Genomics of phage-bacterium-host interaction of mushroom pathogenic Pseudomonads

by

Nathaniel Storey

Submitted in partial fulfilment of the requirement for the degree of Doctor of Philosophy, January 2018.

Declaration

I declare that this is an account of my own research and has not been submitted for a degree at any other university. The use of material from other sources has been properly and fully acknowledged, where appropriate.

Nathaniel Storey

Acknowledgements

"**Tout ce qui est simple est faux, mais tout ce qui ne l'est pas est inutilisable."**
~ Paul Valéry (1871-1945)

With special thanks to my father Mark Storey for his support and encouragement, without which this project would not have been possible.

At the University of Reading I would especially like to thank Geraldine Mulley for her guidance and generosity in assisting me when it was most needed; with further thanks to Rob Jackson and Ben Neuman for both their support and company during my time at the University of Reading.

Furthermore, I would like to thank my fiancée Lyssa Reeve for her never-ending patience and support; and my friends Natalie Tarling and Hashem Ghazi for their assistance and company throughout all the darkest hours and to celebrate the elusive moments of success.

Finally, I would like to thank Saad Mutlk for all his help and humour in the laboratory.

Abstract

The most commonly cultivated mushroom in Europe and North America is the *Agaricus bisporus*, also known as the button mushroom or Portobello mushroom. Bacterial diseases of *Agaricus bisporus* caused by *Pseudomonas* species are a cause of significant crop loss and downgrading of produce, resulting in considerable economic cost. Bacteriophage have long been an attractive option for biocontrol of bacterial contamination of food products, however the precise genetic interactions between phage, bacterium and host are often inadequately explored.

This project aims to explore the genetic interactions between the mushroom pathogenic bacterium *Pseudomonas tolaasii* and *Pseudomonas agarici* with the newly identified bacteriophage *Pseudomonas* phage NV1 and *Pseudomonas* phage ϕNV3. Full genome sequencing has been performed on the *P. tolaasii* strain 2192T and *P. agarici* NCPPB 2472, and the genomes of both mined for potential biosynthetic clusters involved in virulence as well as genes involved in phage resistance. Within the genome of *P. tolaasii* 2192T, putative non-ribosomal peptide synthases have been identified which are hypothesised to be involved in the production of the tolaasin toxin involved in disease symptom appearance on mushroom surfaces. Within the genome of *P. agarici* NCPPB 2472, a biosynthetic cluster was identified that is hypothesised to produce the siderophore achromobactin, an important virulence factor. *P. agarici* NCPPB 2472 was identified as possessing a single Type I-F CRISPR/Cas system, predicted to be involved in the development of phage resistance, as well as complete operon predicted to be involved in the production of the exopolysaccharide alginate.

A third *Pseudomonas* species was identified on the surface of disease free mushrooms which was identified as a potentially new species of *Pseudomonas*, named *Pseudomonas* sp. NS1. The genome of *P.* sp. NS1 was likewise sequenced and mined for potential biosynthetic gene clusters which identified a cluster demonstrated to be involved in the production of White Line Inducing Principle.

The *Pseudomonas* phages NV1 and ϕNV3 were isolated from environmental samples and identified to be narrow host range phage specific for *P. tolaasii* 2192T and *P. agarici* NCPPB 2472 respectively. Both phage NV1 and ϕNV3 were identified as new species of the *Luz24likeviruses* and *phiKMVlikeviruses* respectively. The genomes of both phages were isolated and sequenced, with phage ϕNV3 identified as containing a conserved Signal-Arrest-Release endolysin system, which was confirmed by *in vitro* protein expression. Likewise, the lysis cassette proteins of ϕNV3 were identified and investigated via protein complementation assay *in vitro*. The full growth characteristics and life cycle of phage ϕNV3 has been investigated and reported in this study and a broad-host range mutant of ϕNV3 identified which has allowed the T7-like tail protein of ϕNV3 to be identified as the host specificity determinant.

Transcriptome analyses of non-infected *P. agarici* NCPPB 2472 and *P. agarici* NCPPB 2472 infected with phage ϕNV3 at a multiplicity of infection (MOI) of 1 at 40 min post infection, were performed in triplicate using RNA-seq. A reliable method has been established that will be useful in future studies, although comparative gene expression analysis revealed no significant differences in expression between the two treatments at the multiplicity of infection and time point chosen in this case a significant quantity of phage transcripts were detected, demonstrating active phage infection.

List of Abbreviations and Acronyms

Abs	Absorbance
Amp	Ampicillin
ANIb	Average Nucleotide Identity based on BLAST+
ANIm	Average Nucleotide Identity based on MUMmer
bp	Base pair
C	Celsius
Cas	CRISPR associated
CFU	Colony Forming Units
CRISPR	Clustered Randomly Interspersed Short Palindromic Repeats
crRNA	CRISPR RNA
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
ds	Double Stranded
EDTA	Ethylenediaminetetraacetic Acid
EPS	Extracellular Polysaccharide
ETEC	Enterotoxigenic *E. coli*
Gp	Genome position
GTP	Guanosine-5'-triphosphate
HHP	High Hydrostatic Pressure
Kan	Kanamycin
KB	Kings B medium
LB	Lysogeny Broth
LPS	Lipopolysaccharide
ml	Millilitres
MLSA	Multilocus sequence analysis
MOI	Multiplicity of infection
mRNA	Messenger ribonucleic acid
MUSCLE	Multiple Sequence Comparison by Log-Expectation
NRPS	Non-Ribosomal Peptide Synthase
OD	Optical Density
ORF	Open Reading Frame
Padj	P-value adjusted for multiple testing using the Benjamin-Hochberg method
PBS	Phosphate Buffered Saline
PCA	Principal Component Analysis
PCP	Peptidyl Carrier Protein
PCR	Polymerase Chain Reaction
PFU	Plaque Forming Units
PMBV	Polymyxin B nonapeptide
pv.	Pathovar
R-M	Restriction-Modification
RBP	Receptor Binding Protein
rDNA	Ribosomal Deoxyribonucleic Acid
rpm	Revolutions Per Minute
rRNA	Ribosomal Ribonucleic Acid
SAR	Signal-Arrest-Release
SD	Standard Deviation
SEM	Standard Error of the Mean

sp.	Species
ss	Single Stranded
TBE	Tris/Borate/EDTA
TE	Thioesterase
TEM	Transmission Electron Microscopy
U.K	United Kingdom
UoR	University of Reading
UPEC	Uropathogenic *E. coli*
UV	Ultraviolet
vsd	variance stabilising transformed

Contents

Chapter 1:
Introduction

1.1 Cultivated mushrooms and their importance

The most commonly cultivated mushroom in Europe and North America is *Agaricus bisporus* (Figure 1.1), which is known by many names depending on the colour and maturity of the mushroom but is most frequently known as the 'button mushroom' or 'Portobello mushroom'. Global production of *Agaricus bisporus* is reported to exceed $1.8x10^6$ tonnes per year, of which $8\text{-}9x10^5$ tonnes are cultivated in Europe alone, worth an approximate 2 billion Euros annually [1].

Figure 1.1: Image showing *Agaricus bisporus* mushrooms with unopened caps and a cross section of an immature cap.

Commercial mushroom production is a complex procedure that comprises of several distinct stages, these are: composting, spawning, casing, case run, pinning and harvesting. The material on which *A. bisporus* is cultivated varies but can include horse manure, wheat straw, plant wastes and animal wastes such as feather meal and chicken manure. This mixture is then composted, where it is mixed and wetted to allow aerobic fermentation to take place. When completed the mixture becomes pliable and capable of holding water.

Once the composting has been completed spawning can begin, this involves the spreading of spawning grains (grains previously inoculated with the *A. bisporus* mycelium) and nutrient supplements over the compost. This stage requires high levels of cleanliness; therefore, all equipment and tools must be thoroughly disinfected before use. The mycelia from the spawning

grains will then grow throughout the compost, forming a network of fine root-like threads that fuse together to form a continuous bed, this process normally takes between 14-21 days to complete [2].

1.2 Bacterial diseases of cultivated mushrooms

Bacterial diseases of *Agaricus bisporus* include cavity disease, also known as rapid soft rot disease, that forms large cavities in the cap of the mushroom that may extend to the stipe, and the causal bacterium was initially identified as *Pseudomonas cepacia* [3]. However, *Pseudomonas cepacia* was renamed *Pseudomonas gladioli pv. agaricicola* [4] before finally being classified as *Burkholderia gladioli pv. agaricicola* [5]. The soft rot caused by *B. gladioli pv agaricicola* is believed to be caused by degradative enzymes such as chitinase and protease virulence factors that are secreted by the Type II secretion system [5].

Another soft rot disease of *Agaricus bisporus* has also been reported to be caused by the Gram negative *Janthinobacterium agaricidamnosum*, the symptoms of which include pitting and sticky blotches on the mushroom cap and in some severe cases, complete dissolution of the mushroom [6]. Unlike the soft rot disease caused by *B. gladioli pv. agaricicola*, the symptoms of *J. agaricidamnosum* soft rot are caused by a singular virulence factor, jagaricin, which is produced as a secondary metabolite by non-ribosomal peptide synthetases (NRPS) [7].

Several pathogenic Pseudomonads are known to cause discolouration of *Agaricus bisporus*, termed ‘blotch diseases’. This includes *Pseudomonas tolaasii*, that causes brown blotch [8]; *Pseudomonas reactans* that causes purple or light brown blotch; *Pseudomonas gingeri* that causes a reddish ginger-coloured blotch and *Pseudomonas agarici*, that causes both yellow blotch and a disease of the mushroom gill structures [9].

1.2.1 *Pseudomonas tolaasii*

Of all the bacterial blotch disease causal organisms, *Pseudomonas tolaasii* is the most intensively documented in literature. As previously mentioned *P. tolaasii* is the causal organism of "brown blotch disease" on *Agaricus bisporus* [8] and other cultivated mushrooms such as *Pleurotus ostreatus*, also known as the oyster mushroom [10].

The symptoms of brown blotch disease commence as small (1-4mm) brown or cream lesions on the pileus and stipe that become darker and more sunken as the damage progresses, a large and sunken lesion of 1cm in diameter is shown in Figure 2. The individual lesions may merge to cover the entire surface of the mushroom [1].

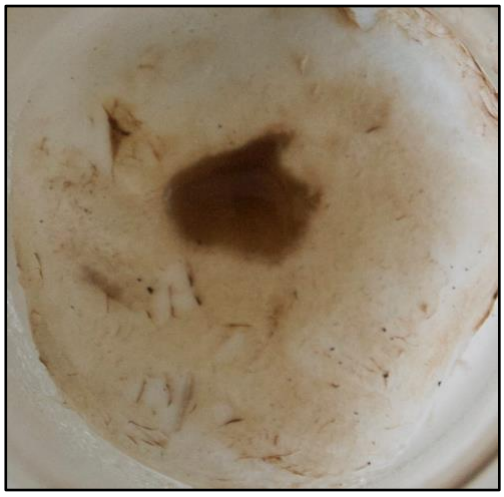

Figure 1.2. A characteristic dark, sunken lesion of approximately 1cm in diameter caused by *P. tolaasii* infection of an *Agaricus bisporus* mushroom.

Mushroom loss due to brown blotch is economically important as *P. tolaasii* was previously estimated to be responsible for 5-10% of crop loss in the U.K and a further downgrading of 10% of mushrooms produced [11]. It has been reported by Wong *et al.* [12] that the primary source of *P. tolaasii* on a mushroom farm is the peat and limestone mix used during the casing process. They have reported that *P. tolaasii* could not be detected in the compost, water supply or mushroom spawn, but was detected in the casing mix [12]. Once *P. tolaasii* has colonized a mushroom bed secondary sources of infection are plenty, including tools and equipment, hands

and clothing of farm workers, airborne dust as well as flies and mites that are frequent pests of mushroom farms [12].

Phylogenetic analysis based on 16S rDNA places *P. tolaasii* in the *Pseudomonas fluorescens* subgroup [13] and has been officially placed in the taxonomic group of fluorescent *Pseudomonas* biotype II [1]. *P. tolaasii* can undergo phenotypic variation, a tactic that is beneficial for its survival in nutrient-poor environments [14]. The wild-type colony morphology is pathogenic, opaque, mucoid, non-fluorescent, able to hydrolyse casein and capable of normal growth on medium containing cetrimide; whereas the phenotypic variant is translucent, non-mucoid, non-pathogenic, fluorescent, does not have the ability to hydrolyse casein or to grow normally on medium containing cetrimide [14, 15], the observable differences in colony morphology are shown in Figure 1.3.

The phenotypic variant of *P. tolaasii* has also been shown to swim faster via flagellar motility and respond more rapidly to chemotactic gradients than the wild-type, as well as produce larger amounts of siderophore, which would confer advantages in nutrient-poor water-saturated environments such as the casing layer in mushroom production [16]. *P. tolaasii* has a distinct polar flagella type that does not conform to the H1 and H2 serotyping methods used to type *Pseudomonas* flagella, similar to *P. fluorescens* which also does not conform, as described by Guillorit-Rondeau *et al.* [17].

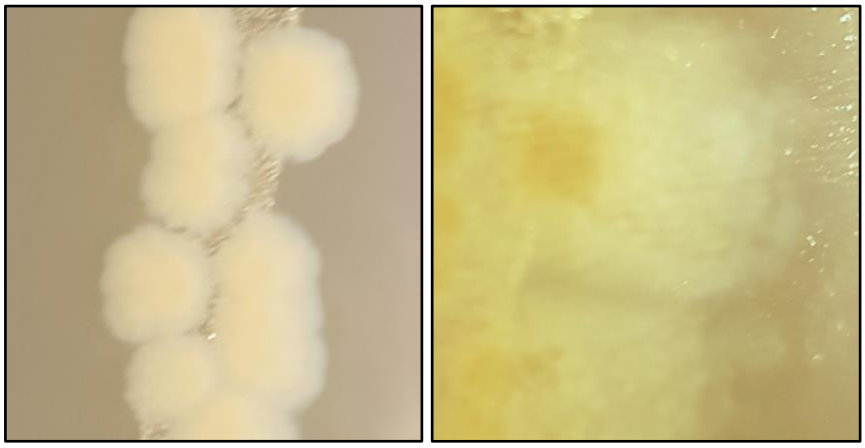

Figure 1.3. Colony morphology of *P. tolaasii* with the pathogenic wild-type colony on the left and the non-pathogenic translucent colony on the right.

The phenotypic shift in *P. tolaasii* is positively regulated by a 99-kDa protein, PheN, which shows amino acid similarities to members of a family of two-component regulatory proteins including LemA of *P. syringae* pv. syringae [18], BvgC of *Bordetella Pertussis* [19] and RcsC of *Escherichia coli* [20] and is required to maintain the wild-type pathogenic phenotype [14]. It is believed that PheN is involved either directly or indirectly with the activation of expression of genes involved in tolaasin synthesis, protease production and mucus production but represses other genes such as those controlling motility, chemotaxis and siderophore production [14].

Sinha *et al.* [21] have demonstrated that inactivation of the *PheN* gene results in the phenotypic shift from pathogenic to non-pathogenic, due to a 661bp duplication within the gene which results in early termination of the *PheN* ORF and the synthesis of a truncated protein lacking the N-terminus periplasmic domain. This N-terminal domain has been deduced to be the environmental sensor, the loss of which is predicted to result in loss of gene function [21]. It has also been shown that phenotypic switching from the pathogenic wild-type to the non-pathogenic phenotype is *recA* dependent whereas the reverse switch appears to be *recA* independent [22]. The precise stimuli that regulate expression are not known; however, it has been suggested that it is a ‘social phenomenon’ possibly due to quorum sensing and

autoregulatory compounds, which occurs when cells are present at a high enough density and in nutrient-poor conditions [14].

The brown blotch disease mushroom pitting symptoms of *P. tolaasii* infection are caused by a secondary metabolite, tolaasin, that is explained in detail in Section 1.3.2. Other enzymes and secondary metabolites of *P. tolaasii* that may play roles in pathogenicity are also outlined in Section 1.3.

1.2.2 *Pseudomonas agarici*

While *P. tolaasii* may be one of the most intensively studied causes of bacterial disease of cultivated mushrooms, *P. agarici* arguably takes the title as the least well studied mushroom pathogen in academic literature. *P. agarici* is a Gram negative, non-spore forming, short rod that can occur singly, in pairs or in chains, and based on 16S rRNA gene topology analysis, belongs to the *P. fluorescens* intragenic cluster [23, 24]. However, further phylogenetic analysis of *P. agarici* involving four 'housekeeping' genes 16S rRNA, *gyrB*, *rpoB* and *rpoD* by Mulet *et al.* failed to place *P. agarici* in any specific *Pseudomonas* subgroup [25].

P. agarici was demonstrated by Young [24] in 1970 to be the causal agent of the disease 'drippy gill' in *A. bisporus* mushrooms. The disease is characterised by the appearance of small, dark pigmented spots on the gills of mature mushrooms which progressively increase to a diameter of 2mm or more, which contain a light cream or grey spot in their centre. The infection of the gills may cause a collapse of the locally affected tissue or in the case of severe infections the bacterial droplets may coalesce between gills and lead to a total collapse of gill tissue [24]. Outside of the gill tissue, the stipes of infected mushrooms also often show fine longitudinal splits that contain large numbers of bacteria, which as the mushroom matures become dark brown [24].

The droplets were also reported to be found inside unbroken veils of immature mushrooms, which appeared to indicate the bacterium was already present within the mycelium [24]; however, it was then shown that *P. agarici* has the ability to degrade the extracellular matrix and pass through protective membranes unnoticed, which allows it to infect the stipe and gills before the veil opens during the maturation process [26, 27]. In cases where the bacteria are observed to be intrahyphal, it is believed that it gains access by a break in the fungal cell wall and thus is opportunistic in this aspect [27].

Chitinase enzymes are found in many *Pseudomonas* species including *Pseudomonas stutzeri* YPL-1 [28], *Pseudomonas aeruginosa* [29] and members of *Pseudomonas fluorescens* biovar I, VI and III, frequently as antagonists of fungi [30]. Chitin is a simple polysaccharide that forms an essential component in the cell walls and septa of all fungi studied to date [31]. *P. agarici* is distinct from other *Pseudomonas* species pathogenic to fungi in its inability to produce the chitinase enzyme, which prevents it from actively penetrating hyphal cell walls [26].

While the majority of large U.K based outbreaks were reported before 1980, outbreaks of *P. agarici* have also been reported in New Zealand in 1991 [26] and in Serbia as recently as 2008 [32].

1.3 Secondary metabolites, non-ribosomal peptide synthesis and bacteriocins in *Pseudomonas*

The *Pseudomonas* genus is well known for its production and use of secondary metabolites, they have been reported to produce compounds such as lipids, phenazines, pyrroles, indoles, amino acids and peptides, pterines and other miscellaneous compounds [33].

One method for production of peptide secondary metabolites in Pseudomonads is via the use of non-ribosomal peptide synthetases (NRPSs), NRPSs are very large multimodular enzymes that synthesise specific peptide products. NRPSs are comprised of multiple modules which each contain three catalytic domains, the first is the adenylation domain (A domain) which is involved in the activation of the amino acid, the second a thiolation or peptidyl carrier protein (PCP domain) which is involved in the extension of the peptide and the third which is the condensation (C domain) domain which is involved with the condensation of the amino acids. A final fourth catalytic domain which is located only in the termination module is the thioesterase (TE) domain [34].

The NRPS produced secondary metabolites perform a wide variety of roles such as in phytopathogenicity, microbial and predatory antagonism, swarming motility and biofilm formation [35, 36].

In Pseudomonads, NRPSs are heavily involved in the production of cyclic-lipopeptides, which are important secondary metabolites. These cyclic lipopeptides produced by Pseudomonads are classified into six groups primarily based on their length and amino acid sequence, these groups are: viscosin, amphisin, tolaasin, syringomycin, putisolvin and syringopeptin [37, 38]. However, some lipopeptides produced do not fit into these groups, these include; entolysin of *Pseudomonas entomophila [39]*, orfamide produced by *Pseudomonas fluorescens* Pf-5 [40]

and linear lipopeptides produced by strains of *Pseudomonas syringae* such as syringofactins of *Pseudomonas syringae pv. tomato* strain DC3000 [36] and a structural variant of syringopeptin, peptin31, produced by *P. syringae* strain 31R1 [40].

A major group of these *Pseudomonas* lipopeptides produced by NRPSs is the viscosin-related nonapeptides, which has six-subtypes that have been identified: viscosin, massetolide, viscosinamide, pseudophomin and white line-inducing principal (WLIP) [41]. Within this group, a minor difference in amino acid sequence confers significantly different biological properties, making predictions on function based on amino acid sequence difficult, such as with the WLIP and viscosin. WLIP produces a precipitate in agar medium when a WLIP producing *Pseudomonas* strain is co-cultured alongside a strain of tolaasin producing *Pseudomonas tolaasii*. However, viscosin, which is extremely similar structurally to WLIP (differing only in the fifth amino acid residue where D-Leu is replaced with L-Leu) or viscosinamide (similar to viscosin but with D-Gln instead of D-Glu at amino acid position 2) fail to produce this precipitate [42].

1.3.1 Tolaasin

While it had been known since 1973 that nutrient broth suspensions of *P. tolaasii* that had been placed on mushroom surfaces, but separated by dialysis membranes, were able to produce disease symptoms [43]; it was not until 1986 that J. T Peng devised a method for isolating the tolaasin toxin. It was reported that addition of a partially purified toxin, which he had identified as a polypeptide between 1 and 10 kDa in weight, reproduced *P. tolaasii* disease symptoms and interacted with the WLIP of *P. reactans [44]*.

The molecular structure of tolaasin was determined by Nutkins *et al.* [45] and was published in 1991, they showed that tolaasin is comprised of 18 amino acids, with a mass of 1,985Da and

due to the presence of two positive charges a the C-terminus, that it is an amphipathic peptide [45].

Nutkins *et al.* reported to have isolated two isomers of tolaasin, Tol I and Tol II, with Tol II being a structural analogue of Tol I, which is identical to Tol I, except for a substitution of homoserine for glycine at amino acid position 16 [45]. Since then, Shirata *et al.* in 1995, identified 8 isomers of tolaasin I, of which 4 and 6 were identified as Tol I and Tol II respectively [46]; and in 2004 Bassarello *et al*. reported the discovery of 5 new minor products, Tol A-E produced by *P. tolaasii* NCPPB 2192 together with Tol I and Tol II, of which only two differ structurally from those hypothesised by Shirata *et al.* [46, 47]

Tolaasin can cause disruption of *Agaricus bisporus* cells via two distinct methods, the first is as a result of its ability to form Zn^{2+}-sensitive ion channels, which is observed at low tolaasin concentrations [48]. The second is due to the surfactant properties of tolaasin, which is observed at higher concentrations. Interestingly, the surfactant properties are reported to be responsible for causing damage on the mushroom caps, whereas it is the ion channel properties that induce cell lysis in cut mushroom tissue [49].

Scherlach *et al.* [50] have demonstrated in *P. costantinii*, which also produces tolaasin, that tolaasin production is dependent on a biosynthetic gene cluster comprised of 5 NRPS proteins. This NRPS cluster is reported to contain 18 biosynthetic 'modules' each comprised of a condensation, adenylation and thiolation domain [50].

1.3.2 Siderophores

Under iron limited conditions the majority of fluorescent Pseudomonads produce yellow-green fluorescent siderophores called pyoverdines [33]. While pyoverdine was first described as a pigment in the late 19th century [51], it was not until 1978 that the link between Fe3+ concentrations and *P. fluorescens* pyoverdine synthesis and its role in the transport of iron was correlated [52]. Pyoverdines are powerful iron(III) scavengers and transporters with a conserved structure comprised of 3 domains; the first is a dihydroxyquinolone chromophore, the second, a variable peptide chain of between 6-12 amino acids in length and finally a third domain comprising of a side chain which is usually either a dicarboxylic acid or dicarboxylic amide [53].

Several genes have been identified to be involved in the biosynthesis of pyoverdine in *P. aeruginosa*, *P. putida* [54] and *P. fluorescens*, including the NRPSs *pvdD* [55] and *pvdIJK* [56] of *P. aeruginosa*. Likewise, Mossialos *et al.* have identified a NRPS of *P. fluorescens* ATC 17400, PvsA that is necessary for pyoverdine biosynthesis [57].

The biosynthesis of siderophores by NRPSs within the *Pseudomonas* is not limited to pyoverdine, with a second siderophore of *P. aeruginosa*, pyochelin, also synthesised by NRPSs [58]. Synthesis of pyochelin in *P. aeruginosa* is dependent on two gene clusters, *pchEF* [59] and *pchDCBA* [60], with the proteins PchE and PchF having been identified as showing significant characteristics of NRPSs.

However not all of *Pseudomonas* siderophores are dependent on NRPSs for biosynthesis. For example, *P. syringae* pv. syringae B728a has been demonstrated by Berti and Thomas [61] to produce a siderophore, achromobactin, via the NRPS-independent synthetases AcsD, AcsA and AcsC from the precursor citrate.

1.3.3 Bacteriocins and phage tail-like bacteriocins

Bacteria possess and produce a diverse range of defence systems against other microorganisms, including broad-killing antibiotics and bacteriocins, which often demonstrate more narrow killing activity [64].

Phage tail-like bacteriocins (PTLBs) are an example of antimicrobial peptides produced by bacterial ribosomal protein synthesis, in contrast to the peptides produced by alternative mechanisms such as NRPSs. While other antimicrobial compounds produced by the Pseudomonads tend to offer antimicrobial action against a broader spectrum of phylogenetically distant species, bacteriocins such as the pyocins produced by *Pseudomonas aeruginosa* frequently possess a narrower target range which often include closely related *Pseudomonas* species [62] [63].

Bacteriocins, such as PTLBs, offer a commercially attractive alternative to chemical preservatives in food production and preservation for a variety of reasons and their use in these situations has been well documented in literature. The primary allure of the use of bacteriocins stems from their sheer variety and abundance, with estimates that over 99% of bacteria are capable of producing one or more bacteriocin [64], offering a large pool of potential commercially viable products. Furthermore, as public opinion on the use of chemical preservatives in food has shifted and the legal and the moral implications of the presence of antibiotics or their associated residues in food the appeal of proteinaceous ‘naturally’ produced compounds such as bacteriocins that have been demonstrated to be sensitive to and degraded by proteases in the stomach [65], is amplified. Finally, the use of bacteriocins, such as the FDA approved nisin that is produced by *Lactococcus lactis* [64, 66], in food to prevent spoilage has been well studied and demonstrated to be both safe and effective.

While the ability of bacteria to produce PTLBs has been known for over 60 years, to date few studies have specifically detailed the scope of their production among bacteria. PTLBs are very large ($1x10^6$-$1x10^7$ Da) protein structures that are encoded in distinct gene clusters, similar to those of the related phage tail proteins encoded in phage structural protein regions, which often likewise contain assembly and/or chaperone proteins necessary for assembly as well as associated regulatory genes [67].

PTLBs are comprised of two broad groups, R-type and F-type, as illustrated in Figure 1.4. The F-type PTLBs are phylogenetically related to the tails of *Siphoviridae* bacteriophage and are non-contractile, comprised only of a tube and no sheath. The R-type PTLBs are related to the tails of *Myoviridae* phage and are contractile, consisting of a central tube surrounded by a sheath which is connected to an additional baseplate structure to which receptor-binding proteins such as tail fibers are bound [69] [67].

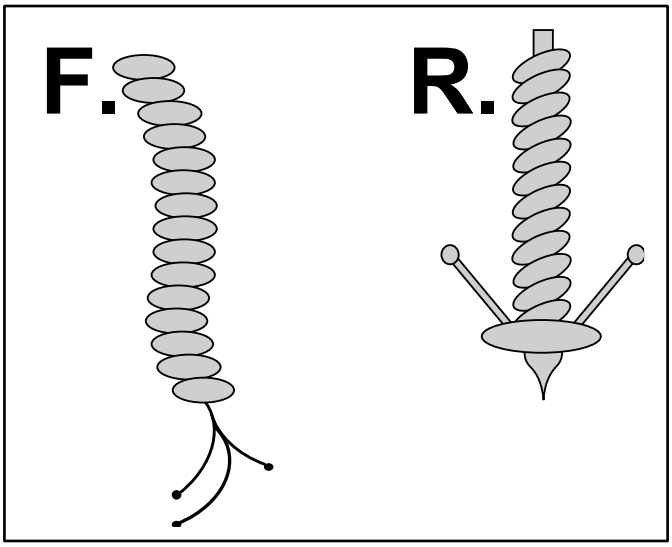

Figure 1.4. Diagrammatic representation of: F. F-Type non-contractile PTLB consisting only of a central tube and no sheath; R, R-Type contractile PTLB possessing both a tube and contractile sheath as well as a baseplate and tail fiber-like proteins. Generated in Microsoft PowerPoint, based on data from [67, 68].

Examples of R-type and F-type PTLBs include the R-type and F-type pyocins produced by *Pseudomonas aeruginosa*, which are reported to be produced by 90% of *P. aeruginosa* strains. It has been demonstrated by Köhler *et al.* [62] that in the case of the R1, R2 and R5 (R-type) pyocins, that the binding receptors are specific residues of the lipopolysaccharide core, accounting for their O-serotype specific binding affinities.

The precise stimuli behind the timing of PTLB expression is not known although it has been reported that expression can be induced by DNA damage [67]. In *Pseudomonas fluorescens* SF4c PTLB expression can be induced by mitomycin C, UV light and hydrogen peroxide [70]. In *Pseudomonas aeruginosa* PAO1, control over expression of the pyocin genes is controlled by the *prtN* gene, which encodes a transcriptional activator, which in turn in controlled by the PrtR repressor protein. PrtR is reported to be inactivated by the presence of the RecA protein which is activated by DNA damage [71].

1.4 Bacteriophage Treatments of *Pseudomonas* infections

Bacteriophage are viruses that infect and frequently kill bacteria, which were first discovered early in the 20th century by Fredrick Twort and Felix d'Hérelle [72]. Bacteriophage offer an attractive biocontrol agent for *Pseudomonas* infections of cultivated mushrooms as some *Pseudomonas* strains such as *Pseudomonas putida*, which colonises the casing layer, are beneficial for mushroom growth and maturation [73]. Therefore, total sterilisation of the casing material could potentially harm mushroom production and a more targeted approach is necessary.

While little research has been conducted on phage capable of infecting *P. agarici*, several phages have been characterized that utilise *P. tolaasii* as a host. Recently Kim *et al.* have

isolated 21 environmental phage of *P. tolaasii* of varying lytic ability [74]. With two phage, hb1a and hb2d, they achieved complete bacterial lysis after 12 hours with a multiplicity of infection (MOI) of 100. Similarly, they also managed to suppress blotch formation on oyster mushroom (*P. ostreatus*) surfaces when phage hb1a was added simultaneously with *P. tolaasii* [74].

1.5 Bacteriophage background

Bacteriophage are thought to be the most numerous organism on Earth, with estimates of the global population of bacteriophages being as high as 10^{31} virus particles [75]. Broadly, bacteriophage can be split in to two groups based on life cycle, either 'lytic' in the case of virulent phage or 'lysogenic' in the case of temperate phage. A lytic or virulent phage lyses the host cell soon after initial infection, whereas a lysogenic or temperate phage may establish a latent infection and become dormant for a period of time [76]. Lysogenic infections most often involve integration of the phage genome into the host chromosome or plasmid as a 'prophage', which is passively copied to daughter cells when the host cell divides. A prophage can be activated in response to environmental stimuli, for example if the host cell is damaged, to produce phage via the lytic cycle [76]. Other temperate phage establish lysogenic infections by forming autonomous plasmids within the host [76, 77].

Taxonomic classification of bacteriophage is based on several factors, including host range, phage particle morphology, genome type and protein structures such as tails, the most important of these factors being the phage particle morphology and genome type. The largest order of bacteriophages is the *Caudovirales*, which are the tailed dsDNA containing phages, that represent 96% of all known phage [78].

The *Caudovirales* comprises of three families, depending on tail structure; the *Myoviridae* that possess contractile tails consisting of a sheath and a central tube; the *Siphovoridae* which possess long, non-contractile tails; and the *Podoviridae* which possess short, non-contractile tails. The *Podoviridae* will be explored in greater detail in Section 1.5.1.

