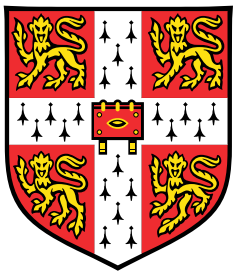
Environmental bacteriophages infecting *Dickeya* and *Serratia* species: receptors and diversity



Andrew Day

Department of Biochemistry
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Homerton College

September 2018

I would like to dedicate this thesis to all of the people in my life who have professed an interest in my work and yet remain convinced that I work on potatoes.

"It is a mistake to think you can solve any major problems just with potatoes"

— Douglas Adams

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. This dissertation contains fewer than 60,000 words, exclusive of tables, footnotes, bibliography, and appendices.

Andrew Day
September 2018

Acknowledgements

My thanks, first and foremost, go to my supervisor, Prof. George Salmond, for his support and guidance throughout my time in Cambridge. His passion for microbiology has inspired me since we first met and I value highly the wealth of knowledge he has brought to this dissertation.

I would like to thank the Biotechnology and Biological Sciences Research Council Doctoral Training Partnership for providing financial support for the four years I have spent in Cambridge and the administrators Suzy, Sarah and Jennifer for organising the training programme.

Special thanks go to Dr Rita Monson and Alison Rawlinson for welcoming me into the lab and providing me with guidance through my project. I will always appreciate their help and advice for all of my problems, lab-based and otherwise.

Numerous people have helped during this project, both scientifically and socially. Thanks to Chin Mei (so wild) for our date nights and making me big in Malaysia. Thanks to my starting buddies, Alex, Bihe and Ray, for getting me through the years, it's been great watching us all grow in confidence. Special mentions goes to Alice and Leo, who did all the hard graft of collecting the phages I have studied. Thanks to all those that had a hand in supporting this project: Jessica, Karin, Lyn, MicrobesNG and the rest of the Salmond lab.

I couldn't have done this without my support network to get me through the rough patches. James, Kathryn and Konnie, despite my lack of contact at times, you've always been there for me and I will be forever grateful. Lewis, thank you for the distractions, the motivation and for always putting a smile on my face. James, I know you are jealous of me becoming a real doctor, but I couldn't have asked for a better supporter. You came into my life when I was at my lowest and helped me pick up the pieces. You believed in me even when I didn't, and my life has been so much richer with you in it.

Last but not least, I am thankful for my parents, who have always supported and encouraged me to follow my dreams and without whom I would not be where I am today.

Abstract

Phytopathogenic *Dickeya* species inflict large economic losses on a variety of crops. A lack of effective chemical control methods has generated interest in the use of bacteriophages (phages) as a novel tool for biocontrol. In the last decade, six phages have been isolated in Belgium and Poland using *Dickeya solani* as the host. Previous work in this laboratory has isolated ninety phages capable of infecting *D. solani*. The majority have been morphologically classified as members of the *Ackermannviridae* family.

In agreement with findings in *Salmonella* and *Klebsiella* species, the capsule of *D. solani* is a likely receptor of *Ackermannviridae* family phages. Analysis of *D. solani* strains carrying reporter fusions suggested that the capsule genes are expressed in response to nutritional stress, however disruption of the capsular polysaccharide cluster did not significantly impact virulence.

Experiments assessing capsular polysaccharide as a putative receptor for *Ackermannviridae* family phages in nosocomial pathogen *Serratia* produced inconclusive results. Phage-resistance due to random transposon mutagenesis identified genes encoding transcription factors and regulators, but none directly linked to capsular polysaccharide production.

Thirteen phages were capable of infecting a wider host range of *Dickeya* species. Morphological and genomic analysis showed that six were *Podoviridae* family members, whilst the other seven were *Myoviridae* family members. These are part of the recently defined 'hairy *Myoviridae*', characterised by a distinct morphology. Another member of this grouping was isolated during this study, but is more closely related to phages of *Erwinia amylovora*.

A subset of the *Ackermannviridae* family phages were shown to be capable of facilitating transduction. This makes them unsuitable for use in the environment due to the risk of deleterious horizontal gene transfer. This is also true for the *Myoviridae* family members, but not for one of the *Podoviridae* family members. This phage could therefore be a promising candidate for therapeutic use.

Table of contents

List of figures

List of tables

Chapter One

Introduction

1.1 *Dickeya* species - an economically damaging threat

1.1.1 The genus *Dickeya*

Global food production must increase by 50% in order to meet the projected demand of the world's population by 2050 [36]. One of the major challenges in agriculture therefore is the reinforcement of global food security to allow maintenance and growth of the world population. Loss of crops to pathogen infection is a limiting factor in food production, and it has been estimated that over a quarter of the global harvest is lost to pests annually [123].

The bacterial genus *Dickeya*, recently reclassified into the novel family Pectobacteriaceae [4], currently consists of up to eleven phytopathogenic species that can cause severe disease in economically important crops including tomato, orchid and potato [12]. Colloquially grouped with *Pectobacterium* species under the name 'soft rot Enterobacteriaceae', all of these species used to be classified as *Erwinia* species and have only been separated taxonomically in the past 20 years [138]. *Dickeya* species have been reported to have a host range across over 35% of angiosperm plant orders [39]. Able to grow at warmer ambient temperatures than *Pectobacterium* species, and cause disease at much lower inocula, *Dickeya* species are an increasing environmental pest [167]. There have been several outbreaks over the past two decades that have highlighted the large economic impact this pathogen can have across both Europe and North America [39], although *Dickeya* species have recently been identified in Pakistan [150] and Australia [179], showing that these pathogens are an increasingly global threat. In potato, one of the top five agricultural products as identified by the United Nations in 2015 [66], *Dickeya* infections are reported to cause losses of up to 30 million euros each year in the Netherlands alone [167] and the loss of up to 30% of crops in Poland [118]. Table 1.1 summarises the hosts and geographical range of *Dickeya* species.

***Dickeya* species**	**Hosts**	**Geographic range**
D. aquatica	No reported hosts	UK and Finland [127]
D. chrysanthemi	Artichoke, aubergine, chicory, potato, sunflower, tomato, *Chrysanthemum* sp., *Parthenium* sp., *Philodendron*, *Vanda* sp.	Europe, Japan [39, 158, 167]
D. dadantii	Aubergine, banana, carrot, peach, strawberry, sweet potato, tomato, *Amorphophallus konjac*, *Anubias barteri*, *Brassica rapa*, *Dieffenbachia* sp., *Musa* sp., *Phalaenopsis* sp., *Philodendron* sp., *Tagetes patula*, *Vanilla planifolia*	Brazil, China, Europe, Japan, Malaysia, Peru, Zimbabwe [39, 158, 167]
D. dianthicola	Artichoke, carnation, chicory, potato, tomato, yacon, *Cichorium intybus*, *Chrysanthemum* sp., *Dahlia* sp., *Dianthus* sp., *Kalanchoe* sp.	Australia, Bangladesh, Europe, Colombia, Japan, New Zealand, Pakistan, Taiwan and the USA [39, 150, 158, 167, 179]
D. fangzhongdai	*Pyrus* sp.	China [39, 166]
D. paradisiaca	Banana, maize, potato, *Musa* sp.	Europe [39, 167]
D. solani	Potato, *Hyacinthus orientalis*	China, Europe, Israel [39, 167]
D. zeae	Banana, maize, potato, pineapple, rice, tobacco, *Brachiaria* sp., *Chrysanthemum* sp., *Calanthe* sp., *Musa* sp., *Setaria* sp.	Australia, China, Europe, Japan, Mexico and Papua New Guinea [39, 158, 167]

Table 1.1 Hosts and geographic range of the genus *Dickeya*. All species with a European host range have been isolated in at least three European countries, but these have not been listed independently to minimise text.

Until 2004 almost all European potato isolates of *Dickeya* were assigned as *Dickeya dianthicola* [167]. A new clade of *Dickeya* in European potato isolates was subsequently identified [94, 126, 154] and in 2014 a new species was proposed; *Dickeya solani* [172]. In many of the countries in which this species has been isolated, introduction of the pathogen was traced to the international trade of seed potatoes and has also been found in hyacinth, leading to suggestions that these bacteria have recently adapted to infect potatoes, with possible transfer via contaminated irrigation water [128]. *D. solani* has also been isolated from the roots of healthy weeds [169], leaf surfaces [141] and insect vectors [146], giving the pathogen multiple reservoirs and methods of dissemination. *D. solani* is able to spread more easily through the plant vascular system and survive at higher temperatures than *D. dianthicola* [167]. Whilst there have been eight isolated cases of *D. solani* reported in England and Wales since 2007, five were found in crops originating from outside of the UK and three in the irrigation sources of these crops [33].

The classical soft rot symptoms of infection by *Dickeya* species are primarily caused by secreted pectinolytic enzymes that degrade pectin in the plant cell wall, macerating plant tissue and causing a wet, foul-smelling rot primarily in storage organs such as tubers and bulbs [79]. When conditions are favourable, infected tubers can rot within three days, with plants dying a few hours after initial wilting symptoms emerge [39]. Most studies of *Dickeya* virulence have been performed in the laboratory strain *D. dadantii* 3937, chosen by the community as the model organism due to its amenability to genetic manipulation [141]. Table 1.2 summarises the main virulence determinants of *D. dadantii* during the extracellular stage, which are assumed to broadly applicable to other *Dickeya* species. A comprehensive description of the virulence factors, including during intracellular growth and their regulation, has been published by Reverchon *et al.* [141].

Initial attachment of the *Dickeya* cells to plant surfaces is mediated by cellulose fibrils [81] and haemagglutinins [144]. Aggregates of the bacteria are protected from dessication on plant surfaces by an extracellular polysaccharide layer [44]. Motility provided by biosurfactant secretion [75] and flagella, coupled with a strong chemotactic response to jasmonic acid, which is produced by wounded plant tissue [18] allow systemic invasion of the host. The virulence of *Dickeya* is largely due to the ability to secrete plant cell wall degrading enzymes such as the isoenzyme forms of various pectinases, but also xylanases, galactanases, cellulases and proteases. These enzymes are largely actively secreted via Type 1 and 2 Secretion Systems [141]. Unlike many other Gram-negative pathogens, the Type 3 Secretion System of *Dickeya* may be less important, with mutants showing a reduced virulence but maintaining the ability to cause rotting disease [183]. Siderophores are important for pathogenesis of *Dickeya* species, to allow growth in iron-limited conditions [63].

Virulence determinant	Function
Pectinases	
PelA-E, I, L, N, W, X and Z	Pectate lyases
PehK, N, V, W and X	Polygalacturonases
RhiE	Rhamnogalacturonate lyase
PemA and B	Pectin methylesterases
PaeX and Y	Pectin acetylesterases
FaeD and T	Feruloyl esterases
Other cell wall degrading enzymes	
XynA	Glucuronoxylanase
GanA	Endogalactanase
CelZ	Cellulase
PrtA,B, C and G	Proteases
Type 3 Secretion System Effectors	
DspE	AvrE superfamily effector
HrpN and W	Aggregation factors
Siderophores and iron metabolism	
Achromobactin and chrysobactin	Siderophores
FntA, Bfr, and Dps	Iron storage proteins
Fur repressor	Iron-sensitive repressor of pectinase genes
Colonisation factors	
NipE	Extracellular necrosis inducing protein
AvrL and M	Avirulence factors
Haemagglutinins	Adhesion
Cellulose fibrils	Adhesion and aggregation
Extracellular polysaccharide	Prevention of dessication
Biosurfactant	Adhesion
Flagella	Adhesion and chemotaxis

Table 1.2 The main virulence determinants of *Dickeya* species involved in adhesion and invasion of the host. These factors have been largely discovered and studied in *D. dadantii* 3937.

Detection and monitoring of *Dickeya* in the environment is hindered by the lack of swift and robust detection assays. In many countries, diagnosis of diseased potatoes is by visual inspection alone, therefore potentially misclassifying infections that have similar symptoms, such as soft rots induced by *Pectobacterium* and *Dickeya* species [39]. In laboratory-based assays, crystal violet pectate (CVP) medium is semi-selective and is the standard technique for identifying soft-rot pathogens [46], however, there are reports that *Dickeya* species can lose the ability to cause the characteristic pits on CVP medium that identify soft rot pathogens [39]. Attempts to move to using PCR-based methods for detection of *Dickeya* have proved promising, but have also suffered a high rate of false positives when applied to environmental samples [171] and in one case deletion of the binding site for a well established primer set [137] caused the pathogen not to be detected in a recent outbreak in the United States [39]. Over thirty different methods have been proposed for identification of pectinolytic bacteria [52]. New methods have been developed recently to accelerate identification and improve efficacy, including capillary electrophoretic techniques, Matrix-Assisted Laser Desorption/Ionization Time-Of-Flight (MALDI-TOF) mass spectrometry [148] and Loop-Mediated Isothermal Amplification [186], which has been shown to reduce detection times to under thirty minutes. These new methods all require rigorous testing for robustness using environmental samples before widespread application.

Due to the large economic damage inflicted by *Dickeya* species, regulations have been put in place in an attempt to limit the spread of these bacteria. The Seed Potatoes (Scotland) Amendment Regulations 2010 has established a zero tolerance policy for all *Dickeya* species in Scotland [138]. The Jamaican Ministry of Industry, Commerce, Agriculture and Fisheries has listed *Dickeya* species as a quarantine pest and prohibited import of seed from affected countries [130]. The European and Mediterranean Plant Protection Organization (EPPO), an intergovernmental body encompassing 52 countries across Europe, North Africa, the Middle East and Central Asia, has designated *D. dianthicola* as a quarantine pest already present in the region, and encourages regulation [138].

Bacterial plant pathogens are a global agricultural problem and lack good control tools other than antibiotics, instead relying on sanitation and exclusion methods [141]. Antibiotic usage is restricted as prevalence in the environment is likely to drive antibiotic resistance among many organisms, which could then transfer into animal pathogens. Bacteria also exhibit high levels of adaptability and are able to colonise new hosts and invade new geographic areas relatively quickly [61]. This is evident in *Dickeya* species, which were initially considered to be restricted to tropical and sub-tropical regions before the emergence of cold-tolerant *D. dianthicola* in the Netherlands three decades ago [82] and subsequent spread of *D. solani* [172]. It is largely accepted that once *Dickeya* species have infected a plant there

is no efficient method of infection control [47]. There is no current chemical or physical treatment available that can readily clear infection, and, despite research into genetically modified potatoes that are resistant to infection [115], no commercially available cultivars currently exist. Plant growth promoting rhizobacteria have also been tested in the hope that they could out-compete or suppress pathogen growth [47]. Despite these many avenues of research, so far only preventative measures, such as separation and screening, have proved effective.

Whilst *Dickeya* species have been largely studied due to their agricultural importance, experiments with *Erwinia chrysanthemi*, a previous name in the complex taxonomy for this group, have yielded multiple fundamental biological discoveries, including protein structures of pectate lyases [187], insights on secretion through Type 2 Secretion Systems [100, 101] and the first demonstration of the role of the Type 3 Secretion Systems in biofilm formation [185].

1.2 *Serratia* species

The genus *Serratia*, until recently a member of the Enterobacteriaceae before taxonomic reclassification into the novel family Yersiniaceae [4], is found in both terrestrial and aquatic environments associated with animals and plants [80]. Composed of fourteen recognised species, it was first described in 1819 in Italy when it was determined to be the cause of polenta turning red [113]. This was due to the production of the pigment prodigiosin, which is produced by many *Serratia* strains, and is thought to be behind many reports of 'bleeding' bread throughout predominantly Christian literature, and as far back as the siege of Tyre in 332 B.C. [71]. The pigmentation conferred by prodigiosin led to the use of *Serratia marcescens* as a tracer organism in experiments that tested the dispersal of microbes. This began in 1906 in the UK Houses of Parliament [15] and was followed by a series of tests in United States Military facilities [104] and led to release of pigmented *Serratia* in locations across the United States between 1950 and 1968, including off the coast of San Francisco and in the underground railway systems of New York [170]. Similar experiments were also carried out in Paris [78] and the United Kingdom [21], all aiming to study the dispersal of potential bioweapons.

Previously thought to be non-pathogenic, hence their use in biological release experiments, it has since been found that *Serratia* species can cause infections in immuno-compromised individuals, and they are an increasing healthcare challenge due to intrinsic and acquired antibiotic resistance [149]. In the most recent data available from the European Centre for Disease Prevention and Control from 2014, *Serratia* species represented 2.5% of bloodstream

infections and 5.3% of pneumonia cases acquired in intensive care units [59]. The majority of isolates are intrinsically resistant to penicillins and tetracyclines and some have been found to have acquired resistance to aminoglycosides, carbapenems and quinolones, which are the front-line antibiotics used to treat *Serratia* infections [104].

Despite the opportunistic nature of *Serratia* species, much of the academic research concerns secondary metabolites such as the pigment prodigiosin, a member of the prodiginine family of molecules with anti-cancer and antibacterial properties [87]. Some strains of *Serratia*, in particular *S. plymuthica*, are also important in the root rhizosphere and have been used to control soil-borne fungal pathogens [57].

1.3 Bacteriophages

1.3.1 Viruses of bacteria

Bacteriophages (phages) are obligate parasites of bacteria that are predicted to outnumber their hosts at least ten-fold [41]. Phages are found in every environment inhabited by bacteria, including soil, the oceans and in the human body [174]. It has been estimated that there are 10^7 viruses in every millilitre of surface sea water [29] and 10^9 virus-like particles per gram of human stool [102]. With an estimated 10^{23} infection events occurring per second globally, phages are an important driver in the evolution of their bacterial hosts as they engage in a biological arms race [160]. Some phages are also capable of facilitating horizontal gene transfer between bacterial cells through accidental packaging of host DNA into new phage particles, which has been shown to be a contributing factor in the spread of antibiotic resistance genes [43]. It has been estimated that, globally, over $2x10^{16}$ phage-mediated gene transfer events occur per second [30].

1.3.2 Taxonomy of bacteriophages

As of August 2018, there are nearly 10,000 complete phage genomes that have been published [114]. The majority of these have been sequenced in the last few years. Until recently, phages were classified morphologically when viewed by transmission electron microscopy (TEM) [2]. This led to definition of three families within the order *Caudovirales*, which includes all tailed phages and comprises 96% of known phages, termed *Myoviridae*, *Podoviridae* and *Siphoviridae*, with the classical examples of each family being the phages T4, T7 and Lambda respectively [69]. The morphological distinctions between these families are largely due to the tail, with *Myoviridae* possessing a straight contractile tail, *Podoviridae* a short

non-contractile tail and *Siphoviridae* a long flexible tail [2]. Schematic representatives of each of these families can be seen in Fig. 1.1.

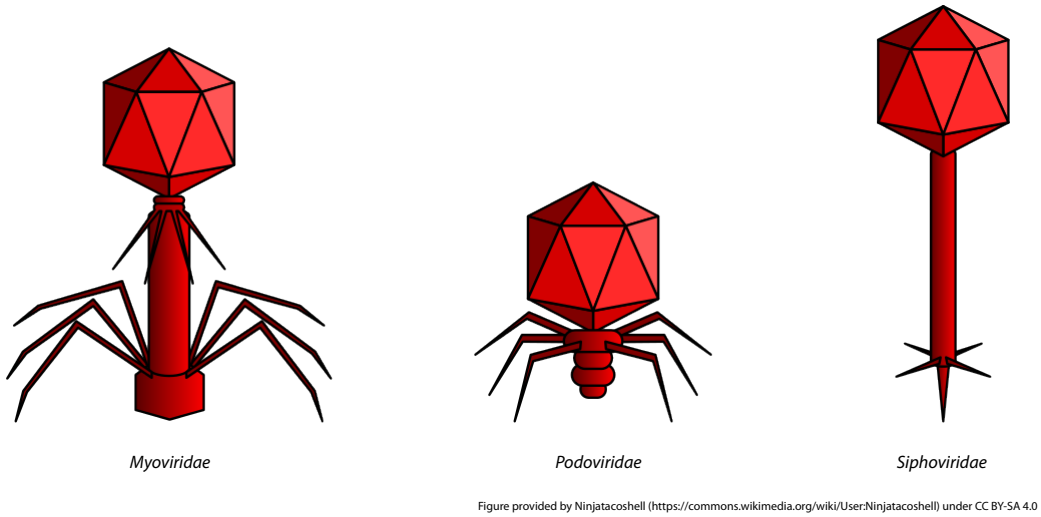


Fig. 1.1 Illustrative representations of three phage families in the order *Caudovirales*. The tails of *Myoviridae* family members are straight and contractile, those of *Podoviridae* family members are straight and non-contractile and those of *Siphoviridae* are flexible and non-contractile.

Whilst morphological classification has proved sufficient for broad grouping of phages, genomic sequencing allows a deeper level of classification to be established. This has resulted in some rearrangement of phage taxonomy, headed by the International Committee on Taxonomy of Viruses (ICTV) and led in 2018 to the elevation of the existing *Myoviridae* genus *Vi1virus* and creation of a fourth family within the *Caudovirales* order, the *Ackermannviridae* family [9]. Genomic data have also allowed classification of phages into lower taxonomic groupings such as genera and species, and there is also, apparently, discussion of abolishing the existing families in *Caudovirales* altogether [11]. Analyses of phages derived from metagenomic sequencing and in environmental samples however still face challenges, as phages do not have a conserved genetic element that can be used for phylogenetic mapping, unlike the bacterial 16S sequences [143].

1.3.3 Life cycle of a bacteriophage

To effectively complete a productive life cycle, a phage must first be capable of adsorbing to a bacterial cell through the interactions between receptor binding proteins of the phage and the receptor(s) of the host [139]. As the first step in the process, this is an important determinant of phage host range [56]. In the literature, phage host ranges are classified as

either narrow or broad, however the actual demarcation between these two categories is a subject of discussion. Recent attempts define narrow host range as infecting only one species, with broad host range including all phages capable of infecting one or more species of bacteria [56]. Receptors for bacteriophages include a wide array of surface-exposed structures, including flagella, lipopolysaccharides and capsular polysaccharide [122]. A more detailed summary of the major phage receptors is listed in Table 1.3 and shows that phages of all three traditional families of the *Caudovirales* order have been found to utilise a variety of receptors, with no obvious correlation between taxonomy and receptor usage. The receptors for *Ackermannviridae* family phages is a subject of this project and will be discussed in later chapters.

The simplified lytic life cycle of a virulent phage is shown in Fig. 1.2. Structures at the base of the phage tail recognise specific receptors on the bacterial cell surface. After adsorption, the phage particle is brought into contact with the host cell surface, followed by phage genome injection into the host. The cellular replication machinery is then hijacked by the phage in order to bias replication towards producing new phage particles [147]. Following a tightly regulated signal [67], the cell is lysed and the progeny phages are released. The genomes of some phages, known as temperate or lysogenic phages, can integrate into the host cell genome during infection, either directly into the chromosome or as self-replicating plasmids and be incorporated in progeny cells for later excision or reactivation and resumption of the viral life cycle [70]. These are known as prophages and can have significant impacts on their host, including enhancing virulence in the well-studied case of the cholera toxin [86]. A recent review of the pangenome of *Dickeya* species has identified likely prophages [72] but, to the best of my knowledge at the time of writing, no experimental work into the role of these prophages has been published. All of the phages discussed in this project have only been observed to follow a lytic life cycle.

1.3.4 Phage therapy

As soon as phages were discovered in the early twentieth century, the possibility of utilising them as a therapeutic for bacterial infection was recognised and investigated, leading to a variety of phage preparations being proposed [147]. However, a lack of rigour in experimentation, coupled with the discovery of antibiotics, led to a decline in the use of phage therapy in Western countries, particularly after the Council on Pharmacy and Chemistry of the American Medical Association concluded that the efficacy of phage therapy was ambiguous [159]. Recently the prevalence of antimicrobial resistance, coupled with a dearth of novel antimicrobials, has generated renewed interest in the use of phages and phage products as antibacterial tools, as well as in bacterial detection methods [151]. A recent clinical

Receptor	**Phage**	**Phage family**
Lipopolysaccharide	Mu	*Myoviridae*
	T7	*Podoviridae*
	T5	*Siphoviridae*
Flagella	SPN3US	*Myoviridae*
	None reported	*Podoviridae*
	Chi	*Siphoviridae*
Outer membrane proteins	T6	*Myoviridae*
	T1	*Podoviridae*
	ϕ80	*Siphoviridae*
Pili	ϕTMA	*Myoviridae*
	MPK7	*Podoviridae*
	DMS3	*Siphoviridae*
Teichoic acids	A511	*Myoviridae*
	phi29	*Podoviridae*
	LL-H	*Siphoviridae*
Capsular polysaccharide	ViI	*Myoviridae*
	ViIII	*Podoviridae*
	ViII	*Siphoviridae*

Table 1.3 Host receptors for bacteriophages in Gram-negative bacteria and examples from each of the three major families in the *Caudovirales* order. The majority of this table is populated with examples from Nobrega *et al.* [122]. ϕTMA was published by Tamakoshi *et al.* [163].

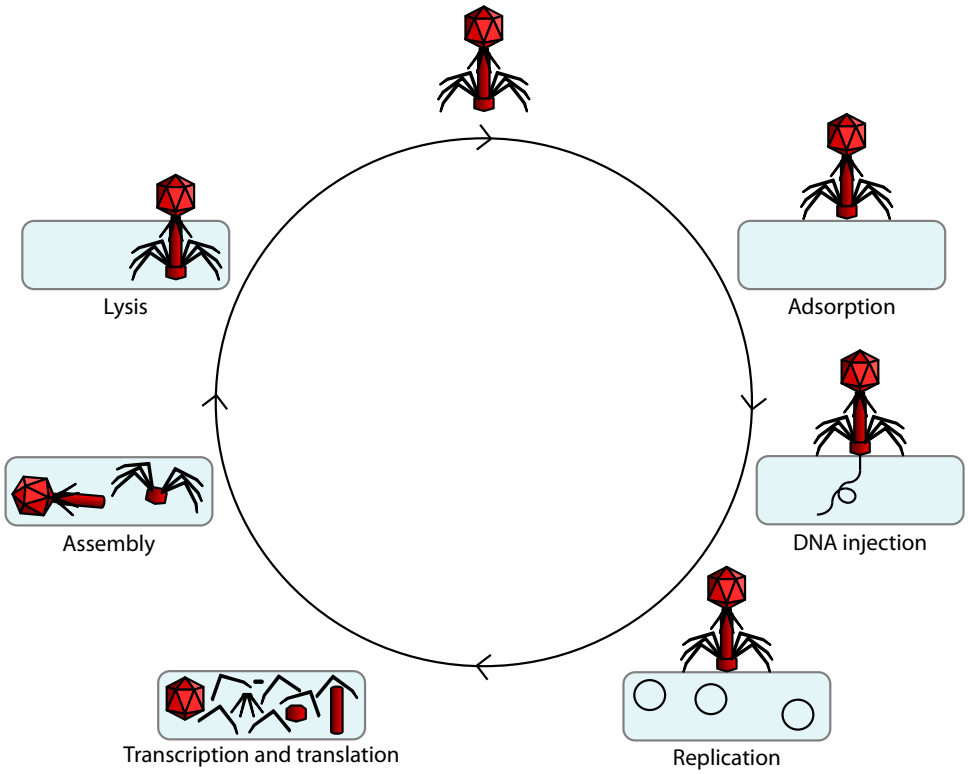


Fig. 1.2 Schematic of the lytic life cycle of a phage. Free phage in the environment encounter bacterial cells and adsorb to receptors on the surface of the host. Once the phage particle is in contact with the host cell surface, the phage genome is injected. The host replication machinery is hijacked by the phage and used to produce structural components of new phage particles as well as copies of the phage genome. The phage progeny assemble inside the host cell before it is lysed, releasing the phages into the extracellular environment. Phage adapted from Ninjatacoshell (https://commons.wikimedia.org/wiki/User:Ninjatacoshell) under CC BY-SA 4.0.

trial investigating the use of phages for treating infected burn wounds, named PhagoBurn (http://www.phagoburn.eu/) concluded in 2017 and results are due to be published soon. The use of phages as medical therapeutics however is hindered by the lack of a regulatory framework, although recent discussions in Belgium have started to investigate this hurdle [134]. Use of phages in food safety and agriculture has faced fewer barriers and therefore there are several commercially available phage products such as PhageGuard Listex (Micreos; the Netherlands) for *Listeria* and Biolyse (APS Biocontrol; United Kingdom) for *Pectobacterium* and *Dickeya* species. There has also been the development of phage-based products, which utilise the lytic enzymes of the phages alone to control bacterial growth, such as the Staphefekt product (Micreos; the Netherlands), which uses a lytic phage enzyme for treatment of *Staphylococcus aureus* skin infections.

The significant economic costs inflicted by *Dickeya* species have stimulated research interest in methods for control of these virulent phytopathogens. Phages have been suggested as potential tools for biocontrol, and are promoted by commercial entities, due to their specificity, environmental persistence and biological 'organic' nature [161]. Several studies have isolated phages capable of infecting *Dickeya* species [8, 13, 48–50, 54, 55, 107]. Their potential use as biocontrol agents has been trialled both in the lab and in the field and these studies showed a partially 'therapeutic' outcome with reduced crop losses [8]. There is a commercial product available, Biolyse, that is a phage cocktail able to target *Pectobacterium* as well as *Dickeya* species. Designed as a washing solution for potatoes during factory processing, it is considered the first, and currently the only, commercial *Dickeya*-targeting biocontrol product. It has been reported that Biolyse has been used in the production of potatoes sold by the UK supermarket chain Tesco [16]. The identities of the phages contained within this cocktail however have not been reported in the public domain.

Whilst there has been much excitement regarding the application of phage therapy, and cut through to the mainstream media [153], phage therapy still faces a series of hurdles. It is commonly accepted that temperate phages, which persist in the host cell, are unsuitable for use as a therapeutic agent [161]. Whilst the biological nature of phages is discussed as a positive, as it allows adaptation and maintenance of the phages in the environment until the pathogen is cleared [53], it should also be noted that the fact that this is a dynamic interaction also results in bacterial resistance to the phages. Bacteria have evolved a variety of mechanisms to evade or abort phage infection at all stages of the life cycle shown in Fig. 1.2 including receptor mutation, restriction/modification, abortive infection and CRISPR-Cas systems [89]. The specificity of phages is touted as an advantage over traditional antibiotics, however, this can also hinder the clearance of polymicrobial infections, for example in burn wounds, and requires an exact characterisation of the pathogen to allow efficient phage

clearance [99]. The problems with specificity and resistance can be mitigated by utilising a cocktail of phages targeting different receptors and a wider range of pathogens [37], however the need to tailor phage cocktails for individual patients also generates significant regulatory hurdles which are only just being investigated [134]. Some, but not all, phages also possess the ability to facilitate horizontal gene transfer, which can drive the spread of antibiotic resistance genes [43], therefore it is accepted that this is an undesirable trait for therapeutic phages [129]. In clinical usage, the reaction of the human immune system to the addition of phage particles is currently unclear, although there are reports that application of phage cocktails in a rodent model system resulted in increased intestinal permeability and inflammatory markers [165].

1.4 Aims

This project began with an investigation into a closely related group of phages, at the time members of the genus *Vi1virus*, but since reclassified as members of the family *Ackermannviridae*. Phages of this family isolated on *D. solani*, one of the more virulent members of the phytopathogenic *Dickeya* genus, were phenotypically and genomically characterised to compare these phages with similar phages isolated elsewhere in Europe.

Based on previous work performed in this laboratory, the receptor for *Ackermannviridae* family phages of *D. solani* was thought to be capsular polysaccharide. This project aimed to test and confirm these findings as well as investigate the capsular polysaccharide of *D. solani*, about which little has been published. Experiments were also performed with *Serratia* species to determine if capsular polysaccharides were also the receptor for *Ackermannviridae* family phages that infect this genus.

Ninety bacteriophages isolated from the environment using *D. solani* were present in this laboratory at the beginning of this project. The vast majority of these were only able to form individual plaques on *D. solani*, and not other species of *Dickeya*, and were identified as members of the same viral family; the *Ackermannviridae*. There were, however, eight phages with a broader host range capable of lysing other *Dickeya* species. These phages were therefore characterised morphologically and genomically to investigate the nature of this broader host range.

Chapter Two

Materials and Methods

2.1 Media, reagents and solutions

All media, solutions and supplements used in this study are listed in Tables 2.1 and 2.2. All solutions were prepared using deionised water unless otherwise stated and, where necessary, sterilised by either autoclaving at 121°C for 20 minutes or filtering through 0.22 μm filters (EMD Millipore).

2.2 Bacterial strains, bacteriophages and plasmids

All bacterial strains and plasmids used in this study are listed in Table 2.3. *Dickeya* and *Serratia* species were routinely grown at 30°C in Lysogeny Broth (LB) or on LB 1.5% (w/v) agar (Formedium) plates, whereas *E. coli* strains were grown at 37°C. All bacterial overnight cultures were prepared by inoculating LB with a single isolated colony before overnight incubation on a rotary wheel. Bacterial growth was assessed by measuring optical density at 600 nm using a Thermo Scientific Spectrophotometer Helios Zeta. Bacterial strain stocks were made by mixing equal volumes of an overnight culture and 50% glycerol in a CryoTube (Thermo Scientific) and were stored at -80°C. Phages used in this project are listed in Table 2.4. All were stored at 4°C in phage buffer over a few drops of $NaHCO_3$ saturated chloroform.

Medium	Ingredients per litre
Lysogeny Broth (LB)	10 g Tryptone 5 g Yeast extract 5 g NaCl
2x LB	20 g Tryptone 10 g Yeast extract 10 g NaCl
1.5% LB agar (LBA)	LB with 15 g agar (Formedium)
0.35% LB top agar	LB with 3.5 g agar (Formedium)
0.35% LB top agarose	LB with 3.5 g agarose (Sigma)
Minimal media	40 mL 1 M K_2HPO_4 14.7 mL 1 M KH_2PO_4 10 mL 10% $(NH_4)_2SO_4$ 820 μL 1 M $MgSO_4$ 10 mL 20% carbon source
Solution	**Components**
50x Tris-acetate-EDTA (TAE) buffer (per litre, pH 8.0)	242 g Tris base 57.1 mL glacial acetic acid 100 mM 0.5 M EDTA
Agarose gel	1% agarose in TAE buffer 500 ng mL^{-1} ethidium bromide
Phage buffer	10 mM Tris-HCl (pH 7.4) 10 mM $MgSO_4$ 0.01% gelatine
DNA loading dye (6x)	30% glycerol 0.25% bromophenol blue

Table 2.1 Media, buffers and solutions used in this project. All solutions were prepared using deionised water unless otherwise stated.

Supplement	**Stock solution**	**Working concentration**
Antibiotics		
Ampicillin (Ap)	100 mg mL^{-1}, stored at -20°C	100 μg mL^{-1}
Chloramphenicol (Cm)	5 mg mL^{-1} in 70% ethanol, stored at -20°C	20 μg mL^{-1}
Kanamycin (Km)	50 mg mL^{-1}, stored at -20°C	50 μg mL^{-1}
Spectinomycin (Sp)	500 mg mL^{-1}, stored at -20°C	50 μg mL^{-1}
Streptomycin (Sm)	50 mg mL^{-1}, stored at -20°C	50 μg mL^{-1}
Tetracycline (Tc)	10 mg mL^{-1} in 50% ethanol, stored at -20°C	10 μg mL^{-1}
Other supplements		
2,6-diaminopimelic acid (DAPA)	30 mM, stored at -20°C	300 μM
5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal)	40 mg mL^{-1} in dimethylformamide, stored at -20°C	40 μg mL^{-1}
4'-Methylumbelliferyl-β-D-galactopyranoside (MUG)	12.5 mg mL^{-1} in DMSO, stored at -20°C	250 μg mL^{-1}
D-glucose	20%, stored at 4°C	0.2%
D-mannose	20%, stored at 4°C	0.2%
L-fucose	20%, stored at 4°C	0.2%

Table 2.2 Supplements used in this project. All solutions were prepared using deionised water unless otherwise stated.

Bacterial strain	Genotype/characteristics	Reference
Dickeya chrysanthemi NCPBB 402	Wild type strain	[136]
Dickeya dadantii subsp. *dieffenbachiae* NCPBB 2976	Wild type strain	[136]
Dickeya dianthicola NCPBB 453	Wild type strain	[135]
Dickeya paradisiaca NCPBB 2511	Wild type strain	[136]
Dickeya solani MK10	Wild type strain	[135]
Dickeya zeae NCPBB 3532	Wild type strain	[136]
Dickeya solani MK10 strain AMD124	Transposon insertion mutant defective for β-galactosidase, Sm^R, Sp^R	[10]
Dickeya solani MK10 Mutant 6	Transposon insertion mutant of strain AMD124 with insertion in *cpsB* gene, Sm^R, Sp^R, Km^R	[10]
Dickeya solani MK10 Mutant 23	Transposon insertion mutant of strain AMD124 with insertion in *wzt* gene, Sm^R, Sp^R, Km^R	[10]
Escherichia coli DH5α	Wild type strain	Invitrogen
Escherichia coli β2163	Wild type strain	[58]
Serratia marcescens MSU97	Wild type strain	[156]
Serratia marcescens Sma12	Wild type strain	Lab collection
Serratia marcescens Sma274	Wild type strain	Lab collection
Serratia plymuthica A153	Wild type strain	[19]

Plasmid	Genotype/characteristics	Reference
pDS1028	*tetA*, *tnp*, *oriR6K*, *cat*; Cm^R, Tc^R	[117]
pECA1039-Km3	EZ::TnTM <*Not*I/KAN-3> mutant in pECA1039 *orf*5; Km^R	[27]
pKRCPN1	*tetA*, *tnp*, *'lacZ*, *oriR6K*, *aph*; Km^R, Tc^R	[117]

Table 2.3 Bacterial strains used in this project

Bacteriophage	Reference
3M	[140]
AD1	This project
AD2	This project
JA10	[10]
JA11	[10]
JA13	[10]
JA15	[10]
JA29	[10]
JA31	[10]
JA32	[10]
JA33	[10]
JA37	[10]
MAM1	[106]
XF4	[65]
XF16	[65]

Table 2.4 Bacteriophages used in this project

2.3 Bacteriophage techniques

2.3.1 Isolation of phages from the environment

Treated sewage effluent was collected from a sewage treatment plant in Cambridge, United Kingdom. River water was collected from multiple sites along the River Cam. Samples were filter sterilised using a 0.22 μm filter (EMD Millipore) before five mL of the sample was added to 2x LB along with 500 μL of an overnight culture of *Dickeya solani* MK10. This mixture was incubated overnight in a 250 mL flask at 30°C with shaking at 250 rpm. One mL of the enriched sample was mixed with 100 μL of chloroform (saturated with $NaHCO_3$) and vortexed to lyse bacterial cells. After sedimentation at 16,000 x g for four minutes, ten μL of a serial dilution series of the supernatant was mixed with 200 μL of an overnight bacterial culture and four mL of LB top agar. This mixture was poured as an overlay on an LBA plate and incubated overnight at 30°C. Single phage plaques were picked with a sterile toothpick, placed into 100 μL phage buffer, and shaken with five μL of chloroform to kill any bacteria. The phages obtained were plaque purified three times. High-titre phage lysates were then obtained by incubation of ten-fold serial dilutions of the phage overnight in an agar overlay. Those plates exhibiting near-confluent lysis (seen as a mosaic-like effect in which the plaques were close to merging) were used for lysate preparation. The top agar was removed from the plate, vortexed with chloroform before sedimentation at 2200 x g for 20

minutes at 4°C. The supernatant was removed and stored over a few drops of chloroform to produce the final lysate.

2.3.2 Determination of phage titre

The titre of a phage lysate is defined as the number of plaque forming units per mL. To determine the phage titre, a serial ten-fold dilution series of the lysate was plated out as above. Plates with between 30 and 300 plaques were used to calculate the titre.

2.3.3 Transmission electron microscopy

High-titre lysates for transmission electron microscopy were obtained as described above using 0.35% (w/v) LB agarose instead of 0.35% (w/v) LB agar overlays. Ten μL of high-titre phage lysates were adsorbed onto 400-mesh copper grids with holey carbon support films (Agar Scientific, Stansted, United Kingdom) for two minutes. The copper grids were discharged in a Quorum/Emitech K100X system (Quorum, Ringmer, United Kingdom) prior to use. Excess phage suspension was removed with filter paper and phage samples were negatively stained by placing the grids for 30 seconds in ten μL of 2% uranyl acetate neutralised with NaOH. The grids were then blotted on filter paper to remove the excess solution and allowed to air dry. Phages were examined by transmission electron microscopy at the Cambridge Advanced Imaging Centre (Department of Physiology, Development and Neuroscience, University of Cambridge) using an FEI Tecnai G2 transmission electron microscope (FEI, OR, USA). The accelerating voltage was 120.0 kV, and images were captured with an AMT XR60B digital camera running Deben software.

2.3.4 Host range

The host range of isolated phages was determined by plating out ten-fold serial dilutions of the phage lysates onto agar overlays containing host *Dickeya* cells and incubating overnight at 30°C. Following best practice to avoid potential confusion with 'lysis from without', only phages that produced individual plaques following serial dilution on three independent occasions were considered as being able to infect the respective host productively through a lytic cycle.

2.3.5 Transduction

To test for transduction, phage lysates were generated as described above on donor bacterial strains carrying the desired marker. All the experiments described in Chapter Six used

kanamycin as the antibiotic for selection. The chromosomal marker for the AD phages was a transposon stably inserted into the *lacZ* gene. Successful transduction was confirmed by kanamycin-resistant recipient colonies that were white on media containing X-gal.

Transduction was performed by mixing phage lysate with an overnight culture of the recipient cells to achieve a multiplicity of infection (the ratio of phage particles to bacterial cells) of 0.01, meaning that for each phage there were one hundred bacterial cells. The mixture was left on the lab bench at room temperature for 20 minutes, followed by incubation on a rotary wheel at 30°C for 30 minutes. The infected culture was then centrifuged and the bacterial pellets washed with LB twice to eliminate any remaining non-adsorbed phage. The bacterial pellets were resuspended in one mL LB and 100 μL aliquots were spread onto LBA plates with selection for the chromosomal or plasmid marker. Appropriate standard controls, which were routinely negative, were used to score for any spontaneous resistance of the recipient strain. 100 μL of the phage lysate was also spread onto LBA plates to confirm lysate sterility.

2.3.6 Adsorption assay

1x10^8 bacterial cells per mL were added to ten mL of LB in a 250 mL conical flask. Flasks were placed in a water bath and allowed to equilibrate to 30°C for five minutes with shaking at 100 rpm. Phages were then added at a multiplicity of infection of 0.1 and samples taken at increasing intervals between 0 and 64 minutes. At each timepoint, 50 μL of the culture was transferred into a chilled Eppendorf tube containing five μL of chloroform and then vortexed. Samples were serially diluted and plated on top agar lawns of *D. solani* to determine phage titres.

2.3.7 Phage genomic DNA extraction

Phage genomic DNA was obtained from high-titre phage lysates using a standard phenol/chloroform method and utilising sterilised silicone grease to facilitate separation of organic and inorganic phases. The purity of genomic DNA was assessed using a 1% agarose gel and was stored at 4°C. The double-stranded DNA content of extractions was assessed using the Quant-iT PicoGreen dsDNA Assay (Thermo Scientific).

2.4 Bacterial growth and virulence assays

2.4.1 Capsule expression assay

Approximation of capsular gene expression was measured throughout bacterial growth. An overnight culture was used to inoculate 25 mL of LB in a 250 mL conical flask to achieve a starting OD_{600} of 0.05. The flasks were then incubated in a water bath at 30°C and shaking at 215 rpm. 100 μL samples were taken every two hours and added to 900 μL LB before measurement of OD_{600}.

Approximation of expression utilised measurement of β-galactosidase activity, which was determined by monitoring the breakdown of the substrate 4'-Methylumbelliferyl-β-D-galactopyranoside (MUG). 100 μL samples of the culture described above taken every two hours were transferred into a 96 well plate and frozen at -80°C until needed. Samples were thawed and ten μL transferred to a new plate and frozen at -80°C for 15 minutes before thawing at 37°C. 100 μL of the reaction mixture (400 μg mL $^{-1}$ lysozyme and 250 μg mL $^{-1}$ MUG in phosphate buffered saline) was added to each well. Wells were monitored in a Gemini XPS plate reader using the following parameters: 360 nm excitation, 450 nm emission, 435 nm cut off, eight reads per well, measured every minute for 30 minutes. Relative fluorescence units (RFU) per minute were calculated during a linear phase of fluorescence increase and were normalised to the OD_{600} to generate a measurement of RFU $OD_{600}{}^{-1}$. Graphs were constructed and ANOVA statistical tests were performed using R Studio [164].

2.4.2 Potato tuber virulence assay

Potato tubers (cultivar Maris Piper) were obtained from Marks and Spencer and surface-sterilised by immersion for ten minutes in 1% Virkon solution (Lanxess) followed by washing in distilled water and air-drying. An inoculation site was bored into each side of the tuber using sterile 200 μL capacity pipette tips. Overnight cultures of the *Dickeya* strains were diluted to an OD_{600} of one in LB and then diluted a further 100,000 fold in LB. Ten μL of this dilution was used to inoculate the potato, corresponding to approximately 100 colony forming units. Inoculation sites were then sealed with silicon grease and wrapped in six alternating layers of tissue dampened with distilled water and clingfilm to prevent dehydration of the tubers. Tubers were then incubated at 30°C. At each timepoint, tubers were unwrapped and bisected across the inoculation sites. Soft rot was removed and weighed. 1 g of rot was resuspended in LB and serial dilutions plated on LB agar plates to assess colony forming

units per mL. Graphs were constructed and ANOVA statistical tests were performed using R Studio [164].

2.5 Recombinant DNA techniques

2.5.1 Generation of phage-resistant mutants

For transposon mutagenesis *E. coli* β2163 containing either the plasmid pKRCPN1 or pDS1028 was used as the donor strain. Recipient strains were *Dickeya solani* MK10 strain AMD124, *Serratia marcescens* Sma12 or *Serratia marcescens* Sma274. Overnight cultures of the donor and recipient were normalised to OD_{600} of 1 and mixed at a ratio of 1:1. 30 μL spots of the mixture were placed onto LBA plates supplemented with DAPA and incubated overnight at 30°C. The spots were scraped off the plate and resuspended in 1 mL LB and serially diluted. These dilutions were plated in a top agar lawn containing ten μL of high-titre phage lysate on plates containing kanamycin in the case of pKRCPN1 donors or chloramphenicol in the case of pDS1028 donors. These plates were then incubated for two days at 30°C. Resultant colonies were picked, streaked and confirmed for resistance to both antibiotic and phage.

2.5.2 Random-primed PCR

DNA amplification via the polymerase chain reaction (PCR) was carried out using Phusion polymerase (Thermo Scientific) and a Veriti PCR machine (Applied Biosystems). Random-primed PCR is an established technique within this laboratory [117], but in brief, combines an oligonucleotide primer specific for the transposon with several oligonucleotide primers that bind randomly throughout the genome. The random-primed PCR protocol is detailed in Table 2.5. For the first round of PCR, a bacterial colony was used as the template and resuspended in the reaction mixture. In the second round the product from the first round was used as the template. The same cycling conditions were used for both rounds. Oligonucleotide primers used in these reactions are listed in Table 2.6.

2.5.3 DNA visualisation and purification

DNA samples were visualised in 1% (w/v) agarose gels prepared as detailed in Table 2.1. DNA samples were mixed 6:1 with DNA Loading Dye before loading onto the gel. HyperLadder (Bioline) or 1kb ladder (NEB) were used as the molecular weight marker in every gel. Bands were visualised with the Syngene GeneGenius Bio-Imaging System

Reaction components in 25 μL

Round 1	
Component	**Volume**
5x HF buffer	5 μL
10 mM dNTPs	0.5 μL
10 μM PF106	0.425 μL
10 μM PF107	0.425 μL
10 μM PF108	0.425 μL
10 μM Specific primer 1 - MAMV1 or REM7	1.25 μL
Phusion	0.2 μL
DNA template	Colony
dH_2O	16.8 μL

Round 2	
Component	**Volume**
5x HF buffer	5 μL
10 mM dNTPs	0.5 μL
10 μM PF109	1.25 μL
10 μM Specific primer 2 - MAMV2 or REM8	1.25 μL
Phusion	0.2 μL
Round 1 product	2.5 μL
dH_2O	14.3 μL

Cycling conditions

Cycling step	**Temperature**	**Time**
1) Initial denaturation	94°C	3 minutes
2) Denaturation	94°C	15 seconds
3) Annealing	42°C + 1°C per cycle	30 seconds
4) Extension	72°C	3 minutes
Repeat steps 2-4 five times		
5) Denaturation	94°C	15 seconds
6) Annealing	55°C	30 seconds
7) Extension	72°C	3 minutes
Repeat steps 5-7 twenty-five times		
6) Final extension	72°C	10 minutes
7) Hold	12°C	∞

Table 2.5 Random-primed PCR protocol

Primer	Sequence (5'-3')	Description
MAMV1	GGAATTGATCCGGTGGATG	Specific for Tn-KRCPN1
MAMV2	GCATAAAGCTTGCTCAATCAATCAC	Specific for Tn-KRCPN1
PF106	GACCACACGTCGACTAGTGCNNNNNNNNNNAGAG	RP PCR
PF107	GACCACACGTCGACTAGTGCNNNNNNNNNNACGCC	RP PCR
PF108	GACCACACGTCGACTAGTGCNNNNNNNNNNGATAC	RP PCR
PF109	GACCACACGTCGACTAGTGC	RP PCR
REM7	CTAGAGTCGACCTGCAGGC	Specific for Tn-DS1028
REM8	CACAGGAACACTTAACGGC	Specific for Tn-DS1028

Table 2.6 Oligonucleotide primers used in this project

(Syngene, Synoptics Ltd.). DNA from agarose gel slices was purified using the GeneJET Gel Extraction Kit (Thermo Scientific) according to the manufacturer's instructions. Plasmid DNA was extracted from bacterial overnight cultures using the GeneJET Plasmid Miniprep Kit (Thermo Scientific), according to the manufacturer's instructions

2.5.4 Short read DNA sequencing and analysis

Purified DNA products were sequenced using GATC Biotech sequencing services. Products were sent with either MAMV1 for KRCPN1-derived mutants or REM7 for DS1028-derived mutants. Transposon insertion locations were identified using NCBI Blast.

2.6 Sequencing and bioinformatic analyses

2.6.1 Genome sequencing and assembly

Phage genomes were sequenced on the Illumina MiSeq Sequencer at MicrobesNG (Birmingham, UK). The reads were trimmed using Trimmomatic [28], assessed for quality using BWA-MEM [98] and assembled using SPAdes 3.7.1 [20] with standard settings. Except for JA10 and AD2, all generated over 140,000 reads and had higher than 100x coverage of the full genome. JA10 generated 3270 reads and had 26x coverage. AD2 generated 1899 reads and had 4.79x coverage. All assembled into one contig except AD2. Annotation was

performed using DNAMaster 5.23.2 [95]. Genome maps were generated using Circos 0.69.6 [90]. Genomes were deposited in Genbank using BankIt (NCBI) and are available under accession numbers MH929319 (3M), MH460463 (AD1), MH460459 (JA10), MH389777 (JA11), MH460460 (JA13), KY942056.1 (JA15), MH460461 (JA29) and MH460462 (JA33).

2.6.2 Comparative genomics

Genomes were compared using NCBI Blast and the Artemis Comparison Tool 13.0.0 [34]. Genomes that were identified and used in these comparisons are listed in Table 2.7. Phylogenetic trees were constructed using MEGA 7.0.26 [92]. Genetic maps were created using Snapgene Viewer (GSL Biotech LLC). CRISPR sequences in bacterial genomes were identified using CRISPRDetect [26] and corresponding sequences in phage genomes were identified using CRISPRTarget [25].

Bacteriophage	**Genbank ID and reference**
BF25/12	KT240186.1 [13]
LIMEstone1	HE600015.1 [8]
Lu11	JQ768459.1 [7]
NCTB	LT598654.1 [132]
PaBG	KF147891.1 [162]
phiRSL1	AB366653 [182]
PP74	KY084243.1 [84]
XF4	KY942057.1 [54]
Y3	KY984068.1 [32]
Yoloswag	KY448244.1 [62]

Table 2.7 Bacteriophage genomes used in this project

2.6.3 Structural modelling

Structural modelling was performed using the I-TASSER Suite [184] with standard settings. Structures were visualised using the CCP4mg molecular-graphics software [112].

Chapter Three

Ackermannviridae family phages of *D. solani*

3.1 Introduction

Dickeya species are major phytopathogens that, whilst largely identified and studied in Europe, have been responsible for recent outbreaks in the United States [103], and have been reported this year for the first time infecting potato in Pakistan [150] and Australia [179], making them a global threat. The significant economic costs inflicted by *Dickeya* species have stimulated research interest in methods for control of these virulent phytopathogens. Bacteriophages have been suggested as potential tools for biocontrol due to their specificity, environmental persistence and biological 'organic' nature [161]. Before this project began, six phages isolated using strains of *D. solani* had been published in the literature: LIMEstone1 and 2 from Belgium [8] and D3, D5, PD10.3 and PD23.1 from Poland [49–51]. These had all been isolated from the potato rhizosphere and the genomes of all, except LIMEstone2, had been published. A summary of these phages is shown in Table 3.1

Morphological comparisons of the six published *D. solani* phages using transmission electron microscopy suggested grouping into a proposed novel genus of phages, the *V1virus*, as part of the *Myoviridae* family [6]. Morphologically, this genus was characterised as having an icosahedral head and rigid contractile tail, as found in all *Myoviridae* family members [69]. The major morphological difference between the genus *V1virus* and the archetypal *Myoviridae* family member T4 are the structures at the base of the tail. Whilst T4 possesses long, slender tail fibres, the *V1virus* genus members instead had short, stubby tail spikes, as shown for LIMEstone1, JA15 and XF4 in Fig. 3.1a-c. These are usually characteristic of another family, the *Podoviridae*, represented by XF28 in Fig. 3.1d.

Bacteriophage	Isolation	Location	Genome size (bp)	Reference
LIMEstone1	2008	Belgium (soil)	152247	[8]
LIMEstone2	2008	Belgium (soil)	N/A	[8]
D5	2012	Poland (soil)	155346	[49]
D3	2013	Poland (soil)	152308	[51]
PD10.3	2013	Poland (soil)	156113*	[50]
PD23.1	2013	Poland (soil)	153365*	[50]

Table 3.1 Members of the *Ackermannviridae* family isolated on *Dickeya solani* prior to 2015. * genomes are marked incomplete, largest scaffold is reported and shows 99% identity to LIMEstone1

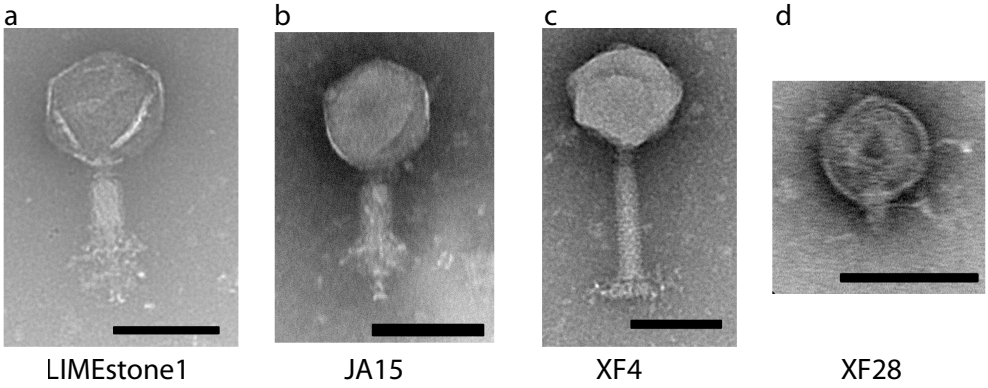


Fig. 3.1 Transmission electron micrographs of four *D. solani* phages. LIMEstone1, JA15 and XF4 are members of the *Ackermannviridae* family, with tails displayed in both a contracted (LIMEstone1 and JA15) and uncontracted state (XF4). XF28 is a member of the *Podoviridae* family. These images were provided by Xinzhe Fang [65] and Jiyoon Ahn [10] from work in this laboratory.

The genus *V1virus* has since been reclassified to form the novel *Caudovirales* family *Ackermannviridae* [9]. This family contains phages capable of infecting Gram-negative bacteria including *Salmonella* and *Shigella* species as well as *D. solani*. Sequencing of 26 members of this family has led to the demarcation of sub-families and genera. The *D. solani* phages LIMEstone1, D3 and D5 have been classified into the genus *Limestonevirus* of the sub-family *Aglimvirinae*. The other two phages, PD10.3 and PD23.1, have not been formally classified into this genus as the genomes in Genbank are marked as incomplete, however, the largest scaffolds reported, as detailed in Table 3.1, are highly similar to the full genomes of the other phages and share 99% identity. I would therefore classify these phages as members of this genus also.

At the beginning of this project, 90 phages had been isolated from the River Cam by previous lab members using *D. solani* as a host . XF1-28 and FX1-23 were isolated in October 2013 [65] and JA1-37 were isolated a year later [10]. Extensive host range testing had shown that 77 of these were capable of forming plaques on *D. solani* only and not other *Dickeya* species. The other thirteen were capable of forming plaques on other *Dickeya* species and will be discussed in Chapter Six. Morphological characterisation using transmission electron microscopy showed that, of 24 that were imaged, three were members of the *Podoviridae* family and 21 were members of the *Ackermannviridae* family. Representatives of these two families can be seen in Fig. 3.1b-d. The representatives of the *Podoviridae* family were all XF phages, whereas the *Ackermannviridae* family members were found in both XF and JA phages. The FX phages were not morphologically characterised.

There is interest in phages of *D. solani* for use as a biological tool to control the virulent phytopathogen. The previously published phages have been assessed for stability, viability and undergone limited field trials [8, 48, 50, 53]. None, however, were tested for the ability to facilitate transduction of genetic material between host cells. The authors of these studies instead reported that the absence of bacterial DNA in the genome sequence of the phages signified a lack of transduction capacity. Whilst there is a logic to this assumption, transduction is an infrequent event [43], therefore the absence of host DNA is not proof of an absence of transduction capability. In fact, testing of the XF and JA phages showed that all identified morphologically as members of the Ackermannviridae family, were capable of transduction of both plasmid and chromosomal markers between *D. solani* cells at a frequency greater than 10^{-6} [10, 65]. LIMEstone1 and LIMEstone2 were obtained from Professor Rob Lavigne in Belgium, and they too were able to efficiently effect transduction of these markers at frequencies greater than 10^{-5}, as were ViI and CBA120, two other members of the *Ackermannviridae* family that infect other hosts. It has therefore been proposed that the capacity for transduction is a characteristic of this family [107].

It was hypothesised that the phages that were classified morphologically as members of the *Ackermannviridae* family would share high similarity, when sequenced, with LIMEstone1. This is due to the high levels of identity exhibited between the genomes of the published phages from Poland with LIMEstone1. Three phages, XF4, 11 and 16, were previously genomically sequenced [65] and a summary of the findings is shown in Table 3.2. They all shared over 90% nucleotide identity with the phage LIMEstone1 and, based on the demarcations proposed by Kuhn *et al.* [91], would all be placed within the *Ackermannviridae* genus *Limestonevirus*. It was reported that variations between the genomes occurred mainly in regions coding for hypothetical proteins and the gene order was highly conserved. Variations

in the tail spike proteins of these phages were reported, and these will be discussed later in this chapter.

Bacteriophage	Isolation	Location	Genome size (bp)	Nucleotide identity to LIMEstone1
XF4	2013	UK (Wastewater)	151,450	94.2%
XF11	2013	UK (Wastewater)	153,309	95.5%
XF16	2013	UK (Wastewater)	154,083	91.1%

Table 3.2 Members of the *Ackermannviridae* family isolated in Cambridge on *Dickeya solani* sequenced prior to 2015.

The conservation of the genomes between the *D. solani*-infecting *Ackermannviridae* family members isolated across three countries and five years would seem to suggest that these phages are successful at maintaining their presence in the environment. Whether they are transient in the environment occurring only in the presence of *D. solani*, or if they maintain their presence via another reservoir(s) remains to be tested. However, this previous work has shown that representatives of this family could be found a year apart from aquatic samples of an environment not thought to contain *D. solani*. Sequencing showed that some of these isolates are members of the same genus as the phages isolated from terrestrial environments in which *D. solani* is present. Despite this high level of genome identity, the host range of these phages differs, with some of the viruses reported to form plaques on *Pectobacterium* species [48] as well as multiple *Dickeya* species [50, 53]. This genus of phages therefore warranted further investigation to determine the reason for this apparently differing host range.

3.2 Results

3.2.1 Genome of the *Dickeya solani* phage JA15

The phage JA15 was isolated in 2014 from a water sample taken in Cambridge [10] and has an *Ackermannviridae* family morphology as shown in Fig 3.1b. In this electron micrograph, the phage is shown whilst the tail is contracted, similar to LIMEstone1 in Fig 3.1a, whilst XF4 in Fig 3.1c shows the tail at its full length. JA15 is phenotypically indistinguishable from XF4; it is capable of facilitating generalised transduction at a frequency of over 10^{-5} transductants per plaque forming unit (PFU) and it forms lysates with a titre of over 10^{10} PFU per mL [10]. It is also only capable of forming plaques on strains of *D. solani* and not

***Dickeya* species**	***Pectobacterium* species**	**Other species**
D. chrysanthemi NCPBB 402	*P. carotovorum* SCRI 193	*Serratia* sp. ATCC 39006
D. dadantii 3937	*P. carotovorum* ATCC 39048	*Serratia plymuthica* A153
D. dianthicola NCPPB 453	*P. atrosepticum* SCRI 1043	*Serratia marcescens* Sma12
D. dadantii subsp. *dieffenbachiae* NCPPB 2976		*Escherichia coli* DH5α
		Pseudomonas aeruginosa PA01
D. zeae NCPPB 3532		*Pantoea agglomerans* 3Rc14
D. paradisiaca NCPPB 2511		*Serratia marcescens* Sma3888
D. dadantii Ech703		*Citrobacter rodentium* ICC169
D. dianthicola IPO 980		*Citrobacter rodentium* DBS100
Dickeya sp. MK7		*Xenorhabdus luminescens* 3Rp5
Dickeya sp. CSL RW240		

Table 3.3 Hosts not lysed by *Ackermannviridae* family phages isolated on *D. solani* including LIMEstone1, XF4, XF16 and JA15. Data in this table referencing genera other than *Dickeya*, *Serratia* and *Pectobacterium* were obtained by Jiyoon Ahn [10].

other species or genera tested, in keeping with other members of the *Limestonevirus* genus as shown in Table 3.3. It was therefore considered highly likely that JA15 would be a member of the *Limestonevirus* genus in keeping with all other *D. solani*-infecting members of the *Ackermannviridae* family. In order to compare JA15 to other *Limestonevirus* phages and determine any variations between them, JA15 was genomically sequenced.

JA15 has a circular double-stranded genome 153650 bp in length, with a GC content of 49.2% and 188 predicted genes. A map of the genome can be seen in Fig 3.2 and more detailed annotation of the open reading frames can be found in Appendix A.6. It shares 97% DNA identity with XF4, with the main areas of difference being regions encoding homing endonucleases and hypothetical proteins. A comparison of this phage with the previously published phage LIMEstone1, showed over 96% nucleotide identity, therefore it would also be placed in the genus *Limestonevirus*. As with XF4, the major areas of difference again consist of genes predicted to encode homing endonucleases and hypothetical proteins. These genes are highlighted in Fig 3.2. Homing endonucleases are mobile genetic elements found throughout phage genomes that facilitate genetic reshuffling [60]. It is therefore unsurprising that the main differences between the phages are located in these sites, however there are no putative annotated genes surrounding these endonucleases which differ. The only other annotated gene that differs between the three phages is a 5'(3') DNase, although the impact of this variation is unclear. Aside from this, each phage has between three and seven open reading frames with no annotated function that differ.

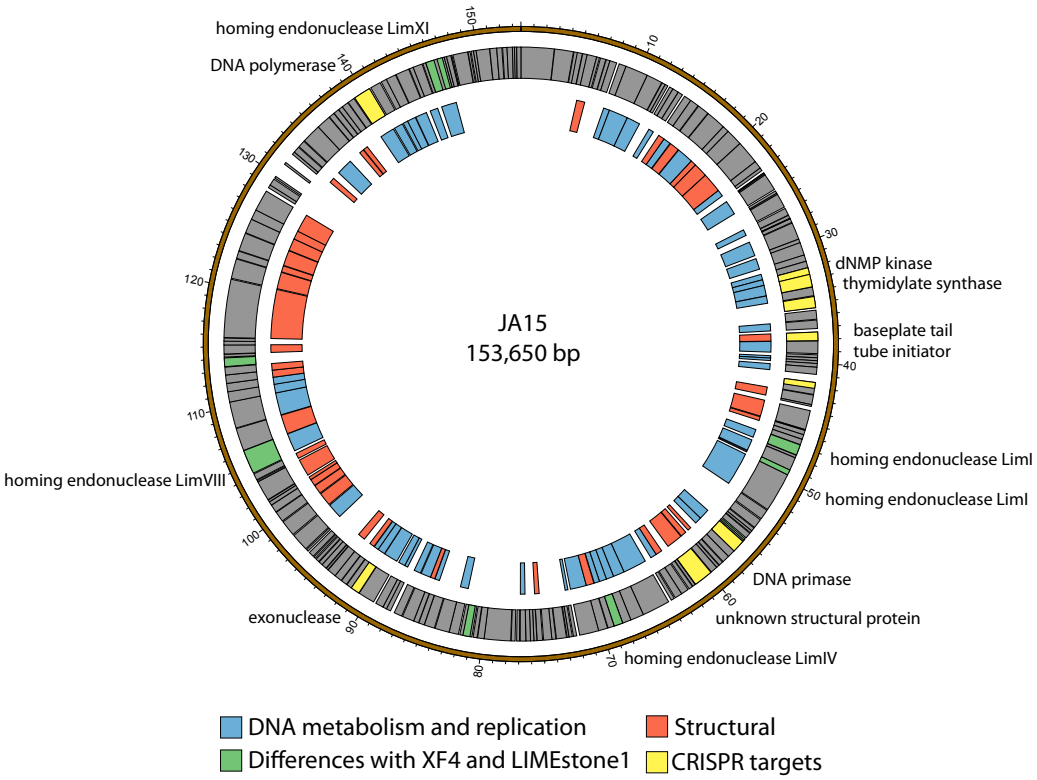


Fig. 3.2 Map of the genome of JA15. The outer grey ring marks open reading frames whilst the middle ring categorises the proposed functions of these genes. The green highlighted areas on the outer ring indicate areas of the JA15 genome that differ from the genome of the LIMEstone1 and XF4 phages and are discussed further in the text. The yellow areas on the outer ring highlight genes targeted by CRISPR systems in *Dickeya* species as discussed later in the chapter.

