Enterobacteria phage SP6 (Summer et al., 2007).

Another class of proteins associated with host lysis are holins. These create openings in the host cell periplasm allowing the phage endolysin to access cell wall peptidoglycan (Barenboim et al., 1999). As CB7 (including other members of *Cr3virus*) does not appear to have a classic lysis cassette, where all the lysis genes are in proximity to each other, this makes the identification of candidate holins difficult. However, in the case of CB7, a hypothetical membrane protein (65 amino acid residues in length) was observed, whose ORF overlaps that of its putative SleB-like protein CB7_83. This protein is predicted to possess two transmembrane domains where its N - and C - termini are situated in the host cell cytoplasm, making it a strong candidate for a class two holin (CB7_82) (Barenboim et al., 1999). A similar protein is found among a number of other phages of the *Cr3virus* genus (phage CR9 being an exception) downstream of their SleB-like protein (Appendix, Table S5.11).

Phage CB7 does not appear to have homologs of rIIA and rIIB, which are present in rV5 and PVP-SE1. The function of these proteins are not well understood but they are thought to have an influence on the regulation of host lysis relating to membrane integrity and energetics (Paddison et al., 1998)

5.3. Discussion

The *Cr3virus* ICTV-defined genus is the first to date, which includes phages that infect *P. atrosepticum*. However, the coming years should see the creation of additional genera with at least two other phage types known which infect that bacterial species (Buttimer et al., 2018a, 2018b). The categorisation of phage within defined genera should assist in understanding their biology and indeed in their selection as biocontrol agents by knowing more about their various shared properties.

Phylogenetic analysis in this study allowed the positioning of *Pectobacterium* phage CB7 within *Cr3virus* (Table 5.2, Figure 5.9, 5.10, 5.11). Examination of this phage and comparative analysis with other phages within the genus show they possess a conserved transcriptional strategy, involving the use of a promoter with elements resembling that of the *E. coli* sigma70 promoter (Appendix, Tables S5.2.-S5.5.).

Moreover, these phages appear to possess conserved strategies of DNA replication, DNA metabolism, host lysis and virion structure as indicated by their shared protein content associated with these processes. However, there also appears to be limited diversity within the genus, such as members not sharing homolog proteins involved in DNA methylation, nucleotide metabolism, virion structure and those involved in peptidoglycan degradation (Appendix, Table S5.7, S5.9, S5.11).

Of the twenty-six structural proteins identified to form the virion of phage CB7 (Table 5.4), twenty were found to be shared with the type phages rV5 and PVP-SE1 of *V5virus* and *SE1virus*, respectively (Appendix, Table S5.10). The differences within protein content among these phages are concentrated within those proteins predicted to form the tail fibre. Such differences likely reflect the adaption of the tail fibre of these phages to allow the recognition of their respective host cell receptors.

Both *Pectobacterium* phage ΦTE and *Cronobacter* phage CR3 are flagellum-dependent for the recognition of their respective host bacterium and the feature is likely shared among all the *Cr3virus* members (Blower et al., 2012; Shin et al., 2012). *Pectobacterium* phage ΦATI described by Evans *et al*. (Evans et al., 2010b) shares this flagellum receptor specificity and is likely a *Cr3virus,* with currently available partial genome sequences of this phage (Genbank accession nos. FN396585, FN396583 and FN396595) possessing significant homology to that of *Cr3virus* phages. It was reported that mutants of *P. atrosepticum* resistant to phage ΦATI infection due to defective flagella had reduced virulence in potato tuber rot assays (Evans et al., 2010a). Such reduced virulence resulting from phage application would be a desirable outcome were host resistance to develop.

However, a closely related phage of the same genus, *Pectobacterium* phage ΦTE has been shown to be capable of causing generalised transduction (Blower et al., 2012). Both phages CB7 (Figure 5.6) and ΦTE have been identified to possess circularly permuted genomes and the feature is likely shared with other members of *Cr3virus* (Blower et al., 2012). Due to the physical nature of these phage genomes, it is likely they employ a headful packaging strategy. Such strategies are known to have a high occurrence of generalised transduction (Casjens and Gilcrease, 2009; Thierauf et al., 2009). Such a genomic feature could hypothetically contribute to the spread of

pathogen virulence factors among different strains of the bacterial host, were transduction to occur after application of the phages for biocontrol strategies (Buttimer et al., 2017b). On the other hand, T4-like phages with similar packaging strategies have been deemed suitable for biocontrol applications suggesting that it's not a major consideration (Bruttin and Brüssow, 2005; Denou et al., 2009).

Chapter 6. Things are getting hairy;

Pectobacterium phage vB_PcaM_CBB

6.1. Introduction

Bacteriophages (phages) the viruses of bacteria are the most abundant biological entities in the biosphere with an estimated number of 10^{31} (Hendrix, 2002; Whitman et al., 1998). The order *Caudovirales* (the tailed phages) make up the greatest majority of the phages types that have been studied and within this order are the phage families of *Myoviridae*, *Siphovirdae* and *Podoviridae* (Ackermann, 2001)*. Myoviridae* have the most sophisticated virion design, possessing a tail capable of contracting on infection and generally having the largest genomes when compared to the other families (Hatfull, 2008). However, only a small number of the known *Myoviridae* phages have genomes greater than 200 kbp; these are often referred to as 'giant' or 'jumbo' phages (Hendrix, 2009). The largest of these isolated to date are *Bacillus* phage G (498 kbp, accession no. JN638751.1), *Cronobacter* phage GAP32 (358 kbp, accession no. NC_019401), *Escherichia* phage PBECO4 (348 Kbp, accession no. NC_027364) and *Klebsiella* phage Rak2 (345 kbp, accession no. NC_019526). The former three phages, GAP32, PBECO4 and Rak2 are recent discoveries, and their genomes have only been presented within the last three years (Abbasifar et al., 2014; Kim et al., 2013; Šimoliūnas et al., 2013). These phages share a number of protein homologs with the T4-like phages, but they lack a number of universal core proteins found among the *Myoviridae* subfamily of *Tevenvirinae*. Due to the lack of sequence identity, the lack of an even distribution of homologs among their genomes with the T4-like phages and the possession of their own species-specific proteins, it has been proposed that these phages should be placed within a new subfamily (Abbasifar et al., 2014). More recently, there have been two additional phages that share homology to this subfamily (termed the Rak2-like phages from here onwards in this chapter) these are *Escherichia* phage 121Q (348 kbp, accession no. NC_025447.1) and *Klebsiella* phage K61-1 (346 kbp, accession no. AB897757).

This article presents morphological, genomic and structural proteomic findings on another newly isolated Rak2-like phage, *Pectobacterium* phage vB_PcaM_CBB. This is a broad host range *Enterobacteriaceae* infecting jumbo phage which possesses highly atypical whisker-like structures along its contractile tail surface, a feature which has not been described in any of the other Rak2-like phage published to date.

6.2. Results

6.2.1. Growth parameters, morphology and host range

Pectobacterium phage vB_PcaM_CBB was isolated from activated sludge from a wastewater treatment plant in Little Island, Co. Cork, Ireland. The bacterial host used was *Pectobacterium carotovorum sbp. carovotorum* strain CBBL19-37-1 which had previously been isolated from blackleg-infected potato crop from Co. Cork. When the phage was plated and examined by plaque assay utilising a 0.4% agar overlay, pinpoint-sized plaques were observed with inconsistent formation across replicate experiments. This problem was overcome by using a 0.2% agarose overlay as described by Serwer *et al.* (2007). The host range of the phage was examined using a number of bacterial genera and species within the family *Enterobacteriaceae* which showed it was capable of forming plaques on strains of *Pectobacterium carotovorum, Pectobacterium atrosepticum*, *Erwinia mallotivora, Cronobacter muytjensii* and *Cronobacter malonaticus.* The phage was also found to inhibit the growth of *Dickeya dianthicola, Dickeya solani, Pantoea agglomerans* and *Erwinia amylovora,* as lysis was observed on low dilutions (neat, 10^{-1}) of tested phage lysate with no plaque formation observed at subsequent higher dilutions (up to 10^{-8}) (Table 6.1).

Table 6.1. Host range of *Pectobacterium* phage CBB on 35 strains of various members of the bacterial family of *Enterobacteriaceae* as determined by spot testing with serial phage dilutions (the experiment was done twice: on separate days). Results were recorded as 2 (plaque formation), 1 (growth inhibition) and 0 (no sensitivity). Host strain marked with *.

Bacteria	Strain	Sensitivity
Cronobacter muytjensii	ATCC 51329 (type strain)	2
Cronobacter malonaticus	DPC 6531	2
Cronobacter sakazakii	ATCC 29004	0
Dickeya chrysanthemi biovar *chrysanthemi*	LMG 2804 (type strain)	0
Dickeya dianthicola	PD 482	0
	PD 2174	1
	GBBC 1538	1
Dickeya solani	sp. PRI 2222	1
	LMG 25865	1
	GBBC 1502	1
	GBBC 1586	1
Enterobacter cloacae	NCTC 11590	0
Enterobacter gergoviae	NCTC 11434 (type strain)	0
Erwinia amylovora	LMG 2024 (type strain)	0
	GBBC 403	1
Erwinia mallotivora	LMG 1271	2
Pantoea agglomerans	LMG 2660	1
	LMG 2570	0
Pantoea stewartii	LMG 2713	0
	LMG 2714	0
	LMG 2712	0

Pectobacterium atrosepticum	DSM 18077 (type strain)	0
	DSM30186	2
	CB BL5-1	1
	CB BL7-1	1
	CB BL11-1	1
	CB BL12-2	1
	CB BL13-1	1
	CB BL14-1	1
	CB BL15-1	1
	CB BL16-1	2
Pectobacterium carotovorum subsp. *carovotorum*	DSM 30168 (type strain)	0
	DSM 30169	0
	DSM 30170	0
	CB BL19-1-37*	2

Examination of the morphology of CBB by transmission electron microscopy showed it belonged to the family of *Myoviridae* with an A1 morphotype (Ackermann, 2001), displaying a large icosahedral head presenting hexagonal and pentagonal outlines in micrographs (Figure 6.1) from which estimations were made on its dimensions (Table 6.2). Capsids had a height of 126.9 ± 4.9 nm and a width of 128.0 ± 6.2 nm. Tails displayed transverse striations with dimensions of 123.0 ± 2.6 nm x 27.1 ± 1.8 nm with a base plate with dimensions of 36.4 ± 2.3 nm x 14.7 ± 1.2 nm. Phage CBB also possesses a neck passage structure (see triangle in Figure 6.1.) and beneath the base plate a set of six short tail fibres (length 28.6 ± 2.7nm) could also be visualised (see arrow in Figure 6.1). Phage particles with contracted tail sheaths were only rarely found indicating high structural stability of the phage. The contractile tail of CBB was found to be highly unusual as it possesses hair-like appendages (whiskers) covering its surface (see open arrows in Figure 6.1). These hairs are highly stable, as they did not dissociate from the phage tails during a 1½ year of storage at 4 ˚C. Only two other phages have been reported to possess these hairy features; these being *Escherichia coli* phage PhAPEC6 (Tsones, 2014) and an unpurified phage observed in a preparation of crushed silkworm larvae referred to as X particle (Ackermann et al., 1994). Other than CBB having the whiskers, its morphology is typical of the Rak2-like phages in that it has a very large head with a comparatively short tail (Drulis-Kawa et al., 2014).
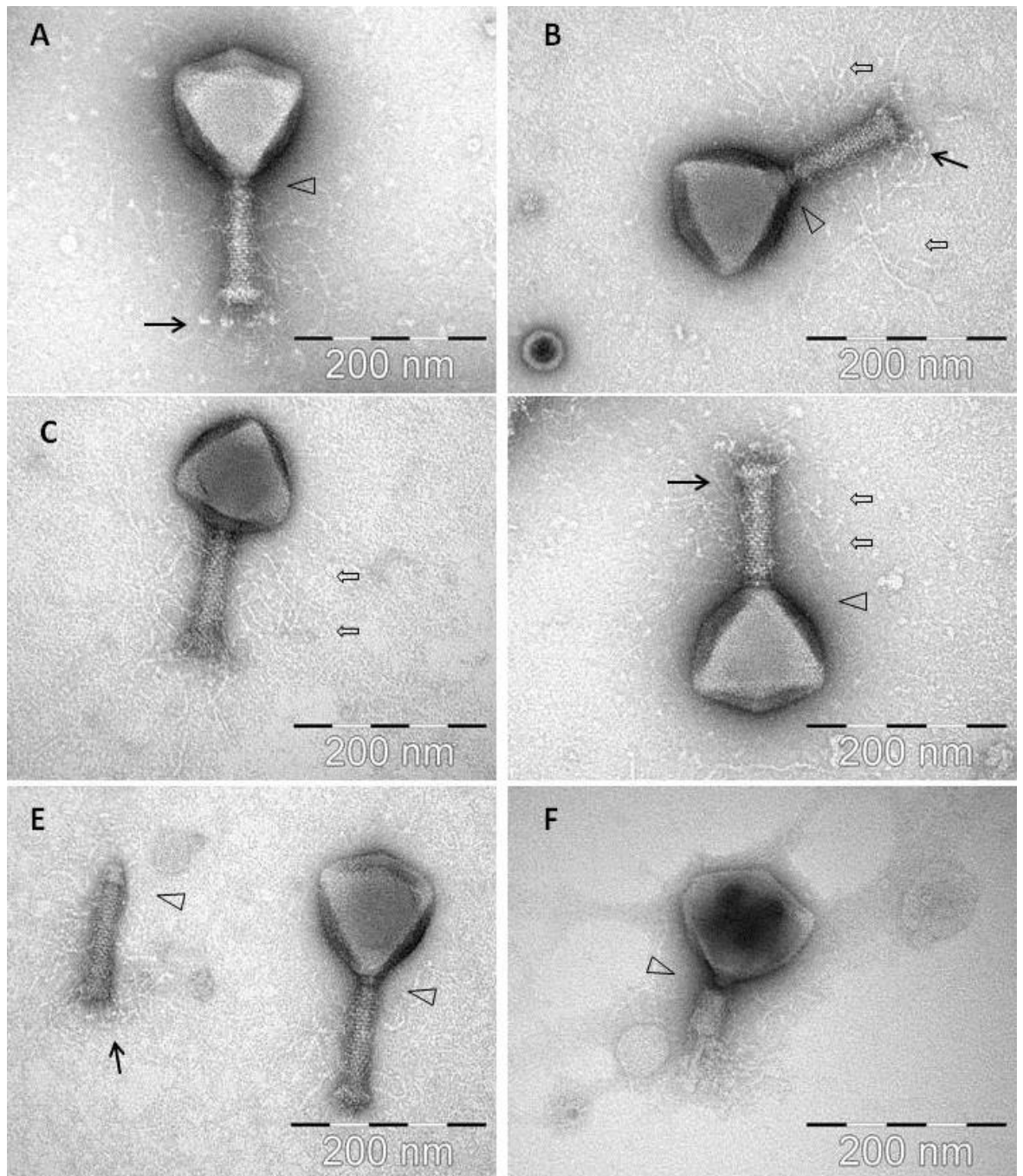
Figure 6.1. Electron micrographs of *Pectobacterium* phage CBB with black arrows indicating baseplate fibres, open arrows indicating hair-like appendages (whiskers) and triangles indicating neck passage structure. (A) CBB virion base plate tail fibres are indicated. (B-D) Fully intact virions with atypical whisker-like structures on the contractile tail surface. (E) CBB virion contractile tail missing a capsid. (F) CBB virion with a contracted tail.

Table 6.2. Estimations of the dimensions of *Pectobacterium* phage CBB derived from micrographs obtained from transmission electron microscopy

Head		Tail		NPS		Baseplate		Tail hairs	Baseplate fibres
height	width	length	width	width	height	width	height	length	length
$n=16$	$n=16$	$n=16$	$n=16$	$n=16$	$n=16$	$n=16$	$n=16$	$n=29$	$n=33$
126.9	128.0	123.0	27.1	21.8	7.4	36.4	14.7	119.3	28.6
± 4.9	± 6.2	± 2.6	± 1.8	± 1 .4	± 0.7	± 2.3	± 1.2	± 10.4	± 2.7

Tail length: incl. NPS (collar) structure and incl. baseplate (but without baseplate fibres)
Tail hairs: measured from flanking tail sheath surface and incl. tiny terminal globular structures
Baseplate fibres: measured incl. terminal globular structures

6.2.2. General genome features

Phage CBB was found to possess a very large genome of 355,922 bp with predicted terminal repeats of 22,456 bp (resulting in total genome size of 378,378 bp). The contig representing its genome was resolved with 1,044,532 reads and with an average coverage of 401x. The existence of terminal repeats was suggested during sequence analysis by the identification of a localised area in the genome which had twice the read depth compared with the rest of the genome. Identification of phage genome ends by this approach has been reported for other phages (Fouts et al., 2013; Li et al., 2014). The existence of the terminal repeats was verified by pulsed-field gel electrophoresis, which indicated that the size of the genome obtained from sequencing is within the correct range (Figure 6.2). However, even when terminal repeats are excluded, the genome is one of the largest published phage genomes to date, behind *Bacillus megaterium* phage G (498 kbp) and *Cronobacter sakazakii* phage vB_CsaM_GAP32 (358,663 bp) (Abbasifar et al., 2014).
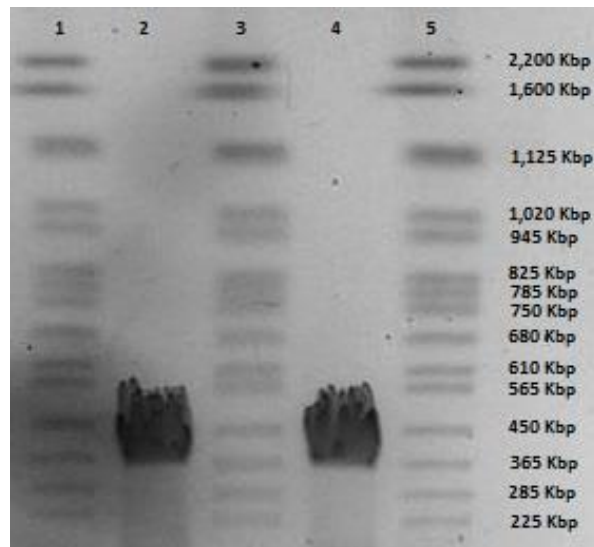
Figure 6.2. PFGE of *Pectobacterium* phage CBB genomic DNA; lanes 1, 3 and 5 yeast chromosome PFG marker (Bio-Rad Laboratories) and lane 2 and 4 genomic DNA of phage CBB.

The average G+C content of the genome of CBB is 36% with below-average skews appearing around ORFs on the minus strand (Figure 6.3). The average G+C content is much lower than that typically found in its hosts species namely, *Pectobacterium atrosepticum* at 50-51% (Bell et al., 2004; Nikolaichik et al., 2014), *P. carotovorum subsp carotovorum* at 52.2% (Park et al., 2012) and *Erwinia mallotivora* at 52.4% (Redzuan et al., 2014). This observation is not as unusual, as virulent phages will usually deviate from their host G+C content having a higher A+T. It has been suggested that this is due to energy and metabolism cost limitations (Rocha and Danchin, 2002). However, the G+C content is quite similar to other Rak2-like phages (35.5%, 34.09% and 35.5% for GAP32, PBECO4 and Rak2, respectively).

The genome of CBB has 554 predicted open reading frames (ORFs) (excluding 51 ORFs within the predicted terminal repeat), of which only 34 are encoded on the minus strand. Putative proteins of twenty-two ORFs were found to share sequence homology with each other, suggesting that their genes may have arisen from duplication events (paralogs) (Appendix, Table S6.1). On the basis of i) observed protein sequence homologies, ii) protein structure homology, iii) lipoprotein and iv) transmembrane analysis it was possible to predict the possible role for 162 of these proteins (Appendix, Table 6.2) with several others being categorised as hypothetical proteins (47),

conserved hypothetical proteins (291), putative lipoproteins (4), and putative membrane proteins (7) or conserved putative membrane proteins (43). No integrase, excisionase or repressor genes were detected in the genome which suggests that this phage follows an exclusively lytic lifestyle.

It is understood that virulent phages can overcome codon utilisation differences from their hosts by using encoded tRNAs, a feature often observed in large genomes (Mesyanzhinov et al., 2002). Phage CBB has a large number of predicted tRNA genes with a possible 33 being identified and concentrated within a ~51Kbp region within the genome. Thirty-one of these genes appear to encode functional tRNA genes, with many of the tRNAs being for amino acids with codons which are highly utilised by the phage (Appendix, Table S6.3, S6.4).

6.2.3. Bacteriophage CBB: a Rak2-like phage and its distant relationship to the *Tevenvirinae* subfamily

Initial BLASTP searches with the predicted ORFs of phage CBB showed that it shared strong homology with the Rak2-like phages. Comparisons among these phages showed that the closest relative to CBB is *Cronobacter sakazaki* phage GAP32 (358,663 bp; 545 ORFs), with 479 homolog proteins followed by *Escherichia coli* 121Q (348, 532; 611 ORFs) with 239 homolog proteins, *E. coli* PBECO4 (348,113 bp; 551 ORFs) with 230 homolog proteins and *Klebsiella* phage RAK2 (345,809 bp; 554 ORFs) with 230 homolog proteins. Phage CBB also shares strong homology with *Klebsiella* phage K64-1 (346,602 bp), however, at present its Genbank file contains an incomplete annotation. This protein homology is illustrated in Figure 6.3. Comparison between the proteomes of these phages shows that they share between 33-38% of their proteins (204 homologous proteins). These shared proteins likely represent the core genome of the Rak2-like phages, given that they appear to be resistant to sequence deviation due to their likely importance for successful phage infection. Nucleotide pairwise comparison (BLASTN) of these phages with CBB show that *Cronobacter* phage GAP32 is the most closely related with 65% identity, followed by *Escherichia* phages 121Q and PBECO4 with 4-5% and with *Klebsiella* phages Rak2 and K64-1 being the least related with 1% identity.

It has been reported that the Rak2-like phages possess a number of proteins that are homologous to those found within the superfamily of T4-like phages (Abbasifar et al., 2014; Šimoliūnas et al., 2013). Phage T4 is the type phage of the genera of *T4virus*, while phage KVP40 is the type phage of *Schizot4virus*. Total protein comparisons of phages CBB, GAP32, PEBEC04, RAK2 with phage T4 (NC_000866) and phage KVP40 (NC_005083.2) show that they share 41 homologous proteins with T4 (14.75% of the total proteins of T4) and 46 homologous proteins with KVP40 (12.07% of the total proteins of KVP40) using Coregenes (BLASTP cut value of 75%). However, despite this correlation to the *Tevenvirinae* subfamily, it is clear that the vast differences suggest these phages should be grouped separately. Nevertheless, a clear evolutionary link is present. There are 38 proteins which are considered to be the core proteins of the T4-like viruses. These proteins range in function from DNA replication, repair and recombination, auxiliary metabolism, gene expression and phage morphogenesis (Petrov et al., 2010). Examination of the CBB genome revealed that it has ORFs for proteins that appear to be homologs to these core proteins, with 21 being identified with functions involved in DNA replication and recombination, auxiliary metabolism as well as phage morphogenesis proteins with homology to those of T4 and KVP40 (reference strains for *T4virus* and *Schizot4virus*, respectively). However, large divergence was observed, especially with morphologically-related proteins, with some having very little homology to those of the type strains (Table 6.3). As well as T4-like core genes, there are a number of genes that have been designated at T4-like quasi-core genes. These are genes whose presence will vary from phage to phage as it is believed that they are not necessary for certain genetic backgrounds (Petrov et al., 2010). In the genome of CBB, it was also possible to identify 12 ORFs specifying putative homologs of these quasi-core proteins (Appendix, Table S6.5).

To further investigate the relationship between CBB and the other Rak2-like phages and the T4-like phages, the amino acid sequence of portal vertex protein (CBB_252) of CBB was used to construct a phylogenetic tree using the top BLASTP hits against CBB_252 (Appendix, Table S6.6). The portal vertex protein has been used as a marker in a number of phylogenetic studies of the T4-like phages (Brewer et al., 2014; Sullivan et al., 2008; Zhong et al., 2002). This tree placed CBB and the other Rak2-like portal

proteins within a well-supported clade showing large divergence from clades representing the *Tevenvirinae* genera of *T4virus*, *Schizot4virus* and clades representing other unclassified T4-like phages. The result suggests that the Rak2-like phages are highly distinct from T4-like phages for which sequence data is available in the public databases (Figure 6.4). The major capsid protein has also been used to conduct similar phylogenetic studies within the T4-like phages (Comeau and Krisch, 2008). A similar phylogenetic study was done with the major capsid protein of CBB (CBB_257) which also suggested the same conclusion to that obtained with the portal vertex protein (Appendix, Figure S6.1).
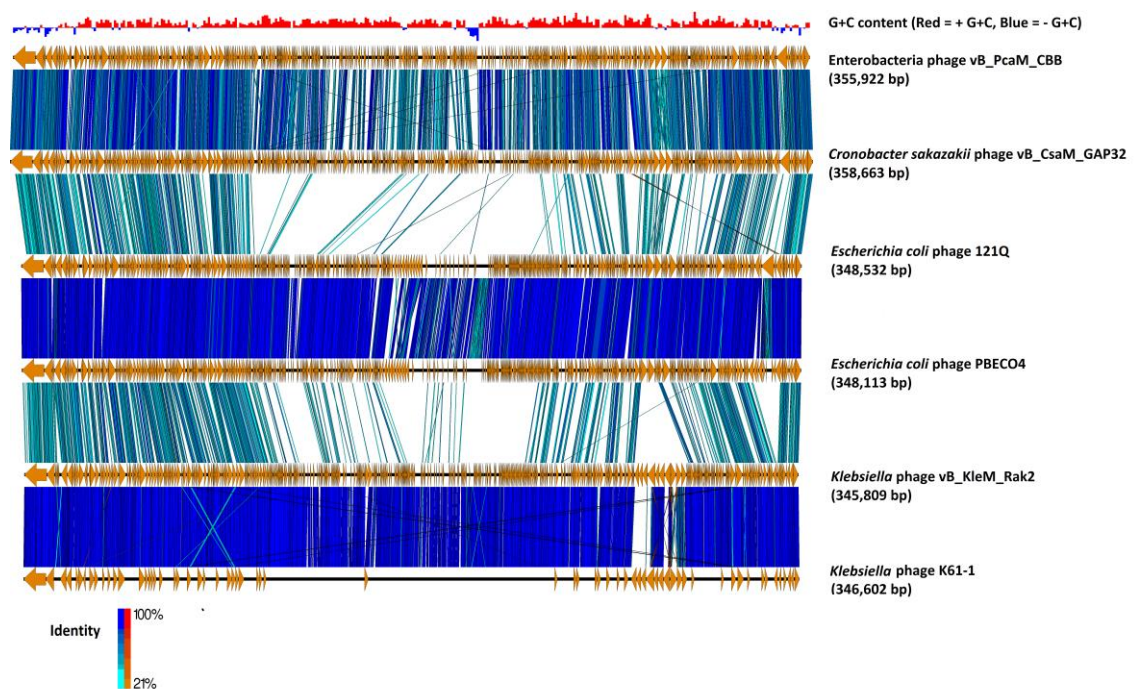
Figure 6.3. Comparison of the genomes of *Pectobacterium* phage CBB (without predicted terminal repeat region) to other potential Rak2-like phages (*Cronobacter sakazakii* phage vB_CsaM_ GAP32, *Escherichia coli* phage 121Q, *E. coli* phage PBECO4, *Klebsiella* phage vB_KleM_RAK2, *Klebsiella* phage K64-1) using currently available annotations employing TBLASTX and visualised with Easyfig (Sullivan et al., 2011). A bar chart shows the G+C skew of the CBB genome, genome maps comprise of orange arrows indicating locations of genes among the different phage genomes; and lines between genome maps indicate the level of homology (blue/turquoise – genes sharing orientation, red/orange – genes in an inverted orientation). To assist in the comparison between genomes the largest gene of each of the phages was positioned as the first gene for each genome.

Table 6.3. Core proteins of T4-like phages identified within the genome of *Pectobacterium* phage CBB with a comparison to homologs in Enterobacter phage T4 and *Vibrio* phage KVP40 (type strains of *T4virus* and *Schizot4virus*, respectively) using BLASTP.

Function	T4 like core protein	CBB homolog	Accession no. of T4 protein	Identity	E-value	Accession no. of KVP40 protein	Identity	E-value
DNA replication, repair and recombination	gp43 - DNA polymerase	CBB_263	NP_049662.1	28%	4.00E-26	NP_899330.1	29%	4.00E-37
	gp44 - sliding clamp complex	CBB_332	NP_049665.1	30%	6.00E-28	NP_899326.1	27%	5.00E-25
	gp41 - helicase-primase complex	CBB_291	NP_049654.1	25%	1.00E-21	NP_899258.1	25%	4.00E-28
	gp32 - single strand binding protein	CBB_277	NP_049854.1	29%	3.00E-08	NP_899253.1	23%	0.005
	gp46 - subunits of recombination nuclease	CBB_246	NP_049669.1	31%	1.00E-23	NP_899322.1	30%	3.00E-21
	gp47 - subunits of recombination nuclease	CBB_245	NP_049672.1	25%	6.00E-17	NP_899320.1	24%	8.00E-14
	UvsW protein - recombination DNA-RNA helicase	CBB_281	NP_049796.1	23%	3.00E-29	NP_899623.1	27%	6.00E-46
Auxiliary metabolism	NrdA - ribonucleotide reductase	CBB_303	NP_049845.1	46%	0	NP_899523.1	47%	0
	NrdB - ribonucleotide reductase	CBB_305	NP_049841.1	40%	3.00E-105	NP_899524.1	45%	9.00E-122
Phage morphogenesis	gp4 - head completion protein	CBB_208	NP_049755.1	34%	1.00E-21	NP_899577.1	42%	2.00E-25
	gp6 - baseplate wedge component	CBB_237	NP_049764.1	42%	0.098	NP_899587.1	27%	3.00E-05
	gp13 - neck protein	CBB_201	NP_049772.1	31%	1.00E-06	NP_899596.1	21%	3.00E-07
	gp15 - tail completion protein	CBB_233	NP_049774.1	20%	3.00E-07	NP_899598.1	21%	2.00E-05

	gp17 - large terminase	CBB_275	NP_049776.1	28%	8.00E-29	NP_899601.1	25%	2.00E-42
	gp18 - tail sheath subunit	CBB_204	NP_049780.1	36%	2.00E-27	NP_899602.1	39%	4.00E-32
	gp20 - head portal vertex protein	CBB_252	NP_049782.1	29%	2.00E-36	NP_899604.1	31%	2.00E-35
	gp21 - prohead core protein	CBB_255	NP_049785.1	31%	2.00E-10	NP_899607.1	37%	1.00E-25
	gp23 - precursor to major head protein	CBB_257	NP_049787.1	35%	3.00E-36	NP_899609.1	34%	8.00E-37
	gp25 - base plate wedge subunit	CBB_238	NP_049800.1	31%	2.00E-10	NP_899586.1	28%	1.00E-08
	gp53 - baseplate wedge component	CBB_242	NP_049756.1	38%	1.00E-04	NP_899579.1	35%	5.00E-07
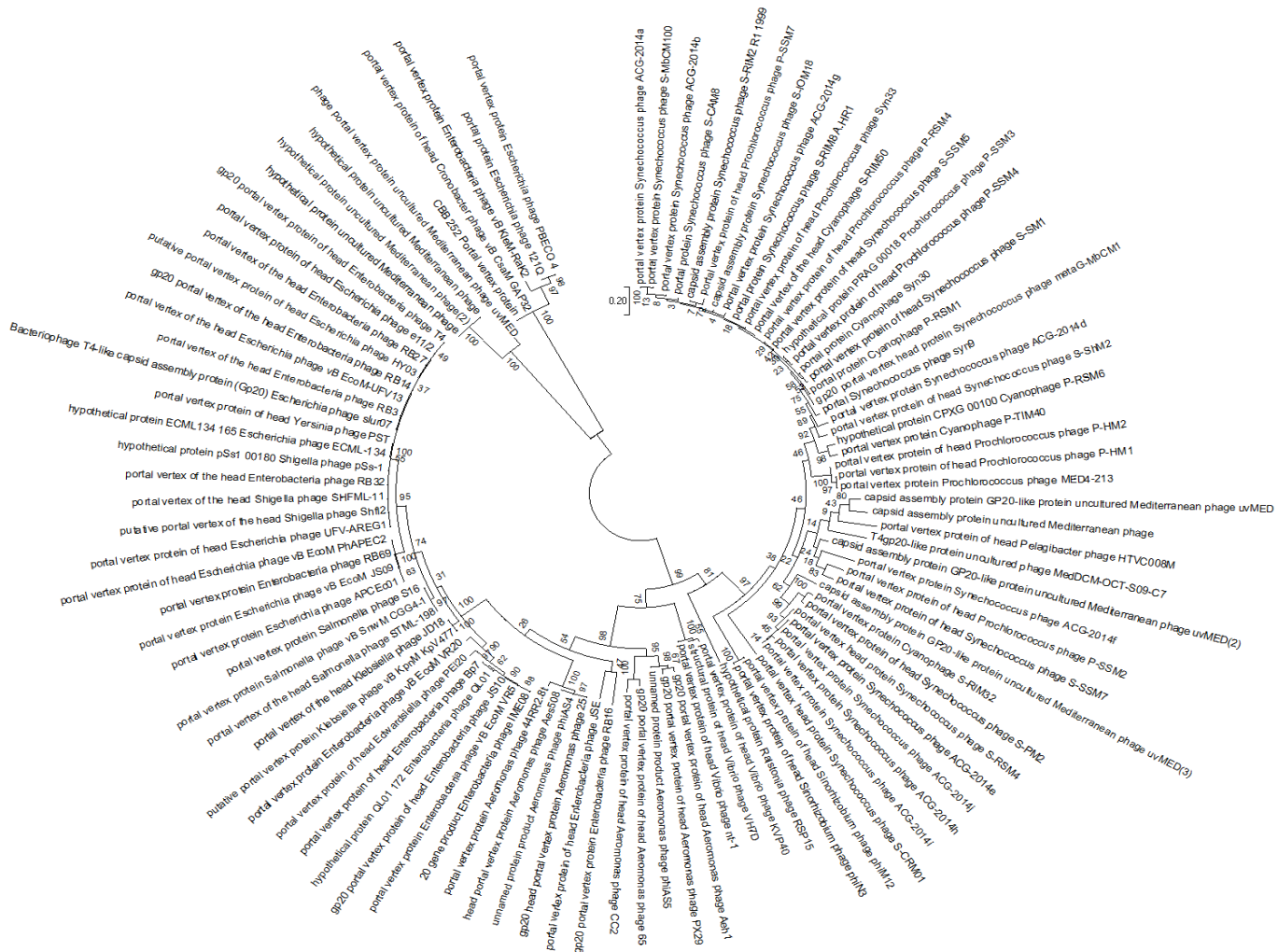
Figure 6.4. Maximum likelihood tree created from the alignment of the conserved region of the portal vertex region of 100 homologous sequences from different T4-like phage to that of the portal vertex protein of *Pectobacterium* phage CBB, found using a BLASTP search. 100 bootstrap replicates were   conducted.

6.2.4. Transcription

Within the genome of CBB, 18 sigma70-like promoters were predicted with a consensus sequence preceded by an UP-like element (Estrem et al., 1998) (Appendix, Table S6.7). Eleven of these promoters were concentrated in a region upstream of *ORFs* CBB_390 to *CBB_546*, with the majority of ORFs within this region encoding short hypothetical proteins (Figure 6.5, part A). Of the few ORFs for proteins of known function, two are most likely to be involved in overriding of the host transcription: the transcriptional regulator (CBB_511) and RNA polymerase sigma factor (CBB_513). The RNA polymerase sigma factor contains an RNA polymerase -70 like domain (IPR014284) containing sigma factor region 2 (IPR013325), region 3 and region 4 (IPR013324). Protein CBB_476 may also have a role in overriding host transcription due to a PrlF antitoxin family domain (IPR031848). It is suspected that the mentioned region CBB_390 to CBB_546 represents the early genes which are expressed in the very early stages of infection and are involved in host take over. It should also be noted that three of the other seven sigma70-like promoters lie on the terminal repeat downstream of this region and could possibly be a continuation of the early gene region with the remaining promoters being located on the opposite end of the genome (Figure 6.4, part B).

Aside from this sigma70-like promoter, a second possible promoter was detected which shows resemblance to a classic sigma70-like promoter with a classic -10 element (TATA). However, it differs with its -35 element with a consensus sequence of TGAAACG instead of TTGACA (Appendix, Table S6.8). This motif was also identified in the GAP32 genome (Abbasifar et al., 2014). Of the 18 examples of this detected motif, 11 were upstream of ORFs CBB_258 to CBB_329 (Figure 6.5, part B), with many of the ORFs within this region found to encode proteins with functions related to DNA synthesis (CBB_263, 280, 328, 290, 291, 315, 277, 324 and 325). There was also a T4-like gp55-like sigma factor (PHA02547, 4.69e-08) predicted for possible late transcription (CBB_244) just downstream of one of these promoters (Porf_244). However, no homologs for the other associated late transcription proteins of phage T4, namely gp33 and gp45, were detected (Geiduschek and Kassavetis, 2010).

A third potential promoter was found by analysing the upstream sequences of ORFs that had been predicted to encode structure-related proteins of CBB in a single submission to MEME. Analysis showed the presence of a consensus sequence of ATAAATA with a concentration of A and T downstream of this motif. This hypothetical promoter was found to be present in 35 locations and has also been predicted to precede structurally related genes in GAP32. It is expected that this sequence plays a role in late gene expression (Appendix, Table S6.9). This motif resembles that of the T4 late promoter sequence, namely TATAAATA (Geiduschek and Kassavetis, 2010).

Apart from the above-mentioned promoters, 74 potential rho factor independent terminators were identified in the CBB genome (Appendix, Table S6.10)
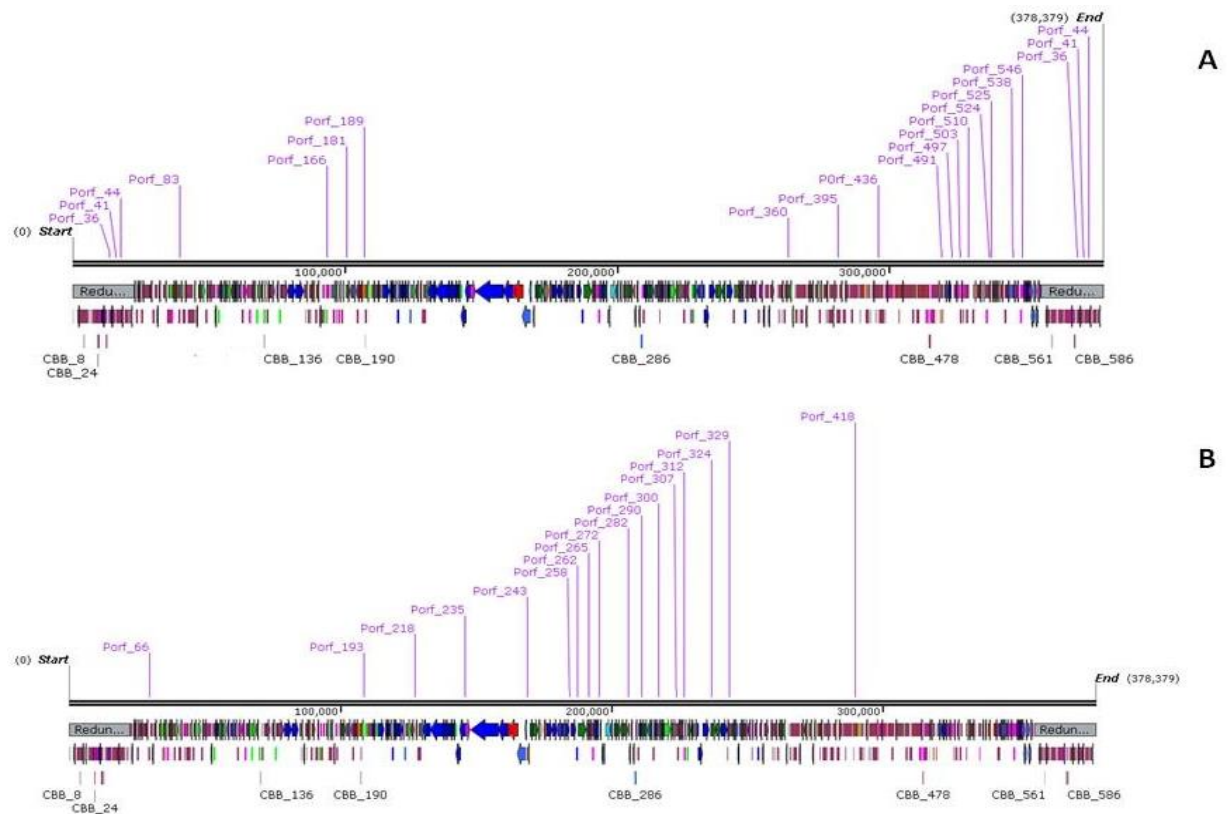
Figure 6.5. Genome maps of *Pectobacterium* phage CBB showing locations of sigma70-like promoters (A) and CBB divergent sigma70-like promoters (B) created using SnapGene.

6.2.5. DNA replication, DNA modification and nucleotide metabolism

There are eight proteins which form the T4 DNA replisome these being DNA polymerase (gp43), clamp loaders (gp44, gp62, gp45), single strand DNA binding

protein (gp32), primase (gp61), helicase (gp41) and loading protein (gp59) (Nelson et al., 2009). CBB was found to have homologs to five of the eight proteins (CBB_263, 332, 277, 290 and 291), but lacks the clamp loaders (gp62, gp45) and loading protein (gp59). Interestingly, the phage possesses a bacterial DNA polymerase III epsilon subunit (CBB_280) with a DnaQ domain (COGO847, 4.32e-21). There was also a protein which resembles the bacterial DNA polymerase III alpha subunit (CBB_328). However, this protein appears to be truncated, being much smaller in size to its closest bacterial hits on BLASTP (172aa vs c.1120aa) such as those from *Oceanobacillus massiliensis* (WP_010651140.1) and *Bacillus megaterium* (WP_026681428.1).

The phage also has T4 homologs for helicase uvsw (CBB_281) and dda (CBB_315), topoisomerase gp52 (CBB_325), RNA ligase rnlA (CBB_83) and DNA ligase gp30 (CBB100), while one of its two ribonuclease H enzymes (CBB_272, 314) is related to T4 rnh. In addition, an array of T4 homologs involved in DNA repair and recombination were identified, including gp46 and gp47 (CBB_246,267), endonuclease VII (CBB_248), uvsX (CBB_278), uvxW (CBB_281) and UV repair endonuclease V (CBB_519). CBB also has other DNA-replication proteins that have no counterpart to phage T4 such as primase-helicase (CBB_291) and gyrase subunit (CBB_324).

Like the other Rak2-like phages, CBB has ORFs coding for a number of enzymes apparently involved in nucleotide metabolism including the aerobic class I NrdA-NrdB (CBB_303, 305) and anaerobic class III NrdD-NrdG (CBB_141, 145) ribonucleotide reductase (RNR) enzymes for the conversion of ribonucleotides to deoxyribonucleotides (Dwivedi et al., 2013). There is also a glutaredoxin (CBB_194) to potentially support with class I RNR function (Sengupta and Holmgren, 2014). CBB has two CMP/dCMP deaminases (CBB_110,135). These generate dUMP, which is a substrate for its thymidylate synthase (CBB_223, T4 homolog td) which generates dTMP. The thymidylate synthase is supported by a dihydrofolate reductase (CBB_313, T4 homolog Frd), which provides the intermediate metabolite tetrahydrofolate. The phage has a thymidine kinase (CBB_203, T4 homolog tk) for the production of TMP, and such dNMPs may potentially be utilised by its deoxynucleoside-monophosphate kinase (CBB_203) for the production of dNDP derivatives. There is also a putative 5', 3'

deoxyribonucleotidase (CBB_380), an enzyme in humans and mice involved in the dephosphorisation of dNMP (Walldén et al., 2007).

Like GAP32, CBB has genes that appear to encode possible nicotinamide-nucleotide adenylyltransferases (CBB_113, 361) as well as a gene encoding nicotinamide phosphoribosyltransferase (CBB_115). These enzymes are involved in the *de novo* synthesis of coenzyme NAD^+ (Schweiger et al., 2001) and their function is supported by a nicotinamide mononucleotide transporter [PnuC] (CBB_361), a membrane-bound protein involved in the uptake of the NAD^+ precursor nicotinamide mononucleotide (Zhu et al., 1991).

DNA methylation and glycosylation are strategies used by phages to provide protection against host restriction (Samson et al., 2013). CBB possesses DNA methylation enzymes for both adenine and cytosine (CBB_72, 157, 399). However, unlike GAP32, CBB lacks a second copy of DNA N-6-adenine methyltransferase (GAP32_519). No enzymes were identified in the CBB genome related to the production of glycosylated hydroxymethylcytosine which are found in the T-even phages (Petrov et al., 2010).

6.2.6. tRNA gene and tRNA related proteins

Phage CBB has a large number of tRNA genes and it appears that in two cases (tRNA gene 13 and 22), correct formation and function of their tRNA gene products is assisted by tyrosyl tRNA synthetase (CBB_132) and tRNAHis guanylyltransferase (CBB_152) respectively, both of which are also encoded on the CBB genome. The phage also appears to assist in tRNA turnover with the capability of releasing tRNA molecules from newly formed peptides by a tRNA peptidyl-tRNA hydrolase (CBB_392). All three gene products have homologs present in GAP32 with just CBB_132 and CBB_392 being shared with Rak2 also. As mentioned CBB has two RNA ligase enzymes (CBB_83, 181) of these CBB_83 may be involved in tRNA repair (Wang et al., 2006).

6.2.7. Translation and post-translation

The phage also appears to be able to assist in the formation of the bacterial translation initiation complex (ribosome, mRNA and tRNA) by possessing its own translation initiation factor IF-3 (CBB_318), with homologs of this protein also being identified in

GAP32, 121Q, PBECO4 and Rak2. In addition, CBB has two ORFs encoding for GroES-like proteins (CBB_131,268). Large phages such as T4 and *Enterobacter* phage RB49 can use host-encoded co-chaperonin GroES to assist in the correct folding of their own structural proteins. However, these two phages also possess their own phage-encoded GroES-like proteins that likely mimic bacterial host GroES proteins (Keppel et al., 2002), this may also be the case for CBB

6.2.8. Terminase and DNA packing

The packaging of bacteriophage T4 DNA into its capsid requires two proteins; the small terminase (gp16) and the large terminase (gp17). Phage CBB has two proteins which possess the 17 terminase domain (PHA02533), namely CBB_274 (E-value 1.28e-09) and CBB_275 (E-value 5.47e-79). However, in the case of CBB_274, this domain is incomplete suggesting it to be a truncated form of a large terminase. This feature was also found to be present with the other Rak2-like phages and is considered to be unusual and has also been identified in the T4 superfamily *Sinorhizobium* phages phiM12 (NC_027204) and phiN3 (NC_028945).

The genomic DNA of phage T4 is packaged into its capsid with the headful packaging strategy, which results in a partially circularly permuted chromosome that is terminally repeated (Casjens and Gilcrease, 2009). The exact physical nature of the genome of CBB currently remains unknown; but as discussed previously, analysis of its genome sequence reads suggests the presence of terminal repeats.

CBB also has an ORF for a T4 homolog to endonuclease VII (CBB_248), which in T4 is involved in DNA packaging as well as recombination and mismatch repair.

6.2.9. Selfish genetic elements

Homing endonucleases are selfish mobile genetic elements with endonuclease activity that only promote the spread of their own encoded gene. These can be found as self-standing genes within introns, as fusions with host proteins or also in self-splicing inteins and their presence is prevalent in a number of phage genomes (Edgell et al., 2010). In CBB, three free-standing homing endonuclease genes of the HNH family were identified, namely homing endonucleases CBB_420, CBB_ 528 and CBB_535 resembling those found in bacteriophages Rak2 and GAP32. Homing endonuclease

CBB_420 lies at the end of tRNA gene 7A, CBB_528 sits between a hypothetical and a putative membrane protein and CBB_535 lies at end of virion structural protein, Inteins are selfish genetic elements which self-cleave from a protein posttranslation. These elements are typically spread by homing endonuclease elements (Tori and Perler, 2011). Nevertheless, no intein-related domains (hedgehog/Hint) were found in the genome of CBB. However, it should be mentioned that inteins can occur in the Rak2-like phages with hedgehog/Hint domain (smart00306) in association with the domain of a homing endonuclease of the LAGLIDADG family (pfam14528). Both were detected in the large terminase of phages Rak2 and K64-1.

6.2.10. Cell wall degrading enzymes

The 35.5 kDa gene product of CBB_187 is predicted to be the endolysin of CBB with homologs identified in GAP32 (gp180), 121Q (gp532) and RAK2 (gp506). *In silco* analysis shows that the protein is modular in structure, possessing three domains, these being an N-terminal transmembrane domain, a possible intermediate cell wall-binding domain showing weak homology to the LysM domain (IPR018392), and a C-terminus lysozyme domain (IPR023347). It is worth mentioning that the cell wall binding domain is very apparent in the 121Q homolog (PB1_121_532) of this protein. The presence of a cell binding domain is unusual in phages infecting Gram-negative bacteria and only a few endolysins with this feature have been reported to date, examples being the *Pseudomonas aeruginosa* phage endolysins KZ144 (phage φKZ), and EL188 (phage EL) (Briers et al., 2009, 2007). However, the cell-binding domains of these differ to those identified in the Rak2-like phages. Another feature of this endolysin is the presence of an N-terminal transmembrane domain, reported in endolysins that follow the signal-arrest-system (SAR), examples being P1 lyz and φKMV gp45 lysin (Xu et al. 2004).

Three identified proteins (CBB_238,239 and 240) are possibly associated with the baseplate structure of the phage tail, and each has a lysozyme domain. Gene product CBB_238 possesses a GPW 25-like domain (PF04965), which is related to gp25 of phage T4. The latter forms part of the T4 baseplate and possesses acidic lysozyme activity (Szewczyk et al., 1986), with protein CBB_238 also being quite similar in size (129aa vs 132aa). CBB_240 a N-terminal gp5 domain (IPR006531) with a C-terminal

lysozyme (IPR002196). The gp5 N-terminal domain is an oligosaccharide-binding (OB) fold and is related to the tail spike (gp5) of T4. However, it is not certain if CBB_240 has this function as it is much larger than gp5 (894aa vs 575aa) and the domain architectures differ, with gp5 possessing an N-terminal OB-fold, centrally located lysozyme and C-terminal triple stranded-helix domains (Kanamaru et al., 2002). CBB_240 may also contain a centrally located M23 family peptidase domain (pfam01551, E-value 6.48E-4) which could be identified in homologs in other Rak2-like phages. Gene product CBB_239 is 18.5 kDa and possesses a T4-type lysozyme domain (IPR001165).

6.2.11. Demonstration of the activity of proteins CBB_187 and CBB_239

6.2.11.1. CBB_187 possess SAR activity

For a number of SAR endolysins, an inducible system with *E.coli* has permitted the demonstration of SAR activity with the detection of a reduction in culture turbidity upon induction of the cloned gene (Briers et al., 2011; Lynch et al., 2013). A similar experiment was used to demonstrate this activity for CBB_187. Expression of the cloned gene was observed to be lethal to cells as the addition of IPTG after four hours of growth caused the OD_{600} to drop from ~0.400 to ~0.200 (Figure 6.5). This result would suggest the protein is a SAR endolysin using the host *sec* system to translocate it to the inner membrane of the cell, thus allowing the lysozyme domain of the protein to reach the periplasm to act on the cell wall of the bacterium. No lytic activity was detected for the truncated transmembrane domain version of the protein (rTM-CBB_187), with induction of the protein in *E.coli* not causing lysis.

An important feature of SAR endolysins is the regulation of their lytic activity while in their membrane-tethered form. The two main mechanisms that have been identified for those with the lysozyme domain are (a) those which regulate activity by the formation of disulphide bridges (as seen with LyzP1 of coliphage P1) and (b) those which utilise steric hindrance of catalytic residues due to the proximity of the bilayer to which the enzyme is embedded, as seen with R21of lambdoid phage 21 (Kuty et al., 2010). As CBB_187 possesses only one cysteine residue (Cys 265), making the formation of disulphide bridges impossible, it would suggest that the mechanism of

enzymatic regulation of the protein is more likely to be similar to that of the latter mechanism.
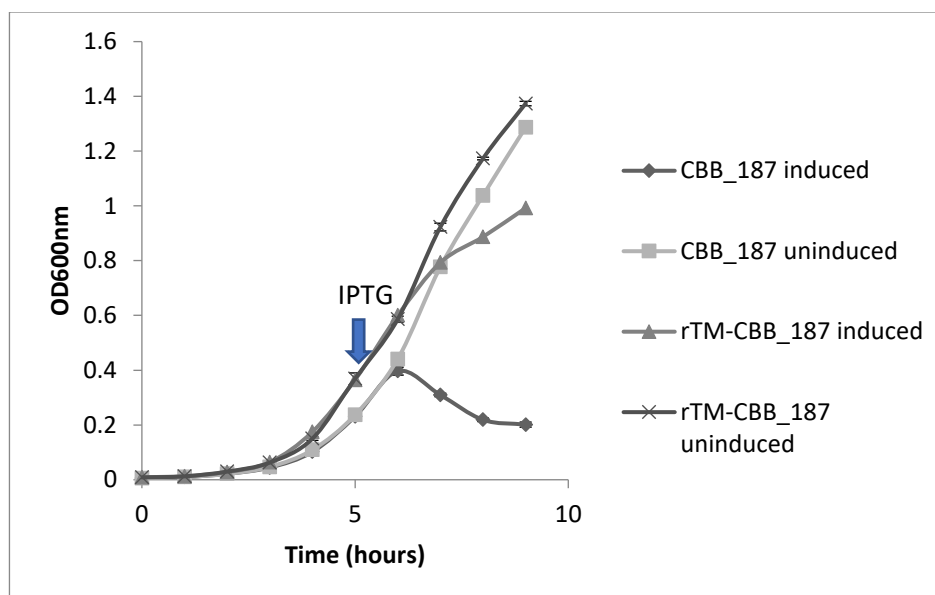
Figure 6.5. Growth curve of *E. coli* Lemo21 (DE3) carrying pET28a with SAR endolysin CBB_187 and its transmembrane domain truncated derivative (rTM-CBB_187). The culture was induced at OD_{600} ~0.400 with 400 μM IPTG (indicated with an arrow). All measurements were carried out in triplicate and averages are represented

6.2.11.2. Zymograms and turbidity reduction assays show CBB_239 can breakdown peptidoglycan

The ability of CBB_239 to degrade peptidoglycan was first demonstrated by zymogram assay. Lysis was detected for all Gram-negative strains tested (Table 6.3), with the observation of a band clearing with a molecular weight between 17-20 kDa with a tail-back of lysis above the band on the zymogram gels. However, no activity was detected for the Gram-positive strains. The turbidity reduction assay also demonstrated the activity of CBB_239. Prior to protein exposure, cells of *P. carotovorum* subsp. *carotovorum* were treated with chloroform-saturated 50 mM Tris-HCl (pH 8) buffer causing outer membrane removal before exposure of the cell wall to the protein CBB_239. CBB_239 (final concentration of 0.15μg/mL) could be seen to reduce the OD_{600nm} over a 10 min period, indicating the breakdown of bacterial cell wall due to the enzymatic activity.

CBB_239 is a glycoside hydrolase with its enzymatic action directed towards the glycan strand, the universal component of peptidoglycan among Gram-negative and positive bacteria. Zymogram assays indicated that the enzyme has a wide host range, with activity being found against peptidoglycan of all tested strains of Gram-negative bacteria (Table 6.3). This level of activity is typical of lysozyme proteins of Gram-negative infecting phages, due to the highly conserved nature of peptidoglycan among these bacteria, possessing the A1γ chemotype (Briers et al., 2007; Oliveira et al., 2016, 2014). No enzymatic activity was found with the peptidoglycan of the Gram-positive species *Listeria monocytogenes* (A1γ chemotype) and *Staphylococcus aureus* (A3α chemotype). The peptidoglycan of both these bacteria is known to exhibit resistance to egg-white lysozyme. This resistance has been determined to be due to the presence of deacetylated glucosamine residues for *L. monocytogenes* and O-acetylated muramic acid residues for *S. aureus* on the glycan strand of the peptidoglycan of these bacteria (Bera et al., 2004; Kamisango et al., 1982). Such peptidoglycan modifications could possibly explain their resistance to CBB_239.

Table 6.3. Lysis spectrum of CBB_239 observed on zymograms using crude peptidoglycan from a number of different Gram-negative and Gram-positive bacteria.

Strain name	Bacteria origin	Peptidoglycan hydrolysis
Pectobacerium atrospeticum DSM 18077	*Solanum tuberosum*	+
P. atrosepticum DSM 30186	*Solanum tuberosum*	+
P. atrosepticum CB BL 13-1	*Solanum tuberosum*	+
P. atrosepticum CB BL 16-1	*Solanum tuberosum*	+
Pectobacterium carotovorum subsp. carotovorum DSM 30168	*Solanum tuberosum*	+
P. carotovorum subsp. carotovorum DSM 30169	*Brassica oleracea* var. *capitata*	+
P. carotovorum subsp. carotovorum BL 30-1-37	*Solanum tuberosum*	+
Erwinia mallotivora LMG 1271	*Mallotus japonicus*	+
Cronobacter mutyjensii ATCC51329	_	+
Cronobacter malanaticus DPC 6531	Brain tumour	+
Cronobacter sakazakii ATCC 29004	_	+
Escherichia coli Lemo21 (DE3)	_	+
Staphylococcus aureus ATCC 6538	_	-
Listeria monocytogenes 10403 S	_	-

Results recorded as +, peptidoglycan degrading activity observed; -, no peptidoglycan degrading activity observed.
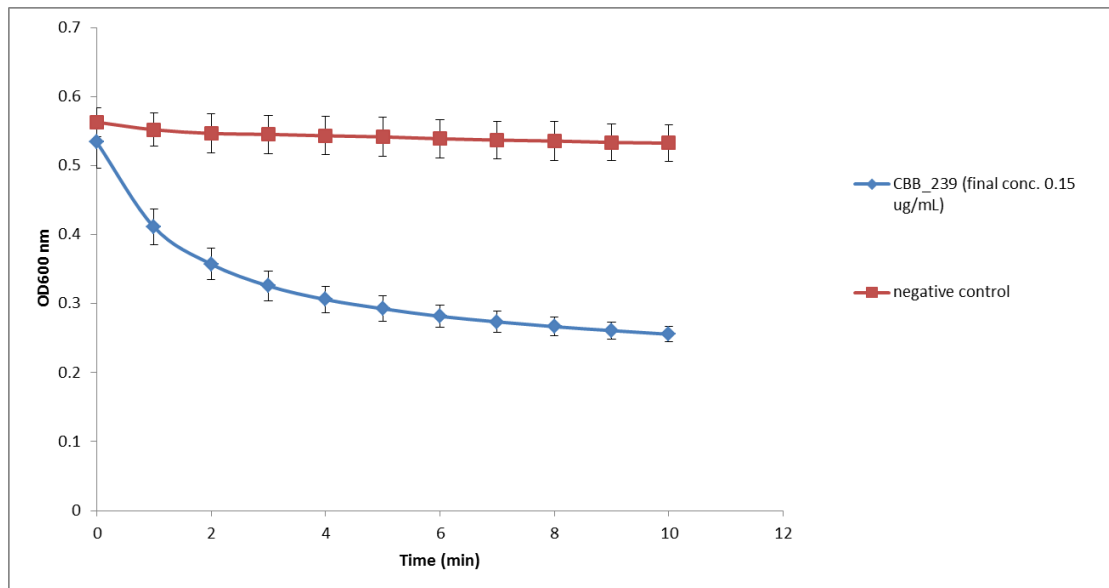
Figure 6.6. Turbidly reduction assay of desalted his-tagged recombinant CBB_239 (final conc. 0.15µg/mL) on *Pectobacterium carotovorum* subsp. *carotovorum* CB BL19-1-37 treated with chloroform saturated 10mM Tris-HCl (pH7) to remove outer membrane. The negative control was a desalting buffer of 10mM sodium phosphate (pH 7.4), 150 mM NaCl. All measurements were carried out in triplicate and averages are represented.

6.2.12. Structural proteins of CBB and other Rak2-like phages

To identify gene products involved in the maturation and structure of the capsid of CBB, an *in-silico* approach was first undertaken to identify homologs of the proteins with known functions. This was followed by the identification of proteins with domains with structure-related roles and also by protein analysis using mass spectrometry (ESI-MS/MS). The identification of phage structural homologs was further supported by structural proteomic analysis data that had previously been conducted on other Rak2-like phages, namely GAP32 and Rak2 (Abbasifar et al., 2014; Šimoliūnas et al., 2013). Using this approach, it was possible to predict 65 gene products with structural roles (excluding the terminase proteins as well as the PhoH and ssDNA binding protein of Rak2). Such genes were identified to have functions in the structure of the capsid (CBB_208, 252, 255, 257), neck (CBB_201), tail (CBB_112, 204, 233), baseplate (CBB_237, 238, 239, 240, 242) and tail fibres (CBB_226, 260, 261). Next, we attempted an experimental verification of structural proteins using SDS-PAGE of the denatured virion proteins (Figure 6.7), followed by trypsin digestion and mass spectrometry (ESI-MS/MS). From this work, 55 proteins were associated with the structure of the phage

capsid (Table 6.5) with sequence coverage ranging from 4.1% to 91.3% and the number of unique peptides ranging from 1 to 30. These results allowed the identification of an additional 18 structural gene products to be associated with the structure of the mature CBB phage particle (Appendix, Table S7.11).

The structure of CBB is very similar to that of GAP32, sharing 76 homologs of the 83 gene products predicted to be involved in CBB morphology. Phage Rak2 however, is less similar to CBB, only sharing 64 homologs of these genes. TEM images taken of these phages shows that tail fibres associated with the baseplates of CBB and GAP32 are morphological very similar with both greatly differing to that of Rak2 (Abbasifar et al., 2014; Šimoliūnas et al., 2013). Analysis of the structural genes also indicates this. There are tail fibre related proteins (CBB_226, 261) in the genomes of GAP32 and CBB which have no homologs in Rak2 and there are also a number of tail fibre/tail spike gene products (RAK_527, 528, 530, 530, 532) in Rak2 which are not represented in CBB and GAP32. These differences most likely represent adaption of tail components correlating with host ranges of these phages.

The only structural protein of GAP32 that had previously been identified by mass spectrometry, which was not found in the genome of CBB, was GAP_001. It resembles a DNA chromosome condensation protein. This protein possesses the AST1 (COG5184) domain of the Conserved Domain Database and the regulator of chromosome condensation 1/ beta-lactamase-inhibitor protein II (IPR009091) domain of Interproscan. Interestingly homologues of this protein were also identified in multiple copies in the genomes of other Rak2-like phages 121Q (121Q_234-238, 240-242 and 244-248) and PBEC04 (ACQ_286, 289, 292 and 294-296).

The majority of the structural genes of CBB are focused around one large region of approximately 70Kbp (CBB_201-261), with small clusters of genes occurring throughout the rest of the genome. There are six structural proteins (CBB_144, 247, 536, 537, 551 and 552) that have been identified by mass spectrometry whose genes are not present on GAP32 genome, and four of these occur in a small ~10kb cluster (CBB_534 to 553).

One of the unusual features found with phage CBB when compared to GAP32 was a number of proteins containing Kelch-like domains for which there were no obvious homologues present in the other Rak2-like phages. These proteins were CBB_550, possessing a Kelch-type beta propeller (IPR015915) domain and CBB_552 and CBB_554, both with a Kelch-type beta propeller (IPR015915) and Kelch repeat type 1 (IPR006652) domains. These domains are comprised of six four-stranded beta propellers and have been identified as important in protein-binding interactions in a number of non-phage-related proteins. Of the three mentioned proteins, CBB_552 has been identified to form part of the mature phage structure. Proteins containing Kelch-like domains have been identified in the proteomes of eukaryotic viruses such as poxvirus. Their presence is unusual in phage but they have been reported in other giant phages that infect *Pseudomonas* and *Yersinia* (Barry et al., 2010; Hertveldt et al., 2005; Skurnik et al., 2012)

Given the high level of morphological and genetic similarity of phage CBB to GAP32, it is logical to assume that the presence of the extra morphological feature of hair-like structures along the tail shaft of phage CBB would be represented by morphological genes not present in GAP32 (other than CBB_370, a homologue of which is found in Rak2), if GAP32 does indeed lack these atypical whisker-like structures observed on CBB. However, it still possible that a number of other structural proteins of CBB may not have been detected in the mass spectrometry assays conducted for this article. At this time, it is not possible to pinpoint the exact gene products responsible for the atypical whisker-like structures observed in phage CBB.
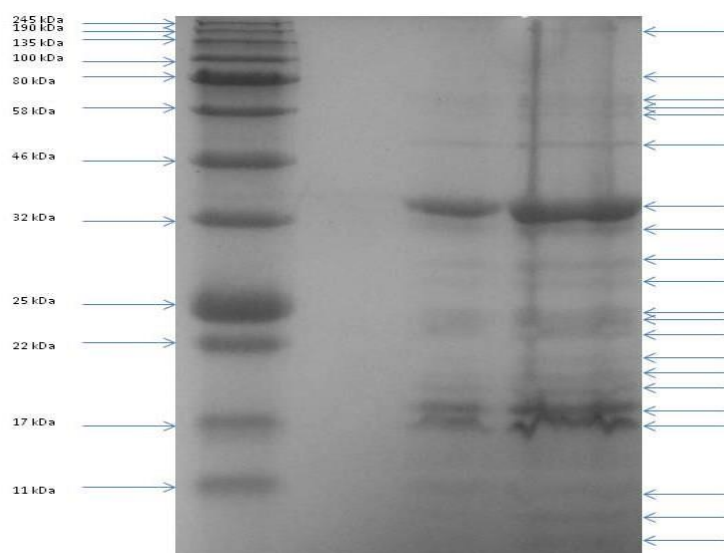
Figure 6.7. SDS PAGE of the structural proteins of *Pectobacterium* phage CBB. Left lane shows migration patterns of molecular mass colour protein standard broad range (New England Biolabs) with right lanes showing that of structural proteins of phage CBB.

Table 6.5. Details of the proteins identified from ES-MS/MS conducted on capsid proteins of *Pectobacterium* phage CBB.

ORF	Predicted function	No. of unique peptides	sequence coverage %
CBB_29	unknown structural protein	8	29,55
CBB_38	unknown structural protein	9	56,61
CBB_104	ATP-dependent Clp protease proteolytic subunit	1	4,51
CBB_112	major tail protein	4	24,30
CBB_118	unknown structural protein	7	60,40
CBB_144	unknown structural protein	5	34,00
CBB_169	unknown structural protein	3	26,30
CBB_183	unknown structural protein	8	23,60
CBB_201	Neck protein	5	32,60
CBB_205	unknown structural protein	6	20,10
CBB_207	unknown structural protein	9	37,10
CBB_210	unknown structural protein	13	60,50
CBB_213	unknown structural protein	5	18,93
CBB_214	unknown structural protein	1	15,40
CBB_224	unknown structural protein	22	47,80
CBB_225	unknown structural protein	16	68,54
CBB_226	long tail fibre proximal subunit	10	6,06
CBB_227	unknown structural protein	5	32,07
CBB_228	unknown structural protein	13	44,09

CBB_230	unknown structural protein	2	12,70
CBB_231	unknown structural protein	12	53,30
CBB_233	tail sheath stabiliser and completion protein	12	31,61
CBB_234	unknown structural protein	12	35,84
CBB_236	unknown structural protein	21	7,72
CBB_237	baseplate wedge	13	12,47
CBB_240	baseplate hub subunit and tail lysozyme	3	4,11
CBB_241	unknown structural protein	12	17,10
CBB_242	baseplate wedge protein	3	39,40
CBB_243	unknown structural protein	8	46,20
CBB_247	unknown structural protein	1	4,78
CBB_251	unknown structural protein	29	50,14
CBB_254	unknown structural protein	3	15,50
CBB_256	unknown structural protein	4	13,23
CBB_257	precursor of major head subunit	21	68,06
CBB_260	putative tail fibre protein	11	24,04
CBB_270	unknown structural protein	12	91,32
CBB_271	unknown structural protein	9	56,40
CBB_276	unknown structural protein	8	21,70
CBB_286	unknown structural protein	4	29,70
CBB_292	unknown structural protein	1	10,90
CBB_294	unknown structural protein	4	33,10
CBB_298	unknown structural protein	10	35,81
CBB_299	unknown structural protein	2	20,00
CBB_320	unknown structural protein	4	37,20
CBB_322	unknown structural protein	30	43,57
CBB_334	unknown structural protein	7	40,00
CBB_494	unknown structural protein	1	5,81
CBB_496	unknown structural protein	4	17,50
CBB_536	unknown structural protein	8	41,40
CBB_537	unknown structural protein	2	11,00
CBB_548	unknown structural protein	5	25,90
CBB_549	unknown structural protein	6	50,70
CBB_551	unknown structural protein	4	24,10
CBB_552	unknown structural protein	2	9,54
CBB_553	unknown structural protein	7	34,80

6.3. Discussion

A number of jumbo phages like CBB have been reported in recent years (Abbasifar et al., 2014; Buttimer et al., 2016; Drulis-Kawa et al., 2014; Kim et al., 2013; Šimoliūnas et al., 2013; Yuan and Gao, 2016). Among them, phage CBB is significant given the identification of atypical whisker-like structures on its contractile tail, for which a function has not yet been assigned. However, given that a number of proteins found on phage particle surfaces (such as those with Ig-like domains (Fraser et al., 2006) have been proposed to play a role in the interaction between the phage and the host cell surface other than tail fibre proteins, it is not difficult to hypothesise that the whisker-like structures may play a similar role. However, this remains to be proven. The likely reason that these phages have been rarely isolated and reported may be due to the difficulty observed with plaque formation in standard plaque assay overlays, as experienced with phage CBB.

CBB is a member of a group of phages represented by phage Rak2 (the first to have been reported), of which, phages PBEC03, 121Q, Gap32 and K61-1 are members. CBB, like the other Rak2-like phages possesses a number of homologous core proteins found in T4-like phages, a number of which play crucial functions, suggesting that these proteins have not been acquired by horizontal gene transfer as seen with some phages (Petrov et al., 2010). However, given the level of divergence and the lack of a number of core T4-like protein homologues, this relationship appears to be distant with current members of *Tevenvirinae*, as suggested by the phylogenetic analysis of portal vertex and major capsid proteins done in this study. In addition, these phages also possess a large number of proteins that have not been identified in other T4-like phages, as illustrated with total protein comparison of the Rak2-like phages to those of phage T4 and KVP40. The currently available sequence annotation and morphological data for the Rak2-like phages give merit to the proposal of placing them within a new subfamily, as suggested by Abbasifar *et al*. (2014). Furthermore, a shared nucleotide pairwise identity of 65%, a similar arrangement of genes and the position of phages CBB and GAP32 within phylograms (Figure 6.3, 6.4 and Appendix, Figure S6.1) suggest the allocation of their own genus within this subfamily (Adriaenssens and Brister, 2017).

Chapter 7. *Erwinia amylovora* phage vB_EamM_Y3 represents another lineage of hairy *Myoviridea*

Isolation, host range, TEM micrographs of *Erwinia* phage Y3 was conducted by Yannick Born.

Genome sequencing of phage Y3 was outsourced to GATC Biotech, Germany.

7.1. Introduction

The order of *Caudovirales* (the tailed phages) consists of three families of bacteriophage (phage), with about 25% consisting of members of the family *Myoviridae* (Ackermann and Prangishvili, 2012). These members possess virions with contractile tails and include the phages with the largest genomes sizes (Hatfull, 2008). Those with genomes greater than 200 kbp are typically referred to as "jumbo phages"; and currently, there are over ninety in the databases with most having Gram-negative hosts (Hendrix, 2009; Yuan and Gao, 2017). Those with the largest genomes identified to date are *Bacillus* phage G (498 kbp) and those belonging to the Rak2-like group such as *Cronobacter sakazakii* phage vB_CsaM_GAP32 (358 kbp), *Serratia marcescens* phage BF (357 kbp) and *Pectobacterium* phage vB_PcaM_CBB (355 kbp) (Abbasifar et al., 2014; Buttimer et al., 2017a; Casey et al., 2017). *Erwinia* phages are of relevance in the context of alternative methods for bacterial crop disease control. Indeed, recent years have seen an intensive investigation into the use of these phages for biocontrol of contagious fire blight of apples and pears. Such research is motivated by increasing reports of resistance to other antibacterial agents (McManus et al., 2002; Nischwitz and Dhiman, 2013; Tancos et al., 2016). The search for new phages for biocontrol applications has resulted in the identification of a number of jumbo phages such as *Erwinia amylovora* jumbo phages Ea35-70 (accession no. NC_023557.1; (Yagubi et al., 2014)), ɸEaH1 (Meczker et al., 2014), ɸEaH2 (Dömötör et al., 2012), including another sixteen jumbo phages which have had their genome sequences only recently published (Esplin et al., 2017).

Another important bacterial pathogen is *Ralstonia solanacearum*, which colonises the xylem and causes wilt in a wide range of economically important crops. Its relevance has resulted in the discovery of *Ralstonia solanacearum* phage ΦRSL1 (231 kbp, accession no. NC_010811.2), described in by Yamada et al (2010) and found to possess very little homology to known myoviruses. In recent times, a number of other jumbo phages have been identified with low but significant homology to phage ΦRSL1 at the protein level. These are *Pseudomonas* phage Lu11 (280 kbp, accession no. NC_017972.1) and PaBG (258 kbp, accession no. NC_022096.1) as well as phage NTCB (258 kbp, accession no. LT598654.1), whose genome sequence was identified in a

metagenomic study of a decaying *Trichodesmium* bloom, and whose host range remains unknown (Adriaenssens et al., 2012a; Pfreundt et al., 2016; Sykilinda et al., 2014). As well as the recently described RNA *Erwinia amylovora* phage vB_EamM_Yoloswag (259 kbp, accession no. KY448244), which can also be assigned to this group (Esplin et al., 2017).

Here we report a newly isolated jumbo phage, vB_Eam_Y3 of *E. amylovora,* which possesses atypical whisker-like structures along the surface of its contractile tail; and has been found to share homology (at the protein level) to Lu11 and PaBG with distant homology to ΦRSL1.

7.2. Results

7.2.1. Growth parameters, host range and morphology

Erwinia amylovora phage vB_EamM_Y3 was isolated from soil, which had been collected at an apple orchard with trees displaying fire blight symptoms in Suree (canton of Lucerne, Switzerland) in 2008. The bacterial host strain used for its isolation was *Erwinia amylovora* strain 1/79, which had originally been isolated from an apple tree in Germany in 1979 (Falkenstein et al., 1988); and its use as a host strain resulted in tiny clear pin-sized plaques. The phage was found to infect a number of strains of *E. amylovora* as well as strains of *Pantoea agglomerans* and *Pantoea vagans* (Table 7.1). Essentially, its host range was similar to other previously identified *E. amylovora* phages (Born et al., 2011).

Examination of the morphology of Y3 (*n*=10) by transmission electron microscopy showed the phage belonged to the family of *Myoviridae* possessing an A1 morphotype (Ackermann, 2007), with an icosahedral head of 129 nm (±4 nm) in diameter and with a contractile tail with a length of 192 nm (±12 nm) (Figure 7.1). The phage was named in accordance with the bacterial virus nomenclature set out by Kropinski et al (2009). The most unusual feature of Y3 was the hair-like structures along the surface of its contractile tail, a characteristic that has only been observed on a small number of phages to date: these being the jumbo *Pseudomonas* phage Lu11 (Adriaenssens et al., 2012a) which shares protein homology with Y3, and those belonging to jumbo phage Rak2-like group such as *Pectobacterium* phage vB_PccM_CBB (Buttimer et al., 2017a)

and *Escherichia* phage 121Q (accession no. NC_025447.1; (Hua, 2016)), and phages of

unknown genome sizes and lineages such as X particle and *Escherichia* phage PhAPEC6

(Ackermann et al., 1994; Tsones, 2014).

Table 7.1. Host range of bacteriophage vB_EamM_Y3 tested on strains of *Erwinia amylovora* and closely related genera and species, as determined by spot testing with dilution.

Species	Strain	Sensitivity
Erwinia amylovora	CFBP 1430	-
	CFBP 1232	+
	Ea153	+
	1/74	-
	1/79	+
	ACW 38899	+
	ACW 56400	+
	ACW 44274	+
	Ea Rac 3075	+
	ACW 55500	+
	ACW 55835	+
	ACW 55955	+
	IPV 1077/7	+
	01SFR-BO	-
	LA 469	-
	LA 071	+
	LA 411	+
	LA 468	+
	LA 477	+
Erwinia persicina	ACW 40943	-
	ACW 40560x	-
	ACW 41072	-
Erwinia billingiae	Pagg B90	-
Pantoea agglomerans	Eh 42	-
	Em 406	-
	Em 283	+
Pantoea vagans	C9-1	+
Pantoea ananatis	351 Lys	-

Results recorded as +, sensitive; −, no infection.
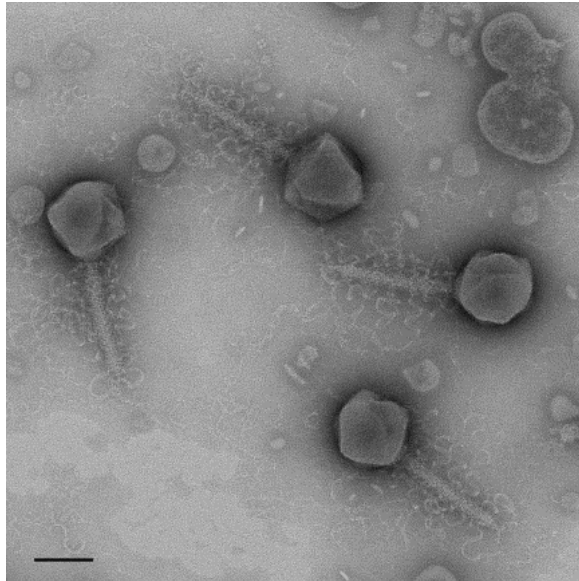
Figure 7.1. Electron micrograph of *Erwinia amylovora* phage Y3. Phage particles were negatively stained with 2% ammonium molybdate. Hair-like structures are evident around the tails. Scale bar represents 100 nm.

7.2.2. General genome features

The genome size of Y3 (261,365 bp) places it among the double-stranded DNA jumbo phages. The assembly pattern of the reads obtained from sequencing indicated that its genome is likely to be circularly permuted with terminal repeats like that of coliphage T4 (Casjens and Gilcrease, 2009). Due to this observation, the start position of its Genbank file was set to match that of *Erwinia* phage Yoloswag, its closest relative available on public databases. The average G+C content of Y3 was found to be 47%, which is lower than that which would typically be found for its host chromosomal DNA at ca.53% (Mann et al., 2013; Powney et al., 2011).

A total of 333 ORFs were identified with putative gene products ranging in size of 39 to 1951 amino acids. These ORFs were largely divided into two major gene clusters (Figure 7.2). The largest region 1 was situated between Y3_75 to 221, with ORFs mostly in the anticlockwise direction and region 2 situated from Y3_001 to 74 and Y3_222 to 333 with ORFs mostly in the clockwise direction. The GC skew correlates well with the ORF orientation of the proposed clusters (Marín and Xia, 2008). The ORFs were found to be tightly organised having little intergenic space, with a number of examples of neighbouring ORFs' start and end points overlapping. Most ORFs were found to start with AUG, while twenty and three ORFs were predicted to start with GUG and UUG,

respectively. Of the gene products of the identified ORFs, seventy could be allocated a possible function (Appendix, Table S7.1) with others identified as putative transmembrane proteins (fourteen) putative lipoproteins (two) and hypothetical proteins (247). No tRNA genes were identified on the genome.

In terms of lysogeny, no integrase, excisionase or repressor genes were identified. However, one gene product (Y3_031) was found to possess the AntA/AntB antirepressor domain (IPR013557), which has been associated with anti-repressor proteins such as that encoded by LO142 of temperate phage 993W (Sandt et al., 2002). These proteins are believed to be involved in the lysis/lysogenic decision-making process of the phage, possibly suggesting a temperate ancestry (Trotter et al., 2006), with a homolog of this protein also evident in Yoloswag.

A potential promoter with a consensus sequence of CTGTAAATA was identified at thirty two locations in the genome of Y3 (Appendix, Figure S7.1, Table S7.2), with a highly similar promoter (ACTGTAAATA[N7]A) also having been predicted in the genome of *Pseudomonas* phage Lu11 (Adriaenssens et al., 2012a) and like that of Lu11, this promotor is mostly situated before ORFs with products predicted to play roles in nucleotide metabolism and DNA replication. In addition, forty-five potential rho factor independent terminators were located throughout the genome of Y3 (Appendix, Table S7.3).

Figure 7.2. Summary of the genomic organisation of the 261,365 bp genome of *Erwinia amylovora* phage Y3. On the outer ring, putative ORFs are represented by purple arrowheads labelled with a predicted function where possible. On the middle ring is shown GC content relative to the mean GC content of the genome. On the inner ring, GC skew is illustrated where green represents positive skew and purple a negative skew.

7.2.3. Phylogenetic position of Y3 within the family *Myovirdea*

At the DNA level, Y3 was found to have little or no homology to phage genomes currently available on public databases, with its closest match being Yoloswag (coverage 6%, identity 75%) using BLASTN analysis. However, during BLASTP analysis, a small number of other jumbo phages were found to share a number of homologs (large terminase, major capsid protein, etc). These were phage NTCB, *Pseudomonas* phages Lu11 and PaBG and *Ralstonia* phage ΦRSL1, which hinted at a possible evolutionary relationship (Figure 7.3). Like Y3 at the DNA level, these five phages (Yoloswag, NCTB, Lu11, PaBG and ΦRSL1) also exhibited no homology to each other indicating that they are at most only distantly related.

To determine the relatedness to each other among above six phages, a phylogenetic analysis was conducted using four shared proteins (portal vertex, DNA polymerase, large terminase and ATP dependent DNA helicase) to generate phylogenetic trees with confident bootstrap values (Figure 7.4). Similar phylogenetic markers have been used in studies of other jumbo phages (Buttimer et al., 2017a; Šimoliūnas et al., 2013; Yuan and Gao, 2017).   These trees showed that among the six, two separate clades were formed: one included Y3 and Yoloswag, and the second was comprised of those phages infecting the genus *Pseudomonas* (Lu11 and PaBG). Phages NCTB and ΦRSL1 were outliers. The NCTB position on the tree is somewhere between the *Erwinia* and *Pseudomonas* phages. Phage ΦRSL1 appeared on a branch that sits outside of all the other five phages implying a far more distant relatedness. This interpretation was further supported by the analysis of the phages proteomes using Coregene (Table 7.2). The *E. amylovora* phages were found to share 225 proteins with each other; the *Pseudomona*s phages shared 136; NTCB shared between 88 and 112 proteins with both the *Erwinia* clade and the *Pseudomonas* clade respectively. ΦRSL1 was still the most distantly related, sharing only 55 to 65 homologues with the other five phages. This distant evolutionary relationship could possibly be explained in the context of host range with the genera *Erwinia* and *Pseudomonas* belonging to the class of *Gammaproteobacteria* while *Ralstonia* belongs to the class *Betaproteobacteria* (Garrity, 2004).