1.5.1 The *Podoviridae*, subfamily *Autographivirinae* and genus *Luz24likeviruses*

The *Podoviridae* are a family of bacteriophage within the *Caudovirales* that are characteristically non-enveloped, with a dsDNA genome, an icosahedral head and a short non-contractile tail, as illustrated in Figure 1.5. As of 2016 the *Podoviridae* family now comprises of three subfamilies, the *Autographivirinae*, *Picovirinae* and the newly added *Sepvirinae* as well as 21 genera, such as the *Luz24likevirus* [79].

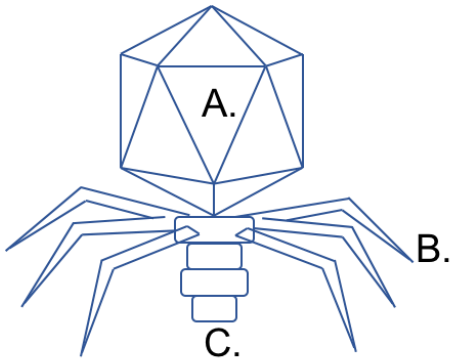

Figure 1.5. Diagram representing standard virion morphology of the *Podoviridae*, visible are A. Icosahedral head, B. Tail Fibers, C. Short Tail. Generated by Microsoft PowerPoint, based on data from Fokine and Rossman [80].

The *Autographivirinae* are a subfamily of the *Podoviridae* and is comprised of 7 confirmed genera to date: *Fri1virus*, *Kp32virus*, *Kp32virus*, *Phikmvvirus*, *Pradovirus*, *Sp6virus* and *T7virus [81]*. The *Autographivirinae* contains 40 species to date of which only 3 are unassigned to a genus, *Prochlorococcus virus PSSP7 and Synechoccus virus P60* and *Syn5* [81]. The bacteriophage *Escherichia virus T7*, also known as *Enterobacteria phage T7* or simply T7 phage, is the most extensively studied phage of the *Autographivirinae* subfamily. Phage T7

was first isolated and identified around 1944 by Demerec and Fano [82] who demonstrated its ability to cause large plaques with a large halo on plates of *E. coli*. Since it was identified it has been the subject of intense study, with the first genetic 'map' of T7 being published in 1969 by William Studier who identified a total of 19 genes [83], from which he identified 12 protein products [84].

In 1983 the complete nucleotide sequence of phage T7 was published by Dunn *et al.* [85], who reported that the genome contained 39,936bp with a low GC content of 49% and coding for 50 genes. The nucleotide sequence of T7 has been subject to several minor revisions in subsequent years, including the addition of a single 'T' nucleotide, bringing the total nucleotide count to 39,937bp [86-88]. It is now predicted that the genome of phage T7 contains up to 57 genes encoding 60 potential proteins, of which only 35 have a known function [89]. The genes of phage T7 are arranged linearly within the genome and are categorised in to three temporal classes [90, 91]:

Class I: These genes are expressed early in in infection and encode proteins that are involved in host conversion.

Class II: These genes are expressed mid-infection and are encode proteins primarily involved in DNA replication.

Class III: These genes are expressed late in infection and encode proteins involved in virion structure and assembly.

Structurally, phage T7 is characteristic of the *Podoviridae*, composed of an icosahedral capsid 'head' ~60nm in diameter [92], a short stubby tail of 23nm in length that is attached to the capsid by way of a head-tail connector protein [90]. Attached towards the head end of the tail are six tail fibers, each composed of three copies of a single protein Gp17, that form a kinked structure [93]. Initial host recognition is performed by these tail fibers, although this stage is

reversible; before a second irreversible host cell binding is hypothesised to be mediated by a tail-tube protein [94]. Upon binding, genomic DNA is injected in to the host by extension of the T7 tail [95]. Only a small fraction of the genomic DNA (~1kb) is ejected at this initial stage of infection and is pulled in to the host cell by the action of the host RNA polymerase, which recognises promotors present early in the phage genomic DNA [96]. Transcription of the phage DNA is halted approximately 19% through, at which point the phage RNA polymerase that has been produced from the initial genes begins transcription, acting to 'pull' the remaining genomic DNA from the phage particle [96, 97]. Temporal regulation of phage T7 genes throughout replication is achieved by promotors specific for T7 RNA polymerase, while the host RNA polymerase is inhibited by a Class II gene Gp2 to prevent the detrimental effects of uncontrolled transcription [91, 98]. The specificity of T7 RNA polymerase for T7 promotors has resulted in the adaptation of T7 RNA polymerase for a number of research applications.

PhiKMV-like viruses

The "*PhiKMV-like viruses*" are a phage genus of the subfamily *Autographivirinae* which contains 4 confirmed species to date, including the type species *Pseudomonas* phage ϕKMV, *Pseudomonas* phage LKA1 and *Pantoea virus LIMElight* [99].

The type strain, *Pseudomonas* phage ϕKMV was first identified in literature in 2003 by Lavigne *et al.* [100], and was reported to infect several clinical isolates of *Pseudomonas aeruginosa* including *P. aeruginosa* PA01. ϕKMV possesses a dsDNA genome of 42,105bp in length with a relatively high GC content of 62.3%, which is predicted to encode 48 ORFs [100]. The genomic structure of ϕKMV is similar to that of phage T7, illustrated in Figure 1.6, in that it comprises of three distinct clusters: 1. Early genes (host conversion), 2. Middle genes (DNA metabolism), 3. Late genes (structural and host cell lysis); with the exception of the

location of the DNA-dependent RNA polymerase which, in ϕKMV, is located in the middle-stage genes, rather than within the early genes as is the case in T7 [100].

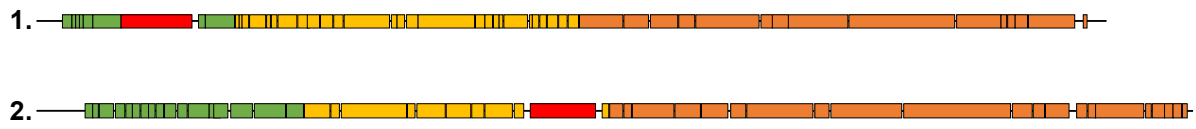

Figure 1.6. Genomic organisation diagram of 1. *Escherichia virus T7*, 2. *Pseudomonas* phage ϕKMV. Early (host conversion) genes are illustrated in green, middle genes (DNA metabolism) in yellow, late (structural and host cell lysis) genes in orange; the location of the DNA-dependent RNA polymerase gene is illustrated in red. Diagram generated in Microsoft PowerPoint from data reported by Lavigne *et al.* [100] and Dunn *et al.* [85].

For ϕKMV, adsorption and infection is dependent on the type IV pili, as reported by Chibeu *et al.* [101]; who identified a spontaneous mutant of *P. aeruginosa* PAO1 that was resistant to ϕKMV infection, named *P. aeruginosa* PAO1KR. Using cosmid complementation they found that a common regulatory mechanism of, or an interaction between, the gene products of *ponA* and *pilMNOPQ* is responsible for the demonstrated resistance, due to atypical type IV pili expression caused by poorly regulated transcription of these genes [101]. Host cell lysis for phage ϕKMV is achieved using a Signal-Anchor-Release (SAR) endolysin system comprising of a holin, endolysin, Rz and Rz1 proteins [102], which will be detailed further in Section 1.5.4.

The phage *Pantoea virus* LIMElight and *Pantoea virus* LIMEzero, infect the bacterium *Pantoea agglomerans* [103], a Gram negative member of the *Enterobacteriaceae* that, similar to the *Pseudomonadaceae*, are ubiquitous in nature and found in water, soil and plant surfaces, as a host [104, 105]. The phages LIMEzero and LIMElight were isolated from soil samples in Belgium and were identified as causing plaques on the *Pantoea agglomerans* strain GBBC2043. However, the plaques of each phage varied in morphology with those of LIMEzero showing plaques of ~2mm with a halo like appearance while LIMElight showed a

clear ~1mm plaque morphology [103]. Both phage have a classic *Podoviridae* virion structure, with EM images reported as showing a head of 60nm in diameter, a short (12nm) stubby tail and a number of tail spikes. While both phage share similarities in terms of morphology and host, there are stark differences in their growth characteristics; while the latent period for phage LIMEzero is relatively short, at 30 min, the latent period for LIMElight is significantly larger at 270 min; similarly, while the adsorption constant of LIMEzero is high, at $8.2x10^{-9}$ ml/min, the adsorption constant of LIMElight is significantly lower at $1.2x10^{-9}$ ml/min [103].

The genomes of LIMElight and LIMEzero are relatively similar in size, at 44,546bp and 43,032bp respectively (54% and 55.4% GC content) and are predicted to encode 57 and 55 ORFs respectively. However, to date, direct terminal repeats of 277bp have only been identified in the genome of LIMElight, while they are expected in the genome of LIMEzero. Both phage possess 18 ORFs that are similar in amino acid sequence to those found in ϕKMV, there is no significant DNA similarity between either LIMEzero or LIMElight and ϕKMV. Phage LIMEzero and LIMElight share the common genome architecture of the *PhiKMVlikevirus* genus, with three distinct regions of early, middle and late genes with the unusual location of the DNA-dependent RNA polymerase, which is located in the middle-stage genes, rather than within the early genes [100, 103].

Luz24likeviruses and UFV-P2

The *Luz24likevirus*, is a genus of phage within the family *Podoviridae*, similar to the subfamily *Autographivirinae*. The *Luz24likevirus* genus to date contains 7 confirmed species, including the type strain *Pseudomonas virus* LUZ24 and *Pseudomonas virus* PaP3 and PaP4 [81].

The virulent type strain, LUZ24, was originally isolated from hospital sewage samples in Belgium, where it was observed to cause zones of lysis on lawns of the *P. aeruginosa strain* Li010. The phage particle morphology of LUZ24, of an icosahedral head of 63nm and a short tail of 12nm in length, is characteristic of the *Podoviridae* family. The LUZ24 genome consists of a linear dsDNA of 45,625bp (52% GC), includes two 184bp direct terminal repeats, and possesses 68 protein encoding ORFS in a bidirectional organisation, with 47 encoded on the positive strand and 21 on the negative [106].

Phage UFV-P2, a tentative member of the *Luz24likeviruses* was isolated from dairy industry waste water in Brazil and shown to be capable of infecting *P. fluorescens* 07A. UFV-P2, similar to LUZ24, possesses a linear dsDNA genome of 45,517bp (51.5%GC) and was initially reported to encode 41 ORFs in a bidirectional organisation with 19 encoded on the positive strand and 22 on the negative strand [107]. However, further analysis revealed the genome to encode a total of 75 ORFs. 55 of these ORFs are encoded on the positive strand and are involved in nucleotide biosynthesis and replication; and 20 encoded on the negative strand involved with virion particle assembly and composition, DNA packaging and host cell lysis [108].

Alignment and pairwise comparisons of the genomic sequences of phage UFV-P2 and other members of the *Podoviridae* showed that UFV-P2 shares a large degree of nucleotide identity with other members of the *Luz24likeviruses*, such as LUZ24, PaP3 and PaP4 (56.8%, 56.7% and 56.2% respectively), however it also shared a 46.4% identity with *Pseudomonas* phage PhiKMV. Analysis of conserved locally collinear blocks within the members of the LUZ24 and phage UFV-P2, as well as the presence of a shared conserved bidirectional genome

organisation, are reported to show that UFV-P2 is highly likely to be a member of the *Luz24likevirus* genus [108].

1.6 Endolysin systems

At the end of the late phase of the bacteriophage infective cycle, once progeny phage have been created, they must escape into the environment. To do this, most dsDNA phage such as the *Autographivirinae,* use a lytic enzyme, or lysin, to achieve lysis of their hosts and release progeny virions out into the external environment. These lysins are termed endolysins to distinguish them from a second type of bacteriophage lysin, a glycosylase, that is associated with the phage tail proteins and is involved in penetration into the host at the start of the infection cycle [109].

Endolysins attack the four major bonds in the peptidoglycan host cell wall via a variety of actions [110]. This catalytic activity can be classified into several groups, the N-acetylmuramidases (lysozymes or muramidases), endo-β-N-acetylglucosaminidases (glucosaminidases), N-acetylmuramoyl-L-alanine amidases (NAM-amidases), endopeptidases and lytic glycosylases. The glucosaminidases and lytic glycosylases cleave the glycan component of the peptidoglycan, whilst the endopeptidases cleave the peptide cross-bridge and the NAM-amidases hydrolyze the amide bond connecting the sugar and peptide constituents of the peptidoglycan. [110-112]. Despite the variety of enzymatic activities, the endolysins weaken the integrity of the cell wall and bacteriolysis occurs when the cell ruptures due to the internal osmotic pressure [113].

1.6.1 Types of host cell lysis systems

The host cell lysis systems of bacteriophage, also termed the 'lysis cassette', can be broadly categorised into two functional classes, canonical and signal-anchor-release (SAR) lysis systems. Both systems require the action of transmembrane proteins, holins, to control the precise timing of host cell lysis, however the mechanisms by which these proteins perform their roles differs dramatically between these two classes [114].

Within the first system, canonical lysis, the majority of endolysins have no intrinsic secretory signals and thus accumulate in the host cytoplasm in their active form, restricted from accessing the peptidoglycan by the host cytoplasmic membrane [109, 115]. In these cases, a second protein is required to allow the endolysin protein to physically access the peptidoglycan, termed a 'holin'.

The holin proteins accumulate in the cytoplasmic membrane until at a genetically predetermined time, they form clusters [116]. These clusters form non-specific lesions, or holes, in the cytoplasmic membrane that allows the endolysin direct physical access to the peptidoglycan layer [117]. The formation of these large, micron scale holes that allow the fully active and pre-folded endolysin access to the peptidoglycan is characteristic of the canonical lysis systems. These holins are also non-specific, that is, they can serve unrelated canonical endolysins of different phage species [116].

The second type of endolysin system is the Signal-Anchor-Release (SAR) lysis system, such as that found in phage ϕKMV and enterobacteriophage P1 [102, 118]. Some phage endolysins possess an N-terminal signal sequence that allows the endolysin to be exported by the host *sec* system to the cytoplasmic membrane where they are tethered and accumulate as inactive

proteins [119]. They become active when their SAR domains leave the membrane and form the mature, fully folded form in the periplasm. This can either occur gradually at a slow rate by spontaneous release, or simultaneously when the cytoplasmic membrane is depolarized [120]. This membrane depolarization is caused at a genetically predetermined time again by the holin, however, these holins which are known as 'pinholins' do not form the large micron scale holes as in the canonical system, but instead form small nanometer scale (~2nm) holes that result in membrane depolarization, rather than allow passage of the endolysin [116]. This depolarization causes the release of the SAR domain of the endolysin, resulting in the refolding of the lysozyme into its active state [102]. Differences in the mechanisms of the canonical and SAR-endolysin systems are illustrated graphically in Figure 1.7.

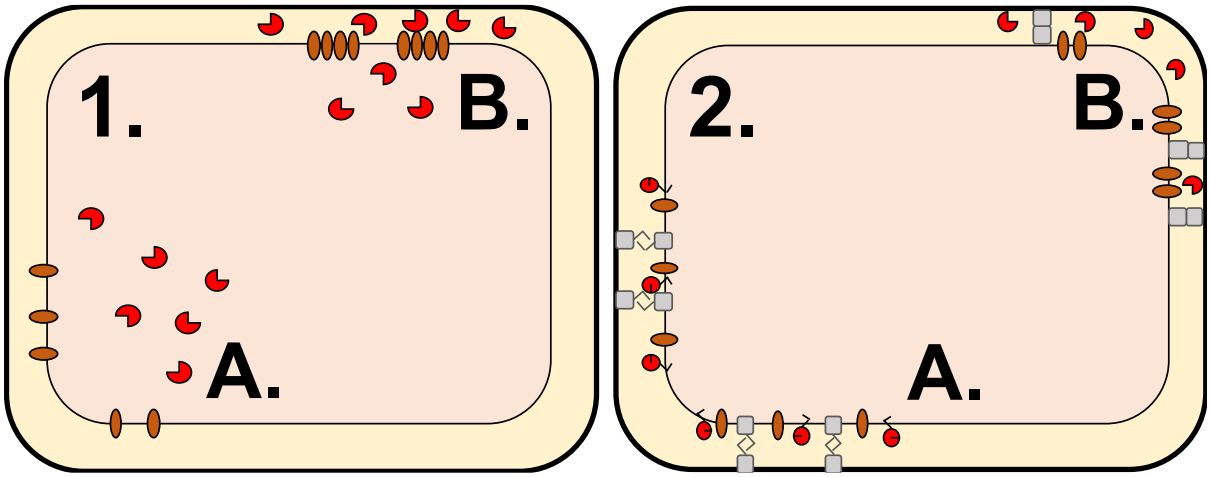

Figure 1.7. Schematic representation of canonical (1.) and SAR-endolysin (2.) lysis systems. Endolysins are illustrated by red circles, holins/pinholins by brown ovals and Rz/Rz1 proteins by grey squares. Within the canonical lysis diagram (1.); A. illustrates the accumulation of active endolysin within the cytoplasm and the accumulation of holin proteins within the inner membrane before reaching a critical concentration in B. showing the formation of holin clusters and micron scale holes allowing the endolysin access to the peptidoglycan. Within the SAR-endolysin lysis diagram (2.); A. illustrates the accumulation of the inactive endolysin proteins, Rz/Rz1 proteins and pinholin proteins within the inner membrane before the critical concentration of pinholin proteins is reached in B. where pinholins aggregate to form pinholes, depolarizing the membrane and releasing active form endolysins, simultaneously the outer and inner membranes are pulled into close proximity by the Rz/Rz1 proteins. Generated in Microsoft PowerPoint, based on data from [121, 122].

Timing of host cell lysis is performed by the holin in both canonical and SAR endolysin systems, however the precise mechanism involved in timing of holin triggering is unknown. The leading model is that the timing of holin triggering is encoded in the physical structure of the holin protein itself and is allele-specific [123]. The holin or pinholin proteins accumulate harmlessly in the cytoplasmic membrane until a critical concentration of proteins is reached, at which point a reduction in membrane potential reaches a triggering threshold causing the holin proteins to aggregate in to two-dimensional 'rafts' [124, 125]. The formation of rafts results in the creation of large holes or pinholes, depending on the lysis system, further reducing membrane potential and instigating further holin aggregation resulting in a complete collapse of the proton motive force (PMF) [126]. This model is validated by the observations that treatment of phage infected cells with compounds that reduce the membrane polarization and PMF results in premature lysis of the host cells [124].

1.6.2 Lysis systems and host cell wall structure

Endolysins instigate host cell lysis by damaging peptidoglycan within the host cell wall. One of the key differences between the Gram positive and Gram negative cell wall structure is the thickness and location of the peptidoglycan layer. In Gram negative bacteria the peptidoglycan lies between the inner cell membrane and outer membrane, within the periplasmic space. It is relatively thin and lacking in surface proteins and carbohydrates. In comparison the peptidoglycan layer in Gram positive bacteria is much thicker, external-facing and is embedded with a variety of surface proteins and carbohydrates [127]. These differences are reflected in the structure and function of the endolysins and lysis systems that have evolved to target Gram negative vs. Gram positive cell walls.

Endolysins that target Gram negative bacteria tend to be single domain globular proteins composed of only a single catalytic domain. However, some endolysins such as that from *Pseudomonas* phage KZ144 have been shown to possess both a lytic domain and an N-terminal cell wall-binding domain [112, 128]. The binding domain of KZ144 is of particular interest as it recognizes a common peptidoglycan composition of all Gram negative species, which provides it with a wide binding spectrum and can be combined with peptidoglycan hydrolases of other phages in order to enhance their specific activity [129].

A further, recently identified, common feature of the lysis cassettes of Gram negative infecting bacteriophage is the presence of Rz/Rz1-like lysis proteins, known as spanins [130]. The *Rz* and *Rz1* genes were initially identified in bacteriophage λ, where it was documented that nonsense mutations in either *Rz* or *Rz1* prevented host cell lysis if the growth medium contained divalent cations at concentrations >5mM [131]. An unusual feature of the phage λ *Rz* and *Rz1* genes which drew initial interest, is that the *Rz1* gene is nested within *Rz* gene in the +1 reading frame [131]. It is now known that the Rz protein resides within the inner membrane and the Rz1 protein within the outer membrane and interact via their associated periplasmic domains, spanning the periplasm and linking the inner and outer membranes [132]. It has been recently demonstrated by Berry *et al.* [133] that the presence of the Rz and Rz1 proteins is essential for complete host cell lysis by phage λ and that expression of λ prophages lacking these proteins prevents lysis and results instead in host cells with a spherical shape. It is believed that fusion of the inner and outer membranes, mediated by a Rz and Rz1 protein complex bringing the two membranes in close proximity following degradation of the peptidoglycan by the endolysin causing host cell lysis [122].

Endolysins that target Gram positive bacteria, which possess no outer membrane but instead have a much thicker peptidoglycan layer that is heavily cross-linked and decorated with surface

proteins and carbohydrates, typically use a modular structure. These modular endolysins typically have one or more catalytic domains and a cell-wall binding domain that often allows strain or species-specific binding [112].

1.6.3 Endolysins in the treatments of bacterial infections

Due to their bacteriolytic activities, endolysins offer an attractive alternative to antibiotics in the treatment of bacterial infections, both within the medical field and food industries. The rise of antibiotic resistance and increasing pressure to reduce the use of antibiotics in agriculture has seen a surge in interest in the use of 'enzybiotics'. The term 'enzybiotic' is a fusion of the words 'enzyme' and 'antibiotic', although it can refer to any enzyme with antimicrobial activity, but most commonly refers to phage associated lysins and was first utilised in this context by Daniel Nelson *et al*. in 2001 [112, 134].

A great deal of research has been performed on endolysins from phage that affect Gram positive bacteria as they can be added exogenously to act as exolysins due to the peptidoglycan, in most cases, being accessible from the outside [111]. For example, Nelson *et al*. reported that with a lysin of Streptococcal phage G, addition of 1,000 units (10ng) of purified lysin *in vitro* is sufficient to sterilize a culture of ~10^7 group A streptococci within 5 seconds [134]. The use of phage endolysins has also been studied *in vivo*, Loeffler *et al.* show that the use of a lysin, Cpl-1 of a pneumococcal bacteriophage has potential as a therapy for pneumococcal bacteraemia in mice. They report that a 2000µg dose of lysin could reduce pneumococcal titres from a median of $\log_{10}$ 4.7 cfu/ml to undetectable levels within 15 minutes; and that this dose given 1 hour post intravenous infection, led to 100% survival at 48 hours, compared to 20% survival of buffer-treated control mice [135].

As these studies show, endolysins offer promise as treatment for bacterial infections, however they are not without their drawbacks. The most pressing being that lysins, as proteins, can trigger an immune response, unlike the majority of antibiotics, that are generally not immunogenic. However, the majority of studies regarding the immunogenicity of phage lytic enzymes are based on their use intravenously, where it has been shown that hyperimmune rabbit serum only modestly inhibits enzymatic activity and that no signs of anaphylaxis or adverse side effects were observed in mice, although human studies would be needed to be certain that there are no adverse effects of endolysin based therapies [135].

While the use of lysins for treating Gram positive infections is relatively straightforward, the use of lysins for treating Gram negative infections is more complex due to the outer membrane shielding the peptidoglycan from their lytic activity. There are several methods that have been proposed to overcome this challenge, including the use of outer membrane permeabilizers. For example, Briers *et al.* have researched the use of various compounds such as ethylene diamino tetraacetic acid disodium salt dehydrate (EDTA), citric acid, poly-L-lysine and polymyxin B nonapeptide (PMBV) as membrane permeabilizers of various *Pseudomonas aeruginosa* strains [136]. They concluded that EDTA was the most suitable, however EDTA has been shown to inhibit blood coagulation at very low concentrations (1.3 $mmol^{-1}$) and thus is unlikely to be used in the treatment of systemic infections or as a treatment for food products [136].

Another approach is the use of high hydrostatic pressure (HHP), a technique that uses a liquid, often water and high pressures (200-500MPa) to inactivate pathogenic and food spoilage organisms [137]. HHP can be combined with lysozymes to inactivate bacteria, and has been shown to inactivate the Gram positive bacteria *Staphylococcus aureus* and *Listeria innocua* in

vitro by Masschalk *et al.* (2002) [138]. In relation to Gram negative organisms, HHP is capable of permeabilizing the outer membrane and allowing access for bacteriophage lysins to act. This has been demonstrated with *P. aeruginosa*, where under high hydrostatic pressure (175MPa) for 15 mins, the bacteriophage endolysins KZ144 and EL188 showed a reduction in bacterial cell numbers of ~3.5 log units, of which two thirds was caused by the lytic enzymes alone (3.5±0.2 log units vs. 1.4±0.2 log units for HHP alone) [139]. As HHP is already used frequently in the food industry it offers a more practical approach to membrane permeabilization, however due to the high pressures involved and large costs associated with producing these pressures, it may not be suitable for all food based applications.

1.7 Host cell receptors and phage host specificity determinants

Bacteriophage specificity is frequently determined by the interaction of a phage receptor-binding protein (RBP) and a specific receptor on the host cell surface, for Gram negative bacteria such as the *Pseudomonas*, these receptors can be extremely diverse and include both protein receptors, lipopolysaccharide receptors [140] and capsular polysaccharides, present in the cell wall or via interaction with a specific bacterial structure such as a capsule, pili and flagellum [141].

Rakhuba *et al.* classify the potential protein receptors of the outer membrane of Gram negative bacteria into 5 classes: 1. Structural proteins interacting with the peptidoglycan layer; 2. Specific and non-specific porins forming membrane channels; 3. Enzymes; 4. Substrate receptors with high affinity; 5. Transport proteins responsible for secretion [141]. Examples include: OmpA, an outer membrane protein of *Escherichia coli* involved in conjugation and resistance to chelating agents and acts as a receptor for bacteriophage K3 binding [142-145]; OmpC, a major outer membrane porin protein of *E. coli* K-12 which acts as a receptor for

phage T4 (however lipopolysaccharide is also required for T4 binding) [146, 147]; and TonA, now named FhuA, and TonB, which are membrane transport proteins of *E. coli* and act as binding receptors for many phage including T1, T5, φ80, and UC-1 [148].

1.7.1 Lipopolysaccharide (LPS) receptors

Lipopolysaccharide (LPS) is ubiquitously expressed by all Gram negative bacteria and is vital to the structural and functional integrity of the bacterial outer wall [149]. LPS is a complex polymer comprised of a lipid moiety attached to a long chain polysaccharide and comprises of three structural regions: 1) Lipid A; 2) An inner and outer polysaccharide core; 3) Distal polysaccharide (O-antigen) [150]. Lipid A, the highly hydrophobic and endotoxically active part of LPS, is different from typical phospholipids found in prokaryotic membranes as it contains six saturated fatty acid chains rather than two saturated or unsaturated chains [149, 151]. The polysaccharide core is branched and contains 6 to 10 sugars and 3-deoxy-D-manno-octulosonic acid and some regions are highly conserved among strains and species. The O-antigen is an immunogenic oligosaccharide which is attached to the polysaccharide core and is comprised of repeating saccharide subunits and differs greatly among Gram negative species [149, 151].

There are several morphologies of colonies caused by LPS truncations; the wild-type which contains intact LPS have the ‘smooth’ morphology, those which have lost the O-antigen portion, termed ‘rough’ and those that have lost both the O-antigen and outer core polysaccharide termed ‘deep rough’[152]. Most phage appear to be specific for a single morphology, either smooth or rough. Those that are specific for the ‘smooth’ type tend to display a very narrow host range as they depend on the O-antigen for binding, which is highly

variable in structure; for example, bacteriophage Sf6 of *Shigella flexneri*, which has a very narrow host range (those with the 3,4 group antigen) [153]. Another example is the "ϕKMV-like" phage LKA1 that infects *P. aeruginosa*, which is dependent on *algC* expression for successful infection [154]; the *algC* gene encodes the enzyme phosphomannomutase and mutants lacking this gene do not express the O-antigen or A-band antigen and the LPS is more similar to that of rough-type colonies [155].

1.7.2 Capsule phage

A large number of bacteria possess an external protective layer or "capsule" that protects them both from the environment and from the host immune system, and can be an important virulence factor of pathogens [156]; for example the capsule of *Pasteurella multocida*, an important livestock pathogen, is a key virulence determinant that prevents phagocytosis by host macrophages [157]. Capsules are typically composed of polysaccharide chains, repeating chains of monosaccharide units linked by glycosidic linkages, known as capsular polysaccharides [158, 159]. While the capsule may offer protection from external stress it also provides an attachment site for many phage species that possess polysaccharide depolymerases as tail-associated proteins [160]. An example of phage that infect via the capsule are the "K Phage" of *Escherichia coli* such as phage phi K1A and phi K1F which are specific for the K1 capsular polysaccharide antigen of *E. coli* [161-163]. However, while some phage such as K1F and K15 appear to have evolved to utilise the K1 capsule as a receptor, the K1 capsule acts to block infection of other phage such as phage T7, which uses a LPS-binding receptor [164]. The use of capsular receptors is not limited to *E. coli* phage, for example the *Camplyobacter jejuni* phage F336, F198, F287, F303 and F326 all utilise a capsular polysaccharide receptor, in this case it is the phase variable O-methyl phosphoramidate moiety [165].

1.7.3 Pili phage

The use of the pili of Gram negative bacteria as a receptor for bacteriophage attachment has been documented as early as 1949 [166]. In order to understand the diversity of phage that target pili, it is first necessary to understand the inherent diversity of pilus types and uses. The pili of Gram negative bacteria, such as those of the *Pseudomonas* genus, can be broadly categorised based on their associated assembly pathways into four categories.

The first category is broadly defined as the 'pili assembled by the chaperone-usher' pathway [167]. This includes the Type I pili that are found on the majority of *E. coli* strains, particularly the uropathogenic *E. coli* (UPEC) strains where binding to the host cells is mediated by the FimH adhesin located at the distal tip of the pilus [168, 169].

The second category is defined as the 'Type IV pili', which are homopolymers of a 15-20kDa pilin subunit and are found in a large variety of Gram negative bacteria including *Pseudomonas aeruginosa*, *Neisseria gonorrhoeae* and enteropathogenic *E. coli* [170, 171]. Type IV pili perform a wide variety of functions including host cell adhesion, DNA uptake in transformation, phage transduction as well as twitching and gliding motility [171]. The assembly of Type IV pili requires a large number of genes, unlike other pilus structures which may require only one or two genes, in the case of *Pseudomonas aeruginosa* studies have shown that Type IV pilus assembly requires at least 35 genes [172]. Due to their ubiquity in Gram negative bacteria Type IV pili present an attractive target and thus a large number of bacteriophage utilise the Type IV pili as a binding site including phage PA1Ø, a member of the *Siphoviridae*, and as previously mentioned, *Pseudomonas* phage ϕKMV infection has been demonstrated to be dependent on Type IV pili. Interestingly, while ϕKMV is limited in host range to strains of *P. aeruginosa*, phage PA1Ø is capable of forming plaques on lawns of *P.*

aeruginosa, *Staphylococcus aureus*, *Staphylococcus epidermis*, *Listeria monocytogenes* among others, despite the common binding receptor [101, 173].

The third category of pili are the curli pili, which were first described in 1989 by Olsén *et al.* as coiled surface structures comprised of a single subunit on the surface of *E. coli* [174]. Curli pili have subsequently been found in both *E. coli* and *Salmonella* species and are implicated in host cell adherence, biofilm formation and surface adhesion [175, 176].

Finally, the fourth category of pili are the CS1 pili, which are a serologically distinct family of pili associated with enterotoxigenic *E. coli* (ETEC) and are also known as the 'alternative chaperone usher family', Class 5 or α-fimbriae [167]. CS1 pili are implicated in mediating adhesion of ETEC cells to the intestinal wall of the host as well as attachment to bovine erythrocytes resulting in hemagglutination [177].

1.7.4 Flagella phage

Bacterial flagella are complicated organelles that protrude out of the bacterial cell surface and can be considered to be the 'propellers' driving bacterial movement, involved in many processes from chemotaxis to exiting biofilms in order to find new habitats [178, 179]. However, similar to pili, as an appendage extruding from the cell surface into the environment they offer a site for bacteriophage binding. For example, the *Salmonella typhimurium* phage iEPS5, a member of the *Siphoviridae*, requires both the presence of flagella and flagellar movement for phage adsorption and successful infection [180]. Similarly, the phage PhiAT1, a member of the *Myoviridae*, which is capable of infecting the phytopathogenic bacterium *Erwinia carotovora* ssp. *Attroseptica*, has been shown to be dependent on the bacterial flagellum for successful infection [181].