3.2.2 Tail spike proteins of *Dickeya solani* phages

As mentioned earlier in this chapter, it was discovered that the phages XF4, 11 and 16 showed variation in their tail spike proteins (TSPs), the structures responsible for recognition of the host bacterial cell [22]. Members of the *Ackermannviridae* family have up to four TSPs [6] and the previously isolated *D. solani*-infecting LIMEstone1 was found to possess three. This is the same for the three XF phages and, whilst XF4 and XF11 share 100% nucleotide identity in their TSPs, the TSPs of XF4 and XF16, named TSP1, TSP2 and TSP4 following the LIMEstone1 nomenclature [8], differ. A translated nucleotide comparison of these three TSPs can be seen in Fig 3.3. Whilst there is overall conservation of TSP1 (Fig 3.3a) between the two phages, sharing 88% amino acid identity, TSP2 (Fig 3.3b) and TSP4 (Fig 3.3c) are clearly quite different. Whilst the N-termini of the proteins are conserved, it would appear that the rest of the protein is quite different. The much shorter length of the

proteins in XF4 when compared to XF16 could suggest that there has been a truncation or extension, particularly for TSP4. TSP2 appears to be a very different protein between the two phages aside from the N-terminal conservation, and threading modelling showed no structural conservation (data not shown). It is therefore very interesting that these two phages are phenotypically identical and share the same host range, despite these differences in two out of the three tail spike proteins.

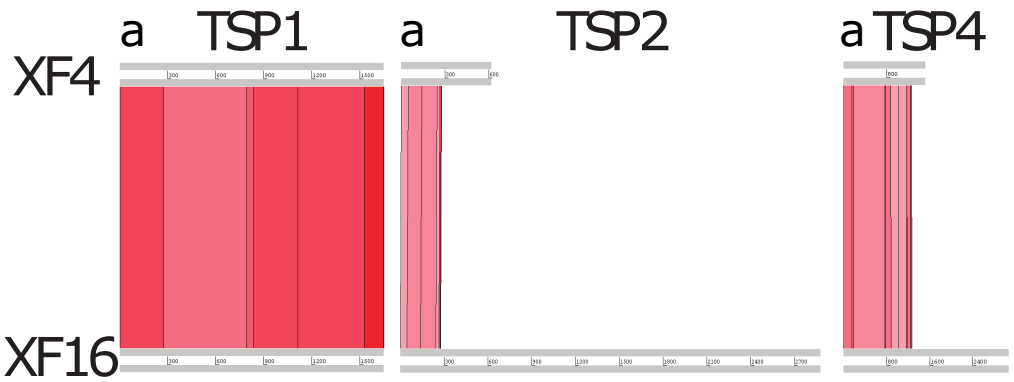


Fig. 3.3 Translated nucleotide comparison of tail spike proteins between XF4 and XF16. Red bars indicate areas of amino acid identity, with darker bars denoting higher identity.

Whilst the genomes of the four phages isolated in Poland have been published, the authors did not investigate the TSPs of these phages in any detail. In order to determine if variations between the TSPs of XF4 and XF16 were a novel finding, or if it were common among the *Ackermannviridae* family phages of *D. solani*, a comparison of TSPs was performed. Phylogenetic trees generated from the amino acid sequence of these proteins are shown in Fig. 3.4.

The same clustering pattern is observed in all three phylogenetic trees. XF16 clusters with the phages D5 and PD10.3, whilst all of the other phages cluster with XF4. All tail spike proteins within each cluster are identical except for the TSP1 XF16 cluster, in which there is single amino acid difference between XF16 and the two Polish phages, a substitution of serine to glycine at position 81. The lack of structural data makes it difficult to determine the significance of this substitution, but a threading model of XF16 TSP1 produced using the I-TASSER suite [184], and shown in Fig. 3.5, suggests that this substitution is located on the surface of the protein. The closest structural homologue is the TSP of another *Ackermannviridae* family member CBA120, which forms a homotrimer around a Zn^{2+} that interacts with a histidine at position 25 [40]. Position 25 has been marked on the predicted XF16 structure in Fig. 3.5 and is instead a phenylalanine, so the assembly of a polymeric structure may occur differently, if at all, as the two proteins share low sequence identity. The substitution at position 81 therefore, based on this model, would have no obvious effect

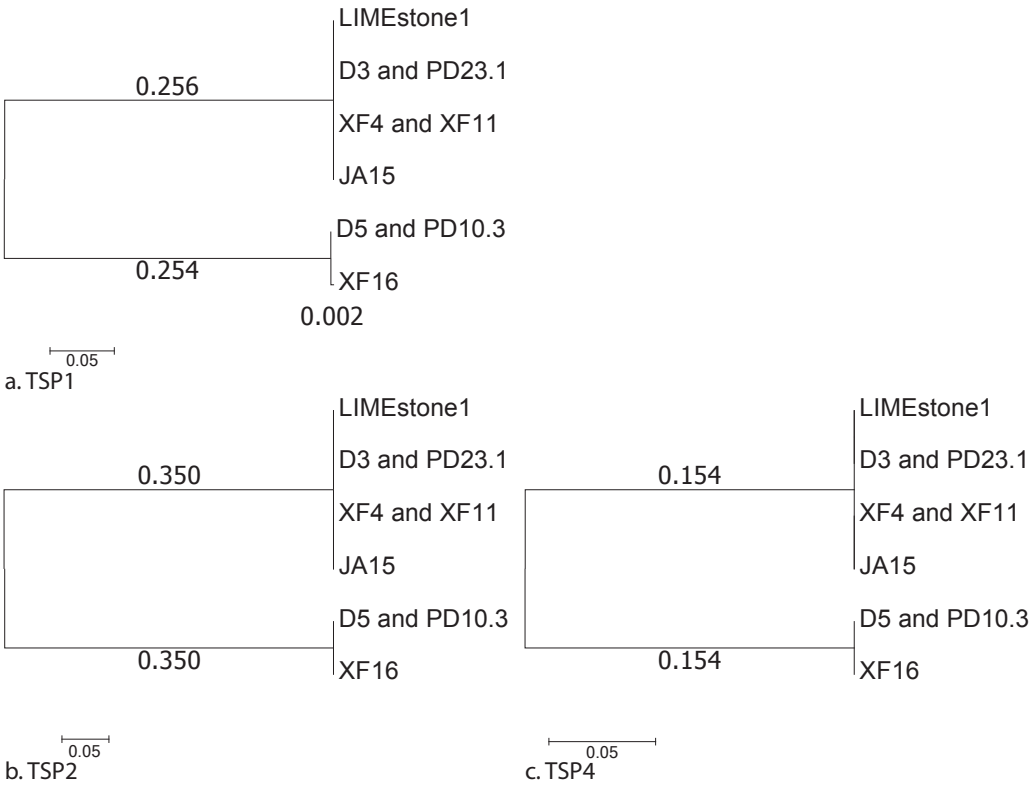


Fig. 3.4 Phylogenetic tree of the amino acid sequences of TSP1 (a), TSP2 (b) and TSP4 (c) from sequenced *D. solani* phages. Tree calculated using the Maximum Likelihood method with branch lengths measured as the number of substitutions per site out of a total of 497, 204 and 504 positions respectively. Shown are the trees with the highest log likelihood which is -2370.15, -969.64 and -2039.29 respectively.

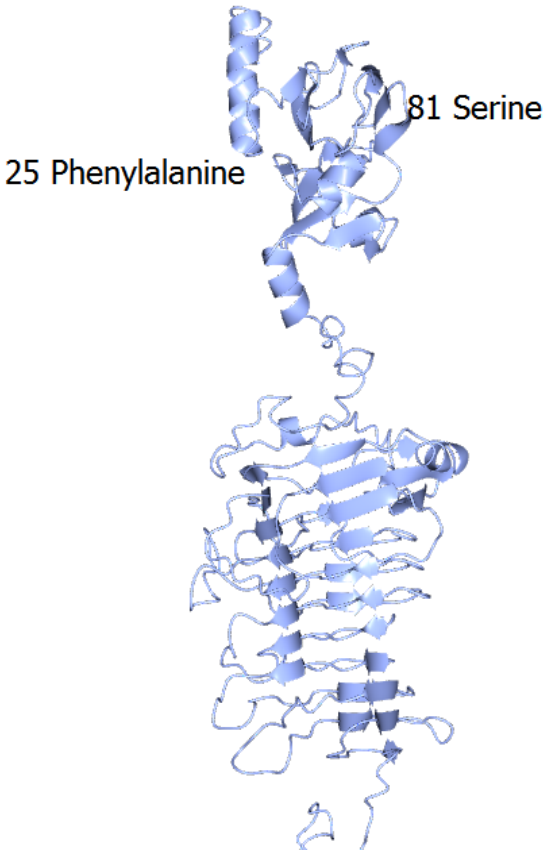


Fig. 3.5 I-TASSER model of TSP1 of XF16. The N-terminal and presumed baseplate binding domain is located at the top of this model, with the presumed host interaction domain(s) located at the bottom. C-score = -0.59, Estimated TM-score = 0.64±0.13 , Estimated RMSD = 8.9±4.6Å

on protein-protein interactions or enzyme-substrate interactions, and would require further experimentation.

3.2.3 Adsorption to *Dickeya* species

So far the only major difference discovered between XF4 and XF16 is at the sequence level. Phenotypically, and morphologically, they have been found to be the same. The phenotypes tested included transduction capability and host range, with both being able to form plaques on *D. solani* only. The formation of plaques is indicative of a full and productive lytic infection cycle, however, it is possible for a phage to begin this cycle and for it to be non-productive for a variety of reasons, often linked to host factors [89]. Whilst the host range of XF4 and XF16 is the same, despite the differences observed in their TSPs, it is possible that these changes may affect the adsorption of the phages to a potential host cell, even though they may not complete the infection cycle [139]. Adsorption experiments were therefore performed with the two phages, testing their adsorption to strains of five other *Dickeya* species as well as *D. solani*.

The black line in the top panel of Fig. 3.6 shows the adsorption pattern of XF4 to *D. solani*. The steep drop in the line in the first few minutes of the time series is indicative of adsorption to the host cells. The subsequent increase in the number of free phage that ended up exceeding the initial titre is likely the result of host cell lysis and release of phage progeny. This is a typical adsorption curve for a host in which the phage is able to complete a full replicative cycle. The phage titres when incubated with the bacteria represented in blue, *D. zeae* and *D. paradisiaca*, showed no real change across the time series. This suggested that these cells were non-permissive for the phage XF4 and do not possess the required receptor(s) for infection to be initiated. XF4 did however appear to adsorb to the species represented in pink, as demonstrated by the steep decline in free phage in the first few timepoints. There was no subsequent increase in phage levels however, even over four hours (data not shown). This suggests that, for these bacteria, the phages were able to adsorb to the cell, but were unable to complete a productive replicative cycle.

The same pattern for all six *Dickeya* species was observed in the lower panel of Fig. 3.6. XF16 was capable of adsorbing to the same three species, *D. dianthicola*, *D. chrysanthemi* and *D. dadantii* subsp. *dieffenbachiae*, but did not complete a replicative cycle. The portion of the *D. solani* curve in which the amount of free phage remained relatively stable between four and thirty-two minutes is characteristic of the latent period, in which replication of new phage particles occurs. This was not observed for XF4 and could suggest that XF16 had a slower replication cycle. The number of phage also does not decrease as much in the lower panel as it does in the upper panel. This could suggest different rates of adsorption and infection.

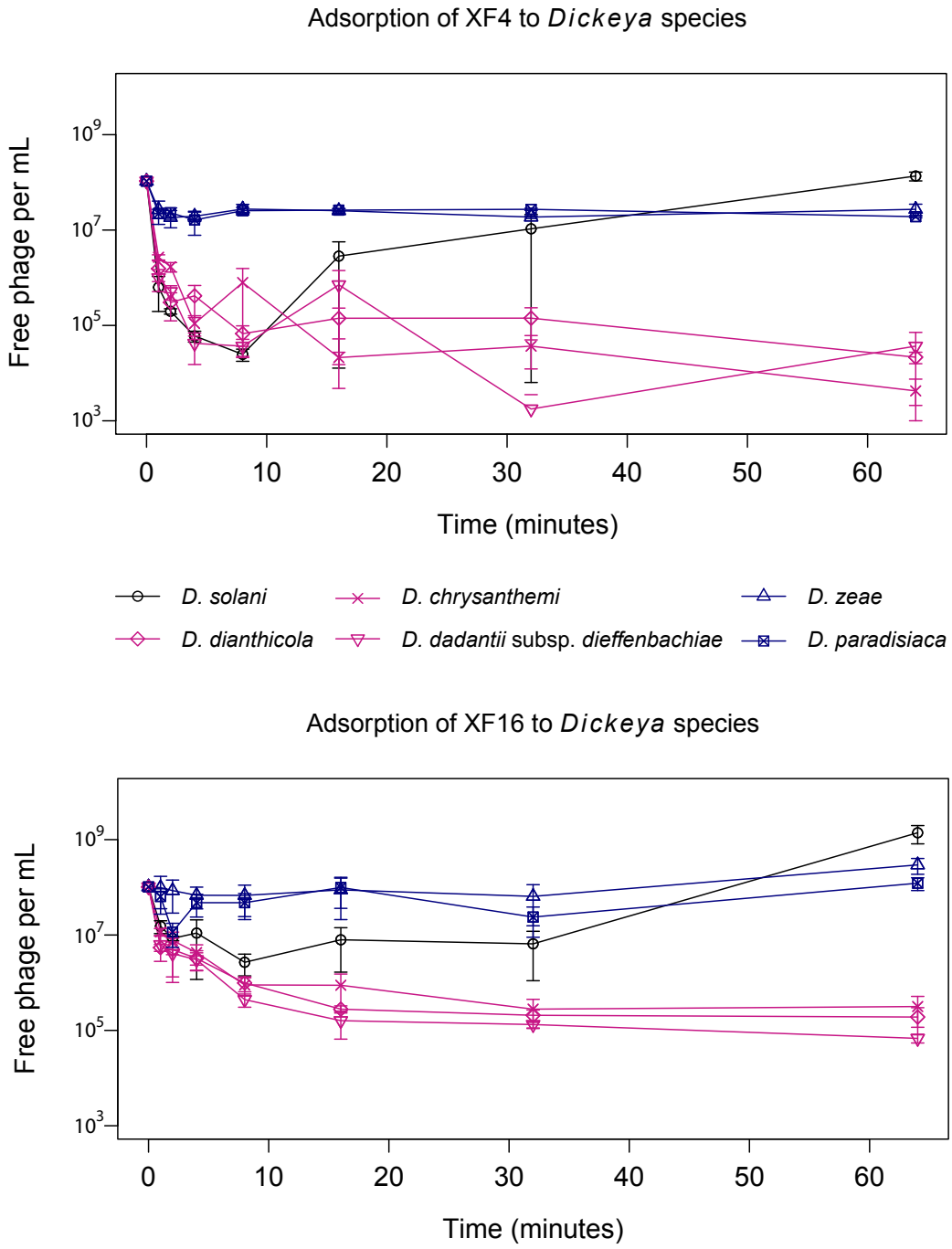


Fig. 3.6 Adsorption data for XF4 and XF16 on six *Dickeya* species. Experiments consist of three repeats, points represent averages and error bars show standard error of the mean. Experiments involved incubation of $1x10^9$ phages with $1x10^{10}$ host bacteria in liquid culture, resulting in a multiplicity of infection of 0.1. Samples were taken at increasing time intervals over an hour and immediately mixed with chloroform to kill bacterial cells. The samples were then serially diluted and plated on a semi-solid agar top lawn of the hosts to determine the phage titre. Adsorption is indicated by a decreasing titre of free phage particles, suggesting that they are no longer in solution and have instead adsorbed to the bacterial cells.

However, these data show that the adsorption of XF4 and XF16 was not significantly affected by the observed differences in TSPs. The effect (if any) of this difference therefore remains to be determined.

3.2.4 CRISPR immunity to XF4 and XF16

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated systems are bacterial immunity systems that function by introducing site-specific double-stranded DNA breaks in invading viral DNA. This specificity is conferred by short sequences, known as spacers, which are found in CRISPR arrays in the bacterial genome and match sequences found in the phage genome. These spacers are transcribed into CRISPR RNAs, which, upon hybridisation with target phage sequences, recruit Cas nucleases which cause the double-stranded breaks [83]. This is an effective immunity system employed by bacteria, first reported in 1993 [116], that has recently become one of the most high-profile areas of biological research as it can be adapted for gene editing [5].

The patterns seen for the adsorption, in which both XF4 and XF16 are able to adsorb to three hosts but not form plaques, are indicative of abortive infection, most likely due to a host factor, which could include CRISPR. Thanks to the high level of interest surrounding CRISPR, it is now possible to screen bacterial genomes for CRISPR arrays bioinformatically. All six hosts used for the adsorption experiments were screened, and all were found to contain CRISPR arrays, with most containing at least four, as summarised in Table 3.4.

Bioinformatic searching of the XF4, XF16 and JA15 genomes to identify any sequences matching the spacers in these six *Dickeya* species revealed that *D. dianthicola* NCPBB 453 possessed three spacers that had 100% nucleotide identity to all three, and that *D. chrysanthemi* NCPBB 402 and *D. dianthicola* NCPBB 453 both possessed three that shared over 82% identity with regions of the genomes. The six spacers found in *D. dianthicola* and the corresponding regions of the three genomes are shown in Fig. 3.7. Two of the spacers which share 100 % nucleotide identity to the phage genomes match genes with no known function, whilst the other matches a putative exonuclease. The other three share a lower but significant nucleotide identity with the genomes, and match genes annotated as a dNMP kinase, a baseplate tail tube initiator and an unknown structural protein. The location of the genes encoding these proteins are marked on the JA15 genome in Fig. 3.2. The four annotated proteins are likely important for the viral lifecycle, therefore it may be difficult for the phages to mutate to evade these CRISPR sequences. These CRISPR arrays also match homologous sequences in other members of the *Limestonevirus* genus including LIMEstone1, D3 and PD10.3.

D. chrysanthemi NCPBB 402				
Array	1	2	3	4
Spacers	25	16	10	24
Location	760417-758887	769121-770109	775725-775097	1589261-1587642

D. dianthicola NCPBB 453					
Array	1	2	3	4	5
Spacers	18	10	4	8	9
Location	1480224-1479002	1513662-1512967	3542969-3542697	3756775-376266	3765940-3766510

D. dadantii* subsp. *dieffenbachiae NCPBB 2976				
Array	1	2	3	4
Spacers	39	5	29	16
Location	385677-387477	387604-387938	701182-702892	703184-704192

D. paradisiaca NCPBB 2511				
Array	1	2	3	4
Spacers	43	62	3	13
Location	816259-813648	824846-828651	3876682-3876469	4132455-4131633

D. solani MK10	
Array	1
Spacers	3
Location	3947816-3948025

D. zeae NCPBB 3532					
Array	1	2	3	4	5
Spacers	8	10	8	16	22
Location	953830-953232	963026-962288	3736657-3737167	3743264-3742276	3752680-3754030

Table 3.4 Summary of the CRISPR arrays and number of spacers found in the genomes of six *Dickeya* species using CRISPRDetect [26].

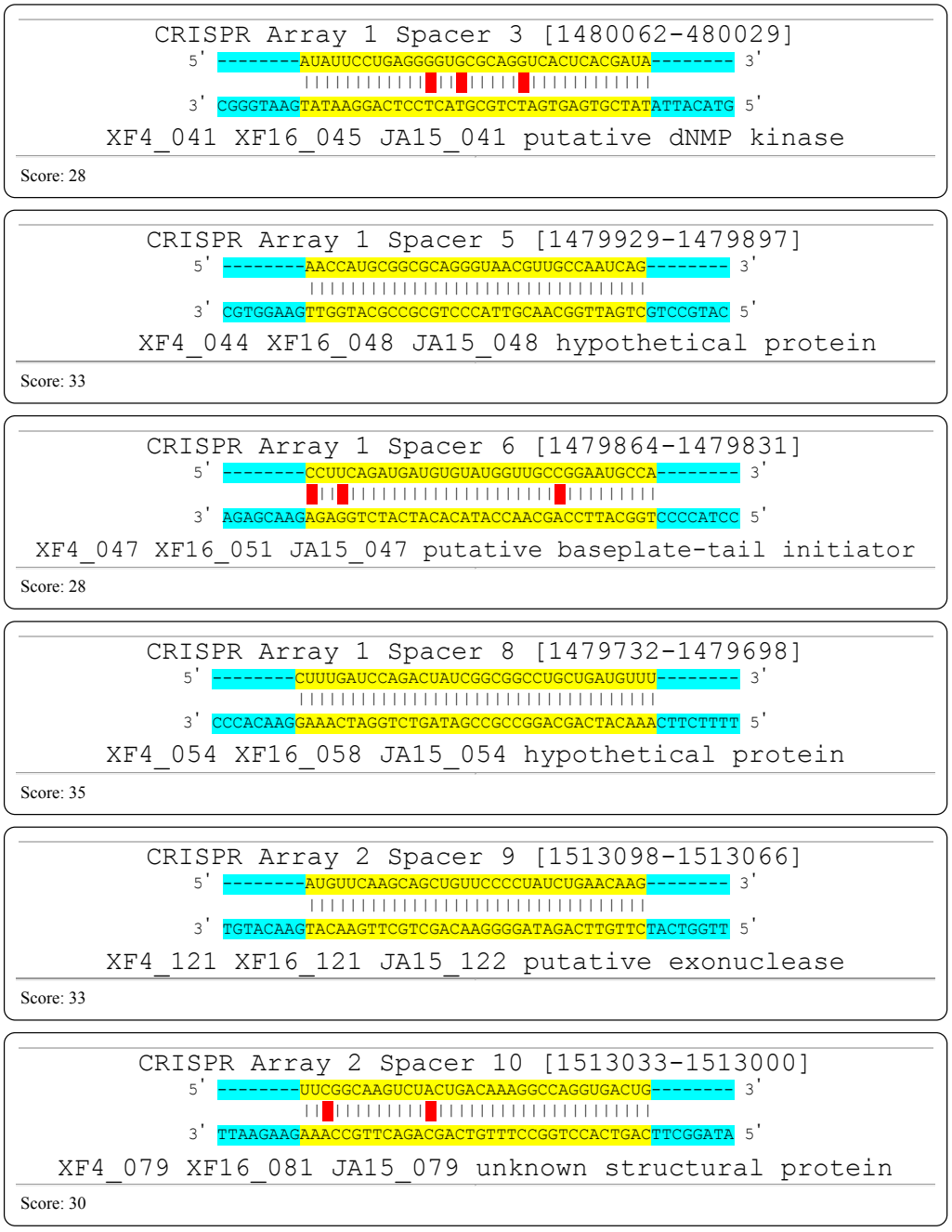


Fig. 3.7 Sequence of six *D. dianthicola* NCPBB 453 CRISPR spacers that match with sequences in the XF4, XF16 and JA15 genomes discovered using CRISPRTarget [25].

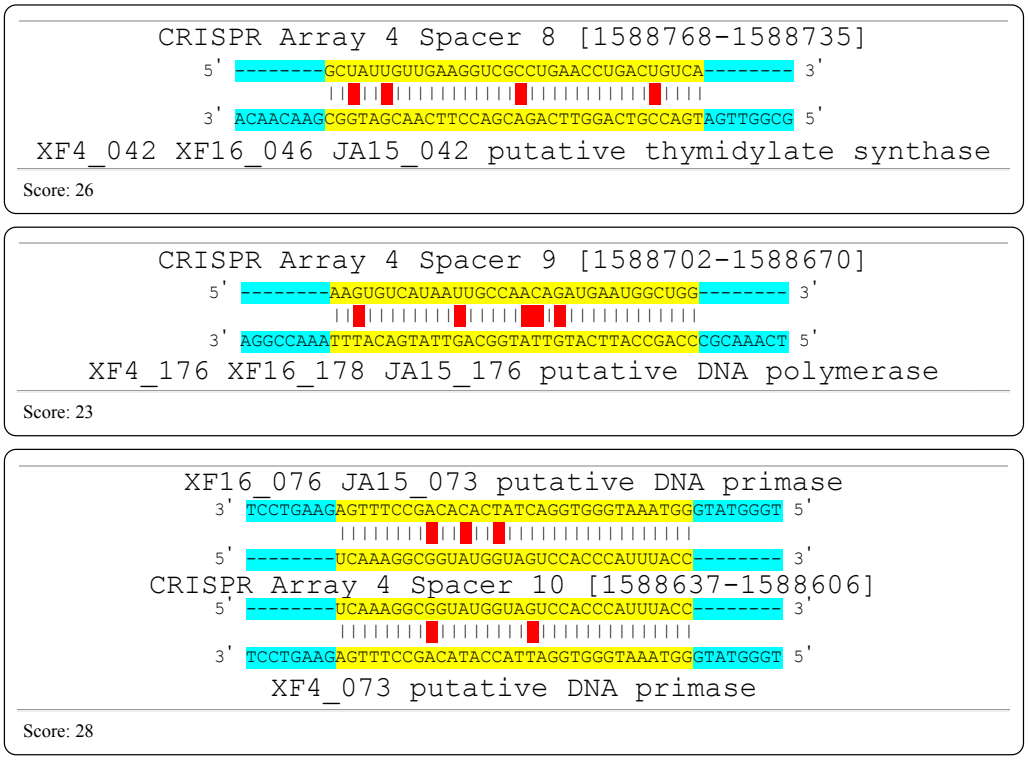


Fig. 3.8 Sequence of three *D. chrysanthemi* NCPBB 402 CRISPR spacers that match with sequences in the XF4 genome discovered using CRISPRTarget [25].

D. chrysanthemi NCPBB 402 possesses three CRISPR arrays which match sequences in the XF4 genome. These are shown in Fig. 3.8 and match genes annotated as a DNA polymerase, a DNA primase and a thymidylate synthase. These are again likely critical for the viral life cycle, limiting the ability of the virus to mutate and escape this bacterial defence. The sequences are also conserved in XF16 and JA15, apart from the DNA primase, the sequence of which is shown in Fig. 3.8 and shows minor nucleotide variations. The locations of the genes encoding these proteins are marked on the JA15 genome in Fig. 3.2.

It is therefore possible that the reason XF4 and XF16 are able to adsorb to *D. chrysanthemi* and *D. dianthicola* but not undergo a full lytic cycle is due to the action of the CRISPR systems of the host bacteria, which can recognise multiple sequences within the phage and cleave the DNA, aborting the infection. This is only a hypothesis however, and requires experimental testing. It remains a mystery why *D. dadantii* subsp. *dieffenbachiae* exhibits the same adsorption phenotype, as none of its CRISPR arrays match sequences within the phage genomes.

3.3 Discussion

Members of the *Ackermannviridae* family have been isolated using *D. solani* in three European countries over several years, from both soil and aquatic environments. Prior to this project, 77 such phages had been isolated from sewage outflow and river water in Cambridge and were all found to be capable of forming plaques on *D. solani* only. Genomic sequencing of four representatives showed a high conservation of gene order and nucleotide identity of over 90% with previous isolates from Poland and Belgium, placing all sequenced *Dickeya*-infecting *Ackermannviridae* family phages in the same genus: *Limestonevirus*. The majority of the differences between the phage genomes were found to be contained within homing endonucleases and hypothetical proteins.

A comparison of the TSPs of XF4 and XF16, which have the same host range, showed them to differ quite significantly. There is homology between the proteins at the N-terminus, but the XF4 proteins are considerably shorter than their XF16 counterparts, which could be the result of a truncation or extension, or they could be different proteins altogether. These two variants of the proteins are found in all other *D. solani*-infecting Ackermannviridae family members, with phages from the UK and Poland clustering into both of the categories. It is however interesting to see that the differences between TSPs are found in phages isolated in different locations (Poland and the UK) and from different environmental sources (soil and water), and yet there is no obvious pattern that links to the reported host range of these phages. Whilst all the other phages are reported to plaque on *D. solani* only, the four phages isolated in Poland are reported to have a wider host range to include other *Dickeya* species and even *Pectobacterium* genus members. The phages D3 and D5 are reported to share identical host ranges among *Dickeya* species [48, 49], yet cluster separately in all three trees in Fig. 3.4. The same is also true for PD10.3 and 23.1, which are reported to infect the same *Dickeya* and *Pectobacterium* species [50]. It would therefore appear that these differences have no obvious phenotypic effect. Further investigation, including structural characterisation, would be needed to fully understand the differences between these two proteins, and it would be interesting to test whether the TSPs are functionally interchangeable between the phages.

Adsorption experiments showed no obvious phenotypic effect of the differences in TSPs as XF4 and XF16 were found to have the same adsorption profile. However, they did reveal that the two phages are capable of adsorbing to three hosts on which they do not form plaques. Bioinformatic searching showed that two of these hosts contain CRISPR arrays that match likely conserved sequences within both phage genomes. This could be a contributing factor to the lack of lysis observed in these hosts following adsorption. The reason for the non-productive infection in the third host remains unknown, as it does not contain matching CRISPR arrays. It would be assumed that this adsorption pattern would be the same in other

members of the *Limestonevirus* genus, as bioinformatic searching showed that the same CRISPR arrays match sequences in these genomes as well (data not shown).

The similarity between the phages isolated from aquatic sources in Cambridge and those isolated from the soil and potato rhizosphere in mainland Europe is remarkable, with a nucleotide identity of over 90%. *D. solani* is not thought to be established in the UK, as the few isolated cases that have been reported were found in seed originating outside of the UK [168], therefore the origin of the phages isolated around Cambridge remains a mystery. Testing so far has shown that these phages are capable of forming individual plaques on strains of *D. solani* only. The isolation of the XF and JA phages a year apart suggests maintenance of the viral population, which would logically require the presence of a permissive host in this environment. This therefore leads to two possibilities for the origin of these phages. The more troublesome prospect is that, despite current thinking and the results of extensive testing by the Scottish government [17], *D. solani* is present in the environment around Cambridge. The lack of similar testing in England makes this a distinct possibility. However, it is also possible that there are other, as yet undiscovered, hosts of these phages present in the environment. For example, a novel species of *Dickeya*, *Dickeya aquatica*, has been isolated from waterways in England [127], and so far has only been identified in aquatic environments. It is, therefore, a formal possibility that this species could be an environmental host for the phages isolated here, but this has not yet been tested.

The majority of the interest in phages of *D. solani* stems from their potential usage as biocontrol agents to reduce the economic losses it inflicts. Many of the previous studies have therefore assessed the stability and potency of these phages, and conducted experiments *in planta* or even in the field [8, 48, 50, 53]. However, work performed previously in this laboratory showed that all the *Ackermannviridae* family members tested, mainly from the XF and JA phages but also including LIMEstone1, are capable of facilitating the transduction of chromosomal and plasmid markers between *D. solani* cells. The potential for this to occur in the field suggests that these phages are not the most suitable for use as biocontrol agents, as it is possible that this would facilitate the transfer of virulence and resistance genes between bacterial cells. It does however make them very useful in a laboratory setting for routine genetic manipulation of *D. solani*.

This project began with a discrepancy surrounding the reported host range of *Ackermannviridae* family phages isolated on *D. solani*. Despite genome identity of over 90%, the host range of four phages isolated in Poland has been reported to include species of *Dickeya* other than *D. solani* [50, 53] and even species of *Pectobacterium* [48]. This was in contrast to the reported host range of the LIMEstone phages discovered in a similar environment in Belgium [8] and that of the phages isolated in Cambridge, all of which were capable of

forming plaques on *D. solani* species only. Critical analysis of the methods detailed in the papers concerning the Polish phages revealed that the reported host range was derived by undiluted spot tests. This involves placing a spot of phage lysate onto a semi-solid agar lawn containing the host bacteria, allowing it to dry and incubating overnight. Observance of a zone of clearing after incubation was taken by the authors to indicate lytic infection. However, it is known that clearing can be due to the phenomenon of 'lysis from without', in which membrane disruption due to high phage titres causes cell lysis instead of phage infection [85]. It is therefore best practice to fully titrate the phage by serially diluting the phage lysate and incubating spots of the dilution series on a top lawn. This allows, in a productive infection, the observation of individual plaques. Presence of individual plaques is the accepted marker for phage infection in this laboratory and in much of the literature. The reported host range of these Polish phages may therefore be questionable. Following the methods of Czajkowski *et al.*, I was also able to see clearing, but not individual plaques, on *Pectobacterium* strains from an undiluted lysate of the phages discussed in this chapter (data not shown). The CRISPR spacers discussed in this chapter also match sequences in the Polish phages, which may conflict with the reports that these phages can infect these hosts [53]. Whilst I cannot say that the reported host range of these four phages is inaccurate or an artefact of the experimental procedure, it would be important to retest these results following best practice. In the absence of this confirmation, I remain to be convinced by these results.

Chapter Four

The capsule of *Dickeya solani*

4.1 Introduction

Infection of a host cell commences upon the adsorption of the phage particle to surface receptors, therefore the host range of a phage is largely determined by the receptors it recognises [139]. Knowledge of the receptor(s) of the phage is key when considering it for use in phage therapy. To prevent the development of resistance, it has been suggested that, following the protocols now in place for antibiotic usage, phages should be used in cocktails of multiple phages targeting different receptors, to reduce the likelihood of phage-resistant mutants developing [68]. Host receptors should therefore be identified in order to better predict the host range of a phage, lowering the risk of unintended off-target effects.

Reports on *Ackermannviridae* family phages have suggested that the bacterial polysaccharide capsule is a key host receptor for these phages. Wetter *et al.* [175], working with *Salmonella enterica*, showed that transfer of the capsular polysaccharide synthesis (cps) gene cluster into a previously non-permissive *E. coli* strain using a plasmid rendered the host permissive to infection by the *Salmonella Ackermannviridae* family phage ViI. Hsu *et al.* [77] discovered a *Klebsiella pnuemoniae Ackermannviridae* family phage unable to infect strains lacking a KN2 capsule and identified a corresponding capsule depolymerase. Previous work in this laboratory also generated mutants of *D. solani* resistant to the phage XF4 using transposon mutagenesis [117], with many of the transposon insertion sites mapping to the predicted cps cluster [10, 65]. The genes disrupted match those thought to be involved in the synthesis pathway of GDP-L-fucose proposed by Wu *et al.* [**?**] in *Helicobacter*, which is shown in Fig. 4.1. Whilst the capsules of *Salmonella* and *Klebsiella* have been well studied due to their role in virulence [125, 133], the capsule of *D. solani* has not; therefore its role in *D. solani* was investigated.

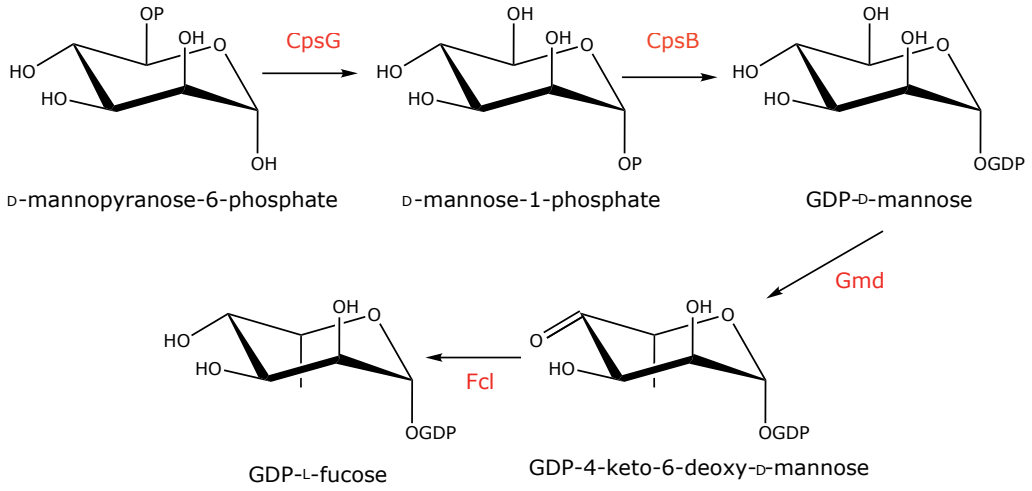


Fig. 4.1 Synthesis pathway for GDP-L-fucose (proposed by Wu *et al.* in *Helicobacter*) annotated with *D. solani* cps cluster encoded proteins.

4.2 Results

4.2.1 Phage-resistant mutants of *D. solani*

Previous work in this laboratory has generated phage-resistant mutants of *D. solani* using transposon mutagenesis. This is an established method that uses a plasmid-transposon (plasposon) capable of stably inserting into the chromosome to randomly mutagenise a host [117]. Phage-resistant mutants are isolated by simultaneously selecting for antibiotic resistance, to detect the transposon, and phage-resistance to acquire mutants in which the transposon disrupted a gene necessary for phage infection. The insertion site for the transposon can then be determined via random-primed PCR, which uses both transposon specific and random primers.

In previous work 21 independent mutants resistant to the phage XF4 were generated, and 12 of them mapped to the same genetic cluster, shown in Fig. 4.2, thought to encode the cps cluster in *D. solani* MK10 [10, 65]. The functional annotations for all eight genes are shown in Table 4.1. The transposons were found to be inserted in the first five genes in the cluster, with none in the final three genes (*fcl*, *wbeA* and *wbpZ*). As shown in Fig. 4.2, there are multiple ribosome binding sites and promoters within this gene cluster, particularly in the final few genes. It is therefore possible that these genes are transcribed separately and that they may not be involved in formation of the receptor(s) for the phage. It is also a formal possibility that these genes have functional homologues elsewhere in the genome, or that lethality might arise upon their disruption.

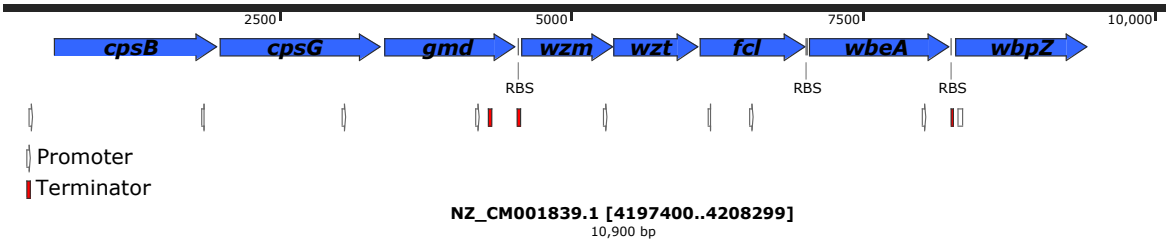


Fig. 4.2 Map of the predicted cps cluster of *D. solani* MK10. Promoters were predicted using BPROM [155], ribosome binding sites and rho-independent terminators were predicted with RegRNA 2.0 [38].

Gene	Functional annotation
cpsB	mannose-1-phosphate guanylyltransferase/mannose-6-phosphate isomerase
cpsG	phosphomannomutase/phosphoglucomutase
gmd	GDP-mannose 4,6-dehydratase
wzm	ABC transporter permease
wzt	ABC transporter ATP-binding protein
fcl	GDP-L-fucose synthase
wbeA	RfaB family glycosyltransferase
wbpZ	RfaB family glycosyltransferase

Table 4.1 Functional annotation of the *D. solani* cps cluster

In the previous mutagenesis screens, the selection for both antibiotic and phage resistance occurred simultaneously, applying a double pressure on the cells. Whilst this double selection is highly efficient and effective at generating phage-resistant mutants, it is possible that this method biases selection towards insertion sites that are more common or more stable and that the original screens may have missed other mutants, either in the three final genes of the cps cluster or elsewhere within the genome. Access to a transposon mutant library within this laboratory allowed testing of this hypothesis, with the screening split into two independent stages. Over 11,000 mutants existed in the library, all of which contained stable insertions of the transposon and were antibiotic resistant. This library was screened for resistance to the phage XF4, with 118 mutants proving to be resistant. The location of the transposons in these mutants was determined by random-primed PCR and revealed 44 independent insertion sites throughout the genome. A summary of the results is shown in Table 4.2.

Consistent with previous findings, the majority of mutants had transposons located in the predicted cps cluster. This encompassed all eight genes in the predicted cps cluster including, for the first time, the final three genes, suggesting that either the simultaneous

Group	**Cps cluster**								**LPS cluster**		**Intergenic**
Gene	*cpsB*	*cpsG*	*gmd*	*wzm*	*wzt*	*fcl*	*wbeA*	*wbpZ*	*rfaL*	*rfaB*	Various
No.	4	2	3	7	7	10	3	2	1	1	4
Total	38								2		4

Table 4.2 Genes disrupted by transposon insertion in mutants of *D. solani* MK10 resistant to XF4. Insertion sites were identified using random-primed PCR.

double selection, or potentially just the smaller sample size of the previous mutagenesis, was the reason that these three genes were not detected in the earlier data. This does however suggest that all eight genes are essential for formation of the phage receptor(s), making it likely that it is a component of the full capsule that acts as the receptor(s).

Six of the phage-resistant mutants hit regions of the genome that did not feature in the previous screens. These included two in a cluster of genes that, based on the annotation, are likely related to LPS biosynthesis. The transposons occur in genes annotated as an O-antigen ligase family protein RfaL (also known as WaaL) and a glycosyltransferase family 1 protein RfaB, which are shown in their genomic context in Fig. 4.3. The *rfa* gene cluster, in *E. coli*, has been found to contain many of the genes for LPS biosynthesis [152], and LPS also functions as a host receptor for many phages, including the classical T4 phage [173]. RfaL is responsible for linking the O-antigen to the LPS core by an unknown process [178], and RfaB is involved in modification of sugars in the inner core [157]. It is therefore possible that LPS is in some way involved in recognition of *D. solani* by the phage XF4.

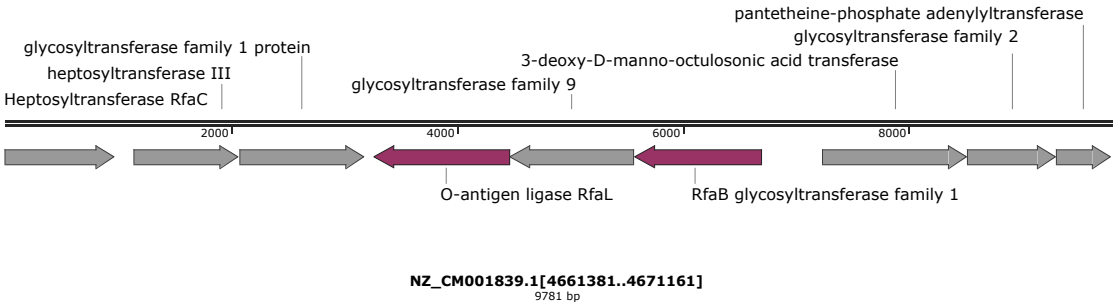


Fig. 4.3 Map of an LPS-related cluster in *D. solani*. Genes highlighted in burgundy are those disrupted by transposons in phage-resistant mutants of *D. solani*.

The other four phage-resistant mutants all have transposons inserted in different intergenic regions of the genome. Based on the location of predicted ribosome binding sites and terminators, these likely affect one gene, as summarised in Table 4.3, but could also have downstream effects on other neighbouring genes shown in Fig. 4.4. The transposon inserted

at base 1719393 is downstream of a putrescine aminotransferase, as shown in Fig. 4.4a. Putrescine transferases are involved in the regulation of polyamines within bacterial cells, which in turn regulate a variety of processes, including protein synthesis and development [35]. The insertion is before the predicted Rho-independent terminator for this protein, therefore it may interrupt synthesis of this enzyme which could have a wide range of effects, consequently it is possible that the phage-resistance phenotype is an indirect outcome of the disruption. The insertion at base 2343784 is upstream of a gene encoding a MarC family protein and the *opp* gene cluster, as shown in Fig. 4.4b. MarC family proteins were thought to be involved in multiple antibiotic resistance, however this has since been disputed [111] and the role for MarC remains undefined. The Opp proteins are involved in oligopeptide transport and are thought to be involved in environmental sensing, nutritional uptake and virulence [189]. The transposon most likely disrupts the gene encoding the MarC family protein alone, but it is unclear as to why disruption of this gene confers phage resistance.

Genome position	**Gene affected**
166390	Phage tail fibre protein
1719393	Putrescine aminotransferase
2185898	Formate C-acetyltransferase
2343784	MarC family protein

Table 4.3 Intergenic regions of the *D. solani* genome disrupted by transposons that result in resistance to the phage XF4.

The transposon inserted at base 2185898 occurs in the middle of a predicted two gene operon shown in Fig. 4.4c concerning formate, featuring a formate transmembrane transporter FocA and a formate C-acetyltransferase (pyruvate formate lyase). The transposon is inserted between the two genes, but likely disrupts only the formate C-acetyltransferase. These enzymes are involved in glucose metabolism in anaerobic conditions [96], and so, again, it is likely that the phage-resistance phenotype is an indirect effect. Base 166390 in the *D. solani* genome is in a region that, in the published genome, is intergenic and surrounded by phage structural genes and a transposase that match similar regions found in other *Dickeya* species. This region of the genome was identified by Golanowska *et al.* as a prophage of 'questionable' nature, lacking a full set, or subset, of prophage genes [72]. Further analysis of this region showed that one of the genes had been truncated in the published genome and actually is predicted to extend into the area disrupted by the transposon. The updated prophage region of the genome is shown in Fig. 4.4d and features a transposase, a tail fibre assembly protein and three phage tail collar fibre proteins, one of which is disrupted by the

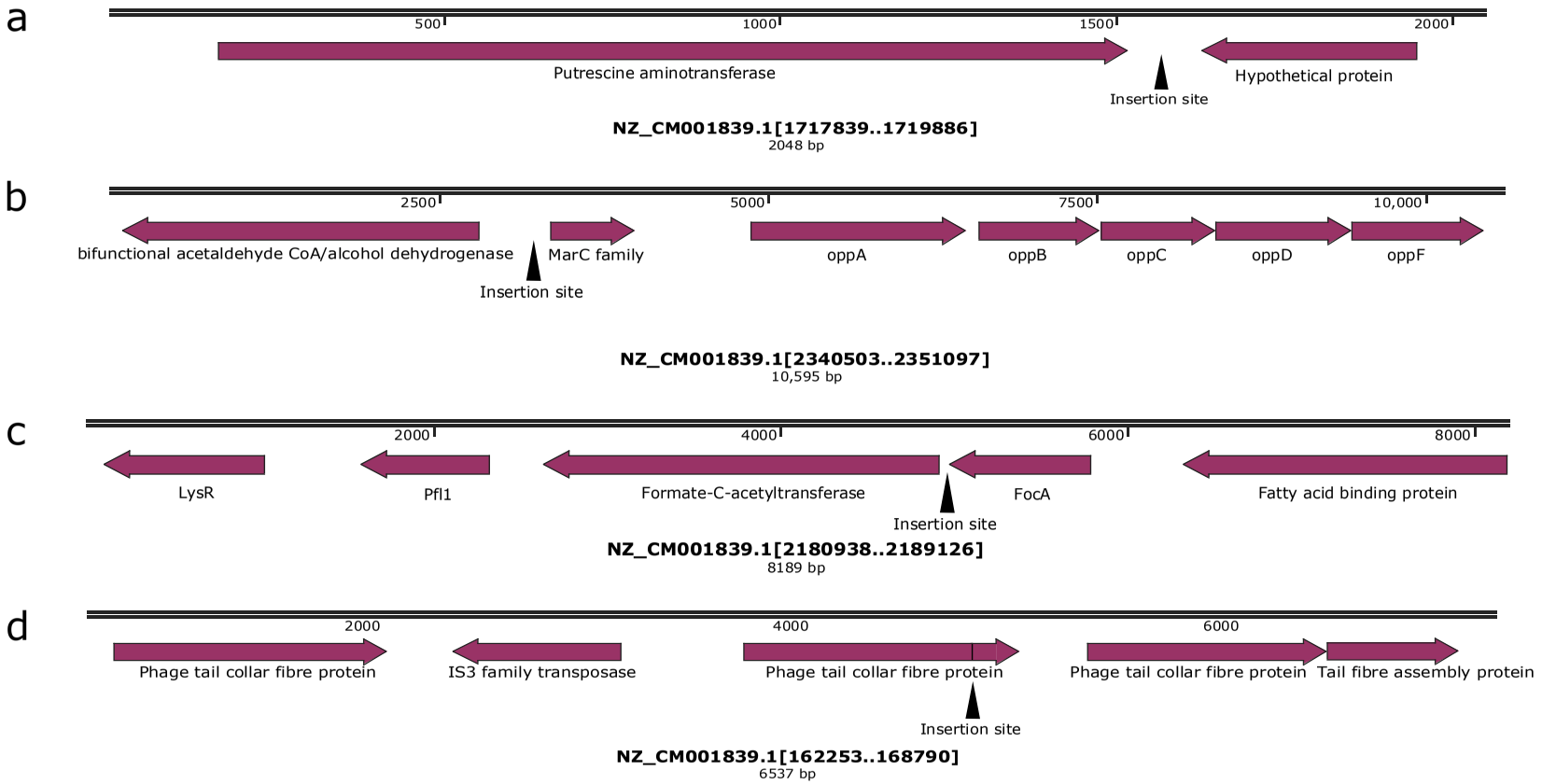


Fig. 4.4 Schematics of the four intergenic transposon insertion sites in the *D. solani* MK10 genome. The surrounding region around each insertion site is included.

transposon. It is interesting that disruption of a prophage gene results in resistance to phage, and, whilst it is possible that this could be an example of superinfection immunity, in which a phage resident in a host cell prevents infection by another phage [1], this would require further investigation.

Selection of *D. solani* phage-resistant mutants in a two-step process has revealed a number of mutants not identified in the two previous one-step screens. Several mutants outside the capsule cluster were identified for the first time, however, none of these mutants were transduced into a naive host cell. It is therefore possible that other mutations occurred during the mutagenesis that were the true cause of the phage-resistance. If these mutants were to be investigated further this would be a critical first step. The vast majority of the mutants however are still contained within the predicted cps cluster, strongly suggesting that the capsule is the key receptor for XF4. Little is known about the capsule of *D. solani* however, and so its expression and role in virulence were investigated.

4.2.2 Expression of the *D. solani* capsule

The phage-resistant mutants were generated using a plasposon which, when integrated, contains a promoterless *lacZ* gene [117]. By measuring the production of β-galactosidase, the gene product of *lacZ*, which cleaves methylumbelliferyl-β-D-galactopyranoside (MUG) to generate the fluorophore 4-methylumbelliferone, the transcription of the genes disrupted by the transposon can be assessed. This MUG assay was performed whilst monitoring the growth of *D. solani* in liquid culture to determine the timing of capsular gene expression. The transposon mutants used had been transduced into a naive Lac$^-$ background to eliminate the possibility of unknown mutations elsewhere in the genome affecting results.

The expression of two genes, *cpsB* and *wzt*, was measured in LB using the appropriate mutants and is shown in Fig. 4.5. These genes were chosen as they represent the synthetic and transport genes respectively (annotations shown in Table 4.1), and may occur in two different operons as shown in Fig. 4.2. The native *lacZ* had been disrupted by another transposon to create the Lac$^-$ strain, which was used as the recipient for the mutagenesis, and is therefore the control as there should be no *lacZ* expression in this strain. The Lac$^-$ control and both mutants grow at the same exponential rate up to four hours and then transition into stationary phase in which the optical density varies slightly between the strains, but not significantly, shown by the open symbols.

Four hours is also the peak in the β-galactosidase activity, shown by the closed symbols in Fig. 4.5 and measured as relative fluorescence normalised to the OD of the culture. The *cpsB* transposon mutant exhibits higher expression of β-galactosidase than the *wzt* mutant, potentially because it is at the beginning of the cluster as opposed to *wzt* which is one gene

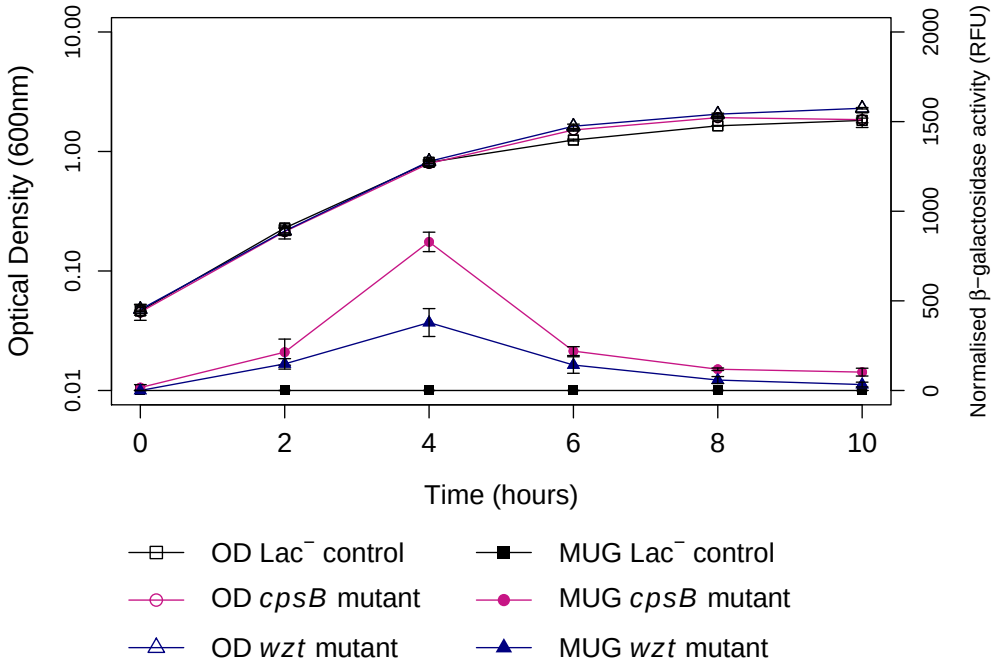


Fig. 4.5 Expression of *D. solani* genes *cpsB* and *wzt* in LB. β-galactosidase activity does not significantly differ between the two mutants with a p-value of 0.11. Experiments consisted of three independent repeats; points represent averages and error bars denote standard error of the mean.

down from a ribosome binding site, as shown in Fig. 4.2. The difference over the time course is not statistically significant however, with a p-value of 0.11. This does however suggest that, in rich media such as LB, capsule expression occurs at the transition between the exponential and stationary phases and is therefore likely a response to nutrient limitation. The reduction in the β-galactosidase activity following this peak is assumed to be due to turnover of the enzyme itself and should not be interpreted as a reduction in gene expression. A true measurement of gene expression would require isolation of mRNA and quantification using PCR-based methods.

Bacteria do not generally exist in rich media such as LB, therefore it is unlikely that, in the environment, the capsule is expressed at such a defined point as suggested by Fig. 4.5. To investigate the effect of a more realistic, nutrient poor, environment, the same experiment was performed in minimal medium with glucose as the sole carbon source and the results can be seen in Fig. 4.6. As expected, β-galactosidase activity was less uniform in minimal medium than in LB, and it is at a higher level, peaking at over 2000 Relative Fluorescence

Units (RFU) compared to below 1000 RFU in LB. The levels of expression between the two genes also significantly differ, with a p-value of $2.2\text{x}10^{-16}$. These data would suggest constitutive expression of the cps cluster during all growth phases. This would agree with the hypothesis that the capsule is expressed as a response to nutrient stress, which the cells were experiencing throughout growth in minimal medium. The reduction in activity over the time course is again likely due to enzyme turnover. The cultures used in these experiments were pre-grown in LB, a rich medium, before washing and transfer to minimal medium. It is therefore likely that the replicable initial large increase in activity observed is a response to this nutritional down-shift.

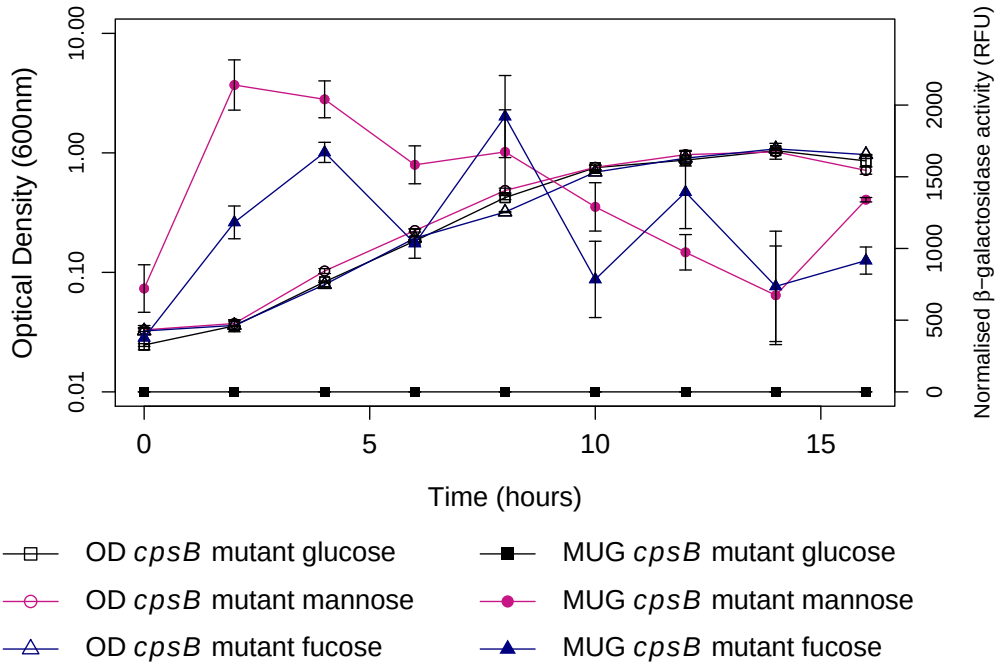


Fig. 4.6 Expression of *D. solani* genes *cpsB* and *wzt* in minimal media with glucose as the sole carbon source. The β-galactosidase activity between the two mutants significantly differs, with a p-value of $2.2\text{x}10^{-16}$. Experiments consisted of three independent repeats; points represent averages and error bars denote standard error of the mean.

4.2.3 Capsule expression in response to different carbon sources

Some of the genes encoded by the predicted *D. solani* cps cluster have functional homologues in GDP-L-fucose biosynthetic clusters in other bacteria such as *Helicobacter* [180] as well

as eukaryotes including *Caenorhabditis elegans* and *Drosophila melanogaster* [142]. The *Helicobacter* pathway proposed by Wu *et al.* [180] is shown in Fig. 4.1 with the *D. solani* homologues annotated. This shows that proteins encoded by the *cpsB*, *cpsG*, *gmd* and *fcl* genes are likely involved in this pathway. Based on this proposed pathway, it was questioned whether using mannose or fucose as a carbon source might feed into this pathway and alter the expression of the cps genes. It was hypothesised that, if there were an effect, it would mostly impact *cpsB*, as the cognate enzyme acts upstream of GDP-D-mannose and GDP-L-fucose as shown in Fig. 4.1. Another MUG assay was therefore performed, comparing the Lac^- control with the *cpsB* mutant using glucose, fucose or mannose as the sole carbon source in minimal media. The resultant data are shown in Fig. 4.7.

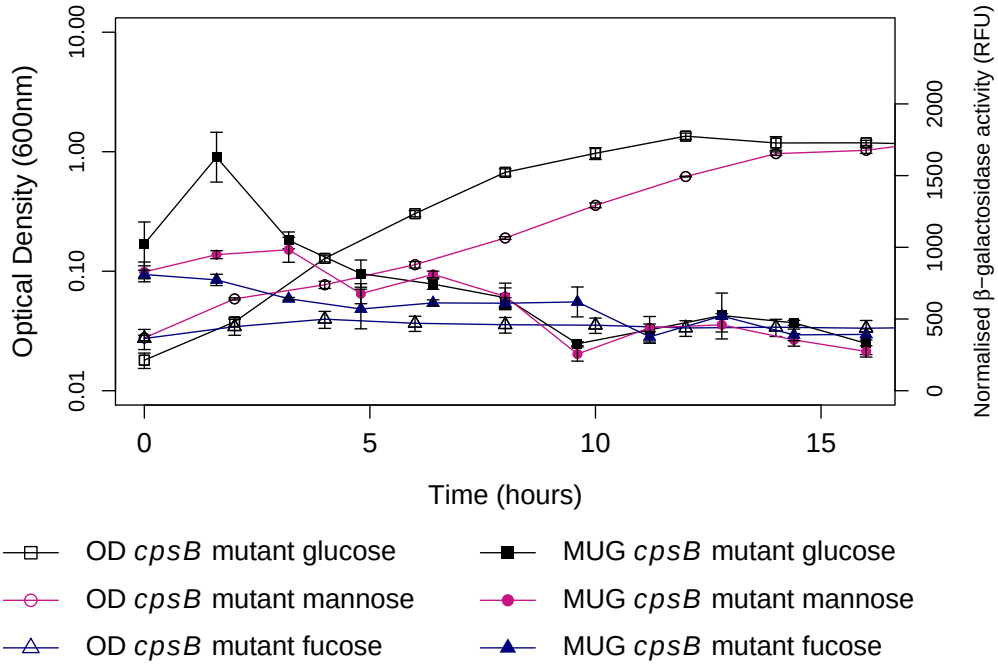


Fig. 4.7 Expression of *cpsB* in minimal media with glucose, fucose or mannose as the sole carbon source. The β-galactosidase activity between the two mutants does not significantly differ, with a p-value of 0.16. Experiments consisted of three independent repeats; points represent averages and error bars denote standard error of the mean.

These data suggest that *D. solani* is unable to grow on fucose as a sole carbon source, as the optical density stays relatively constant throughout the time series. It is able to grow on mannose and reaches the same final optical density as when grown on glucose, but grows at a slower rate, reaching stationary phase after fourteen instead of twelve hours, although

this is not statistically significant. Matching the previous experiments in minimal media, the expression of β-galactosidase is highest in the initial phase of growth and slowly decreases over the time series. The expression is lower in the cells utilising mannose as a carbon source and does not exhibit the initial spike seen in the cells using glucose, although overall the curves do not differ significantly, as an ANOVA comparison has a p-value of 0.16. It is possible however that the absence of this replicable initial spike is due to the incorporation of exogenous mannose into the GDP-L-fucose pathway, as hypothesised.

4.2.4 The capsule as a virulence determinant in *D. solani*

The capsule of a bacterial cell can be considered a virulence factor, involved in adhesion, immune evasion and environmental resilience [176], therefore the loss of the capsule can expose the cell to environmental degradation or clearance by host responses. In antimicrobial therapy it is advantageous for the target of the antibacterial agent to be a factor that is critical for virulence or survival of the bacterial cell, as evolution of the bacteria to resist the therapy may simultaneously make the cell less virulent [110]. In terms of *D. solani*, if the capsule were found to be a virulence determinant, phages that utilise the capsule as a receptor could be an effective phage therapy.

Fig. 4.5 and 4.6 show that, in liquid media, there is no effect on growth rate of *D. solani* when the cps cluster is disrupted, suggesting that the lack of a capsule does not hinder growth. This is not surprising however, as rich media are unlikely examples of environmental growth conditions for the bacteria. To test the role of the capsule in virulence in a more biologically relevant setting, assays were performed using potato tubers, a plant susceptible to *D. solani* infection. Individual potatoes were injected with 100 cells of the Lac^- ($capsule^+$) mutant at one end and 100 cells of the *cpsB* ($capsule^-$) mutant at the other. The potatoes were incubated over several days and at each timepoint the amount of rot was weighed and colony forming units (CFU mL^{-1}) were calculated from 1 g of rot. The resulting data are shown in Fig. 4.8.

The first thing to note is that data could not be obtained after 48 hours, because after this point the rot was so advanced that it had extended from the two sites to meet in the middle of the potato and so rendered discrimination between the two bacteria impossible. The data show that the CFU follow a similar trend in both bacteria, increasing after initial application before stabilisation around $1x10^8$ CFU mL^{-1}. Rot is not generated in significant quantities for the first 24 hours post-infection, but is then produced at an increasing rate up to 48 hours. This suggests that, in the potato, it takes around 24 hours for the bacteria to become established within the tuber and for virulence genes to be fully expressed. These data agree with the finding that expression of the plant cell wall degrading enzymes, which

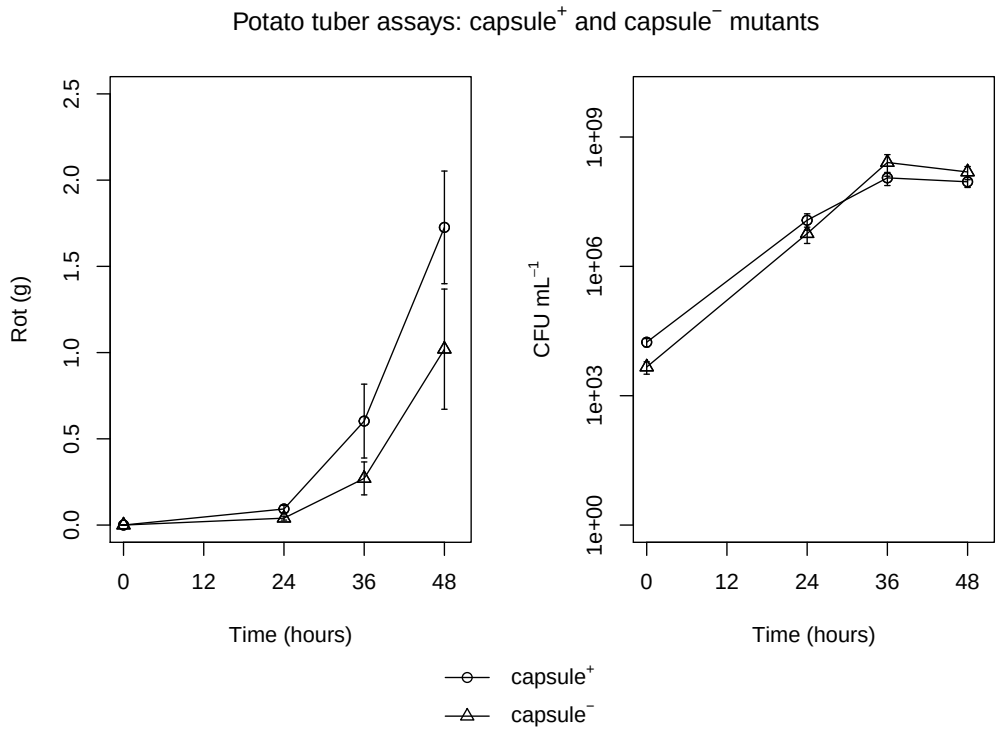


Fig. 4.8 Grams of rot generated by capsule$^+$ and capsule$^-$ *D. solani* and the corresponding CFU mL^{-1} calculated from 1 g of rot. The amount of rot generated between the two strains does not significantly differ, with a p-value of 0.17. Experiments consisted of three independent repeats; points represent averages and error bars denote standard error of the mean.

produce much of the rot, are under control of the density-dependent Vfm quorum sensing system [120]. Whilst there does appear to be an increased amount of rot generated by the capsule$^+$ cells, it is not statistically significant compared to the rot generated by the capsule$^-$ cells, as an ANOVA comparison elicits a p-value of 0.17. These data therefore show that, although the capsule may have some impact on disease aggression, it does not seem to be critical for virulence of *D. solani* in these conditions.