Finally, phylogenetic analysis was performed using the large terminase and portal protein of all jumbo myoviruses of *E. amylovora* that have had their genome sequences submitted to Genbank to date. Both these proteins have been used as phylogenetic markers in such studies of distantly related phages of *Enterobacteriaceae* (Grose and Casjens, 2014). The resulting phylogram shows that there is a minimum of four clades of these phages infecting this bacterium. With Y3, along with phage Yoloswag, forming a novel clade that is highly distinct from the other three clades identified, as the branch that these phages reside on in phylograms sits outside of those representing the other clades (Appendix, Table S7.4 and Figure 7.5).
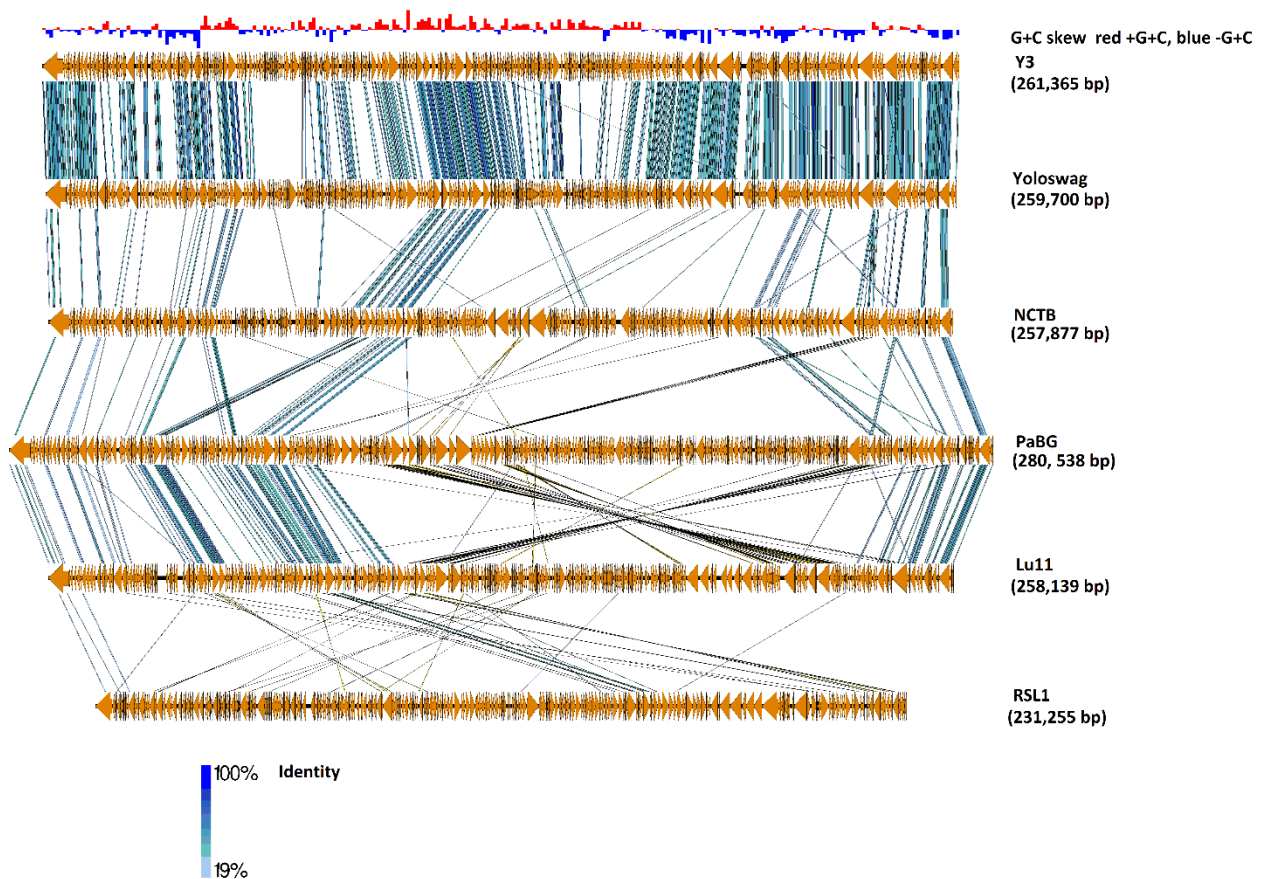
Figure 7.3. Comparison of the genomes of phage Y3 with jumbo phages *Erwinia amylovora* phage Yoloswag, phage NCTB, *Pseudomonas* phages Lu11 and PaBG and *Ralstonia* phage ΦRSL1 using currently available annotations employing TBLASTX and visualised with Easyfig. The blue/red bar chart (top) shows the G+C skew of the Y3 genome. The genome maps comprise of arrows indicating locations and orientation of ORFs. Lines between genome maps indicate the level of homology. To assist in the comparison, the largest gene of Y3 was set to the first position of its map with the corresponding homologous gene being set to the first position for the other phage genomes. Level of amino acid identity is shown via the gradient scale.

173

(A)

Erwinia phage Y3
1Erwinia phage Yoloswag (AQT28493.1)
Phage NCTB (SBV38265.1)
Pseudomonas phage PaBG (YP 008433438.1)
Pseudomonas phage Lu11 (YP 006382721.1)
Ralstonia phage phiRSL1 (YP 001950090.1)

0.10

(B)

Y3 DNA polymerase
Erwinia phage Yoloswag (AQT28647.1)
Phage NCTB (SBV38480.1)
Pseudomonas phage PaBG (YP 008433527.1)
Pseudomonas phage Lu11 (YP 006382627.1)
Ralstonia phage phiRSL1 (YP 001950134.1)

0.20

(C)

Erwinia phage Y3
Erwinia phage Yoloswag (AQT28491.1)
Phage NCTB (SBV38263.1)
Pseudomonas phage Lu11 (YP 006382718.1)
Pseudomonas phage PaBG (YP 008433440.1)
Ralstonia phage phiRSL1 (YP 001950095.1)

0.20

174

Figure 7.4. Phylogenetic analysis using the portal vertex protein (phylogram A), DNA polymerase (B), large terminase (C) and ATP dependent DNA helicase (D) using the six phages Y3, Yoloswag, NCTB, Lu11, PaBG and ΦRSL1. The amino acid sequences were compared using MUSCLE. The tree was constructed using the maximum likelihood algorithm. The percentages of replicate trees were assessed with the bootstrap test (1000).
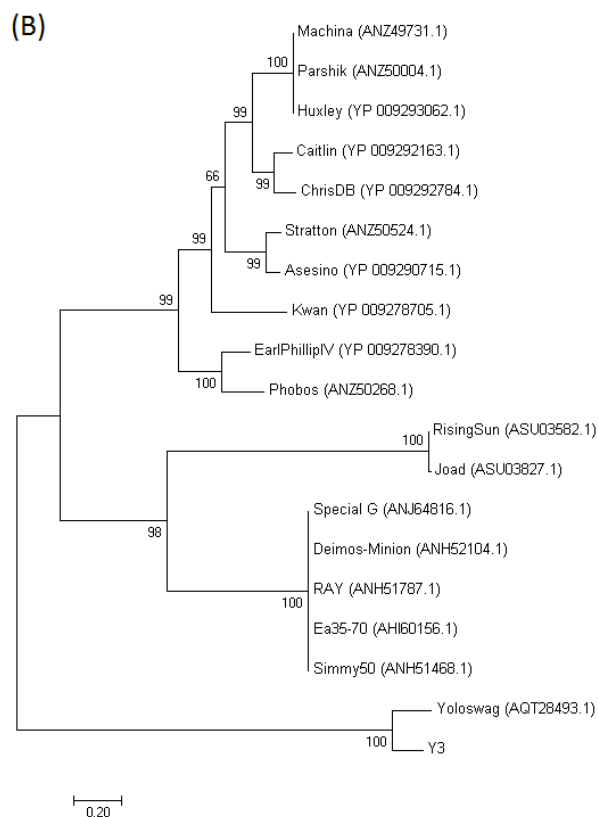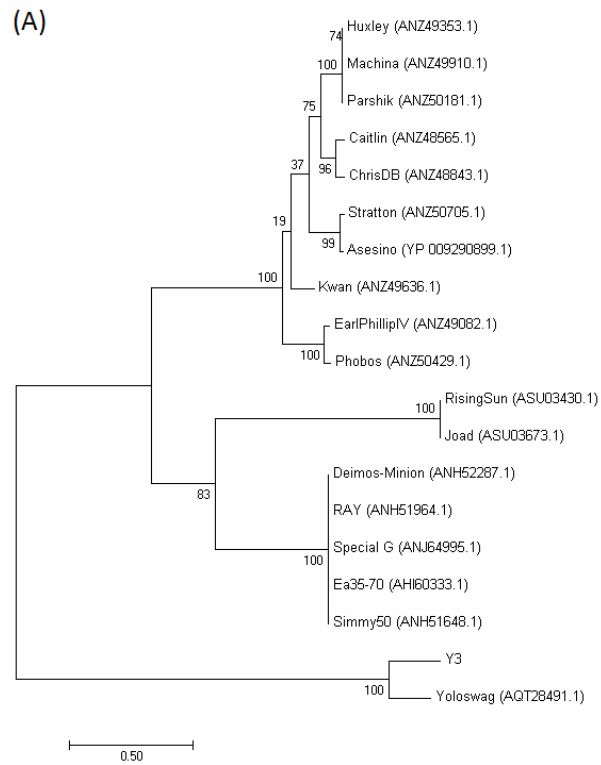
Figure 7.5. Phylogenetic analysis using the large terminase (phylogram A) and portal protein (B) of jumbo *Erwinia amylovora* myoviruses with genome sequences submitted to Genbank to date. The amino acid sequences were compared using MUSCLE. The tree was constructed using the maximum likelihood algorithm. The percentages of replicate trees were assessed with the bootstrap test (1000)

Table 7.2. Homologous proteins shared between the phages related to *Erwinia amylovora* phage Y3 (Yoloswag, NTCB, *Pseudomonas* phages PaBG and Lu11 and *Ralstonia* phage ΦRSL1).

	Y3	Yoloswag	NCTB	PaBG	Lu11	ΦRSL1
Y3 (333)		225 (67.57%)	112 (33.63%)	105 (31.53%)	87 (26.13%)	55 (16.52%)
Yoloswag (334)	225 (67.37%)		113 (33.83%)	102 (30.54%)	87 (26.05%)	53 (15.87%)
NCTB (301)	112 (37.21%)	113 (37.54%)		95 (31.56%)	88 (29.24%)	53 (17.61)
PaBG (308)	105 (34.09%)	102 (33.12%)	95 (30.84%)		136 (44.16%)	65 (21.1%)
Lu11 (391)	87 (22.25%)	87 (22.25%)	88 (22.51)	136 (34.78%)		62 (15.86%)
ΦRSL1 (343)	55 (16.03%)	53 (15.45%)	53 (15.45)	65 (18.95%)	62 (18.08%)	

The number in bracket next phage name is total predict proteins with percentages representing the portion of shared proteins within the proteome of the phage in question. Analyses were conducted using Coregenes with the BLASTP threshold set to 75%.

7.2.4. Morphogenesis proteins

The majority of the putative structural proteins of Y3 were situated in a cluster with a length of 65 kbp (Y3_104 -158) within region 1, with the majority of these shared with phages Yolosway, NTCB, PaBG and Lu11. These putative proteins were the tail tube (Y3_132, 134), the tail sheath (Y3_135, 158), the baseplate (113, 114, 124), the tail fibres (Y3_104, 108, 109, 110), the tail-collar fibre protein (Y3_110) and the major capsid protein (Y3_139). A putative portal protein (Y3_006) was located outside of this region, being adjacent to the large terminase (Y3_004). Interestingly, the homology of the portal protein to that of *Bacillus* phage G (gp14, 5E-8, 23%) was found to be high when using BLASTP. Proteins Y3_106 and Y3_143 also shared strong homology with gp431 (2e-9, 28%) and gp19 (3e-09, 33%) respectively of *Bacillus* phage G. Y3_106 is situated between ORFs predicted to form the tail fibres that possess collagen domains (Ghosh et al., 2012). This gene product was predicted to be a tail fibre hinge protein due to the presence of an ILEI domain (PF15711.4, 1.7E-7). This domain belongs to the FAM3 superfamily to which the T4-like phage tail hinge gp35 protein is distantly related (Guo et al., 2006). Three of the structural proteins that have been identified in phage ΦRSL1 (ΦRSL1_135, 136, 140) as confirmed by N-terminal sequencing (Yamada

et al., 2010), were found to have homologues in phage Y3 (namely Y3_135, 139, 140, 158) providing further evidence that they play a role in the virion structure of Y3.

7.2.5. Nucleotide metabolism, DNA replication and modification

Typical of large myoviruses, Y3 has a number of proteins with functions predicted to influence the nucleotide pool within its host (Hendrix, 2009). These include a CMP deaminase (Y3_035) allowing the hydrolysis of dCMP to dUMP and a putative dUTPase (Y3_300) allowing the conversion of dUTP to dUMP. The dUMP can then be acted upon by a thymidylate synthase (Y3_045), which can catalyse the methylation of dUMP to dTMP, which a dTMP kinase (Y3_118) can, in turn, convert to dTDP. Another feature typical of large phages is the possession of a number of proteins involved in DNA replication. It is worth mentioning that smaller phages typically use a lesser number of proteins to recruit the host replication machinery to carry out this role (Hendrix, 2002). Phage Y3 was identified to possess helicases (Y3_001, 168, 176), primases (Y3_024, 227), a DNA ligase (Y3_148), topoisomerase subunits (Y3_327, 329), as well as class 1 polymerases (Y3_011, 173) and the delta component (PF13177.5, 5.7E-26) of a class III polymerase (Y3_209). Y3 also has proteins involved in DNA repair and maintenance including RecA (Y3_243), UV damage endonuclease UvsE (Y3_287) and MmcB-like DNA repair protein (Y3_119) as well as those involved in DNA recombination (Y3_186) and DNA degradation (Y3_229). DNA methylation is a well-known strategy used by some phages to protect against host restriction endonucleases (Samson et al., 2013) and accordingly, Y3 appears to be able to methylate both cytosine and adenine (Y3_29, 50).

7.2.6. Other metabolic functions of Y3

While the genome of phage Y3 does not possess tRNA genes, it does possess proteins potentially involved in their maturation such as RNA 2'-phosphotransferase (Y3_272) and a carboxy-S-adenosyl-L-methionine synthase (Y3_017). Curiously, the phage was found to possess an asparagine synthase (Y3_020), which would allow the conversion of aspartate to asparagine. Such auxiliary phage genes that produce proteins that compliment a host's metabolism have been identified in a number phages of marine environments; and the role of their encoded proteins are understood to be the removal of potential bottlenecks in their host metabolism that could potentially hinder the lytic cycle (Hurwitz and U'Ren, 2016).

7.2.7. Selfish genetic elements

Inteins are selfish genetic elements that self-cleave and remove themselves from a translated protein (Tori and Perler, 2011). Proteins identified with these elements in phage Y3 are the large terminase (Y3_004), C-5 cytosine methyltransferase (Y3_050) and one of the four predicted tail fibre proteins (Y3_110). Also identified was a homing endonuclease of the HNH family (Y3_179), situated between two hypothetical proteins. These homing endonucleases are mobile genetic elements with endonuclease activity that only promote the spread of their own encoding gene (Edgell et al., 2010).

7.2.8. Cell wall degrading enzymes

Analysis of the protein Y3_301 showed that it possesses an N-terminal transmembrane domain (12-34 aa) with a high alanine content (23%), with one basic residue (arginine), and a C-terminal soluble lytic transglycosylase (SLT) domain (IPR008258; 108-218 aa). The latter domain degrades peptidoglycan by cleavage of the β-1,4 glycosidic bond that results in the formation of a 1,6-anhydrobond in the muramic acid residue. A homolog of this protein was also found in phage Yoloswag (YOLOSWAG_297). Due to the composition of the domains in this protein, it was suspected to be an endolysin that followed the signal arrest (SAR) system like gp45 of φKMV and lyz of phage P1 (Briers et al., 2011; Xu et al., 2004). Analysis of the protein using SignalP 3.0, showed that it possessed an N-terminal signal sequence with a probability of 0.729. However, the more recent version of SignalP did not identify this domain suggesting that the N-terminal secretory signal is not removed after export, but rather remains tethered to the inner membrane after translocation by the host *sec* translocon system.

For a number of these types of endolysin, an inducible system with *E. coli* permits demonstration of SAR endolysin activity using a detectable reduction in culture turbidity upon induction (Briers et al., 2011; Lynch et al., 2013). A similar experiment was conducted here for protein Y3_301. Accordingly, gene *Y3_301* was cloned into pET28a and transformed into *E coli* Lemo21 (DE3). Expression of the cloned gene was lethal to cells, as the addition of IPTG after 4 hours of growth caused the OD_{600} to drop from ~0.480 to ~0.320 at 8 hours (Figure 7.6). No lytic activity was detected after the removal of the signal sequence as seen with the induction of the truncated

transmembrane domain Y3_301 (rTM-Y3_301). As the toxic activity of this protein was found to be dependent on the presence of its transmembrane domain, it could be concluded that this protein behaves as a typical SAR endolysin, requiring the transmembrane domain to act as a signal sequence for its translocation enabling the catalytic domain of the protein to reach the periplasm where it acts on the peptidoglycan of the bacterial cell resulting in lysis.

An important feature of SAR endolysins is the regulation of their enzymatic activity while in their membrane-tethered form. The two main mechanisms that have been identified for those with the lysozyme domain are (a) those which utilise steric hindrance of catalytic residues due to the proximity of the bilayer to which the enzyme is embedded (as seen with R_{21} of lambdoid phage 21) and (b) those which regulate activity by the formation of disulphide bridges as seen with Lyz_{P1} of coliphage P1 (Kuty et al., 2010). Little is understood of how SAR endolysins with transglycosylase domains regulate their activity. However, the formation of disulphide bridges is a possible explanation of how Y3_301 might regulate itself. Five cysteine residues were identified in its amino acid sequence, one situated within the transmembrane domain (Cys23) and the four-remaining located within its catalytic domain (Cys127, 147, 182 & 188).

A second protein, namely Y3_197 was found to possess a soluble lytic transglycosylase (SLT) domain (PF05838), with a peptidoglycan binding domain (IPR018537). A homolog of this protein was also found in phage Yoloswag (YOLOSWAG_189). To date, we have not succeeded in demonstrating its predicted enzymatic activity after cloning in *E. coli* (data not shown).
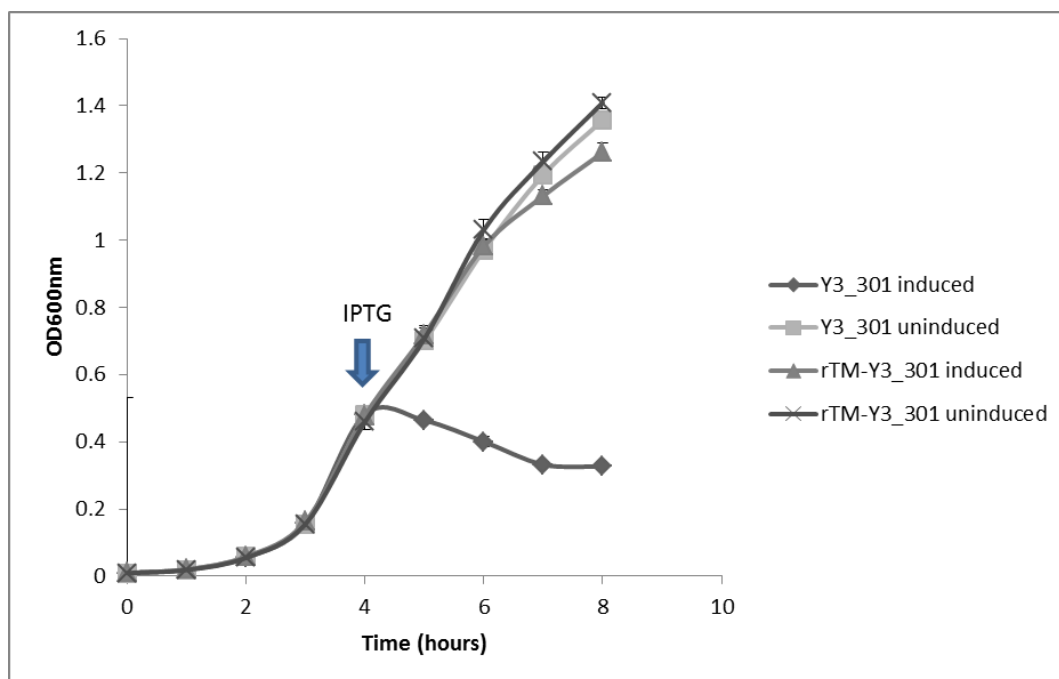
Figure 7.6. Growth curve of *E. coli* Lemo21 (DE3) carrying pET28a with the SAR endolysin Y3_301 and its truncated derivative without the transmembrane domain (rTM-Y3_301). The culture was induced at OD_{600} ~0.450 with 400 µM IPTG (indicated with an arrow). All OD measurements were taken in triplicate and the averages are represented.

7.3. Discussion

As mentioned above, a number of hairy *Myoviridae* jumbo phages to date have been placed in the Rak2-like group. The possession of whisker-like structures is shared with the unrelated *E. amylovora* phage Y3, suggesting that these may be more widespread in the *Myoviridae* family than previously thought. For the Rak2-like *Escherichia* phage 121Q, this feature has been linked to a protein which resembles a phage tail fibre (Hua, 2016).   The exact gene products responsible for the atypical whisker-like structures of Y3 is yet unknown. However, a few genes have been predicted to encode tail fibre structures (Y3_104, 106, 108, 109, 110), and any of these could potentially be responsible. The role of these atypical hairs of these phages is still unknown but it has been speculated to play a role in enhancing phage ability to detect and infect its host by facilitating the phage adsorption processes (Buttimer et al., 2017a; Hua, 2016).

Phylogenetic studies show that phages Y3 and Yoloswag establish a novel clade of jumbo phages infecting *E. amylovora* (Figure 7.5). In the context of shared homologous proteins, Lavigne et al (2009) defined phage subfamily and genus boundaries at ≥20%

and ≥40%, respectively. Based on these cut-off values, it would place phages Y3 and Yoloswag within a genus within a subfamily to which NTCB, PaBG and Lu11 are also members and which is peripherally related to ΦRSL1 (Table 7.2).

Little has been described about SAR endolysins in jumbo phages to date. However, a small number of jumbo phages have been predicted to possess them. These include members of the Rak2-like group such as *Klebsiella* phage vB_KleM-RaK2 (gp506), *Pectobacterium* phage CBB (CBB_187), *Cronobacter* phage GAP32 (gp180) and coliphage Q121 (gp532) (Abbasifar et al., 2014; Buttimer et al., 2016); and upon examination within our study, *Erwinia* phages Ea35-70 (Ea357_029), Deimos-Minion (accession no. ANH52127.1; DM_29) and Simmy50 (accession no. KU886223.1; gp_30). Phages Y3 and Yoloswag can be included in this group. These endolysins are known to be highly prevalent in phages that infect *Enterobacteriaceae* (Oliveira et al., 2013); and this may also be the case for their jumbo phages, as a number of such endolysins have already been identified among them.

Thesis summary

Note, Table 8.1 below condenses the information in this summary.

Pectobacterium atrosepticum and Pectobacterium carotovorum subsp. *carotovorum* are species that belong to a group of bacteria, collectively known as the soft rot *Enterobacteriaceae* (SRE). They harm production of a number of ornamental plants and arable crops. Both *P. atrosepticum* and *P. carotovorum* subsp. *carotovorum* impact potato production causing potato blackleg disease and tuber soft rot (Mansfield et al., 2012; Pérombelon, 2002). This thesis describes recently isolated bacteriophages of three distinct phage genera that infect *P. atrosepticum*, namely *Cbunavirus*, *Phimunavirus* and *Cr3virus*. Both *Cbunavirus* and *Phimunavirus* are newly established genera that were identified and formalised in the work of this thesis and for which taxonomy proposals have been submitted to the International Committee for the Taxonomy of Viruses (ICTV). This taxonomic classification was based on shared nucleotide identity values of ≤95% but ≥70% (using the BLASTN algorithm), genome synteny and their position within phylograms based on conserved proteins and/or total proteomes (TBLASTX) (Adriaenssens and Brister, 2017). Additionally, papers were also published describing these phages of *Cbunavirus* and *Phimunavirus*. The thesis also describes two new jumbo phages (phages CBB and Y3). The first was isolated against *P. carotovorum* subsp. *carotovorum* and was found to be closely related (at the protein level) to a group of jumbo phages that infect genera belonging to the family of *Enterobacteriaceae,* collectively named after the first described representative of this group, *Klebsiella* phage Rak2 (345 kbp, accession no. NC_019526). The second jumbo phage studied in the thesis infects the phytopathogen *Erwinia amylovora*, the causative agent of fire blight, a destructive disease that impacts pear and apple trees (Bonn and van der Zwet, 2000; Mansfield et al., 2012). This latter phage was found to be related at the protein level to a group of jumbo phages of *Erwinia* and *Pseudomonas*, all distantly related to *Ralstonia* phage ΦRSL1 (231 kbp, accession no. NC_017972) and collectively named after this phage. The significance of both jumbo phages is discussed in more detail later in this summary, although it's worthy of mention at this point that each of them merited a published paper.

The phages belonging to the genera of *Cbunavirus*, *Phimunavirus* and *Cr3virus* were isolated from soil obtained from potato grading machinery from two commercial potato farmers in

Co. Cork, Ireland in 2013. Eleven isolates were obtained and subjected to genome restriction digestion analysis using BamHI, SspI and BglII (Appendix, Figure S5.1 and Figures 3.1, 5.1), nine of these isolates were found to produce similar banding profiles. Three of these nine isolates, namely *Pectobacterium* phages vB_PatP_ CB1 (phage CB1), vB_PatP_CB3 (phage CB3) and vB_PatP_CB4 (phage CB4), being closely related were collectively referred to as the CB1-like phages and were analysed in detail. This analysis established phage CB1 as the type phage of the genus of *Cbunavirus*. The two remaining phage types isolated were *Pectobacterium* phage vB_PatP_ CB5 (phage CB5), a phage of the newly described genus of *Phimunavirus*; and *Pectobacterium* phage vB_PatM_CB7 (phage CB7), a phage of the genus *Cr3virus*.

TEM analysis showed all five phages (CB1, CB3, CB4, CB5 and CB7) were of the order *Caudovirales*, possessing virions with tails (Ackermann, 2001) (Figures 3.2, 4.2, 5.5). The CB1-like phages and phage CB5 were identified to possess a morphology associated with the phage family *Podoviridae*, possessing icosahedral capsids with diameters of 67.8 ± 4.4 – 71.8 ± 1.7 nm and 63.1±3.6 nm, with a short non-contractile tail with a length of 23.9 ± 1.7 – 25.8 nm and 13.1 ± 1.8 nm, respectively. The dimensions of these are similar to related phages in the scientific literature (Eriksson et al., 2015; Lavigne et al., 2003; Wittmann et al., 2015). Phage CB7 was identified to have the morphology associated with the phage family *Myoviridae*. This phage is larger than the above-mentioned *Podoviridae* (CB1, CB3, CB4 and CB5) possessing an icosahedral capsid with a diameter of 84.05 ± 3.88 nm, with a contractile tail with a length of 123.39 ± 2.63 nm. A similar tail length was reported for *Cr3virus* member *Pectobacterium* phage ɸTE, although the capsid dimensions of CB7 are somewhat smaller than that of ɸTE at 98.4 ± 3 nm (Blower et al., 2012).

Given the biocontrol potential of bacteriophages, attention was devoted to defining their host ranges on representative phytopathogenic bacterial strains. The above five phages were all isolated using two strains of *P. atrosepticum*, namely, stain DSM 18077 (for phage CB1) and strain DSM 30186 (for phages CB3, CB4, CB5 and CB7). *P. atrosepticum* strain DSM 18077 is the type strain of this bacterial species. Host range examination against different species within SRE, showed that their infection was limited to *P. atrosepticum* with the CB1-like phages having the largest host range within this species, each being able to infect a minimum of 10 strains of *P. atrosepticum* while phages CB5 and CB7 could infect 3 and 5

strains, respectively (Tables 3.1, 4.1., 5.1). This information is important for potential inclusion of the phages in future bacterial phytopathogen control measures.

Further analysis of these five phages included the examination of the latent periods and phage burst sizes on their respective host strains under standard conditions (LB medium, incubation temperature of 25 °C). The CB1-like phages have the longest latent period at 60-65 min, CB7 had an intermediate length latent period at 55 min and CB5 had the shortest at 45 min. There was also a similar pattern with burst size among the phages, with CB1-like phages having the largest burst size of 158-246 PFU/cell, phage CB7 an intermediate burst size of 154 PFU/cell and phage CB5 the smallest at 43 PFU/cell (Appendix, Figure S3.1 and Figure 4.1, 5.2). Interestingly, phage CB5 was observed to be difficult to cultivate and propagate under laboratory conditions, and its low burst size was deemed to be likely the reason why it was never possible to obtain a lysate with a titre exceeding $1x10^9$ PFU/mL regardless of whether harvesting phage from double agar overlay plates with SM buffer or standard propagating in broth. The overall data obtained was broadly similar to that reported for other SRE phages (Adriaenssens et al., 2012b).

Phage virion stability should be taken into consideration when creating phage preparations for biocontrol applications, as it is important to understand factors that could negatively impact phage titre during storage but also upon application. Stability was thus established for the CB1-like phages and CB7. Over an incubation period of one-hour, the CB1-like phages were found to have no significant titre losses between -18 to 50 °C, while phage CB7 appeared to exhibit slightly greater thermal resistance, but again with no significant titre loss between -18 to 55 °C (Appendix, Figure S3.2 and Figure 5.3). Regarding pH, incubation for a 24-hour period showed that for the CB1-like phages and phage CB7 no significant titre losses occurred between pH values of 5 to 11 (Appendix, Figure S3.3 and Figure 5.4). Phage CB5, on the other hand, was found to be sensitive to the universal buffer used for pH stability experiments. When the phage was incubated in each of the separate components of this buffer, significate titre losses were found to occur in 10 mM trisodium citrate (data not shown). Phage viability loss due to sodium citrate has been previously reported (Kuo et al., 1971).

Genome sequencing was conducted on the five phages CB1, CB3, CB4, CB5 and CB7. Among these, the largest genome with a size of 142,778 kbp was obtained for CB7, an intermediate

size genome was obtained for each of the CB1-like phages between 75,394 - 75,973 kbp, and the smallest genome being obtained for phage CB5 at 44,262 kbp. Examination of genome sequence reads for both the CB1-like phages and phage CB5 revealed a location with a doubling of sequence reads compared to the average across their entire genome, indicating the presence of direct terminal repeats (DTRs). DTRs with a size of 647-648 bp and 287 bp were determined for the CB1-like phages and phage CB5 respectively, hinting their genomes to be linear. These genome properties are similar to those reported of related phages (Blower et al., 2017; Ohmori et al., 1988). Examination of the genome of phage CB7 did not reveal this feature, indicating that its genome is likely circular permutated. This conclusion was also indicated by a Bal-31 exonuclease time course treatment of the CB7 genome followed by restriction digestion with BglII, showing no bands to be preferentially degraded upon examination by agarose gel electrophoresis (Figure 5.6). Additionally, related phages of phage CB7 (those of *Cr3virus*) are also believed to have circularly permutated genomes (Blower et al., 2012). The largest number of open reading frames (ORFs) were identified for phage CB7 with 253 ORFs, an intermediate number identified for the CB1-like phages between 97 - 102 ORFs, with the smallest number identified for phage CB5 with 60 ORFs. Genes for tRNAs were only identified for phages CB7 and CB4, each having one and two genes, respectively. All five phages were found to have no genes associated with lysogeny, indicating these phages follow a fully lytic lifestyle. Additionally, genes associated with virulence factors or known toxins were also not identified on their genomes.

Of all the phages identified to infect *P. atrosepticum* in the research of this thesis, the CB1-like phages were identified to have a number of the desirable features for biocontrol applications against this phytopathogen. Such as good biostability, ease at which these phages could be prepared in the lab (phage lysates with titres exceeding 1×10^9 PFU/mL could be routinely prepared), lack of undesirable genes (those for lysogeny and virulence factors) and the possession of a wide host range against *P. atrosepticum,* collectively being able to infect 15 of the 19 (79%) tested strains of the species. Thus, the disease control capability of a phage mixture of CB1, CB3 and CB4 was investigated by tuber rot assays. At a multiplicity of infection (MOI) of 1, these phages were found to suppress soft rot formation on tubers of the *Rooster* potato variety by a mixed infection using *P. atrosepticum* strains

DSM 18077 and DSM 30186 (Figure 3.7). However, further testing is warranted to investigate the biocontrol potential of these phages in greater detail. Such as an investigation of the minimum MOI which is required for these phages to inhibit disease formation. To investigate if these phages can inhibit soft rot formation caused by *P. atrosepticum* strains most relevant to the Irish potato industry, but additionally strains of this bacterium relevant to the potato industry in other parts of the world. It should also be investigated if these phages can suppress the formation of blackleg of potato crop in the field, with further testing of phage biostability to factors such as UV, agricultural chemicals (such as copper compounds and fertiliser) as well stability in soil; as such factors are known to affect phage stability (Iriarte et al., 2007; Straub et al., 1992; Sykes et al., 1981)

The first phage of this thesis summary mentioned two jumbo phages. The first, *Pectobacterium* phage vB_PcaM_CBB (phage CBB) was isolated from activated sludge obtained from a wastewater treatment plant in Little Island, Co. Cork, Ireland in 2015. The bacterial host used was *P. carotovorum* subsp. *carotovorum* strain CB BL19-1-37 which had previously been isolated from a potato plant showing signs of blackleg from a commercial potato farm in Co. Cork, Ireland in 2013. Examination of this phage by TEM analysis showed that it belongs to the family *Myoviridae*, possessing a large capsid with a width of 128.0 ± 6.2 nm and a relatively short contractile tail with a length of 123.0 ± 2.6 nm (Figure 6.1). The length of the tail of this phage is comparable to phage CB7, but its capsid is significantly larger. Furthermore, this analysis revealed that this phage possesses atypical hair-like appendages on the surface of its contractile tail. Host range analysis showed the phage is not only able to infect its host species of *P. carotovorum* subsp. *carotovorum* but also *P. atrosepticum*, *Cronobacter muytjensii*, *Cronobacter malonaticus* and *Erwinia mallotivora* (Table 6.1). However (and explaining why this phage was not included in the above-mentioned biocontrol study), the phage was found to be quite difficult to propagate under laboratory conditions. To allow the phage to create reproducible plaques using the double overlay method for enumeration, a dilute LB overlay method using 0.2% (w/v) agarose was developed and employed. Additionally, the phage could not be propagated in broth but only by harvesting phage from double agar overlay plates with SM buffer. Genome sequencing revealed phage CBB to have an atypically large genome of 355,922 kbp in size and is suspected to be linear with fixed ends with DTRs of 22,456 kbp. Furthermore, it has one of

the largest phage genomes sequenced to date. The largest described in the literature are those of *Bacillus* phage G (498 kbp, accession no. JN638751.1) and *Cronobacter* phage GAP32 (358 kbp, accession no. NC_019401), with phage CBB being related to the former. A phage with a genome greater than 200 kbp can be termed as a jumbo phage. The genome of phage CBB was identified to contain 554 ORFs (not taking into consideration its DTRs) with 33 tRNA genes. These values are significantly larger than those of any of the *P. atrosepticum* specifically infecting phages. It is tempting to speculate if a large number of genes encoding proteins and tRNAs of phage CBB contributes to its wide host range.

The second jumbo phage described in this thesis is *Erwinia* phage vB_EamM_Y3 (phage Y3). This phage was isolated from soil from an apple orchard with trees displaying symptoms of fire blight in the canton of Lucerne in Switzerland in 2008. The TEM analysis of the phage showed it belongs to the family of *Myoviridae* (Figure 7.1), possessing a large capsid with a width of 129 ± 4 nm and a contractile tail with a length of 192 ± 12 nm. The capsid diameters of this phage are similar to that of phage CBB, but its tail is significantly longer. Furthermore, as seen with phage CBB, the analysis also showed this phage to possess atypical hair-like appendages on the surface of its contractile tail. However, the hairs appeared to be much thicker than those of phage CBB. Like phage CBB, this phage is also capable of infecting outside of its genus *E. amylovora*, in that it could also lyse *Panthoea agglomerans* and *Panthoea vagans* (Table 7.1). Genome sequencing of this phage revealed a genome of 261,365 kbp, but analysis of sequence reads failed to identify DTRs as determined for CB1-like phages, phage CB5 and phage CBB. It is believed its genome is circularly permutated like that of phage CB7. The genome was identified to contain 333 ORFs with no tRNA genes. Furthermore, phylogenetic studies with this phage and *Erwinia* phage Yoloswag (259 kbp, accession no. KY448244) showed that these phages form a newly identified clade of jumbo phages infecting *E. amylovora* (Figure 7.5).

Detailed molecular analysis of all of the phages in this thesis revealed a variety of interesting properties. For example, selfish genetic elements were identified in the genomes of all phages examined in this thesis. The genomes of phages CB5 and CBB were identified to contain three HNH endonucleases with phage Y3 found to contain five and CB1-like have between four and six. However none of these contained inteins. In comparison to these phages, phage CB7 was determined to contain a significantly greater number of

endonucleases: twenty-one HNH and two LADLIDADG (Table 5.3). Five of the HNH endonucleases form introns that divide the CB7 genes encoding the helicase, ribonucleotide-diphosphate reductase, nicotinamide phosphoribosyltransferase and DNA polymerase. These, apart from the HNH associated intron of the nicotinamide phosphoribosyl transferase gene, were demonstrated to be splicing competent by utilising reverse transcriptase PCR (Figure 5.14). Furthermore, phage CB7 was found to have three inteins, two of which contain the LAGLIDADG homing endonucleases (Table 5.3). This number of endonucleases in a phage genome is one of the largest described to date. Other phages that have been reported to have many endonucleases are *Escherichia* phage T4 with fifteen, and *Dickeya* phage Limestone with fourteen (Adriaenssens et al., 2012b).

Mass spectrometry (ESI-MS/MS) was also conducted on the virions of phages CB1, CB7 and CBB. This analysis revealed that the virions of these three phages are formed from a minimum of ten, twenty-six and fifty-five proteins, respectively (Table 3.3, 5.4, 6.5). It is interesting to observe that the genome size of these phages appears to correlate with the complexity of the structure of their virions.

Three of all the phage types, namely those represented by CB5, CBB and Y3 have signal arrest release (SAR) endolysins (ORFs CB5_59, CBB_187 and Y3_301). They have an N terminal transmembrane domain on their respective endolysins enabling the host *sec* system to transport the protein to the inner membrane (Briers et al., 2011; Xu et al., 2004). Interestingly, ORF CBB_187 of phage CBB is predicted to also have an intermediate cell wall binding domain. Both of the endolysins of CB5 and CBB are predicted to have lysozyme activity, while that of phage Y3 is unusual in that it is predicted to have lytic transglycosylase activity. It was possible to demonstrate the SAR activity of both CBB_187 and Y3_301 in the laboratory by cloning the respective ORFs into *Escherichia coli* and overexpressing them, resulting in cell lysis (Figures 6.5, 7.6).

Similar to phages CB5 and CBB, the endolysin of the CB1-like phages (CB1_60, CB3_64, CB4_63) is also predicted to have lysozyme activity. However, their protein lacks the N-terminal transmembrane domain, thus they are not likely to have SAR activity. As for phage CB7, the exact protein that functions as the endolysin remains to be identified. The top two endolysin candidates here are CB7_83, which resembles SleB (an N-acetylmuramyl-L-alanine amidase), a protein responsible for hydrolysis of the spore cortex in *Bacillus subtilis*; and

CB7_190 which is predicted to be a peptidase with HHpred analysis showing homology to the endolysin of *Escherichia* phage T5.

Table 8.1. Feature comparison of the bacteriophages described in this thesis.

Features	CB1-like (*Pectobacterium* phages CB1, CB3, CB4)	*Pectobacterium* phage CB5	*Pectobacterium* phage CB7	*Pectobacterium* phage CBB	*Erwinia* phage Y3
Host range	Pa (infects a minimum of 10 strains each)	Pa (infects a minimum of 3 strains)	Pa (infects a minimum of 5 strains)	Pa, Pcc, Cmu, Cma, Em	Ea, Pa, Pv
Taxonomy	*Podoviridae, CBunavirus*	*Podoviridae, Autographivirinae, Phiunavirus*	*Myoviridae, Vequintavirinae, Cr3virus*	*Myoviridae,* 'Rak2-like'	*Myoviridae,* 'ΦRSL1-like'
Latent period/burst size	60-65 min/ 158-246 PFU/cell	45 min/ 43 PFU/cell	55 min/ 154 PFU/cell	-	-
Genome size	75, 394 – 75,973 kbp	44, 262 kbp	142, 778 kbp	355, 922 kbp	261,365 kbp
DTRs	647-648 bp	287 bp	-	22,456 bp	-
ORFs no. (excluding DTRs)	97-102	60	253	554	333
tRNA gene no.	2 identified for phage CB4	0	1	33	0
Homing endonucleases	3-6 (HNH family)	5 (HNH family)	23 (HNH and LAGLIDADG family)	3 (HNH family)	3 (HNH family)
Endolysin	Lysozyme	Lysozyme (SAR)	SleB or Peptidase	Lysozyme (SAR)	Soluble lytic transglycosylase (SAR)
Minimum no. of proteins determined to form phage virion by ESI-MS/MS	10 (determined using phage CB1)	-	26	55	-
Biocontrol potential	Yes	Unlikely	Unlikely	unlikely	unlikely

Bacteria recorded as Pa, *P. atrosepticum*; Pcc, *P. carotovorum* subsp. *carotovorum*; Cmu, *C. muytjensii*; Cma, *C. malonaticus*; Em, *E. malotivora*; Ea, *E. amylovora*; Pa, *P. agglomerans*; Pv, *P. vagans.*

A number of proteins identified among the phages described in this thesis merited further investigation due to their predicted enzymatic activity or function within the respective phage lifecycle (Table 8.2).

Several peptidoglycan degrading proteins were identified among the phages and these have potential applications as antibacterial agents (Rodríguez-Rubio et al., 2013; Yang et al., 2014). Examples are those previously described of in phage CB7, namely proteins CB7_83 and CB7_190. There is also a putative endolysin in phage CB5 (namely ORF CB_51), a predicted SAR endolysin with lysozyme activity. SAR endolysins have been identified in other phages of genera belonging to the subfamily of *Autographivirinae, s*uch as the SAR endolysin KMV45 of *Pseudomonas* phage ɸKMV (Briers et al., 2011)*.* It would be interesting to identify if the SAR endolysin is present in the genus of *Phimunavirus*. SAR endolysin activity has been confirmed with gp61 of *Escherichia* phage N4 (Stojković and Rothman-Denes, 2007). However, for the related CB1-like phages (such as CB1_60 of *Pectobacterium* phage CB1) with predicted lysozyme activity, their endolysin does not have the predicted N-terminal transmembrane domain. It would be interesting in a future project to confirm that such SAR activity is not associated with phages of the genus of *Cbunavirus*. Furthermore, CB5_51 of phage CB5 was predicted (by HHpred analysis) to be a virion-associated peptidoglycan degrading protein with lysozyme activity. Such proteins have also been identified among *Autographivirinae* members *Pseudomonas* phage ɸKMV and *Escherichia* phage T7 (Briers et al., 2006; Moak and Molineux, 2000).

Proteins with a putative role in cell surface receptor binding were also identified. Such proteins can potentially be exploited for the detection of problematic bacteria (Javed et al., 2013). Examples of these are the putative tail fibres of phage CB7 (CB7_10, 36) and the putative tail spike proteins CB1_61 and CB5_60 of phages CB1 and CB5, respectively. The latter proteins are suspected to degrade or modify surface polysaccharide on recognition of host cells. ORF CB5_60 encodes a P22 tail spike domain (IPR015331), derived from the tail spike of *Salmonella* phage P22 shown to break down saccharides on binding to host cell surface (Andres et al., 2010). In addition, ORF CB1_62 possesses a GDSL lipase/esterase (IPR001087), suggesting the protein may be like that of N4-like *Escherichia* phage G7C, as the tail spike of this

phage (gp63.1) shares this domain and has been shown to deacetylate host surface polysaccharides (Prokhorov et al., 2017). Such proteins have been shown to modify host cell surfaces altering their susceptibility to phage infection. Indeed, the tail associated depolymerase of *Erwinia* phage L1 was shown to be capable of degrading saccharides associated with the EPS of *E. amylovora* enabling infection by *Erwinia* phage Y2 (Born et al., 2013).

Overall, the thesis has conducted a detailed analysis of seven new and interesting phages that infect the soft-rot *Enterobacteriaceae* (SRE). At a scientific level, the study has led to new taxonomic designations with several interesting molecular aspects being identified in one or other of the phages. The various characteristics were compared among the phages in the thesis and also compared with phages already in the public domain. Properties of relevance to potential biocontrol applications of the phages were also described. Given their scientific interest and application potential, all of the phages were deposited in DSMZ (Deutsche Sammlung von Mikroorganismen und Zellkulturen - German Collection of Microorganisms and Cell Cultures GmbH), one of the top culture collections in the world.

Table 8.1. List of identified proteins on the of the bacteriophages described in this thesis that merit further investigation.

Phage	Protein locus	Molecular mass (kDa)	Amino acids	Description
Pectobacterium phage CB1	CB1_60	17.4	158	Lysozyme
Pectobacterium phage CB1	CB1_61	105	976	tail spike (putative SGNH esterase - hydrolase)
Pectobacterium phage CB5	CB5_51	97.9	904	internal virion protein B (putative lysozyme)
Pectobacterium phage CB5	CB5_59	21.9	201	putative SAR (lysozyme)
Pectobacterium phage CB5	CB5_60	61.9	585	tail spike (putative pectate lysate)
Pectobacterium phage CB7	CB7_10	111.9	1099	putative fibre protein
Pectobacterium phage CB7	CB7_36	81.9	793	putative fibre protein
Pectobacterium phage CB7	CB7_83	19.9	176	SleB-like protein
Pectobacterium phage CB7	CB7_190	14.5	127	putative L-alanyl-D-glutamate peptidase

Appendix 1

Table S2.1. Bacteria strains used in the isolation and testing of host range of *Pectobacterium* phages CB1, CB3, CB4, CB5, CB7 and CBB

Bacteria	strain	Isolation source
Cronobacter muytjensii	ATCC 51329 (type strain)	Unknown
Cronobacter malonaticus	DPC 6531	Brain tumour
Cronobacter sakazakii	ATCC 29004	Unknown
Dickeya chrysanthemi bv. chrysanthemi	LMG 2804 (type strain)	*Chrysanthemum*
Dickeya dianthicola	PD 482	*Solanum tuberosum* cv. Ostara
	PD 2174	-
	GBBC 1538	-
Dickeya solani	sp. PRI 2222	-
	LMG 25865	*Solanum tuberosum* cv. Première
	GBBC 1502	-
	GBBC 1586	-
Enterobacter cloacae	NCTC 11590	Unknown
Enterobacter gergoviae	NCTC 11434 (type strain)	Human urinary tract
Erwinia amylovora	LMG 2024 (type strain)	Pear (*Pyris communis*)
	GBBC 403	*Crataegus sp.*
Erwinia mallotivora	LMG 1271	*Mallotus japonicus*
Pantoea agglomerans	LMG 2660	*Wisteria floribunda*
	LMG 2570	*Sorbus sp.*
Pantoea stewartii	LMG 2713	*Zea mays*
	LMG 2714	*Zea mays*
	LMG 2712	*Zea mays*
Pectobacterium atrosepticum	DSM 18077 (type strain)	*Solanum tuberosum*
	DSM 30184	*Solanum tuberosum* cv. Bodenkraft
	DSM 30185	Solanum tuberosum
	DSM 30186	*Solanum tuberosum* cv. Maritta
	CB BL1-1	*Solanum tuberosum* cv. British Queen
	CB BL2-1	*Solanum tuberosum* cv. British Queen
	CB BL3-1	*Solanum tuberosum* cv. British Queen
	CB BL4-1	*Solanum tuberosum* cv. British Queen
	CB BL5-1	*Solanum tuberosum* cv. British Queen
	CB BL7-1	*Solanum tuberosum* cv. Golden wonder
	CB BL9-1	*Solanum tuberosum* cv. Golden wonder
	CB BL11-1	*Solanum tuberosum* cv. Rooster
	CB BL12-2	*Solanum tuberosum* cv. Golden wonder
	CB BL13-1	*Solanum tuberosum* cv. Golden wonder

	CB BL14-1	*Solanum tuberosum* cv. Golden wonder
	CB BL15-1	*Solanum tuberosum* cv. Golden wonder
	CB BL16-1	*Solanum tuberosum* cv. Golden wonder
	CB BL18-1	*Solanum tuberosum* cv. Golden wonder
	CB BL19-1	*Solanum tuberosum* cv. Golden wonder
Pectobacterium carotovorum subsp. *carotovorum*	DSMZ 30168 (type strain)	*Solanum tuberosum*
	DSMZ 30169	*Brassica oleracea* var. capitate
	DSMZ 30170	*Solanum tuberosum* "Maritta"
	CB BL19-1-37	*Solanum tuberosum* cv. Golden wonder

Table S2.2. Genbank details of N4-like phages used in phylograms and Gegenees analysis of the CB1-like phages.

Phage	Genbank accession number	DNA polymerase	virion RNA polymerase
Achromobacter phage JWAlpha	KF787095.1	KF787095.1	YP_009004769.1
Achromobacter phage JWDelta	KF787094.1	KF787094.1	AHC56581.1
Achromobacter phage phiAxp-3	NC_028908.1	NC_028908.1	YP_009208706.1
Acinetobacter phage Presley	KF669658.1	KF669658.1	YP_009007647.1
Dinoroseobacter phage DFL12phi1	KJ621082.2	KJ621082.2	YP_009043702.1
Dinoroseobacter phage vBDshPR2C	KJ803031.1	KJ803031.1	AID16877.1
Enterobacter phage EcP1	NC_019485.1	NC_019485.1	YP_007003173.1
Erwinia phage Ea9-2	NC_023579.1	YP_009007430	YP_009007447.1
Erwinia phage vB_EamP_Gutmeister	KX098391.1	ANJ65360.1	ANJ65375.1
Erwinia phage vB_EamP_Rexella	KX098390.1	ANJ65282.1	ANJ65299.1
Erwinia phage vB_EamP-S6	NC_019514.1	NC_019514.1	YP_007005815.1
Escherichia phage ECBP1	JX415535.1	YP_006908814.1	YP_006908827.1
Escherichia phage N4	NC_008720.1	YP_950517.1	YP_950528.1
Escherichia phage Pollock	NC_027381.1	NC_027381.1	YP_009152160.1
Escherichia phage vB_EcoP_G7C	HQ259105.1	YP_004782168.1	YP_004782180.1
Escherichia phage vB_EcoP_PhAPEC7	KF562340.1	YP_009056171.1	YP_009056186.1
Pseudoalteromonas phage pYD6-A	JF974296.1	JF974296.1	YP_007674286.1
Pseudomonas phage inbricus	MG018928.1	ATW58103.1	ATW58114.1
Pseudomonas phage KPP21	LC064302.1	LC064302.1	YP_009218950.1
Pseudomonas phage LIT1	FN422399.1	YP_003358435.1	YP_003358468.1
Pseudomonas phage LUZ7	FN422398.1	FN422398.1	YP_003358355.1
Pseudomonas phage RWG	KM411958.1	AIZ94788.1	AIZ94822.1
Pseudomonas phage vB_Pae575P-3	KX171209.1	ANT44317.1	ANT44349.1
Pseudomonas phage ZC03	KU356690.1	KU356690.1	AMD43402.1
Pseudomonas phage ZC08	KU356691.1	KU356691.1	AMD43541.1
Roseophage DSS3P2	FJ591093.1	FJ591093.1	YP_002899070.1
Roseophage EE36P1	FJ591094.1	FJ591094.1	YP_002898988.1
Salmonella phage FSL SP-058	NC_021772.1	NC_021772.1	YP_008239463.1
Salmonella phage FSL SP-076	KC139520.1	KC139520.1	YP_008240191.1
Sulfitobacter phage phiCB2047-B	NC_020862	NC_020862	YP_007675808.1
Vibrio phage JA-1	KC438282.1	KC438282.1	YP_008126816.1
Vibrio phage JSF3	KY065148.1	APD18062.1	APD18049.1
Vibrio phage phi 1	KP280062.1	KP280062.1	YP_009198592.1
Vibrio phage pVa5	KX889068.1	KX889068.1	APC46019.1
Vibrio phage pVco-5	KY612839.1	ARM71084.1	ARM71100.1
Vibrio phage VBP32	HQ634196.1	HQ634196.1	YP_007676574.1
Vibrio phage VBP47	HQ634194.1	HQ634194.1	YP_007674140.1
Vibrio phage VCO139	KC438283.1	KC438283.1	AGI61882.1

Table S2.3. Details of proteins used in the phylogenetic analysis of 52 phages from the subfamily *Autographivirinae* and *Pectobacterium* phage CB5

Phage	Genome accession no.	Major capsid accession no.
Acinetobacter phage Abp1	NC_021316.1	YP_008058231.1
Acinetobacter phage Fri1	NC_028848.1	YP_009203047.1
Acinetobacter phage Petty	NC_023570.1	YP_009006529.1
Acinetobacter phage phiAB1	NC_028675.1	YP_009189372.1
Acinetobacter phage vB_ApiP_P2	MF033351.1	ASN73550.1
Aeromonas phage phiAS7	NC_019528.1	YP_007007808.1
Cronobacter phage vB_CskP_GAP227	NC_020078.1	YP_007348355.1
Dickeya phage BF25/12	KT240186.1	ALA46504.1
Enterobacteria phage J8-65	NC_025445.1	YP_009101383.1
Enterobacteria phage K30	NC_015719.1	YP_004678755.1
Enterobacteria phage T7	NC_001604.1	NP_041997.1
Erwinia amylovora phage Era103	NC_009014.1	YP_001039668.1
Escherichia phage phiKT	NC_019520.1	YP_007006600.1
Escherichia virus K1-5	NC_008152.1	YP_654132.1 31
Escherichia virus K1E	NC_007637.1	YP_425009.1
Klebsiella phage F19	NC_023567.2	YP_009006057.1
Klebsiella phage K11	NC_011043.1	YP_002003823.1
Klebsiella phage KP32	NC_013647.1	YP_003347548.1
Klebsiella phage KP34	NC_013649.2	YP_003347636.1
Klebsiella phage NTUH-K2044-K1-1	NC_025418.1	YP_009098373.1
Klebsiella phage vB_KpnP_SU503	NC_028816.1	YP_009199922.1
Klebsiella phage vB_KpnP_SU552A	NC_028870.1	YP_009204828.1
Kluyvera phage Kvp1	FJ194439.1	ACJ14590.1
Pantoea phage LIMElight	NC_019454.1	YP_007002894.1
Pantoea phage LIMEzero	NC_015585.1	YP_004539113.1
Pectobacterium phage PhiM1	JX290549.1	AFQ22523.1
Pectobacterium phage PP16	NC_031068.1	YP_009286812.1
Pectobacterium phage PP90	NC_031096.1	YP_009289647.1
Pectobacterium phage PPWS1	LC063634.2	BAS69556.1
Pectobacterium_phage_Peat1	NC_029081.1	YP_009224669.1+YP_009224670.1
Pseudomonad phage gh-1	AF493143.1	AAO73167.1
Pseudomonas phage Bf7	NC_016764.1	YP_005098192.1
Pseudomonas phage LKA1	NC_009936.1	YP_001522884.1
Pseudomonas phage LKD16	NC_009935.1	YP_001522824.1
Pseudomonas phage LUZ19	NC_010326.1	YP_001671977.1
Pseudomonas phage MPK6	NC_022746.1	YP_008766800.1

Pseudomonas phage MPK7	NC_022091.1	YP_008431345.1
Pseudomonas phage phi-2	NC_013638.1	YP_003345495.1
Pseudomonas phage phikF77	NC_012418.1	YP_002727855.1
Pseudomonas phage phiKMV	NC_005045	NP_877471.1
Pseudomonas phage PT2	NC_011107.1	YP_002117817.1
Pseudomonas phage PT5	EU056923.1	ABW23115.1
Ralstonia phage RSB1	NC_011201.1	YP_002213721.1
Ralstonia phage RSB3	NC_022917.1	YP_008853924.1
Ralstonia virus phiAp1	KY117485.1	APU03181.1
Salmonella phage SP6	NC_004831.2	NP_853592.1
Vibrio phage VP93	NC_012662	YP_002875653.1
Xanthomonas phage f20-Xaj	KU595432.1	AMM44667.1
Xanthomonas phage f30-Xaj	KU595433.1	AMM44714.1
Xylella phage Prado	NC_022987.1	YP_008859419.1
Yersinia phage phi80-18	HE956710.2	CCI88880.2
Yersinia phage phiR8-01	HE956707.2	CCI88417.2
Pectobacterium phage vB_Pat_CB5	KY953156	ARW59018

Table S3.1. Results of physiological, biochemical, *Pectobacterium* genus (*pel* gene) specific and *Pectobacterium atrosepticum* and *Pectobacterium carotovorum* subsp. *carotovorum* species-specific PCRs and MALDI-TOF mass spectrometry on isolates obtained from potato stem samples symptomatic for blackleg from farms in Co. Cork, Ireland.

Strains	Isolation source	Cavity formation on CVP medium (27 °C, 48 h)	Growth on NA at 37 °C	Production of reducing substances from sucrose	Acid production from alpha-methyl glucoside	PCR - *Pel* gene (Darrasse *et al* 1994)	PCR - *P. atrosepticum* (De Boer & Ward 1995)	PCR - *P. carotovorum* subsp. *carotovorum* (Kang *et al* 2003)	MALDI-TOF MS	Identity
CB BL1-1	Co. Cork, Clonakilty, cultivar British queens	+	-	+	+	+	+	NA	Pa	Pa
CB BL2-1		+	-	+	+	+	+	NA	Pa	Pa
CB BL3-1		+	-	+	+	+	+	NA	Pa	Pa
CB BL4-1		+	-	+	+	+	+	NA	Pa	Pa
CB BL5-1		+	-	+	+	+	+	NA	Pa	Pa
CB BL7-1	Co. Cork, Dunmanway, cultivar Golden wonders	+	-	+	+	+	+	NA	Pa	Pa
CB BL9-1		+	-	+	+	+	+	NA	Pa	Pa
CB BL11-1	Co. Cork, Dunmanway, cultivar Roosters	+	-	+	+	+	+	NA	Pa	Pa
CB BL12-2	Co. Cork, Skibbereen, cultivar Golden wonders	+	-	+	+	+	+	NA	Pa	Pa
CB BL13-1		+	-	+	+	+	+	NA	Pa	Pa
CB BL14-1		+	-	+	+	+	+	NA	Pa	Pa
CB BL15-1		+	-	+	+	+	+	NA	Pa	Pa
CB BL16-1		+	-	+	+	+	+	NA	Pa	Pa
CB BL18-1		+	-	+	+	+	+	NA	Pa	Pa
CB BL19-1		+	-	+	+	+	+	NA	Pa	Pa
CB BL19-1-37		+	+	-	-	+	-	+	NA	Pcc

Results recorded as +, positive; –, negative; NA, not available; Pa, *P. atrosepticum*; Pcc, *P. carotovorum* subsp. *carotovorum*
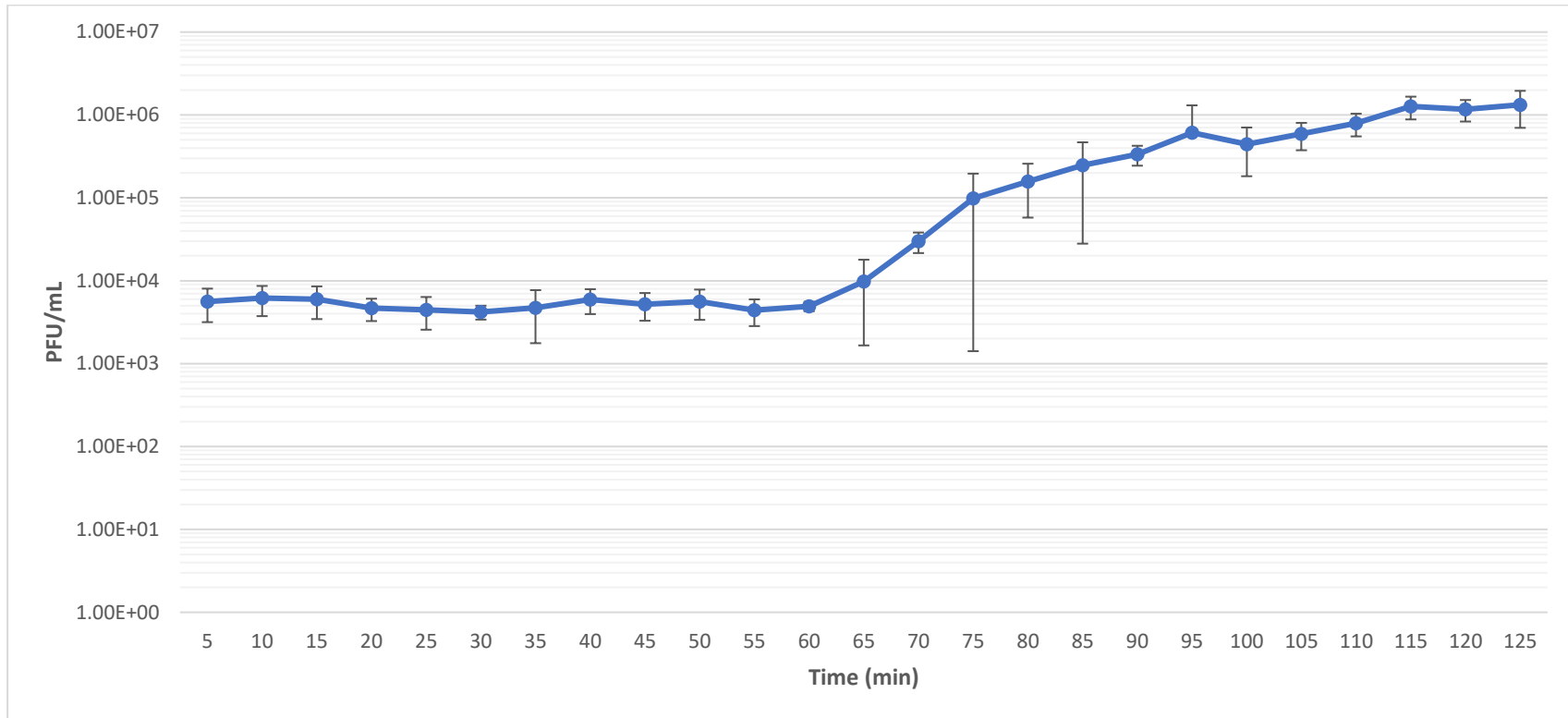
(B)
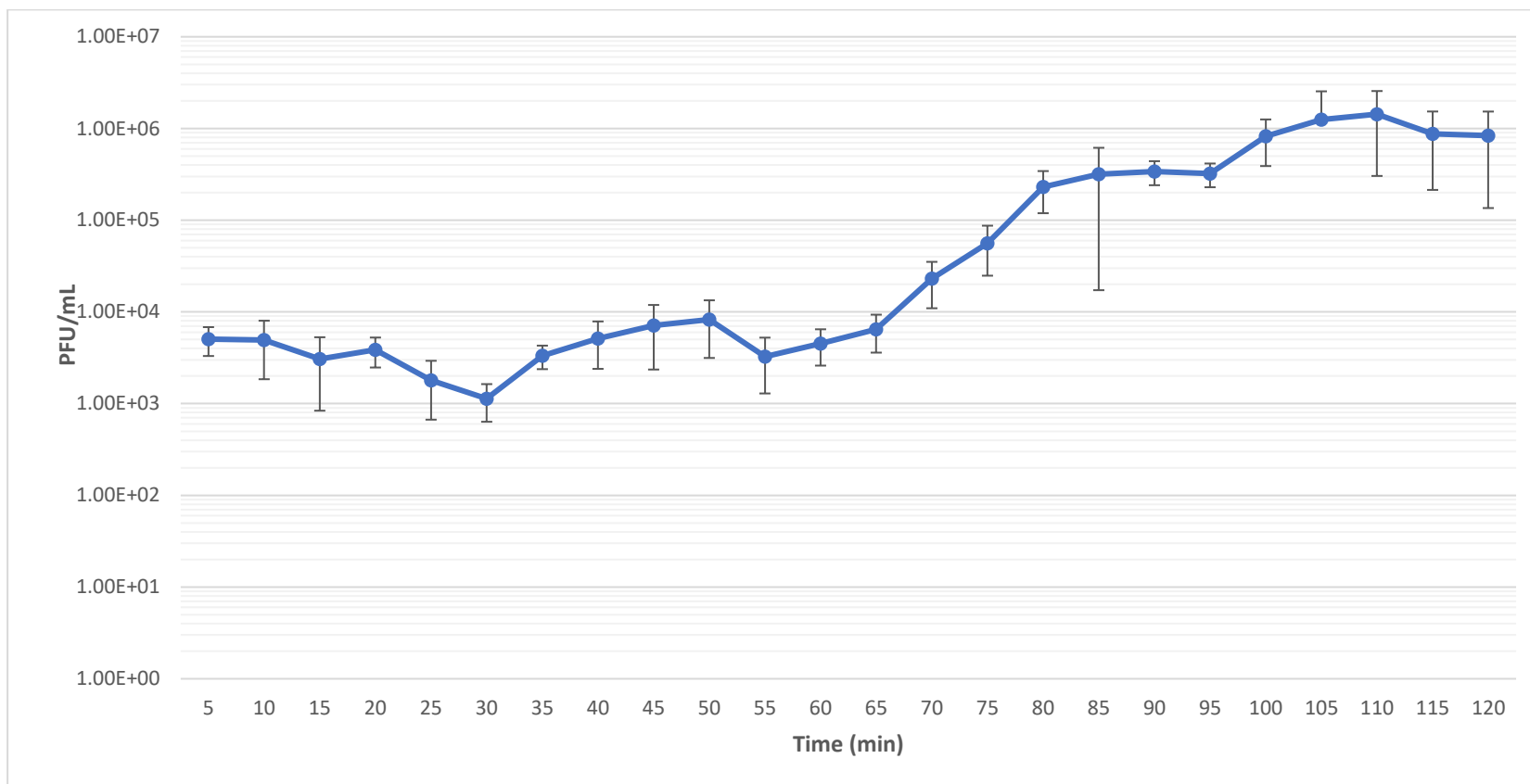
(C)
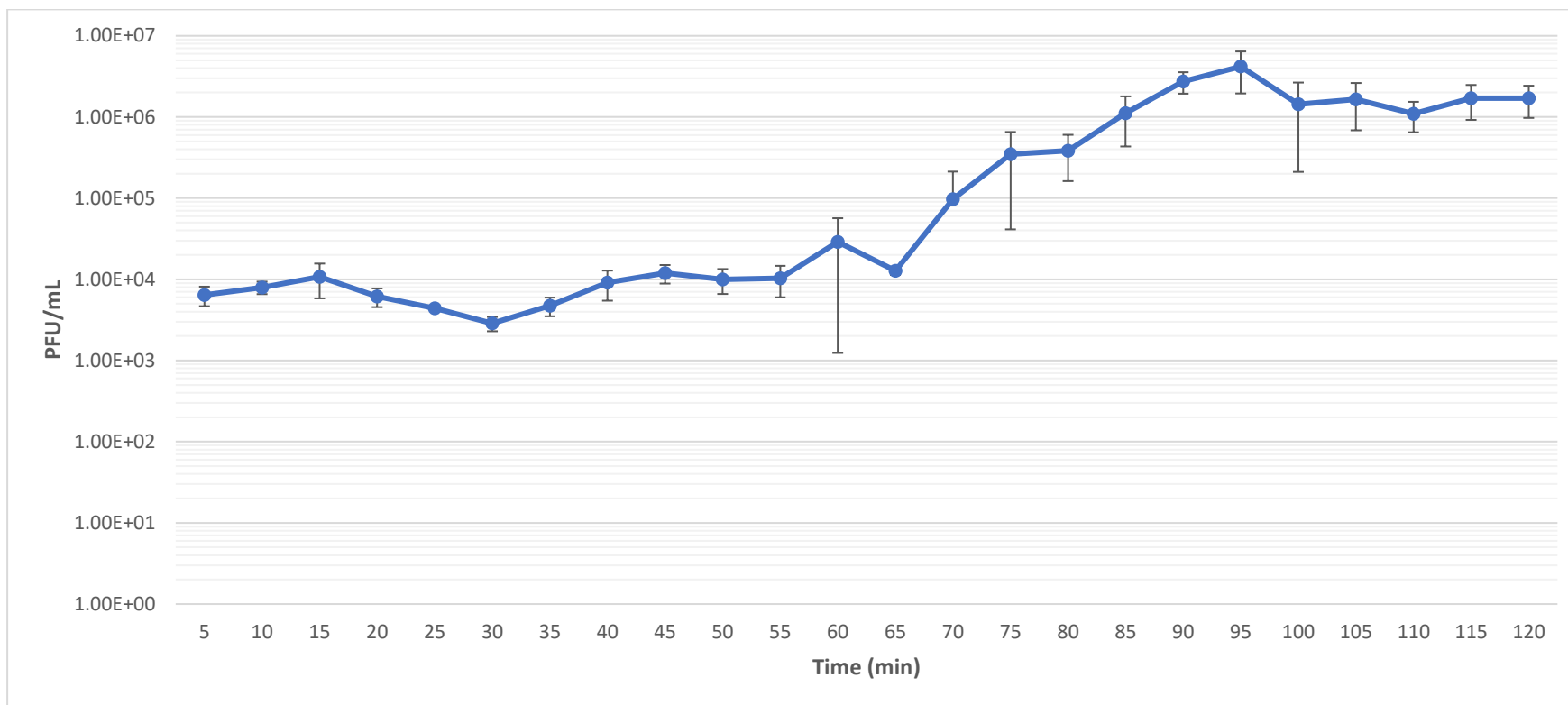
Figure S3.1. Single step growth curve growth analysis of phage vB_PatP_CB1 infection of *P. atrosepticum* strain DSM 18077 (A), phage vB_PatP CB3 infection of *P. atrosepticum* strain DSM 30186 (B) and phage vB_PatP CB4 infection of *P. atrosepticum* DSM 30186 (C). Each assay was independently repeated in triplicate and the results were averaged.
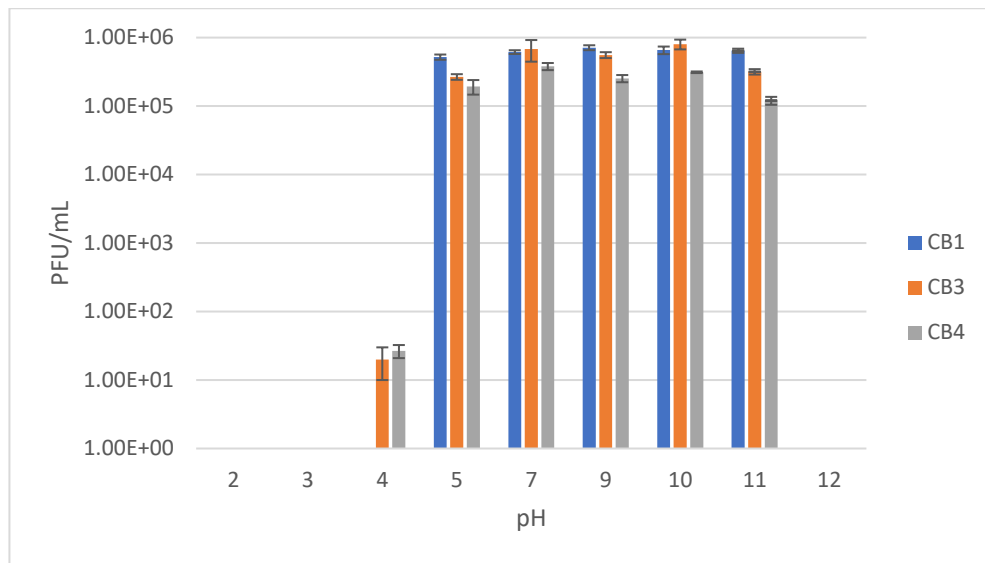
Figure S3.2. Stability of *Pectobacterium* phages CB1, CB3 and CB4 to various pH values upon 24 hours of exposure. Error bars represent the standard deviation of biological repeats (*n*=3).
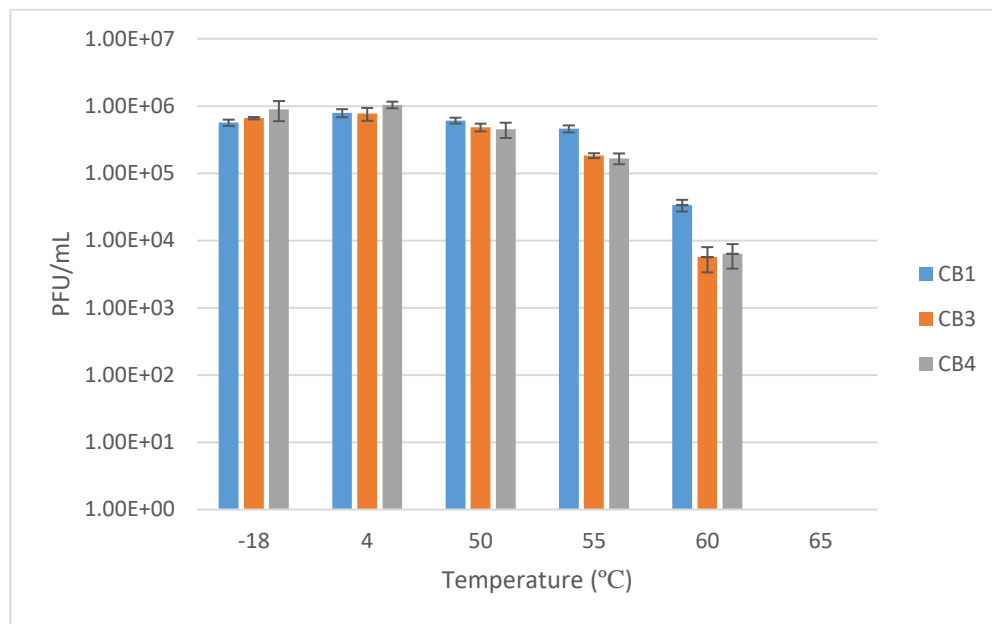
Figure S3.3. Stability of *Pectobacterium* phages CB1, CB3 and CB4 to various temperatures upon one-hour exposures. Error bars represent the standard deviation of biological repeats (*n*=3).