1.7.5 Phage receptor binding proteins

The previous section focused on the host receptor to which bacteriophage adsorb. This section describes the phage associated proteins involved in host recognition and binding, the phage receptor binding proteins (RBPs), with a primary focus on the RBPs of the *Podoviridae*, such as the *Escherichia coli* phage T7.

In the majority of *Caudovirales*, that is the tailed bacteriophages, the first stage of recognition and binding to the host cell is performed via the tail-spike proteins or tail fibers, this stage is reversible [94, 182]. Positive recognition by these tail fibers will then lead to a second, irreversible binding to the host membrane followed by DNA ejection by the phage tail proteins [94, 182].

In the case of the initial reversible binding of phage T7 to the host lipopolysaccharide receptor in the *E. coli* outer membrane is mediated via six tail fibers, each comprised of three copies of the protein gp17 [94, 183]. While the tail fibers of the *E. coli* phage T7 simply recognise the host cell surface receptor, in some cases such as with the *Salmonella* phage P22 (*Podoviridae*); not only do the tail spike proteins bind to the O-antigen moiety of the LPS, it also possesses an enzymatic ability that allows it to cleave the O-antigen, although the current purpose of this is unclear it is hypothesised that it may facilitate access to the membrane and a secondary receptor [184].

1.8 Host resistance and phage response

Bacteria and phage have been evolving alongside each other for millions of years and so bacteria have evolved a wide variety of mechanisms for phage resistance, and in response the phage have similarly evolved a variety of mechanisms by which to overcome or avoid these resistance mechanisms. For bacteriophages to be a viable form of antimicrobial treatment or for the identification of novel antimicrobials produced by phages, a greater understanding of the mechanisms behind these adaptations of both host and phage is needed. A brief outline of the various mechanisms of host resistance and the phage responses is outlined below.

1.8.1 Prevention/inhibition of adsorption

Bacteriophage adsorption and successful infection is dependent on access to the host cell surface receptor, therefore the most effective method of resistance to infection is to prevent the interaction between the bacteriophage receptor binding protein and the host cell surface receptor. This can be via several methods, such as blocking the receptor, modification of the receptor or even preventing the receptor from being synthesised, which will be outlined in further detail below.

The first method, blocking of the phage binding receptor has been documented as far back as 1974 where it was reported by Nordström and Forsgren [185] that *Staphylococcus aureus* strains that expressed higher levels of 'protein A', which is located in the outermost part of the cell wall, were more resistant to staphyloccocal phage 80 than those deficient in the protein, and it was hypothesised that this was due to masking of the phage binding receptor. Similarly, Riede and Eschbach [186] have demonstrated that the *E. coli* outer membrane protein OmpA,

which acts as a receptor protein for a variety of bacteriophage including phage K3, can be blocked by expression of the outer membrane lipoprotein TraT, preventing phage adsorption.

The second method, modification of the phage binding receptor, is more likely to occur as a method of superinfection exclusion (detailed further later in this section) in temperate phage, rather than prevention of initial phage infection. For example, lysogenization of *Salmonella enterica* serovar Typhimurium by the temperate bacteriophage P22 results in a change in serotype by modification of the LPS O-antigen, which results in the prevention of further binding by phage P22 and other serovar Typhimurium phage [187]. Similarly, the temperate *E. coli* phage PhiV10 is capable of modifying the host O-antigen to prevent further O-antigen dependent phage infection; however, in this case it is due to the expression of a specific enzyme, a serotype converting O-antigen acetyltransferase, encoded within the PhiV10 genome [188].

The third method is a total removal of the phage binding receptor, for example the O1 *Vibrio cholerae* phage K139 is dependent on the O1 antigen for host cell binding; in response several K139 resistant strains have been identified which completely lack the O1 side chain, caused by an insertion of a transposable element (IS*1004*) into the biosynthetic gene cluster encoding the O antigen [188].

Methods such as blocking or removal of phage binding receptors, by which bacteria acquire resistance, are often permanent and frequently come at a significant cost to the bacteria. For example, the O-antigen of UPEC provides a significant survival advantage to these bacteria and thus modification or loss could have significant knock on effects [189]. One method bacteria use to overcome this is phase-variable transient resistance mechanisms. Phase

variation is frequently described as an ‘all-or-none’ control mechanism; this is because the expression of the factor(s) is either completely on or completely off. However, they are usually reversible, as well as being heritable and are in general randomly switched on and off at high frequency (between 10^{-5} to 10^{-1} per generation), although the frequency can be impacted by environmental signals. This rapid switching results in a phenotypically diverse population, which when applied to bacteriophage binding receptors can be extremely beneficial in maintaining resistant populations [190, 191]. For example, as previously mentioned in Section 1.7.2, the *C. jejuni* phage F336, among others, utilises a capsular polysaccharide as a binding receptor and it has been reported by Sørensen *et al.* that resistance of *C. jejuni* to infection by phage F336 was due to the loss of the capsular polysaccharide receptor caused by phase variation of the gene encoding the receptor [165]. Another example is reported by Kim and Ryu [192] who have demonstrated that the *Salmonella enterica* subspecies *enterica* serovar Typhimurium bacterium is capable of acquiring resistance to phage SPC35 through phase-variable modification of the surface O-antigen.

1.8.2 Exopolysaccharides

Exopolysaccharides (EPSs) are polysaccharide substances secreted into the external environment by many bacterial species. These polysaccharides can be covalently bound cohesive layers as capsular exopolysaccharides which are found in a wide range of bacteria including *Acinetobacter baumannii* [193], *Pseudomonas putida* [194] and even species as diverse as Cyanobacteria [195]; or excreted fully in to the external environment as a viscous material such as in alginate production by *Pseudomonas aeruginosa* [196] or EPS by lactic acid producing bacteria used in the dairy industry [197]. These EPSs can perform a wide variety of roles, from avoidance of host immune systems [193], surface adhesion, biofilm formation [198] and even binding of heavy metals [199].

These EPS compounds or capsules can block access to or mask the phage binding receptor, however many phage have evolved tools to overcome this. For example, in the case of the K1 polysaccharide capsule of *E. coli*, which has been shown to block phage T7 infection, can be broken down by endosialidases produced by the K1-specific phages K1F and K1-5 that can hydrolyze the K1 structure [164]. Similarly, in the case of the *Pseudomonas* phage PT-6, a member of the *Podoviridae*, produces a 37kDa alginate lysase that is capable of reducing viscosity of secreted alginate, presumable to facilitate access to the phage binding receptors [200].

1.8.3 Superinfection exclusion systems

Superinfection exclusion is the process by which a bacteriophage prevents secondary infection of its host by further phage; while not being a true method of host resistance, is still a relevant method by which a phage is prevented from infecting a host cell and so will be covered very briefly. A further example of a superinfection exclusion system is the temperate *Pseudomonas* phage D3112 which utilises a protein, D3112 protein gp05, to inhibit the Type IV pili assembly protein PilB, with the effect of a loss of surface pilation and consequently a loss of twitching motility [201]. Most notably, this loss of surface pili has been demonstrated to prevent the further infection of the host *Pseudomonas aeruginosa* cell by other phage that utilise the Type IV pili as a binding receptor and enabling the hosts to become resistant to superinfection [202].

1.8.4 Restriction modification systems

After succeeding in binding to the host cell and ejecting its genome, a phage may still have to overcome bacterial defence systems in order to complete the infection process. One of the most

well documented is the restriction-modification (R-M) systems that are ubiquitous among prokaryotes [203].

R-M systems are frequently referred to as an aspect of the prokaryotic 'immune system' in that they recognise and attack foreign DNA that enters the cell and are usually comprised of enzymes with two distinct actions: 1) an endonuclease; and 2) a methyltransferase. The restriction endonuclease recognises a specific DNA sequence (a restriction site) as a target for cleavage. The methyltransferase adds methyl groups to either adenine or cytosine within the same target sequence, which prevents cleavage by the restriction endonuclease. Working in tandem, these two enzymes constitute a basic 'self' and 'non-self' recognition system [204]. The methyltransferase modifies the bacterial genome, effectively labelling it as a 'self' sequence. When foreign DNA that enters the cell, such as phage genomes or plasmids, the non-methylated ('non-self') restriction sites are susceptible to cleavage by the bacterial restriction endonuclease.

R-M systems are classified into four major groups (Type I-IV) depending on their subunit composition, cleavage position, cofactor requirement and recognition site and each will be briefly outlined below [204, 205].

Type I

Type I R-M systems are multifunctional hetero-oligomeric enzymes that typically comprise of two restriction subunits (R subunits), a specificity subunit (S subunit) and two methyltransferase (M subunits) that utilise S-adenosylmethionine (SAM) to methylate adenosine within the target sequence [204]. The target sequence recognised by Type I R-M systems are asymmetric and comprised of two separate components and a nonspecific spacer

sequence and the reaction requires SAM, ATP and Mg^{2+} as cofactors [204, 206]. An example of a Type I R-M enzyme is EcoKI, originally isolated from *E. coli* K-12 [207].

Type II

Type II R-M systems are among the most widely documented with over 3600 systems identified to date [204]. Type II restriction endonucleases and methyltransferases recognize the same short, palindromic sequence that is normally between 4-8bp in length; all Type II restriction endonucleases cleave within or adjacent to its recognition site, are usually Mg^{2+} dependent and do not require ATP [203, 204]. An example of a Type II restriction endonuclease is the well-known *Eco*RI enzyme [208].

Type III

Type III R-M systems are comprised of two subunits, a 'Mod' subunit that is involved in site recognition and methylation, and a 'Res' subunit that is involved with DNA and ATP hydrolysis [209]. While the Mod subunit can act independently as a methyltransferase, the Res subunit is only active in a Res(2)Mod(2) complex with the Mod subunit [210]. Type II R-M systems recognize two inversely oriented, asymmetric recognition sequences of 5-6bp in length and require both Mg^{2+} and ATP for cleavage [204]. An example of a Type III R-M system is the *Eco*P1I system of the P1 prophage [211].

Type IV

Type IV R-M systems are unusual in that they target DNA that have been modified, for example DNA that has been methylated [204]. The Type IV R-M systems usually combine both the methyltransferase and restriction endonuclease activities into a single polypeptide chain [212]. While the endonuclease action is reported to be positively impacted by SAM and

Mg^{2+} is required for activity, ATP does not impact on either enzymatic action [204, 212]. Type IV enzymes typically recognise asymmetrical DNA sequences [212]. An example of a Type IV R-M system is the McrBC system isolated from *E. coli* K12 [213].

These R-M systems present a barrier to successful phage infection, however phage themselves have, perhaps unsurprisingly, evolved a multitude of ways in which to overcome this challenge that can be broadly categorised into passive and active systems, as outlined below.

Passive evasion of R-M systems

A phage can be protected from the R-M defence system if the host methyltransferase is fast-acting that it rapidly methylates the injected phage genomic DNA before it is cleaved by the restriction endonuclease. This will also protect the progeny phage when infecting other bacterial cells that express the same R-M system [214]. Alternatively, the phage can acquire mutations that remove the restriction site from the genome sequence: a phage with less restriction sites will have an advantage over a phage with more. As previously mentioned, many restriction endonucleases require specific orientation of their recognition sites and therefore a shift in orientation will confer resistance to this R-M system; this is apparent in the genome of phage T7 where the recognition sites of the *Eco*P15I (a Type III R-M system related to the previously mentioned *Eco*P1I system [210]) are in the same orientation rather than in the inverse orientation required for recognition by the restriction endonuclease [215]. Some phage instead resort to the incorporation of unusual bases into their DNA as a method of avoiding R-M systems, such as the *E. coli* phage Mu whose genome contains a modified purine that confers resistance to a variety of restriction enzymes [216, 217]; or the *Shigella* phage DDVI which contains a methylated guanine which provides a protective function [218].

Active evasion of R-M systems

Active evasion systems involve proteins produced by the phage and can involve methods such as subverting the activities of the R-M systems or by direct inhibition of these systems. An example of subversion would be the activity of the phage λ gene *ral*, which has been shown to prevent restriction of the phage genome in *E. coli* K and *E. coli* B by enhancing the methylation modification properties of their associated R-M genes, thus allowing the genome to be methylated and protected before the host restriction endonuclease can act [219]. Another method by which phage can subvert the host R-M systems is masking of the recognition sites by phage encoded proteins; as is the case with the DarA and DarB proteins of coliphage P1, which are injected into the host along with the phage genome and bind to the phage DNA effectively masking the R-M system recognition sites [220]. Direct inhibition of the R-M systems is a method employed by many phage; for example phage T7 utilises a protein 'ocr' which is similar in structure to the DNA target of the *Eco*KI R-M system and binds to the restriction enzymes, blocking and overwhelming them, thus allowing successful phage infection [221].

1.8.5 Abortive-infection systems

Abortive infection systems, also known as 'Abi(s)' are another bacterial defence mechanism, sometimes referred to as the 'innate immune system' of bacteria. Similar to R-M systems they are a form of post-infection response, however in the case of Abi systems they act by killing the infected cell in an effort to limit the spread of phage throughout the bacterial population, and has therefore been referred to as 'altruistic cell suicide' [222]. These Abi systems are widespread among bacteria and have been documented in a wider variety of species including,

Erwinia carotovora subspecies *atroseptica* [223], *Lactococus lactis* [224], *E. coli* [225] and *Shigella dysenteriae* [226].

Abi systems frequently operate on the basis of a toxin/antitoxin (TA) pair, under normal conditions both the toxin and the antitoxin are expressed, preventing the action of the toxin; upon phage infection the expression of both is suppressed. Antitoxins are less stable than their associated toxins, therefore when both cease being synthesized the antitoxins break down before the toxins, allowing them to act, thus killing the infected cell and preventing completion of prophage development [222-224]. For example, the ToxIN system of the phytopathogen *Erwinia carotovora* subspecies *atroseptica*, the action of the growth-inhibiting protein ToxN is suppressed by the RNA product of the *toxI* gene (an antitoxic RNA); upon phage infection the ratio of ToxI:ToxN is disrupted, which allows ToxN to act and restrict growth [223].

1.8.6 CRISPR/Cas Systems

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and their associated genes (*cas*) are often regarded as the 'adaptive' immune system of prokaryotes. The first CRISPR elements were described in 1987 by Ishino *et al*. [227] although their significance was not noted until early 2002 when it was discovered by Jansen *et al.* [228] that these repeat elements were present in both Archaea and Bacteria, that they were direct repeats between 21-37bp separated by spacer regions of DNA and the presence of the *cas* genes was identified, although the exact function of the system remained elusive. In 2005 Bolotin *et al*. [229] reported that these spacer regions were comprised of extrachromosomal DNA from both phage and plasmid sources, with a ratio of approximately 3:1 phage to plasmid origin; they also reported a correlation between the number of spacer regions to resistance to phage infection although the mechanism that mediated this effect continued to remain elusive. In 2007

Barrangou *et al*. [230] demonstrated in *S. thermophilus* that spontaneous resistance to phage infection corresponded to the addition of 1 to 4 spacers within the CRISPR region and that the addition of these spacers was polarised to one end of the CRISPR region; they also demonstrated through selective addition and deletion of spacers that resistance to particular phage could be specifically modulated by addition of phage homologous DNA within the spacer region. Since 2007, CRISPR/Cas systems have undergone a significant amount of research due to their potential uses in genome editing technologies [231].

In order to understand the mechanisms behind the CRISPR/Cas systems it is first necessary to understand the diversity of the CRISPR/Cas sytems in prokaryotes. CRISPR/Cas systems are broadly categorised into two classes, Class 1 and Class 2, although these are then further categorised into five types and sixteen subtypes [232]. Each of the two main classes has a signature Cas protein, for Type I the signature protein is Cas3 and in Class 2 the signature protein is the Cas9 protein [233]. Type 1 is the most common and is divided into seven subtypes, Type I-A to I-F and I-U [234].

It is now known that the activity of the CRISPR/Cas system can be split in to three distinct stages; the first stage is known as the activation stage, the second is known as the ‘expression and processing’ stage and the third stage is known as the interference stage, illustrated below in Figure 1.8 [235].

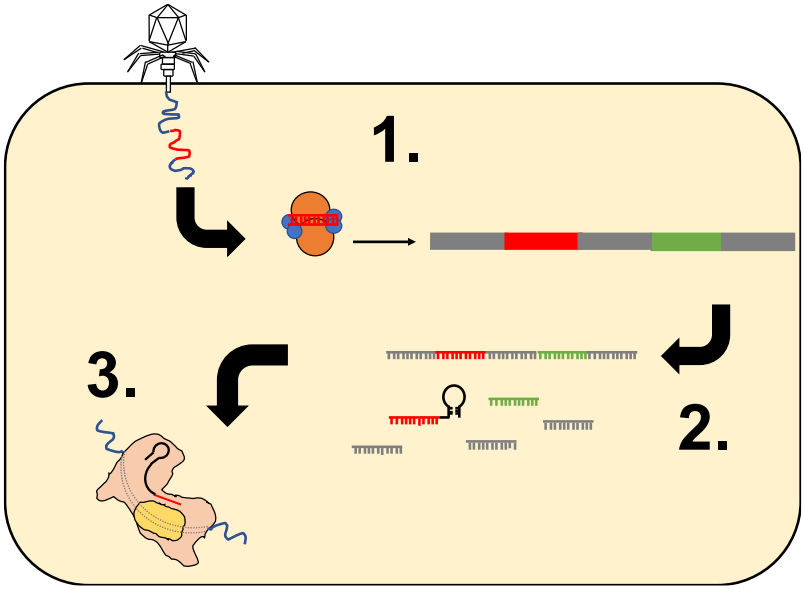

Figure 1.8. Schematic representation of the stages of CRISPR/Cas system. 1. Activation stage: Acquisition of proto-spacers (red) and integration in to the CRISPR loci. 2. Expression and Processing stage: Transcription of the CRISPR locus and processing in to crRNAs. 3. Interference stage: Targeting of foreign DNA/RNA by crRNA and Cas proteins. Generated in Microsoft PowerPoint and adapted from [236, 237].

The first stage, known as the activation stage and as described by Barrangou *et al.* [230], is the stage in which short spacers (known as proto-spacers) that are acquired from extrachromosomal phage or plasmid DNA are integrated in to the CRISPR loci. In most CRISPR/Cas systems the two distinct Cas proteins Cas1 and Cas2 are involved in the acquisition and integration of new spacer by forming an integrase complex [238], however in Type 1 subtype I-F CRISPR systems, such as found in *Pseudomonas aeruginosa* PA14 the Cas2 protein is in fact a fusion of the Cas2 and Cas3 proteins into a single protein (Cas 2/3) [239]. Some CRISPR/Cas systems also require the presence of short motifs adjacent to the proto-spacer, known as proto-spacer adjacent motifs (PAMs), for acquisition of these proto-spacer sequences [240].

The second stage, the ‘expression and processing’ stage is where the CRISPR locus is transcribed into a long transcript, known as the precursor transcript (pre-crRNA), which is then

processed into small crRNAs by endoribonucleases that cleave the repeat regions, either bacterial RNase III [241] or, as is the case for Type 1 CRISPR/Cas systems, an endoribonuclease subunit of a larger protein complex known as Cascade (CRISPR-associated complex for antiviral defence) [235].

The final stage, the interference stage, is where the 'invading' DNA or RNA is targeted by a ribonucleoprotein complex comprised of the crRNA guide and Cas proteins which cleaves the target DNA/RNA [232].

Phage CRISPR/Cas Resistance

Similar to the resistance mechanisms of phage again R-M systems, phage resistance mechanisms against CRISPR/Cas systems can be broadly categorised into active and passive mechanisms. Passive mechanisms involve mutations in either the phage proto-spacer sequence or PAM that confer protection from the host CRISPR/Cas system; it has been shown that a single base mutation in the PAM region or certain positions within the proto-spacer can confer resistance to CRISPR/Cas degradation, these can be substitutions or deletions [242]. It is worth noting, however, that it has also been demonstrated that the CRISPR/Cas systems can be tolerant of single nucleotide mismatches at certain positions within the proto-spacer sequence [243]. It has also been reported by Datsenko *et al*. [244] that these substitutions within the proto-spacer or PAM that confer initial protection from the CRISPR/Cas system actually stimulate the acquisition of new spacers.

Active mechanisms involve proteins produced by phage in order to inactivate the host CRISPR/Cas system. An example of this is the recently reported anti-CRISPR proteins identified by Bondy-Denomy *et al*. [245] in temperate phage of *P. aeruginosa*. Bondy-Denomy *et al*. have identified 5 proteins produced by the aforementioned phage that inhibit the Type 1

subtype I-F CRISPR/Cas system found in *P. aeruginosa* strains (although not the closely related subtype I-E systems); these anti-CRISPR proteins have been shown not to protect specific DNA sequences or to impact on the expression of the CRISPR loci or *cas* genes, but to exert their inhibitory effect at a stage after the formation of the crRNA-Cas complex [245].

1.9 Aims and Objectives

The primary aim of this study is to: 1) perform genetic analyses of *P. tolaasii* NCPPB 2192T and *P. agarici* NCPPB 2472 in order to better understand how they colonise and cause disease symptoms on *Agaricus bisporus* mushrooms, and; 2) to isolate and characterise bacteriophage that are capable of infecting these bacteria. Characterisation and genomic sequencing of the identified bacteriophage should elucidate the interactions between the phage and the host bacterium, with the potential for the whole phage or phage proteins to be used as a future treatment of *P. tolaasii* or *P. agarici* infections. These aims will be achieved by working through the following objectives:

1. Extract and sequence genomic DNA of both *P. tolaasii* NCPPB 2192T and *P. agarici* NCPPB 2472. Perform assemblies and genome mining for potential virulence factors and phage susceptibility or resistance factors.
2. Isolate, identify and characterise bacteriophage capable of infecting *P. tolaasii* NCPPB 2192T and *P. agarici* NCPPB 2472, including full genome sequencing. This will provide the basis for analysis into the interactions between phage and host bacterium and allow the identification of phage proteins with the potential for use in treatment of *P. tolaasii* and *P. agarici* infections.
3. Investigate the genetic basis of phage/host interactions. Full genomic sequencing and characterisation of both the phage and bacterial host achieved in the previous objectives should allow for the analysis of phage/host interactions at a genetic and transcriptomic level using RNA sequencing (RNA-seq).

Chapter 2:
Materials and Methods

2.1 Bacterial strains, phage preparations and growth conditions

2.1.1 Bacterial strains

Bacterial strains used in this study are outlined below in Table 2.1.

Table 2.1. Bacterial strains used in this study. (R.W.J.: Gift from Professor Robert Jackson)

Strain:	Source
P. tolaasii NCPPB 2192T	R.W.J.
P. agarici NCPPB 2472	R.W.J.
P. sp NS1	*Agaricus bisporus* mushroom surface. This study
P. syringae pv. morsprunorum R1 5244	R.W.J.
P. syringae pv. syringae 9097	R.W.J.
P. syringae pv. morsprunorum R2 5255	R.W.J.
P. fluorescens Pfo-1	R.W.J.
P. aeruginosa 14207	R.W.J.
Bl21-AI™ One Shot® Chemically Competent *E. coli*	Invitrogen
One Shot® TOP10 Chemically Competent *E. coli*	Invitrogen

2.1.2 Proof of pathogenicity

In order to demonstrate that the bacteria *P. tolaasii* 2192T and *P. agarici* 2472 were capable of causing the symptoms of disease associated with brown blotch and drippy gill respectively on *Agaricus bisporus* mushrooms, a series of inoculations were performed on excised 'button' (immature, unopened) mushroom caps sourced from a local supermarket, that showed no prior signs of browning or lesions.

A single non-fluorescent colony of either *P. tolaasii* NCPPB 2192T or *P. agarici* NCPPB 2472 was taken from a streak plate on KB agar and used to inoculate 10 ml of KB, before overnight incubation at 28°C with shaking. This overnight culture was then pelleted at 1,500 *g* for 30 minutes and the pellet resuspended in 2ml Phosphate Buffered Saline (PBS) (NaCl 8 gL^{-1}; KCl 0.2 gL^{-1}; Na_2HPO_4 1.44 gL^{-1}; KH_2PO_4 0.24 gL^{-1}). 10 µl of the bacterial suspension was then applied to either the mushroom cap for *P. tolaasii* or for *P. agarici* the mushroom hymenial tissue was removed and the suspension applied to the exposed gill tissue. Care was taken to

avoid damaging the tissue during pipetting. The inoculated mushrooms were then placed into sterile beakers which were sealed with cling-film before being incubated at 28°C until disease symptoms appeared.

2.1.3 Sampling and sample preparation for phage isolation

Samples of water were taken from several environmental sources including the River Thames, sampling locations shown in Figure 2.1, and un-treated sewage. 10 ml of these samples was combined with an equal volume of overnight culture of either *P. tolaasii* NCPPB 2192T or *P. agarici* NCPPB 2472, followed by overnight incubation at 28°C with 220 rpm shaking. Post-incubation the bacteria were pelleted by centrifugation at 1,500 *g* and the supernatant was removed and filtered with a 0.45 µm pore-size syringe injection filter (Millipore) and stored at 4°C.

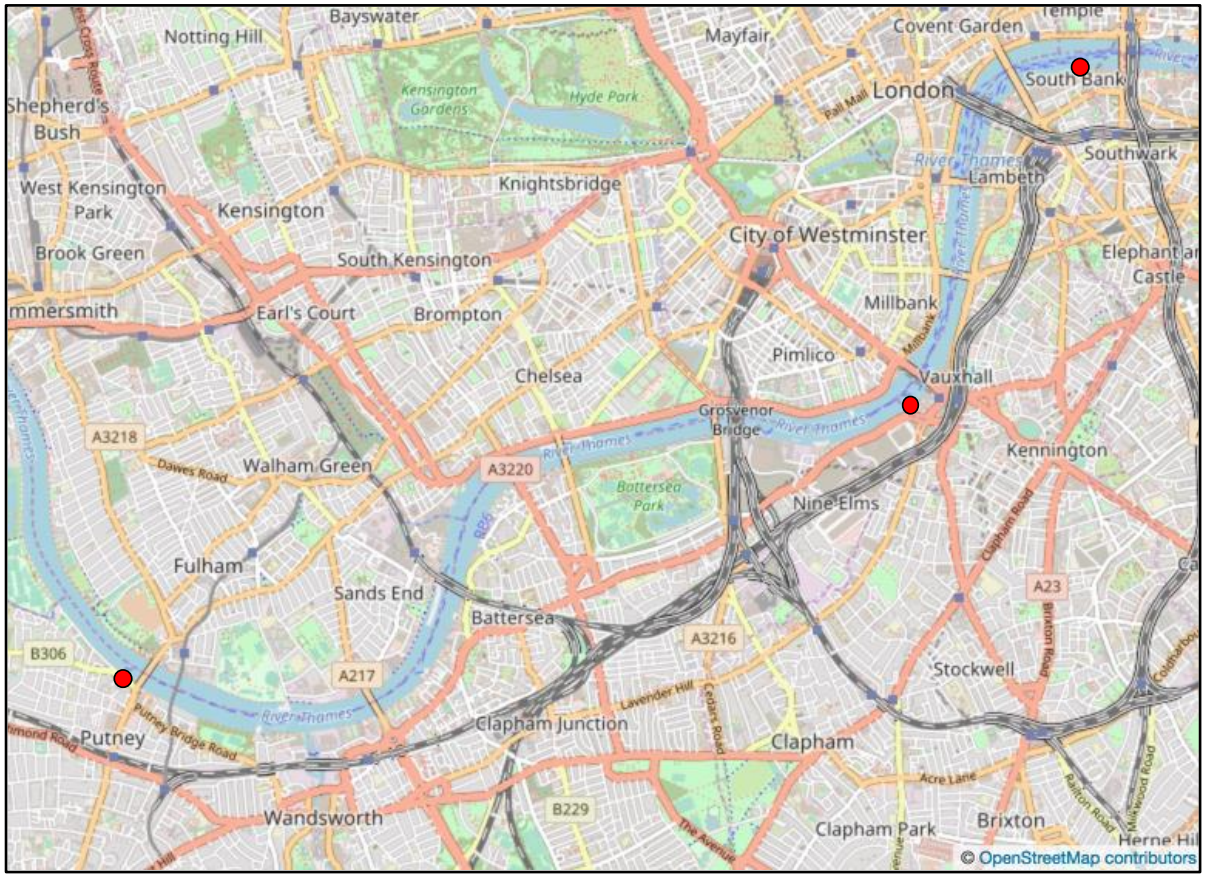

Figure 2.1. Map of South-West London with bacteriophage sampling locations marked (red) © OpenStreetMap contributors.

Environmental samples were taken from locations around the University of Reading campus that correspond to likely habitats for wild mushroom growth for example, partially or fully shaded leaf litter and soil, sampling locations shown in Figure 2.2.

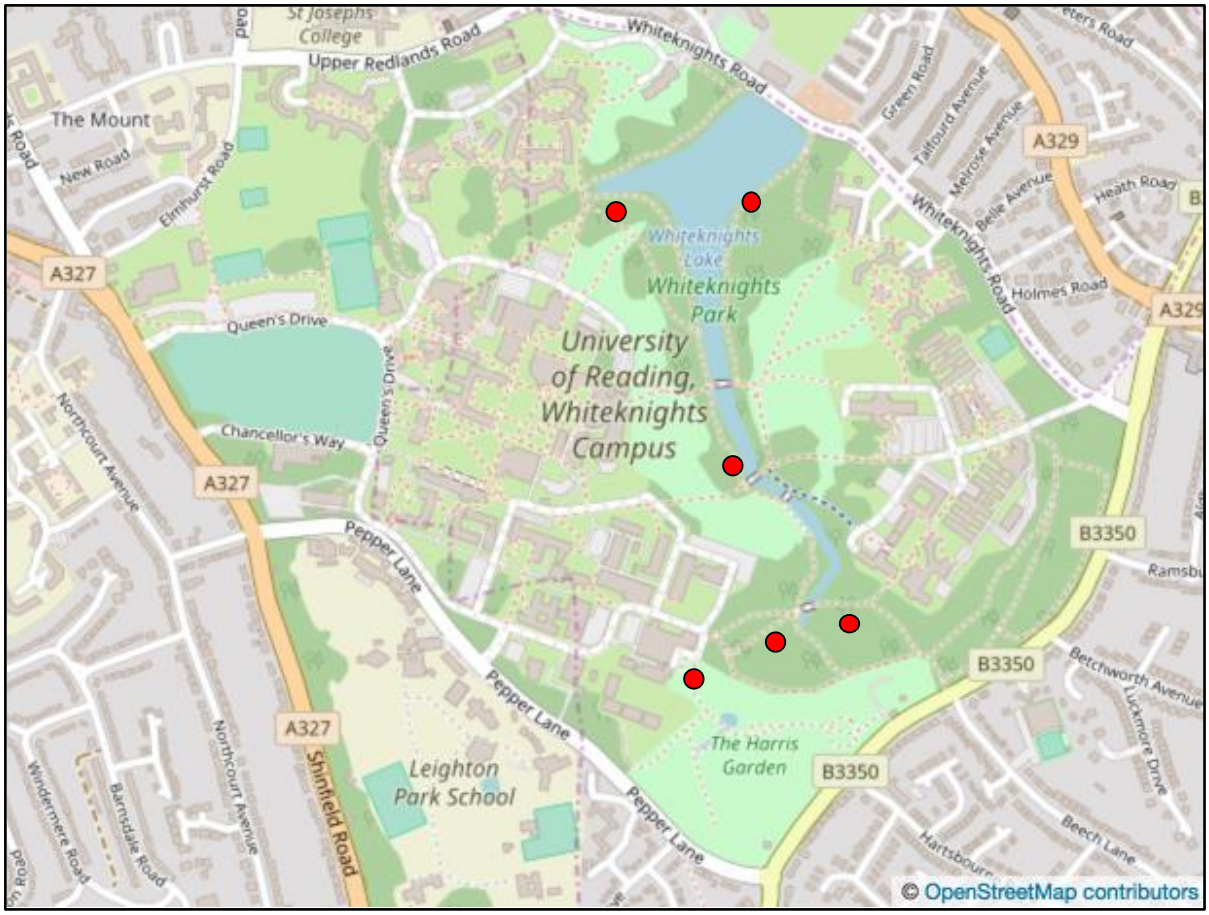

Figure 2.2. Map of the University of Reading Campus with bacteriophage sampling locations marked (red) © OpenStreetMap contributors.