4.2.5 Cps clusters of other *Dickeya* species

A search of the literature shows no studies concerning the capsular polysaccharide of *Dickeya* species. Whilst it is possible that capsular polysaccharide has been investigated in *Erwinia chrysanthemi*, a previous taxonomic complex to which *Dickeya* belonged, it is difficult to determine whether these studies were conducted with bacteria that are now known as *Dickeya*. To the best of my knowledge, the only study concerning exopolysaccharide of

bacteria identified as *Dickeya* investigated LPS of *D. solani* and reported that it was composed of the rare monosaccharide 6-deoxyaltrose [124]. The capsule of *Dickeya* species therefore warrants further research. Given the data already shown in this and previous chapters, and the availability of genome sequences for representatives of several *Dickeya* species, comparison of the putative cps clusters across strains of six *Dickeya* species were made.

Using the presence of a *cpsB* gene homologue as a marker for the capsule cluster, putative cps clusters were identified from strains of six *Dickeya* species that were available for analysis; these are listed in Table 4.4. A comparison of the cps clusters shows that two of the *Dickeya* species have the same gene order and share high identity with the previously described *D. solani* cps cluster. A comparison is shown in Fig. 4.9. This agrees with the adsorption data presented in the previous chapter, which shows that phages are able to adsorb to these two hosts as well as *D. solani*.

***Dickeya* species**	**Genbank reference**	**Predicted cps cluster**
D. chrysanthemi NCPBB 402	NZ_CM001974.1	3868937-3889482
D. dadantii subsp. dieffenbachiae NCPBB 2976	NZ_CM001978.1	4087623-4098461
D. dianthicola NCPBB 453	NZ_CM001841.1	4000898-4011716
D. paradisiaca NCPBB 2511	NZ_CM001857.1	3797963-3819383
D. solani MK10	NZ_CM001839.1	4197400-4208299
D. zeae NCPBB 3532	NZ_CM001980.1	683393-710138

Table 4.4 Genbank references for six *Dickeya* species and location of the predicted cps clusters.

A comparison of the cps cluster of *D. solani* MK10 with the clusters from *D. chrysanthemi* 402, *D. zeae* 3531 and *D. paradisiaca* 2511 however shows a greater difference. Schematic maps of these cps clusters can be seen in Fig. 4.10. A comparison of the clusters of *D. solani* with *D. chrysanthemi* shows that, whilst the *D. chrysanthemi* strain possesses the same eight genes as *D. solani* MK10, there are also additional genes, including a transcriptional regulator upstream and eight genes downstream that include transporter and synthesis genes. Whilst these transporter genes are functional homologues of the *wzm* and *wzt* genes, they share low nucleotide identity. Whether these additional genes are involved in capsular polysaccharide synthesis therefore is unclear. The adsorption data presented in the previous chapter could be interpreted to suggest that they are not, as *D. solani* phages were still able to adsorb to this host.

The predicted cps cluster of *D. zeae* 3531 has the additional gene homologues seen in the *D. chrysanthemi* 402 cluster, and features homologues of all eight genes present in the *D.*

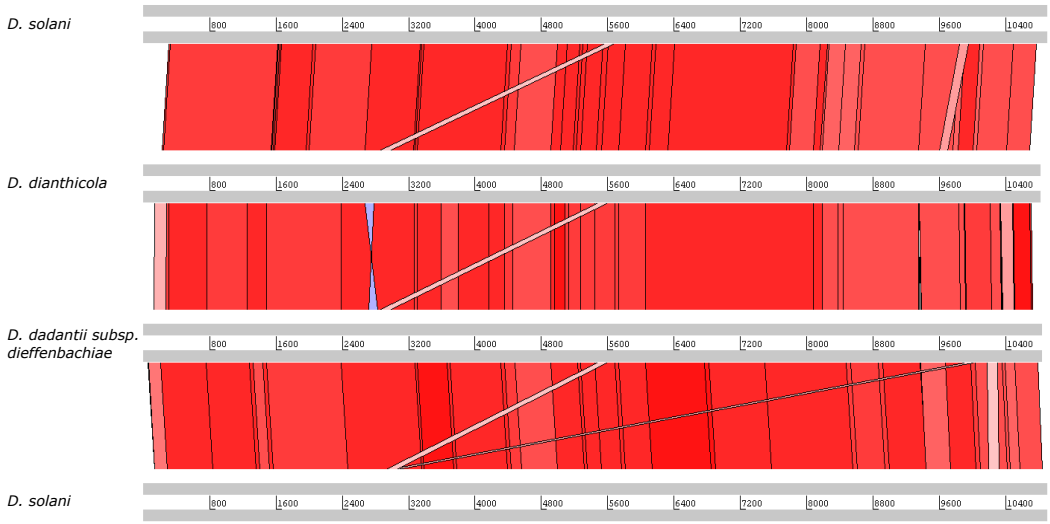


Fig. 4.9 Translated nucleotide comparison of the predicted capsular polysaccharide synthesis clusters of three *Dickeya* species. Red bars mark areas of conservation, with darker colours showing higher conservation. Blue bars highlight areas of inversion.

solani MK10 cluster. These genes however appear in a different order, and have a series of eight genes in the middle encoding mostly hypothetical products. In the previous chapter, the adsorption data showed that *D. solani* phages could not adsorb to this host, suggesting that this rearrangement and insertion, compared to the *D. solani* MK10 cluster, could have significant effects on the capsular polysaccharide of the bacteria. This is also true of the cps cluster of *D. paradisiaca* 2511, which is even more divergent as it lacks homologues of the *wzm* and *wzt* genes which constitute the export machinery for capsular polysaccharide [176]. In this host capsular polysaccharide is presumably exported via a different mechanism, potentially using the other transport genes present in the cluster.

Comparison of the predicted cps clusters from these six hosts agrees with the previous adsorption data and shows that, whilst the *D. solani* MK10 cluster is relatively small and assumed to be self-contained, at least for synthesis and export to the periplasm, other species of *Dickeya* possess more complicated cps clusters. Based on the hypothesis that the capsular polysaccharide is the receptor for *Ackermannviridae* family phages, it would appear that the phages presented here could have the potential to form plaques on a broader range of *Dickeya* species than has been found. As discussed in the previous chapter, host systems such as CRISPR likely contribute to the lack of plaque formation when tested in these specific hosts, but it remains possible that other strains of these *Dickeya* species would be permissive to *Ackermannviridae* family phages.

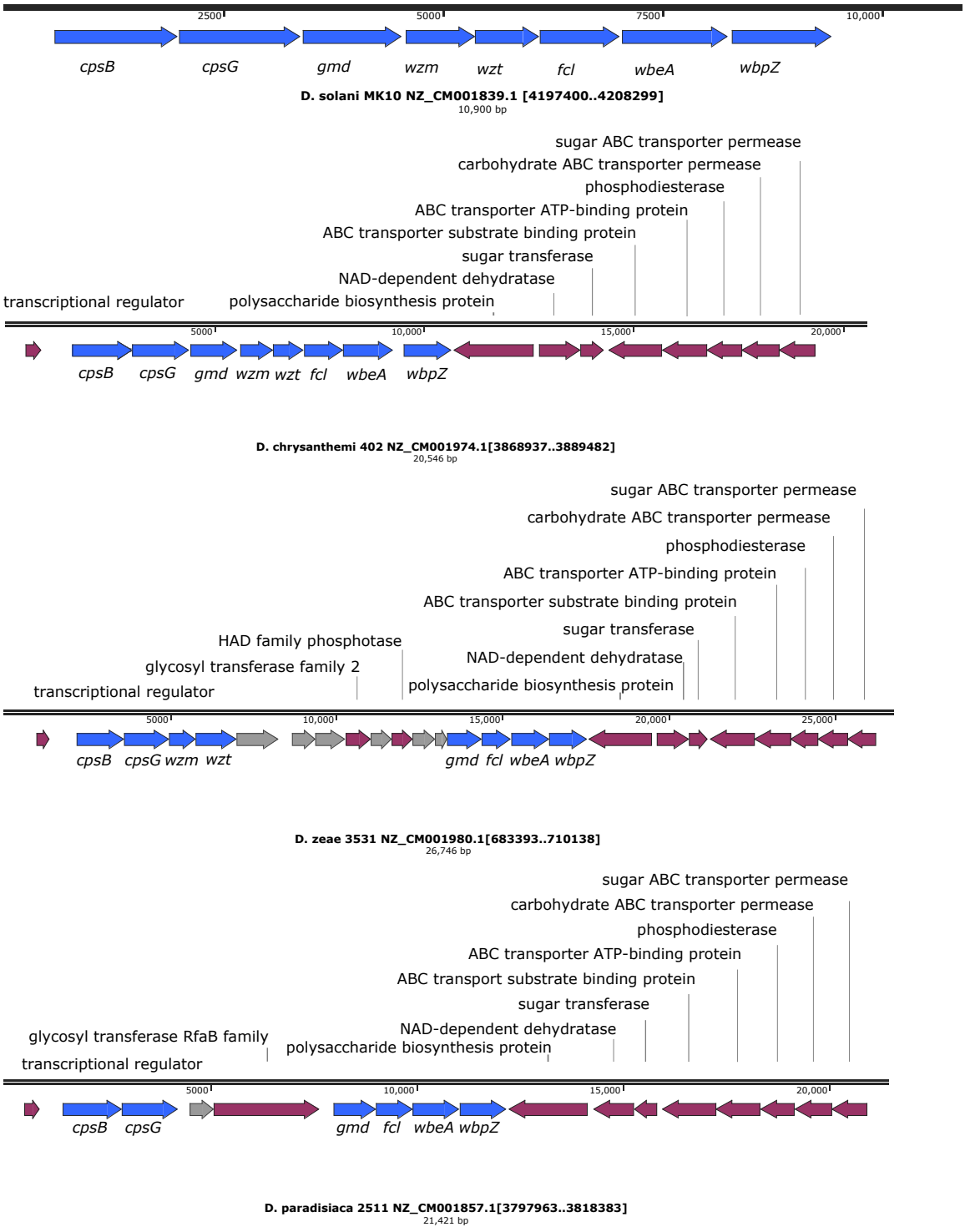


Fig. 4.10 Maps of the predicted cps clusters from four *Dickeya* species. Functional homologues based on the *D. solani* gene cluster are marked in blue. Hypothetical proteins are marked in grey and other annotated genes in burgundy.

4.3 Discussion

The phage-resistant mutants generated in this study, as well as work performed previously in this laboratory [10, 65], strongly suggest that the capsular polysaccharide of *D. solani* is responsible for recognition of the host cell by the *Ackermannviridae* family phages. Nearly 80% of the phage-resistant mutants generated in multiple rounds of transposon mutagenesis were found in the same gene cluster, predicted to encode the capsular polysaccharide synthesis (cps) pathway, with disruption to any of the eight genes resulting in resistance to the phage. This is in agreement with reports concerning *Ackermannviridae* family phages of *Klebsiella* and *Salmonella* species [77, 175] which showed that the capsule was necessary for infection.

Whilst the majority of phage-resistant mutants had transposons inserted in the predicted cps cluster, two had mutations in an LPS-related cluster disrupting genes encoding for the proteins RfaL and RfaB. These two proteins are involved in assembly and modification of the LPS, which is a known receptor for other phages, including the well-studied T4 phage [173]. It is therefore possible that LPS also interacts with the phage to facilitate infection. If this were the case however, it would be expected that mutants defective in these genes might appear more often, as deletions of *rfa* genes are not lethal in other bacteria [88]. It may be more likely that these genes are indirectly affecting the interactions between the receptor(s) and the phage. It is worth noting that these genes are on the opposite strand compared to the other surrounding LPS-related genes shown in Fig. 4.3, suggesting an inversion event, although whether this has significance is unknown. It is also worth highlighting, as shown in Table 4.1, that the final two genes in the cps cluster, *wbeA* and *wbpZ*, are also members of the RfaB family of glycosyltransferases, although it does not appear that they are functionally redundant as disruption of any one of the three genes leads to phage-resistance. Four other transposon mutants carried mutations that were found elsewhere in the genome, in intergenic locations. Whilst it is not certain which genes these transposons disrupt, bioinformatic prediction suggests a variety of genes largely involved in metabolism or in a prophage element. The reason behind these insertions causing phage-resistance is unclear, but are assumed to be an indirect effect. It should also be noted that these mutations were not transduced into a naive genetic background, and so it is possible that other undetected mutations in the genomes of these mutants are the true cause of the phage-resistance.

The data presented here establish the capsule of *D. solani* as the receptor for *Ackermannviridae* family phages. There is however a paucity of information regarding the capsule of *Pectobacteriaceae* family members. A recent publication concerning *Pectobacterium* investigates the exopolysaccharide of the bacteria, but leaves investigation of the capsular polysaccharide specifically open [73]. The majority of published experimental work on *D. solani* has concentrated on virulence factors such as plant cell wall degrading enzymes and

the underlying systems controlling virulence [39]. Therefore it seems likely that this work starts to provide the first insights into the capsule of this virulent phytopathogen.

The predicted cps cluster is a relatively simple assembly of eight genes, six synthetic and two transport, that mirrors well studied clusters such as the *Salmonella enterica* serovar Typhi ViaB cluster [175]. The functional homologues of several genes in the cluster are found in many other prokaryotic and eukaryotic species and have been identified as a GDP-L-fucose synthesis pathway [180]. Expression of cps genes appears to be a response to nutritional stress, either upon entry to stationary phase in rich media, or throughout growth in minimal media. This is in agreement with investigations of exopolysaccharide production in the related genus *Pectobacterium* [73]. These experiments were performed in liquid media, which may not be physiologically relevant for the bacteria, however, they suggest that expression of capsular genes is likely constitutive in the environment. With regards to application in phage therapy, this is advantageous as constant expression of the capsule would render the bacteria susceptible to phage infection throughout growth. In an exploratory investigation, the use of mannose as a sole carbon source reduced initial expression levels of the first gene in the cluster, suggesting that exogenous mannose may be incorporated into the capsular synthesis pathway and that it may be similar to the pathway proposed in other organisms.

When considering a phage for therapeutic use, it is considered advantageous for the receptor to be a virulence factor. This argument is predicated on the idea that bacteria are constantly evolving in the 'biological arms race' and as such there is a strong evolutionary pressure for the bacteria to mutate the receptor so that it is no longer susceptible to the phage being used. If this receptor is a virulence factor, mutation may have the added effect of reducing the virulence of the bacterial cell, thereby reducing crop losses in the case of *D. solani* [110]. Investigation of the capsule and its role in virulence showed that, in a potato tuber assay, there was no significant effect of capsule loss. It is however likely that the capsule is more important for adherence to the potato surface and prevention of desiccation [177], neither of which were tested in these experiments. Further experimentation involving whole plants in a more environmentally relevant setting would therefore be needed to ascertain the role of the capsule in virulence.

Comparison of the predicted cps clusters between six *Dickeya* strains present in this laboratory showed that all shared functional homologues with proteins in the *D. solani* cps cluster. However, some species had a more expansive set of genes that could cause differences in the capsular polysaccharide, as suggested by data from the previous chapter. The presence of genes with functional homologues in the GDP-L-fucose pathway is conserved across all six strains however, suggesting that this pathway, and therefore potentially GDP-L-fucose, is

an integral part of *Dickeya* capsular polysaccharide. Structural analysis of this polysaccharide would be an important future experiment to investigate this hypothesis, and could also shed light on the specific components of the capsule recognised by the phage.

Chapter Five

Ackermannviridae family phages of *Serratia* species

5.1 Introduction

The phage family *Ackermannviridae*, discussed in previous chapters in relation to *D. solani*, also contains phages capable of infecting many members of the recently defined Enterobacterales order including *Salmonella* [76], *Shigella* [14] and *Serratia* [109] species. *Serratia* species are members of the family Yersiniaceae and are found in both terrestrial and aquatic environments associated with animals and plants [80]. Previously thought to be non-pathogenic, it has been found that they can cause infections in immunocompromised individuals and are an increasing healthcare challenge due to intrinsic and acquired antibiotic resistance [149]. Much of the academic research into *Serratia* concerns secondary metabolites such as the pigment prodigiosin, which has been shown to have anti-cancer and antibacterial properties [87]. The pigmentation conferred by prodigiosin led to the use of *Serratia marcescens* as a tracer organism in a variety of tests including simulation of bio-weapon dispersal in Paris and San Francisco in the mid-twentieth century [104].

Work in this laboratory involves *Serratia* species, and analysis of two *Ackermannviridae* family members, 3M and MAM1, capable of infecting them. In the previous chapter a mechanism of interaction was proposed between *Ackermannviridae* family phages and their hosts. The availability of *Ackermannviridae* family phages capable of infecting a different host allowed for direct testing of this model to investigate if it were applicable to the family as a whole.

3M was isolated from river water in Spain over 25 years ago and found to be a generalised transducer [140], a characteristic of the family as described for *Dickeya* species [107], and

electron microscopy showed an *Ackermannviridae*-like morphology. MAM1 was isolated from river water in Cambridge, UK in 2011, and was also shown to be a transducer and possess the same morphology [106]. The genome of MAM1 has been published [105] and shows translated nucleotide homology with other *Ackermannviridae* family phages of around 50%. 3M had not been sequenced before this project began.

3M was isolated using *Serratia marcescens* strain 2170, but in this laboratory is maintained on the host *Serratia marcescens* 274 (Sma274), which was also found to be permissive during the original isolation [140]. The transduction capability of 3M has been previously used in this laboratory for horizontal gene transfer between Sma274 and another host *Serratia marcescens* 12 (Sma12) [45], and 3M has been found to infect another *S. marcescens*, MSU97, which was plant-associated [108]. 3M therefore has a host range within *S. marcescens* strains but was not found to infect other species when tested. MAM1 was originally isolated using another *Serratia* species, *S. plymuthica* A153 [105], a wheat rhizosphere isolate that produces many secondary metabolites [97]. Host range testing showed that MAM1 was also capable of forming plaques on Sma12 and Sma274, as well as Sma2170, the original host of 3M [106]. It was also shown that it could plaque on *Kluyvera cryocrescens* 2Kr27, a potato rhizosphere isolate [23], making it the first *Ackermannviridae* family member in this laboratory able to cross genera. It was also shown that MAM1 could facilitate transduction of plasmids across these genera at a frequency of $2.7\text{x}10^{-7}$ [106]. MAM1 therefore has the broadest host range of any *Ackermannviridae* family phage available in this laboratory. A summary of the host range of the two phages is shown in Table 5.1. This overlapping host range makes these phages an intriguing target for study, as it could suggest that either the phages target different receptors or that unknown host factors are responsible for the difference in host range.

Host	**MAM1**	**3M**
Serratia marcescens MSU97	-	+
Serratia marcescens 12	+	+
Serratia marcescens 274	+	+
Serratia marcescens 2170	+	+
Serratia plymuthica A153	+	-
Kluyvera cryocrescens 2Kr27	+	-

Table 5.1 Host range of MAM1 and 3M. + denotes observed or reported host range on these strains. - denotes strains which were not observed to be susceptible to infection.

The aim of this work was therefore to determine if the findings from experiments with *Dickeya* phages were also applicable to *Serratia*-infecting *Ackermannviridae* family phages, as well as investigating the similarities between 3M and MAM1.

5.2 Results

5.2.1 Phage-resistant mutants of MAM1 and 3M

In the previous chapter random transposon mutagenesis was used to show that the capsule is likely the key host factor determining the host range of *D. solani*-infecting *Ackermannviridae*, agreeing with published work on phages of the same family infecting *Salmonella* [175] and *Klebsiella* [77] species. A screen performed previously using MAM1 and A153 showed that phage-resistant mutants had transposons located in a cluster predicted to encode capsular polysaccharide synthesis (cps) genes [106]. This cluster is shown in Fig. 5.1a and is nearly four times larger than the cps cluster in *Dickeya* discussed in the previous chapter, at just under 40 kb.

As detailed by Whitfield [176], capsular polysaccharide, in *E. coli*, is split into four groups determined by serological, genetic and biochemical criteria. The major obvious genetic differences are within the export apparatus, with Groups 1 and 4 utilising Wza, Wzc and Wzx proteins, whereas Groups 2 and 3 export using Wzm and Wzt (otherwise known as KpsM and KpsT). This classification system has been used for other Enterobacterales, and it would appear that the cps cluster in the *Serratia* species shown in Fig. 5.1a is different from that of *Dickeya* species. The presence of Wza, Wzc and Wzx suggests that it is either a Group 1 or 4 cluster [106], in contrast to the *D. solani* cps cluster shown in the previous chapter, in which the presence of the *wzm* and *wzt* genes mark it as a Group 2 or 3 cluster. Whilst this differs, this is not a unique finding, as the cps cluster identified as the receptor for *Ackermannviridae* family phages in *Salmonella* is a Group 2 or 3 cluster [175] whereas the identified cluster in *Klebsiella* is a Group 1 or 4 cluster [77]. The type of capsule cluster that the host bacteria possesses therefore seems to have little obvious effect on recognition by this family of phages.

Whilst the host range of MAM1 and 3M overlaps, it is not the same, which questions the idea that the capsule is the sole target for phages of the *Ackermannviridae* family, as if this were true host range would not be expected to differ greatly. It is possible that host immunity factors could play a role in restricting the host range of the phages, if their genomes are different, or it could suggest that the capsule is not the receptor. Exploiting the overlapping host range, transposon mutagenesis of Sma12 and Sma274 was used to investigate the receptors of MAM1 and 3M individually. It was expected that transposons would disrupt similar regions of the genome as those found in *S. plymuthica* A153, as bioinformatic searching showed the presence of similar cps clusters in both Sma12 and Sma274. The genome of Sma274 is unpublished and fragmented, meaning that the annotated gene homologues for the predicted cps cluster were scattered across multiple contigs, hindering analysis. The genome sequence

for Sma12 was kindly provided by Sarah Coulthurst and the cluster for Sma12 assembled in one contig and is shown in Fig. 5.1b.

Random mutagenesis of Sma274 using MAM1 as the phage was performed to see if the capsular genes were disrupted at similar frequencies as was found in *D. solani* and to see if the phage could identify the parts of the cps cluster scattered across the Sma274 contigs. This method has been described in the previous chapter, but in brief involved conjugation of an DAPA-auxotrophic donor strain of *E. coli* containing a suicide plasposon with the recipient *Serratia*. After conjugation for two days, the cells were plated in a semi-solid agar lawn containing kanamycin to select for the transposon, the phage to select for phage-resistance, and the absence of DAPA to counter select the donor *E. coli*. Any colonies that grew were streaked on antibiotic media and tested for phage resistance to confirm the phenotype. Determination of the insertion sites used random primed PCR, coupling a transposon-specific primer and a series of random primers over two rounds of PCR.

138 mutants were generated and 13 were sequenced by random primed PCR. Some of these sequences proved to be siblings; they came from the same conjugation and had the same insertion site. A summary of the eight unique insertion sites hit in this mutagenesis screen can be seen in Table 5.2. The main finding was that none of the genes disrupted were located in genes obviously linked to capsular polysaccharide production. Four of the eight insertion sites were intergenic, but there is no obvious link between any of the surrounding genes and the capsule. The genes disrupted include a response regulator, a transcriptional regulator and an indole-3-pyruvate decarboxylase. These are all involved in signalling or regulation, therefore it could be hypothesised that they are part of a network that modulates expression of the cps genes in Sma274. However, if the capsule were the receptor for MAM1, it would be expected to have at least one mutant in a functional homologue of a gene in the previously identified cluster, considering the high frequency of obtaining these mutants in *D. solani*. It is possible, as discussed in the previous chapter, that the selection method used to generate these mutants, which selects for both antibiotic and phage-resistance simultaneously, places a large selective burden on the bacteria and may bias the results. Future work would benefit from utilising a two-step screening process to remove this bias. The data presented here however suggest that the receptor recognised by MAM1 on Sma274 may not be the capsule, contrary to findings with the same phage and the host *S. plymuthica* A153.

Whilst these data suggest that the capsule may not be the receptor for MAM1 in Sma274, they do not rule it out. It is possible that, considering that many of the genes disrupted are regulatory in nature, that only genes which modulate capsular polysaccharide expression were affected. The same experiment with this phage in the original host *S. plymuthica* A153 generated multiple mutants with transposons located in a hypothesised cps cluster,

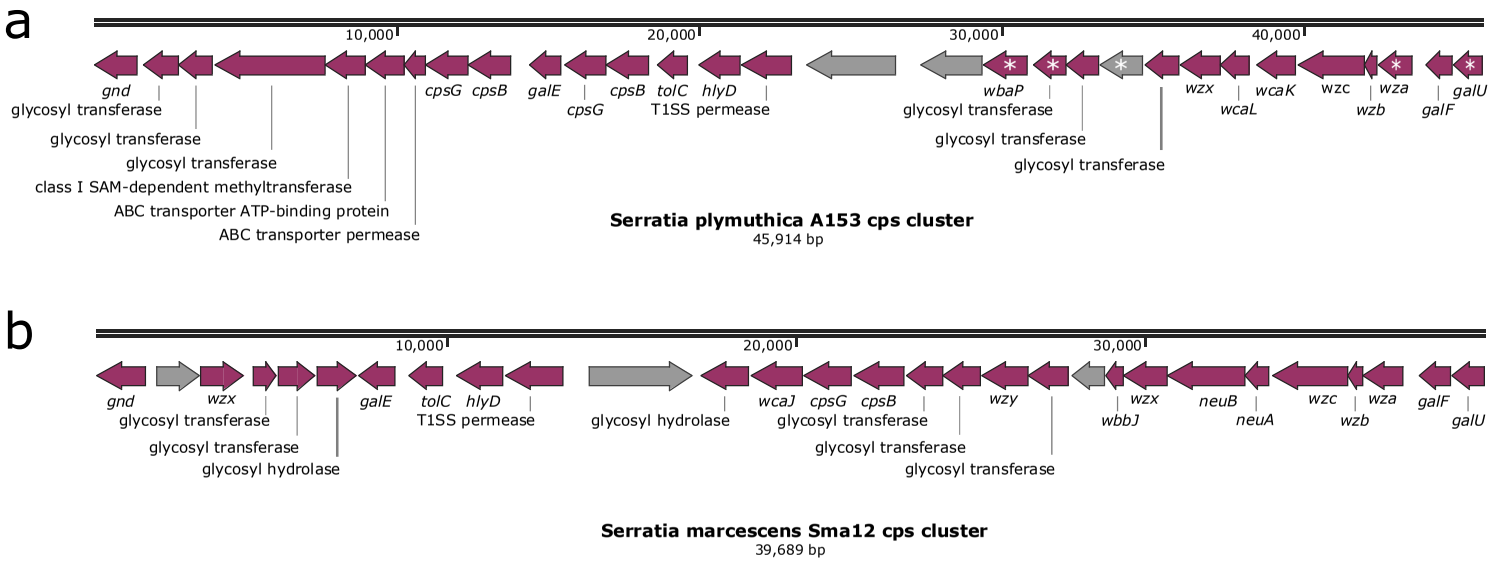


Fig. 5.1 Proposed cps clusters of *Serratia plymuthica* A153 and *Serratia marcescens* Sma12. Hypothetical proteins are marked in grey.
a) shows gene cluster disrupted in *Serratia plymuthica* A153 resulting in resistance to the *Ackermannviridae* family phage MAM1 with * denoting genes disrupted [106].
b) shows the predicted cps cluster in Sma12.

Serratia marcescens Sma274

Insertion	**Gene disrupted**	**Likely role of disrupted gene**
728371	*ipdC* Sma274_00728	PDC1 family indole-3-pyruvate decarboxylase
1290891	Intergenic between hypothetical proteins Sma274_01249 and 01250	Unknown
1599158	Upstream of tRNA binding CsaA-like protein YgjH Sma274_01513	Unknown
2487550	Downstream of phage late control gene D Sma274_02318	Unknown
2569550	*rssB*	CheY-like response regulator
4293300	Sma274_03970	Pectate lyase superfamily protein
4594484	Sma274_04267	LysR family HTH-type transcriptional regulator
4641697	Intergenic between 5‘ nucleotidase and flavin mononucleotide-dependent oxidoreductase Sma274_04220 and 04221	Unknown

Serratia marcescens Sma12

Insertion	**Gene disrupted**	**Likely role of disrupted gene**
327504	Sma12_02930	DUF1471 domain-containing protein
2348632	*wcaA* Sma12_22160	Cell wall biosynthesis family 2 glycosyl transferase

Table 5.2 Transposon insertion sites in Sma274 and Sma12 mutants resistant to MAM1

therefore the results in Sma274 may be anomalous. To investigate whether these findings are specific to the host Sma274, or whether this is a broader pattern, MAM1-resistant mutants of another host, Sma12, were generated via the same method. 14 mutants were generated, with three sequenced by random prime PCR. This screen only generated mutants with insertions at two locations, as shown in Table 5.2. One of these disrupted a gene encoding a hypothetical DUF1471 domain-containing protein located between the Kef glutathione-regulated potassium-efflux system gene and a dihydrofolate reductase gene, but in different operons. The function of this gene is therefore unclear but again appears tied to metabolism and regulation. The other insertion site was in a gene encoding for a glycosyl transferase with homology to WcaA proteins, which are a component of the Group 1/4 colanic acid cps synthesis cluster in *E. coli* [176]. In the Sma12 genome this gene is apparently orphaned, as shown in Fig. 5.2, lying between two genes related to biosynthesis of purines and tryptophan respectively; *purT*, a phosphoribosylglycinamide formyltransferase, and *trpE*, an anthranilate/para-aminobenzoate synthase. There is no WcaA homologue in either of the predicted *Serratia* cps clusters shown in Fig. 5.1 and so, whilst it is possible that it acts as part of the capsular synthesis pathway, its role is unclear.

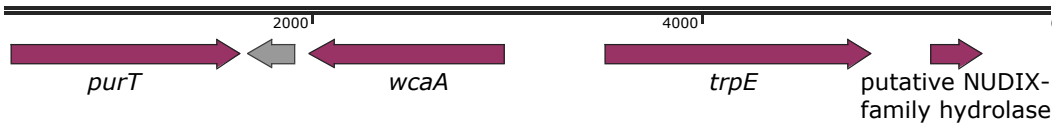


Fig. 5.2 Map of the region containing the *wcaA* gene disrupted by a transposon in a phage-resistant mutant of Sma12. Two genes either side of the disrupted gene are shown with their annotations, with the grey open reading frame being hypothetical. *purT* encodes a phosphoribosylglycinamide formyltransferase, which is involved in the purine biosynthesis pathway. *trpE* encodes an anthranilate/para-aminobenzoate synthase, which plays a role in both the tryptophan and folate biosynthetic pathways.

The previous data suggest that MAM1 may not utilise the capsular polysaccharide of two *Serratia marcescens* strains Sma274 and Sma12 to facilitate infection, as no cps-related genes were hit in either mutagenesis. This is a novel finding that does not agree with previous experimentation with *Ackermannviridae* family phages in three other genera of hosts, as well as the same phage in another species of *Serratia*. The phage 3M is also capable of infecting the strains Sma12 and Sma274. Therefore to investigate whether the observations in MAM1 are unique, or whether this is characteristic of *Serratia*-infecting *Ackermannviridae* family phages, random transposon mutagenesis was also performed with both hosts and 3M. 121 mutants of Sma12 were generated, with 20 sequenced, and 26 mutants of Sma274, with 18 sequenced. The genes disrupted in these experiments are shown in Table 5.3. Again no genes were disrupted that appear obviously linked to capsular polysaccharide synthesis. Those that

Serratia marcescens Sma274

Insertion	**Gene disrupted**	**Likely role of disrupted gene**
306716	Sma274_00300	Diguanylate cyclase
1063143	*viuB* Sma274_01031	NADPH-dependent ferric siderophore reductase
1090604	*yafV* Sma274_01055	Amidohydrolase
1655527	Upstream of *rssB* Sma274_01564	CheY-like response regulator
2807925	Sma274_02618	Hypothetical protein
2811778	Sma274_02620	ImpA domain type VI secretion-associated protein
3082190	*btuB* Sma274_02859	Vitamin B12/cobalamin outer membrane transporter
3279822	Sma274_03048	LysR family HTH-type transcriptional regulator
3714432	Upstream of *livH* Sma274_03535	Branched-chain amino acid transporter permease
3885303	Sma274_03592	Prophage tail fibre N-terminal
4186048	Sma274_03877	Molybdopterin oxidoreductase

Serratia marcescens Sma12

Insertion	**Gene disrupted**	**Likely role of disrupted gene**
1830382	Sma12_17260	FadH2 NADPH-dependent 2,4-dienoyl-CoA reductase
2842535	Sma12_26830	TonB-dependent outer membrane receptor
3222311	Sma12_30250	Putative haemagglutinin/ haemolysin
4334340	*ulaG* Sma12_41390	L-ascorbate-6-phosphate lactamase

Table 5.3 Transposon insertion sites in Sma274 and Sma12 mutants resistant to 3M

were disrupted included a variety of metabolic and regulatory genes, including a LysR family transcriptional regulator and RssB, which had also been found in the screen with MAM1. Other genes disrupted included prophage-like genes and a putative haemagglutinin, but no capsular polysaccharide biosynthetic genes. These data appear to suggest that, like MAM1, 3M does not utilise capsular polysaccharide as a receptor in *Serratia marcescens*.

5.2.2 Genome of the phage 3M

The genome of MAM1 has been published [105] and exhibits translated nucleotide identity to other *Ackermannviridae* family phages of around 50%. 3M does not have a published genome sequence despite being isolated much earlier [140]. In order to determine the similarity between MAM1 and 3M, and to investigate the reason for their overlapping host range, 3M was genomically sequenced. The genome of 3M is 159,398 bp in length with a GC content of 51.4% and encodes 201 open readings frames (ORFs) and 2 tRNAs, all of which are characteristic of other members of the family *Ackermannviridae*. A map of the genome can be seen in Fig. 5.3 and a breakdown of the ORFs can be found in Appendix A.1.

The 3M genome shares 86% nucleotide identity with MAM1, and only significantly differs in seven genes with annotation and twelve that are hypothetical. The annotated genes are listed in Table 5.4 and highlighted in Fig. 5.3, with four of them predicted to encode the tail spike proteins of the two phages. The ssDNA binding protein occurs in both phages at the same genomic location but varies slightly in sequence, whilst the two homing endonucleases listed in Table 5.4 occur at different locations and share no identity. The alpha hydrolase is encoded by one of four genes present in the 3M genome that do not have homologues in the MAM1 genome, with the other three genes having no annotation. It is unlikely that the differences observed in the homing endonucleases and ssDNA binding proteins impact significantly on the host range of the phage.

MAM1 and 3M both possess four tail spike proteins; the major source of genome differences between the two phages. Amino acid identity between the four proteins is listed in Table 5.5. TSP1 of the two phages share 91% identity, with TSP2 sharing 64%. However, TSP3 of 3M shares less than 40% identity with TSP4 of MAM1, and 3M TSP4 has no real identity with any MAM1 TSP.

A translated nucleotide comparison of the TSPs is shown in Fig. 5.4. In agreement with the amino acid identity shown in Table 5.5, this shows that there is broad conservation of TSP1 and TSP2 between the two phages and that the identity between 3M TSP3 and MAM1 TSP4 is contained within the N-termini of the two proteins. No structural data exists for these proteins, but threading modelling with the I-TASSER suite was used to provide some insight. The models for 3M TSP3 and MAM1 TSP4 are shown in Fig. 5.5 and are

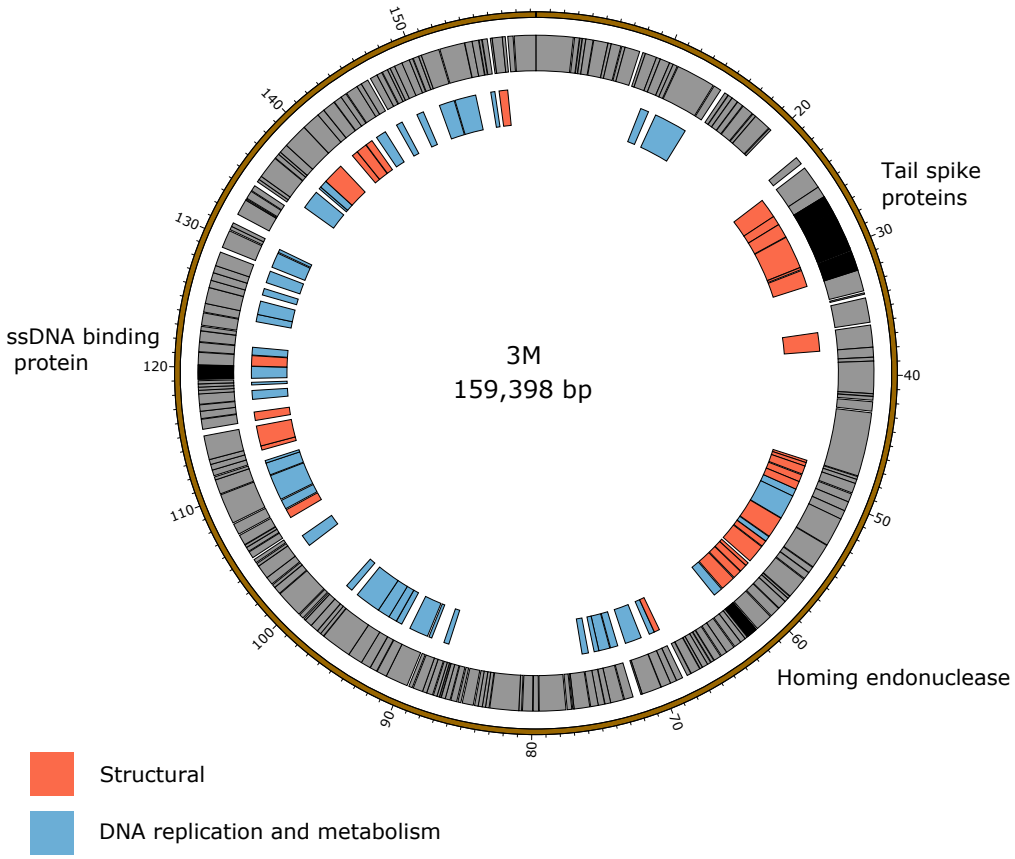


Fig. 5.3 Map of the 3M genome. The outer grey ring marks open readings frames with those highlighted in black discussed in more detail in the text. The inner ring categorises the open reading frames based on their proposed annotations.

clearly very different. The N-termini of the two models are located at the top of Fig. 5.5 and likely interact with the phage baseplate, which could explain the conservation. The closest structural homologue of 3M TSP3 is the only *Ackermannviridae* family TSP that has been structurally characterised: TSP1 from the phage CBA120, which infects *E. coli* [40]. Whilst the sequence identity between these proteins is only 18%, the structural model has over 96% coverage. As is visible in Fig. 5.5, MAM1 TSP4 is different, with the closest structural homologue being the S-layer protein RsaA from the Alphaproteobacterium *Caulobacter crescentus* [24]. S-layers are proteinaceous arrays frequently found on the surface of bacteria and archaea [64]. Whilst the predicted structure of MAM1 TSP4 shares overall coverage of nearly 99% with RsaA, the relevance of this homology is unclear.

MAM1 gene	3M gene	Annotated function
MAM1_034	3M_032	Tail spike protein
MAM1_035	3M_033	Tail spike protein
MAM1_036	3M_034	Tail spike protein
MAM1_037	3M_035	Tail spike protein
MAM1_053	-	Homing endonuclease
-	3M_063	Homing endonuclease
MAM1_142	3M_149	ssDNA binding protein
MAM1_175	-	Alpha hydrolase

Table 5.4 Annotated genes which differ between the phages MAM1 and 3M

%	MAM1 TSP1	MAM1 TSP2	MAM1 TSP3	MAM1 TSP4
3M TSP1	91.0	2.1	5.3	7.4
3M TSP2	13.4	64.0	13.4	9.4
3M TSP3	5.3	0.08	5.3	39.5
3M TSP4	5.6	2.0	5.6	13.2

Table 5.5 Percentage amino acid identity between tail spike proteins of MAM1 and 3M

It is therefore possible that the overlapping host range of these two phages is the result of the shared identity of some, but not all, TSPs between the two phages. Host recognition of the *Serratia* species Sma12 and Sma274 could be facilitated by TSP1 and TSP2, whereas recognition of the other species could be dependent on TSP3 and TSP4. This is merely a hypothesis however, and further experimental work is required to test this model.

5.2.3 MAM1, 3M and other *Ackermannviridae*

Phage phylogenetic trees, whilst not as easy as those using bacterial 16S sequences, utilise conserved genes such as the terminase large subunit and major capsid protein. Due to the (comparatively) low level of similarity between MAM1/3M and the other *Ackermannviridae* family members, as well as the unexpected results of the phage-resistant mutant screens, it would seem that these two phages are quite different from the rest. A phylogenetic tree for the major capsid protein can be seen in 5.6a and for the terminase in Fig. 5.6b. In the proposal for the novel family *Ackermannviridae*, two sub-families and four genera were identified [91]. The four genera; *Ag3virus*, *Limestonevirus*, *Cba120virus* and *ViIvirus*, are included in both figures, with the *Limestonevirus* genus containing known phages of the family that infect *Dickeya*. The previously discussed phages MAM1 and 0507-KN2-1, as

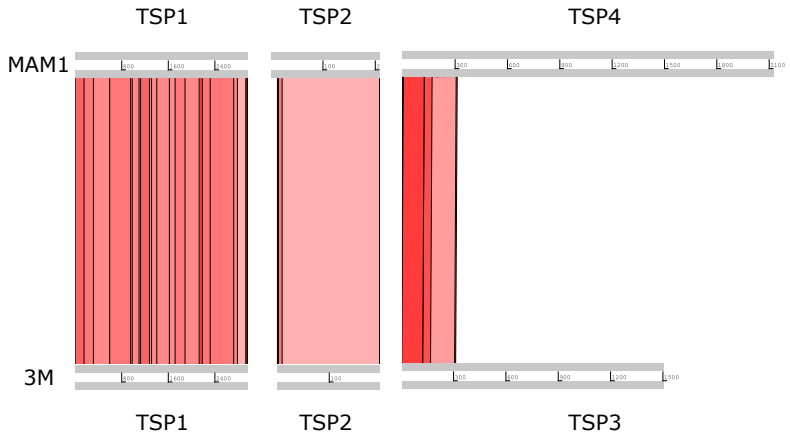


Fig. 5.4 Translated nucleotide comparison of the TSPs of MAM1 and 3M. Red bars mark areas of translated nucleotide identity with darker colours indicating higher identity. MAM1 TSP3 and 3M TSP4 share no significant identity.

well as the *Serratia rubidea* phage IME250 [181] and *Erwinia amylovora* phage Ea2809 [93] were recognised as members of the *Ackermannviridae* family when it was proposed [91], but were not placed within genera. It was possible to include 3M in these trees, along with two other *Serratia marcescens* phages 2050H1 (Genbank reference MF285619.1) and KSP90 [109]. The four *Serratia* phages cluster together in the Fig. 5.6a. The genome of 2050H1 shares over 95% nucleotide identity with MAM1, and over 87% identity with 3M. KSP90 has not been fully sequenced, but it is likely to also share high identity. The absence of sequence data prevents KSP90 from being included in Fig. 5.6b, but the same clustering is observed for the other three phages.

5.3 Discussion

Whilst it was expected that *Ackermannviridae* family phages infecting *Serratia* species would behave similarly to those infecting *Dickeya* species, the data presented here show that this is not the case. Transposon mutagenesis has not produced phage-resistant mutants with insertions in the predicted capsule cluster. This could be interpreted to suggest that the *Serratia* phages MAM1 and 3M do not utilise the capsule of their hosts for recognition of two *Serratia marcescens* strains. However, the transposon mutagenesis gives no clear picture as to the true receptor(s) for these phages, and the absence of a cps mutant does not prove that the capsule is not responsible. This also conflicts with previous work using MAM1 with another species of *Serratia*, *S. plymuthica*, which suggested that capsular polysaccharide was the receptor. The genes disrupted in the phage-resistant mutants of two

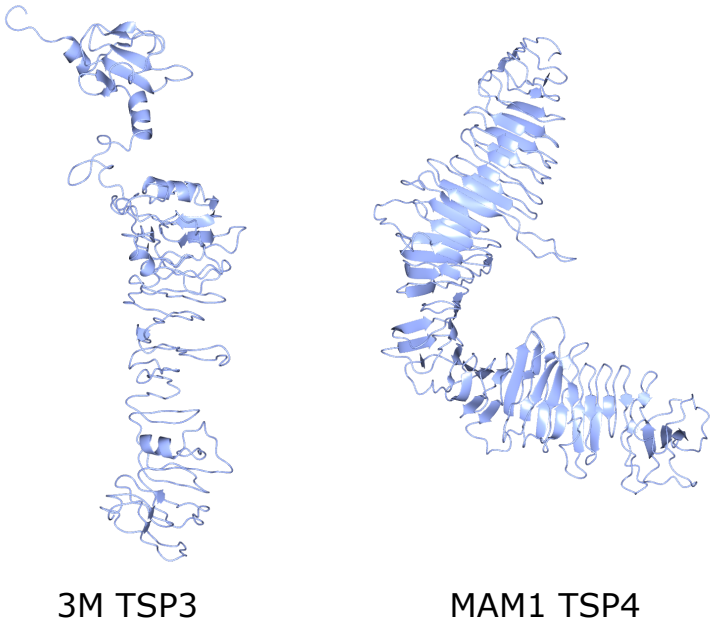


Fig. 5.5 Threading models of 3M TSP3 and MAM1 TSP4 amino acid sequences using I-TASSER. These two proteins share structural homology at the N-terminal only, which is located at the top of these models. The model for 3M TSP3 has a C-score of -1.71, an estimated TM-score of 0.51±0.15 and an estimated RMSD of 11.4±4.5. The model for MAM1 TSP4 has a C-score of -0.43, an estimated TM-score of 0.66±0.13 and an estimated RMSD of 9.1±4.6.

Serratia marcescens strains included multiple regulators, which could suggest that direct disruption of the genes responsible for the receptor renders the cells non-viable, and therefore the way to generate phage-resistance in this screen is by indirect interference with a regulator of capsular expression. This mutagenesis screen also used a selection step in which presence of the transposon and resistance to the phage were selected for simultaneously. As discussed in the previous chapter, whilst this method has been shown to produce results in multiple bacterial species [117], it is possible that it biases the screen towards insertion sites in which the transposon is more easily able to integrate. Selection for phage-resistance against a library of transposon mutants may therefore yield different results. Targeted mutagenesis and disruption of capsule cluster genes could also be used to determine the role of the capsule as a receptor.

The predicted cps clusters in Fig. 5.1 both contain a T1SS gene cluster consisting of three genes; a T1SS permease, *hlyD* and *tolC*. It is therefore possible that disruption of some parts of the cps cluster could impact this secretion system, which could be lethal or hinder growth. All of the phage-resistant mutants in *S. plymuthica* had insertions several kilobases

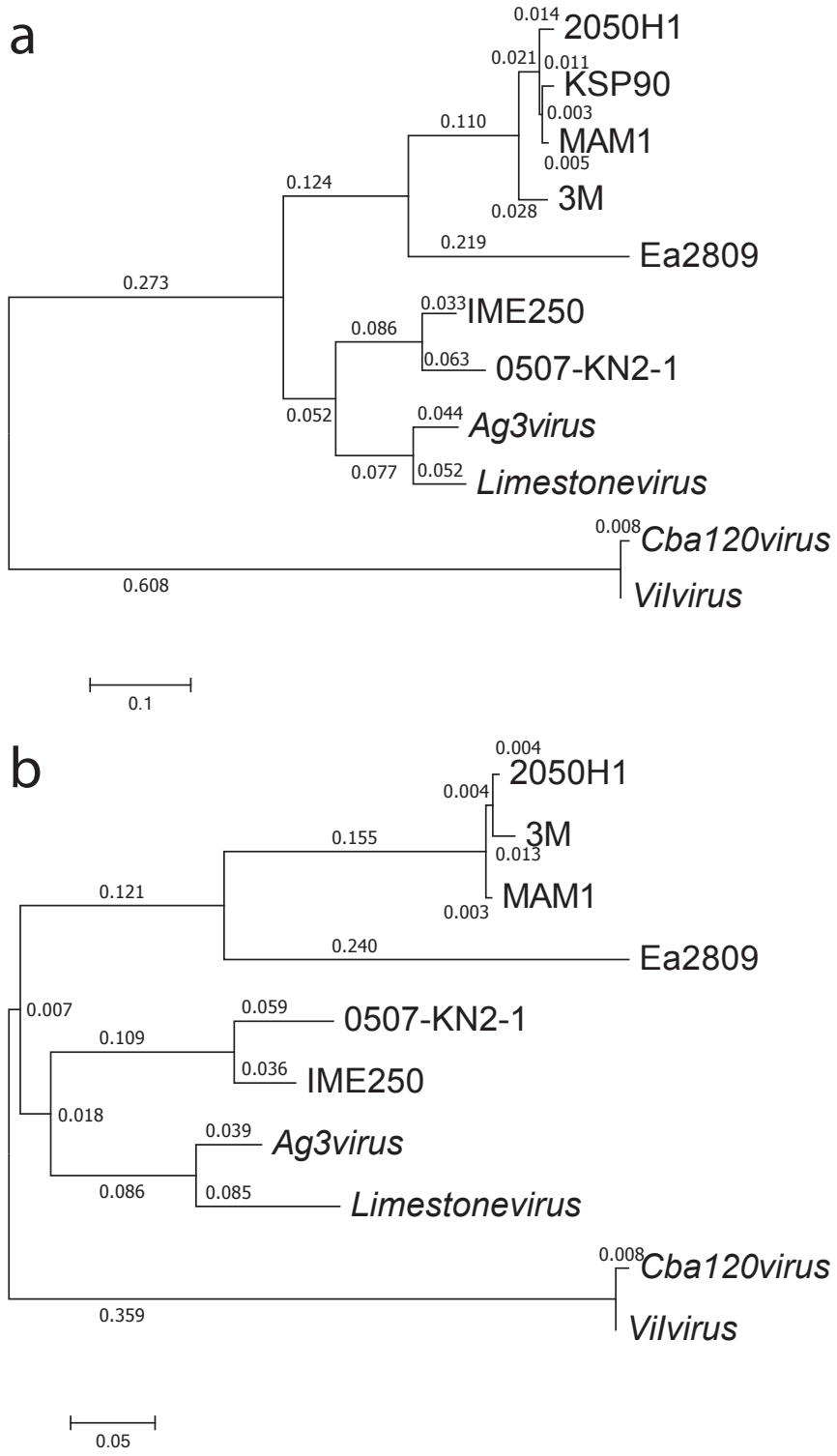


Fig. 5.6 Phylogeny of *Ackermannviridae* family a) major capsid protein and b) terminase using the Maximum Likelihood method. The trees with the highest log likelihood (-3647.71 and -4972.24) are shown. The trees are drawn to scale, with branch lengths measured in the number of substitutions per site. All positions containing gaps and missing data were eliminated. There were a total of 370 and 532 positions respectively in the final datasets. The genera *Ag3virus*, *Limestonevirus*, *Cba120virus* and *ViIvirus* include multiple phages but have been condensed for readability, with the sequence of the corresponding protein in the type phage (AG3, LIMEstone1, CBA120 and ViI) for each genus used to calculate the tree.

upstream of the T1SS cluster, as shown in Fig. 5.1a, which could support this hypothesis. This still does not explain however the absence of similar mutants in Sma12. The only major functional difference between the two is the presence of the *neuAB* genes in the Sma12 cluster, which are involved in sialic acid biosynthesis. It is not unusual for these genes to be located in a cps cluster however and their disruption is unlikely to be lethal [74]. The true nature of the *Ackermannviridae* family phage receptor(s) in *S. marcescens* therefore remains unknown.

MAM1 and 3M have an overlapping host range across *Serratia* species, and this is reflected in their tail spike proteins, two of which share high identity, whilst the other two share little to no identity. Paired with the inconclusive data regarding the receptor(s) for these phages, it would seem to suggest that these phages are different from members of the same family discussed in previous chapters. This is confirmed when comparing the genomes of these phages. The genome of 3M shares high overall identity with that of MAM1 and other *Serratia*-infecting *Ackermannviridae* and they form a clade quite distinct from other members of the family. This leads to the tentative suggestion that these phages inhabit a separate sub-family and genus. This is despite isolation dates ranging over thirty years and including locations as far apart as Japan and Spain. The phages 2050H1 and MAM1 are even likely members of the same species despite being isolated over 5000 miles apart. It would therefore be interesting to investigate whether the data presented here are replicated in these *Serratia*-infecting phages with their cognate hosts, or whether they are unique to the *Serratia marcescens* strains tested. The overlapping host range of the phages also requires further investigation, as this may be due to host factors, but could be a result of the variation in tail spike proteins. Adsorption experiments may prove illuminating in this regard.

Based on criteria used by Kuhn *et al.* to demarcate genera and sub-families in the *Ackermannviridae* [91], this would place MAM1, 3M and 2050H1 in a separate genus, which, based on precedent, would be titled *Mam1virus*. Kuhn *et al.* also propose a species demarcation of 95% nucleotide identity, which would class MAM1 and 2050H1 as the same species. The proposed novel genus *Mam1virus* would be distinct from the other genera (and sub-families) within the *Ackermannviridae* family. Whilst establishing the family, Kuhn *et al.* showed that the other genera shared at least 4% nucleotide identity and 52% translated nucleotide identity, whereas MAM1 shared, at best 1% nucleotide identity and less than 40% translated nucleotide identity with phages of other genera. The proposed *Mam1virus* genus would therefore be a novel genus of the *Ackermannviridae* family currently comprised of phages isolated on *Serratia* species.

Chapter Six

Wider diversity of *Dickeya solani* phages

6.1 Introduction

The majority of phages discussed so far are members of the recently defined family *Ackermannviridae* (formerly the genus *Vi1virus*) [9]. The International Committee for the Taxonomy of Viruses (ICTV) (as of April 2018) recognises 23 members of this family including the *Dickeya* phages LIMEstone1 [8] and RC2014 [49] (referred to in the literature, and henceforth, as D5, but published in Genbank as RC2014). The other *D. solani* phages listed in Table 6.1 could also be assigned to this family, as they share morphology and 99% nucleotide identity with LIMEstone1.

Bacteriophage	Isolation	Location	Genome size (bp)	Reference
LIMEstone1	2008	Belgium (soil)	152247	[8]
D3	2013	Poland (soil)	152308	[51]
D5	2012	Poland (soil)	155346	[49]
PD10.3	2013	Poland (soil)	156113*	[50]
PD23.1	2013	Poland (soil)	153365*	[50]
XF4	2013	UK (waterway)	151519	[54]
XF11	2013	UK (waterway)	153309	[65]
XF16	2013	UK (waterway)	154083	[65]
JA15	2014	UK (waterway)	153757	[54]

Table 6.1 Members of the *Ackermannviridae* family isolated on *D. solani*. * genomes are marked incomplete, largest scaffold is reported and shows 99% match to LIMEstone1

The majority of the *Ackermannviridae* family members isolated in either Belgium or the UK (LIMEstone, XF, FX and JA phages) are only capable of forming plaques on *D. solani* and not isolates of other genera or *Dickeya* species [54]. However, it was found that thirteen

of the ninety phages in the lab were capable of forming plaques on *Dickeya* species other than *D. solani*. These are also phytopathogens, and so these phages with a broader host-range may be more useful as biocontrol agents, as well as for understanding phage-host interactions. Some of these phages were therefore characterised phenotypically, morphologically and genomically.

6.2 Results

6.2.1 Phenotypic characteristics of phages with a wider host range

Host range

There are ninety phages in this laboratory (XF1-28, FX1-23 and JA1-39) isolated on *D. solani* which have been tested against, and shown not to form plaques on, a wide range of hosts including *Pectobacterium* and *Serratia* species, as listed in Table 6.2. Previous work in this laboratory [10] also showed that these phages were unable to infect representatives of more distant genera such as *Pantoea*, *Escherichia* and *Pseudomonas*. Eight of the JA phages and five of the XF phages however were able to form plaques on other species of *Dickeya* as shown in Table 6.3. These phages therefore warranted further investigation to determine the reason for this broader host range. Unfortunately the five XF phages could no longer be propagated for use in experimentation, therefore they were not further analysed.

***Dickeya* species**	***Pectobacterium* species**	***Serratia* species**
D. species MK7	*P. atrosepticum* SCRI collection	*S. marcescens* Sma12
D. dianthicola 3534	*P. carotovorum* SCRI collection	*S. marcescens* Sma274
D. dianthicola IPO980		
D. dadantii Ech703		
D. species CSL RW240		

Table 6.2 Bacterial strains unable to be infected by *D. solani* phages

6.2.2 Morphology

Classification of phage into taxonomic families is traditionally performed using electron microscopy [2]. To investigate whether the eight broader host range phages were members of the *Ackermannviridae* family, in keeping with the other imaged *Dickeya* phages, or whether they were something else, as their host range suggested, all eight were imaged. The images

***Dickeya* species**	XF24, 25, 26 and JA10	XF27 and 28	JA11, 31, 32, 33 and 37	JA13	JA29
D. dadantii subsp. dieffenbachiae	+	-	+	+	+
D. paradisiaca	-	-	+	+	+
D. dianthicola	+	+	+	-	-
D. zeae	-	-	+	+	-
D. chrysanthemi	+	+	-	-	-

Table 6.3 Broader host range of eight phages capable of infecting other species of *Dickeya*. + denotes isolated plaque formation of the phages on the respective host. - denotes the absence of individual plaque formation.

are shown in Figure 6.1 along with an image of the *Ackermannviridae* family member XF4 for comparison.

Unexpectedly, as can be seen in Fig. 6.1b, the phage JA10 turned out to be a member of the *Podoviridae* family when imaged, characterised by an icosahedral head and a short non-contractile tail. This is not the first member of this family that has been isolated in this laboratory, as XF24-28 were also shown to be podoviruses [65], and XF24-26 exhibit the same host range as JA10, as seen in Table 6.3, likely making them similar. Unfortunately, no viable viral particles could be recovered from any of the XF lysates, preventing further comparison.

The other seven phages look morphologically similar; possessing an icosahedral head and long contractile tails. The structures at the base of the tail, which morphologically distinguish members of the *Myoviridae* family (tail fibres) from the *Ackermannviridae* family (tail spikes), could not be determined. The head diameter was around 120 nm with a tail length in the region of 150 nm. This is significantly larger than the defined dimensions of members of the *Ackermannviridae* family that have a head of 90 nm and a tail around 110 nm [6], which can be seen by comparing these seven phages to Fig. 6.1a. This suggested that these phages were perhaps members of the *Myoviridae* family.

The indistinct morphology of the tail appendages of the JA jumbo phages (best seen in Fig. 3.1c) has also been identified in other phages. When first described in the *Escherichia* phage 121Q [3] this morphology was presumed to be an artefact of microscopy involving damage to the tail. It was also thought that the dimensions, at the time reported to be a head diameter of 150 nm and a tail length of 165 nm, were overstated. However, this morphology has since been directly reported in the *Pseudomonas putida* phage Lu11 [7], the *Pectobacterium*

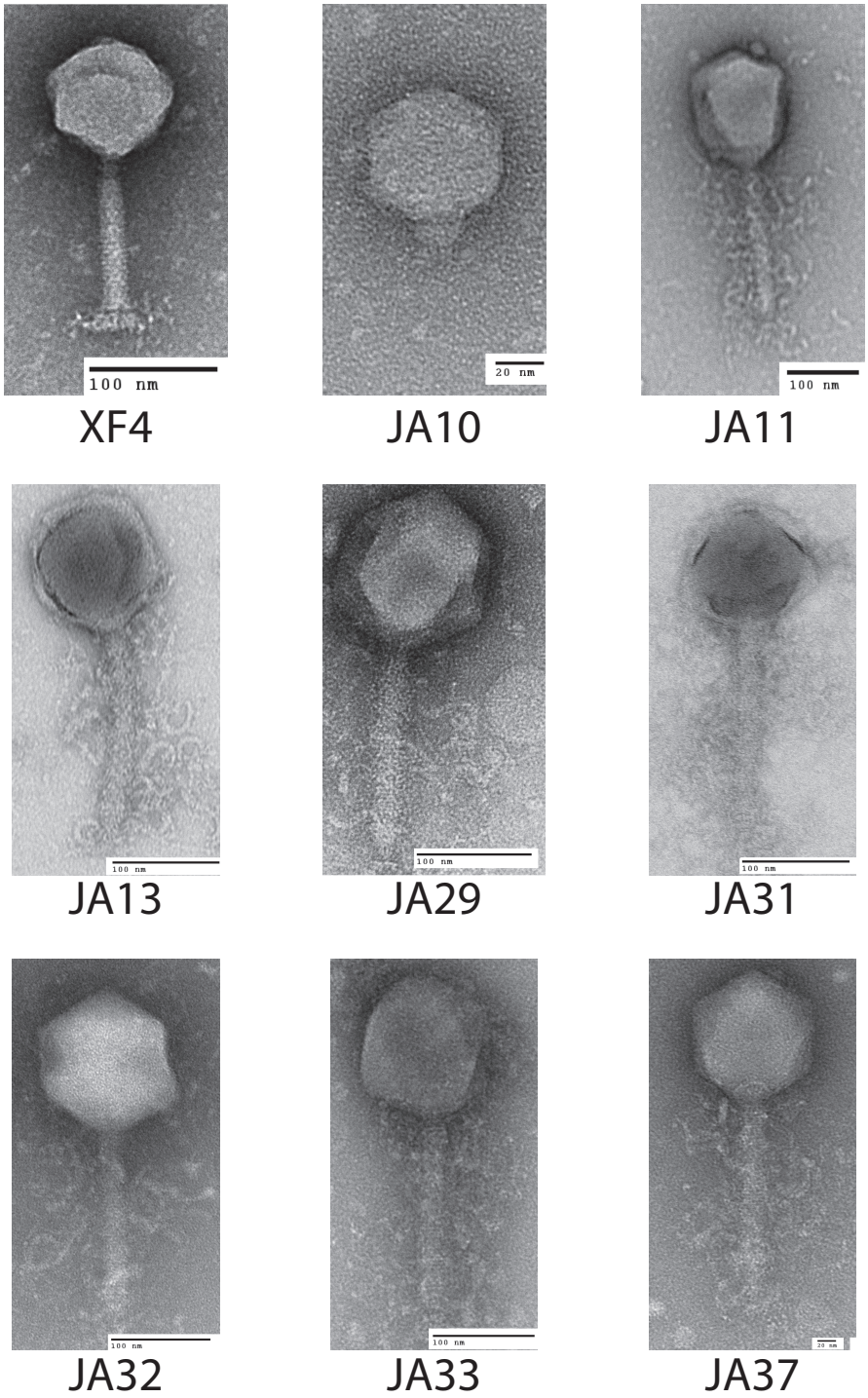


Fig. 6.1 Electron micrographs of XF4, a previously described member of the *Ackermannviridae* family along with eight *D. solani* phages with a broader host range. The image of XF4 was provided by Jiyoon Ahn [10].

carotovorum phage CBB [31] and the *Erwinia amylovora* phage Y3 [32], and has been dubbed the 'hairy *Myoviridae*' morphology [31].

Transduction

Previous work in this laboratory [10] showed that all of these JA phages except JA10 and JA11 were able to effect transduction of chromosomal markers between *D. solani* cells at a frequency of between $1x10^{-5}$ and $3x10^{-4}$ transductants per plaque forming unit. It was also shown that JA13, 29 and 37 could transfer plasmid markers between *Dickeya* species at frequencies ranging from $5x10^{-9}$ to $4x10^{-4}$ depending on the host. It is surprising that JA11 was found to be incapable of facilitating transduction, due to the phenotypic similarity with the other JA phages, and especially considering the genomic data presented later in this chapter. Due to this observation, both JA10 and JA11 were retested for their ability to facilitate transduction. In these experiments JA11 proved capable of effecting transfer at frequencies similar to the other JA phages, but JA10 showed no capacity for transduction.

6.2.3 Genomics of broader host range phages

The reason for the interest in the broader host range phages is due to the phenotypic host range data, but to gain a better understanding of the interactions between the phages and their hosts, genomic data is invaluable and increasingly available. Based on the phenotypic and morphological characterisation of the eight phages so far, it would appear that five of them (JA11, JA31, JA32, JA33 and JA37) are highly similar. Table 6.4 shows the isolation dates of the eight phages, and shows that of these five phages, three (JA31, JA32 and JA33) came from the same sample. It is therefore possible that they are siblings, meaning that they are essentially the same phage. JA10 and JA11 however also came from the same enrichment, and they are clearly not the same phage. In the interests of getting the largest dataset possible, all eight of the phages were sent for full genome sequencing.

Podoviridae JA10

The genome of the podovirus JA10 is shown in Fig. 6.2. It is 40,131 bp in length and has 50 predicted genes, the annotations of which are listed in Appendix A.3. The closest match in the database is an as yet unpublished *D. solani* phage Ninurta (Genbank reference: MH059639) isolated from organic waste in Denmark that shares 95% DNA identity with JA10. The closest published phage, covering 18% of the genome with 74% nucleotide identity, is a *Pectobacterium parmentieri* phage PP74 isolated from potato washing waste water in Russia in 2015 [84]. It shares no nucleotide identity with the only other published

Phage	Isolation date	Enrichment
JA10	03/11/14 Sample 3	12 hours
JA11	03/11/14 Sample 3	12 hours
JA13	03/11/14 Sample 2	12 hours
JA29	18/11/14	8 hours
JA31	11/11/14 Sample 1	8 hours
JA32	11/11/14 Sample 1	8 hours
JA33	11/11/14 Sample 1	12 hours
JA37	11/11/14 Sample 3	24 hours

Table 6.4 Isolation dates of broader host range *Dickeya* phages. Samples were taken five minutes apart from a treated sewage outflow in Cambridge. Enrichments with *D. solani* were carried out over 24 hours with samples taken at four hour intervals.

Dickeya-infecting podovirus BF25/12 [13]. PP74 has been designated as a T7-like virus and a member of the *Autographivirinae* subfamily, with a conserved core genome. A translated nucleotide comparison of JA10 with T7 shows that most of the predicted genes are conserved. This consists of almost all the genes with a proposed function, including the T3/T7-like RNA polymerase, structural capsid genes and DNA packaging machinery. JA10 is therefore a member of the *Autographivirinae* subfamily.

There are several genes encoded by the JA10 genome which do not have significant homology in the phage T7, which are highlighted in Fig. 6.2. Four of these seven genes annotated as encoding the S-adenosyl-L-methionine hydrolase, bacterial RNA polymerase inhibitor, minor capsid protein and the first tail assembly protein, have a gene of similar function at this position in T7. These variations between JA10 and T7 in these predicted proteins therefore could be a determinant of host specificity. The marked tail fibre protein, which share a common N-terminal region but differ at the C-terminus between the two phages, is also likely involved in host range specificity. The DNA ligase highlighted in Fig. 6.2 neighbours a conserved ligase and consists of fewer than 200 codons. This is therefore a possible result of a recombination duplication event and may not be functional. The final tail assembly gene, close to the end of the JA10 genome, has no functional homologue in T7.