Table S3.2. Genomic annotation of *Pectobacterium* phage CB1

ORF	Start (bp)	Stop (bp)	Length (aa)	Molecular mass (kDa)	Function
CB1_01	141	389		8.64	putative membrane protein
CB1_02	915	1196	93	10.69	conserved hypothetical protein
CB1_03	1262	1762	166	18.74	hypothetical protein
CB1_04	1831	2070	79	9.32	hypothetical protein
CB1_05	2140	2247	35	3.77	putative membrane protein
CB1_06	2244	2594	116	13.7	hypothetical protein
CB1_07	2738	2908	56	6.31	putative membrane protein
CB1_08	2954	3229	91	10.11	hypothetical protein
CB1_09	3232	3447	71	8.08	hypothetical protein
CB1_10	3444	3758	104	12.06	hypothetical protein
CB1_11	3755	4246	163	19.11	conserved hypothetical protein
CB1_12	4239	4619	126	14.04	N4 gp14-like protein
CB1_13	4790	4906	38	4.49	hypothetical protein
CB1_14	4910	5179	89	10.22	hypothetical protein
CB1_15	5176	5469	97	11.28	hypothetical protein
CB1_16	5469	5792	107	12.57	hypothetical protein
CB1_17	5789	6094	101	11.33	hypothetical protein
CB1_18	6094	6390	98	11.29	hypothetical protein
CB1_19	6463	6744	93	10.61	hypothetical protein
CB1_20	6823	7119	98	10.45	conserved hypothetical protein
CB1_21	7165	7608	147	16.98	hypothetical protein
CB1_22	7618	8601	327	37.78	RNAP1
CB1_23	8612	9892	426	47.79	RNAP2
CB1_24	9937	10470	177	20.81	HNH endonuclease
CB1_25	10473	10715	80	8.96	hypothetical protein
CB1_26	10718	10960	80	9.1	hypothetical protein
CB1_27	11206	11415	69	7.99	hypothetical protein
CB1_28	11402	11989	195	22.06	hypothetical protein
CB1_29	11970	12164	64	7.58	hypothetical protein
CB1_30	12161	12475	104	11.72	hypothetical protein
CB1_31	12731	12949	72	8.82	hypothetical protein
CB1_32	13011	13112	33	3.97	hypothetical protein
CB1_33	13109	13471	120	13.67	HNH endonuclease
CB1_34	13468	13854	128	13.59	putative lipoprotein
CB1_35	13944	15014	356	40.03	N4 gp24-like protein
CB1_36	15057	15665	202	22.06	hypothetical protein
CB1_37	15680	16267	195	22.92	hypothetical protein
CB1_38	16236	17387	383	43.01	N4 gp25-like protein
CB1_39	17389	17856	155	17.27	HNH endonuclease
CB1_40	17874	19229	451	51.51	helicase
CB1_41	19299	19753	174	20.01	hypothetical protein

CB1_42	19750	20088	112	12.45	N4 gp32 like
CB1_43	20347	20853	168	19.19	HNH endonuclease
CB1_44	20850	23276	808	92.07	rIIA-like protein
CB1_45	23273	25108	611	65.39	rIIB-like protein
CB1_46	25120	27901	903	101.88	DNA polymerase I
CB1_47	27902	28087	61	6.7	putative membrane protein
CB1_48	28116	28307	63	7.47	hypothetical protein
CB1_49	28285	28758	157	18.56	conserved hypothetical protein
CB1_50	28766	29113	115	13.14	hypothetical protein
CB1_51	29172	29492	106	11.65	putative membrane protein
CB1_52	29530	30042	170	19.21	HNH endonuclease
CB1_53	30039	30845	268	30.4	DNA adenine methylase
CB1_54	30852	31037	61	6.87	hypothetical protein
CB1_55	31034	31303	89	10.04	hypothetical protein
CB1_56	31527	31760	77	8.8	hypothetical protein
CB1_57	32260	31805	151	15.59	putative tail protein
CB1_58a	32722	32321	133		putative Rz
CB1_58	32840	32343	165	18.21	putative Rz1
CB1_59	33051	32815	78	9.19	holin
CB1_60	33524	33048	158	17.45	lysozyme
CB1_61	36651	33721	976	104.98	putative tail spike protein
CB1_62	37382	36663	239	26.01	putative tail protein
CB1_63	38247	37480	255	28.66	thymidylate synthase
CB1_64	38255	38422	55	6.27	conserved hypothetical protein
CB1_65	38448	38972	174	19.82	conserved hypothetical protein
CB1_66	38972	39961	329	37.92	N4 gp42-like protein
CB1_67	39972	42134	720	82.1	primase
CB1_68	42142	42693	183	20.51	ryanodine receptor like protein
CB1_69	42793	43005	70	7.94	hypothetical protein
CB1_70	43005	43742	245	27.72	N4 gp44-like protein
CB1_71	43729	44250	173	20.22	HNH endonuclease
CB1_72	44268	45032	254	27.69	ssDNA binding protein
CB1_73	45096	45431	111	12.06	hypothetical protein
CB1_74	45424	45987	187	20.22	putative holiday junction resolvase
CB1_75	45993	46781	262	27.45	hypothetical protein
CB1_76	46791	46970	59	6.36	putative membrane protein
CB1_77	58065	46984	3693	399.92	virion associated RNA polymerase (N4 gp50-like)
CB1_78	59234	58068	288	41.64	unknown structural protein
CB1_79	59647	59234	137	14.2	structural protein (N4 gp52-like)
CB1_80	62430	59656	924	101.4	N4 gp53-like protein
CB1_81	63282	62494	262	28.82	structural protein (N4 gp54-like)
CB1_82	63875	63294	193	21.85	N4 gp55-like protein
CB1_83	65136	63964	390	43.33	major capsid protein (N4 gp56-like)
CB1_84	66365	65151	163	44.18	tail measure protein (N4 gp57-like)

CB1_85	66782	66456	404	12.29	conserved hypothetical protein
CB1_86	69087	66850	745	83.31	portal protein (N4 gp59-like)
CB1_87	69155	69703	182	20.23	putative nucleoside triphosphate pyrophosphohydrolase
CB1_88	70214	70459	81	9.33	conserved hypothetical protein
CB1_89	70648	71040	130	14.12	N4 gp48-like
CB1_90	71977	71072	301	33.72	structural protein (N4 gp67-like)
CB1_91	73599	71986	537	61.46	large terminase subunit
CB1_92	73765	73592	57	6.43	hypothetical protein
CB1_93	74444	73758	228	25.42	N4 gp69 -like
CB1_94	74500	74760	86	9.56	conserved hypothetical protein
CB1_95	75086	74781	101	10.91	hypothetical protein
CB1_96	75191	75295	34	3.74	putative membrane protein
CB1_97	75385	75783	132	14.17	hypothetical protein

Table S3.3. Genomic annotation of *Pectobacterium* phage CB3

ORF	start	stop	Length (aa)	Molecular weight (kDa)	Function
CB3_01	45	443	132	14.17	hypothetical protein
CB3_02	1011	1292	93	10.68	conserved hypothetical protein
CB3_03	1341	1841	166	18.69	conserved hypothetical protein
CB3_04	1874	2011	45	4.87	putative membrane protein
CB3_05	2008	2358	116	13.7	hypothetical protein
CB3_06	2355	2501	46	5.58	hypothetical protein
CB3_07	2502	2672	56	6.31	putative membrane protein
CB3_08	2718	2960	90	9.04	hypothetical protein
CB3_09	2957	3250	97	10.95	conserved hypothetical protein
CB3_10	3228	3470	80	9.34	hypothetical protein
CB3_11	3470	4123	217	25.81	hypothetical protein
CB3_12	4120	4491	123	13.7	hypothetical protein
CB3_13	4476	4778	100	11.71	hypothetical protein
CB3_14	4782	5051	89	10.26	hypothetical protein
CB3_15	5048	5341	97	11.22	hypothetical protein
CB3_16	5341	5664	107	12.57	hypothetical protein
CB3_17	5661	5951	96	10.99	hypothetical protein
CB3_18	5930	6250	106	11.96	hypothetical protein
CB3_19	6250	6546	98	11.43	hypothetical protein
CB3_20	6619	6900	93	10.61	hypothetical protein
CB3_21	6979	7275	98	10.48	conserved hypothetical protein
CB3_22	7321	7599	92	10.71	conserved hypothetical protein
CB3_23	7609	8022	137	16.09	hypothetical protein
CB3_24	8032	9015	327	37.71	N4 RNAP1-like protein
CB3_25	9026	10306	426	47.83	N4 RNAP2-like protein
CB3_26	10309	10500	63	7.7	hypothetical protein
CB3_27	10547	10822	91	10.88	hypothetical protein
CB3_28	10870	11064	64	7.14	hypothetical protein
CB3_29	11067	11312	81	9.37	hypothetical protein
CB3_30	11558	11767	69	8.04	hypothetical protein
CB3_31	11754	12341	195	22.06	hypothetical protein
CB3_32	12322	12516	64	7.58	hypothetical protein
CB3_33	12513	12821	102	11.48	hypothetical protein
CB3_34	13087	13302	71	8.69	hypothetical protein
CB3_35	13364	13465	33	3.97	hypothetical protein
CB3_36	13462	13824	120	13.67	HNH endonuclease
CB3_37	13821	14207	128	13.59	putative lipoprotein
CB3_38	14297	15367	356	39.98	N4 gp24-like protein

CB3_39	15410	16018	202	21.99	hypothetical protein
CB3_40	16069	16620	183	21.54	hypothetical protein
CB3_41	16571	17740	389	43.76	N4 gp25-like protein
CB3_42	17758	19053	431	48.93	DNA helicase
CB3_43	19034	19240	68	8.01	DNA helicase
CB3_44	19240	19764	174	20.01	hypothetical protein
CB3_45	19761	20129	122	13.59	N4 gp32-like protein
CB3_46	20129	20320	63	7.13	hypothetical protein
CB3_47	20392	20898	168	19.19	HNH endonuclease
CB3_48	20895	23321	808	91.86	rIIA-like protein
CB3_49	23318	25417	699	74.58	rIIB-like protein
CB3_50	25499	28210	903	101.89	DNA polymerase I
CB3_51	28211	28396	61	6.76	putative lipoprotein
CB3_52	28425	28616	63	7.47	conserved hypothetical protein
CB3_53	28594	29067	157	18.53	nucleotidase
CB3_54	29075	29422	115	13.16	hypothetical protein
CB3_55	29481	29801	106	11.65	putative membrane protein
CB3_56	29839	30351	170	19.16	HNH endonuclease
CB3_57	30348	31154	268	30.28	DNA adenine methylase
CB3_58	31161	31346	61	6.87	putative HNN endonuclease
CB3_59	31343	31612	89	10.04	conserved hypothetical protein
CB3_60	31836	31069	77	8.87	hypothetical protein
CB3_61	32569	32114	151	15.59	tail protein
CB3_62a	33031	32630	133	15.1	putative rz
CB3_62	33149	32652	165	18.27	putative rz1
CB3_63	33360	33124	78	9.19	holin
CB3_64	33833	33357	158	17.45	lysozyme
CB3_65	33877	33975	32	3.31	hypothetical protein
CB3_66	34020	36960	976	105.02	putative tail spike protein
CB3_67	37691	36972	239	26.02	tail fiber protein
CB3_68	37789	38556	255	28.66	thymidylate synthase
CB3_69	38564	38731	55	6.27	conserved hypothetical protein
CB3_70	38757	39281	174	19.82	conserved hypothetical protein
CB3_71	39281	40270	329	37.94	N4 gp42-like protein
CB3_72	40281	42443	720	82.1	primase
CB3_73	42451	43002	183	20.55	Ryanodine receptor Ryr-like protein
CB3_74	43102	43314	70	7.94	hypothetical protein
CB3_75	43314	44051	245	27.72	N4 gp44-like protein
CB3_76	44038	44559	173	20.22	Putative HNH homing endonuclease
CB3_77	44577	45341	254	27.69	ssDNA binding protein

CB3_78	45405	45740	111	11.99	hypothetical protein
CB3_79	45733	46296	187	20.22	putative holiday junction resolvase
CB3_80	46302	47090	262	27.41	hypothetical protein
CB3_81	47100	47279	68	6.36	putative membrane protein
CB3_82	11082	47293	3693	400.09	virion associated RNA polymerase
CB3_83	59543	58377	373	41.67	Structural protein
CB3_84	59956	59543	137	14.22	N4 gp52-like protein
CB3_85	62739	59965	924	101.53	N4 gp53-like protein
CB3_86	63591	62803	262	28.82	N4 gp54-like protein
CB3_87	64184	63603	193	21.85	N4 gp55-like protein
CB3_88	65445	64273	390	43.33	major virion protein
CB3_89	66641	65460	394	42.85	tape measure protein
CB3_90	66964	66638	108	12.29	conserved hypothetical protein
CB3_91	69269	67032	745	83.31	N4 gp59-like protein
CB3_92	69337	69885	182	20.23	putative nucleoside triphosphate pyrophosphohydrolase
CB3_93	70397	70311	28	3.24	hypothetical protein
CB3_94	70396	70641	81	9.33	conserved hypothetical protein
CB3_95	70830	71222	130	14.12	N4 gp48-like protein
CB3_96	72159	71254	301	33.72	putative structural protein
CB3_97	73781	72168	537	61.45	terminase large subunit
CB3_98	73947	73774	57	6.43	putative membrane protein
CB3_99	73940	74626	228	25.42	N4 gp69-like protein
CB3_100	74682	74942	86	9.56	conserved hypothetical protein
CB3_101	75268	74963	101	11.02	hypothetical protein
CB3_102	75567	75965	132	14.17	hypothetical protein

Table S3.4. Genomic annotation of *Pectobacterium* phage CB4

ORF	Start (bp)	Stop (bp)	Length (aa)	Molecular weight (kDa)	Function
CB4_001	141	389	82	8.6	putative membrane protein
CB4_002	958	1239	93	10.68	conserved hypothetical protein
CB4_003	1288	1788	166	18.69	conserved hypothetical protein
CB4_004	1851	1958	35	3.77	putative membrane protein
CB4_005	1955	2305	116	13.7	hypothetical protein
CB4_006	2302	2448	48	5.58	hypothetical protein
CB4_007	2449	2619	56	6.31	putative membrane protein
CB4_008	2665	2907	80	9.04	hypothetical protein
CB4_009	2904	3197	97	10.95	conserved hypothetical protein
CB4_010	3175	3417	80	9.34	hypothetical protein
CB4_011	3417	4070	217	25.81	hypothetical protein
CB4_012	4067	4438	123	13.7	N4 gp14-like protein
CB4_013	4423	4725	100	11.71	hypothetical protein
CB4_014	4729	4998	89	10.26	hypothetical protein
CB4_015	4995	5288	97	11.22	hypothetical protein
CB4_016	5288	5611	107	12.57	hypothetical protein
CB4_017	5608	5898	96	10.99	hypothetical protein
CB4_018	5877	6197	106	11.96	hypothetical protein
CB4_019	6197	6493	98	11.43	hypothetical protein
CB4_020	6566	6847	93	10.61	hypothetical protein
CB4_021	6926	7222	98	10.48	hypothetical protein
CB4_022	7268	7546	92	10.71	conserved hypothetical protein
CB4_023	7556	7969	137	16.09	hypothetical protein
CB4_024	7979	8962	327	37.71	N4 RNAP1-like protein
CB4_025	8973	10253	426	47.83	N4 RNAP2-like protein
CB4_026	10256	10447	63	7.7	hypothetical protein
CB4_027	10494	10769	91	10.88	hypothetical protein
CB4_028	10769	11011	90	8.94	hypothetical protein
CB4_029	11014	11259	81	9.37	hypothetical protein
CB4_030	11505	11714	69	8.04	hypothetical protein
CB4_031	11701	12288	195	22.06	hypothetical protein
CB4_032	12269	12463	64	7.58	hypothetical protein
CB4_033	12460	12768	102	11.48	hypothetical protein
CB4_034	13034	13249	71	8.69	hypothetical protein
CB4_035	13311	13412	33	3.97	hypothetical protein
CB4_036	13409	13771	120	13.67	HNH endonuclease
CB4_037	13768	14154	128	13.59	putative lipoprotein
CB4_038	14244	15314	356	40.03	N4 gp24-like protein

CB4_039	15357	15965	202	22.14	hypothetical protein
CB4_040	15980	16567	195	23.01	hypothetical protein
CB4_041	16518	17687	289	43.71	N4 gp25-like protein
CB4_042	17705	19060	451	51.55	DNA helicase
CB4_043	19060	19584	174	19.98	hypothetical protein
CB4_044	19581	19919	112	12.45	N4 gp32-like protein
CB4_045	19919	20107	62	7.04	hypothetical protein
CB4_046	20178	20684	168	19.17	HNH endonuclease
CB4_047	20681	23107	808	91.81	rIIA-like protein
CB4_048	23104	25203	699	74.58	rIIB-like protein
CB4_049	25354	27996	880	99.26	DNA polymerase I
CB4_050	28033	28182	49	5.28	putative membrane protein
CB4_051	28211	28402	63	7.47	conserved hypothetical protein
CB4_052	28380	28853	157	18.48	nucleotidase
CB4_053	28861	29208	115	13.13	hypothetical protein
CB4_054	29267	29587	106	11.65	putative membrane protein
CB4_055	29625	30137	170	19.16	HNH endonuclease
CB4_056	30134	30940	268	30.28	DNA adenine methylase
CB4_057	30947	31132	61	6.87	hypothetical protein
CB4_058	31129	31398	89	10.04	conserved hypothetical protein
CB4_059	31622	31855	77	8.85	hypothetical protein
CB4_060	32355	31900	151	15.59	tail protein
CB4_61a	32817	32418	133	15.1	putative rz
CB4_061	32935	32438	165	18.27	putative rz1
CB4_062	33146	32910	78	9.19	holin
CB4_063	33619	33143	158	17.45	lysozyme
CB4_064	33816	35627	603	65.39	putative tail spike protein
CB4_065	36873	35611	420	44.3	conserved hypothetical protein
CB4_066	37604	36885	239	25.98	tail protein
CB4_067	38194	38814	206	23.46	thymidylate synthase
CB4_068	38819	38986	55	6.32	hypothetical protein
CB4_069	39012	39536	174	19.88	N4 gp41-like protein
CB4_070	39536	40525	329	37.93	N4 gp42-like protein
CB4_071	40536	42698	720	82.1	primase
CB4_072	42706	43257	183	20.55	conserved hypothetical protein
CB4_073	43357	43569	70	7.95	hypothetical protein
CB4_074	43569	44306	245	27.71	N4 gp44-like protein
CB4_075	44293	44814	173	20.2	HNH endonuclease
CB4_076	44832	45596	245	27.69	ssDNA binding protein
CB4_077	45660	45995	111	11.95	hypothetical protein
CB4_078	45988	46551	187	20.23	putative holiday junction resolvase

CB4_079	46557	47345	262	27.45	hypothetical protein
CB4_080	47355	47534	59	6.36	putative membrane protein
CB4_081	58629	47548	3693	400.08	virion associated RNA polymerase
CB4_082	59798	58632	388	41.73	structural protein
CB4_083	60211	59798	137	14.23	N4 gp52-like protein
CB4_084	62994	60220	924	101.39	N4 gp53-like protein
CB4_085	63846	63058	262	28.82	N4 gp54-like protein
CB4_086	63858	64439	193	21.85	N4 gp55-like protein
CB4_087	65700	64528	390	43.33	major capsid protein
CB4_088	66896	65715	393	43.03	tape measure protein
CB4_089	67219	66893	108	12.29	conserved hypothetical protein
CB4_090	69524	67287	745	83.26	portal protein
CB4_091	69593	70162	189	21.22	putative nucleoside triphosphate pyrophosphohydrolase
CB4_092	70768	71013	81	9.33	conserved hypothetical protein
CB4_093	71203	71595	130	14.12	N4 gp48-like protein
CB4_094	72532	71627	301	33.57	putative tail protein
CB4_095	74154	72541	537	61.46	terminase large subunit
CB4_096	74344	74147	65	7.53	hypothetical protein
CB4_097	75023	74337	228	25.45	N4 gp69-like protein
CB4_098	75079	75339	86	9.56	conserved hypothetical protein
CB4_099	75665	75360	101	10.96	hypothetical protein
CB4_100	76114	76362	82	8.6	putative membrane protein

Table S3.5. Identified ORFs and tRNA gene variations between the genomes of *Pectobacterium* phages CB1, CB3 and CB4. With genome comparisons made using BLASTN with ACT. Bold with italics (shared feature), without bold with italics (unique feature to phage in question).

CB1	CB3	CB4	Variation
CB1_4 (hypothetical)	_	_	CB1_4 has no homolog in CB3+CB4
CB1_8 (hypothetical)	_	_	CB1_8 has no homolog in CB3+CB4
CB1_10 (hypothetical)	*CB3_9 (hypothetical)*	*CB4_9*	ORFs of CB1 not shared with CB3 +CB4
CB1_11 (hypothetical)	*CB3_10 (hypothetical)*	*CB4_10*	
_	*CB3_11 (hypothetical)*	*CB4_11*	
_	*CB3_17 (hypothetical)*	*CB4_17*	CB3/CB4_17 homolog not present in CB1
CB1_21 (hypothetical)	*CB3_22 (hypothetical)*	*CB4_22*	ORF of CB1 are not shared with CB3+CB4
_	*CB3_23 (hypothetical)*	*CB4_23*	
CB1_24 (HNH)	*CB3_26 (hypothetical)*	*CB4_26*	ORFs of CB1 are not shared with CB3+CB4
_	*CB3_27 (hypothetical)*	*CB4_27*	
CB1_26 (hypothetical)	*CB3_29 (hypothetical)*	*CB4_29*	ORF of CB1 different in part to CB3+CB4
CB1_39 (HNH)	_		CB1 has no homolog in CB3+CB4
CB1_42 (N4 gp32-like)	CB3_45 (N4 gp32-like)	*CB4_44 (N4 gp32-like)*	N4 gp32-like gene of CB3 different to CB1+CB4
CB1_45 (rIIB)	*CB3_49 (rIIB)*	*CB4_48 (rIIB)*	Difference of CB1 ORF length compared to CB1+CB4
CB1_61 (tail spike)	*CB3_66 (tail spike)*	CB4_64, 65 (tail spike)	CB4 homolog split into two ORFs
CB1_63 (thymidylate synthase)	*CB3_68 (thymidylate synthase)*	CB4_67 (thymidylate synthase)	ORF of CB1 and CB3 different to CB4
		2 tRNA genes (tRNA1 and tRNA2)	CB4 has tRNA genes not present in CB1+CB3

Table S3.6. Homologs of the eighteen core proteins described by Li et al. 2016 found present in the genomes of *Pectobacterium* phages CB1, CB3 and CB4.

No.	N4-like Core genes	CB1 homolog	CB3 homolog	CB4 homolog
1	RNA P1	CB1_22	CB3_24	CB4_24
2	RNA P2	CB1_23	CB3_25	CB4_25
3	gp24 N4	CB1_35	CB3_38	CB4_38
4	gp25 N4	CB1_38	CB3_41	CB4_41
5	DNA P/ gp39 N4	CB1_46	CB3_50	CB4_49
6	gp42 N4	CB1_66	CB3_71	CB4_70
7	DNA primase	CB1_67	CB3_72	CB4_71
8	gp44 N4	CB1_70	CB3_75	CB4_74
9	SSB/gp45 N4	CB1_72	CB3_77	CB4_76
10	vRNAP	CB1_77	CB3_82	CB4_81
11	gp53 N4	CB1_80	CB3_85	CB4_84
12	gp54 N4	CB1_81	CB3_86	CB4_85
13	gp55 N4	CB1_82	CB3_87	CB4_86
14	MCP/ gp56 N4	CB1_83	CB3_88	CB4_87
15	gp57 N4	CB1_84	CB3_89	CB4_88
16	94kDa protein/ gp59 N4	CB1_86	CB3_91	CB4_90
17	terminase A	CB1_91	CB3_97	CB4_95
18	gp69	CB1_93	CB3_99	CB4_97

Table S3.7. Putative single-stranded hairpin promoters predicted in the genomes of *Pectobacterium* phages CB1, CB3 and CB4 identified assisted with QuikFold.

Phage	Promotor	Coordinates	Sequence
CB1	Porf1_1	76 - 105 bp	GTGTTGAACCGGTATCCGGTACAGTACCGT
	Porf1_2	823 - 852 bp	ATGGCATCCATGCATCATGGCATCCATATG
	Porf1_5	2085 - 2114 bp	TTGGACCGAGCGTATCGCTCAGCCCACTTA
CB3	Porf3_1	130 - 159 bp	GTGTTGAACCGGTATCCGGTACAGTACCGT
	Porf3_2	919 - 948 bp	ATGGCATCCATGCATCATGGCATCCATATG
	Porf3_4	1850 - 1879 bp	TCGGACCGCACGTATCGTGCAGCCCACTTA
CB4	Porf4_1	76 - 105 bp	GTGTTGAACCGGTATCCGGTACAGTACCGT
	Porf4_2	866 - 895 bp	ATGGCATCCATGCATCATGGCATCCATATG
	Porf4_4	1797 - 1826 bp	TCGGACCGCACGTATCGTGCAGCCCACTTA

Table S3.8. High ΔG rho-independent terminators predicted in the genome *Pectobacterium* phage vB_PatP_CB1 identified using ARNold and QuikFold.

Terminator	Coordinates	Sequence	ΔG kcal/mol
Torf1_1	397 - 428	GCCTACTCTTCGGAGTAGGCTTATTCTTTTCT	-17.2
Torf1_7	2,906 - 2,935	TAACCCCTTCGGGGGTTATCTTATTTTTTA	-13.6
Torf1_20	7,127 - 7,158	GCCATCCCTTCGGGGATGGCTGTTTTATTGAG	-19.3
Torf1_26	11,034 - 11,070	CGGACTCCCTAAGATGGGGAGTCCGTATTTTTTCATA	-19
Torf1_45	25,131 - 25,163	GCCAGCCCTTCGGGGGCTGGTTTTTTAATATCAT	-18.8
Torf1_47	28,088 - 28,113	GCCCCTTCGGGGGCTTTTTTGAGGCT	-13.7
Torf1_51	29,497 - 29,525	AGCCCCTAACGGGGCTTTTTTATTGAGGT	-12.8
Torf1_56	31,765 - 31,793	GCCCACCTAGTGTGGGCTTATATTAATCT	-10.7
Torf1_57	31,751 - 31,782, complement	AGCCCACACTAGGTGGGCTTTTTTATAGCATC	-11.8
Torf1_61	33,531 - 33,561, complement	GGGAGCCTAATGGCTCCCTTTTTAATCTGGA	-15.4
Torf1_72	45,051 - 45,078	GCCCACTTCGGTGGGCTTTTTTATCTAT	-14.7
Torf1_76	46,958 - 46,987	GCCCCTCGATTGAGGGGCTTTATTTTTTAG	-14.5
Torf1_77	46,944 - 46,976, complement	AGCCCCTCAATCGAGGGGCTTTTTATTACGGTA	-15.4
Torf1_81	62,438 - 62,467, complement	GGGAGCTTAAGGCTCCCTTTTCATTGTGAG	-11.9
Torf1_83	63,896 - 63,928, complement	GCCGGGGATAATTCCCCGGCTTTTTTATATCTA	-18.1
Torf1_87	69,766 - 69,796	CTCCCTCTTCAGAGGGAGcTTTTAAACCTGA	-12.1
Torf1_88	70,620 - 70,649	GCCCCACTTCGGTGGGGCTTTTTCCGTTAT	-18
Torf1_89	71,047 - 71,076	GCCCCCGATTAAGGGGGCTTTTTTATTACA	-14.6

Table S3.9. High ΔG rho-independent terminators predicted in the genome *Pectobacterium* phage vB_PatP_CB3 identified using ARNold and QuikFold.

Terminator	Coordinates	Sequence	ΔG kcal/mol
Torf3_1	451 - 482	GCCTACTCTTCGGAGTAGGCTTATTCTTTTCT	-17.2
Torf3_7	2,670 - 2,699	TAACCCCTTCGGGGGTTATCTTATTTTTTA	-13.6
Torf3_21	7,283 - 7,314	GCCATCCCTTCGGGGATGGCTGTTTTATTGAG	-19.3
Torf3_29	11,386 - 11,422	CGGACTCCCCATCTTAGGGAGTCCGTATTTTTTCATA	-17.1
Torf3_57	20,329 - 20,357	GCCAGATTTAATCTGGCTTTTTCTTTTTA	-9.4
Torf3_49	25,440 - 25,472	GCCAGCCCTTCGGGGCTGGTTTTTTAATATCAT	-18.8
Torf3_51	28,397 - 28, 422	GCCCCTTCGGGGGCTTTTTTGAGGCT	-13.7
Torf3_55	29,806 - 29, 834	AGCCCCTAACGGGGCTTTTTTATTGAGGT	-12.8
Torf3_60	32,074 - 32,102	GCCCACCTAGTGTGGGCTTATATTAATCT	-10.7
Torf3_65	33,840 - 33,870, complement	GGGAGCCTAATGGCTCCCTTTTTAATCTGGA	-15.4
Torf3_77	45,360 - 45, 387	GCCCACTTCGGTGGGCTTTTTTATCTAT	-14.7
Torf3_81	47,267 - 47,296	GCCCCTCGATTGAGGGGCTTTATTTTTTAG	-14.5
Torf3_82	47, 253 - 47, 285, complement	AGCCCCTCAATCGAGGGGCTTTTTATTACGGTA	-15.4
Torf3_86	62,747 - 62,776, complement	GGGAGCTTAAGGCTCCCTTTTCATTGTGAG	-11.9
Torf3_88	64,205 - 64,237, complement	GCCGGGGATAATTCCCCGGCTTTTTTATATCTA	-18.4
Torf3_92	69,948 - 69,978	CTCCCTCTTCAGAGGGAGcTTTTAAACCTGA	-12.1
Torf3_94	70,802 - 70,831	GCCCCACTTCGGTGGGGCTTTTTCCGTTAT	-18
Torf3_95	71,229 - 71,258	GCCCCCGATTAAGGGGGCTTTTTTATTACA	-14.6

Table S3.10. High ΔG rho-independent terminators predicted in the genome *Pectobacterium* phage vB_PatP_CB4 identified using ARNold and QuikFold.

Terminator	Coordinates	Sequence	ΔG kcal/mol
Torf4_1	396 - 429	GCCTCATCCTTCGGGATGGGGCTATCTCTTTTCT	-19.6
Torf4_7	2,617 - 2,646	TAACCCCTTCGGGGGTTATCTTATTTTTTA	-13.6
Torf4_21	7,230 - 7,261	GCCATCCCTTCGGGGATGGCTGTTTTATTGAG	-19.3
Torf4_29	11,333 - 11,369	CGGACTCCCCATCTTAGGGAGTCCGTATTTTTTCATA	-17.1
Torf4_45	20,116 - 20,144	GCCAGATTTAATCTGGCTTTTCTTTTTAA	-9.4
Torf4_48	25,226 - 25,258	GCCAGCCCTTCGGGGCTGGTTTTTTAATATCAT	-18.8
Torf4_50	28,183 - 28,208	GCCCCTTCGGGGGCTTTTTTGAGGCT	-13.7
Torf4_54	29,592 - 29,620	AGCCCCTAACGGGGCTTTTTTATTGAGGT	-12.8
Torf4_59	31,860 - 31,888	GCCCACCTAGTGTGGGCTTATATTAATCT	-10.7
Torf4_64	33,826 - 33,656, complement	GGGAGCCTAATGGCTCCCTTTTTAATCTGGA	-15.4
Torf4_76	45,615 - 45,642	GCCCACTTCGGTGGGCTTTTTTATCTAT	-14.7
Torf4_81	47,508 - 47,540, complement	AGCCCCTCAATCGAGGGGCTTTTTATTACGGTA	-15.4
Torf4_80	47,522 - 47,551	GCCCCTCGATTGAGGGGCTTTATTTTTTAG	-14.5
Torf4_85	63,002 - 63,031, complement	GGGAGCTTAAGGCTCCCTTTTCATTGTGAG	-18.1
Torf4_91	70,203 - 70,233	CTCCCTCTTCGGAGGGAGcTTTTAAACCTGA	-15.2
Torf4_92	71,175 - 71,204	GCCCCACTTCGGTGGGGCTTTTTCCGTTAT	-18
Torf4_93	71,602 - 71,631	GCCCCCGATTAAGGGGGCTTTTTTATTACA	-14.6

Table S3.11. Shared high ΔG putative rho-independent terminators among *Pectobacterium* phages CB1, CB3 and CB4

no.	CB1	CB3	CB4
1	Torf1_01	Torf3_01	Torf4_01
2	Torf1_07	Torf3_07	Torf4_07
3	Torf1_20	Torf3_21	Torf4_21
4	Torf1_26	Torf3_29	Torf4_29
5	Torf1_45	Torf3_49	Torf4_48
6	Torf1_47	Torf3_51	Torf4_50
7	Torf1_51	Torf3_55	Torf4_54
8	Torf1_56	Torf3_60	Torf4_59
9	Torf1_61	Torf3_65	Torf4_64
10	Torf1_72	Torf3_77	Torf4_76
11	Torf1_76	Torf3_81	Torf4_80
12	Torf1_77	Torf3_82	Torf4_81
13	Torf1_81	Torf3_86	Torf4_85
14	Torf1_87	Torf3_92	Torf4_91
15	Torf1_88	Torf3_94	Torf4_92
16	Torf1_89	Torf3_95	Torf4_93
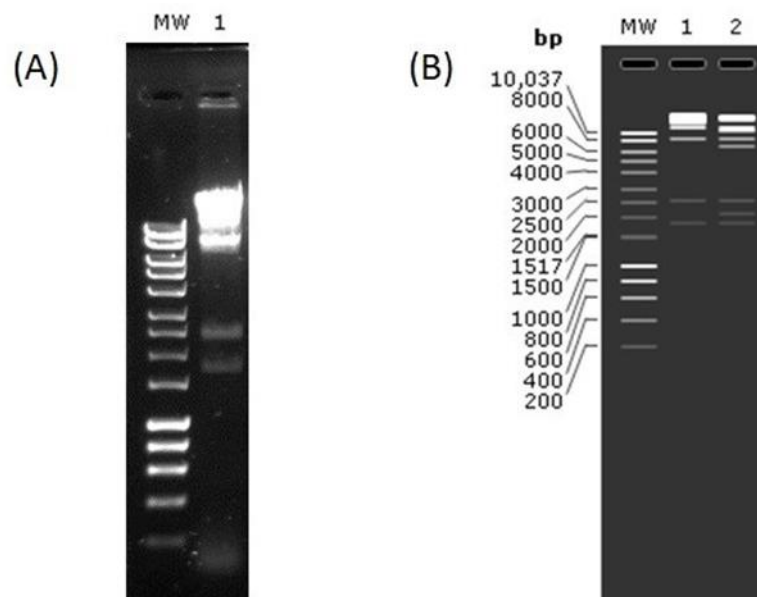
Figure S3.4. (A) Genomic DNA of *Pectobacterium* phage vB_PatP_CB1, which had been digested with restriction enzyme ClaI (lane 1), with DNA marker (Hyperladder 1kb, Bioline) (lane MW). (B) *In silico* digest of CB1 redundant genomic DNA with ClaI with Dam methylation (lane 2); non-Dam methylation (lane 3); DNA marker (Hyperladder 1kb, Bioline) (lane MW). Gel concentration 1 % w/v agarose. Image B was generated using Snapgene.
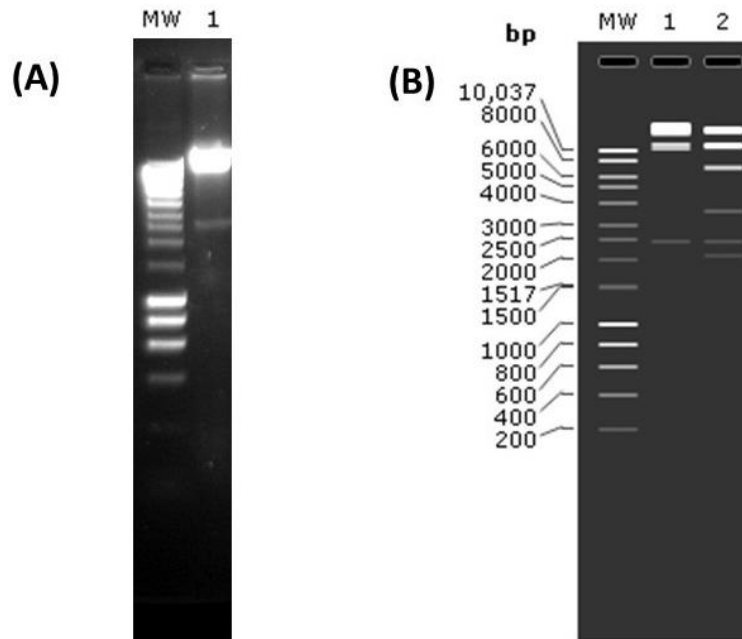
Figure S3.5. (A) Genomic DNA of *Pectobacterium* phage vB_PatP_CB3, which had been digested with restriction enzyme ClaI (lane 1), with DNA marker (Hyperladder 1kb, Bioline) (lane MW). (B) *In silico* digest of CB3 redundant genomic DNA with ClaI with Dam methylation (lane 2); non-Dam methylation (lane 3); DNA marker (Hyperladder 1kb, Bioline) (lane MW). Gel concentration 1 % w/v agarose. Image B was generated using Snapgene.
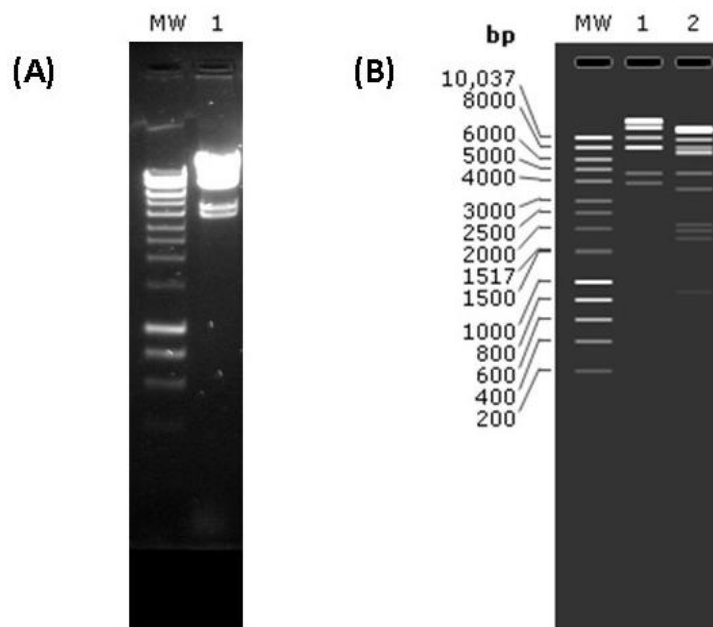
Figure S3.6. (A) Genomic DNA of *Pectobacterium* phage vB_PatP_CB4, which had been digested with restriction enzyme ClaI (lane 1), with DNA marker (Hyperladder 1kb, Bioline) (lane MW). (B) *In silico* digest of CB4 redundant genomic DNA with ClaI with Dam methylation (lane 2); non-Dam methylation (lane 3); DNA marker (Hyperladder 1kb, Bioline) (lane MW). Gel concentration 1 % w/v agarose. Image B was generated using Snapgene.
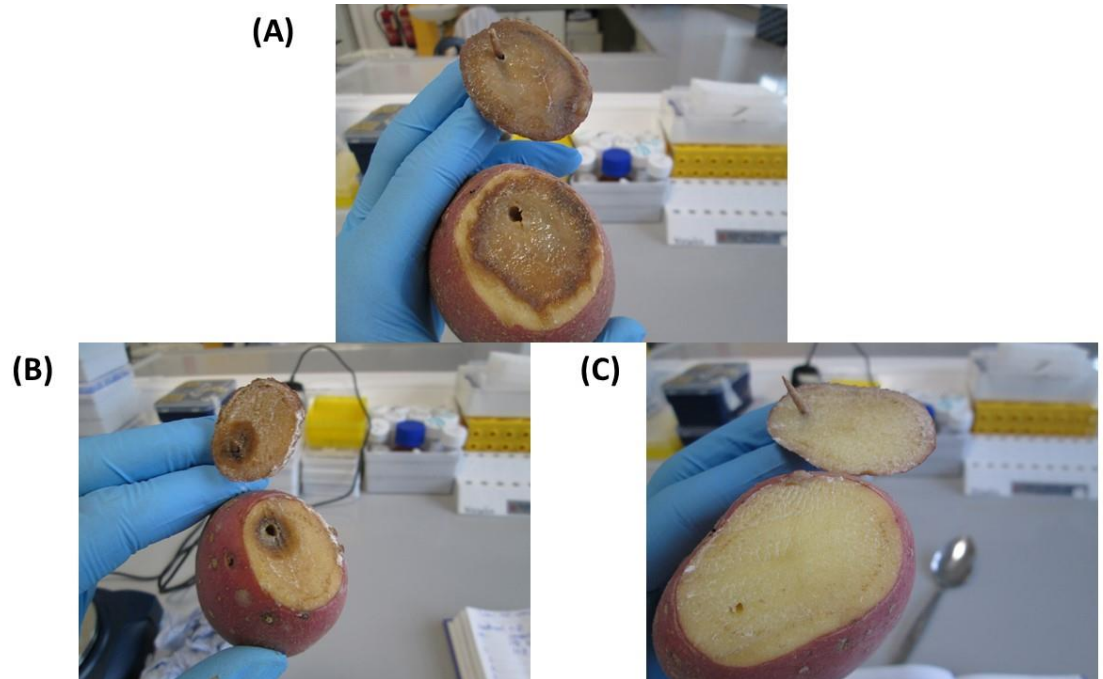
Figure S3.7. Pictures of the typically observed outcomes for the tuber rot assays. A: tuber treated with bacterial strains DSM 18077 + DSM 30186 and SM buffer. B: tuber treated with bacterial strains DSM 18077 + DSM 30186 and phage mixture (CB1 + CB3 + CB4). C: tuber treated with water and SM buffer.
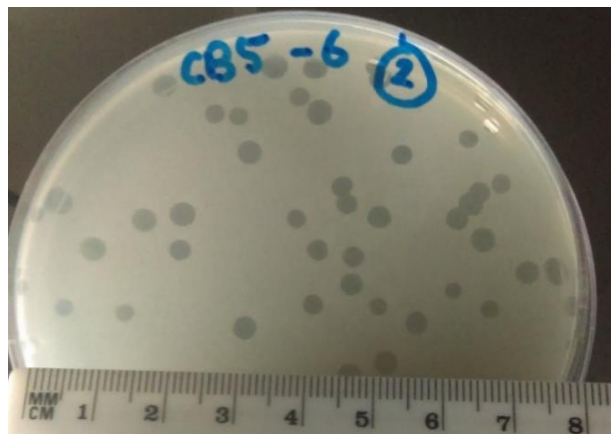
Figure S4.1. *Pectobacterium* phage CB5 plaque morphology on 0.4% w/v LB overlay using host strain *P. atrosepticum* DSM 30186 (12 hr incubation).
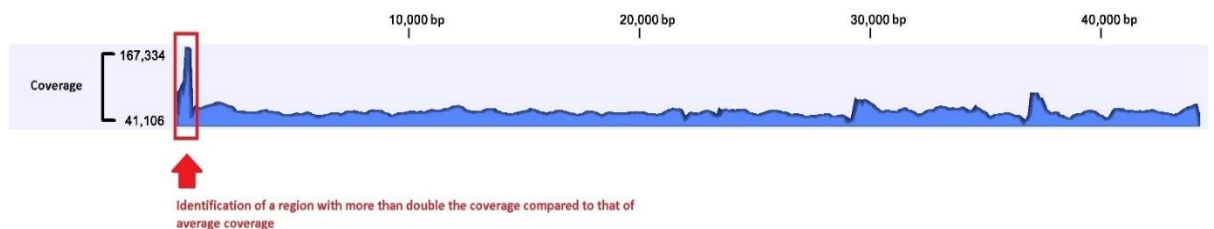
Figure S4.2. Coverage map showing the distribution of reads when mapped back to the contig representing the genome of *Pectobacterium* phage CB5 obtained from genome assembly. Map created with CLC Genomics Workbench.

Table S4.1. Genomic annotation of *Pectobacterium* phage CB5

ORF	Start (bp)	Stop (bp)	Length (aa)	Molecular weight (kDa)	Function
CB5_1	617	450	55	6.67	hypothetical protein
CB5_2	786	944	52	6.02	hypothetical protein
CB5_3	1172	1474	100	11.15	hypothetical protein
CB5_4	1461	1652	63	7.14	hypothetical protein
CB5_5	1652	1780	42	5.25	hypothetical protein
CB5_6	1829	2398	189	20.38	hypothetical protein
CB5_7	2462	2725	87	9.67	hypothetical protein
CB5_8	2802	3050	82	9.87	hypothetical protein
CB5_9	3047	3325	92	10.72	hypothetical protein
CB5_10	3408	3584	58	6.6	hypothetical protein
CB5_11	3581	3760	59	7.07	hypothetical protein
CB5_12	3757	4101	114	13.2	hypothetical protein
CB5_13	4107	4982	291	32.6	hypothetical protein
CB5_14	4995	5852	285	5.81	hypothetical protein
CB5_15	5842	6813	323	35.83	Putative peptidase
CB5_16	6800	7234	144	15.81	hypothetical protein
CB5_17	7227	7664	145	16.92	HNH endonuclease
CB5_18	7661	8464	267	30.25	putative DNA primase
CB5_19	8645	8782	45	5.09	hypothetical protein
CB5_20	8784	8927	47	5.14	hypothetical protein
CB5_21	8927	9088	53	5.77	hypothetical protein
CB5_22	9088	10350	420	46.64	putative DNA helicase
CB5_23	10337	10792	151	17.61	HNH endonuclease
CB5_24	10792	11322	176	19.57	putative nucleotidyl transferase
CB5_25	11366	11596	76	8.69	hypothetical protein
CB5_26	11586	11768	60	6.91	hypothetical protein
CB5_27	11761	11943	60	6.55	hypothetical protein
CB5_28	11997	14339	780	88.55	putative DNA polymerase
CB5_29	14329	14805	158	18.26	HNH endonuclease
CB5_30	14817	15656	279	29.7	hypothetical protein
CB5_31	15670	16716	348	39.15	putative 5' exonuclease
CB5_32	16673	16840	55	6.46	hypothetical protein
CB5_33	16837	17082	81	9.01	hypothetical protein
CB5_34	17075	17500	141	15.97	DNA endonuclease VI
CB5_35	17481	17702	73	8.12	hypothetical protein
CB5_36	17695	18669	324	36.18	putative metallophosphoesterase
CB5_37	18779	19102	107	12.02	hypothetical protein
CB5_38	19213	19395	60	6.83	hypothetical protein
CB5_39	19402	21846	814	91.61	DNA-dependent RNA polymerase

CB5_40	21939	22136	65	7.54	hypothetical protein
CB5_41	22133	22585	150	16.67	HNH endonuclease
CB5_42	22578	22997	139	15.85	hypothetical protein
CB5_43	22990	23316	108	11.2	putative structural protein
CB5_44	23319	24830	503	55.63	putative head-tail connector protein
CB5_45	24830	25699	289	30.47	putative scaffolding protein
CB5_46	25717	26700	327	36.28	putative capsid protein
CB5_47	26760	27317	185	21.13	putative tubular protein A
CB5_48	27327	27809	160	18.47	HNH endonuclease
CB5_49	27829	30129	766	83.39	putative tubular protein B
CB5_50	30126	30740	204	21.53	putative internal virion protein A
CB5_51	30750	33464	904	97.96	putative internal virion protein B
CB5_52	33474	37265	1263	135.4	putative internal virion protein C
CB5_53	37275	38879	534	55.84	tail fiber protein
CB5_54	38888	39175	95	10.38	putative DNA maturase B
CB5_55	39184	41001	605	68.34	putative DNA maturase B
CB5_56	41010	41198	62	6.34	hypothetical protein
CB5_57	41208	41597	129	13.24	putative Rz1A protein
CB5_58	41587	41856	89	10.16	putative holin
CB5_59	41840	42445	201	21.93	putative lysozyme
CB5_60	42460	44217	586	61.92	putative tail spike protein

(A)
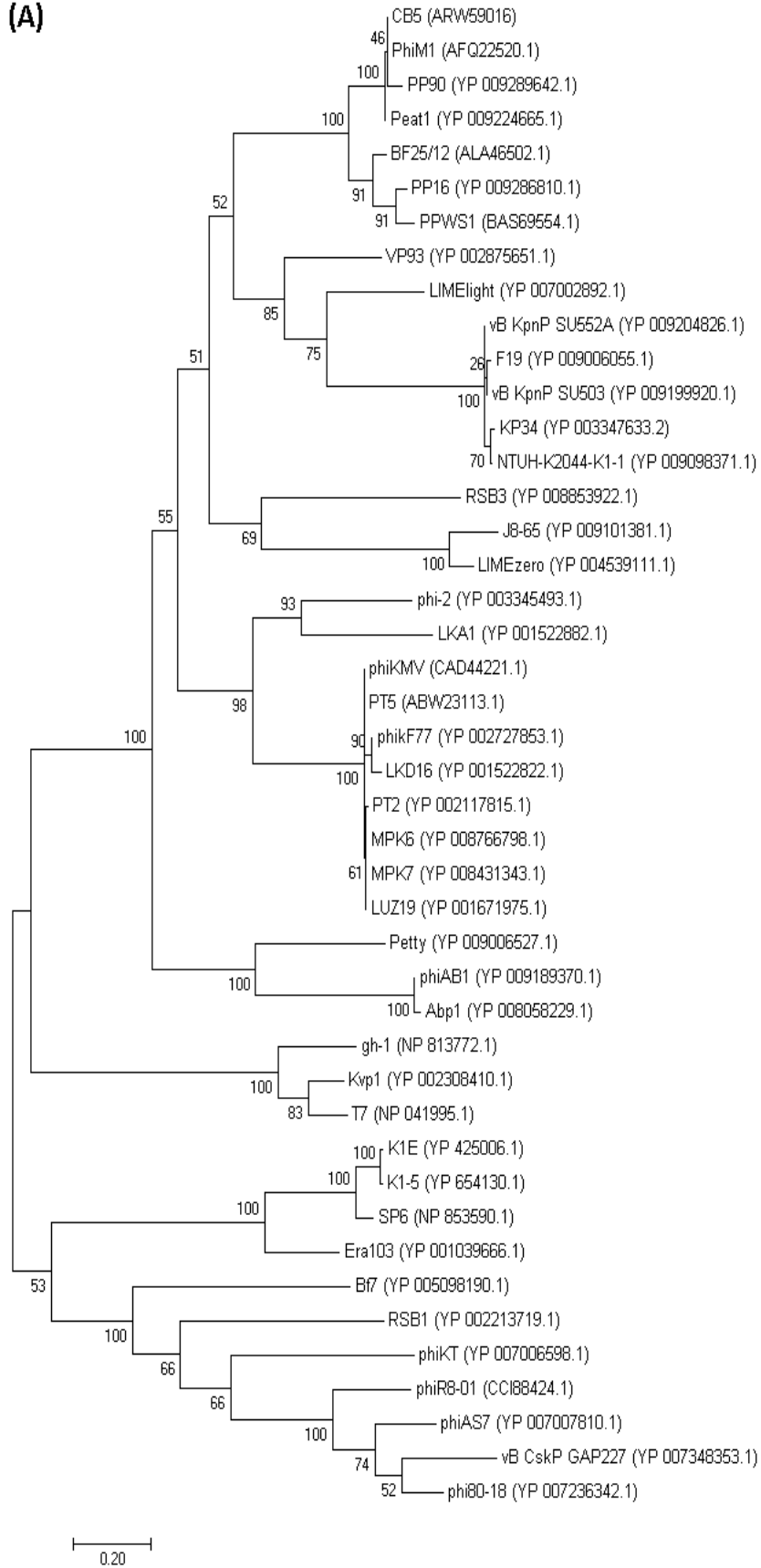
(B)
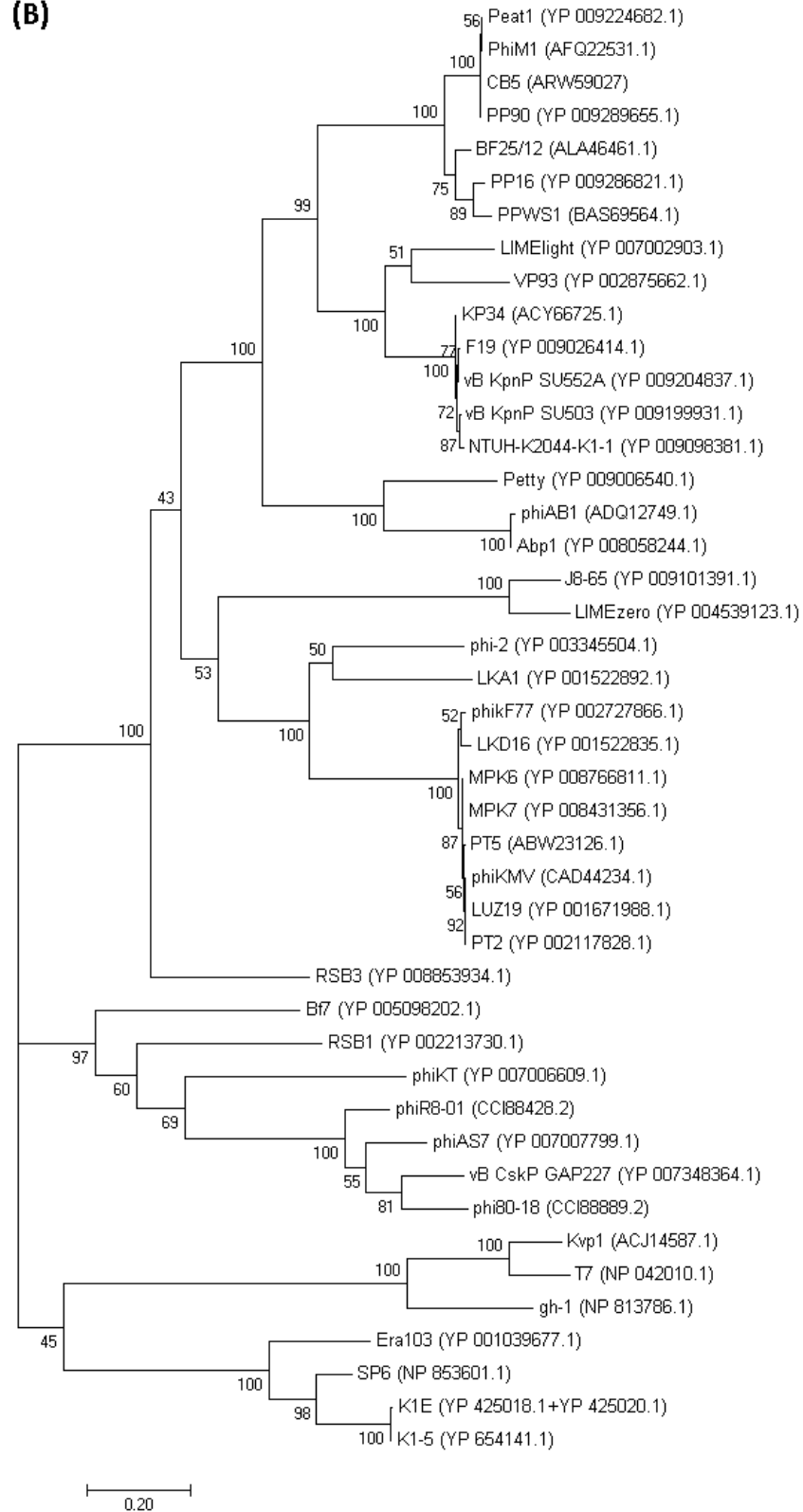
(C)
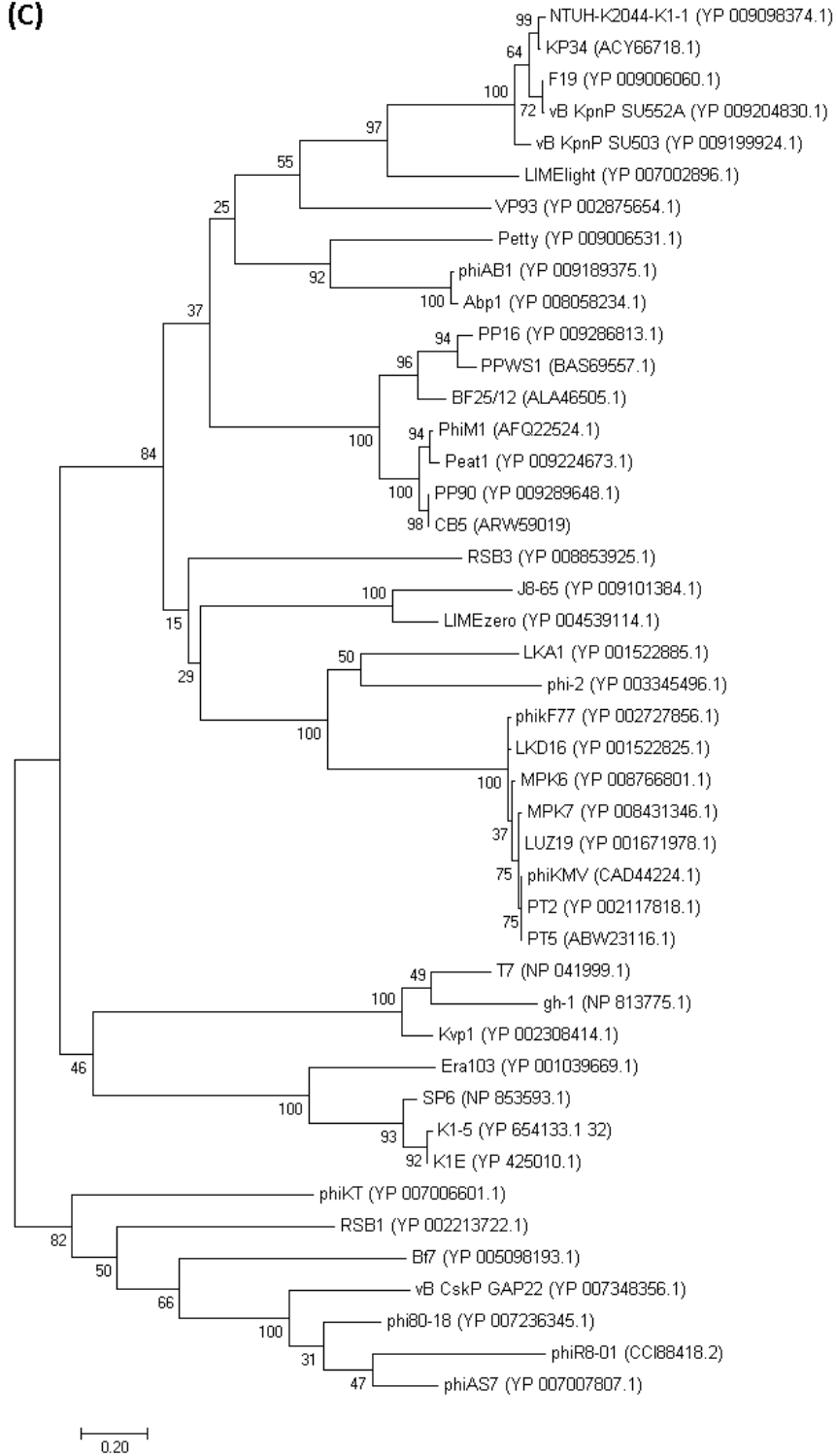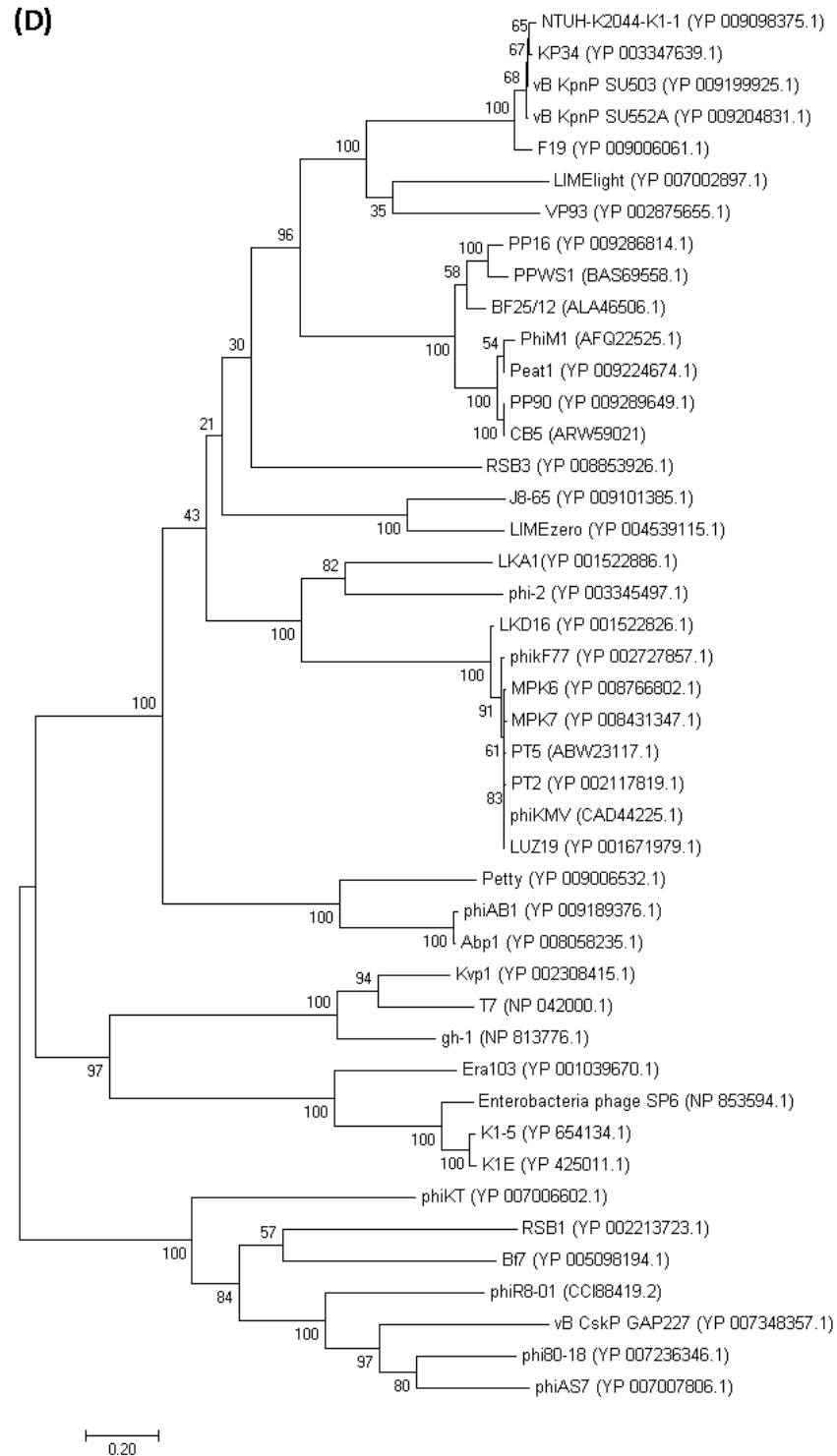
Figure S4.3. Phylogenetic analyses of amino sequences of the head-tail connecting protein (A), terminase (B), tail tube protein A (C) and tail tube protein B (D) of *Pectobacterium* phage CB5 and 52 members of the *Autographivirinae* subfamily, using maximum likelihood (Whelan and Goldman substitution model), with 100 bootstrap replicates.

Table S4.2. Taxonomy output from VICTOR analysis of 52 phages from the subfamily *Autographivirinae* and *Pectobacterium* phage CB5, OPTSIL clustering of taxon boundaries based on the D4 formula at genus and subfamily level. Numbers represent the genus and subfamily which phages have been allocated.

Phage	genus	subfamily
Enterobacteria phage SP6 (NC_004831)	1	1
Escherichia virus K1E (NC_007637)	1	1
Escherichia virus K1-5 (NC_008152)	1	1
Erwinia amylovora phage Era103 (NC_009014)	1	1
Pseudomonad phage gh-1 (AF493143)	3	2
Kluyvera phage Kvp1 (FJ194439)	3	2
Enterobacteria phage T7 (NC_001604)	3	2
Klebsiella phage K11 (NC_011043)	3	2
Klebsiella phage KP32 (NC_013647)	3	2
Escherichia phage K30 (NC_015719)	3	2
Ralstonia phage RSB1 (NC_011201)	2	3
Pseudomonas phage Bf7 (NC_016764)	4	3
Escherichia phage phiKT (NC_019520)	5	3
Yersinia phage phiR8-01 (HE956707)	6	3
Yersinia phage phi80-18 (HE956710)	6	3
Aeromonas phage phiAS7 (NC_019528)	6	3
Cronobacter phage vB_CskP_GAP227 (NC_020078)	6	3
Xanthomonas phage f20-Xaj (KU595432)	9	3
Xanthomonas phage f30-Xaj (KU595433)	9	3
Xylella phage Prado (NC_022987)	9	3
Pseudomonas phage PT5 (EU056923)	7	4
Pseudomonas phage phiKMV (NC_005045)	7	4
Pseudomonas phage LKD16 (NC_009935)	7	4
Pseudomonas phage LKA1 (NC_009936)	7	4
Pseudomonas phage LUZ19 (NC_010326)	7	4
Pseudomonas phage PT2 (NC_011107)	7	4
Pseudomonas phage phikF77 (NC_012418)	7	4
Pseudomonas phage phi-2 (NC_013638)	7	4
Pseudomonas phage MPK7 (NC_022091)	7	4
Pseudomonas phage MPK6 (NC_022746)	7	4
Ralstonia virus phiAp1 (KY117485)	8	4
Ralstonia phage RSB3 (NC_022917)	8	4
Pantoea phage LIMEzero (NC_015585)	10	4
Enterobacteria phage J8-65 (NC_025445)	10	4
Acinetobacter phage vB_ApiP_P2 (MF033351)	11	4
Acinetobacter phage Abp1 (NC_021316)	11	4
Acinetobacter phage Petty (NC_023570)	11	4
Acinetobacter phage phiAB1 (NC_028675)	11	4
Acinetobacter phage Fri1 (NC_028848)	11	4

Vibrio phage VP93 (NC_012662)	12	4
Klebsiella phage KP34 (NC_013649)	12	4
Pantoea phage LIMElight (NC_019454)	12	4
Klebsiella phage F19 (NC_023567)	12	4
Klebsiella phage NTUH-K2044-K1-1 (NC_025418)	12	4
Klebsiella phage vB_KpnP_SU503 (NC_028816)	12	4
Klebsiella phage vB_KpnP_SU552A (NC_028870)	12	4
Pectobacterium phage PhiM1 (JX290549)	13	4
Dickeya phage BF25/12 (KT240186)	13	4
Pectobacterium phage PPWS1 (LC063634)	13	4
Pectobacterium phage Peat1 (NC_029081)	13	4
Pectobacterium phage PP16 (NC_031068)	13	4
Pectobacterium phage PP90 (NC_031096)	13	4
Pectobacterium phage CB5 (KY953156)	13	4

Table S4.3. Shared early region ORFs between PhiM1-like phages as determined by Coregenes

no.	Early gene region	Peat1	CB5	phiM1	PP90
1	hypothetical protein	AXI77_gp61	CB5_3	PhiM1_03	PP90_2
2	hypothetical protein	AXI77_gp01	CB5_6	PhiM1_04	PP90_3
3	hypothetical protein	AXI77_gp07	CB5_12	PhiM1_08	PP90_7
4	hypothetical protein	AXI77_gp08	CB5_13	PhiM1_09	PP90_8
5	peptidase	AXI77_gp09	CB5_15	PhiM1_10	PP90_9
6	hypothetical protein	AXI77_gp10	CB5_16	PhiM1_11	PP90_10

Table S4.4. Proteins of PhiM1-like phage involved in DNA replication, repair and related metabolism

no.	DNA replication & nucleotide metabolism	Peat1	CB5	phiM1	PP90
1	DNA primase	AXI77_gp12 + AXI77_gp13	CB5_18	PhiM1_12	PP90_12
2	DNA helicase	AXI77_gp17	CB5_22	PhiM1_14	PP90_15
3	DNA polymerase	AXI77_gp21	CB5_28	PhiM1_14	PP90_21
4	5' exonuclease	AXI77_gp23	CB5_31	PhiM1_22	PP90_24
5	endonuclease VII	AXI77_gp25	CB5_34	PhiM1_25	PP90_27
6	DNA dependent RNA polymerase	AXI77_gp30	CB5_39	PhiM1_31	PP90_32
7	5' kinase / 3'phosphatase	_	_	PhiM1_28	PP90_29
8	nucleatidyl transferase	_	CB5_24	_	_

Table S4.5. Structural proteins of the PhiM1-like phages

no.	Structural protein	Peat1	CB5	phiM1	PP90
1	head tail connector	AXI77_gp34	CB5_44	PhiM1_35	PP90_38
2	scaffolding protein	AXI77_gp37	CB5_45	PhiM1_36	PP90_40 +PP90_41
3	major capsid	AXI77_gp38+AXI77_gp39	CB5_46	PhiM1_38	PP90_43
4	tubular protein A	AXI77_gp41+AXI77_gp42	CB5_47	PhiM1_39	PP90_44
5	tubular protein B	AXI77_gp43	CB5_49	PhiM1_40	PP90_45
6	internal virion protein A	AXI77_gp44	CB5_50	PhiM1_41	PP90_46
7	internal virion protein B	AXI77_gp45	CB5_51	PhiM1_42	PP90_47
8	internal virion protein C	AXI77_gp46+AXI77_gp47	CB5_52	PhiM1_43	PP90_48
9	tail fibre	AXI77_gp48	CB5_53	PhiM1_44	PP90_49
10	large terminase	AXI77_gp50	CB5_54	PhiM1_45	PP90_50
11	small terminase	AXI77_gp51	CB5_55	PhiM1_46	PP90_51
12	tail spike	AXI77_gp57	CB5_60	PhiM1_52	PP90_56

Table S4.6. Proteins of lysis cassette of the PhiM1-like phages

no.	Lysis protein	Peat1	CB5	phiM1	PP90
1	U-spanin	AXI77_gp54	CB5_57	PhiM1_49	PP90_53
2	Holin	AXI77_gp55	CB5_58	PhiM1_50	PP90_54
3	Endolsyin	AXI77_gp56	CB5_59	PhiM1_51	PP90_55
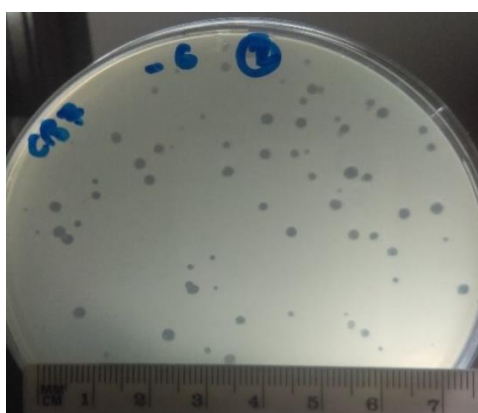
Figure S5.1. *Pectobacterium* phage CB7 plaque morphology on 0.4% w/v LB overlay using host strain *P. atrosepticum* DSM 30186 (12 hr incubation).

Table S5.1. Genome annotation of *Pectobacterium* phage CB7

ORF	Start	Stop	Length (aa)	Molecular weight (kDa)	Predicted function
CB7_1	1	384	127	14.12	large terminase (Part 2)
CB7_2	678	1178	166	18.59	HNH homing endonuclease
CB7_3	1348	2457	369	41.94	large terminase (Part 1)
CB7_4	2476	3999	507	55.37	putative portal protein
CB7_5	4069	4644	191	21.35	putative prohead core protein protease
CB7_6	4641	5750	369	40.57	structural protein
CB7_7	5768	6217	149	15.6	putative head stabilization/decoration protein
CB7_8	6237	7235	332	37.6	putative major head protein
CB7_9	7356	7970	204	22.67	Conserved hypothetical protein
CB7_10	8031	11330	1099	111.95	virion structural protein
CB7_11	11369	12019	216	24.54	structural protein
CB7_12	12021	12515	164	17.19	structural protein
CB7_13	12614	13141	175	20.14	structural protein
CB7_14	13204	13665	153	17.4	Conserved hypothetical protein
CB7_15	13676	14116	146	16.58	structural protein
CB7_16	14116	14841	241	26.47	structural protein
CB7_17	14902	16323	473	50.69	structural protein
CB7_18	16327	16806	159	17.2	putative tail tuber protein
CB7_19	16878	17354	158	17.21	conserved hypothetical protein
CB7_20	17441	17623	60	6.95	conserved hypothetical protein
CB7_21	17701	17817	38	4.33	conserved hypothetical protein
CB7_22	17821	18771	316	34.88	conserved hypothetical protein
CB7_23	18771	21203	810	87.72	putative tape measure protein
CB7_24	21272	22126	284	31.67	structural protein
CB7_25	22129	22503	124	14.38	structural protein
CB7_26	22503	23495	330	36.5	putative tail protein
CB7_27	23505	24248	247	26.42	putative baseplate assembly protein
CB7_28	24257	24790	177	20.45	putative tail lysozyme
CB7_29	24880	26364	494	54.37	putative baseplate assembly protein
CB7_30	26374	27018	214	24.2	putative baseplate protein
CB7_31	27029	28693	554	59.21	phage Tail Collar Domain containing protein
CB7_32	28699	29214	171	19.4	tail fiber assembly protein
CB7_33	29256	29594	112	13.09	conserved hypothetical membrane protein
CB7_34	29581	30049	152	17.41	conserved hypothetical membrane protein
CB7_35	30073	30333	86	9.89	conserved hypothetical protein
CB7_36	30346	32727	793	81.89	putative tail fiber protein
CB7_37	32766	34007	413	43.48	structural protein

CB7_38	34544	34041	167	19.02	HNH homing endonuclease
CB7_39	35770	34544	408	43.15	conserved hypothetical protein
CB7_40	36320	35946	124	14.68	conserved hypothetical protein
CB7_41	37133	36441	230	24.64	conserved hypothetical protein
CB7_42	37519	37130	129	14.07	hypothetical protein
CB7_43	37850	37521	109	12.77	conserved hypothetical protein
CB7_44	38239	37853	128	14.68	hypothetical membrane protein
CB7_45	38445	38236	69	8.16	conserved hypothetical protein
CB7_46	38426	38707	93	11.03	conserved hypothetical protein
CB7_47	38707	38910	67	7.72	conserved hypothetical protein
CB7_48	39006	39491	161	17.95	HNH homing endonuclease
CB7_49	39493	39843	116	12.89	DNA polymerase (part 3)
CB7_50	40327	39944	127	14.91	HNH homing endonuclease
CB7_51	40809	40474	111	12.78	DNA polymerase (part 2)
CB7_52	41393	40944	149	17.72	HNH homing endonuclease
CB7_53	43520	41523	665	75.98	DNA polymerase (part 1)
CB7_54	44064	43498	188	21.95	HNH homing endonuclease
CB7_55	44848	44132	238	26.53	structural protein
CB7_56	45027	44848	59	6.59	conserved hypothetical protein
CB7_57	45277	45020	85	9.6	conserved hypothetical protein
CB7_58	45741	45277	154	18	conserved hypothetical protein
CB7_59	46120	45719	133	15.04	restriction alleviation protein Lar
CB7_60	46479	46117	120	13.63	conserved hypothetical protein
CB7_61	47814	47176	212	23.88	conserved hypothetical protein
CB7_62	47980	47801	59	6.43	HNH homing endonuclease
CB7_63	48778	48077	233	26.1	putative DNA methylase
CB7_64	48771	49034	87	9.77	conserved hypothetical protein
CB7_65	49031	49210	59	6.65	conserved hypothetical protein
CB7_66	49383	49207	58	6.4	hypothetical membrane protein
CB7_67	49989	49717	90	10.42	conserved hypothetical protein
CB7_68	50179	49982	65	7.35	conserved hypothetical protein
CB7_69	50388	50176	70	8.65	conserved hypothetical protein
CB7_70	50550	50407	47	5.25	hypothetical membrane protein
CB7_71	50846	50547	99	11.11	conserved hypothetical protein
CB7_72	50843	51127	94	10.88	hypothetical protein
CB7_73	53328	51214	704	79.27	putative primase/helicase
CB7_74	53767	53492	91	10.29	conserved hypothetical protein
CB7_75	53994	53803	63	7.37	conserved hypothetical protein
CB7_76	54206	53991	71	8.48	conserved hypothetical protein
CB7_77	54348	54184	54	6.23	conserved hypothetical protein
CB7_78	54571	54341	76	8.24	conserved hypothetical protein
CB7_79	54782	54564	72	7.74	conserved hypothetical protein
CB7_80	57250	54779	823	92.19	putative helicase (with intein)
CB7_81	57435	57250	61	6.56	conserved hypothetical protein
CB7_82	57643	57446	65	7.16	hypothetical membrane protein

CB7_83	58170	57640	176	19.88	SleB
CB7_84	58600	58220	126	14.05	conserved hypothetical protein
CB7_85	59135	58593	180	19.79	putative phosphatase
CB7_86	59125	59337	70	8.12	conserved hypothetical protein
CB7_87	59566	59396	56	6.31	conserved hypothetical protein
CB7_88	59793	59563	76	9.07	conserved hypothetical protein
CB7_89	59969	59871	32	3.79	conserved hypothetical protein
CB7_90	60105	59959	48	5.28	conserved hypothetical protein
CB7_91	60332	60162	56	6.7	conserved hypothetical protein
CB7_92	60774	60325	149	17.08	conserved hypothetical protein
CB7_93	61006	60818	62	7.25	conserved hypothetical protein
CB7_94	61329	61006	107	11.67	conserved hypothetical protein
CB7_95	61534	61316	72	7.71	conserved hypothetical protein
CB7_96	61878	61534	114	13.31	conserved hypothetical protein
CB7_97	62117	61875	80	9.16	hypothetical protein
CB7_98	62438	62175	87	10.08	hypothetical membrane protein
CB7_99	62703	62422	93	10.68	conserved hypothetical protein
CB7_100	62962	62693	89	9.94	putative lipoprotein
CB7_101	63192	62959	77	8.81	conserved hypothetical protein
CB7_102	63371	63189	60	6.83	conserved hypothetical protein
CB7_103	64002	63475	175	20.5	conserved hypothetical protein
CB7_104	64126	65367	413	48.03	conserved hypothetical protein
CB7_105	65357	65527	56	6.63	conserved hypothetical protein
CB7_106	65517	65720	67	7.78	conserved hypothetical protein
CB7_107	65717	66247	176	20.32	conserved hypothetical protein
CB7_108	66313	67551	412	46.86	MoxR
CB7_109	67583	68095	170	19.49	HNH homing endonuclease
CB7_110	68097	69647	516	58.39	conserved hypothetical protein
CB7_111	69656	70036	126	14.39	conserved hypothetical protein
CB7_112	70033	70546	167	19.44	hypothetical protein
CB7_113	70768	71010	80	8.93	conserved hypothetical protein
CB7_114	71198	71320	40	4.49	conserved hypothetical protein
CB7_115	71320	71670	116	13.72	hypothetical membrane protein
CB7_116	71784	71981	65	7.16	conserved hypothetical protein
CB7_117	71971	72510	179	20.87	conserved hypothetical protein
CB7_118	72503	72712	69	8.02	conserved hypothetical protein
CB7_119	72790	73020	76	8.49	conserved hypothetical protein
CB7_120	73208	73438	76	8.8	conserved hypothetical protein
CB7_121	73440	73721	93	10.58	conserved hypothetical protein
CB7_122	73718	73975	85	9.45	conserved hypothetical protein
CB7_123	74035	74871	278	31.59	conserved hypothetical protein
CB7_124	76170	75748	140	16.32	conserved hypothetical protein
CB7_125	76169	76495	108	11.96	conserved hypothetical protein
CB7_126	76575	76796	73	7.82	conserved hypothetical protein
CB7_127	76843	77190	115	12.82	conserved hypothetical protein

CB7_128	77290	77559	89	10.1	conserved hypothetical protein
CB7_129	77846	78076	76	8.19	hypothetical membrane protein
CB7_130	78087	78539	17	16.98	conserved hypothetical protein
CB7_131	78903	79289	128	14.09	conserved hypothetical protein
CB7_132	79299	79610	103	12	hypothetical membrane protein
CB7_133	79613	79747	44	5.5	hypothetical protein
CB7_134	79824	80063	79	8.56	conserved hypothetical protein
CB7_135	80159	80596	145	16.43	conserved hypothetical protein
CB7_136	80661	80840	59	6.91	hypothetical protein
CB7_137	80892	81149	85	9.54	conserved hypothetical protein
CB7_138	81363	81542	59	6.6	putative lipoprotein
CB7_139	81605	81901	98	10.73	conserved hypothetical protein
CB7_140	82114	82320	68	7.54	conserved hypothetical protein
CB7_141	82770	83315	181	20.69	conserved hypothetical protein
CB7_142	83415	83615	66	7.47	conserved hypothetical protein
CB7_143	85143	85628	161	18.35	HNH homing endonuclease
CB7_144	86073	85861	70	8	conserved hypothetical protein
CB7_145	86386	86066	106	12.17	conserved hypothetical protein
CB7_146	86702	86379	107	12.2	conserved hypothetical protein
CB7_147	87047	86712	111	12.48	conserved hypothetical protein
CB7_148	87220	87044	58	6.44	conserved hypothetical protein
CB7_149	87567	87343	74	8.51	conserved hypothetical protein
CB7_150	87713	87567	48	5.55	conserved hypothetical protein
CB7_151	87964	87713	83	9.35	conserved hypothetical protein
CB7_152	88438	87968	156	18.17	conserved hypothetical protein
CB7_153	88638	88438	66	7.45	conserved hypothetical protein
CB7_154	88868	88638	76	8.61	hypothetical protein
CB7_155	89147	88881	88	10.69	conserved hypothetical protein
CB7_156	89503	89144	119	13.93	conserved hypothetical protein
CB7_157	89853	89626	75	8.83	conserved hypothetical protein
CB7_158	90395	89850	181	20.58	conserved hypothetical protein
CB7_159	90564	90376	62	6.71	conserved hypothetical protein
CB7_160	90853	90524	109	12.51	conserved hypothetical protein
CB7_161	91080	90853	75	8.78	conserved hypothetical protein
CB7_162	88438	87968	79	8.92	conserved hypothetical protein
CB7_163	91510	91397	37	4.01	conserved hypothetical protein
CB7_164	92010	91510	166	19.08	conserved hypothetical protein
CB7_165	92361	92023	112	13.06	conserved hypothetical protein
CB7_166	92706	92497	69	7.28	conserved hypothetical protein
CB7_167	93005	92685	106	11.57	conserved hypothetical protein
CB7_168	93453	92992	153	17.69	conserved hypothetical protein
CB7_169	93971	93456	171	19.93	conserved hypothetical protein
CB7_170	94176	94027	49	5.67	conserved hypothetical protein
CB7_171	94860	94417	147	16.45	hypothetical membrane protein
CB7_172	95659	95066	197	22.54	conserved hypothetical protein

CB7_173	95921	95661	86	9.93	conserved hypothetical protein
CB7_174	96430	95930	166	18.69	conserved hypothetical protein
CB7_175	97542	96434	372	42.96	putative nucleotidyltransferase
CB7_176	98030	97629	133	14.6	hypothetical membrane protein
CB7_177	98428	98030	132	15.33	conserved hypothetical protein
CB7_178	98724	98425	99	10.7	hypothetical membrane protein
CB7_179	100178	98853	441	50.54	tRNA nucleotidyl transferase
CB7_180	100681	100172	169	19.86	HNH homing endonuclease
CB7_181	101010	100678	110	13.22	conserved hypothetical protein
CB7_182	101439	101095	114	12.72	putative HNH endonuclease
CB7_183	102106	101567	179	20.5	putative DNA methylase, N-6 adenine-specific
CB7_184	103008	102106	300	33.97	ClpP ATP-dependent protease subunit
CB7_185	103484	103008	158	18.38	conserved hypothetical protein
CB7_186	103812	103495	105	12.08	conserved hypothetical protein
CB7_187	104282	103812	156	17.88	putative NrdG
CB7_188	106381	104279	700	78.38	putatice NrdD
CB7_189	107127	106378	249	28.36	phosphate starvation-inducible phoH-like protein
CB7_190	107526	107143	127	14.48	endolysin
CB7_191	107948	107535	137	15.62	conserved hypothetical protein
CB7_192	108246	107959	95	10.96	glutaredoxin
CB7_193	108470	108246	74	8.37	conserved hypothetical protein
CB7_194	108691	108470	73	8.16	conserved hypothetical protein
CB7_195	109088	108702	128	14.27	ribonucleotide reductase small subunit, NrdB (part 2)
CB7_196	109953	109246	235	27.18	HNH homing endonuclease
CB7_197	111732	110083	549	63.26	ribonucleotide reductase small subunit, NrdB (part 1)
CB7_198	112000	111830	56	7.12	conserved hypothetical protein
CB7_199	112107	111997	36	4.22	hypothetical protein
CB7_200	112339	112109	76	8.87	hypothetical protein
CB7_201	114789	112375	804	90.46	ribonucleotide reductase large subunit, NrdA (part 2)
CB7_202	115425	114934	163	18.61	HNH homing endonuclease
CB7_203	116424	115543	293	33.04	ribonucleotide reductase large subunit, NrdA (part 1)
CB7_204	116445	116675	76	8.84	conserved hypothetical protein
CB7_205	116678	117172	164	18.61	conserved hypothetical protein
CB7_206	117939	117259	226	25.68	thymidylate synthase
CB7_207	118235	117942	97	10.8	conserved hypothetical protein
CB7_208	118822	118640	60	7.16	conserved hypothetical protein
CB7_209	119433	118822	203	23.35	putative 5'(3')-deoxyribonucleotidase
CB7_210	120079	119423	218	24.76	conserved hypothetical protein

CB7_211	121247	120165	360	41.45	putative DNA polymerase/exonuclease
CB7_212	121799	121263	178	20.66	HNH homing endonuclease
CB7_213	122029	121799	76	9.13	conserved hypothetical protein
CB7_214	122534	122022	170	19.34	endoVII packaging and recombination endonuclease
CB7_215	123171	122536	211	24.48	conserved hypothetical protein
CB7_216	123751	123128	207	22.88	conserved hypothetical protein
CB7_217	124064	123732	110	12.82	conserved hypothetical protein
CB7_218	125179	124064	371	42.43	exonuclease
CB7_219	125676	125179	165	19.48	HNH homing endonuclease
CB7_220	126179	125676	167	19.58	HNH homing endonuclease
CB7_221	126681	126169	170	19.59	HNH homing endonuclease
CB7_222	127216	126665	183	21.28	conserved hypothetical protein
CB7_223	127628	127290	112	12.51	conserved hypothetical protein
CB7_224	127993	127628	121	13.76	conserved hypothetical protein
CB7_225	128205	127993	70	8.42	conserved hypothetical protein
CB7_226	128422	128213	69	7.35	conserved hypothetical membrane protein
CB7_227	128940	128419	173	20.49	conserved hypothetical protein
CB7_228	129323	129039	94	10.79	conserved hypothetical protein
CB7_229	129705	129397	102	11.64	conserved hypothetical protein
CB7_230	130088	129717	123	14.64	HNH homing endonuclease
CB7_231	130393	130217	58	6.48	conserved hypothetical protein
CB7_232	130755	130393	120	13.84	pyrophosphatase
CB7_233	131250	130810	146	17.42	hypothetical membrane protein
CB7_234	132626	131247	459	52.12	DNA ligase
CB7_235	132817	132623	64	7.49	conserved hypothetical protein
CB7_236	133299	132904	131	14.6	conserved hypothetical protein
CB7_237	133622	133296	108	12.22	conserved hypothetical protein
CB7_238	133918	133619	99	11.73	conserved hypothetical protein
CB7_239	134025	133909	38	4.57	conserved hypothetical protein
CB7_240	134348	134025	107	12.56	conserved hypothetical protein
CB7_241	134670	135041	123	14.37	conserved hypothetical protein
CB7_242	135096	135986	296	32.89	ribose-phosphate pyrophosphokinase
CB7_243	135970	136527	185	20.73	HNH homing endonuclease
CB7_244	136538	137374	278	31.7	nicotinamide phosphoribosyl transferase (part 2)
CB7_245	134356	137937	193	21.56	HNH homing endonuclease
CB7_246	137934	138932	332	37.41	nicotinamide phosphoribosyl transferase (part 1)
CB7_247	138978	139169	63	7	conserved hypothetical protein
CB7_248	139229	139480	83	9.64	conserved hypothetical protein
CB7_249	139638	140030	130	14.8	HNH homing endonuclease
CB7_250	140312	140686	124	13.54	conserved hypothetical protein

CB7_251	141337	141891	184	19.68	structural protein
CB7_252	141988	142359	123	14.11	conserved hypothetical membrane protein
CB7_253	142359	142673	104	11.7	putative lipoprotein

Table S5.2. Sigma 70 – like putative promoter sequence detected in the genome of *Pectobacterium* phage CB7

Promoter	Start	Stop	Sequence
Porf56	45103	45065	TACAACATTGTTGTCGCCAAAGGCGGGGACCGCAATAAC
Porf79	54853	54815	GACATTATTGTTGACCAAGGTGTCCGTAAGTTCTATCAT
Porf97	62188	62188	TAAAATTCGGTTGACAGTAGAGATCGGGGCCTGTAATAT
Porf125	76101	76139	AAATCATTTGTTGACAACCTGGCCGGGCCGGGCCACAAT
Porf126	76489	76527	TAAATAGTTGTTGACACCGGGGCCACGTGGCCCCATACT
Porf128	77200	77238	GTTTTAGGGGTTGACGCCCGGCGGCGTTACCCCCATAAT
Porf129	77,776	77814	AAAATAATTGTTGACGCCGGGCACCGTCCCGGCATACAA
Porf131	78811	78849	AAAATAGTTGTTGACTCTTGCCCCGTCCCGCTGGCATAA
Porf135	80067	80105	AAAAAGGGGCTTGACGTTGTGAAACGTTAGGCCCTATAA
Porf136	80590	80628	AAAATAAAAGTTGACACCGGGGCCACCTGGCCCCATAAT
Porf138	81288	81326	GAAAATAAACTTGACGCCGGGACGGTTCCCGGCTTATAA
Porf139	81536	81574	AAAATAAATGTTGACGCCGGGCACCGTCCCGGCTATACT
Porf140	82045	82083	AAAAATTAGTTTGACACCCGGCCCGGAGGGGCCGATAAT

Table S5.3. Sigma 70-like promoters detected in the genome of *Cronobacter* phage CR3 (accession no. JQ691612)
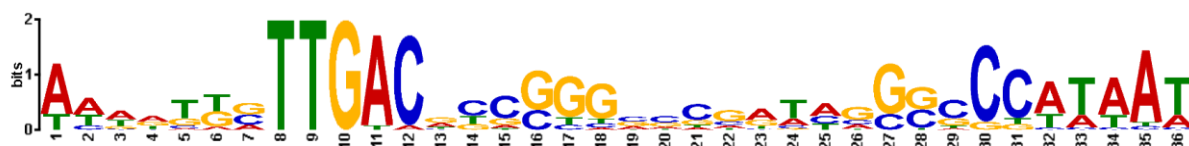
Promoter	Start	Stop	Sequence
pcr3_53	41845	41810	AAAATTGTTGACTTCCTTCCTGATTTCCCGCATAAT
pcr3_54	42349	42314	TTTTTAATTGACAGCGGGGCCGAAAGGCCCCATAAT
pcr3_59	45981	45946	AAAATTCTTGTATTCCGGGGGCTTCGGCCCCTTTTC
pcr3_128	80533	80568	ATTTTTCTTGACCTGCCAGGCAAGCCGGGCCACAAT
pcr3_129	80932	80967	ATAATGCTTGACGGGGCCAGCGATAGGCCCCATAAT
pcr3_130	81471	81506	AAGAGTATTGACACCGGGGCCACCTGGCCCCATACT
pcr3_132	82181	82216	TTTGGGGTTGACGCCCGGCCAGGATGGGTCCATAAT
pcr3_133	82513	82548	ATAATTGTTGACCGGGGCCGGGTTAGGCCCCATAAT
pcr3_135	91606	91571	TCTTTTGTTGACACCGGGGCCGATAGGCCCCATAAT
pcr3_135	83275	83310	ATTTTGGTTGACACGGGGCCGGGTTAGGCCCCATAA
pcr3_139	84806	84841	AAGGGGCTTGACGTCGTGAAACGTTAGGCCCTATAA
pcr3_141	85882	85917	AAAGCGCTTGACGCCGGGACGGTTCCCGGCTTATAA
pcr3_143	86335	86370	ATAATTCTTGACGCCGGGCACCGTCCCGGCTTACAA
pcr3_144	87078	87113	TAAATGATTGACATCCGGCCCGAAAGGGCCGATAAT
pcr3_145	87356	87391	AAAGTAGTTGACGGCGGGGCCGATAGGCCCCATAAT
pcr3_147	87898	87933	ACTATTGTTGACGCCGGGCACCGTCCCGGCCTATAA
pcr3_150	89365	89400	TCCGGGGTTGACACCCGGCGCAATAGGGCCGATAAT
pcr3_157	91952	91987	AATAATATTGACTCAGGGGCCGAAAGGCCCCATAAT
pcr3_242	133519	133484	AAAGTTGTTGACACACTGAGTTAAAACGAGCATAAT
pcr3_243	133519	133484	AAACCGCTTGACGTTGGTTTCTGCCGGGTCCATAAT
pcr3_248	136375	136340	AAGTGGCTTGACGCCGGGATCGATAAGTGCCTTAAT
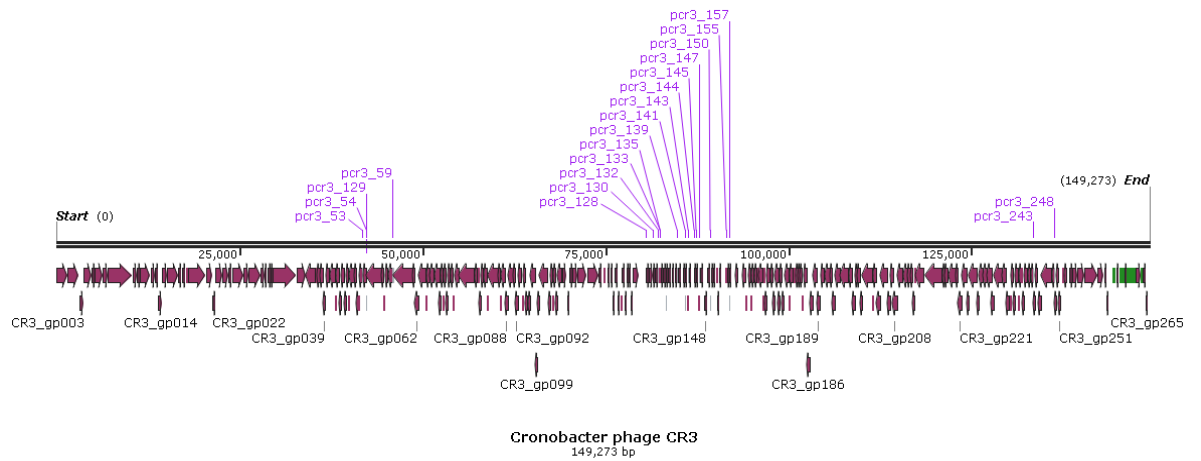
Figure S5.2. Genome map of *Cronobacter* phage CR3 (accession no. JQ691612) showing locations of Sigma 70-like. The map comprises of arrows indicating the location of ORFs. Genome map created with Snapgene.