The environmental samples were suspended in Nanopure water to a total volume of 50 ml in a 50 ml falcon tube before being centrifuged for 10 min at 1,500 *g*. The supernatant was then decanted into a new 50 ml falcon tube before being passed through a 0.45 µm pore-size syringe injection filter (Millipore). The filtered samples were then pooled to a total volume of 1 ml which was then added to 5 ml KB cultures of *P. agarici* NCPPB 2472 and *P. tolaasii* NCPPB 2192T then incubated overnight at 28°C. The overnight cultures were then centrifuged at 1,500 *g* for 30 min to pellet the bacterial cells and the supernatant was either used immediately for phage isolation or stored at 4°C.

2.1.4 Phage isolation

100 µl of the supernatant from the environmental sample preparations outlined in Section 2.1.3, were serially diluted eightfold in PBS. 100 µl of each serial dilution was then plated on KB agar plates and overlaid with 0.7% agar containing either *P. agarici* NCPPB 2472 or *P. tolaasii* NCPPB 2192T. The plates were then incubated overnight at 28°C and observed for the presence of visible plaques.

If visible plaques were present, a new overnight culture of either *P. tolaasii* or *P. agarici* was inoculated with a stab of an isolated phage plaque before being incubated overnight at 28°C. The overnight culture was then spun at 1,500 *g* for 30 minutes to pellet cells and the supernatant filtered with a 0.45 µm pore-size syringe injection filter and 100 µl of filtered supernatant was eightfold serially diluted and each dilution plated of KB agar and overlaid with 0.7% agarose containing either *P. agarici* or *P. tolaasii* before incubation at 28°C overnight. The plates were then checked for uniform plaque morphology and if there was variable plaque morphology the process was repeated with each distinct plaque morphology until a uniform morphology was attained for each inoculation.

2.1.5 Plaque assay protocols

Plaque Assay

Standard plaque assay procedure involved the addition of 1 ml of diluted phage sample in PBS or phage stock to 5 ml of 0.7% agar cooled to approximately 37°C in a sterile 50 ml falcon tube, this was then mixed by gently inverting the tube several times. The phage/agar mix was then poured onto a pre-prepared hard KB or LB agar plate and allowed to solidify at room temperature in a laminar flow cabinet before being inverted and incubated at 28°C overnight.

Droplet Assay

Droplet assays procedure involved preparing a 10ml KB culture of either *P. tolaasii* or *P. agarici* which was then incubated overnight at 28°C with 220 rpm shaking. 5 ml of the overnight culture was then combined with 45 ml of molten overlay agar in a 50 ml falcon tube and mixed by inverting the tube several times. 5 ml of the culture/overlay agar mix was poured on to a pre-prepared hard KB or LB agar plate, which was allowed to solidify in a laminar flow cabinet at room temperature. A phage sample or stock was then passed through an eightfold serial dilution in sterile PBS, with mixing by vortexing between each dilution stage. 10 µl of each dilution was then pipetted on to the surface of the solidified plates and returned to the laminar flow cabinet to allow the droplets to dry, followed by inversion and overnight incubation at 28°C.

2.1.6 Phage stocks

Low titre phage stock

Standard, low titre ($<1x10^{10}$ pfu/ml) phage stocks were created by addition of 1ml of previous phage stock to 1ml of overnight bacterial culture in 10 ml of KB broth, before incubation overnight at 28°C with 200 rpm shaking. The bacterial cells were then pelleted out by centrifugation at 1,500 *g* for 30 minutes, and the phage containing supernatant clarified by filtration through a 0.45 µm pore-size syringe filter. The filtered phage supernatant was then stored at 4°C.

High titre phage stocks

The standard phage stock was serially diluted eightfold and each dilution used to create 3 repeat plaque assay plates and the plates then incubated for 24 hours at 28°C. After incubation 5 ml of PBS was added to the surface of the dilution plates in triplicate with almost confluent lysis

and the plates left for 90 minutes at room temperature with agitation every 15 minutes. Following this the PBS from all 3 plates was combined together and filtered through a 0.45 µm pore-size syringe filter.

2.1.7 Phage host range assay

Phage host range assays were performed using the plaque assay protocol and 1ml of undiluted low titre phage stock on *Pseudomonas* species shown in Table 2.2.

Table 2.2. Bacterial strains used in phage host range assay. (R.W.J.: Gift from Professor Robert Jackson)

Strain:	**Source**
P. tolaasii NCPPB 2192T	R.W.J.
P. agarici NCPPB 2472	R.W.J.
P. sp NS1	*Agaricus bisporus* mushroom surface. This study
P. syringae pv. morsprunorum R1 5244	R.W.J.
P. syringae pv. syringae 9097	R.W.J.
P. syringae pv. morsprunorum R2 5255	R.W.J.
P. fluorescens Pfo-1	R.W.J.
P. aeruginosa 14207	R.W.J.

2.1.8 Chloroform sensitivity assay

A 100 µl sample of phage culture of known concentration ($1x10^8$ pfu/ml) was incubated with 5 µl of chloroform. The phage culture containing chloroform was then eightfold serially diluted with PBS and 10 µl drops of each dilution was spotted on to a pre-prepared soft agar overlay plate of *P. agarici* or *P. tolaasii* and incubated for 24 hours at 28°C and observed for plaque formation.

2.1.9 Single step growth curve

A 10 ml culture of *P. agarici* in KB was incubated overnight at 28°C with shaking at 220 rpm. 1 ml of the stationary overnight culture was taken and transferred to a fresh 10 ml of KB and incubated at 28°C with shaking for 2 hours. The OD 600 was measured and the culture diluted

to a concentration of $1x10^8$ cfu/ml (~0.028 Abs). 2.5 ml of *P. agarici* at $1x10^8$ cfu/ml was then combined with 25 µl of ϕNV3 stock ($1x10^9$ pfu/ml) to give a MOI of 0.1.

The mixture of *P. agarici* and ϕNV3 was then mixed by vortexing and incubated for 10 minutes on ice to allow for adequate phage adsorption, before being added to 247.5 ml of KB for a final volume of 250 ml to sufficiently dilute the culture and prevent further adsorption events. The culture was then transferred to a water bath at 28°C with frequent shaking. A 1 ml sample was transferred to a 1.5 ml Eppendorf every 1.5 minutes for 15 minutes then every 5 minutes for a further 75 minutes. These 1 ml samples were centrifuged at 10,000 *g* for 2 minutes to pellet any *P. agarici* cells that may otherwise form infective centres and 100 µl of the phage-containing supernatant transferred to an Eppendorf containing 900 µl PBS, on ice. The samples were then passed down a threefold serial dilution in PBS and placed back on ice. 10 µl droplets from each time point were placed on pre-made soft agar plates of *P. agarici* and repeated in duplicate and allowed to dry at room temperature before incubation for 24 hours at 28°C to allow plaques to form.

2.1.10 Adsorption assay

A 10 ml inoculation of a single plate colony of *P. agarici* in KB was incubated overnight at 28°C with shaking at 220 rpm, 1 ml of the stationary phase overnight culture was taken and transferred to a fresh 10 ml of KB and incubated at 28°C with shaking until cells reached mid log-phase (~0.4 Abs). Dilution tubes were prepared containing 1 ml of chloroform saturated PBS, which were placed on ice. The mid-log phase culture was then used to create a 50 ml culture in pre-warmed KB at 28°C containing $1x10^8$ cfu/ml as determined by OD 600 (0.028 Abs) to which $5x10^8$ pfu of phage ϕNV3 was added and mixed by vortexing to give a final MOI of 0.1. The culture was then incubated at 28°C with regular mixing and at time intervals of 1 minute for 10 minutes, 1 ml of culture was taken from the ϕNV3/*P.agarici* culture and added to the chloroform saturated PBS.

The chloroform saturated samples were then fivefold serially diluted and 10 µl of appropriate dilutions of each time point were spotted in triplicate on to a pre-prepared *P. agarici* containing soft agar overlay plate and incubated overnight at 28°C.

2.1.11 Kill curve assay

Kill curves were performed using either *P. tolaasii* 2192T or *P. agarici* NCPPB 2472 taken from stationary phase overnight cultures grown in 10 ml KB broth at 28°C. 1 ml of overnight culture was taken and added to a fresh tube of 10 ml KB broth and incubated at 28°C. After 2 hours OD 600 was measured, and culture diluted to the appropriate starting OD 600 of 0.092 Abs for *P. tolaasii* or 0.028 Abs for *P. agarici*, that is equal to approximately $1x10^8$ cfu/ml.

Phage stocks were then diluted to the appropriate starting concentrations to produce MOI's of 0.1, 0.01 and 0.001 ($1x10^8$, $1x10^7$, $1x10^6$ pfu/ml respectively) and 100 µl of these phage dilutions added to 1 ml of the $1x10^8$ cfu/ml bacterial solutions in a 1.5 ml Eppendorf tube and vortexed to mix. These mixes were left to stand at room temperature for five minutes to allow adequate diffusion of the phage particles throughout the solution and then 100 µl aliquots were transferred to 10 individual wells in a 96 well Cellstar® Cell Culture Plate with one row per MOI. The plates were then transferred to a Tecan GENios microplate reader at 28°C, and OD 595 readings taken ever 5 minutes over a period of 12 hours. Data was visualised using GraphPad Prism 5, Version 5.0b.

2.1.12 Phage/host co-evolution assay

The phage/host co-evolution assays in this study were performed using a method adapted from Betts *et al.* [246]. The initial phage/host incubation was performed in 6 ml of KB in a 30 ml universal tube with 5 µl of ancestral ϕNV3 phage (~$5x10^5$ pfu) with 10 µl of ancestral *P. agarici* (~$2x10^6$ cfu), which was incubated at 27°C for 24 hours. From this overnight phage/bacteria culture 100 µl was taken and transferred to a new universal containing 6 ml of

fresh KB and incubated for 24 hours at 27°C. This process was repeated for a total of 12 sequential transfers. For each of the transfers 1 ml of bacteria/phage mixture was removed and stored in 40% glycerol at -80°C; phage was isolated from each transfer by adding 100 µl of chloroform to 900 µl of bacteria/phage culture, mixed by vortexing for several seconds followed by centrifugation for 4 minutes at 10,000 *g* before the supernatant of purified phage was removed and stored at -80°C.

Phage infection capacity for bacterial transfers T1, T3, T5, T7 and T9 was measured by streaking of 16 arbitrarily sampled bacterial colonies from streak plates of each transfer that were incubated at 27°C for 72 hours on KB agar. Each bacterial transfer was tested against the ancestral, present and future phage transfers (T-1, T0, T+1). To each of the 16 arbitrarily sampled bacterial streaks 5 µl of purified phage was placed in the centre of each line. This process was repeated 5 times per phage time point (ancestral, present and future). After incubation for 72 hours at 28°C a bacterial colony was considered resistant if its growth continued uninterrupted by the addition of phage.

2.1.13 Electron microscopy

Copper coated carbon-formvar TEM grids were prepared via floating on 10 µl drops of filtered phage lysate solutions of roughly 10^6-10^8 pfu/ml for 10 minutes. Excess liquid was removed by touching the side of the grid to filter paper. 10 µl of 1.5% (w/v) uranyl acetate stain was then spotted on to the grid for 20-30 seconds, excess stain was removed by touching the side of the grid to filter paper and the grids were left to dry at room temperature in a fume hood before being stored at room temperature in a grid box. Grids were then viewed under a Phillips CM200 Transmission Electron Microscope (TEM) operated at 80 kV.

2.2 Growth and microbiological media

P. agarici and *P. tolaasii* were routinely cultured in both Kings B (KB) broth (Proteose peptone #3 20 gL^{-1}; K_2HPO_4 1.5 gL^{-1}; $MgSO_4$ $7H_2O$ 1.5 gL^{-1}; Glycerol 0.01% (v/v)) and Lysogeny Broth (LB) (Tryptone 10 gL^{-1}; Yeast Extract 5 gL^{-1}; NaCl 10 gL^{-1}). For plating, bacteriological agar was added to broth media at a concentration of 15 gL^{-1}.

2.3 Antibiotics

All stock antibiotics used were dissolved in nanopure water before filter sterilisation through a 0.22 µm filter and storage at -20°C. Working concentrations for ampicillin was 100 $\mu g.ml^{-1}$ and kanamycin was 50 $\mu g.ml^{-1}$.

2.4 Molecular biological techniques

2.4.1 Phage DNA extraction and purification

Phage genomic DNA was isolated and purified from 1 ml of 0.45 µm filtered phage lysate containing between $1x10^9$ pfu/ml-$1x10^{10}$ pfu/ml using a Norgen Biotek Phage DNA Isolation Kit using the standard manufacturers protocol, supplemented with a 30 minute incubation with 4 µl of Proteinase K (20 mg/ml) at 50°C. DNA concentration was measured using a NanoDrop™ 2000, using standard manufacturers protocol (ThermoFisher Scientific).

2.4.2 Bacterial genomic DNA extraction and purification

Bacterial genomic DNA was isolated and purified with a Sigma-Aldrich GenElute™ Bacterial Genomic DNA Kit from 1.5 ml aliquots of 10 ml cultures of either *P. agarici*, *P. tolaasii* or *Pseudomonas sp. NS1* grown overnight in KB at 28°C with shaking and eluted in nanopure water. DNA concentrations were quantified by NanoDrop™ 2000 (ThermoFisher Scientific).

2.4.3 PCR protocols

Polymerase chain reaction (PCR) was used to amplify specific target regions of DNA for cloning and sequencing. Primers were designed using ApE v2.0.47 (http://biologylabs.utah.edu/jorgensen/wayned/ape/) and manufactured by Integrated DNA Technologies, all primers were diluted with nuclease-free water to a stock concentration of 100 pmol/µl. Standard PCR was performed with GoTaq® Green Master Mix manufactured by Promega® in 20 µl total reaction volumes (1 µl forward primer, 1 µl reverse primer, 3 µl template DNA, 7 µl nuclease-free water, 10 µl GoTaq® Green Master Mix); reaction conditions were as follows: Denaturation at 95°C for 2 min; Cycle 30x at 95°C for 30 seconds, 58°C for 45 seconds, 72°C for 1 min/kb of predicted product; Final elongation at 72°C for 5 min.

Colony PCR protocol reaction conditions were as per standard PCR protocol, with extended initial denaturation (5 min) and 3 µl of liquid culture/stock or 3 µl nanopure water with colony scraping to replace template DNA.

Self-priming primer PCR protocol was as per standard PCR, with 2 µl of a single primer and 8 µl of nuclease-free water.

2.4.4 Agarose gel electrophoresis

PCR products were separated by agarose gel electrophoresis. Agarose powder was dissolved to the required concentration (1.5% (w/v) for PCR products <1000 bp, 1% (w/v) for PCR products >1000 bp or unknown size and 0.7% (w/v) for agarose gel extraction) in 0.5X Tris/Borate/EDTA (TBE) buffer (5X TBE stock contains: Tris base 54 gL^{-1}; Boric acid 27.5 gL^{-1}; 0.02% (v/v) EDTA). 1 µl of 10,000X Biotium GelRed® was added to 100 ml agarose/TBE solution for nucleic acid staining. 1 µl of sample DNA was mixed with 1 µl of

5X Bioline Sample Loading Buffer and 3 µl of nanopure water. The sample mix and ladder (Bioline HyperLadder™ 1) were loaded into the pre-cast gel and run at 80-120V for the desired time (45-90 min). Completed gels were imaged using a Syngene G:BOX Chemi.

2.4.5 DNA gel recovery

DNA was extracted from 0.7% agarose gels using a Qiagen QIAquick Gel Extraction Kit as per manufacturers protocol. Final elution was performed using 50 µl of Qiagen EB buffer.

2.4.6 Genomic DNA sequencing

De novo paired-end sequencing was performed by Source Bioscience on a single lane of an Illumina MiSeq® with a read length of 50 bp. Whole genome sequencing libraries were generated by Source Bioscience using an Illumina TruSeq® DNA Nano kit. The forward adapter sequence was: AGATCGGAAGAGCACACGTCTGAACTCCAGTCA and the reverse adapter sequence was: AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT. Returned data was adapter and quality trimmed using Skewer v0.1 [247] and trimmed reads were assembled into contigs with SPAdes v3.6.2 [248].

The phage ΦNV3 genomic sequence was confirmed with extensive primer walking (Primer sequences available in Appendix 2, Tables Ap2.3 and Ap2.4) combined with Sanger Sequencing performed by Source Bioscience.

2.4.7 *P. agarici* contig stitching

Mapped contig junctions were assigned a numerical reaction number and primers were designed to allow for PCRs to amplify the unknown junction region as shown in Figure 2.3, a full list of primers used is available in Appendix 2 Tables Ap2.1 and Ap2.2.

Figure 2.3. Diagram of the primer design process for contig junction stitching.

For contig stitching reactions related to predictions of contig order and orientation produced by alignment of *P. agarici* NCPPB 2472 sequenced contigs to *P. fluorescens* A506 outlined in Section 2.5.1 see Table 2.6.

Table 2.6. Reaction list for contig junctions generated from mapping *P. agarici* NCPPB 2472 sequenced contigs with *P. fluorescens* A506 (GCA_000262325.2).

Reaction Number:	Contig 1:	Contig 2:	Primer 1:	Primer 2:
1	C3	C21	Seq 6	Seq 33
2	C21	C38	Seq 34	Seq 66
3	C38	C26	Seq 65	Seq 42
4	C26	C35	Seq 41	Seq 59
5	C35	C16	Seq 60	Seq 23
6	C16	C17	Seq 24	Seq 25
7	C17	C59	Seq 26	Seq 98
8	C59	C14	Seq 97	Seq 19
9	C14	C10	Seq 20	Seq 11
10	C10	C1	Seq 12	Seq 2
11	C1	C33	Seq 1	Seq 56
12	C33	C27	Seq 55	Seq 43
13	C27	C34	Seq 44	Seq 57
14	C34	C28	Seq 58	Seq 45
15	C28	C2	Seq 46	Seq 4
16	C2	C39	Seq 3	Seq 67
17	C39	C11	Seq 68	Seq 13
18	C11	C9	Seq 14	Seq 9
19	C9	C42	Seq 10	Seq 74
20	C42	C18	Seq 73	Seq 27
21	C18	C32	Seq 28	Seq 53
22	C32	C29	Seq 54	Seq 47
23	C29	C6	Seq 48	Seq 7
24	C6	C30	Seq 8	Seq 49
25	C30	C24	Seq 50	Seq 40
26	C24	C15	Seq 39	Seq 22
27	C15	C13	Seq 21	Seq 17
28	C13	C46	Seq 18	Seq 81
29	C46	C12	Seq 82	Seq 16
30	C12	C23	Seq 15	Seq 38

For contig stitching reactions related to predictions of contig order and orientation produced by alignment of *P. agarici* sequenced contigs to *P. agarici* NCPPB 2289, outlined in Section 2.5.1, see Table 2.7.

Table 2.7. Reaction list for contig junctions generated from mapping *P. agarici* NCPPB 2472 sequenced contigs with *P. agarici* NCPPB 2289.

Reaction Number:	**Contig 1:**	**Contig 2:**	**Primer 1:**	**Primer 2:**
Rb1	C46	C73	Seq 82	Seq 108
Rb2	C73	C44	Seq 107	Seq 77
Rb3	C44	C12	Seq 78	Seq 15
Rb4	C12	C49	Seq 16	Seq 87
Rb5	C49	C58	Seq 88	Seq 95
Rb6	C58	C13	Seq 96	Seq 18
Rb7	C13	C57	Rseq 17	Seq 93
Rb8	C57	C53	Seq 94	Seq 113
Rb9	C53	C52	Seq 114	Seq 91
Rb10	C52	C62	Seq 92	Seq 102
Rb11	C62	C23	Seq 101	Seq 38
Rb12	C23	C35	Seq 37	Seq 59
Rb13	C35	C22	Seq 60	Seq 35
Rb14	C22	C37	Seq 36	Seq 63
Rb15	C37	C1	Seq 64	Seq 1
Rb16	C1	C65	Seq 2	Seq 104
Rb17	C65	C19	Seq 103	Rseq 30
Rb18	C19	C15	Rseq 29	Rseq 21
Rb19	C15	C24	Seq 22	Seq 39
Rb20	C24	C10	Rseq 40	Seq 11
Rb21	C10	C28	Seq 12	Rseq 45
Rb22	C28	C11	Rseq 46	Seq 14
Rb23	C11	C42	Seq 13	Rseq 74
Rb24	C42	C26	Seq 73	Rseq 42
Rb25	C26	C2	Seq 41	Rseq 3
Rb26	C2	C16	Rseq 4	Rseq 23
Rb27	C16	C27	Rseq 24	Rseq 43
Rb28	C27	C3	Seq 44	Rseq 6
Rb29	C3	C33	Seq 5	Rseq 55
Rb30	C33	C21	Rseq 56	Seq 34
Rb31	C21	C18	Rseq 33	Rseq 28
Rb32	C18	C38	Seq 27	Rseq 65

Rb33	C38	C14	Seq 66	Seq 20
Rb34	C14	C29	Seq 19	Seq 47
Rb35	C29	C6	Seq 48	Seq 7
Rb36	C6	C32	Seq 8	Rseq 53
Rb37	C32	C20	Seq 54	Seq 31
Rb38	C20	C34	Seq 32	Seq 57
Rb39	C34	C61	Rseq 58	Seq 99
Rb40	C61	C39	Seq 100	Seq 68
Rb41	C39	C59	Seq 67	Rseq 97
Rb42	C59	C9	Seq 98	Seq 9
Rb43	C9	C47	Rseq 10	Seq 84
Rb44	C47	C17	Seq 83	Seq 26
Rb45	C17	C30	Rseq25	Rseq50
Rb46	C30	C31	Seq 49	Seq 52

All designed primers were first run using the Self Priming Primer PCR protocol to check for multiple bands and self-priming due to inverted repeat regions in the unknown inter-contig sequence, if signs of self-priming were evident the primers were re-designed and moved ~500 bp back from their previous position and re-run using the Self Priming Primer PCR until a single band was achieved.

PCRs were then run on a 1% agarose gel, and those showing a single band the reactions were then repeated 2x in 160 µl volumes before being gel extracted and purified as per the described protocol in 2.4.6. The DNA concentrations were then analysed using a NanoDrop™ 2000, any samples below the required concentration for Sanger sequencing were then concentrated via rotary evaporator at 40°C for 90 min.

The samples were then sent for Sanger sequencing by Source Bioscience and the returned reads were analysed, aligned and assembled into a single sequence with ApE v2.0.47.

2.4.8 Vectors, cloning and plasmid construction

The DNA vectors and plasmids used for cloning phage and bacterial genes for expression and knockout creation are listed below in Table 2.8.

Table 2.8. Vectors and plasmids used in this study. (R.W.J: Gift from Professor Robert Jackson, UoR)

Plasmid	Features	Source
pEXP5-CT/TOPO®	T7 promoter, T7 forward binding site, Ribosome binding site, TOPO® Cloning site, C-terminal polyhistidine (6xHis) tag, T7 transcription terminator, T7 reverse priming site, *bla* promoter, Ampicillin resistance gene (ß-lactamase), pUC origin of replication (*ori*)	Invitrogen
pCR™2.1- TOPO®	LacZα fragment, M13 reverse priming site, Multiple cloning site, T7 promoter/priming site, M13 Forward (-20) priming site, f1 origin, Kanamycin resistance ORF, Ampicillin resistance ORF, pUC origin	Invitrogen
pBBR1MCS-2	pBBR1 oriV, pBBR1 Rep, LacZα promoter, NeoR/KanR, M13 Forward priming site, M13 Reverse priming site, Multiple cloning site.	R.W.J.

Endolysin constructs

To allow for expression of the NV1 cell lysis proteins, two constructs were created using the primers in Table 2.9 that contain the native stop codon and the vector pEXP5-CT/TOPO® (Table 2.8). PCR amplification of the desired genes was performed using phage colony PCR and the product checked by agarose gel electrophoresis. The band corresponding to the desired product was extracted using the DNA gel recovery method and sent for Sanger Sequencing by Source Bioscience using the stock T7-promoter forward and reverse primers. After the returned sequence was confirmed to be the desired product by alignment in ApE v2.0.47, the products were TOPO® cloned into the pEXP5-CT/TOPO® vector as per the manufacturer's protocol using 4 µl of DNA. Following the cloning reaction, 2 µl of the products were transformed in to One Shot® TOP10 Chemically Competent *E. coli*, as per the manufacturer's protocol, and

plated on to selective LB agar plates containing ampicillin (100 µg/ml), then incubated at 37°C overnight.

Colonies that grew on the overnight plates were checked for the presence of the respective NV1 holin and lysozyme genes via colony PCR with a negative control containing water instead of bacterial cells and a positive control of NV1 DNA, with the products visualised via agarose gel electrophoresis.

The colonies that were determined to contain the required inserts were then used to inoculate universals containing 10ml of ampicillin (100 µg/ml) infused LB broth and incubated for 24 hours at 37°C. The constructs were then purified form the cultures using a QIAprep Spin Miniprep kit and the concentrations of the returned DNA analysed by Nanodrop™ 2000 (ThermoFisher Scientific). The sequence of the products was then confirmed by Sanger-Sequencing (Source Bioscience) with stock T7 forward primers supplied by Source Bioscience.

Table 2.9. Primers used in phage NV1 lysis protein constructs.

Name:	**Length:**	**Tm:**	**Sequence:**
NV1HolinF	21	54	ATGGTATTCTTTGCTGCGTAC
NV1HolinR	21	54	TCATTTGGTTGCTCTCCATTC
NV1LysoF	19	55	ATGTCGCGGATCTCACTAC
NV1LysoR	26	54	TTATCCTTTATAAAGGTCGTACAGTG

The method for producing the constructs for the ϕNV3 lysis proteins was the same as for the NV1 constructs, however instead using the primers in Table 2.10, 5 separate constructs were created. Each of the lysis proteins was cloned separately in to pEXP5-CT/TOPO® (pinholin, lysozyme, Rzl-like protein), as well as one construct containing both the pinholin and lysozyme from a DNA fragment produced from primer pair Nv3aPin-F and 3aLys_ex_R, and one

construct containing all the lysis proteins from a DNA fragment produced from the primer pair Nv3aPin-F and 3aRz1_ex_R.

Table 2.10. Primers used in ϕNV3 lysis protein constructs.

Name:	Length:	Tm:	Sequence:
Nv3aPin-F	22	61	ATGCAACTAGACACCACGAGCG
Nv1Lyso_ex_F	24	61	ATGTCGCGGATCTCACTACTGAGG
Nv3aRzl-F	22	62	ATGCGCTACGCCATCATTGCTG
3aLys_ex_R	23	60	ATCCAGCAATAGTGGCGAGTACG
3aPin_ex_R	22	61	TTCCTTAGGTTCATCCGCAGCC
3aRz1_ex_R	26	61	ATCCTCCAGTGGGTATCCATTAAGGG

A full table of all lysis protein constructs used in this study are outlined in Table 2.11.

Table 2.11: Table of all lysis protein constructs used in this study.

Name:	Contains:
pEXP5-NV1Hol	Phage NV1 Holin
pEXP5-NV1Lys	Phage NV1 Lysozyme
pEXP5-NV3Pin	Phage ϕNV3 Pinholin
pEXP5-NV3Lys	Phage ϕNV3 Endolysin
pEXP5-NV3Rzl	Phage ϕNV3 Rzl/Rzl-like protein
PEXP5-NV3PinLys	Phage ϕNV3 Pinholin and Endolysin
pEXP5-NV3Full	Phage ϕNV3 Pinholin, Endolysin and Rzl/Rzl-like protein.

2.4.9 Protein expression

Expression of the phage lysis proteins was performed by first transforming the 5-10 ng of lysis protein constructs, outlined in Table 2.8, in a volume of 1-5 µl into BL21-AI™ One Shot® Chemically Competent *E. coli* using the manufacturers basic protocol using 5-10 ng of DNA in a volume of 1-5 µl; all growth media was supplemented with 0.1% (w/v) glucose due to the toxic nature of the proteins encoded by the constructs. Following incubation of the transformation plates at 37°C overnight, colonies were picked from the plates and cultured overnight at 37°C in 10 ml LB supplemented with 0.1% glucose and ampicillin. Each overnight

culture was aliquoted in to 10, 1 ml volumes in 2 ml Eppendorf tubes and supplemented with 300 µl of 40% (v/v) glycerol solution and immediately frozen at -80°C. One of each aliquot was analysed by PCR using the appropriate primers from Table 2.6 and Table 2.7 corresponding to the relevant lysis proteins, including a negative control containing water and a positive control of the corresponding phage DNA. Bands corresponding to the desired product were excised from the gels using the gel extraction protocol and sent for Sanger Sequencing by Source Bioscience with the appropriate primers from Table 2.6 and Table 2.7 corresponding to the relevant lysis proteins. Any aliquots that did not contain the required phage sequence were discarded.

For the protein expression assay, 1 ml of the required -80°C aliquots of BL21-AI cells containing phage lysis protein sequences was pipetted in to 10 ml of 0.1% glucose supplemented ampicillin infused LB broth and incubated at 37°C with 220 rpm shaking until the OD 600 reached approximately 0.4 Abs. 100 µl of each culture was then pipetted into 12 wells of a Cellstar® 96 well cell culture plate. To 6 of each of the culture containing wells, 2 µl of 10% L-Arabinose solution was added to induce protein expression and the plate immediately transferred to a Tecan GENios microplate reader pre-warmed to 37°C and maintained at 37°C while OD 595 readings were taken every 5 minutes for 150 cycles with rotary shaking between readings.

2.4.10 *P. agarici* knockout vector construction and competent cell preparation

The vector used in the creation of the knockout construct was Invitrogen™ pCR2.1™-TOPO® TA, containing both ampicillin and kanamycin resistance markers. The gene fragment used was from the *P. agarici* NCPPB 2472 Type IV pilus assembly ATPase PilB gene (Locus Tag:

AWM79_RS16250), amplified by PCR using the primers P.ag-IVPiliF and P.ag-IVPili-R to create an insert of 840 bp in length with an 'A' overhang for TOPO® cloning.

A map of the completed construct is shown below (Figure 2.4).

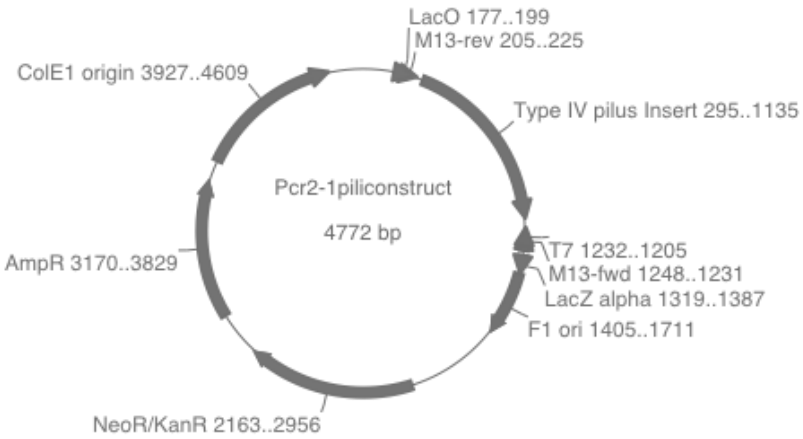

Figure 2.4. Map of the pCR2.1/TypeIV pilus construct indicating the location of the AmpR and KanR resistance markers. Generated using ApE v2.0.47.

The insert was then purified by gel extraction and imaged on an agarose gel, the concentration of the insert was determined to be 43.2 ng/µl via NanoDrop™ 2000 (ThermoFisher Scientific). The insert was then cloned in to the pCR2.1™-TOPO® TA vector using the method outlined in the manufacturers recommendations as follows: 4 µl of insert and 1 µl of pCR2.1™-TOPO® TA vector were combined in a TOPO® cloning reaction with 1 µl of salt solution (1.2 M NaCl; 0.06 M $MgCl_2$), followed by transformation of 2 µl of the TOPO® cloning reaction into chemically competent One Shot® TOP10 and plating on to selective agar containing ampicillin. After overnight incubation at 37°C, 6 colonies were selected and analysed by colony PCR for the presence of the gene insert, these reactions were separated on a 1% agarose gel by gel electrophoresis and imaged using a Syngene G:BOX Chemi. The colonies that were demonstrated to contain the insert were then inoculated into 10ml LB supplemented with ampicillin and incubated overnight at 37°C, 220 rpm. Following overnight incubation, plasmid

DNA was purified from all cultures and sent for Sanger sequencing (Souce Bioscience) using T7 forward and reverse sequencing primers, to analyse the quality of the DNA insert.