Novel jumbo *Myoviridae*

Sequencing of the other seven genomes showed that, although they had been isolated independently, several shared 100% identity at the nucleotide level. JA11, 31 and 32 grouped together, as did JA33 and 37. As discussed earlier and shown in Table 6.4, of these only JA31 and JA32 were isolated from the same sample at the same time, with JA11 isolated over a

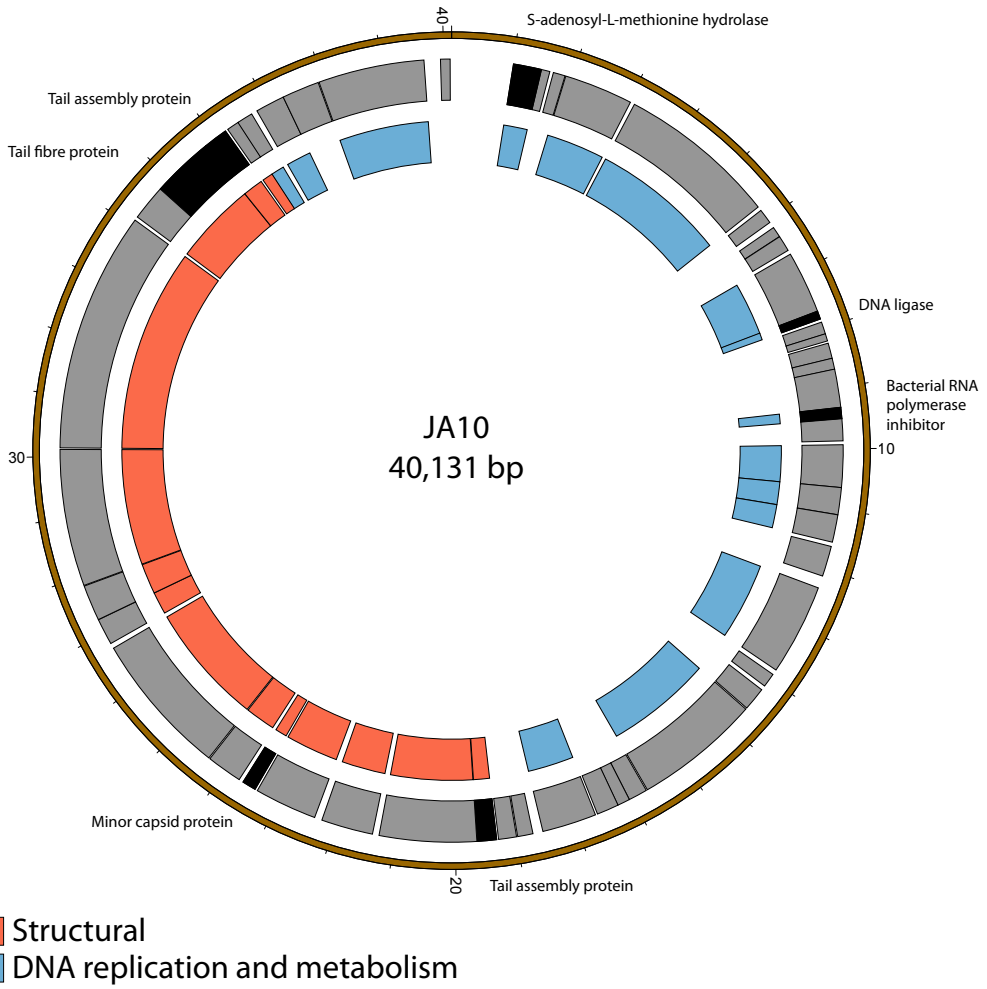


Fig. 6.2 JA10 genome map. The outer grey ring marks open reading frames, whilst the inner ring categorises the proposed functions of these predicted genes. Highlighted genes are discussed further in the text.

week earlier. JA33 and JA37 were isolated on the same day at different times. JA31, 32 and 37 were therefore excluded from further analysis.

The genome size for the four remaining phages is between 253 kbp and 256 kbp. A summary is shown in Table 6.5. These phages are therefore jumbo phages, defined as phages with a genome over 200 kbp [188]. The genomes are significantly larger than the conserved size of the *Ackermannviridae* genomes, which are around 150 kbp, and larger than most sequenced phages. As of August 2018, there were nearly 10,000 recorded phage genome sequences in Genbank [114] and these JA phages would be the 60th-63rd largest sequenced. Many of the over 300 predicted open reading frames in each genome do not match any annotated genes; the majority of those that do share any identity with known genes are from the *Erwinia amylovora* phages Yoloswag [62] and Y3 [32]. The annotations for these ORFs are listed in Appendices A.4, 5, 7 and 8. These are largely structural genes and genes involved in DNA metabolism and replication.

Phage	**Genome size (bp)**	**GC content (%)**	**Open Reading Frames**
JA10	40,131	51.5	50
JA11	255,356	44.5	321
JA13	254,061	44.5	323
JA29	253,327	43.8	319
JA33	255,356	44.5	321

Table 6.5 Summary of the broader host range phages genomes. Annotations of the ORFs are listed in Appendices A.3, 4, 5, 7 and 8 respectively.

The jumbo phages exhibit very low nucleotide identity with any published genomes, although they do possess many of the genes believed to form the T4 'core genome' [131]. As mentioned, the closest match for some of the genes during annotation was the *Erwinia amylovora* phage Y3. A translated nucleotide comparison of JA11 with Y3 is shown in Fig. 6.3. This shows that there is conservation of most of the annotated genes between the two phages. The majority of the genes that do not share identity are predicted to encode hypothetical proteins, aside from a putative DNA adenine methylase. Y3 also possesses a putative DNA adenine methylase in the same genomic context, but the translated nucleotide identity of the two genes is less than 15%. This likely reflects the different hosts of the two phages, as phage-encoded methyltransferases are thought to offer protection against host restriction-modification systems [119].

Variation within the JA jumbo phages

The gene order of the four JA jumbo phages is largely conserved. Over three quarters of the predicted ORFs are annotated as encoding hypothetical proteins, and many of the differences between the phages is contained within these ORFs as shown in Fig. 6.4. JA29 is the most different from the others, sharing 86% nucleotide identity with JA11, and JA13 shares 95% nucleotide identity with JA11. JA11 and JA33 share 99% identity, with the major difference being the insertion of 126 bp in both genomes at different positions, and of different sequences. These insertions are in non-coding regions however, therefore their relevance is unclear. The only other differences are in two genes: one with no predicted function and the other containing a putative discoidin domain, with an alanine to threonine substitution in the middle of the domain. Discoidin domains are present in eukaryotic agglutination factors and therefore the possible biological role for this in a phage genome, and the effect of the substitution, is not immediately obvious.

Whilst most of the differences between the four phages are located in genes with no predicted function, there are five with annotation that are present in all of the JA phages. These five were analysed further to see how close JA11 and JA33 (which share 100% amino acid identity in all five of these putative proteins) were to JA13 and JA29. They consist of two related to DNA replication, two potential transcription initiation factors and one likely structural protein. All five are highlighted in Fig. 6.4 and listed in Table 6.6.

There are variations in two DNA-related genes: a DNA primase and a DNA helicase. The helicase shows the most variation between the phages, as it appears to have undergone insertion or deletion between some of the phages. A comparison of this region of the genome can be seen in Fig. 6.5. There are two ORFs annotated as putative helicases in JA11 and JA33, which both share homology with one ORF in JA13 and JA29. Whether the ORFs are able to function independently as helicases, or whether this duplication or deletion has rendered them non-functional, is unknown. A DksA/TraR family protein and a ssDNA binding protein, both likely transcription factors, differ in one amino acid between JA11 and JA13, and share lower identity with the JA29 homologue, particularly the ssDNA binding protein, which differs in 32 positions. A VgrG-like family protein, a component of the T6SS thought to be phage-derived as it is capable of assembling into a structure similar to a phage tail spike [42] shares, at best, only 33% amino acid identity with the closest hit in the *E. amylovora* phage Y3, and so these may define a relatively novel VgrG-like protein group. JA11 and JA13 differ by one conservative substitution in this protein, whilst JA29 differs in 87 amino acids, 13 of which are conservative substitutions. It is possible that differences in this predicted protein are a contributing factor to the differing host range of these phages.

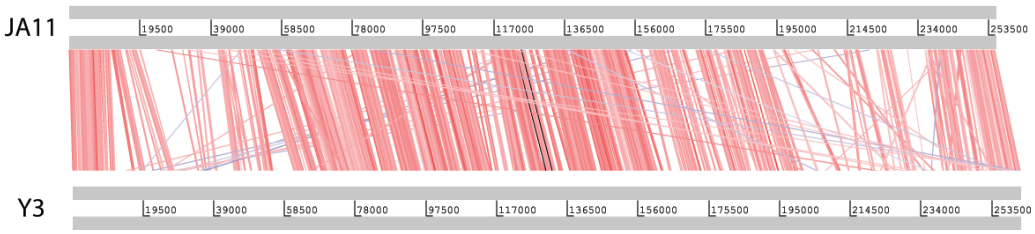


Fig. 6.3 Translated nucleotide comparison of the genomes of JA11 and Y3. Red bars mark areas of conservation, with darker colours showing higher conservation. Blue bars highlight areas of inversion.

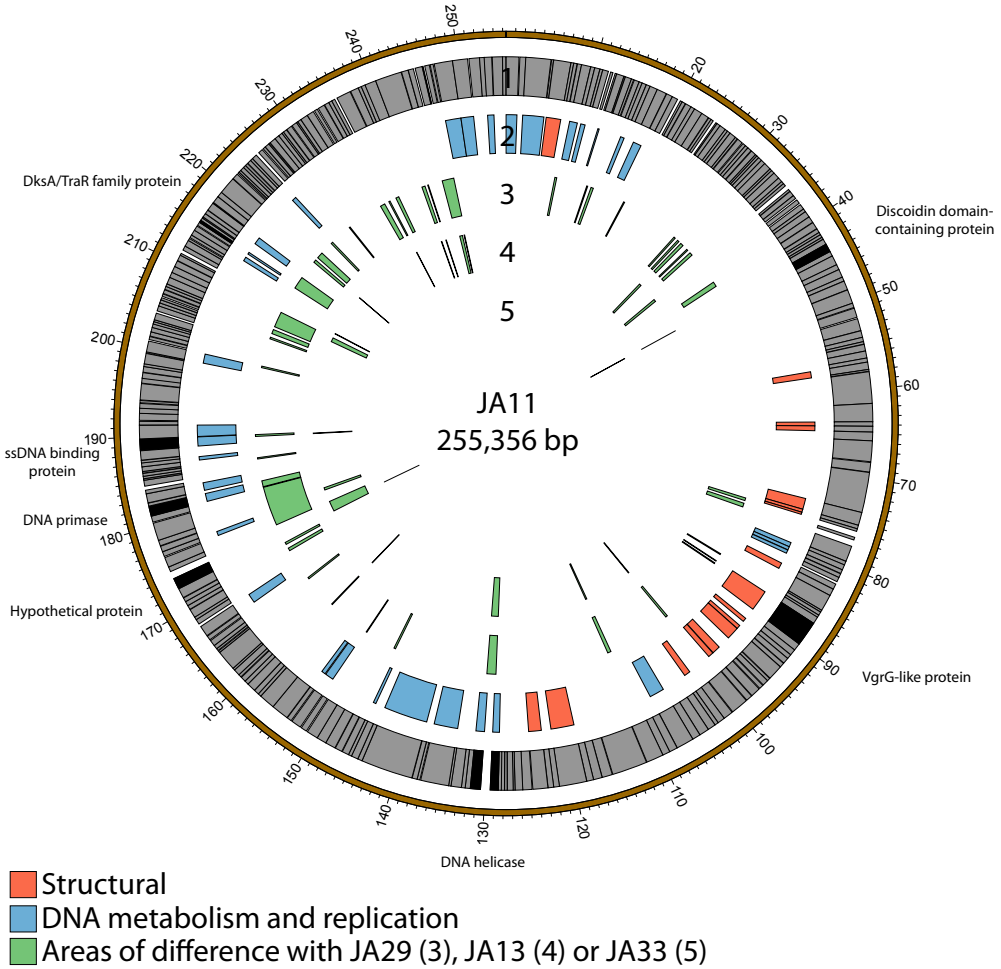


Fig. 6.4 JA11 genome map. The outer grey ring (1) marks open reading frames whilst the second ring categorises the proposed functions of these genes. The inner rings highlight the areas of the genome that differ from the genomes of JA29, JA13 and JA33 (third to fifth ring respectively). Highlighted genes are discussed further in the text. The genome map was generated using Circos.

Function	Gene	Length	Amino acids that differ with JA11
Tail fibre one	JA11_090	272	
	JA13_090		0
	JA29_093		6
Tail fibre two	JA11_094	164	
	JA13_095		0
	JA29_096		7
Tail fibre three	JA11_95	210	
	JA13_096		0
	JA29_098		1
DNA primase	JA11_208	350	
	JA13_208		0
	JA29_210		7
DNA helicase	JA11_155 + 156	*	
	JA13_156		*
	JA29_158		*
DksA/TraR family protein	JA11_264	85	
	JA13_267		1
	JA29_265		5
ssDNA binding protein	JA11_221	402	
	JA13_222		1
	JA29_223		32
VgrG-like protein	JA11_117	931	
	JA13_118		1
	JA29_120		87

Table 6.6 Summary of the annotated genes which differ between JA11, JA13 and JA29. * see Fig. 6.5

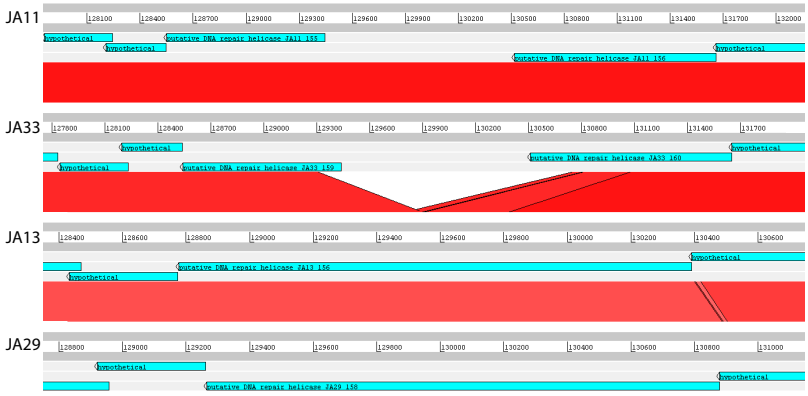


Fig. 6.5 Translated nucleotide comparison of putative DNA helicases between JA11 (top), JA33 (second), JA13 (third) and JA29 (bottom). Red bars mark areas of amino acid identity, with darker colours showing higher identity.

The significance of the differences observed between the four JA phages is currently unclear. It is somewhat surprising to find variations in genes involved in transcription initiation and DNA replication, as the genomes of these phages are relatively similar in both size and GC content, as summarised in Table 6.5. It is therefore possible that these differences do not significantly alter the function of these proteins. The variation in the VgrG-like proteins is more logical, as the different host ranges of these phages may be related to differences in tail spike proteins and other host recognition factors. When looking at the predicted tail fibres in these phages, there is no difference between JA11, 13 and 33 in the three predicted tail fibre proteins as shown in Table 6.6. JA29 shows minor variations in all three, which may contribute to the difference in host range between these phages. To determine the impact of these differences, and to investigate why JA11 and JA13 have a different host range despite having identical tail fibres, would require further experimental work.

6.2.4 Novel environmental isolates: AD phages

All of the JA phages described in this chapter were isolated in November 2014. Isolation of XF phages from the same location in 2013 produced mainly members of the *Ackermannviridae* family and a few *Podoviridae* family members. Whilst there is clearly some maintenance of viral populations, as members of the two families have been isolated on both occasions, the jumbo phages presented here are a novel grouping. To gain further insight into the viral populations in the River Cam and see whether the previously isolated families of phages were maintained, further samples were taken in October 2017. Two phages were isolated on *D. solani* and are named AD1 and AD2. When viewed under microscopy, AD1 (Fig. 6.6c)

appeared to have a morphology similar to that of the JA jumbo phages, with a head diameter of 120 nm, tail length of 150 nm and unclear structures at the base of the tail. AD2 on the other hand (Fig. 6.6d) had a head diameter of 90 nm and a (potentially partially-contracted) 70 nm tail, putting it closer to the previously imaged *Ackermannviridae* family members. The structures at the end of the tail were inconclusive.

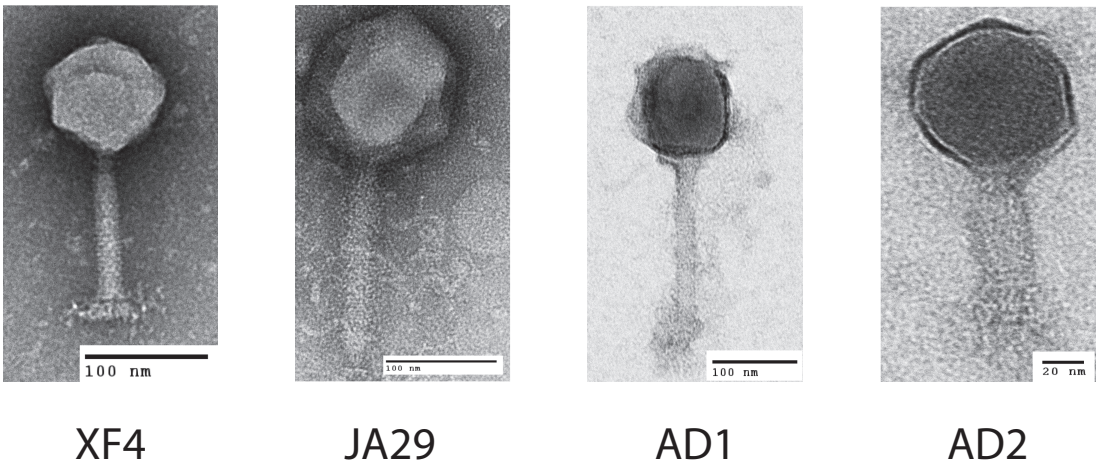


Fig. 6.6 Electron micrographs of the previously discussed XF4 and JA29, along with the novel environmental isolates AD1 and AD2. AD1 exhibits similar morphology to JA29 and AD2 is similar to XF4.

Host range testing showed that both AD phages were only capable of forming plaques on *D. solani* and not strains of other *Dickeya* species. However, both phages proved capable of facilitating the transduction of a chromosomal marker between *D. solani* cells at a frequency greater than 10^{-6}. This morphological and phenotypic evidence suggested that AD2 was a member of the *Ackermannviridae* family and that AD1 was different from any of the phages already discovered, but was potentially a jumbo phage. Both of these phages therefore were genomically sequenced to determine whether this were the case.

As was suspected, the genome of AD2 showed that it is a member of the *Ackermannviridae* family. It shares 98% nucleotide identity with previously published *D. solani Ackermannviridae* such as XF4 and LIMEstone1 [54], although full coverage of the genome was not achieved. A nucleotide comparison of the contigs scaffolded onto the XF4 genome is shown in Fig. 6.7. This therefore suggests a maintenance of the Ackermannviridae viral population in the waterways around Cambridge, as representatives have been isolated on three separate occasions over the span of four years.

AD1, as expected, has a large genome of 261,658 bp, confirming that it is a jumbo phage, shown in Fig. 6.8. Annotations of the ORFs are listed in Appendix A.2. However,

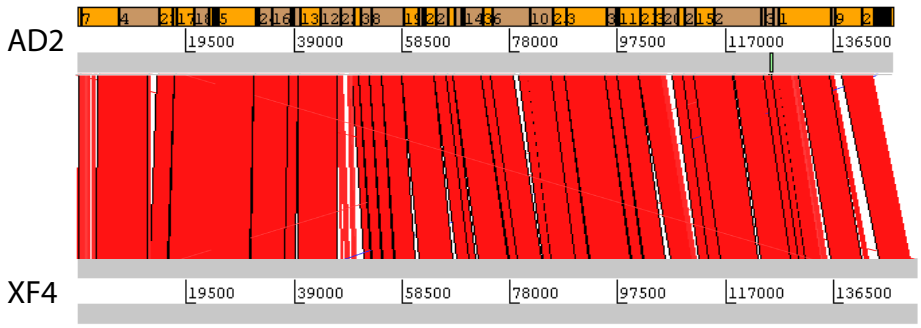


Fig. 6.7 AD2 contigs scaffolded onto the XF4 genome. Contigs generated through sequencing of AD2 could not be fully resolved into a complete genome and have instead been arranged using the existing full genome of XF4 as a scaffold. AD2 contigs map to cover 93% of the XF4 genome with an identity of 98%.

unexpectedly, it has low nucleotide identity with the JA jumbo phages, despite sharing a similar gene order. In fact, a translated nucleotide comparison of JA11 and AD1, as shown in Fig. 6.9c, shows that JA11 is about as similar to AD1 as it is to Y3, and a comparison of AD1 and Y3 (Fig. 6.9b) shows them to be more similar to each other than to JA11. AD1 therefore defines another new phage 'group' distinct from the JA jumbo phages.

Phylogeny of the 'hairy *Myoviridae*' phages

In their recent publication, Buttimer *et al.* discussed the phylogenetic position of Y3 considering its low level of nucleotide identity to existing genomes [32]. Two potential subgroups within the 'hairy *Myoviridae*' have emerged; the Rak2-like phages, which includes the previously mentioned *Pectobacterium* phage CBB [31], and the as yet unnamed group that encompasses the phages discussed here. This group was established as it was found that Y3 had homologues including terminase, polymerase and helicase genes in several other phages reported or suspected to have the 'hairy *Myoviridae*' morphology. A comparison of the tail sheath proteins of these phages with those reported here shows clear clustering, and can be seen in Fig. 6.10a. As expected, the three JA phages cluster tightly with little variation between them. As reported by Buttimer *et al.*, the *Pseudomonas*-infecting phages PaBG [162] and Lu11 [7] form a clade, whilst the *Ralstonia solanacearum* phage phiRSL1 [182] and the metagenomically-derived NCTB [132] are single nodes within the tree. As suggested by the translated nucleotide comparison in Fig. 6.9b, Y3 and AD1 form a clade that puts AD1 closer to *Erwinia*-infecting phages than to the other *D. solani* phages. Intriguingly, AD1 is placed closer phylogenetically to Y3 than the other *Erwinia*-infecting phage Yoloswag. All of the phages except phiRSL1 have two annotated tail sheath proteins, and the same

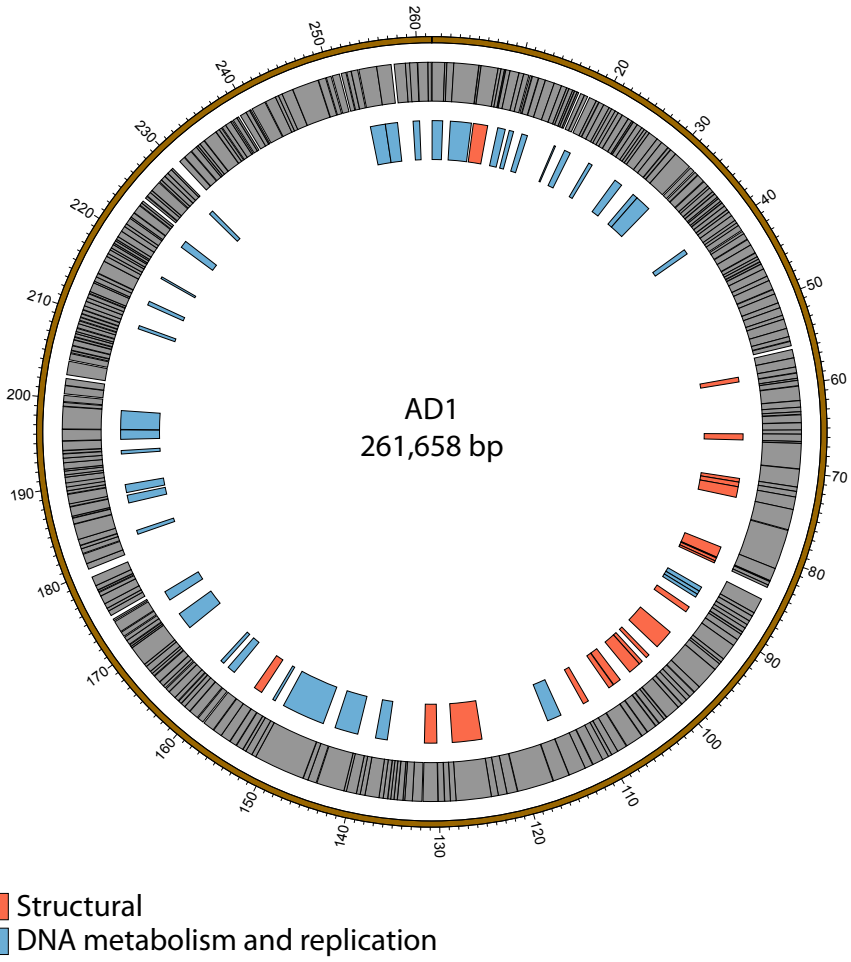


Fig. 6.8 AD1 genome map. The outer grey ring (1) marks open reading frames whilst the inner ring (2) categorises the proposed functions of these genes. The genome map was generated using Circos.

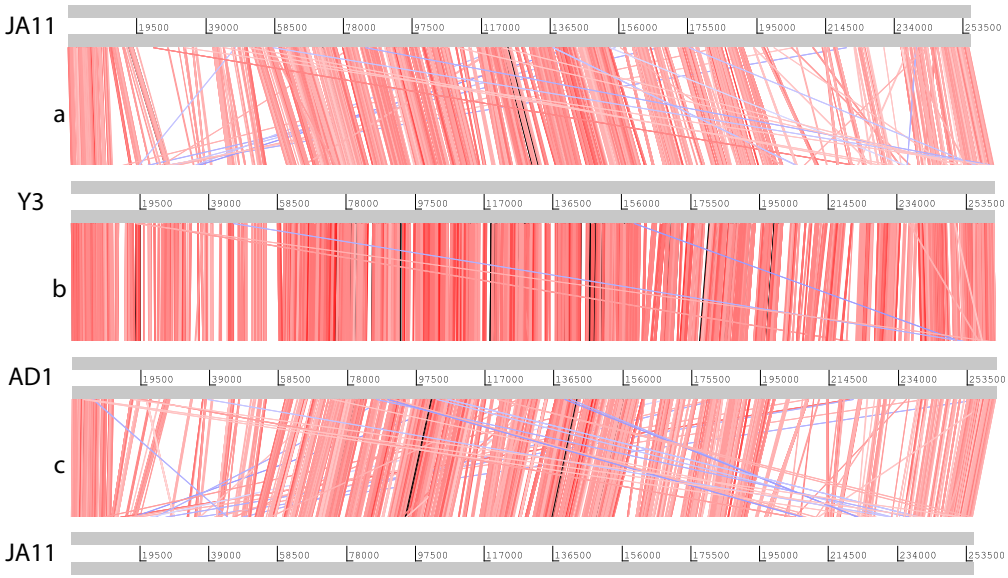


Fig. 6.9 Translated nucleotide comparison of the genomes of JA11, Y3 and AD1. Red bars mark areas of conservation, with darker colours showing higher conservation. Blue bars highlight areas of inversion.

phylogeny is seen with both (data not shown). The same clustering is seen when using the sequence of the large terminase subunit of the phages, shown in Fig. 6.10b.

The gene order between JA11, AD1 and Y3 is highly conserved. All three genomes contain over 300 open reading frames, with each containing only between one and three unique annotated genes. These unique genes are listed in Table 6.7 and are all DNA or metabolism-related. There are also five genes common to AD1 and Y3 that are not found in JA11. Whilst phylogenetic clustering, as shown in Fig 6.10, groups AD1 and Y3 closer than Y3 and Yoloswag, it is interesting to note that the two unique genes possessed by Y3 have homologues in Yoloswag. These two phages were both isolated from apple orchards using *Erwinia amylovora*, therefore it is surprising that they differ phylogenetically. It is possible that these unique genes are a determinant of the host range of these phages. A phylogenetic comparison of three tail fibre genes found in each genome is shown in Fig 6.11. This again shows a definite separation between Yoloswag and the other two phages, particularly when comparing Yoloswag_102 with Y3_104 and AD1_102, which occupy the same syntenic position. This also suggests the possibility that AD1 may be capable of forming plaques on *Erwinia* species, but this has not yet been tested due to the availability of strains.

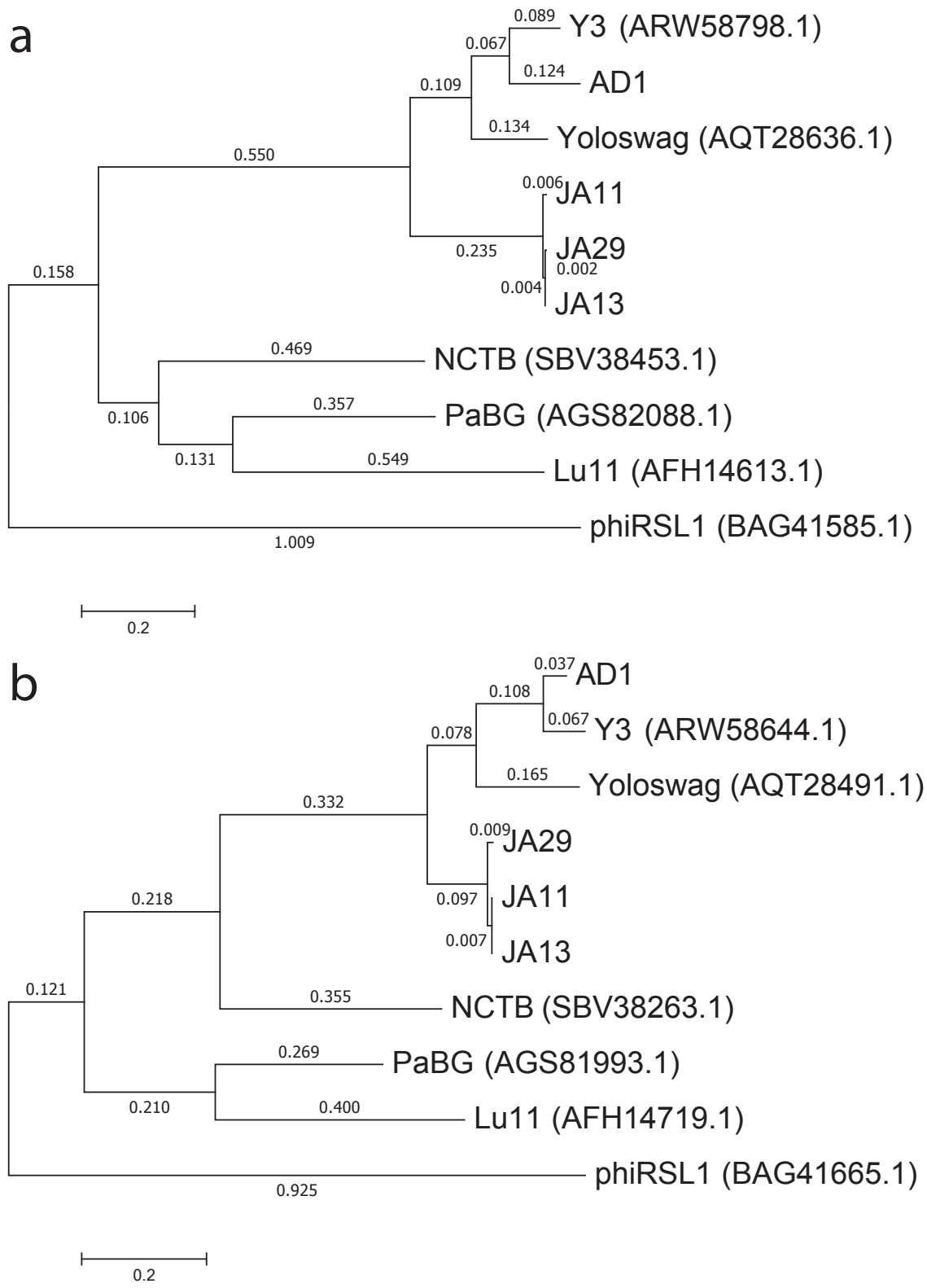


Fig. 6.10 Phylogenetic trees of the a) tail sheath protein and b) large terminase subunit from 'hairy' *Myoviridae* phages. Trees were calculated using the Maximum Likelihood method in MEGA. The trees with the highest log likelihood (-7514.90 and -8677.52) are shown. The trees are drawn to scale, with branch lengths measured in the number of substitutions per site (next to the branches). Both analyses involved 10 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 501 and 642 positions in the final datasets respectively.

Genome	Gene	Gene annotation
JA11	JA11_30	DNA adenine methylase
AD1	AD1_017	DUF1611-domain containing protein
	AD1_258	XRE family transcriptional regulator
Y3	Y3_020	Oxygenase
	Y3_031	AntA/B antirepressor domain-containing protein
Common to AD1 and Y3	AD1_047	Transcriptional repressor
	Y3_049	
	AD1_048	DNA-cytosine methyltransferase
	Y3_050	
	AD1_018	Asparagine synthase
	Y3_018	
	AD1_267	Radical SAM superfamily protein
	Y3_272	
	AD1_016	Methyltransferase
	Y3_017	

Table 6.7 Unique annotated genes found in the genomes of JA11, AD and Y3, as well as genes common to AD1 and Y3 but not present in JA11.

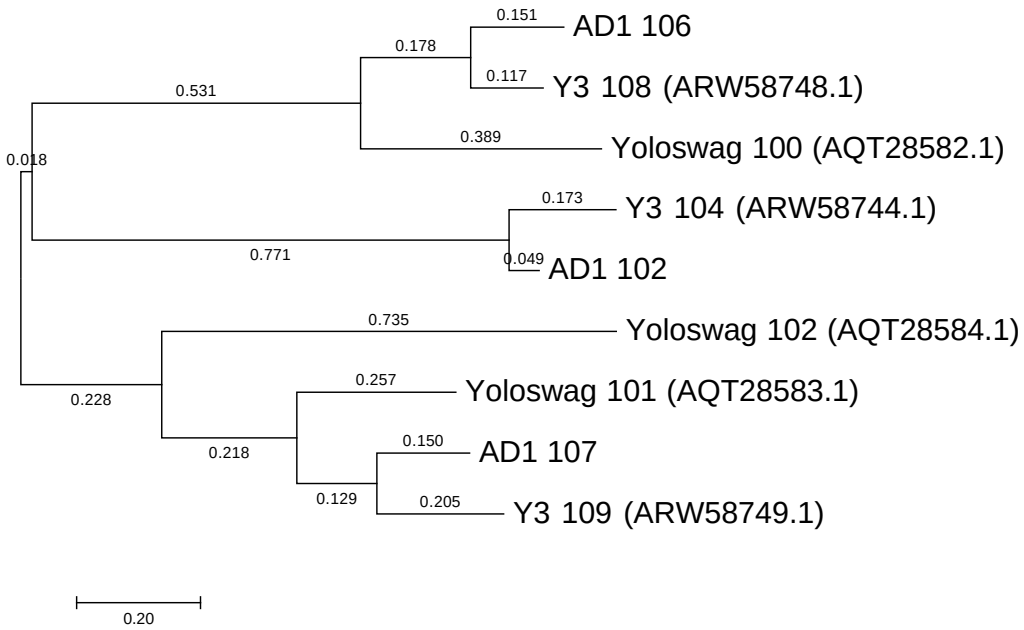


Fig. 6.11 Phylogenetic tree of the tail fibre proteins from 'hairy' *Myoviridae* phages. Tree calculated using the Maximum Likelihood method in MEGA. The tree with the highest log likelihood (-2426.84) is shown. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site (next to the branches). The analysis involved 9 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 158 positions in the final dataset.

6.3 Discussion

Previously isolated phages of *D. solani* have been found almost exclusively to be members of the *Ackermannviridae* family. This has been a consistent feature of phage isolations spanning multiple European countries across the last decade, including from both soil and water samples. There was, therefore, the question of whether this indicated a special relationship between *Ackermannviridae* family phages and *D. solani*. The phages discussed in this chapter show that this was the result of an extrapolation from a limited viral sample set. Whilst *Ackermannviridiae* family members have indeed been found in every group of samples taken over the course of four years, it can now be seen that there are at least four groups of *D. solani* phages present in waterways around Cambridge. Representatives from three of the four families of phages within the *Caudovirales* order have been isolated. Phages that have been isolated on other species of *Dickeya* that are capable of forming plaques on *D. solani* have also been recently described [13], including members of the fourth *Caudovirales* family, the *Siphoviridae*. It is therefore apparent that we are only superficially defining the level of phage diversity present in the environment, consistent with the notion that double-stranded DNA phages alone have been predicted to outnumber their bacterial hosts by a factor of ten to one [41].

All the phages presented here, apart from JA10, appear to be representatives of a recently described 'hairy *Myoviridae*' subfamily [32]. To the best of our knowledge, these are the first reported members of this subfamily isolated on *Dickeya* species. Many of the previously reported members of this subfamily were also isolated on soil and plant-associated bacteria such as *Pseudomonas putida* [7] and *Erwinia amylovora* [62, 32]. Whether there is a link between this group of phages and these bacteria, or whether the recent increase in isolation of phages using plant-associated hosts is skewing this view remains to be determined. The proteins responsible for the 'hairs' that typify this grouping remain unknown, although the identified tail fibre proteins are likely candidates for further investigation.

Chapter Seven

Discussion

7.1 *Ackermannviridae* family bacteriophages

The family *Ackermannviridae* has only been recently described following the elevation and renaming of the *Myoviridae* genus *Vi1virus* [9]. Members of the family share a distinct morphology and gene synteny, and have hosts across the recently proposed Enterobacterales order [4] including *Salmonella*, *Serratia* and *Dickeya* species. Studies in *Salmonella* and *Klebsiella* showed that the receptor for *Ackermannviridae* family phages in these hosts is capsular polysaccharide [77, 175]. Previous work in this laboratory had suggested that this was the case for *Dickeya solani* [10, 65] and *Serratia plymuthica* [106].

This suggestion was given more weight during this project for *D. solani*, for the first time demonstrating that insertion of a transposon into any one of the eight genes in the predicted capsular polysaccharide synthesis cluster resulted in resistance to *Ackermannviridae* family phages. When the same experiments were performed in *Serratia marcescens* however this proved not to be the case. Resistance to the phage was the result of transposon insertion into a myriad of genes, none of which were obviously linked to capsular polysaccharide. This was true for all four combinations of two *S. marcescens* strains and two *Ackermannviridae* family phages, MAM1 and 3M, suggesting that, at least for these combinations, that capsular polysaccharide may not be the receptor. Many of the genes disrupted in these phage-resistant mutants, however, were regulators. This could suggest that direct disruption of the gene(s) encoding the receptor in these hosts hinders growth of the cells such that only mutants in which transposons result in indirect disruption of these genes remain viable. I would consider this a more likely explanation, considering that similar experiments in *S. plymuthica* did highlight the capsular polysaccharide synthesis cluster as the receptor for MAM1 [106]. A comparison of the proposed cluster from *S. plymuthica* and one of the *S. marcescens* strains however showed no major functional differences that would support this hypothesis.

Phylogenetic mapping of conserved genes among the family *Ackermannviridae*, for example the major capsid protein in Fig. 7.1, shows that the *Serratia* phages tested, MAM1 and 3M, form a distinct clade with the other *Serratia*-infecting *Ackermannviridae* family members KSP90 and 2050H1. Based on the guidelines established by the International Committee on Taxonomy of Viruses [9], this clade could be proposed as a novel genus *Mam1virus* and would place MAM1 and 2050H1 in the same species. In contrast to the *Dickeya*-infecting *Ackermannviridae* family phages found both in Cambridge in this laboratory and also elsewhere in Europe, these *Serratia* phages have differing host ranges. The overlapping nature of the host range of MAM1 and 3M is an intriguing mystery and, whilst this project has shown that they share high identity in two of their four tail spike proteins, whether this explains the overlap remains unclear. MAM1 is also able to infect and facilitate transduction in *Kluyvera cryocrescens* 2Kr27, a member of a different bacterial family the Enterobacteriaceae, and therefore appears much more promiscuous than other *Ackermannviridae* family phages. To the best of my knowledge this is the only *Ackemannviridae* family phage that has been shown to have a multi-familial host range, and therefore makes it, and the other members of this novel proposed genus, very interesting for further research.

7.2 Determining host range of *Ackermannviridae* family members isolated on *D. solani*

The majority of *Dickeya* phages characterised before this project began are capable of forming individual plaques on *D. solani* only [8, 10, 65]. Czajkowski *et al.* [48] have reported that phages D3 and D5 are capable of infecting multiple species of *Dickeya*. This conclusion was based on simple spot test assays in which undiluted spots of phage lysate were tested on bacterial top lawns and incubated, with any resultant clearing taken to show infection. It is known that applications of high titre lysates of phage to bacterial cells can cause the phenomenon of 'lysis-from-without', in which cells lyse due to membrane disruption instead of productive phage infection [85]. Consequently, confirmation of host range requires serial dilution of the phage lysate to visualise individual plaques on a host. Until these confirmatory data are provided, I remain unconvinced by the reported host range of these phages, especially considering the genome identity of nearly 100 % with other *Ackermannviridae* family members.

The phages PD10.3 and 23.1 were also isolated by Czajkowski *et al.* [50] and are reported to infect both *Dickeya* and *Pectobacterium* species, although host range was determined by the same method as D3 and D5. They do however report adsorption and burst size data

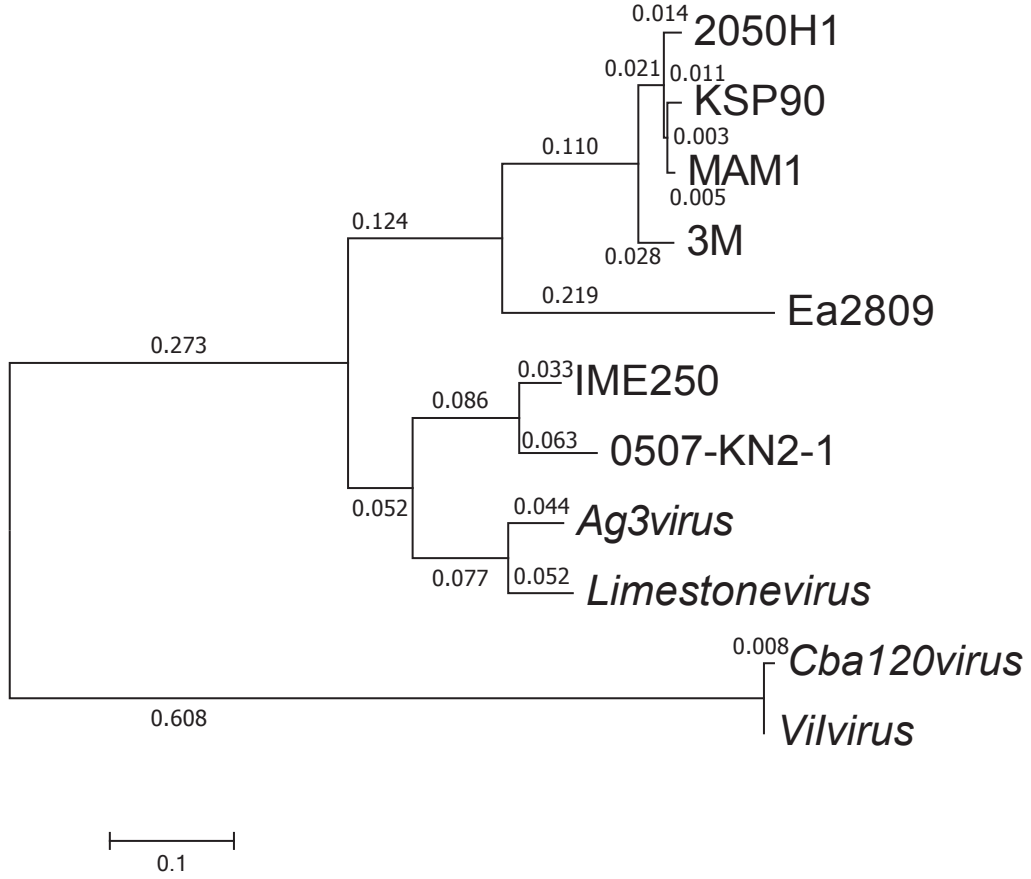


Fig. 7.1 Phylogeny of *Ackermannviridae* family major capsid protein using the Maximum Likelihood method. The tree with the highest log likelihood (-3647.71) is shown. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. All positions containing gaps and missing data were eliminated. There were a total of 370 positions in the final dataset. The genera *Ag3virus*, *Limestonevirus*, *Cba120virus* and *ViIvirus* include multiple phages but have been condensed for readability, with the sequence of the corresponding protein in the type phage (AG3, LIMEstone1, CBA120 and ViI) for each genus used to calculate the tree.

for both phages on the two genera. The genomes of both have been sequenced and are reported as incomplete. However, the largest scaffold of both is similar to the size of other *Ackermannviridae* family members, as shown in Table 6.1 and these scaffolds share 99% identity with the full genome of LIMEstone1. The morphology of these two phages also clearly places them within the *Ackermannviridae* family. It is therefore intriguing that these phages are so similar yet have such different host ranges, and it would be interesting to see data confirming the broader host range of these two phages, because, if this host range were confirmed, it would be an exciting and surprising discovery. However, I am currently sceptical of this reported host range, and advise caution, as my own experimentation has disproved host range data obtained previously in the lab that was found to be insufficiently rigorous (data not shown). Host range tests of my phages following the method of Czajkowski *et al.* have also, by the standards of this method, shown a much broader host range than I know to be true (data not shown). I would therefore caution against assigning host range to phages without rigorous experimentation involving plaque formation data of these phages.

7.3 Increased diversity of *Dickeya* bacteriophages

Almost all *D. solani* bacteriophages published before this project began are members of the family *Ackermannviridae* [8, 10, 48, 50, 65]. This has been a consistent finding of isolations in multiple European countries across the past decade from both terrestrial and aquatic environments, and seemed to suggest a strong link between this family of phages and *D. solani*. The phages presented here however show that, whilst it is still possible to isolate *Ackermannviridae* family phages, at least three other groups of *D. solani* phages are present in the waterways around Cambridge, and this laboratory now possesses representatives from three out of the four families of the tailed bacteriophage order *Caudovirales*. These three groups, one *Podoviridae* and two sub-groups of the 'hairy *Myoviridae*' [32], are all novel, as they share little homology with any published phages. Whilst the *Podoviridae* family member JA10 possesses functional homologues of many classical T7-like genes and is otherwise consistent with well defined members of this sub-family, the members of the 'hairy *Myoviridae*' grouping are all jumbo phages with a genome over 250 kb with little similarity to any well studied phages. Phylogenetic mapping shows that the broader host range phages (JA11, 13, 29, 31, 32, 33 and 37) all share high similarity despite their varying host ranges. The more recently isolated *D. solani*-specific phage AD1 however shares little identity with these phages, and instead is more closely related to the *Erwinia amylovora* phages Y3 and Yoloswag. Whether this phage is capable of forming plaques on *E. amylovora* has not been tested, but the results, either way, would be intriguing. It is therefore apparent that we are

only superficially defining the level of phage diversity present in the environment, consistent with the notion that double-stranded DNA phages alone have been predicted to outnumber their bacterial hosts by a factor of ten to one [41].

7.4 Phage therapy and *D. solani*

Research into phages of plant pathogens such as *D. solani* is largely conducted in an effort to investigate the potential for phage-based biocontrol tools. There is currently one commercial phage-based product, Biolyse, available that targets *Dickeya* species, but the phages contained within the product have not been described in the public domain. Published *D. solani* phages have undergone some suitability testing, including persistence [53] and field trials [8]. However, as reported here and previously [10, 107], all of the *Dickeya* phages tested in this lab, including a phage from Belgium that has undergone field trials [8], are capable of facilitating transduction of chromosomal and plasmid markers at high efficiency. Testing of the capability for transduction for the Polish phages has not been reported, although logic predicts that they will also be shown capable when tested due to the high level of genome identity. I would however suggest that, whilst this ability makes the phages very useful for genetic manipulation in a research setting, it renders them unsuitable for use in the environment. Echoing the caution of the European Medicines Agency, among others, who have stated that it is 'important to ensure that therapeutic phages do not carry out generalized transduction' [129], I would not advise use of *Ackermannviridae* family phages in the environment. It has also been shown that all the JA phages except JA10 are also capable of facilitating transduction with high efficiency, which would also render them unsuitable for use in the field. The *Podoviridae* family member JA10 however proved incapable of facilitating transduction when tested, and, as can be seen in Table 7.1 is capable of forming plaques on strains of three other *Dickeya* species as well as *D. solani*. This phage is therefore a more attractive candidate for therapeutic use, but would require extensive testing to confirm stability, efficacy and persistence before it could be commercialised.

It was demonstrated that *Ackermannviridae* phages of *D. solani* had the capacity to adsorb to three other *Dickeya* species but were incapable of forming individual plaques, likely due to host factors such as CRISPR. It still remains formally possible that there are other *Dickeya* isolates outside of those tested here that are permissive for these phages. However, the novel phages presented here are capable of a wider host range among *Dickeya* species tested in this laboratory, as shown in Table 7.1. This could make them more attractive for use in phage therapy, as they are capable of acting on a wider range of pathogens. However, whether the broader host range is beneficial is a point for discussion, even though the majority of the

literature is supportive of broad host range phages along the same lines as broad spectrum antibiotics [145]. A commonly-cited advantage of phage therapy is the specificity of the phage allows avoidance of 'off-target' effects on the normal bacterial flora, preventing the microbial dysbiosis that can be caused by traditional antibiotics [121]. However, this project, among many other reports, shows that phages can have an unexpectedly broad host range and that host range data is only as good as the array of hosts included. Whilst host range testing is normally performed against a selection of common lab strains, the phage MAM1 proved capable of forming plaques on a species of *Kluyvera*, and this was only discovered due to screening of MAM1 against a variety of environmental isolates that were thought to be *Serratia* species (Miguel Matilla, personal communication). This project has also shown that, despite a reported host range restricted to *D. solani* only, two *Dickeya* phages were capable of adsorbing to strains of three other *Dickeya* species, but were unable to form plaques. The true host range of a phage is therefore likely unknowable, however, I would posit that a broader host range is not necessarily the advantage it may seem, as it increases the likelihood of 'off-target' effects.

***Dickeya* species**	XF24, 25, 26 and JA10	XF27 and 28	JA11, 31, 32, 33 and 37	JA13	JA29
D. dadantii subsp. dieffenbachiae	+	-	+	+	+
D. paradisiaca	-	-	+	+	+
D. dianthicola	+	+	+	-	-
D. zeae	-	-	+	+	-
D. chrysanthemi	+	+	-	-	-

Table 7.1 Broader host range of eight phages capable of infecting other species of *Dickeya*. + denotes isolated plaque formation of the phages on the respective host. - denotes no observed plaque formation.

7.5 Future directions

The family *Ackermannviridae* is the newest family of the order *Caudovirales*. Whilst the ViI phage has been the most well studied due to its inclusion in the classical *Salmonella* typing set [175], further investigation has been limited. The data presented here illuminate the possible receptor for the majority of these phages, agreeing with previously published work, but also demonstrate that this is not necessarily applicable for all members of the family. Genome size and synteny remains common to all members, as does the capacity to facilitate

transduction between host cells in all those tested. Whilst this makes them unsuitable for use in environmental application, they have potential for use in the lab as genetic tools. The work by Wetter *et al.* [175] demonstrated that the ViI phage could form plaques on a previously non-permissive host after expression of the *Salmonella* capsule cluster in that host. This suggests that the ViI phage has the capability to replicate inside and lyse non-permissive hosts, it merely requires a receptor to enter. Future work should test if this is true for other *Ackermannviridae* family members, as this could prove useful for genetic manipulation, particularly in less tractable hosts.

The *Ackermannviridae* family phage MAM1 has the broadest reported host range, able to infect bacteria across two families of the Enterobacterales order. Unlike members of this family isolated against *D. solani*, all of which have the same host range when tested, another member of the newly proposed *Mam1virus* genus, 3M, has an overlapping, but different host range. The host recognition apparatus, consisting of four tail spike proteins, shares little to no identity in two of the four predicted proteins, whereas the other two are largely conserved. Due to the high level of genome conservation in the rest of the genome between these two phages, it would be interesting to investigate whether the two differing proteins are the cause of the different host ranges. Replacement of one or both of these proteins from MAM1 into 3M, or vice versa, would prove illuminating. Further investigation into the host receptor for these phages is also needed due to the conflicting results arising from the mutagenesis screens performed in this project. Screening of an unbiased mutant library for phage-resistance may prove sufficient, otherwise direct genetic manipulation of the predicted cps cluster in *Serratia marcescens* strains may be needed to confirm or deny its role as the receptor.

The discovery of *D. solani* phages that were not members of the *Ackermannviridae* family and instead part of the 'hairy *Myoviridae*' occurred in the final few months of this project, therefore they have not been studied in great depth in laboratory experiments. The data presented previously [10] shows that they are still capable of forming plaques on *D. solani* cells containing a transposon in the capsule cluster, thereby suggesting that they do not use the capsule as the receptor. Mutagenesis experiments using these phages to select for resistant mutants could therefore be performed to determine the receptors for these phages. The designation of these phages as 'hairy' is due to their morphology when viewed under transmission electron microscopy, in which the tails are less distinct than for other phages and appear to be wrapped in a bundle of proteinaceous fibres. Whilst this morphology has been seen multiple times since the first reports in 1983 [3], the nature of these fibres, or their physiological or structural relevance, is yet to be determined. It has been suggested that they are likely tail fibres [32], and this could be investigated via genetic knockouts, structural proteomics and electron cryomicroscopy. The phylogenetic clustering of the *Dickeya* phage

AD1 with the *Erwinia* phages Y3 and Yoloswag also merits further investigation, particularly into whether this phage is capable of interacting with *Erwinia* species.

As already discussed, the majority of phages discussed here would be deemed unsuitable by organisations such as the European Medicines Agency for use as a biocontrol tool due to their propensity to transfer genes between host cells [129]. This does raise the question as to the identity of the phage(s) contained within the commercially available Biolyse product, which purports to clear infections of both *Dickeya* and *Pectobacterium* species. Due to its commercial nature there are no publications detailing the phages included in this cocktail but I would expect that best practice was followed and capacity for transduction was tested during development. The lack of information in the public domain concerning this cocktail is concerning however, as other studies isolating *Dickeya* phages did not routinely test for this capability [8, 13, 50, 51, 53]. One of the phages presented here however, the *Podoviridae* family member JA10, does not appear to have this capability under the conditions tested, and therefore may be more suitable, as it is capable of forming plaques on strains of four *Dickeya* species. Five other *Podoviridae* were also found, but were not investigated further as the lysates were no longer viable. JA10 would require testing for stability, efficacy and persistence before application in the field, but the methods for this are relatively well established. However I would again raise the question as to whether the broad host range is truly beneficial and would encourage extensive and rigorous host range testing before environmental application.

In 2015 this project began with the aim of studying bacteriophages belonging to the genus *Vi1virus*, situated taxonomically in the order *Caudovirales* and family *Myoviridae*, and their interaction with their hosts comprising two genera of the Enterobacteriales order and Enterobacteriaceae family, *Serratia* and *Dickeya*. Advances in genomic sequencing however has led to large scale taxonomic reclassification of a variety of microbes. In 2016 the order Enterobacteriales was renamed Enterobacterales and the number of families expanded from one to seven [4], with the result that the genus *Dickeya* is now a member of the family Pectobacteriaceae, whereas *Serratia* now belongs to the Yersiniaceae family. In April 2018 the bacteriophage genus *Vi1virus* was raised to become the novel *Caudovirales* family Ackermannviridae [9] and a debate is currently under way over proposals for wholesale reorganisation of bacteriophage phylogeny and abolition of current families to create a new framework [11]. Given the current rate of change in taxonomic reclassifications, the groupings discussed in this dissertation may soon become obsolete. However, the interactions between these phages and their hosts, whatever they may be named, still remains an important area for research before application in therapeutic, industrial or environmental settings.

References

[1] Stephen T. Abedon. Bacteriophage secondary infection. *Virologica Sinica*, 30(1): 3–10, February 2015. ISSN 1674-0769. doi: 10.1007/s12250-014-3547-2.

[2] Hans-W. Ackermann. Bacteriophage Electron Microscopy. In *Advances in Virus Research*, pages 1–32. Elsevier, 2012. doi: 10.1016/B978-0-12-394621-8.00017-0.

[3] Hans-W. Ackermann and Thé-M. Nguyen. Sewage coliphages studied by electron microscopy. *Applied and Environmental Microbiology*, 45(3):1049–59, March 1983. ISSN 0099-2240.

[4] Mobolaji Adeolu, Seema Alnajar, Sohail Naushad, and Radhey S Gupta. Genome-based phylogeny and taxonomy of the 'Enterobacteriales': proposal for Enterobacterales ord. nov. divided into the families Enterobacteriaceae, Erwiniaceae fam. nov., Pectobacteriaceae fam. nov., Yersiniaceae fam. nov., Hafniaceae fam. nov., Morgane. *International Journal of Systematic and Evolutionary Microbiology*, 66(12): 5575–5599, December 2016. ISSN 1466-5026. doi: 10.1099/ijsem.0.001485.

[5] Mazhar Adli. The CRISPR tool kit for genome editing and beyond. *Nature Communications*, 9(1):1911, December 2018. ISSN 2041-1723. doi: 10.1038/s41467-018-04252-2.

[6] Evelien M Adriaenssens, Hans Wolfgang Ackermann, Hany Anany, Bob Blasdel, Ian F Connerton, David Goulding, Mansel W Griffiths, Steven P Hooton, Elizabeth M Kutter, Andrew M Kropinski, Ju Hoon Lee, Martine Maes, Derek Pickard, Sangryeol Ryu, Zargham Sepehrizadeh, S Sabouri Shahrbabak, Ana L Toribio, and Rob Lavigne. A suggested new bacteriophage genus: "Viunalikevirus". *Archives of Virology*, 157 (10):2035–2046, October 2012. ISSN 03048608. doi: 10.1007/s00705-012-1360-5.

[7] Evelien M. Adriaenssens, Wesley Mattheus, Anneleen Cornelissen, Olga Shaburova, Victor N. Krylov, Andrew M. Kropinski, and Rob Lavigne. Complete Genome Sequence of the Giant *Pseudomonas* Phage Lu11. *Journal of Virology*, 86(11):6369–6370, June 2012. ISSN 0022-538X. doi: 10.1128/JVI.00641-12.

[8] Evelien M. Adriaenssens, Johan van Vaerenbergh, Dieter Vandenheuvel, Vincent Dunon, Maurice Ceyssens, PieterJanuaryand de Proft, Andrew M. Kropinski, Jean Paul Noben, Martine Maes, and Rob Lavigne. T4-related bacteriophage LIMEstone isolates for the control of soft rot on potato caused by '*Dickeya solani*'. *PLoS ONE*, 7(3), 2012. ISSN 19326203. doi: 10.1371/journal.pone.0033227.

[9] Evelien M. Adriaenssens, Johannes Wittmann, Jens H. Kuhn, Dann Turner, Matthew B. Sullivan, Bas E. Dutilh, Ho Bin Jang, Leonardo J. van Zyl, Jochen Klumpp, Malgorzata Lobocka, Andrea I. Moreno Switt, Janis Rumnieks, Robert A. Edwards, Jumpei Uchiyama, Poliane Alfenas-Zerbini, Nicola K. Petty, Andrew M. Kropinski, Jakub Barylski, Annika Gillis, Martha R. C. Clokie, David Prangishvili, Rob Lavigne, Ramy Karam Aziz, Siobain Duffy, Mart Krupovic, Minna M. Poranen, Petar Knezevic, Francois Enault, Yigang Tong, Hanna M. Oksanen, and J. Rodney Brister. Taxonomy of prokaryotic viruses: 2017 update from the ICTV Bacterial and Archaeal Viruses Subcommittee. *Archives of Virology*, 163(4):1125–1129, April 2018. ISSN 0304-8608. doi: 10.1007/s00705-018-3723-z.

[10] Jiyoon Ahn. *Isolation and characterisation of new bacteriophages of the emerging phytopathogen , Dickeya solani*. Master's, University of Cambridge, 2015.

[11] Pakorn Aiewsakun, Evelien M. Adriaenssens, Rob Lavigne, Andrew M. Kropinski, and Peter Simmonds. Evaluation of the genomic diversity of viruses infecting bacteria, archaea and eukaryotes using a common bioinformatic platform: steps towards a unified taxonomy. *Journal of General Virology*, pages 1–13, July 2018. ISSN 0022-1317. doi: 10.1099/jgv.0.001110.

[12] Špela Alič, Tina Naglič, M. Tušek-Žnidarič, Matjaž Peterka, Maja Ravnikar, and Tanja Dreo. Putative new species of the genus *Dickeya* as major soft rot pathogens in *Phalaenopsis* orchid production. *Plant Pathology*, 66(8):1357–1368, October 2017. ISSN 00320862. doi: 10.1111/ppa.12677.

[13] Špela Alič, Tina Naglič, Magda Tušek-Žnidarič, Maja Ravnikar, Nejc Rački, Matjaž Peterka, and Tanja Dreo. Newly Isolated Bacteriophages from the *Podoviridae*, *Siphoviridae*, and *Myoviridae* Families Have Variable Effects on Putative Novel *Dickeya* spp. *Frontiers in Microbiology*, 8:1–14, 2017. ISSN 1664-302X. doi: 10.3389/fmicb.2017.01870.

[14] Hany Anany, Erika J Lingohr, Andre Villegas, Hans-Wolfgang Ackermann, Yi-Min She, Mansel W Griffiths, and Andrew M Kropinski. A *Shigella boydii* bacteriophage which resembles *Salmonella* phage ViI. *Virology Journal*, 8:242, 2011. ISSN 1743-422X. doi: 10.1186/1743-422X-8-242.

[15] Anonymous. *Serratia* septicaemia. *BMJ*, 4(5686):756–757, December 1969. ISSN 0959-8138. doi: 10.1136/bmj.4.5686.756.

[16] Anonymous. A natural solution to takle potential soft rot, 2012. URL http://www.branston.com/news/a-natural-solution-to-takle-potential-soft-rot/.

[17] Anonymous. 2017 Inspection of Growing Crops of Seed and Ware Potatoes: Survey for the Presence of *Dickeya* spp., 2017. URL http://www.gov.scot/Topics/farmingrural/Agriculture/plant/18273/PotatoHealthControls/PotatoQuarantineDiseases/Dickeya/DickeyaGrowingCropResults2017.

[18] María Antúnez-Lamas, Ezequiel Cabrera-Ordóñez, Emilia. López-Solanilla, Rosa Raposo, Oswaldo Trelles-Salazar, Andrés. Rodríguez-Moreno, and Pablo Rodríguez-Palenzuela. Role of motility and chemotaxis in the pathogenesis of *Dickeya dadantii*

3937 (ex *Erwinia chrysanthemi* 3937). *Microbiology*, 155(2):434–442, February 2009. ISSN 1350-0872. doi: 10.1099/mic.0.022244-0.

[19] Boel Åström and Berndt Gerhardson. Differential reactions of wheat and pea genotypes to root inoculation with growth-affecting rhizosphere bacteria. *Plant and Soil*, 109(2):263–269, July 1988. ISSN 0032-079X. doi: 10.1007/BF02202093.

[20] Anton Bankevich, Sergey Nurk, Dmitry Antipov, Alexey A. Gurevich, Mikhail Dvorkin, Alexander S. Kulikov, Valery M. Lesin, Sergey I. Nikolenko, Son Pham, Andrey D. Prjibelski, Alexey V. Pyshkin, Alexander V. Sirotkin, Nikolay Vyahhi, Glenn Tesler, Max a. Alekseyev, and Pavel a. Pevzner. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*, 19(5):455–477, 2012. ISSN 1066-5277. doi: 10.1089/cmb.2012.0021.

[21] Antony Barnett. Millions were in germ war tests. *The Guardian*, April 2002. URL https://www.theguardian.com/politics/2002/apr/21/uk.medicalscience.

[22] Ulrich Baxa, Stefan Steinbacher, Stefan Miller, Andrej Weintraub, Robert Huber, and Robert Seckler. Interactions of phage P22 tails with their cellular receptor, *Salmonella* O-antigen polysaccharide. *Biophysical Journal*, 71(4):2040–2048, October 1996. ISSN 00063495. doi: 10.1016/S0006-3495(96)79402-X.

[23] Gabriele Berg, Nicolle Roskot, Anette Steidle, Leo Eberl, Angela Zock, and Kornelia Smalla. Plant-Dependent Genotypic and Phenotypic Diversity of Antagonistic Rhizobacteria Isolated from Different *Verticillium* Host Plants. *Applied and Environmental Microbiology*, 68(7):3328–3338, July 2002. ISSN 0099-2240. doi: 10.1128/AEM.68.7.3328-3338.2002.

[24] Tanmay A. M. Bharat, Danguole Kureisaite-Ciziene, Gail G. Hardy, Ellen W. Yu, Jessica M. Devant, Wim J. H. Hagen, Yves V. Brun, John A. G. Briggs, and Jan Löwe. Structure of the hexagonal surface layer on *Caulobacter crescentus* cells. *Nature Microbiology*, 2(April):17059, April 2017. ISSN 2058-5276. doi: 10.1038/nmicrobiol.2017.59.

[25] Ambarish Biswas, Joshua N. Gagnon, Stan J.J. Brouns, Peter C. Fineran, and Chris M. Brown. CRISPRTarget. *RNA Biology*, 10(5):817–827, May 2013. ISSN 1547-6286. doi: 10.4161/rna.24046.

[26] Ambarish Biswas, Raymond H.J. Staals, Sergio E. Morales, Peter C. Fineran, and Chris M. Brown. CRISPRDetect: A flexible algorithm to define CRISPR arrays. *BMC Genomics*, 17(1):356, December 2016. ISSN 1471-2164. doi: 10.1186/s12864-016-2627-0.

[27] Tim R Blower, Terry J Evans, Rita Przybilski, Peter C Fineran, and George P C Salmond. Viral evasion of a bacterial suicide system by RNA-based molecular mimicry enables infectious altruism. *PLoS Genetics*, 8(10):e1003023, January 2012. ISSN 1553-7404. doi: 10.1371/journal.pgen.1003023.

[28] Anthony M. Bolger, Marc Lohse, and Bjoern Usadel. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15):2114–2120, 2014. ISSN 14602059. doi: 10.1093/bioinformatics/btu170.

[29] Mya Breitbart. Marine Viruses: Truth or Dare. *Annual Review of Marine Science*, 4(1):425–448, January 2012. ISSN 1941-1405. doi: 10.1146/annurev-marine-120709-142805.