Table S5.4. Sigma 70-like promoters detected in the genome of *Cronobacter* phage CR8 (accession no. KC954774).

Promoter	Start	Stop	Sequence
pcr8_54	42743	42706	ACAAAATTGTTGACTTCCTTCCTGATTTCCCGCATAAT
pcr8_55	43247	43210	ATTTTTTAATTGACAGCGGGGCCGAAAGGCCCCATAAT
pcr8_86	59955	59918	ACATTATTGTTGACCAAGGTGTCCGTAAGTTCTATCAT
pcr8_132	81405	81442	AAATTTTTCTTGACCTGCCAGGCAAGCCGGGCCACAAT
pcr8_133	81,804	81841	AAATAGTGCTTGACGGGGCCGATGATAGGCCCCATAAT
pcr8_134	82342	82379	AAAAGACTATTGACATCGGGGCCACCTGGCCCCATACT
pcr8_136	83054	83091	TTTTAGGGGTTGACGCCCGGCGGCGTTACCCTCATAAT
pcr8_137	83414	83451	AAATAATTGTTGACGCCGGGAACCGTCCCGGCTTATAA
pcr8_138	83740	83777	AAATAATTGTTGACCGGGGCCGGGTTAGGCCCCATAAT
pcr8_140	84503	84540	TAATTTTGGTTGACACGGGGCCGGGTTAGGCCCCATAA
pcr8_144	86044	86081	TCCCGGGGCTTGACGTCCGGCCCCGGTTCCTCCATAAT
pcr8_146	87117	87154	TAAAAGCGCTTGACGCCGGGACGGTTCCCGGCTTATAA
pcr8_148	87570	87607	AAATAATTGTTGACGCCGGGCACCGTCCCGGCTTACAA
pcr8_149	88313	88350	TTTAAATGATTGACATCCGGCCCAAAAGGGCCGATAAT
pcr8_150	88587	88624	AAAAAGTAGTTGACGGCGGGGCCGATAGGCCCCATAAT
pcr8_151	88906	88943	AAATAATGCTTGACGGGGCCAACGTTAGGCCCCATAAT
pcr8_154	90372	90409	GATCCGGGGTTGACACCCGGCGCAATAGGGCCGATAAT
pcr8_160	92418	92455	AATCTTTTGTTGACACCGGGGCCGATAGGCCCCATAAT
pcr8_161	92798	92835	AAAATAATATTGACTCAGGGGCCGAAAGGCCCCATAAT
pcr8_246	134079	134042	AAAAAGTTGTTGACACACTGAGTTAAAACGAGCATAAT
pcr8_250	136935	136898	AAAAGTGGCTTGACGCCGGGATCGGTAAGTGCCTTAAT

240
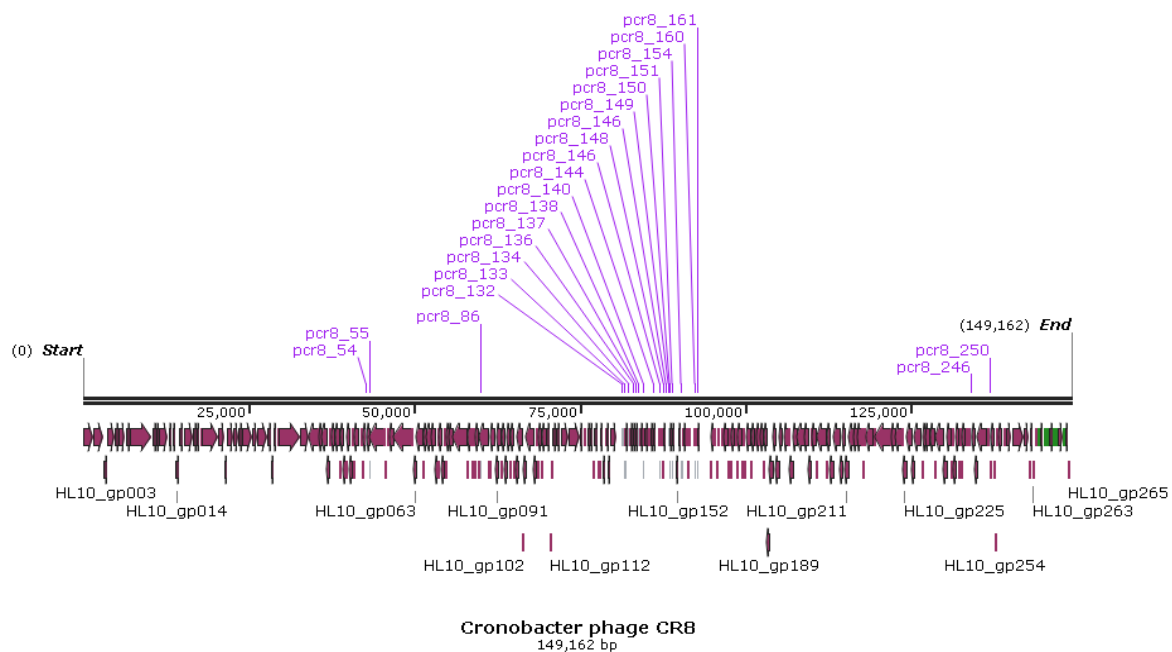
Figure S5.3. Genome map of *Cronobacter* phage CR8 (accession no. KC954774) showing locations of Sigma 70-like. The map comprises of arrows indicating the location of ORFs. Genome map created with Snapgene.

Table S5.5. Sigma 70-like promoters detected in the genome of *Cronobacter* phage CR9 (accession no. JQ691611).
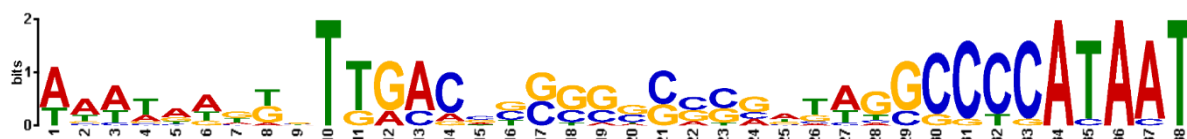
| Promoter	Start	Stop	Sequence
54	39759	39722	ACAAAAGTATTGACTTCCCTCCCAAGTTCCCTCATAAT
55	40262	40225	TTTTCTGATTGACAACCGGGGCCGGAAGGCCCCATAAT
146	83055	83092	AAATTTTTCTTGACCTGCCAGGCAAGCCGGGCCACAAT
148	83952	83989	AAATATTGCTTGACGGGGCCGATGATAGGCCCCATAAT
149	84490	84527	AAAAGACTATTGACATCGGGGCCACCTGGCCCCATACT
151	85201	85238	TTTTAGGGGTTGACGCCCAGCGGCGTTACCCTCATAAT
152	85532	85569	AAATAATTGTTGACCGGGGGCCGGATAGGCCCCATAAT
155	86296	86333	AATTTAGGTTGACACGGGGCCGGGTTAGGCCCCATAAT
158	87840	87877	TTCCGGGGCTTGACGCCCGGCCCCGGTTCCCCCATAAT
160	88865	88902	AAATAATTGTTGACCGGGGCCGGGTTAGGCCCCATAAT
164	90355	90392	AAAAGTAGTTGACGCCGGGGCCGCTTAGGCCCCATAAT
165	90676	90713	AAATAAATCTTGACGGGGCCGACGATAGGCCCCATAAT
168	92144	92181	TGTGAAGGGTTGACACCTGCCGCGATAGGGCCGATAAT

241
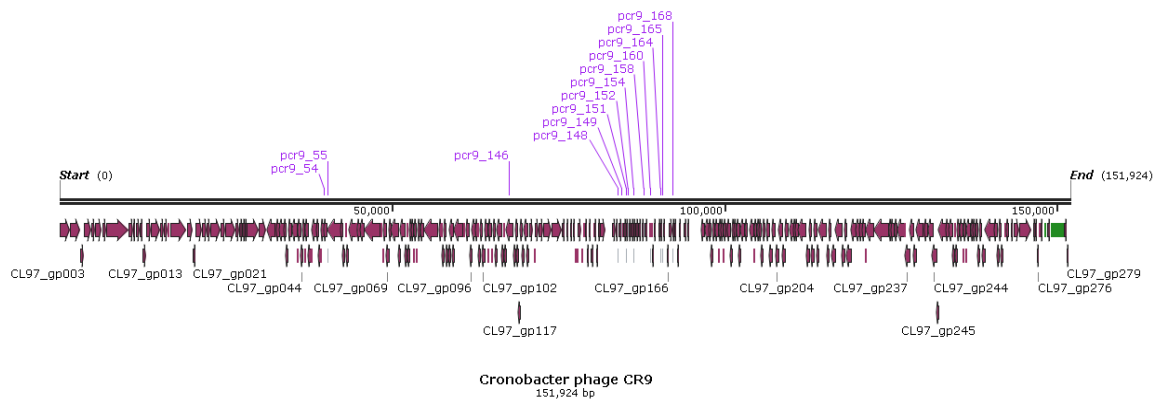
Figure S5.4. Genome map of *Cronobacter* phage CR9 (accession no. JQ691611) showing locations of sigma 70-like. The map comprises of arrows indicating the location of ORFs. Genome map created with Snapgene.

Table S5.6. High ΔG rho-independent terminators predicted in the genome *Pectobacterium* phage CB7 identified using ARNold and QuikFold.

Terminator	Start	Stop	Sequence	ΔG (kcal/mol)
Torf8	7244	7274	GGTCACCTTCGGGTGGCCTTTTTTATTGACC	-15.30
Torf12	12521	12551	AGCCGCCTCCGGGCGGCTTTCTTTTTGTCTG	-13.70
Torf20	17625	17660	GGCTGGAGTCGAAAGACTCTGGCCTTTTCTTTTTTG	-20.80
Torf23	21209	21235	GGGGCTTCGGCCCCcTTTTTTATTGCC	-13.30
Torf28	24794	24825	AGGGAGCCTCTGGCTCCCTTTTCTTTTAATTC	-13.80
Torf37	34032	34003	GGCACCCGACTGGGTGCCTTTTTTATTACT	-14.30
Torf40	35813	35784	GGCCTCAATACTGGGGCCTTGTTTATCCTC	-11.80
Torf48	39005	38978	CCGCCCTTCGGGGCGGTTTTTTATTATC	-15.50
Torf73	51208	51178	AGGTCCCTTCTGGGGCCTTTTTTCGTTTGAC	-12.00
Torf84	58211	58179	GGCTCAGTGATAAACTGGGCCTTTCTTATTTAG	-13.70
Torf108	67570	67603	GGATCATCTTCGGATGGTCCtTTTTGCATTTGGG	-14.60
Torf118	72718	72718	GGGGTTGACTTCGGTCAACCCCTTTTGCTATTCTT	-20.10
Torf126	76806	76834	GGCCCCCGCCGGGGGCCTTTCGTGTGGAG	-15.40
Torf133	79794	79822	GGGGCCGCTTAGGCCCCTTATTTTTGATC	-14.40
Torf141	83315	83346	GCCCGCCACCCGGCGGGCTTTTTGTTGCATC	-16.20
Torf142	83634	83663	GCCCGGCATAGTCCGGGCTTTTTGTTGCCT	-12.00
Torf144	85824	85795	GACCTCCTTCGGGAGGTCTTTTTGTATCTT	-16.10
Torf149	87342	87313	ACCGCCCTCTGGGCGGTTTTCATGTTTAAG	-11.80
Torf163	91391	91365	CCGCCTTCGGGCGGgTTTTTTTTTTAT	-12.90
Torf171	94413	94413	GGTCGATCGTCGAAAGGCGGTCGGCCTTTCTTTTTGGA	-22.20
Torf182	101088	101059	GGCCACCTTCGGGTGGCCTTTTTTATTGCA	-18.00
Torf198	111826	111796	ACCGCCCTTCGGGGCGGTTTTTCTATCTTGA	-16.70
Torf211	120158	120127	GCCGCCCTTCGGGGCGGTtTTCTTTTTGACTT	-17.00
Torf228	129035	129003	AACCGCCCTTCGGGGCGGTTTTTCTTTATCTGC	-17.60
Torf241	135032	135060	GGCCCCTTGACGGGGCCTTTTTCTTTTTG	-13.60
Torf246	138939	138967	CCGCCCTTCGGGGCGGTTTTCTTCTCTCA	-15.50
Torf253	142685	142719	GGCTCTCCATTCGTGGAGGGCCTTTTCTTTTAGGG	-16.40

Table S5.7. Comparison of potential gene products of CB7 identified in DNA replication, methylation and nucleotide metabolism with other members of *Cr3virus* using ACT (TBLASTX)

Phages	CB7	ΦTE	DU_PP_I	DU_PP_IV	CR3	CR8	CR9	PEB02
Accession no.		JQ015307	MF979569	MF979563	JQ691612	KC954774	JQ691611	KT353109
DNA polymerase	CB7_49,51,53	phiTE_6, 8, 10	P1A145kb_p52	P12B145kb_p52	CR3_gp060	HL10_gp061	CL97_gp063, 65	ADU18_174
cytosine methylase	CB7_63	_	_	_	CR3_gp067	HL10_gp068	CL97_gp073	ADU18_183
putative primase/helicase	CB7_73	phiTE_23	P1A145kb_p70	P12B145kb_p71	CR3_gp077	HL10_gp078	CL97_gp084	ADU18_194
putative helicase	CB7_80	phiTE_30	P1A145kb_p79	P12B145kb_p80	CR3_gp084	HL10_gp087	CL97_gp093	ADU18_203
putative nucleotidyl transferase	CB7_175	phiTE_132	P1A145kb_p185	P12B145kb_p186	CR3_gp195	HL10_gp198	CL97_gp211	ADU18_34
tRNA nucleotidyl transferase	CB7_179	phiTE_136	P1A145kb_p191	P12B145kb_p192	CR3_gp201	HL10_gp204	CL97_gp217	ADU18_40
adenine methylase	CB7_183	phiTE_140	P1A145kb_p194	P12B145kb_p195	CR3_gp204	HL10_gp207	CL97_gp0220	ADU18_43
NrdG	CB7_187	phiTE_144	_	_	_	_	_	_
NrdD	CB7_188	phiTE_145	_	_	_	_	_	_
NrdB	CB7_195, 197	phiTE_152	P1A145kb_p206	P12B145kb_p207	CR3_gp215	HL10_gp218	CL97_gp230	ADU18_55
NrdA	CB7_201, 203	phiTE_154	P1A145kb_p209	P12B145kb_p210	CR3_gp216	HL10_gp220	CL97_gp231	ADU18_56
thymidylate synthase	CB7_206	phiTE_159	P1A145kb_p212 (poor homologyl)	P12B145kb_p213 (poor homology)	CR3_gp219 (poor homology)	HL10_gp223 (poor homology protein level)	CL97_gp235 (poor homology)	ADU18_59 (poor homology)
5'(3') deoxyribonuclease	CB7_209	phiTE_162	P1A145kb_p214	P12B145kb_p215	CR3_gp221	HL10_gp225	CL97_gp237	ADU18_61

DNA polymerase/exonuclease	CB7_211	phiTE_164	P1A145kb_p218	P12B145kb_p219	CR3_gp224	HL10_gp228	CL97_gp240	ADU18_64
EndoVII	CB7_214	phiTE_167	P1A145kb_p221	P12B145kb_p222	CR3_gp226	HL10_gp230	CL97_gp242	ADU18_66
endonuclease	CB7_218	phiTE_171	P1A145kb_p225	P12B145kb_p226	CR3_gp230	HL10_gp234	CL97_gp246	ADU18_70
pyrophophatase	CB7_232	phiTE_185	P1A145kb_p242	P12B145kb_p243	CR3_gp245	HL10_gp247	CL97_gp262	ADU18_84
DNA ligase	CB7_234	phiTE_187	P1A145kb_p244	P12B145kb_p245	CR3_gp247	HL10_gp249	CL97_gp264	ADU18_86
ribose phosphate pyrophosphokinase	CB7_242	phiTE_193	P1A145kb_p252	P12B145kb_p253	CR3_gp256	HL10_gp258	CL97_gp272	ADU18_97
nicotinamide phosphoribosyl transferase	CB7_244, 246	phiTE_195, 197	P1A145kb_p254, 256	P12B145kb_p255, 257	CR3_gp257	HL10_gp259	CL97_gp273	ADU18_98

Table S5.8. Comparison of potential gene products of CB7 identified in DNA replication, methylation and nucleotide metabolism with that of *Escherichia* phage rv5 and *Salmonella* phage PVP-SE1 using ACT (TBLASTX)

Phages	CB7	rV5	PVP-SE1
Accession no.	KY514263	DQ832317	GU070616
DNA polymerase	CB7_49,51,53	ORF228	16
cytosine methylase	CB7_63	ORF231	9
putative primase/helicase	CB7_73	ORF230	13
putative helicase	CB7_80	ORF237	3
putative nucleotidyl transferase	CB7_175	_	166
tRNA nucleotidyl transferase	CB7_179	_	159
adenine methylase	CB7_183	ORF123	152
NrdG	CB7_187	ORF117	143
NrdD	CB7_188	ORF112	142
NrdB	CB7_195, 197	ORF110	139
NrdA	CB7_201, 203	ORF109	138
thymidylate synthase	CB7_206	ORF106 (little homology at protein level)	134 (little homology at protein level)
5'(3') deoxyribonuclease	CB7_209	ORF104	131
DNA polymerase/exonuclease	CB7_211	ORF99	128
EndoVII	CB7_214	ORF97	126
endonuclease	CB7_218	ORF94	123
pyrophosphatase	CB7_232	_	_
DNA ligase	CB7_234	_	_
ribose phosphate pyrophosphokinase	CB7_242	ORF80	97
nicotinamide phosphoribosyl transferase	CB7_244, 246	_	96

Table S5.9. Shared structural proteins of phages of *Cr3virus* determined using ACT (TBLASTX).

Phages	CB7	ΦTE	DU_PP_I	DU_PP_IV	CR3	CR8	CR9	PBES 02
Accession no.	KY514263	JQ015307	MF979560	MF979563	JQ691612	KC954774	JQ691611	KT353109
putative portal protein	CB7_4	phiTE_209	P1A145kb_p2	P12B145kb_p2	CR3_gp2	HL10_gp2	CL97_gp2	ADU18_109
prohead core protein protease	CB7_5	phiTE_210	P1A145kb_p3	P12B145kb_p3	CR3_gp3	HL10_gp3	CL97_gp3	ADU18_110
unknown structural protein	CB7_6	phiTE_211	P1A145kb_p4	P12B145kb_p4	CR3_gp4	HL10_gp4	CL97_gp4	ADU18_111
head stabilization/decoration protein	CB7_7	phiTE_212	P1A145kb_p5	P12B145kb_p5	CR3_gp5	HL10_gp5	CL97_gp5	ADU18_112
putative major head protein	CB7_8	phiTE_213	P1A145kb_p6	P12B145kb_p6	CR3_gp6	HL10_gp6	CL97_gp6	ADU18_113
putative tail fibre protein	CB7_10	phiTE_215	P1A145kb_p8	P12B145kb_p8	CR3_gp8	HL10_gp8	CL97_gp8	ADU18_115
unknown structural protein	CB7_11	phiTE_210	P1A145kb_p9	P12B145kb_p9	CR3_gp9	HL10_gp9	CL97_gp9	ADU18_116
unknown structural protein	CB7_12	phiTE_217	P1A145kb_p10	P12B145kb_p10	CR3_gp10	HL10_gp10	CL97_gp10	ADU18_117
unknown structural protein	CB7_13	phiTE_218	P1A145kb_p11	P12B145kb_p11	CR3_gp12	HL10_gp12	CL97_gp11	ADU18_119
unknown structural protein	CB7_15	phiTE_220	P1A145kb_p13	P12B145kb_p13	CR3_gp14	HL10_gp14	CL97_gp13	ADU18_121
unknown structural protein	CB7_16	phiTE_211	P1A145kb_p14	P12B145kb_p14	CR3_gp15	HL10_gp15	CL97_gp14	ADU18_122
putative tail sheath protein	CB7_17	phiTE_222	P1A145kb_p15	P12B145kb_p15	CR3_gp16	HL10_gp16	CL97_gp15	ADU18_123
putative tail tube protein	CB7_18	phiTE_223	P1A145kb_p16	P12B145kb_p16	CR3_gp17	HL10_gp17	CL97_gp16	ADU18_124
putative tape measure protein	CB7_23	phiTE_228	P1A145kb_p21	P12B145kb_p21	CR3_gp20	HL10_gp20	CL97_gp19	ADU18_127
unknown structural protein	CB7_24	phiTE_229	P1A145kb_p22	P12B145kb_p22	CR3_gp21	HL10_gp21	CL97_gp20	ADU18_128
unknown structural protein	CB7_25	phiTE_230	P1A145kb_p23	P12B145kb_p23	CR3_gp22	HL10_gp22	CL97_gp21	ADU18_129
putative tail protein	CB7_26	phiTE_231	P1A145kb_p24	P12B145kb_p24	CR3_gp23	HL10_gp23	CL97_gp22	ADU18_130
putative baseplate protein	CB7_27	phiTE_232	P1A145kb_p25	P12B145kb_p25	CR3_gp24	HL10_gp24	CL97_gp23	ADU18_131
putative tail lysozyme	CB7_28	phiTE_233	P1A145kb_p26	P12B145kb_p26	CR3_gp25	HL10_gp25	CL97_gp24	ADU18_132
putative baseplate wedge protein	CB7_29	phiTE_234	P1A145kb_p27	P12B145kb_p27	CR3_gp26	HL10_gp26	CL97_gp25	ADU18_133
putative baseplate protein	CB7_30	phiTE_235	P1A145kb_p28	P12B145kb_p28	CR3_gp27	HL10_gp27	CL97_gp26	ADU18_134
putative tail collar protein	CB7_31	phiTE_236	P1A145kb_p29	P12B145kb_p29	CR3_gp28	HL10_gp28	CL97_gp27	ADU18_135
putative tail fibre protein	CB7_32	phiTE_237	P1A145kb_p30	P12B145kb_p30	CR3_gp29	HL10_gp29	CL97_gp28	ADU18_136

tail chaparone protein	CB7_36	phiTE_241	P1A145kb_p34	P12B145kb_p34	CR3_gp33	HL10_gp33	CL97_gp32	ADU18_140
unknown structural protein	CB7_37	phiTE_237	P1A145kb_p35	P12B145kb_p35	CR3_gp34	HL10_gp34	CL97_gp33	ADU18_141
unknown structural protein	CB7_55	-	-	-	-	-	-	-
unknown structural protein	CB7_251	phiTE_203	-	-	-	-	-	-

Table S5.10. Shared structural proteins (using currently available annotation) of phages of *Vequintavirinae* (*Pectobacterium* phage CB7, *Escherichia* phage rv5 and *Salmonella* phage PVP-SE1) determined using ACT (TBLASTX). Proteins in bold and italics are those with have been determined to be presence in mature virion of phage in question by ESI-MS/MS analysis as determined by Santos et al 2011 and Kropinski et al 2013

Pectobacterium phage CB7 Accession no. KY514263	*Escherichia* phage rV5 Accession no. DQ832317	*Salmonella* phage PVP-SE1 Accession no. NC_016071
CB7_4, putative portal protein	*gp64, portal protein*	*PVP-SE1_gp75, conserved hypothetical protein*
CB7_5, putative prohead core protein protease	gp63, conserved hypothetical protein	PVP-SE1_gp74, hypothetical protein
CB7_6, unknown structural protein	gp62, conserved hypothetical protein	*PVP-SE1_gp73, hypothetical protein*
CB7_7, putative head stabilization/decoration protein,	*gp61, putative head stabilization/decoration protein*	*PVP-SE1_gp72, head stablization/ decorative protein*
CB7_8, putative major head protein	*gp60, major capsid protein*	*PVP-SE1_gp71, putative major head protein*
CB7_10, putative tail fibre protein	_	*PVP-SE1_gp69, hypothetical protein*
CB7_11, unknown structural protein	_	PVP-SE1_gp68, hypothetical protein
_	_	*PVP-SE1_gp67, hypothetical protein*
CB7_12, unknown structural protein	_	_
CB7_13, unknown structural protein	gp57, hypothetical protein	*PVP-SE1_gp65, conserved hypothetical protein*
CB7_15, unknown structural protein	gp55, conserved hypothetical protein	*PVP-SE1_gp63, conserved hypothetical protein*
CB7_16, unknown structural protein	gp54, hypothetical protein	*PVP-SE1_gp62, hypothetical protein*
CB7_17, putative tail sheath protein	*gp53, tail sheath protein*	*PVP-SE1_gp61, structural protein*

CB7_18, putative tail tube protein	gp52, tail tube protein	PVP-SE1_gp60, structural protein
CB7_23, putative tape measure protein	gp49, putative tail protein	PVP-SE1_gp57, hypothetical protein
CB7_24, unknown structural protein	gp48, hypothetical protein	PVP-SE1_gp56, conserved hypothetical protein
CB7_25, unknown structural protein	gp47, hypothetical protein	PVP-SE1_gp55, conserved hypothetical protein
CB7_26, putative tail protein	gp46, conserved hypothetical protein	PVP-SE1_gp54, conserved hypothetical protein
CB7_27, putative baseplate protein	gp45, tail baseplate protein	PVP-SE1_gp53, putative baseplate assembly protein
CB7_28, putative tail lysozyme	gp44, conserved hypothetical protein	PVP-SE1_gp52, conserved hypothetical protein
_	gp43, tail fibre protein	PVP-SE1_gp51, possible tail fibre protein
_	gp42, putative tail protein	PVP-SE1_gp50, conserved hypothetical protein
_	_	PVP-SE1_gp49, conserved hypothetical protein
_	gp41, tail fibre protein	_
_	gp37, putative tail protein	_
CB7_29, putative baseplate wedge protein	gp36, tail baseplate protein	PVP-SE1_gp48, baseplate component
CB7_30, putative baseplate protein	gp35, hypothetical protein	PVP-SE1_gp47, conserved hypothetical protein
CB7_31, putative tail collar protein (N-terminus only)	gp33, putative tail fibre protein	PVP-SE1_gp46, putative tail fibre protein
_	gp32, tail fibre protein	PVP-SE1_gp45, putative tail fibre assembly protein
_	gp30, tail fibre protein	PVP-SE1_gp46, hypothetical protein
putative tail fibre protein,	gp28, putative tail fibre	PVP-SE1_gp41, putative tail

CB7_36	protein (weakening homology towards C terminus)	**fibre protein**
unknown structural protein, CB7_37	_	**PVP-SE1_gp40, hypothetical with Ig-like domain (weak homology)**
_	gp27, hypothetical protein	
unknown structural protein, CB7_55	_	
unknown structural protein, CB7_251	_	
_	*gp133, unknown structural protein*	
_	_	*PVP-SE1_gp19, hypothetical protein*
_	_	*PVP-SE1_gp10, conserved hypothetical protein*

Table S5.11. Putative cell wall degrading and cell lysis proteins of phages in the subfamily *Vequintavirinae.* Determined using ACT (TBLASTX).

Phage	rV5	PVP-SE1	CB7	ΦTE	DU_PP_I	DU_PP_IV	CR3	CR8	CR9	PBES 02
Accession no.	DQ832317	NC_016071	KY514263	NC_020201	MF979560	MF979563	NC_017974	NC_024354	NC_023717	KT353109
Rz	gp67	gp78	CB7_252	phiTE_204	P1A145kb_p266	P12B145kb_p267	CR3_264	CR8_268	CL97_gp280	ADU18_0106
Rz1	Gp66	Gp79	CB7_253	phiTE_205	P1A145kb_p267	P12B145kb_p268	CR3_265	CR8_269	CL97_gp281	ADU18_0107
SleB	Gp2	Gp243	CB7_83	phiTE_33	P1A145kb_p083	P12B145kb_p084	CR3_87	CR8_90	CL97_gp95	ADU18_0206
putative class II holin	_	_	CB7_82	phiTE_032	P1A145kb_p081	P12B145kb_p082	CR3_086	CR8_089	CL97_gp094	ADU18_0205
T4 gp25 - like	Gp44	Gp52	CB7_28	phiTE_233	P1A145kb_p026	P12B145kb_p026	CR3_25	CL97_gp_024	CL97_gp024	ADU18_0132
CB7_ 190 - like			CB7_190	phiTE_147	_	_	_	_	_	_
CR3_ 210 - like			_	_	P1A145kb_p201	P12B145kb_p202	CR3_210	CL97_gp226	CL97_gp226	ADU18_0050

251

Table S6.1. Proteins in the genome of *Pectobacterium* phage CBB that share homology (Detected using BLASTP).

Proteins that share homology
CBB_47 (CBB_600) to CBB_50 (CBB_603)
CBB_147 to CBB_148 to CBB_149
CBB_160 to CBB_161
CBB_438 to CBB_439
CBB_480 to CBB_481 to CBB_482
CBB_536 to CBB_537
CBB_539 to CBB_540 to CBB_541 to CBB_543
CBB_549 to CBB_553
CBB_356 to CBB_10 (CBB_563)

Table S6.2. Genome annotation of *Pectobacterium* phage CBB

ORF	Start	Stop	Length (aa)	Molecular weigth (KDa)	Function
CBB_1	20	197	55	6.65	hypothetical protein
CBB_2	1427	1615	62	7.00	hypothetical protein
CBB_3	1800	2477	225	25.17	conserved hypothetical protein
CBB_4	2627	2896	89	10.13	conserved hypothetical protein
CBB_5	3032	3316	94	10.83	conserved hypothetical protein
CBB_6	3352	3741	129	14.96	conserved hypothetical protein
CBB_7	3906	3793	57	6.30	hypothetical protein
CBB_8	3965	4117	50	5.58	hypothetical protein
CBB_9	4165	4437	90	10.08	conserved hypothetical protein
CBB_10	4501	4947	148	16.80	conserved hypothetical protein
CBB_11	5024	5479	151	17.13	conserved hypothetical protein
CBB_12	5560	5685	41	4.54	hypothetical protein
CBB_13	5758	5985	75	8.43	hypothetical protein
CBB_14	6059	6481	140	16.39	conserved hypothetical protein
CBB_15	6663	6998	111	12.23	conserved hypothetical protein
CBB_16	7068	7235	55	6.68	conserved hypothetical protein
CBB_17	7319	7498	59	6.51	conserved hypothetical protein
CBB_18	7989	8243	123	14.45	conserved hypothetical protein
CBB_19	7989	8243	84	9.99	conserved hypothetical protein
CBB_20	8248	8406	52	5.88	hypothetical protein
CBB_21	8696	8929	77	8.56	conserved hypothetical protein
CBB_22	8953	9111	52	6.06	putative membrane protein
CBB_23	9123	9326	67	7.51	conserved hypothetical protein
CBB_24	9326	9502	58	6.86	conserved hypothetical protein
CBB_25	9581	9784	67	7.56	conserved hypothetical protein
CBB_26	9814	10164	116	13.40	conserved hypothetical protein
CBB_27	10391	10642	83	9.44	putative membrane protein

CBB_28	10674	10832	52	5.87	hypothetical protein
CBB_29	10841	11716	291	31.54	structural protein
CBB_30	11856	12053	65	8.04	conserved hypothetical protein
CBB_31	12017	12142	41	4.86	conserved hypothetical protein
CBB_32	12132	12326	64	8.12	hypothetical protein
CBB_33	12370	12723	117	13.52	putative membrane protein
CBB_34	12720	13109	129	14.27	hypothetical protein
CBB_35	13181	13372	63	7.27	conserved hypothetical protein
CBB_36	13435	13821	128	14.27	conserved hypothetical protein
CBB_37	13941	14183	80	9.53	conserved hypothetical protein
CBB_38	14240	14758	172	18.82	structural protein
CBB_39	14822	15160	112	12.64	conserved hypothetical protein
CBB_40	15403	15750	115	12.92	hypothetical protein
CBB_41	15841	16059	72	8.27	conserved hypothetical protein
CBB_42	16235	16837	200	23.42	conserved hypothetical protein
CBB_43	17042	17374	110	12.84	conserved hypothetical protein
CBB_44	17584	17826	80	9.16	conserved hypothetical protein
CBB_45	17906	18793	295	33.78	conserved hypothetical protein
CBB_46	18869	19207	112	13.26	conserved hypothetical protein
CBB_47	19411	19641	76	8.64	conserved hypothetical protein
CBB_48	19725	20072	115	13.49	conserved hypothetical protein
CBB_49	20272	20463	63	7.18	hypothetical protein
CBB_50	20533	20709	58	6.43	conserved hypothetical protein
CBB_51	21113	21382	89	10.35	conserved hypothetical protein
CBB_52	21452	22009	185	20.69	hypothetical protein
CBB_53	22520	23107	195	23.35	conserved hypothetical protein
CBB_54	23110	23721	203	23.78	conserved hypothetical protein
CBB_55	23696	24328	210	25.13	conserved hypothetical protein
CBB_56	24459	25058	199	23.04	conserved hypothetical protein
CBB_57	25055	25576	173	20.85	conserved hypothetical protein

CBB_58	25537	26139	200	23.71	conserved hypothetical protein
CBB_59	26146	26559	137	15.94	conserved hypothetical protein
CBB_60	26573	27016	147	17.35	conserved hypothetical protein
CBB_61	27096	27572	158	18.37	conserved hypothetical protein
CBB_62	27576	27746	56	6.48	Conserved hypothetical protein
CBB_63	27751	28359	202	23.72	conserved hypothetical protein
CBB_64	28360	29247	295	35.49	conserved hypothetical protein
CBB_65	29231	29551	105	12.91	conserved hypothetical protein
CBB_66	29595	29993	132	14.79	conserved hypothetical protein
CBB_67	30018	30368	116	13.41	conserved hypothetical protein
CBB_68	30409	30729	106	12.54	putative membrane protein
CBB_69	30733	31506	257	29.85	conserved hypothetical protein
CBB_70	31506	32300	264	30.72	conserved hypothetical protein
CBB_71	32990	33514	174	21.26	conserved hypothetical protein
CBB_72	33527	34069	180	20.59	DNA N-6-adenine-methyltransferase
CBB_73	34091	34630	179	20.43	cytitidyltransferase
CBB_74	34614	34895	93	10.90	conserved hypothetical protein
CBB_75	34942	35490	182	21.01	conserved hypothetical protein
CBB_76	35528	35686	52	6.20	conserved hypothetical protein
CBB_77	35673	36374	233	27.34	conserved hypothetical protein
CBB_78	36371	36574	67	7.65	putative membrane protein
CBB_79	36555	37667	370	42.67	conserved hypothetical protein
CBB_80	37709	38440	243	28.09	conserved hypothetical protein
CBB_81	38485	38784	99	11.32	conserved hypothetical protein
CBB_82	38781	39077	98	11.45	conserved hypothetical protein
CBB_83	39142	40290	382	44.10	RNA ligase 1 and tail fiber attachment catalyst
CBB_84	40325	40561	78	9.27	conserved hypothetical protein
CBB_85	40554	40774	74	8.71	conserved hypothetical protein
CBB_86	40791	41015	74	8.52	conserved hypothetical protein

CBB_87	41003	41935	310	35.99	polynucleotide 5'-kinase and 3'-phosphatase
CBB_88	41991	42200	69	7.71	conserved hypothetical protein
CBB_89	42213	42380	55	6.47	putative membrane protein
CBB_90	42370	42642	90	10.90	conserved hypothetical protein
CBB_91	42639	42977	112	12.94	conserved hypothetical protein
CBB_92	42979	43206	75	8.47	putative membrane protein
CBB_93	43187	43435	82	9.58	conserved hypothetical protein
CBB_94	43432	43788	118	14.06	conserved hypothetical protein
CBB_95	43788	44033	81	9.78	conserved hypothetical protein
CBB_96	44077	44715	212	24.93	conserved hypothetical protein
CBB_97	44715	45494	259	30.02	putative serine/threonine protein phosphatase
CBB_98	45494	45952	152	17.25	conserved hypothetical protein
CBB_99	45963	47201	412	47.48	conserved hypothetical protein
CBB_100	47228	48643	471	53.23	DNA ligase
CBB_101	48640	49065	141	16.67	conserved hypothetical protein
CBB_102	49043	49834	263	30.56	metallophosphoesterase
CBB_103	49836	50264	142	14.64	conserved hypothetical protein
CBB_104	50278	50970	230	25.46	ATP-dependent Clp protease proteolytic subunit
CBB_105	50987	51274	95	10.46	conserved hypothetical protein
CBB_106	51280	51735	151	17.75	conserved hypothetical protein
CBB_107	51781	52335	184	22.46	conserved hypothetical protein
CBB_108	52320	52847	175	20.60	conserved hypothetical protein
CBB_109	52837	53319	160	19.09	conserved hypothetical protein
CBB_110	53306	53761	151	16.85	CMP/dCMP deaminase
CBB_111	53763	54032	89	10.40	putative membrane protein
CBB_112	54740	54084	218	21.75	major tail protein
CBB_113	54982	56079	365	42.17	bifunctional nicotinamide mononucleotide adenylyltransferase/ADP-ribose

					pyrophosphatase
CBB_114	56123	56641	172	20.20	conserved hypothetical protein
CBB_115	56702	58189	495	55.68	nicotinamide phosphoribosyltransferase
CBB_116	58232	58453	73	8.39	putative membrane protein
CBB_117	58455	59402	315	37.08	conserved hypothetical protein
CBB_118	59615	60061	148	16.64	structural protein
CBB_119	60073	60315	80	9.60	conserved hypothetical protein
CBB_120	60328	61119	263	29.72	putative Sir2-like protein
CBB_121	61135	62034	299	34.45	conserved hypothetical protein
CBB_122	62034	62336	100	11.13	putative membrane protein
CBB_123	62407	63846	479	53.57	PhoH family protein
CBB_124	63894	64076	60	7.03	conserved hypothetical protein
CBB_125	64251	64943	230	27.02	conserved hypothetical protein
CBB_126	64940	65146	68	7.26	putative membrane protein
CBB_127	65153	65302	49	5.62	putative membrane protein
CBB_128	65377	65922	181	22.37	conserved hypothetical protein
CBB_129	65949	66791	200	31.87	conserved hypothetical protein
CBB_130	66793	67407	204	22.68	conserved hypothetical protein
CBB_131	67407	67688	93	10.26	co-chaperonin GroES
CBB_132	67698	6893	421	47.60	tyrosyl-tRNA synthetase
CBB_133	68985	69171	68	7.80	putative lipoprotein
CBB_134	69317	69958	213	25.45	conserved hypothetical protein
CBB_135	69948	70313	121	13.72	dCMP deaminase
CBB_136	70300	70620	106	12.88	hypothetical protein
CBB_137	70613	71239	208	24.62	hypothetical protein
CBB_138	71229	71486	85	10.21	conserved hypothetical protein
CBB_139	71483	71683	66	7.74	hypothetical protein
CBB_140	71696	72247	183	21.16	conserved hypothetical protein
CBB_141	72380	74217	611	69.42	anaerobic NTP reductase large subunit

CBB_142	74304	74621	105	12.19	hypothetical protein
CBB_143	74628	75074	148	18.00	conserved hypothetical protein
CBB_144	75120	75884	254	25.22	hypothetical protein
CBB_145	75881	76354	157	17.91	anaerobic ribonucleoside-triphosphate reductase activating protein
CBB_146	76347	76523	58	6.57	conserved hypothetical protein
CBB_147	76602	79310	902	98.07	conserved hypothetical protein
CBB_148	79320	82070	916	98.70	structural protein
CBB_149	82083	84833	916	100.67	structural protein
CBB_150	84966	85526	186	22.00	conserved hypothetical protein
CBB_151	85483	85752	89	10.61	hypothetical protein
CBB_152	85749	86528	259	29.85	putative tRNA-His guanylyltransferase
CBB_153	87000	87302	154	19.59	conserved hypothetical protein
CBB_154	87000	87302	100	11.23	conserved hypothetical protein
CBB_155	87310	87789	159	18.38	conserved hypothetical protein
CBB_156	87786	88418	210	25.22	conserved hypothetical protein
CBB_157	88427	89287	286	33.21	putative N-4 cytosine-specific methyltransferase
CBB_158	89284	89619	111	12.97	conserved hypothetical protein
CBB_159	89628	90878	416	48.92	nucleotidyltransferase
CBB_160	90878	91294	138	16.13	conserved hypothetical protein
CBB_161	91298	91801	167	19.46	conserved hypothetical protein
CBB_162	91791	92360	189	21.48	conserved hypothetical protein
CBB_163	92332	92748	138	15.80	acyl carrier protein
CBB_164	92821	93189	122	14.09	structural protein
CBB_165	93186	93500	104	11.90	structural protein
CBB_166	93574	93963	129	15.24	conserved hypothetical protein
CBB_167	93974	94228	84	10.36	conserved hypothetical protein
CBB_168	94890	95255	121	14.61	conserved hypothetical protein
CBB_169	95287	95727	146	17.18	structural protein
CBB_170	95771	96202	143	16.86	conserved hypothetical protein

CBB_171	96192	96695	167	19.77	GTP cyclohydrolase
CBB_172	96676	96963	95	11.16	conserved hypothetical protein
CBB_173	96976	97545	189	21.41	thymidine kinase
CBB_174	97599	98108	169	19.78	putative membrane protein
CBB_175	98105	98713	202	23.04	putative membrane protein
CBB_176	98710	99069	119	13.58	conserved hypothetical protein
CBB_177	99071	99205	44	4.71	putative membrane protein
CBB_178	99207	99806	199	23.57	hypothetical protein
CBB_179	99808	100128	106	12.32	hypothetical protein
CBB_180	100131	100678	182	21.27	conserved hypothetical protein
CBB_181	100759	101995	418	47.45	RNA ligase
CBB_182	101998	102606	202	24.17	conserved hypothetical protein
CBB_183	102737	104251	504	55.82	structural protein
CBB_184	104304	104444	46	5.03	putative membrane protein
CBB_185	104498	104920	140	17.28	conserved hypothetical protein
CBB_186	104913	105488	191	21.94	conserved hypothetical protein
CBB_187	105603	106538	311	35.38	lysozyme
CBB_188	106561	106899	112	12.88	sigma 54 modulation protein/ribosomal protein S30EA
CBB_189	106983	107384	133	16.15	conserved hypothetical protein
CBB_190	107371	107514	47	5.63	hypothetical protein
CBB_191	107501	107716	71	8.66	conserved hypothetical protein
CBB_192	107713	108828	371	42.75	multifunctional tRNA nucleotidyl transferase/2'3'-cyclic phosphodiesterase/2'nucleotidase/phosphatase
CBB_193	108894	109316	140	16.50	nudix hydrolase
CBB_194	109342	109578	140	8.70	glutaredoxin
CBB_195	109580	109993	137	15.18	conserved hypothetical protein
CBB_196	109996	110340	114	13.50	conserved hypothetical protein
CBB_197	110388	110630	80	9.39	conserved hypothetical protein
CBB_198	110675	111175	166	18.85	conserved hypothetical protein

CBB_199	111202	111531	109	11.97	conserved hypothetical protein
CBB_200	111541	111903	120	13.48	conserved hypothetical protein
CBB_201	111951	112736	261	29.62	neck protein
CBB_202	112774	113028	84	9.77	conserved hypothetical protein
CBB_203	113067	113708	213	24.82	putative deoxynucleoside monophosphate kinase
CBB_204	113797	116451	884	95.91	tail sheath monomer
CBB_205	116558	117247	229	25.39	structural protein
CBB_206	117331	118107	258	29.03	structural protein
CBB_207	118118	119212	364	39.83	structural protein
CBB_208	119212	119670	152	17.69	head completion protein
CBB_209	119672	120235	187	21.53	conserved hypothetical protein
CBB_210	120263	120808	181	20.95	structural protein
CBB_211	120811	121371	186	20.24	structural protein
CBB_212	121384	122820	478	53.17	structural protein
CBB_213	122838	123767	309	33.36	structural protein
CBB_214	123767	124177	136	14.86	structural protein
CBB_215	124181	124555	124	14.33	conserved hypothetical protein
CBB_216	124693	126036	447	50.67	ATPase
CBB_217	126078	127430	450	51.29	putative membrane protein
CBB_218	127539	127775	218	8.64	conserved hypothetical protein
CBB_219	127777	128469	230	26.15	conserved hypothetical protein
CBB_220	128469	128777	102	11.58	conserved hypothetical protein
CBB_221	128764	129351	195	22.36	conserved hypothetical protein
CBB_222	129358	129687	109	12.86	conserved hypothetical protein
CBB_223	129650	130579	309	35.33	thymidylate synthase
CBB_224	132533	130611	640	68.10	structural protein
CBB_225	133259	132621	212	24.19	structural protein
CBB_226	138084	137934	1553	153.27	long tail fiber proximal subunit
CBB_227	138944	138084	286	30.02	structural protein

CBB_228	140166	139024	380	41.68	structural protein
CBB_229	140272	140769	165	19.05	structural protein
CBB_230	140780	141388	202	21.80	structural protein
CBB_231	141399	142505	368	40.39	structural protein
CBB_232	142976	142509	148	17.72	MutT/NUDIX hydrolase family protein
CBB_233	144323	142986	445	51.35	tail sheath stabilizer and completion protein
CBB_234	145816	144320	498	57.00	structural protein
CBB_235	144320	145816	765	84.34	ATP-dependent Clp protease ATP-binding subunit clpA
CBB_236	158274	148252	3340	379.94	structural protein
CBB_237	161820	158356	1154	131.71	baseplate wedge
CBB_238	162285	161896	129	14.83	base plate protein
CBB_239	162784	162296	162	18.49	baseplate hub subunit and tail lysozyme
CBB_240	165504	162820	894	97.09	baseplate hub subunit and tail lysozyme
CBB_241	167891	165504	795	88.83	structural protein
CBB_242	168217	167903	104	12.23	baseplate wedge protein
CBB_243	169931	169071	245	27.69	structural protein
CBB_244	169071	169931	286	33.56	sigma factor for late transcription
CBB_245	169918	170949	343	39.92	recombination endonuclease subunit
CBB_246	170962	173103	713	82.05	recombination endonuclease subunit
CBB_247	173132	173866	244	27.28	structural protein
CBB_248	174344	173868	158	18.15	DNA endonuclease VII
CBB_249	175307	175131	58	6.93	conserved hypothetical protein
CBB_250	176185	175328	285	32.55	structural protein
CBB_251	176250	178418	722	78.41	structural protein
CBB_252	178435	180171	578	65.18	portal vertex protein of head
CBB_253	180229	180459	76	8.63	conserved hypothetical protein
CBB_254	181414	182043	274	30.25	structural protein
CBB_255	181414	182043	209	23.24	prohead core scaffold and protease

CBB_256	182112	183236	274	42.55	structural protein
CBB_257	183315	184490	391	42.11	precursor of major head subunit
CBB_258	184644	185060	138	15.98	conserved hypothetical protein
CBB_259	185089	185400	103	12.02	conserved hypothetical protein
CBB_260	185445	187064	539	56.65	putative tail fiber protein
CBB_261	187057	187470	137	16.06	tail fiber assembly protein
CBB_262	187916	190948	100	11.31	conserved hypothetical protein
CBB_263	187916	190948	1010	115.94	DNA polymerase
CBB_264	190964	191476	170	19.32	conserved hypothetical protein
CBB_265	191541	192272	243	26.74	conserved hypothetical protein
CBB_266	192319	192933	204	23.92	putative phosphoesterase or phosphohydrolase
CBB_267	192945	193154	69	8.22	conserved hypothetical protein
CBB_268	193160	193459	90	10.74	co-chaperonin GroES
CBB_269	193452	194171	239	28.50	structural protein
CBB_270	194219	194875	218	22.43	structural protein
CBB_271	194891	195451	186	19.23	structural protein
CBB_272	195646	196632	328	38.17	RNaseH ribonuclease
CBB_273	196632	197132	166	19.04	conserved hypothetical protein
CBB_274	197176	197883	235	27.52	terminase DNA packaging enzyme large subunit
CBB_275	197870	199519	549	62.89	DNA terminase packaging enzyme large subunit
CBB_276	199556	200845	429	48.67	structural protein
CBB_277	201073	202116	347	38.62	ssDNA binding protein
CBB_278	202175	203305	376	41.44	recA-like protein
CBB_279	203307	203771	154	17.96	conserved hypothetical protein
CBB_280	203817	204662	281	32.53	putative DNA polymerase III epsilon subunit
CBB_281	204685	206169	494	56.09	NA-DNA + DNA-DNA helicase
CBB_282	204685	206169	145	17.16	conserved hypothetical protein
CBB_283	206670	207173	167	19.15	conserved hypothetical protein

CBB_284	207173	208018	281	32.25	conserved hypothetical protein
CBB_285	208005	208724	239	28.82	conserved hypothetical protein
CBB_286	208721	209068	115	13.30	structural protein
CBB_287	209034	209684	216	25.11	putative membrane protein
CBB_288	209668	210021	117	14.23	conserved hypothetical protein
CBB_289	210079	211050	323	36.68	structural protein
CBB_290	211113	212159	348	40.38	putative DNA primase
CBB_291	212167	213663	498	55.75	DNA primase-helicase
CBB_292	213666	213935	89	10.37	structural protein
CBB_293	213937	214095	52	5.60	conserved hypothetical protein
CBB_294	214097	214495	132	14.70	structural protein
CBB_295	214541	215485	314	34.25	conserved hypothetical protein
CBB_296	215485	215703	72	8.36	conserved hypothetical protein
CBB_297	215700	216503	267	30.22	conserved hypothetical protein
CBB_298	216536	217180	214	23.65	structural protein
CBB_299	217193	217528	111	12.15	structural protein
CBB_300	217568	218269	233	26.64	restriction endonuclease type II-like
CBB_301	218274	218825	183	20.26	conserved hypothetical protein
CBB_302	218854	219432	192	20.83	structural protein
CBB_303	219508	221838	776	87.79	ribonucleotide-diphosphate reductase alpha subunit
CBB_304	221881	222054	257	30.80	conserved hypothetical protein
CBB_305	222664	223785	373	43.37	ribonucleotide-diphosphate reductase subunit beta
CBB_306	223824	224012	62	7.09	putative membrane protein
CBB_307	224073	224306	77	9.14	conserved hypothetical protein
CBB_308	224318	224698	126	14.78	structural protein
CBB_309	224698	225414	238	26.30	conserved hypothetical protein
CBB_310	225436	226089	217	24.21	structural protein
CBB_311	226101	226811	236	26.33	putative membrane protein
CBB_312	226944	227318	124	13.97	putative membrane protein

CBB_313	227315	227833	172	19.68	dihydrofolate reductase
CBB_314	227833	228267	144	16.20	ribonuclease H
CBB_315	228264	229604	446	51.65	DNA helicase Dda
CBB_316	229672	230217	181	19.94	structural protein
CBB_317	230236	230430	64	7.30	conserved hypothetical protein
CBB_318	230504	231010	168	18.74	translation initiation factor IF-3
CBB_319	231071	231604	177	20.09	ATP-dependent Clp protease
CBB_320	231666	232199	177	20.64	structural protein
CBB_321	231189	233799	536	61.65	structural protein
CBB_322	233799	236342	847	96.34	structural protein
CBB_323	236433	236717	94	10.58	putative membrane protein
CBB_324	236813	238786	657	74.42	DNA gyrase subunit B
CBB_325	238837	240243	468	53.17	DNA topoisomerase II
CBB_326	240331	243014	789	87.98	structural protein
CBB_327	242709	243014	101	11.84	conserved hypothetical protein
CBB_328	242974	243492	172	20.57	DNA polymerase III alpha subunit
CBB_329	243539	243913	124	13.94	conserved hypothetical protein
CBB_330	243924	244244	106	11.70	conserved hypothetical protein
CBB_331	244288	244578	96	11.15	conserved hypothetical protein
CBB_332	244581	245534	317	36.15	sliding clamp loader subunit
CBB_333	245543	246163	206	24.06	conserved hypothetical protein
CBB_334	246221	246790	189	21.94	structural protein
CBB_335	246790	247473	227	25.70	structural protein
CBB_336	247473	248267	264	30.76	conserved hypothetical protein
CBB_337	248404	248943	179	19.20	conserved hypothetical protein
CBB_338	248977	249177	66	7.02	conserved hypothetical protein
CBB_339	249236	250432	398	45.96	conserved hypothetical protein
CBB_340	250525	251667	380	42.63	toxic ion resistance protein
CBB_341	251734	252066	110	13.01	conserved hypothetical protein
CBB_342	252066	252740	224	25.57	putative metallopeptidase

CBB_343	252797	253894	365	42.11	putative membrane protein
CBB_344	253891	254445	184	20.71	structural protein
CBB_345	254558	254737	59	6.44	conserved hypothetical protein
CBB_346	254749	255096	115	13.26	putative membrane protein
CBB_347	255145	255777	210	25.48	conserved hypothetical protein
CBB_348	255777	256262	161	18.99	conserved hypothetical protein
CBB_349	256322	256046	574	64.97	conserved hypothetical protein
CBB_350	258785	259024	79	9.26	hypothetical protein
CBB_351	259027	259467	146	17.29	conserved hypothetical protein
CBB_352	259469	259672	67	7.74	conserved hypothetical protein
CBB_353	259672	260019	115	13.81	conserved hypothetical protein
CBB_354	260021	260335	104	12.35	conserved hypothetical protein
CBB_355	260322	260672	116	13.29	hypothetical protein
CBB_356	260659	261378	239	27.60	conserved hypothetical protein
CBB_357	361368	261616	82	9.91	conserved hypothetical protein
CBB_358	361621	261815	64	7.65	conserved hypothetical protein
CBB_359	261832	262851	339	39.69	conserved hypothetical protein
CBB_360	262861	263232	123	14.27	conserved hypothetical protein
CBB_361	263250	264299	349	40.33	nicotinamide-nucleotide adenylyltransferase
CBB_362	264296	265099	267	30.12	nicotinamide mononucleotide transporter PnuC
CBB_363	365101	265212	103	11.83	hypothetical protein
CBB_364	265432	265674	80	9.22	conserved hypothetical protein
CBB_365	265721	265924	67	8.38	conserved hypothetical protein
CBB_366	265921	266358	145	17.42	conserved hypothetical protein
CBB_367	266355	266660	101	11.61	hypothetical protein
CBB_368	266657	266968	103	11.70	conserved hypothetical protein
CBB_369	266977	267321	114	13.01	conserved hypothetical protein
CBB_370	267305	267658	117	13.82	structural protein
CBB_371	267658	268017	119	13.97	conserved hypothetical protein

CBB_372	268028	268333	101	11.83	conserved hypothetical protein
CBB_373	268330	268779	149	17.30	conserved hypothetical protein
CBB_374	268895	269311	138	15.96	conserved hypothetical protein
CBB_375	269337	270125	275	29.73	conserved hypothetical protein
CBB_376	270125	271390	421	49.56	conserved hypothetical protein
CBB_377	271430	272620	396	45.37	conserved hypothetical protein
CBB_378	272817	273290	157	18.24	conserved hypothetical protein
CBB_379	273314	273646	110	12.84	conserved hypothetical protein
CBB_380	273650	274237	195	21.98	putative 5'3'-deoxyribonucleotidase
CBB_381	274231	274755	184	21.05	conserved hypothetical protein
CBB_382	274908	275486	192	21.86	conserved hypothetical protein
CBB_383	276036	276362	108	12.90	conserved hypothetical protein
CBB_384	276721	276870	49	5.31	conserved hypothetical protein
CBB_385	276870	277124	84	9.80	conserved hypothetical protein
CBB_386	277124	277815	263	30.18	conserved hypothetical protein
CBB_387	277908	278285	125	15.24	conserved hypothetical protein
CBB_388	278242	278703	153	17.78	conserved hypothetical protein
CBB_389	278703	278972	89	10.60	conserved hypothetical protein
CBB_390	278969	279484	171	20.33	conserved hypothetical protein
CBB_391	279438	279893	151	17.79	hypothetical protein
CBB_392	279874	280302	142	16.92	peptidyl-tRNA hydrolase
CBB_393	280575	280934	119	13.85	putative membrane protein
CBB_394	280924	281157	77	8.75	conserved hypothetical protein
CBB_395	281139	281300	53	6.27	hypothetical protein
CBB_396	281303	281557	84	10.07	conserved hypothetical protein
CBB_397	281564	281926	120	13.63	conserved hypothetical protein
CBB_398	281910	282374	154	18.72	conserved hypothetical protein
CBB_399	282396	283256	286	32.94	DNA-cytosine methyltransferase
CBB_400	283249	283560	103	11.73	conserved hypothetical protein
CBB_401	283563	283904	113	13.37	conserved hypothetical protein

CBB_402	283901	284605	234	27.62	conserved hypothetical protein
CBB_403	284616	284972	118	14.28	conserved hypothetical protein
CBB_404	284976	285500	174	20.74	hypothetical protein
CBB_405	285502	285699	65	7.88	conserved hypothetical protein
CBB_406	286108	286341	77	8.78	conserved hypothetical protein
CBB_407	286338	286478	46	5.33	conserved hypothetical protein
CBB_408	286581	286904	107	12.32	conserved hypothetical protein
CBB_409	286950	287381	143	16.39	conserved hypothetical protein
CBB_410	287391	287585	64	7.45	conserved hypothetical protein
CBB_411	287587	287937	116	13.34	conserved hypothetical protein
CBB_412	287948	288283	111	13.02	conserved hypothetical protein
CBB_413	288587	288790	67	7.68	hypothetical protein
CBB_414	288793	288909	38	4.50	hypothetical protein
CBB_415	288912	289204	97	11.07	putative membrane protein
CBB_416	289202	289468	88	10.43	hypothetical protein
CBB_417	289425	289715	96	11.02	conserved hypothetical protein
CBB_418	289786	289986	66	7.38	conserved hypothetical protein
CBB_419	290507	290740	77	8.78	conserved hypothetical protein
CBB_420	290851	291255	134	15.88	HNH
CBB_421	291456	291608	50	5.85	putative membrane protein
CBB_422	291615	291815	66	7.92	conserved hypothetical protein
CBB_423	291815	292030	71	7.79	putative membrane protein
CBB_424	292027	292224	65	7.46	putative membrane protein
CBB_425	292224	292664	146	16.95	conserved hypothetical protein
CBB_426	292771	292983	70	7.96	hypothetical protein
CBB_427	292983	293228	81	9.06	hypothetical protein
CBB_428	293231	293458	75	8.91	hypothetical protein
CBB_429	293458	293718	86	9.76	hypothetical protein
CBB_430	293781	293984	67	7.37	hypothetical protein
CBB_431	293985	294785	266	31.18	conserved hypothetical protein

CBB_432	294787	295089	100	11.17	hypothetical protein
CBB_433	295082	295489	135	16.01	conserved hypothetical protein
CBB_434	295486	295713	75	8.65	hypothetical protein
CBB_435	295860	296153	97	11.65	conserved hypothetical protein
CBB_436	296228	296510	93	10.10	conserved hypothetical protein
CBB_437	296513	296710	65	7.77	conserved hypothetical protein
CBB_438	296694	297059	121	14.35	conserved hypothetical protein
CBB_439	297059	297415	118	13.86	conserved hypothetical protein
CBB_440	297566	297799	77	8.77	conserved hypothetical protein
CBB_441	287796	298419	207	22.95	hypothetical protein
CBB_442	298580	298416	54	6.21	putative membrane protein
CBB_443	298638	299012	124	13.92	putative membrane protein
CBB_444	299204	299788	194	22.99	conserved hypothetical protein
CBB_445	299807	300313	168	19.31	conserved hypothetical protein
CBB_446	300462	300677	71	8.28	conserved hypothetical protein
CBB_447	300741	300983	80	9.42	conserved hypothetical protein
CBB_448	300996	301466	156	17.67	putative membrane protein
CBB_449	301822	301499	107	12.97	hypothetical protein
CBB_450	301908	302054	48	5.50	conserved hypothetical protein
CBB_451	302871	303272	133	15.96	conserved hypothetical protein
CBB_452	303296	303505	69	8.15	conserved hypothetical protein
CBB_453	303601	303825	74	8.40	conserved hypothetical protein
CBB_454	303825	304223	132	15.67	conserved hypothetical protein
CBB_455	304960	305559	199	22.79	conserved hypothetical protein
CBB_456	305594	305947	117	13.68	conserved hypothetical protein
CBB_457	306246	306680	144	16.17	conserved hypothetical protein
CBB_458	306761	307072	103	12.15	conserved hypothetical protein
CBB_459	307418	307648	76	9.02	conserved hypothetical protein
CBB_460	307810	308058	82	9.86	hypothetical protein
CBB_461	308384	308710	108	12.71	conserved hypothetical protein

CBB_462	308798	309034	78	9.29	conserved hypothetical protein
CBB_463	309027	309512	161	19.18	conserved hypothetical protein
CBB_464	309514	309717	67	8.19	conserved hypothetical protein
CBB_465	309819	310202	127	15.22	hypothetical protein
CBB_466	310195	310458	87	10.19	conserved hypothetical protein
CBB_467	310740	310958	72	8.66	conserved hypothetical protein
CBB_468	310951	311301	116	14.25	conserved hypothetical protein
CBB_469	311332	311727	131	15.68	conserved hypothetical protein
CBB_470	311949	312266	105	12.42	conserved hypothetical protein
CBB_471	312370	312621	83	9.83	conserved hypothetical protein
CBB_472	312784	313227	147	17.05	conserved hypothetical protein
CBB_473	313316	313441	41	4.56	conserved hypothetical protein
CBB_474	313431	313574	47	5.31	conserved hypothetical protein
CBB_475	313561	313809	82	9.26	conserved hypothetical protein
CBB_476	313917	314393	158	18.14	PrlF antitoxin like protein
CBB_477	314383	314901	172	19.72	conserved hypothetical protein
CBB_478	314891	315019	42	5.11	conserved hypothetical protein
CBB_479	315019	315333	104	11.86	putative membrane protein
CBB_480	315333	315671	112	12.40	putative membrane protein
CBB_481	315671	315988	105	11.60	putative lipoprotein
CBB_482	315998	316297	99	10.85	putative membrane protein
CBB_483	316356	316493	45	5.33	conserved hypothetical protein
CBB_484	316468	316671	67	7.91	conserved hypothetical protein
CBB_485	316707	317192	161	18.55	conserved hypothetical protein
CBB_486	317245	317511	88	10.26	hypothetical protein
CBB_487	317511	317957	148	16.31	putative lipoprotein
CBB_488	318005	318136	43	4.84	conserved hypothetical protein
CBB_489	318191	318814	207	24.12	conserved hypothetical protein
CBB_490	318898	319257	119	14.22	conserved hypothetical protein
CBB_491	319369	319557	62	7.20	conserved hypothetical protein

CBB_492	319565	319762	65	7.44	conserved hypothetical protein
CBB_493	319847	320068	73	8.43	hypothetical protein
CBB_494	320745	321431	228	24.72	structural protein
CBB_495	321487	322026	179	20.99	S-adenosyl-L-methionine-dependent methyltransferase
CBB_496	323175	322042	377	40.94	structural protein
CBB_497	323350	323925	191	21.62	conserved hypothetical protein
CBB_498	323937	324242	101	11.57	putative membrane protein
CBB_499	324246	324728	160	18.21	putative membrane protein
CBB_500	324788	325309	173	19.65	conserved hypothetical protein
CBB_501	325354	325824	156	18.63	putative membrane protein
CBB_502	325834	326283	149	16.81	putative membrane protein
CBB_503	326355	326600	81	9.24	conserved hypothetical protein
CBB_504	326593	327318	241	28.55	putative membrane protein
CBB_505	327311	327430	39	3.96	conserved hypothetical protein
CBB_506	327433	327867	144	16.72	putative membrane protein
CBB_507	327921	328448	175	20.65	conserved hypothetical protein
CBB_508	328450	328773	107	12.01	conserved hypothetical protein
CBB_509	328773	329156	125	14.78	conserved hypothetical protein
CBB_510	329214	329405	63	6.99	conserved hypothetical protein
CBB_511	330631	330861	76	8.77	transcriptional regulator
CBB_512	330858	331007	49	5.55	conserved hypothetical protein
CBB_513	331103	332091	330	37.92	subfamily RNA polymerase sigma-70 subunit
CBB_514	332108	312818	236	26.63	conserved hypothetical protein
CBB_515	332827	333018	63	7.20	conserved hypothetical protein
CBB_516	333073	334047	324	37.22	conserved hypothetical protein
CBB_517	334461	334634	98	11.47	hypothetical protein
CBB_518	334461	334634	57	6.57	hypothetical protein
CBB_519	334700	335134	144	16.95	endonuclease V N-glycosylase UV repair enzyme

CBB_520	335198	335383	61	6.99	conserved hypothetical protein
CBB_521	335458	335853	131	15.54	conserved hypothetical protein
CBB_522	335850	336044	64	7.15	putative membrane protein
CBB_523	336041	336274	77	8.82	conserved hypothetical protein
CBB_524	336949	337752	267	31.25	conserved hypothetical protein
CBB_525	337822	338832	336	39.56	conserved hypothetical protein
CBB_526	338899	339423	174	19.85	conserved hypothetical protein
CBB_527	339433	340227	264	30.74	conserved hypothetical protein
CBB_528	340294	340860	188	21.78	HNH
CBB_529	340862	341182	106	12.48	putative membrane protein
CBB_530	341299	342369	356	40.50	conserved hypothetical protein
CBB_531	342424	342612	62	7.21	conserved hypothetical protein
CBB_532	342609	342866	85	10.40	conserved hypothetical protein
CBB_533	342863	342030	55	6.23	conserved hypothetical protein
CBB_534	343063	343335	90	10.45	structural protein
CBB_535	343417	344337	306	35.70	conserved hypothetical protein
CBB_536	344368	345027	219	24.75	structural protein
CBB_537	345042	345689	215	24.14	structural protein
CBB_538	345789	346571	260	29.12	conserved hypothetical protein
CBB_539	345789	346571	107	12.12	putative membrane protein
CBB_540	346911	347099	62	7.10	putative membrane protein
CBB_541	347102	347263	53	5.89	putative membrane protein
CBB_542	347277	348047	256	29.98	conserved hypothetical protein
CBB_543	348094	348444	116	13.14	putative membrane protein
CBB_544	348444	348782	112	13.12	putative membrane protein
CBB_545	348782	349132	116	13.29	putative membrane protein
CBB_546	349199	349495	98	10.86	putative membrane protein
CBB_547	349547	350170	207	22.41	putative lipoprotein
CBB_548	350849	350196	217	23.81	structural protein
CBB_549	351530	350859	223	24.84	structural protein

CBB_550	352396	351539	285	31.83	kelch-like protein
CBB_551	353100	352396	234	25.38	structural protein
CBB_552	353952	353119	280	31.25	structural protein
CBB_553	354701	354030	223	25.40	structural protein
CBB_554	355464	354667	265	29.91	kelch-like protein
CBB_555	357350	357538	62	7.00	hypothetical protein
CBB_556	357723	358400	225	25.17	conserved hypothetical protein
CBB_557	358550	358819	89	10.13	conserved hypothetical protein
CBB_558	358955	359239	94	10.83	conserved hypothetical protein
CBB_559	359275	359664	129	14.96	conserved hypothetical protein
CBB_560	359889	359716	57	6.30	hypothetical protein
CBB_561	359888	360040	50	5.58	hypothetical protein
CBB_562	360088	360360	90	10.08	conserved hypothetical protein
CBB_563	360424	360870	148	16.80	conserved hypothetical protein
CBB_564	360947	361402	151	17.13	conserved hypothetical protein
CBB_565	361483	361608	41	4.54	hypothetical protein
CBB_566	361681	361908	75	8.43	hypothetical protein
CBB_567	361982	362404	140	16.39	conserved hypothetical protein
CBB_568	362586	362921	111	12.23	conserved hypothetical protein
CBB_569	362991	363158	55	6.68	conserved hypothetical protein
CBB_570	363242	363421	59	6.51	conserved hypothetical protein
CBB_571	363494	363865	123	14.45	conserved hypothetical protein
CBB_572	363912	364166	84	9.99	conserved hypothetical protein
CBB_573	365171	365329	52	5.88	hypothetical protein
CBB_574	364619	364852	77	8.56	conserved hypothetical protein
CBB_575	364876	365034	52	6.06	putative membrane protein
CBB_576	365046	365249	67	7.51	conserved hypothetical protein
CBB_577	365282	365425	47	6.86	conserved hypothetical protein
CBB_578	365504	365707	67	7.56	conserved hypothetical protein
CBB_579	365737	366087	116	13.40	conserved hypothetical protein

CBB_580	366314	366565	83	9.44	putative membrane protein
CBB_581	366597	366755	52	5.87	hypothetical protein
CBB_582	366764	367639	291	31.54	structural protein
CBB_583	367779	367976	65	8.04	conserved hypothetical protein
CBB_584	367940	368065	41	4.86	conserved hypothetical protein
CBB_585	368055	368249	64	8.12	hypothetical protein
CBB_586	368293	368646	117	13.52	putative membrane protein
CBB_587	368643	369032	129	14.27	hypothetical protein
CBB_588	369104	369295	63	7.27	conserved hypothetical protein
CBB_589	369358	369744	128	14.27	conserved hypothetical protein
CBB_590	369864	370106	80	9.53	conserved hypothetical protein
CBB_591	370163	370681	172	18.82	conserved hypothetical protein
CBB_592	370745	371083	112	12.64	conserved hypothetical protein
CBB_593	371326	371673	115	12.92	hypothetical protein
CBB_594	371764	371982	72	8.27	conserved hypothetical protein
CBB_595	372158	372760	200	23.42	conserved hypothetical protein
CBB_596	372965	373297	110	12.84	conserved hypothetical protein
CBB_597	373507	373749	80	9.16	conserved hypothetical protein
CBB_598	373829	374716	295	33.78	conserved hypothetical protein
CBB_599	374792	375130	112	13.26	conserved hypothetical protein
CBB_600	375334	375564	76	8.64	conserved hypothetical protein
CBB_601	375648	375995	115	13.49	conserved hypothetical protein
CBB_602	376195	376386	63	7.18	hypothetical protein
CBB_603	376456	376632	58	6.43	conserved hypothetical protein
CBB_604	377036	377305	89	10.35	conserved hypothetical protein
CBB_605	377375	377932	185	20.69	hypothetical protein

Table S6.3. tRNA genes in the genome of *Pectobacterium* phage CBB (Detected using tRNAscan –SE and ARAGORN).

Feature	Location	Function	Predection program
tRNA gene 1	268,792 bp - 268,878 bp	Ser, anticodon GCT	tRNAscan-SE
tRNA gene 1A	272,675 bp - 272,749 bp	tRNA-Trp(cca)	ARAGORN
tRNA gene 1B	275,819 bp - 275,892 bp	tRNA-Thr(tgt)	ARAGORN
tRNA gene 2	276,549 bp - 276,669 bp	Val, anticodon AAC (intron present 276,585 bp – 276,634 bp)	tRNAscan-SE
tRNA gene 3	280,358 bp - 280,440 bp	Leu, anticodon TAG	tRNAscan-SE
tRNA gene 4	288,346 bp – 288,417 bp	Arg, anticodon TCT	tRNAscan-SE
tRNA gene 5	289,995 bp - 290,067 bp	Sup, anticodon CTA or tRNA-Pyl(cta)	TRNAscan -SE ARAGORN
tRNA gene 6	290,284 bp – 290,367 bp	Leu, anticodon TAA	tRNAscan-SE
tRNA gene 7	290,373 bp – 290,482 bp	Val, anticodon GAC (intron present 290,373 bp – 290,482 bp)	tRNAscan-SE
tRNA gene 7A	290,748 bp - 290,824 bp	tRNA-Met(cat)	ARAGORN
tRNA gene 8	292,673 bp - 292,754 bp	Leu, anticodon TAA	tRNAscan-SE
tRNA gene 9	295,724 bp – 295,837 bp	Pseudo tRNA	tRNAscan-SE
tRNA gene 10	297,425 bp - 297,546 bp	Phe, anticodon GAA, (intron present 297,459 bp – 297,510 bp)	tRNAscan-SE
tRNA gene 11	299,090 bp - 299,176 bp	Leu, anticodon CAA	tRNAscan-SE
tRNA gene 12	302,062 bp – 302,132 bp	Glu, anticodon TTC	tRNAscan-SE
tRNA gene 13	302,143 bp - 302,231 bp	tRNA-Tyr(gta)	ARAGORN
tRNA gene 14	302,564 bp - 302,654 bp	tRNA-Ser(tga)	ARAGORN
tRNA gene 15	302,662 bp - 302,780 bp	Ser, anticodon AGA (intron present 302,696 bp – 302,744 bp)	tRNAscan-SE
tRNA gene 16	304, 228 bp – 304,298 bp	Ile, anticodon GAT	tRNAscan-SE
tRNA gene 16A	307,123 bp - 307,195 bp	tRNA-Asn(gtt)	ARAGORN
tRNA gene 17	307,717 bp – 307,789 bp	Gln, anticodon TTG	tRNAscan-SE

tRNA gene 17A	308,195 bp - 308,267 bp	tRNA-Gly(tcc)	ARAGORN
tRNA gene 18	309,726 bp - 309,787 bp	Asp, anticodon GTC	tRNAscan-SE
tRNA gene 19	310, 643 bp – 310,718 bp	Arg, anticodon ACG	tRNAscan-SE
tRNA gene 20	311,782 bp – 311,853 bp	Pro, anticodon TGG	tRNAscan-SE
tRNA gene 21	311,859 bp - 311,931 bp	Pro, anticodon TGG	tRNAscan-SE
tRNA gene 22	312,275 bp – 312,348 bp	His, anticodon GTG	tRNAscan-SE
tRNA gene 23	312, 637 bp – 312,710 bp	Phe, anticodon GAA	tRNAscan-SE
tRNA gene 24	313,823 bp – 313,895 bp	Lys, anticodon TTT	tRNAscan-SE
tRNA gene 25	318,802 bp – 318,876 bp	Cys, anticodon GCA	tRNAscan-SE
tRNA gene 26	319,754 bp – 319,826 bp	Met, anticodon CAT	tRNAscan-SE
tRNA gene 26A	320,334 bp- 320,411 bp	tRNA-Met(cat)	ARAGORN
tRNA gene 27	320, 479 bp – 320,511 bp	Ala, anticondon TGC	tRNAscan-SE

Table S6.4. *Pectobacterium* phage CBB tRNA genes and codon usage (CBB_1 - CBB_554)

aa	Frequency	Codon	Occurrence	Codon Freq.	Freq./aa	tRNA gene occurrence
Ala	5.83%	GCA	3058	2.86%	49.06%	1
		GCC	313	0.29%	5.02%	
		GCG	584	0.55%	9.37%	
		GCT	2278	2.13%	36.55%	
Arg	4.09%	CGA	380	0.36%	8.69%	
		CGC	432	0.40%	9.88%	
		CGG	130	0.12%	2.97%	
		CGT	2587	2.42%	59.17%	1
		AGA	684	0.64%	15.65%	1
		AGG	159	0.15%	3.64%	
Asm	6.16%	AAC	2506	2.34%	38.06%	1
		AAT	4079	3.81%	61.94%	
Asp	6.65%	GAC	1525	1.43%	21.44%	1
		GAT	5588	5.22%	78.56%	
Cys	1.14%	TGC	287	0.27%	23.62%	1
		TGT	928	0.87%	76.38%	
Gln	3.60%	CAA	2750	2.57%	71.39%	1
		CAG	1102	1.03%	28.61%	
Glu	7.35%	GAA	6766	6.33%	86.13%	1
		GAG	1090	1.02%	13.87%	
Gly	5.83%	GGA	1316	1.23%	21.12%	1
		GGC	565	0.53%	9.07%	
		GGG	336	0.31%	5.39%	
		GGT	4015	3.75%	64.43%	
His	1.97%	CAC	593	0.55%	28.13%	1
		CAT	1515	1.42%	71.87%	
Ile	6.88%	ATA	1017	0.95%	13.83%	
		ATC	1620	1.51%	22.02%	1
		ATT	4719	4.41%	64.15%	
Leu	7.75%	CTA	1751	1.64%	21.13%	1
		CTC	264	0.25%	3.19%	
		CTG	715	0.67%	8.63%	
		CTT	1750	1.64%	21.12%	
		TTA	2521	2.36%	30.43%	2
		TTG	1284	1.20%	15.50%	1
Lys	6.58%	AAA	5314	4.97%	75.47%	1
		AAG	1727	1.61%	24.53%	
Met	2.35%	ATG	2512	2.35%	100.00%	3
Phe	4.58%	TTC	2109	1.97%	43.08%	2
		TTT	2786	2.60%	56.92%	

Pro	3.28%	CCA	1951	1.82%	55.70%	2
		CCC	86	0.08%	2.46%	
		CCG	329	0.31%	9.39%	
		CCT	1137	1.06%	32.46%	
Ser	6.63%	TCA	1883	1.76%	26.54%	1
		TCC	329	0.31%	4.64%	
		TCG	313	0.29%	4.41%	
		TCT	1836	1.72%	25.88%	1
		AGC	544	0.51%	7.67%	1
		AGT	2189	2.05%	30.86%	
Thr	6.41%	ACA	2141	2.00%	31.23%	1
		ACC	677	0.63%	9.87%	
		ACG	477	0.45%	6.96%	
		ACT	3561	3.33%	51.94%	
Trp	1.21%	TGG	1289	1.21%	100.00%	1
Tyr	4.89%	TAC	1752	1.64%	33.53%	1
		TAT	3473	3.25%	66.47%	
Val	6.85%	GTA	3095	2.89%	42.22%	
		GTC	365	0.34%	4.98%	1
		GTG	666	0.62%	9.08%	
		GTT	3205	3.00%	43.72%	1

Table S6.5. Quasicore proteins of the T4 like phage identified within the genome of *Pectobacterium* phage CBB with sequence comparison to homologs in Enterobacter phage T4 and Vibrio phage KVP40 (reference strains of *T4virus* and *Schizot4virus*, respectively) using BLASTP.

Function	T4 like core protein	CBB homolog	Accession no. of T4 protein	Identity	E-value	Accession no. of KVP40	Identity	E-value
DNA replication, repair and recombination	gp49-Endonclease VII	CBB_288	NP_049692.1	30	3.01E-06	NP_899379.1	32	1.00E-11
	gp30 - DNA ligase	CBB_100	NP_049813.1	30	1.00E-47	NP_899305.1	33	2.00E-63
	RNaseH	CBB_172	NP_049859.1	29	1.00E-13	NP_899249.1	24	1.00E-13
	gp52- Type II DNA toppisomerase subunit	CBB_325	NP_049875.1	27	1.00E-33	NP_899529.1	30	4.00E-42
	Dda- short range DNA helicase	CBB_315	NP_049632.1	30	4.00E-30	NP_899402.1	28	1.18E+04
Auxillary metabolism	NrdG - subunit of anaerobic ribonucleotide reductase complex	CBB_145	NP_049688.1	48	4.00E-51	NP_899263.1	43	3.00E-47
	cd - dCMP deaminase	CBB_110	NP_049828.1	41	3.00E-38	NP_899367.1	39	1.00E-33
	Frd - Dihydrofolate reductase	CBB_313	NP_049850.1	30	2.00E-13	NP_899254.1	36	6.00E-18

	Td - thymidylate synthtase	CBB_233	NP_049848.1	43	4.00E-18	NP_899279.1	34	2.00E-57
	Tk - thymidine kinase	CBB_173	NP_049819.1	44	2.00E-51	NP_899436.1	46	3.00E-46
	gp1 - dDMP kinase	CBB_203	NP_049852.1	29	3.00E-04	NP_899338.1	67	0.59

Table S6.6. Phage portal vertex proteins used for phylogenetic study of CBB and the GAP32-like phage.

The conserved region of CBB portal vertex protein used to construct the phylogenetic tree.