The next step was to create competent *P. agarici* NCPPB 2472 cells for transformation of the knockout plasmid. A positive control plasmid, pBBR1MCS-2, containing a kanamycin resistance marker (Figure 2.5), was used to determine the competency of the *P. agarici* cells.

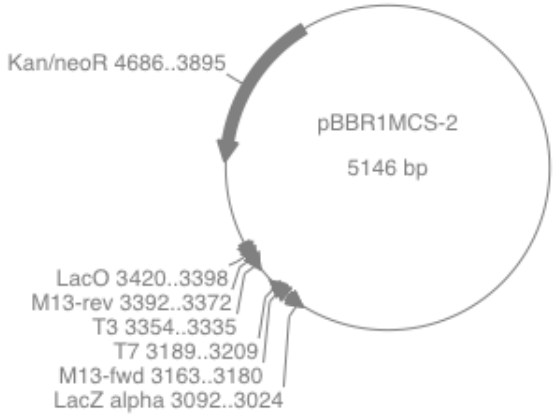

Figure 2.5. Map of the control plasmid pBBR1MCS-2 indicating the location of the Kan/neoR resistance marker. Generated using ApE v2.0.47.

The method for preparing electrocompetent *Pseudomonas* cells was adapted from the method described by Choi *et al.* [249]. Six millilitres of overnight *P. agarici* culture grown in LB medium was split into three 2ml microcentrifuge tube and were centrifuged at 10,000 *g* for 4 min at room temperature. The cell pellet was then washed with 2 ml room temperature 300mM sucrose, then centrifuged at 10,000 *g* for 4 min. The wash was step repeated and the excess sucrose solution was removed by pipette following centrifugation and the resultant cell pellet was resuspended in 100 µl 300mM sucrose solution.

For electroporation, 2 µl of 246 ng/µl pCR2.1™-TOPO® Pili construct or 2.7 µl of 183ng/µl pBBR1MCS-2 positive control vector was used. DNA was mixed with 100 µl of fresh electrocompetent cells was transferred to the electroporation cuvette. The cells were pulsed (settings: 25µF; 200Ω, 2.5kV) before 1 ml of room temperature LB medium was added and

the tube was transferred to a 28°C incubator for 2 hours. The cells were then centrifuged at 10,000 *g* for 4 minutes and the supernatant was discarded. The cells were then resuspended in 100 µl LB medium and plated on to a LB agar containing kanamycin and incubated at 28°C for 48 hours. A negative control of competent *P. agarici* was also performed, without electroporation.

The protocol for preparation of chemically competent *Pseudomonas* cells was adapted from Chuanchuen *et al.* [250]. 10 ml *P. agarici* cells were grown overnight in LB medium until they reached saturation. A 2 ml microcentrifuge was pre-chilled on ice for ~20min, then 2 ml of overnight culture was transferred to the tube followed by 1 minute centrifugation at 10,000 *g*. The cell pellet was resuspended in 2 ml 0.1M $MgCl_2$ at 4°C. The cell suspension was then centrifuged again at 10,000 *g* for 1 minute. The supernatant was removed, and the cell pellets were resuspended in 1 ml transformation salts with glycerol solution (75mM $CaCl_2$, 6mM $MgCl_2$, 15% glycerol) at 4°C. The cell suspension was then incubated on ice for 10 min then centrifuged at 10,000 *g* for 1 min. After decanting the supernatant, the cell pellets were then resuspended in 200 µl transformation salts solution at 4°C with glycerol and flash frozen in liquid nitrogen before being transferred to -80°C freezer for storage.

For transformation, 2 µl of 246 ng/µl PCR2.1-TOPO Pili construct or 2.7 µl of 183 ng/µl pBBR1MCS-2 was used as a positive control. The DNA and competent cells were combined in a 2 ml microcentrifuge tube, a negative control using nanopure water was also performed. The cell/DNA suspension was then incubated on ice for 15 min, followed by a 2 min heat-shock at 37°C, after which 500 µl LB broth was immediately added to the tubes. The tubes were then incubated at 28°C for two hours. After incubation, 200 µl was taken of each transformation and plated on to LB agar containing kanamycin.

2.4.11 RNA Extraction, purification and sequencing

P. agarici NCPPB 2472 was inoculated in triplicate in 10 ml KB and incubated at 28°C overnight, then subcultured 1:10 in fresh KB for ~2 hours until cells reached mid log phase (0.4 Abs at 600 nm). Each culture was diluted to 0.28 Abs (approximately $1x10^9$ cfu) and two 1 ml aliquots of cells per culture were transferred to 2 ml reaction tubes. To 3 of the 6 reaction tubes 100 µl of $1x10^{10}$ pfu/ml ϕNV3a in PBS was added, to the remaining 3 reaction tubes 100 µl of PBS was added. The reaction tubes were incubated at room temperature for 5 mins before being transferred to a 28°C shaking incubator at 220 rpm for 40 min. Following incubation, the tubes were spun at 10,000 *g* for 5 minutes and the resultant cell pellet was resuspended in 2 ml RNA*later*® before being stored at 4°C.

RNA was extracted from the samples stored in RNA*later*® using an Ambion® PureLink® RNA Mini Kit. Homogenisation of the bacterial cells was achieved using MP Biomedical Lysing Matrix B 2 ml tubes in a MP Biomedical FastPrep®-24 benchtop homogeniser at maximum speed for 45 seconds and immediately placed on ice. RNA was extracted using a PureLink™ RNA Mini Kit. DNA was removed from the samples by on-column DNase treatment using the PureLink™ DNase Set. The extracted RNA was concentrated using a Thermo Scientific GeneJET RNA Cleanup and Concentration Micro Kit and stored at -80°C.

Initial RNA concentration was determined by NanoDrop™ 2000 (ThermoFisher Scientific), RNA integrity and concentration were analysed using an Agilent RNA 6000 Nano Kit in an Agilent 2100 Bioanalyzer.

Sequencing was performed by Source Bioscience using a NextSeq® 500 v2 using one Mid-output flow cell at 75 bp paired-end reads, library preparation was performed using a TruSeq™ Stranded mRNA kit with Ribo-Zero™.

2.5 Bioinformatic techniques

2.5.1 Contig mapping and scaffold construction

P. agarici NCPPB 2472

Mapping of *P. agarici* contigs provided by Source Bioscience was performed using both *P. fluorescens* A506 (GCA_000262325.2) and *P. agarici* 2289 (GCA_000280785.1) as a reference sequence using CONTIGuator 2.7.4 [251].

P. tolaasii NCPPB 2192T

Mapping of the *P. tolaasii* contigs provided by Source Bioscience was performed using *P. tolaasii* PMS117 as a reference genome, CONTIGuator 2.7.4 [251] was unable to return any results and mapping was instead performed using MeDuSa v1.6 [252].

Pseudomonas sp. NS1

Mapping of *Pseudomonas sp NS1* contigs provided by Source Bioscience was performed using the *Pseudomonas* strain that had been identified as sharing the highest nucleotide identity by BLASTN [253] analysis of all provided contigs, *Pseudomonas azotoformans* S4 (GCA_001579805.1), as a reference sequence using CONTIGuator 2.7.4 [251].

2.5.2 Phage phylogeny

For analysis of phage phylogeny of NV1 and ϕNV3, the evolutionary history was inferred using the Maximum Parsimony method over the entire genome sequence [254]. For NV1 a

total of 19 genomic sequences were analysed, which are shown in Table 2.12, based on analysis previously performed by Monique Eller *et al.* [108].

The phage genomic sequences were first aligned using ClustalW [255], then the Maximum Parsimony tree was calculated using Mega6.06 [256] with 500 Bootstrap Test replicates [257] , the MP tree was obtained using the Subtree-Pruning-Regrafting algorithm [258] with search level 1 in which the initial trees were obtained by the random addition of sequences (10 replicates), all codon positions were included (1^{st}, 2^{nd},3^{rd} and Noncoding) and all positions containing gaps and missing data were eliminated. The final tree was rooted to *Enterobacteria phage T7*.

Table 2.12. Phage strain sequences used in NV1 phylogenetic relationship analysis.

Strain:	**Accession:**
Enterobacteria phage T7	NC_001604.1
Pseudomonas aeruginosa phage PaP3	NC_004466.2
Pseudomonas phage DL54	KR054029.1
Pseudomonas phage gh-1	NC_004665.1
Pseudomonas phage LUZ7	NC_013691.1
Pseudomonas phage LUZ24	NC_010325.1
Pseudomonas phage MR299-2	JN254801.1
Pseudomonas phage PaP4	KC294142.1
Pseudomonas phage phi-2	NC_013638.1
Pseudomonas phage PhiCHU	KP233880.1
Pseudomonas phage phiIBB-PAA2	KF856712.1
Pseudomonas phage phiIBB-PF7A	NC_015264.1
Pseudomonas phage ϕKMV	NC_005045.1
Pseudomonas phage tf	NC_017971.2
Pseudomonas phage TL	HG518155.1
Pseudomonas phage UFV-P2	JX863101.2
Pseudomonas phage vB PaeP C1-14 Or	HE983844.1
Pseudomonas phage vB PaeP p2-10 Or1	HF543949.1

The same method of Maximum Parsimony analysis was performed on ϕNV3 using 18 phage genomic sequences of *Autographivirinae* and *Podoviridae* members, shown in Table 2.13, including NV1 and ϕNV3 (not included in Table 2.13). The final tree was rooted to *Enterobacteria phage K1E*.

Table 2.13. Phage strain sequences used in ϕNV3 phylogenetic relationship analysis.

Strain:	**Accession:**
Enterobacteria phage K1-5	AY370674.1
Enterobacteria phage K1E	NC_007637.1
Enterobacteria phage SP6	AY370673.1
Erwinia amylovora phage Era103	NC_009014.1
Kluyvera phage Kvp1	NC_011534.1
Pseudomonas phage ϕKMV	NC_005045.1
Pseudomonas phage DL62	KR054031.1
Pseudomonas phage gh-1	NC_004665.1
Pseudomonas phage LKA1	NC_009936.1
Pseudomonas phage LKD16	NC_009935.1
Pseudomonas phage LUZ19	AM910651.1
Pseudomonas phage MPK7	JX501340.1
Pseudomonas phage phikF77	FN263372.1
Pseudomonas phage phiNFS	KU743887.1
Pseudomonas phage PT2	NC_011107.1
Pseudomonas phage UFV-P2	JX863101.2

2.5.3 Bacterial phylogeny

Bacterial phylogeny was inferred using multi locus sequence analysis (MLSA) of 5 genes, 23s *rlmJ*, 16s rDNA fragments, *rpoD, gyrB, rpoB.*

The 16S Fragment were generated by identifying the region of DNA between the pseudomonas specific primers PA-GS-F and PA-GS-R [259] PA-GS-F (5′-GACGGGTGAGTAATGCCTA3-′) and PA-GS-R (5′-CACTGGTGTTCCTTCCTATA-3′) *in silico* with Artemis v16.0.0.1 genome browser.

All genes were aligned individually for all bacterial strains using MUSCLE [260] and the alignments then transferred to Mega6.06 [256], where the gaps were removed and the

alignments for all the genes were then concatenated into a single alignment session for phylogenetic analysis.

Table 2.14. *Pseudomonas* strain sequences used for phylogenetic relationship analysis.

Strain:	**Accession:**
P. azotoformans LMG 21611	NZ_LT629702.1
P. azotoformans F77	NZ_CP019856.1
P. azotoformans S4	NZ_CP014546.1
P. fluorescens F113	NC_016830.1
P. fluorescens L228	NZ_CP015639.1
P. tolaasii 2192T	NZ_CP020369.1
P. tolaasii PMS117	NZ_AJXG01000263.1
P. tolaasii 6264	NZ_AKYY00000000.1
P. protegens CHA0	NC_021237.1
P. cedrina BS2981	LT629753.1
P. fulva 12-X	NC_015556.1
P. putida KT2440	NC_007005.1
P. syringae B728a	NC_002516.2
P. aeruginosa PA01	NC_002516.2
P. chloroaphis PA23	NZ_CP008696.1
P. agarici NCPPB 2472	NZ_CP014135.1
P. agarici NCPPB 2289	NZ_JH730872.1
P. agarici LMG2112	NZ_FOAR01000001.1
E. coli K-12 subset MG1655	NC_000913.3

Phylogenetic relationships between 15 of the strains in Table 2.14, chosen to most accurately place each of *P. agarici* NCPPB 2472, *P. tolaasii* NCPPB 2192T and *P.* sp NS1, were calculated using the Maximum Likelihood method based on the Tamura-Nei model [261]. The initial tree for the heuristic search was obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood approach and then selecting the topology with superior log likelihood value.

2.5.4 Bacterial secondary metabolite and NRPS prediction

Putative biosynthetic gene clusters were predicted from the full genomic sequence files of *P. tolaasii*, *P. agarici* and *Pseudomonas sp.* NS1 using antiSMASH 3.0 [262]. These putative clusters were then checked manually using both BLASTN and BLASTP [253] and checked for predicted function.

In order to identify ORF's potentially encoding large NRPS's all ORF's over 1000 aa in length were identified in Artemis and analysed with BLASTP [253]. ORFs identified as potentially encoding NRPSs were then analysed using InterPro v65.0 [263] for the presence of condensation, adenylation and thiolation domains.

2.5.5 RNA sequencing analysis

The returned RNA sequencing reads of *P. agarici* samples were aligned to the partially assembled *P. agarici* NCPPB 7472 genome sequence using Burrows-Wheeler Aligner (Illumina Basespace). The resulting alignment files in BAM format were imported into the Bioconductor v3.6 [264, 265] package in RStudio (Version 1.0.143) running R version 3.4.0, for further analysis. The gene annotation model (gff) used for read counting was created from genomic feature file (gff3) produced by the NCBI Prokaryotic Pipeline for *P. agarici* NCPPB 2472.

Read counting in Bioconductor v3.6 [264, 265] was performed using the *summarizeOverlaps* function of the *GenomicAlignments* package [266], reads were only counted once to each gene with settings for paired-end reads and non-strand specific reads. Non-phage treated samples (P4, P5, P6) were used as the reference level in DESeq2 [267]. The count matrix was filtered to remove results with counts <10.

Scatterplots were created using log2, regularized-logarithm transformation (*rlog*) [267] and the variance stabilizing transformation (*vst*) for negative binomial data with a dispersion-mean trend [267]. Euclidean distances were created using the *rlog*-transformed values to ensure equal contribution from all genes. Distances were visualised in a heat map created with the '*pheatmap'* package, colour was added using the *RColorBrewer* package. Poisson Distance was calculated with the *PoiClaClu* package [268] to take in to account the variance structure of counts using the non-normalized count matrix, and imaged again with *pheatmap* and *RColorBrewer*. The PCA plot data was generated for the *rlog*-transformed values using *plotPCA* function in the DESeq2 package and plotted using the *ggplot2* package [269].

2.5.6 Phage terminal repeat identification

Identification and analysis of putative phage terminal repeats was performed using PhageTerm v1.0.11 [270] on the Galaxy web platform [271]. The seed length was set as 20, the peak surrounding region as 20 and the coverage limit as 250.

Chapter 3:
Comparative genomics of *Agaricus bisporus* pathogenic *Pseudomonas* strains.

3.1 Introduction

Agaricus bisporus are an important agricultural crop, with 8-9x10^5 tonnes are cultivated in Europe alone, worth an approximate 2 billion Euros annually [1]. In this study I will be isolating, characterising and comparing the genomes of two *Agaricus bisporus* pathogenic *Pseudomonas* species, *Pseudomonas tolaasii* NCPPB 2192T, *Pseudomonas agarici* NCPPB 2472 and a third non-pathogenic species recently identified which I have provisionally named *Pseudomonas* sp. NS1(2017); with the aim of identifying potential virulence factors and phage resistance/susceptibility factors.

P. tolaasii is the causal organism of the mushroom blotch disease "brown blotch disease", the symptoms of which commence as small (1-4 mm) brown or cream lesions on the pileus and stipe of the mushroom that become darker and more sunken as the damage progresses [1]. Mushroom loss due to brown blotch disease caused by *P. tolaasii* is economically significant as it has been estimated to be responsible for a 5-10% crop loss in the U.K and a further downgrading of 10% of mushrooms produced [11].

P. agarici is the causal agent of the disease 'drippy gill' in *A. bisporus* mushrooms, a disease characterised by cloudy white droplets of bacteria on the gill tissue of mushrooms, that can progress to cause total destruction of the affected tissue [24]. The majority of the large U.K based outbreaks were reported before 1980, although outbreaks of *P. agarici* have also been reported in New Zealand in 1991 [26] and in Serbia as recently as 2008 [32], which indicates that *P. agarici* has the potential to cause economic loss to cultivated mushroom farmers.

While other members of the *P. fluorescens* group have been well characterised at the genome level, such as the type strains *P. fluorescens* and *P. tolaasii*; *P. agarici* has not been the subject

of in depth characterisation and analysis, with only two partial genomes published (*P. agarici* strain NCPPB 2289 and *P. agarici* strain LMG2112). To date no scientific literature has described the genome of *P. agarici* NCPPB 2472 and this study aims to address this shortfall.

3.2 *Pseudomonas tolaasii* 2192T

3.2.1 Proof of pathogenicity

In order to determine whether the provided *P. tolaasii* 2192T strain was capable of producing brown blotch disease symptoms on the host *A. bisporus* cap surface, a pathogenicity assay was performed. After 72 hours, the mushrooms infected with *P. tolaasii* 2192T begun to display darkening of the tissue and pitting of the surface, as shown in Figure 3.1 confirming the ability of the *P. tolaasii* strain to cause brown blotch disease.

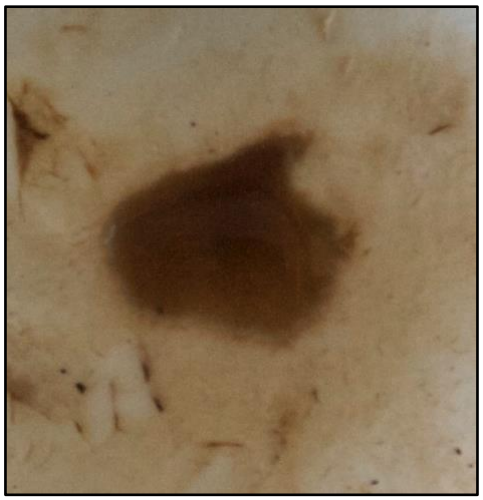

Figure 3.1. *A. bisporus* mushroom cap showing brown discolouration and pitting associated with brown blotch disease caused by *P. tolaasii* infection.

3.2.2 Phylogeny

To infer the phylogenetic relationship of *P. tolaasii* NCPPB 2192T among the fluorescent Pseudomonads a Maximum-Likelihood analysis was performed on the DNA sequence returned from whole-genome sequencing using 5 housekeeping genes (*RlmJ*, *rrs*, *RpoD*, *GyrB* and *RpoB*) using the Tamura-Nei model [261].

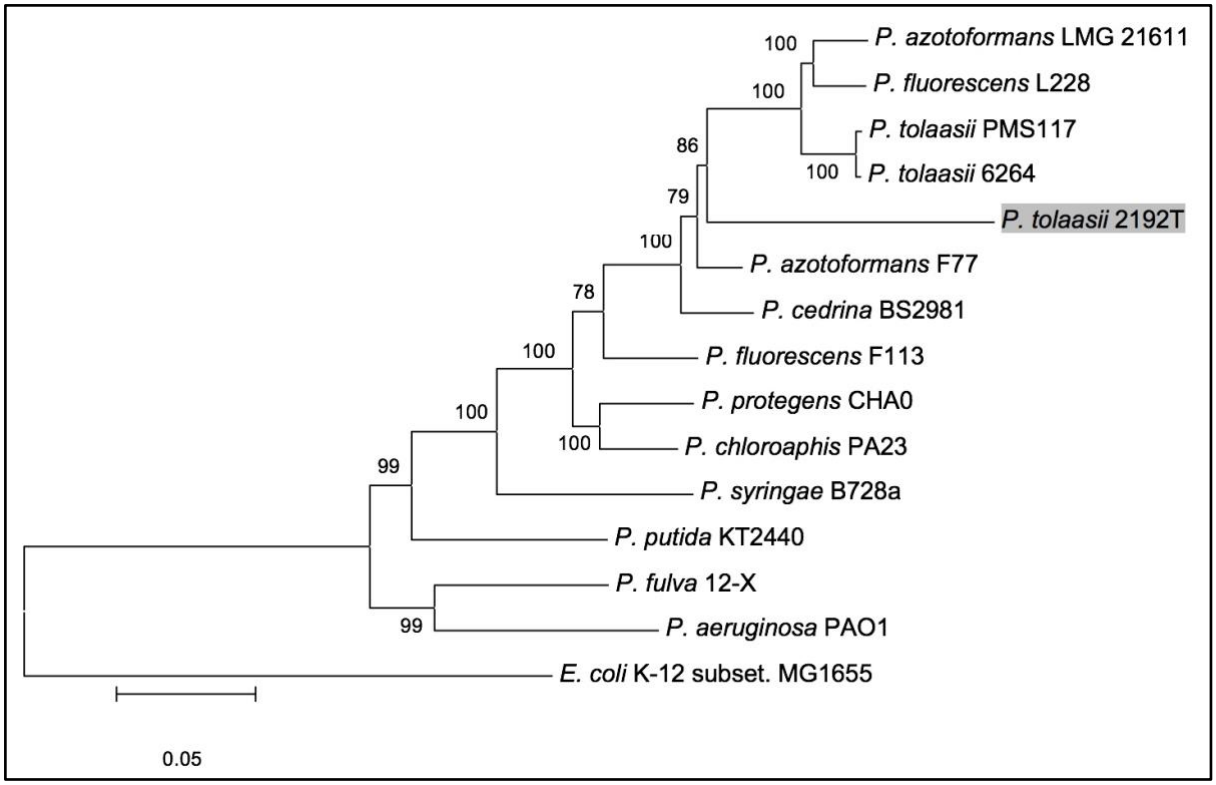

Figure 3.2. Molecular phylogenetic analysis of *P. tolaasii* 2192T (highlighted) by Maximum Likelihood method. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 15 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 8768 positions in the final dataset. Evolutionary analyses were conducted in MEGA6.06

The results of this analysis shown in Figure 3.2 agree with the phylogeny reported by Yamamoto *et al.* [272]; with *P. tolaasii* 2192T clustering with the closely related *P. azotoformans* and *P. fluorescens* which are all members of the Pseudomonas intragenic cluster II, *P. fluorescens* complex, *P. fluorescens* lineage. However, it is unusual that *P. tolaasii* 2192T is located on a separate branch to *P. tolaasii* 6264 and PMS117, indicating that it may have diverged from the other members of the *P. tolaasii* species.

3.2.3 Sequencing and genome organisation

For the whole-genome sequencing of *P. tolaasii*, a total of 150 µl of genomic DNA at a concentration of 79 ng/µl (11.85 µg total) was sent to Source Bioscience, who prepared a whole-genome sequencing library using an Illumina TruSeq Nano kit, before paired end sequencing was performed on an Illumina MiSeq® with a read length of 50 bp.

In total, the sequencing run returned 536.3 Mbases of sequence, with an average coverage of approximately 40-fold. The reads were initially assembled in to 262 contigs by Source Bioscience, with a total sequence length of 6,808,027 bp. These contigs were then aligned to *Pseudomonas tolaasii* PMS117 with the MeDuSa v1.6 software, which returned a single scaffold, and the resulting sequence was annotated using the NCBI Prokaryotic Genome Annotation Pipeline. The full genome is available under the accession code NZ_CP020369.

Of the 6,286 ORFs identified, of which 6,065 were identified as being protein coding sequences and 63 RNA encoding genes were identified including 56 tRNAs, 3 complete rRNAs (5S, 16S, 23S) and 4 ncRNAs. The total gene count is lower than that of the closely related *P. tolaasii* PMS117 which is predicted to encode 6,482 genes, but higher than *P. tolaasii* 6264 which is predicted to encode 6,093 genes; a full comparison is shown in Table 3.1.

Table 3.1. Comparison of the genomes and predicted genes of *P. tolaasii* strains 2192T, PMS117 and 6264.

Strain	Total (bp)	ORFs	Encoding	Genes (RNA)	tRNA	rRNA	ncRNA	GC (%)
2192T	6,856,683	6,286	6,065	63	56	3	4	60.5
PMS117	7,007,821	6,482	6,011	60	53	3	4	60.2
6264	6,233,813	6,093	5,265	63	54	5	4	61

A full circular map illustrating the genome of *P. tolaasii* 2192T including G+C content, location of the tRNA genes and gene orientation is shown in Figure 3.3.

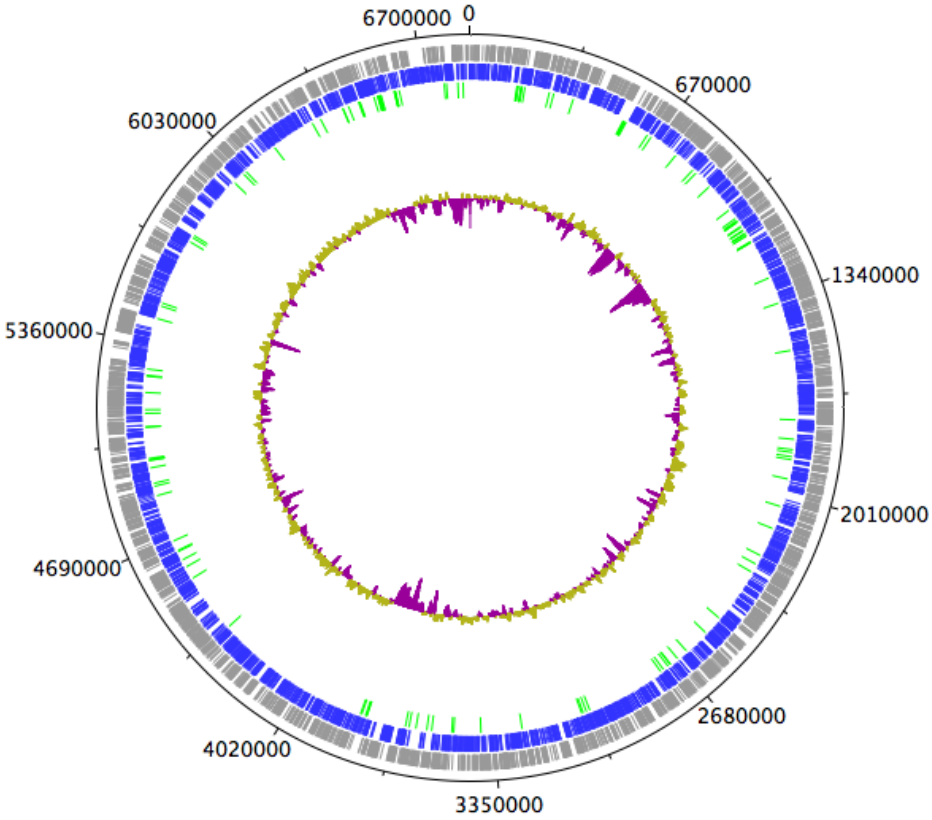

Figure 3.3. Circular map of the *P. tolaasii* NCPPB 2192T genome with the genes encoded on the positive strand in the first lane (grey), the negative strand in the second lane (blue) and the tRNA genes in the third lane (green). The deviation in GC content from the total average is illustrated in the centre circle with yellow bars denoting an above average content and purple bars, a below average GC content. Figure generated using DNAPlotter function of the Artemis v16.0.0.1 genome browser.

3.2.4 Secondary metabolites and NRPSs

The genome of *P. tolaasii* NCPPB 2192T was mined for potential biosynthetic gene clusters, which revealed a total of 17 complete and incomplete putative NRPS encoding ORFs. These ranged from 21,173 bp to 675 bp in length. Six putative NRPSs were identified that showed both amino acid and nucleotide similarity to a reported tolaasin biosynthetic cluster of *P. costantinii* [50] as shown in Table 3.2. The ORF RS02650 predicted by the NCBI prokaryotic pipeline to encode a frameshifted single protein was identified as to be comprised of the C-terminus of a NRPS, between bases 531441 and 534953 on the leading strand, with amino acid

and nucleotide similarity to the *P. costantinii taaC* gene and associated protein product; and a complete ORF between bases 535013 and 545227 on the leading strand, with amino acid and nucleotide similarity to the *P. costantinii taaD* gene and associated protein product.

Table 3.2. Predicted NRPS of *P. tolaasii* 2192T with similarity to the tolaasin biosynthesis cluster of *P. costantinii* as identified by BLASTP [253], Predicted function of the gene product in *P. tolaasii* 2192T and the associated protein in *P. costantinii*, including the locus tag of the encoding gene are shown.

Locus Tag:	**Predicted Function:**	***P. costantinii* Protein**	**Locus Tag:**	**Cover:**	**Ident:**
RS30620	NRPS	TaaA	CCJ67636.1	97%	85%
RS22935	NRPS	TaaB	CCJ67637.1	99%	72%
RS23980	NRPS	TaaC (N)	CCJ67638.1	99%	86%
RS02650	NRPS	TaaC (C)	CCJ67638.1	87%	82%
RS02650	NRPS	TaaD	CCJ67639.1	98%	77%
RS24725	NRPS	TaaE	CCJ67640.1	100%	85%

Within the tolaasin biosynthetic cluster of *P. costantinii* that the gene encoding the protein TaaE is upstream of genes encoding two further proteins, a macrolide-specific efflux protein MacA and macrolide export ATP-binding/permease protein MacB (CCJ67641.1 and CCJ67642.1), it was observed that the gene encoding the predicted protein in *P. tolaasii* 2192T with amino acid identity to TaaE (RS24725) was likewise upstream of genes predicted to encode both a macrolide-specific efflux protein macA and macrolide export ATP-binding/permease protein MacB (RS24720 and RS24715). Similarly, the gene encoding the protein TaaA in *P. costantinii* was found to be preceded by a cyclic peptide transporter family protein (CCJ67635.1), it was also observed that the gene encoding the predicted protein in *P. tolaasii* 2192T with amino acid identity to TaaA (RS30620) was likewise preceded by a gene predicted to encode a cyclic peptide export ABC transporter (RS30625).

Two of the further NRPSs identified *in silico* were also predicted to be incorrectly separated parts of a single large NRPS, with similarity to a known biosynthesis gene, these were RS16165

and RS16170 (1527 aa and 224 aa in size respectively), which were found to show moderate amino acid identity by BLASTP [253] to the known pyoverdine synthesis genes PvdI of *Pseudomonas syringae* pv. daphniphylli (57% identity) Accession: KPX09284.1 and the pyoverdine sidechain peptide synthetase of *P. fluorescens* (68% identity) Accession: WP_003191936.1, respectively. These genes were also found to be upstream of a further incomplete NRPS, RS16160, that showed 63% amino acid identity to pyoverdine synthesis protein PvdJ of *P. aeruginosa* 62 (Accession: ERX81199.1) and downstream of a TonB-dependent siderophore receptor (RS16175).

Similarly, a 1948 aa NRPS was identified (RS12420) that is also hypothesised to be involved in siderophore synthesis, due to its location downstream of a gene predicted to encode a siderophore synthetase (RS12390), MFS transporter (RS12395) and TonB-dependent receptor (RS12400); as well as being located upstream from a TonB-dependent siderophore receptor (RS12430).

Phage tail-like bacteriocin

A biosynthetic gene cluster hypothesised to be involved in the synthesis of a phage tail-like bacteriocin (PTLB) was identified, shown in Table 3.3.

Table 3.3. Predicted PTLB gene cluster of *P. tolaasii* 2192T.