[30] Harald Brüssow, Carlos Canchaya, and Wolf-Dietrich Hardt. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiology and Molecular Biology Reviews*, 68(3):560–602, September 2004. ISSN 1092-2172. doi: 10.1128/MMBR.68.3.560-602.2004.

[31] Colin Buttimer, Hanne Hendrix, Hugo Oliveira, Aidan Casey, Horst Neve, Olivia McAuliffe, R. Paul Ross, Colin Hill, Jean-Paul Noben, Jim O'Mahony, Rob Lavigne, and Aidan Coffey. Things Are Getting Hairy: Enterobacteria Bacteriophage vB_PcaM_CBB. *Frontiers in Microbiology*, 8:1–16, January 2017. ISSN 1664-302X. doi: 10.3389/fmicb.2017.00044.

[32] Colin Buttimer, Yannick Born, Alan Lucid, Martin J Loessner, Lars Fieseler, and Aidan Coffey. *Erwinia amylovora* phage vB_EamM_Y3 represents another lineage of hairy *Myoviridae*. *Research in Microbiology*, May 2018. ISSN 09232508. doi: 10.1016/j.resmic.2018.04.006.

[33] G Cahill, K Fraser, M J Kowalewska, D M Kenyon, and G S Saddler. Recent findings from the *Dickeya* survey and monitoring programme. *Proceedings Crop Protection in Northern Britain*, pages 171–176, 2010.

[34] Tim J Carver, Kim M Rutherford, Matthew Berriman, Marie-Adele Rajandream, Barclay G Barrell, and Julian Parkhill. ACT: the Artemis Comparison Tool. *Bioinformatics*, 21(16):3422–3, 2005. ISSN 1367-4803. doi: 10.1093/bioinformatics/bti553.

[35] Hyung Jin Cha, Jae-Hee Jeong, Catleya Rojviriya, and Yeon-Gil Kim. Structure of Putrescine Aminotransferase from *Escherichia coli* Provides Insights into the Substrate Specificity among Class III Aminotransferases. *PLoS ONE*, 9(11):e113212, November 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0113212.

[36] Sukumar Chakraborty and Adrian C. Newton. Climate change, plant diseases and food security: an overview. *Plant Pathology*, 60(1):2–14, February 2011. ISSN 00320862. doi: 10.1111/j.1365-3059.2010.02411.x.

[37] Benjamin K Chan, Stephen T Abedon, and Catherine Loc-Carrillo. Phage cocktails and the future of phage therapy. *Future Microbiology*, 8(6):769–783, June 2013. ISSN 1746-0913. doi: 10.2217/fmb.13.47.

[38] Tzu-Hao Chang, Hsi-Yuan Huang, Justin Bo-Kai Hsu, Shun-Long Weng, Jorng-Tzong Horng, and Hsien-Da Huang. An enhanced computational platform for investigating the roles of regulatory RNA and for identifying functional RNA motifs. *BMC Bioinformatics*, 14 Suppl 2:S4, 2013. ISSN 1471-2105. doi: 10.1186/1471-2105-14-S2-S4.

[39] Amy O. Charkowski. The Changing Face of Bacterial Soft-Rot Diseases. *Annual Review of Phytopathology*, 56(1), August 2018. ISSN 0066-4286. doi: 10.1146/annurev-phyto-080417-045906.

[40] Chen Chen, Patrick Bales, Julia Greenfield, Ryan D. Heselpoth, Daniel C. Nelson, and Osnat Herzberg. Crystal Structure of ORF210 from *E. coli* O157:H1 Phage CBA120 (TSP1), a Putative Tailspike Protein. *PLoS ONE*, 9(3):e93156, 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0093156.

[41] Sandra Chibani-Chennoufi, Anne Bruttin, M.-L. Dillmann, and H. Brussow. Phage-Host Interaction: an Ecological Perspective. *Journal of Bacteriology*, 186(12):3677–3686, June 2004. ISSN 0021-9193. doi: 10.1128/JB.186.12.3677-3686.2004.

[42] Francesca R. Cianfanelli, Juliana Alcoforado Diniz, Manman Guo, Virginia De Cesare, Matthias Trost, and Sarah J. Coulthurst. VgrG and PAAR Proteins Define Distinct Versions of a Functional Type VI Secretion System. *PLoS Pathogens*, 12(6):e1005735, June 2016. ISSN 1553-7374. doi: 10.1371/journal.ppat.1005735.

[43] Anna Colavecchio, Brigitte Cadieux, Amanda Lo, and Lawrence D. Goodridge. Bacteriophages Contribute to the Spread of Antibiotic Resistance Genes among Foodborne Pathogens of the Enterobacteriaceae Family – A Review. *Frontiers in Microbiology*, 8:1–13, June 2017. ISSN 1664-302X. doi: 10.3389/fmicb.2017.01108.

[44] Guy Condemine, Arnaud Castillo, Fabrice Passeri, and Corine Enard. The PecT Repressor Coregulates Synthesis of Exopolysaccharides and Virulence Factors in *Erwinia chrysanthemi*. *Molecular Plant-Microbe Interactions*, 12(1):45–52, January 1999. ISSN 0894-0282. doi: 10.1094/MPMI.1999.12.1.45.

[45] Sarah J. Coulthurst, Neil R. Williamson, Abigail K. P. Harris, David R. Spring, and George P. C. Salmond. Metabolic and regulatory engineering of *Serratia marcescens*: mimicking phage-mediated horizontal acquisition of antibiotic biosynthesis and quorum-sensing capacities. *Microbiology*, 152(7):1899–1911, July 2006. ISSN 1350-0872. doi: 10.1099/mic.0.28803-0.

[46] Diane Cuppels and Arthur Kelman. Evaluation of selective media for isolation of soft-rot bacteria from soil and plant tissues. *Phytopathology*, 64:468–475, 1973.

[47] Robert Czajkowski, M. C. M. Pérombelon, J. A. van Veen, and Jean Martin van der Wolf. Control of blackleg and tuber soft rot of potato caused by *Pectobacterium* and *Dickeya* species: a review. *Plant Pathology*, 60(6):999–1013, December 2011. ISSN 00320862. doi: 10.1111/j.1365-3059.2011.02470.x.

[48] Robert Czajkowski, Zofia Ozymko, and Ewa Lojkowska. Isolation and characterization of novel soilborne lytic bacteriophages infecting *Dickeya* spp. biovar 3 (‘ *D. solani* ’). *Plant Pathology*, 63(4):758–772, August 2014. ISSN 00320862. doi: 10.1111/ppa.12157.

[49] Robert Czajkowski, Zofia Ozymko, Szymon Zwirowski, and Ewa Lojkowska. Complete genome sequence of a broad-host-range lytic *Dickeya* spp. bacteriophage phiD5. *Archives of Virology*, 159(11):3153–3155, November 2014. ISSN 0304-8608. doi: 10.1007/s00705-014-2170-8.

[50] Robert Czajkowski, Zofia Ozymko, Victor de Jager, Joanna Siwinska, Anna Smolarska, Adam Ossowicki, Magdalena Narajczyk, and Ewa Lojkowska. 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(3):e0119812, January 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0119812.

[51] Robert Czajkowski, Zofia Ozymko, Joanna Siwinska, Adam Ossowicki, Victor de Jager, Magdalena Narajczyk, and Ewa Lojkowska. The complete genome, structural proteome, comparative genomics and phylogenetic analysis of a broad host lytic bacteriophage phiD3 infecting pectinolytic *Dickeya* spp. *Standards in Genomic Sciences*, 10(1):68, 2015. ISSN 1944-3277. doi: 10.1186/s40793-015-0068-z.

[52] Robert Czajkowski, M.C.M. Pérombelon, Sylwia Jafra, Ewa Lojkowska, Marta Potrykus, Jean Martin van der Wolf, and Wojciech Sledz. Detection, identification and differentiation of *Pectobacterium* and *Dickeya* species causing potato blackleg and tuber soft rot: a review. *Annals of Applied Biology*, 166(1):18–38, January 2015. ISSN 00034746. doi: 10.1111/aab.12166.

[53] Robert Czajkowski, Anna Smolarska, and Zofia Ozymko. The viability of lytic bacteriophage ΦD5 in potato-associated environments and its effect on *Dickeya solani* in potato (*Solanum tuberosum* L.) plants. *PLoS ONE*, 12(8):e0183200, August 2017. ISSN 1932-6203. doi: 10.1371/journal.pone.0183200.

[54] Andrew Day, Jiyoon Ahn, Xinzhe Fang, and George P C Salmond. Environmental Bacteriophages of the Emerging Enterobacterial Phytopathogen, *Dickeya solani*, Show Genomic Conservation and Capacity for Horizontal Gene Transfer between Their Bacterial Hosts. *Frontiers in Microbiology*, 8:1654, August 2017. ISSN 1664-302X. doi: 10.3389/fmicb.2017.01654.

[55] Andrew Day, Jiyoon Ahn, and George P. C. Salmond. Jumbo Bacteriophages Are Represented Within an Increasing Diversity of Environmental Viruses Infecting the Emerging Phytopathogen, *Dickeya solani*. *Frontiers in Microbiology*, 9:1–15, September 2018. ISSN 1664-302X. doi: 10.3389/fmicb.2018.02169.

[56] Patrick A. de Jonge, Franklin L. Nobrega, Stan J.J. Brouns, and Bas E. Dutilh. Molecular and Evolutionary Determinants of Bacteriophage Host Range. *Trends in Microbiology*, pages 1–13, September 2018. ISSN 0966842X. doi: 10.1016/j.tim.2018.08.006.

[57] David de Vleesschauwer and Monica Höfte. Using *Serratia plymuthica* to control fungal pathogens of plants. *CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources*, 2(046), September 2007. ISSN 17498848. doi: 10.1079/PAVSNNR20072046.

[58] Gaëlle Demarre, Anne-Marie Guérout, Chiho Matsumoto-Mashimo, Dean A. Rowe-Magnus, Philippe Marlière, and Didier Mazel. A new family of mobilizable suicide plasmids based on broad host range R388 plasmid (IncW) and RP4 plasmid (IncPα) conjugative machineries and their cognate *Escherichia coli* host strains. *Research in Microbiology*, 156(2):245–255, March 2005. ISSN 09232508. doi: 10.1016/j.resmic.2004.09.007.

[59] ECDC. Annual Epidemiological Report 2016 – Healthcare-associated infections acquired in intensive care units. Technical Report May, European Centre for Disease

Prevention and Control, 2016. URL https://ecdc.europa.eu/en/publications-data/healthcare-associated-infections-intensive-care-units-annual-epidemiological.

[60] David R. Edgell, Ewan A. Gibb, and Marlene Belfort. Mobile DNA elements in T4 and related phages. *Virology Journal*, 7(1):290, 2010. ISSN 1743-422X. doi: 10.1186/1743-422X-7-290.

[61] Anneke Engering, Lenny Hogerwerf, and Jan Slingenbergh. Pathogen–host–environment interplay and disease emergence. *Emerging Microbes & Infections*, 2(2):e5–e5, February 2013. ISSN 2222-1751. doi: 10.1038/emi.2013.5.

[62] Ian N D Esplin, Jordan A Berg, Ruchira Sharma, Robert C Allen, Daniel K Arens, Cody R Ashcroft, Shannon R Bairett, Nolan J Beatty, Madeline Bickmore, Travis J Bloomfield, T Scott Brady, Rachel N Bybee, John L Carter, Minsey C Choi, Steven Duncan, Christopher P Fajardo, Brayden B Foy, David A Fuhriman, Paul D Gibby, Savannah E Grossarth, Kala Harbaugh, Natalie Harris, Jared A Hilton, Emily Hurst, Jonathan R Hyde, Kayleigh Ingersoll, Caitlin M Jacobson, Brady D James, Todd M Jarvis, Daniella Jaen-Anieves, Garrett L Jensen, Bradley K Knabe, Jared L Kruger, Bryan D Merrill, Jenny A Pape, Ashley M. Payne Anderson, David E Payne, Malia D Peck, Samuel V Pollock, Micah J Putnam, Ethan K Ransom, Devin B Ririe, David M Robinson, Spencer L Rogers, Kerri A Russell, Jonathan E Schoenhals, Christopher A Shurtleff, Austin R Simister, Hunter G Smith, Michael B Stephenson, Lyndsay A Staley, Jason M Stettler, Mallorie L Stratton, Olivia B Tateoka, P J Tatlow, Alexander S Taylor, Suzanne E Thompson, Michelle H Townsend, Trever L Thurgood, Brittian K Usher, Kiara V Whitley, Andrew T. Ward, Megan E H Ward, Charles J Webb, Trevor M Wienclaw, Taryn L Williamson, Michael J Wells, Cole K Wright, Donald P Breakwell, Sandra Hope, and Julianne H Grose. Genome Sequences of 19 Novel *Erwinia amylovora* Bacteriophages. *Genome Announcements*, 5(46):e00931–17, November 2017. ISSN 2169-8287. doi: 10.1128/genomeA.00931-17.

[63] D Expert. WITHHOLDING AND EXCHANGING IRON: Interactions Between *Erwinia* spp. and Their Plant Hosts. *Annual Review of Phytopathology*, 37(1):307–334, September 1999. ISSN 0066-4286. doi: 10.1146/annurev.phyto.37.1.307.

[64] Robert P. Fagan and Neil F. Fairweather. Biogenesis and functions of bacterial S-layers. *Nature Reviews Microbiology*, 12(3):211–222, March 2014. ISSN 1740-1526. doi: 10.1038/nrmicro3213.

[65] Xinzhe Fang. *Characterisation of new enterobacterial phages and their responses to the ToxIN$_{Pa}$ system*. Master's, University of Cambridge, 2014.

[66] FAO. *FAO Statistical Pocketbook 2015*. United Nations, 2015. ISBN 9789251088029. doi: 978-92-5-108802-9. URL http://www.fao.org/3/a-i4691e.pdf.

[67] Sofia Fernandes and Carlos São-José. Enzymes and Mechanisms Employed by Tailed Bacteriophages to Breach the Bacterial Cell Barriers. *Viruses*, 10(8):396, July 2018. ISSN 1999-4915. doi: 10.3390/v10080396.

[68] Andrey A. Filippov, Kirill V. Sergueev, Yunxiu He, Xiao-Zhe Huang, Bryan T. Gnade, Allen J. Mueller, Carmen M. Fernandez-Prada, and Mikeljon P. Nikolich.

Bacteriophage-Resistant Mutants in *Yersinia pestis*: Identification of Phage Receptors and Attenuation for Mice. *PLoS ONE*, 6(9):e25486, September 2011. ISSN 1932-6203. doi: 10.1371/journal.pone.0025486.

[69] Andrei Fokine and Michael G Rossmann. Molecular architecture of tailed double-stranded DNA phages. *Bacteriophage*, 4(2):e28281, April 2014. ISSN 2159-7081. doi: 10.4161/bact.28281.

[70] Sylvain Gandon. Why Be Temperate: Lessons from Bacteriophage λ. *Trends in Microbiology*, 24(5):356–365, May 2016. ISSN 0966842X. doi: 10.1016/j.tim.2016.02.008.

[71] Eugene R. L. Gaughran. DIVISION OF MICROBIOLOGY: FROM SUPERSTITION TO SCIENCE: THE HISTORY OF A BACTERIUM*. *Transactions of the New York Academy of Sciences*, 31(1 Series II):3–24, January 1969. ISSN 00287113. doi: 10.1111/j.2164-0947.1969.tb02887.x.

[72] Malgorzata Golanowska, Marta Potrykus, Agata Motyka-Pomagruk, Michal Kabza, Giovanni Bacci, Marco Galardini, Marco Bazzicalupo, Izabela Makalowska, Kornelia Smalla, Alessio Mengoni, Nicole Hugouvieux-Cotte-Pattat, and Ewa Lojkowska. Comparison of Highly and Weakly Virulent *Dickeya solani* Strains, With a View on the Pangenome and Panregulon of This Species. *Frontiers in Microbiology*, 9:1940, August 2018. ISSN 1664-302X. doi: 10.3389/fmicb.2018.01940.

[73] Vladimir Gorshkov, Bakhtiyar Islamov, Polina Mikshina, Olga Petrova, Gennady Burygin, Elena Sigida, Alexander Shashkov, Amina Daminova, Marina Ageeva, Bulat Idiyatullin, Vadim Salnikov, Yuriy Zuev, Tatyana Gorshkova, and Yuri Gogolev. *Pectobacterium atrosepticum* exopolysaccharides: identification, molecular structure, formation under stress and *in planta* conditions. *Glycobiology*, 27(11):1016–1026, November 2017. ISSN 0959-6658. doi: 10.1093/glycob/cwx069.

[74] Jason Gunawan, Dave Simard, Michel Gilbert, Andrew L. Lovering, Warren W. Wakarchuk, Martin E. Tanner, and Natalie C J Strynadka. Structural and Mechanistic Analysis of Sialic Acid Synthase NeuB from *Neisseria meningitidis* in Complex with Mn^{2+}, Phosphoenolpyruvate, and N-Acetylmannosaminitol. *Journal of Biological Chemistry*, 280(5):3555–3563, February 2005. ISSN 0021-9258. doi: 10.1074/jbc.M411942200.

[75] Florence Hommais, Christine Oger-Desfeux, Frédérique Van Gijsegem, Sandra Castang, Sandrine Ligori, Dominique Expert, William Nasser, and Sylvie Reverchon. PecS Is a Global Regulator of the Symptomatic Phase in the Phytopathogenic Bacterium *Erwinia chrysanthemi* 3937. *Journal of Bacteriology*, 190(22):7508–7522, November 2008. ISSN 0021-9193. doi: 10.1128/JB.00553-08.

[76] Steven P T Hooton, Andrew R Timms, Joanna Rowsell, Ray Wilson, and Ian F Connerton. *Salmonella Typhimurium*-specific bacteriophage ΦSH19 and the origins of species specificity in the Vi01-like phage family. *Virology Journal*, 8:498, 2011. ISSN 1743-422X. doi: 10.1186/1743-422X-8-498.

[77] Chun-Ru Hsu, Tzu-Lung Lin, Yi-Jiun Pan, Pei-Fang Hsieh, and Jin-Town Wang. Isolation of a Bacteriophage Specific for a New Capsular Type of *Klebsiella pneumoniae*

and Characterization of Its Polysaccharide Depolymerase. *PLoS ONE*, 8(8):e70092, August 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0070092.

[78] Martin Hugh-Jones. Wickham Steed and German biological warfare research. *Intelligence and National Security*, 7:379–402, 1992. doi: 10.1080/02684529208432176.

[79] Nicole Hugouvieux-Cotte-Pattat, Guy Condemine, and Vladimir E. Shevchik. Bacterial pectate lyases, structural and functional diversity. *Environmental Microbiology Reports*, 6(5):427–440, October 2014. ISSN 17582229. doi: 10.1111/1758-2229.12166.

[80] Atsushi Iguchi, Yutaka Nagaya, Elizabeth Pradel, Tadasuke Ooka, Yoshitoshi Ogura, Keisuke Katsura, Ken Kurokawa, Kenshiro Oshima, Masahira Hattori, Julian Parkhill, Mohamed Sebaihia, Sarah J. Coulthurst, Naomasa Gotoh, Nicholas R. Thomson, Jonathan J. Ewbank, and Tetsuya Hayashi. Genome Evolution and Plasticity of *Serratia marcescens*, an Important Multidrug-Resistant Nosocomial Pathogen. *Genome Biology and Evolution*, 6(8):2096–2110, August 2014. ISSN 1759-6653. doi: 10.1093/gbe/evu160.

[81] Courtney E. Jahn, Dija A. Selimi, Jeri D. Barak, and Amy O. Charkowski. The *Dickeya dadantii* biofilm matrix consists of cellulose nanofibres, and is an emergent property dependent upon the type III secretion system and the cellulose synthesis operon. *Microbiology*, 157(10):2733–2744, October 2011. ISSN 1350-0872. doi: 10.1099/mic.0.051003-0.

[82] Jaap D. Janse. Characterization and Classification of *Erwinia chrysanthemi* Strains from Several Hosts in The Netherlands. *Phytopathology*, 78(6):800, 1988. ISSN 0031949X. doi: 10.1094/Phyto-78-800.

[83] Fuguo Jiang and Jennifer A Doudna. CRISPR–Cas9 Structures and Mechanisms. *Annual Review of Biophysics*, 46(1):505–529, May 2017. ISSN 1936-122X. doi: 10.1146/annurev-biophys-062215-010822.

[84] Anastasia Kabanova, Mikhail Shneider, Eugenia Bugaeva, Vo Thi Ngoc Ha, Kirill Miroshnikov, Aleksei Korzhenkov, Eugene Kulikov, Stepan Toschakov, Alexander Ignatov, and Konstantin Miroshnikov. Genomic characteristics of vB_PpaP_PP74, a T7-like *Autographivirinae* bacteriophage infecting a potato pathogen of the newly proposed species *Pectobacterium* parmentieri. *Archives of Virology*, 163(6):1691–1694, June 2018. ISSN 0304-8608. doi: 10.1007/s00705-018-3766-1.

[85] Mohammadali Khan Mirzaei and Anders S. Nilsson. Isolation of Phages for Phage Therapy: A Comparison of Spot Tests and Efficiency of Plating Analyses for Determination of Host Range and Efficacy. *PLoS ONE*, 10(3):e0118557, January 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0118557.

[86] Eun Jin Kim, Dokyung Lee, Se Hoon Moon, Chan Hee Lee, and Dong Wook Kim. CTX Prophages in *Vibrio cholerae* O1 Strains. *Journal of Microbiology and Biotechnology*, 24(6):725–731, June 2014. ISSN 1017-7825. doi: 10.4014/jmb.1403.03063.

[87] Önder Kimyon, Theerthankar Das, Amaye I. Ibugo, Samuel K. Kutty, Kitty K. Ho, Naresh Tebben, Januaryand Kumar, and Mike Manefield. *Serratia* Secondary Metabolite Prodigiosin Inhibits *Pseudomonas aeruginosa* Biofilm Development by Producing

Reactive Oxygen Species that Damage Biological Molecules. *Frontiers in Microbiology*, 7:1–15, June 2016. ISSN 1664-302X. doi: 10.3389/fmicb.2016.00972.

[88] Qingke Kong, Jiseon Yang, Qing Liu, Praveen Alamuri, Kenneth L. Roland, and Roy Curtiss. Effect of Deletion of Genes Involved in Lipopolysaccharide Core and O-Antigen Synthesis on Virulence and Immunogenicity of *Salmonella enterica* Serovar Typhimurium. *Infection and Immunity*, 79(10):4227–4239, October 2011. ISSN 0019-9567. doi: 10.1128/IAI.05398-11.

[89] Eugene V. Koonin, Kira S. Makarova, and Yuri I. Wolf. Evolutionary Genomics of Defense Systems in Archaea and Bacteria. *Annual Review of Microbiology*, 71(1):233–261, September 2017. ISSN 0066-4227. doi: 10.1146/annurev-micro-090816-093830.

[90] Martin Krzywinski, Jacqueline Schein, Inanç Birol, Joseph Connors, Randy Gascoyne, Doug Horsman, Steven J Jones, and Marco A Marra. Circos: An information aesthetic for comparative genomics. *Genome Research*, 19(9):1639–1645, September 2009. ISSN 10889051. doi: 10.1101/gr.092759.109.

[91] Jens Kuhn, Andrew Kropinski, Hany Anany, Igor Tolstoy, Elizabeth Kutter, and Evelien Adriaenssens. To create a new bacteriophage family, *Ackermannviridae*, containing two (2) new subfamilies including four (4) genera. *ICTV TaxoProp*, 2017. doi: 10.13140/RG.2.2.29173.88800.

[92] Sudhir Kumar, Glen Stecher, and Koichiro Tamura. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, 33(7):msw054, 2016. ISSN 1537-1719. doi: 10.1093/molbev/msw054.

[93] Alexander L. Lagonenko, Olga Sadovskaya, Leonid N. Valentovich, and Anatoly N. Evtushenkov. Characterization of a new ViI-like *Erwinia amylovora* bacteriophage phiEa2809. *FEMS Microbiology Letters*, 362(7):fnv031–fnv031, April 2015. ISSN 1574-6968. doi: 10.1093/femsle/fnv031.

[94] Jaana Laurila, Virpi Ahola, Ari Lehtinen, Tiina Joutsjoki, Asko Hannukkala, Anne Rahkonen, and Minna Pirhonen. Characterization of *Dickeya* strains isolated from potato and river water samples in Finland. *European Journal of Plant Pathology*, 122 (2):213–225, 2008. ISSN 09291873. doi: 10.1007/s10658-008-9274-5.

[95] Jeffrey Lawrence. DNA Master, 2012. URL http://cobamide2.bio.pitt.edu/.

[96] Martina Leibig, Manuel Liebeke, Diana Mader, Michael Lalk, Andreas Peschel, and F. Gotz. Pyruvate Formate Lyase Acts as a Formate Supplier for Metabolic Processes during Anaerobiosis in *Staphylococcus aureus*. *Journal of Bacteriology*, 193(4): 952–962, February 2011. ISSN 0021-9193. doi: 10.1128/JB.01161-10.

[97] Jolanta J. Levenfors, Rikard Hedman, Christian Thaning, Berndt Gerhardson, and Christopher J. Welch. Broad-spectrum antifungal metabolites produced by the soil bacterium *Serratia plymuthica* A153. *Soil Biology and Biochemistry*, 36(4):677–685, April 2004. ISSN 00380717. doi: 10.1016/j.soilbio.2003.12.008.

[98] Heng Li. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint*, 00(00):3, 2013. ISSN 2169-8287. doi: arXiv:1303.3997[q-bio.GN].

[99] Derek M Lin, Britt Koskella, and Henry C Lin. Phage therapy: An alternative to antibiotics in the age of multi-drug resistance. *World Journal of Gastrointestinal Pharmacology and Therapeutics*, 8(3):162, 2017. ISSN 2150-5349. doi: 10.4292/wjgpt.v8.i3.162.

[100] Magdalen Lindeberg, George P. C. Salmond, and Alan Collmer. Complementation of deletion mutations in a cloned functional cluster of *Erwinia chrysanthemi out* genes with *Erwinia carotovora out* homologues reveals OutC and OutD as candidate gatekeepers of species-specific secretion of proteins via the type II pathway. *Molecular Microbiology*, 20(1):175–190, April 1996. ISSN 0950-382X. doi: 10.1111/j.1365-2958.1996.tb02499.x.

[101] Magdalen Lindeberg, Carol M. Boyd, Noel T. Keen, and Alan Collmer. External loops at the C terminus of *Erwinia chrysanthemi* pectate lyase C are required for species-specific secretion through the Out type II pathway. *Journal of Bacteriology*, 180(6):1431–1437, 1998. ISSN 00219193.

[102] Marzanna Łusiak-Szelachowska, Beata Weber-Dąbrowska, Ewa Jończyk-Matysiak, Renata Wojciechowska, and Andrzej Górski. Bacteriophages in the gastrointestinal tract and their implications. *Gut Pathogens*, 9(1):44, December 2017. ISSN 1757-4749. doi: 10.1186/s13099-017-0196-7.

[103] Xing Ma, Allison Schloop, Bryan Swingle, and Keith L. Perry. *Pectobacterium* and *Dickeya* Responsible for Potato Blackleg Disease in New York State in 2016. *Plant Disease*, 102(9):1834–1840, September 2018. ISSN 0191-2917. doi: 10.1094/PDIS-10-17-1595-RE.

[104] Steven D. Mahlen. *Serratia* Infections: from Military Experiments to Current Practice. *Clinical Microbiology Reviews*, 24(4):755–791, October 2011. ISSN 0893-8512. doi: 10.1128/CMR.00017-11.

[105] Miguel A Matilla and George P C Salmond. Complete genome sequence of *Serratia plymuthica* bacteriophage ΦMAM1. *Journal of Virology*, 86(24):13872–3, 2012. ISSN 1098-5514. doi: 10.1128/JVI.02702-12.

[106] Miguel A Matilla and George P C Salmond. Bacteriophage phiMAM1, a viunalikevirus, is a broad-host-range, high-efficiency generalized transducer that infects environmental and clinical isolates of the enterobacterial genera *Serratia* and *Kluyvera*. *Applied and Environmental Microbiology*, 80(20):6446–57, October 2014. ISSN 1098-5336. doi: 10.1128/AEM.01546-14.

[107] Miguel A Matilla, Xinzhe Fang, and George P C Salmond. Viunalikeviruses are environmentally common agents of horizontal gene transfer in pathogens and biocontrol bacteria. *The ISME Journal*, 8(10):2143–7, 2014. ISSN 1751-7370. doi: 10.1038/ismej.2014.150.

[108] Miguel A. Matilla, Zulema Udaondo, Tino Krell, and George P. C. Salmond. Genome Sequence of *Serratia marcescens* MSU97, a Plant-Associated Bacterium That Makes Multiple Antibiotics. *Genome Announcements*, 5(9):e01752–16, March 2017. ISSN 2169-8287. doi: 10.1128/genomeA.01752-16.

[109] Kenshi Matsushita, Jumpei Uchiyama, Shin-ichiro Kato, Takako Ujihara, Hiroshi Hoshiba, Shigeyoshi Sugihara, Asako Muraoka, Hiroshi Wakiguchi, and Shigenobu Matsuzaki. Morphological and genetic analysis of three bacteriophages of *Serratia marcescens* isolated from environmental water. *FEMS Microbiology Letters*, 291(2): 201–8, 2009. ISSN 1574-6968. doi: 10.1111/j.1574-6968.2008.01455.x.

[110] Jaclyn McCutcheon, Danielle Peters, and Jonathan Dennis. Identification and Characterization of Type IV Pili as the Cellular Receptor of Broad Host Range *Stenotrophomonas maltophilia* Bacteriophages DLP1 and DLP2. *Viruses*, 10(6):338, June 2018. ISSN 1999-4915. doi: 10.3390/v10060338.

[111] Patrick F. McDermott, Laura M. McMurry, Isabelle Podglajen, Joann L. Dzink-Fox, Thamarai Schneiders, Michael P. Draper, and Stuart B. Levy. The *marC* Gene of *Escherichia coli* Is Not Involved in Multiple Antibiotic Resistance. *Antimicrobial Agents and Chemotherapy*, 52(1):382–383, January 2008. ISSN 0066-4804. doi: 10.1128/AAC.00930-07.

[112] Stuart McNicholas, Elisabeth Potterton, Keith S. Wilson, and Martin E. M. Noble. Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallographica Section D Biological Crystallography*, 67(4):386–394, April 2011. ISSN 0907-4449. doi: 10.1107/S0907444911007281.

[113] C P Merlino. Bartolomeo Bizio's Letter to the most Eminent Priest, Angelo Bellani, Concerning the Phenomenon of the Red Colored Polenta [Translated from the Italian]. *Journal of Bacteriology*, 9(6):527–43, 1924. ISSN 0021-9193.

[114] Andrew Millard. Bacteriophage Genomes, 2018. URL http://millardlab.org/bioinformatics/bacteriophage-genomes/.

[115] Sara Mohan, Sathiyamoorthy Meiyalaghan, Julie M. Latimer, Michelle L. Gatehouse, Katrina S. Monaghan, Bhanupratap R. Vanga, Andrew R. Pitman, E. Eirian Jones, Anthony J. Conner, and Jeanne M. E. Jacobs. GSL2 over-expression confers resistance to *Pectobacterium atrosepticum* in potato. *Theoretical and Applied Genetics*, 127(3): 677–689, March 2014. ISSN 0040-5752. doi: 10.1007/s00122-013-2250-2.

[116] Francisco J. M. Mojica, Guadalupe Juez, and Francisco Rodriguez-Valera. Transcription at different salinities of *Haloferax mediterranei* sequences adjacent to partially modified PstI sites. *Molecular Microbiology*, 9(3):613–621, August 1993. ISSN 0950-382X. doi: 10.1111/j.1365-2958.1993.tb01721.x.

[117] Rita Monson, Debra S Smith, Miguel A Matilla, Kevin Roberts, Elizabeth Richardson, Alison Drew, Neil Williamson, Josh Ramsay, Martin Welch, and George P C Salmond. A Plasmid-Transposon Hybrid Mutagenesis System Effective in a Broad Range of Enterobacteria. *Frontiers in Microbiology*, 6:1442, December 2015. ISSN 1664-302X. doi: 10.3389/fmicb.2015.01442.

[118] Agata Motyka, Sabina Zoledowska, Wojciech Sledz, and Ewa Lojkowska. Molecular methods as tools to control plant diseases caused by *Dickeya* and *Pectobacterium* spp: A minireview. *New Biotechnology*, 39(October 2016):181–189, October 2017. ISSN 18716784. doi: 10.1016/j.nbt.2017.08.010.

[119] James Murphy, Jochen Klumpp, Jennifer Mahony, Mary O'Connell-Motherway, Arjen Nauta, and Douwe van Sinderen. Methyltransferases acquired by lactococcal 936-type phage provide protection against restriction endonuclease activity. *BMC Genomics*, 15(1):831, 2014. ISSN 1471-2164. doi: 10.1186/1471-2164-15-831.

[120] William Nasser, Corinne Dorel, Julien Wawrzyniak, Frédérique Van Gijsegem, Marie-Christine Groleau, Eric Déziel, and Sylvie Reverchon. Vfm a new quorum sensing system controls the virulence of *Dickeya dadantii*. *Environmental Microbiology*, 15 (3):865–880, March 2013. ISSN 14622912. doi: 10.1111/1462-2920.12049.

[121] Anders S. Nilsson. Phage therapy—constraints and possibilities. *Upsala Journal of Medical Sciences*, 119(2):192–198, May 2014. ISSN 0300-9734. doi: 10.3109/03009734.2014.902878.

[122] Franklin L. Nobrega, Marnix Vlot, Patrick A. de Jonge, Lisa L. Dreesens, Hubertus J. E. Beaumont, Rob Lavigne, Bas E. Dutilh, and Stan J. J. Brouns. Targeting mechanisms of tailed bacteriophages. *Nature Reviews Microbiology*, August 2018. ISSN 1740-1526. doi: 10.1038/s41579-018-0070-8.

[123] Erich-Christian Oerke. Crop losses to pests. *The Journal of Agricultural Science*, 144 (01):31, February 2006. ISSN 0021-8596. doi: 10.1017/S0021859605005708.

[124] Karolina Ossowska, Małgorzata Czerwicka, Wojciech Sledz, Sabina Zoledowska, Agata Motyka, Małgorzata Golanowska, Guy Condemine, Ewa Lojkowska, and Zbigniew Kaczyński. The uniform structure of O-polysaccharides isolated from *Dickeya solani* strains of different origin. *Carbohydrate Research*, 445:40–43, June 2017. ISSN 00086215. doi: 10.1016/j.carres.2017.04.001.

[125] Michelle K. Paczosa and Joan Mecsas. *Klebsiella pneumoniae*: Going on the Offense with a Strong Defense. *Microbiology and Molecular Biology Reviews*, 80(3):629–661, September 2016. ISSN 1092-2172. doi: 10.1128/MMBR.00078-15.

[126] Neil Parkinson, David Stead, Janice Bew, John Heeney, Leah Tsror, and John Elphinstone. *Dickeya* species relatedness and clade structure determined by comparison of *recA* sequences. *International Journal of Systematic and Evolutionary Microbiology*, 59(10):2388–2393, 2009. ISSN 14665026. doi: 10.1099/ijs.0.009258-0.

[127] Neil Parkinson, Paul DeVos, Minna Pirhonen, and John Elphinstone. *Dickeya aquatica* sp. nov., isolated from waterways. *International Journal of Systematic and Evolutionary Microbiology*, 64(Pt 7):2264–2266, July 2014. ISSN 1466-5026. doi: 10.1099/ijs.0.058693-0.

[128] Neil Parkinson, Leighton Pritchard, Ruth Bryant, Ian Toth, and John Elphinstone. Epidemiology of *Dickeya dianthicola* and *Dickeya solani* in ornamental hosts and potato studied using variable number tandem repeat analysis. *European Journal of Plant Pathology*, 141(1):63–70, 2014. ISSN 15738469. doi: 10.1007/s10658-014-0523-5.

[129] Eric Pelfrene, Elsa Willebrand, Ana Cavaleiro Sanches, Zigmars Sebris, and Marco Cavaleri. Bacteriophage therapy: a regulatory perspective. *Journal of Antimicrobial Chemotherapy*, 71(8):2071–2074, August 2016. ISSN 0305-7453. doi: 10.1093/jac/dkw083.

[130] Pest Risk Analysis Unit. Import Risk Analysis : ' *Dickeya solani* '. Technical report, Ministry of Agriculture & Fisheries, Kingston, 2010. URL http://www.moa.gov.jm/PlantHealth/.

[131] Vasiliy M Petrov, Swarnamala Ratnayaka, James M Nolan, Eric S Miller, and Jim D Karam. Genomes of the T4-related bacteriophages as windows on microbial genome evolution. *Virology Journal*, 7:292, 2010. ISSN 1743-422X. doi: 10.1186/1743-422X-7-292.

[132] Ulrike Pfreundt, Dina Spungin, Shengwei Hou, Björn Voß, Ilana Berman-Frank, and Wolfgang R. Hess. Genome of a giant bacteriophage from a decaying *Trichodesmium* bloom. *Marine Genomics*, 33:21–25, June 2017. ISSN 18747787. doi: 10.1016/j.margen.2017.02.001.

[133] Derek Pickard, Robert a. Kingsley, Christine Hale, Keith Turner, Karthikeyan Sivaraman, Michael Wetter, Gemma Langridge, and Gordon Dougan. A genomewide mutagenesis screen identifies multiple genes contributing to Vi capsular expression in *Salmonella enterica* serovar Typhi. *Journal of Bacteriology*, 195(6):1320–1326, 2013. ISSN 00219193. doi: 10.1128/JB.01632-12.

[134] Jean-Paul Pirnay, Gilbert Verbeken, Pieter-Jan Ceyssens, Isabelle Huys, Daniel De Vos, Charlotte Ameloot, and Alan Fauconnier. The Magistral Phage. *Viruses*, 10(2):64, February 2018. ISSN 1999-4915. doi: 10.3390/v10020064.

[135] Leighton Pritchard, Sonia Humphris, Steve Baeyen, Martine Maes, Johan Van Vaerenbergh, John Elphinstone, Gerry Saddler, and Ian Toth. Draft Genome Sequences of Four *Dickeya dianthicola* and Four *Dickeya solani* Strains. *Genome Announcements*, 1(4):e00087–12, 2013. ISSN 2169-8287. doi: 10.1128/genomeA.00087-12.

[136] Leighton Pritchard, Sonia Humphris, Gerry S Saddler, John G Elphinstone, Minna Pirhonen, and Ian K Toth. Draft genome sequences of 17 isolates of the plant pathogenic bacterium *Dickeya*. *Genome Announcements*, 1(6):e00978–13, November 2013. ISSN 2169-8287. doi: 10.1128/genomeA.00978-13.

[137] Leighton Pritchard, Sonia Humphris, Gerry S. Saddler, Neil M. Parkinson, V. Bertrand, John G. Elphinstone, and Ian K. Toth. Detection of phytopathogens of the genus *Dickeya* using a PCR primer prediction pipeline for draft bacterial genome sequences. *Plant Pathology*, 62(3):587–596, June 2013. ISSN 00320862. doi: 10.1111/j.1365-3059.2012.02678.x.

[138] Leighton Pritchard, Rachel H. Glover, Sonia Humphris, John G. Elphinstone, and Ian K. Toth. Genomics and taxonomy in diagnostics for food security: Soft-rotting enterobacterial plant pathogens. *Analytical Methods*, 8(1):12–24, 2016. ISSN 17599679. doi: 10.1039/c5ay02550h.

[139] D V Rakhuba, E I Kolomiets, E Szwajcer Dey, and G I Novik. Bacteriophage Receptors, Mechanisms of Phage Adsorption and Penetration into Host Cell. *Polish Journal of Microbiology*, 59(3):145–155, 2010.

[140] M. Regue, C. Fabregat, and M. Vinas. A generalized transducing bacteriophage for *Serratia marcescens*. *Research in Microbiology*, 142(1):23–27, 1991. ISSN 09232508. doi: 10.1016/0923-2508(91)90093-P.

[141] Sylvie Reverchon, Georgi Muskhelisvili, and William Nasser. Virulence Program of a Bacterial Plant Pathogen: The *Dickeya* Model. In *Progress in Molecular Biology and Translational Science*, volume 142, pages 51–92. Elsevier Inc., 2016. ISBN 9780128093856. doi: 10.1016/bs.pmbts.2016.05.005.

[142] Simone Rhomberg, Christina Fuchsluger, Dubravko Rendić, Katharina Paschinger, Verena Jantsch, Paul Kosma, and Iain B.H. Wilson. Reconstitution in vitro of the GDP-fucose biosynthetic pathways of *Caenorhabditis elegans* and *Drosophila melanogaster*. *FEBS Journal*, 273(10):2244–2256, 2006. ISSN 1742464X. doi: 10.1111/j.1742-4658.2006.05239.x.

[143] Forest Rohwer and Rob Edwards. The Phage Proteomic Tree: a Genome-Based Taxonomy for Phage. *Journal of Bacteriology*, 184(16):4529–4535, August 2002. ISSN 0021-9193. doi: 10.1128/JB.184.16.4529-4535.2002.

[144] Clemencia M Rojas, Jong Hyun Ham, Wen-Ling Deng, Jeff J Doyle, and Alan Collmer. HecA, a member of a class of adhesins produced by diverse pathogenic bacteria, contributes to the attachment, aggregation, epidermal cell killing, and virulence phenotypes of *Erwinia chrysanthemi* EC16 on *Nicotiana clevelandii* seedlings. *Proceedings of the National Academy of Sciences of the United States of America*, 99 (20):13142–7, 2002. ISSN 0027-8424. doi: 10.1073/pnas.202358699.

[145] Alexa Ross, Samantha Ward, and Paul Hyman. More Is Better: Selecting for Broad Host Range Bacteriophages. *Frontiers in Microbiology*, 7:1–6, September 2016. ISSN 1664-302X. doi: 10.3389/fmicb.2016.01352.

[146] Simeon Rossmann, Merete Wiken Dees, Juliana Perminow, Richard Meadow, and May Bente Brurberg. Soft rot Enterobacteriaceae are carried by a large range of insect species in potato fields. *Applied and Environmental Microbiology*, 84(12), April 2018. ISSN 10985336. doi: 10.1128/AEM.00281-18.

[147] George P. C. Salmond and Peter C. Fineran. A century of the phage: past, present and future. *Nature Reviews Microbiology*, 13(12):777–786, 2015. ISSN 1740-1526. doi: 10.1038/nrmicro3564.

[148] Jiří Šalplachta, Anna Kubesová, Jaroslav Horký, Hana Matoušková, Marie Tesařová, and Marie Horká. Characterization of *Dickeya* and *Pectobacterium* species by capillary electrophoretic techniques and MALDI-TOF MS. *Analytical and Bioanalytical Chemistry*, 407(25):7625–7635, 2015. ISSN 16182650. doi: 10.1007/s00216-015-8920-y.

[149] Luisa Sandner-Miranda, Pablo Vinuesa, Alejandro Cravioto, and Rosario Morales-Espinosa. The Genomic Basis of Intrinsic and Acquired Antibiotic Resistance in the Genus *Serratia*. *Frontiers in Microbiology*, 9:1–16, May 2018. ISSN 1664-302X. doi: 10.3389/fmicb.2018.00828.

[150] S Sarfraz, K Riaz, S Oulghazi, J Cigna, M W Alam, Y Dessaux, and D Faure. First Report of *Dickeya dianthicola* Causing Blackleg Disease on Potato Plants in Pakistan. *Plant Disease*, pages PDIS–04–18–0551, August 2018. ISSN 0191-2917. doi: 10.1094/PDIS-04-18-0551-PDN.

[151] Mathias Schmelcher and Martin J Loessner. Application of bacteriophages for detection of foodborne pathogens. *Bacteriophage*, 4(2):e28137, April 2014. ISSN 2159-7081. doi: 10.4161/bact.28137.

[152] C a Schnaitman and J D Klena. Genetics of lipopolysaccharide biosynthesis in enteric bacteria. *Microbiological Reviews*, 57(3):655–682, 1993. ISSN 0146-0749.

[153] Alexandra Sifferlin. Scientists Are Using a New Weapon to Fight Drug-Resistant Bacteria—Viruses. *Time*, June 2018. URL http://time.com/5316516/phage-therapy-virus-san-diego/.

[154] Monika Sławiak, Jose R C M van Beckhoven, Adrianus G C L Speksnijder, Robert Czajkowski, Grzegorz Grabe, and Jan M. van der Wolf. Biochemical and genetical analysis reveal a new clade of biovar 3 *Dickeya* spp. strains isolated from potato in Europe. *European Journal of Plant Pathology*, 125(2):245–261, 2009. ISSN 09291873. doi: 10.1007/s10658-009-9479-2.

[155] V Solovyev and A Salamov. Automatic Annotation of Microbial Genomes and Metagenomic Sequences. In *Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies*, pages 61–78. Nova Science Publishers, 2011. ISBN 978-1-61668-682-6.

[156] Gary Strobel, Jia-Yao Li, Fumio Sugawara, James Harper, Hiroyuki Koshino, and W. M. Hess. Oocydin A, a chlorinated macrocyclic lactone with potent anti-oomycete activity from *Serratia marcescens*. *Microbiology*, 145(12):3557–3564, December 1999. ISSN 1350-0872. doi: 10.1099/00221287-145-12-3557.

[157] Jing Su, Dommo Timbely, Minmin Zhu, Xiaomei Hua, Biao Liu, Yanjun Pang, Hengguan Shen, Jinliang Qi, and Yonghua Yang. RfaB, a galactosyltransferase, contributes to the resistance to detergent and the virulence of *Salmonella enterica* serovar Enteritidis. *Medical Microbiology and Immunology*, 198(3):185–194, August 2009. ISSN 0300-8584. doi: 10.1007/s00430-009-0115-8.

[158] Radix Suharjo, Hiroyuki Sawada, and Yuichi Takikawa. Phylogenetic study of Japanese *Dickeya* spp. and development of new rapid identification methods using PCR–RFLP. *Journal of General Plant Pathology*, 80(3):237–254, May 2014. ISSN 1345-2630. doi: 10.1007/s10327-014-0511-9.

[159] William C Summers. Bacteriophage Therapy. *Annual Review of Microbiology*, 55(1): 437–451, October 2001. ISSN 0066-4227. doi: 10.1146/annurev.micro.55.1.437.

[160] Curtis A. Suttle. Marine viruses — major players in the global ecosystem. *Nature Reviews Microbiology*, 5(10):801–812, October 2007. ISSN 1740-1526. doi: 10.1038/nrmicro1750.

[161] Antonet Svircev, Dwayne Roach, and Alan Castle. Framing the Future with Bacteriophages in Agriculture. *Viruses*, 10(5):218, April 2018. ISSN 1999-4915. doi: 10.3390/v10050218.

[162] Nina N Sykilinda, Alexander A Bondar, Anna S Gorshkova, Lidia P Kurochkina, Eugene E Kulikov, Mikhail M Shneider, Vassily A Kadykov, Natalia V Solovjeva, Marsel R Kabilov, Vadim V Mesyanzhinov, Valentin V Vlassov, Valentin V Drukker, and Konstantin A Miroshnikov. Complete Genome Sequence of the Novel Giant *Pseudomonas* Phage PaBG. *Genome Announcements*, 2(1):e00929–13–e00929–13, January 2014. ISSN 2169-8287. doi: 10.1128/genomeA.00929-13.

[163] Masatada Tamakoshi, Aya Murakami, Motoki Sugisawa, Kenji Tsuneizumi, Shigeki Takeda, Toshihiko Saheki, Takashi Izumi, Toshihiko Akiba, Kaoru Mitsuoka, Hidehiro Toh, Atsushi Yamashita, Fumio Arisaka, Masahira Hattori, Tairo Oshima, and Akihiko Yamagishi. Genomic and proteomic characterization of the large *Myoviridae* bacteriophage ΦTMA of the extreme thermophile *Thermus thermophilus*. *Bacteriophage*, 1 (3):152–164, 2011. ISSN 2159-7073. doi: 10.4161/bact.1.3.16712.

[164] RStudio Team. RStudio: Integrated Development Environment for R, 2015. URL http://www.rstudio.com/.

[165] George Tetz and Victor Tetz. Bacteriophage infections of microbiota can lead to leaky gut in an experimental rodent model. *Gut Pathogens*, 8(1):33, December 2016. ISSN 1757-4749. doi: 10.1186/s13099-016-0109-1.

[166] Yanli Tian, Yuqiang Zhao, Xiaoli Yuan, Jianping Yi, Jiaqin Fan, Zhigang Xu, Baishi Hu, Solke H. De Boer, and Xiang Li. *Dickeya fangzhongdai* sp. nov., a plant-pathogenic bacterium isolated from pear trees (*Pyrus pyrifolia*). *International Journal of Systematic and Evolutionary Microbiology*, 66(8):2831–2835, August 2016. ISSN 1466-5026. doi: 10.1099/ijsem.0.001060.

[167] I. K. Toth, J. M. van der Wolf, G. Saddler, E. Lojkowska, V. Hélias, M. Pirhonen, L. Tsror Lahkim, and J. G. Elphinstone. *Dickeya* species: an emerging problem for potato production in Europe. *Plant Pathology*, 60(3):385–399, June 2011. ISSN 00320862. doi: 10.1111/j.1365-3059.2011.02427.x.

[168] I. K. Toth, G Cahill, J. G. Elphinstone, S Humphris, and S J Wale. An update on the Potato Council Scottish government-funded blackleg project - Year 2. *Proceedings Crop Protection in Northern Britain 2016*, pages 203–204, 2016.

[169] L. Tsror [Lahkim], S. Lebiush, O. Erlich, B. Ben-Daniel, and J. van der Wolf. First report of latent infection of *Cyperus rotundus* caused by a biovar 3 *Dickeya* sp.(Syn. *Erwinia chrysanthemi*) in Israel. *New Disease Reports*, 22(2):14, September 2010. ISSN 2044-0588. doi: 10.5197/j.2044-0588.2010.022.014.

[170] United States Congress Senate Committee on Human Resources Subcommittee on Health and Scientific Research. Biological testing involving human subjects by the Department of Defense, 1977 : hearings before the Subcommittee on Health and Scientific Research of the Committee on Human Resources, United States Senate, Ninety-fifth Congress, first session, on examina, 1977. URL https://www.ncbi.nlm.nih.gov/nlmcatalog/7800210.

[171] J. M. van der Wolf, B. H. de Haas, R. van Hoof, E. G. de Haan, and G. W. van den Bovenkamp. Development and evaluation of Taqman assays for the differentiation of

Dickeya (sub)species. *European Journal of Plant Pathology*, 138(4):695–709, April 2014. ISSN 0929-1873. doi: 10.1007/s10658-013-0343-z.

[172] Jan M. van der Wolf, Els H. Nijhuis, Malgorzata J. Kowalewska, Gerry S. Saddler, Neil Parkinson, John G. Elphinstone, Leighton Pritchard, Ian K. Toth, Ewa Lojkowska, Marta Potrykus, Malgorzata Waleron, Paul de Vos, Ilse Cleenwerck, Minna Pirhonen, Linda Garlant, Valérie Hélias, Joël F. Pothier, Valentin Pflüger, Brion Duffy, Leah Tsror, and Shula Manulis. *Dickeya solani* sp. nov., a pectinolytic plant-pathogenic bacterium isolated from potato (*Solanum tuberosum*). *International Journal of Systematic and Evolutionary Microbiology*, 64(PART 3):768–774, 2014. ISSN 14665026. doi: 10.1099/ijs.0.052944-0.

[173] Ayaka Washizaki, Tetsuro Yonesaki, and Yuichi Otsuka. Characterization of the interactions between *Escherichia coli* receptors, LPS and OmpC, and bacteriophage T4 long tail fibers. *MicrobiologyOpen*, 5(6):1003–1015, December 2016. ISSN 20458827. doi: 10.1002/mbo3.384.

[174] Markus G. Weinbauer. Ecology of prokaryotic viruses. *FEMS Microbiology Reviews*, 28(2):127–181, May 2004. ISSN 1574-6976. doi: 10.1016/j.femsre.2003.08.001.

[175] Michael Wetter, David Goulding, Derek Pickard, Michael Kowarik, Charles J. Waechter, Gordon Dougan, and Michael Wacker. Molecular Characterization of the *viaB* Locus Encoding the Biosynthetic Machinery for Vi Capsule Formation in *Salmonella* Typhi. *PLoS ONE*, 7(9):e45609, September 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0045609.

[176] Chris Whitfield. Biosynthesis and Assembly of Capsular Polysaccharides in *Escherichia coli*. *Annual Review of Biochemistry*, 75(1):39–68, June 2006. ISSN 0066-4154. doi: 10.1146/annurev.biochem.75.103004.142545.

[177] Chris Whitfield and Ian S. Roberts. Structure, assembly and regulation of expression of capsules in *Escherichia coli*. *Molecular Microbiology*, 31(5):1307–1319, 1999. ISSN 0950-382X. doi: 10.1046/j.1365-2958.1999.01276.x.

[178] Chris Whitfield and M. Stephen Trent. Biosynthesis and Export of Bacterial Lipopolysaccharides. *Annual Review of Biochemistry*, 83(1):99–128, June 2014. ISSN 0066-4154. doi: 10.1146/annurev-biochem-060713-035600.

[179] D Wright, A Bwye, M Banovic, J Baulch, C Wang, S Hair, N Hammond, B Coutts, and M Kehoe. First Report of *Dickeya dianthicola* in Potatoes in Australia. *Plant Disease*, pages PDIS–01–18–0094–PDN, August 2018. ISSN 0191-2917. doi: 10.1094/PDIS-01-18-0094-PDN.

[180] B Wu, Y Zhang, and P G Wang. Identification and characterization of GDP-D-mannose 4,6-dehydratase and GDP-L-fucose synthetase in a GDP-L-fucose biosynthetic gene cluster from *Helicobacter pylori*. *Biochemical and Biophysical Research Communications*, 285(2):364–71, July 2001. ISSN 0006-291X. doi: 10.1006/bbrc.2001.5137.

[181] Shaozhen Xing, Taping Ma, Xianglilan Zhang, Yong Huang, Zhiqiang Mi, Qiang Sun, Xiaoping An, Hang Fan, Shuhui Wu, Lin Wei, and Yigang Tong. First complete genome sequence of a virulent bacteriophage infecting the opportunistic pathogen

Serratia rubidaea. *Archives of Virology*, 162(7):2021–2028, July 2017. ISSN 0304-8608. doi: 10.1007/s00705-017-3300-x.

[182] Takashi Yamada, Souichi Satoh, Hiroki Ishikawa, Akiko Fujiwara, Takeru Kawasaki, Makoto Fujie, and Hiroyuki Ogata. A jumbo phage infecting the phytopathogen *Ralstonia solanacearum* defines a new lineage of the *Myoviridae* family. *Virology*, 398(1):135–147, March 2010. ISSN 00426822. doi: 10.1016/j.virol.2009.11.043.

[183] Ching-Hong Yang, Marina Gavilanes-Ruiz, Yasushi Okinaka, Regine Vedel, Isabelle Berthuy, Martine Boccara, Jason Wei-Ta Chen, Nicole T. Perna, and Noel T. Keen. *hrp* genes of *Erwinia chrysanthemi* 3937 Are Important Virulence Factors. *Molecular Plant-Microbe Interactions*, 15(5):472–480, May 2002. ISSN 0894-0282. doi: 10.1094/MPMI.2002.15.5.472.

[184] Jianyi Yang and Yang Zhang. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Research*, 43(W1):W174–W181, July 2015. ISSN 0305-1048. doi: 10.1093/nar/gkv342.

[185] M.-N. Yap, Ching-hong Yang, Jeri D Barak, Courtney E Jahn, and Amy O Charkowski. The *Erwinia chrysanthemi* Type III Secretion System Is Required for Multicellular Behavior. *Journal of Bacteriology*, 187(2):639–648, January 2005. ISSN 0021-9193. doi: 10.1128/JB.187.2.639-648.2005.

[186] Jarred Yasuhara-Bell, Glorimar Marrero, Mohammad Arif, Asoka de Silva, and Anne M. Alvarez. Development of a Loop-Mediated Isothermal Amplification Assay for the Detection of *Dickeya* spp. *Phytopathology*, 107(11):1339–1345, November 2017. ISSN 0031-949X. doi: 10.1094/PHYTO-04-17-0160-R.

[187] M. Yoder, N. Keen, and Frances Jurnak. New domain motif: the structure of pectate lyase C, a secreted plant virulence factor. *Science*, 260(5113):1503–1507, June 1993. ISSN 0036-8075. doi: 10.1126/science.8502994.

[188] Yihui Yuan and Meiying Gao. Jumbo bacteriophages: An overview. *Frontiers in Microbiology*, 8(MAR):1–9, 2017. ISSN 1664302X. doi: 10.3389/fmicb.2017.00403.

[189] Feng Zheng, Zhu-Qing Shao, Xina Hao, Qianqian Wu, Chaolong Li, Hongfen Hou, Dan Hu, Changjun Wang, and Xiuzhen Pan. Identification of oligopeptide-binding protein (OppA) and its role in the virulence of *Streptococcus suis* serotype 2. *Microbial Pathogenesis*, 118(December 2017):322–329, May 2018. ISSN 08824010. doi: 10.1016/j.micpath.2018.03.061.