Sequence –
DMQPHQAMAYVERVKNDIHQRRIPSNKGGSTSLMDAAYNPLSILEDYFFPQTAEGRGSSVETLPGGDNLG
QIDDLRYFNNKLIRGLQIPASYLPMGPDDGGVALFGDGATQAMASELRFNNECMRYQRIISRIFDEEFKRYM
IKNGYNISASSFEVTFNPPMNFAANRKAEMDAKLIQTYM

Phage	Portal vertex protein - Accession no.
Enterobacteria_phage_IME08	YP_003734309.1
Escherichia_phage_slur07	YP_009197287.1
uncultured_Mediterranean_phage_uvMED	BAR36277.1
uncultured_Mediterranean_phage_uvMED(2)	BAR31405.1
uncultured_Mediterranean_phage_uvMED(3)	BAR29473.1
Synechococcus_phage_S-IOM18	YP_008126421.1
Synechococcus_phage_S-RIM2_R1_1999	YP_007675582.1
uncultured_Mediterranean_phage	ANS04762.1
Enterobacteria phage RB14	YP_002854503.1
Aeromonas_phage_25	YP_656382.1
Synechococcus_phage_metaG-MbCM1	YP_007001618.1
Enterobacteria_phage_RB16	YP_003858509.1
Aeromonas_phage_65	YP_004300925.1
Aeromonas_phage_Aeh1	NP_944108.1
Aeromonas_phage_PX29	YP_009011644.1
Enterobacteria_phage_JS10	YP_002922513.1
Aeromonas_phage_Aes508	YP_007010833.1
uncultured Mediterranean phage	ANS05306.1
Cyanophage_P-RSM6	YP_007675137.1
Escherichia_phage_ECML-134	YP_009102640.1
Prochlorococcus_phage_P-SSM3	YP_008129949.1
Shigella_phage_pSs-1	YP_009110988.1
Enterobacteria_phage_QL01	YP_009202903.1
Ralstonia_phage_RSP15	BAU39992.1
uncultured_Mediterranean_phage	ANS05306.1
uncultured_Mediterranean_phage(2)	ANS05306.1
uncultured_Mediterranean_phage_uvMED	BAR34134.1
Cyanophage_P-RSM1	YP_007877738.1
Cyanophage_Syn30	YP_007877943.1
Escherichia_phage_121Q	YP_009102185.1
Synechococcus_phage_S-CAM8	YP_008125640.1
Synechococcus_phage_S-RIM8_A.HR1	YP_007518198.1
Synechococcus_phage_syn9	YP_717798.1
Synechococcus_phage_S-CRM01	YP_004508471.1
Synechococcus_phage_S-RSM4	YP_003097343.1

Cyanophage_S-RIM50	AMO42907.1
Enterobacteria_phage_RB27	YP_009102372.1
Enterobacteria_phage_RB3	YP_009102372.1
Enterobacteria_phage_RB32	YP_803110.1
Escherichia_phage_vB_EcoM-UFV13	ANA50202.1
Salmonella_phage_STML-198	YP_009148150.1
Shigella_phage_SHFML-11	ANN86599.1
Aeromonas_phage_44RR2.8t	NP_932511.1
Cyanophage_P-TIM40	YP_009188207.1
Cyanophage_S-RIM32	AMO43137.1
Enterobacteria_phage_RB69	NP_861872.1
Enterobacteria_phage_vB_EcoM_VR20	YP_009207360.1
Enterobacteria_phage_vB_EcoM_VR5	YP_009205862.1
Enterobacteria_phage_vB_KleM-RaK2	YP_007007244.1
Escherichia_phage_APCEc01	YP_009225085.1
Escherichia_phage_PBECO_4	SCA80472.1
Escherichia_phage_vB_EcoM_JS09	YP_009030623.1
Aeromonas_phage_CC2	YP_007010339.1
Cronobacter_phage_vB_CsaM_GAP32	YP_006987350.1
Edwardsiella_phage_PEi20	YP_009190346.1
Enterobacteria_phage_Bp7	YP_007004276.1
Enterobacteria_phage_JSE	YP_002922232.1
Escherichia_phage_e11/2	YP_009030776.1
Escherichia_phage_UFV-AREG1	ANH50304.1
Escherichia_phage_vB_EcoM_PhAPEC2	YP_009056756.1
Pelagibacter_phage_HTVC008M	YP_007517949.1
Prochlorococcus_phage_P-HM1	YP_004322541.1
Prochlorococcus_phage_P-HM2	YP_004323487.1
Prochlorococcus_phage_P-RSM4	YP_004323264.1
Prochlorococcus_phage_P-SSM2	YP_214363.1
Prochlorococcus_phage_P-SSM4	YP_214665.1
Prochlorococcus_phage_P-SSM7	YP_004324951.1
Prochlorococcus_phage_Syn33	YP_004323727.1
Sinorhizobium_phage_phiM12	YP_009142980.1
Sinorhizobium_phage_phiN3	YP_009212309.1
Synechococcus_phage_S-PM2	YP_195138.1
Synechococcus_phage_S-ShM2	YP_004322786.1
Synechococcus_phage_S-SM1	YP_004323020.1
Synechococcus_phage_S-SSM5	YP_004324725.1
Synechococcus_phage_S-SSM7	YP_004324197.1
Vibrio_phage_KVP40	BAA77377.1
Vibrio_phage_nt-1	YP_008125180.1
Yersinia_phage_PST	YP_009153767.1
Prochlorococcus_phage_MED4-213	YP_007673752.1
Salmonella_phage_S16	YP_007501199.1

Salmonella_phage_vB_SnwM_CGG4-1	ANA49508.1
Synechococcus_phage_ACG-2014a	AIX28031.1
Synechococcus_phage_ACG-2014b	YP_009140680.1
Synechococcus_phage_ACG-2014d	AIX21202.1
Synechococcus_phage_ACG-2014e	YP_009134611.1
Synechococcus_phage_ACG-2014f	AIX42351.1
Synechococcus_phage_ACG-2014g	YP_009133666.1
Synechococcus_phage_ACG-2014h	YP_009008243.1
Synechococcus_phage_ACG-2014i	YP_009140894.1
Synechococcus_phage_ACG-2014j	YP_009134101.1
Synechococcus_phage_S-MbCM100	YP_009007965.1
Shigella_phage_Shfl2	YP_004415062.1
Escherichia_phage_HY03	AKJ72700.1
Vibrio_phage_VH7D	YP_009006286.1
uncultured_phage_MedDCM-OCT-S09-C7	ADD95604.1
Aeromonas_phage_phiAS4	YP_003969129.1
Aeromonas_phage_phiAS5	YP_003969303.1
Klebsiella phage vB_KpnM_KpV477	ANT40618.1
Klebsiella phage JD18	YP_009190743.1
Enterobacteria phage T4	NP_049782.1

Figure S6.1. Maximum likelihood tree created from the alignment of the major capsid protein of 100 homologous sequences from different T4-like phages to that of the major capsid protein of phage CBB found using a BLASTP search. 100 bootstrap replicates were conducted.

Table S6.7. Promoter 1 of Phage CBB found using MEME

Promoter	location	Sequence
Porf_36	13,365-13,402	AAAAATAGCTTGCATCACTCGTAATTATGAAGTATAAT
Porf_41	15,759-15,796	AAAAAGGAGTTGACAACAGTTGGCTCCTTTGCTATACT
Porf_44	17,515-17,552	AAAAATTCCCTTGCTTTCTTTCTTTCGTGTGGTATACT
Porf_83	39,070-39,107	AATACTGACTTGACACGAGGCAGCAAGAGTGTTACAAT
Porf_166	93,493-93,530	AAAAATAGCTTGATTTTGAGCTACATAATTGATATAAT
Porf_181	100,667`-100,704	ACAAATTGCTTGACATGGCAGTGAGTATCATATATCAT
Porf_189	106,894-106,931	GAATAAACCTTGACAGTACGATAAAAAAGTGCTACATT
Porf_360	262,781-262,818	AAATTTACCTTGACAAAATCAACTCCATCCCTGATAAT
Porf_395	281,079-281,116	AAAAATGACTTGGCAGGAGATTAATATTCCTGTACAAT
Porf_436	296,151-296,188	TAAAAATCCTTGACTTGCGTCGGTTGACCTGATAGAAT
Porf_491	319,303-319,340	CAAAACACCTTGACAGTGGCATCCTCTCATGGCAAAAT
Porf_497	323,285-323,322	AGATTTTGCTTGACAAGACGGCATTTTTTGTTTATAAT
Porf_503	326,287-326,324	AAATACCAGTTGACTCAACGTGGGTTAACTGGTATACT
Porf_510	329,125-329,162	AATTTGGCCTTGCAATCAGCGAACGCATGTGATATATT
Porf_524	336,865-336,902	AGAAAATCCTTGACTTGCTTCAGCAAGGCGAGTACACT
Porf_525	337,752-337,789	AAAAATACATTGACGTGGCACATTTTTTGTGCCATAAT
Porf_538	345,724-345,761	TATTTAGGCTTGACACGATGTTAAATCTTTGATATAAT
Porf_546	349,130-349,167	TAATTTGGCTTGACTTAGGCTTTCGAACTCCGTATAAT

Table S6.8. Promoter 2 of Phage CBB found using MEME
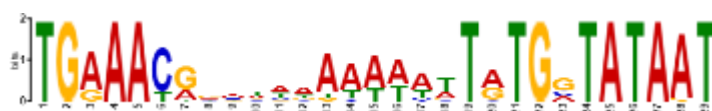
Promoter	Location	Sequence
Porf_66	29,539-29,567	TGAAACAGAATAATTTTTTATGGTATAAT
Porf_113	54,897-54,925	TGAAACGTAAGGAATTTTTCTGCTATAAT
Porf_193	108,832-108,860	TGAAACAGCTTTAAAAATTATGGTATAGT
Porf_218	127,475-127,503	TGAAACGTACCAAAATTTTATGGTATAAT
Porf_235	145,858-145,886	TGGAACGACCTAAAAAAATGTGCTATAAT
Porf_244	168,994-169,022	TGGAACGTTTTTTAAAATTGTGGTATAAT
Porf_258	184,600-184,628	TGAAACGTATACATTTATTATGGTATAAT
Porf_262	187,472-187,500	TGAAACGAAAATAAAACGTATGATATAAT
Porf_265	191,483-191,511	TGAAACGGGTACATTTTTTATGATATAAT
Porf_272	195,587-195,615	TGAAACTGCACATAAAAATGTGGTATAAT
Porf_282	206,162-206,190	TGGAATGAGTTAAATTTTTGTGATATAAT
Porf_290	211,052-211,080	TGGAATGCATTAAAAAAATGTGGTATAAT
Porf_300	217,526-217,554	TGAAACAATAAAAAAACATGTGCTATAAT
Porf_307	224,015-224,043	TGAAACGCCAAAATAAAATATGGTATAAT
Porf_312	226,886-226,914	TGAAACAGCTTGGAATATTATGATATAAT
Porf_324	236,752-236,780	TGAAACAAAGAAACAAATTATGCTATAAT
Porf_329	243,482-243,510	TGAAATGGTAAAAAAAATTATGGTATAAT
Porf_418	289,731-289,759	TGAAACGGCGATAAAAACTGTGGTATAAT

Table S6.9. Promoter 3 of Phage CBB found using MEME based on analysis of 100bps upstream regions of structural genes

Promoter	Location	Sequence
Porf_112	54,769-54,782	ATAAATATTAATAT
Porf_165	93,137-93,150	ATGAATACCGCAAA
Porf_169	95,252-95,265	CTAAATACGGTATC
Porf_201	111,914-111,927	CTAAATATGATATA
Porf_204	133,744-113,757	ATAAATAACATTAT
Porf_205	116,515-116528	ATAAATACATTAAA
Porf_206	177,288-117,301	CTAAATATAATATA
Porf_210	120,228-120,241	ATAAATAAAATAAA
Porf_213	122,794-122,807	CTAAATTCATTAAA
Porf_224	132,558-132,571	ATAAATACTGGATA
Porf_226	137,986-137,999	ATAAATACTTGAAA
Porf_227	138,970-138,983	ATAAATATAAGAAA
Porf_228	140,195-140,208	ATAAATATTTGTAA
Porf_229	140,234-140,247	ATAAATAAGTAAAA
Porf_234	145,841-145,854	ATAAATACGGTATA
Porf_236	158,306-158,319	ATAAATAAAACAAA
Porf_237	161,845-161,858	ATAAATATTTCTAT
Porf_239	162,810-162,823	ATAAATACTTATAC
Porf_243	168,976-168,989	ATAAATACTTGATA
Porf_250	176,224-176,237	ATAAATAACTAATA
Porf_251	176,198-176211	ATAAATACGACTAT
Porf_256	182,056-182,069	ATAAATATCTTTAC
Porf_260	185,404-185,417	ATAAATATTTCTAT
Porf_271	195,498-195,511	ATAAATAACATTAA
Porf_276	199,519-199,532	ATAAATATATCTAT
Porf_289	210,031-210,044	ATAAATATTGTATA
Porf_298	216,496-216,509	CTAAATAAGGATAT
Porf_302	218,818-218,831	CTAAATAATTCTAT
Porf_320	231,630-231,643	ATAAATACTCAAAA
Porf_496	323,194-323,207	ATAAATACTTTATA
Porf_534	343,023-343,036	ATAAATAAATGAAA
Porf_537	345,717-345,730	CTAAATAATTAATA
Porf_551	353,119-353,132	CTAAATTTTACAAA
Porf_552	353,975-353,988	GTAAATACATAATA
Porf_553	354,727-354,740	GTAAATACAGGATA

Table S7.10. Terminators of Phage CBB detected using ARNold and confirmed using Mfold Quikfold

Terminator	Coordinates	Sequence	ΔG kcal/mol
Torf8	3,939-3,974	GGGGGAACATGAAAGTTCCCCTgTTTTTTTCCTTGA	-13.8
Torf18	7,952 - 7,980	GGGGAACATTGTTCCCCTTTTTAAGTGAG	-10.9
Torf26	10,171 - 10,208	AGTACTGGTGAAATATCCAGTACTgTTTTTATGTACAG	-11.4
Torf39	15,175 - 15,205	AGGTTGACTTCGGTCGGCCTtTTTTGTTATACTT	-11.8
Torf47	19,234 - 19,268	GGGCAACCAAATAACGGTTGCCCTTTTTTGTATTC	-15.3
Torf52	22,047 - 22, 081	GCCGCTGTCATAATGACGGCGGCTTACTTTATGGG	-18
Torf67	30,374 - 30,401	GGGGCTTAATTGCCCCTTTTCTATAGGA	-11.1
Torf70	32,322 - 32, 354	GGACGGTACAGTGTATCGTCCTTTTTTAATTGA	-12.2
Torf80	38,447 - 38,475	GCCCCGAAATTCGGGGCTTTCTTTTTAAA	-13.4
Torf87	41,945 - 41,972	GGGATACTTCGGTATCCCTTCTTTTTTAAT	-14
Torf106	51,742- 51,774	AGGCCACTTAGTTGTGGCCTTTTTCTTTTAAGG	-12.6
Torf111	54,040 - 54,067	GGAGCCATATGGCTCCTTTTTTATTTTA	-11.9
Torf113	56,086 - 56, 116	GGAGTACTTCGGTACTCCaTTTTAAGTAAGA	-14.4
Torf123	63,856 - 63, 887	GGGGATACTTAGGTATCCCCTTTTTAATTGGA	-14.3
Torf124	64,214 - 64,242	GGCCACGCAATGTGGCCTTTTTTATTGGA	-11
Torf146	76-523 - 76,553	AGCACCTGCAAGGGTGCTTTTTTTGTTTATA	-10
Torf149	84,839 - 84, 869	AGCGGTCACTTGACCGCTTTTTTCTTTTGTG	-11.6
Torf152	86,569 - 86,596	GGGAGTGTTAGCTCCCTTTTACATTTCA	-10
Torf163	92,758 - 92, 787	CCCGTCATAAATGGCGGGTTTTCTATTAAA	-10.4
Torf165	93,533 - 93,565	AGGAAATCGTGAGATTTCCTTTTTTCGTTTGGA	-12.3
Torf173	97,558 - 97,590	TGGGCTGATATAATCAGTCCATATTTTTATGGG	-11

Torf183	104,263 - 104, 292	GCCTCACTCCGGTGGGGCTTTTTCATTTAA	-12
Torf188	106,936 - 106, 971	GGACAGGCAGCGATGCTCTGTCCTTTTTTATTTTAA	-13.8
Torf197	110,637 - 110, 666	GGGCATCTTAGGATGCCCTTTTTTCAATTG	-13
Torf202	113,016 - 113,052	GGAAGAATCTTAATAGGGTTCTTCCTTTTTTCGTTTA	-13.2
Torf204	116,418 - 116,485	GGAGCCATATGGCTCCTTTTTTGTTTTC	-11.9
Torf205	117,258 - 117, 289	GCCCTGGACTTGTCTAGGGCTTTTTTATTGCT	-14.7
Torf215	124,614 - 124, 647	AGCACTGAAAACTCAGTGCTtTTCTTTTTATCAA	-10.9
Torf216	126, 042 - 126, 072	AGGGGCATTATGCCCCTtTTGTTTTATGAGG	-11.2
Torf224	130,573 - 130,603 (complement)	AGGGAGCTTTTGCTCCCTTTCTTTTTACACC	-12.6
Torf223	130,586 - 130, 616	AGGGAGCAAAAGCTCCCTTTTTTTGTTATTG	-12.1
Torf225	132,578 - 132, 611	AGGTCAACTTATGTTGGCCTtTTGTTTTTTGTAG	-10.6
Torf228	138, 985 - 139, 017 (complement)	AGGGAGATACTTATCTCCCTTTTTTCATTGTGA	-11.2
Torf236	148, 214 - 148, 255 (complement)	GGGAGCCATATGGCTCCCTTTTTCTTTATTT	-15.2
Torf235	148,227 - 148, 256	GGGAGCCATATGGCTCCCTTTTTTGTTATA	-15.2
Torf237	158,318 - 158, 348 (complement)	GGGCATCTTCGGATGCCCTTTATTCTTGCAT	-16
Torf246	173, 110 - 173, 140	AGGAGCCATTAGGCTCCTTTTTTTTATGATAT	12.1
Torf247	173, 095 - 173, 128 (complement)	AAGGAGCCTAATGGCTCCTTtTTGTTTATAGGTC	13.5
Torf257	184, 538 - 184, 566	GGACGCTTTATGCGTCCTTTTCGCATTTC	-11.3
Torf261	187,504 - 187, 543	GGCATCCTAGCGGAGAGTTAGGATGCTTTTTTTATACATG	-17.1
Torf262	187,863 - 187, 892	TGCCCCAGGATTGGGGCGTTTTGTTTTAAG	-11.6
Torf265	192,281 - 192, 313	AGGGAGCGTAATGCTCCCTcTTTTTATTTGGAT	-11.7

Torf269	194, 180 - 194, 209	GGAGAACATATGTTCTCCTTTTTCTTTTTT	-10.7
Torf270	174, 166 - 194, 198 (complement)	AGGAGAACATATGTTCTCCTTTTTTTATTACTC	-11.2
Torf277	202, 130 - 202, 164	TTGGGGCAACGGATTTGCCCCGATTTCATTTCATG	-13.7
Torf279	203, 777 - 203, 808	AGGACGCTTATTGCGTCCTTTTTGTTTTTTAG	-11.8
Torf303	221, 845 - 221, 873	TCCCGTCTATGACGGGATTTTTTATCAGG	-10.4
Torf305	223,790 - 223, 816	GGAGGCTTTGCCTCCTTTTTCATGGAT	-10
Torf310	225,406 - 225, 434 (complement)	GGGAGCTTACGCTCCCTTTTTTATGCATT	-11.5
Torf323	236,722 - 236, 757	TAACGACGCTCCAGAGCGTCGTTATTTTTTTGAAAC	-15
Torf325	240,252 - 240,279	GGACGCATAAAGCGTCCTTTATTTTTAAT	-10.8
Torf330	244,250 - 244, 279	GCCCTGCATTGTCAGGGCTTTCTTTTTGTC	-10.5
Torf333	246,154 - 246, 182	AGGGGAGTAATCCCCTTTTCCGTTAGTA	-11
Torf338	249,187 - 249, 219	TGGGGAATGAAAGTTCCCCGcTTTTTAATTTAA	-13.5
Torf340	251, 689 - 251, 720	AGCTCCCAACACTTGGGGGCTTTTTTCATATG	-12.1
Torf344	254, 455 - 254, 489	AGCTGGCCTTAGTTGGTCAGCTTTTTTAAGTAAGA	-13.2
Torf364	265, 682 - 265, 712	CCCTGTACTAAGTACAGGGTTTTTAATTGAG	-11.4
Torf446	300, 688 - 300, 723	TGCCGCCAACATTGTTGGCGGTAcTTTTAAGTTAGA	-15.4
Torf448	301, 467 - 301, 490	GGGCGAAAGCCCTTTTTTAATTGA	-10.8
Torf468	311, 288 - 311, 323	GGGATTTTCCGTAAGGGAAATCCCTTTTTTCGTTTA	-17.8
Torf488	318, 141 - 318, 173	AGGGAGCGGATTGCTCCCTgTTTTATAATGCAT	-12.3
Torf494	321, 450 - 321, 478	GGCAGCCATTGGCTGCCTTTTTTCGTTTA	-13.1
Torf495	322, 016 - 322, 046	GGCACTCATAAAGAGTGCCTTTTGTTTTACG	-12.6
Torf496	322,001 - 322, 035 (complement)	AGGCACTCTTTATGAGTGCCTtTTTTTCTTGTATA	-13.6

Torf500	325, 314 - 325, 350	TTCGGGGATTAACTATCCCCGAAacTTTTAATGGTGA	-14.6
Torf515	333, 025 - 333, 050	GGGTTGACTTCGGTCAGCCCTTTATTTTAATT	-16.8
Torf518	334,640 - 334, 671	GGGTTGACTTCGGTCAGCCCTTTTTTGTATAA	-16.8
Torf520	335, 395 - 335, 426	GGGTTGACTTCGGTCAGCCCTTTTTGCTATAC	-16.8
Torf525	338, 856 - 338, 888	TGCCAAACATCGTGTTTGGCATTTTTATTAAA	-11.2
Torf542	348,061 - 348, 089	ACTCCCTACGGGGAGTTTTTTATTGAGGC	-11.6
Torf546	349, 505 - 349, 535	CCGCCAGCATTGTTGGCGGTTTTTTAATTGC	-12.5
Torf547	350, 176 - 350, 206	AAGCCCCAAAAGGGGCTTTTTTAATTTGTGA	-11.8
Torf548	350,166 - 350, 191	GCCCCAAAAGGGGCTTTTTTAATTTG	-10.4
Torf553	353, 992 - 354, 024 (complement)	AGCTCAACTTAGGTTGGGCTTTTTCTGTTTTTG	-10.1

Table S7.11. Comparison of identified structural proteins of phages CBB, to those identified for Gap32 (Abbasifar, et al. 2014) and RaK2 (Šimoliūnas et al 2014)

CBB ORFs	GAP32 ORFs	Rak2 ORFs	CBB-MS detected	GAP32-MS detected	Rak2-MS detected
_	hypothetical protein GAP32_001	_		yes	
structural protein CBB_029/ CBB_582	hypothetical protein GAP32_025	_	yes	yes	
structural protein CBB_038/CBB_591	hypothetical protein GAP32_032	hypothetical protein ORF275	yes	yes	yes
ATP-dependent Clp protease proteolytic subunit CBB_104	ATP-dependent Clp protease proteolytic subunit GAP32_103	putative ATP-dependent Clp protease proteolytic subunit RaK2_00489	yes		
_		hypothetical protein ORF493			yes
major tail protein CBB_112	Ig domain-containing protein GAP32_111	putative structural protein ORF496	yes	yes	yes
putative membrane protein CBB_118	hypothetical protein GAP32_118	_	yes	yes	
PhoH family protein CBB_123	PhoH family protein GAP32_122	hypothetical protein ORF506			yes
_		hypothetical protein ORF526w			yes
_		putative tail fibre protein ORF527w			yes
_		tail fibre fragment ORF528w			yes

291

_		hypothetical protein ORF529			yes
_		hypothetical protein, ORF530w			yes
_		hypothetical protein, ORF531w			yes
_		tail spike protein ORF532			yes
_		hypothetical protein ORF533			yes
_		hypothetical protein ORF534			yes
hypothetical protein CBB_144	_	_	yes		
hypothetical protein CBB_148	putative membrane protein GAP32_144	_		yes	
structural protein CBB_149	hypothetical protein GAP32_145	_		yes	
putative membrane protein CBB_164	putative membrane protein GAP32_160	hypothetical protein ORF010			yes
hypothetical protein CBB_165	hypothetical protein GAP32_161	hypothetical protein ORF011			yes
conserved hypothetical protein CBB_169	hypothetical protein GAP32_165	hypothetical protein RaK2_00015	yes		
conserved hypothetical protein CBB_183	hypothetical protein GAP32_176	_	yes		
neck protein CBB_201	neck protein GAP32_195	hypothetical protein ORF038	yes		yes
tail sheath monomer	tail sheath monomer	hypothetical protein		yes	yes

CBB_204	GAP32_198	ORF041			
structural protein CBB_205	hypothetical protein GAP32_199	hypothetical protein ORF042	yes	yes	yes
structural protein CBB_206	hypothetical protein GAP32_200	hypothetical protein ORF043		yes	yes
conserved hypothetical protein CBB_207	hypothetical protein GAP32_201	RaK2_00044	yes		
head completion protein CBB_208	head completion protein GAP32_202	head completion protein RaK2_0004			
structural protein CBB_210	hypothetical protein GAP32_204	hypothetical protein ORF047	yes	yes	yes
structural protein CBB_211	hypothetical protein GAP32_205	hypothetical protein ORF048		yes	yes
structural protein CBB_212	hypothetical protein GAP32_206	hypothetical protein ORF049		yes	yes
conserved hypothetical protein CBB_213	hypothetical protein GAP32_207	hypothetical protein ORF050	yes		yes
conserved hypothetical protein CBB_214	hypothetical protein GAP32_208	hypothetical protein ORF051	yes		yes
structural protein CBB_224	hypothetical protein GAP32_218	hypothetical protein ORF060w	yes	yes	yes
structural protein CBB_225	hypothetical protein GAP32_219	_	yes	yes	
long tail fibre proximal subunit CBB_226	long tail fibre proximal subunit GAP32_220	_	yes	yes	
structural protein CBB_227	hypothetical protein GAP32_221	hypothetical protein ORF061w	yes	yes	yes

structural protein CBB_228	hypothetical protein GAP32_222	hypothetical protein RaK2_00062	yes	yes	
structural protein CBB_229	hypothetical protein GAP32_223	hypothetical protein ORF063		yes	yes
structural protein CBB_230	hypothetical protein GAP32_224	hypothetical protein RaK2_00064	yes	yes	
structural protein CBB_231	hypothetical protein GAP32_225	hypothetical protein ORF065	yes	yes	yes
tail sheath stabilizer and completion protein CBB_233	tail sheath stabilizer and completion protein GAP32_227	hypothetical protein RaK2_00067	yes	yes	
structural protein CBB_234	hypothetical protein GAP32_228	hypothetical protein ORF068w	yes	yes	yes
structural protein CBB_236	hypothetical protein GAP32_230	hypothetical protein ORF070w	yes	yes	yes
baseplate wedge CBB_237	baseplate wedge GAP32_232	hypothetical protein ORF071w	yes	yes	yes
base plate protein CBB_238	base plate protein GAP32_232	hypothetical protein ORF072w		yes	yes
baseplate hub subunit and tail lysozyme CBB_239	subunit and tail lysozyme GAP32_233	T4-like lysozyme ORF073w			
baseplate hub subunit and tail lysozyme CBB_240	baseplate hub subunit and tail lysozyme GAP32_234	putative virion protein and lysozyme murein RaK2_00078	yes	yes	
conserved hypothetical protein CBB_241	hypothetical protein GAP32_235	hypothetical protein ORF079w	yes		yes
baseplate wedge	hypothetical protein	hypothetical protein	yes		

protein CBB_242	GAP32_236	ORF080w			
conserved hypothetical protein CBB_243	hypothetical protein GAP32_237	hypothetical protein ORF081	yes	yes	yes
hypothetical protein CBB_247	_	_	yes		
hypothetical protein CBB_250	hypothetical protein GAP32_243	baseplate hub subunit ORF087			yes
structural protein CBB_251	hypothetical protein GAP32_244	hypothetical protein ORF088	yes	yes	
portal vertex protein of head CBB_252	portal vertex protein of head GAP32_245	portal vertex protein ORF089		yes	yes
conserved hypothetical protein CBB_254	hypothetical protein GAP32_247	hypothetical protein ORF091	yes		yes
prohead core scaffold and protease CBB_255	prohead core scaffold and protease GAP32_248	prohead core scaffolding protein and protease ORF092		yes	yes
conserved hypothetical protein CBB_256	hypothetical protein GAP32_249	hypothetical protein ORF093	yes		
precursor of major head subunit CBB_257	precursor of major head subunit GAP32_250	major capsid protein ORF094	yes	yes	yes
putative tail fibre protein CBB_260	putative tail fibre protein GAP32_254	hypothetical protein ORF098	yes	yes	yes
tail fibre assembly protein CBB_261	tail fibre assembly protein GAP32_255	_			
hypothetical protein CBB_269	hypothetical protein GAP32_263	hypothetical protein ORF105			yes
structural protein CBB_270	hypothetical protein GAP32_264	hypothetical protein ORF106w	yes	yes	yes

structural protein CBB_271	hypothetical protein GAP32_265	hypothetical protein ORF107w	yes	yes	yes
conserved hypothetical protein CBB_276	hypothetical protein GAP32_270	hypothetical protein ORF112	yes		
ssDNA binding protein CBB_277	ssDNA binding protein GAP32_271	hypothetical protein ORF113			yes
S-adenosyl-L-methionine-dependent methyltransferase CBB_286	hypothetical protein GAP32_281	hypothetical protein ORF123	yes		
structural protein CBB_289	hypothetical protein GAP32_284	hypothetical protein ORF126		yes	
conserved hypothetical protein CBB_292	hypothetical protein GAP32_287	hypothetical protein ORF129	yes		
conserved hypothetical protein CBB_294	hypothetical protein GAP32_289	hypothetical protein ORF131	yes		
structural protein CBB_298	hypothetical protein GAP32_293	hypothetical protein ORF135	yes	yes	yes
conserved hypothetical protein CBB_299	hypothetical protein GAP32_294	hypothetical protein ORF136	yes		
structural protein CBB_302	hypothetical protein GAP32_297	hypothetical protein ORF139		yes	yes
hypothetical protein CBB_308	hypothetical protein GAP32_303	hypothetical protein ORF143			yes
putative membrane protein CBB_310	hypothetical protein GAP32_305	hypothetical protein ORF145w			yes
structural protein CBB_316	hypothetical protein GAP32_311	hypothetical protein ORF151		yes	yes

conserved hypothetical protein CBB_320	hypothetical protein GAP32_315	hypothetical protein RaK2_00155	yes	yes	
structural protein CBB_321	hypothetical protein GAP32_316	_		yes	
structural protein CBB_322	hypothetical protein GAP32_317	hypothetical protein ORF156	yes	yes	yes
structural protein CBB_326	hypothetical protein GAP32_321	hypothetical protein ORF160		yes	yes
conserved hypothetical protein CBB_334	hypothetical protein GAP32_329	hypothetical protein ORF168	yes		
structural protein CBB_335	hypothetical protein GAP32_330	hypothetical protein ORF169		yes	
structural protein CBB_344	hypothetical protein GAP32_339	hypothetical protein ORF175		yes	yes
hypothetical protein CBB_370	_	hypothetical protein ORF182			yes
conserved hypothetical protein CBB_494	hypothetical protein GAP32_473	hypothetical protein ORF233	yes		yes
structural protein CBB_496	hypothetical protein GAP32_475	putative structural protein RaK2_00496	yes	yes	
structural protein CBB_534	hypothetical protein GAP32_515	_		yes	
hypothetical protein CBB_536	_	_	yes		
hypothetical protein CBB_537	_	_	yes		
structural protein CBB_548	hypothetical protein GAP32_530	_	yes	yes	

structural protein CBB_549	hypothetical protein GAP32_531	_	yes	yes	
laminin_G_3 family protein CBB_551	_	_	yes		
kelch-like protein CBB_552	_	_	yes		
conserved hypothetical protein CBB_553	hypothetical protein GAP32_002	_	yes		

Table S7.1. Genomic annotation of *Erwinia* phage Y3

ORF	Start (bp)	End (bp)	Length (aa)	Molecular weight (kDa)	Function
Y3_001	1	1497	498	56.32	replicative DNA helicase (DnaB)
Y3_002	1507	1767	86	10	hypothetical protein
Y3_003	1748	2551	267	30.28	hypothetical protein
Y3_004	2538	5369	943	105.71	putative terminase
Y3_005	5416	5628	70	7.65	hypothetical protein
Y3_006	5636	7576	646	72.53	putative portal protein
Y3_007	7621	7914	97	10.75	hypothetical protein
Y3_008	7911	8294	127	13.92	hypothetical protein
Y3_009	8338	8460	40	4.52	hypothetical protein
Y3_010	8457	8945	162	18.33	hypothetical protein
Y3_011	9002	9973	323	36.76	DNA polymerase I
Y3_012	9983	10498	171	19.38	hypothetical protein
Y3_013	10498	11082	194	21.81	ADP-ribose-binding protein
Y3_014	11098	11430	110	12.36	hypothetical protein
Y3_015	11437	12189	250	28.48	putative membrane protein
Y3_016	12201	12590	129	14.77	hypothetical protein
Y3_017	12641	13411	256	28.9	carboxy-S-adenosyl-L-methionine synthase
Y3_018	13437	14309	290	32.85	asparagine synthase
Y3_019	14376	14525	49	5.19	hypothetical protein
Y3_020	14594	15232	212	24.2	oxygenase
Y3_021	15324	15932	202	22.34	hypothetical protein
Y3_022	15943	16206	87	9.71	hypothetical protein
Y3_023	16215	16670	151	17.1	hypothetical protein
Y3_024	16685	16936	83	9.81	putative DNA primase
Y3_025	16902	17183	93	10.23	hypothetical protein
Y3_026	17186	17608	140	15.78	hypothetical protein
Y3_027	17605	17997	130	14.6	hypothetical protein
Y3_028	18007	18426	139	15.87	hypothetical protein
Y3_029	18608	19450	280	32.39	DNA adenine methylase
Y3_030	19460	20170	236	27.14	hypothetical protein
Y3_031	20178	20549	123	13.85	AntA/AntB antirepressor domain-containing protein
Y3_032	20551	21201	216	24.87	hypothetical protein
Y3_033	21203	21718	171	19.52	hypothetical protein
Y3_034	21687	22103	138	16.29	hypothetical protein
Y3_035	22088	22606	172	18.98	CMP deaminase
Y3_036	22713	23390	225	25.08	hypothetical protein
Y3_037	23399	23878	159	17.31	hypothetical protein
Y3_038	23878	24033	51	5.46	hypothetical protein
Y3_039	24082	24543	153	17.31	hypothetical protein
Y3_040	24521	25051	176	19.9	hypothetical protein

Y3_041	25032	25706	224	25.91	putative membrane protein
Y3_042	25699	26013	104	12.17	hypothetical protein
Y3_043	25997	26332	111	12.92	hypothetical protein
Y3_044	26329	26634	101	11.38	hypothetical protein
Y3_045	26646	27746	366	42.02	Thymidylate synthase
Y3_046	27743	28390	215	24.63	hypothetical protein
Y3_047	28390	29271	293	33.81	hypothetical protein
Y3_048	29321	29656	111	12.82	hypothetical protein
Y3_049	29703	30407	234	25.35	putative transcriptional repressor protein
Y3_050	30409	32538	709	79.84	C-5 cytosine methyltransferase
Y3_051	32684	33256	190	21.71	hypothetical protein
Y3_052	33256	33453	65	7.38	putative membrane protein
Y3_053	33696	33574	40	4.54	hypothetical protein
Y3_054	33753	34154	133	15.12	hypothetical protein
Y3_055	34215	34952	245	27.39	hypothetical protein
Y3_056	35021	35359	112	12.73	hypothetical protein
Y3_057	35310	35759	149	16.04	hypothetical protein
Y3_058	35780	36064	94	11.32	hypothetical protein
Y3_059	36201	36725	174	20.32	hypothetical protein
Y3_060	36722	37123	133	15	ASCH domain-containing protein
Y3_061	37120	37755	211	23.89	hypothetical protein
Y3_062	38306	38716	136	15.69	hypothetical protein
Y3_063	38732	38962	76	8.38	putative membrane protein
Y3_064	38959	39078	39	4.22	putative membrane protein
Y3_065	39075	39587	170	19.5	hypothetical protein
Y3_066	39577	40257	226	25.11	GTP pyrophosphokinase
Y3_067	40229	40579	116	13.46	hypothetical protein
Y3_068	40548	40934	128	14.88	hypothetical protein
Y3_069	40897	41514	205	23.36	hypothetical protein
Y3_070	41511	41849	112	13.24	hypothetical protein
Y3_071	42000	42338	112	12.87	hypothetical protein
Y3_072	42355	43011	218	24.83	hypothetical protein
Y3_073	43072	43767	231	24.67	hypothetical protein
Y3_074	43767	44588	273	30.34	hypothetical protein
Y3_075	45113	44649	154	17.24	putative lipoprotein
Y3_076	45403	45110	97	10.87	hypothetical protein
Y3_077	45721	45416	101	11.3	putative membrane protein
Y3_078	46017	45778	79	9.17	hypothetical protein
Y3_079	46650	46051	199	20.94	hypothetical protein
Y3_080	47516	46650	288	31.12	hypothetical protein
Y3_081	48062	47526	178	18.7	hypothetical protein
Y3_082	49054	48062	330	35.24	hypothetical protein
Y3_083	49701	49180	173	18.03	hypothetical protein
Y3_084	50792	49698	364	40.31	hypothetical protein

Y3_085	51383	50802	193	20.72	hypothetical protein
Y3_086	52186	51383	267	28.61	hypothetical protein
Y3_087	52754	52296	152	16.19	hypothetical protein
Y3_088	53839	52754	361	38.55	hypothetical protein
Y3_089	54785	53841	314	33.35	hypothetical protein
Y3_090	55282	54785	165	17.4	putative membrane protein
Y3_091	56367	55360	335	36.53	hypothetical protein
Y3_092	57401	56379	340	36.01	hypothetical protein
Y3_093	58471	57413	352	38.25	hypothetical protein
Y3_094	59133	58552	193	20.28	hypothetical protein
Y3_095	59365	59144	73	8.06	hypothetical protein
Y3_096	60307	59378	309	33.34	hypothetical protein
Y3_097	60917	60321	198	21.67	hypothetical protein
Y3_098	62610	60997	537	54.92	putative virion protein
Y3_099	63377	62634	247	28.88	hypothetical protein
Y3_100	63880	63374	168	18.49	putative membrane protein
Y3_101	65055	64234	273	30.57	hypothetical protein
Y3_102	65957	65130	275	29.95	hypothetical protein
Y3_103	66417	65941	158	17.11	hypothetical protein
Y3_104	67238	66420	272	28.02	putative tail fibre protein
Y3_105	67452	67249	67	7.45	hypothetical protein
Y3_106	70346	67470	958	106.02	putative tail fiber hinge protein
Y3_107	72409	70391	672	74.33	hypothetical protein
Y3_108	72912	72421	163	16.07	putative tail fiber protein
Y3_109	73554	72922	210	20.26	phage tail fiber protein
Y3_110	74991	73555	478	50.11	putative phage tail protein, containing intein domain
Y3_111	77315	74988	775	83.33	putative phage tail-collar fibre protein
Y3_112	81929	77403	1508	166.89	hypothetical protein
Y3_113	83392	81926	488	54.34	putative baseplate wedget subunit protein
Y3_114	83811	83392	139	16.01	baseplate protein
Y3_115	84101	83811	96	9.81	putative baseplate spike protein
Y3_116	87129	85369	586	64.59	hypothetical protein
Y3_117	87572	87138	144	16.54	hypothetical protein
Y3_118	88183	87569	204	23.02	dTMP kinase
Y3_119	88686	88195	163	19.11	DNA repair protein MmcB-like
Y3_120	89134	88676	152	17.34	NUDIX hydrolase
Y3_121	89643	89131	170	18.45	hypothetical protein
Y3_122	89970	89686	94	10.38	hypothetical protein
Y3_123	90668	89973	231	26.34	hypothetical protein
Y3_124	91466	90678	262	28.65	putative baseplate protein
Y3_125	93358	91463	631	69.24	hypothetical protein
Y3_126	93795	93361	144	16.13	hypothetical protein

Y3_127	94085	93795	96	10.49	hypothetical protein
Y3_128	94876	94064	270	28.14	hypothetical protein
Y3_129	97565	94887	892	94.13	putative chitosanase
Y3_130	98380	97565	271	31.93	hypothetical protein
Y3_131	99099	98416	227	25.85	hypothetical protein
Y3_132	99621	99109	170	18.78	putative tail tube protein
Y3_133	100297	99626	223	25.23	hypothetical protein
Y3_134	100853	100344	169	18.94	putative tail tube protein
Y3_135	102549	100864	561	61.62	phage tail sheath protein
Y3_136	102950	102606	114	12.7	hypothetical protein
Y3_137	103624	102950	224	24.93	hypothetical protein
Y3_138	104104	103691	137	14.86	hypothetical protein
Y3_139	105260	104160	366	39.84	putative major capsid protein
Y3_140	106028	105321	235	24.79	structural protein
Y3_141	108157	106082	691	73.94	hypothetical protein
Y3_142	109347	108235	370	41.12	hypothetical protein
Y3_143	110128	109349	259	29.14	putative prohead core protein protease
Y3_144	110555	110139	138	15.62	hypothetical protein
Y3_145	111337	110558	259	28.14	hypothetical protein
Y3_146	112475	111303	390	44.39	glycosyl transferase family 2
Y3_147	113451	112534	305	35.5	hypothetical protein
Y3_148	115400	113454	648	71.07	DNA ligase
Y3_149	117322	115442	626	70.97	hypothetical protein
Y3_150	121573	117389	1394	153.9	hypothetical protein
Y3_151	122802	121633	389	41.77	hypothetical protein
Y3_152	123800	122823	325	34.35	hypothetical protein
Y3_153	124588	123803	261	29.15	hypothetical protein
Y3_154	128607	124600	1335	141.4	putative major tail protein
Y3_155	129236	128604	210	23.16	hypothetical protein
Y3_156	129885	129247	212	24.38	hypothetical protein
Y3_157	130628	129933	231	25.84	hypothetical protein
Y3_158	132357	130678	559	62.16	tail sheath protein
Y3_159	133488	132445	347	38.49	hypothetical protein
Y3_160	134333	133485	282	32.3	hypothetical protein
Y3_161	134542	134333	69	7.75	hypothetical protein
Y3_162	135167	134553	204	22.94	hypothetical protein
Y3_163	135626	135255	123	14.47	hypothetical protein
Y3_164	135922	135632	96	10.35	hypothetical protein
Y3_165	136387	135962	141	15.63	hypothetical protein
Y3_166	136910	136395	171	19.51	hypothetical protein
Y3_167	137194	136910	94	10.54	hypothetical protein
Y3_168	138792	137194	532	60.92	DNA repair helicase
Y3_169	139340	138795	181	21.37	hypothetical protein
Y3_170	139774	139340	144	16.21	hypothetical protein

Y3_171	140949	139771	392	43.9	hypothetical protein
Y3_172	141145	140927	72	7.97	hypothetical protein
Y3_173	144375	141211	1054	121.34	DNA polymerase A
Y3_174	145684	144458	408	43.8	hypothetical protein
Y3_175	146242	145694	182	20.63	putative lysozyme
Y3_176	152139	146284	1951	217.11	putative ATP-dependent DNA helicase
Y3_177	152636	152139	165	18.37	hypothetical protein
Y3_178	153429	152647	260	29.77	hypothetical protein
Y3_179	153791	153426	121	13.54	HNH family endonuclease
Y3_180	154658	153828	276	30.38	hypothetical protein
Y3_181	155525	154719	268	28.94	hypothetical protein
Y3_182	156738	155578	386	42.73	putative head to tail joining protein
Y3_183	157846	156740	368	41.61	hypothetical protein
Y3_184	159129	157846	427	46.7	hypothetical protein
Y3_185	159998	159321	225	23.88	hypothetical protein
Y3_186	161195	160119	358	40.94	phage recombination related endonuclease
Y3_187	162263	161268	331	34.79	hypothetical protein
Y3_188	162837	162310	175	20.3	putative ssDNA binding protein
Y3_189	163698	162856	280	29.91	hypothetical protein
Y3_190	164702	163767	311	34.69	hypothetical protein
Y3_191	165107	164742	121	13.1	DUF2778 domain-containing protein
Y3_192	165312	165172	46	5.64	hypothetical protein
Y3_193	165824	165312	170	19.92	hypothetical protein
Y3_194	166879	165824	351	36.98	hypothetical protein
Y3_195	167490	166888	200	21.83	hypothetical protein
Y3_196	167822	167490	110	11.98	hypothetical protein
Y3_197	168394	167825	189	21.17	putative glycosyl hydrolase
Y3_198	168804	168469	111	12.96	putative membrane protein
Y3_199	169517	168804	237	26.76	hypothetical protein
Y3_200	169756	169508	82	9.18	hypothetical protein
Y3_201	170067	169759	102	12.03	hypothetical protein
Y3_202	172400	170067	777	89.06	putative ATPase
Y3_203	172600	172400	66	7.75	hypothetical protein
Y3_204	172809	172600	69	8.11	hypothetical protein
Y3_205	173344	172907	145	16.54	hypothetical protein
Y3_206	174127	173357	256	27.83	hypothetical protein
Y3_207	174084	174221	45	5.03	hypothetical protein
Y3_208	174802	174269	177	19.44	hypothetical protein
Y3_209	176195	174903	430	47.66	putative DNA polymerase III
Y3_210	176547	176350	65	7.61	hypothetical protein
Y3_211	177605	177018	195	22.71	hypothetical protein

Y3_212	177937	177605	110	12.85	hypothetical protein
Y3_213	178513	177941	190	21.57	hypothetical protein
Y3_214	179486	178515	323	36.32	hypothetical protein
Y3_215	179971	179495	158	17.25	hypothetical protein
Y3_216	180207	179983	74	8.76	hypothetical protein
Y3_217	180709	180212	165	19.78	hypothetical protein
Y3_218	181279	180755	174	20.47	hypothetical protein
Y3_219	182899	181283	538	62.27	hypothetical protein
Y3_220	183944	183288	218	24.15	hypothetical protein
Y3_221	185064	183955	369	41.36	hypothetical protein
Y3_222	185163	185711	182	20.97	putative Holliday junction resolvase
Y3_223	185721	186110	129	14.32	hypothetical protein
Y3_224	186145	186816	223	25.93	hypothetical protein
Y3_225	186809	188575	588	60.82	putative surface adhesion protein
Y3_226	188632	189291	219	24.47	hypothetical protein
Y3_227	190437	189346	363	39.78	phage-associated DNA primase
Y3_228	190503	190844	113	13.51	hypothetical protein
Y3_229	190810	191895	361	41.35	putative exonuclease
Y3_230	191934	192302	122	14.24	hypothetical protein
Y3_231	192311	192817	168	17.48	hypothetical protein
Y3_232	192801	193331	176	20.47	hypothetical protein
Y3_233	193345	193890	181	20.28	hypothetical protein
Y3_234	193894	194118	74	8.2	hypothetical protein
Y3_235	194115	194450	111	12.62	hypothetical protein
Y3_236	194440	194658	72	7.88	hypothetical protein
Y3_237	194664	195347	227	26.21	hypothetical protein
Y3_238	195347	195871	174	19.27	hypothetical protein
Y3_239	195864	196331	155	17.53	putative cyclic phosphodiesterase
Y3_240	196387	196704	105	12	hypothetical protein
Y3_241	196750	197787	345	40.23	hypothetical protein
Y3_242	197832	199094	420	48.28	putative single stranded DNA binding protein
Y3_243	199160	201652	830	92.86	RecA protein
Y3_244	201708	202085	125	13.56	hypothetical protein
Y3_245	202160	202879	239	26.69	hypothetical protein
Y3_246	202986	204005	339	39.28	hypothetical protein
Y3_247	204051	204812	253	29.09	hypothetical protein
Y3_248	204984	205106	40	4.74	hypothetical protein
Y3_249	205256	206188	310	33.92	hypothetical protein
Y3_250	206329	206889	186	20.6	hypothetical protein
Y3_251	207116	207754	212	23.13	hypothetical protein
Y3_252	207860	208111	83	9.23	hypothetical protein
Y3_253	208179	208835	218	24.07	hypothetical protein
Y3_254	208887	209414	175	20.57	hypothetical protein

Y3_255	209471	209740	89	10.05	hypothetical protein
Y3_256	209785	210126	113	12.85	hypothetical protein
Y3_257	210203	210562	119	13.81	hypothetical protein
Y3_258	210563	210832	89	9.74	hypothetical protein
Y3_259	210829	211194	121	14.49	hypothetical protein
Y3_260	211277	211762	161	19.28	hypothetical protein
Y3_261	211750	212124	124	14.45	hypothetical protein
Y3_262	212128	212442	104	12.22	hypothetical protein
Y3_263	212534	212977	147	16.79	hypothetical protein
Y3_264	213043	213369	108	12.05	hypothetical protein
Y3_265	213436	213792	118	13.07	hypothetical protein
Y3_266	213857	214276	139	16.31	hypothetical protein
Y3_267	214337	214846	169	19.66	hypothetical protein
Y3_268	214981	215181	66	7.72	hypothetical protein
Y3_269	215184	215624	146	17.05	hypothetical protein
Y3_270	215692	216732	346	39.06	hypothetical protein
Y3_271	216707	216928	73	8.29	putative membrane protein
Y3_272	216921	217466	181	20.65	RNA 2'-phosphotransferase
Y3_273	217459	217857	132	15.52	hypothetical protein
Y3_274	217933	219276	447	51.67	GTP 3',8-cyclase MoaA
Y3_275	219288	219842	184	20.99	hypothetical protein
Y3_276	219842	220174	110	12.99	putative membrane protein
Y3_277	220174	220500	108	12.13	hypothetical protein
Y3_278	220504	220980	158	18.13	hypothetical protein
Y3_279	220985	221242	85	9.3	C4-type zinc finger domain-containing protein
Y3_280	221245	221610	121	14.59	hypothetical protein
Y3_281	221585	221944	119	13.77	hypothetical protein
Y3_282	221952	222287	111	13.04	hypothetical protein
Y3_283	222339	222704	121	13.57	hypothetical protein
Y3_284	222714	223430	238	26.94	hypothetical protein
Y3_285	223405	223602	65	7.27	putative membrane protein
Y3_286	223599	224159	186	21.09	hypothetical protein
Y3_287	224216	225190	324	37.76	UV damage endonuclease UvsE
Y3_288	225282	225575	97	11.05	hypothetical protein
Y3_289	225577	225840	87	10.09	hypothetical protein
Y3_290	226002	226295	97	11.12	hypothetical protein
Y3_291	226273	226521	82	9.34	hypothetical protein
Y3_292	226592	227101	169	18.45	hypothetical protein
Y3_293	227101	227403	100	11.7	hypothetical protein
Y3_294	227390	228460	356	40.29	hypothetical protein
Y3_295	228572	228874	100	11.58	hypothetical protein
Y3_296	228883	229443	186	20.77	hypothetical protein
Y3_297	229458	229904	148	16.5	hypothetical protein
Y3_298	229901	230083	60	6.6	putative lipoprotein

Y3_299	230097	230360	87	9.68	hypothetical protein
Y3_300	230360	230896	178	19.53	putative dUTPase
Y3_301	231516	232193	225	25.11	lytic transglycosylase
Y3_302	232208	232882	224	25.25	hypothetical protein
Y3_303	232954	233895	313	36.76	hypothetical protein
Y3_304	233970	234662	230	25.91	putative membrane protein
Y3_305	234665	234883	72	8.57	putative membrane protein
Y3_306	234888	235220	110	12.6	hypothetical protein
Y3_307	235207	235749	180	20.81	hypothetical protein
Y3_308	235787	237085	432	49.85	hypothetical protein
Y3_309	237135	237566	143	15.52	hypothetical protein
Y3_310	237576	237875	99	11.52	hypothetical protein
Y3_311	237993	238457	154	17.4	hypothetical protein
Y3_312	238458	239090	210	24.86	hypothetical protein
Y3_313	239102	239293	63	7.32	hypothetical protein
Y3_314	239418	239756	112	12.48	hypothetical protein
Y3_315	240074	241612	512	55.71	hypothetical protein
Y3_316	241698	242933	411	45.3	hypothetical protein
Y3_317	242990	243319	109	12.79	hypothetical protein
Y3_318	243316	243795	159	17.38	hypothetical protein
Y3_319	243785	245065	426	49.51	hypothetical protein
Y3_320	245088	248105	1005	109.22	hypothetical protein
Y3_321	248161	249009	282	32.09	hypothetical protein
Y3_322	249105	249665	186	19.33	hypothetical protein
Y3_323	249818	250597	259	28.56	hypothetical protein
Y3_324	250812	251375	187	19.14	hypothetical protein
Y3_325	251476	251973	165	18.01	hypothetical protein
Y3_326	251954	252718	254	28.7	hypothetical protein
Y3_327	252846	254885	679	75.4	DNA topoisomerase 4 subunit B
Y3_328	254887	255189	100	10.83	hypothetical protein
Y3_329	255189	256841	550	60.07	DNA topoisomerase 4 subunit A
Y3_330	257187	258455	422	47.57	hypothetical protein
Y3_331	258935	259780	281	31.04	hypothetical protein
Y3_332	259838	260968	376	43.38	hypothetical protein
Y3_333	260986	261357	123	14.14	hypothetical protein
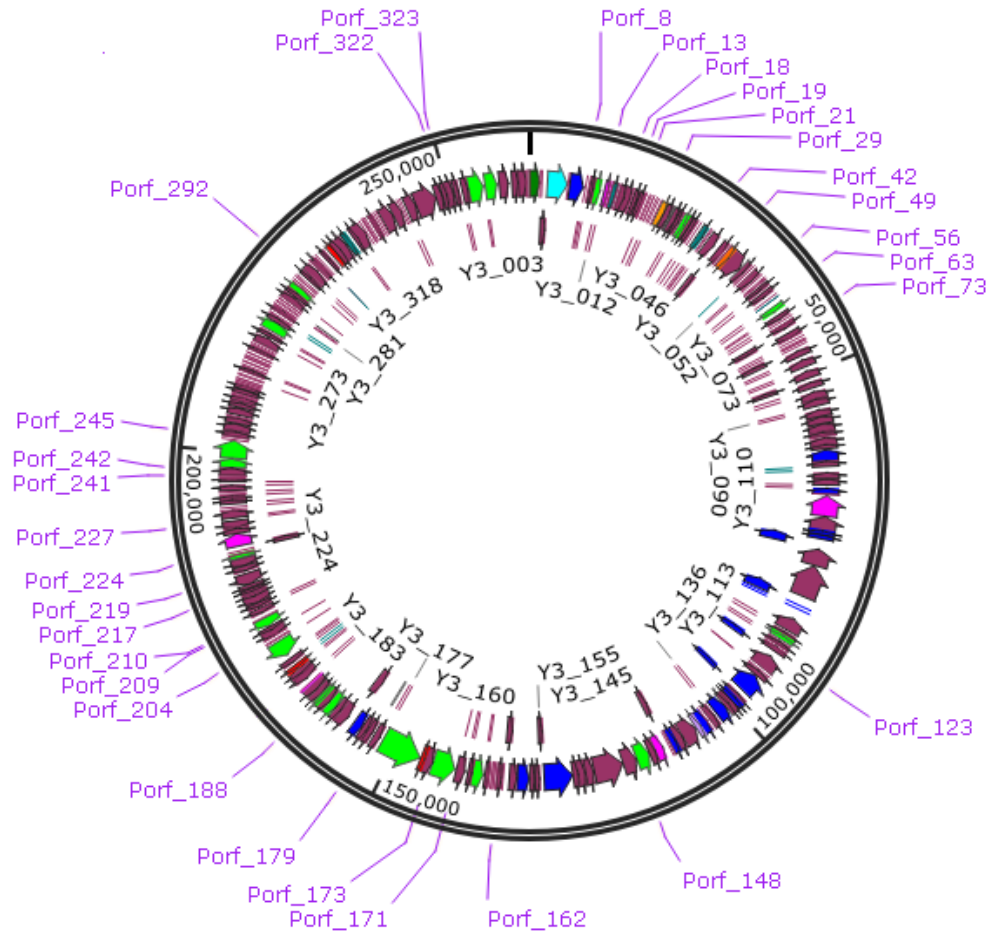
Figure S7.1. Genome map of *Erwinia* phage Y3 showing locations of predicted promotors with the consensus sequence of CTGTAAATA in the genome. Arrows indicate locations of ORFs with colours indicating the following. Purple: hypothetical protein, green: nucleotide metabolism and DNA replication-related, blue: virion structural related, pink: miscellaneous.

Figure S7.2. Predicted promotor in *Erwinia amylovora* phage vB_PcaM_ Y3 shared with *Pseudomonas* phage Lu11

Promoter	Location	Sequence
Porf_008	7816-7828	AATTCTGTAAATA
Porf_014	10448-10460	AATACTGTAAATA
Porf_018	13410-13422	AAATCTGTAAACA
Porf_019	14296-14308	GAATCTGTAAATA
Porf_021	15281-15293	AAGTCTGTAAATA
Porf_029	18577-18589	AAATCTGTAAATA
Porf_042	25623-25635	TATTCTGTAAATG
Porf_049	29646-29678	ATAACTGTAAATA
Porf_056	34950-34962	TAATCTGTAAATA
Porf_063	38678-38690	GAATCTGTAAATA
Porf_073	43002-43014	AAATCTGTAAATA
Porf_123	90713-90701	GTTACTGTAAATG
Porf_148	115443-115431	AATGCTGTAAATA
Porf_162	135206-135194	AAAACTGTAAATA
Porf_171	140998-140986	AAGGCTGTAAATG
Porf_173	144417-144405	AAATCTGTAAATA
Porf_179	153829-153817	AGAACTGTAAATA
Porf_188	162856-162844	GTATCTGTAAATA
Porf_204	172861-172849	AAAACTGTAAATA
Porf_209	176243-176231	AATACTGTAAATA
Porf_210	176601-176589	TTAACTGTAAATA

Porf_217	180756-180744	GATTCTGTAAACT
Porf_219	182939-182927	ATATCTGTAAATA
Porf_224	186101-186113	GGATCTGTAAATA
Porf_228	190479-190491	TTATCTGTAAATA
Porf_241	14296-14308	GAATCTGTAAATA
Porf_242	197786-197798	AAAACTGTAAACA
Porf_245	202076-202088	TAATCTGTAAATA
Porf_291	226504-226516	GAGTCTGTAAATA
Porf_293	222295-222307	GATTCTGTAAATC
Porf_322	249032-249044	TAATCTGTAAACA
Porf_323	259759-249771	GAATCTGTAAATC

Table S7.3. High ΔG rho-independent terminators predicted in the genome *Erwinia amylovora* phage vB_EamM_Y3 identified using ARNold and QuikFold.

Terminator	Coordinates	Sequence	ΔG kcal/mol
Torf019	14,545 - 14577	GGGGCTGCTCATTGCAGTCCCTTTTTACTTTCC	-15.8
Torf028	18433 - 18465	AGGGTGCCTTAGGGTGCCCTTTTCTGATCGTAT	-12.7
Torf038	24040 - 24074	TGGGGTGGCTGTGTGCTGCCCCATTTTTGTTTGGA	-16.3
Torf055	34975 - 35007	TGGGTGGCCTTGTGCTGCCCATTTTTGTTTCTA	-11.8
Torf061	37762 - 37793	TGGGTGGCTTCGGCTGCCCATTTTGCATTTGG	-15.8
Torf061a	38232 - 38265	GCTGGGAAGTATTTCTCAGCaaTTTTAGTGGCAG	-10.4
Torf075	44598 - 44638 (complement)	GGGCGACTCTCTTGATTGAGGGCCGCCCTTTTTCGATCTTG	-17.9
Torf074	44610 - 44652	AGGGCGGCCCTCAATCAAGAGAGTCGCCCTTTTTATTTGCTAC	-18.4
Torf078	45738 - 45770 (compl ement)	AGGGTGGCTTCGGCTGCCCTTTCTTTTTGTCCA	-15.3
Torf083	49134 - 49173 (complement)	ACAGTGGTGCTCATTCAGGCCACTGTcTTTAATTTAAATT	-10.4
Torf094	58504 - 58545 (complement)	GGGGCATGTATCTTCTGAGGTACTGCCCCTTTCGCATTTCCA	-18.3
Torf098	60928 - 60960 (complement)	GGGTGGGCCTGTGCTCACCCTTTTTACTATTCA	-14.3
Torf107	70356 - 70385 (complement)	AGGGGCTTCGGCCCCTcTTCTCGATCTTAA	-14.2
Torf112	77354 - 77388 (complement)	GGAGCTAGATGTTCCTGGCTCCcTTTTGACTTTCA	-13.3
Torf116	84142 - 84176 (complement)	CGGGTGGCCATTAGGCTGCCCGTTTTTGTTTCTCG	-16.6
Torf134	100306 - 100337 (complement)	GGGCCAGATGTTTTGGCCCTTTTTCGTTTCTG	-12.3
Torf138	103634 - 103667 (complement)	AGGGTGGCTTCGGCTGCCCTtTTTTGTATCTGGA	-15.3
Torf150	117370 - 117399 (complement)	GCGGGGGTTAAATCCCGCTTTTCCTTTTCG	-12.8
Torf158	130637 - 130667 (complement)	GGAGGGCTTCGGTCCTCCTTTTTCGTATTGT	-15.5
Torf159	132404 - 132436 (complement)	AGGGTGCCTGCGGGTGCCCTTTTTACGTTTTGG	-12.7
Torf163	135211 - 135245 (complement)	GAGCTGGGCTTCGGTCCGGCTCTTTTACCTTCTGC	-19.1
Torf174	144417 - 144448 (complement)	TGGGTGGCTTCGGCTGCCCATTTTCATTTCTA	-15.8

Torf181	154674 - 154705 (complement)	GGGAGCCTCGCGGGGGTTCCCTTTTTCGTTTCT	-14.8
Torf185	159253 - 159285 (complement)	AGGGTGGCTTCGGCTGCCCTTTTTGTTTTTCTA	-15.3
Torf187	161228 - 161261 (complement)	TGGGCGGCCTTCGGGTTGCCCATTTTTCTGTAAA	-17.7
Torf192	165128 - 165166 (complement)	GGGCTTGCATCGAGAGGTGTAGGCCCTTTTGTTTTTTAT	-21
Torf198	168431 - 168463 (complement)	AGGGTGGCTTCGGCTGCCCTTTTTGTTTCTGAC	-15.3
Torf205	172871 - 172902 (complement)	GGGCTGCCTAATGGTGGCCCTTCTTTCTATCT	-15.3
Torf209	174860 - 174894 (complement)	TGGGTGGCCAATGAGCTGCCCATTTTTATTTCAGA	-12.2
Torf227	189302 - 189340 (complement)	TGGGAGACTCCTTTCGGGGCCTCCCATTTTTTTGTCTAA	-15.6
Torf226	189315 - 189352	TGGGAGGCCCCGAAAGGAGTCTCCCATTTTACTATGCC	-19.9
Torf239	196342 - 196375	GCGGTGGCCATTGTGCTGCCGCTTTTGTCATTTG	-13.5
Torf242	199108 - 199141	TGGGGTGGCATAAGCTGCCCCATTCTTTTCTCTA	-15.8
Torf243	201667 - 201697	GGCGGGCTTCGGCTCGCCcTTTCTTTAACAG	-16.1
Torf245	202919 - 202952	GGGAGATCATGATGGTCTCCCaTTTTTTTTATTT	-14.3
Torf248	205154 - 205194	GGGCTGTCAGAAATCAATCTGATGGCCCTTTCTCTTTTATA	-18.2
Torf250	206912 - 206943	GGGGTGCCTTCGGGTGCCCCTTTTGCGTCTCT	-18.5
Torf252	208127 - 208156	GAGACGTTACGGCGTCTCTTTTCTTTTATA	-10.2
Torf254	209432 - 209461	GCCTTGCGCAAGCGGGGCTTTTCTGATCTG	-13.5
Torf256	210134 - 210165	GGGCTGCCTAACGGTGGCCCTTTTTCGTTTCT	-16.7
Torf259	211224 - 211252	GGGGCTCGCTGAGTCCCTTTCATTTAGGA	-10.6
Torf266	214284 - 214322	GGGATGCTCACAAAGGTGGGTGTCCCcTTTTACTTTCTG	-17
Torf269	215643 - 215678	GCACTGGTCTGAAAAGGTCAGTGCTTTTCCTACATT	-16.6
Torf302	232894 -	CGGGAGACCTTAACTGGTTTCCCGTTTTTCATTCTT	-16

	232929			
Torf310	237944 - 237973	GCGTAGCAGATGCTGCGCTTTTTTACTTTA		-12.2

Table S7.4. List of jumbo *Erwinia amylovora* phage with genome sequences deposited in Genbank to date.

Phage	Genome accession no.	Terminase accession no.	Portal vertex accession no	Reference
vB_EamM_Deimos-Minion	KU886225	ANH52287.1	ANH52104.1	(Esplin et al., 2017)
vB_EamM_RAY	KU886224	ANH51964.1	ANH51787.1	(Esplin et al., 2017)
vB_EamM_Special G	KU886222	ANJ64995.1	ANJ64816.1	(Esplin et al., 2017)
Ea35-70	KF806589	AHI60333.1	AHI60156.1	(Yagubi et al., 2014)
vB_EamM_Simmy50	KU886223	ANH51648.1	ANH51468.1	(Esplin et al., 2017)
vB_EamM_Caitlin	KX397365	ANZ48565.1	YP_009292163.1	(Esplin et al., 2017)
vB_EamM_ChrisDB	KX397366	ANZ48843.1	YP_009292784.1	(Esplin et al., 2017)
vB_EamM_EarlPhillipIV	KX397367	ANZ49082.1	YP_009278390.1	(Esplin et al., 2017)
vB_EamM_Huxley	KX397368	ANZ49353.1	YP_009293062.1	(Esplin et al., 2017)
vB_EamM_Kwan	KX397369	ANZ49636.1	YP_009278705.1	(Esplin et al., 2017)
vB_EamM_Machina	KX397370	ANZ49910.1	ANZ49731.1	(Esplin et al., 2017)
vB_EamM_Parshik	KX397371	ANZ50181.1	ANZ50004.1	(Esplin et al., 2017)
vB_EamM_Phobos	KX397372	ANZ50429.1	ANZ50268.1	(Esplin et al., 2017)
vB_EamM_Stratton	KX397373	ANZ50705.1	ANZ50524.1	(Esplin et al., 2017)
vB_EamM_RisingSun	MF459646	ASU03430.1	ASU03582.1	(Esplin et al., 2017)
vB_EamM_Joad	MF459647	ASU03673.1	ASU03827.1	(Esplin et al., 2017)
vB_EamM_Asesino	NC_031107.1	YP_009290899.1	YP_009290715.1	unpublished
vB_EamM_Yoloswag	KY448244	AQT28491.1	AQT28493.1	(Esplin et al., 2017)
vB_EamM_Y3	KY984068	ARW58644	ARW58646	This work

References

Abbasifar, R., Griffiths, M.W., Sabour, P.M., Ackermann, H.W., Vandersteegen, K., Lavigne, R., Noben, J.P., Alanis Villa, A., Abbasifar, A., Nash, J.H.E., Kropinski, A.M., 2014. Supersize me: *Cronobacter sakazakii* phage GAP32. Virology 460–461, 138–146. doi:10.1016/j.virol.2014.05.003

Abbasifar, R., Kropinski, A.M., Sabour, P.M., Ackermann, H.-W., Alanis Villa, A., Abbasifar, A., Griffiths, M.W., 2013. The genome of *Cronobacter sakazakii* bacteriophage vB_CsaP_GAP227 suggests a new genus within the *Autographivirinae*. Genome Announc. 1. doi:10.1128/genomeA.00122-12

Abedon, S.T., Lejeune, J.T., 2005. Why bacteriophage encode exotoxins and other virulence factors. Evol. Bioinform. Online 1, 97–110.

Abedon, S.T., Thomas-Abedon, C., Thomas, A., Mazure, H., 2011. Bacteriophage prehistory: Is or is not Hankin, 1896, a phage reference? Bacteriophage 1, 174–178. doi:10.4161/bact.1.3.16591

Ackermann, H.-W., 2007. 5500 Phages examined in the electron microscope. Arch. Virol. 152, 227–43. doi:10.1007/s00705-006-0849-1

Ackermann, H.-W., Prangishvili, D., 2012. Prokaryote viruses studied by electron microscopy. Arch. Virol. 157, 1843–1849. doi:10.1007/s00705-012-1383-y

Ackermann, H.W., 2001. Frequency of morphological phage descriptions in the year 2000. Brief review. Arch. Virol. 146, 843–57.

Ackermann, H.W., Auclair, P., Basavarajappa, S., Konjin, H.P., Savanurmath, C., 1994. Bacteriophages from Bombyx mori. Arch. Virol. 137, 185–90.

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

Addy, H.S., Askora, A., Kawasaki, T., Fujie, M., Yamada, T., 2012. Loss of virulence of the phytopathogen *Ralstonia solanacearum* through infection by φRSM filamentous phages. Phytopathology 102, 469–77. doi:10.1094/PHYTO-11-11-0319-R

Adriaenssens, E.M., Brister, J.R., 2017. How to Name and Classify Your Phage: An Informal Guide. Viruses 9. doi:10.3390/v9040070

Adriaenssens, E.M., C Clokie, M.R., Sullivan, M.B., Gillis, A., Kuhn, J.H., Kropinski, A.M., 2017. Taxonomy of prokaryotic viruses: 2016 update from the ICTV bacterial and archaeal viruses subcommittee. Arch. Virol. 162, 1153–1157. doi:10.1007/s00705-016-3173-4

Adriaenssens, E.M., Mattheus, W., Cornelissen, A., Shaburova, O., Krylov, V.N., Kropinski, A.M., Lavigne, R., 2012a. Complete genome sequence of the giant *Pseudomonas* phage Lu11. J. Virol. 86, 6369–70. doi:10.1128/JVI.00641-12

Adriaenssens, E.M., Van Vaerenbergh, J., Vandenheuvel, D., Dunon, V., Ceyssens, P.-J., De Proft, M., Kropinski, A.M., Noben, J.-P., Maes, M., Lavigne, R., 2012b. T4-Related Bacteriophage LIMEstone Isolates for the Control of Soft Rot on Potato Caused by 'Dickeya solani.' PLoS One 7, e33227. doi:10.1371/journal.pone.0033227

Ågren, J., Sundström, A., Håfström, T., Segerman, B., 2012. Gegenees: fragmented alignment of multiple genomes for determining phylogenomic distances and genetic signatures unique for specified target groups. PLoS One 7, e39107. doi:10.1371/journal.pone.0039107

Ahern, S.J., Das, M., Bhowmick, T.S., Young, R., Gonzalez, C.F., 2014. Characterization of novel

virulent broad-host-range phages of *Xylella fastidiosa* and *Xanthomonas*. J. Bacteriol. 196, 459–71. doi:10.1128/JB.01080-13

Ahmad, A.A., Ogawa, M., Kawasaki, T., Fujie, M., Yamada, T., 2014. Characterization of bacteriophages Cp1 and Cp2, the strain-typing agents for *Xanthomonas axonopodis* pv. *citri*. Appl. Environ. Microbiol. 80, 77–85. doi:10.1128/AEM.02310-13

Ahmad, M.H., Morgan, V., 1994. Characterization of a cowpea (*Vigna unguiculata*) rhizobiophage and its effect on cowpea nodulation and growth. Biol. Fertil. Soils 18, 297–301. doi:10.1007/BF00570632

Alexander, S.A., Kim, S.H., Waldenmaier, C.M., 1999. First Report of Copper-Tolerant *Pseudomonas syringae* pv. *tomato* in Virginia. Plant Dis. 83, 964–964. doi:10.1094/PDIS.1999.83.10.964C

Alič, Š., Naglič, T., Tušek-Žnidarič, M., Ravnikar, M., Rački, N., Peterka, M., Dreo, T., 2017. Newly isolated bacteriophages from the *Podoviridae*, *Siphoviridae*, and *Myoviridae* families have variable effects on putative novel *Dickeya* spp. Front. Microbiol. 8, 1870. doi:10.3389/fmicb.2017.01870

Andres, D., Baxa, U., Hanke, C., Seckler, R., Barbirz, S., 2010. Carbohydrate binding of *Salmonella* phage P22 tailspike protein and its role during host cell infection. Biochem. Soc. Trans. 38, 1386–1389. doi:10.1042/BST0381386

Anonymous, 2013. Commission Implementing Regulation (EU)No. 485/2013 of 24 May 2013 amending Implementing Regulation (EU) No. 540/2011, as regards the conditions of approval of the active substances clothianidin, thiamethoxam and imidacloprid, and prohibiting the use and. Off. J. Eur. Union L 129, 12–26.

Anonymous, 2000. Council Directive 2000/29/EC of 8 May 2000 on protective measure against the introduction into the Community of organisms harmful to plants or plant products and against their spread within the Commnity L169:1-12.

APS Biocontrol Ltd, n.d. APS Biocontrol [WWW Document]. URL http://apsbiocontrol.com/ (accessed 3.24.16).

Ayers, A.R., Ayers, S.B., Goodman, R.N., 1979. Extracellular Polysaccharide of *Erwinia amylovora*: a Correlation with Virulence. Appl. Environ. Microbiol. 38, 659–66.

Bae, J.Y., Wu, J., Lee, H.J., Jo, E.J., Murugaiyan, S., Chung, E., Lee, S.-W., 2012. Biocontrol potential of a lytic bacteriophage PE204 against bacterial wilt of tomato. J. Microbiol. Biotechnol. 22, 1613–20.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., Noble, W.S., 2009. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 37, W202-8. doi:10.1093/nar/gkp335

Balogh, B., 2006. Characterization and use of bacteriophages assciated with citrus bacterial pathogens for disease control. Unversity of Florida.

Balogh, B., Canteros, B.I., Stall, R.E., Jones, J.B., 2008. Control of citrus canker and citrus bacterial spot with bacteriophages. Plant Dis. 92, 1048–1052. doi:10.1094/PDIS-92-7-1048

Balogh, B., Jones, J.B., Momol, M.T., Olson, S.M., Obradovic, A., King, P., Jackson, L.E., 2003. Improved efficacy of newly formulated bacteriophages for management of bacterial spot on tomato. Plant Dis. 87, 949–954. doi:10.1094/PDIS.2003.87.8.949

Barenboim, M., Chang, C.Y., dib Hajj, F., Young, R., 1999. Characterization of the dual start motif of a class II holin gene. Mol. Microbiol. 32, 715–27.

Barry, M., van Buuren, N., Burles, K., Mottet, K., Wang, Q., Teale, A., 2010. Poxvirus exploitation of the ubiquitin-proteasome system. Viruses 2, 2356–80. doi:10.3390/v2102356

Basit, H.A., Angle, J.S., Salem, S., Gewaily, E.M., 1992. Phage coating of soybean seed reduces nodulation by indigenous soil bradyrhizobia. Can. J. Microbiol. 38, 1264–1269. doi:10.1139/m92-208

Becker, S.A., Palsson, B.?, Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O., 2005. The RAST Server: Rapid Annotations using Subsystems Technology. BMC Microbiol. 5, 8. doi:10.1186/1471-2180-5-8

Bell, K.S., Sebaihia, M., Pritchard, L., Holden, M.T.G., Hyman, L.J., Holeva, M.C., Thomson, N.R., Bentley, S.D., Churcher, L.J.C., Mungall, K., Atkin, R., Bason, N., Brooks, K., Chillingworth, T., Clark, K., Doggett, J., Fraser, A., Hance, Z., Hauser, H., Jagels, K., Moule, S., Norbertczak, H., Ormond, D., Price, C., Quail, M.A., Sanders, M., Walker, D., Whitehead, S., Salmond, G.P.C., Birch, P.R.J., Parkhill, J., Toth, I.K., 2004. Genome sequence of the enterobacterial phytopathogen *Erwinia carotovora* subsp. *atroseptica* and characterization of virulence factors. Proc. Natl. Acad. Sci. U. S. A. 101, 11105–10. doi:10.1073/pnas.0402424101

Berry, J.D., Rajaure, M., Young, R., 2013. Spanin function requires subunit homodimerization through intermolecular disulfide bonds. Mol. Microbiol. 88, 35–47. doi:10.1111/mmi.12167

Besemer, J., Lomsadze, A., Borodovsky, M., 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res. 29, 2607–18.

Blower, T.R., Chai, R., Przybilski, R., Chindhy, S., Fang, X., Kidman, S.E., Tan, H., Luisi, B.F., Fineran, P.C., Salmond, G.P.C., 2017. Evolution of *Pectobacterium* bacteriophage ΦM1 to escape two bifunctional type III toxin-antitoxin and abortive infection systems through mutations in a single viral gene. Appl. Environ. Microbiol. 83, AEM.03229-16. doi:10.1128/AEM.03229-16

Blower, T.R., Evans, T.J., Przybilski, R., Fineran, P.C., Salmond, G.P.C., 2012. Viral evasion of a bacterial suicide system by RNA-based molecular mimicry enables infectious altruism. PLoS Genet. 8, e1003023. doi:10.1371/journal.pgen.1003023

Bonn, W.G., van der Zwet, T., 2000. Distribution and economic importance of fire blight, in: Vanneste, J.L. (Ed.), Fire Blight: The Disease and Its Causative Agent Erwinia Amylovora. CABI, p. 37.

Bord Bia the Irish Food Board, 2015. Potato industry welcomes €1 million marketing boost. [WWW Document]. Bord Bia. Irish Food Board. URL https://www.bordbia.ie/corporate/press/2015/pages/Potatoindustry1millionmarketingboost.aspx (accessed 1.22.18).

Born, Y., Bosshard, L., Duffy, B., Loessner, M.J., Fieseler, L., 2015. Protection of *Erwinia amylovora* bacteriophage Y2 from UV-induced damage by natural compounds.

Bacteriophage 5, e1074330. doi:10.1080/21597081.2015.1074330

Born, Y., Fieseler, L., Klumpp, J., Eugster, M.R., Zurfluh, K., Duffy, B., Loessner, M.J., 2013. The tail-associated depolymerase of *Erwinia amylovora* phage L1 mediates host cell adsorption and enzymatic capsule removal, which can enhance infection by other phage. Environ. Microbiol. 1. doi:10.1111/1462-2920.12212

Born, Y., Fieseler, L., Marazzi, J., Lurz, R., Duffy, B., Loessner, M.J., 2011. Novel virulent and broad-host-range *Erwinia amylovora* bacteriophages reveal a high degree of mosaicism and a relationship to *Enterobacteriaceae* phages. Appl. Environ. Microbiol. 77, 5945–54. doi:10.1128/AEM.03022-10

Boulé, J., Sholberg, P.L., Lehman, S.M., O'gorman, D.T., Svircev, A.M., 2011. Isolation and characterization of eight bacteriophages infecting *Erwinia amylovora* and their potential as biological control agents in British Columbia, Canada. Can. J. Plant Pathol. 33, 308–317. doi:10.1080/07060661.2011.588250

Branston, 2012. A natural solution to takle potential soft rot | Branston Limited [WWW Document]. URL http://www.branston.com/news/a-natural-solution-to-takle-potential-soft-rot/ (accessed 3.24.16).

Brewer, T.E., Stroupe, M.E., Jones, K.M., 2014. The genome, proteome and phylogenetic analysis of *Sinorhizobium meliloti* phage ΦM12, the founder of a new group of T4-superfamily phages. Virology 450–451, 84–97. doi:10.1016/j.virol.2013.11.027

Briers, Y., Lavigne, R., Plessers, P., Hertveldt, K., Hanssens, I., Engelborghs, Y., Volckaert, G., 2006. Stability analysis of the bacteriophage φKMV lysin gp36C and its putative role during infection. Cell. Mol. Life Sci. 63, 1899–1905. doi:10.1007/s00018-006-6183-7

Briers, Y., Peeters, L.M., Volckaert, G., Lavigne, R., 2011. The lysis cassette of bacteriophage φKMV encodes a signal-arrest-release endolysin and a pinholin. Bacteriophage 1, 25–30. doi:10.4161/bact.1.1.14868

Briers, Y., Schmelcher, M., Loessner, M.J., Hendrix, J., Engelborghs, Y., Volckaert, G., Lavigne, R., 2009. The high-affinity peptidoglycan binding domain of *Pseudomonas* phage endolysin KZ144. Biochem. Biophys. Res. Commun. 383, 187–191. doi:10.1016/j.bbrc.2009.03.161

Briers, Y., Volckaert, G., Cornelissen, A., Lagaert, S., Michiels, C.W., Hertveldt, K., Lavigne, R., 2007. Muralytic activity and modular structure of the endolysins of *Pseudomonas aeruginosa* bacteriophages φKZ and EL. Mol. Microbiol. 65, 1334–1344. doi:10.1111/j.1365-2958.2007.05870.x

Brunner, M., Pootjes, C.F., 1969. Bacteriophage Release in a Lysogenic Strain of *Agrobacterium tumefaciens*. J. Virol. 3, 181–186.

Bruttin, A., Brüssow, H., 2005. Human volunteers receiving *Escherichia coli* phage T4 orally: a safety test of phage therapy. Antimicrob. Agents Chemother. 49, 2874–8. doi:10.1128/AAC.49.7.2874-2878.2005

Bubán, T., Orosz-Kovács, Z., 2003. The nectary as the primary site of infection by *Erwinia amylovora* (Burr.) Winslow et al.: a mini review. Plant Syst. Evol. 238, 183–194. doi:10.1007/s00606-002-0266-1

Burstein, D., Sun, L., Brown, C., Sharon, I., Anantharaman, K., Probst, A., Thomas, B., Banfield, J., 2016. Major bacterial lineages are essentially devoid of CRISPR-Cas viral defense systems. Nat. Commun. In Press, 1–8. doi:10.1038/ncomms10613

Buttimer, C., Hendrix, H., Lucid, A., Neve, H., Noben, J.-P., Franz, C., O'Mahony, J., Lavigne, R., Coffey, A., 2018a. Novel N4-Like Bacteriophages of *Pectobacterium atrosepticum*. Pharmaceuticals 11, 45. doi:10.3390/ph11020045

Buttimer, C., Hendrix, H., Oliveira, H., Casey, A., Neve, H., McAuliffe, O., Ross, R.P., Hill, C., Noben, J.-P., O'Mahony, J., Lavigne, R., Coffey, A., 2017a. Things are getting hairy: Enterobacteria bacteriophage vB_PcaM_CBB. Front. Microbiol. 8, 44. doi:10.3389/fmicb.2017.00044

Buttimer, C., Lucid, A., Neve, H., Franz, C., O'Mahony, J., Turner, D., Lavigne, R., Coffey, A., 2018b. *Pectobacterium atrosepticum* phage vB_PatP_CB5: A member of the proposed genus 'Phimunavirus.' Viruses 10, 394. doi:10.3390/v10080394

Buttimer, C., McAuliffe, O., Ross, R.P.P., Hill, C., O'Mahony, J., Coffey, A., O'Mahony, J., Coffey, A., 2017b. Bacteriophages and bacterial plant diseases. Front. Microbiol. 8, 34. doi:10.3389/fmicb.2017.00034

Buttimer, C., O'Sullivan, L., Elbreki, M., Neve, H., McAuliffe, O., Ross, R.P., Hill, C., O'Mahony, J., Coffey, A., 2016. Genome Sequence of Jumbo Phage vB_AbaM_ME3 of *Acinetobacter baumanni*. Genome Announc. 4, e00431-16. doi:10.1128/genomeA.00431-16

CABI, 2016. *Erwinia amylovora* (fireblight) [WWW Document]. URL http://www.cabi.org/isc/datasheet/21908

Calendar, R., 2006. The bacteriophages. Oxford University Press.

Camesano, T.A., 2015. Nanotechnology to aid chemical and biological defense. Springer.

Carter, R.H., Demidenko, A.A., Hattingh-Willis, S., Rothman-Denes, L.B., 2003. Phage N4 RNA polymerase II recruitment to DNA by a single-stranded DNA-binding protein. Genes Dev. 17, 2334–45. doi:10.1101/gad.1121403

Carver, T.J., Rutherford, K.M., Berriman, M., Rajandream, M.-A., Barrell, B.G., Parkhill, J., 2005. ACT: the Artemis comparison tool. Bioinformatics 21, 3422–3423. doi:10.1093/bioinformatics/bti553

Casey, E., Fitzgerald, B., Mahony, J., Lugli, G.A., Ventura, M., van Sinderen, D., 2017. Genome Sequence of *Serratia marcescens* Phage BF. Genome Announc. 5, e00211-17. doi:10.1128/genomeA.00211-17

Casjens, S.R., Gilcrease, E.B., 2009. Determining DNA packaging strategy by analysis of the termini of the chromosomes in tailed-bacteriophage virions, in: Clokie, M.R.J., Kropinski, A.M. (Eds.), Bacteriophages: Methods and Protocols, Volume 2 Molecular and Applied Aspects. Humana Press, Totowa, NJ, pp. 91–111. doi:10.1007/978-1-60327-565-1_7

Ceyssens, P.-J., Brabban, A., Rogge, L., Lewis, M.S., Pickard, D., Goulding, D., Dougan, G., Noben, J.-P., Kropinski, A., Kutter, E., Lavigne, R., 2010. Molecular and physiological analysis of three *Pseudomonas aeruginosa* phages belonging to the "N4-like viruses." Virology 405, 26–30. doi:10.1016/J.VIROL.2010.06.011

Ceyssens, P.-J., Hertveldt, K., Ackermann, H.-W., Noben, J.-P., Demeke, M., Volckaert, G., Lavigne, R., 2008. The intron-containing genome of the lytic *Pseudomonas* phage LUZ24 resembles the temperate phage PaP3. Virology 377, 233–238. doi:10.1016/J.VIROL.2008.04.038

Ceyssens, P.-J., Lavigne, R., Mattheus, W., Chibeu, A., Hertveldt, K., Mast, J., Robben, J., Volckaert, G., 2006. Genomic analysis of *Pseudomonas aeruginosa* phages LKD16 and

LKA1: establishment of the ΦKMV subgroup within the T7 supergroup. J. Bacteriol. 188, 6924–6931. doi:10.1128/JB.00831-06

Chan, B.K., Abedon, S.T., Loc-Carrillo, C., 2013. Phage cocktails and the future of phage therapy. Future Microbiol. 8, 769–783. doi:10.2217/fmb.13.47

Chan, J.Z.-M., Millard, A.D., Mann, N.H., Schäfer, H., 2014. Comparative genomics defines the core genome of the growing N4-like phage genus and identifies N4-like Roseophage specific genes. Front. Microbiol. 5, 506. doi:10.3389/fmicb.2014.00506

Chang, C.C., Sheen, J., Bligny, M., Niwa, Y., Lerbs-Mache, S., Stern, D.B., 1999. Functional analysis of two maize cDNAs encoding T7-like RNA polymerases. Plant Cell 11, 911–26.