Predicted Function:	Start:	End:	Locus Tag:	Protein Id:
pyocin R2, holin	3523422	3523766	RS16310	WP_016970764.1
hypothetical protein	3523787	3524302	RS16315	WP_016970765.1
phage baseplate assembly protein V	3524306	3524914	RS16320	WP_016970766.1
phage baseplate protein	3524927	3525259	RS16325	WP_016970767.1
baseplate J protein	3525256	3526251	RS16330	WP_016970768.1
phage tail protein I	3526248	3526976	RS16335	WP_016970769.1

tail fiber protein	3526973	3527506	RS16340	WP_016970770.1
hypothetical protein	3527533	3528171	RS16345	WP_016970771.1
hypothetical protein	3528233	3528439	RS16350	WP_020372583.1
tail protein	3528700	3529866	RS16355	WP_016970774.1
phage tail protein	3529866	3530372	RS16360	WP_016970775.1

Downstream of the predicted PTLB gene cluster, a further protein (RS16305) was later identified, which was initially classified as an XRE family transcriptional regulator, however it was then noted that it showed 100% amino acid identity to the repressor protein PrtR of *P. fluorescens* Strain: PCL1751 and is therefore likely to be involved in transcriptional regulation of the above identified PTLB in tandem with another protein, the pyocin activator protein PrtN (RS06590), identified in the genome of *P. tolaasii* 2192T.

3.2.5 Phage resistance systems

No CRISPR-associated repeats or proteins were identified in the *P. tolaasii* 2192T genome or within those of *P. tolaasii* PMS117 or 6264. However, multiple restriction endonuclease proteins, another well documented system by which bacteria are able to resist phage infection [203], were identified within the genome of *P. tolaasii* 2192T, shown in Table 3.4.

The first predicted 'restriction endonuclease' (RS17360) contains an McrB conserved domain, which is associated with the GTP-binding regulatory subunit [273]. The RS17360 predicted restriction endonuclease gene is upstream of a 'hypothetical protein' (RS17355) that contains a PDDEXK_7 conserved domain, which is a domain of the PD-(D/E)XK nuclease family and it is speculated that it could function as a methylation-dependent restriction enzyme [274]; therefore it is likely that both RS17360 and RS17355 are two subunits of a single uncharacterised Type IV methylation-dependent restriction endonuclease system.

Table 3.4. Predicted restriction endonuclease proteins of *P. tolaasii* 2192T.

Protein:	Start:	End:	Locus Tag:
Restriction endonuclease	3766583	3768847	RS17360
Restriction endonuclease subunit R	3783281	3786340	RS17410
Restriction endonuclease subunit M	<5750878	5751093	RS26390
Restriction endonuclease subunit M	<6602607	6602822	RS30090
Restriction endonuclease subunit M	6602929	>6603507	RS30095

The second predicted restriction endonuclease (RS17410) is predicted to be a Type III restriction endonuclease and is directly downstream of a site-specific DNA-methyltransferase (RS17420), with a second site-specific DNA-methyltransferase (RS17440) approximately 7,700 bp upstream of the first.

The incomplete methyltransferase encoded by RS26390, predicted as a 'restriction endonuclease subunit M' shows significant amino acid identity to a known prophage (Prophage PSPPH02) adenine modification methyltransferase and therefore has no associated restriction endonuclease. Likewise, the incomplete methyltransferases encoded by RS30090 and RS30095 are likely to be of phage origin; the protein encoded by RS30090 shows 100% amino acid identity to a protein of the prophage PSPPHO2 (as predicted by BLASTP [253]) and the protein encoded by RS30095 shows a high degree of similarity to prophage PSPPH06 (95% Ident as predicted by BLASTP [253]). Furthermore, there are several ORFs downstream of both these genes that encode phage related proteins, including RS30070 (baseplate assembly protein) and RS30080 (microvirus H family protein).

A biosynthetic operon comprised of 15 genes (Table 3.5) was identified in the genome of *P. tolaasii* 2192T, hypothesised to be involved in the production of alginate due to the presence of a gene encoding the alginate biosynthesis protein Alg44 (RS11530), which has been demonstrated to be required for alginate biosynthesis in *Pseudomonas aeruginosa* [275] and AlgE, also reported to be involved alginate biosynthesis in *Pseudomonas* [276].

Table 3.5. Predicted alginate biosynthesis proteins of *P. tolaasii* 2192T

.

Predicted Protein Function	Locus Tag:
Nucleotide sugar dehydrogenase	RS11545
Glycosyl transferase	RS11540
Alginate biosynthesis protein Alg44	RS11530
Sel1 repeat family protein	RS11525
Alginate biosynthesis protein AlgE	RS11520
Right-handed parallel beta-helix repeat-containing protein	RS11515
Alginate O-acetyltransferase	RS11510
Mannuronate-specific alginate lyase	RS11505
MBOAT family protein	RS11500
Alginate O-acetyltransferase	RS11495
Alginate O-acetyltransferase	RS11490
Mannose-1-phosphate guanylyltransferase/mannose-6-Phosphate isomerase	RS11485
DUF3077 domain-containing protein	RS11480
Hypothetical protein	RS11475
Short-chain dehydrogenase	RS11470

3.3 *Pseudomonas agarici* NCPPB 2472

3.3.1 Proof of pathogenicity

In order to determine the pathogenicity of *P. agarici* NCPPB 2472, the gill tissue of previously un-opened and disease-free *Agaricus bisporus* was inoculated with bacterial culture.

Figure 3.4. Gill tissue of *Agaricus bisporus* inoculated with *P. agarici* NCPPB 2472 showing characteristic symptoms of 'drippy gill' disease.

After 72 hours of incubation at 28°C, cloudy-white bacterial droplets appeared on the mushroom gill tissue, as shown in Figure 3.4. This shows that *P. agarici* NCPPB 2472 is capable of infecting and causing drippy gill disease symptoms on *Agaricus bisporus* mushroom gill tissue.

3.3.2 Phylogeny

To infer the phylogenetic relationship of *P. agarici* NCPPB 2472 among the fluorescent Pseudomonads a Maximum-Likelihood analysis was performed on the DNA sequence returned from whole-genome sequencing using 5 housekeeping genes (*RlmJ*, *rrs*, *RpoD*, *GyrB* and *RpoB*) using the Tamura-Nei model [261].

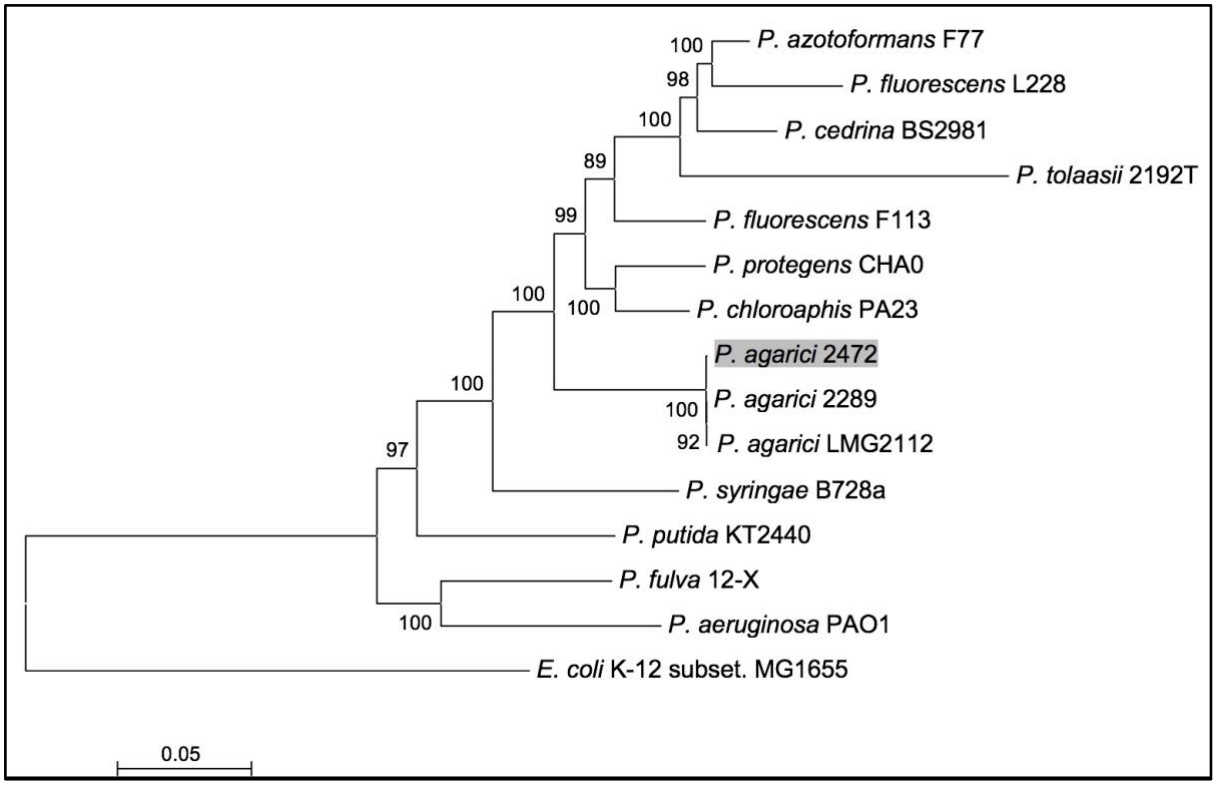

Figure 3.5. Molecular phylogenetic analysis of *P. agarici* NCPPB 2472 (highlighted) by Maximum Likelihood method. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 15 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 9117 positions in the final dataset. Evolutionary analyses were conducted in MEGA6.06.

The results in Figure 3.5 show that *P. agarici* 2472 clusters significantly in a tight monophyletic branch with other strains of *P. agarici*, which indicates that they are likely to be closely related. This confirms the results reported by Yamamoto *et al.*[272] who also reported that the *P. agarici* strains used in their study formed a tight monophyletic branch. Similarly, the results in Figure 3.5 show that the *P. agarici* branch is closer to the branch containing *P. chloroaphis* than to that containing *P. fluorescens*, which again is similar to the results reported by Yamamoto *et al.* [272], who suggest that *P. agarici* forms part of the *P. chloroaphis* lineage within the *P. fluorescens* complex of the *Pseudomonas* intrageneric cluster II, while *P. tolaasii* forms part of *the P. fluorescens* lineage.

3.3.3 Sequencing and genome organisation

Sequencing of the genome of *P. agarici* by Source Bioscience returned 5,618,560 bp in a total of 232 contigs with an average contig length of 24,218 bp, a maximum contig length of 607,529 bp and a minimum of 211 bp. These contigs were then mapped to two related *Pseudomonas* strains, *Pseudomonas fluorescens* A506 and *Pseudomonas agarici* NCPPB 2289 using CONTIGuator 2.7.4 [251], in order to orient the returned contigs correctly and allow a genome scaffold to be created that would minimise the number of contig junctions that require stitching. Mapping of the contigs to *P. fluorescens* A506 reduced the number of contigs from 232 to 32 and the total sequence length to 5,314,908 bp. Primers were then created for each contig junction to allow PCR amplification of the unknown region between them. The ACT comparison of the genome of *P. fluorescens* A506 and the 32 aligned contigs of the sequenced *P. agarici* is shown in Figure 3.6, the total degree of nucleotide similarity between the two is not high however there are several large regions of similarity that allow a confident prediction of contig orientation and ordering.

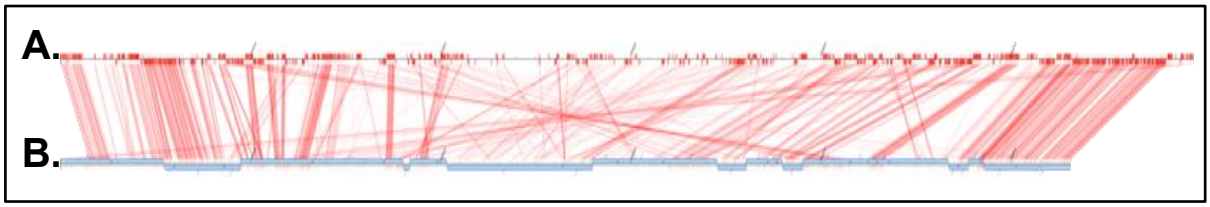

Figure 3.6. Alignment map of *P. agarici* NCPPB 2472 contigs (B.) mapped against the genome of *P. fluorescens* A506 (A.) generated using WebACT. Red bars indicate collinear regions of similarity between the mapped sequences.

The returned *P. agarici* NCPPB 2472 contigs were also mapped to the genome sequence of *P. agarici* NCPPB 2289 and this reduced the number of contigs from 232 to 48 and the total sequence to 5,501,881 bp. The ACT comparison of the genome of *P. agarici* NCPPB 2289 and the 48 aligned contigs is shown in Figure 3.7, the total degree of nucleotide between the two is very high, with large regions of collinear similarity that would indicate a very confident prediction of contig orientation and ordering.

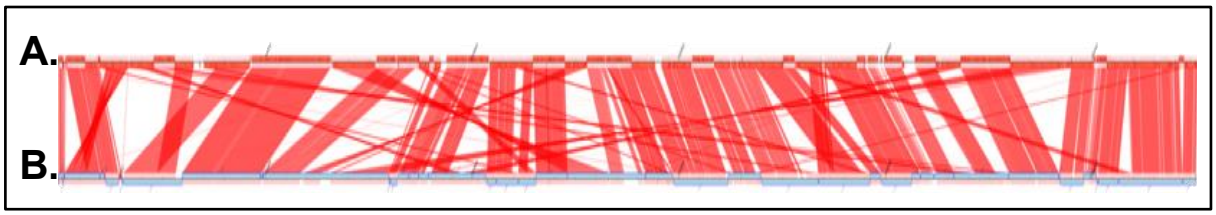

Figure 3.7. Alignment map of *P. agarici* NCPPB 2472 contigs (B.) mapped against the genome of *P. agarici* NCPPB 2289 (A.) generated using WebACT. Red bars indicate collinear regions of similarity between the mapped sequences.

The *P. agarici* NCPPB 2472 contigs mapped to the genome sequence of *P. agarici* NCPPB 2289 were then annotated using the NCBI Prokaryotic Genome Annotation Pipeline, the results are available with the accession code NZ_CP014135.

Of the 4,901 ORFs identified, 4,673 were identified as being protein coding and 67 RNA encoding genes; including 59 tRNAs, 4 complete rRNAs (5S, 16S, 23S) and 4 ncRNAs. The total gene count is lower than the counts for both the closely related strains *P. agarici* NCPPB 2289 and LMG 2112, a full comparison is shown in Table 3.6.

Table 3.6. Comparison of the genomes and predicted genes of *P. agarici* strains NCPPB 2472, 2289 and LMG 2112.

Strain	Total (bp)	ORFs	Encoding	Genes (RNA)	tRNA	rRNA	ncRNA	GC (%)
2472	5,502,003	4,901	4,673	67	59	4	4	58.89
2289	5,511,390	5,009	4,414	62	54	4	4	59
LMG 2112	5,508,100	5,002	4,738	68	58	6	4	59

A full circular genome map was created using the DNAPlotter function of Artemis v16.0.0.1 in order to visualise the returned data and is shown in Figure 3.8.

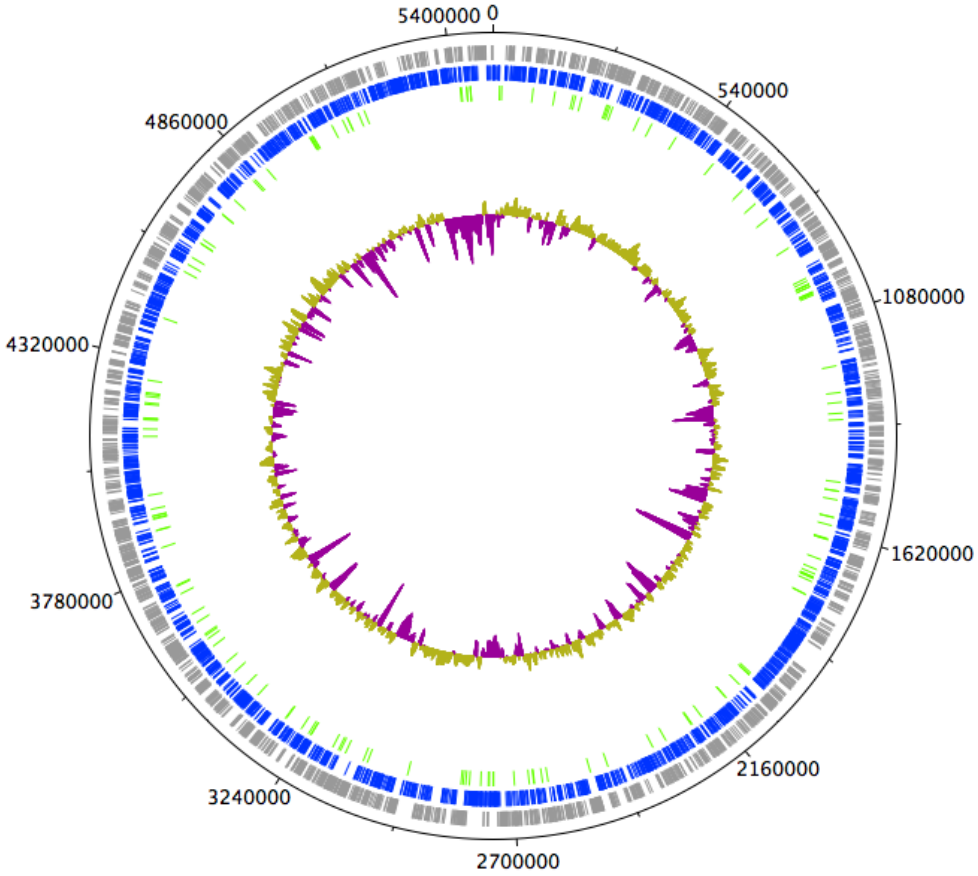

Figure 3.8. Circular map of the *P. agarici* NCPPB 2472 genome with the genes encoded on the positive strand in the first lane (grey), the negative strand in the second lane (blue) and the tRNA genes in the third lane (green). The deviation in GC content from the total average is illustrated in the centre circle with yellow bars denoting an above average content and purple bars, a below average GC content. Figure generated using DNAPlotter function of the Artemis v16.0.0.1 genome browser.

3.3.4 Contig stitching

Contig stitching using PCR amplification of the regions between contigs was performed in order to both confirm the bioinformatic predictions of genome organisation and to identify and sequence any regions not covered by the initial sequencing run. 125 PCRs were performed corresponding to the reactions previously outlined in Table 2.6 and Table 2.7. Of the 30 reactions that were performed based on the alignment of *P. agarici* and *P. fluorescens* A506 that are shown in the agarose gel image of Figure 3.9, 8 returned products that were

successfully sequenced and allowed stitching of their associated contigs (Figure 3.9, lanes 1, 2, 18, 19, 23, 26, 28 and 30). These successful reactions varied considerably in number of bases they added, from a maximum of 5366 bp, to a minimum of 737 bp, the full results are outlined in Table 3.7 and the returned sequences are available in Appendix 1.

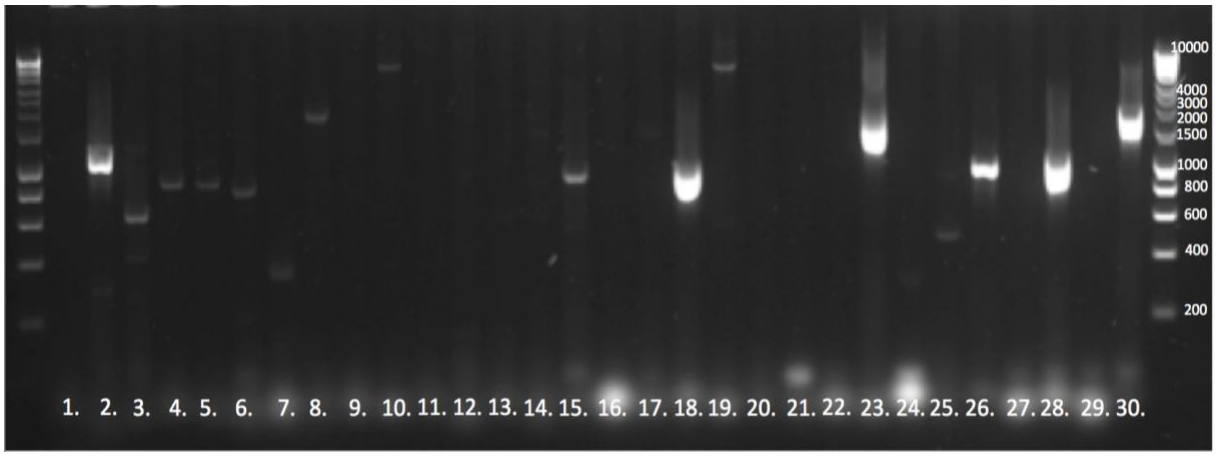

Figure 3.9. Agarose gel image of all PCR products from Table 2.6 from contig junctions generated from mapping of *P. agarici* NCPPB 2472 with *P. fluorescens* A506, lane numbers (1-30) correspond to respective reaction numbers (R1-R30).

Of the 46 reactions that were performed based on the alignment of *P. agarici* 2472 and *P. agarici 2289* shown in Table 2.7, only 3 returned products that were successfully sequenced and allowed stitching of their associated contigs, Rb13, Rb19, Rb40, lanes 13,19 and 40 of Figure 3.10 respectively. Of these three products, two corresponded to regions of overlap between the contigs and only one product returned novel sequence.

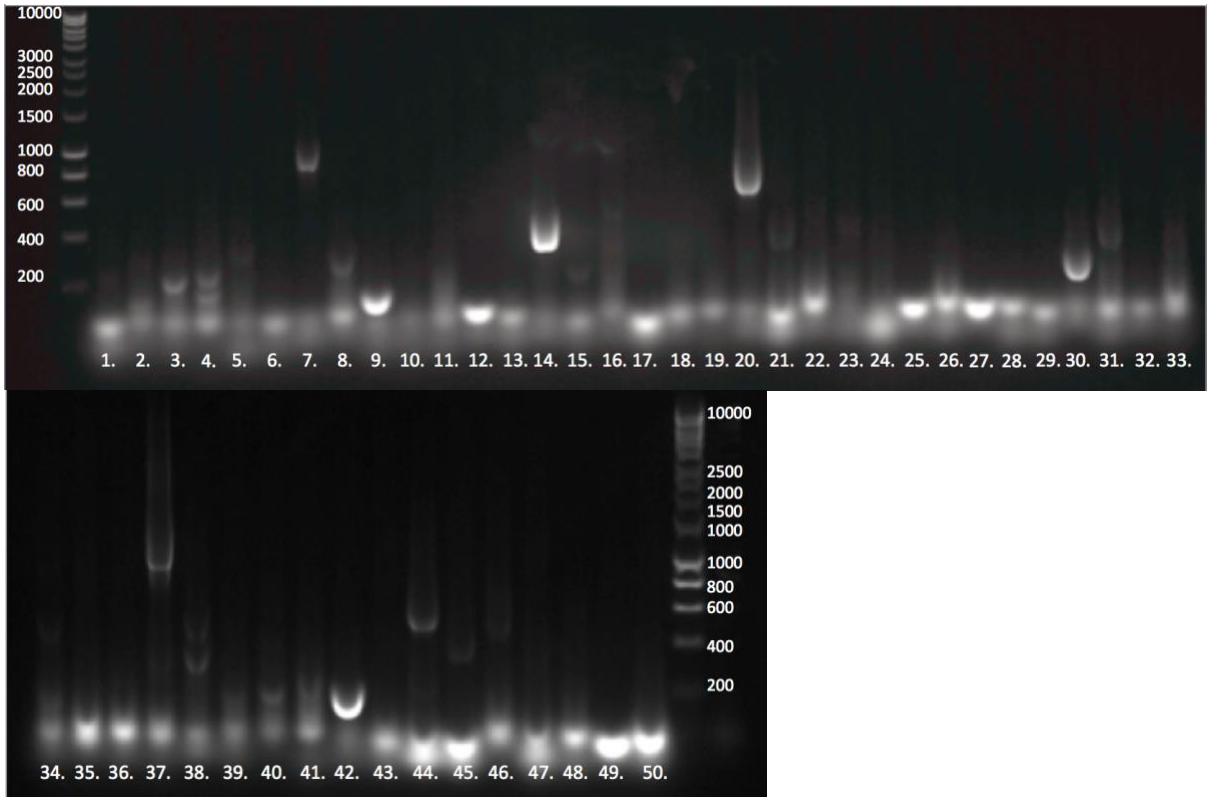

Figure 3.10. Gel image of all PCRs from Table 2.7 from contig junctions generated from mapping of *P. agarici* NCPPB 2472 with *P. agarici* NCPPB 2289, lane numbers (1-50) correspond to the respective reaction numbers (Rb1-Rb50).

The large discrepancy in reactions that returned products from which sequence data was obtained would appear to show that the genome organisation of *P. agarici* NCPPB 2472 is more similar to that of *P. fluorescens* A506 than *P. agarici* NCPPB 2289, despite the significantly larger degree of collinear similarity between the genomes of *P. agarici* NCPPB 2472 and *P. agarici* NCPPB 2289. However due to limitations of Sanger sequencing several of the products returned from the PCRs of contig alignments of *P. agarici* NCPPB 2472 and *P. agarici* NCPPB 2289 were not able to be successfully sequenced and with other methods may in fact contain valid sequence.

Table 3.7. Table of all completed contig junction stitches and associated sequence additions or overlaps.

Reaction:	Contig 1:	Contig 2:	Bp Added:	Overlap:	GC%
R1	3	21	5366	No	52
R2	21	38	847	No	58
R18	11	9	737	No	63
R19	9	42	3203	No	64
R23	29	6	983	No	60
R26	24	15	849	No	58
R28	13	46	882	No	57
R30	12	23	1597	No	59
Rb13	22	35	N/A	Yes	N/A
Rb19	15	24	848	No	58
Rb40	61	39	N/A	Yes	N/A

3.3.5 Secondary metabolites and NRPSs

The genome of *P. agarici* was mined for potential biosynthetic gene clusters and NRPSs. In total 10 NRPS/PKSs were predicted in three distinct biosynthetic clusters; however, it was identified that two of these identified (RS0070/RS10745) were in fact part of larger NRPSs (RS12925 and RS00065 respectively) that had been split due to assembly errors of the full genome sequence data. The first cluster is comprised of 5 NRPS proteins, as shown in Table 3.8.

Table 3.8. Predicted NRPS proteins of *P. agarici* NCPPB 2472.

No.	Predicted Function:	Locus: Tag
1	non-ribosomal peptide synthetase	RS12910
2	non-ribosomal peptide synthetase	RS12915
3	non-ribosomal peptide synthetase	RS12920
4	non-ribosomal peptide synthetase	RS12925
5	non-ribosomal peptide synthetase	RS00065
6	non-ribosomal peptide synthetase (part of RS12925)	RS00070
7	non-ribosomal peptide synthetase (part of RS00065)	RS10745

The second is comprised of a single NRPS (RS20800) and a polyketide synthase (RS20810); the final biosynthetic cluster appears to possess only a single NRPS (RS10630). While all the NRPS showed a high degree of conservation with other NRPSs of the *Pseudomonas*, it was not possible to predict the function of the three biosynthetic clusters from *in silico* analysis.

Phage tail-like bacteriocin

A biosynthetic gene cluster hypothesised to be involved in the synthesis of a phage tail-like bacteriocin was identified, shown in Table 3.9.

Table 3.9. Predicted potential pyocin proteins of *P. agarici* NCPPB 2472.

Predicted Function:	Start:	End:	Locus Tag:	Protein Id:
pyocin R2, holin	2500773	2501108	AWM79_10925	WP_060782816.1
hypothetical protein	2501271	2501855	AWM79_10930	WP_060782817.1
DUF2635 domain-containing protein	2501852	2502034	AWM79_10935	WP_060782818.1
phage tail protein	2502034	2503530	AWM79_10940	WP_060782819.1
tail protein	2503597	2503944	AWM79_10945	WP_017131070.1
hypothetical protein	2503941	2504237	AWM79_10950	WP_017131069.1
phage tail protein	2504368	2505999	AWM79_10955	N/A
hydroxyacid dehydrogenase	2505980	2507245	AWM79_10960	WP_060782820.1
baseplate protein	2507249	2508295	AWM79_10965	WP_060782821.1
hypothetical protein	2508348	2508857	AWM79_10970	WP_060782822.1
hypothetical protein	2508857	2509255	AWM79_10975	WP_060782823.1
baseplate J protein	2509245	2510285	AWM79_10980	WP_017131062.1
phage tail protein	2510273	2510872	AWM79_10985	WP_017131061.1
hypothetical protein	2510884	2512029	AWM79_10990	WP_017131060.1
acyltransferase	2512800	2514725	AWM79_10995	WP_060782824.1
pyocin R, lytic enzyme	2515164	2515727	AWM79_11000	WP_017131058.1
lysozyme	2515709	2516245	AWM79_11005	WP_017131057.1

Similar to the PTLB cluster identified in *P. tolaasii* 2192T an XRE family transcriptional regulator was identified downstream of the PTLB gene cluster (AWM79_10920) which may function in a role similar to the PrtR protein as well as an incomplete gene predicted to encode the pyocin activator protein PrtN (AWM79_18570) which may be involved in transcriptional regulation of the PTLB cluster.

Siderophore Biosynthesis

A biosynthetic cluster was identified that is predicted to produce the siderophore achromobactin, shown in Table 3.10.

Table 3.10. Predicted achromobactin biosynthesis proteins of *P. agarici* NCPPB 2472.

No.	Predicted Function	Locus Tag:
1	RNA polymerase sigma factor	RS04660
2	sugar ABC transporter substrate-binding protein	RS04665
3	TonB-dependent siderophore receptor	RS04670
4	aspartate aminotransferase family protein	RS04675
5	AcsD protein	RS04680
6	diaminopimelate decarboxylase	RS04685
7	MFS transporter	RS04690
8	AcsC protein	RS04695
9	siderophore biosynthesis protein SbnG	RS04700
10	AcsA protein	RS04705

Three proteins (RS04680, RS04695 and RS04705) showed high amino acid identity to the proteins AcsD, AcsC and AcsA respectively, which have been demonstrated by Berti and Thomas [61] to be involved in the biosynthesis of achromobactin by *P. syringae* pv. syringae B728a. A fourth protein (RS04685), which is predicted to function as a diaminopimelate decarboxylase, I hypothesise to perform the function of the *P. syringae* pv. syringae B728a protein AcsE.

The gene cluster also shows significant similarities in ORF organisation to the achromobactin biosynthesis cluster of *P. syringae* pv. syringae B728a, as is illustrated in Figure 3.11. All ORFs show high (72-91%) amino acid identity to the corresponding ORFSs of *P. syringae* pv. syringae B728a, a full table of results including BLASTP [253] amino acid identity is available in Appendix 4, Table Ap4.5. The predicted function of this cluster in the production of a siderophore is further confirmed by the presence of a predicted TonB-dependent siderophore receptor (RS04670) and predicted siderophore biosynthesis protein SbnG (RS04700).

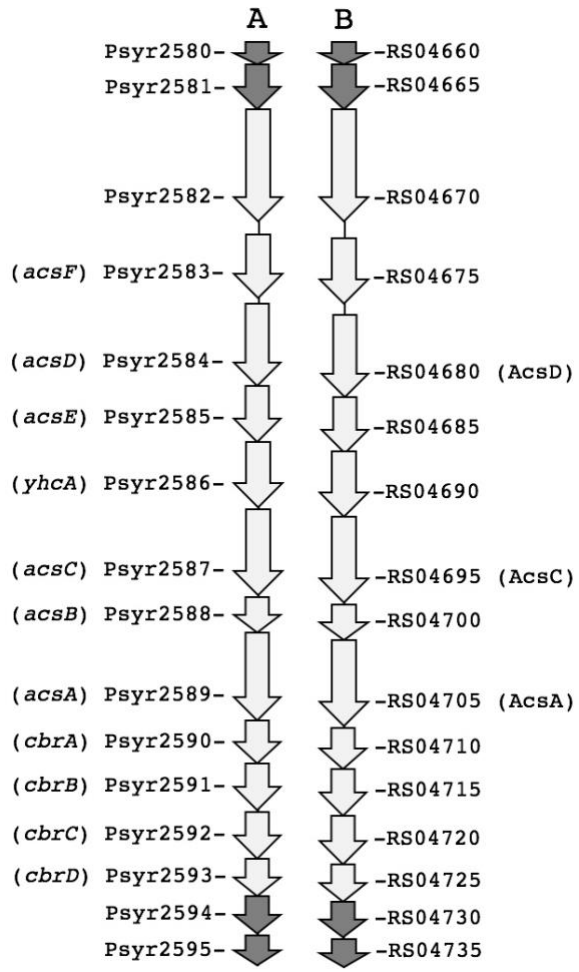

Figure 3.11. Comparison of the achromobactin biosynthetic cluster of (A) *P. syringae* pv. syringae B728a, including gene nomenclature proposed by Berti and Thomas [61] and (B) putative achromobactin biosynthetic cluster of *P. agarici* NCPPB 2472. Gene labels are adjacent to their respective ORFs. Figure generated using Microsoft PowerPoint.