Appendix One

Genome annotation tables

A.1 3M genome annotation table

ORF	Start	End	Annotation
1	3	2813	RIIA protein
2	2907	3272	hypothetical protein
3	3311	3472	hypothetical protein
4	3504	3761	hypothetical protein
5	3763	4311	hypothetical protein
6	4361	5446	hypothetical protein
7	5456	5704	hypothetical protein
8	5704	6450	hypothetical protein
9	6520	6777	hypothetical protein
10	6774	7913	hypothetical protein
11	8166	8375	hypothetical protein
12	8372	9256	hypothetical protein
13	9253	9558	hypothetical protein
14	9566	10312	putative deoxyribonucleotidase
15	10317	10616	hypothetical protein
16	10606	10941	hypothetical protein
17	11010	14003	putative DNA polymerase
18	14087	14665	hypothetical protein
19	15041	15337	hypothetical protein
20	15364	15705	hypothetical protein

21	15716	16198	hypothetical protein
22	16225	16680	hypothetical protein
23	16736	17614	hypothetical protein
24	17680	17943	hypothetical protein
25	17950	19260	hypothetical protein
26	19324	19491	hypothetical protein
27	19727	19652	tRNA-Met
28	20724	20635	tRNA-Tyr
29	22308	22895	hypothetical protein
30	23301	25073	putative baseplate wedge subunit
31	25060	25911	putative baseplate wedge subunit
32	25913	27175	putative tail spike protein
33	27224	30187	putative tail spike protein
34	30202	30396	Hypothetical protein
35	30487	31989	putative tail spike protein
36	31986	33602	hypothetical protein
37	33754	33861	hypothetical protein
38	34148	36007	hypothetical protein
39	36221	37891	putative tail spike protein head-binding protein
40	37893	38639	putative pectate lyase
41	38650	38937	hypothetical protein
42	38939	41389	hypothetical protein
43	41398	41583	hypothetical protein
44	41604	41882	hypothetical protein
45	41999	42709	hypothetical protein
46	42838	47685	putative virulence-associated VriC protein
47	47743	47970	putative structural protein
48	47963	48295	putative capsid protein
49	48285	49037	putative neck protein
50	49124	49768	putative neck protein
51	49765	50490	putative proximal tail sheath stabilisation protein
52	50490	51188	putative terminase DNA packaging enzyme small subunit
53	51169	53376	putative terminase DNA packaging enzyme large subunit
54	53419	55329	putative tail sheath protein

55	55383	55853	putative GIY-YIG homing endonuclease
56	55877	56410	putative tail tube protein
57	56481	58118	putative portal vertex protein
58	58167	58343	hypothetical protein
59	58355	58663	putative prohead core protein
60	58674	59339	putative prohead core protein protease
61	59386	60162	putative prohead core scaffold protein
62	60252	61577	putative major capsid protein
63	61683	62522	putative homing endonuclease
64	62573	62773	hypothetical protein
65	62770	63198	hypothetical protein
66	63201	63746	hypothetical protein
67	63831	64496	hypothetical protein
68	64498	64959	hypothetical protein
69	65040	65597	hypothetical protein
70	65642	65872	hypothetical protein
71	65921	66190	hypothetical protein
72	66187	66630	hypothetical protein
73	66641	66832	hypothetical protein
74	66871	67593	hypothetical protein
75	67623	67931	hypothetical protein
76	68321	68815	putative tail completion and sheath stabiliser protein
77	68825	69310	putative UvsY DNA repair/recombination protein
78	69315	70058	putative exonuclease
79	70090	71589	putative UvsW DNA helicase
80	71590	71694	hypothetical protein
81	72306	72974	putative slidiing clamp DNA polymerase accessory protein
82	73045	74040	putative clamp loader subunit DNA polymerase accessory protein
83	74044	74475	putative clamp loader subunit DNA polymerase accessory protein
84	74509	75000	hypothetical protein
85	74997	75572	putative nucleoside triphosphate pyrophosphohydrolase
86	75641	76831	hypothetical protein
87	76833	76934	hypothetical protein
88	77006	77347	hypothetical protein

89	77445	79454	hypothetical protein
90	79512	79907	hypothetical protein
91	79962	80717	hypothetical protein
92	80788	80994	hypothetical protein
93	80997	83183	hypothetical protein
94	83238	83507	hypothetical protein
95	83511	83756	hypothetical protein
96	83753	84079	hypothetical protein
97	84066	84338	putative acyl carrier protein
98	84441	85286	hypothetical protein
99	85394	85624	hypothetical protein
100	85716	86165	hypothetical protein
101	86227	86607	hypothetical protein
102	86604	86909	hypothetical protein
103	86937	87053	hypothetical protein
104	87062	87256	hypothetical protein
105	87249	87524	hypothetical protein
106	87597	88061	putative superinfection exclusion protein
107	88064	88393	hypothetical protein
108	88390	89130	hypothetical protein
109	89216	89488	putative histone family DNA-binding protein
110	89630	91348	putative ATP-dependent DNA helicase
111	91349	92116	hypothetical protein
112	92164	92703	putative ribonuclease H
113	92719	93483	putative late transcription sigma factor
114	93474	94592	putative recombination-related endonuclease
115	94595	96919	putative recombination endonuclease subunit
116	96951	97247	hypothetical protein
117	97228	97542	hypothetical protein
118	97555	98166	putative RegB endoribonuclease
119	98166	98762	hypothetical protein
120	98759	98902	hypothetical protein
121	99005	99334	hypothetical protein
122	99420	101501	hypothetical protein

123	101554	102120	hypothetical protein
124	102131	102373	hypothetical protein
125	102460	103068	hypothetical protein
126	103065	104132	putative DNA primase subunit
127	104134	104361	hypothetical protein
128	104452	104727	hypothetical protein
129	104903	105412	hypothetical protein
130	105472	105795	hypothetical protein
131	105805	106014	hypothetical protein
132	106054	106848	putative peptidoglycan binding protein
133	106959	107795	putative PhoH-like phosphate starvation-inducible protein
134	107906	110206	putative NrdA ribonucleoside-diphosphate reductase alpha subunit
135	110280	111398	putative NrdB ribonucleoside-diphosphate reductase beta subunit
136	111395	111628	putative glutaredoxin
137	111728	112171	hypothetical protein
138	112168	112611	hypothetical protein
139	112608	112988	putative baseplate wedge subunit
140	112988	114808	putative baseplate hub subunit and tail lysozyme
141	115312	116073	putative baseplate hub subunit
142	116126	116680	hypothetical protein
143	116681	117163	hypothetical protein
144	117209	118003	putative RuvC-like holliday junction resolvase
145	117984	118274	hypothetical protein
146	118261	118506	hypothetical protein
147	118499	118735	putative later promoter transcription factor
148	118747	118983	hypothetical protein
149	119092	120153	putative ssDNA binding protein
150	120179	121117	putative baseplate tail tube protein
151	121168	121866	putative DNA end protector protein
152	121924	122730	hypothetical protein
153	122750	123079	hypothetical protein
154	123171	124103	hypothetical protein
155	124171	124830	putative kinase

156	124830	126044	putative thymidylate synthase
157	126056	126619	hypothetical protein
158	126616	127209	putative dUTP diphosphatase
159	127211	127780	hypothetical protein
160	127768	128844	putative UvsX RecA-like recombination protein
161	129235	130581	putative DNA primase-helicase subunit
162	130679	130957	putative GTPase-activator protein
163	130947	131243	hypothetical protein
164	131971	133191	hypothetical protein
165	133195	133287	hypothetical protein
166	133352	133963	hypothetical protein
167	134030	134461	hypothetical protein
168	134822	135043	hypothetical protein
169	135040	135321	hypothetical protein
170	135321	137018	putative DNA ligase
171	137062	137436	hypothetical protein
172	137433	137633	putative transcriptional regulator
173	137638	138321	putative DNA helicase loader
174	138324	140213	putative tail length tape measure protein
175	140224	141618	hypothetical protein
176	141611	142165	putative baseplate wedge subunit
177	142177	143142	putative baseplate tail tube cap protein
178	143182	143811	putative head completion protein
179	143803	144243	hypothetical protein
180	144309	145244	putative deoxycytidylate deaminase
181	145237	145614	hypothetical protein
182	145617	146180	putative alpha hydrolase
183	146405	146968	putative metallophosphatase
184	146968	147369	hypothetical protein
185	147429	147854	hypothetical protein
186	147854	148078	hypothetical protein
187	148078	148446	hypothetical protein
188	148443	149072	putative DexA exonuclease
189	149045	149650	hypothetical protein

190	149647	149913	hypothetical protein
191	149967	150389	hypothetical protein
192	150398	150616	hypothetical protein
193	150613	151998	putative DNA topoisomerase/gyrase small subunit
194	152077	153978	putative DNA topoisomerase/gyrase large subunit
195	154006	154569	hypothetical protein
196	154566	155057	hypothetical protein
197	155103	155300	hypothetical protein
198	155345	155716	putative histone-like protein
199	155953	156093	hypothetical protein
200	156093	156869	putative tail fibre protein
201	156857	157084	hypothetical protein
202	157274	157672	hypothetical protein
203	157754	159385	hypothetical protein

Table A.1 Annotation table for 3M (Genbank reference MH929319)

A.2 AD1 genome annotation table

ORF	Start	End	Annotation
1	34	1455	putative DNA helicase DnaB
2	1464	1706	hypothetical protein
3	1699	2496	hypothetical protein
4	2483	5311	putative terminase
5	5351	5566	hypothetical protein
6	5574	7517	putative portal protein
7	7517	7852	hypothetical protein
8	7849	8235	hypothetical protein
9	8264	8398	hypothetical protein
10	8398	8895	hypothetical protein
11	8892	9920	putative DNA polymerase I
12	9931	10473	hypothetical protein
13	10473	11066	putative O-acetyl-ADP-ribose deacetylase
14	11208	11423	hypothetical protein
15	11492	12181	putative membrane protein
16	12233	13003	putative methyltransferase
17	13015	14070	putative DUF1611 domain-containing protein
18	14063	14932	putative asparagine synthase
19	14934	15554	hypothetical protein
20	15567	15830	hypothetical protein
21	15838	16293	hypothetical protein
22	16290	16499	hypothetical protein
23	16529	16822	hypothetical protein
24	16830	17012	putative DNA primase
25	17265	17717	hypothetical protein
26	17732	18115	hypothetical protein
27	18354	19196	putative DNA adenine methylase
28	19204	19515	hypothetical protein
29	19568	20152	hypothetical protein
30	20152	20850	hypothetical protein
31	20843	21181	hypothetical protein
32	21159	21638	hypothetical protein

33	21619	22029	hypothetical protein
34	22014	22535	putative CMP deaminase
35	22561	23328	hypothetical protein
36	23339	23827	hypothetical protein
37	23814	23978	hypothetical protein
38	24020	24547	hypothetical protein
39	24551	25189	putative membrane protein
40	25164	25511	hypothetical protein
41	25511	25849	hypothetical protein
42	25849	26199	hypothetical protein
43	26199	27305	putative thymidylate synthase
44	27302	27952	hypothetical protein
45	27952	28842	hypothetical protein
46	28893	29228	hypothetical protein
47	29274	29984	putative transcriptional repressor
48	29986	32115	putative DNA-cytosine methyltransferase
49	32288	32827	hypothetical protein
50	32995	33405	hypothetical protein
51	33464	34369	hypothetical protein
52	34450	35109	hypothetical protein
53	35118	35471	hypothetical protein
54	35407	35931	hypothetical protein
55	35957	36265	hypothetical protein
56	36297	36764	hypothetical protein
57	36761	36961	hypothetical protein
58	36958	37356	putative ASCH domain-containing protein
59	37353	37568	hypothetical protein
60	37614	38039	hypothetical protein
61	38047	38529	hypothetical protein
62	38559	38885	hypothetical protein
63	38885	39400	hypothetical protein
64	39390	40064	putative GTP pyrophosphokinase
65	40039	40413	hypothetical protein
66	40385	40768	hypothetical protein

67	40740	41390	hypothetical protein
68	41541	41888	hypothetical protein
69	41842	42567	hypothetical protein
70	42634	43335	hypothetical protein
71	43335	44333	hypothetical protein
72	44391	44855	putative lipoprotein
73	44852	45163	hypothetical protein
74	45163	45471	putative membrane protein
75	45527	45766	hypothetical protein
76	45802	46401	hypothetical protein
77	46401	47303	hypothetical protein
78	47313	47846	hypothetical protein
79	47848	48897	hypothetical protein
80	48964	49488	hypothetical protein
81	49485	50561	hypothetical protein
82	50571	51149	hypothetical protein
83	51149	52051	hypothetical protein
84	52051	52506	hypothetical protein
85	52508	53596	hypothetical protein
86	53598	54554	hypothetical protein
87	54554	55051	putative membrane protein
88	55119	55613	hypothetical protein
89	56052	57002	hypothetical protein
90	57012	58055	hypothetical protein
91	58065	58700	putative structural protein
92	58761	59342	hypothetical protein
93	59354	59575	hypothetical protein
94	59587	60183	hypothetical protein
95	60241	61860	hypothetical protein
96	61885	62628	hypothetical protein
97	62625	63134	putative membrane protein
98	63140	63424	hypothetical protein
99	63489	64352	hypothetical protein
100	64388	65212	hypothetical protein

101	65196	65672	hypothetical protein
102	65674	66495	putative tail fibre protein
103	66507	66710	hypothetical protein
104	66720	69593	putative ILEI domain-containing protein
105	69637	71661	hypothetical protein
106	71671	72162	putative tail fibre protein
107	72172	72804	putative tail fibre protein
108	72804	74243	putative tail protein
109	74240	76579	hypothetical protein
110	76665	81200	hypothetical protein
111	81197	82660	putative baseplate wedge subunit protein
112	82662	82787	putative baseplate wedge subunit
113	82789	83211	putative baseplate protein
114	83211	83501	putative baseplate spike protein
115	84958	86721	hypothetical protein
116	86730	87140	hypothetical protein
117	87161	87775	putative dTMP kinase
118	87785	88276	putative MmcB-like DNA repair protein
119	88266	88724	putative NUDIX hydrolase
120	88721	89224	hypothetical protein
121	89263	89553	hypothetical protein
122	89550	90275	hypothetical protein
123	90286	91074	putative baseplate protein
124	91071	92966	hypothetical protein
125	92969	93403	hypothetical protein
126	93403	93696	hypothetical protein
127	93681	94487	hypothetical protein
128	94500	97139	putative VGRG protein
129	97139	97918	hypothetical protein
130	97984	98667	hypothetical protein
131	98678	99190	putative tail tube protein
132	99193	99867	hypothetical protein
133	99919	100431	putative tail tube protein
134	100443	102128	putative tail sheath protein

135	102185	102529	hypothetical protein
136	102529	103218	hypothetical protein
137	103284	103706	hypothetical protein
138	103759	104859	putative major capsid protein
139	104919	105626	putative structural protein
140	105682	107676	hypothetical protein
141	107753	108880	hypothetical protein
142	108880	109659	putative prohead core protein protease
143	109670	110080	hypothetical protein
144	110082	110849	hypothetical protein
145	110818	112026	putative glycosyl transferase
146	112088	113005	hypothetical protein
147	113008	114954	putative DNA ligase
148	114994	116853	hypothetical protein
149	116916	121043	hypothetical protein
150	121099	122280	hypothetical protein
151	122299	123285	hypothetical protein
152	123293	124078	hypothetical protein
153	124089	128099	putative major tail protein
154	128096	128728	hypothetical protein
155	128739	129377	hypothetical protein
156	129412	130107	hypothetical protein
157	130155	131831	putative tail sheath protein
158	131946	133031	hypothetical protein
159	133034	133900	hypothetical protein
160	133897	134091	hypothetical protein
161	134101	134721	hypothetical protein
162	134742	135104	hypothetical protein
163	135120	135491	hypothetical protein
164	135481	135804	hypothetical protein
165	135842	136333	hypothetical protein
166	136336	136818	hypothetical protein
167	136818	138419	putative DNA repair helicase
168	138422	138967	hypothetical protein

169	138967	139401	hypothetical protein
170	139410	140588	hypothetical protein
171	140554	140847	hypothetical protein
172	140844	144005	putative DNA polymerase I
173	144095	145072	hypothetical protein
174	145082	145633	hypothetical protein
175	145677	151523	putative ATP-dependent DNA helicase
176	151523	152017	hypothetical protein
177	152028	152810	hypothetical protein
178	152807	153172	putative HNH family endonuclease
179	153206	154024	hypothetical protein
180	154084	154884	hypothetical protein
181	154931	156094	putative head to tail joining protein
182	156096	157127	hypothetical protein
183	157202	158485	hypothetical protein
184	158658	159347	hypothetical protein
185	159466	160539	putative recombination related endonuclease
186	160580	161335	hypothetical protein
187	161339	161845	putative ssDNA binding protein
188	161885	162250	putative DUF2778 domain-containing protein
189	162332	162841	hypothetical protein
190	162842	163753	hypothetical protein
191	163772	164233	hypothetical protein
192	164371	164706	hypothetical protein
193	164709	165281	putative glycosyl hydrolase
194	165318	166163	hypothetical protein
195	166250	166591	putative membrane protein
196	166567	166884	hypothetical protein
197	166886	167599	hypothetical protein
198	167602	167922	hypothetical protein
199	167922	170252	putative exonuclease
200	170252	170452	hypothetical protein
201	170452	170664	hypothetical protein
202	170744	171265	hypothetical protein

203	171265	171714	hypothetical protein
204	171845	172381	hypothetical protein
205	172487	173776	putative DNA polymerase III
206	173919	174116	hypothetical protein
207	174591	175178	hypothetical protein
208	175178	175522	hypothetical protein
209	175527	176099	hypothetical protein
210	176099	177073	hypothetical protein
211	177085	177519	hypothetical protein
212	177611	177805	hypothetical protein
213	177845	178369	hypothetical protein
214	178373	179425	hypothetical protein
215	180393	181052	hypothetical protein
216	181062	182183	hypothetical protein
217	182283	182831	putative holliday junction resolvase
218	182839	183204	hypothetical protein
219	183251	183943	hypothetical protein
220	183936	185705	putative inverse autotransporter beta-barrel domain-containing protein
221	185759	186409	hypothetical protein
222	186466	186645	hypothetical protein
223	186647	187705	putative DNA primase
224	187770	188105	hypothetical protein
225	188068	189153	putative exonuclease
226	189231	189560	hypothetical protein
227	189568	190032	hypothetical protein
228	189992	190525	hypothetical protein
229	190458	191114	hypothetical protein
230	191120	191401	hypothetical protein
231	191382	191813	hypothetical protein
232	191813	192028	hypothetical protein
233	192031	192759	hypothetical protein
234	192752	193285	hypothetical protein
235	193278	193745	putative cylic phosphodiesterase

236	193797	194108	hypothetical protein
237	194154	195191	hypothetical protein
238	195242	196516	putative ssDNA binding protein
239	196574	199099	putative RecA protein
240	199149	199568	hypothetical protein
241	199668	200369	hypothetical protein
242	200546	201472	hypothetical protein
243	201517	202284	hypothetical protein
244	202718	204256	hypothetical protein
245	204427	205131	hypothetical protein
246	205236	205475	hypothetical protein
247	205532	206035	hypothetical protein
248	206098	206712	hypothetical protein
249	206761	207021	hypothetical protein
250	207100	207468	hypothetical protein
251	207465	207710	hypothetical protein
252	207703	208065	hypothetical protein
253	208154	208639	hypothetical protein
254	208627	208944	hypothetical protein
255	209037	209354	hypothetical protein
256	209419	209772	hypothetical protein
257	209838	210242	hypothetical protein
258	210300	210752	putative XRE family transcriptional regulator
259	210745	211284	hypothetical protein
260	211297	211737	hypothetical protein
261	211734	212177	hypothetical protein
262	212259	212486	hypothetical protein
263	212483	213538	hypothetical protein
264	213535	213744	hypothetical protein
265	213737	214270	putative RNA NAD 2
266	214273	214686	hypothetical protein
267	214742	216229	putative radical SAM superfamily protein
268	216244	216861	hypothetical protein
269	216863	217219	hypothetical protein

270	217219	217698	hypothetical protein
271	217689	217952	putative DksA/TraR family C4-type zinc finger protein
272	217955	218383	hypothetical protein
273	218380	218565	hypothetical protein
274	218565	218882	hypothetical protein
275	218879	219076	hypothetical protein
276	219073	219426	hypothetical protein
277	219426	220145	hypothetical protein
278	220155	220784	hypothetical protein
279	220781	221152	hypothetical protein
280	221154	221912	hypothetical protein
281	221887	222090	hypothetical protein
282	222077	222628	hypothetical protein
283	222687	223667	putative UV damage repair endonuclease
284	223756	224058	hypothetical protein
285	224060	224323	hypothetical protein
286	224752	225003	hypothetical protein
287	225078	225590	hypothetical protein
288	225590	225877	hypothetical protein
289	225893	227065	hypothetical protein
290	227178	227480	hypothetical protein
291	227489	228043	hypothetical protein
292	228056	228514	hypothetical protein
293	228525	229094	putative dUTPase
294	229087	229302	hypothetical protein
295	230406	231092	hypothetical protein
296	231161	232102	hypothetical protein
297	232180	232884	hypothetical protein
298	232890	233117	hypothetical protein
299	233120	233428	hypothetical protein
300	233428	234003	hypothetical protein
301	234056	235657	hypothetical protein
302	235693	236127	hypothetical protein
303	236136	236435	hypothetical protein

304	236542	237006	hypothetical protein
305	237008	237703	hypothetical protein
306	237703	237894	hypothetical protein
307	238035	238517	hypothetical protein
308	238522	238737	hypothetical protein
309	238992	239723	hypothetical protein
310	239723	240172	hypothetical protein
311	240190	240342	hypothetical protein
312	240505	242136	hypothetical protein
313	242225	243457	hypothetical protein
314	243454	243834	hypothetical protein
315	243831	244292	hypothetical protein
316	244282	245562	hypothetical protein
317	245590	248613	hypothetical protein
318	248671	249501	hypothetical protein
319	249602	250147	hypothetical protein
320	250297	251118	hypothetical protein
321	251337	251873	hypothetical protein
322	251971	252471	hypothetical protein
323	252496	253236	hypothetical protein
324	253374	255404	putative DNA gyrase subunit B
325	255406	257052	putative DNA topoisomerase 4 subunit A
326	257399	258673	hypothetical protein
327	258666	259208	hypothetical protein
328	259141	260013	putative DNA topoisomerase 4 subunit A
329	260068	261207	hypothetical protein
330	261225	261596	hypothetical protein

Table A.2 Annotation table for AD1 (Genbank reference MH460463)

A.3 JA10 genome annotation table

ORF	Start	End	Annotation
1	1014	1478	putative S-adenosyl-L-methionine hydrolase,phage-associated protein
2	1480	1620	hypothetical protein
3	1680	1877	hypothetical protein
4	1890	3029	putative protein kinase
5	3102	5744	putative T3/T7-like RNA polymerase
6	5815	6060	hypothetical protein
7	6162	6338	hypothetical protein
8	6341	6604	hypothetical protein
9	6676	7713	putative DNA ligase
10	7706	7843	putative ligase
11	7897	8100	hypothetical protein
12	8097	8219	hypothetical protein
13	8257	8514	hypothetical protein
14	8511	8696	hypothetical protein
15	8693	9331	hypothetical protein
16	9324	9512	putative bacterial RNA polymerase inhibitor
17	9505	9873	hypothetical protein
18	9936	10637	putative ssDNA-binding protein
19	10637	11083	putative endonuclease
20	11085	11540	putative lysozyme
21	11612	12115	hypothetical protein
22	12320	13831	DNA primase/helicase
23	13918	14151	hypothetical protein
24	14228	14614	hypothetical protein
25	14637	16721	DNA polymerase
26	16742	17056	hypothetical protein
27	17056	17265	hypothetical protein
28	17262	17627	hypothetical protein
29	17663	18562	putative exonuclease
30	18728	18982	hypothetical protein
31	19001	19297	hypothetical protein

32	19328	19642	putative tail-assembly protein
33	19653	21257	putative head-to-tail joining protein
34	21366	22223	putative capsid and scaffold protein
35	22348	23382	putative structural protein
36	23424	23654	putative minor capsid protein
37	23728	24315	putative tail tubular protein A
38	24337	26721	putative tail tubular protein B
39	26809	27246	putative internal core protein
40	27249	27842	putative tail protein
41	27854	30118	putative tail protein
42	30144	34118	putative internal (core) protein
43	34181	35800	putative tail fibre protein
44	35803	36198	putative tail fibre assembly protein
45	36243	36452	putative holin lysis protein
46	36445	36708	putative DNA packaging protein A
47	36807	37292	putative endopeptidase
48	37292	37909	hypothetical protein
49	37925	39679	putative DNA packaging protein
50	39949	40107	hypothetical protein

Table A.3 Annotation table for JA10 (Genbank reference MH460459)

A.4 JA11 genome annotation table

ORF	Start	End	Annotation
1	54	1433	putative replicative DNA helicase DnaB
2	1531	2205	hypothetical protein
3	2189	5002	putative terminase
4	5068	5283	hypothetical protein
5	5296	7239	putative portal protein
6	7239	7547	hypothetical protein
7	7549	7923	hypothetical protein
8	7971	8435	hypothetical protein
9	8432	9469	putative DNA polymerase I
10	9481	9930	hypothetical protein
11	9936	10523	putative O-acetyl-ADP-ribose deacetylase
12	10532	10846	hypothetical protein
13	10852	11496	putative membrane protein
14	11757	12023	putative DNA primase
15	12272	12451	putative DNA adenine methylase
16	12460	13302	hypothetical protein
17	13318	13461	hypothetical protein
18	13470	14192	hypothetical protein
19	14185	14439	hypothetical protein
20	14520	15098	hypothetical protein
21	15098	15463	hypothetical protein
22	15450	15944	putative CMP deaminase
23	15954	16406	hypothetical protein
24	16409	17281	hypothetical protein
25	17274	18377	putative thymidylate synthase
26	18419	19129	hypothetical protein
27	19129	19776	hypothetical protein
28	20125	20364	hypothetical protein
29	20354	20758	hypothetical protein
30	20804	21508	hypothetical protein
31	21480	21950	hypothetical protein
32	21943	22902	hypothetical protein

33	22883	23746	hypothetical protein
34	23796	24131	hypothetical protein
35	24109	24396	hypothetical protein
36	24751	24906	hypothetical protein
37	24914	25327	hypothetical protein
38	25324	26043	hypothetical protein
39	26033	26344	hypothetical protein
40	26341	26700	hypothetical protein
41	26709	27461	hypothetical protein
42	27463	27723	hypothetical protein
43	27785	28426	hypothetical protein
44	28546	28911	hypothetical protein
45	28901	29260	hypothetical protein
46	29271	29738	hypothetical protein
47	29799	30392	hypothetical protein
48	30395	30607	hypothetical protein
49	30604	30873	hypothetical protein
50	30881	31312	hypothetical protein
51	31312	31839	hypothetical protein
52	31814	32653	hypothetical protein
53	32662	32916	hypothetical protein
54	32913	33314	putative ASCH domain-containing protein
55	33290	34048	hypothetical protein
56	34035	34565	hypothetical protein
57	34558	34875	hypothetical protein
58	34862	35185	hypothetical protein
59	35717	35962	hypothetical protein
60	36000	36584	putative bifunctional (p)ppGpp synthetase/guanosine-3
61	36588	36983	hypothetical protein
62	37191	37652	hypothetical protein
63	37849	38448	hypothetical protein
64	38445	38810	hypothetical protein
65	38807	39514	hypothetical protein
66	39516	40238	hypothetical protein

67	40275	41066	hypothetical protein
68	41105	41572	putative membrane protein
69	41569	41850	hypothetical protein
70	41858	42187	putative membrane protein
71	42198	42437	hypothetical protein
72	42455	43087	hypothetical protein
73	43087	43995	hypothetical protein
74	44006	44560	hypothetical protein
75	44557	45639	hypothetical protein
76	45636	46229	hypothetical protein
77	46229	47152	hypothetical protein
78	47152	47619	hypothetical protein
79	47633	48652	hypothetical protein
80	48679	50253	hypothetical protein
81	50254	50751	putative membrane protein
82	50834	51886	hypothetical protein
83	52049	53425	putative T1SS secreted agglutinin RTX
84	53444	54193	hypothetical protein
85	54190	54681	putative membrane protein
86	54683	54958	hypothetical protein
87	54994	55740	hypothetical protein
88	55743	56540	hypothetical protein
89	56533	56991	hypothetical protein
90	56993	57811	putative tail fibre protein
91	57823	58023	hypothetical protein
92	58068	61601	putative ILEI domain-containing protein
93	61610	63631	hypothetical protein
94	63644	64138	putative tail fibre protein
95	64152	64784	putative tail fibre protein
96	64795	65769	putative tail protein
97	65766	68087	hypothetical protein
98	68175	69245	hypothetical protein
99	69289	73812	hypothetical protein
100	73809	75272	putative baseplate wedge subunit

101	75269	75682	putative baseplate protein
102	75682	75972	putative baseplate spike
103	76665	76976	hypothetical protein
104	77713	79491	hypothetical protein
105	79501	79941	hypothetical protein
106	79926	80519	putative dTMP kinase
107	80524	80961	putative MmcB-like DNA repair protein
108	81002	81454	putative NUDIX hydrolase
109	81451	81957	hypothetical protein
110	82255	82881	hypothetical protein
111	82894	83679	putative baseplate protein/tail-associated lysozyme
112	83676	85559	hypothetical protein
113	85563	85784	hypothetical protein
114	85791	86189	hypothetical protein
115	86189	86482	hypothetical protein
116	86455	87258	hypothetical protein
117	87258	90050	putative VgrG-like protein/endolysin
118	90050	90892	hypothetical protein
119	90885	91571	hypothetical protein
120	91582	92094	putative tail tube protein
121	92097	92765	hypothetical protein
122	92808	93323	putative tail tube protein
123	93338	95023	putative tail sheath protein
124	95081	95425	hypothetical protein
125	95427	96089	hypothetical protein
126	96161	96634	hypothetical protein
127	96726	97820	putative major capsid protein
128	97873	98577	putative structural protein
129	98647	100605	putative ATPase
130	100689	101792	hypothetical protein
131	101789	102610	putative prohead core protein protease
132	102617	103039	hypothetical protein
133	103044	103931	hypothetical protein
134	103939	105039	putative glycosyl transferase

135	105050	105964	hypothetical protein
136	105964	107934	putative DNA ligase
137	107958	108881	hypothetical protein
138	108878	109720	hypothetical protein
139	109720	111651	hypothetical protein
140	111711	115673	hypothetical protein
141	115735	116889	hypothetical protein
142	116908	117882	hypothetical protein
143	117886	118650	hypothetical protein
144	118666	121749	putative major tail protein/T1SS secreted agglutinin RTX
145	121739	122383	hypothetical protein
146	122380	123006	hypothetical protein
147	123028	124704	putative tail sheath protein
148	124792	125838	hypothetical protein
149	125826	126680	hypothetical protein
150	126685	126894	hypothetical protein
151	126891	127523	hypothetical protein
152	127560	127841	hypothetical protein
153	127858	128241	hypothetical protein
154	128207	128545	hypothetical protein
155	128551	129444	putative DNA repair helicase
156	130521	131657	putative DNA repair helicase
157	131660	132223	hypothetical protein
158	132225	132653	hypothetical protein
159	132650	133825	hypothetical protein
160	133812	134081	hypothetical protein
161	134078	137233	putative DNA polymerase I
162	137328	137645	hypothetical protein
163	137648	138202	hypothetical protein
164	138246	144080	putative ATP-dependent DNA helicase
165	144091	144591	hypothetical protein
166	144604	145371	hypothetical protein
167	145371	145730	putative HNH family endonuclease
168	145737	146495	hypothetical protein

169	146556	147359	hypothetical protein
170	147412	148572	putative head to tail joining protein
171	148575	149672	hypothetical protein
172	149672	150949	hypothetical protein
173	151090	151752	hypothetical protein
174	151888	153018	putative recombination-related endonuclease
175	153133	153657	putative ssDNA binding protein
176	153704	155668	hypothetical protein
177	155665	157452	hypothetical protein
178	157463	157828	putative DUF2778 domain-containing protein
179	157828	158394	hypothetical protein
180	158413	159405	hypothetical protein
181	159408	159875	hypothetical protein
182	159872	160447	putative glycosyl hydrolase
183	160527	160850	hypothetical protein
184	160850	161563	hypothetical protein
185	161566	161817	hypothetical protein
186	161817	164117	putative exonuclease
187	164120	164320	hypothetical protein
188	164401	164637	hypothetical protein
189	164654	165088	hypothetical protein
190	165091	165768	hypothetical protein
191	165889	166368	hypothetical protein
192	166448	167668	putative DNA polymerase III
193	168011	168586	hypothetical protein
194	168586	168918	hypothetical protein
195	168908	169885	hypothetical protein
196	169928	170395	hypothetical protein
197	170402	170896	hypothetical protein
198	170908	171372	hypothetical protein
199	171421	172509	hypothetical protein
200	172547	173764	hypothetical protein
201	174627	175178	hypothetical protein
202	175181	176299	hypothetical protein

203	176443	176991	putative holliday junction resolvase
204	177004	177645	hypothetical protein
205	177704	178342	hypothetical protein
206	178339	180324	putative inverse autotransporter beta-barrel domain-containing protein
207	180446	181123	hypothetical protein
208	181159	182214	putative DNA primase
209	182280	182615	hypothetical protein
210	182653	183669	putative exonuclease
211	183677	184078	hypothetical protein
212	184532	185194	hypothetical protein
213	185197	185361	hypothetical protein
214	185373	185645	hypothetical protein
215	185642	185959	hypothetical protein
216	186002	186178	hypothetical protein
217	186181	186702	hypothetical protein
218	186699	187148	putative cyclic phosphodiesterase
219	187132	187413	hypothetical protein
220	187449	188486	hypothetical protein
221	188547	189752	putative ssDNA binding protein
222	189807	191312	putative RecA protein
223	191354	191755	hypothetical protein
224	191872	192651	hypothetical protein
225	192648	193226	hypothetical protein
226	193210	193743	hypothetical protein
227	193824	194072	hypothetical protein
228	194136	195953	hypothetical protein
229	196006	196599	hypothetical protein
230	196538	197179	hypothetical protein
231	197227	197868	hypothetical protein
232	197868	198185	hypothetical protein
233	198163	199332	putative methyltransferase
234	199388	199534	hypothetical protein
235	199513	200736	putative DNA adenine methylase

236	200913	201248	hypothetical protein
237	201311	201790	hypothetical protein
238	201790	202305	hypothetical protein
239	202295	202720	hypothetical protein
240	202689	202964	hypothetical protein
241	202964	203182	hypothetical protein
242	203184	203942	hypothetical protein
243	204209	204802	hypothetical protein
244	204786	205136	hypothetical protein
245	205145	205495	hypothetical protein
246	205507	205755	hypothetical protein
247	205926	206198	hypothetical protein
248	206263	206691	hypothetical protein
249	206751	207098	hypothetical protein
250	207095	207388	hypothetical protein
251	207398	208183	hypothetical protein
252	208180	208788	hypothetical protein
253	208785	209504	hypothetical protein
254	209506	210186	hypothetical protein
255	210235	210906	hypothetical protein
256	210903	211121	Hypothetical protein
257	211543	212043	hypothetical protein
258	212045	212761	hypothetical protein
259	212761	213003	hypothetical protein
260	213123	213461	hypothetical protein
261	213464	214084	hypothetical protein
262	214081	214623	putative RNA 2'-phosphotransferase
263	214726	215118	hypothetical protein
264	215128	215385	putative DksA/TraR family C4-type zinc finger protein
265	215388	215696	hypothetical protein
266	215699	215839	hypothetical protein
267	215839	216273	hypothetical protein
268	216292	216741	hypothetical protein
269	216777	217712	putative UvsE UV damage repair endonuclease

270	217703	217804	hypothetical protein
271	217813	218622	hypothetical protein
272	218633	219241	hypothetical protein
273	219242	221191	hypothetical protein
274	221244	221903	hypothetical protein
275	222039	222218	hypothetical protein
276	222222	222470	hypothetical protein
277	222474	222932	hypothetical protein
278	222935	223231	hypothetical protein
279	223244	223537	hypothetical protein
280	223592	224035	hypothetical protein
281	224038	224328	hypothetical protein
282	224328	224894	putative dUTPase
283	225272	225790	putative lytic transglycosylase
284	225937	226230	hypothetical protein
285	226246	226758	hypothetical protein
286	226771	226914	hypothetical protein
287	226957	227664	hypothetical protein
288	227664	228188	hypothetical protein
289	228185	228640	hypothetical protein
290	228641	228847	hypothetical protein
291	228854	229258	hypothetical protein
292	229309	229932	hypothetical protein
293	230033	230638	hypothetical protein
294	230635	230811	hypothetical protein
295	230808	231215	hypothetical protein
296	231226	232059	hypothetical protein
297	232076	232858	hypothetical protein
298	233013	233363	hypothetical protein
299	233514	234257	hypothetical protein
300	234313	234843	hypothetical protein
301	234850	235227	hypothetical protein
302	235227	235592	hypothetical protein
303	235878	237491	hypothetical protein

304	237564	238490	hypothetical protein
305	238490	238852	hypothetical protein
306	238861	239337	hypothetical protein
307	239309	240577	hypothetical protein
308	240577	243549	hypothetical protein
309	243604	244356	hypothetical protein
310	244379	244957	hypothetical protein
311	245162	245953	hypothetical protein
312	245985	246080	hypothetical protein
313	246258	246776	hypothetical protein
314	246870	247211	hypothetical protein
315	247330	249390	putative DNA topoisomerase IV/gyrase subunit B
316	249390	251072	putative DNA topoisomerase 4 subunit A
317	251226	252416	hypothetical protein
318	252406	252948	hypothetical protein
319	252920	253759	Hypothetical protein
320	253821	254954	hypothetical protein
321	254962	255318	hypothetical protein

Table A.4 Annotation table for JA11 (Genbank reference MH389777)

A.5 JA13 genome annotation table

ORF	Start	End	Annotation
1	74	1453	putative replicative DNA helicase DnaB
2	1446	2225	hypothetical protein
3	2209	5022	putative terminase
4	5215	5430	Hypothetical protein
5	5443	7386	putative portal protein
6	7386	7694	hypothetical protein
7	7696	8070	hypothetical protein
8	8118	8582	hypothetical protein
9	8579	9616	putative DNA polymerase I
10	9628	10077	hypothetical protein
11	10083	10670	putative O-acetyl-ADP-ribose deacetylase
12	10679	10993	hypothetical protein
13	10999	11643	putative membrane protein
14	11904	12170	putative DNA primase
15	12193	12414	hypothetical protein
16	12419	12598	hypothetical protein
17	12607	13449	putative DNA adenine methylase
18	13465	13608	hypothetical protein
19	13617	14339	hypothetical protein
20	14332	14586	hypothetical protein
21	14667	15245	hypothetical protein
22	15245	15610	hypothetical protein
23	15597	16091	CMP deaminase
24	16101	16553	hypothetical protein
25	16556	17428	hypothetical protein
26	17421	18524	thymidylate synthase
27	18566	19276	hypothetical protein
28	19276	19923	hypothetical protein
29	20272	20511	hypothetical protein
30	20501	20905	hypothetical protein
31	20951	21655	hypothetical protein
32	21627	22097	hypothetical protein

33	22090	23049	hypothetical protein
34	23030	23893	hypothetical protein
35	23943	24278	hypothetical protein
36	24256	24543	hypothetical protein
37	24644	24895	hypothetical protein
38	24898	25053	hypothetical protein
39	25061	25474	hypothetical protein
40	25471	26190	hypothetical protein
41	26180	26491	hypothetical protein
42	26488	26844	hypothetical protein
43	26853	27605	hypothetical protein
44	27607	27867	hypothetical protein
45	27930	28571	hypothetical protein
46	28691	29056	hypothetical protein
47	29046	29405	hypothetical protein
48	29416	29883	hypothetical protein
49	29944	30537	hypothetical protein
50	30540	30752	hypothetical protein
51	30749	31018	hypothetical protein
52	31027	31488	hypothetical protein
53	31463	32302	hypothetical protein
54	32311	32565	hypothetical protein
55	32562	32963	putative ASCH domain-containing protein
56	32939	33697	hypothetical protein
57	33684	34214	hypothetical protein
58	34207	34524	hypothetical protein
59	34511	34834	hypothetical protein
60	35155	36198	hypothetical protein
61	36195	36779	putative bifunctional (p)ppGpp synthetase/guanosine-3
62	36783	37178	hypothetical protein
63	37386	37847	hypothetical protein
64	38044	38643	hypothetical protein
65	38640	39005	hypothetical protein
66	39002	39709	hypothetical protein

67	39711	40433	hypothetical protein
68	40470	41261	hypothetical protein
69	41300	41767	putative membrane protein
70	41764	42045	hypothetical protein
71	42053	42382	putative membrane protein
72	42393	42632	hypothetical protein
73	42650	43282	hypothetical protein
74	43282	44190	hypothetical protein
75	44201	44755	hypothetical protein
76	44752	45834	hypothetical protein
77	45831	46424	hypothetical protein
78	46424	47347	hypothetical protein
79	47347	47814	hypothetical protein
80	47828	48847	hypothetical protein
81	48874	50448	hypothetical protein
82	50449	50946	putative membrane protein
83	51029	52081	putative T1SS secreted agglutinin RTX
84	52244	53620	hypothetical protein
85	53639	54388	putative membrane protein
86	54385	54876	hypothetical protein
87	54878	55153	hypothetical protein
88	55189	55935	hypothetical protein
89	55938	56735	hypothetical protein
90	56728	57186	hypothetical protein
91	57188	58006	putative tail fibre protein
92	58018	58218	Hypothetical protein
93	58263	61796	putative ILEI domain-containing protein
94	61805	63826	hypothetical protein
95	63839	64333	putative tail fibre protein
96	64347	64979	putative tail fibre protein
97	64990	65964	putative tail protein
98	65961	68282	hypothetical protein
99	68370	69440	hypothetical protein
100	69484	74007	hypothetical protein

101	74004	75467	putative baseplate wedge subunit protein
102	75464	75877	putative baseplate protein
103	75877	76167	putative baseplate spike protein
104	76860	77171	hypothetical protein
105	77908	79686	hypothetical protein
106	79696	80136	hypothetical protein
107	80121	80714	putative dTMP kinase
108	80719	81156	putative MmcB-like DNA repair protein
109	81197	81649	putative NUDIX hydrolase
110	81646	82152	hypothetical protein
111	82450	83076	hypothetical protein
112	83089	83871	putative baseplate protein/tail-associated lysozyme
113	83871	85754	hypothetical protein
114	85758	85979	hypothetical protein
115	85986	86384	hypothetical protein
116	86384	86677	hypothetical protein
117	86650	87453	hypothetical protein
118	87453	90245	putative VgrG-like protein/endolysin
119	90245	91087	hypothetical protein
120	91080	91766	hypothetical protein
121	91777	92289	putative tail tube protein
122	92292	92960	hypothetical protein
123	93003	93518	putative tail tube protein
124	93533	95218	putative tail sheath protein
125	95276	95620	hypothetical protein
126	95622	96284	hypothetical protein
127	96355	96828	hypothetical protein
128	96920	98014	putative major capsid protein
129	98067	98771	putative structural protein
130	98841	100799	putative ATPase
131	100883	101929	hypothetical protein
132	101986	102804	putative prohead core protein protease
133	102811	103233	hypothetical protein
134	103238	104125	hypothetical protein

135	104133	105233	putative glycosyl transferase
136	105244	106158	hypothetical protein
137	106158	108128	putative DNA ligase
138	108152	109075	hypothetical protein
139	109072	109914	hypothetical protein
140	109914	111878	hypothetical protein
141	111938	115900	hypothetical protein
142	115962	117116	hypothetical protein
143	117135	118109	hypothetical protein
144	118113	118877	hypothetical protein
145	118893	121976	putative major tail protein/T1SS secreted agglutinin RTX
146	121966	122610	hypothetical protein
147	122607	123233	hypothetical protein
148	123255	124931	putative tail sheath protein
149	125019	126065	hypothetical protein
150	126053	126907	hypothetical protein
151	126912	127121	hypothetical protein
152	127118	127750	hypothetical protein
153	127787	128068	hypothetical protein
154	128085	128468	hypothetical protein
155	128434	128772	hypothetical protein
156	128778	130388	putative DNA repair helicase
157	130391	130954	hypothetical protein
158	130956	131384	hypothetical protein
159	131381	132556	hypothetical protein
160	132543	132812	hypothetical protein
161	132809	135964	putative DNA polymerase I
162	136060	136377	hypothetical protein
163	136380	136934	hypothetical protein
164	136978	142812	putative ATP-dependent DNA helicase
165	142823	143323	hypothetical protein
166	143336	144103	hypothetical protein
167	144103	144462	putative HNH family endonuclease
168	144469	145227	hypothetical protein

169	145288	146091	hypothetical protein
170	146144	147304	putative head to tail joining protein
171	147307	148404	hypothetical protein
172	148404	149681	hypothetical protein
173	149822	150484	hypothetical protein
174	150620	151750	putative recombination-related endonuclease
175	151865	152389	putative ssDNA binding protein
176	152436	154400	hypothetical protein
177	154397	156184	hypothetical protein
178	156195	156560	putative DUF2778 domain-containing protein
179	156560	157126	hypothetical protein
180	157145	158173	hypothetical protein
181	158176	158643	hypothetical protein
182	158640	159215	putative glycosyl hydrolase
183	159295	159618	hypothetical protein
184	159618	160331	hypothetical protein
185	160334	160585	hypothetical protein
186	160585	162885	putative exonuclease
187	162888	163088	hypothetical protein
188	163169	163405	hypothetical protein
189	163422	163856	hypothetical protein
190	163859	164536	hypothetical protein
191	164657	165136	hypothetical protein
192	165216	166436	putative DNA polymerase III
193	166779	167354	hypothetical protein
194	167354	167686	hypothetical protein
195	167676	168653	hypothetical protein
196	168696	169163	hypothetical protein
197	169170	169664	hypothetical protein
198	169676	170140	hypothetical protein
199	170189	171328	hypothetical protein
200	171366	172220	hypothetical protein
201	173464	174015	hypothetical protein
202	174018	175136	hypothetical protein

203	175280	175828	putative holliday junction resolvase
204	175841	176536	hypothetical protein
205	176598	177215	hypothetical protein
206	177212	179197	putative inverse autotransporter beta-barrel domain-containing protein
207	179319	179996	hypothetical protein
208	180032	181087	putative DNA primase
209	181153	181488	hypothetical protein
210	181526	182542	putative exonuclease
211	182550	182951	hypothetical protein
212	182902	183489	hypothetical protein
213	183405	184067	hypothetical protein
214	184070	184234	hypothetical protein
215	184246	184518	hypothetical protein
216	184515	184832	hypothetical protein
217	184875	185051	hypothetical protein
218	185054	185575	putative cyclic phosphodiesterase
219	185572	186021	hypothetical protein
220	186005	186286	hypothetical protein
221	186322	187359	hypothetical protein
222	187420	188607	putative ssDNA binding protein
223	188662	190167	putative RecA recombinase
224	190209	190610	hypothetical protein
225	190728	191507	hypothetical protein
226	191504	192082	hypothetical protein
227	192066	192599	hypothetical protein
228	192680	192928	hypothetical protein
229	192992	194809	hypothetical protein
230	194862	195455	hypothetical protein
231	195403	196035	hypothetical protein
232	196035	196724	hypothetical protein
233	196724	197041	hypothetical protein
234	197019	198188	putative methyltransferase
235	198202	198390	hypothetical protein

236	198369	199592	putative DNA adenine methylase
237	199769	200104	hypothetical protein
238	200167	200646	hypothetical protein
239	200646	201161	hypothetical protein
240	201151	201576	hypothetical protein
241	201545	201820	hypothetical protein
242	201820	202038	hypothetical protein
243	202040	202798	hypothetical protein
244	203065	203658	hypothetical protein
245	203642	203992	hypothetical protein
246	204001	204351	hypothetical protein
247	204363	204611	hypothetical protein
248	204782	205054	hypothetical protein
249	205119	205547	hypothetical protein
250	205608	205955	hypothetical protein
251	205952	206245	hypothetical protein
252	206255	207040	hypothetical protein
253	207037	207645	hypothetical protein
254	207642	208361	hypothetical protein
255	208361	209023	Hypothetical protein
256	209094	209765	hypothetical protein
257	209762	209980	Hypothetical protein
258	209989	210198	hypothetical protein
259	210201	210407	hypothetical protein
260	210404	210904	hypothetical protein
261	210906	211622	hypothetical protein
262	211622	211864	hypothetical protein
263	211984	212322	hypothetical protein
264	212325	212945	hypothetical protein
265	212942	213484	putative RNA 2'-phosphotransferase
266	213587	213979	hypothetical protein
267	213989	214246	putative DksA/TraR family C4-type zinc finger protein
268	214249	214557	hypothetical protein
269	214560	214700	hypothetical protein

270	214700	215134	hypothetical protein
271	215154	215603	hypothetical protein
272	215639	216574	putative UvsE UV damage endonuclease
273	216565	216666	hypothetical protein
274	216675	217484	hypothetical protein
275	217492	218103	hypothetical protein
276	218104	220080	hypothetical protein
277	220133	220792	hypothetical protein
278	220928	221107	hypothetical protein
279	221111	221359	hypothetical protein
280	221363	221821	hypothetical protein
281	221824	222120	hypothetical protein
282	222133	222426	hypothetical protein
283	222481	222924	hypothetical protein
284	222927	223217	hypothetical protein
285	223217	223783	putative dUTPase
286	224161	224679	putative lytic transglycosylase
287	224826	225119	hypothetical protein
288	225135	225647	hypothetical protein
289	225660	225803	hypothetical protein
290	225846	226553	Hypothetical protein
291	226553	227077	hypothetical protein
292	227074	227529	hypothetical protein
293	227530	227736	hypothetical protein
294	227743	228147	hypothetical protein
295	228196	228819	hypothetical protein
296	228920	229525	hypothetical protein
297	229522	229698	hypothetical protein
298	229695	230102	hypothetical protein
299	230113	230946	hypothetical protein
300	230963	231745	hypothetical protein
301	231900	232250	hypothetical protein
302	232401	233144	hypothetical protein
303	233200	233730	hypothetical protein

304	233736	234113	hypothetical protein
305	234113	234478	hypothetical protein
306	234763	236376	hypothetical protein
307	236449	237375	hypothetical protein
308	237375	237737	hypothetical protein
309	237746	238222	hypothetical protein
310	238194	239462	hypothetical protein
311	239462	242410	hypothetical protein
312	242465	243217	hypothetical protein
313	243240	243818	hypothetical protein
314	244023	244805	hypothetical protein
315	244983	245501	hypothetical protein
316	245595	245936	hypothetical protein
317	246055	248115	putative DNA topoisomerase IV/gyrase subunit B
318	248115	249797	putative DNA topoisomerase 4 subunit A
319	249951	251141	hypothetical protein
320	251131	251673	hypothetical protein
321	251660	252484	Hypothetical protein
322	252546	253679	hypothetical protein
323	253687	254043	hypothetical protein

Table A.5 Annotation table for JA13 (Genbank reference MH460460)

A.6 JA15 genome annotation table

ORF	Start	End	Annotation
1	3	2813	RIIA protein
2	2907	3272	hypothetical protein
3	3311	3472	hypothetical protein
4	3504	3761	hypothetical protein
5	3763	4311	hypothetical protein
6	4361	5446	hypothetical protein
7	5456	5704	hypothetical protein
8	5704	6450	hypothetical protein
9	6520	6777	hypothetical protein
10	6774	7913	hypothetical protein
11	8166	8375	hypothetical protein
12	8372	9256	hypothetical protein
13	9253	9558	hypothetical protein
14	9566	10312	putative deoxyribonucleotidase
15	10317	10616	hypothetical protein
16	10606	10941	hypothetical protein
17	11010	14003	putative DNA polymerase
18	14087	14665	hypothetical protein
19	15041	15337	hypothetical protein
20	15364	15705	hypothetical protein
21	15716	16198	hypothetical protein
22	16225	16680	hypothetical protein
23	16736	17614	hypothetical protein
24	17680	17943	hypothetical protein
25	17950	19260	hypothetical protein
26	19324	19491	hypothetical protein
27	19727	19652	tRNA-Met
28	20724	20635	tRNA-Tyr
29	22308	22895	hypothetical protein
30	23301	25073	putative baseplate wedge subunit
31	25060	25911	putative baseplate wedge subunit
32	25913	27175	putative tail spike protein

33	27224	30187	putative tail spike protein
34	30202	30396	Hypothetical protein
35	30487	31989	putative tail spike protein
36	31986	33602	hypothetical protein
37	33754	33861	hypothetical protein
38	34148	36007	hypothetical protein
39	36221	37891	putative tail spike protein head-binding protein
40	37893	38639	putative pectate lyase
41	38650	38937	hypothetical protein
42	38939	41389	hypothetical protein
43	41398	41583	hypothetical protein
44	41604	41882	hypothetical protein
45	41999	42709	hypothetical protein
46	42838	47685	putative virulence-associated VriC protein
47	47743	47970	putative structural protein
48	47963	48295	putative capsid protein
49	48285	49037	putative neck protein
50	49124	49768	putative neck protein
51	49765	50490	putative proximal tail sheath stabilisation protein
52	50490	51188	putative terminase DNA packaging enzyme small subunit
53	51169	53376	putative terminase DNA packaging enzyme large subunit
54	53419	55329	putative tail sheath protein
55	55383	55853	putative GIY-YIG homing endonuclease
56	55877	56410	putative tail tube protein
57	56481	58118	putative portal vertex protein
58	58167	58343	hypothetical protein
59	58355	58663	putative prohead core protein
60	58674	59339	putative prohead core protein protease
61	59386	60162	putative prohead core scaffold protein
62	60252	61577	putative major capsid protein
63	61683	62522	putative homing endonuclease
64	62573	62773	hypothetical protein
65	62770	63198	hypothetical protein
66	63201	63746	hypothetical protein

67	63831	64496	hypothetical protein
68	64498	64959	hypothetical protein
69	65040	65597	hypothetical protein
70	65642	65872	hypothetical protein
71	65921	66190	hypothetical protein
72	66187	66630	hypothetical protein
73	66641	66832	hypothetical protein
74	66871	67593	hypothetical protein
75	67623	67931	hypothetical protein
76	68321	68815	putative tail completion and sheath stabiliser protein
77	68825	69310	putative UvsY DNA repair/recombination protein
78	69315	70058	putative exonuclease
79	70090	71589	putative UvsW DNA helicase
80	71590	71694	hypothetical protein
81	72306	72974	putative slidiing clamp DNA polymerase accessory protein
82	73045	74040	putative clamp loader subunit DNA polymerase accessory protein
83	74044	74475	putative clamp loader subunit DNA polymerase accessory protein
84	74509	75000	hypothetical protein
85	74997	75572	putative nucleoside triphosphate pyrophosphohydrolase
86	75641	76831	hypothetical protein
87	76833	76934	hypothetical protein
88	77006	77347	hypothetical protein
89	77445	79454	hypothetical protein
90	79512	79907	hypothetical protein
91	79962	80717	hypothetical protein
92	80788	80994	hypothetical protein
93	80997	83183	hypothetical protein
94	83238	83507	hypothetical protein
95	83511	83756	hypothetical protein
96	83753	84079	hypothetical protein
97	84066	84338	putative acyl carrier protein
98	84441	85286	hypothetical protein
99	85394	85624	hypothetical protein
100	85716	86165	hypothetical protein

101	86227	86607	hypothetical protein
102	86604	86909	hypothetical protein
103	86937	87053	hypothetical protein
104	87062	87256	hypothetical protein
105	87249	87524	hypothetical protein
106	87597	88061	putative superinfection exclusion protein
107	88064	88393	hypothetical protein
108	88390	89130	hypothetical protein
109	89216	89488	putative histone family DNA-binding protein
110	89630	91348	putative ATP-dependent DNA helicase
111	91349	92116	hypothetical protein
112	92164	92703	putative ribonuclease H
113	92719	93483	putative late transcription sigma factor
114	93474	94592	putative recombination-related endonuclease
115	94595	96919	putative recombination endonuclease subunit
116	96951	97247	hypothetical protein
117	97228	97542	hypothetical protein
118	97555	98166	putative RegB endoribonuclease
119	98166	98762	hypothetical protein
120	98759	98902	hypothetical protein
121	99005	99334	hypothetical protein
122	99420	101501	hypothetical protein
123	101554	102120	hypothetical protein
124	102131	102373	hypothetical protein
125	102460	103068	hypothetical protein
126	103065	104132	putative DNA primase subunit
127	104134	104361	hypothetical protein
128	104452	104727	hypothetical protein
129	104903	105412	hypothetical protein
130	105472	105795	hypothetical protein
131	105805	106014	hypothetical protein
132	106054	106848	putative peptidoglycan binding protein
133	106959	107795	putative PhoH-like phosphate starvation-inducible protein

134	107906	110206	putative NrdA ribonucleoside-diphosphate reductase alpha subunit
135	110280	111398	putative NrdB ribonucleoside-diphosphate reductase beta subunit
136	111395	111628	putative glutaredoxin
137	111728	112171	hypothetical protein
138	112168	112611	hypothetical protein
139	112608	112988	putative baseplate wedge subunit
140	112988	114808	putative baseplate hub subunit and tail lysozyme
141	115312	116073	putative baseplate hub subunit
142	116126	116680	hypothetical protein
143	116681	117163	hypothetical protein
144	117209	118003	putative RuvC-like holliday junction resolvase
145	117984	118274	hypothetical protein
146	118261	118506	hypothetical protein
147	118499	118735	putative later promoter transcription factor
148	118747	118983	hypothetical protein
149	119092	120153	putative ssDNA binding protein
150	120179	121117	putative baseplate tail tube protein
151	121168	121866	putative DNA end protector protein
152	121924	122730	hypothetical protein
153	122750	123079	hypothetical protein
154	123171	124103	hypothetical protein
155	124171	124830	putative kinase
156	124830	126044	putative thymidylate synthase
157	126056	126619	hypothetical protein
158	126616	127209	putative dUTP diphosphatase
159	127211	127780	hypothetical protein
160	127768	128844	putative UvsX RecA-like recombination protein
161	129235	130581	putative DNA primase-helicase subunit
162	130679	130957	putative GTPase-activator protein
163	130947	131243	hypothetical protein
164	131971	133191	hypothetical protein
165	133195	133287	hypothetical protein
166	133352	133963	hypothetical protein

167	134030	134461	hypothetical protein
168	134822	135043	hypothetical protein
169	135040	135321	hypothetical protein
170	135321	137018	putative DNA ligase
171	137062	137436	hypothetical protein
172	137433	137633	putative transcriptional regulator
173	137638	138321	putative DNA helicase loader
174	138324	140213	putative tail length tape measure protein
175	140224	141618	hypothetical protein
176	141611	142165	putative baseplate wedge subunit
177	142177	143142	putative baseplate tail tube cap protein
178	143182	143811	putative head completion protein
179	143803	144243	hypothetical protein
180	144309	145244	putative deoxycytidylate deaminase
181	145237	145614	hypothetical protein
182	145617	146180	putative alpha hydrolase
183	146405	146968	putative metallophosphatase
184	146968	147369	hypothetical protein
185	147429	147854	hypothetical protein
186	147854	148078	hypothetical protein
187	148078	148446	hypothetical protein
188	148443	149072	putative DexA exonuclease
189	149045	149650	hypothetical protein
190	149647	149913	hypothetical protein
191	149967	150389	hypothetical protein
192	150398	150616	hypothetical protein
193	150613	151998	putative DNA topoisomerase/gyrase small subunit
194	152077	153978	putative DNA topoisomerase/gyrase large subunit
195	154006	154569	hypothetical protein
196	154566	155057	hypothetical protein
197	155103	155300	hypothetical protein
198	155345	155716	putative histone-like protein
199	155953	156093	hypothetical protein
200	156093	156869	putative tail fibre protein

201	156857	157084	hypothetical protein
202	157274	157672	hypothetical protein
203	157754	159385	hypothetical protein

Table A.6 Annotation table for JA15 (Genbank reference KY942056.1)

A.7 JA29 genome annotation table

ORF	Start	End	Annotation
1	4	1467	putative replicative DNA helicase DnaB
2	1565	2239	hypothetical protein
3	2223	5036	putative terminase
4	5100	5315	hypothetical protein
5	5328	7271	putative portal protein
6	7271	7579	hypothetical protein
7	7581	7955	hypothetical protein
8	8002	8463	hypothetical protein
9	8460	9497	putative DNA polymerase I
10	9509	9958	hypothetical protein
11	9964	10548	putative O-acetyl-ADP-ribose deacetylase
12	10564	10878	hypothetical protein
13	10884	11528	putative membrane protein
14	11789	12055	putative DNA primase
15	12226	12483	hypothetical protein
16	12492	13334	putative DNA adenine methylase
17	13348	13491	hypothetical protein
18	13499	14218	hypothetical protein
19	14211	14444	hypothetical protein
20	14518	15096	hypothetical protein
21	15135	15461	hypothetical protein
22	15448	15942	putative CMP deaminase
23	15952	16404	hypothetical protein
24	16407	17282	hypothetical protein
25	17275	18039	putative thymidylate synthase
26	18173	18973	putative GIY-YIG family homing endonuclease
27	19151	19504	putative thymidylate synthase
28	19545	20270	hypothetical protein
29	20270	20917	hypothetical protein
30	21259	21498	hypothetical protein
31	21488	21892	hypothetical protein
32	21938	22642	hypothetical protein

33	22614	23084	hypothetical protein
34	23077	24042	hypothetical protein
35	24023	24886	hypothetical protein
36	24936	25271	hypothetical protein
37	25249	25536	hypothetical protein
38	25577	25888	hypothetical protein
39	25891	26046	hypothetical protein
40	26054	26467	hypothetical protein
41	26464	27183	hypothetical protein
42	27173	27484	hypothetical protein
43	27481	27840	hypothetical protein
44	27849	28601	hypothetical protein
45	28603	28863	hypothetical protein
46	28924	29565	hypothetical protein
47	29688	30053	hypothetical protein
48	30043	30402	hypothetical protein
49	30413	30880	hypothetical protein
50	30942	31532	hypothetical protein
51	31535	31756	hypothetical protein
52	31753	32022	hypothetical protein
53	32031	32462	hypothetical protein
54	32470	33309	hypothetical protein
55	33321	33722	putative ASCH domain-containing protein
56	33698	34531	hypothetical protein
57	34518	35048	hypothetical protein
58	35041	35361	hypothetical protein
59	35348	35671	hypothetical protein
60	35691	36032	hypothetical protein
61	36200	36445	putative bifunctional (p)ppGpp synthetase/guanosine-3' 5'-bis diphosphate 3'-pyrophosphohydrolase protein
62	36483	37067	hypothetical protein
63	37071	37466	hypothetical protein
64	37435	37665	hypothetical protein
65	37673	38134	hypothetical protein

66	38328	38927	hypothetical protein
67	38924	39289	hypothetical protein
68	39286	39996	hypothetical protein
69	39998	40702	hypothetical protein
70	40737	41492	hypothetical protein
71	41531	41998	putative membrane protein
72	41995	42276	hypothetical protein
73	42284	42613	putative membrane protein
74	42624	42863	hypothetical protein
75	42881	43513	hypothetical protein
76	43513	44421	hypothetical protein
77	44432	44986	hypothetical protein
78	44983	46065	hypothetical protein
79	46062	46658	hypothetical protein
80	46658	47581	hypothetical protein
81	47581	48048	hypothetical protein
82	48062	49081	hypothetical protein
83	49108	50682	hypothetical protein
84	50683	51180	putative membrane protein
85	51263	52315	hypothetical protein
86	52478	53854	putative T1SS secreted agglutinin RTX
87	53873	54622	hypothetical protein
88	54619	55110	putative membrane protein
89	55112	55387	hypothetical protein
90	55424	56170	hypothetical protein
91	56173	56970	hypothetical protein
92	56963	57421	hypothetical protein
93	57423	58241	putative tail fibre protein
94	58253	58453	hypothetical protein
95	58498	62031	putative ILEI domain-containing protein
96	62040	64061	hypothetical protein
97	64074	64568	putative tail fibre protein
98	64582	65214	putative tail fibre protein
99	65225	66199	putative tail protein

100	66196	68517	hypothetical protein
101	68604	69674	hypothetical protein
102	69720	74243	hypothetical protein
103	74240	75703	putative baseplate wedge subunit protein
104	75700	76113	putative baseplate protein
105	76113	76403	putative baseplate spike protein
106	77095	77406	hypothetical protein
107	78410	80188	hypothetical protein
108	80198	80638	hypothetical protein
109	80623	81216	putative dTMP kinase
110	81221	81709	putative MmcB-like DNA repair protein
111	81699	82151	putative NUDIX hydrolase
112	82148	82654	hypothetical protein
113	82952	83578	hypothetical protein
114	83588	84376	putative baseplate protein/tail-associated lysozyme
115	84373	86256	hypothetical protein
116	86260	86481	hypothetical protein
117	86488	86874	hypothetical protein
118	86874	87167	hypothetical protein
119	87140	87943	hypothetical protein
120	87943	90741	putative VgrG-like protein/endolysin
121	90741	91583	hypothetical protein
122	91576	92262	hypothetical protein
123	92273	92785	putative tail tube protein
124	92788	93456	hypothetical protein
125	93499	94014	putative tail tube protein
126	94029	95714	putative tail sheath protein
127	95772	96116	hypothetical protein
128	96118	96780	hypothetical protein
129	96851	97324	hypothetical protein
130	97420	98514	putative major capsid protein
131	98568	99272	putative structural protein
132	99337	101325	putative ATPase
133	101409	102512	hypothetical protein

134	102512	103330	putative prohead core protease
135	103337	103759	hypothetical protein
136	103764	104651	hypothetical protein
137	104659	105759	putative glycosyl transferase
138	105770	106684	hypothetical protein
139	106684	108654	putative DNA ligase
140	108678	109601	hypothetical protein
141	109598	110440	hypothetical protein
142	110440	112365	hypothetical protein
143	112425	116387	hypothetical protein
144	116451	117605	hypothetical protein
145	117624	118598	hypothetical protein
146	118602	119366	hypothetical protein
147	119382	122465	putative major tail protein/T1SS secreted agglutinin RTX
148	122455	123099	hypothetical protein
149	123096	123722	hypothetical protein
150	123744	125420	putative tail sheath protein
151	125508	126554	hypothetical protein
152	126542	127396	hypothetical protein
153	127401	127610	hypothetical protein
154	127607	128239	hypothetical protein
155	128275	128556	hypothetical protein
156	128573	128956	hypothetical protein
157	128922	129260	hypothetical protein
158	129266	130876	putative DNA repair helicase
159	130879	131439	hypothetical protein
160	131441	131869	hypothetical protein
161	131866	133041	hypothetical protein
162	133028	133297	hypothetical protein
163	133294	136449	putative DNA polymerase I
164	136544	136861	hypothetical protein
165	136864	137418	hypothetical protein
166	137462	143296	putative ATP-dependent DNA helicase
167	143307	143807	hypothetical protein

168	143820	144587	hypothetical protein
169	144587	144946	putative HNH family endonuclease
170	144953	145717	hypothetical protein
171	145777	146580	hypothetical protein
172	146633	147793	putative head to tail joining protein
173	147796	148893	hypothetical protein
174	148893	150167	hypothetical protein
175	150308	150973	hypothetical protein
176	151109	152239	putative recombination-related endonuclease
177	152353	152754	putative ssDNA binding protein
178	152924	154885	hypothetical protein
179	154882	156669	hypothetical protein
180	156680	157045	putative DUF2778 domain-containing protein
181	157045	157611	hypothetical protein
182	157630	158691	hypothetical protein
183	158695	159162	hypothetical protein
184	159159	159734	putative glycosyl hydrolase
185	159814	160137	hypothetical protein
186	160137	160850	hypothetical protein
187	160841	161104	hypothetical protein
188	161104	163404	putative exonuclease
189	163407	163607	hypothetical protein
190	163692	163925	hypothetical protein
191	163925	164359	hypothetical protein
192	164362	165039	hypothetical protein
193	165159	165638	hypothetical protein
194	165718	166938	putative DNA polymerase III
195	167276	167851	hypothetical protein
196	167851	168183	hypothetical protein
197	168173	169150	hypothetical protein
198	169195	169671	hypothetical protein
199	169679	170173	hypothetical protein
200	170185	170406	hypothetical protein
201	170696	171835	hypothetical protein

202	171873	173510	hypothetical protein
203	173970	174521	hypothetical protein
204	174524	175642	hypothetical protein
205	175786	176334	putative holliday junction resolvase
206	176347	176988	hypothetical protein
207	177047	177685	hypothetical protein
208	177682	179667	putative autotransporter beta-barrel domain-containing protein
209	179789	180469	hypothetical protein
210	180503	181567	putative DNA primase
211	181633	181968	hypothetical protein
212	182006	183022	putative exonuclease
213	183030	183431	hypothetical protein
214	183885	184547	hypothetical protein
215	184550	184714	hypothetical protein
216	184714	184998	hypothetical protein
217	184995	185318	hypothetical protein
218	185362	185538	hypothetical protein
219	185541	186059	hypothetical protein
220	186056	186505	putative cyclic phosphodiesterase
221	186489	186770	hypothetical protein
222	186806	187843	hypothetical protein
223	187904	189088	putative ssDNA binding protein
224	189143	190648	putative RecA protein
225	190690	191091	hypothetical protein
226	191209	191988	hypothetical protein
227	191985	192563	hypothetical protein
228	192547	193080	hypothetical protein
229	193161	193409	hypothetical protein
230	193475	195292	hypothetical protein
231	195345	195938	hypothetical protein
232	195877	196518	hypothetical protein
233	196566	197207	hypothetical protein
234	197207	197524	hypothetical protein
235	197502	198671	hypothetical protein

236	198685	198885	putative methyltransferase
237	198864	200087	putative DNA adenine methylase
238	200265	200600	hypothetical protein
239	200661	201140	hypothetical protein
240	201140	201655	hypothetical protein
241	201645	202070	hypothetical protein
242	202117	202314	hypothetical protein
243	202314	202532	hypothetical protein
244	202535	203293	hypothetical protein
245	203560	204153	hypothetical protein
246	204137	204487	hypothetical protein
247	204496	204846	hypothetical protein
248	204857	205105	hypothetical protein
249	205274	205546	hypothetical protein
250	205608	206021	hypothetical protein
251	206085	206429	hypothetical protein
252	206511	206924	hypothetical protein
253	206970	207260	hypothetical protein
254	207270	208061	hypothetical protein
255	208058	208375	hypothetical protein
256	208490	209161	hypothetical protein
257	209158	209376	Hypothetical protein
258	209798	210298	hypothetical protein
259	210300	211016	hypothetical protein
260	211016	211258	hypothetical protein
261	211270	211716	hypothetical protein
262	211719	212339	hypothetical protein
263	212336	212878	putative RNA 2'-phosphotransferase
264	212983	213813	hypothetical protein
265	213855	214112	putative DksA/TraR family C4-type zinc finger protein
266	214115	214252	hypothetical protein
267	214252	214674	hypothetical protein
268	214692	215102	hypothetical protein
269	215163	216098	putative UvsE UV damage endonuclease

270	216089	216190	hypothetical protein
271	216199	217011	hypothetical protein
272	217019	217630	hypothetical protein
273	217631	219796	hypothetical protein
274	219847	220512	hypothetical protein
275	220661	220840	hypothetical protein
276	220844	221092	hypothetical protein
277	221096	221554	hypothetical protein
278	221557	221853	hypothetical protein
279	221863	222150	hypothetical protein
280	222205	222645	hypothetical protein
281	222681	222974	hypothetical protein
282	222977	223543	putative dUTPase
283	223921	224439	putative lytic transglycosylase
284	224587	224883	hypothetical protein
285	224895	225407	hypothetical protein
286	225606	226313	hypothetical protein
287	226313	226840	hypothetical protein
288	226837	227289	hypothetical protein
289	227290	227496	hypothetical protein
290	227503	227907	hypothetical protein
291	227955	228578	hypothetical protein
292	228678	229283	hypothetical protein
293	229280	229456	hypothetical protein
294	229453	229860	hypothetical protein
295	229871	230704	hypothetical protein
296	230722	231504	hypothetical protein
297	231658	232008	hypothetical protein
298	232158	232895	hypothetical protein
299	232963	233343	hypothetical protein
300	233343	233708	hypothetical protein
301	233996	235612	hypothetical protein
302	235685	236614	hypothetical protein
303	236614	236988	hypothetical protein

304	236985	237461	hypothetical protein
305	237433	238701	hypothetical protein
306	238701	241661	hypothetical protein
307	241717	242469	hypothetical protein
308	242492	243067	hypothetical protein
309	243273	244073	hypothetical protein
310	244245	244766	hypothetical protein
311	244859	245200	hypothetical protein
312	245320	247395	putative DNA topoisomerase IV/gyrase subunit B
313	247395	249077	putative DNA topoisomerase 4 subunit A
314	249227	250417	hypothetical protein
315	250407	250949	hypothetical protein
316	250921	251760	Hypothetical protein
317	251822	252955	hypothetical protein
318	252963	253319	hypothetical protein

Table A.7 Annotation table for JA29 (Genbank reference MH460461)

A.8 JA33 genome annotation table

ORF	Start	End	Annotation
1	44	1423	putative replicative DNA helicase DnaB
2	1521	2195	hypothetical protein
3	2179	4992	putative terminase
4	5058	5273	hypothetical protein
5	5286	7229	putative portal protein
6	7229	7537	hypothetical protein
7	7539	7913	hypothetical protein
8	7961	8425	hypothetical protein
9	8422	9459	putative DNA polymerase I
10	9471	9920	hypothetical protein
11	9926	10513	putative O-acetyl-ribose deacetylase
12	10522	10836	hypothetical protein
13	10842	11486	putative membrane protein
14	11747	12013	putative DNA primase
15	12262	12441	hypothetical protein
16	12450	13292	putative DNA adenine methylase
17	13308	13451	hypothetical protein
18	13460	14182	hypothetical protein
19	14175	14429	hypothetical protein
20	14510	15088	hypothetical protein
21	15088	15453	hypothetical protein
22	15440	15934	putative CMP deaminase
23	15944	16396	hypothetical protein
24	16399	17271	hypothetical protein
25	17264	18367	putative thymidylate synthase
26	18409	19119	hypothetical protein
27	19119	19766	hypothetical protein
28	20115	20354	hypothetical protein
29	20344	20748	hypothetical protein
30	20794	21498	hypothetical protein
31	21470	21940	hypothetical protein
32	21933	22892	hypothetical protein

33	22873	23736	hypothetical protein
34	23786	24121	hypothetical protein
35	24099	24386	hypothetical protein
36	24741	24896	hypothetical protein
37	24904	25317	hypothetical protein
38	25314	26033	hypothetical protein
39	26023	26334	hypothetical protein
40	26331	26690	hypothetical protein
41	26699	27451	hypothetical protein
42	27453	27713	hypothetical protein
43	27775	28416	hypothetical protein
44	28536	28901	hypothetical protein
45	28891	29250	hypothetical protein
46	29261	29728	hypothetical protein
47	29789	30382	hypothetical protein
48	30385	30597	hypothetical protein
49	30594	30863	hypothetical protein
50	30871	31302	hypothetical protein
51	31302	31829	hypothetical protein
52	31804	32643	hypothetical protein
53	32652	32906	hypothetical protein
54	32903	33304	putative ASCH domain-containing protein
55	33280	34038	hypothetical protein
56	34025	34555	hypothetical protein
57	34548	34865	hypothetical protein
58	34852	35175	hypothetical protein
59	35707	35952	hypothetical protein
60	35990	36574	putative bifunctional (p)ppGpp synthetase/guanosine-3
61	36578	36973	hypothetical protein
62	37181	37642	hypothetical protein
63	37839	38438	hypothetical protein
64	38435	38800	hypothetical protein
65	38797	39504	hypothetical protein
66	39506	40228	hypothetical protein