Chatterjee, S., Almeida, R.P.P., Lindow, S., 2008. Living in two Worlds: The Plant and Insect Lifestyles of *Xylella fastidiosa*. Annu. Rev. Phytopathol. 46, 243–271. doi:10.1146/annurev.phyto.45.062806.094342

Chattopadhyay, D., Chattopadhyay, S., Lyon, W.G., Wilson, J.T., 2002. Effect of Surfactants on the Survival and Sorption of Viruses. Environ. Sci. Technol. 36, 4017–4024. doi:10.1021/es0114097

Chevalier, B.S., Stoddard, B.L., 2001. Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. Nucleic Acids Res. 29, 3757–3774. doi:10.1093/nar/29.18.3757

Choi, K.H., McPartland, J., Kaganman, I., Bowman, V.D., Rothman-Denes, L.B., Rossmann, M.G., 2008. Insight into DNA and protein transport in double-stranded DNA viruses: the structure of bacteriophage N4. J. Mol. Biol. 378, 726–36. doi:10.1016/j.jmb.2008.02.059

Choi, M., Miller, A., Cho, N.Y., Rothman-Denes, L.B., 1995. Identification, cloning, and characterization of the bacteriophage N4 gene encoding the single-stranded DNA-binding protein. A protein required for phage replication, recombination, and late transcription. J. Biol. Chem. 270, 22541–7.

Chopin, M.-C., Chopin, A., Bidnenko, E., 2005. Phage abortive infection in lactococci: variations on a theme. Curr. Opin. Microbiol. 8, 473–9. doi:10.1016/j.mib.2005.06.006

Coffey, A., Ross, R.P., 2002. Bacteriophage-resistance systems in dairy starter strains: molecular analysis to application. Antonie Van Leeuwenhoek 82, 303–21.

Comeau, A.M., Krisch, H.M., 2008. The capsid of the T4 phage superfamily: The evolution, diversity, and structure of some of the most prevalent proteins in the biosphere. Mol. Biol. Evol. 25, 1321–1332. doi:10.1093/molbev/msn080

Considine, D.M., Considine, G.D., 1995. Foods and Food Production Encyclopedia. Springer US, Boston, MA. doi:10.1007/978-1-4684-8511-0

Coons, G., Kotila, J., 1925. The transmissible lytic principle (bacteriophage) in relation to plant pathogens. Phytopathology 15, 357–70.

Cruz, L., Cruz, J., Eloy, M., Oliveira, H., Vaz, H., Tenreiro, R., 2010. First Report of Bacterial Speck of Tomato Caused by *Pseudomonas syringae* pv. *tomato* Race 1 in Portugal. Plant Dis. 94, 1504–1504. doi:10.1094/PDIS-06-10-0415

Czajkowski, R., 2016. Bacteriophages of soft rot *Enterobacteriaceae* —a minireview. FEMS Microbiol. Lett. 363, fnv230. doi:10.1093/femsle/fnv230

Czajkowski, R., Ozymko, Z., de Jager, V., Siwinska, J., Smolarska, A., Ossowicki, A., Narajczyk,

M., Lojkowska, E., 2015. Genomic, proteomic and morphological characterization of two novel broad host lytic bacteriophages ΦPD10.3 and ΦPD23.1 infecting pectinolytic *Pectobacterium* spp. and *Dickeya* spp. PLoS One 10, e0119812. doi:10.1371/journal.pone.0119812

Czajkowski, R., Ozymko, Z., Lojkowska, E., 2014. Isolation and characterization of novel soilborne lytic bacteriophages infecting *Dickeya* spp. biovar 3 (" *D. solani* "). Plant Pathol. 63, 758–772. doi:10.1111/ppa.12157

Czajkowski, R., Pérombelon, M.C.M., van Veen, J. a., van der Wolf, J.M., 2011. Control of blackleg and tuber soft rot of potato caused by *Pectobacterium* and *Dickeya species*: a review. Plant Pathol. 60, 999–1013. doi:10.1111/j.1365-3059.2011.02470.x

D'Herelle, F., 1917. Sur un microbe invisible antagoniste des *Bacillies dysentériques*. C R Acad Sci 165, 373–375.

Darrasse, A., Priou, S., Kotoujansky, A., Bertheau, Y., 1994. PCR and restriction fragment length polymorphism of a pel gene as a tool to identify *Erwinia carotovora* in relation to potato diseases. Appl. Environ. Microbiol. 60, 1437–43.

Das, M., Bhowmick, T.S., Ahern, S.J., Young, R., Gonzalez, C.F., 2015. Control of Pierce's Disease by Phage. PLoS One 10, e0128902. doi:10.1371/journal.pone.0128902

De Boer, S.H., 2004. Blackleg of potato. Plant Heal. Instr. doi:10.1094/PHI-I-2004-0712-01

De Boer, S.H., Ward, L.J., 1995. PCR detection of *Erwinia carotovora* subsp *atroseptica* associated with potato tissue. Phytopathology. doi:10.1094/Phyto-85-854

de León Door, A.P., Romo Chacón, A., Acosta Muñiz., C., 2013. Detection of streptomycin resistance in *Erwinia amylovora* strains isolated from apple orchards in Chihuahua, Mexico. Eur. J. Plant Pathol. 137, 223–229. doi:10.1007/s10658-013-0241-4

Delcher, A.L., Harmon, D., Kasif, S., White, O., Salzberg, S.L., 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. 27, 4636–41. doi:10.1093/nar/27.23.4636

Delitheos, A., Tiligada, E., Yannitsaros, A., Bazos, I., 1997. Antiphage activity in extracts of plants growing in Greece. Phytomedicine 4, 117–24. doi:10.1016/S0944-7113(97)80055-4

Denny, T., 2007. Plant pathogenic Ralstonia species, in: Plant-Associated Bacteria. Springer Netherlands, Dordrecht, pp. 573–644. doi:10.1007/978-1-4020-4538-7_16

Denou, E., Bruttin, A., Barretto, C., Ngom-Bru, C., Brüssow, H., Zuber, S., 2009. T4 phages against *Escherichia coli* diarrhea: potential and problems. Virology 388, 21–30. doi:10.1016/j.virol.2009.03.009

Doffkay, Z., Dömötör, D., Kovács, T., Rákhely, G., 2015. Bacteriophage therapy against plant , animal and human pathogens. Acta Biol. Szeged. 59, 291–302.

Dömötör, D., Becságh, P., Rákhely, G., Schneider, G., Kovács, T., 2012. Complete genomic sequence of *Erwinia amylovora* phage PhiEaH2. J. Virol. 86, 10899. doi:10.1128/JVI.01870-12

Drulis-Kawa, Z., Olszak, T., Danis, K., Majkowska-Skrobek, G., Ackermann, H.-W., 2014. A giant *Pseudomonas* phage from Poland. Arch. Virol. 159, 567–72. doi:10.1007/s00705-013-1844-y

Dwivedi, B., Xue, B., Lundin, D., Edwards, R.A.R., Breitbart, M., Grath, S.M., VanSinderen, D.,

Weinbauer, M., Rassoulzadegan, F., Wommack, K., Colwell, R., Chen, F., Lu, J., Chenard, C., Suttle, C., Goldsmith, D., Crosti, G., Dwivedi, B., McDaniel, L., Al, E., 2013. A bioinformatic analysis of ribonucleotide reductase genes in phage genomes and metagenomes. BMC Evol. Biol. 13, 33. doi:10.1186/1471-2148-13-33

Eddy, S.R., 2011. Accelerated Profile HMM Searches. PLoS Comput. Biol. 7, e1002195. doi:10.1371/journal.pcbi.1002195

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797. doi:10.1093/nar/gkh340

Edgell, D.R., Belfort, M., Shub, D.A., 2000. Barriers to intron promiscuity in bacteria. J. Bacteriol. 182, 5281–9. doi:10.1128/JB.182.19.5281-5289.2000

Edgell, D.R., Gibb, E.A., Belfort, M., 2010. Mobile DNA elements in T4 and related phages. Virol. J. 7, 290. doi:10.1186/1743-422X-7-290

Elizabeth, S. V., Bender, C.L., 2007. The phytotoxin coronatine from *Pseudomonas syringae* pv. tomato DC3000 functions as a virulence factor and influences defence pathways in edible brassicas. Mol. Plant Pathol. 8, 83–92. doi:10.1111/j.1364-3703.2006.00372.x

Enviroinvest, n.d. Business activity [WWW Document]. URL http://biotechnologia.enviroinvest.hu/business-activity.html (accessed 3.24.16).

Eriksson, H., Maciejewska, B., Latka, A., Majkowska-Skrobek, G., Hellstrand, M., Melefors, Ö., Wang, J.T., Kropinski, A.M., Drulis-Kawa, Z., Nilsson, A.S., 2015. A suggested new bacteriophage genus, "Kp34likevirus", within the *Autographivirinae* subfamily of *Podoviridae*. Viruses 7, 1804–1822. doi:10.3390/v7041804

Erskine, J.M., 1973. Characteristics of *Erwinia amylovora* bacteriophage and its possible role in the epidemiology of fire blight. Can. J. Microbiol. 19, 837–845. doi:10.1139/m73-134

Esplin, I.N.D., Berg, J.A., Sharma, R., Allen, R.C., Arens, D.K., Ashcroft, C.R., Bairett, S.R., Beatty, N.J., Bickmore, M., Bloomfield, T.J., Brady, T.S., Bybee, R.N., Carter, J.L., Choi, M.C., Duncan, S., Fajardo, C.P., Foy, B.B., Fuhriman, D.A., Gibby, P.D., Grossarth, S.E., Harbaugh, K., Harris, N., Hilton, J.A., Hurst, E., Hyde, J.R., Ingersoll, K., Jacobson, C.M., James, B.D., Jarvis, T.M., Jaen-anieves, D., Jensen, G.L., Knabe, B.K., Kruger, J.L., Merrill, B.D., Pape, J.A., Anderson, A.M.P., Payne, D.E., Peck, M.D., Pollock, S. V, Putnam, M.J., Ransom, E.K., Ririe, D.B., Robinson, D.M., Rogers, S.L., Russell, K.A., Schoenhals, J.E., Shurtleff, C.A., Simister, A.R., Smith, H.G., Stephenson, M.B., Staley, L.A., Stettler, J.M., Stratton, M.L., Tateoka, O.B., Tatlow, P.J., Taylor, A.S., Thompson, S.E., Townsend, M.H., Thurgood, T.L., Usher, B.K., Whitley, K. V, Ward, A.T., Ward, M.E.H., Webb, C.J., Wienclaw, T.M., Williamson, T.L., Wells, M.J., Wright, C.K., Breakwell, D.P., Hope, S., Grose, J.H., 2017. Genome Sequences of 19 Novel *Erwinia amylovora* Bacteriophages. Genome Announc. 5, 1–3.

Estrem, S.T., Gaal, T., Ross, W., Gourse, R.L., 1998. Identification of an UP element consensus sequence for bacterial promoters. Proc. Natl. Acad. Sci. U. S. A. 95, 9761–6.

Evans, T.J., Coulthurst, S.J., Komitopoulou, E., Salmond, G.P.C., 2010. Two mobile *Pectobacterium atrosepticum* prophages modulate virulence. FEMS Microbiol. Lett. 304, 195–202. doi:10.1111/j.1574-6968.2010.01901.x

Evans, T.J., Ind, A., Komitopoulou, E., Salmond, G.P.C., 2010a. Phage-selected lipopolysaccharide mutants of *Pectobacterium atrosepticum* exhibit different impacts on virulence. J. Appl. Microbiol. 109, 505–14. doi:10.1111/j.1365-2672.2010.04669.x

Evans, T.J., Trauner, A., Komitopoulou, E., Salmond, G.P.C., 2010b. Exploitation of a new flagellatropic phage of *Erwinia* for positive selection of bacterial mutants attenuated in plant virulence: towards phage therapy. J. Appl. Microbiol. 108, 676–85. doi:10.1111/j.1365-2672.2009.04462.x

Falkenstein, H., Bellemann, P., Walter, S., Zeller, W., Geider, K., 1988. Identification of *Erwinia amylovora*, the Fireblight Pathogen, by Colony Hybridization with DNA from Plasmid pEA29. Appl. Environ. Microbiol. 54, 2798–2802.

Farris, J.S., 1972. Estimating phylogenetic trees from distance matrices. Am. Nat. 106, 645–668. doi:10.1086/282802

Fatmi, M., 2003. Use of Oxos, a Complex of Hydrogen Peroxide, Acetic Acid and Silver Ion, to Control Bacterial Speck of Tomato (Pseudomonas syringae pv. tomato) and Angular Leaf Spot of Melon (P. s. pv. lachrymans), in: *Pseudomonas Syringae* and Related Pathogens. Springer Netherlands, Dordrecht, pp. 459–466. doi:10.1007/978-94-017-0133-4_50

Feeney, M.A., Ke, N., Beckwith, J., 2012. Mutations at several loci cause increased expression of ribonucleotide reductase in *Escherichia coli*. J. Bacteriol. 194, 1515–1522. doi:10.1128/JB.05989-11

Fineran, P.C., Blower, T.R., Foulds, I.J., Humphreys, D.P., Lilley, K.S., Salmond, G.P.C., 2009. The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. Proc. Natl. Acad. Sci. U. S. A. 106, 894–9. doi:10.1073/pnas.0808832106

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J., Bateman, A., 2015. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 44, D279-85. doi:10.1093/nar/gkv1344

Flaherty, J.E., Harbaugh, B.K., Jones, J.B., Somodi, G.C., Jackson, L.E., 2001. H-mutant bacteriophages as a potential biocontrol of bacterial blight of Geranium. HortScience 36, 98–100.

Fouts, D.E., Klumpp, J., Bishop-Lilly, K.A., Rajavel, M., Willner, K.M., Butani, A., Henry, M., Biswas, B., Li, M., Albert, M., Loessner, M.J., Calendar, R., Sozhamannan, S., 2013. Whole genome sequencing and comparative genomic analyses of two *Vibrio cholerae* O139 Bengal-specific Podoviruses to other N4-like phages reveal extensive genetic diversity. Virol. J. 10, 165. doi:10.1186/1743-422X-10-165

Frampton, R.A., Pitman, A.R., Fineran, P.C., 2012. Advances in bacteriophage-mediated control of plant pathogens. Int. J. Microbiol. 2012, 326452. doi:10.1155/2012/326452

Frampton, R.A., Taylor, C., Holguín Moreno, A. V, Visnovsky, S.B., Petty, N.K., Pitman, A.R., Fineran, P.C., 2014. Identification of bacteriophages for biocontrol of the kiwifruit canker phytopathogen *Pseudomonas syringae* pv. *actinidiae*. Appl. Environ. Microbiol. 80, 2216–28. doi:10.1128/AEM.00062-14

Fraser, J.S., Yu, Z., Maxwell, K.L., Davidson, A.R., 2006. Ig-like domains on bacteriophages: A tale of promiscuity and deceit. J. Mol. Biol. 359, 496–507. doi:10.1016/j.jmb.2006.03.043

Friedrich, N.C., Torrents, E., Gibb, E.A., Sahlin, M., Sjö Berg, B.-M., Edgell, D.R., 2007. Insertion of a homing endonuclease creates a genes-in-pieces ribonucleotide reductase that retains function. Proc Natl Acad Sci U S A. 104, 6176–6181.

Fujiwara, A., Fujisawa, M., Hamasaki, R., Kawasaki, T., Fujie, M., Yamada, T., 2011. Biocontrol of *Ralstonia solanacearum* by treatment with lytic bacteriophages. Appl. Environ.

Microbiol. 77, 4155–62. doi:10.1128/AEM.02847-10

García, P., Martínez, B., Obeso, J.M., Rodríguez, A., 2008. Bacteriophages and their application in food safety. Lett. Appl. Microbiol. 47, 479–485. doi:10.1111/j.1472-765X.2008.02458.x

Gardan, L., 2003. Elevation of three subspecies of *Pectobacterium carotovorum* to species level: *Pectobacterium atrosepticum* sp. nov., *Pectobacterium betavasculorum* sp. nov. and *Pectobacterium wasabiae* sp. nov. Int. J. Syst. Evol. Microbiol. 53, 381–391. doi:10.1099/ijs.0.02423-0

Garrity, G.M., 2004. Bergey's manual of systematic bacteriology. Vol. 2, The proteobacteria, Bergey's manual of systematic bacteriology. Vol. 2, The proteobacteria. Springer, New York.

Geiduschek, E.P., Kassavetis, G.A., 2010. Transcription of the T4 late genes. Virol. J. 7, 288. doi:10.1186/1743-422X-7-288

Ghosh, N., McKillop, T.J., Jowitt, T.A., Howard, M., Davies, H., Holmes, D.F., Roberts, I.S., Bella, J., 2012. Collagen-Like Proteins in Pathogenic *E. coli* Strains. PLoS One 7, e37872. doi:10.1371/journal.pone.0037872

Glucksmann, M.A., Markiewicz, P., Malone, C., Rothman-Denes, L.B., 1992. Specific sequences and a hairpin structure in the template strand are required for N4 virion RNA polymerase promoter recognition. Cell 70, 491–500. doi:10.1016/0092-8674(92)90173-A

Göker, M., García-Blázquez, G., Voglmayr, H., Tellería, M.T., Martín, M.P., 2009. Molecular taxonomy of phytopathogenic fungi: a case study in *Peronospora*. PLoS One 4, e6319. doi:10.1371/journal.pone.0006319

Goode, M.J., Sasser, M., 1980. Prevention-the key to controlling bacterial spot and bacterial speck of tomato. Plant Dis. 64, 831. doi:10.1094/PD-64-831

Goyal, S.M., Gerba, C.P., 1979. Comparative adsorption of human enteroviruses, simian rotavirus, and selected bacteriophages to soils. Appl. Environ. Microbiol. 38, 241–247.

Goyer, C., 2005. Isolation and characterization of phages Stsc1 and Stsc3 infecting *Streptomyces scabiei* and their potential as biocontrol agents. Can. J. Plant Pathol. 27, 210–216. doi:10.1080/07060660509507218

Grant, J.R., Stothard, P., 2008. The CGView Server: a comparative genomics tool for circular genomes. Nucleic Acids Res. 36, W181–W184. doi:10.1093/nar/gkn179

Grazziotin, A.L., Koonin, E. V., Kristensen, D.M., 2017. Prokaryotic virus orthologous groups (pVOGs): A resource for comparative genomics and protein family annotation. Nucleic Acids Res. 45, D491–D498. doi:10.1093/nar/gkw975

Griffiths, A.J., Miller, J.H., Suzuki, D.T., Lewontin, R.C., Gelbart, W.M., 2000. Transduction, in: An Introduction to Genetic Analysis. 7th Edition. W. H. Freeman.

Grose, J.H., Casjens, S.R., 2014. Understanding the enormous diversity of bacteriophages: the tailed phages that infect the bacterial family *Enterobacteriaceae*. Virology 468–470, 421–43. doi:10.1016/j.virol.2014.08.024

Guo, J., Cheng, H., Zhao, S., Yu, L., 2006. GG: A domain involved in phage LTF apparatus and implicated in human MEB and non-syndromic hearing loss diseases. FEBS Lett. 580, 581–584. doi:10.1016/j.febslet.2005.12.076

Hardies, S.C., Thomas, J.A., Black, L., Weintraub, S.T., Hwang, C.Y., Cho, B.C., 2016.

Identification of structural and morphogenesis genes of *Pseudoalteromonas* phage φRIO-1 and placement within the evolutionary history of *Podoviridae*. Virology 489, 116–27. doi:10.1016/j.virol.2015.12.005

Hatfull, G.F., 2008. Bacteriophage genomics. Curr. Opin. Microbiol. 11, 447–53. doi:10.1016/j.mib.2008.09.004

Hauben, L., Moore, E.R., Vauterin, L., Steenackers, M., Mergaert, J., Verdonck, L., Swings, J., 1998. Phylogenetic position of phytopathogens within the *Enterobacteriaceae*. Syst. Appl. Microbiol. 21, 384–97. doi:10.1016/S0723-2020(98)80048-9

Hedtke, B., Börner, T., Weihe, A., 1997. Mitochondrial and chloroplast phage-type RNA polymerases in *Arabidopsis*. Science (80-.). 277, 809–811. doi:10.1126/science.277.5327.809

Hélias, V., Hamon, P., Huchet, E., Wolf, J.V.D., Andrivon, D., 2012. Two new effective semiselective crystal violet pectate media for isolation of *Pectobacterium* and *Dickeya*. Plant Pathol. 61, 339–345. doi:10.1111/j.1365-3059.2011.02508.x

Hendrix, R.W., 2009. Jumbo bacteriophages. Curr. Top. Microbiol. Immunol. 328, 229–240. doi:10.1007/978-3-540-68618-7-7

Hendrix, R.W., 2002. Bacteriophages: Evolution of the majority. Theor. Popul. Biol. 61, 471–480. doi:10.1006/tpbi.2002.1590

Hertveldt, K., Lavigne, R., Pleteneva, E., Sernova, N., Kurochkina, L., Korchevskii, R., Robben, J., Mesyanzhinov, V., Krylov, V.N., Volckaert, G., 2005. Genome Comparison of *Pseudomonas aeruginosa* large phages. J. Mol. Biol. 354, 536–545. doi:10.1016/j.jmb.2005.08.075

Hinton, D.M., 2010. Transcriptional control in the prereplicative phase of T4 development. Virol. J. 7, 289. doi:10.1186/1743-422X-7-289

Hirano, S.S., Upper, C.D., 2000. Bacteria in the leaf ecosystem with emphasis on *Pseudomonas syringae* -a pathogen, ice nucleus, and epiphyte. Microbiol. Mol. Biol. Rev. 64, 624–53. doi:10.1128/MMBR.64.3.624-653.2000

Hirano, S.S., Upper, C.D., 1990. Population Biology and Epidemiology of *Pseudomonas syringae*. Annu. Rev. Phytopathol. 28, 155–177. doi:10.1146/annurev.py.28.090190.001103

Hirata, H., Kashihara, M., Horiike, T., Suzuki, T., Dohra, H., Netsu, O., Tsuyumu, S., 2016. Genome Sequence of *Pectobacterium carotovorum* Phage PPWS1, Isolated from Japanese Horseradish [*Eutrema japonicum* (Miq.) Koidz] Showing Soft-Rot Symptoms. Genome Announc. 4. doi:10.1128/genomeA.01625-15

Hirst, J.M., Riche, H.H., Bascomb, C.L., 1961. Copper accumulation in the soils of apple orchards near Wisbech. Plant Pathol. 10, 105–108. doi:10.1111/j.1365-3059.1961.tb00127.x

Hopkins, D.L., Purcell, A.H., 2002. *Xylella fastidiosa* : Cause of Pierce's Disease of Grapevine and Other Emergent Diseases. Plant Dis. 86, 1056–1066.

Hua, J., 2016. Capsid structure and DNA packing in jumbo bacteriophage. University of Pittsburgh.

Hurwitz, B.L., U'Ren, J.M., 2016. Viral metabolic reprogramming in marine ecosystems. Curr. Opin. Microbiol. 31, 161–168. doi:10.1016/j.mib.2016.04.002

Ikeda, T.M., Gray, M.W., 1999. Identification and characterization of T3/T7 bacteriophage-like RNA polymerase sequences in wheat. Plant Mol. Biol. 40, 567–78.

Iriarte, F.B., Balogh, B., Momol, M.T., Smith, L.M., Wilson, M., Jones, J.B., 2007. Factors affecting survival of bacteriophage on tomato leaf surfaces. Appl. Environ. Microbiol. 73, 1704–11. doi:10.1128/AEM.02118-06

Iriarte, F.B., Obradović, A., Wernsing, M.H., Jackson, L.E., Balogh, B., Hong, J.A., Momol, M.T., Jones, J.B., Vallad, G.E., 2012. Soil-based systemic delivery and phyllosphere in vivo propagation of bacteriophages: Two possible strategies for improving bacteriophage persistence for plant disease control. Bacteriophage 2, 215–224. doi:10.4161/bact.23530

Ivankov, D.N., Payne, S.H., Galperin, M.Y., Bonissone, S., Pevzner, P.A., Frishman, D., 2013. How many signal peptides are there in bacteria? Environ. Microbiol. 15, 983–90. doi:10.1111/1462-2920.12105

Janse, J.D., Obradovic, A., 2010. *Xylella fastidiosa*: Its biology, diagnosis, control and risks. J. Plant Pathol. 92, 1–35–S1.48. doi:10.4454/JPP.V92I1SUP.2504

Javed, M.A., Poshtiban, S., Arutyunov, D., Evoy, S., Szymanski, C.M., 2013. Bacteriophage Receptor Binding Protein Based Assays for the Simultaneous Detection of *Campylobacter jejuni* and *Campylobacter coli*. PLoS One 8, e69770. doi:10.1371/journal.pone.0069770

Jenkins, J., Mayans, O., Pickersgill, R., 1998. Structure and Evolution of Parallel β-Helix Proteins. J. Struct. Biol. 122, 236–246. doi:10.1006/jsbi.1998.3985

Jones, D.T., Taylor, W.R., Thornton, J.M., 1992. The rapid generation of mutation data matrices from protein sequences. Bioinformatics 8, 275–282. doi:10.1093/bioinformatics/8.3.275

Jones, J.B., Lacy, G.H., Bouzar, H., Minsavage, G.V., Stall, R.E., Schaad, N.W., 2005. Bacterial Spot - Worldwide Distribution, Importance and Review. Acta Hortic. 27–34. doi:10.17660/ActaHortic.2005.695.1

Jones, J.B., Lacy, G.H., Bouzar, H., Stall, R.E., Schaad, N.W., 2004. Reclassification of the xanthomonads associated with bacterial spot disease of tomato and pepper. Syst. Appl. Microbiol. 27, 755–62. doi:10.1078/0723202042369884

Juncker, A.S., Willenbrock, H., von Heijne, G., Brunak, S., Nielsen, H., Krogh, A., 2003. Prediction of lipoprotein signal peptides in Gram-negative bacteria. Protein Sci. 12, 1652–1662. doi:10.1110/ps.0303703

Kala, S., Cumby, N., Sadowski, P.D., Hyder, B.Z., Kanelis, V., Davidson, A.R., Maxwell, K.L., 2014. HNH proteins are a widespread component of phage DNA packaging machines. Proc. Natl. Acad. Sci. U. S. A. 111, 6022–7. doi:10.1073/pnas.1320952111

Kalischuk, M., Hachey, J., Kawchuk, L., 2015. Complete genome sequence of phytopathogenic *Pectobacterium atrosepticum* bacteriophage Peat1. Genome Announc. 3. doi:10.1128/genomeA.00760-15

Kanamaru, S., Leiman, P.G., Kostyuchenko, V.A., Chipman, P.R., Mesyanzhinov, V. V, Arisaka, F., Rossmann, M.G., 2002. Structure of the cell-puncturing device of bacteriophage T4. Nature 415, 553–7. doi:10.1038/415553a

Kang, H.W., Kwon, S.W., Go, S.J., 2003. PCR-based specific and sensitive detection of *Pectobacterium carotovorum* ssp. *carotovorum* by primers generated from a URP-PCR fingerprinting-derived polymorphic band. Plant Pathol. 52, 127–133. doi:10.1046/j.1365-3059.2003.00822.x

Katharios, P., Kalatzis, P.G., Kokkari, C., Sarropoulou, E., Middelboe, M., 2017. Isolation and characterization of a N4-like lytic bacteriophage infecting *Vibrio splendidus*, a pathogen of fish and bivalves. PLoS One 12, e0190083. doi:10.1371/journal.pone.0190083

Keppel, F., Rychner, M., Georgopoulos, C., 2002. Bacteriophage-encoded cochaperonins can substitute for *Escherichia coli's* essential GroES protein. EMBO Rep. 3, 893–8. doi:10.1093/embo-reports/kvf176

Khalil, I., Irorere, V., Radecka, I., Burns, A., Kowalczuk, M., Mason, J., Khechara, M., 2016. Poly-γ-glutamic acid: Biodegradable polymer for potential protection of beneficial viruses. Materials (Basel). 9, 28. doi:10.3390/ma9010028

Khayi, S., Cigna, J., Chong, T.M., Quêtu-Laurent, A., Chan, K.-G., Hélias, V., Faure, D., 2016. Transfer of the potato plant isolates of *Pectobacterium wasabiae* to *Pectobacterium parmentieri* sp. nov. Int. J. Syst. Evol. Microbiol. 66, 5379–5383. doi:10.1099/ijsem.0.001524

Kim, M.H., Park, S.W., Kim, Y.K., 2011. Bacteriophages of *Pseudomonas tolaasii* for the biological control of brown blotch disease. J. Appl. Biol. Chem. 54, 99–104. doi:10.3839/jksabc.2011.014

Kim, M.S., Hong, S.S., Park, K., Myung, H., 2013. Genomic analysis of bacteriophage PBECO4 infecting *Escherichia coli* O157:H7. Arch. Virol. 158, 2399–403. doi:10.1007/s00705-013-1718-3

King, A.M.Q., Adams, M.J., Carstens, E.B., Lefknowitz, E.J. (Eds.), 2012. Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier.

Klumpp, J., Dorscht, J., Lurz, R., Bielmann, R., Wieland, M., Zimmer, M., Calendar, R., Loessner, M.J., 2008. The terminally redundant, nonpermuted genome of *Listeria* bacteriophage A511: a model for the SPO1-like myoviruses of Gram-positive bacteria. J. Bacteriol. 190, 5753–65. doi:10.1128/JB.00461-08

Koczan, J.M., Lenneman, B.R., McGrath, M.J., Sundin, G.W., 2011. Cell surface attachment structures contribute to biofilm formation and xylem colonization by *Erwinia amylovora*. Appl. Environ. Microbiol. 77, 7031–9. doi:10.1128/AEM.05138-11

Kolozsváriné Nagy, J., Schwarczinger, I., Künstler, A., Pogány, M., Király, L., 2015. Penetration and translocation of *Erwinia amylovora*-specific bacteriophages in apple - a possibility of enhanced control of fire blight. Eur. J. Plant Pathol. 142, 815–827. doi:10.1007/s10658-015-0654-3

Kotila, J., Coons, G., 1925. Investigations on the blackleg disease of potato. Michigan Agric. Exp. Stn. Tech. Bull. 67, 3–29.

Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L.., 2001. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. J. Mol. Biol. 305, 567–580. doi:10.1006/jmbi.2000.4315

Kropinski, A.M., Prangishvili, D., Lavigne, R., 2009. Position paper: The creation of a rational scheme for the nomenclature of viruses of Bacteria and Archaea. Environ. Microbiol. 11, 2775–2777. doi:10.1111/j.1462-2920.2009.01970.x

Kropinski, A.M., Waddell, T., Meng, J., Franklin, K., Ackermann, H.W., Ahmed, R., Mazzocco, A., Yates, J., Lingohr, E.J., Johnson, R.P., 2013. The host-range, genomics and proteomics of *Escherichia coli* O157:H7 bacteriophage rV5. Virol. J. 10, 1. doi:10.1186/1743-422X-10-76

Krupovic, M., Dutilh, B.E., Adriaenssens, E.M., Wittmann, J., Vogensen, F.K., Sullivan, M.B., Rumnieks, J., Prangishvili, D., Lavigne, R., Kropinski, A.M., Klumpp, J., Gillis, A., Enault, F., Edwards, R.A., Duffy, S., Clokie, M.R.C., Barylski, J., Ackermann, H.-W., Kuhn, J.H., 2016. Taxonomy of prokaryotic viruses: update from the ICTV bacterial and archaeal viruses subcommittee. Arch. Virol. 161, 1095–1099. doi:10.1007/s00705-015-2728-0

Kulikov, E., Kropinski, A.M., Golomidova, A., Lingohr, E., Govorun, V., Serebryakova, M., Prokhorov, N., Letarova, M., Manykin, A., Strotskaya, A., Letarov, A., 2012. Isolation and characterization of a novel indigenous intestinal N4-related coliphage vB_EcoP_G7C. Virology 426, 93–99. doi:10.1016/J.VIROL.2012.01.027

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis Version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–4. doi:10.1093/molbev/msw054

Kuo, T.T., Chiang, C.C., Chen, S.Y., Lin, J.H., Kuo, J.L., 1994. A long lytic cycle in filamentous phage Cf1tv infecting *Xanthomonas campestris* pv. *citri*. Arch. Virol. 135, 253–64.

Kuo, T.T., Chow, T.Y., Lin, Y.T., Yang, C.M., Li, H.W., 1971. Specific dissociation of phage Xp12 by sodium citrate. J. Gen. Virol. 10, 199–202. doi:10.1099/0022-1317-10-2-199

Kutin, R.K., Alvarez, A., Jenkins, D.M., 2009. Detection of *Ralstonia solanacearum* in natural substrates using phage amplification integrated with real-time PCR assay. J. Microbiol. Methods 76, 241–246. doi:10.1016/j.mimet.2008.11.008

Kuty, G.F., Xu, M., Struck, D.K., Summer, E.J., Young, R., 2010. Regulation of a phage endolysin by disulfide caging. J. Bacteriol. 192, 5682–7. doi:10.1128/JB.00674-10

Landthaler, M., Begley, U., Lau, N.C., Shub, D.A., 2002. Two self-splicing group I introns in the ribonucleotide reductase large subunit gene of *Staphylococcus aureus* phage Twort. Nucleic Acids Res. 30, 1935–43.

Lang, J.M., Gent, D.H., Schwartz, H.F., 2007. Management of *Xanthomonas* leaf blight of onion with bacteriophages and a plant activator. Plant Dis. 91, 871–878. doi:10.1094/PDIS-91-7-0871

Laslett, D., Canback, B., 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 32, 11–16. doi:10.1093/nar/gkh152

Lavigne, R., Burkal'tseva, M. V, Robben, J., Sykilinda, N.N., Kurochkina, L.P., Grymonprez, B., Jonckx, B., Krylov, V.N., Mesyanzhinov, V. V, Volckaert, G., 2003. The genome of bacteriophage phiKMV, a T7-like virus infecting *Pseudomonas aeruginosa*. Virology 312, 49–59.

Lavigne, R., Seto, D., Mahadevan, P., Ackermann, H.-W., Kropinski, A.M., 2008. Unifying classical and molecular taxonomic classification: analysis of the *Podoviridae* using BLASTP-based tools. Res. Microbiol. 159, 406–414. doi:10.1016/j.resmic.2008.03.005

Lee, D.H., Kim, J.-B., Lim, J.-A., Han, S.-W., Heu, S., 2014. Genetic Diversity of *Pectobacterium carotovorum* subsp. *brasiliensis* Isolated in Korea. Plant Pathol. J. 30, 117–124. doi:10.5423/PPJ.OA.12.2013.0117

Lee, H.J., Kim, W. Il, Kwon, Y.C., Cha, K.E., Kim, M., Myung, H., 2016. A newly isolated bacteriophage, PBES 02, infecting *Cronobacter sakazakii*. J. Microbiol. Biotechnol. 26, 1629–1635. doi:10.4014/jmb.1605.05020

Lefort, V., Desper, R., Gascuel, O., 2015. FastME 2.0: a comprehensive, accurate, and fast distance-Based phylogeny inference program. Mol. Biol. Evol. 32, 2798–2800.

doi:10.1093/molbev/msv150

Lerat, S., Simao-Beaunoir, A.-M., Beaulieu, C., 2009. Genetic and physiological determinants of *Streptomyces scabies* pathogenicity. Mol. Plant Pathol. 10, 579–85. doi:10.1111/j.1364-3703.2009.00561.x

Li, B., Zhang, S., Long, L., Huang, S., 2016. Characterization and complete genome sequences of three N4-Like Roseobacter phages isolated from the south China sea. Curr. Microbiol. 73, 409–418. doi:10.1007/s00284-016-1071-3

Li, J., Wang, N., 2014. Foliar application of biofilm formation-inhibiting compounds enhances control of citrus canker caused by *Xanthomonas citri* subsp. *citri*. Phytopathology 104, 134–42. doi:10.1094/PHYTO-04-13-0100-R

Li, S., Fan, H., An, X., Fan, H., Jiang, H., Chen, Y., Tong, Y., 2014. Scrutinizing virus genome termini by high-throughput sequencing. PLoS One 9, e85806. doi:10.1371/journal.pone.0085806

Lim, J.-A., Heu, S., Park, J., Roh, E., 2017. Genomic characterization of bacteriophage vB_PcaP_PP2 infecting *Pectobacterium carotovorum* subsp. *carotovorum*, a new member of a proposed genus in the subfamily *Autographivirinae*. Arch. Virol. 162, 2441–2444. doi:10.1007/s00705-017-3349-6

Lim, J.-A., Jee, S., Lee, D.H., Roh, E., Jung, K., Oh, C., Heu, S., 2013. Biocontrol of *Pectobacterium carotovorum* subsp. *carotovorum* using bacteriophage PP1. J. Microbiol. Biotechnol. 23, 1147–53.

Lingohr, E., Frost, S., Johnson, R.P., 2009. Determination of bacteriophage genome size by pulsed-field gel electrophoresis. Methods Mol. Biol. 502, 19–25. doi:10.1007/978-1-60327-565-1_3

Łobocka, M.B., Rose, D.J., Plunkett, G., Rusin, M., Samojedny, A., Lehnherr, H., Yarmolinsky, M.B., Blattner, F.R., 2004. Genome of bacteriophage P1. J. Bacteriol. 186, 7032–68. doi:10.1128/JB.186.21.7032-7068.2004

Loconsole, G., Potere, O., Boscia, D., Altamura, G., Djelouah, K., Elbeaino, T., Frasheri, D., Lorusso, D., Palmisano, F., Pollastro, P., Silletti, M.R., Trisciuzzi, N., Valentini, F., Savino, V., Saponari, M., 2014. Detection of *Xylella fastidiosa* in olive trees by molecular and serological methods. J. Plant Pathol. 96, 7–14. doi:10.4454/JPP.V96I1.041

Lohr, L., 2001. Factors affecting international demand and trade in organic food products, in: Regmi, A. (Ed.), Changing Structure of Global Food Consumption and Trade. DIANE Publishing, p. 67.

Loveland, J.., Ryan, J.., Amy, G.., Harvey, R.., 1996. The reversibility of virus attachment to mineral surfaces. Colloids Surfaces A Physicochem. Eng. Asp. 107, 205–221. doi:10.1016/0927-7757(95)03373-4

Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25, 955–64.

Lu, C., Warchol, K.M., Callahan, R.A., 2014. Sub-lethal exposure to neonicotinoids impaired honey bees winterization before proceeding to colony collapse disorder. Bull. Insectology 67, 125–130.

Lu, M.J., Henning, U., 1994. Superinfection exclusion by T-even-type coliphages. Trends Microbiol. 2, 137–9.

Lundin, D., Torrents, E., Poole, A.M., Sjöberg, B.-M., 2009. RNRdb, a curated database of the universal enzyme family ribonucleotide reductase, reveals a high level of misannotation in sequences deposited to Genbank. BMC Genomics 10, 589. doi:10.1186/1471-2164-10-589

Lynch, K.H., Abdu, A.H., Schobert, M., Dennis, J.J., 2013. Genomic characterization of JG068, a novel virulent podovirus active against *Burkholderia cenocepacia*. BMC Genomics 14, 574. doi:10.1186/1471-2164-14-574

Ma, B., Hibbing, M.E., Kim, H.-S., Reedy, R.M., Yedidia, I., Breuer, J.J., Breuer, J.J., Glasner, J.D., Perna, N.T., Kelman, A., Charkowski, A.O., 2007. Host range and molecular phylogenies of the soft rot enterobacterial genera *Pectobacterium* and *Dickeya*. Phytopathology 97, 1150–63. doi:10.1094/PHYTO-97-9-1150

MacConnell, S., 2002. Blight, slugs and weather threaten potato crop [WWW Document]. Irish Exam. URL https://www.irishtimes.com/news/blight-slugs-and-weather-threaten-potato-crop-1.1061622 (accessed 1.3.19).

Mallmann, W., Hemstreet, C., 1924. Isolation of an inhibitory substance from plants. Agric. Res. 28, 599-02.

Mann, R.A., Smits, T.H.M., Bühlmann, A., Blom, J., Goesmann, A., Frey, J.E., Plummer, K.M., Beer, S. V., Luck, J., Duffy, B., Rodoni, B., 2013. Comparative genomics of 12 strains of *Erwinia amylovora* identifies a pan-genome with a large conserved core. PLoS One 8, e55644. doi:10.1371/journal.pone.0055644

Mansfield, J., Genin, S., Magori, S., Citovsky, V., Sriariyanum, M., Ronald, P., Dow, M.A.X., Verdier, V., Beer, S. V, Machado, M.A., Toth, I.A.N., Salmond, G., Foster, G.D., Lipm, I.P., Tolosan, F.-C., 2012. Top 10 plant pathogenic bacteria in molecular plant pathology 13, 614–629. doi:10.1111/J.1364-3703.2012.00804.X

Mariano, R.L.R., Silveira, N.S.S., Michereff, S.J., 1998. Bacterial wilt in Brazil: Current status and control methods, in: Bacterial Wilt Disease. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 386–393. doi:10.1007/978-3-662-03592-4_59

Marín, A., Xia, X., 2008. GC skew in protein-coding genes between the leading and lagging strands in bacterial genomes: New substitution models incorporating strand bias. J. Theor. Biol. 253, 508–513. doi:10.1016/j.jtbi.2008.04.004

Marraffini, L.A., Sontheimer, E.J., 2008. CRISPR interference limits horizontal targeting DNA. Science (80-.). 322, 1843–1845.

Marrone, P.G., 2014. The market and potential for biopesticides, in: Gross, A.D., Coats, J.R., Duke, S.O., Seiber, J.N. (Eds.), Biopesticides: State of the Art and Future Opportunities, ACS Symposium Series. American Chemical Society, Washington, DC, pp. 245–258. doi:10.1021/bk-2014-1172

Mayerhofer, G., Schwaiger-Nemirova, I., Kuhn, T., Girsch, L., Allerberger, F., 2009. Detecting streptomycin in apples from orchards treated for fire blight. J. Antimicrob. Chemother. 63, 1076–7. doi:10.1093/jac/dkp055

McCarter, S.M., 1983. Survival of *Pseudomonas syringae* pv. tomato in association with tomato seed, soil, host tissue, and epiphytic weed hosts in Georgia. Phytopathology 73, 1393. doi:10.1094/Phyto-73-1393

McDonald, B.A., Linde, C., 2002. Pathogen population genetics , evolutionary potential, and durable resistance. Annu. Rev. Phytopathol. 40, 349–379.

doi:10.1146/annurev.phyto.40.120501.101443

McKenna, F., El-Tarabily, K.A., Hardy, G.E.S.J., Dell, B., 2001. Novel in vivo use of a polyvalent *Streptomyces* phage to disinfest *Streptomyces scabies*-infected seed potatoes. Plant Pathol. 50, 666–675. doi:10.1046/j.1365-3059.2001.00648.x

McManus, P.S., Stockwell, V.O., Sundin, G.W., Jones, A.L., 2002. Antibiotic use in plant agriculture. Annu. Rev. Phytopathol. 40, 443–465. doi:10.1146/annurev.phyto.40.120301.093927

Meczker, K., Dömötör, D., Vass, J., Rákhely, G., Schneider, G., Kovács, T., 2014. The genome of the *Erwinia amylovora* phage PhiEaH1 reveals greater diversity and broadens the applicability of phages for the treatment of fire blight. FEMS Microbiol. Lett. 350, 25–7.

Meier-Kolthoff, J.P., Auch, A.F., Klenk, H.-P., Göker, M., 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 14, 60. doi:10.1186/1471-2105-14-60

Meier-Kolthoff, J.P., Göker, M., 2017. VICTOR: genome-based phylogeny and classification of prokaryotic viruses. Bioinformatics 33, 3396–3404. doi:10.1093/bioinformatics/btx440

Meier-Kolthoff, J.P., Hahnke, R.L., Petersen, J., Scheuner, C., Michael, V., Fiebig, A., Rohde, C., Rohde, M., Fartmann, B., Goodwin, L.A., Chertkov, O., Reddy, T., Pati, A., Ivanova, N.N., Markowitz, V., Kyrpides, N.C., Woyke, T., Göker, M., Klenk, H.-P., 2014. Complete genome sequence of DSM 30083T, the type strain (U5/41T) of *Escherichia coli*, and a proposal for delineating subspecies in microbial taxonomy. Stand. Genomic Sci. 9, 2. doi:10.1186/1944-3277-9-2

Mesyanzhinov, V. V, Robben, J., Grymonprez, B., Kostyuchenko, V.A., Bourkaltseva, M. V, Sykilinda, N.N., Krylov, V.N., Volckaert, G., 2002. The genome of bacteriophage φKZ of *Pseudomonas aeruginosa*. J. Mol. Biol. 317, 1–19. doi:10.1006/jmbi.2001.5396

Mikiciński, A., Sobiczewski, P., Puławska, J., Maciorowski, R., 2016. Control of fire blight (*Erwinia amylovora*) by a novel strain 49M of *Pseudomonas graminis* from the phyllosphere of apple (Malus spp.). Eur. J. Plant Pathol. 145, 265–276. doi:10.1007/s10658-015-0837-y

Mitchell, A., Chang, H.-Y., Daugherty, L., Fraser, M., Hunter, S., Lopez, R., McAnulla, C., McMenamin, C., Nuka, G., Pesseat, S., Sangrador-Vegas, A., Scheremetjew, M., Rato, C., Yong, S.-Y., Bateman, A., Punta, M., Attwood, T.K., Sigrist, C.J.A.A., Redaschi, N., Rivoire, C., Xenarios, I., Kahn, D., Guyot, D., Bork, P., Letunic, I., Gough, J., Oates, M., Haft, D., Huang, H., Natale, D.A., Wu, C.H., Orengo, C., Sillitoe, I., Mi, H., Thomas, P.D., Finn, R.D., 2014. The InterPro protein families database: the classification resource after 15 years. Nucleic Acids Res. 43, D213-21. doi:10.1093/nar/gku1243

Moak, M., Molineux, I.J., 2000. Role of the Gp16 lytic transglycosylase motif in bacteriophage T7 virions at the initiation of infection. Mol. Microbiol. 37, 345–355. doi:10.1046/j.1365-2958.2000.01995.x

Molina, A.M., Pesce, A., Shito, G.C., 1965. Un nuovo batteriofago attivo sul ceppo K12 di E. coli. I. Caratteristiche biologiche. Boll. Ist. Sieroter. 44, 329–337.

Müller, I., Lurz, R., Geider, K., 2012. Tasmancin and lysogenic bacteriophages induced from *Erwinia tasmaniensis* strains. Microbiol. Res. 167, 381–387. doi:10.1016/j.micres.2012.01.005

Murphy, J., Mahony, J., Ainsworth, S., Nauta, A., van Sinderen, D., 2013. Bacteriophage orphan

DNA methyltransferases: Insights from their bacterial origin, function, and occurrence. Appl. Environ. Microbiol. 79, 7547–7555. doi:10.1128/AEM.02229-13

Naville, M., Ghuillot-Gaudeffroy, A., Marchais, A., Gautheret, D., 2011. ARNold: a web tool for the prediction of Rho-independent transcription terminators. RNA Biol. 8, 11–3.

Nelson, S.W., Zhuang, Z., Spiering, M.M., Benkovic, S.J., 2009. T4 Phage Replisome, in: Viral Genome Replication. Springer US, Boston, MA, pp. 337–364. doi:10.1007/b135974_16

Nikolaichik, Y., Gorshkov, V., Gogolev, Y., Valentovich, L., Evtushenkov, A., 2014. Genome sequence of *Pectobacterium atrosepticum* strain 21A. Genome Announc. 2. doi:10.1128/genomeA.00935-14

Nischwitz, C., Dhiman, C., 2013. Streptomycin resistance of *Erwinia amylovora* isolated from apple in Utah [WWW Document]. Plant Heal. Prog. doi:10.1094/PHP-2013-1025-01-RS

Norelli, J.L., Jones, A.L., Aldwinckle, H.S., 2003. Fire blight management in the twenty first century using new technologies. Plant Dis 87, 756–765.

Nowicki, G., Walkowiak-Nowicka, K., Zemleduch-Barylska, A., Mleczko, A., Frąckowiak, P., Nowaczyk, N., Kozdrowska, E., Barylski, J., 2017. Complete genome sequences of two novel autographiviruses infecting a bacterium from the *Pseudomonas fluorescens* group. Arch. Virol. 162, 2907–2911. doi:10.1007/s00705-017-3419-9

O'Flynn, G., Ross, R.P., Fitzgerald, G.F., Coffey, A., 2004. Evaluation of a cocktail of three bacteriophages for biocontrol of *Escherichia coli* O157:H7. Appl. Environ. Microbiol. 70, 3417–24. doi:10.1128/AEM.70.6.3417-3424.2004

Obradovic, A., Jones, J.B., Momol, M.T., Balogh, B., Olson, S.M., 2004. management of tomato bacterial spot in the field by foliar applications of bacteriophages and SAR inducers. Plant Dis. 88, 736–740. doi:10.1094/PDIS.2004.88.7.736

Ohmori, H., Haynes, L.L., Rothman-Denes, L.B., 1988. Structure of the ends of the coliphage N4 genome. J. Mol. Biol. 202, 1–10.

Okabe, N., Goto, M., 1963. Bacteriophages of Plant Pathogens. Annu. Rev. Phytopathol. 1, 397–418. doi:10.1146/annurev.py.01.090163.002145

Oliveira, H., Melo, L.D.R., Santos, S.B., Nóbrega, F.L., Ferreira, E.C., Cerca, N., Azeredo, J., Kluskens, L.D., 2013. Molecular aspects and comparative genomics of bacteriophage endolysins. J. Virol. 87, 4558–70. doi:10.1128/JVI.03277-12

Oliveira, H., São-José, C., Azeredo, J., 2018. Phage-Derived Peptidoglycan Degrading Enzymes: Challenges and Future Prospects for In Vivo Therapy. Viruses 10. doi:10.3390/v10060292

OmniLytics, 2006. OmniLytics Receives OMRI Listing for AgriPhage | OmniLytics | News [WWW Document]. URL http://www.omnilytics.com/news/news015.html (accessed 3.23.16).

OmniLytics, 2004. AgriPhage Product Overview | OmniLytics [WWW Document]. URL http://www.ns.omnilytics.com/products/agriphage/agriphage_info/agriphage_overview.html

Ordax, M., Marco-Noales, E., López, M.M., Biosca, E.G., 2006. Survival strategy of *Erwinia amylovora* against copper: induction of the viable-but-nonculturable state. Appl. Environ. Microbiol. 72, 3482–8. doi:10.1128/AEM.72.5.3482-3488.2006

Paddison, P., Abedon, S.T., Dressman, H.K., Gailbreath, K., Tracy, J., Mosser, E., Neitzel, J., Guttman, B., Kutter, E., 1998. The roles of the bacteriophage T4 r genes in lysis inhibition

and fine-structure genetics: a new perspective. Genetics 148, 1539–50.

Pang, T., Savva, C.G., Fleming, K.G., Struck, D.K., Young, R., 2009. Structure of the lethal phage pinhole. Proc. Natl. Acad. Sci. 106, 18966–18971. doi:10.1073/pnas.0907941106

Park, T.-H., Choi, B.-S., Choi, A.-Y., Choi, I.-Y., Heu, S., Park, B.-S., 2012. Genome sequence of *Pectobacterium carotovorum* subsp. *carotovorum* strain PCC21, a pathogen causing soft rot in Chinese cabbage. J. Bacteriol. 194, 6345–6. doi:10.1128/JB.01583-12

Parkinson, N., Bryant, R., Bew, J., Elphinstone, J., 2011. Rapid phylogenetic identification of members of the *Pseudomonas syringae* species complex using the rpoD locus. Plant Pathol. 60, 338–344. doi:10.1111/j.1365-3059.2010.02366.x

Parma, D.H., Snyder, M., Sobolevski, S., Nawroz, M., Brody, E., Gold, L., 1992. The rex system of bacteriophage-Lambda - tolerance and altruistic cell-death. Genes Dev. 6, 497–510. doi:10.1101/gad.6.3.497

Pérombelon, M.C.M., 2002. Potato diseases caused by soft rot erwinias: an overview of pathogenesis. Plant Pathol. 51, 1–12. doi:10.1046/j.0032-0862.2001.Shorttitle.doc.x

Perombelon, M.C.M., Van Der Wolf, J.M., 2002. Methods for the detection and quantification of *Erwinia carotovora* subsp. *atroseptica* (*Pectobacterium carotovorum* subsp. *atrosepticum*) on potatoes: a laboratory manual, Scottish Crop Research Institute Occasional Publication No. 10. Scottish Crop Research Institute Occasional Publication, Invergowrie, Dundee, Scotland.

Petrov, V.M., Ratnayaka, S., Nolan, J.M., Miller, E.S., Karam, J.D., 2010. Genomes of the T4-related bacteriophages as windows on microbial genome evolution. Virol. J. 7, 292. doi:10.1186/1743-422X-7-292

Pfreundt, U., Spungin, D., Hou, S., Voß, B., Berman-Frank, I., Hess, W.R., 2016. Genome of a giant bacteriophage from a decaying *Trichodesmium* bloom. Mar. Genomics. doi:10.1016/j.margen.2017.02.001

Pickard, D.J.J., 2009. Preparation of bacteriophage lysates and pure DNA., in: Clokie, M., Kropinski, A.M. (Eds.), Bacteriophages: Methods and Protocols, Volume 1: Isolation, Chatacterization, and Interactions. Humana Press, pp. 3–9.

Pietrzak, U., McPhail, D.C., 2004. Copper accumulation, distribution and fractionation in vineyard soils of Victoria, Australia. Geoderma 122, 151–166. doi:10.1016/j.geoderma.2004.01.005

Piqué, N., Miñana-Galbis, D., Merino, S., Tomás, J.M., 2015. Virulence Factors of *Erwinia amylovora*: A Review. Int. J. Mol. Sci. 16, 12836–12854. doi:10.3390/ijms160612836

Pirhonen, M., Heino, P., Helander, I., Harju, P., Palva, E.T., 1988. Bacteriophage T4 resistant mutants of the plant pathogen *Erwinia carotovora*. Microb. Pathog. 4, 359–367. doi:10.1016/0882-4010(88)90063-0

Pirhonen, M., Palva, E.T., 1988. Occurrence of bacteriophage T4 receptor in *Erwinia carotovora*. MGG Mol. Gen. Genet. 214, 170–172. doi:10.1007/BF00340198

Pope, W.H., Jacobs-Sera, D., Best, A.A., Broussard, G.W., Connerly, P.L., Dedrick, R.M., Kremer, T.A., Offner, S., Ogiefo, A.H., Pizzorno, M.C., Rockenbach, K., Russell, D.A., Stowe, E.L., Stukey, J., Thibault, S.A., Conway, J.F., Hendrix, R.W., Hatfull, G.F., 2013. Cluster J Mycobacteriophages: Intron splicing in capsid and tail genes. PLoS One 8, e69273. doi:10.1371/journal.pone.0069273

Powney, R., Smits, T.H.M., Sawbridge, T., Frey, B., Blom, J., Frey, J.E., Plummer, K.M., Beer, S. V, Luck, J., Duffy, B., Rodoni, B., 2011. Genome sequence of an *Erwinia amylovora* strain with pathogenicity restricted to *Rubus* plants. J. Bacteriol. 193, 785–6. doi:10.1128/JB.01352-10

Preston, G.M., 2000. *Pseudomonas syringae* pv. *tomato*: the right pathogen, of the right plant, at the right time. Mol. Plant Pathol. 1, 263–275.

Prins, H., Breukers, A., 2008. In de puree? De gevolgen van aantasting door *Erwinia* voor de pootaardappelsector in kaart gebracht, In. LEI Report Den Haag, The Netherlands.

Prokhorov, N.S., Riccio, C., Zdorovenko, E.L., Shneider, M.M., Browning, C., Knirel, Y.A., Leiman, P.G., Letarov, A. V., 2017. Function of bacteriophage G7C esterase tailspike in host cell adsorption. Mol. Microbiol. 105, 385–398. doi:10.1111/mmi.13710

Przybilski, R., Richter, C., Gristwood, T., Clulow, J.S., Vercoe, R.B., Fineran, P.C., 2011. Csy4 is responsible for CRISPR RNA processing in *Pectobacterium atrosepticum*. RNA Biol. 8, 517–528. doi:10.4161/rna.8.3.15190

Qiao, J., Qiao, X., Sun, Y., Mindich, L., 2010. Role of host protein glutaredoxin 3 in the control of transcription during bacteriophage Phi2954 infection. Proc. Natl. Acad. Sci. U. S. A. 107, 6000–4. doi:10.1073/pnas.1000383107

Rambaut, A., 2006. FigTreeFigTree 1.4.3 - a graphical viewer of phylogenetic trees and a program for producing publication-ready figures [WWW Document]. URL http://tree.bio.ed.ac.uk/software/figtree/ (accessed 5.2.18).

Rao, Y.P., Srivastava, D.N., 1973. Application of phages in investigation of epidemiology of bacterial blight disease of rice. In S. P. Raychandhari (Ed.), in: Proceedings of the Indian National Science Academy: Epidemiology, Forecasting and Control of Plant Diseases, 37. pp. 314–321.

Ravensdale, M., Blom, T.J., Gracia-Garza, J.A., Svircev, A.M., Smith, R.J., 2010. Bacteriophages and the control of *Erwinia carotovora* subsp. *carotovora*. Can. J. Plant Pathol. 29, 121–130.

Ray, D.K., Mueller, N.D., West, P.C., Foley, J.A., 2013. Yield trends are insufficient to double global crop production by 2050. PLoS One 8, e66428. doi:10.1371/journal.pone.0066428

Redzuan, R.A., Abu Bakar, N., Rozano, L., Badrun, R., Mat Amin, N., Mohd Raih, M.F., 2014. Draft genome sequence of Erwinia mallotivora BT-MARDI, causative agent of Papaya dieback disease. Genome Announc. 2, e00375-14. doi:10.1128/genomeA.00375-14

Rezzonico, F., Smits, T.H.M., Duffy, B., 2011. Diversity, evolution, and functionality of clustered regularly interspaced short palindromic repeat (CRISPR) regions in the fire blight pathogen *Erwinia amylovora*. Appl. Environ. Microbiol. 77, 3819–29. doi:10.1128/AEM.00177-11

Ritchie, D.F., Dittapongpitch, V., 1991. Copper- and streptomycin-resistant strains and host differentiated races of *Xanthomonas campestris* pv. *vesicatoria* in North Carolina. Plant Dis. 75, 733–736. doi:10.1094/PD-75-0733

Rocha, E.P.C., Danchin, A., 2002. Base composition bias might result from competition for metabolic resources. Trends Genet. 18, 291–294. doi:10.1016/S0168-9525(02)02690-2

Rodrigues, C.M., Takita, M.A., Coletta-Filho, H.D., Olivato, J.C., Caserta, R., Machado, M.A., de Souza, A.A., 2008. Copper resistance of biofilm cells of the plant pathogen *Xylella*

fastidiosa. Appl. Microbiol. Biotechnol. 77, 1145–57. doi:10.1007/s00253-007-1232-1

Rodríguez-Rubio, L., Martínez, B., Donovan, D.M., Rodríguez, A., García, P., 2013. Bacteriophage virion-associated peptidoglycan hydrolases: potential new enzybiotics. Crit. Rev. Microbiol. 39, 427–434. doi:10.3109/1040841X.2012.723675

Rombouts, S., Volckaert, A., Venneman, S., Declercq, B., Vandenheuvel, D., Allonsius, C.N., Van Malderghem, C., Jang, H.B., Briers, Y., Noben, J.P., Klumpp, J., Van Vaerenbergh, J., Maes, M., Lavigne, R., 2016. Characterization of novel bacteriophages for biocontrol of bacterial blight in Leek caused by *Pseudomonas syringae* pv. *porri*. Front. Microbiol. 7, 279. doi:10.3389/fmicb.2016.00279

Roucourt, B., Lavigne, R., 2009. The role of interactions between phage and bacterial proteins within the infected cell: a diverse and puzzling interactome. Environ. Microbiol. 11, 2789–2805. doi:10.1111/j.1462-2920.2009.02029.x

Russo, N.L., Burr, T.J., Breth, D.I., Aldwinckle, H.S., 2008. Isolation of Streptomycin-resistant isolates of *Erwinia amylovora* in New York. Plant Dis. 92, 714–718. doi:10.1094/PDIS-92-5-0714

Ryan, R.P., Vorhölter, F.-J., Potnis, N., Jones, J.B., Van Sluys, M.-A., Bogdanove, A.J., Dow, J.M., 2011. Pathogenomics of *Xanthomonas*: understanding bacterium–plant interactions. Nat. Rev. Microbiol. 9, 344–355. doi:10.1038/nrmicro2558

Sambrook, J., Russell, D.W., 2001a. Preparing stocks of bacteriophage λ by plate lysis and elution., in: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, p. 2.34.

Sambrook, J., Russell, D.W., 2001b. Plating bacteriophage λ., in: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, p. 2.25.

Sambrook, J., Russell, D.W., 2001c. Picking bacteriophage λ plaques., in: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, p. 2.32.

Sambrook, J., Russell, D.W., 2001d. Purification of bacteriophage λ particles by isopycnic centrifugation through CsCl gradients., in: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, p. 2.47.

Sambrook, J., Russell, D.W., 2001e. Extraction of bacteriophage λ DNA from large-scale cultures using proteinase K and SDS., in: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, p. 2.56.

Sambrook, Russell, Sambrook, J., Russell, D.W., 2001. SDS-polyacrylamide gel electrophoresis of proteins, in: Molecular Cloning: A Laboratory Manual, Volume 1. Cold Spring Harbor Laboratory Press, p. 8.40-51.

Samson, J.E., Magadán, A.H., Sabri, M., Moineau, S., 2013. Revenge of the phages: defeating bacterial defences. Nat. Rev. Microbiol. 11, 675–687. doi:10.1038/nrmicro3096

Samson, R., Legendre, J.B., Christen, R., Fischer-Le Saux, M., Achouak, W., Gardan, L., 2005. Transfer of *Pectobacterium chrysanthemi* (Burkholder et al. 1953) Brenner et al. 1973 and *Brenneria paradisiaca* to the genus *Dickeya* gen. nov. as *Dickeya chrysanthemi* comb. nov. and *Dickeya paradisiaca* comb. nov. and deli. Int. J. Syst. Evol. Microbiol. 55, 1415–27. doi:10.1099/ijs.0.02791-0

Sanchez Perez, A., Mejia, L., Fegan, M., Allen, C., 2008. Diversity and distribution of *Ralstonia solanacearum* strains in Guatemala and rare occurrence of tomato fruit infection. Plant

Pathol. 57, 320–331. doi:10.1111/j.1365-3059.2007.01769.x

Sandt, C.H., Hopper, J.E., Hill, C.W., 2002. Activation of Prophage eib Genes for Immunoglobulin-Binding Proteins by Genes from the IbrAB Genetic Island of *Escherichia coli* ECOR-9. J. Bacteriol. 184, 3640–3648. doi:10.1128/JB.184.13.3640-3648.2002

Santos, S.B., Kropinski, A.M., Ceyssens, P.-J., Ackermann, H.-W., Villegas, A., Lavigne, R., Krylov, V.N., Carvalho, C.M., Ferreira, E.C., Azeredo, J., 2011. Genomic and proteomic characterization of the broad-host-range *Salmonella* phage PVP-SE1: Creation of a new phage genus. J. Virol. 85, 11265–11273. doi:10.1128/JVI.01769-10

Sato, M., 1983. Phage induction from lysogenic strains of *Pseudomonas syringae* pathovar mori by the extract from mulberry leaves. An. Phytopathol. Soc. Japan 49, 259–261.

Savva, C.G., Dewey, J.S., Moussa, S.H., To, K.H., Holzenburg, A., Young, R., 2014. Stable micron-scale holes are a general feature of canonical holins. Mol. Microbiol. 91, 57–65. doi:10.1111/mmi.12439

Scherf, J.M., Milling, A., Allen, C., 2010. Moderate temperature fluctuations rapidly reduce the viability of *Ralstonia solanacearum* race 3, biovar 2, in infected geranium, tomato, and potato plants. Appl. Environ. Microbiol. 76, 7061–7. doi:10.1128/AEM.01580-10

Schiffenbauer, M., Stotzky, G., 1982. Adsorption of coliphages T1 and T7 to clay minerals. Appl. Environ. Microbiol. 43, 590–6.

Schlegel, S., Löfblom, J., Lee, C., Hjelm, A., Klepsch, M., Strous, M., Drew, D., Slotboom, D.J., de Gier, J.-W., 2012. Optimizing membrane Protein overexpression in the *Escherichia coli* strain Lemo21(DE3). J. Mol. Biol. 423, 648–659. doi:10.1016/J.JMB.2012.07.019

Schofield, D., Bull, C.T., Rubio, I., Wechter, W.P., Westwater, C., Molineux, I.J., 2013. "Light-tagged" bacteriophage as a diagnostic tool for the detection of phytopathogens. Bioengineered 4, 50–4. doi:10.4161/bioe.22159

Schoonejans, E., Expert, D., Toussaint, A., 1987. Characterization and virulence properties of *Erwinia chrysanthemi* lipopolysaccharide-defective, phi EC2-resistant mutants. J. Bacteriol. 169, 4011–7.

Schweiger, M., Hennig, K., Lerner, F., Niere, M., Hirsch-Kauffmann, M., Specht, T., Weise, C., Oei, S.L., Ziegler, M., 2001. Characterization of recombinant human nicotinamide mononucleotide adenylyl transferase (NMNAT), a nuclear enzyme essential for NAD synthesis. FEBS Lett. 492, 95–100. doi:10.1016/S0014-5793(01)02180-9

Semenova, E., Nagornykh, M., Pyatnitskiy, M., Artamonova, I.I., Severinov, K., 2009. Analysis of CRISPR system function in plant pathogen *Xanthomonas oryzae*. FEMS Microbiol. Lett. 296, 110–6. doi:10.1111/j.1574-6968.2009.01626.x

Sengupta, R., Holmgren, A., 2014. Thioredoxin and glutaredoxin-mediated redox regulation of ribonucleotide reductase. World J. Biol. Chem. 5, 68–74. doi:10.4331/wjbc.v5.i1.68

Serwer, P., Hayes, S.J., Thomas, J. a, Hardies, S.C., 2007. Propagating the missing bacteriophages: a large bacteriophage in a new class. Virol. J. 4, 21. doi:10.1186/1743-422X-4-21

Shcherbakov, V.P., Plugina, L., Shcherbakova, T., 2011. Endonuclease VII is a key component of the mismatch repair mechanism in bacteriophage T4. DNA Repair (Amst). 10, 356–362. doi:10.1016/j.dnarep.2010.12.006

Shigehisa, R., Uchiyama, J., Kato, S., Takemura-Uchiyama, I., Yamaguchi, K., Miyata, R., Ujihara,

T., Sakaguchi, Y., Okamoto, N., Shimakura, H., Daibata, M., Sakaguchi, M., Matsuzaki, S., 2016. Characterization of *Pseudomonas aeruginosa* phage KPP21 belonging to family *Podoviridae* genus N4-like viruses isolated in Japan. Microbiol. Immunol. 60, 64–67. doi:10.1111/1348-0421.12347

Shin, H., Lee, J.-H., Kim, Y., Ryu, S., 2012. Complete genome sequence of *Cronobacter sakazakii* bacteriophage CR3. J. Virol. 86, 6367–8. doi:10.1128/JVI.00636-12

Šimoliūnas, E., Kaliniene, L., Truncaite, L., Zajančkauskaite, A., Staniulis, J., Kaupinis, A., Ger, M., Valius, M., Meškys, R., 2013. *Klebsiella* phage vB_KleM-RaK2 — A giant singleton virus of the family *Myoviridae*. PLoS One 8. doi:10.1371/journal.pone.0060717

Skurnik, M., Hyytiäinen, H.J., Happonen, L.J., Kiljunen, S., Datta, N., Mattinen, L., Williamson, K., Kristo, P., Szeliga, M., Kalin-Mänttäri, L., Ahola-Iivarinen, E., Kalkkinen, N., Butcher, S.J., 2012. Characterization of the genome, proteome, and structure of yersiniophage φR1-37. J. Virol. 86, 12625–42. doi:10.1128/JVI.01783-12

Smitley, D.R., McCarter, S.M., 1982. Spread of *Pseudomonas syringae* pv. tomato and role of epiphytic populations and environmental conditions in disease development. Plant Dis. 66, 713. doi:10.1094/PD-66-713

Smolarska, A., Rabalski, L., Narajczyk, M., Czajkowski, R., 2018. Isolation and phenotypic and morphological characterization of the first *Podoviridae* lytic bacteriophages φA38 and φA41 infecting *Pectobacterium parmentieri* (former *Pectobacterium wasabiae*). Eur. J. Plant Pathol. 150, 413–425. doi:10.1007/s10658-017-1289-3

Söding, J., Biegert, A., Lupas, A.N., 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. 33, W244-8. doi:10.1093/nar/gki408

Stoddard, B.L., 2005. Homing endonuclease structure and function. Q. Rev. Biophys. 38, 49–95. doi:10.1017/S0033583505004063

Stojković, E.A., Rothman-Denes, L.B., 2007. Coliphage N4 N-acetylmuramidase defines a new family of murein hydrolases. J. Mol. Biol. 366, 406–419. doi:10.1016/j.jmb.2006.11.028

Strange, R.N., Scott, P.R., 2005. Plant disease: A threat to global food security. Annu. Rev. Phytopathol. 43, 83–116. doi:10.1146/annurev.phyto.43.113004.133839

Straub, T.M., Pepper, I.L., Gerba, C.P., 1992. Persistence of viruses in desert soils amended with anaerobically digested sewage sludge. Appl. Environ. Microbiol. 58, 636–41.

Sulakvelidze, A., Alavidze, Z., Morris, J.G., 2001. Bacteriophage therapy. Antimicrob. Agents Chemother. 45, 649–59. doi:10.1128/AAC.45.3.649-659.2001

Sullivan, M.B., Coleman, M.L., Quinlivan, V., Rosenkrantz, J.E., Defrancesco, A.S., Tan, G., Fu, R., Lee, J.A., Waterbury, J.B., Bielawski, J.P., Chisholm, S.W., 2008. Portal protein diversity and phage ecology. Environ. Microbiol. 10, 2810–23. doi:10.1111/j.1462-2920.2008.01702.x

Sullivan, M.J., Petty, N.K., Beatson, S.A., 2011. Easyfig: a genome comparison visualizer. Bioinformatics 27, 1009–1010. doi:10.1093/bioinformatics/btr039

Summer, E.J., Berry, J., Tran, T.A.T., Niu, L., Struck, D.K., Young, R., 2007. Rz/Rz1 lysis gene equivalents in phages of Gram-negative hosts. J. Mol. Biol. 373, 1098–1112. doi:10.1016/J.JMB.2007.08.045

Suttle, C.A., 1994. The significance of viruses to mortality in aquatic microbial communities.

Microb. Ecol. 28, 237–43. doi:10.1007/BF00166813

Svircev, A.M., Lehman, S.M., Kim, W.S., Barszcz, E., Schneider, K.E., Castle, A.J., Zeller, W., Ullrich, C., 2006. Control of the fire blight pathogen with bacteriophages., in: Proceedings of the 1st International Symposium on Biological Control of Bacterial Plant Diseases, Seeheim, Darmstadt, Germany, 23-26 October, 2005. Biologische Bundesanstalt für Land- und Forstwirtschaft, pp. 259–261.

Sykes, I.K., Lanning, S., Williams, S.T., 1981. The effect of pH on soil actinophage. Microbiology 122, 271–280. doi:10.1099/00221287-122-2-271

Sykilinda, N.N., Bondar, A.A., Gorshkova, A.S., Kurochkina, L.P., Kulikov, E.E., Shneider, M.M., Kadykov, V.A., Solovjeva, N. V., Kabilov, M.R., Mesyanzhinov, V. V., Vlassov, V. V., Drukker, V. V., Miroshnikov, K.A., 2014. Complete genome sequence of the novel giant *Pseudomonas* phage PaBG. Genome Announc. 2, e00929-13-e00929-13. doi:10.1128/genomeA.00929-13

Szewczyk, B., Bienkowska-Szewczyk, K., Kozloff, L.M., 1986. Identification of T4 gene 25 product, a component of the tail baseplate, as a 15K lysozyme. Mol. Gen. Genet. 202, 363–7.

Tanaka, H., Negishi, H., Maeda, H., 1990. Control of tobacco bacterial wilt by an avirulent strain of *Pseudomonas solanacearum* M4S and its bacteriophage. Japanese J. Phytopathol. 56, 243–246. doi:10.3186/jjphytopath.56.243

Tancos, K.A., Villani, S., Kuehne, S., Borejsza-Wysocka, E., Breth, D., Carol, J., Aldwinckle, H.S., Cox, K.D., 2016. Prevalence of streptomycin-resistant *Erwinia amylovora* in New York apple orchards. Plant Dis. 100, 802–809.

Thieme, F., Koebnik, R., Bekel, T., Berger, C., Boch, J., Büttner, D., Caldana, C., Gaigalat, L., Goesmann, A., Kay, S., Kirchner, O., Lanz, C., Linke, B., McHardy, A.C., Meyer, F., Mittenhuber, G., Nies, D.H., Niesbach-Klösgen, U., Patschkowski, T., Rückert, C., Rupp, O., Schneiker, S., Schuster, S.C., Vorhölter, F.-J., Weber, E., Pühler, A., Bonas, U., Bartels, D., Kaiser, O., 2005. Insights into genome plasticity and pathogenicity of the plant pathogenic bacterium *Xanthomonas campestris* pv. *vesicatoria* revealed by the complete genome sequence. J. Bacteriol. 187, 7254–66. doi:10.1128/JB.187.21.7254-7266.2005

Thierauf, A., Perez, G., Maloy, and S., 2009. Generalized transduction. Humana Press, pp. 267–286. doi:10.1007/978-1-60327-164-6_23

Thomas, R., 1935. A bacteriophage in relation to Stewart's disease of corn. Phytopathology 25, 371–2.

Tori, K., Perler, F.B., 2011. Expanding the definition of class 3 inteins and their proposed phage origin. J. Bacteriol. 193, 2035–2041. doi:10.1128/JB.01407-10

Toth, I.K., Bell, K.S., Holeva, M.C., Birch, P.R.J., 2003. Soft rot erwiniae: from genes to genomes. Mol. Plant Pathol. 4, 17–30. doi:10.1046/j.1364-3703.2003.00149.x

Toth, I.K., Bertheau, Y., Hyman, L.J., Laplaze, L., López, M.M., McNicol, J., Niepold, F., Persson, P., Salmond, G.P., Sletten, A., van Der Wolf, J.M., Pérombelon, M.C., 1999. Evaluation of phenotypic and molecular typing techniques for determining diversity in *Erwinia carotovora* subsp. *atroseptica*. J. Appl. Microbiol. 87, 770–81.

Toth, I.K., Mulholland, V., Cooper, V., Bentley, S., Shih, Y., Perombelon, M.C.M., Salmond, G.P.C., 1997. Generalized transduction in the potato blackleg pathogen *Ewinia carotovora* subsp . *atroseptica* by bacteriophage φM1. Microbiology 143, 2433–2438.

Toth, I.K., van der Wolf, J.M., Saddler, G., Lojkowska, E., Hélias, V., Pirhonen, M., Tsror Lahkim, L., Elphinstone, J.G., 2011. *Dickeya* species: an emerging problem for potato production in Europe. Plant Pathol. 60, 385–399. doi:10.1111/j.1365-3059.2011.02427.x

Trotter, M., McAuliffe, O., Callanan, M., Edwards, R., Fitzgerald, G.F., Coffey, A., Ross, R.P., 2006. Genome analysis of the obligately lytic bacteriophage 4268 of *Lactococcus lactis* provides insight into its adaptable nature provides insight into its adaptable nature provides insight into its adaptable nature. Gene 366, 189–199. doi:10.1016/j.gene.2005.09.022

Tsones, J., 2014. Use of bacteriophages as antimicrobials against avian pathogenic *Escherichia coli* infections in poultry : from isolation to therapeutic. KU Leuven and Vrije Universiteit Brussel.

Tsror (Lahkim), L., Erlich, O., Lebiush, S., Hazanovsky, M., Zig, U., Slawiak, M., Grabe, G., van der Wolf, J.M., van de Haar, J.J., 2008. Assessment of recent outbreaks of *Dickeya* sp. (syn. *Erwinia chrysanthemi*) slow wilt in potato crops in Israel. Eur. J. Plant Pathol. 123, 311. doi:10.1007/s10658-008-9368-0

Tumber, K.P., Alston, J.M., Fuller, K.B., 2014. Pierce's disease costs California $104 million per year. Calif. Agric 68, 20–29.

Turner, D., Reynolds, D., Seto, D., Mahadevan, P., 2013. CoreGenes3.5: a webserver for the determination of core genes from sets of viral and small bacterial genomes. BMC Res. Notes 6, 140. doi:10.1186/1756-0500-6-140

Twort, F.W., 1915. An investigation on the nature of ultramicroscopic viruses. Lancet ii:1241.

UN, 2013. World must sustainably produce 70 per cent more food by mid-century – UN report [WWW Document]. URL http://www.un.org/apps/news/story.asp?NewsID=46647#.Vvxj0uIrLIU (accessed 3.30.16).

Van den Bossche, A., Ceyssens, P.-J., De Smet, J., Hendrix, H., Bellon, H., Leimer, N., Wagemans, J., Delattre, A.-S., Cenens, W., Aertsen, A., Landuyt, B., Minakhin, L., Severinov, K., Noben, J.-P., Lavigne, R., 2014. Systematic identification of hypothetical bacteriophage proteins targeting key protein complexes of *Pseudomonas aeruginosa*. J. Proteome Res. 13, 4446–4456. doi:10.1021/pr500796n

van der Merwe, R.G., van Helden, P.D., Warren, R.M., Sampson, S.L., Gey van Pittius, N.C., 2014. Phage-based detection of bacterial pathogens. Analyst 139, 2617–26. doi:10.1039/c4an00208c

van der Wolf, J.M., de Haan, E.G., Kastelein, P., Krijger, M., de Haas, B.H., Velvis, H., Mendes, O., Kooman-Gersmann, M., van der Zouwen, P.S., 2017. Virulence of *Pectobacterium carotovorum* subsp. *brasiliense* on potato compared with that of other *Pectobacterium* and *Dickeya* species under climatic conditions prevailing in the Netherlands. Plant Pathol. 66, 571–583. doi:10.1111/ppa.12600

Vanneste, J.L., Eden-Green, S., 2000. Migration of *Erwinia amylovora* in host plant tissues, in: Vanneste, J.L. (Ed.), Fire Blight: The Disease and Its Causative Agent. *Erwinia Amylovora*. CABI, p. 73.

Waleron, M., Waleron, K., Lojkowska, E., 2013. Occurrence of *Pectobacterium wasabiae* in potato field samples. Eur. J. Plant Pathol. 137, 149–158. doi:10.1007/s10658-013-0227-2

Walldén, K., Rinaldo-Matthis, A., Ruzzenente, B., Rampazzo, C., Bianchi, V., Nordlund, P., 2007.

Crystal structures of human and murine deoxyribonucleotidases: insights into recognition of substrates and nucleotide analogues. Biochemistry 46, 13809–18. doi:10.1021/bi7014794

Wang, J.-F., Olivier, J., Thoquet, P., Mangin, B., Sauviac, L., Grimsley, N.H., 2000. Resistance of Tomato Line Hawaii7996 to *Ralstonia solanacearum* Pss4 in Taiwan Is Controlled Mainly by a Major Strain-Specific Locus. Mol. Plant-Microbe Interact. 13, 6–13. doi:10.1094/MPMI.2000.13.1.6

Wang, L.K., Schwer, B., Shuman, S., 2006. Structure-guided mutational analysis of T4 RNA ligase 1. RNA 12, 2126–34. doi:10.1261/rna.271706

Whelan, S., Goldman, N., 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol. Biol. Evol. 18, 691–9.

Whitman, W.B., Coleman, D.C., Wiebe, W.J., 1998. Perspective Prokaryotes: The unseen majority. Proc Natl Acad Sci U S A 95, 6578–6583.

Wilkinson, L., 2001. Félix d'Herelle and the origins of molecular biology, Medical History. Cambridge University Press.

Wittmann, J., Grose, J.H., Yagubi, A.I., Svircev, A.M., Kropinski, A.M., 2016a. To create a new genus, *Ea92virus*, including 2 (two) new species within the family *Podoviridae* [WWW Document]. URL https://talk.ictvonline.org/files/ictv_official_taxonomy_updates_since_the_8th_report/m/prokaryote-official/6772 (accessed 1.22.18).

Wittmann, J., Klumpp, J., Moreno Switt, A.I., Yagubi, A., Ackermann, H.-W., Wiedmann, M., Svircev, A., Nash, J.H.E., Kropinski, A.M., 2015. Taxonomic reassessment of N4-like viruses using comparative genomics and proteomics suggests a new subfamily - "*Enquartavirinae*." Arch. Virol. 160, 3053–3062. doi:10.1007/s00705-015-2609-6

Wittmann, J., Kropinski, A.M., Adriaenssens, E.M., Ackermann, H.-W., Lavigne, R., Kuhn, J.H., Uchiyama, J., 2016b. To create a new genus, Luz7virus, including 2 (two) new species within the family *Podoviridae* [WWW Document]. doi:10.13140/RG.2.2.14877.51689

Wommack, K.E., Colwell, R.R., 2000. Virioplankton: viruses in aquatic ecosystems. Microbiol. Mol. Biol. Rev. 64, 69–114.

Woods, T.L., Israel, H.W., Sherf, A.F., 1981. Isolation and partial characterization of a bacteriophage of *Erwinia stewartii* from the corn flea beetle, *Chaetocnema pulicaria*. Prot. Ecol. 3, 229–236.

Xin, X.-F., He, S.Y., 2013. *Pseudomonas syringae* pv. *tomato* DC3000: a model pathogen for probing disease susceptibility and hormone signaling in plants. Annu. Rev. Phytopathol. 51, 473–98. doi:10.1146/annurev-phyto-082712-102321

Xu, M., Struck, D.K., Deaton, J., Wang, I.-N., Young, R., 2004. A signal-arrest-release sequence mediates export and control of the phage P1 endolysin. Proc. Natl. Acad. Sci. 101, 6415–6420. doi:10.1073/pnas.0400957101

Yagubi, A.I., Castle, A.J., Kropinski, A.M., Banks, T.W., Svircev, A.M., 2014. Complete Genome Sequence of *Erwinia amylovora* Bacteriophage vB_EamM_Ea35-70. Genome Announc. 2, e00413-14. doi:10.1128/genomeA.00413-14

Yamada, T., 2013. Filamentous phages of *Ralstonia solanacearum*: double-edged swords for

pathogenic bacteria. Front. Microbiol. 4, 325. doi:10.3389/fmicb.2013.00325

Yamada, T., Satoh, S., Ishikawa, H., Fujiwara, A., Kawasaki, T., Fujie, M., Ogata, H., 2010. A jumbo phage infecting the phytopathogen *Ralstonia solanacearum* defines a new lineage of the *Myoviridae* family. Virology 398, 135–147. doi:10.1016/j.virol.2009.11.043

Yamamoto, N., Fraser, D., Mahler, H.R., 1968. Chelating agent shock of bacteriophage T5. J. Virol. 2, 944–950.