3.3.6 Phage resistance systems

The genome of *P. agarici* was found to contain a single CRISPR-associated repeat region, between 1180063-1181413 bp on the positive strand. The *P. agarici* CRISPR/Cas system was identified as comprising of a cluster of 6 Type I-F CRISPR-associated proteins located downstream of the repeat region, outlined in Table 3.11.

Table 3.11. Predicted CRISPR-associated proteins of *P. agarici* NCPPB 2472.

Location:	Locus Tag:	Gene:	Protein ID:
1171497-1172474	RS05180	Type I-F CRISPR-associated endonuclease Cas1	WP_060782314.1
1172471-1175881	RS05185	Type I-F CRISPR-associated helicase Cas3	WP_060782315.1
1175960-1177330	RS05190	Type I-F CRISPR-associated protein Csy1	WP_060782316.1
1177323-1178306	RS05195	Type I-F CRISPR-associated protein Csy2	WP_060782317.1
1178340-1179365	RS05200	Type I-F CRISPR-associated protein Cas7f/Csy3	WP_060782318.1
1179369-1179932	RS05205	Type I-F CRISPR-associated endoribonuclease Cas6/Csy4	WP_060782319.1

The Cas proteins identified in Table 3.11 are characteristic of a Type 1 Subtype I-F CRISPR system, similar to that reported in *Pseudomonas aeruginosa* PA14 by Rollins *et al.* [239], which require a CRISPR RNA-guided surveillance complex (also known as a Csy complex), in this case formed of the proteins RS05190-RS05205, to recruit the Cas2/3 (RS05185) *trans*-acting nuclease for degradation of the target DNA.

The repeat consensus sequence associated with the CRISPR region was identified as being 28bp in length composed of the sequence: GTTCACTGCCGTATAGGCAGCTCAGAAA.

In total 34 spacer regions were identified of 32bp in length which are shown in Table 3.12.

Table 3.12. CRISPR-associated repeat region spacer sequences of *P. agarici* NCPPB 2472.

No (5'-3'):	Sequence
1	TCGCGGATTGGATGGATTAGCGGTAGACGCCC
2	TGCCGTCAGCGCTAACACCATTCAGAACGCGT
3	ACTTGAGCCCGGCGAAATTGTGCACCACAAAA
4	ATGCCGCGGTCTGAACCAGCGTTGCTTGAAGG
5	TGCCAGTAATCGGCTAAGCGGTAGTGCATCGG
6	TACTACCGGCGACCCAGTAACCGATGGTATCC
7	GGTTAAAAGGATGATCCTTTTGATCTTGGTAT
8	TCCTTGCGAATCGCGGGTTGGAGTACCACACG
9	ATAAATCCAGCAACACCAGAAGGCCCGGCTCT
10	TATCTGCGCGAGGAAAGCAGCTATACGCAGCC
11	TGCTGGAGATTAAATAATGACTGGTAATAACT
12	TCAACGTCGCCGGCTTCTTCATGAATGCCGAC
13	TTGGTGGTGTCCAGTGGGATGCGCAACTGTTG
14	TTCAGGCGGGGCGCCAAACGCTGACGCCCTAC
15	GCATCGCCTACTTTGCGCACGGCGGTCAATAC
16	ACCACGTTCAAGAGCTGTTGCAGGGCGTCTCG
17	TTGTCGACGAACCACATTTGCCCCCACAGATC
18	ATACGGCGGACTTCGCTGCCCATTCCGGTTGC
19	GCCGAGAGCATCGACATTATTGAAGCACTGGG
20	TGAACGCCGCAGCCAAGTCGATGAAGCTCATG
21	GCGAAGGCCGCAGGCGAGATAGTCATGTGGCAT
22	TTTGCACTGGCTTGAGATACTCCTGCCAAACC
23	GGAGTACACGCAGGACGCTGTGGTCACGACTGA
24	TTCACCTCAAACTCGCCCTGGATCTCATCCAT
25	TTCCGTCGGCTCCGGACATCCCTACGGACTCGA
26	TTCACCGCTCAGTCGGTCAAGGCTCTGACCAA
27	TGGCCGACGGCTCGTCATCGTCGCCTGGAAGG
28	AGGCCATGAACAAGAAAAACGTTCAACTCAAG
29	AAGAAAGGCAAGAAGGGCAAGAAGGGTGGACG
30	GTCGGGCTCAGCGATCAAAACAACTTCAGGAC
31	TGGGATCGCGAGATTGAAGCGCACGCCTTCAC
32	GAAACGAAGCGCGATTTGTCTGGCATGAGCCT
33	GTGATGGTTGAGCCGGCACGACCGCGCTGACC
34	TCGGGTGCGCGGGGTGGTGAAGCTACCCCACA

Only one spacer sequence, No. 34, showed any nucleotide identity to a known sequence as identified by BLASTN [253], although a single base substitution was present of the first G to

an A in the matched sequences. Three matches were found corresponding to plasmid sequences; a plasmid present in *P. frederiksbergensis* strain AS1, *P. fluorescens* strain PC20 (plasmid pNAH20) and *P. putida* NCIB 9816-4 (plasmid pDTG1) and the single base pair substitution was conserved in all matches. These plasmids were found to encode naphthalene degrading enzymes, including the plasmid originating from *P. frederiksbergensis* strain AS1 which has been demonstrated to be a naphthalene-degrading bacterium [277]. None of the spacer regions showed similarity to the sequences of the phage identified during this study.

Restriction Endonuclease Systems

In total 7 ORFs that were predicted to encode restriction endonucleases were identified within the *P. agarici* 2472 genome. These are outlined in Table 3.13.

Table 3.13. Predicted restriction endonuclease proteins of *P. agarici* NCPPB 2472.

Protein:	Start:	End:	Locus Tag:
Restriction endonuclease	<1266141	1266375	RS05570
Restriction endonuclease subunit R	1277574	1280486	RS05645
Restriction endonuclease	4212986	4213897	RS18565
Type I restriction endonuclease subunit R	4213907	4217116	RS18570
Restriction endonuclease subunit S	4217117	4217992	RS24540
Restriction endonuclease subunit S	4239877	4241145	RS18660
Type I restriction endonuclease subunit R	4242778	4245888	RS18670

Of these 7 ORFs, 3 were predicted to encode the restriction (R) subunits (RS05645, RS18570, RS18670) and 2 to encode the specificity (S) subunits (RS24540, RS18670).

Two SAM-dependent DNA methyltransferase encoding ORFs were identified that are associated with 4 of the restriction endonuclease ORFs identified, one (RS18580) which is

associated with the restriction endonuclease subunits RS18565 and RS18570, and a second (RS18665) which is located between the S and R restriction endonuclease subunits RS18660 and RS18670.

Alginate

I have identified an operon comprised of 12 genes within the genome of *P. agarici 2472*, shown in Table 3.14, hypothesised to be involved in the production of alginate.

All 12 ORFS within this operon are orthologous to the 12 proteins (AlgD, Alg8, Alg44, AlgK, AlgE, AlgG, AlgX, AlgL, AlgI, AlgJ, AlgF, and AlgA) required for alginate production and export in *Pseudomonas* [276].

Table 3.14. Predicted alginate biosynthesis proteins of *P. agarici* NCPPB 2472.

ORF No.	Predicted Function:	Locus Tag:	Ortholog:
1	GDP-mannose 6-dehydrogenase	RS03600	AlgD
2	glycosyl transferase	RS03595	Alg8
3	alginate biosynthesis protein Alg44	RS03590	Alg44
4	alginate biosynthesis protein	RS03585	AlgK
5	alginate biosynthesis protein AlgE	RS03580	AlgE
6	poly(beta-D-mannuronate) C5 epimerase	RS03575	AlgG
7	alginate O-acetyltransferase	RS03570	AlgX
8	mannuronate-specific alginate lyase	RS03565	algL
9	poly(beta-D-mannuronate) O-acetylase	RS03560	AlgI
10	alginate O-acetyltransferase	RS03555	AlgJ
11	alginate O-acetyltransferase	RS03550	AlgF
12	mannose-1-phosphate guanylyltransferase/mannose-6-phosphate isomerase	RS03545	AlgA

3.4 *Pseudomonas sp. NS1(2017)*

The strain that I have provisionally named '*Pseudomonas* sp. NS1(2017)' was first identified from plate counts of *P. tolaasii* washed from the surface of *Agaricus bisporus* mushrooms showing no previous symptoms of disease, where it was found to both bear a strikingly similar colony morphology to *P. tolaasii* and to rapidly outgrow *P. tolaasii* NCPPB 2192T cultures on KB agar, further analysis revealed it to also be Gram negative.

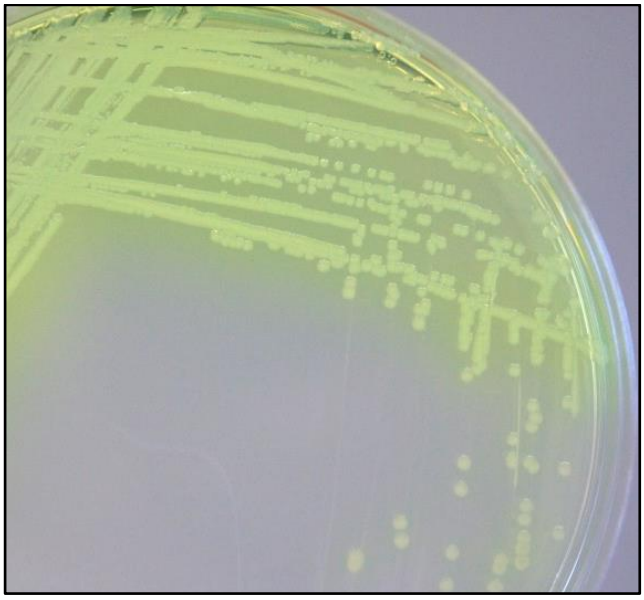

Figure 3.12. Streak plate of *Pseudomonas* sp. NS1 on KB agar after 72 hours at 28°C.

Pseudomonas sp. NS1 was observed to produce a fluorescent pigment on KB agar similar to that of *P. tolaasii*, as shown in Figure 3.12. It was hypothesised from the plaque morphology and Gram staining results that *P.* sp NS1 may be a member of the fluorescent Pseudomonads or potentially a faster growing strain of *P. tolaasii* and so further investigation was undertaken.

3.4.1 Phylogeny

To infer the phylogenetic relationship of *Pseudomonas sp.* NS1 among the *Pseudomonas* genus and assist in taxonomic placement, a Maximum-Likelihood analysis was performed on the DNA sequence returned from whole-genome sequencing using 5 housekeeping genes (23s *rlmJ*, 16S *rrs*, *rpoD*, *gyrB* and *rpoB*) of 14 *Pseudomonas* strains, as well as *E. coli* K12 for

rooting the phylogenetic tree, using the Tamura-Nei model [261]. The final tree with the highest log likelihood (-55462.6960) is shown in Figure 3.13.

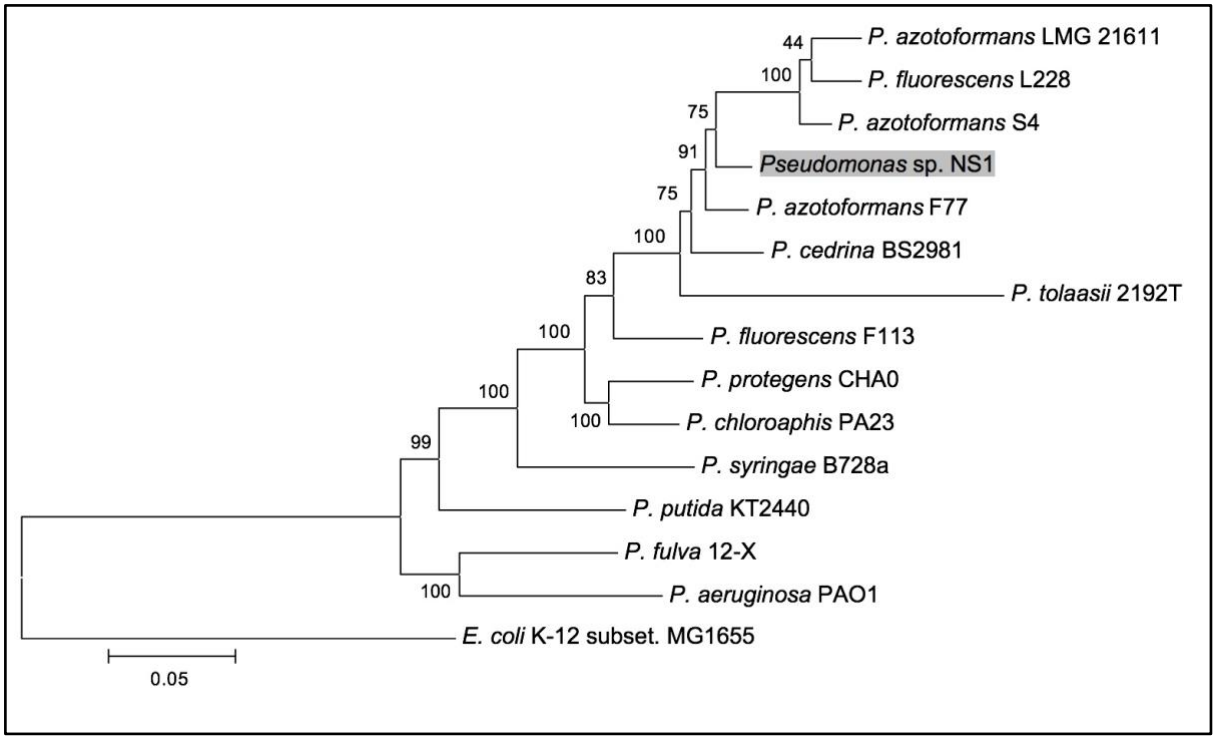

Figure 3.13. Molecular phylogenetic analysis of *Pseudomonas* sp. NS1 (highlighted) by Maximum Likelihood method. The percentage of trees in which the associated taxa clustered together is shown next to the branches. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 15 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 9210 positions in the final dataset. Evolutionary analyses were conducted in MEGA6.06.

As shown in Figure 3.13, *Pseudomonas sp.* NS1 is clustered among members *Pseudomonas azotoformans* species, however it is on a separate branch to the closely related *P. azotoformans S4*. It is therefore likely that *P. sp* NS1 forms part of the *P. fluorescens* lineage of the *P. fluorescens* complex of the *Pseudomonas* intrageneric cluster II as reported by Yamamoto *et al.* [272].

3.4.2 Sequencing and genome organisation

In total, the Illumina MiSeq® full genome sequencing run performed by Source Bioscience of *P.* sp NS1 returned 527.6 Mbases with a coverage of approximately 75 fold. The reads were assembled in to 546 contigs by Source Bioscience, with a total sequence length of 7,069,489 bp. The contigs returned by Source Bioscience were mapped to the genome of *Pseudomonas azotoformans* S4 using CONTIGuator 2.7.4 [251] which returned 83 contigs, containing a total of 6,702,516 bp. The final mapped contigs were aligned to the genome using ACT, the results are shown in Figure 3.14.

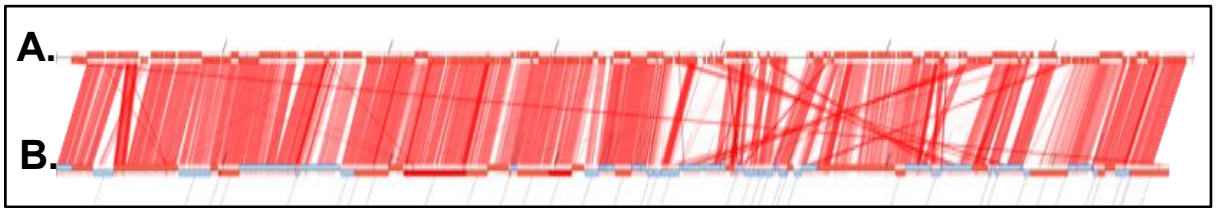

Figure 3.14. Alignment of *Pseudomonas* sp NS1 contigs (B.) to the genomic sequence of *Pseudomonas azotoformans* S4 (A.), figure generated by CONTIGuator 2.7.4 [251].

As is evident in Figure 3.14, the mapped contigs of *P.* sp NS1 and *P. azotoformans* S4 show large areas of collinear similarity, as indicated by red bars, this is borne out by the total nucleotide identity as determined by BLASTN [253] of 98% over 84% of the total assembled genome.

The returned contigs, in a single scaffold, were annotated using the NCBI Prokaryotic Genome Annotation Pipeline, the full results are available under the accession code CP022960.

Of the 6,241 ORFs identified, 6,073 were identified as being protein coding and 72 RNA encoding genes were identified; including 60 tRNAs, 8 complete rRNAs (5S, 16S, 23S) and 4 ncRNAs. The total gene count is lower than the count of the closely related *P. azotoformans* S4 (6,241 vs. 6,324) and *P. fluorescens* LMG 5329 (6,241 vs. 6,429), correlating with the lower overall genome size.

Table 3.15. Comparison of the genomes and predicted genes of *P.* sp NS1, *P. azotoformans* S4 and *P. fluorescens* LMG 5329.

Strain	Total (bp)	ORFs	Encoding	Genes (RNA)	tRNA	rRNA	ncRNA	GC (%)
NS1(2017)	6,702,516	6,241	6,073	72	60	8	4	61.08
S4	6,859,618	6,324	5,991	94	70	19	5	60
LMG 5329	6,870,020	6,429	6,216	63	57	2	4	60.5

A full circular genome map was created using the DNAPlotter function of Artemis v16.0.0.1 in order to visualise the returned data and is shown in Figure 3.15.

To further aid in taxonomic classification and to calculate the probability that *P.* sp NS1 was a strain of either *P. azotoformans* or *P. fluorescens* the full draft genome was analysed using the JSpecies [278] work package. The results of the pairwise comparison of the draft genome of *P.* sp NS1 using average nucleotide identity showed that it was most closely related to *P. fluorescens* LMG 5329 and *P. azotoformans* S4 with scores of 93.43% and 93.32% respectively (TETRA score 0.99858 and 0.99892 respectively). Full results of all analysis conducted are available in Appendix 4 Tables Ap4.2 and Ap4.3.

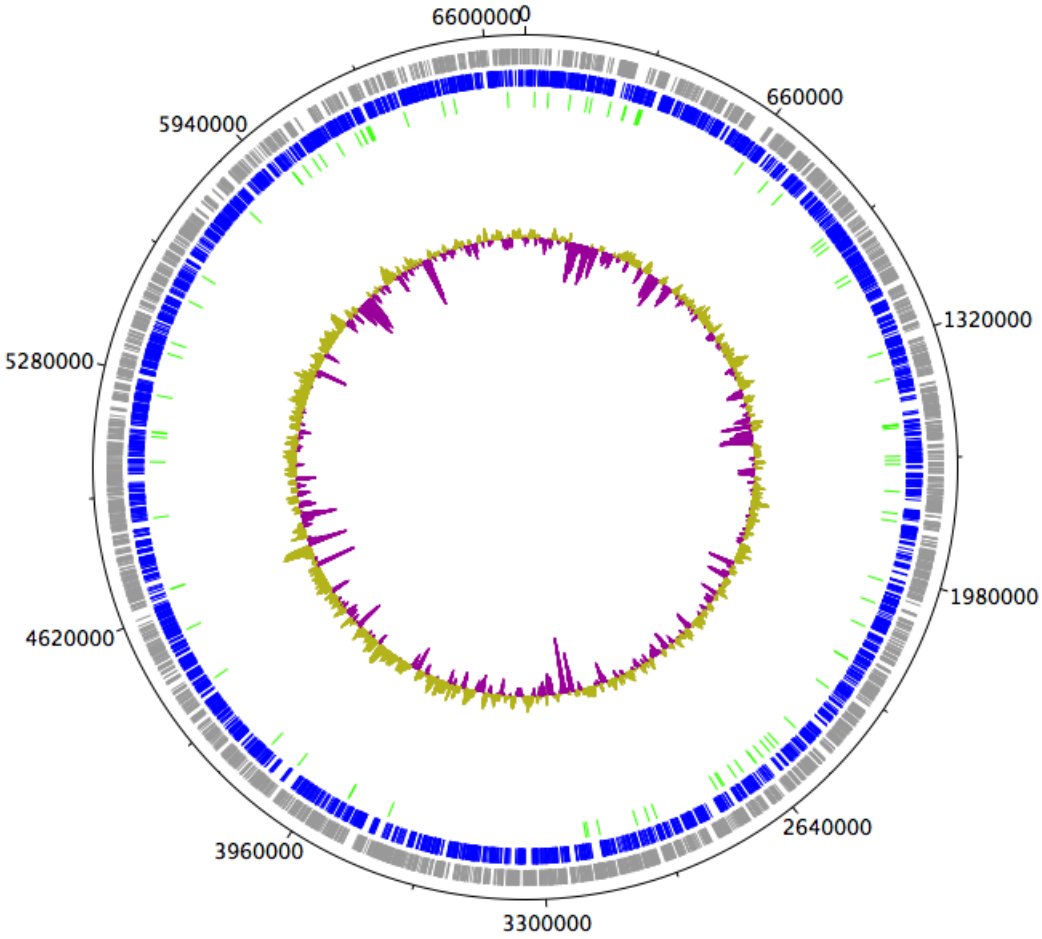

Figure 3.15. Circular map of the *Pseudomonas* sp. NS1 genome with the genes encoded on the positive strand in the first lane (grey), the negative strand in the second lane (blue) and the tRNA genes in the third lane (green). The deviation in GC content from the total average is illustrated in the centre circle with yellow bars denoting an above average content and purple bars, a below average GC content. Figure generated using DNAPlotter function of the Artemis v16.0.0.1 genome browser.

3.4.3 Secondary metabolites and NRPSs

In total 10 complete or partial NRPSs were identified within the genome of *P.* sp NS1, with 8 of these located in two distinct clusters. A NRPS containing biosynthetic gene cluster containing 6 whole and partial NRPSs as shown in Table 3.16, was predicted by antiSMASH v3.0 [262] to potentially produce the lipopeptide poaeamide.

Table 3.16. Predicted proteins of the WLIP production cluster 2 of *P.* sp. NS1.

ORF No.	Predicted Function:	Locus Tag:
1	MATE family efflux transporter	CI807_19185
2	TetR/AcrR family transcriptional regulator	CI807_19190
3	MFS transporter	CI807_19195
4	toxin-antitoxin system HicB family antitoxin	CI807_19200
5	type II toxin-antitoxin system HicA family toxin	CI807_19205
6	non-ribosomal peptide synthetase	CI807_19210
7	peptide synthase	CI807_19215
8	non-ribosomal peptide synthetase	CI807_19220
9	non-ribosomal peptide synthetase	CI807_19225
10	non-ribosomal peptide synthetase	CI807_19230
11	TonB-dependent siderophore receptor	CI807_19235
12	cyclic peptide transporter	CI807_19240
13	N(5)-hydroxyornithine transformylase PvdF	CI807_19245
14	chromophore maturation protein PvdO	CI807_19250
15	aminotransferase	CI807_19255
16	peptidase M19	CI807_19260
17	PvdJ/PvdD/PvdP-like protein	CI807_19265
18	non-ribosomal peptide synthetase	CI807_19270
19	non-ribosomal peptide synthetase	CI807_19275
20	macrolide transporter subunit MacA	CI807_19280
21	MacB family efflux pump subunit	CI807_19285
22	LuxR family transcriptional regulator	CI807_19290
23	methionine gamma-lyase	CI807_19295
24	Lrp/AsnC family transcriptional regulator	CI807_19300

However, BLASTN [253] analysis of the NRPSs within the cluster showed that it has a significantly higher nucleotide identity to the lipopeptide production system gene cluster for white line-inducing principle (WLIP) of *Pseudomonas fluorescens* strain LMG 5329 cluster 2.

This would indicate that it is more likely this gene cluster encodes the WLIP, a lipopeptide of the viscosin group, than poaeamide [279].

A second NRPS (CI807_26350) similarly identified by antiSMASH v3.0, when analysed by BLASTN [253] showed significant nucleotide identity to the WLIP production cluster 1 of *Pseudomonas fluorescens* LMG 5329. However, small amino acid differences in the sequence of the NRPSs can confer significant changes in function of the final peptide, therefore sequence alone is not enough to conclude that these NRPS clusters produce WLIP [279].

Further *in silico* analysis showed that the NRPSs in the initial cluster, as shown in Table 3.16, not only showed nucleotide identity, but also a conserved pattern of gene encoding ORFS when compared to the WLIP production cluster 2 of *Pseudomonas fluorescens* strain LMG 5329. A direct visual comparison of gene organisation in the identified cluster of *Pseudomonas* sp. NS1 and the WLIP production cluster 2 of *Pseudomonas fluorescens* strain LMG 5329 is shown in Figure 3.16.

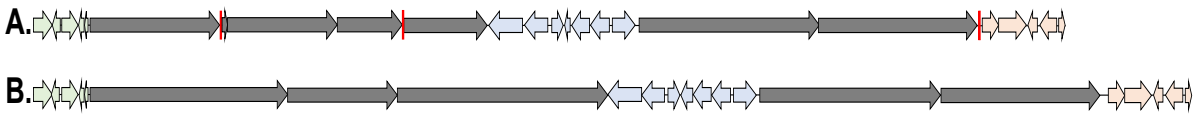

Figure 3.16. Diagram of gene organisation of WLIP production cluster 2 of A) *Pseudomonas* sp. NS1 and B) *Pseudomonas fluorescens* strain LMG 5329; NRPSs illustrated in grey, 'Nx100' regions illustrated in red, ORFS 1-5 of *P.* sp NS1 illustrated in green, ORFS 11-17 of *P.* sp NS1 in blue and ORFS 20-24 of *P.* sp NS1 in orange with their corresponding orthologs in B) highlighted using the same colour scheme. Figure generated using Microsoft PowerPoint.

The first NRPSs within the cluster (ORFS 6-10) are incomplete due to regions of 100 N's inserted during the assembly of the contigs are shown in Figure 3.16, line A. as red bars and appear to be fragmented parts of the 3 large NRPSs found in *P. fluorescens* LMG5329 in Figure 3.16 B.

The ORFS 18 and 19 in Table 3.16, was predicted to putatively encode the WLIP synthetase B and C respectively. ORF 18 (CI807_19270) shows 97% amino acid identity to the WLIP synthetase B NRPS of *P. fluorescens* LMG 5329 and ORF 19 (CI807_19275) shows 96% amino acid identity to WLIP synthetase C NRPS of *P. fluorescens* LMG 5329. As shown in, Figure 3.16, ORFS 18 and 19 are flanked by pyoverdine synthesis genes upstream (illustrated in blue) and macrolide transporters downstream (illustrated in orange) an organisation pattern which has been identified by Rokni-Zadeh *et al.* [279] to be conserved in both the WLIP NRPS system of *P. fluorescens* strain LMG 5329 and the viscosin system of *Pseudomonas fluorescens* SBW25.

In order to determine if *Pseudomonas* sp. NS1 produced WLIP, *P.* sp NS1 was cultured alongside *P. tolaasii* 2192T; a distinct white precipitate formed within the agar, shown in Figure 3.17, which confirms that the biosynthetic gene clusters identified most likely encode the NRPS system for WLIP production rather than any other of the viscosin-related nonapeptides.

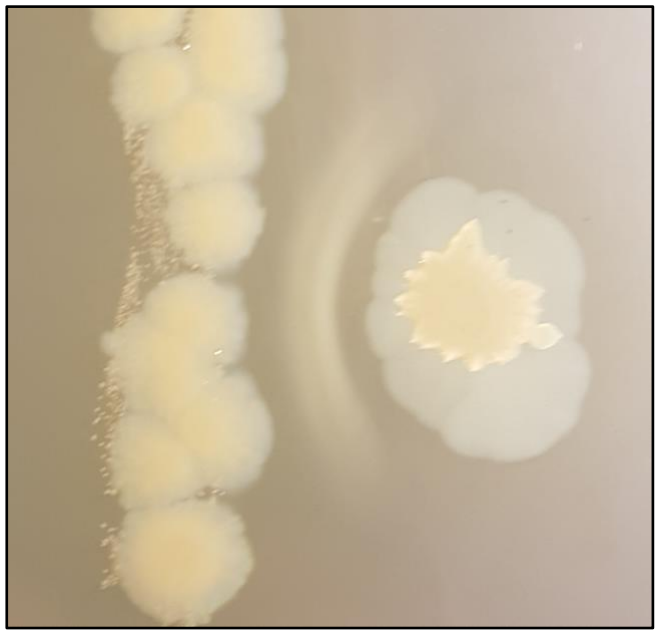

Figure 3.17. Co-culturing of *P. tolaasii* NCPPB 2192T, left of image, and *Pseudomonas sp.* NS1, right of image, on KB agar after 72 hours showing white co-precipitate of tolaasin and WLIP.

3.4.4 Phage resistance systems

Alginate

Similar to both *P. tolaasii* 2192T and *P. agarici* 2472, an operon predicted to encode the genes required for alginate biosynthesis was identified within the genome of *P.* sp. NS1 that could possibly act to mask cell surface receptors from phage. In comparison to the operon in *P. agarici* which contained predicted proteins orthologous to the 12 proteins required for alginate synthesis, the operon in *P.* sp. NS1 contains 13 predicted proteins, shown in Table 3.17, and lacks a protein orthologous to AlgE but also contains an additional hypothetical protein in ORF position 12.

Table 3.17. Predicted alginate biosynthesis proteins of *P.* sp. NS1.

ORF No.	Predicted Function:	Locus Tag:	Ortholog:
1	nucleotide sugar dehydrogenase	CI807_11825	AlgD
2	glycosyl transferase	CI807_11820	Alg8
3	hemolysin D	CI807_11810	Alg44
4	alginate biosynthesis protein	CI807_11805	AlgK
5	transcriptional regulator	CI807_11800	/
6	poly(beta-D-mannuronate) C5 epimerase	CI807_11795	AlgG
7	alginate O-acetyltransferase	CI807_11790	AlgX
8	mannuronate-specific alginate lyase	CI807_11785	AlgL
9	MBOAT family protein	CI807_11780	AlgI
10	alginate O-acetyltransferase	CI807_11775	AlgJ
11	alginate O-acetyltransferase	CI807_11770	AlgF
12	hypothetical protein	CI807_11765	/
13	mannose-1-phosphate guanylyltransferase/mannose-6-phosphate isomerase	CI807_11760	AlgA

Restriction Endonuclease Systems

No CRISPR-associated repeats or proteins were identified in within the genome of *P.* sp. NS1, however 4 predicted restriction endonuclease proteins were identified, as shown in Table 3.18.

Table 3.18. Predicted restriction endonuclease proteins of *P.* sp. NS1.

Protein:	Start:	End:	Locus Tag:
restriction endonuclease subunit M	3169222	3170016	CI807_15035
restriction endonuclease subunit S	4531120	4531992	CI807_21265
restriction endonuclease (McrC)	5903408	5904724	CI807_27500
restriction endonuclease (McrB)	5904721	5906754	CI807_27505

The third and fourth proteins (CI807_27500 and CI807_27505) contain McrC and McrB conserved domains respectively, the McrB conserved domain is associated with the GTP-binding regulatory subunit of the methylcytosine specific McrBC restriction endonuclease of *E. coli* K-12 [273]. As the McrBC restriction endonuclease cleaves foreign methylated DNA, it does not possess an associated methyltransferase [280].