67	40265	41056	hypothetical protein
68	41095	41562	putative membrane protein
69	41559	41840	hypothetical protein
70	41848	42177	putative membrane protein
71	42188	42427	hypothetical protein
72	42445	43077	hypothetical protein
73	43077	43985	hypothetical protein
74	43996	44550	hypothetical protein
75	44547	45629	hypothetical protein
76	45626	46219	hypothetical protein
77	46219	47142	hypothetical protein
78	47142	47609	hypothetical protein
79	47623	48642	hypothetical protein
80	48669	50243	hypothetical protein
81	50244	50741	putative membrane protein
82	50824	51876	hypothetical protein
83	52039	53415	putative T1SS secreted agglutinin RTX
84	53434	54183	hypothetical protein
85	54180	54671	putative membrane protein
86	54673	54948	hypothetical protein
87	54984	55730	hypothetical protein
88	55733	56530	hypothetical protein
89	56523	56981	hypothetical protein
90	56983	57801	putative tail fibre protein
91	57813	58013	hypothetical protein
92	58058	61591	putative ILEI domain-containing protein
93	61600	63621	hypothetical protein
94	63634	64128	putative tail fibre protein
95	64142	64774	putative tail fibre protein
96	64785	65759	putative tail protein
97	65756	68077	hypothetical protein
98	68165	69235	hypothetical protein
99	69279	73802	hypothetical protein
100	73799	75262	putative baseplate wedge subunit protein

101	75259	75672	putative baseplate protein
102	75672	75962	putative baseplate spike
103	76655	76966	hypothetical protein
104	77703	79481	hypothetical protein
105	79491	79931	hypothetical protein
106	79916	80509	putative dTMP kinase
107	80514	80951	putative MmcB-like DNA repair protein/ transcription elongation factor
108	80992	81444	putative NUDIX hydrolase domain-containing protein
109	81441	81947	hypothetical protein
110	82245	82871	hypothetical protein
111	82884	83669	putative baseplate protein/tail-associated lysozyme
112	83666	85549	hypothetical protein
113	85553	85774	hypothetical protein
114	85781	86179	hypothetical protein
115	86179	86340	hypothetical protein
116	86445	87248	hypothetical protein
117	87248	90040	putative VgrG-like protein/endolysin
118	90040	90882	hypothetical protein
119	90875	91561	hypothetical protein
120	91572	92084	putative tail tube protein
121	92087	92755	hypothetical protein
122	92798	93313	putative tail tube protein
123	93328	95013	putative tail sheath protein
124	95071	95415	hypothetical protein
125	95417	96079	hypothetical protein
126	96151	96624	hypothetical protein
127	96716	97810	putative major capsid protein
128	97863	98567	putative structural protein
129	98637	100595	putative ATPase
130	100679	101782	hypothetical protein
131	101779	102600	putative prohead core protein protease/endolysin
132	102607	103029	hypothetical protein
133	103034	103921	hypothetical protein

134	103929	105029	putative glycosyl transferase
135	105040	105954	hypothetical protein
136	105954	107924	putative DNA ligase
137	107948	108871	hypothetical protein
138	108868	109710	hypothetical protein
139	109710	111641	hypothetical protein
140	111701	115663	hypothetical protein
141	115725	116879	hypothetical protein
142	116898	117872	hypothetical protein
143	117876	118640	hypothetical protein
144	118656	121739	putative major tail protein/T1SS secreted agglutinin RTX
145	121729	122373	hypothetical protein
146	122370	122996	hypothetical protein
147	123018	124694	putative tail sheath protein
148	124782	125828	hypothetical protein
149	125816	126670	hypothetical protein
150	126675	126884	hypothetical protein
151	126881	127513	hypothetical protein
152	127550	127831	hypothetical protein
153	127848	128231	hypothetical protein
154	128197	128535	hypothetical protein
155	128541	129434	putative DNA repair helicase
156	130511	131647	putative DNA repair helicase
157	131650	132213	hypothetical protein
158	132215	132643	hypothetical protein
159	132640	133815	hypothetical protein
160	133802	134071	hypothetical protein
161	134068	137223	putative DNA polymerase I
162	137318	137635	hypothetical protein
163	137638	138192	hypothetical protein
164	138236	144070	putative ATP-dependent DNA helicase
165	144081	144581	hypothetical protein
166	144594	145361	hypothetical protein
167	145361	145720	putative HNH family endonuclease

168	145727	146485	hypothetical protein
169	146546	147349	hypothetical protein
170	147402	148562	putative head to tail joining protein
171	148565	149662	hypothetical protein
172	149662	150939	hypothetical protein
173	151080	151742	hypothetical protein
174	151878	153008	putative recombination-related endonuclease
175	153123	153647	putative ssDNA binding protein
176	153694	155658	hypothetical protein
177	155655	157442	hypothetical protein
178	157453	157818	putative DUF2778 domain-containing protein
179	157818	158384	hypothetical protein
180	158403	159395	hypothetical protein
181	159398	159865	hypothetical protein
182	159862	160437	putative glycosyl hydrolase
183	160517	160840	hypothetical protein
184	160840	161553	hypothetical protein
185	161556	161807	hypothetical protein
186	161807	164107	putative exonuclease
187	164110	164310	hypothetical protein
188	164391	164627	hypothetical protein
189	164644	165078	hypothetical protein
190	165081	165758	putative SAM-dependent methyltransferase
191	165879	166358	hypothetical protein
192	166438	167658	putative DNA ploymerase III
193	168001	168576	hypothetical protein
194	168576	168908	hypothetical protein
195	168898	169875	hypothetical protein
196	169918	170385	hypothetical protein
197	170392	170886	hypothetical protein
198	170898	171362	hypothetical protein
199	171411	172499	hypothetical protein
200	172537	173754	hypothetical protein
201	174617	175168	hypothetical protein

202	175171	175347	hypothetical protein
203	175379	176416	hypothetical protein
204	176560	177108	putative holliday-junction resolvase
205	177121	177762	hypothetical protein
206	177821	178459	hypothetical protein
207	178456	180441	putative inverse autotransporter beta-barrel domain-containing protein
208	180563	181240	hypothetical protein
209	181276	182331	putative DNA primase
210	182397	182732	hypothetical protein
211	182770	183786	putatuve exonuclease
212	183794	184195	hypothetical protein
213	184649	185311	hypothetical protein
214	185314	185478	hypothetical protein
215	185490	185762	hypothetical protein
216	185759	186076	hypothetical protein
217	186119	186295	hypothetical protein
218	186298	186819	hypothetical protein
219	186816	187265	putative cyclic phosphodiesterase
220	187249	187530	hypothetical protein
221	187566	188603	hypothetical protein
222	188664	189869	putative ssDNA binding protein
223	189924	191429	putative RecA protein
224	191471	191872	hypothetical protein
225	191989	192768	hypothetical protein
226	192765	193343	hypothetical protein
227	193327	193860	hypothetical protein
228	193941	194189	hypothetical protein
229	194253	196070	hypothetical protein
230	196123	196716	hypothetical protein
231	196655	197296	hypothetical protein
232	197344	197985	hypothetical protein
233	197985	198302	hypothetical protein
234	198280	199449	putative methyltransferase

235	199505	199651	hypothetical protein
236	199630	200853	putative DNA adenine methylase
237	201030	201365	hypothetical protein
238	201428	201907	hypothetical protein
239	201907	202422	hypothetical protein
240	202412	202837	hypothetical protein
241	202806	203081	hypothetical protein
242	203081	203299	hypothetical protein
243	203301	204059	hypothetical protein
244	204326	204919	hypothetical protein
245	204903	205253	hypothetical protein
246	205262	205612	hypothetical protein
247	205624	205872	hypothetical protein
248	206043	206315	hypothetical protein
249	206380	206808	hypothetical protein
250	206868	207215	hypothetical protein
251	207212	207505	hypothetical protein
252	207515	208300	hypothetical protein
253	208297	208905	hypothetical protein
254	208902	209621	hypothetical protein
255	209623	210303	hypothetical protein
256	210352	211023	hypothetical protein
257	211020	211238	hypothetical protein
258	211660	212160	hypothetical protein
259	212162	212878	hypothetical protein
260	212878	213120	hypothetical protein
261	213240	213578	hypothetical protein
262	213581	214201	hypothetical protein
263	214198	214740	putative RNA 2
264	214843	215235	hypothetical protein
265	215245	215502	putative DksA/TraR family C4-type zinc finger protein
266	215505	215813	hypothetical protein
267	215816	215956	hypothetical protein
268	215956	216390	hypothetical protein

269	216409	216858	hypothetical protein
270	216894	217829	putative UvsE UV damage repair endonuclease
271	217820	217921	hypothetical protein
272	217930	218739	hypothetical protein
273	218750	219358	hypothetical protein
274	219359	221308	hypothetical protein
275	221361	222020	hypothetical protein
276	222156	222335	hypothetical protein
277	222339	222587	hypothetical protein
278	222591	223049	hypothetical protein
279	223052	223348	hypothetical protein
280	223361	223654	hypothetical protein
281	223709	224152	hypothetical protein
282	224155	224445	hypothetical protein
283	224445	225011	putative dUTPase
284	225389	225907	putative lytic transglycosylase
285	226054	226347	hypothetical protein
286	226363	226875	hypothetical protein
287	226888	227031	hypothetical protein
288	227074	227781	hypothetical protein
289	227781	228305	hypothetical protein
290	228302	228757	hypothetical protein
291	228758	228964	hypothetical protein
292	228971	229375	hypothetical protein
293	229426	230049	hypothetical protein
294	230150	230755	hypothetical protein
295	230752	230928	hypothetical protein
296	230925	231332	hypothetical protein
297	231343	232176	hypothetical protein
298	232193	232975	hypothetical protein
299	233130	233480	hypothetical protein
300	233631	234374	hypothetical protein
301	234430	234960	hypothetical protein
302	234967	235344	hypothetical protein

303	235344	235709	hypothetical protein
304	235995	237608	hypothetical protein
305	237681	238607	hypothetical protein
306	238607	238969	hypothetical protein
307	238978	239454	hypothetical protein
308	239426	240694	hypothetical protein
309	240694	243666	hypothetical protein
310	243721	244473	hypothetical protein
311	244496	245074	hypothetical protein
312	245279	246070	hypothetical protein
313	246248	246766	hypothetical protein
314	246860	247201	hypothetical protein
315	247320	249380	putative DNA topoisomerase IV/gyrase subunit B
316	249380	251062	putative DNA topoisomerase 4 subunit A
317	251216	252406	hypothetical protein
318	252396	252938	hypothetical protein
319	252910	253749	hypothetical protein
320	253811	254944	hypothetical protein
321	254952	255308	hypothetical protein

Table A.8 Annotation table for JA33 (Genbank reference MH460462)

Appendix Two

Published papers

Experimental work described in this dissertation has directly contributed to two published papers that are included in full for reference. These papers are listed below. The first paper includes work that has been described in Chapter Three and the second paper is largely derived from data presented in Chapter Six.

- Andrew Day, Jiyoon Ahn, Xinzhe Fang, and George P C Salmond. Environmental Bacteriophages of the Emerging Enterobacterial Phytopathogen, *Dickeya solani*, Show Genomic Conservation and Capacity for Horizontal Gene Transfer between Their Bacterial Hosts. *Frontiers in Microbiology*, 2017. doi: 10.3389/fmicb.2017.01654
- Andrew Day, Jiyoon Ahn, and George P. C. Salmond. Jumbo Bacteriophages Are Represented Within an Increasing Diversity of Environmental Viruses Infecting the Emerging Phytopathogen, *Dickeya solani*. *Frontiers in Microbiology*, 2018. doi: 10.3389/fmicb.2018.02169

frontiers in Microbiology

ORIGINAL RESEARCH
published: 30 August 2017
doi: 10.3389/fmicb.2017.01654

Environmental Bacteriophages of the Emerging Enterobacterial Phytopathogen, *Dickeya solani*, Show Genomic Conservation and Capacity for Horizontal Gene Transfer between Their Bacterial Hosts

*Andrew Day, Jiyoon Ahn, Xinzhe Fang and George P. C. Salmond **

Department of Biochemistry, University of Cambridge, Cambridge, United Kingdom

OPEN ACCESS

Edited by:
Helene Sanfacon, Agriculture and Agri-Food Canada, Canada

Reviewed by:
Evelien M. Adriaenssens, University of Liverpool, United Kingdom
Ananda Shankar Bhattacharjee, Bigelow Laboratory for Ocean Sciences, United States

***Correspondence:**
George P. C. Salmond
gpcs2@cam.ac.uk

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *24 May 2017*
Accepted: *15 August 2017*
Published: *30 August 2017*

Citation:
Day A, Ahn J, Fang X and Salmond GPC (2017) Environmental Bacteriophages of the Emerging Enterobacterial Phytopathogen, Dickeya solani, Show Genomic Conservation and Capacity for Horizontal Gene Transfer between Their Bacterial Hosts. Front. Microbiol. 8:1654. doi: 10.3389/fmicb.2017.01654

Dickeya solani is an economically important phytopathogen widespread in mainland Europe that can reduce potato crop yields by 25%. There are no effective environmentally-acceptable chemical systems available for diseases caused by *Dickeya*. Bacteriophages have been suggested for use in biocontrol of this pathogen in the field, and limited field trials have been conducted. To date only a small number of bacteriophages capable of infecting *D. solani* have been isolated and characterized, and so there is a need to expand the repertoire of phages that may have potential utility in phage therapy strategies. Here we describe 67 bacteriophages from environmental sources, the majority of which are members of the viral family *Myoviridae*. Full genomic sequencing of two isolates revealed a high degree of DNA identity with *D. solani* bacteriophages isolated in Europe in the past 5 years, suggesting a wide ecological distribution of this phage family. Transduction experiments showed that the majority of the new environmental bacteriophages are capable of facilitating efficient horizontal gene transfer. The possible risk of unintentional transfer of virulence or antibiotic resistance genes between hosts susceptible to transducing phages cautions against their environmental use for biocontrol, until specific phages are fully tested for transduction capabilities.

Keywords: *Dickeya solani*, bacteriophage, environmental viruses, phytopathogen, horizontal gene transfer

1. INTRODUCTION

The enterobacterial genus, *Dickeya*, currently consists of six phytopathogenic species that can cause severe disease in economically important crops, including tomato, chicory, and potato (Reverchon and Nasser, 2013). The first report of *Dickeya* (previously known as *Erwinia chrysanthemi*) infecting European potatoes came from the Netherlands in the 1970s (Maas Geesteranus HP, 1972). Until 2004, almost all European potato isolates of *Dickeya* were assigned as *Dickeya dianthicola*, which has a broad host range across both nutritional and ornamental species (Toth et al., 2011). In the past

10 years however, three groups have independently identified a new clade of *Dickeya* in European potato isolates (Laurila et al., 2008; Parkinson et al., 2009; Sławiak et al., 2009). In 2014 this led to the classification of a new species; *Dickeya solani* (van der Wolf et al., 2014).

Dickeya solani is more aggressive than other *Dickeya* species, able to spread more easily through the plant vascular system and survive at higher temperatures than *D. dianthicola* (Toth et al., 2011). It is currently the predominant potato pathogen in Europe and in 2010 Scotland became the first country to introduce specific legislation aimed at preventing the establishment of *D. solani* in its seed industry (Mansfield et al., 2012).

In Israel a reduction in yield of up to 25% was observed in potatoes exposed to *Dickeya* species (Tsror et al., 2009). This imposes a significant economic cost, and consequently has led to research into methods for control of these virulent phytopathogens. In the absence of any effective chemical control systems, bacterial viruses (bacteriophages; phages) have been suggested as potential biocontrol tools and several studies have isolated phages capable of infecting *Dickeya* species (Adriaenssens et al., 2012b; Czajkowski et al., 2014, 2015; Matilla et al., 2014). Their potential use as biocontrol agents has been trialed both in the lab and in the field (Adriaenssens et al., 2012b) and these studies showed a "therapeutic" outcome with an increase in yield of the potato crop. Because of the potential utility of specific phages as therapeutic agents in potato soft rot control experiments, there is value in investigating a wider range of *Dickeya* phages. However, prior work has shown that a previously isolated *D. solani* phage is capable of generalized transduction of both chromosomal and plasmid markers (Matilla et al., 2014). The European Medicines Agency, among others, has stated that it is "important to ensure that therapeutic phages do not carry out generalized transduction" (Pelfrene et al., 2016), and therefore this is an important consideration as some *Dickeya* phages may not have been fully tested for generalized transduction capacity before field trials. This study therefore aimed to isolate and characterize a larger repertoire of new environmental phages against *D. solani* and investigate their potential for generalized transduction.

2. RESULTS

2.1. Isolation and Classification

Sixty-seven phages were isolated using standard enrichment techniques from both treated sewage effluent and river water between 2013 and 2015 using *D. solani* MK10 as the host organism. Transmission electron microscopy (TEM) showed two different morphological groups, a selection of which are shown in **Figure 1** alongside the previously characterized phage LIMEstone1 (Adriaenssens et al., 2012b).

Of 24 phages imaged, all possessed an icosahedral head and a tail, placing them in the order *Caudovirales*. Three possessed short tails, characteristic of the family *Podoviridae* (such as ϕXF28 in the last panel of **Figure 1**) whilst the rest possessed longer contractile tails and belong to the family *Myoviridae*. The 21 *Myoviridae* members did not appear to possess the tail fibers characteristic of the family. Instead, short tail spikes were observed, (first three panels of **Figure 1**), and these are generally associated with the family *Podoviridae*.

2.2. Transduction

Other *Dickeya* phages with similar morphology have been described and were shown to be efficient generalized transducing phages (Matilla et al., 2014). Due to the transduction capability of certain phages shown by Matilla et al., the 67 newly-isolated phages were also tested for ability to affect horizontal gene transfer. Of these, 51 (including the 21 phages with the non-classical morphology) proved capable of transducing chromosomal markers. Twelve of the isolates, all of which had the non-classical morphology, were also tested for generalized transduction, and proved capable of transferring plasmids between *Dickeya* species. The results of transduction of the plasmid pBR322 by three of these phages are shown in **Figure 2**.

2.3. Host Range

Based on the bacterial strains tested, LIMEstone1 is capable of forming plaques on strains of *D. solani* only (Adriaenssens et al., 2012b). The phages isolated during this study were tested against a variety of *Dickeya* strains, listed in **Table 1**, to determine their host range. The majority of the phages presented here exhibited the same host range as LIMEstone1 and were only capable of forming plaques on *D. solani* strains but not isolated representatives of other *Dickeya* species used in this study. However, eight of the phage isolates had a wider host range extending to species such as *Dickeya dieffenbachiae*, *Dickeya paradisiaca*, and *Dickeya zeae* (**Table 2**).

2.4. Genetic Comparisons

Two key genes known to be conserved between these phages, those for DNA polymerase (DNAP) and tail spike protein 1 (TSP1), were sequenced for several of the newly-isolated phages. These nucleotide sequences were then compared to those of LIMEstone1 as shown in **Table 3**. All of the phages in **Table 3** were able to form plaques on *D. solani*. The corresponding amino acid sequences were compared between these phages and phylogenetic trees were created as shown in **Figure 3** (DNAP) and **Figure 4** (TSP1). These show that, in agreement with **Table 3**, the DNAP genes of ϕXF4 and ϕXF11 grouped together away from the other phages with a branch length of 0.053. The other phages formed two clusters that differed in a single amino acid. The TSP1 genes formed two distinct clusters, with the genes from ϕXF16 and ϕJA1 forming their own cluster with a branch length of 0.092, whereas the other phages all had identical TSP1 genes. The final 20 amino acids were trimmed from the TSP1 genes as the sequencing data for some of the phages was insufficient for tree construction.

2.5. Genomic Sequencing of Two New *Dickeya solani* Phages

ϕXF4 and ϕJA15 were isolated over a year apart yet they shared the same host range and PCR amplification and preliminary sequence analysis showed 100% DNA identity in the TSP1 genes, although they differ in their DNAP genes. The full genomes of both phages were then sequenced. Both consist of circular

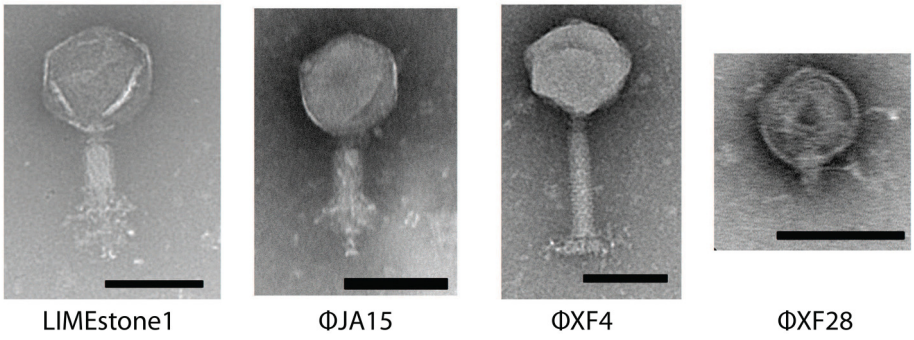


FIGURE 1 | Transmission electron micrographs of four *Dickeya* phages. LIMEstone1 is a previously characterized *Dickeya* phage (Adriaenssens et al., 2012b). ϕJA15 and ϕXF4 are phages of the *Myoviridae* family exhibiting tail spikes as opposed to tail fibers. ϕXF28 is a phage of the *Podoviridae* family. Scale bars represent 100 nm.

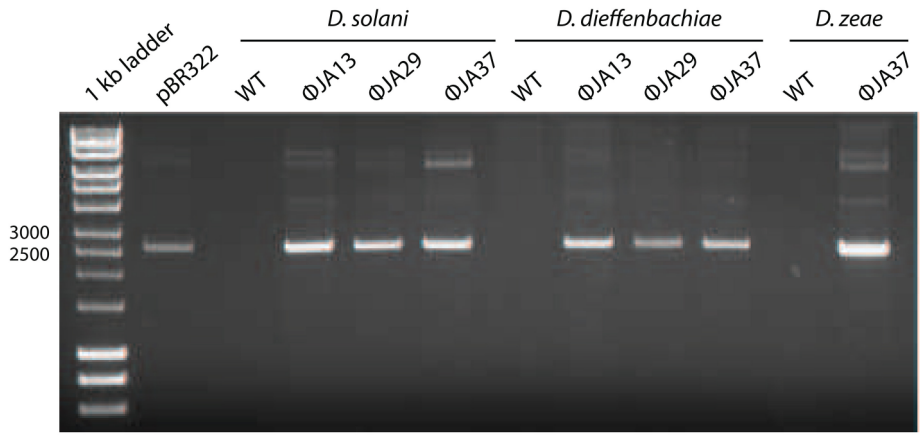


FIGURE 2 | pBR322 was first used to transform *Dickeya solani* MK10 before the creation of lysates of ϕJA13, ϕJA29, or ϕJA37 on the recombinant host. These lysates were then used to infect *Dickeya solani*, *D. dieffenbachiae*, or *D. zeae* as appropriate and transductants selected on LB agar containing ampicillin. The plasmids from these transductants were extracted and analyzed by gel electrophoresis. Co-migration of the plasmid DNA samples and absence of the plasmid from the wild type (WT) controls confirmed successful plasmid transduction.

double-stranded DNA of 151,519 and 153,757 bp, respectively. The genome of ϕXF4 has a G+C content of 49.4% and contained 185 predicted genes with lengths ranging between 116 and 4,838 nucleotides, as shown in **Figure 5**. ϕJA15 has a G+C content of 49.2% and contained 188 predicted genes of lengths between 122 and 4,838 nucleotides, as shown in **Figure 6**. Full annotation tables for the two genomes can be found in Tables S1, S2, respectively.

2.6. Genomic Comparison

The genomes of the two new phages shared 97% DNA identity, with the main areas of difference being regions encoding endonucleases and hypothetical proteins. A comparison of both of these phages with the previously published phage LIMEstone1, showed over 95% DNA identity, with the major areas of difference consisting of genes thought to be involved in DNA replication (such as homing endonucleases and polymerases) as well as introns located throughout all three genomes. These areas of difference are highlighted in **Figures 5**, **6**.

3. DISCUSSION

The Scottish government tests all seed crops imported from outside Scotland plus 10% of Scottish-origin crops for *D. solani* and did not find any positive samples in 2016 (Scottish Government, 2016). These data are consistent with a view that *D. solani* is not yet environmentally widespread within the UK, although there have been isolated cases of *D. solani* reported in England and Wales since 2007 (Cahill et al., 2010) in crops originating from outside of the UK (Toth et al., 2016). The relative ease with which we have isolated environmental phages that infect *D. solani* therefore seems counter-intuitive given the reported paucity of the pathogen in the environment. In a restricted host range screen these phages did not form plaques on eight isolates of other Gram-negative laboratory strains such as *Pectobacterium carotovorum*, *Pectobacterium atrosepticum*, *Serratia plymuthica*, *Serratia marcescens*, *Escherichia coli*, and *Pantoea agglomerans* (data not shown). The apparently contradictory observations from phage isolations and host distribution beg the ecological question as to why *D. solani*

TABLE 1 | Bacterial strains, bacteriophages, and primers used in this study.

Bacterial strain	References
Dickeya solani MK10	Pritchard et al., 2013a
Dickeya dianthicola NCPBB 453	Pritchard et al., 2013a
Dickeya dieffenbachiae NCPBB 2976	Pritchard et al., 2013b
Dickeya paradisiaca NCPBB	Pritchard et al., 2013b
Dickeya zeae NCPBB 3532	Pritchard et al., 2013b
Dickeya chrysanthemi NCPBB 402	Pritchard et al., 2013b
Bacteriophages	**Reference**
LIMEstone1	Adriaenssens et al., 2012b
Primer name	**Sequence**
oJA1	GGTTGAGGTTCATTTCTTGC
oJA2	AACGACAGGAGATTCTTYAT
oJA14	AACCACTGTTGGATTTGTCACAAGC
oJA15	AACGTCCAGTAGGGTGGAGCAT

TABLE 2 | Extended host range of three groups of the isolated phages capable of infecting other species of *Dickeya*.

Dickeya species	ϕJA10, 11, and 32	ϕJA13, 33, and 37	ϕJA29 and 31
D. dieffenbachiae NCPBB 2976	+	+	+
D. paradisiaca NCPBB 2511	−	−	+
D. dianthicola NCPBB 453	+	−	−
D. zeae NCPBB 3532	−	+	+
D. chrysanthemi NCPBB 402	+	−	−

[+]*Denotes isolated plaque formation of the phages on the respective host.*

TABLE 3 | Nucleotide comparison of two conserved genes between the previously characterized LIMEstone1 and a selection of isolated phages.

	% Identity to LIMEstone1	
Phage	**DNAP**	**TSP1**
ϕXF4	90.6	100
ϕXF11	90.6	100
ϕXF16	100	83.9
ϕJA1	100	83.8
ϕJA15	99.7	100
ϕJA17	99.7	99.9
ϕJA19	99.7	100
ϕJA21	99.7	100
ϕJA23	99.7	100

phages are easy to find by simple enrichments. We would suggest that either *D. solani* is present in the environment around Cambridge and is not being detected, or that there is an, as yet unknown, alternative host(s) for these phages present in the environment.

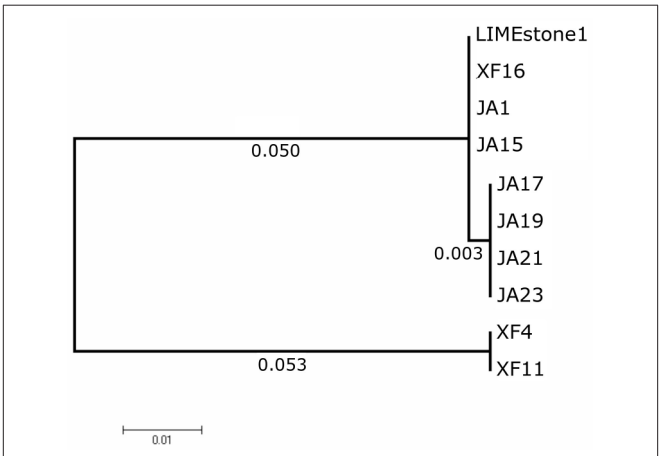


FIGURE 3 | Phylogenetic tree of the DNA Polymerase (DNAP) gene from 10 phages of *Dickeya solani*. Tree was constructed using the Maximum Likelihood method with 1,134 positions in the final dataset, and the tree shown has the highest log likelihood (−2033.82).

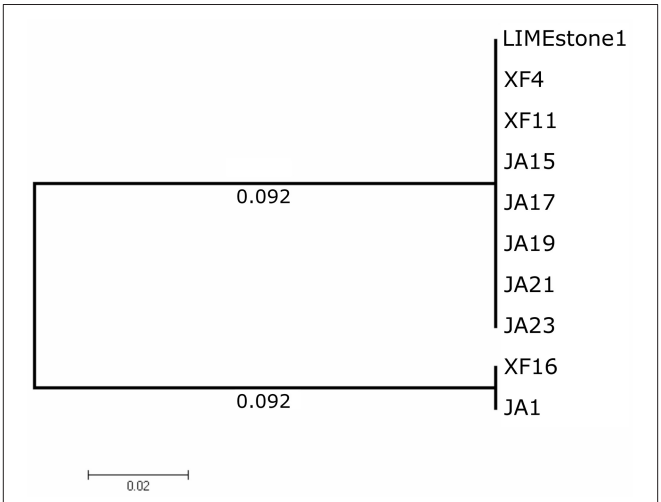


FIGURE 4 | Phylogenetic tree of the Tail Spike Protein 1 (TSP1) gene from 10 phages of *Dickeya solani*. Tree was constructed using the Maximum Likelihood method with 1,593 positions in the final dataset, and the tree shown has the highest log likelihood (−3105.91). The final 20 amino acids of the TSP1 gene were trimmed from the alignment as the sequencing data for some of the phages was insufficient.

All but three of the phages imaged by TEM were morphologically characterized as *Myoviridae* due to the presence of an icosahedral head and a contractile tail. Classical *Myoviridae* members, such as the coliphage T4 possess long slender tail fibers attached to the baseplate that participate in adsorption of the phage to the bacterial host. The imaged phages do not appear to have tail fibers, but instead possess shorter, clustered structures more akin to the tail spikes present in members of the *Podoviridae*. Genome analysis of the phages ϕXF4 and ϕJA15 showed genes encoding potential tail spike proteins, which show 100% sequence identity to putative

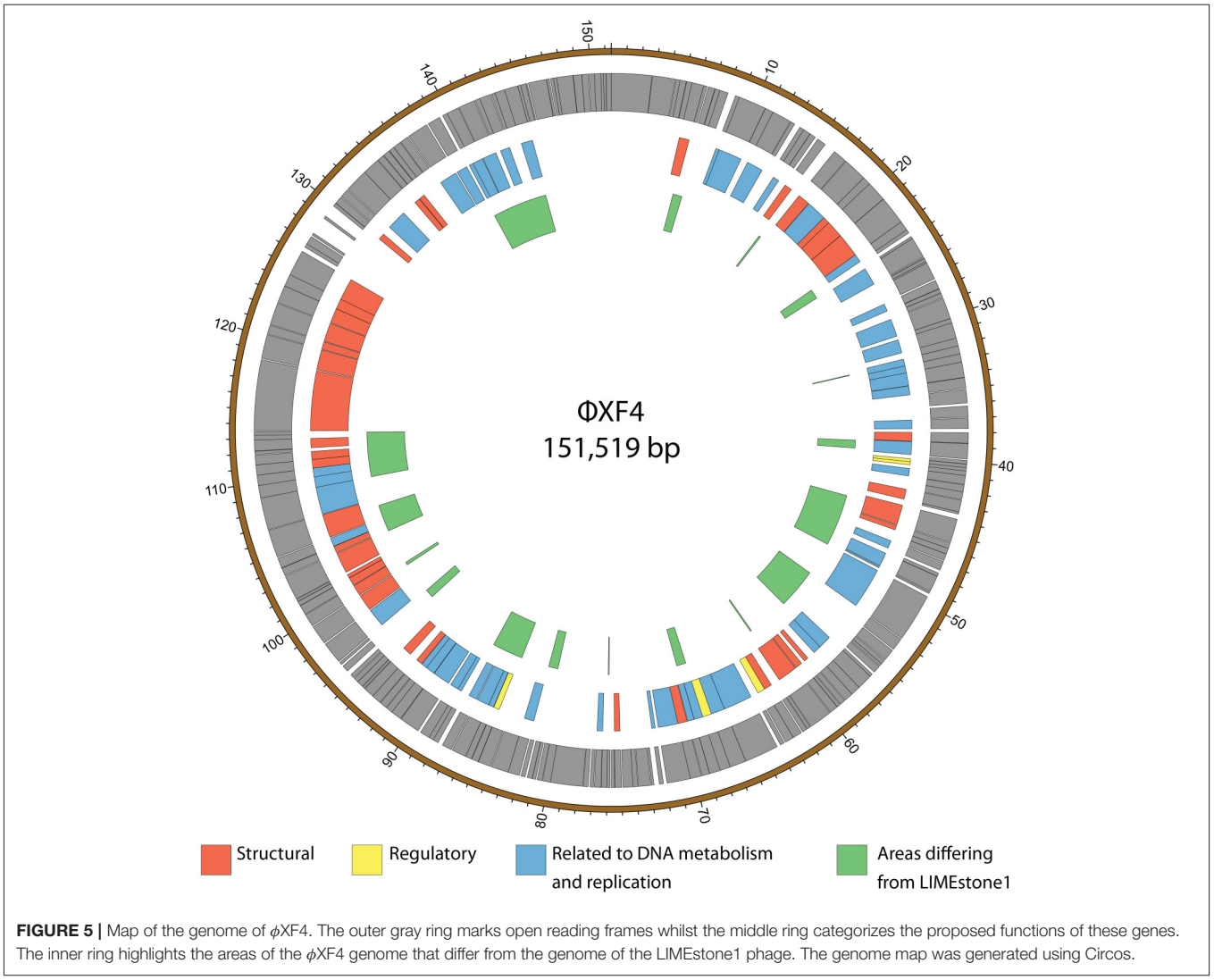


FIGURE 5 | Map of the genome of ϕXF4. The outer gray ring marks open reading frames whilst the middle ring categorizes the proposed functions of these genes. The inner ring highlights the areas of the ϕXF4 genome that differ from the genome of the LIMEstone1 phage. The genome map was generated using Circos.

tail spike protein genes in LIMEstone1. This combination of a *Myoviridae*-like morphology with tail spikes has been reported previously as a feature of the novel viral genus termed viunalikevirus (Adriaenssens et al., 2012a), named after the ViI *Salmonella* typing phage, and includes virulent phages capable of infecting a wide range of Gram-negative hosts. Members of the genus share a high degree of genome order and identity, with the major region of divergence being the genes encoding the tail spike proteins. Although, we cannot state that all phages that exhibit this morphology are definitively members of the viunalikevirus genus, we conclude it is likely. This does pose the question of whether there is some particular connection between *D. solani* and viunalikeviruses, or whether the environment around Cambridge is a particularly abundant source of this genus of phages.

A comparison of two genes from a subset of the phages isolated here, along with LIMEstone1, showed that, whilst there were some differences in DNA polymerase and TSP1 genes, these did not translate into a difference in host range and that in general these phages are highly similar. Several *D. solani* phages have now been isolated from environmental sources, including the LIMEstone phages (Adriaenssens et al., 2012b) as well as ϕD3 (Czajkowski et al., 2015) and ϕD5 (Czajkowski et al., 2014). They are all remarkably similar upon comparison of their genomes, and the two phages discussed in this study show 95% DNA identity with LIMEstone1. Much of the variation comes within the many introns and homing endonucleases found throughout the genomes of the phages, whereas structural elements are largely conserved. These introns and homing endonucleases vary in their sequence, but their position within the genomes of the phages, and thus the gene order, remains the same. It is interesting that these phages have been isolated independently in countries across Europe, and even from different environments; soil (LIMEstone1 and 2, ϕD3, and ϕD5) and water (ϕXF4 and ϕJA15), and yet they share such conservation despite their wide geographical separation.

The lytic activity of these phages, coupled with the high economic burden of *D. solani* crop phytopathogenesis, have

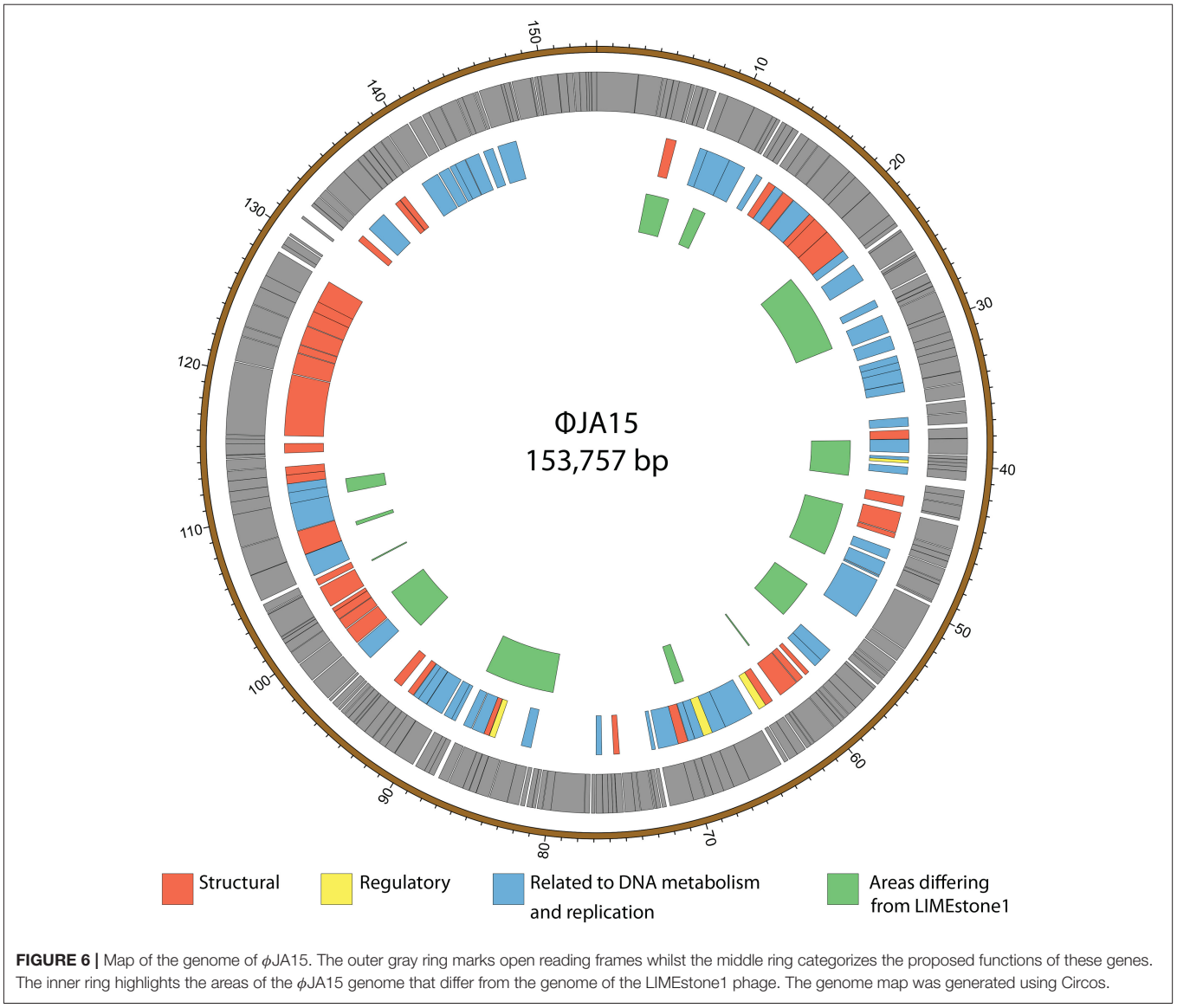


FIGURE 6 | Map of the genome of ϕJA15. The outer gray ring marks open reading frames whilst the middle ring categorizes the proposed functions of these genes. The inner ring highlights the areas of the ϕJA15 genome that differ from the genome of the LIMEstone1 phage. The genome map was generated using Circos.

highlighted the *D. solani* phages as potential biocontrol agents, and limited field trials have been performed (Adriaenssens et al., 2012b). Nevertheless, although these phages may have potential phage therapy features, we have shown that, 51 were capable of effecting horizontal gene transfer. We showed previously that eight candidate viunalikeviruses (including the LIMEstone phages and ϕXF4) were efficient generalized transducers of plasmid markers (Matilla et al., 2014). Consequently, we proposed that generalized transduction capacity is a characteristic trait of the viunalikevirus genus. This feature is important if these virulent phages are to be used therapeutically, as there could be potential (albeit low) for collateral transfer of bacterial virulence genes or drug resistance plasmids into unintended hosts, with unknown consequences—depending on ecological selection pressures. We therefore consider it prudent to caution against further field trials until the specific phage(s) involved are fully tested for generalized transduction capabilities.

4. MATERIALS AND METHODS

4.1. Bacterial Strains, Phages, Culture Media, and Growth Conditions

All bacterial strains used in this study are listed in **Table 1**. *Dickeya* species were routinely grown at 30°C in Luria broth (LB) or on LB agar plates (1.5%, wt/vol, agar). Phages were stored at 4°C in phage buffer (10 mM Tris-HCl, pH 7.4, 10 mM $MgSO_4$, 0.01%, wt/vol, gelatin) over a few drops of $NaHCO_3^-$ saturated chloroform.

4.2. Isolation of Phages

Treated sewage effluent was collected from a sewage treatment plant in Cambridge, United Kingdom (Matilla and Salmond, 2014). River water was collected from multiple sites along the River Cam. Samples were filter sterilized before 5 mL of the sample was added to 2x LB along with 500 µL of an overnight

culture of *D. solani* MK10. This mixture was incubated overnight in a 250 mL flask at 30°C with shaking at 250 rpm. One milliliter of the enriched sample was mixed with 100 μL of chloroform (saturated with sodium hydrogen carbonate) and vortexed to lyse bacterial cells. The sample was centrifuged at 16,000 x g for 4 min and 10 μL of a serial dilution series of the supernatant was mixed with 200 μL of an overnight bacterial culture and 4 ml of LB top agar. This mixture was poured as an overlay on an LBA plate and incubated overnight at 30°C. Single phage plaques were picked with a sterile toothpick, placed into 100 μL phage buffer, and shaken with 40 μL of chloroform to kill any bacteria. The phages obtained were plaque purified three times. High-titre phage lysates were then obtained as described previously (Petty et al., 2006). Briefly, 10-fold serial dilutions of the phage were incubated overnight in an agar overlay as already described. Those plates exhibiting confluent lysis (seen as a mosaic-like effect in which the plaques are close to merging) were used for lysate preparation. The top agar was removed from the plate, vortexed with chloroform before sedimentation at 2,200 x g for 20 min at 4°C. The supernatant was removed and vortexed with a few drops of chloroform to produce the final lysate.

4.3. Transmission Electron Microscopy

High-titre lysates for transmission electron microscopy were obtained as described previously (Petty et al., 2006) using 0.35% (w/v) LB agarose instead of 0.35% (w/v) LB agar overlays. Ten microliters of high-titer phage lysates were adsorbed onto 400-mesh copper grids with holey carbon support films (Agar Scientific, Stansted, United Kingdom) for 30 min. The copper grids were discharged in a Quorum/Emitech K100X system (Quorum, Ringmer, United Kingdom) prior to use. After 1 min, excess phage suspension was removed with filter paper and phage samples were negatively stained by placing the grids for 5 min in 10 μL of 2% phosphotungstic acid (PTA) neutralized with sodium hydroxide, or with 10 μL of 2% uranyl acetate for 2 min. The grids were then blotted on filter paper to remove the excess solution and allowed to air dry for 10 min. Phages were examined by transmission electron microscopy in the Multi-Imaging Centre (Department of Physiology, Development and Neuroscience, University of Cambridge) using an FEI Tecnai G2 transmission electron microscope (FEI, OR, USA). The accelerating voltage was 120.0 kV, and images were captured with anAMT XR60B digital camera running Deben software.

4.4. Determination of Host Range

The host range of isolated phages was determined by plating out 10-fold serial dilutions of the phage lysates, onto agar overlays containing the six species of *Dickeya* listed in **Table 1**. To avoid potential confusion with "lysis from without," only phages that produced lysis at low dilution and individual plaques were considered as being able to infect the host.

4.5. Transduction

To test for transduction, phage lysates were generated on donor bacterial strains carrying the desired plasmid or chromosomal marker as already described. Transduction was performed by mixing phage lysate with an overnight culture of the recipient host to achieve a multiplicity of infection of 0.1, meaning that for each phage there were 10 bacterial cells. The mixture was left on the lab bench at room temperature for 20 min, followed by incubation on a rotary wheel at 30°C for 30 min. The infected culture was then centrifuged and the bacterial pellets washed with LB twice to eliminate any remaining non-adsorbed phage. The bacterial pellets were resuspended in 1 mL LB and 100 μL aliquots were spread onto LBA plates with drug selection for the chromosomal or plasmid marker. Appropriate standard controls, which were routinely negative, were used to score for any spontaneous resistance of the recipient strain. One hundred microliters of the phage lysate was also spread onto LBA plates to confirm lysate sterility.

4.6. Gene Amplification and Analysis

Genomic DNA was extracted using Phase Lock Gel tubes (5 Prime, Hamburg, Germany) following manufacturer's instructions for isolation of Lambda DNA. DNA Polymerase (DNAP) and Tail Spike Protein 1 (TSP1) genes were amplified using Phusion polymerase (ThermoScientific, MA, USA) following standard protocols. TSP1 was amplified using the primers oJA1 and oJA2, and DNAP by the primers oJA14 and oJA15, listed in **Table 1**. PCR products were sequenced by GATC Biotech AG (Konstanz, Germany). Sequences were compared using NCBI Blast and the Artemis Comparison Tool (Carver et al., 2005). Phylogenetic trees were constructed using MEGA 7.0.26 (Kumar et al., 2016).

4.7. Genome Sequencing and Analysis

ϕXF4 was sequenced on the Illumina Bench Top MiSeq Sequencer (Illumina, CA, USA) at the DNA Sequencing Facility, Department of Biochemistry, University of Cambridge, UK. The resulting 138,803 reads were trimmed, quality assessed and assembled using Geneious 7.1.5 (Biomatters Ltd.), leading to higher than 100x coverage of the full genome. Gaps or single nucleotide polymorphisms were further filled or verified by Sanger sequencing to produce one contig. ϕJA15 was sequenced on the Illumina MiSeq Sequencer at MicrobesNG (Birmingham, UK). The 454,086 reads were trimmed using Trimmomatic (Bolger et al., 2014), assessed for quality using BWA-MEM (Li, 2013) and assembled using SPAdes 3.7.1 (Bankevich et al., 2012) with standard settings, leading to higher than 100x coverage of the full genome and producing one contig. Annotation of both genomes was performed using Prokka 1.11 (Seemann, 2014) using standard settings and LIMEstone1 (NC_019925.1) as a scaffold. Genome maps were generated using Circos (Krzywinski et al., 2009). Genomes were deposited in Genbank using Sequin (NCBI) and are available under accession numbers KY942057 (XF4) and KY942056 (JA15). Genomes were compared using NCBI Blast and the Artemis Comparison Tool (Carver et al., 2005).

AUTHOR CONTRIBUTIONS

Analyzed the data, conceived and designed the experiments: AD, JA, XF, and GS. Performed the experiments: AD, JA, and XF. Wrote the paper: AD and GS.

FUNDING

This work was supported by the BBSRC, UK. AD was supported by a Cambridge Doctoral Training Partnership Award from the BBSRC, UK.

ACKNOWLEDGMENTS

Sequencing of ϕJA15 was conducted by MicrobesNG (http://www.microbesng.uk), which is supported by the BBSRC (grant number BB/L024209/1). We are grateful to Ian Toth (James Hutton Institute, Scotland) for generous provision of *Dickeya* strains. This work was done under DEFRA license: 50864/197900/3. We thank Alison Rawlinson for technical support and Rita Monson and Miguel Matilla for helpful advice.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2017.01654/full#supplementary-material

REFERENCES

Adriaenssens, E. M., Ackermann, H. W., Anany, H., Blasdel, B., Connerton, I. F., Goulding, D., et al. (2012a). A suggested new bacteriophage genus: "Viunalikevirus". *Arch. Virol.* 157, 2035–2046. doi: 10.1007/s00705-012-1360-5

Adriaenssens, E. M., van Vaerenbergh, J., Vandenheuvel, D., Dunon, V., Ceyssens, P. J., de Proft, M., et al. (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

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Cahill, G., Fraser, K., Kowalewska, M. J., Kenyon, D. M., and Saddler, G. S. (2010). "Recent findings from the *Dickeya* survey and monitoring programme," in *Proceedings Crop Protection in Northern Britain* (Dundee), 171–176.

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

Czajkowski, R., Ozymko, Z., Siwinska, J., Ossowicki, A., de Jager, V., Narajczyk, M., et al. (2015). The complete genome, structural proteome, comparative genomics and phylogenetic analysis of a broad host lytic bacteriophage phiD3 infecting pectinolytic *Dickeya* spp. *Stand. Genomic Sci.* 10:68. doi: 10.1186/s40793-015-0068-z

Czajkowski, R., Ozymko, Z., Zwirowski, S., and Lojkowska, E. (2014). Complete genome sequence of a broad-host-range lytic *Dickeya* spp. bacteriophage phiD5. *Arch. Virol.* 159, 3153–3155. doi: 10.1007/s00705-014-2170-8

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33:msw054. doi: 10.1093/molbev/msw054

Laurila, J., Ahola, V., Lehtinen, A., Joutsjoki, T., Hannukkala, A., Rahkonen, A., et al. (2008). Characterization of *Dickeya* strains isolated from potato and river water samples in Finland. *Eur. J. Plant Pathol.* 122, 213–225. doi: 10.1007/s10658-008-9274-5

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint, arXiv:1303.3997.

Maas Geesteranus HP (1972). Natrot en zwartbenigheid bij aardappelen. *Bedrijfsontwikkeling* 3, 941–945.

Mansfield, J., Genin, S., Magori, S., Citovsky, V., Sriariyanum, M., Ronald, P., et al. (2012). Top 10 plant pathogenic bacteria in molecular plant pathology. *Mol. Plant Pathol.* 13, 614–629. doi: 10.1111/j.1364-3703.2012.00804.x

Matilla, M. A., Fang, X., and Salmond, G. P. C. (2014). Viunalikeviruses are environmentally common agents of horizontal gene transfer in pathogens and biocontrol bacteria. *ISME J.* 8, 2143–2147. doi: 10.1038/ismej.2014.150

Matilla, M. A., and Salmond, G. P. C. (2014). Bacteriophage phiMAM1, a viunalikevirus, is a broad-host-range, high-efficiency generalized transducer that infects environmental and clinical isolates of the enterobacterial genera *Serratia* and *Kluyvera*. *Appl. Environ. Microbiol.* 80, 6446–6457. doi: 10.1128/AEM.01546-14

Parkinson, N., Stead, D., Bew, J., Heeney, J., Tsror, L., and Elphinstone, J. (2009). *Dickeya* species relatedness and clade structure determined by comparison of recA sequences. *Int. J. Syst. Evol. Microbiol.* 59, 2388–2393. doi: 10.1099/ijs.0.009258-0.

Pelfrene, E., Willebrand, E., Cavaleiro Sanches, A., Sebris, Z., and Cavaleri, M. (2016). Bacteriophage therapy: a regulatory perspective. *J. Antimicrob. Chemother.* 71, 2071–2074. doi: 10.1093/jac/dkw083

Petty, N. K., Foulds, I. J., Pradel, E., Ewbank, J. J., and Salmond, G. P. C. (2006). A generalized transducing phage (phiIF3) for the genomically sequenced *Serratia marcescens* strain Db11: a tool for functional genomics of an opportunistic human pathogen. *Microbiology* 152, 1701–1708. doi: 10.1099/mic.0.28712-0

Pritchard, L., Humphris, S., Baeyen, S., Maes, M., Van Vaerenbergh, J., Elphinstone, J., et al. (2013a). Draft genome sequences of four *Dickeya dianthicola* and four *Dickeya solani* strains. *Genome Announc.* 1:e00087–12. doi: 10.1128/genomeA.00087-12

Pritchard, L., Humphris, S., Saddler, G. S., Elphinstone, J. G., Pirhonen, M., and Toth, I. K. (2013b). Draft genome sequences of 17 isolates of the plant pathogenic bacterium *Dickeya*. *Genome Announc.* 1:e00978–13. doi: 10.1128/genomeA.00978-13

Reverchon, S., and Nasser, W. (2013). *Dickeya ecology*, environment sensing and regulation of virulence programme. *Environ. Microbiol. Rep.* 5, 622–636. doi: 10.1111/1758-2229.12073

Scottish Government (2016). *2016 Inspection of Growing Crops of Seed and Ware Potatoes: Survey for the Presence of Dickeya spp.* Available online at: http://www.gov.scot/Topics/farmingrural/Agriculture/plant/18273/PotatoHealthControls/PotatoQuarantineDiseases/Dickeya/Report2015/2016Report

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Sławiak, M., van Beckhoven, J. R. C. M., Speksnijder, A. G. C. L., Czajkowski, R., Grabe, G., and van der Wolf, J. M. (2009). Biochemical and genetical analysis reveal a new clade of biovar 3 *Dickeya* spp. strains isolated from potato in Europe. *Eur. J. Plant Pathol.* 125, 245–261. doi: 10.1007/s10658-009-9479-2

Toth, I. K., Cahill, G., Elphinstone, J. G., Humphris, S., and Wale, S. J. (2016). "An update on the Potato Council/Scottish government-funded blackleg project - Year 2," in *Proceedings Crop Protection in Northern Britain 2016* (Dundee), 203–204.

Toth, I. K., van der Wolf, J. M., Saddler, G., Lojkowska, E., Hélias, V., Pirhonen, M., et al. (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

Tsror, L., Erlich, O., Lebiush, S., Hazanovsky, M., Zig, U., Slawiak, M., et al. (2009). Assessment of recent outbreaks of *Dickeya* sp. (syn. *Erwinia chrysanthemi*) slow wilt in potato crops in Israel. *Eur. J. Plant Pathol.* 123, 311–320. doi: 10.1007/s10658-008-9368-0

van der Wolf, J. M., Nijhuis, E. H., Kowalewska, M. J., Saddler, G. S., Parkinson, N., Elphinstone, J. G., et al. (2014). *Dickeya solani* sp. nov., a pectinolytic plant-pathogenic bacterium isolated from potato (Solanum tuberosum). *Int. J. Syst. Evol. Microbiol.* 64(Pt 3), 768–774. doi: 10.1099/ijs.0.052944-0

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 12 September 2018
doi: 10.3389/fmicb.2018.02169

Jumbo Bacteriophages Are Represented Within an Increasing Diversity of Environmental Viruses Infecting the Emerging Phytopathogen, *Dickeya solani*

*Andrew Day, Jiyoon Ahn and George P. C. Salmond**

Department of Biochemistry, University of Cambridge, Cambridge, United Kingdom

Dickeya species are economically important phytopathogens widespread in mainland Europe that can reduce crop yields by 25%. There are no effective environmentally-acceptable chemical systems available for diseases caused by *Dickeya*. Bacteriophages have been suggested for use in biocontrol of these pathogens in the field, and limited field trials have been conducted. To date the majority of bacteriophages capable of infecting *Dickeya solani*, one of the more aggressive species, are from the same family, the *Ackermannviridae*, many representatives of which have been shown to be unsuitable for use in the field due to their capacity for generalized transduction. Members of this family are also only capable of forming individual plaques on *D. solani*. Here we describe novel bacteriophages from environmental sources isolated on *D. solani*, including members of two other viral families; *Myoviridae* and *Podoviridae*, most of which are capable of forming plaques on multiple *Dickeya* species. Full genomic sequencing revealed that the *Myoviridae* family members form two novel clusters of jumbo bacteriophages with genomes over 250 kbp, with one cluster containing phages of another phytopathogen *Erwinia amylovora*. Transduction experiments showed that the majority of the new environmental bacteriophages are also capable of facilitating efficient horizontal gene transfer, however the single *Podoviridae* family member is not. This particular phage therefore has potential for use as a biocontrol agent against multiple species of *Dickeya*.

Keywords: *Dickeya solani*, bacteriophage, environmental viruses, phytopathogen, horizontal gene transfer, phage therapy, *Ackermannviridae*, jumbo bacteriophage

OPEN ACCESS

Edited by:
William Michael McShan, University of Oklahoma Health Sciences Center, United States

Reviewed by:
Victor Krylov, I.I. Mechnikov Research Institute of Vaccines and Sera (RAS), Russia
Elizabeth Martin Kutter, The Evergreen State College, United States
Olivia McAuliffe, Teagasc, The Irish Agriculture and Food Development Authority, Ireland

****Correspondence:***
George P. C. Salmond
gpcs2@cam.ac.uk

Specialty section:
This article was submitted to Virology, a section of the journal Frontiers in Microbiology

Received: *25 June 2018*
Accepted: *23 August 2018*
Published: *12 September 2018*

Citation:
Day A, Ahn J and Salmond GPC (2018) Jumbo Bacteriophages Are Represented Within an Increasing Diversity of Environmental Viruses Infecting the Emerging Phytopathogen, Dickeya solani. Front. Microbiol. 9:2169. doi: 10.3389/fmicb.2018.02169

1. INTRODUCTION

The genus *Dickeya*, recently reclassified into the novel family Pectobacteriaceae (Adeolu et al., 2016), currently consists of 11 phytopathogenic species that can cause severe disease in economically important crops including tomato, orchid, and potato (Alic et al., 2017a). Until 2004, almost all European potato isolates of *Dickeya* were assigned as *Dickeya dianthicola* (Toth et al., 2011). In 2008/2009, a new clade of *Dickeya* in European potato isolates was identified (Laurila et al., 2008; Parkinson et al., 2009; Sławiak et al., 2009) and in 2014 a new species was proposed; *Dickeya solani* (van der Wolf et al., 2014).

Dickeya solani is able to spread more easily through the plant vascular system and survive at higher temperatures than *D. dianthicola* (Toth et al., 2011). It is currently the predominant potato pathogen in Europe, with reductions in yield of up to 25% reported in potatoes exposed to *Dickeya* species (Tsror et al., 2009). Whilst there have been isolated cases of *D. solani* reported in England and Wales since 2007 (Cahill et al., 2010), these were all found in crops originating from outside of the UK (Toth et al., 2016). It is currently yet to become established in the UK, and to mitigate the significant economic cost inflicted by this virulent phytopathogen, the Scottish government has introduced specific legislation aimed at preventing the establishment of *D. solani* in its seed industry (Mansfield et al., 2012).

The significant economic costs inflicted by *Dickeya* species have stimulated research interest in methods for control of these virulent phytopathogens. Bacterial viruses (bacteriophages; phages) have been suggested as potential tools for biocontrol due to their specificity, environmental persistence and biological "organic" nature (Iriarte et al., 2007; Czajkowski et al., 2017; Svircev et al., 2018). Several studies have isolated phages capable of infecting *Dickeya* species (Adriaenssens et al., 2012c; Czajkowski et al., 2014a,b, 2015a; Matilla et al., 2014; Alič et al., 2017b; Day et al., 2017). Their potential use as biocontrol agents has been trialed both in the lab and in the field (Adriaenssens et al., 2012c) and these studies showed a partially "therapeutic" outcome with reduced crop losses. There is a commercial product available, Biolyse™, from APS Biocontrol Ltd that is a phage cocktail able to target *Pectobacterium* as well as *Dickeya* species. Designed as a washing solution for potatoes during factory processing, to our knowledge it is the first, and currently the only, commercial *Dickeya*-targeting biocontrol product. It has been reported that Biolyse™ has been used by the UK supermarket chain Tesco (Branston, 2012). The identities of the phages contained within this cocktail however have not been reported.

All of the *Dickeya* phages isolated so far, and 96% of all known phages, are members of the order *Caudovirales* (Fokine and Rossmann, 2014), which currently consists of four families. Apart from the *Siphoviridae* family member BF-CIM1/14 recently described by Alič et al. (2017b) and three *Podoviridae* family members reported in our recent publication (Day et al., 2017), the vast majority of *D. solani* phages characterized so far share a high degree of similarity and have been designated members of the *Ackermannviridae* family (formerly known as the *Vi1virus* genus; Adriaenssens et al., 2018) based on morphology and genomic comparisons. A summary of these phages is shown in **Table 1**. This has generated research interest, as these phages have been isolated from both soil and water samples and in three separate European countries; Belgium, Poland, and the United Kingdom. Host range testing has shown that the phages isolated in Belgium and the majority isolated in the UK are capable of forming plaques on strains of *D. solani* only (Adriaenssens et al., 2012c; Day et al., 2017). The phages isolated in Poland are reported to infect multiple species of *Dickeya* and *Pectobacterium* (Czajkowski et al., 2014a,b, 2015a), however, host range testing to the level of individual plaque formation has not been reported—an important criterion that allows exclusion of false positives (Khan Mirzaei and Nilsson, 2015). The high degree of morphological and genomic similarity between these phages and the other *Ackermannviridae* family members makes the reported broader host range that spans genera an intriguing prospect, assuming plaque formation data supporting this broader host range can be confirmed.

Sixty-seven phages were described in our recent publication (Day et al., 2017), 59 of which were only capable of forming plaques on *D. solani* species. When two were genomically sequenced they showed a high degree of similarity with the previously published *D. solani* phages of the *Ackermannviridae* family. The remaining eight phages were capable of forming plaques on other species of *Dickeya*, including *Dickeya zeae*, *Dickeya chrysanthemi*, and *Dickeya paradisiaca*. These particular phages warranted further investigation, as an expanded host range can be helpful for further application in phage therapy. The aim of this study therefore was to genomically sequence these phages to determine their similarity to previously published *D. solani* phages and their suitability for use in phage therapy.

2. RESULTS

2.1. Host Range and Plaque Morphology

The plaques of the previously described *D. solani* phages that are members of the *Ackermannviridae* family have tended to be clear, defined, and easy to distinguish from the bacterial top lawn (Adriaenssens et al., 2012c; Czajkowski et al., 2015a). This is the case for the other *Ackermannviridae* family members

TABLE 1 | Members of the *Ackermannviridae* family isolated on *Dickeya solani*.

Bacteriophage	Isolation	Location	Genome size (bp)	References
LIMEstone1	2008	Belgium (soil)	152,247	Adriaenssens et al., 2012c
D5	2012	Poland (soil)	155,346	Czajkowski et al., 2014b
PD10.3	2013	Poland (soil)	156,113*	Czajkowski et al., 2015a
PD23.1	2013	Poland (soil)	153,365*	Czajkowski et al., 2015a
D3	2013	Poland (soil)	152,308	Czajkowski et al., 2015b
XF4	2013	UK (waterway)	151,519	Day et al., 2017
JA15	2014	UK (waterway)	153,757	Day et al., 2017

**Genomes are marked incomplete, largest scaffold is reported and exhibits 99% DNA identity to LIMEstone1.*

isolated in this laboratory, and is also true of one of the broader host range phages JA10. However, the other seven have an indistinct, turbid plaque morphology that is often hard to distinguish from the bacterial top lawn (data not shown). We believe this to be the reason why the host range of the eight *D. solani* phages shown in **Table 2** is different from the host range reported in the previous publication (Day et al., 2017). This first became apparent due to confusing results generated by related, unpublished experiments that suggested a variation in the host range from that published in our previous paper (Day et al., 2017). Rigorous retesting has confirmed that the host range data presented in **Table 2** are accurate and that the previous interpretations were incorrect. The efficiency of plating data in **Table 2** shows that most of the phages are able to adsorb at a similar efficiency to all species of *Dickeya*, apart from JA29 which has an efficiency 10^{-4} lower on *D. paradisiaca* and *D. dadantii* subsp. *dieffenbachiae*.

2.2. Morphological Classification

Bacteriophages have traditionally been classified based on morphological characteristics viewed under electron microscopy (Ackermann, 2012). The majority of *D. solani* phages isolated to date are members of the *Ackermannviridae* family, which share common morphological characteristics. A representative of this family, XF4, is shown in **Figure 1A**. These phages possess an icosahedral head with a diameter of around 90 nm, a contractile tail around 110 nm in length and structures at the base of the tail that have been described as "stars" or "prongs" and have been identified as tail spikes (Adriaenssens et al., 2012a). Apart from these tail spikes, this is classical morphology of the phage family *Myoviridae*, therefore the combination of a contractile tail and tail spikes are the morphological markers of the family *Ackermannviridae*.

All eight of the expanded host range phages were viewed under transmission electron microscopy. Seven of them had an icosahedral head and long tail, with the structures at the base of the tail remaining unclear, with two representatives shown in **Figures 1C,D**. This marks them as either members of the *Myoviridae* or *Ackermannviridae* families, however, these phages were significantly larger than the previously viewed members of the *Ackermannviridae* family, which can be seen by comparing **Figure 1A** with **Figure 1D**. The head diameters were over 120 nm, with a tail length of around 150 nm, which suggested that these phages were not *Ackermannviridae* family members.

The indistinct morphology of the tail appendages of the JA jumbo phages (best seen in **Figure 1C**) has also been identified in other phages. When first described in the *Escherichia* phage 121Q (Ackermann and Nguyen, 1983) this morphology was presumed to be an artifact of microscopy involving damage to the tail. It was also thought that the dimensions, at the time reported to be a head diameter of 150 nm and a tail length of 165 nm, were overstated. However, this morphology has since been directly reported in the *Pseudomonas putida* phage Lu11 (Adriaenssens et al., 2012b), the *Pectobacterium carotovorum* phage CBB (Buttimer et al., 2017), and the *Erwinia amylovora* phage Y3 (Buttimer et al., in press), and has been dubbed the "hairy *Myoviridae*" morphology (Buttimer et al., 2017).

Unexpectedly, as can be seen in **Figure 1B**, the phage JA10 could be classified as a member of the *Podoviridae* family when imaged, characterized by an icosahedral head and a short non-contractile tail. Whilst this is not the first member of this family that we have isolated (Day et al., 2017), this is the first isolate we have studied in further depth. The genome of JA10 was therefore sequenced to investigate the similarity between it and previously published *Dickeya*-infecting *Podoviridae* family members (Ališ et al., 2017b).

2.3. Genome Sequence of *Podoviridae* Family Member JA10

The genome of JA10 is 40,131 bp, has 50 predicted genes, and is shown in **Figure 2**. The closest match in the database is an as yet unpublished *D. solani* phage Ninurta (Genbank reference: MH059639) isolated from organic waste in Denmark that shares 95% DNA identity with JA10. The closest published phage is the *Pectobacterium parmentieri* phage PP74 isolated from potato washing waste water in Russia in 2015 (Kabanova et al., 2018), which shares less than 14% nucleotide identity. It shares no DNA sequence identity with the other sequenced *Dickeya*-infecting *Podoviridae* family member BF25/12 (Ališ et al., 2017b). PP74 has been designated as a T7-like virus and a member of the *Autographivirinae* subfamily, with a conserved core genome. A translated nucleotide comparison of JA10 with the type phage T7 showed that most of the predicted genes are conserved (data not

TABLE 2 | Broader host range of eight phages capable of infecting other species of *Dickeya* as well as *Dickeya solani*.