Yang, H., Yu, J., Wei, H., 2014. Engineered bacteriophage lysins as novel anti-infectives. Front. Microbiol. 5, 542. doi:10.3389/fmicb.2014.00542

Young, R., 2014. Phage lysis: Three steps, three choices, one outcome. J. Microbiol. 52, 243–258. doi:10.1007/s12275-014-4087-z

Yu, J.G., Lim, J.A., Song, Y.R., Heu, S., Kim, G.H., Koh, Y.J., Oh, C.S., 2015. Isolation and characterization of bacteriophages against *Pseudomonas syringae* pv. *actinidiae* causing bacterial canker disease in kiwifruit. J. Microbiol. Biotechnol. 26, 385–393. doi:10.4014/jmb.1509.09012

Yuan, Y., Gao, M., 2017. Jumbo bacteriophages: An overview. Front. Microbiol. 8, 403. doi:10.3389/fmicb.2017.00403

Yuan, Y., Gao, M., 2016. Proteomic analysis of a novel *Bacillus* jumbo phage revealing glycoside hydrolase as structural component. Front. Microbiol. 7, 745. doi:10.3389/fmicb.2016.00745

Yuliar, Nion, Y.A., Toyota, K., 2015. Recent trends in control methods for bacterial wilt diseases caused by *Ralstonia solanacearum*. Microbes Environ. 30, 1–11. doi:10.1264/jsme2.ME14144

Zhong, Y., Chen, F., Wilhelm, S.W., Poorvin, L., Hodson, R.E., 2002. Phylogenetic diversity of marine Cyanophage Isolates and natural virus communities as revealed by sequences of viral capsid assembly protein gene g20. Appl. Environ. Microbiol. 68, 1576–1584. doi:10.1128/AEM.68.4.1576-1584.2002

Zhu, N., Olivera, B.M., Roth, J.R., 1991. Activity of the nicotinamide mononucleotide transport system is regulated in *Salmonella typhimurium*. J. Bacteriol. 173, 1311–20.

Zhuang, J., Jin, Y., 2003. Virus retention and transport as influenced by different forms of soil organic matter. J. Environ. Qual. 32, 816–23.

Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31, 3406–15.

Appendix 2

This appendix contains peer reviewed papers published during the duration of the work of this thesis.

Comparison of *Staphylococcus* Phage K with Close Phage Relatives Commonly Employed in Phage Therapeutics

Jude Ajuebor [1], **Colin Buttimer** [1] (iD), **Sara Arroyo-Moreno** [1], **Nina Chanishvili** [2], **Emma M. Gabriel** [1], **Jim O'Mahony** [1], **Olivia McAuliffe** [3], **Horst Neve** [4], **Charles Franz** [4] and **Aidan Coffey** [1,5,*] (iD)

[1] Department of Biological Sciences, Cork Institute of Technology, Bishopstown, Cork T12 P928, Ireland; jude.ajuebor@mycit.ie (J.A.); colin.buttimer@mycit.ie (C.B.); sara.arroyo-moreno@mycit.ie (S.A.-M.); emma.gabriel@mycit.ie (E.M.G.); Jim.OMahony@cit.ie (J.O.)

[2] Eliava Institute of Bacteriophages, Microbiology and Virology, Tbilisi 0160, Georgia; nina.chanishvili@pha.ge

[3] Teagasc, Moorepark Food Research Centre, Fermoy, Cork P61 C996, Ireland; Olivia.McAuliffe@teagasc.ie

[4] Department of Microbiology and Biotechnology, Max Rubner-Institut, DE-24103 Kiel, Germany; horst.neve@mri.bund.de (H.N.); charles.franz@mri.bund.de (C.F.)

[5] Alimentary Pharmabiotic Centre, University College, Cork T12 YT20, Ireland

[*] Correspondence: aidan.coffey@cit.ie; Tel.: +353-214-335-486

Abstract: The increase in antibiotic resistance in pathogenic bacteria is a public health danger requiring alternative treatment options, and this has led to renewed interest in phage therapy. In this respect, we describe the distinct host ranges of *Staphylococcus* phage K, and two other K-like phages against 23 isolates, including 21 methicillin-resistant *S. aureus* (MRSA) representative sequence types representing the Irish National MRSA Reference Laboratory collection. The two K-like phages were isolated from the *Fersisi* therapeutic phage mix from the Tbilisi Eliava Institute, and were designated B1 (vB_SauM_B1) and JA1 (vB_SauM_JA1). The sequence relatedness of B1 and JA1 to phage K was observed to be 95% and 94% respectively. In terms of host range on the 23 *Staphylococcus* isolates, B1 and JA1 infected 73.9% and 78.2% respectively, whereas K infected only 43.5%. Eleven open reading frames (ORFs) present in both phages B1 and JA1 but absent in phage K were identified by comparative genomic analysis. These ORFs were also found to be present in the genomes of phages (Team 1, vB_SauM-fRuSau02, Sb_1 and ISP) that are components of several commercial phage mixtures with reported wide host ranges. This is the first comparative study of therapeutic staphylococcal phages within the recently described genus *Kayvirus*.

Keywords: phage isolation; bacteriophage; phage resistance; MRSA; *Staphylococcus*; *Kayvirus*

1. Introduction

Staphylococcus aureus (*S. aureus*) is an opportunistic and important pathogen in clinical and health-care settings causing a wide variety of diseases commonly involving the skin, soft tissue, bone, and joints [1]. It is also a well-known causative agent of prosthetic joint infections (PJI), cardiac device infections, and intravascular catheter infections [1]. *S. aureus* pathogenicity is due, in part, to its ability to acquire and express a wide array of virulence factors, as well as antimicrobial resistance determinants [2], an example of which involves the acquisition of the staphylococcal cassette chromosome (SCCmec) leading to the development of methicillin resistance in *S. aureus* [3]. Methicillin-resistant *S. aureus* (MRSA) was first reported in 1961 [4], and has since been observed to cause serious infections in hospitals worldwide. Reports of MRSA clones resistant to the majority of antibiotics are a growing concern [5]. As such, new treatment options are needed.

Bacteriophages (phages) are biological entities composed of either DNA or RNA enclosed within a protein coat [6]. They are highly specific, with most phages capable of infecting only a single bacterial species [6,7], and studies on these viruses have been performed since the late 19th century [8]. The phage infection process usually begins with the recognition of the receptor on the bacterial cell surface by its receptor binding protein [9]. In natural environments bacterial hosts have evolved many mechanisms to protect themselves from phage attack to include; adsorption blocking, DNA injection blocking, restriction-modification system (R/M), abortive infection, and the clustered regularly interspaced short palindromic repeats (CRISPR)-Cas systems [10,11]. In turn, phages have evolved several strategies for overcoming these systems to ensure their survival in the phage-host co-evolutionary race [12–14].

The use of phages as therapeutics to eliminate pathogenic bacteria dates back to experiments conducted by Felix d'Herelle in 1919 at a French hospital to treat dysentery [15]. Since then, a wide range of phage therapy trials have been undertaken, many with very promising results [15,16]. Pyophage and Intesti-phage are among the commercial phage mixtures currently produced at the Eliava Institute. Metagenomic studies on these phage mixtures have been reported [17,18] and the staphylococcal phages Sb-1 and ISP are key components of Pyophage [19,20]. Other phages isolated from these commercial phages mixes have also been reported [21–24]. Phages like vB_SauM-fRuSau02 was isolated from a phage mix produced by Microgen (Moscow, Russia) [21] and Team 1 was isolated from PhageBioDerm, a wound healing preparation consisting of a biodegradable polymer impregnated with an antibiotic and lytic phages [22–24]. These phages all possess a wide host range against a number of clinically relevant *S. aureus* isolates, demonstrating the efficacy of such commercial phage mixtures in treating a range of bacterial infections [19–24].

In this paper, we employed another phage mixture from the Eliava Institute, namely the Fersisi phage mix. Fersisi is a relatively new combination developed approximately 15–20 years ago on the basis of Pyophage, although with fewer phage components. Two phages from this mix were designated B1 (vB_SauM_B1) and JA1 (vB_SauM_JA1). Phage K, on the other hand, is a well-known phage being the type phage of the recently designated genus *Kayvirus* of the subfamily *Spounavirinae* [25]. The exact origin of phage K is unknown, but descriptions of the phage are made as far back as 1949 [26,27]. An initial host range study involving this phage reported it to be ineffective against many MRSA strains [26]. Thus, phages B1 and JA1 were compared (on the basis of their host range) to phage K to explore possible host range differences and it was observed that both phages had broader host ranges. A comparative study was performed on their genomes and the genomes of similar phages from other commercial phage mixtures (Team 1, vB_SauM-fRuSau02, Sb_1 and ISP) with reported wide host ranges, to provide molecular insight into the differences in host range encountered in this study.

2. Results and Discussion

2.1. Origin of Phages B1 and JA1

Phages B1 and JA1 were isolated from the Fersisi commercial phage mixtures; batch 010112 (B1) and F-062015 (JA1). This product is used in the treatment of staphylococcal and streptococcal infections. For the isolation of B1, phage enrichment was carried out using staphylococcal host cultured from the sonicate fluid of a hospital patient suffering from PJI. DPC5246 was subsequently used as propagating host for B1, as a prophage was encountered in the PJI strain. Phage enrichment in the isolation of JA1 was done using the Cork Institute of Technology (CIT) collection strain *S. aureus* CIT281189. Both the PJI strain and CIT281189 were insensitive to phage K.

2.2. Morphology and Host Range of Phages K, B1 and JA1

Phages B1 and JA1 exhibited typical characteristics of phages belonging to the *Myoviridae* family, similar to the reported morphology of phage K [26]. All three phages possessed an A1 morphology [28], displaying an icosahedral head as well as a long contractile tail. They also contained a structure previously described as knob-like appendages by O'Flaherty et al. [26], extending from their base plates (likely

"clumped/aggregated" base plate appendices) and clearly visible in Figure 1. Estimations were made on the dimensions of these phages (Table 1). Capsid heights were estimated as 92.9 ± 4.0 nm (B1), 87.0 ± 2.1 nm (JA1) and 92.9 ± 3.8 nm (K). Tail dimension were also estimated as $233.0 \pm 4.4 \times 23.4 \pm 1.2$ nm (B1), $231.5 \pm 4.7 \times 22.7 \pm 0.9$ nm (JA1), and $227.5 \pm 5.5 \times 23.8 \pm 1.0$ nm (K), and base plates/knobs complexes were estimated as $30.1 \pm 1.8 \times 47.2 \pm 3.7$ nm (B1), $32.5 \pm 7.9 \times 45.8 \pm 1.4$ nm (JA1), and $36.6 \pm 5.1 \times 41.7 \pm 2.6$ nm (K). Owing to the similar morphology of all three phages, a host range study was conducted to explore possible differences in host spectra across a number of hospital isolates. Twenty-one of these isolates represented the entire collection of MRSA sequence-types identified in Ireland by the National MRSA Reference Laboratory (Dublin, Ireland), and includes the commonly encountered ST22-MRSA-IV, which has been predominant in Irish hospitals since the late 1990s [29]. The other two *S. aureus* strains used in this study were included as additional phage propagation strains. Host range was assessed by plaque assay technique on lawns of various MRSA strains listed in Table 2. The efficiency of plaquing (EOP) was used to represent the degree to which each of the phages studied infected all 23 staphylococcal strains. Phage JA1 had the broadest host range, forming plaques on 18 out of the 23 staphylococcal strains examined. B1 also had a broad host range and was capable of forming plaques on 17 isolates (with some in common with the 18 lysed by phage JA1). Phage K had the narrowest host range, forming plaques on only 10 of the isolates (including its propagating strain DPC5246). All 23 staphylococcal strains were effectively lysed by at least one of the three phages, with the exception of E1139 (IV) ST45 and E1185 (IV) ST12, whose EOP were significantly low at 3.88×10^{-6} and 1.16×10^{-6} respectively; as well as 3488 (VV) ST8, which was resistant to all three phages. Plaque size ranged from 0.5 mm to 1.5 mm, with a halo occurring in some instances (Table 3 and Supplementary Materials, Figure S1). The wide host range encountered in this study is common among K-like phages and has been reported for other staphylococcal K-like phages, such as JD007, which infected 95% of *S. aureus* isolates obtained from several hospitals in Shanghai, China [30].
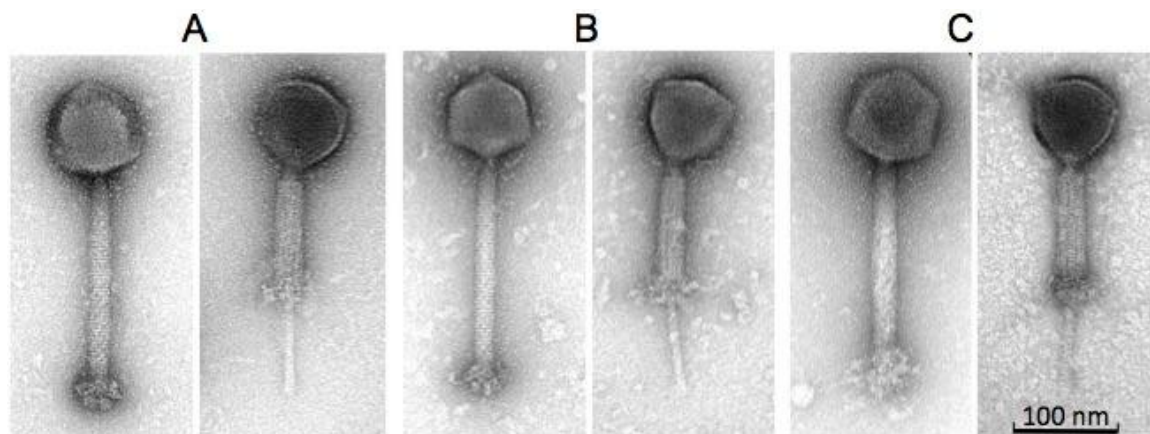
Figure 1. Transmission electron micrographs of phages B1 (**A**), JA1 (**B**), and K (**C**) showing their icosahedral capsid and their long contractile tail (both extended and contracted).

Table 1. Dimensions of staphylococcal phages B1, JA1, and K derived from micrographs obtained from transmission electron microscopy.

Phages	Head (nm)	Tail Length (nm) (incl. "knob")	Tail Width (nm)	Baseplate "knob" Length (nm)	Baseplate "knob" Width (nm)
B1	92.9 ± 4.0 (*n* = 11)	233.0 ± 4.4 (*n* = 12)	23.4 ± 1.2 (*n* = 12)	30.1 ± 1.8 (*n* = 12)	47.2 ± 3.7 (*n* = 10)
JA1	87.0 ± 2.1 (*n* = 9)	231.5 ± 4.7 (*n* = 9)	22.7 ± 0.9 (*n* = 9)	32.5 ± 7.9 (*n* = 9)	45.8 ± 1.4 (*n* = 9)
K	92.9 ± 3.8 (*n* = 16)	227.5 ± 5.5 (*n* = 16)	23.8 ± 1.0 (*n* = 16)	36.6 ± 5.1 (*n* = 16)	41.7 ± 2.6 (*n* = 16)

Table 2. Host ranges of staphylococcal phages B1, JA1, and K against methicillin-resistant *Staphylococcus aureus* (MRSA) strains from the Irish National Reference Laboratory (St. James's Hospital Dublin, Ireland) including the efficiency of plaquing (EOP) of these strains.

S. aureus Strain	Phage K	Phage B1	Phage JA1
DPC5246*	1.00 ± 0.0	1.00 ± 0.0	8.98×10^{-1} ± 0.8
CIT281189*	No infection	No infection	1.00 ± 0.0
0.0066 (IIIV) ST239	No infection	No infection	2.59 ± 2.5
0.1206 (IV) ST250	No infection	3.89×10^{-1} ± 0.3	1.35 ± 1.2
0.1239 (III) ST239	No infection	1.46×10^{-1} ± 0.1	4.17×10^{-2} ± 0.0
0.1345 (II) ST5	No infection	No infection	2.08×10^{-1} ± 0.1
0073 (III) ST239	No infection	3.21×10^{-1} ± 0.2	No infection
0104 (III) ST239	No infection	3.95×10^{-1} ± 0.2	1.82 ± 1.6
0220 (II) ST5	3.03×10^{-1} ± 0.1	2.17×10^{-1} ± 0.2	2.38×10^{-1} ± 0.2
0242 (IV) ST30	4.43×10^{-1} ± 0.1	5.23×10^{-1} ± 0.5	4.90×10^{-1} ± 0.3
0308 (IA) ST247	1.40 ± 0.2	1.36 ± 1.3	1.71 ± 1.6
3045 (IIV) ST8	No infection	4.93×10^{-2} ± 0.0	1.69 ± 0.7
3144 (IIV) ST8	No infection	1.21 ± 1.0	2.17 ± 1.2
3488 (VV) ST8	No infection	No infection	No infection
3581 (IA) ST247	No infection	No infection	9.26×10^{-1} ± 0.7
3594 (II) ST36	4.38×10^{-1} ± 0.1	8.67×10^{-1} ± 0.4	1.06 ± 0.7
3596 (IIV) ST8	2.49×10^{-4} ± 0.0	1.29 ± 0.9	3.59 ± 2.7
E1038 (IIV) ST8	1.27×10^{-4} ± 0.0	2.02×10^{-1} ± 0.2	1.89 ± 1.4
E1139 (IV) ST45	No infection	3.88×10^{-6} ± 0.0	No infection
E1174 (IV) ST22	7.03×10^{-1} ± 0.7	3.11×10^{-1} ± 0.2	No infection
E1185 (IV) ST12	1.16×10^{-6} ± 0.0	No infection	No infection
E1202 (II) ST496	No infection	4.79×10^{-1} ± 0.2	9.49×10^{-1} ± 0.8
M03/0073 (III) ST239	1.76 ± 0.5	1.51 ± 0.8	2.30 ± 0.7

* *S. aureus* strains for phage propagation; data is represented as means ± standard deviations based on triplicate measurements.

Table 3. Zone sizes and morphologies of B1, JA1, and K plaques formed on MRSA strains collected from the Irish National MRSA Reference Laboratory (St. James's Hospital Dublin, Ireland).

S. aureus Strain	Phage K	Phage B1	Phage JA1
DPC5246	2 mm	1 mm with halo to 2 mm	1 mm with halo to 2 mm
CIT281189	No plaques	No plaques	1.5 mm
0.0066 (IV) ST239	No plaques	No plaques	1 mm
0.1206 (IV) ST250	No plaques	2 mm	0.5 mm with halo to 1 mm
0.1239 (III) ST239	No plaques	0.5 mm, faint plaques	1 mm
0.1345 (II) ST5	No plaques	No plaques	1 mm
0073 (III) ST239	No plaques	0.5 mm	No plaques
0104 (III) ST239	No plaques	0.5 mm	1 mm
0220 (II) ST5	0.5 mm	1 mm	1 mm
0242 (IV) ST30	1 mm	1.5 mm	1.5 mm
0308 (IA) ST247	1 mm	1 mm	0.5 mm, faint plaques
3045 (IIV) ST8	No plaques	1 mm	1 mm

Table 3. *Cont.*

S. *aureus* Strain	Phage K	Phage B1	Phage JA1
3144 (IIV) ST8	No plaques	1.5 mm, faint plaques	1 mm
3488 (VV) ST8	No plaques	0.5 mm, faint plaques	0.5 mm with halo to 1 mm
3581 (IA) ST247	No plaques	No plaques	1 mm
3594 (II) ST36	1.5 mm	1 mm	1.5 mm
3596 (IIV) ST8	0.5 mm	0.5 mm with halo to 1.5 mm	0.5 mm with halo to 1.5 mm
E1038 (IIV) ST8	0.5 mm, faint plaques	0.5 mm, faint plaques	1.5 mm
E1139 (IV) ST45	No plaques	0.5 mm, faint plaques	No plaques
E1174 (IV) ST22	0.5 mm, faint plaques	0.5 mm	No plaques
E1185 (IV) ST12	0.5 mm, faint plaques	No plaques	No plaques
E1202 (II) ST496	No plaques	1 mm	0.5 mm
M03/0073 (III) ST239	2 mm	0.5 mm with halo to 1.5 mm	0.5 mm with halo to 1.5 mm

2.3. Phage Adsorption on Phage Resistant Isolates

While some level of phage insensitivity was encountered against all three phages, phage K was the frequently insensitive virion to the *S. aureus* strains tested, and thus, was chosen to evaluate whether or not adsorption inhibition played a role in its insensitivity. Phage K was able to adsorb to all phage-insensitive strains to approximately the same extent as the propagating strain DPC5246. This rules out the possibility of adsorption inhibition playing a role in the narrow host range encountered with phage K in comparison to both phages B1 and JA1 (Supplementary Materials, Figure S2). Additionally, adsorption studies with phages B1 and JA1 indicated that adsorption did not play a role in the differences observed.

2.4. Genome Comparison between Phages B1, JA1, and K

The genome of phage K is 139,831 bp in size with long terminal repeats (LTRs) of 8486 bp [31]. Genomes of similar sizes were obtained for phages B1 and JA1, these being 140,808 bp and 139,484 bp, respectively. Examination of sequence reads allowed the identification of LTRs for these phages, due to the identification of a region within their genomes with roughly double the average number of reads, these regions being 8076 bp and 7651 bp in size for phages B1 and JA1, respectively. This approach to the determination of terminal repeats has been utilized for a number of phages [32–34]. The sequences of all three phages, when analyzed, contained the 12 bp inverted repeat sequences 5′-TAAGTACCTGGG-3′ and 5′-CCCAGGTACTTA-3′, which separates the LTRs from the non-redundant part of the phage DNA, and are characteristic of K-like phages [22,35]. Thus, the entire packaged genome sizes are 148,884 bp (B1), 147,135 bp (JA1), and 148,317 bp (K). Phage K possessed 212 ORFs in its genome [31,36], whereas phages B1 and JA1 possessed 219 (Supplementary Materials, Table S1) and 215 ORFs (Supplementary Materials, Table S2) respectively.

Nucleotide pairwise sequence alignment based on BLASTN revealed phages B1 and JA1 (including their LTRs) to be 99% identical to each other, thus can be considered different isolates of the same phage species [37]. On the other hand, phages B1 and JA1 (including their LTRs) showed 95% and 94% identity (respectively) to phage K, placing these phages on the boundary of speciation.

The examination of 100 bp sequences upstream of each ORFs on the non-redundant genome of these phages, using MEME [38], identified 44 and 43 RpoD-like promoters for phages B1 and JA1, respectively. It was observed that these promoters where heavily concentrated in regions with ORFs encoding short hypothetical proteins and those with functions associated with nucleotide metabolism and DNA replication, rather than those associated with virion structure (Supplementary Materials, Tables S3 and S4). A similar finding was also reported with K-like phage vB_SauM-fRuSau02 [21]. Additionally, 30 Rho-independent terminators were identified on the non-redundant genomes for both B1 and JA1 (Supplementary Materials, Tables S5 and S6).

Four ORFs present in phage B1 were observed to be absent in JA1 (Table 4). These ORFs encoded two putative terminal repeat-encoded proteins (PhageB1_009, 016) and two proteins of unknown function (phageB1_202, 203). Although both B1 and JA1 had similar content of ORFs with 1% difference between

their genomes, both phages varied in their host range on the *S. aureus* strains they infected. This variation is likely attributed to the difference encountered in their genome. Additionally, multiple ORFs present in phage K but absent in both B1 and JA1 were encountered (Figure 2, Table 5). Furthermore, ORFs present in both phages B1 and JA1 but absent in K were also encountered (Figure 2, Table 6). These ORFs are discussed below.

Table 4. List of missing ORFs present in phage B1 but absent in phage JA1.

ORFs	Amino Acid Numbers	Protein Size (kDa)	Predicted Function
PhageB1_009	112	13.5	Terminal repeat encoded protein
PhageB1_016	107	12.4	Terminal repeat encoded protein
PhageB1_202	32	3.5	Unknown
PhageB1_203	104	11.6	Unknown
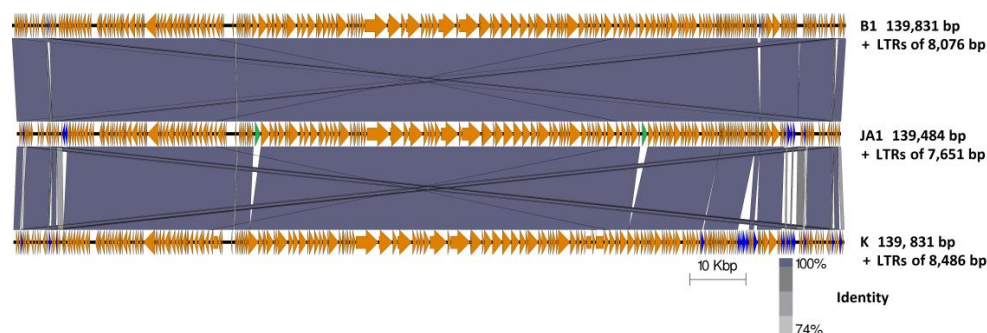
Figure 2. Genome comparison of phages B1, JA1, and K (including their long terminal repeats) using currently available annotations employing BLASTN and visualized with Easyfig. Regions of sequence similarity are connected by the shaded area, using a grey scale; genome maps consisting of orange arrows indicating the location of ORFs along the phage genomes, with unshared ORFs highlighted in blue with those indicating unshared homing endonuclease highlighted in green.

Table 5. List of missing ORFs and their predicted putative functions absent in both phages B1 and JA1 but present in phage K.

ORFs	Amino Acid Number	Protein Size (kDa)	Predicted Function
PhageK_004	108	12.7	Unknown
PhageK_016*	107	12.4	Unknown
PhageK_019	57	4.7	Unknown
PhageK_020	89	10.2	Unknown
PhageK_168	185	21.7	Predicted to contain a transmembrane region based on InterProScan
PhageK_187	101	11.7	Unknown
PhageK_188	123	13.8	Predicted to contain a transmembrane region based on InterProScan
PhageK_189	78	9.2	Unknown
PhageK_190	175	20.6	Predicted as a putative metallophoshatase
PhageK_191	106	12.9	Unknown
PhageK_192	76	8.9	Predicted to contain a transmembrane region based on InterProScan
PhageK_196	226	25.8	Unknown
PhageK_205	83	9.7	Unknown
PhageK_206	98	11.2	Unknown
PhageK_208	99	11.6	Unknown
PhageK_209	75	8.9	Unknown
PhageK_211	117	13.9	Predicted to possess a transmembrane region based on InterProScan
PhageK_212	128	15.6	Unknown

* ORF that phage JA1 does not share with phage K.

Table 6. List of missing ORFs and their predicted function absent in phage K but present in phages B1 and JA1.

ORFs	Amino Acid Number	Protein Size (kDa)	Predicted Function
PhageJA1_003 (PhageB1_003)	96	11.3	Unknown
PhageJA1_020 (PhageB1_022)	161	19.1	Unknown
PhageJA1_021 (PhageB1_023)	135	16.5	Unknown
PhageJA1_084 (PhageB1_087)	323	39.6	Predicted as a putative endonuclease interrupting the terminase large subunit [PhageJA1_083 (PhageB1_086) and PhageJA1_085 (PhageB1_088)]
PhageJA1_152 (PhageB1_155)	322	38.3	Predicted as a putative endonuclease containing a LAGLIDADG-like domain and an Intein splicing domain and interrupts the DNA repair protein [PhageJA1_151 (PhageB1_154) and PhageJA1_153 (PhageB1_156)]
PhageJA1_206 (PhageB1_212)	73	8.9	Unknown
PhageJA1_208 (PhageB1_214)	169	20.3	HHpred indicates homology to cell wall hydrolases
PhageJA1_209 (PhageB1_215)	109	12.6	Unknown
PhageJA1_211 (PhageB1_217)	104	12.0	Unknown
PhageJA1_212 (PhageB1_218)	55	6.5	Unknown
PhageJA1_213 (PhageB1_219)	33	3.7	Predicted to possess a transmembrane region based on InterProScan

2.4.1. Characteristic Features of Phage K ORFs Absent in Both JA1 and B1

Seventeen ORFs present in phage K were absent in both phages B1 and JA1, with one additional ORF found not to be shared between JA1 and K. These ORFs are listed in Table 5. No function could be assigned to these with the exception of phageK_190, which based on NCBI conserved domain search possessed a metallophosphatase-like domain (cd07390; E value; 3.94×10^{-30}) and is a member of the metallophosphatase (MPP) superfamily. Families within this superfamily of enzymes are functionally diverse, involved in the cleavage of phosphoester bonds, and include Mre11/SbcD-like exonucleases, Dbr1-like RNA lariat debranching enzymes, YfcE-like phosphodiesterases, purple acid phosphatases (PAPs), YbbF-like UDP-2,3-diacylglucosamine hydrolases, and acid sphingomyelinases (ASMases) [39].

2.4.2. Characteristic Features of Phages B1 and JA1 ORFs Absent in Phage K

Eleven ORFs present in both phages B1 and JA1 were absent in phage K (Table 6). No putative function could be assigned to the majority of these ORFs based on BLASTP, InterProScan or HHpred analysis, with the exception of phageJA1_084 (phageB1_087) and phageJA1_152 (phageB1_155), which encoded homing endonucleases interrupting both the terminase large subunit and the DNA repair protein, respectively. These homing endonucleases are site-specific DNA endonucleases capable of initiating DNA breaks leading to repair and recombination event that results in the integration of this endonuclease ORF into a gene that was previously lacking it [40]. The presence of these mobile genetic elements is common among known staphylococcal phages of the subfamily *Spounavirinae*, and these endonucleases ORFs are known to insert themselves into essential phage genes [21,41]. Additionally, HHpred analysis indicated ORFs PhageJA1_208 and PhageB1_214 to possess remote homology to cell-degrading proteins. The majority of these ORFs were found to be located next to the genome termini of JA1 and B1, with genes located in this region having been previously reported in similar phages to be expressed early in phage development [35]. Such proteins are usually involved in subversion of the host's machinery to aid phage takeover [42,43].

2.4.3. Comparison of Phages K, B1, and JA1 with other Similar Therapeutic Phages (Team1, vB_SauM-fRuSau02, Sb-1 and ISP)

Four additional staphylococcal phages that originate in commercial phage therapeutic mixtures are Team1, vB_SauM-fRuSau02, Sb-1 and ISP, as discussed earlier [19–24]. These phages were also reported to possess wide host ranges towards a number of clinically relevant *S. aureus* strains. Although similar, these phages have several feature differences from each other and from phages B1 and JA1. Comparison of nucleotide identities (BLASTN) with phage K shows that they belong to the genus *Kayvirus* (Supplementary Materials, Table S7) possessing genomes of similar sizes, apart from Sb-1, being smaller than would be expected, suggesting the genome submission may have been incomplete (Figure 3). Additionally, the arrangement of ORFs is quite similar. Furthermore, tRNA genes of these phages were also examined. All seven phages were found to possess the same four tRNA genes for methionine, tryptophan, phenylalanine, and aspartic acid (Supplementary Materials, Table S8). The eleven ORFs which were present in B1 and JA1 but absent in K (Table 6, Supplementary Materials, Figure S3) were similarly present in Team 1, vB_SauM-fRuSau02, Sb-1 and ISP. And likewise, the ORFs present in K, but absent in both B1 and JA1, were also missing in these phages. However, vB_SauM-fRuSau02 possesses a much shorter putative tail protein (RS_159) of 73 amino acids compared to the phage K counterpart (PhageK_151) of 170 amino acids. Non-hypothetical proteins that differed between these phages were a membrane protein (Phage B1_180, PhageJA1_177, and Phage_170) and an ATPase-like protein (Protein id: CCA65911.1 for phage ISP). Other ORFs that differed among these phages were mostly hypothetical proteins.
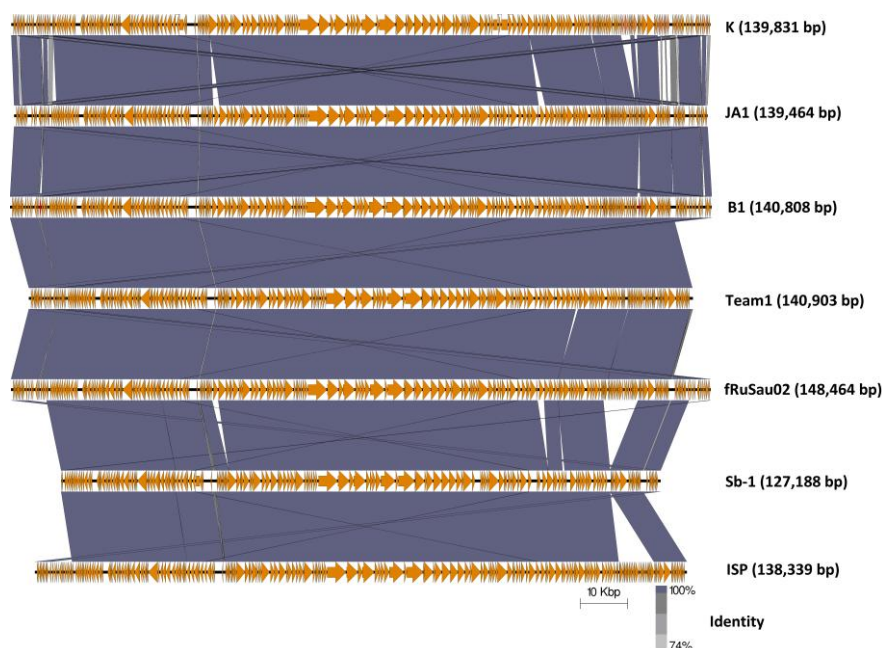
Figure 3. Genome comparison of phage K with the six staphylococcal phages employed in commercial phages mixture (B1, JA1, Team 1 [22–24], vB_SauM-fRuSau02 [21], Sb-1 [19] and ISP [20]) using currently available annotations employing BLASTN and visualized with Easyfig.

S. aureus employ several defense strategies against viral attack [10,44] and these, such as restriction modification systems [45] and CRISPR-Cas systems [46], may vary from strain to strain. These defenses along with several variations encountered at the genetic level across phages B1, JA1, and K may explain the differences in host ranges observed in this study.

3. Materials and Methods

3.1. Bacterial Strains, Phage and Growth Requirement

Phages B1 and JA1 were isolated from a commercial phage cocktail purchased from the George Eliava Institute of Bacteriophage, Microbiology and Virology, Tbilisi, Georgia. The MRSA strains utilized in this study were all acquired from the Irish National MRSA Reference Laboratory, Dublin, Ireland [2] with the exception of DPC5246 and CIT281189, which are routine propagation strains utilized in our laboratory [26,36]. These strains were routinely cultured in Brain Heart Infusion broth (BHI; Sigma, St. Louis, MO, USA) at 37 °C with shaking or on BHI plates containing 1.5% (*w/v*) bacteriological agar (Sigma). All strains were stocked in BHI containing 40% glycerol and stored at −80 °C.

3.2. CsCl Gradient Purification

Isopycnic centrifugation through CsCl gradients was performed as previously described [47], with a number of modifications. A high titer phage lysate (>1 × 10^9 plaque forming units [PFU] mL^{-1}), was precipitated using polyethylene glycol (15% *w/v* PEG8000, 1 M NaCl) at 4 °C overnight and centrifuged, after which the pellet was resuspended in TMN buffer (10 mM Tris-HCl pH 7.4, 10 mM MgSO$_4$·7H$_2$O, 0.5 M NaCl). The resulting phage preparation was placed onto a CsCl step gradient composed of 1.3, 1.5, and 1.7 g/mL layers and spun in a 100 Ti rotor (Beckman Coulter, Brea, CA, USA) at 200,480 *g* for 3 h at 4 °C. The resulting phage preparations were dialyzed in Tris-HCl buffer (10 mM, pH 7.5) at 4 °C.

3.3. Phage Host Range and Adsorption Study

Host range assay was performed for phages B1, JA1, and K using the plaque assay plating technique (Tables 2 and 3). This was done in triplicate for three independent experiments. The efficiency of plaquing (EOP) was determined by dividing the phage titer on each test strain by the phage titer of the reference strain (*S. aureus* DPC5246, in the case of phages B1 and K, and *S. aureus* CIT281189 for phage JA1) [48]. An adsorption assay was performed according to the protocol previously described elsewhere with some modification [49]. Briefly, MRSA strains were grown to an optical density (OD) of 0.2 at 600 nm (estimated cell count at 10^8 colony forming unit (cfu) mL^{-1}) and 100 μL of cells were mixed with 100 μL of respective phage titered at approximately 1 × 10^7 PFU/mL for a multiplicity of infection (MOI) of 0.1. The resulting mixtures were incubated at room temperature for 5 min to allow for phage adsorption. The bound phages were separated from the free phages by centrifugation at 14,000 rpm for 5 min. Adsorption of the phage on each strain was determined by subtracting the number of unbound phage (per mL) from the total input PFU/mL. Adsorption efficiency was expressed as a percentage relative to the propagating strain DPC5246.

3.4. Transmission Electron Microscopy

Electron microscopic analysis was performed following negative staining of the CsCl gradient prepared phages on freshly prepared carbon films with 2% (*w/v*) uranyl acetate. Electron micrographs were taken using a Tecnai 10 transmission electron microscope (FEI Thermo Fisher, Eindhoven, the Netherlands) at an acceleration voltage of 80 kV with a MegaView G2 CDD camera (EMSIS, Muenster, Germany).

3.5. Phage DNA Isolation

Phage DNA extraction was performed on CsCl purified high titer phages. These were initially treated with MgCl$_2$ followed by pre-treatment with DNase and RNase for 60 min at 37 °C. Following that subsequent treatment with SDS, EDTA and proteinase K with further incubation for 60 min at 55 °C. DNA extractions were then performed on the pre-treated samples with phenol/chloroform/isoamyl

alcohol (25:24:1 *v/v/v*) and chloroform/isoamyl alcohol (24:1 *v/v*). DNA precipitation was achieved using sodium acetate and 95% ethanol. DNA quality and quantity were estimated using a Nanodrop (ND-1000) and visualized following agarose gel electrophoresis

3.6. Phage DNA Sequencing

DNA sequencing was performed with a high throughput Illumina HiSeq system sequencing (GATC Biotech, Konstanz, Germany). Library preparation was performed by DNA fragmentation together with adapter ligation. The libraries were then measured and quantified on a Fragment Analyzer and then sequenced to generate 2 × 300 bp paired-end reads. *De novo* assembly was performed using CLC Bio Genomics Workbench v8.0 (Aarhus, Denmark).

3.7. Bioinformatic Analysis

Open reading frames (ORFs) for the sequenced phages were predicted with Glimmer [50] and GenemarkS [51]. Putative functions were assigned to these ORFs using BLASTP (https://blast.ncbi.nlm. nih.gov/Blast.cgi?PAGE=Proteins), HHpred (https://toolkit.tuebingen.mpg.de/#/tools/hhpred; [52]) and InterProscan (http://www.ebi.ac.uk/interpro/search/sequence-search; [53]). Transfer RNA was predicted using tRNAscan-SE (http://lowelab.ucsc.edu/tRNAscan-SE/; [54]) and ARAGORN (http://130.235.46.10/ARAGORN/; [55]). Potential promoters were predicted using MEME (Multiple Em for Motif Elicitation) (http://meme-suite.org/tools/meme; [38]), followed by manual curation. Potential Rho-independent terminators were identified using ARNold (http://rna.igmors.u-psud. fr/toolbox/arnold; [56]) with Mfold QuikFold (http://unafold.rna.albany.edu/?q=DINAMelt/ Quickfold; [57]) using RNA energy rules 3.0 to verify predictions. Artemis Comparison Tool (ACT) was used for the identification of feature variations between the genomes of phages, with homology being assessed with BLASTN [58] Genome comparison maps between phages were visualized using the Easyfig visualization tool [59]. K-like *Staphylococcus* phages used in comparative studies were K (KF766114), Team 1 (KC012913), vB_SauM-fRuSau02 (MF398190), Sb-1 (HQ163896) and ISP (FR852584).

3.8. Nucleotide Sequence Accession Number

The genome sequence for phages B1 and JA1 were deposited into GenBank under the accession numbers MG656408 and MF405094, respectively.

4. Conclusions

Host range of three highly similar phages was performed in this study, and it was identified that phages B1 and JA1 from the Fersisi commercial phage mix had a much broader host range in comparison to phage K on a representative Irish bank of clinical MRSA sequence type isolates. Comparisons of their genomes lead to the identification of several ORFs absent in phage K, but present in both phages B1 and JA1. These ORFs were also identified in several other staphylococcal phages sourced from commercial phage mixtures (B1, JA1, Team 1 [22–24], vB_SauM-fRuSau02 [21], Sb-1 [19] and ISP [20]), also with a reported wide host range. The exact role of these ORFs is currently unknown. However, these ORFs along with several variations encountered at the genetic level between these phages may, in part, explain their different host range. Unfortunately, information is lacking on the influences of various phage resistance systems, which may be active in *Staphylococcus aureus*. Phage research also needs to focus more on elucidation of the functions of hypothetical proteins to allow greater understanding of how phages overcome such systems.

References

1. Tong, S.Y.C.; Davis, J.S.; Eichenberger, E.; Holland, T.L.; Fowler, V.G. *Staphylococcus aureus* infections: Epidemiology, pathophysiology, clinical manifestations, and management. *Clin. Microbiol. Rev.* **2015**, *28*, 603–661. [CrossRef] [PubMed]

2. Shore, A.C.; Rossney, A.S.; O'Connell, B.; Herra, C.M.; Sullivan, D.J.; Humphreys, H.; Coleman, D.C. Detection of staphylococcal cassette chromosome mec-associated DNA segments in multiresistant methicillin-susceptible *Staphylococcus aureus* (MSSA) and identification of *Staphylococcus epidermidis* ccrAB4 in both methicillin-resistant *S. aureus* and MSSA. *Antimicrob. Agents Chemother.* **2008**, *52*, 4407–4419. [CrossRef] [PubMed]

3. Hiramatsu, K.; Cui, L.; Kuroda, M.; Ito, T. The emergence and evolution of methicillin-resistant *Staphylococcus aureus*. *Trends Microbiol.* **2001**, *9*, 486–493. [CrossRef]

4. Jevons, M.P. "Celbenin"—Resistant Staphylococci. *BMJ* **1961**, *1*, 124–125. [CrossRef]

5. Klein, E.; Smith, D.L.; Laxminarayan, R. Hospitalizations and Deaths Caused by Methicillin-Resistant *Staphylococcus aureus*, United States, 1999–2005. *Emerg. Infect. Dis.* **2007**, *13*, 1840–1846. [CrossRef] [PubMed]

6. O'Flaherty, S.; Ross, R.P.; Coffey, A. Bacteriophage and their lysins for elimination of infectious bacteria. *FEMS Microbiol. Rev.* **2009**, *33*, 801–819. [CrossRef] [PubMed]

7. Schmelcher, M.; Loessner, M.J. Application of bacteriophages for detection of foodborne pathogens. *Bacteriophage* **2014**, *4*, e28137. [CrossRef] [PubMed]

8. Wittebole, X.; De Roock, S.; Opal, S.M. A historical overview of bacteriophage therapy as an alternative to antibiotics for the treatment of bacterial pathogens. *Virulence* **2014**, *5*, 226–235. [CrossRef] [PubMed]

9. Bertozzi Silva, J.; Storms, Z.; Sauvageau, D. Host receptors for bacteriophage adsorption. *FEMS Microbiol. Lett.* **2016**, *363*, 1–11. [CrossRef] [PubMed]

10. Hyman, P.; Abedon, S.T. Bacteriophage Host Range and Bacterial Resistance. In *Advances in Applied Microbiology*, 1st ed.; Laskin, A.I., Sarislani, S., Gadd, G.M., Eds.; Elsevier Inc.: Cambridge, MA, USA, 2010; Volume 70, pp. 217–248, ISBN 9780123809919.

11. Labrie, S.J.; Samson, J.E.; Moineau, S. Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.* **2010**, *8*, 317–327. [CrossRef] [PubMed]

12. Hall, A.R.; Scanlan, P.D.; Buckling, A. Bacteria-Phage Coevolution and the Emergence of Generalist Pathogens. *Am. Nat.* **2011**, *177*, 44–53. [CrossRef] [PubMed]

13. Hall, J.P.J.; Harrison, E.; Brockhurst, M.A. Viral host-adaptation: Insights from evolution experiments with phages. *Curr. Opin. Virol.* **2013**, *3*, 572–577. [CrossRef] [PubMed]

14. Samson, J.E.; Magadán, A.H.; Sabri, M.; Moineau, S. Revenge of the phages: Defeating bacterial defences. *Nat. Rev. Microbiol.* **2013**, *11*, 675–687. [CrossRef] [PubMed]

15. Sulakvelidze, A.; Alavidze, Z.; Morris, J.G. Bacteriophage therapy. *Antimicrob. Agents Chemother.* **2001**, *45*, 649–659. [CrossRef] [PubMed]

16. Abedon, S.T.; Kuhl, S.J.; Blasdel, B.G.; Kutter, E.M. Phage treatment of human infections. *Bacteriophage* **2011**, *1*, 66–85. [CrossRef] [PubMed]

17. Villarroel, J.; Larsen, M.V.; Kilstrup, M.; Nielsen, M. Metagenomic analysis of therapeutic PYO phage cocktails from 1997 to 2014. *Viruses* **2017**, *9*. [CrossRef] [PubMed]

18. Zschach, H.; Joensen, K.G.; Lindhard, B.; Lund, O.; Goderdzishvili, M.; Chkonia, I.; Jgenti, G.; Kvatadze, N.; Alavidze, Z.; Kutter, E.M.; et al. What can we learn from a metagenomic analysis of a georgian bacteriophage cocktail? *Viruses* **2015**, *7*, 6570–6589. [CrossRef] [PubMed]

19. Kvachadze, L.; Balarjishvili, N.; Meskhi, T.; Tevdoradze, E.; Skhirtladze, N.; Pataridze, T.; Adamia, R.; Topuria, T.; Kutter, E.; Rohde, C.; et al. Evaluation of lytic activity of staphylococcal bacteriophage Sb-1 against freshly isolated clinical pathogens. *Microb. Biotechnol.* **2011**, *4*, 643–650. [CrossRef] [PubMed]

20. Vandersteegen, K.; Mattheus, W.; Ceyssens, P.J.; Bilocq, F.; de Vos, D.; Pirnay, J.P.; Noben, J.P.; Merabishvili, M.; Lipinska, U.; Hermans, K.; et al. Microbiological and molecular assessment of bacteriophage ISP for the control of *Staphylococcus aureus*. *PLoS ONE* **2011**, *6*, e24418. [CrossRef] [PubMed]

21. Leskinen, K.; Tuomala, H.; Wicklund, A.; Horsma-Heikkinen, J.; Kuusela, P.; Skurnik, M.; Kiljunen, S. Characterization of vB_SauM-fRuSau02, a Twort-Like Bacteriophage Isolated from a Therapeutic Phage Cocktail. *Viruses* **2017**, *9*, 258. [CrossRef] [PubMed]

22. El Haddad, L.; Abdallah, N.B.; Plante, P.L.; Dumaresq, J.; Katsarava, R.; Labrie, S.; Corbeil, J.; St-Gelais, D.; Moineau, S. Improving the safety of *Staphylococcus aureus* polyvalent phages by their production on a *Staphylococcus xylosus* strain. *PLoS ONE* **2014**, *9*, e102600. [CrossRef] [PubMed]

23. Markoishvili, K.; Tsitlanadze, G.; Katsarava, R.; Morris, J.G.; Sulakvelidze, A. A novel sustained-release matrix based on biodegradable poly(ester amide)s and impregnated with bacteriophages and an antibiotic shows promise in management of infected venous stasis ulcers and other poorly healing wounds. *Int. J. Dermatol.* **2002**, *41*, 453–458. [CrossRef] [PubMed]

24. Jikia, D.; Chkhaidze, N.; Imedashvili, E.; Mgaloblishvili, I.; Tsitlanadze, G.; Katsarava, R.; Morris, J.G.; Sulakvelidze, A. The use of a novel biodegradable preparation capable of the sustained release of bacteriophages and ciprofloxacin, in the complex treatment of multidrug-resistant *Staphylococcus aureus*-infected local radiation injuries caused by exposure to Sr90. *Clin. Exp. Dermatol.* **2005**, *30*, 23–26. [CrossRef] [PubMed]

25. Adriaenssens, E.M.; Clokie, C.M.R.; Sullivan, M.B.; Gillis, A.; Jens Kuhn, B.H.; Kropinski, A.M. Taxonomy of prokaryotic viruses: 2016 update from the ICTV bacterial and archaeal viruses subcommittee. *Arch. Virol.* **2017**, *162*, 1153–1157. [CrossRef] [PubMed]

26. O'Flaherty, S.; Ross, R.P.; Meaney, W.; Fitzgerald, G.F.; Elbreki, M.F.; Coffey, A. Potential of the Polyvalent Anti-Staphylococcus Bacteriophage K for Control of Antibiotic-Resistant Staphylococci from Hospitals. *Appl. Environ. Microbiol.* **2005**, *71*, 1836–1842. [CrossRef] [PubMed]

27. Rountree, P.M. The serological differentiation of staphylococcal bacteriophages. *J. Gen. Microbiol.* **1949**, *3*, 164–173. [CrossRef] [PubMed]

28. Ackermann, H.W. Frequency of morphological phage descriptions in the year 2000. *Arch. Virol.* **2001**, *146*, 843–857. [CrossRef] [PubMed]

29. Rossney, A.S.; Lawrence, M.J.; Morgan, P.M.; Fitzgibbon, M.M.; Shore, A.; Coleman, D.C.; Keane, C.T.; O'Connell, B. Epidemiological typing of MRSA isolates from blood cultures taken in Irish hospitals participating in the European Antimicrobial Resistance Surveillance System (1999–2003). *Eur. J. Clin. Microbiol. Infect. Dis.* **2006**, *25*, 79–89. [CrossRef] [PubMed]

30. Cui, Z.; Feng, T.; Gu, F.; Li, Q.; Dong, K.; Zhang, Y.; Zhu, Y.; Han, L.; Qin, J.; Guo, X. Characterization and complete genome of the virulent Myoviridae phage JD007 active against a variety of *Staphylococcus aureus* isolates from different hospitals in Shanghai, China. *Virol. J.* **2017**, *14*, 26. [CrossRef] [PubMed]

31. Gill, J.J. Revised Genome Sequence of *Staphylococcus aureus* Bacteriophage K. *Genome Announc.* **2014**, *2*, 12–13. [CrossRef] [PubMed]

32. Buttimer, C.; Hendrix, H.; Oliveira, H.; Casey, A.; Neve, H.; McAuliffe, O.; Paul Ross, R.; Hill, C.; Noben, J.P.; O'Mahony, J.; et al. Things are getting hairy: Enterobacteria bacteriophage vB_PcaM_CBB. *Front. Microbiol.* **2017**, *8*, 1–16. [CrossRef] [PubMed]

33. Li, S.; Fan, H.; An, X.; Fan, H.; Jiang, H.; Chen, Y.; Tong, Y. Scrutinizing virus genome termini by high-throughput sequencing. *PLoS ONE* **2014**, *9*, e85806. [CrossRef] [PubMed]

34. Fouts, D.E.; Klumpp, J.; Bishop-Lilly, K.A.; Rajavel, M.; Willner, K.M.; Butani, A.; Henry, M.; Biswas, B.; Li, M.; Albert, M.J.; et al. Whole genome sequencing and comparative genomic analyses of two *Vibrio cholerae* O139 Bengal-specific Podoviruses to other N4-like phages reveal extensive genetic diversity. *Virol. J.* **2013**, *10*, 165. [CrossRef] [PubMed]

35. Łobocka, M.; Hejnowicz, M.S.; Dąbrowski, K.; Gozdek, A.; Kosakowski, J.; Witkowska, M.; Ulatowska, M.I.; Weber-Dąbrowska, B.; Kwiatek, M.; Parasion, S.; et al. Genomics of staphylococcal Twort-like phages—Potential therapeutics of the post-antibiotic era. *Adv. Virus Res.* **2012**, *83*, 143–216. [CrossRef] [PubMed]

36. O'Flaherty, S.; Coffey, A.; Edwards, R.; Meaney, W.; Fitzgerald, G.F.; Ross, R.P. Genome of Staphylococcal Phage K: A New Lineage of Myoviridae Infecting Gram-Positive Bacteria with a Low G+C Content. *J. Bacteriol.* **2004**, *186*, 2862–2871. [CrossRef] [PubMed]

37. Adriaenssens, E.M.; Rodney Brister, J. How to name and classify your phage: An informal guide. *Viruses* **2017**, *9*, 70. [CrossRef] [PubMed]

38. Bailey, T.L.; Boden, M.; Buske, F.A.; Frith, M.; Grant, C.E.; Clementi, L.; Ren, J.; Li, W.W.; Noble, W.S. MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res.* **2009**, *37*, W202–W208. [CrossRef] [PubMed]

39. Matange, N.; Podobnik, M.; Visweswariah, S.S. Metallophosphoesterases: structural fidelity with functional promiscuity. *Biochem. J.* **2015**, *467*, 201–216. [CrossRef] [PubMed]

40. Gogarten, J.P.; Hilario, E. Inteins, introns, and homing endonucleases: recent revelations about the life cycle of parasitic genetic elements. *BMC Evol. Biol.* **2006**, *6*, 94. [CrossRef] [PubMed]

41. Vandersteegen, K.; Kropinski, A.M.; Nash, J.H.E.; Noben, J.-P.; Hermans, K.; Lavigne, R. Romulus and Remus, two phage isolates representing a distinct clade within the Twortlikevirus genus, display suitable properties for phage therapy applications. *J. Virol.* **2013**, *87*, 3237–3247. [CrossRef] [PubMed]

42. Wei, P.; Stewart, C.R. A Cytotoxic Early Gene of *Bacillus subtilis* Bacteriophage SPOl. *J. Bacteriol.* **1993**, *175*, 7887–7900. [CrossRef] [PubMed]

43. Stewart, C.R.; Gaslightwala, I.; Hinata, K.; Krolikowski, K.A.; Needleman, D.S.; Peng, A.S.Y.; Peterman, M.A.; Tobias, A.; Wei, P. Genes and regulatory sites of the "host-takeover module" in the terminal redundancy of *Bacillus subtilis* bacteriophage SPO1. *Virology* **1998**, *246*, 329–340. [CrossRef] [PubMed]

44. Seed, K.D. Battling Phages: How Bacteria Defend against Viral Attack. *PLoS Pathog.* **2015**, *11*, e1004847. [CrossRef] [PubMed]

45. Roberts, G.A.; Houston, P.J.; White, J.H.; Chen, K.; Stephanou, A.S.; Cooper, L.P.; Dryden, D.T.F.; Lindsay, J.A. Impact of target site distribution for Type i restriction enzymes on the evolution of methicillin-resistant *Staphylococcus aureus* (MRSA) populations. *Nucleic Acids Res.* **2013**, *41*, 7472–7484. [CrossRef] [PubMed]

46. Cao, L.; Gao, C.H.; Zhu, J.; Zhao, L.; Wu, Q.; Li, M.; Sun, B. Identification and functional study of type III—A CRISPR-Cas systems in clinical isolates of *Staphylococcus aureus*. *Int. J. Med. Microbiol.* **2016**, *306*, 686–696. [CrossRef] [PubMed]

47. Sambrook, J.; Russell, D.W. Purification of bacteriophage lamda particles by isopycnic centrifugation through CsCl gradients. In *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: New York, NY, USA, 2001; Volume 1, p. 247, ISBN 0879695773.

48. Gutiérrez, D.; Vandenheuvel, D.; Martínez, B.; Rodríguez, A.; Lavigne, R.; García, P. Two Phages, phiIPLA-RODI and phiIPLA-C1C, Lyse Mono- and Dual-Species Staphylococcal Biofilms. *Appl. Environ. Microbiol.* **2015**, *81*, 3336–3348. [CrossRef] [PubMed]

49. Li, X.; Koç, C.; Kühner, P.; Stierhof, Y.-D.; Krismer, B.; Enright, M.C.; Penadés, J.R.; Wolz, C.; Stehle, T.; Cambillau, C.; et al. An essential role for the baseplate protein Gp45 in phage adsorption to *Staphylococcus aureus*. *Sci. Rep.* **2016**, *6*, 26455. [CrossRef] [PubMed]

50. Delcher, A. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* **1999**, *27*, 4636–4641. [CrossRef] [PubMed]

51. Besemer, J.; Lomsadze, A.; Borodovsky, M. GeneMarkS: A self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* **2001**, *29*, 2607–2618. [CrossRef] [PubMed]

52. Söding, J.; Biegert, A.; Lupas, A.N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **2005**, *33*, W244–W248. [CrossRef] [PubMed]

53. Mitchell, A.; Chang, H.-Y.; Daugherty, L.; Fraser, M.; Hunter, S.; Lopez, R.; McAnulla, C.; McMenamin, C.; Nuka, G.; Pesseat, S.; et al. The InterPro protein families database: The classification resource after 15 years. *Nucleic Acids Res.* **2015**, *43*, D213–D221. [CrossRef] [PubMed]

54. Lowe, T.M.; Eddy, S.R. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **1997**, *25*, 955–964. [CrossRef] [PubMed]

55. Laslett, D.; Canback, B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* **2004**, *32*, 11–16. [CrossRef] [PubMed]

56. Naville, M.; Ghuillot-Gaudeffroy, A.; Marchais, A.; Gautheret, D. ARNold: A web tool for the prediction of rho-independent transcription terminators. *RNA Biol.* **2011**, *8*, 11–13. [CrossRef] [PubMed]

57. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **2003**, *31*, 3406–3415. [CrossRef] [PubMed]

58. Carver, T.J.; Rutherford, K.M.; Berriman, M.; Rajandream, M.-A.; Barrell, B.G.; Parkhill, J. ACT: The Artemis comparison tool. *Bioinformatics* **2005**, *21*, 3422–3423. [CrossRef] [PubMed]

59. Sullivan, M.J.; Petty, N.K.; Beatson, S.A. Easyfig: A genome comparison visualizer. *Bioinformatics* **2011**, *27*, 1009–1010. [CrossRef] [PubMed]

CrossMark

Investigating the biocontrol and anti-biofilm potential of a three phage cocktail against *Cronobacter sakazakii* in different brands of infant formula

Lorraine Endersen[a], Colin Buttimer[a], Eoghan Nevin[a], Aidan Coffey[a], Horst Neve[b], Hugo Oliveira[c], Rob Lavigne[d], Jim O'Mahony[a],*

[a] *Department of Biological Sciences, Cork Institute of Technology, Cork, Ireland*
[b] *Department of Microbiology and Biotechnology, Max Rubner-Institute, Federal Research Institute of Nutrition and Food, 1, Kiel, Germany*
[c] *CEB-Centre of Biological Engineering, LIBRO – Laboratório de Investigação em Biofilmes Rosário Oliveira, University of Minho, 4710-057 Braga, Portugal*
[d] *Laboratory of Gene Technology, KU-Leuven, Leuven, Belgium*

A B S T R A C T

In recent years, the microbiological safety of powdered infant formula has gained increasing attention due to the identification of contaminating *C. sakazakii* and its epidemiological link with life-threatening neonatal infections. Current intervention strategies have fallen short of ensuring the production of infant formula that is free from *C. sakazakii*. In this study, we describe the isolation and characterisation of three bacteriophages (phages) and their application as a phage cocktail to inhibit the growth of *C. sakazakii* in different brands of infant formula, while also assessing the phages ability to prevent biofilm formation. All three phages, isolated from slurry, possess a relatively broad host range, verified by their ability to infect across genera and species. When all three phages were combined and used as part of a phage cocktail, 73% coverage was obtained across all *Cronobacter* strains tested. Optimum thermo-tolerance and pH stability were determined between 4 °C–37 °C, and pH 6–8, respectively, well within the normal range of application of infant formula. Genome sequencing and analysis revealed all the phages to be free from lysogenic properties, a trait which renders each favourable for phage therapy applications. As such, the combined-phage preparation (3×10^8 pfu/mL) was found to possess a strong bactericidal effect on *C. sakazakii*/*C. sakazakii* LUX cells ($\leq 10^4$ cfu/mL), resulting in a significant reduction in cell numbers, to below the limit of detection (< 10 cfu/mL). This was observed following a 20 h challenge in different brands of infant formula, where samples in the absence of the phage cocktail reached concentrations of $\sim 10^9$ cfu/mL. The phage cocktail also demonstrated promise in preventing the establishment of biofilm, as biofilm formation could not be detected for up to 48 h post treatment. These results highlight the potential application of this phage preparation for biocontrol of *C. sakazakii* contamination in reconstituted infant formula and also as a preventative agent against biofilm formation.

1. Introduction

Cronobacter spp. (formally *Enterobacter sakazakii*) consists of a diverse group of Gram-negative, facultatively anaerobic, motile bacilli belonging to the *Enterobacteriaceae* family (Iversen et al., 2007). The genus comprises seven species: *Cronobacter sakazakii, Cronobacter malonaticus, Cronobacter turicensis, Cronobacter universalis, Cronobacter muytjensi, Cronobacter dublinensis* and *Cronobacter condiment* (Brady et al., 2013; Joseph et al., 2012). In recent years, *C. sakazakii* has gained significant attention as an emerging food-borne pathogen due to the associated link between infectious disease and the consumption of contaminated foods, in particular, reconstituted infant milk formula.

While *C. sakazakii* is responsible for causing severe clinical infections in immunocompromised individuals of all ages, it is pre-term, low-birth weight infants who are most at risk (FAO/WHO 2008; Healy et al., 2010). Clinical symptoms of infection in infants include meningitis, bacteraemia and severe forms of necrotising enterocolitis, with case fatality rates ranging between 40 and 80% (Friedemann, 2009). These high mortality rates and the fact that many survivors are very often left with chronic neurological and developmental disorders, highlights the damaging effect this organism has on infant health (Forsythe, 2005; Lai, 2001). Accordingly, The International Commission for Microbiological Specifications for Foods has ranked *C. sakazakii* as a "Severe hazard for restricted populations, life threatening or substantial chronic sequelae

of long duration", placing the organism in the same category as *Clostridium botulinum*, *Cryptosporidium parvum* and *Listeria monocytogenes* (types A and B) (ICMSF, 2002).

C. sakazakii is ubiquitous in nature with many studies indicating that plant material is its primary niche (in particular vegetables, fruits, cereals, wheat, rice, herbs and spices). However, other more pertinent sources also found to harbour this pathogen include powdered infant formula (PIF) and milk powder manufacturing environments (Friedemann, 2007; Kandhai et al., 2004).

C. sakazakii possess physiological traits which affords its ability to survive in such environments and thus permit PIF to serve as a prime vehicle for transmission to the immunocompromised infant. These traits include, (1) resistance to desiccation and osmotic stress (Breeuwer et al., 2003), (2) an extended temperature growth range (Iversen and Forsythe, 2004), (3) thermo-tolerance compared to other *Enterobacteriaceae* found in PIF (Nazarowec-White and Farber, 1997a, 1997b), and (4) the ability to form biofilms on a range of different materials including polycarbonate which is often used to make babies bottles (Iversen and Forsythe, 2004).

PIF is not manufactured as a sterile preparation and hence can become contaminated with *C. sakazakii* during production. Although the organism is effectively inactivated during pasteurisation (Nazarowec-White and Farber, 1997a, 1997b), contamination is likely to occur from the addition of non-sterile ingredients during manufacture. Indeed, it has been suggested that PIF ingredients originating from plant material, which have not been heat treated are a potential source of *C. sakazakii* contamination (Healy et al., 2010). Other possible sources of contamination include the use of non-sterile equipment during processing or reconstitution, and from temperature abuse of the reconstituted formula itself (Al-Nabulsi et al., 2009).

Capsular polysaccharides on the outer surface of *C. sakazakii* cells play a central role in biofilm formation, giving the organism the ability to attach and colonise a variety of surfaces including stainless steel, glass, latex, polycarbonate, silicon and polyvinyl chloride (PVC) (Iversen et al., 2004). The bacterial cells embedded in a matrix of exopolymeric substances, are physiologically distinct from their planktonic counterparts. These cells demonstrate changes in growth rate and gene transcription and often exhibit a significantly higher tolerance to antibiotics (≤ 1000 times higher) and other sanitising agents. The presence of persister cells, the reduced metabolic activity present in the inner layers of the biofilm and the decreased penetration of antibiotics through the exopolymeric matrix, all contribute to this increased resistance (Donlan and Costerton, 2002; Keren et al., 2004). As a result, complete elimination is very often compromised and re-infection commonly occurs.

Consequently, the microbiological safety of PIF is under continuous scrutiny as a result of contaminating *C. sakazakii* and its epidemiological link with life-threatening neonatal infections (Forsythe, 2005; Himelright, 2002; Hunter and Bean, 2013). The destructive economic impact the pathogen has on healthcare systems and PIF production facilities due to contaminated product recalls is also apparent (Chenu and Cox, 2009). Indeed, there is a growing requirement for the development of new and effective mechanisms to further prevent *C. sakazakii* contamination in both food, and the food processing environment.

Bacteriophages (phages) and their derivatives are well recognised for their antibacterial properties, demonstrating promise as natural, safe and effective alternatives for the prevention, treatment and/or eradication of foodborne pathogens in a range of different foods and food processing environments. These include, decontamination of livestock, sanitation of contact surfaces and equipment, in addition to biocontrol of raw meats, fresh foods and vegetables (Endersen et al., 2014; Goodridge and Bisha, 2011), cheese (Carlton et al., 2005), ready-to-eat (RTE) foods (Bigot et al., 2011), skim milk (Ellis et al., 1973; Endersen et al., 2013), and reconstituted infant formula (Kim et al., 2007), all of which demonstrate their applicability for use at each stage

of the food production process. ListShield™, approved for use in the US in 2006 as a processing aid to control *Listeria monocytogenes* in meat and poultry products, marked the arrival of the first phage-based product to the commercial marketplace in the Western world (Bren, 2006). Following this significant development, several phage related products have been approved for use in the US, including preparations active against the prominent foodborne pathogens, *E. coli* O157:H7, *Salmonella*, and additional preparations against L. *monocytogenes* (Endersen et al., 2014; Goodridge and Bisha, 2011). The pioneering anti-listeria phage-based preparation is now registered in Europe as an organic food additive and has also been approved for use as a food processing aid by the Food Standards Australia & New Zealand, highlighting the fact that phage-based preparations are continuing to gain global acceptance as safe and effective alternatives for the biocontrol of harmful foodborne pathogens (Fsanz, 2012; Hodgson, 2013).

Here we report the isolation and characterisation of three phages against the opportunistic, foodborne, infant formula pathogen, *C. sakazakii*. In addition, we demonstrate the efficacy of the phages, used as part of a phage cocktail, at inhibiting the growth of *C. sakazakii* in four different brands of reconstituted infant milk formula while also assessing the phages ability to prevent biofilm formation.

2. Materials and methods

2.1. Bacterial strains and growth media

The following strains were used in this study: *C. sakazakii* ATCC BAA 894, *C. sakazakii* ATCC BAA 894 LUX, *C. sakazakii* DPC 6258, *C. muytjensi* ATCC 51329, *C. sakazakii* ATCC 29004, and *C. sakazakii* DPC 8155. These were sourced from the Dairy Products Research Centre, DPC, Moorepark, Fermoy, Co. Cork, Ireland. Strains were stored at −80 °C and routinely grown on Luria-Bertani (LB) agar, in LB broth and in some cases supplemented with 1% or 2.5% D-(+)-glucose at 37 °C. HiChrome™ *Cronobacter* spp. Agar, Modified (14,763, Sigma Aldrich) was used for selective growth of *C. sakazakii* when necessary. Selective growth of *C. sakazakii* ATCC BAA 894 LUX was achieved by the addition of 500 μg/mL erythromycin to LB broth/agar.

2.2. Isolation of phages

Phages were isolated from slurry, sourced from a cattle farmer in Clonakilty, West Cork, Ireland, using methods described previously (Carlson, 2005). Briefly, each sample (5 mL) was added to equal volumes of LB broth (Sigma Aldrich, UK), supplemented with 10 mM CaCl₂ (Sigma Aldrich) and inoculated with a mid-log phase *C. sakazakii* ATCC BAA 894 culture. Samples were incubated overnight at 37 °C with shaking. Samples were centrifuged at 4000g for 15 min to pellet cells and debris. The supernatant was filtered through a 0.45 μM filter and the filtrate was re-enriched with mid-log phage *C. sakazakii* culture two more times. The supernatant obtained from the final enrichment step was filter-sterilised and tested for the presence of viable infective phages by adding 100 μL of the filtrate to 100 μL of early log phase *C. sakazakii* cells in a 5 mL LB 0.4% w/v overlay agar tube tempered to 50 °C and subsequently poured onto the surface of a LB 1.5% w/v agar plate. Plates were incubated at 37 °C overnight and then examined for phage plaques.

2.3. Phage purification and amplification

Presumptive phages were purified by successive single plaque isolation and were routinely propagated on *C. sakazakii* ATCC BAA 894 as previously described (O'Flaherty et al., 2005). Briefly, a single isolated plaque was aseptically picked from a lawn of *C. sakazakii* ATCC BAA 894 on LB agar using a sterile capillary tube and added to 100 μL of early-log phase culture. The sample was incubated at 37 °C overnight. The resulting lysate was centrifuged, filter sterilised and serially

diluted followed by standard plaque assay on *C. sakazakii*, as described previously. Single plaque purification was repeated twice, after which purified phage was recovered. The phage titre was assessed using a spot plaque assay method, where 10 μL of serial dilutions of each phage suspension were spotted onto *C. sakazakii* seeded indicator plates. Plates were incubated at 37 °C overnight and spot dilutions were subsequently examined for the presence of plaques. High titre phage suspensions were obtained by adding 5 mL of phage buffer (50 mM Tris pH 8, 150 mM NaCl, 10 mM MgCl₂, 2 mM CaCl₂) to a standard plaque assay plate revealing confluent lysis. Phages were recovered by aseptically running sterile hockey sticks across the surface of the agar plate to physically recover the phages, incubating for 2 h at 37 °C with shaking. The buffer was then recovered from the plate, centrifuged to pellet debris and filter sterilised using a 0.45 μm filter. Again, phage titres were assessed using a spot plaque assay technique as described previously. The purified high-titre phage solutions were stored at 4 °C.

2.4. Phage DNA preparation

Phage DNA isolation and restriction analysis were carried out as previously described (Kropinski and Clokie, 2009). Briefly, 18 μL of 1 mg/mL DNAse I (Fischer Scientific, 11873795) and 8 μL of 12.5 mg/ mL RNase A (Sigma-Aldrich) were added to 1.8 mL aliquots of phage lysate. Samples were mixed and incubated at 37 °C for 30 min in order to remove host genomic contaminants. Subsequently, 92 μL of 10% SDS and 18 μL of 10 mg/mL proteinase K (Roche) were added. Following further 30 min incubation at 65 °C, proteins were removed by two chloroform steps. Equal volumes of chloroform: isoamyl alcohol: phenol (24:24:1) (Sigma-Aldrich, Dublin, Ireland) was added, mixed thoroughly and centrifuged at 1500g for 5 min (repeated twice). The upper layer was carefully removed and transferred to a new sterile tube. This step was repeated using equal volumes of chloroform: isoamyl alcohol (24:1), centrifuged for 5 min at 6000g. DNA was precipitated from the sample using 0.3 M Sodium Acetate solution pH 5.2 and 100% ethanol (2/3 volume of extract) following incubation at room temperature for 20 min. Samples were centrifuged for 20 min at 14000g. The DNA was washed twice with 70% ethanol. DNA pellets were dried in a Rotovap (DNA speed Vac, DNA 110, Savant, Instruments. Inc. Holbrook, NY) on low power for 2–3 min to remove existing ethanol. DNA was resuspended in TE buffer pH 7.4 and dissolved at 55 °C for 1–2 h. The quality and quantity of DNA was assessed by use of Nanodrop spectrophotometer (Nanodrop ND 1000) and running DNA on an agarose gel by electrophoresis followed by visualisation.

2.5. Genome sequencing and annotation

Phage DNA was sequenced with the Illumina MiSeq system at VIB Nucleomics Core, Belgium for phages vB_CsaM_leB and vB_CsaM_leN and at the University of Liverpool, Centre for Genomic Research, UK for phage vB_CsaM_leE (herein referred to as leB, leN, and leE, respectively). Libraries for phage leB and leN were created with a custom NEBNext® Ultra™ DNA Kit and for phage leE the TruSeq DNA Nano LT library Sample Preparation Kit was used. The quality of each library preparation was assessed using the Agilent Bioanalyzer and Qubit measurements before being sequenced with paired end reads of 2 × 250 bp for phage leE and 2 × 150 bp for phage leB and leN. Reads were assembled for phage leB and leN with the CLC Bio Genomics Workbench v7.0 (Aarhus, Denmark) and for phage leE with the Spades genome assembler v.3.10 (St. Petersburg, Russia). Contigs were resolved with average coverage of 964 ×, 2,182 × and 2030 × for phages leB, leN and leE, respectively. Open reading frames (ORFs) encoding potential proteins were predicted with Glimmer (http://www.ncbi.nlm.nih.gov/genomes/MICROBES/glimmer_3.cgi) (Delcher et al., 1999) and GenemarkS (http://exon.gatech.edu/genemark/genemarks.cgi) (Besemer et al., 2001). Possible functions of these proteins were predicted with BLASTP (http://blast.ncbi.nlm.nih.gov/Blast.

cgi?PAGE = Proteins) with tRNA genes being predicted with ARAGORN (http://130.235.46.10/ARAGORN/) (Laslett and Canback, 2004). Linear genome map comparison of the phages was created using Easyfig (Sullivan et al., 2011) with comparison of genome sequences utilising TBLASTX. Coregenes (Turner et al., 2013) was used for total proteome comparisons between phages leB, leE, and leN to *Enterobacteria* phage T4 (BLASTP threshold was set at 75%) with PHACTs being used to predict the lifestyles of phages (http://www.phantome.org/PHACTS/index.php) (McNair et al., 2012).

2.6. Accession numbers

The genomes of all three phages were submitted to Genbank under accession numbers KX443552, KX431559 B and KX431560 for phage leE, leB and leN, respectively.

2.7. Electron microscopy

Transmission electron microscopy (TEM) was performed as described previously (Kelly et al., 2012b, Sambrook and Russell, 2006). Briefly, a purified phage stock was prepared using CsCl₂ gradient to achieve a titre in excess of 10⁸ pfu/mL. The sample was negatively stained with 2% (w/v) uranyl acetate on freshly prepared carbon films. Grids were analysed using a Tecnai 10 transmission electron microscope (FEI Company, Eindhoven, the Netherlands) at an acceleration voltage of 80 kV. Micrographs were taken with a MegaView G2 CCD-camera (EMSIS, Münster, Germany).

2.8. Host range profiles of phages leB, leE, and leN

The host range of each phage was examined against 21 different *Cronobacter/Enterobacter* strains, many of clinical origin, which were sourced from the Dairy Products Research Centre, DPC, Moorepark, Fermoy, Co. Cork, Ireland. Host range was determined by standard plaque assay technique as described previously (O'Flaherty et al., 2005).

2.9. Effect of thermal treatment on phage infectivity

For each phage, 3 × 250 μL phage suspensions (~1.0 × 10⁸ pfu/mL) were each placed either in the fridge at 4 °C or in water baths set to 37 °C, 45 °C, 55 °C, 60 °C, 72 °C and 90 °C respectively. Following 1 h incubation, 10 μL volumes of serially diluted samples were spotted on *C. sakazakii* ATCC BAA 894 seeded agar. Plates were left to dry and placed in the incubator at 37 °C overnight. A thermocouple (Kane-May KM330, Inlec, UK) was placed in a control tube containing 250 μL of phage buffer to ensure the desired temperature of the phage suspension was achieved for each experiment.

2.10. Effect of pH on phage infectivity

For each phage, 3 × 250 μL aliquots of each phage suspension (~1.0 × 10⁸ pfu/mL) were exposed to different pH values ranging from pH 2.0 to pH 10.0. Phages were suspended in MP buffer [50 mM Tris pH 8, 150 mM NaCl, 10 mM MgCl₂, 2 mM CaCl₂], which had been adjusted to desired pH values of 2.0, 4.0, 6.0, 7.0, 8.0, and 10.0, respectively. The samples was incubated at 37 °C for 1 h, serially diluted and spotted (10 μL) onto *C. sakazakii* ATCC BAA 894 seeded agar. Plates were left to dry and the plates were incubated at 37 °C overnight.

2.11. Preparation of phage cocktail

The phage cocktail was prepared by combining equal volumes of each individual purified phage solution at ~10⁹ pfu/mL. The high titre phage preparation was stored at 4 °C until needed.

2.12. Investigation of the antibacterial potential of the three phage cocktail against C. Sakazakii ATCC BAA 894 in four different brands of reconstituted infant milk formula

The three phage cocktail ($\sim 3 \times 10^8$ pfu/mL) was inoculated into each of the four reconstituted infant milk formulae containing $\sim 1 \times 10^4$ cfu/mL of *C. sakazakii* ATCC BAA 894 to a final volume of 10 mL, and incubated with shaking at 37 °C for 20 h. Standard plate counts were performed on HiChrome™ *Cronobacter* spp. agar every 2 h over the 20 h period to quantify surviving cells. This was achieved by setting up a staggered experiment 12 h apart to account for the hours that could not be monitored during the night. PIF was prepared aseptically using sterile water (10% w/v) and mixed to ensure homogeneity. Positive controls were represented as samples containing the inoculated culture at $\sim 1 \times 10^4$ cfu/mL + MP buffer only. Negative controls were represented as samples containing reconstituted infant formula only. All experiments were performed in triplicate and bacterial concentrations were expressed as mean cfu/mL counts and standard deviation. In addition, parallel experiments were performed using a LUX labelled *C. sakazakii* ATCC BAA 894 strain supplemented with 500 µg/mL erythromycin in a microtitre plate assay and luminescence was monitored using the *in vivo* IVIS Imaging System (Xenogen Inc).

2.13. Biofilm formation

Static microtitre plate assays based on previous studies (Cerca et al., 2007) were used to investigate biofilm formation but with modifications. Briefly, an overnight pure culture of the following strains (*C. sakazakii* ATCC BAA 894, *C. sakazakii* ATCC 29004, *C. sakazakii* DPC 6528, *C. sakazakii* DPC 8155 and *C. muytjensi* ATCC 51329) were each suspended in LB broth to an optical density equivalent of 0.5 McFarland units. A selection of different broths (TSB, BHI and LB) supplemented with 1% and 2.5% D-(+)-glucose and four different brands of infant formula were tested for their ability to aid in biofilm formation. Cells were subsequently diluted into the appropriate solution to give a starting inoculum of $\sim 1 \times 10^4$ cfu/mL. 200 µL of the test sample (BHIglucose(g), LBg, TSBg or PIF) containing the culture was added to a sterile 96 well plate. 200 µL of test media alone was added as a negative control. All wells were seeded in triplicate. Microtitre plates were incubated at 37 °C for 48 h to allow biofilm formation. Following incubation, media was removed and wells were washed three times with 200 µL of saline (Ringer's) solution using a multichannel pipette (Gilson) to remove media and planktonic cells. Microtitre plates were placed in a 50 °C incubator in an inverted position and allowed to dry for 1 h. Formed biofilms were stained with 200 µL of 1% crystal violet solution (Sigma Aldrich, UK) for 10 min at room temperature. The wells were washed to remove excess dye and left to dry before visually comparing the intensity of staining in the wells.

2.14. Biofilm prevention with phage cocktail

Following the identification of two *C. sakazakii* strains (*C. sakazakii* ATCC BAA 894, and *C. sakazakii* DPC 6528) that were capable of forming strong biofilms when grown in different types of infant formula (outlined above, *2.13.*), the ability of the phage cocktail to prevent biofilm formation by these strains was investigated using a plate staining and live cell count assay.

2.14.1. Plate staining assay

The plate staining assay was performed as per the biofilm formation method described above with the following adjustments. At the start of the assay, 50 µL of the phage cocktail ($\sim 3 \times 10^8$ pfu/mL) was added to 200 µL of *C. sakazakii* culture in one set of wells in the microtitre plate, while 50 µL of MP buffer was added to a second set of wells also containing 200 µL of *C. sakazakii* culture. PIF alone was used as a negative control. The plates were incubated at 37 °C for 48 h, after which the media was poured off and biofilms were stained as described above. The stain was removed and the wells were washed gently. The plates were left to dry, after which 30% glacial acetic acid was added to solubilise the stain. The biofilm prevention ability of the phage cocktail was assessed by examining the optical density of the wells spectrophotometrically at OD_{590nm}.

2.14.2. Live cell count assay

The live cell count assay was performed as per the plate staining method with the following modifications. At the end of the 48 h incubation period, the media was removed and biofilms were washed three times with ringers to remove media and planktonic cells. Biofilms were resuspended by mechanical scraping in 100 µL of saline solution (Ringer's) and immediately diluted and plated on to HiChrome™ *Cronobacter* spp. agar plates to determine cfu/mL counts.

3. Results

3.1. Phage isolation and morphology

Following routine enrichments with the host bacterial strain *C. sakazakii* ATCC BAA 894, phage leB, leE and leN were isolated from slurry samples obtained from a cattle farmer in Clonakilty, Co. Cork, Ireland. All three phages formed clear plaques on the host strain following overnight incubation at 37 °C. Plaque size ranged from 2.0 mm to 4.5 mm in diameter on 0.4% semi-solid agar. Each phage was purified by successive single plaque isolation and propagation on the host strain. High titre phage suspensions of $\sim 10^9$ pfu/mL, were recovered from LB agar plates using 5 ml of MP buffer and stored at 4 °C until needed. Morphological characterisation of phages using TEM indicated that each belonged to the family *Myoviridae* (Fig. 1). The micrographs revealed phage leB, leE, and leN, to have moderately
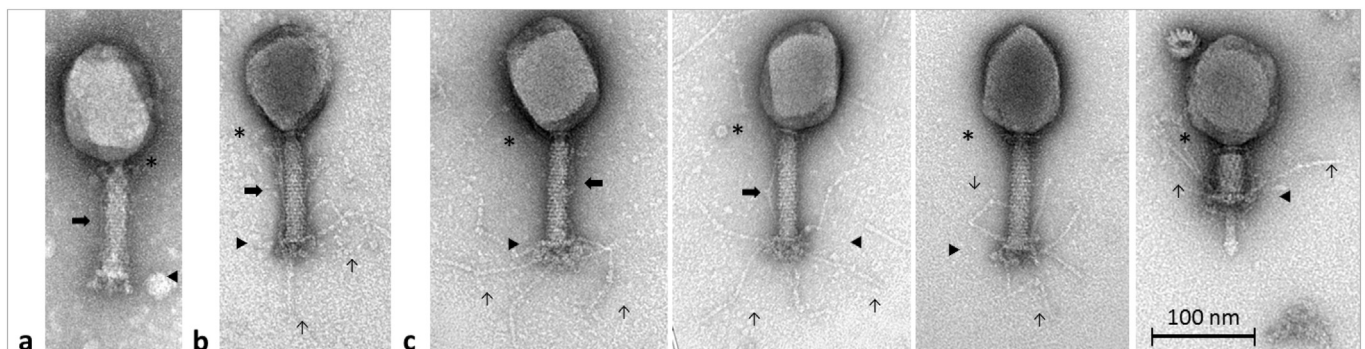

Fig. 1. Electron micrographs of leB (**a**), leN (**b**), and of four phage particles of leE (**c**), with either extended or contracted (right) tail sheaths. Structural details: collar (✱), tail fibers folded upwards along the tail sheath (➞), extended tail fibers with typical knee-like hinges (↑), baseplates (►). Uranyl acetate (2%) was used to stain the phages for transmission electron microscopy.

Table 1
Phage dimensions.

Phage	Head (nm)	Tail length (nm) incl. collar and baseplate	Free tail fibers (nm)
leB	119.4 ± 3.6 × 89.7 ± 3.6 ($n = 6$)	116.0 ± 1.5 ($n = 5$)	Not visible on micrographs
leE	115.1 ± 3.5 × 90.3 ± 6.9 ($n = 15$)	113.8 ± 2.0 ($n = 11$)	111.5 ± 7.4 ($n = 15$)
leN	113.0 ± 4.1 × 91.9 ± 2.5 ($n = 16$)	114.6 ± 2.4 ($n = 18$)	114.0 ± 7.9 ($n = 15$)

elongated heads accompanied by contractile tails, indicating that each belongs to the family *Myoviridae*, subtype A2 (Ackermann, 2001). Phage dimensions are represented in Table 1.

3.2. Results from genome sequencing and annotation

Through sequencing, genomes with sizes of 177,907 bp, 181,570 bp and 179,516 bp were obtained for phages leB, leE and leN, respectively, with a G + C% content of 45%. Following bioinformatic annotation a total number of 281 ORFs, 284 ORFs and 286 ORFs were predicted for phages leB, leE and leN, respectively, with it being possible to annotate about 40% (107 ORFs, 111 ORFs and 109 ORFs for leB, leE and leN, respectively) of their ORFs with a possible function (Supplementary Table 1). Additionally, one tRNA gene for glycine was predicted to be present in the genomes of all three phages (see supplemental data).

The genomes of these phage were found to be highly similar to each other (99% identity with 94–99% coverage) with their closest known relatives at the nucleotide level being found to be T4 like phages *Cronobacter* phage GAP161 [JN882287.1] and *Citrobacter* phage Margaery [KT381880.1] (98–99% identity with 92–97% coverage).

At the proteomic level phages leB, leE and leN were found to have strong homology with Enterobacteria phage T4 [NC_00866] (Fig. 2), with phage leB sharing 116 homologs and phages leE and leN sharing 117 homologs with that of T4..It has been suggested that phages which share more > 40% of their proteins fall within phage genus boundaries (Lavigne et al., 2009). As these phages were found to share > 40% of their gene products they could be placed within the genus T4 virus.

PHACTs software confidently predicted that all these phages had lytic lifestyles, with average probability values exceeded 0.5 for phages leB (0.658), leE(0.663) and leN(0.659). Consistent with this prediction was that genes for integrases, excisionases or phage repressors were not identified in their genomes. Furthermore, no genes for toxins were detected.

3.3. Phage host range

Host range results revealed that our phages infected 12 (leB), 12 (leE) and 13 (leN) of the 21 strains tested. All phages appeared to predominantly infect strains within the *C. sakazakii* species possessing the ability to plaque and lyse 10 (leB), 10 (leE) and 11 (leN) of 15 *C. sakazakii* strains tested. In addition, as evidenced by the presence of plaques, all three phages were capable of infecting across genera and species, lysing *C. muytjensi* strain ATCC 51329 and *E. cloacae* strain NCTC 11933. However, when tested against other *Enterobacter/Cronobacter* species (*E. gergorviae* NCTC 11431, *E. cloacae* NCTC 11590, *C. malonaticus* DPC 6531 and *E. aerogenes* NCTC 10006), lysis was not evident. The combination of all three phages resulted in 73% coverage across the *Cronobacter* strains tested (Table 2).

3.4. Effect of temperature and pH on phage infectivity

All phages were tested for their ability to retain infectivity following exposure to a range of different temperature extremes. In general, it was observed that all phage retained infectivity between 4 and 50 °C. While a ~2.5 log reduction in titre was observed for phage leN following 1 h exposure to 50 °C, the titres of phage leB and leE were reduced by ≤ 6 logs during this temperature challenge. No viable phages could be recovered from the lysates exposed for 1 h to 60 °C, 72 °C or 90 °C (Fig. 3A).

The effect of pH on phage infectivity was investigated by exposing phage leB, leE and leN to a range of acidic, neutral and basic environments for 1 h. In general, all three phages retained activity over pH 6 to pH 10. However, when exposed pH 10 for 1 h, a 5 log, 3.5 log and 6 log reduction in phage titre was observed for phages leB, leE, and leN, respectively. No viable phages could be recovered from lysates following exposure to both pH 2 and pH 4 for 1 h (Fig. 3B).
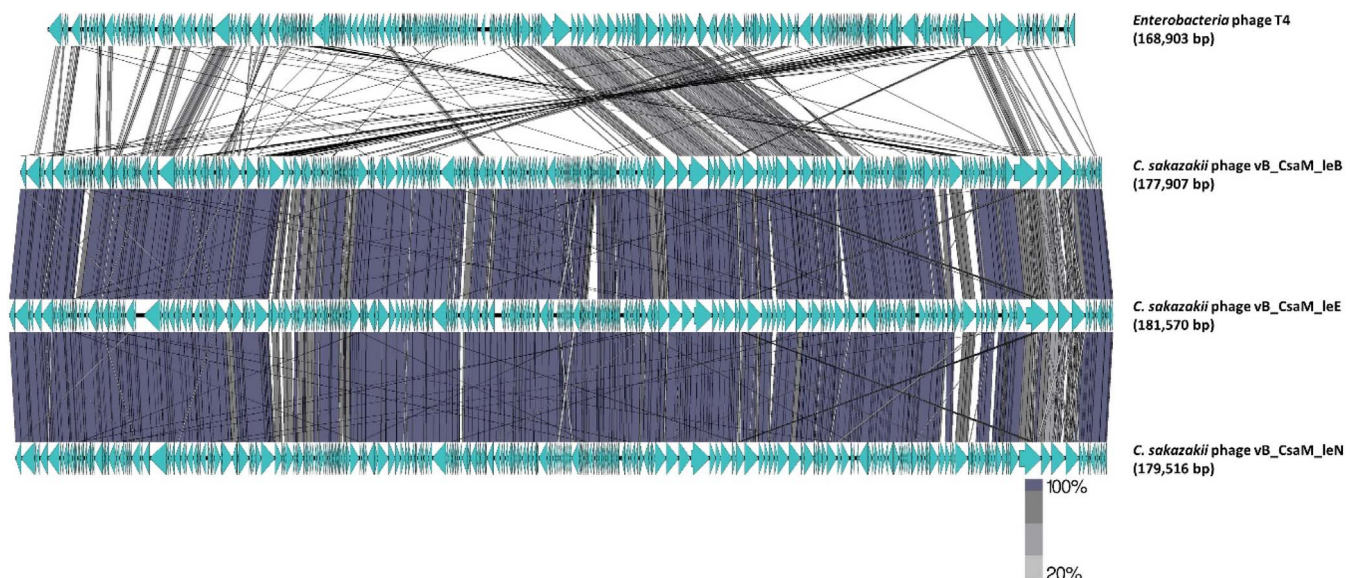
Fig. 2. Linear map comparison of the genomes of *Enterobacteria* phage T4 and of *Cronobacter sakazakii* phage leB, leE and leN, created using EasyFig, utilising tBLASTx to make comparison between genomes. The genome maps are comprised of blue arrows indicating locations of genes among the different phage genomes and lines between genome maps indicate level of homology. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2

Host range of *Cronobacter* phages leB, leE, and leN, where phage lysis is represented in cfu/mL. (−) denotes – no lytic activity. Strains were obtained from the Dairy Products Research Centre, DPC, Moorepark, Fermoy, Co. Cork, Ireland.

Strain name	Origin	Phage lysis (cfu/mL)		
		leB	leE	leN
C. sakazakii ATCC BAA 894	Powdered infant formula	1.35E + 08	1.20E + 08	1.01E + 08
C. sakazakii DPC 6529	Tracheal aspirate	–	–	5.10E + 02
C. sakazakii DPC 6522	Blood	8.00E + 01	1.19E + 02	2.10E + 02
C. sakazakii DPC 6523	Cerebral spinal fluid	–	–	–
C. sakazakii NCT 8155	Tin of dried powdered milk	1.20E + 08	1.17E + 09	1.05E + 09
C. muytjensi ATCC 51329	–	2.20E + 09	1.20E + 08	1.90E + 07
C. sakazakii DPC 6524	Stool	–	–	–
C. sakazakii DSM 4485	Child's throat	7.50E + 05	5.00E + 01	1.70E + 02
C. malonaticus DPC 6531	Brain tumour	–	–	–
C. sakazakii ATCC 29004	–	1.89E + 07	2.70E + 08	4.50E + 08
C. sakazakii DPC 6530	Bronchial alveolar lavage	4.20E + 02	3.30E + 04	6.80E + 04
E. cloacae NCTC 11933	Human/clinical isolate	1.00E + 01	1.57E + 08	1.00E + 02
C. sakazakii DPC 6527	Blood	4.00E + 01	3.10E + 02	2.00E + 01
E. gergorviae NCTC 11431	Human urinary tract	–	–	–
E. cloacae NCTC 11590	–	–	–	–
E. aerogenes NCTC 10006	Sputum	–	–	–
C. sakazakii DPC 6526	Blood	1.10E + 02	1.98E + 09	4.00E + 01
C. sakazakii ATCC 12868	–	–	–	–
C. sakazakii DPC 6528	Cerebral spinal fluid	7.80E + 07	6.50E + 08	8.70E + 07
C. sakazakii DPC6525	Urine	–	–	–
C. sakazakii NCTC 11467	Human throat	1.76E + 08	3.10E + 02	2.00E + 01

3.5. Investigation of the antibacterial potential of the three phage cocktail against C. sakazakii ATCC BAA 894 in four different brands of reconstituted infant milk formula

Cronobacter phages leB, leE, and leN were combined as part of a phage cocktail ($\sim 3 \times 10^8$ pfu/mL) to assess their ability to inhibit the growth of *Cronobacter sakazakii* ATCC BAA 894 in four different brands of infant formula. Bacterial counts of $\sim 1 \times 10^4$ cfu/mL were used for each biocontrol experiment, as this bacteria:phage concentration ratio was determined as the maximum inhibitory limit to successfully prevent outgrowth of *C. sakazakii* following the 20 h challenge. This contaminating concentration of cells is well in excess of the expected levels of *C. sakazakii* contamination typically found in reconstituted powdered infant formula (< 1 cell/100 g of PIF) (van Acker et al., 2001) and hence represents a clear indication of phage biocontrol potential. Following a 5 h incubation period, *C. sakazakii* concentrations were reduced to below the detection limit (< 10 cfu/mL) in samples challenged with the phage cocktail. This level of inactivation was maintained over the 20 h challenge, where samples in the absence of the phage cocktail had reached bacterial concentrations in excess of 10^9 cfu/mL (Fig. 4). Parallel experiments, where the LUX labelled *C.*

sakazakii ATCC BAA 894 strain was employed to monitor cell viability, gave a similar result. Luminescence could not be detected from the wells challenged with the phage cocktail over the 20 h period, where significant relative light units (RLUs) were detected in the wells containing *C. sakazakii* LUX cells + MP buffer only (Fig. 5). At the end of the assay all wells were serially diluted and plated to determine the number of viable cells present in the sample. While the wells containing *C. sakazakii* LUX strain alone had reached a concentration of $\sim 10^9$ cfu/mL, bacterial concentrations were reduced below the limit of detection (< 10 cfu/mL) in the wells challenged with the phage cocktail.

3.6. Biofilm formation

Static microtitre plate assays were used to investigate biofilm formation by selected *Cronobacter* strains. Strain selection was dependant on individual phage infection profiles when tested against the 21 *Cronobacter* strains. Based on phage susceptibility testing (Table 2), five strains were selected for biofilm formation assessment. Biofilm formation experiments were performed in BHI, TSB and LB broths, supplemented with 1% and 2.5% glucose and also in four different brands of
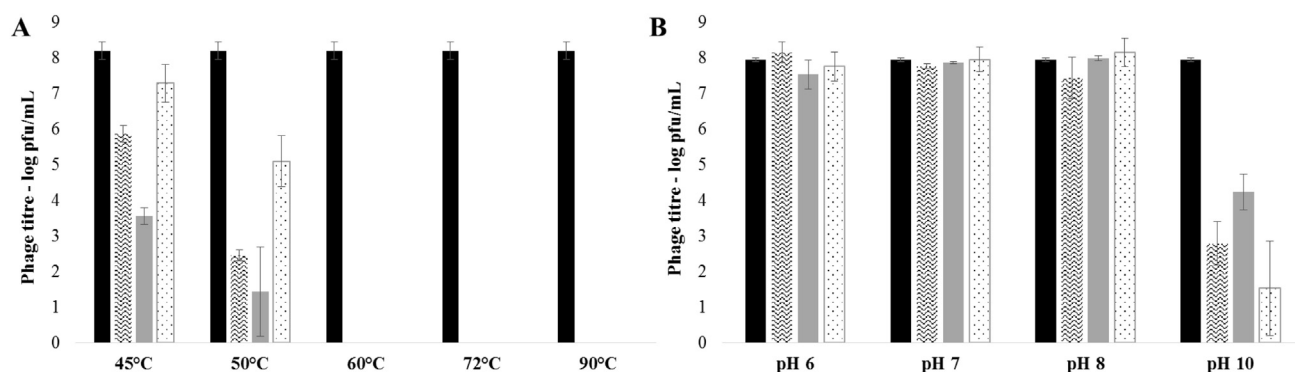
Fig. 3. (A) Effect of temperature on phage viability following 1 h exposure to 4 °C, 37 °C, 45 °C, 50 °C, 60 °C, 72 °C, and 90 °C. (■) represent an average of initial phage titres for all three phages tested. (▨), (▦), and (▫) columns represent phage leB, leE, leN 1 h temperature challenge experiments, respectively. Assays were performed in triplicate and phage titres were expressed as the mean ± standard deviation. (B) Effect of pH on phage viability following 1 h exposure to acidic and alkaline environments. (■) represent an average of initial phage titres for all three phages tested. (▨), (▦), and (▫) represent leB, leE, and leN, 1 h pH challenge experiments. Assays were performed in triplicate and phage titres were expressed as the mean ± standard deviation.
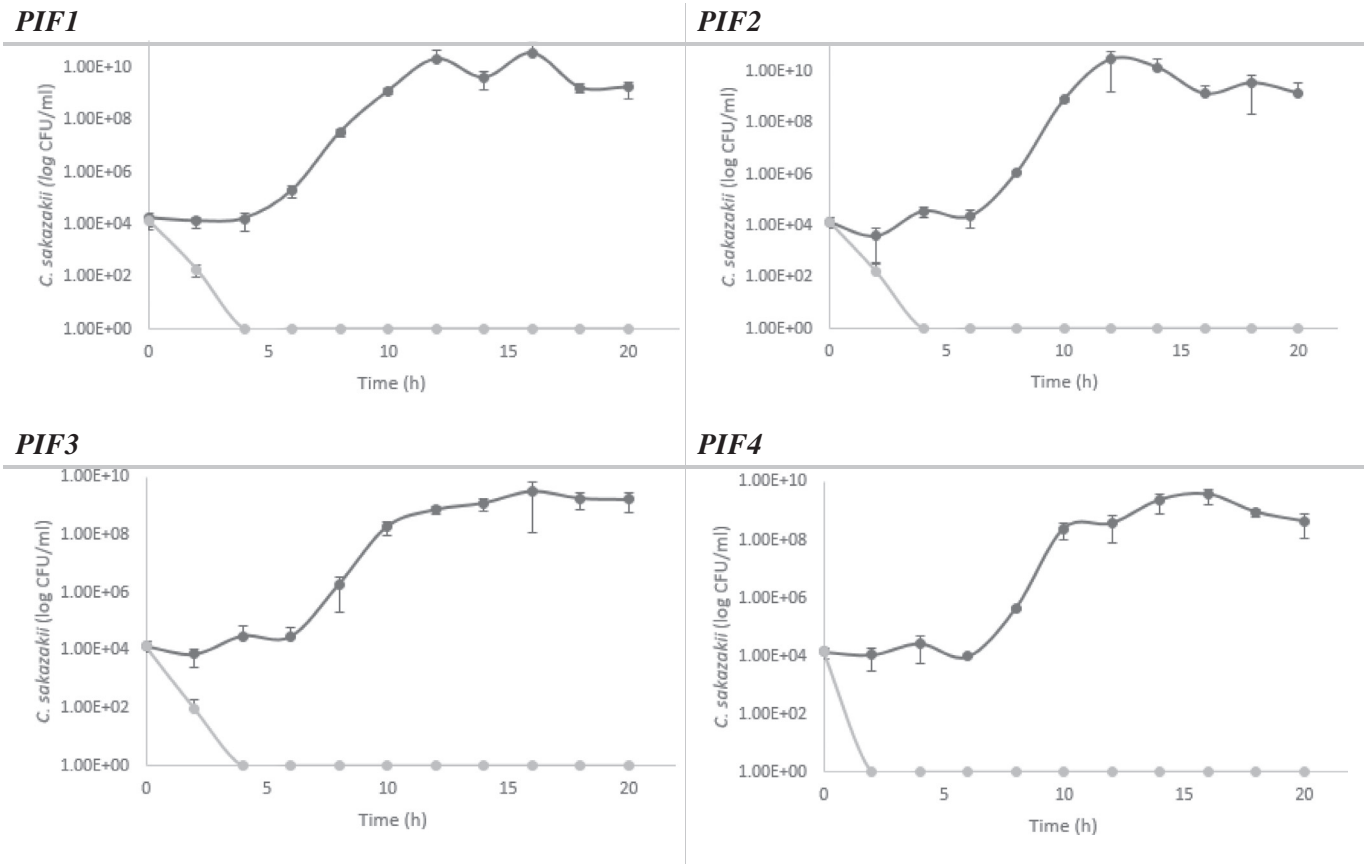
Fig. 4. Growth inhibition of *C. sakazakii* ATCC BAA 894 using the phage cocktail in four different brands of reconstituted infant milk formula (PIF1, PIF2, PIF3, and PIF4). Dark grey lines represent *C. sakazakii* grown in the absence of the phage cocktail. Light grey lines represent *C. sakazakii* grown in the presence of the phage cocktail ($\sim 3 \times 10^8$ pfu/mL). Growth inhibition of *C. sakazakii* was determined by cfu/mL counts. Assays were performed in triplicate and bacterial concentrations were expressed as the mean $\pm$ standard deviation.
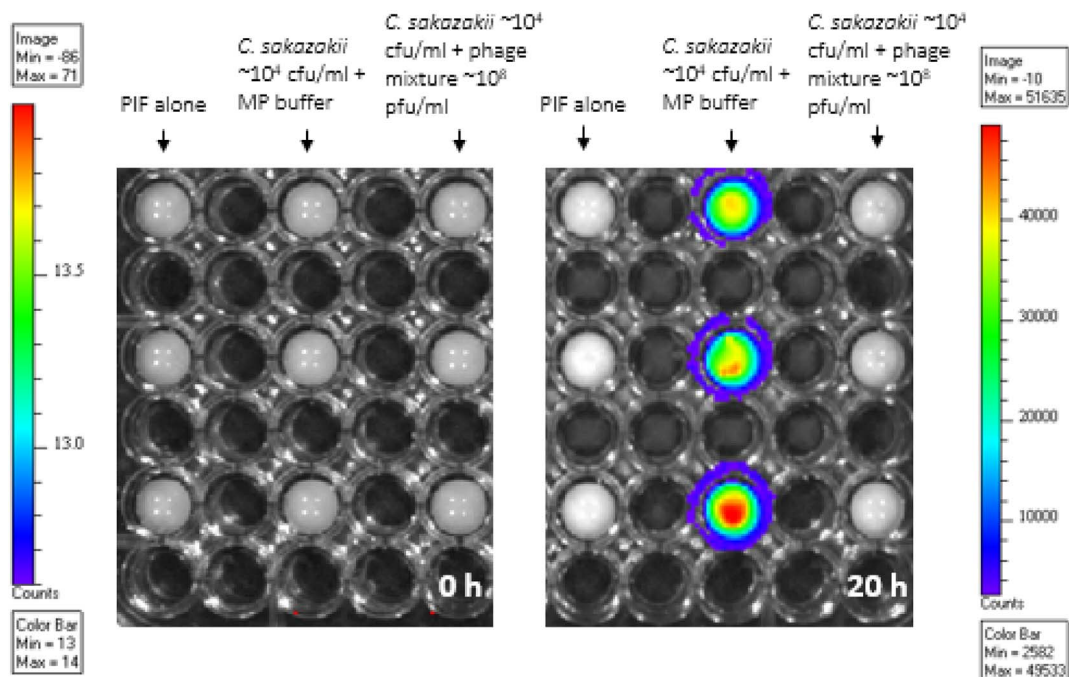
Fig. 5. Bioluminescence assay demonstrating the biocontrol potential of the phage cocktail following a 20 h challenge against *C. sakazakii* ATCC BAA 894 LUX in PIF1 brand of reconstituted infant milk formula. PIF1 is representative of PIF2, PIF3, and PIF4. Lane 1 contains PIF alone. Lane 2 contains *C. sakazakii* LUX + MP buffer. Lane 3 contains *C. sakazakii* LUX + the phage cocktail at $\sim 3 \times 10^8$ pfu/mL.
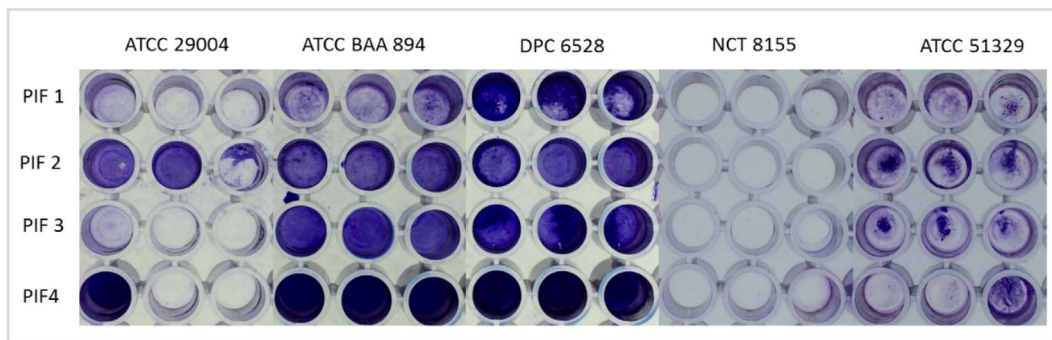
Fig. 6. Biofilm formation by *Cronobacter* spp. after 48 h incubation at 37 °C in different brands of reconstituted infant formulae. Confirmation of strong biofilm formation by *C. sakazakii* ATCC BAA 894 and *C. sakazakii* DPC 6528 is indicated by intense crystal violet staining in wells. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

reconstituted infant milk formulae. Wells were inoculated with the selected *Cronobacter* strains and incubated at 37 °C for 48 h, followed by the addition of a 1% crystal violet stain to allow for direct visualisation of biofilms. It was found that the five *Cronobacter* strains were incapable of forming biofilms in BHI, TSB and LB alone and when supplemented with different concentrations of glucose (results not shown). However, *C. sakazakii* ATCC BAA 894 and *C. sakazakii* DPC 6528 were found to be strong biofilm producers when grown in each of the four different brands of PIF tested (Fig. 6).

3.7. Biofilm prevention - plate staining and live cell count assay

The first step of biofilm formation is initial attachment of bacteria to a solid surface. Accordingly, the effect of a high-titre, three-phage preparation ($\sim 3 \times 10^8$ pfu/mL) was assessed for its ability to prevent biofilm formation by *C. sakazakii* ATCC BAA 894 and *C. sakazakii* DPC 6528 in different brands of infant formula. Static microtitre plate assays were used and monitored by plate staining and live cell count assays.

Plates were stained following the 48 h biofilm prevention challenge. The stain was subsequently solubilised and measured at OD_{590nm} to further assess the ability of the phage cocktail to prevent biofilm formation. This data is presented in Fig. 7. Results revealed the phage cocktail to be very effective at preventing biofilm formation, which was evident when compared to the control wells where strong biofilm formation was observed. In addition, parallel plate count experiments were performed in conjunction with the plate staining and solubilisation assays. Following the 48 h biofilm prevention challenge, bacterial concentrations were reduced to below the limit of detection (< 10 cfu/mL) in the wells containing the phage cocktail, when compared to the control wells, where bacterial growth had reached levels of $\sim 10^9$ cfu/mL, again confirming the strong anti-biofilm potential of the phage cocktail against *C. sakazakii* in different brands of infant formula.
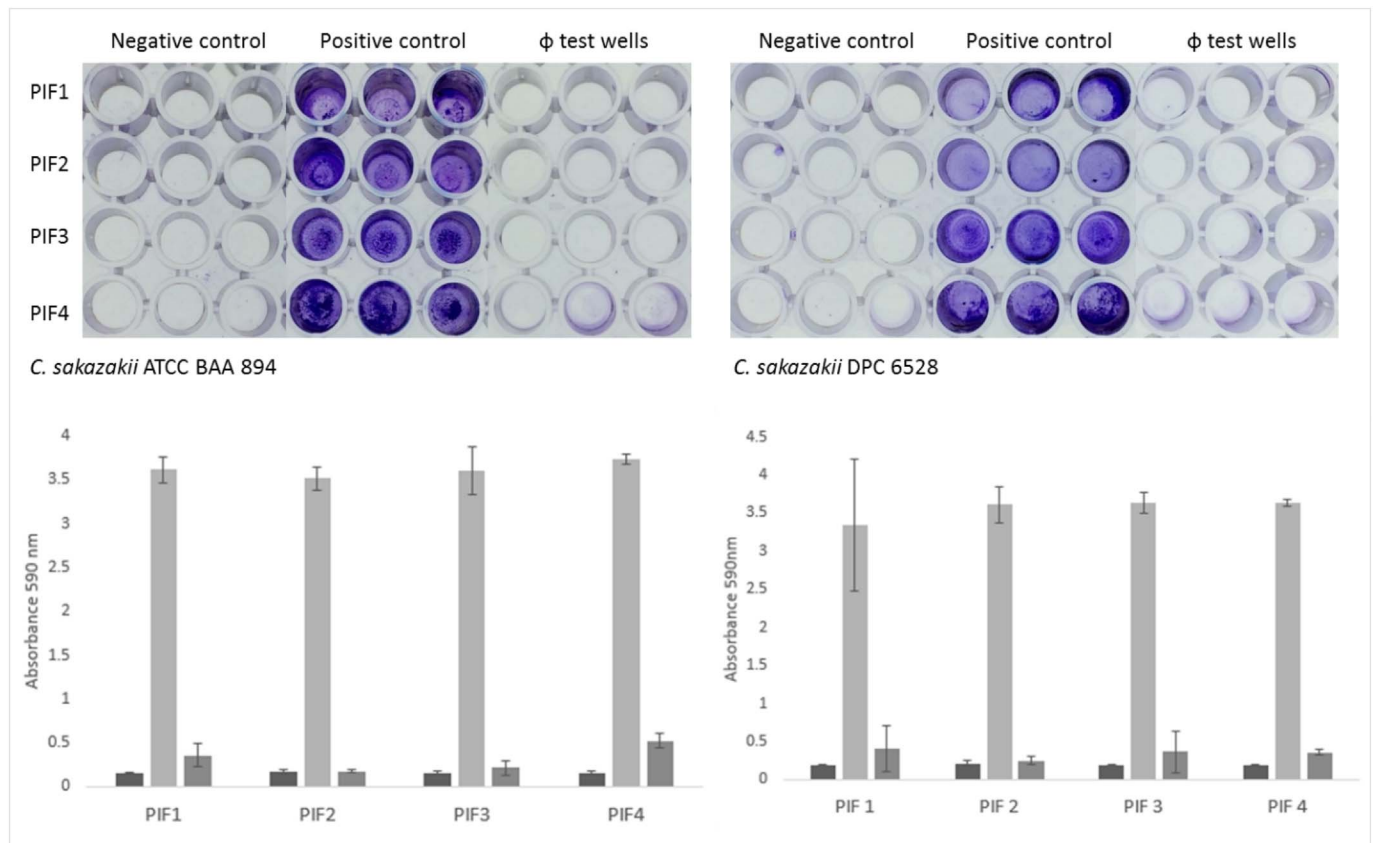
Fig. 7. Biofilm plate staining and solubilisation assays demonstrating the efficacy of the phage cocktail at preventing biofilm formation by *C. sakazakii* ATCC BAA 894 and *C. sakazakii* DPC 6528 following a 48 h challenge. () PIF only, () PIF containing *C. sakazakii* cells and MP buffer, and () PIF containing *C. sakazakii* cells and the three phage cocktail.

4. Discussion

Consumption of contaminated reconstituted infant milk formula and concomitant *C. sakazakii* infections in infants and neonates has resulted in international efforts to improve existing pathogen control processes. Current intervention strategies have fallen short of providing powdered infant formula that is free from *C. sakazakii* contamination. Many novel technologies have been proposed in an effort to address this problem, some of which include, the use of hot water, gamma irradiation and high pressure processing to inactivate the pathogen prior to ingestion (Edelson-Mammel and Buchanan, 2004; Gonzalez et al., 2006; Lee et al., 2006). The natural biotherapeutic properties of phages are well recognised throughout the world, with many studies providing promising evidence of their biocontrol potential against several leading and emerging foodborne pathogens, while also demonstrating their suitability for use at each stage of the farm-to-fork continuum (Endersen et al., 2014; Goodridge and Bisha, 2011).

The rationale and motivation for performing the current study relates to the aforementioned reconstituted infant milk formula dilemma, and its associated health risks, in addition to the possibility of employing naturally occurring phages to alleviate this problem. We describe the isolation and characterisation of three novel *Cronobacter* phages and subsequent utilisation of a three phage preparation for biocontrol of *C. sakazakii* in different brands of infant formula. In addition to the activity and stability of the phages in infant formula, the anti-biofilm potential of the phage cocktail was also demonstrated. Three *Cronobacter* phages were isolated following routine enrichments of a variety of different environmental samples, including soil, slurry, activated sludge, river water, moss, grass, and wheat. While multiple phages were isolated from these sources, three phages which demonstrated particular promise for use as biocontrol agents against *C. sakazakii* were selected for further characterisation. Sewage has been recognised as a primary niche for many *Enterobacteriaceae*, so it is not uncommon that these and other *Cronobacter* phages were isolated from effluent environments (Kim et al., 2007; Zuber et al., 2008).

It is well documented that temperate phages from *Enterobacteriaceae* are known to harbour important bacterial virulence genes that can be readily transferred between bacteria. Determining phages to be absent of lysogenic capabilities is a primary consideration that must be taken into account if they are to be used in food related applications (Brüssow et al., 2004; Faruque and Mekalanos, 2012). As these phages have been determined to have lytic lifestyles and do not possess genes for toxic proteins they meet the required properties of phage intended for phage therapy applications. In addition, these phages have been determined to belong to the phage genus of.T4 virus,.phages from this genus have been deemed safe for phage therapy applications (Bruttin and Brüssow, 2005; Denou et al., 2009).

Phage susceptibility was assessed using 21 strains from *Cronobacter/ Enterobacter* genera. The broad host range attributed to phages leB, leE, and leN correlates well with previous findings. Loessner et al. (1993), demonstrated the broad host range of *Enterobacter cloncae* isolated phages, which were capable of cross infecting *Panteo agglomerans, C. sakazakii, E. coli* and *Serratia marcescens* (Loessner et al., 1993). In addition Zuber and co-authors also observed the broad host range capabilities of five combined *E. sakazakii* phages with an infection profile extending across several genera (Zuber et al., 2008).

Determining the environmental stability of phages considered for biocontrol in food and the food processing environment is essential. Common environmental extremes often associated with the food and food industry include low pH and high temperature pressures. Establishing the stress limits of each phage, permits appropriate and optimal application, and more effective biocontrol results overall. Each of the three phages were challenged for their ability to initiate host infection following exposure to a range of pH and temperature extremes. In general, all three phages retained maximum infectivity between pH 6 and pH 8. These findings compare well with those

previously demonstrated, where the researchers found that *E. coli* phage T7 demonstrated optimum pH stability between pH 6 and pH 8 (Kerby et al., 1949). Similarly, in another study, phage T4 demonstrated optimum pH stability from pH 6–7.4 (Kłak et al., 2010), whereas *coli*phage λ demonstrated very high stability across a wide pH range, where no significant decrease in titre was observed from pH 3–11 (Jepson and March, 2004). Temperature plays a fundamental role in the survival of phages. Optimal attachment, penetration, multiplication and the length of the latent period (in the case of temperate phages) are often dictated by temperature (Nasser and Oman, 1999; Olson et al., 2004; Yates et al., 1985). Higher than optimal temperatures are thought to extend the latent period, while lower than optimal temperatures are often thought to result in a reduced multiplication rate (Tey et al., 2009). In this current study, phages leB, leE, and leN were found to retain infectivity from 4 to 50 °C, with optimum infectivity occurring between 4 and 37 °C. Yang et al. (2010) evaluated the temperature stability of *Acinetobacter* phage AB1 which showed a higher tolerance to extreme temperatures, compared to phage leB, leE, and leN. Phage AB1 was very heat stable at 50 °C and 60 °C, as only a slight drop in phage titre was observed following a 1 h challenge. At 70 °C a significant number of the phages were inactivated, with only 0.52% of viable phages remained in the sample following 1 h incubation. Similarly, at 90 °C, > 99% of the phages had lost their infective ability following 15 min exposure to this temperature (Yang et al., 2010). In another study, the thermo-tolerance of staphylococcal phages was evaluated in milk. Researchers found that these phages were most stable when maintained at 4 °C for 8 h, while observing a 20–30% increase in phage inactivation when challenged at 22 °C and 37 °C, respectively. While both phages survived very short exposure to high temperature extremes (72 °C for 15 s), phage titres were reduced below the limit of detection after 1 min (Garcia et al., 2009).

In recent years, the application of phages for biocontrol of harmful pathogens in both pre- and postharvest food environments has become a noteworthy treatment alternative due to the increasing worldwide problem of antibiotic resistance and the reduction in the development of novel technologies for pathogen control. In this study, phage mediated killing and growth inhibition was demonstrated using a three-phage cocktail against *C. sakazakii* in four different brands of infant formula. Overall, it was found that the phage cocktail was effective at reducing contaminating *C. sakazakii* cells below the limit of detection (< 10 cfu/mL) when bacterial concentration were used at a starting concentration of $\leq 1 \times 10^4$ cfu/mL. This bacteria to phage concentration ratio was the maximum inhibitory limit to successfully prevent outgrowth of *C. sakazakii* over the 20 h challenge. It is interesting to note that despite slight differences in formula constitution, the growth rate of *C. sakazakii* was not affected for all formulas tested, nor did it impede phage infection. Our biocontrol results were similar to those outline by Zuber et al. (2008), who found that a high dose of 10^8 pfu/mL of phage could effectively sterilise broth contaminated with both high (10^6 cfu/mL) and low (10^2 cfu/mL) pathogen counts (Zuber et al., 2008). In addition, Kim et al., 2007 reported inhibition of *C. sakazakii* at a range of different incubating temperatures when high titre 10^9 pfu/mL single phage suspensions were used (Kim et al., 2007). Both Zuber et al. (2008), and Kim et al. (2007) also reported that when challenging low numbers of the contaminating pathogen with a reduced dose of the phage preparation, a delay in phage replication was evident up until bacterial counts had reached $\sim 10^4$ cfu/mL. Only after crossing this threshold did phage replication occur but overall did not result in complete sterilisation of the sample. This would suggest that high concentrations of phage-based preparations are needed for complete sterilisation of a low quantity of target cells (Kasman et al., 2002; Weld et al., 2004). It is reported that the level of *C. sakazakii* contamination in PIF is very low (< 1 bacterial cell/100 g) (Holý and Forsythe, 2014), therefore improper handling is thought to contribute to bacterial growth and subsequent development of infection among low birth weight infants and neonates. The

contaminating levels of *C. sakazakii* used in this study were much higher than levels typically found in PIF which demonstrates the efficacy of this phage cocktail for biocontrol of *C. sakazakii* in reconstituted infant formula.

Many phage-based intervention strategies have been documented against biofilm formation (Doolittle et al., 1995; Kelly et al., 2012a). Biofilms are a significant source of repeat contamination in dairy manufacturing plants, threatening the quality and safety of the manufactured produce, in addition to causing extensive economic losses to the food manufacturing industry due to produce recalls and corrosion of equipment (Gram et al., 2007). In addition, the persistence of foodborne pathogens, like *C. sakazakii*, on feeding areas, on the surfaces of equipment used in formula preparation, and within enteral feeding tubes, are also thought to contribute to neonatal outbreaks (Kim et al., 2007). Due to the high fatality rate associated with *C. sakazakii* infection among infants and neonates (up to 80%), the destructive nature of the organism towards survivors, the ability of *C. sakazakii* to form biofilms and because of the problems associated with biofilm treatment and removal, it is clear that biofilm prevention is the preferable biocontrol option. We evaluated the biocontrol potential of the phage cocktail for its ability to prevent biofilm formation. Results indicated that the phage cocktail was very effective in preventing the establishment of biofilm where biofilm formation could not be detected for both strains tested for up to 2 days post-treatment. These results were similar to those outlined by Kelly et al. (2012a, 2012b), who demonstrated the efficacy of a modified phage cocktail for the prevention of *Staphylococcus aureus* biofilm formation. These researchers observed complete inhibition of biofilm formation over a 48 h time period (Kelly et al., 2012a). In addition, other studies have described the use of phages in an attempt to reduce biofilm formation. Hydrogel-coated catheters were impregnated with phage preparations and resulted in an approximate 90% reduction in both *Proteus mirabilis* and *Escherichia coli* formed biofilms when compared with un-treated controls (Carson et al., 2010). The anti-biofilm formation potential demonstrated by the phage cocktail in this study is promising and may have potential in future applications as perhaps a stand-alone treatment or used in conjunction with alternative antibacterials for the control and/or eradication of *C. sakazakii* contamination in food and the food production environment.

5. Conclusion

International efforts to develop novel strategies to improve the microbiological safety of PIF are ongoing. Although many advances in pathogen control have been made in recent years, cost and efficiency as well as the safety of preparation and nutritional stability are a problem. *C. sakazakii* contamination of PIF still remains an issue for public health concern. Phages have long been recognised for their inherent ability to control pathogens in food and food related environments, where other treatments have failed to do so. More and more, these non-destructive particles are continually gaining acceptance around the world as a viable alternative to antibiotics, and other treatment methods following the rise of multi-drug resistant bacteria. Accordingly, this study has presented three phages active against the infant formula associated pathogen, *C. sakazakii*. The phages were characterised in terms of temperature and pH stability and were combined to demonstrate the efficacy of using a three phage cocktail for biocontrol of *C. sakazakii* in four different brands of infant formula. In addition, their ability to effectively inhibit biofilm formation was also established. Our results highlight the promising potential of these phages for biocontrol of *C. sakazakii* contamination in infant formula and also as preventative biotherapeutic agents against biofilm formation. It should be noted that the isolation of additional phages capable of inactivating more resistant strains of *C. sakazakii* is a worthwhile endeavour. While 73% coverage is impressive, it could certainly be improved upon. Additional studies are now required in order to further this concept into practical use.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ijfoodmicro.2017.04.009.

References

Ackermann, H.W., 2001. Frequency of morphological phage descriptions in the year 2000. Arch. Virol. 146, 843–857.

Al-Nabulsi, A.A., Osaili, T.M., Al-Holy, M.A., Shaker, R.R., Ayyash, M.M., Olaimat, A.N., Holley, R.A., 2009. Influence of desiccation on the sensitivity of *Cronobacter* spp. to lactoferrin or nisin in broth and powdered infant formula. Int. J. Food Microbiol. 136, 221–226.

Besemer, J., Lomsadze, A., Borodovsky, M., 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res. 29, 2607–2618.

Bruttin, A., Brüssow, H., 2005. Human volunteers receiving *Escherichia coli* phage T4 orally: a safety test of phage therapy. Antimicrob. Agents Chemother. 49, 2874–2878.

Bigot, B., Lee, W.-J., McIntyre, L., Wilson, T., Hudson, J., Billington, C., Heinemann, J., 2011. Control of *Listeria monocytogenes* growth in a ready-to-eat poultry product using a bacteriophage. Food Microbiol. 28, 1448–1452.

Brady, C., Cleenwerck, I., Venter, S., Coutinho, T., De Vos, P., 2013. Taxonomic evaluation of the genus *Enterobacter* based on multilocus sequence analysis (MLSA): Proposal to reclassify *E. nimipressuralis* and *E. amnigenus* into *Lelliottia* gen. nov. as *Lelliottia nimipressuralis* comb. nov. and *Lelliottia amnigena* comb. nov., respectively, *E. gergoviae* and *E. pyrinus* into *Pluralibacter* gen. nov. as *Pluralibacter gergoviae* comb. nov. and *Pluralibacter pyrinus* comb. nov., respectively, *E. cowanii*, *E. radicincitans*, *E. oryzae* and *E. arachidis* into *Kosakonia* gen. nov. as *Kosakonia cowanii* comb. nov., *Kosakonia radicincitans* comb. nov., *Kosakonia oryzae* comb. nov. and *Kosakonia arachidis* comb. nov., respectively, and *E. turicensis*, *E. helveticus* and *E. pulveris* into *Cronobacter* as *Cronobacter zurichensis* nom. nov., *Cronobacter helveticus* comb. nov. and *Cronobacter pulveris* comb. nov., respectively, and emended description of the genera *Enterobacter* and *Cronobacter*. Syst. Appl. Microbiol. 36, 309–319.

Breeuwer, P., Lardeau, A., Peterz, M., Joosten, H., 2003. Desiccation and heat tolerance of *Enterobacter sakazakii*. J. Appl. Microbiol. 95, 967–973.

Bren, L., 2006. Bacteria-eating virus approved as food additive. FDA Consum. 41, 20–22.

Brüssow, H., Canchaya, C., Hardt, W.D., 2004. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. Microbiol. Mol. Biol. Rev. 68, 560–602.

Carlson, K., 2005. Appendix: working with bacteriophages: common techniques and methodological approaches. In: Bacteriophages: Biology and Applications, pp. 437–494.

Carlton, R., Noordman, W., Biswas, B., De Meester, E., Loessner, M., 2005. Bacteriophage P100 for control of *Listeria monocytogenes* in foods: genome sequence, bioinformatic analyses, oral toxicity study, and application. Regul. Toxicol. Pharmacol. 43, 301–312.

Carson, L., Gorman, S.P., Gilmore, B.F., 2010. The use of lytic bacteriophages in the prevention and eradication of biofilms of *Proteus mirabilis* and *Escherichia coli*. FEMS Immunol. Med. Microbiol. 59, 447–455.

Cerca, N., Oliveira, R., Azeredo, J., 2007. Susceptibility of *Staphylococcus epidermidis* planktonic cells and biofilms to the lytic action of staphylococcus bacteriophage K. Lett. Appl. Microbiol. 45, 313–317.

Chenu, J., Cox, J., 2009. Cronobacter ('Enterobacter sakazakii'): current status and future prospects. Lett. Appl. Microbiol. 49, 153–159.

Donlan, R.M., Costerton, J.W., 2002. Biofilms: survival mechanisms of clinically relevant microorganisms. Clin. Microbiol. Rev. 15, 167–193.

Doolittle, M., Cooney, J., Caldwell, D., 1995. Lytic infection of *Escherichia coli* biofilms by bacteriophage T4. Can. J. Microbiol. 41, 12–18.

Delcher, A.L., Harmon, D., Kasif, S., White, O., Salzberg, S.L., 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. 27, 4636–4641.

Denou, E., Bruttin, A., Barretto, C., Ngom-Bru, C., Brüssow, H., Zuber, S., 2009. T4 phages against *Escherichia coli* diarrhea: potential and problems. Virology 388, 21–30.

Edelson-Mammel, S.G., Buchanan, R.L., 2004. Thermal inactivation of *Enterobacter sakazakii* in rehydrated infant formula. J. Food Prot. 67, 60–63.

Ellis, D., Whitman, P., Marshall, R., 1973. Effects of homologous bacteriophage on growth of *Pseudomonas fragi* WY in milk. Appl. Microbiol. 25, 24–25.

Endersen, L., Coffey, A., Neve, H., McAuliffe, O., Ross, R.P., O'Mahony, J.M., 2013. Isolation and characterisation of six novel mycobacteriophages and investigation of their antimicrobial potential in milk. Int. Dairy J. 28, 8–14.

Endersen, L., O'Mahony, J., Hill, C., Ross, R.P., McAuliffe, O., Coffey, A., 2014. Phage therapy in the food industry. Annu. Rev. Food Sci. Technol. 5, 327–349.

Faruque, S.M., Mekalanos, J.J., 2012. Phage-bacterial interactions in the evolution of

toxigenic vibrio cholerae. Virulence 3, 556–565.

Food and Agriculture Organization/World Health Organization (FAO/WHO), 2008. Enterobacter sakazakii (Cronobacter spp.) in powdered follow-up formulae. MRA series. Available at. http://www.who.int/foodsafety/publications/micro/MRA_followup. pdf Date last accessed 24/04/17.

Forsythe, S.J., 2005. *Enterobacter sakazakii* and other bacteria in powdered infant milk formula. Matern. Child Nutr. 1, 44–50.

Friedemann, M., 2007. *Enterobacter sakazakii* in food and beverages (other than infant formula and milk powder). Int. J. Food Microbiol. 116, 1–10.

Friedemann, M., 2009. Epidemiology of invasive neonatal *Cronobacter* (*Enterobacter sakazakii*) infections. Eur. J. Clin. Microbiol. Infect. Dis. 28, 1297–1304.

Fsanz, F., 2012. Approval report–application A1045 in. Bacteriophage preparation P100 as processing aid.

Garcia, P., Madera, C., Martinez, B., Rodriguez, A., Suarez, J.E., 2009. Prevalence of bacteriophages infecting *Staphylococcus aureus* in dairy samples and their potential as biocontrol agents. J. Dairy Sci. 92, 3019–3026.

Gonzalez, S., Flick, G., Arritt, F., Holliman, D., Meadows, B., 2006. Effect of high-pressure processing on strains of *Enterobacter sakazakii*. J. Food Prot. 69, 935–937.

Goodridge, L.D., Bisha, B., 2011. Phage-based biocontrol strategies to reduce foodborne pathogens in foods. Bacteriophage 1, 130–137.

Gram, L., Bagge-Ravn, D., Ng, Y.Y., Gymoese, P., Vogel, B.F., 2007. Influence of food soiling matrix on cleaning and disinfection efficiency on surface attached *Listeria monocytogenes*. Food Control 18, 1165–1171.

Healy, B., Cooney, S., O'Brien, S., Iversen, C., Whyte, P., Nally, J., Callanan, J.J., Fanning, S., 2010. Cronobacter (Enterobacter sakazakii): an opportunistic foodborne pathogen. Foodborne Pathog. Dis. 7, 339–350.

Himelright, I., 2002. *Enterobacter sakazakii* infections associated with the use of powdered infant formula-Tennessee, 2001. J. Am. Med. Assoc. 287, 2204–2205.

Hodgson, K., 2013. Bacteriophage therapy. Microbiol. Aust. 34, 28–31.

Holý, O., Forsythe, S., 2014. *Cronobacter* spp. as emerging causes of healthcare-associated infection. J. Hosp. Infect. 86, 169–177.

Hunter, C., Bean, J., 2013. *Cronobacter*: an emerging opportunistic pathogen associated with neonatal meningitis, sepsis and necrotizing enterocolitis. J. Perinatol. 33, 581–585.

ICMSF, 2002. Microorganisms in Foods 7. Microbiological testing in food safety management. Kluwer Academic / Plenum Publishers, New York, USA.

Iversen, C., Forsythe, S., 2004. Isolation of *Enterobacter sakazakii* and other Enterobacteriaceae from powdered infant formula milk and related products. Food Microbiol. 21, 771–777.

Iversen, C., Lane, M., Forsythe, S., 2004. The growth profile, thermotolerance and biofilm formation of *Enterobacter sakazakii* grown in infant formula milk. Lett. Appl. Microbiol. 38, 378–382.

Iversen, C., Lehner, A., Mullane, N., Bidlas, E., Cleenwerck, I., Marugg, J., Fanning, S., Stephan, R., Joosten, H., 2007. The taxonomy of *Enterobacter sakazakii*: proposal of a new genus *Cronobacter* gen. nov. and descriptions of *Cronobacter sakazakii* comb. nov. *Cronobacter sakazakii* subsp. *sakazakii*, comb. nov., *Cronobacter sakazakii* subsp. *malonaticus* subsp. nov., *Cronobacter turicensis* sp. nov., *Cronobacter muytjensii* sp. nov., *Cronobacter dublinensis* sp. nov. and *Cronobacter* genomospecies 1. BMC Evol. Biol. 7, 64.

Jepson, C.D., March, J.B., 2004. Bacteriophage lambda is a highly stable DNA vaccine delivery vehicle. Vaccine 22, 2413–2419.

Joseph, S., Cetinkaya, E., Drahovska, H., Levican, A., Figueras, M.J., Forsythe, S.J., 2012. *Cronobacter condimenti* sp. nov., isolated from spiced meat, and *Cronobacter universalis* sp. nov., a species designation for *Cronobacter* sp. genomospecies 1, recovered from a leg infection, water and food ingredients. Int. J. Syst. Evol. Microbiol. 62, 1277–1283.

Kandhai, M.C., Reij, M.W., Gorris, L.G., Guillaume-Gentil, O., van Schothorst, M., 2004. Occurrence of *Enterobacter sakazakii* in food production environments and households. Lancet 363, 39–40.

Kasman, L.M., Kasman, A., Westwater, C., Dolan, J., Schmidt, M.G., Norris, J.S., 2002. Overcoming the phage replication threshold: a mathematical model with implications for phage therapy. J. Virol. 76, 5557–5564.

Kelly, D., McAuliffe, O., Ross, R., Coffey, A., 2012a. Prevention of *Staphylococcus aureus* biofilm formation and reduction in established biofilm density using a combination of phage K and modified derivatives. Lett. Appl. Microbiol. 54, 286–291.

Kelly, D., O'Sullivan, O., Mills, S., McAuliffe, O., Ross, R.P., Neve, H., Coffey, A., 2012b.

Genome sequence of the phage clP1, which infects the beer spoilage bacterium *Pediococcus damnosus*. Gene 504, 53–63.

Kerby, G., Gowdy, R., Dillon, E., Dillon, M., Csâky, T., Sharp, D., Beard, J., 1949. Purification, pH stability and sedimentation properties of the T7 bacteriophage of *Escherichia coli*. J. Immunol. 63, 93–107.

Keren, I., Kaldalu, N., Spoering, A., Wang, Y., Lewis, K., 2004. Persister cells and tolerance to antimicrobials. FEMS Microbiol. Lett. 230, 13–18.

Kim, K.-P., Klumpp, J., Loessner, M.J., 2007. *Enterobacter sakazakii* bacteriophages can prevent bacterial growth in reconstituted infant formula. Int. J. Food Microbiol. 115, 195–203.

Kłak, M., Międzybrodzki, R., Bubak, B., Jończyk, E., Weber-Dąbrowska, B., Górski, A., 2010. Studies on the gastrointestinal transit and blood penetration of a therapeutic staphylococcal bacteriophage. Abstract no. 209. In: First International Congrogress. Viruses of Microbes, Paris, .

Kropinski, A.M., Clokie, M.R., 2009. Bacteriophages: methods and protocols. In: Molecular and Applied Aspects. Volume 2 Humana Press.

Lai, K.K., 2001. *Enterobacter sakazakii* infections among neonates, infants, children, and adults: case reports and a review of the literature. Medicine 80, 113–122.

Laslett, D., Canback, B., 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 32, 11–16.

Lee, J., Oh, S., Kim, J., Yook, H., Byun, M., 2006. Gamma radiation sensitivity of *Enterobacter sakazakii* in dehydrated powdered infant formula. J. Food Prot. 69, 1434–1437.

Loessner, M.J., Neugirg, E., Zink, R., Scherer, S., 1993. Isolation, classification and molecular characterization of bacteriophages for *Enterobacter* species. J. Gen. Microbiol. 139, 2627–2633.

Nasser, A.M., Oman, S.D., 1999. Quantitative assessment of the inactivation of pathogenic and indicator viruses in natural water sources. Water Res. 33, 1748–1752.

Nazarowec-White, M., Farber, J., 1997a. *Enterobacter sakazakii*: a review. Int. J. Food Microbiol. 34, 103–113.

Nazarowec-White, M., Farber, J., 1997b. Thermal resistance of *Enterobacter sakazakii* in reconstituted dried-infant formula. Lett. Appl. Microbiol. 24, 9–13.

McNair, K., Bailey, B.A., Edwards, R.A., 2012. PHACTS, a computational approach to classifying the lifestyle of phages. Bioinformatics 28, 614–618.

O'Flaherty, S., Ross, R., Flynn, J., Meaney, W., Fitzgerald, G., Coffey, A., 2005. Isolation and characterization of two anti-staphylococcal bacteriophages specific for pathogenic *Staphylococcus aureus* associated with bovine infections. Lett. Appl. Microbiol. 41, 482–486.

Olson, M.R., Axler, R.P., Hicks, R.E., 2004. Effects of freezing and storage temperature on MS2 viability. J. Virol. Methods 122, 147–152.

Lavigne, R., Darius, P., Summer, E.J., Seto, D., Mahadevan, P., Nilsson, A.S., Ackermann, H.W., Kropinski, A.M., 2009. Classification of Myoviridae bacteriophages using protein sequence similarity. BMC Microbiol. 9, 224.

Sambrook, J., Russell, D.W., 2006. Purification of bacteriophage λ particles by isopycnic centrifugation through CsCl gradients. Cold Springs Harb. Protoc (Pdb-prot 3968).

Sullivan, M.J., Petty, N.K., Beatson, S.A., 2011. Easyfig: a genome comparison visualizer. Bioinformatics 27, 1009–1010.

Tey, B.T., Ooi, S.T., Yong, K.C., Ng, M.Y.T., Ling, T.C., Tan, W.S., 2009. Production of fusion m13 phage bearing the di-sulphide constrained peptide sequence (C-WSFFSNI-C) that interacts with hepatitis B core antigen. Afr. J. Biotechnol. 8, 268–273.

Turner, D., Reynolds, D., Seto, D., Mahadevan, P., 2013. CoreGenes3.5: A webserver for the determination of core genes from sets of viral and small bacterial genomes. BMC Res. 6, 140.

van Acker, J., de Smet, F., Muyldermans, G., Bougatef, A., Naessens, A., Lauwers, S., 2001. Outbreak of necrotizing enterocolitis associated with *Enterobacter sakazakii* in powdered milk formula. J. Clin. Microbiol. 39, 293–297.

Weld, R.J., Butts, C., Heinemann, J.A., 2004. Models of phage growth and their applicability to phage therapy. J. Theor. Biol. 227, 1–11.

Yang, H., Liang, L., Lin, S., Jia, S., 2010. Isolation and characterization of a virulent bacteriophage AB1 of *Acinetobacter baumannii*. BMC Microbiol. 10, 131.

Yates, M.V., Gerba, C.P., Kelley, L.M., 1985. Virus persistence in groundwater. Appl. Environ. Microbiol. 49, 778–781.

Zuber, S., Boissin-Delaporte, C., Michot, L., Iversen, C., Diep, B., Brüssow, H., Breeuwer, P., 2008. Decreasing *Enterobacter sakazakii* (*Cronobacter* spp.) food contamination level with bacteriophages: prospects and problems. Microb. Biotechnol. 1, 532–543.

CrossMark
← click for updates

REVIEW

Bacteriophage-based tools: recent advances and novel applications [version 1; referees: 3 approved]

Lisa O'Sullivan[1], Colin Buttimer[1], Olivia McAuliffe[2], Declan Bolton[3], Aidan Coffey[1]

[1]Department of Biological Sciences, Cork Institute of Technology, County Cork, Ireland
[2]Biotechnology Department, Teagasc, Moorepark Food Research Centre, Fermoy, County Cork, Ireland
[3]Division of Food Safety, Teagasc, Food Research Centre, Ashtown, County Dublin, Ireland

Abstract

Bacteriophages (phages) are viruses that infect bacterial hosts, and since their discovery over a century ago they have been primarily exploited to control bacterial populations and to serve as tools in molecular biology. In this commentary, we highlight recent diverse advances in the field of phage research, going beyond bacterial control using whole phage, to areas including biocontrol using phage-derived enzybiotics, diagnostics, drug discovery, novel drug delivery systems and bionanotechnology.

Open Peer Review

Referee Status: ✓ ✓ ✓

	Invited Referees		
	1	2	3
version 1 published 29 Nov 2016	✓	✓	✓

F1000 Faculty Reviews are commissioned from members of the prestigious F1000 Faculty. In order to make these reviews as comprehensive and accessible as possible, peer review takes place before publication; the referees are listed below, but their reports are not formally published.

1 **Rob Lavigne**, Katholieke Universiteit Leuven Belgium

2 **Sam R Nugen**, University of Massachusetts USA

3 **Laurent Debarbieux**, Institut Pasteur France

Discuss this article

Comments (0)

Corresponding author: Aidan Coffey (aidan.coffey@cit.ie)

Competing interests: The authors declare that they have no competing interests.

Introduction

Bacteriophages (phages) are viruses that specifically infect bacteria. After their discovery in 1915 by Twort and 1917 by d'Herelle, these agents were initially used to treat bacterial infections, although widespread acceptance was limited owing to lack of understanding of phage biology and the development of antibiotic therapy in the 1940s[1]. With antibiotic resistance becoming problematic in the late twentieth century[2], there was a renewed interest in phage therapy research. Alongside this application, and indeed the fundamental role that phage research played in the understanding of molecular biology, phage research has led to the development of new technologies not only for therapy and biocontrol but also for bacterial detection, drug delivery, drug discovery, and nanotechnology.

Antibacterials and biocontrol

In addition to the well-documented cases of using wild-type phages as tools to eliminate pathogenic bacteria in infected humans[3] and in foods[4], the phage-encoded peptidoglycan hydrolases called endolysins have also been exploited in purified form to rapidly lyse bacterial cells[5]. The Gram-positive phage endolysins generally contain at least one enzymatic domain and a cell-wall-binding domain. Chimeric endolysins have recently been developed by fusing enzymatic domains to alternative cell-wall-binding determinants, thus altering endolysin behaviour and host range[6]. In the case of Gram-negative bacteria, the outer membrane is a barrier to exogenously added endolysin reaching the peptidoglycan target. Thus, the fusion of polycationic peptides to the Gram-negative endolysin facilitates outer membrane penetration allowing these new so-called Artilysin®s access to the Gram-negative peptidoglycan[7]. Recent research has also reported a phage endolysin (from a *Streptococcus pyogenes* phage) with the ability to cross mammalian cell membranes. Its endolysin, PlyC, was found to consist of two subunits, one of which is proposed to bind to the eukaryotic cell membrane, facilitating entry by endocytosis[8]. These are major breakthroughs in endolysin research, and, with further investigation and testing, similar enzymes may be discovered/engineered and used in the future to, respectively, treat infections caused by Gram-negative bacteria and intracellular bacterial infections.

A recent advance in the area of antibiotic therapy has been the exploitation of phages to control antibiotic-resistant bacteria. Phages have been engineered to deliver CRISPR-Cas nucleases into antibiotic-resistant bacterial cells, and, in doing so, researchers have been able to harness the specific DNA-cleaving capacity of CRISPRs to knock out antibiotic resistance sequences, rendering resistant organisms antibiotic sensitive[9]. The use of phages as delivery vehicles ensures the specificity required in biocontrol. The wider exploitation of phages as delivery systems is discussed below.

Bacterial diagnostics

Phage virions and their encoded proteins can also be useful for the detection and specific identification of bacteria. The simplest of these is where a standard number of specific phages are incubated with a food material or some other test sample. If the bacterial target is present and viable, detectable phage numbers will increase through amplification on the pathogen. Modifications of this method can generate results more rapidly, and in the case of *Yersinia pestis*, Sergueev *et al.*, for example, developed a quantitative real-time PCR to detect the increase in phage DNA instead of traditional plaque assays[10]. Reporter phages can also detect bacteria through infection without needing cell lysis and progeny phages. In this case, the phage genomes are modified to carry a bioluminescence or fluorescence gene that the phage alone cannot express. Upon injection of the phage DNA into its host, active bioluminescent or fluorescent proteins are synthesized, facilitating visual detection. Recently, Zhang *et al.* engineered an *Escherichia coli* 0157:H7 reporter phage containing Luciferase NanoLuc (Nluc)[11] and with it detected as few as five CFU of the *E. coli* by bioluminescence in a complex food matrix within nine hours[12].

Reporter phage assays have also been adapted to assess drug sensitivity in the target bacterium. A *Mycobacterium tuberculosis* (TB) fluorophage, φ²GFP10, has been shown to detect TB in the complex matrix of a sputum sample, but also when rifampin or kanamycin are included in the assay, fluorescence was shown to be detectably diminished in sensitive cells in comparison with antibiotic-resistant variants[13]. Advantages of using whole phages for the detection of bacteria are that only viable bacterial cells are detected, bacterial host specificity is excellent, and phage cultivation is relatively inexpensive (however, lytic activity of a reporter phage should ideally be inactivated to ensure that the bacterial targets are not prematurely destroyed).

Phage receptor-binding proteins (RBPs) have also been used successfully in bacterial detection and identification. The receptor-binding domain of the RBP in *Campylobacter* phage NCTC12673 was used to create a simple glass slide agglutination test for *Campylobacter*, and when fused to green fluorescent protein, the receptor-binding domain allowed the detection of *Campylobacter jejuni* and *Campylobacter coli* isolates through fluorescent microscopy[14]. Phages, because of their vast diversity, provide a plentiful source of host-specific proteins to create simple identification tests such as the agglutination assay mentioned above specifically for *Campylobacter*. In this regard, whole phage and phage RBPs have been successfully attached to biosensing surfaces for bacterial detection, allowing for high specificity. Of the two, the RBPs are simpler and easier to attach. In addition, they can be recombinantly produced and are reported to have better stability than antibodies[15]. Optimization of phage densities and attachment to biosensing surfaces is still ongoing[16].

In the context of detection, the phage endolysins (discussed earlier) can also have a role when used instead of traditional DNA extraction reagents. It was shown that the peptidoglycan of *Staphylococcus aureus* is degraded more rapidly by staphylococcal endolysin ClyH than by lysostaphin, thus shortening the DNA sample preparation for real-time PCR when the endolysin was employed[17]. Phage display, which involves genetically modifying a phage virion so that a foreign peptide is displayed on the surface (discussed further below), can also be exploited in bacterial detection systems. Lee *et al.* created a phage that displayed two different peptides, one with an affinity to gold nanoparticles and another with specificity to a target protein. By measuring the

ultraviolet absorbance of this phage, they could detect as little as 25 femtomoles of their target antigen[18]. These modified phages have also been incorporated into systems capable of in-the-field real-time detection using engineered phage displaying peptides capable of binding to a magnetoelastic resonator as well as the target analytes, such as bacteria and endospores[19–21].

Drug discovery and phage-based drug delivery systems

Since phage display was first described in 1985 by Smith[22], it has seen numerous applications in the identification of receptor and ligand interactions of infectious diseases and cancers[23,24], with these developments allowing for drug discovery[25] and vaccine design[26]. Phage display is now allowing the modification of phages into vehicles (nanocarriers) for chemotherapeutic drug delivery by the attachment of a drug to the phage surface and presentation of peptides on the surface of that phage with specificity to a ligand of interest. Such constructs have even been designed to target non-host bacteria, including mammalian cells[27]. These phages, displaying therapeutic peptides, can even be designed to pass the blood–brain barrier[28], and such constructs could thus have potential in the treatment of diseases such as Alzheimer's and Parkinson's. Phages with an affinity to specific cell receptors, such as those overexpressed in cancer cells, may be exploited beyond drug delivery to allow for simultaneous target detection by displaying diagnostic reporter molecules or by detection of bound phage DNA by real-time PCR[29,30].

Empty phage capsids are also being employed as carriers, with studies demonstrating the potential to encapsulate RNA molecules, peptides, and therapeutic compounds[31–33]. Phage capsids or virus-like particles (VLPs) have also been modified to present ligands on their surface to allow the delivery of encapsulated RNA-guided endonucleases to specific cell types for *in situ* genome editing[34]. When phages are used as nanocarriers to deliver chemotherapeutic drugs for cancer treatment, drug half-life is extended and toxicity is focused only on the site of interest, lessening damage to body tissues[35]. Capsid-based carriers have also been developed by fusing drug-loaded liposomes to capsid proteins displaying peptides with binding specificity to a particular target[36].

Biotechnology

Genetically modified filamentous phages have been used in material synthesis to construct nanowires and films for semiconductor applications[37], piezoelectric energy generation[38], and photo-response properties[39]. These materials have been used to create devices such as ion batteries and catalysts[40,41], with phage M13-based nanowires also being constructed into scaffolding to allow guided cell growth for human tissue formation[42].

Phage-derived enzymes, which have formed part of a core toolbox in traditional molecular biology, are now being applied to novel concepts. Phage RNA polymerase and ribonuclease H are being used to create *in vitro* genetic circuits that have potential future applications in nanodevices and the regulation of processes within artificial cells[43]. Recombinases are seeing use in these constructions by extending memory capabilities to these circuits[44]. These enzymes are also being used to create novel tools for bacterial genome editing and accelerated evolution[45,46]. It is noteworthy that many past phage-dedicated reviews have not satisfactorily encompassed the recent advances of phage applications in nanomedicine; a recent excellent article comments in a comprehensive way on the many roles and opportunities of phages as nano-therapeutics, bioimaging probes, biomimetic biomaterials, tissue regenerative scaffolds, matrices for directing stem cell fate, and probes for detecting disease biomarkers, among numerous others[47].

Summary

This commentary provides a snapshot of the increasing diversity of phage research in recent years and shows that it is advancing rapidly and that new applications are being reported frequently. Since the discovery of phages a century ago, their research focus has diversified from applying these agents to simply treat bacterial infections to a broad range of useful functions including biocontrol, diagnostics, drug discovery, and drug delivery as well as several applications in nanomedicine.

Competing interests

The authors declare that they have no competing interests.

References
F1000 recommended

1. Wittebole X, De Roock S, Opal SM: **A historical overview of bacteriophage therapy as an alternative to antibiotics for the treatment of bacterial pathogens.** *Virulence.* 2014; **5**(1): 226–35.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

2. Davies J, Davies D: **Origins and evolution of antibiotic resistance.** *Microbiol Mol Biol Rev.* 2010; **74**(3): 417–33.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

3. Abedon ST, Kuhl SJ, Blasdel BG, *et al.*: **Phage treatment of human infections.** *Bacteriophage.* 2011; **1**(2): 66–85.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

4. Micreos: **LISTEX: Nature's solution for Listeria.** 2007.
Reference Source

5. Loeffler JM, Nelson D, Fischetti VA: **Rapid killing of *Streptococcus pneumoniae* with a bacteriophage cell wall hydrolase.** *Science.* 2001; **294**(5549): 2170–2172.
PubMed Abstract | **Publisher Full Text**

6. Dong Q, Wang J, Yang H, *et al.*: **Construction of a chimeric lysin Ply187N-V12C with extended lytic activity against staphylococci and streptococci.** *Microb Biotechnol.* 2015; **8**(2): 210–20.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

7. Briers Y, Walmagh M, van Puyenbroeck V, *et al.*: **Engineered endolysin-based**

"Artilysins" to combat multidrug-resistant gram-negative pathogens. *MBio.* 2014; **5**(4): e01379–14.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

8. Shen Y, Barros M, Vennemann T, *et al.*: **A bacteriophage endolysin that eliminates intracellular streptococci.** *eLife.* 2016; **5**: pii: e13152.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

9. [F] Bikard D, Euler CW, Jiang W, *et al.*: **Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials.** *Nat Biotechnol.* 2014; **32**(11): 1146–50.
PubMed Abstract | **Publisher Full Text** | **Free Full Text** | **F1000 Recommendation**

10. Sergueev KV, He Y, Borschel RH, *et al.*: **Rapid and sensitive detection of *Yersinia pestis* using amplification of plague diagnostic bacteriophages monitored by real-time PCR.** *PLoS One.* 2010; **5**: e11337.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

11. Hall MP, Unch J, Binkowski BF, *et al.*: **Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate.** *ACS Chem Biol.* 2012; **7**(11): 1848–57.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

12. Zhang D, Coronel-Aguilera CP, Romero PL, *et al.*: **The Use of a Novel NanoLuc -Based Reporter Phage for the Detection of Escherichia coli O157:H7.** *Sci Rep.* 2016; **6**: 33235.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

13. Jain P, Hartman TE, Eisenberg N, *et al.*: **φ²GFP10, a high-intensity fluorophore, enables detection and rapid drug susceptibility testing of *Mycobacterium tuberculosis* directly from sputum samples.** *J Clin Microbiol.* 2012; **50**(4): 1362–9.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

14. Javed MA, Poshtiban S, Arutyunov D, *et al.*: **Bacteriophage receptor binding protein based assays for the simultaneous detection of *Campylobacter jejuni* and *Campylobacter coli*.** *PLoS One.* 2013; **8**(7): e69770.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

15. Singh A, Poshtiban S, Evoy S: **Recent advances in bacteriophage based biosensors for food-borne pathogen detection.** *Sensors (Basel).* 2013; **13**(2): 1763–86.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

16. Olsson AL, Wargenau A, Tufenkji N: **Optimizing Bacteriophage Surface Densities for Bacterial Capture and Sensing in Quartz Crystal Microbalance with Dissipation Monitoring.** *ACS Appl Mater Interfaces.* 2016; **8**(22): 13698–706.
PubMed Abstract | **Publisher Full Text**

17. Hu Y, Yang H, Wang J, *et al.*: **Comparison between a chimeric lysin ClyH and other enzymes for extracting DNA to detect methicillin resistant *Staphylococcus aureus* by quantitative PCR.** *World J Microbiol Biotechnol.* 2016; **32**(1): 1.
PubMed Abstract | **Publisher Full Text**

18. Lee JH, Domaille DW, Cha JN: **Amplified protein detection and identification through DNA-conjugated M13 bacteriophage.** *ACS Nano.* 2012; **6**(6): 5621–6.
PubMed Abstract | **Publisher Full Text**

19. Li S, Li Y, Chen H, *et al.*: **Direct detection of *Salmonella typhimurium* on fresh produce using phage-based magnetoelastic biosensors.** *Biosens Bioelectron.* 2010; **26**(4): 1313–9.
PubMed Abstract | **Publisher Full Text**

20. Park M, Park JW, Wikle HC, *et al.*: **Evaluation of phage-based magnetoelastic biosensors for direct detection of *Salmonella* Typhimurium on spinach leaves.** *Sensors and Actuators B: Chemical.* 2013; **176**: 1134–40.
Publisher Full Text

21. Shen W, Lakshmanan RS, Mathison LC, *et al.*: **Phage coated magnetoelastic micro-biosensors for real-time detection of *Bacillus anthracis* spores.** *Sensors and Actuators B: Chemical.* 2009; **137**(2): 501–6.
Publisher Full Text

22. Smith GP: **Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface.** *Science.* 1985; **228**(4705): 1315–7.
PubMed Abstract | **Publisher Full Text**

23. Mullen LM, Nair SP, Ward JM, *et al.*: **Phage display in the study of infectious diseases.** *Trends Microbiol.* 2006; **14**(3): 141–7.
PubMed Abstract | **Publisher Full Text**

24. Yip YL, Ward RL: **Application of phage display technology to cancer research.** *Curr Pharm Biotechnol.* 2002; **3**(1): 29–43.
PubMed Abstract | **Publisher Full Text**

25. Omidfar K, Daneshpour M: **Advances in phage display technology for drug discovery.** *Expert Opin Drug Discov.* 2015; **10**(6): 651–69.
PubMed Abstract | **Publisher Full Text**

26. Gao J, Wang Y, Liu Z, *et al.*: **Phage display and its application in vaccine design.** *Ann Microbiol.* 2010; **60**(1): 13–9.
Publisher Full Text

27. Vaks L, Benhar I: ***In vivo* characteristics of targeted drug-carrying filamentous bacteriophage nanomedicines.** *J Nanobiotechnology.* 2011; **9**: 58.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

28. Ksendzovsky A, Walbridge S, Saunders RC, *et al.*: **Convection-enhanced delivery of M13 bacteriophage to the brain.** *J Neurosurg.* 2012; **117**(2): 197–203.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

29. Hosoya H, Dobroff AS, Driessen WH, *et al.*: **Integrated nanotechnology platform for tumor-targeted multimodal imaging and therapeutic cargo release.** *Proc Natl Acad Sci U S A.* 2016; **113**(7): 1877–82.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

30. Brasino M, Cha JN: **Real-time femtomolar detection of cancer biomarkers from photoconjugated antibody-phage constructs.** *Analyst.* 2016.
PubMed Abstract | **Publisher Full Text**

31. Wei B, Wei Y, Zhang K, *et al.*: **Development of an antisense RNA delivery system using conjugates of the MS2 bacteriophage capsids and HIV-1 TAT cell-penetrating peptide.** *Biomed Pharmacother.* 2009; **63**(4): 313–8.
PubMed Abstract | **Publisher Full Text**

32. DePorter SM, McNaughton BR: **Engineered M13 bacteriophage nanocarriers for intracellular delivery of exogenous proteins to human prostate cancer cells.** *Bioconjug Chem.* 2014; **25**(9): 1620–5.
PubMed Abstract | **Publisher Full Text**

33. Stephanopoulos N, Tong GJ, Hsiao SC, *et al.*: **Dual-surface modified virus capsids for targeted delivery of photodynamic agents to cancer cells.** *ACS Nano.* 2010; **4**(10): 6014–20.
PubMed Abstract | **Publisher Full Text**

34. Qazi S, Miettinen HM, Wilkinson RA, *et al.*: **Programmed Self-Assembly of an Active P22-Cas9 Nanocarrier System.** *Mol Pharm.* 2016; **13**(3): 1191–6.
PubMed Abstract | **Publisher Full Text**

35. Lu RM, Chen MS, Chang DK, *et al.*: **Targeted drug delivery systems mediated by a novel Peptide in breast cancer therapy and imaging.** *PLoS One.* 2013; **8**(6): e66128.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

36. Wang T, D'Souza GG, Bedi D, *et al.*: **Enhanced binding and killing of target tumor cells by drug-loaded liposomes modified with tumor-specific phage fusion coat protein.** *Nanomedicine (Lond).* 2010; **5**(4): 563–74.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

37. Mao C, Flynn CE, Hayhurst A, *et al.*: **Viral assembly of oriented quantum dot nanowires.** *Proc Natl Acad Sci U S A.* 2003; **100**(12): 6946–51.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

38. Lee BY, Zhang J, Zueger C, *et al.*: **Virus-based piezoelectric energy generation.** *Nat Nanotechnol.* 2012; **7**(6): 351–6.
PubMed Abstract | **Publisher Full Text**

39. Murugesan M, Abbineni G, Nimmo SL, *et al.*: **Virus-based photo-responsive nanowires formed by linking site-directed mutagenesis and chemical reaction.** *Sci Rep.* 2013; **3**: 1820.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

40. Lee Y, Kim J, Yun DS, *et al.*: **Virus-templated Au and Au/Pt Core/shell Nanowires and Their Electrocatalytic Activities for Fuel Cell Applications.** *Energy Environ Sci.* 2012; **5**(8): 8328–34.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

41. Nam KT, Kim D, Yoo PJ, *et al.*: **Virus-enabled synthesis and assembly of nanowires for lithium ion battery electrodes.** *Science.* 2006; **312**(5775): 885–8.
PubMed Abstract | **Publisher Full Text**

42. Yoo SY, Chung W, Kim TH, *et al.*: **Facile patterning of genetically engineered M13 bacteriophage for directional growth of human fibroblast cells.** *Soft Matter.* 2011; **7**(2): 363–8.
Publisher Full Text

43. Kim J, Winfree E: **Synthetic *in vitro* transcriptional oscillators.** *Mol Syst Biol.* 2011; **7**: 465.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

44. Siuti P, Yazbek J, Lu TK: **Synthetic circuits integrating logic and memory in living cells.** *Nat Biotechnol.* 2013; **31**(5): 448–52.
PubMed Abstract | **Publisher Full Text**

45. Enyeart PJ, Chirieleison SM, Dao MN, *et al.*: **Generalized bacterial genome editing using mobile group II introns and Cre-*lox*.** *Mol Syst Biol.* 2013; **9**: 685.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

46. [F] Esvelt KM, Carlson JC, Liu DR: **A system for the continuous directed evolution of biomolecules.** *Nature.* 2011; **472**(7344): 499–503.
PubMed Abstract | **Publisher Full Text** | **Free Full Text** | **F1000 Recommendation**

47. Sunderland K, Yang M, Mao C: **Phage-Enabled Nanomedicine: From Probes to Therapeutics in Precision Medicine.** *Angew Chem Int Ed Engl.* 2016.
PubMed Abstract | **Publisher Full Text**

Open Peer Review

Current Referee Status: ✔ ✔ ✔

Editorial Note on the Review Process

F1000 Faculty Reviews are commissioned from members of the prestigious F1000 Faculty and are edited as a service to readers. In order to make these reviews as comprehensive and accessible as possible, the referees provide input before publication and only the final, revised version is published. The referees who approved the final version are listed with their names and affiliations but without their reports on earlier versions (any comments will already have been addressed in the published version).

The referees who approved this article are:

Version 1

1 **Laurent Debarbieux**, Department of Microbiology, Interactions Bacteriophages Bacteria in Animals, Institut Pasteur, Paris, France
   Competing Interests: No competing interests were disclosed.

2 **Sam R Nugen**, Department of Food Science, University of Massachusetts, Amherst, MA, USA
   Competing Interests: No competing interests were disclosed.

3 **Rob Lavigne**, Laboratory of Gene Technology, Katholieke Universiteit Leuven, Leuven, Belgium
   Competing Interests: No competing interests were disclosed.

CrossMark
click for updates

Genome Sequence of Jumbo Phage vB_AbaM_ME3 of *Acinetobacter baumanni*

Colin Buttimer,ᵃ Lisa O'Sullivan,ᵃ Mohamed Elbreki,ᵃ Horst Neve,ᵇ Olivia McAuliffe,ᶜ R. Paul Ross,ᶜ Colin Hill,ᵈ Jim O'Mahony,ᵃ Aidan Coffeyᵃ

Department of Biological Sciences, Cork Institute of Technology, Co. Cork, Irelandᵃ; Department of Microbiology and Biotechnology, Max Rubner-Institut, Kiel, Germanyᵇ; Biotechnology Department, Teagasc, Moorepark Food Research Centre, Fermoy, Co. Cork, Irelandᶜ; Department of Microbiology, University College Cork, Co. Cork, Irelandᵈ

Bacteriophage (phage) vB_AbaM_ME3 was previously isolated from wastewater effluent using the propagating host *Acinetobacter baumannii* DSM 30007. The full genome was sequenced, revealing it to be the largest *Acinetobacter* bacteriophage sequenced to date with a size of 234,900 bp and containing 326 open reading frames (ORFs).

*A*cinetobacter baumannii has emerged in recent times as an important nosocomial pathogen. Health care-acquired *A. baumannii* infections include pneumonia and urinary tract and bloodstream infections (1). There is only a small number of bacteriophages (phages) with genomes greater than 200 kbp (termed "jumbo" phages) that have had their genomes sequenced to date. Most of their encoded proteins do not have any homologues in current sequence databases, and the diversity of these phages has been great enough that it has limited comparative genomics studies (2).

A phage with the ability to lyse *A. baumannii* strain DSM 30007 was isolated from effluent obtained from a wastewater treatment plant in Cork, Ireland. Transmission electron microscopy revealed that the phage belonged to the *Myoviridae* family, and according to nomenclature proposed by Kropinski et al. was named vB_AbaM_ME3 (ME3) (3). A high titer phage suspension was concentrated by ultracentrifugation, and DNA extraction was performed as previously described (4). DNA was sequenced using the 454 FLX Titanium PLUS Sequencing approach (LGC Genomics, Mannheim, Germany). Open reading frames (ORFs) were identified using GLIMMER and GenemarkS (5, 6), with possible function of these ORFs' proteins being predicted with BLASTp, pFam, InterProScan, THMHMM v.2.0, LipoP v1.0 (7–11), with tRNAscan.SE 1.21 being used to locate any tRNA present in the genome (12).

To date, this is the largest *Acinetobacter* phage genome sequenced, with a size of 234,900 bp (the genome ends of ME3 are not known). The overall %G+C is 40%, similar to that of its host (13). The genome was predicted to have 326 ORFs with four tRNA genes.

On the basis of homology, putative functions were assigned to 77 ORFs, with 19 ORFs annotated as putative membrane proteins, two ORFs annotated as putative lipoproteins, and the remaining 228 ORFs being annotated as hypothetical proteins.

Phage ME3 is an orphan phage, however, it has eight ORFs encoding structural proteins that share homology to those of the novel *Bacillus* phage 0305phi8-36 (GenBank accession number NC_009760.1), although showing high levels of divergence (percentage identity of 26% to 34%). The major head protein (ME3_22), portal protein (ME3_19), and tail sheath subunit (ME3_29) are examples of such proteins. Until now, these proteins of 0305phi8-36 have only been found to share homology with those of phage-like elements found in the genomes of *B. thuringiensis* serovar *israelensis* and *B. weihenstephanensis* (14). With regard to these structural proteins and the large terminase subunit (ME3_13), phages ME3 and 0305phi8-36 may share a distant ancestor.

Phage ME3 appears to encode its own DNA replication machinery including DNA polymerase subunits (ME3_60 and 61), thymidylate synthase enzymes (ME3_107 and 108), helicases, and enzymes involved in DNA degradation and repair. ME3 also possesses two cell wall degrading enzymes, ME3_8, a lysozyme with proven lytic activity against *A. baumannii* and ME3_113, a putative cell wall hydrolase. Phage ME3 also has a curiously large number of genes associated with Ter-stress response (ME3_286, 284, 289, 290, and 291) and a massive protein of 5,419 amino acids (ME3_104) possessing domains relating to host specificity and binding (IPR015406, IPR008979).

Accession number(s). The full genome sequence of *A. baumannii* phage vB_AbaM_ME3 was deposited in GenBank under the accession number KU935715.

REFERENCES

1. **Fournier PE, Richet H, Weinstein RA.** 2006. The epidemiology and control of *Acinetobacter baumannii* in health care facilities. Clin Infect Dis 42:692–699. http://dx.doi.org/10.1086/500202.

2. **Hendrix RW**. 2009. Jumbo bacteriophages. Curr Top Microbiol Immunol **328**:229–240. http://dx.doi.org/10.1007/978-3-540-68618-7_7.

3. **Kropinski AM, Prangishvili D, Lavigne R**. 2009. Position paper: the creation of a rational scheme for the nomenclature of viruses of bacteria and archaea. Environ Microbiol **11**:2775–2777. http://dx.doi.org/10.1111/j.1462-2920.2009.01970.x.

4. **Keary R, McAuliffe O, Ross RP, Hill C, O'Mahony J, Coffey A**. 2014. Genome analysis of the staphylococcal temperate phage DW2 and functional studies on the endolysin and tail hydrolase. Bacteriophage **4**:e28451. http://dx.doi.org/10.4161/bact.28451.

5. **Delcher AL, Bratke KA, Powers EC, Salzberg SL**. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics **23**:673–679. http://dx.doi.org/10.1093/bioinformatics/btm009.

6. **Besemer J, Lomsadze A, Borodovsky M**. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res **29**:2607–2618. http://dx.doi.org/10.1093/nar/29.12.2607.

7. **Gish W, States DJ**. 1993. Identification of protein coding regions by database similarity search. Nat Genet **3**:266–272. http://dx.doi.org/10.1038/ng0393-266.

8. **Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A**. 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res **44**:D279–D285. http://dx.doi.org/10.1093/nar/gkv1344.

9. **Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R**. 2005. InterProScan: protein domains identifier. Nucleic Acids Res **33**:—W116–W120. http://dx.doi.org/10.1093/nar/gki442.

10. **Krogh A, Larsson B, von Heijne G, Sonnhammer EL**. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol **305**:567–580. http://dx.doi.org/10.1006/jmbi.2000.4315.

11. **Juncker AS, Willenbrock H, von Heijne G, Nielsen H, Brunak S, Krogh A**. 2003. Prediction of lipoprotein signal peptides in gram-negative bacteria. Protein Sci **12**:1652–1662. http://dx.doi.org/10.1110/ps.0303703.

12. **Lowe TM, Eddy SR**. 1997. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res **25**:955–964. http://dx.doi.org/10.1093/nar/25.5.0955.

13. **Davenport KW, Daligault HE, Minogue TD, Bruce DC, Chain PS, Coyne SR, Jaissle JG, Koroleva GI, Ladner JT, Li PE, Palacios GF, Scholz MB, Teshima H, Johnson SL**. 2014. Draft genome assembly of *Acinetobacter baumannii* ATCC 19606. Genome Announc **2**(4):e00832-14. http://dx.doi.org/10.1128/genomeA.00832-14.

14. **Thomas JA, Hardies SC, Rolando M, Hayes SJ, Lieman K, Carroll CA, Weintraub ST, Serwer P**. 2007. Complete genomic sequence and mass spectrometeric analysis of highly diverse, atypical *Bacillus thuringiensis* phage 0305phi8-36. Virology **127**:358–366.