***Dickeya* species**	**JA10**	**JA11, 31, 32, 33 and 37**	**JA13**	**JA29**
D. solani	1	1	1	1
D. dadantii subsp. *dieffenbachiae*	1.00×10^{-1}	$6.50 \times 10^{-1} \pm 6.30 \times 10^{-1}$	7.50×10^{0}	5.40×10^{-4}
D. paradisiaca	–	$6.30 \times 10^{-1} \pm 1.39 \times 10^{-1}$	5.80×10^{-1}	3.5×10^{-5}
D. dianthicola	–	$5.50 \times 10^{-1} \pm 3.25 \times 10^{-1}$	–	–
D. zeae	–	$1.12 \times 10^{0} \pm 5.80 \times 10^{-1}$	4.40×10^{-1}	–
D. chrysanthemi	1.30×10^{-1}	–	–	–

Efficiency of plating is relative and is calculated by dividing the titre of the phage on the host question by the titer of the phage on the original host D. solani. –Denotes no observable plaque formation of the phage on this host.

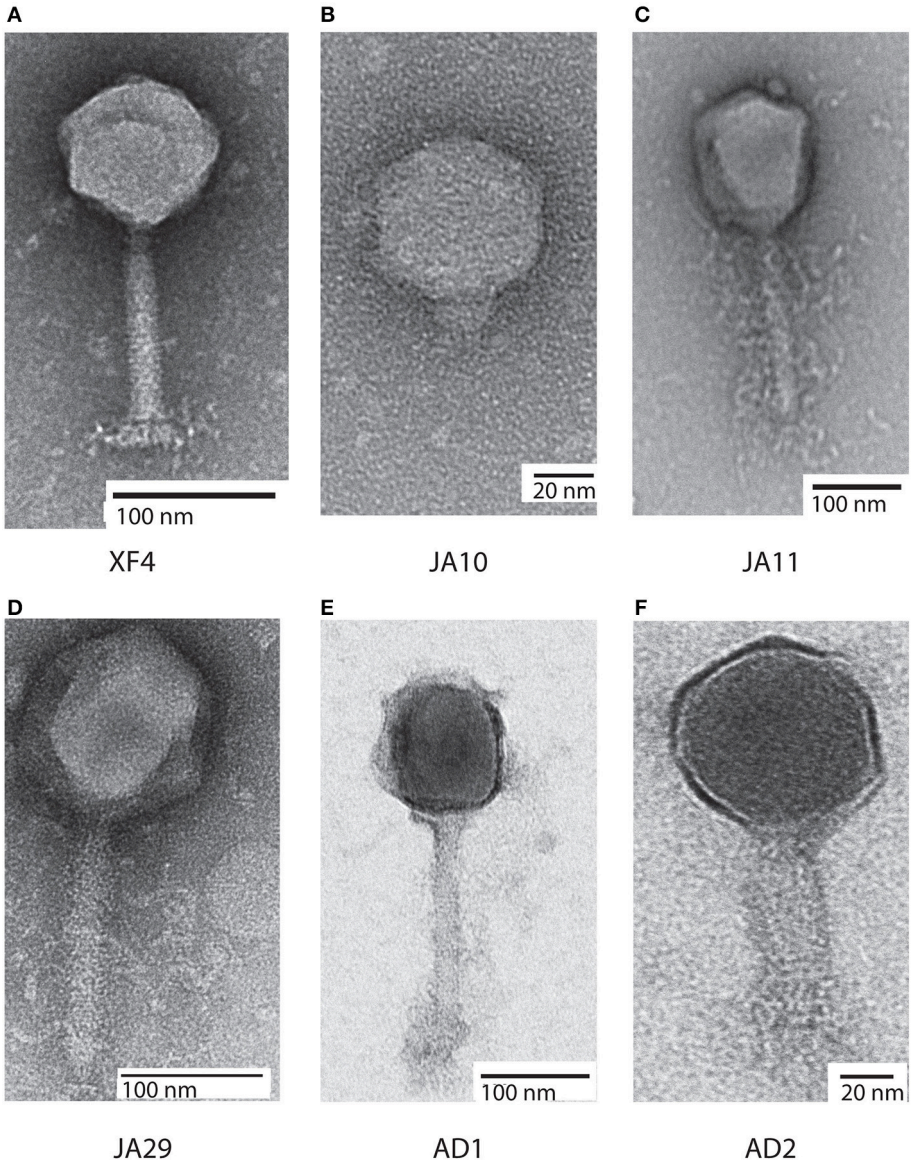


FIGURE 1 | Transmission electron micrographs of six *Dickeya solani* phages. XF4 **(A)** is a member of the *Ackermannviridae* family, exhibiting an icosahedral head of around 90 nm, a tail length of 110 nm and tail spikes at the base of the tail. JA10 **(B)** is a member of the *Podoviridae* family, characterized by a short stubby tail. JA11, JA29, and AD1 **(C–E)** are members of the *Myoviridae* family with unclear tail appendages. AD2 **(F)** has a similar morphology to XF4 and has a partially-contracted tail.

shown), including almost all the genes with a proposed function, such as the T3/T7-like RNA polymerase, structural capsid genes and DNA packaging machinery. JA10 is therefore also a member of the *Autographivirinae* subfamily.

There are several putative proteins encoded in the JA10 genome that do not have significant homology in the T7 genome and these are highlighted in **Figure 2**. Four of these seven putative proteins, the S-adenosyl-L-methionine hydrolase, bacterial RNA polymerase inhibitor, minor capsid protein, and the first tail assembly protein, have a protein of similar function encoded at this position in T7. The variation between JA10 and T7 in these putative proteins is therefore likely a determinant of host specificity. The marked tail fiber protein, which share a common N-terminal region but differ at the C-terminus between the two phages, is also likely a contributor to host range specificity. The DNA ligase highlighted in **Figure 2** neighbors a conserved ligase and consists of fewer than 200 amino acids. This is therefore a possible result of a recombination duplication event and may not be a functional protein. The final tail assembly protein, close to the end of the JA10 genome, has no functional homologue in T7.

2.4. Novel Jumbo "Hairy *Myoviridae*" Phages

Sequencing of the other seven genomes showed that, although the phages were isolated independently, several shared 100%

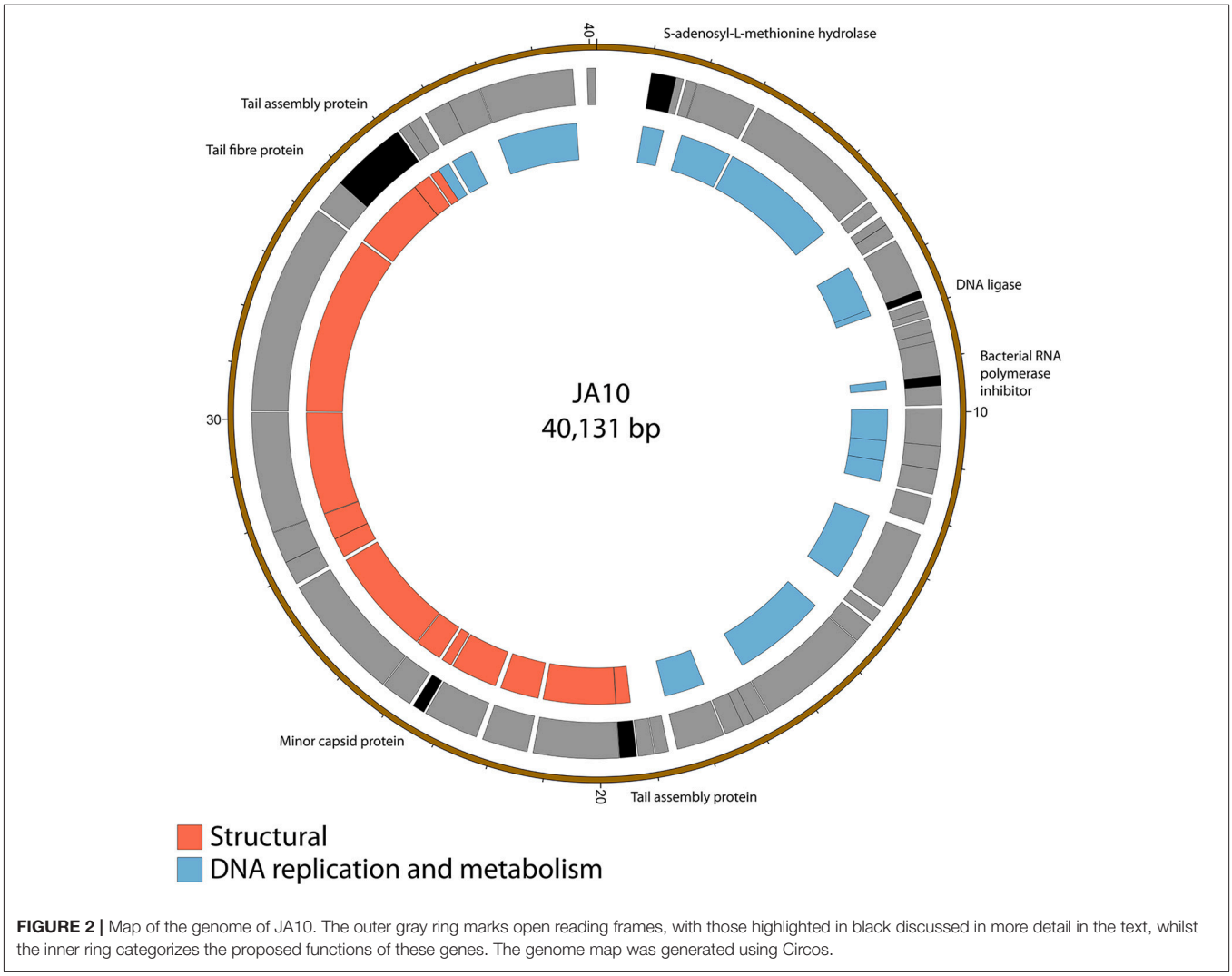


FIGURE 2 | Map of the genome of JA10. The outer gray ring marks open reading frames, with those highlighted in black discussed in more detail in the text, whilst the inner ring categorizes the proposed functions of these genes. The genome map was generated using Circos.

identity at the nucleotide level. JA11, 31, and 32 grouped together, as did JA33 and 37. JA31, 32, and 37 were therefore excluded from further analysis. The genome size for the four remaining phages is between 253 and 256 kbp. A summary of the characteristics of these genomes, as well as JA10, is shown in **Table 3**. These phages are therefore jumbo phages, defined as phages with a genome over 200 kbp (Yuan and Gao, 2017). The genomes are significantly larger than the conserved size of the *Ackermannviridae* family genomes, which are around 150 kbp, and larger than most sequenced phages. As of July 2018, there were 9,351 recorded phage genome sequences in Genbank (http://millardlab.org/bioinformatics/bacteriophage-genomes/) and these JA phages would be the 62nd–65th largest sequenced. Many of the over 300 predicted open reading frames in each genome do not match any known genes; the majority of those that do share identity with known genes are from the *E. amylovora* phages Yoloswag (Esplin et al., 2017) and Y3 (Buttimer et al., in press). These are largely structural genes and genes involved in DNA metabolism and replication.

TABLE 3 | Summary of the genomes of the broader host range phages.

Phage	Genome size (bp)	GC content (%)	Open reading frames
JA10	40,131	51.5	50
JA11	255,356	44.5	321
JA13	254,061	44.5	323
JA29	253,323	43.8	318
JA33	255,356	44.5	321

2.5. Variation Within the JA Phages

The gene order of the four JA jumbo phages is largely conserved. Over three quarters of the predicted ORFs are annotated as encoding hypothetical proteins, and many of the differences between the phages are contained within these ORFs as shown in **Figure 3**. JA29 is the most dissimilar to the others, sharing 86% nucleotide identity with JA11, whereas the nucleotide identity between JA11 and JA13 is 95%. JA11 and JA33 are 99% identical, with the major difference being the insertion of 126 bp in both

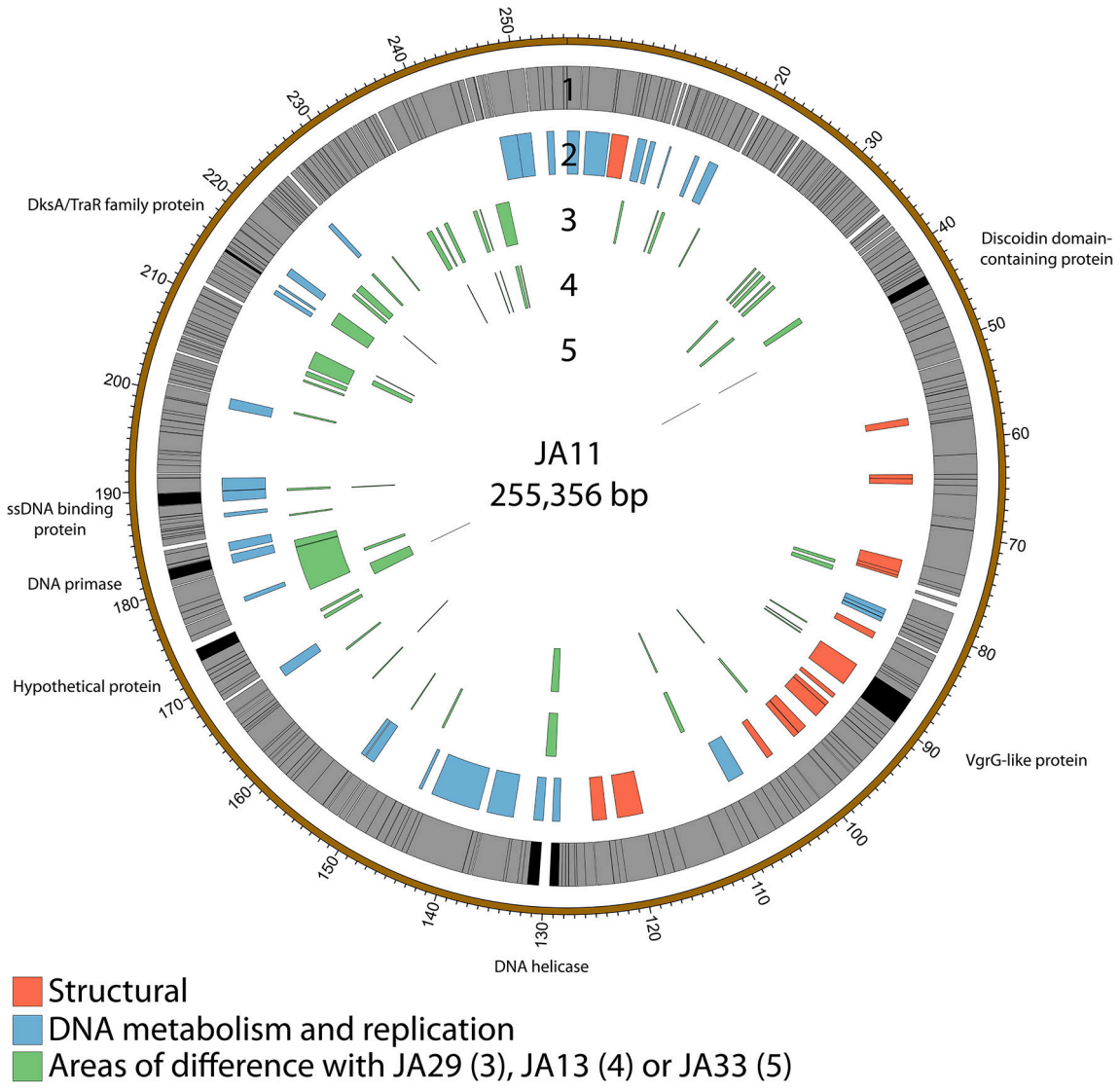


FIGURE 3 | Map of the genome of JA11. The outer gray ring marks open reading frames, with those highlighted in black discussed in more detail in the text. The second ring categorizes the proposed functions of these genes, whilst the inner rings highlight the areas of the genome that differ from the genomes of JA29, JA13, and JA33 (third to fifth ring, respectively). The genome map was generated using Circos.

genomes at different positions, and of different sequences. These insertions are in non-coding regions however, therefore their biological relevance is unclear. The only other differences are in two putative proteins. One is a hypothetical protein, whilst the other contains a putative discoidin domain, with the substitution between JA11 and JA33 (alanine to threonine) in the middle of the domain. Discoidin domains are present in eukaryotic agglutination factors and therefore the possible biological role for this in a phage genome, and the effect of the substitution, is not immediately obvious.

JA11, 13, and 29 have differing host ranges, as listed in **Table 2**. To investigate whether this was caused by variations within the tail fibers of the phages, the amino acid sequences of each of the three putative tail fiber proteins was compared between the phages. JA11 and JA13 possess identical tail fibers, whereas JA29 shows variations of several amino acids in each protein, as listed in **Table 4**. Whilst these variations could explain the differing host range of JA11 and JA29, it does not explain the difference in host range observed for JA11 and JA13.

Whilst most of the differences between the phages are located in genes with no predicted function, there are a few annotated that are present in all of the JA phages. These include encoding a DNA helicase, two potential transcription factors and one structural protein, all highlighted in **Figure 3** and listed in **Table 4**.

There are variations in two DNA related genes: a DNA primase and a DNA helicase. The helicase shows the most variation between the phages, as it appears to have undergone

TABLE 4 | Summary of the annotated genes which differ between JA11, JA13, and JA29.

Function	Gene	Length	Differences with JA11
Tail fiber one	JA11_90	272	
	JA13_090		0
	JA29_093		6
Tail fiber two	JA11_94	164	
	JA13_095		0
	JA29_096		7
Tail fiber three	JA11_95	210	
	JA13_096		0
	JA29_098		1
DNA primase	JA11_208	350	
	JA13_208		0
	JA29_210		7
DNA helicase	JA11_155 + 156	*	
	JA13_156		*
	JA29_158		*
DksA/TraR family protein	JA11_264	85	
	JA13_267		1
	JA29_265		5
ssDNA binding protein	JA11_221	402	
	JA13_222		1
	JA29_223		32
VgrG-like protein	JA11_117	931	
	JA13_118		1
	JA29_120		87

*See ***Figure 4***.

insertion or deletion between some of the phages. A comparison of this region of the genome can be seen in **Figure 4**. There are two ORFs annotated as putative helicases in JA11 and JA33, which both share homology with one ORF in JA13 and JA29. Whether the two ORFs are able to function independently as a helicase, or whether this duplication has rendered them non-functional, is unknown. A DksA/TraR family protein and a ssDNA binding protein, both likely transcription factors, both differ in one amino acid between JA11 and JA13, and share lower identity with the JA29 homolog, particularly the ssDNA binding protein, which differs in 32 positions. A VgrG-like family protein, a component of the T6SS thought to be phage-derived as it is capable of assembling into a structure similar to a phage tail spike (Cianfanelli et al., 2016) shares, at best, only 33% amino acid identity with the closest hit in the *E. amylovora* phage Y3, and so these may define a relatively novel VgrG-like protein group. JA11 and JA13 differ by one conservative substitution in this protein, whilst JA29 differs in 87 amino acids, 13 of which are conservative substitutions. It is possible that differences in this predicted protein are a contributing factor to the differing host range of these phages.

The biological significance of the differences observed between the four JA phages is currently unclear. It is somewhat surprising to find variations in genes involved in transcription initiation and DNA replication, as the genomes of these phages are relatively similar in both size and GC content, as summarized in **Table 3**. It is therefore possible that these differences do not significantly alter the function of these proteins. The variation in the VgrG-like proteins is more logical, as the different host ranges of these phages may be related to differences in tail spikes and other host recognition factors. To determine the impact of these differences, and to investigate why JA11 and JA13 have a different host range despite having apparently identical tail fibers, requires further experimental work.

2.6. More Recent Isolates: AD Phages

All of the JA phages were isolated from the River Cam in November 2014. Isolation of XF phages from the same location a year earlier produced mainly members of the *Ackermannviridae* family and a few *Podoviridae* family members (Day et al., 2017). Whilst there is clearly some maintenance of viral populations, as members of the two families have been isolated on both occasions, the jumbo phages presented here are a novel grouping. To gain further insight into the viral populations in the River Cam, further samples were taken in October 2017. Two phages were isolated on *D. solani* and are named AD1 and AD2. Whilst both were only capable of forming plaques on *D. solani* and not on strains of other species, microscopy showed that they had differing morphologies. AD1 (**Figure 1E**) appears to have a "hairy *Myoviridae*" morphology similar to that of the JA jumbo phages, with a head diameter of 120 nm, tail length of 150 nm, and unclear structures at the base of the tail. AD2 on the other hand (**Figure 1F**) has a head diameter of 90 nm and a (potentially partially-contracted) 70 nm tail, leading to a tentative classification as a member of the *Ackermannviridae* family. The structures at the end of the tail are inconclusive.

Genome sequencing of the two AD phages showed that, as suggested by microscopy, AD2 is a member of the *Ackermannviridae* family. It shares 99% nucleotide identity with previously published *D. solani Ackermannviridae* such as XF4 and LIMEstone1 (Day et al., 2017), although full coverage of the genome was not achieved (data not shown). AD1, as expected, has a large genome of 261,658 bp, confirming that it is a jumbo phage, shown in **Figure 5**. However, unexpectedly, it has low nucleotide sequence identity with the JA jumbo phages, despite sharing a similar gene order. In fact, a translated nucleotide comparison of JA11 and AD1, as shown in **Figure 6C**, shows that JA11 is about as similar to AD1 as it is to Y3 (**Figure 6A**), and a comparison of AD1 and Y3 (**Figure 6B**) shows them to be more similar to each other than to JA11 at the amino acid level.

2.7. Phylogeny of the "Hairy *Myoviridae*" Phages

In their recent publication, Buttimer et al. discussed the phylogenetic position of Y3 considering its low level of nucleotide identity to existing genomes (Buttimer et al., in press). Two potential subfamilies within the "hairy *Myoviridae*" have emerged; the Rak2-like phages, which includes the previously mentioned *Pectobacterium* phage CBB (Buttimer et al., 2017),

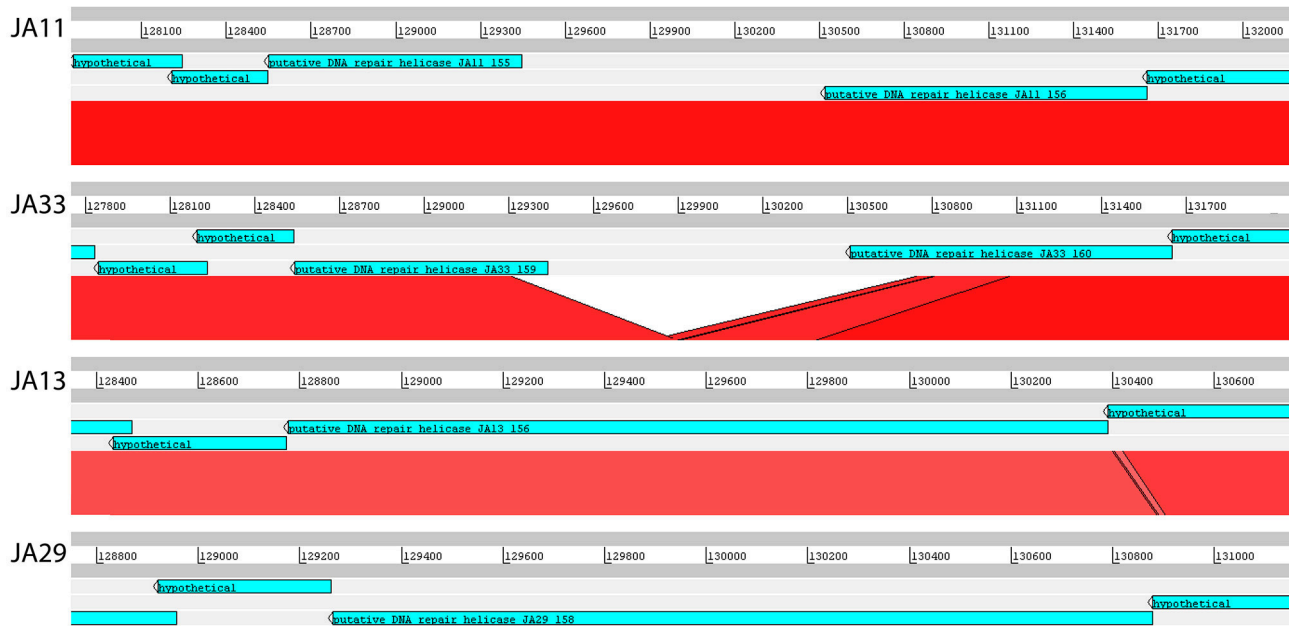


FIGURE 4 | Translated nucleotide comparison of a putative DNA helicase between JA11 (top), JA33 (second), JA13 (third), and JA29 (bottom). Red bars mark areas of amino acid identity, with darker colors showing higher identity. Figure produced using the Artemis Comparison Tool.

and the as yet unnamed subfamily that encompasses the phage discussed here. This was established as it was found that Y3 had homologs including terminase, polymerase and helicase genes in several other phages reported or suspected to have the "hairy *Myoviridae*" morphology. A comparison of the large terminase subunit of these phages with those reported here shows clear clustering, and can be seen in **Figure 7**. As expected, the three JA phages cluster tightly with little variation between them. As reported by Buttimer et al. the *Pseudomonas*-infecting phages PaBG (Sykilinda et al., 2014) and Lu11 (Adriaenssens et al., 2012b) form a clade, whilst the *Ralstonia solanacearum* phage phiRSL1 (Yamada et al., 2010) and the metagenomically-derived NCTB (Pfreundt et al., 2017) are single nodes within the tree. As suggested by the translated nucleotide comparison in **Figure 6B**, Y3 and AD1 form a clade that puts AD1 closer to *Erwinia*-infecting phages than to the other *D. solani* phages. Intriguingly, AD1 is placed closer phylogenetically to Y3 than the other *E. amylovora*-infecting phage Yoloswag (Esplin et al., 2017). The same clustering is seen when using the sequence of the tail sheath proteins of the phages. All of the phages except phiRSL1 have two annotated tail sheath proteins, and the same phylogeny is seen with both (data not shown).

The gene order between JA11, AD1, and Y3 is highly conserved. All three genomes contain over 300 open reading frames, with each containing only between one and three unique annotated genes. These unique genes are listed in **Table 5** and are all DNA or metabolism-related. There are also five genes common to AD1 and Y3 that are not found in JA11. Whilst phylogenetic clustering, as shown in **Figure 7**, groups AD1 and Y3 closer than Y3 and Yoloswag, it is interesting to note that the two unique genes possessed by Y3 have homologs in Yoloswag. These two phages were both isolated from apple orchards using *E. amylovora*, therefore it is surprising that they differ phylogenetically. It is possible that these unique genes are a determinant of the host range of these phages. A phylogenetic comparison of three tail fiber genes found in each genome is shown in **Figure 8**. This again shows a definite separation between Yoloswag and the other two phages, particularly when comparing Yoloswag_102 with Y3_104 and AD1_102, which occupy the same genomic position. This also suggests the possibility that AD1 may be capable of forming plaques on *Erwinia* species, but we have not been able to test this as we do not have access to these hosts.

2.8. Jumbo Phages Are Capable of Transduction

In a recent publication (Day et al., 2017) we tested the ability of the JA phages to facilitate transduction of chromosomal markers and plasmids between *Dickeya* cells. We can reconfirm that all of the JA jumbos are capable of transduction of a chromosomal marker between *D. solani* cells at a frequency of between 1×10^{-6} and 3×10^{-4} and report that the AD phages are capable of transduction of chromosomal markers at similar frequencies. The broader host range of the JA jumbo phages also allows transduction of plasmids between *Dickeya* species, as shown for a representative of each host range group in **Table 6**. JA10, the member of the *Podoviridae* family, proved incapable of transduction under the conditions tested. This makes JA10 a more promising candidate for phage therapy, and suggests the other phages may not be suitable due to the potential risk of transduction in the field.

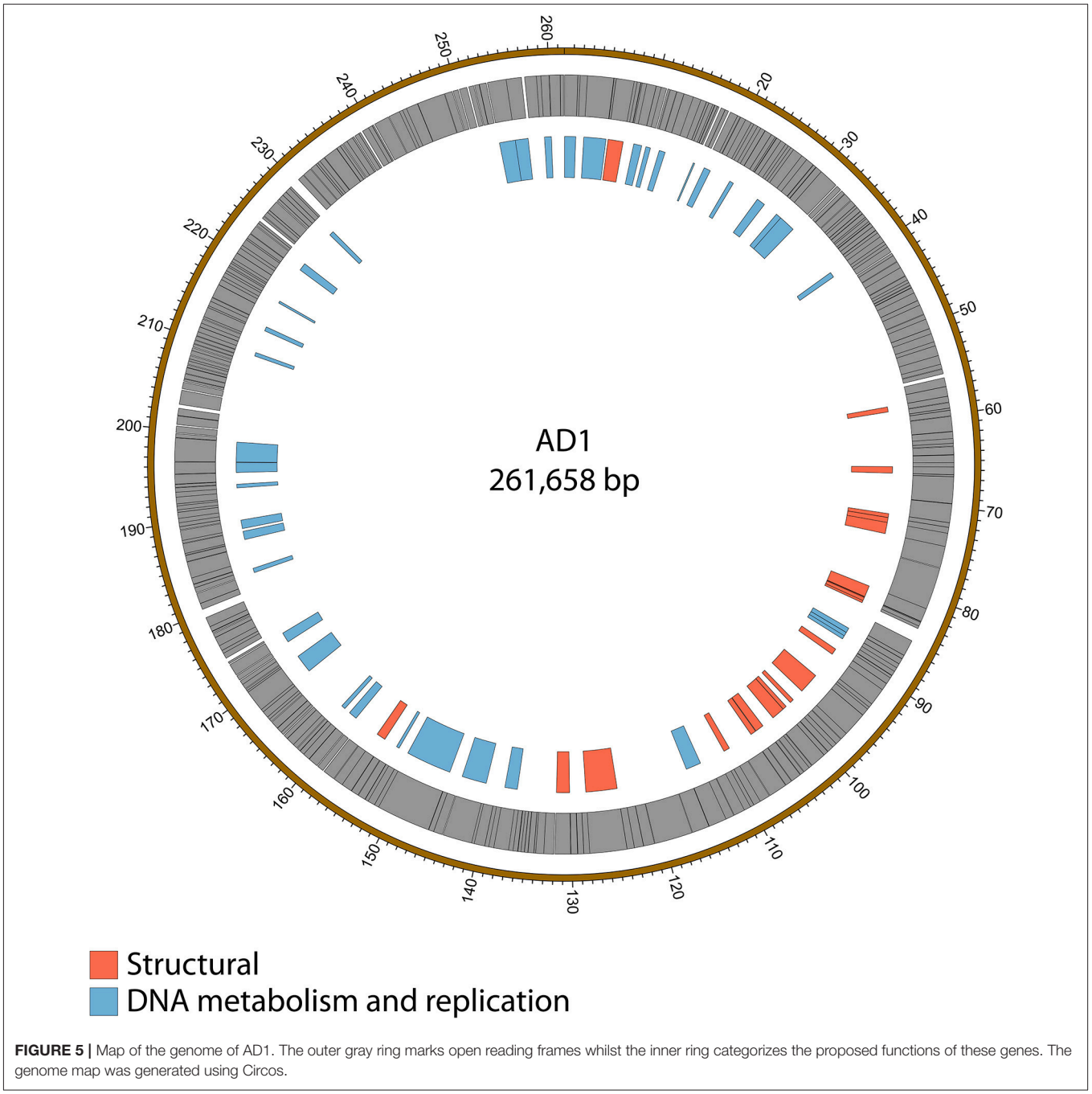


FIGURE 5 | Map of the genome of AD1. The outer gray ring marks open reading frames whilst the inner ring categorizes the proposed functions of these genes. The genome map was generated using Circos.

3. DISCUSSION

All seed crops imported into Scotland, along with 10% of Scottish-origin crops, are tested for *D. solani* each year. In 2017, the most recent year for which there are data, 663 samples were tested and none were positive for *Dickeya* species (Scottish Government, 2017), which has been the finding since 2010 when rigorous testing began. Whilst there have been isolated cases of *D. solani* infection reported in England and Wales since 2007 (Cahill et al., 2010), these have all originated in crops from outside of the UK (Toth et al., 2016), therefore it is not thought that *Dickeya* is established in the UK. This begs the question as to why we have been able to isolate phages capable of infecting *D. solani* with relative ease from the River Cam. We would hypothesize that either *D. solani* is present in the environment, but has not been confirmed by testing, or there is another permissive, but currently unknown, environmental host(s) for these phage. A novel species of *Dickeya*, *Dickeya aquatica*, was isolated from waterways in England (Parkinson et al., 2014), and so far has only been identified in waterway environments. It is, therefore, a formal possibility that this species could be an environmental host

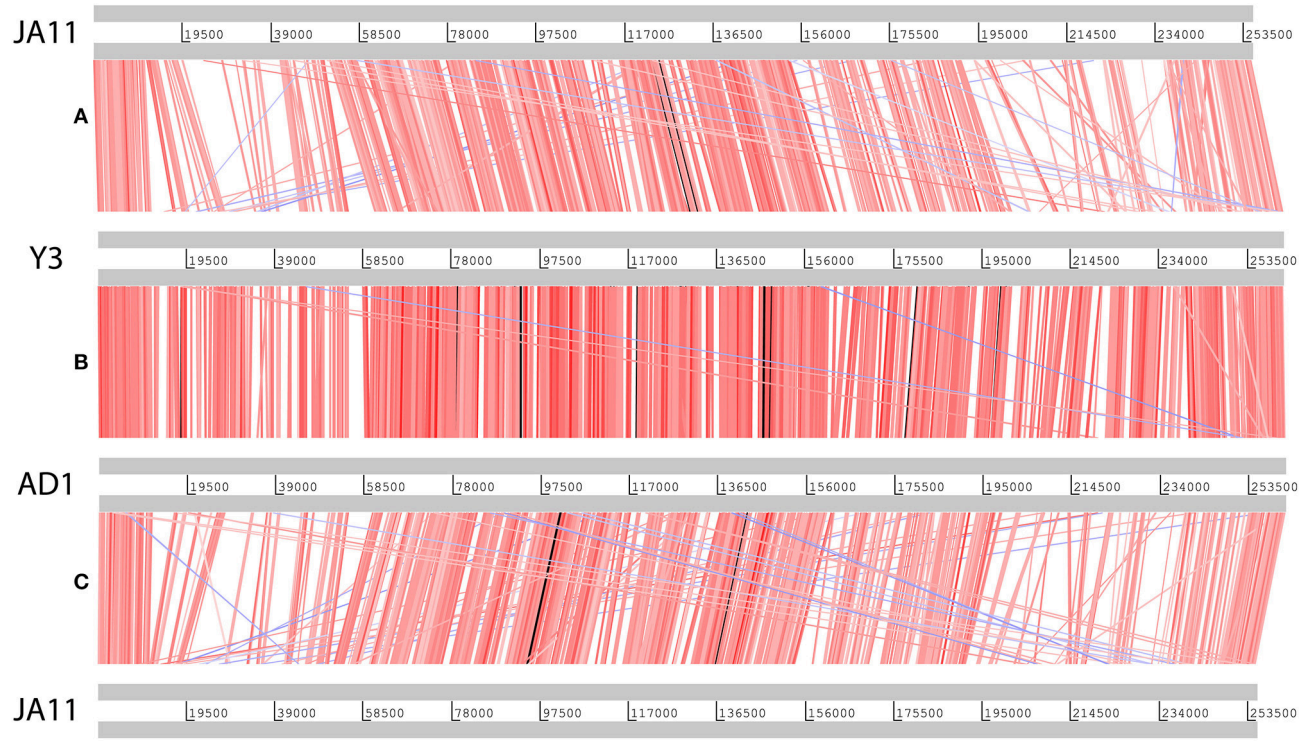


FIGURE 6 | Translated nucleotide comparison of the genomes of JA11 **(A)**, Y3 **(B)**, and AD1 **(C)**. Red bars mark areas of amino acid identity, with darker colors showing higher identity. Blue bars highlight areas of inversion. Figure produced using the Artemis Comparison Tool.

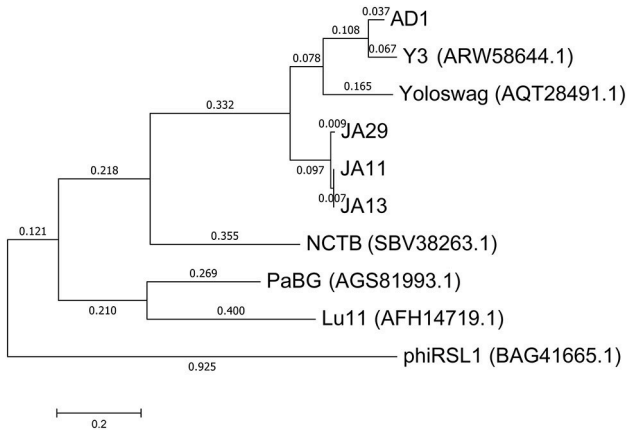


FIGURE 7 | Phylogenetic tree of the large terminase subunit constructed using the Maximum Likelihood method in MEGA. All positions containing gaps and missing data were eliminated, with 642 positions in the final dataset. The tree shown has the highest log likelihood (−8677.52) and is drawn to scale, with branch lengths measured in the number of substitutions per site.

for the phages isolated here, but this has not yet been tested.

Previously isolated phages of *D. solani* have been found almost exclusively to be members of the *Ackermannviridae* family. This has been a consistent feature of isolations spanning multiple European countries across the last decade, including from both soil and water samples, as shown in **Table 1**. We questioned in our recent publication whether this indicated a special relationship between *Ackermannviridae* and *D. solani* (Day et al., 2017). The phages presented here confirm that that was the result of an extrapolation from a limited viral sample set. There are at least four groups of *D. solani* phages present in waterways around Cambridge, spanning three of the four *Caudovirales* families. Phages that have been isolated on other species of *Dickeya* that are capable of forming plaques on *D. solani* have also been recently described (Alič et al., 2017b), including the fourth *Caudovirales* family, *Siphoviridae*. It is therefore apparent that we are only superficially defining the level of phage diversity present in the environment, consistent with the notion that double-stranded DNA phages alone have been predicted to outnumber their bacterial hosts by a factor of 10 to 1 (Chibani-Chennoufi et al., 2004).

All the phages presented here, apart from JA10 and AD2, appear to be representatives of a recently described "hairy *Myoviridae*" subfamily (Buttimer et al., in press). To the best of our knowledge, these are the first reported members of this subfamily isolated using *Dickeya* species. Many of the previously reported members of this subfamily were also isolated on plant-associated bacteria such as *Pseudomonas putida* (Adriaenssens et al., 2012b) and *E. amylovora* (Esplin et al., 2017; Buttimer et al., in press). Whether there is a link between this group of phages and plant-associated bacteria, or whether the recent

TABLE 5 | Unique annotated genes found in the genomes of JA11, AD, and Y3, as well as genes common to AD1 and Y3 but not present in JA11.

Genome	Gene	Gene annotation
JA11	JA11_30	DNA adenine methylase
AD1	AD1_017	DUF1611-domain containing protein
	AD1_258	XRE family transcriptional regulator
Y3	Y3_020	Oxygenase
	Y3_031	AntA/B antirepressor domain-containing protein
Common to AD1 and Y3	AD1_047	Transcriptional repressor
	Y3_049	
	AD1_048	DNA-cytosine methyltransferase
	Y3_050	
	AD1_018	Asparagine synthase
	Y3_018	
	AD1_267	Radical SAM superfamily protein
	Y3_272	
	AD1_016	Methyltransferase
	Y3_017	

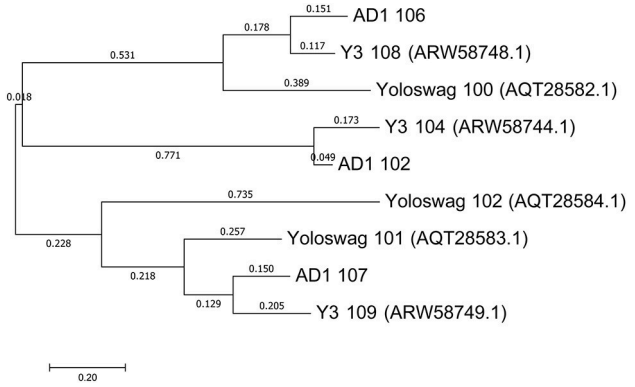


FIGURE 8 | Phylogenetic tree of three tail fiber proteins constructed using the Maximum Likelihood method in MEGA. All positions containing gaps and missing data were eliminated, with 158 positions in the final dataset. The tree shown has the highest log likelihood (−2426.84) and is drawn to scale, with branch lengths measured in the number of substitutions per site.

TABLE 6 | Transduction frequencies of the plasmid pECA1039-Km3 from donor *Dickeya solani* cells into other *Dickeya* species.

Phage	Recipient host	Transduction frequency
JA13	*Dickeya solani*	3.18×10^{-4}
	Dickeya dadantii subsp. *dieffenbachiae*	5.04×10^{-9}
	Dickeya paradisiaca	5.68×10^{-5}
JA29	*Dickeya solani*	1.88×10^{-4}
	Dickeya dadantii subsp. *dieffenbachiae*	6.29×10^{-9}
JA37	*Dickeya solani*	2.31×10^{-4}
	Dickeya dadantii subsp. *dieffenbachiae*	8.43×10^{-9}
	Dickeya paradisiaca	7.74×10^{-5}
	Dickeya zeae	4.27×10^{-6}

Phages are representatives of each host range group of the JA jumbo phages as detailed in ***Table 2****.*

increase in isolation of phages using phytopathogens is skewing this view remains to be determined. The proteins responsible for the "hairs" that typify this grouping remain unknown, although the identified tail fiber proteins are likely candidates for further investigation.

Infection of seed crops with *D. solani* and related species inflicts a high economic burden, and therefore there is great interest in the use of virulent (lytic) bacteriophages as potential biocontrol agents. There have been multiple tests of the stability, environmental viability and efficacy of *Dickeya* phages (Adriaenssens et al., 2012c; Alič et al., 2017b; Czajkowski et al., 2017) in which promising results have been reported. The *Dickeya* phages able to form individual plaques through productive lytic cycle replication on multiple *Dickeya* species, reported both here and by Alič et al. (2017b), are potentially more promising for use as biocontrol agents, as they would be able to act against a wider set of phytopathogens. However, we have described here, and previously (Day et al., 2017), that the majority of the *D. solani* phages (including our phages and the LIMEstone phages) are able to facilitate generalized transduction between host cells. The phages isolated by Czajkowski et al. (2014a,b, 2015a) were not reported to have been tested for generalized transduction, but, due to their classification as *Ackermannviridae* family members, and the finding that all members of this family tested to date are capable of facilitating transduction, we predict that they are likely to be capable of doing so. Alič et al. did not report testing of their phages for transduction capacity, and it is possible that they may not be transducers, but we would echo the caution of the European Medicines Agency, among others, who have stated that it is "important to ensure that therapeutic phages do not carry out generalized transduction" (Pelfrene et al., 2016). However, the results presented here do suggest that JA10, a podovirus capable of infecting three *Dickeya* species other than *D. solani*, could offer some promise as a potentially therapeutic candidate, as it has not shown transduction capabilities when tested.

Czajkowski et al. (2014a) have reported that phages D3 and D5 are capable of infecting multiple species of *Dickeya*. This conclusion was based on simple spot test assays in which undiluted spots of phage lysate were tested on bacterial top lawns and incubated, with any resultant clearing taken to show infection. It is known that applications of high titre lysates of phage to bacterial cells can cause the phenomenon of "lysis-from-without," in which cells lyse due to membrane disruption instead of productive phage infection (Khan Mirzaei and Nilsson, 2015). Consequently, confirmation of host range requires serial dilution of the phage lysate to visualize individual plaques on a host, and this confirmatory data would be helpful when assessing the reported host range of these phages. This reasonable caution is reinforced, particularly when we consider the genome identity of nearly 100 % with other *Ackermannviridae* family members.

The phages PD10.3 and 23.1 (Czajkowski et al., 2015a) are also reported to infect both *Dickeya* and *Pectobacterium* species, although host range was determined by the same method as D3 and D5. However, adsorption and burst size data for both

phages are reported on the two genera. The genomes of both phages have been sequenced and are reported as incomplete. Curiously, the largest scaffold of both is similar to the size of other *Ackermannviridae* family members (shown in **Table 1**) and these scaffolds share 99% identity with the full genome of LIMEstone1. The morphology of these two phages also clearly places them within the *Ackermannviridae* family. It is therefore intriguing that phages that are so similar have such different host ranges, and so further confirmatory data on the broader host range of these two phages could be biologically illuminating. Notwithstanding such observations, interpretation of *Dickeya* phage host range data should be treated with caution. We offer a salutary lesson, based on our own data, reported here, which more rigorously reinterprets previously reported host range data. Tests of the host range of our phages following the method of Czajkowski et al. have also suggested a much broader host range than we now know to be true (data not shown). Consequently, we would caution against assigning host range to phages without rigorous experimental data involving plaque formation in line with the comments of others (Khan Mirzaei and Nilsson, 2015).

TABLE 7 | Bacterial strains and bacteriophage genomes used in this study.

Bacterial strain	References
Dickeya chrysanthemi NCPBB 402	Pritchard et al., 2013b
Dickeya dadantii subsp. *dieffenbachiae* NCPBB 2976	Pritchard et al., 2013b
Dickeya dianthicola NCPBB 453	Pritchard et al., 2013a
Dickeya paradisiaca NCPBB 2511	Pritchard et al., 2013b
Dickeya solani MK10	Pritchard et al., 2013a
Dickeya zeae NCPBB 3532	Pritchard et al., 2013b
Dickeya solani MK10 pECA1039-Km3	This study
Bacteriophage genome	**Genbank ID and references**
BF25/12	KT240186.1 (Alič et al., 2017b)
LIMEstone1	HE600015.1 (Adriaenssens et al., 2012c)
Lu11	JQ768459.1 (Adriaenssens et al., 2012b)
NCTB	LT598654.1 (Pfreundt et al., 2017)
PaBG	KF147891.1 (Sykilinda et al., 2014)
phiRSL1	AB366653 (Yamada et al., 2010)
PP74	KY084243.1 (Kabanova et al., 2018)
XF4	KY942057.1 (Day et al., 2017)
Y3	KY984068.1 (Buttimer et al., in press)
Yoloswag	KY448244.1 (Esplin et al., 2017)

4. MATERIALS AND METHODS

4.1. Bacterial Strains, Phages, Culture Media, and Growth Conditions

All bacterial strains used in this study are listed in **Table 7**. *Dickeya* species were routinely grown at 30°C in Luria broth (LB) or on LB agar plates (1.5%, wt/vol, agar). Phages were stored at 4°C in phage buffer (10 mM Tris-HCl, pH 7.4, 10 mM $MgSO_4$, 0.01%, wt/vol, gelatine) over a few drops of $NaHCO_3$ saturated chloroform.

4.2. Isolation of Phages

Treated sewage effluent was collected from a sewage treatment plant in Cambridge, United Kingdom (Matilla and Salmond, 2014). River water was collected from multiple sites along the River Cam. Samples were filter sterilized using a 0.22 μm filter before 5 mL of the sample was added to 2x LB along with 500 μL of an overnight culture of *D. solani* MK10. This mixture was incubated overnight in a 250 mL flask at 30°C with shaking at 250 rpm. One milliliter of the enriched sample was mixed with 100 μL of chloroform (saturated with $NaHCO_3$) and vortexed to lyse bacterial cells. After sedimentation at 16,000 × g for 4 min, 10 μL of a serial dilution series of the supernatant was mixed with 200 μL of an overnight bacterial culture and 4 mL of LB top agar. This mixture was poured as an overlay on an LBA plate and incubated overnight at 30°C. Single phage plaques were picked with a sterile toothpick, placed into 100 μL phage buffer, and shaken with 40μL of chloroform to kill any bacteria. The phages obtained were plaque purified three times. High-titer phage lysates were then obtained as described previously (Petty et al., 2006). Briefly, 10-fold serial dilutions of the phage were incubated overnight in an agar overlay. Those plates exhibiting confluent lysis (seen as a mosaic-like effect in which the plaques are close to merging) were used for lysate preparation. The top agar was removed from the plate, vortexed with chloroform before sedimentation at 2,200 × g for 20 min at 4°C. The supernatant was removed and vortexed with a few drops of chloroform to produce the final lysate.

4.3. Determination of Host Range

The host range of isolated phages was determined by plating out ten-fold serial dilutions of the phage lysates onto agar overlays containing host *Dickeya* cells and incubating overnight at 30°C. Following best practice to avoid potential confusion with "lysis from without," only phages that produced individual plaques following serial dilution on three independent occasions were considered as being able to infect the respective host productively through a lytic cycle.

4.4. Transmission Electron Microscopy

High-titre lysates for transmission electron microscopy were obtained as described above using 0.35% (w/v) LB agarose instead of 0.35% (w/v) LB agar overlays. Ten μL of high-titre phage lysates were adsorbed onto 400-mesh copper grids with holey carbon support films (Agar Scientific, Stansted, United Kingdom) for 2 min. The copper grids were discharged in a Quorum/Emitech K100X system (Quorum, Ringmer, United Kingdom) prior to use. Excess phage suspension was removed with filter paper and phage samples were negatively stained by placing the grids for 30 s in ten μL of 2% uranyl acetate neutralized with NaOH. The grids were then blotted on filter paper to remove the excess solution and allowed to air dry. Phages were examined by transmission electron microscopy at Cambridge Advanced Imaging Centre (Department of Physiology, Development and Neuroscience, University of Cambridge) using an FEI Tecnai G2 transmission

electron microscope (FEI, OR, USA). The accelerating voltage was 120.0 kV, and images were captured with an AMT XR60B digital camera running Deben software.

4.5. Genome Sequencing and Analysis

Phage genomes were sequenced on the Illumina MiSeq Sequencer at MicrobesNG (Birmingham, UK). The reads were trimmed using Trimmomatic (Bolger et al., 2014), assessed for quality using BWA-MEM (Li, 2013) and assembled using SPAdes 3.7.1 (Bankevich et al., 2012) with standard settings. Except for JA10 and AD2, all generated over 140,000 reads and had higher than 100x coverage of the full genome. JA10 generated 3,270 reads and had 26x coverage. AD2 generated 1,899 reads and had 4.79x coverage. All assembled into one contig except AD2. Annotation was performed using DNAMaster 5.23.2 (Lawrence, 2012). Genome maps were generated using Circos 0.69.6 (Krzywinski et al., 2009). Genomes were deposited in Genbank using BankIt (NCBI) and are available under accession numbers MH460459 (JA10), MH389777 (JA11), MH460460 (JA13), MH460461 (JA29), MH460462 (JA33), and MH460463 (AD1). Genomes were compared using NCBI Blast, MEGA 7.0.26 (Kumar et al., 2016) and the Artemis Comparison Tool 13.0.0 (Carver et al., 2005). Annotation tables are found in **Tables S1**–**S6**.

4.6. Transduction

To test for transduction, phage lysates were generated as described above on donor bacterial strains carrying the desired plasmid or chromosomal marker. All the experiments used kanamycin as the antibiotic for selection. The chromosomal marker for the JA phages was a transposon stably inserted into a protease gene. Successful transduction was therefore confirmed by a protease-negative, kanamycin-resistant phenotype in the recipient cells. The chromosomal marker for the AD phages was a transposon stably inserted into the *lacZ* gene. Successful transduction was confirmed by kanamycin-resistant recipient colonies that were white on media containing X-gal. The plasmid marker was pECA1039-Km3 and successful transduction was confirmed by plasmid extraction and gel electrophoresis under standard conditions as described previously (Day et al., 2017).

Transduction was performed by mixing phage lysate with an overnight culture of the recipient cells to achieve a multiplicity of infection of 0.01, meaning that for each phage there were one hundred bacterial cells, apart from the transduction of pECA1039-Km3 into *Dickeya dadantii* subsp. *dieffenbachiae*, which required an MOI of 0.1. The mixture was left on the lab bench at room temperature for 20 min, followed by incubation on a rotary wheel at 30°C for 30 min. The infected culture was then centrifuged and the bacterial pellets washed with LB twice to eliminate any remaining non-adsorbed phage. The bacterial pellets were resuspended in 1 mL LB and 100 μL aliquots were spread onto LBA plates with selection for the chromosomal or plasmid marker. Appropriate standard controls, which were routinely negative, were used to score for any spontaneous resistance of the recipient strain. One hundred microliters of the phage lysate was also spread onto LBA plates to confirm lysate sterility.

AUTHOR CONTRIBUTIONS

AD, JA, and GPCS conceived and designed the experiments, analyzed the data. AD and JA performed the experiments. AD wrote the paper. AD and GPCS edited the paper.

FUNDING

This work was supported by the BBSRC, UK. AD was supported by a Cambridge Doctoral Training Partnership Award from the BBSRC, UK.

ACKNOWLEDGMENTS

Sequencing was conducted by MicrobesNG (http://www.microbesng.uk), which is supported by the BBSRC (grant number BB/L024209/1). Microscopy was performed at Cambridge Advanced Imaging Centre with the help of Karin Müller and Lyn Carter. We are grateful to Sonia Humphris and Ian Toth (James Hutton Institute, Scotland) for generous provision of *Dickeya* strains. This work was done under DEFRA license: 50864/197900/3. We thank Alison Rawlinson, Sarah Barker, and Diana Breitmaier for technical support and Rita Monson for helpful advice.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2018.02169/full#supplementary-material

REFERENCES

Ackermann, H.-W. (2012). "Bacteriophage electron microscopy," in *Advances in Virus Research*, eds M. Lobocka and W. T. Szybalski (Elsevier), 1–32.

Ackermann, H. W., and Nguyen, T. M. (1983). Sewage coliphages studied by electron microscopy. *Appl. Environ. Microbiol.* 45, 1049–59.

Adeolu, M., Alnajar, S., Naushad, S., and Gupta, R. S. (2016). Genome-based phylogeny and taxonomy of the "Enterobacteriales:" proposal for *Enterobacterales* ord. nov. divided into the families *Enterobacteriaceae, Erwiniaceae* fam. nov., *Pectobacteriaceae* fam. nov., *Yersiniaceae* fam. nov., *Hafniaceae* fam. nov., Morgane. *Int. J. Syst. Evol. Microbiol.* 66, 5575–5599. doi: 10.1099/ijsem.0.001485

Adriaenssens, E. M., Ackermann, H. W., Anany, H., Blasdel, B., Connerton, I. F., Goulding, D., et al. (2012a). A suggested new bacteriophage genus: "Viunalikevirus." *Arch. Virol.* 157, 2035–2046. doi: 10.1007/s00705-012-1360-5

Adriaenssens, E. M., Mattheus, W., Cornelissen, A., Shaburova, O., Krylov, V. N., Kropinski, A. M., et al. (2012b). Complete genome sequence of the giant *Pseudomonas* phage Lu11. *J. Virol.* 86, 6369–6370. doi: 10.1128/JVI.00641-12

Adriaenssens, E. M., Van Vaerenbergh, J., Vandenheuvel, D., Dunon, V., Ceyssens, P. J., De Proft, M., et al. (2012c). T4-related bacteriophage LIMEstone isolates for the control of soft rot on potato caused by "*Dickeya solani*." *PLoS ONE* 7:et002. doi: 10.1371/journal.pone.0033227.t002

Adriaenssens, E. M., Wittmann, J., Kuhn, J. H., Turner, D., Sullivan, M. B., Dutilh, B. E., et al. (2018). Taxonomy of prokaryotic viruses: 2017 update

from the ICTV Bacterial and Archaeal Viruses Subcommittee. *Arch. Virol.* 163, 1125–1129. doi: 10.1007/s00705-018-3723-z

Alic, Š., Naglič, T., Tušek-Žnidarič, M., Peterka, M., Ravnikar, M., and Dreo, T. (2017a). Putative new species of the genus *Dickeya* as major soft rot pathogens in *Phalaenopsis* orchid production. *Plant Pathol.* 66, 1357–1368. doi: 10.1111/ppa.12677

Alič, Š., Naglič, T., Tušek-Žnidarič, M., Ravnikar, M., Rački, N., Peterka, M., etal. (2017b). 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

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Branston (2012). *A Natural Solution to Takle Potential Soft Rot*. Available online at: http://www.branston.com/news/a-natural-solution-to-takle-potential-soft-rot/

Buttimer, C., Hendrix, H., Oliveira, H., Casey, A., Neve, H., McAuliffe, O., et al. (2017). Things are getting hairy: enterobacteria bacteriophage vB_PcaM_CBB. *Front. Microbiol.* 8:44. doi: 10.3389/fmicb.2017.00044

Buttimer, C., Born, Y., Lucid, A., Loessner, M. J., Fieseler, L., and Coffey, A. (in press). *Erwinia amylovora* phage vB_EamM_Y3 represents another lineage of hairy *Myoviridae*. *Res. Microbiol*. doi: 10.1016/j.resmic.2018.04.006

Cahill, G., Fraser, K., Kowalewska, M. J., Kenyon, D. M., and Saddler, G. S. (2010). "Recent findings from the *Dickeya* survey and monitoring programme," in *Proceedings Crop Protection in Northern Britain* (Edinburgh), 171–176.

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

Chibani-Chennoufi, S., Bruttin, A., Dillmann, M.-L., and Brussow, H. (2004). Phage-host interaction: an ecological perspective. *J. Bacteriol.* 186, 3677–3686. doi: 10.1128/JB.186.12.3677-3686.2004

Cianfanelli, F. R., Alcoforado Diniz, J., Guo, M., De Cesare, V., Trost, M., and Coulthurst, S. J. (2016). VgrG and PAAR proteins define distinct versions of a functional type VI secretion system. *PLOS Pathogens* 12:e1005735. doi: 10.1371/journal.ppat.1005735

Czajkowski, R., Ozymko, Z., de Jager, V., Siwinska, J., Smolarska, A., Ossowicki, A., et al. (2015a). 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., and Lojkowska, E. (2014a). 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., Ozymko, Z., Siwinska, J., Ossowicki, A., de Jager, V., Narajczyk, M., et al. (2015b). The complete genome, structural proteome, comparative genomics and phylogenetic analysis of a broad host lytic bacteriophage phiD3 infecting pectinolytic *Dickeya* spp. *Stand. Genomic Sci.* 10:68. doi: 10.1186/s40793-015-0068-z

Czajkowski, R., Ozymko, Z., Zwirowski, S., and Lojkowska, E. (2014b). Complete genome sequence of a broad-host-range lytic *Dickeya* spp. bacteriophage ΦD5. *Arch. Virol.* 159, 3153–3155. doi: 10.1007/s00705-014-2170-8

Czajkowski, R., Smolarska, A., and Ozymko, Z. (2017). The viability of lytic bacteriophage ΦD5 in potato-associated environments and its effect on *Dickeya solani* in potato (*Solanum tuberosum* L.) plants. *PLoS ONE* 12:e0183200. doi: 10.1371/journal.pone.0183200

Day, A., Ahn, J., Fang, X., and Salmond, G. P. C. (2017). Environmental bacteriophages of the emerging enterobacterial Phytopathogen, *Dickeya solani*, show genomic conservation and capacity for horizontal gene transfer between their bacterial hosts. *Front. Microbiol.* 8:1654. doi: 10.3389/fmicb.2017.01654

Esplin, I. N. D., Berg, J. A., Sharma, R., Allen, R. C., Arens, D. K., Ashcroft, C. R., et al. (2017). Genome sequences of 19 novel *Erwinia amylovora* bacteriophages. *Genome Announc.* 5:e00931–17. doi: 10.1128/genomeA.00931-17

Fokine, A., and Rossmann, M. G. (2014). Molecular architecture of tailed double-stranded DNA phages. *Bacteriophage* 4:e28281. doi: 10.4161/bact.28281

Iriarte, F. B., Balogh, B., Momol, M. T., Smith, L. M., Wilson, M., and Jones, J. B. (2007). Factors affecting survival of bacteriophage on tomato leaf surfaces. *Appl. Environ. Microbiol.* 73, 1704–1711. doi: 10.1128/AEM.02118-06

Kabanova, A., Shneider, M., Bugaeva, E., Ha, V. T. N., Miroshnikov, K., Korzhenkov, A., et al. (2018). Genomic characteristics of vB_PpaP_PP74, a T7-like *Autographivirinae* bacteriophage infecting a potato pathogen of the newly proposed species *Pectobacterium parmentieri*. *Arch. Virol.* 163, 1691–1694. doi: 10.1007/s00705-018-3766-1

Khan Mirzaei, M., and Nilsson, A. S. (2015). Isolation of phages for phage therapy: a comparison of spot tests and efficiency of plating analyses for determination of host range and efficacy. *PLoS ONE* 10:e0118557. doi: 10.1371/journal.pone.0118557

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33:msw054. doi: 10.1093/molbev/msw054

Laurila, J., Ahola, V., Lehtinen, A., Joutsjoki, T., Hannukkala, A., Rahkonen, A., et al. (2008). Characterization of *Dickeya* strains isolated from potato and river water samples in Finland. *Eur. J. Plant Pathol.* 122, 213–225. doi: 10.1007/s10658-008-9274-5

Lawrence, J. (2012). DNA Master. Available online at: cobamide2.bio.pitt.edu

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv*.

Mansfield, J., Genin, S., Magori, S., Citovsky, V., Sriariyanum, M., Ronald, P., et al. (2012). Top 10 plant pathogenic bacteria in molecular plant pathology. *Mol. Plant Pathol.* 13, 614–29. doi: 10.1111/j.1364-3703.2012.00804.x

Matilla, M. A., Fang, X., and Salmond, G. P. C. (2014). Viunalikeviruses are environmentally common agents of horizontal gene transfer in pathogens and biocontrol bacteria. *ISME J.* 8, 2143–2147. doi: 10.1038/ismej.2014.150

Matilla, M. A., and Salmond, G. P. C. (2014). Bacteriophage phiMAM1, a viunalikevirus, is a broad-host-range, high-efficiency generalized transducer that infects environmental and clinical isolates of the enterobacterial genera *Serratia* and *Kluyvera*. *Appl. Environ. Microbiol.* 80, 6446–6457. doi: 10.1128/AEM.01546-14

Parkinson, N., DeVos, P., Pirhonen, M., and Elphinstone, J. (2014). *Dickeya aquatica* sp. nov., isolated from waterways. *Int. J. Syst. Evol. Microbiol.* 64(Pt 7), 2264–2266. doi: 10.1099/ijs.0.058693-0

Parkinson, N., Stead, D., Bew, J., Heeney, J., Tsror, L., and Elphinstone, J. (2009). *Dickeya* species relatedness and clade structure determined by comparison of recA sequences. *Int. J. Syst. Evol. Microbiol.* 59, 2388–2393. doi: 10.1099/ijs.0.009258-0

Pelfrene, E., Willebrand, E., Cavaleiro Sanches, A., Sebris, Z., and Cavaleri, M. (2016). Bacteriophage therapy: a regulatory perspective. *J. Antimicrob. Chemother.* 71, 2071–2074. doi: 10.1093/jac/dkw083

Petty, N. K., Foulds, I. J., Pradel, E., Ewbank, J. J., and Salmond, G. P. C. (2006). A generalized transducing phage (phiIF3) for the genomically sequenced *Serratia marcescens* strain Db11: a tool for functional genomics of an opportunistic human pathogen. *Microbiology* 152, 1701–1708. doi: 10.1099/mic.0.28712-0

Pfreundt, U., Spungin, D., Hou, S., Voß, B., Berman-Frank, I., and Hess, W. R. (2017). Genome of a giant bacteriophage from a decaying *Trichodesmium* bloom. *Mar. Genomics* 33, 21–25. doi: 10.1016/j.margen.2017.02.001

Pritchard, L., Humphris, S., Baeyen, S., Maes, M., Van Vaerenbergh, J., Elphinstone, J., et al. (2013a). Draft genome sequences of four *Dickeya dianthicola* and four *Dickeya solani* strains. *Genome Announc.* 1:e00087–12. doi: 10.1128/genomeA.00087-12

Pritchard, L., Humphris, S., Saddler, G. S., Elphinstone, J. G., Pirhonen, M., and Toth, I. K. (2013b). Draft genome sequences of 17 isolates of the plant pathogenic bacterium *Dickeya*. *Genome Announc.* 1:e00978–13. doi: 10.1128/genomeA.00978-13

Scottish Government (2017). *2017 Inspection of Growing Crops of Seed and Ware Potatoes: Survey for the Presence of Dickeya* spp. Available online at: http://www.gov.scot/Topics/farmingrural/Agriculture/plant/18273/PotatoHealthControls/PotatoQuarantineDiseases/Dickeya/DickeyaGrowingCropResults2017

Sławiak, M., van Beckhoven, J. R. C. M., Speksnijder, A. G. C. L., Czajkowski, R., Grabe, G., and van der Wolf, J. M. (2009). Biochemical and genetical analysis

reveal a new clade of biovar 3 *Dickeya* spp. strains isolated from potato in Europe. *Eur. J. Plant Pathol.* 125, 245–261. doi: 10.1007/s10658-009-9479-2

Svircev, A., Roach, D., and Castle, A. (2018). Framing the future with bacteriophages in agriculture. *Viruses* 10:218. doi: 10.3390/v10050218

Sykilinda, N. N., Bondar, A. A., Gorshkova, A. S., Kurochkina, L. P., Kulikov, E. E., Shneider, M. M., et al. (2014). Complete genome sequence of the novel giant *Pseudomonas* phage PaBG. *Genome Announc.* 2:e00929–13. doi: 10.1128/genomeA.00929-13

Toth, I. K., Cahill, G., Elphinstone, J. G., Humphris, S., and Wale, S. J. (2016). "An update on the Potato Council/Scottish government-funded blackleg project - Year 2," in *Proceedings Crop Protection in Northern Britain 2016* (Edinburgh), 203–204.

Toth, I. K., van der Wolf, J. M., Saddler, G., Lojkowska, E., Hélias, V., Pirhonen, M., et al. (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

Tsror, L., Erlich, O., Lebiush, S., Hazanovsky, M., Zig, U., Slawiak, M., et al. (2009). Assessment of recent outbreaks of *Dickeya* sp. (syn. *Erwinia chrysanthemi*) slow wilt in potato crops in Israel. *Eur. J. Plant Pathol.* 123, 311–320. doi: 10.1007/s10658-008-9368-0

van der Wolf, J. M., Nijhuis, E. H., Kowalewska, M. J., Saddler, G. S., Parkinson, N., Elphinstone, J. G., et al. (2014). *Dickeya solani* sp. nov., a pectinolytic plant-pathogenic bacterium isolated from potato (*Solanum tuberosum*). *Int. J. Syst. Evol. Microbiol.* 64(Pt 3), 768–774. doi: 10.1099/ijs.0.052944-0

Yamada, T., Satoh, S., Ishikawa, H., Fujiwara, A., Kawasaki, T., Fujie, M., et al. (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

Yuan, Y., and Gao, M. (2017). Jumbo bacteriophages: an overview. *Front. Microbiol.* 8:403. doi: 10.3389/fmicb.2017.00403

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.