3.5 Discussion

The *Pseudomonas* genus is large, complex, and rapidly expanding, with a current total of 238 species and 18 subspecies recognised and published in the List of Prokaryotic Names with Standing in Nomenclature [281]; a significant increase from 144 species and 10 subspecies recognised in 2014 [282]. Since the genus was first described in 1894 [283] based on morphology and metabolism, the taxonomic status has undergone many revisions with the advent of new technologies, such as the use of nucleic acid homologies [284] and 16S rRNA gene sequences [13, 23].

The use of 16S rDNA (*rrs* gene) alone to resolve phylogenetic relationships, however, is limited for a variety of reasons. These include the intraspecies heterogeneity of *rrs* sequences frequently resulting from multiple copies of the *rrs* gene within a genome [285], which can impact the accuracy of sequencing; and also the lack of *rrs* sequence diversity within closely related strains or even species due to the slow evolution of the *rrs* gene [286].

Instead, a multilocus sequence analysis (MLSA) method using several genes of varying rates of evolution, including protein encoding genes that are reported to evolve faster than rRNA [287], can be used to infer phylogeny. It has previously been demonstrated that the housekeeping genes *rrs*, *gyrB* and *rpoD* are suitable candidates for accurate phylogenetic and taxonomic analysis for both Pseudomonads and other bacterial genus [272, 288, 289].

For this study, a single concatenated sequence comprised of a single rRNA gene, *rrs*, as well as the *rpoD*, *rpoB*, *RlmJ* and *gyrB* genes was used to provide both a slow-evolving rDNA sequence and fast-evolving protein-encoding DNA sequence to most accurately analyse the

phylogenetic relationship of *P. tolaasii* 2192T both among the *Pseudomonas* genus and within other closely related *P. tolaasii* strains.

The *gyrB* gene encodes one of two subunits (*gyrA*/*gyrB*) of the prokaryotic DNA gyrase protein, which is responsible for general DNA supercoiling as well as chromosome partitioning and is ubiquitous among bacteria [290]. The *rpoD* gene encodes the RNA polymerase σ^{70} factor, that allows promoter-specific transcription initiation of RNA polymerase [291]. The *rpoB* gene encodes the β-subunit of the DNA dependent RNA polymerase [292]. The *RlmJ* gene encodes the methyltransferase RlmJ which catalyses the methylation of 23S rRNA during ribosome biogenesis and is ubiquitous among bacteria [293].

In the case of *P. tolaasii* 2192T, the results of the phylogenetic analysis shown in Figure 3.2 agree with the phylogeny reported by Yamamoto *et al.* [272] which utilised a more limited gene selection (*gyrB* and *rpoD* alone); with *P. tolaasii* 2192T clustering with the closely related *P. azotoformans* and *P. fluorescens* which are all members of the *Pseudomonas* intragenic cluster II, *P. fluorescens* complex, *P. fluorescens* lineage. However, it is unusual that *P. tolaasii* 2192T is located on a separate branch to *P. tolaasii* 6264 and PMS117 indicating, as previously mentioned, that it may have diverged earlier from the other members of the *P. tolaasii* species. Likewise, with *P. agarici* NCPPB 2472 the phylogenetic results attained in this study, shown in Figure 3.5, confirm the previous work on phylogenetic relationships of *P. agarici* strains by Yamamoto *et al.* [272]; with *P. agarici* NCPPB 2472 showing tight clustering with *P. agarici* strains NCPPB 2280 and LMG 2112 in a monophyletic branch within the *P. chloroaphis* lineage, itself within the *P. fluorescens* complex of the *Pseudomonas* intrageneric cluster II.

The agreement of the results of the phylogenetic analysis in this study of both *P. tolaasii* 2192T and *P. agarici* NCPPB 2472 with the results previously published by Yamamoto *et al.* [272] increases confidence in the results of the analysis performed on *P.* sp. NS1, shown in Figure 3.13, which show that *P.* sp. NS1 clusters with members of the *P. azotoformans* species, within part of the *P. fluorescens* lineage of the *P. fluorescens* complex of the *Pseudomonas* intrageneric cluster II as reported by Yamamoto *et al.* [272]. To further aid in taxonomic classification and to calculate the probability that *P.* sp. NS1 was a strain of either *P. azotoformans* or *P. fluorescens* the full draft genome was analysed using the JSpecies [278] work package, a form of in silico DNA-DNA hybridisation, which identified that *P.* sp. NS1 was most closely related to *P. fluorescens* LMG 5329 and *P. azotoformans* S4, with scores of 93.43% and 93.32% respectively (TETRA score 0.99858 and 0.99892 respectively). From this we can deduce that it is likely that *P.* sp. NS1 is a novel strain of either the *P. azotoformans* or *P. fluorescens* species.

In order to fully characterise the *Pseudomonas* species used in this study at the genetic level, full genome sequencing was performed, and the returned sequences assembled in to scaffolds based on mapping of the contigs returned by Source Bioscience to closely related species. In the case of *P. agarici* NCPPB 2472, contig mapping was performed utilising both *P. agarici* NCPPB 2289 and *P. fluorescens* A506, to provide two data sets from which to design PCRs in order to confirm contig ordering and 'stitch' the remaining contig junctions with sequence from amplified inter-contig regions. The initial contig mapping to both *P. fluorescens* A506 and *P. agarici* NCPPB 2289, illustrated in Figures 3.6 and 3.7 shows that *P. agarici* NCPPB 2472 shares significantly larger areas of collinear similarity with *P. agarici* NCPPB 2289. However, during the process of stitching these contigs via PCR and Sanger sequencing the resultant products, the results of which are shown in Figures 3.9 and 3.10, it was clear that the reactions

based on contig mapping to *P. fluorescens* A506 returned a larger number of significantly clearer bands on agarose gel than the reactions based on the contig mapping to *P. agarici* NCPPB 2289. This is further evident in the number of the reactions that returned valid sequencing data, with the reactions based on mapping to *P. fluorescens* A506 returning 8 sequenced products from 30 reactions (26%), compared to 3 sequenced products from the 46 reactions based on alignment to *P. agarici* NCPPB 2289 (7%). Difficulties were encountered due to a large number of self-priming primers during the contig stitching PCRs, possibly due to the presence of inverted repeats within the contig junction sequence. The results of the contig stitching PCRs would appear to show that the genome arrangement of *P. agarici* NCPPB 2472 is more similar to that of *P. fluorescens* A506 than to the closely related strain *P. agarici* NCPPB 2289, despite the larger regions of collinear similarity.

The assembled genomic sequences of *P. tolaasii* 2192T, *P. agarici* NCPPB 2472 and *P.* sp. NS1 were then mined for potential biosynthetic gene clusters involved in the production of virulence factors, including toxins such as tolaasin [44], or siderophores such as pyoverdines [33]. Likewise, the genomic sequences were analysed for phage resistance mechanisms that may impede the use of bacteriophage as a potential treatment of infections of Agaricus bisporus mushrooms caused by either *P. tolaasii* or *P. agarici*.

Within the genome of *P. tolaasii* NCPPB 2192T I identified 6 ORFS, shown in Table 3.2, with significant amino acid identity (72-86%) to the proteins TaaA-TaaE of *P. costantinii* which have been demonstrated by Scherlach *et al.* [50] to be involved in the production of Tolaasin I. *P. tolaasii* NCPPB 2192T has been reported to produce Tol I Tol II and Tol A-E by Bassalero *et al.* [47], therefore we can conclude that based on the amino acid identity and the similarities in gene organisation outlined that is likely the NRPSs identified in Table 3.2 are involved in

the production of Tolaasin I. However, the biosynthetic cluster identified by Scherlach *et al.* [50] occurs as a single cluster whereas the NRPSs identified in this study are located sporadically throughout the genome of *P. tolaasii* 2192T and also include at least one NRPS that has been incorrectly assembled (RS02650); it is likely that this is due to errors in the initial contig assembly stage of the full genome sequencing caused by the repetitive nature of NRPS sequence and that the NRPSs identified should occur as a single cluster.

Similarly, two of the further NRPSs identified within the genome of *P. tolaasii* 2192T, RS16165 and RS16170 were initially split during *in silico* assembly of the returned contigs but were found to be likely to instead consist of a single large NRPS. This larger NRPS of approximately 1751 aa in total showed amino acid identity, as identified by BLASTP [253], to the known pyoverdine biosynthesis genes PvdJ of *Pseudomonas syringae* pv. daphniphylli and the pyoverdine sidechain peptide synthetase of *P. fluorescens*. Likewise, an ORF downstream of these NRPSs was also found to show amino acid identity to the PvdJ protein of *P. aeruginosa* 62. Pyoverdines are fluorescent siderophores that are important virulence factors in many fluorescent *Pseudomonas* species [33, 294]. The presence of a TonB-dependent siderophore upstream of these putative NRPSs and their similarity to known pyoverdine synthesis proteins indicates this cluster is likely to be involved in the synthesis of pyoverdine or related chromophore.

Within the genome of *P. agarici* NCPPB 2472 I identified the presence of multiple clusters of NRPSs, including a single large cluster comprised of 5 NRPSs outlined in Table 3.8, as well as two smaller clusters composed of a single NRPS each. However, it was not possible to predict the function of the NRPS containing biosynthetic gene clusters via *in silico* analysis of the nucleotide or amino acid composition, although it is likely that they produce a siderophore. However, a large biosynthetic cluster was identified within the genome of *P. agarici* NCPPB

2472, shown in Table 3.10, that is predicted to encode the proteins involved in the synthesis of the siderophore achromobactin. Achromobactin production was first identified by Munzinger *et al.* [295] in the bacterium formally known as *Erwinia chrysanthemi* (now *Dickeya dadantii* [296]), as a siderophore derived from the precursor citrate. Berti and Thomas [61] reported in 2009 that *P. syringae* pv. syringae B728a is capable of achromobactin synthesis and that it is produced via NRPS-independent synthetases; demonstrating that the enzymes AcsD, AcsA and AcsC are capable of converting citrate to achromobactin. They have likewise demonstrated that a 12 ORF region of the *P. syringae* pv. syringae B728a genome is responsible for the biosynthesis of achromobactin [61]. I have identified a biosynthetic gene cluster in *P. agarici* NCPPR 2472, as shown in Table 3.10, that is predicted to encode the enzymes required for achromobactin synthesis (AcsD, AcsA and AcsC). This gene cluster shows striking similarity in ORF organisation to the cluster of *P. syringae* pv. syringae B728a, as shown in Figure 3.11, as well as a high degree of amino acid identity between the predicted gene products, ranging from 72-91% (by comparison the amino acid identity between the achromobactin synthesis proteins of *P. syringae* pv. syringae B728a and *Dickeya dadantii* is reported as between 64-72% [61]). I therefore suggest that *P. agarici* NCPPB is likely to produce the siderophore achromobactin under iron-limited conditions. Both *Dickeya dadantii* and *P. syringae* pv. syringae B728a produce a second siderophore in addition to achromobactin, in the case of *Dickeya dadantii* this is chrysobactin and for *P. syringae* pv. syringae B728a it is pyoverdine [61, 295, 297]. It has been suggested that in both cases that the primary siderophore, achromobactin, is required for epiphytical growth and that their secondary siderophores are reserved for active infection [61, 298, 299]. It is therefore possible that one of the identified NRPS clusters of *P. agarici* NCPPB 2472 is involved in production of a siderophore, such as pyoverdine, utilised in active infection of *Agaricus bisporus* and that the predicted achromobactin siderophore is utilised during non-infective growth in the environment.

Within the genome of *P.* sp NS1 I identified 10 complete or partial NRPSs, 8 of which reside in two distinct clusters. A single cluster containing 6 NRPSs was identified, shown in Table 3.16, which showed nucleotide identity to the WLIP production cluster 2 of *Pseudomonas fluorescens* strain LMG 5329. WLIP, a subtype of the viscosin-related nonapeptides [41], produces a precipitate in agar medium when a WLIP producing *Pseudomonas* strain is co-cultured alongside a strain of tolaasin producing *Pseudomonas tolaasii*; however minor differences in amino acid sequence can confer significantly differing biological properties, for example viscosin, which differs from WLIP in only one amino acid residue fails to produce a precipitate when co-cultured alongside *P. tolaasii* [42].

A second, ORFan, NRPS (CI807_26350) was identified that showed nucleotide identity to WLIP production cluster 1 of *Pseudomonas fluorescens* LMG 5329. It was then further identified that the 6 NRPS containing cluster with similarity to the WLIP production cluster 2 of *Pseudomonas fluorescens* strain LMG 5329 also shared a conserved gene organisation, as illustrated in Figure 3.16, which has been identified by Rokni-Zadeh *et al.* [279] to be conserved in both the WLIP NRPS system of *P. fluorescens* strain LMG 5329 and the viscosin system of *Pseudomonas fluorescens* SBW25.

The conserved gene organisation and further amino acid identity analysis which identified high levels of similarity between the NRPSs within the *P.* sp. NS1 cluster and the WLIP synthetases of *P. fluorescens* LMG 5329 led to the hypothesis that the identified gene cluster was likely to encode the proteins required for WLIP synthesis. This hypothesis was tested by co-culturing *P.* sp. NS1 alongside *P. tolaasii* 2192T, revealing a white precipitate as illustrated in Figure 3.17, confirming that *P.* sp. NS1 is capable of WLIP production and that is most likely produced by the NRPSs and associated proteins identified in Table 3.16.

Although a wide variety of secondary metabolites involved in host colonisation and invasion by *Pseudomonas* species are produced by NRPS's, many are produced by more traditional assembly routes; therefore, further biosynthetic clusters that may confer advantages on the host surface were identified and characterised, including putative phage tail-like bacteriocin (PTLB) encoding genes. PTLBs are bacteriocins of which there are three broad types, F, R and S [68]. Both the F and R type PTLBs resemble bacteriophage tail proteins, with the R type being most similar to contractile phage tails [300, 301] and the F type being flexible non-contractile rods similar to flexible phage tails [302] and are likely both of phage origin [303]. PTLBs are very common in *Pseudomonas* strains, with 90% *P. aeruginosa* strains being reported to produce either the F or R type [304]. Putative PTLB encoding operons were identified in both *P. tolaasii* 2192T and *P. agarici* NCPPB 2472, shown in Table 3.3 and 3.9 respectively, with the operon for *P. agarici* NCPPB 2472 consisting of a total of 19 ORFs and the smaller operon for *P. tolaasii* NCPPB 2192T consisting of only 13 ORFs in comparison.

As previously mentioned the genomes of *P. tolaasii* NCPPB 2192T, *P. agarici* NCPPB 2472 and *P.* sp. NS1 were also mined for the presence of phage resistance genes that may prevent or hinder the possible use of bacteriophage in the treatment of *Agaricus bisporus* mushrooms.

Within the genome of *P. agarici* NCPPB 2472 I have identified multiple phage resistance systems, including a complete CRISPR/Cas system. No confirmed CRISPRs or CRISPR-associated proteins were identified in the genomes of the closely related *P. agarici* NCPPB 2289, *P. agarici* LMG 2112 or within either of the genome sequences of *P. tolaasii* NCPPB 2192T or *P.* sp. NS1. The Cas proteins of *P. agarici* NCPPB 2472 are characteristic of a Type 1 Subtype I-F CRISPR system, Subtype I-F CRISPR systems have also been documented in

Pseudomonas aeruginosa PA14 [239], *Pseudomonas aeruginosa* strain UCBPP-PA14 [305], *Pectobacterium atrosepticum* [237] as well as certain strains of *E. coli*, where it was reported to be more frequently found in strains susceptible to antimicrobials, possibly due to interference in resistance plasmid acquisition [306]. The repeat consensus identified in *P. agarici* NCPPB 2472 was 28bp in length and a total of 33 spacer regions were associated with the identified repeats. Of the 34 spacer sequences, only a single sequence showed any similarity to known sequence, as identified by BLASTN [253], with a single base pair substitution. This sequence corresponded to 3 plasmid sequences present in the *Pseudomonas* strains: *P. frederiksbergensis* strain AS1, *P. fluorescens* strain PC20 (plasmid pNAH20) and *P. putida* NCIB 9816-4 (plasmid pDTG1). The identified plasmid sequences were found to encode naphthalene degrading enzymes, including the plasmid that was identified from *P. frederiksbergensis* strain AS1, which been demonstrated to be a naphthalene-degrading bacterium [277]. None of the spacer regions showed any nucleotide identity to the genomic sequences of any of the phage used in the course of our study.

While only *P. agarici* NCPPB 2472 was identified as possessing a CRISPR/Cas system, all of the *Pseudomonas* strains I have sequenced and mined possessed restriction endonuclease systems, which are a well-documented method by which bacteria are able to resist phage infection [203].

Within the genome of *P. tolaasii* 2192T several resistance endonucleases were identified, as shown in Table 3.4. The restriction endonucleases identified included a probable methylation-dependent restriction endonuclease, RS17360, and a Type III restriction endonuclease, RS17410. This shows that *P. tolaasii* 2192T possesses defence mechanisms against phage infection, including by phages with previously methylated nucleotides. Similarly, within the genome of *P.* sp. NS1, two restriction endonuclease proteins (CI807_27500 and

CI807_27505), shown in Figure 3.18, were identified as containing McrC and McrB conserved domains respectively which are associated with the methylcytosine specific McrBC restriction endonuclease of *E. coli* K-12 [273]; indicating that similar to *P. tolaasii* NCPPB 2192T, *P.* sp NS1 is capable of cleaving previously methylated nucleotides. In comparison while multiple restriction endonucleases were identified within the genome of *P. agarici* NCPPB 2472, shown in Table 3.13, including two Type I restriction endonuclease R subunits (RS18570 and RS18670), no methylation-dependent restriction endonucleases were identified.

A second potential phage resistance system was likewise identified in all three *Pseudomonas* species in this study, alginate. As previously mentioned, while not a specific phage resistance system, alginate can act to mask host receptors from phage and therefore can offer an advantage to the producing bacterium [200]. For *P. tolaasii* NCPPB 2192T a complete operon of 15 genes, shown in Table 3.5, was identified that was predicted to be involved in the production of alginate due to the presence of a gene encoding the alginate biosynthesis protein Alg44 (RS11530), which has been demonstrated to be required for alginate biosynthesis in *Pseudomonas aeruginosa* [275] and AlgE, also reported to be involved alginate biosynthesis in *Pseudomonas* [276]. Within the genome of *P. agarici* NCPPB 2472 I have identified a 12 ORF operon which is comprised of proteins orthologous to the 12 proteins (AlgD, Alg8, Alg44, AlgK, AlgE, AlgG, AlgX, AlgL, AlgI, AlgJ, AlgF, and AlgA) required for alginate production and export in *Pseudomonas* [276]. Therefore, it is highly likely that *P. agarici* NCPPB 2472 is capable of producing alginate. Similar to the operon of *P. agarici* NCPPB 2472, a cluster of 13 genes was identified in the genome of *P.* sp NS1 that contains proteins orthologous to AlgD, Alg8, Alg44, AlgK, AlgG, AlgX, AlgL, AlgI, AlgJ, AlgF and AlgA, which are required for alginate production and export in *Pseudomonas* [276]; although unlike the operon identified in *P. agarici* NCPPB 2472, it lacks a protein orthologous to AlgE and also contains an additional

hypothetical protein in ORF position 12. However, the presence of 11 out of the 12 genes required for alginate production and export indicates this cluster is highly likely to be involved in the production of alginate in *P.* sp. NS1.

Chapter 4:
Isolation and Identification of Bacteriophage Infecting *P. agarici* and *P. tolaasii*

4.1 Introduction

While the previous chapter has characterised the *P. tolaasii* NCPPB 2192T and *P. agarici* NCPPB 2472 genomes, including the presence of multiple phage resistance mechanisms, this chapter will focus on the identification and characterisation of bacteriophage capable of overcoming these resistance factors and utilising the bacteria as hosts; with the aim of elucidating the genetic and molecular basis of phage and host interaction and the possible use of the whole phage or phage proteins in the treatment of *Pseudomonas* infections of *Agaricus bisporus* mushrooms.

While the majority of phage research involving *Pseudomonas* species revolves around the human pathogen *P. aeruginosa* [34], some research has been performed on bacteriophage capable of infecting *P. tolaasii*, for example recently Kim *et al.* have isolated 21 environmental phage of *P. tolaasii* of varying lytic ability [74] . To date no research has been published on phage capable of infection *P. agarici*.

To enable the characterisation of the interaction of the phage-host relationship it was first necessary to isolate phage capable of infecting the host bacterium from environmental sources. Water samples were taken from the River Thames at select locations, as well as environmental samples from locations within the University of Reading campus and untreated sewage water samples. From these samples, two bacteriophages capable of causing lysis on *P. tolaasii* NCPPB 2192T and *P. agarici* NCPPB 2472 plates were identified; the first, named *Pseudomonas phage* NV1 was capable of infecting *P. tolaasii* 2192T and the second, named *Pseudomonas phage* ϕNV3, was capable of infecting *P. agarici* NCPPB 2472.

4.2 Isolation and characterisation of *Pseudomonas* phage NV1

4.2.1 Phage morphology

To aid in initial phage classification and guide future characterisation of phage NV1, a phage isolated from River Thames water, TEM imaging was performed, the resultant electron micrograph is shown below in Figure 4.1. TEM was performed on clarified phage lysates with negative staining using uranyl acetate.

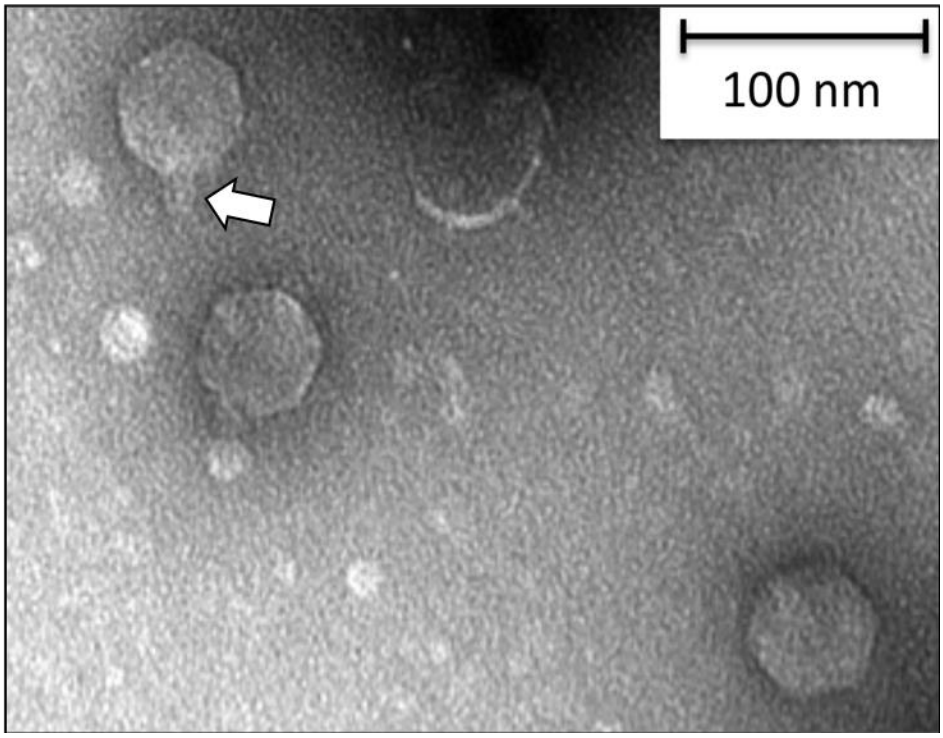

Figure 4.1. Electron micrograph at 80 kV of phage NV1 virion morphology, showing the icosahedral head and short tail (arrow) characteristic of the *Podoviridae*.

TEM was performed on clarified phage lysates with negative staining using uranyl acetate. From the presence of a tail in Figure 4.1 we can confidently identify phage NV1 as a member of the *Caudovirales.* The head of NV1 is approximately 58 nm (± 0.76 SD) in diameter and the short tail is approximately 26 nm (2.6 SD) in length, no tail fibers were visible on any TEM images returned. The presence of the short stubby tail and icosahedral head suggest that phage NV1 is a member of the *Podoviridae* family of the *Caudovirales* order.

Phage NV1 plaque morphology on *P. tolaasii* plaque assay plates show very small zones of lysis (< 2 mm) with a hazy appearance. The host range of NV1 is narrow, as shown in Table 4.1, being capable of only infecting *P. tolaasii* NCPPB 2192T of the 8 *Pseudomonas* strains tested.

Table 4.1. Host range assay results of phage NV1.

Strain:	**NV1 Infection**
P. tolaasii NCPPB 2192T	+
P. agarici NCPPB 2472	/
P. sp NS1	/
P. syringae pv. morsprunorum R1 5244	/
P. syringae pv. syringae 9097	/
P. syringae pv. morsprunorum R2 5255	/
P. fluorescens Pfo-1	/
P. aeruginosa 14207	/

4.2.2 Sequencing and genome organisation

The results of the TEM imaging of phage NV1 indicated that it belonged to the *Podoviridae*, a family of phage within Group I of the Baltimore classification scheme, and therefore most likely possessed a dsDNA genome. The genetic material of phage NV1 was extracted using a Norgen Biotek Phage DNA Isolation Kit, returning 75 µl of DNA at a concentration 128.9 ng/µl as determined by NanoDrop™ 2000. The genomic DNA was sent for sequencing by Source Bioscience on an Illumina MiSeq®, with paired-end reads of 300 bp. The returned sequence for NV1 was a single contig of 45,059 bp in length, with a coverage of 55.8x, BLASTN [253] analysis of the returned sequence showed NV1 was closely related to the *Pseudomonas fluorescens* phage UFV-P2 (84% Identity across 86% of the genome), which is illustrated in the dot matrix shown in Figure 4.2, and so the genome of NV1 was oriented to be collinear to that of UFV-P2. The complete genome of phage NV1 is available on GenBank with the accession code: MG845684.1.

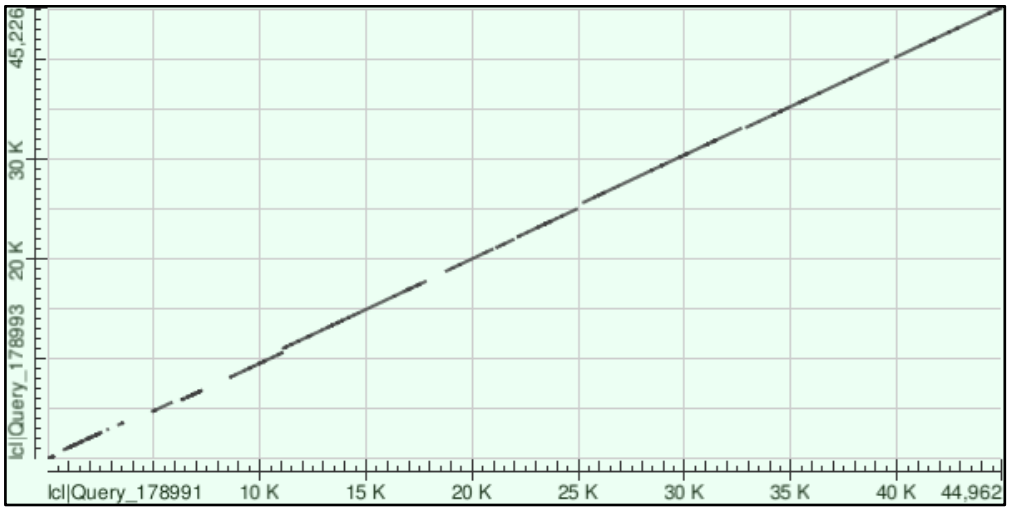

Figure 4.2. Dot matrix view identifying regions of similarity based upon the BLASTN [253] alignment. Phage NV1 sequence is represented on the X-axis and the numbers represent the bases of the genome sequence. Phage UFV-P2 is represented on the Y-axis. Alignments are shown in the plot as lines. Positive strand and protein matches are slanted from the bottom left to the upper right corner, negative strand matches are slanted from the upper left to the lower right. The number of lines shown in the plot is the same as the number of alignments identified by BLASTN[253].

The dsDNA genome of NV1 is approximately 45,059 bp in length, which is slightly smaller than that of the closely related UFV-P2, at 45,517 bp, consequently NV1 encodes 64 predicted ORFs compared to the larger 75 of UFV-P2 [108]. 57 of the 64 encoding ORFs (89%) found in NV1 show significant amino acid identity to encoding ORFs identified in UFV-P2. The genome of NV1 has a slightly higher GC content at 52.87% than is reported for UFV-P2 at 51.5% [108], however it shares a similar bidirectional organization, illustrated graphically in Figure 4.3 with the early host conversion genes as well as the viral DNA replication genes being located within the positive strand, with the exception of the putative phage holin. The putative late stage genes encoding proteins involved in virion structure and assembly are encoded in the negative strand.

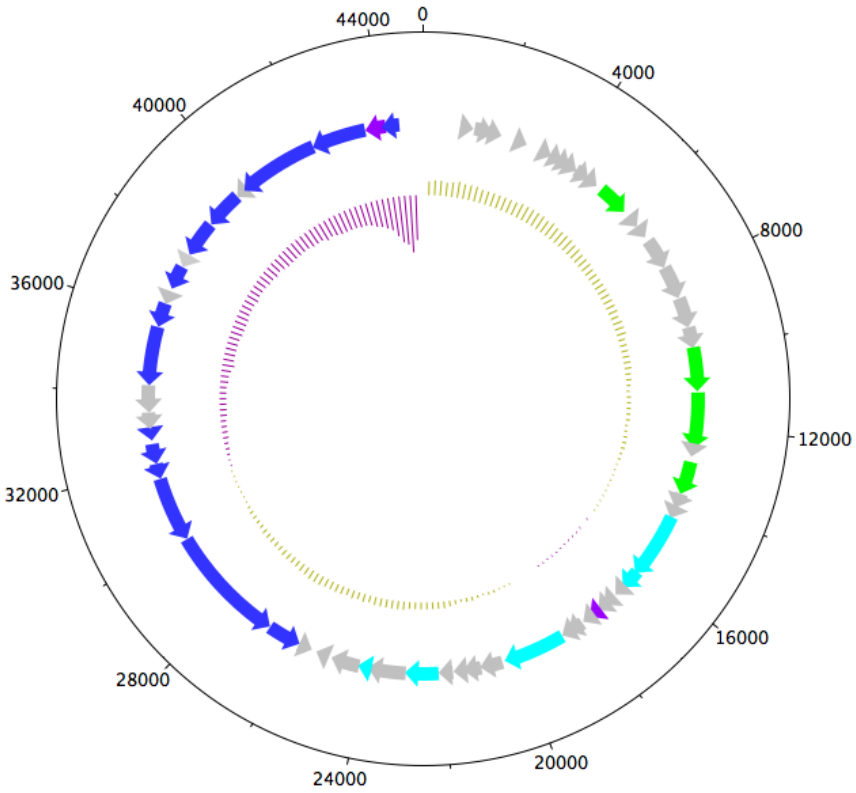

Figure 4.3. Graphical plot of the genome structure of *Pseudomonas phage* NV1, with the early host conversion genes in green, the DNA replication genes in light blue, the virion structure genes in dark blue, hypothetical proteins in grey and the cell lysis proteins in purple. The deviation in GC content from the total average is illustrated in the centre circle with yellow bars denoting an above average content and purple bars illustrating a below average GC content. Figure generated using DNAPlotter function of the Artemis v16.0.0.1 genome browser.

Of the 64 predicted ORF's, only 25 showed amino acid similarity to proteins of known or predicted function as identified by BLASTP [253], these are shown in Table 4.2. The majority of those with predicted functions are those that are involved in virion structure, assembly and DNA packaging which show high levels of amino acid sequence conservation among other closely related phage.

Similarly, the ORF's encoding for the replication genes at NV1_p21, p24, p25, p29 and p33 putatively encoding an ATP-grasp enzyme, DNA primase/helicase, holin and two exons of the phage DNA polymerase respectively show, not only a high degree of amino acid sequence

conservation among closely related phage strains, but a very high nucleotide identity to UFV-P2 (78-88% Ident) and unusual organisation of a putative holin located between the exons of the DNA polymerase [108].

The early host conversion genes, however, show the least conservation both among closely related phage strains and between phage NV1 and UFV-P2; consequently, the function of most is not deducible from the amino acid sequences, a full table of predicted NV1 ORF'S is available in Appendix 3, Table Ap3.1.

Table 4.2. ORFs of phage NV1 with BLASTP [253] similarity to proteins of known or predicted function.

Locus Tag	Start Position	Size (aa)	Total GC (%)	Predicted Function (putative):
NV1_p11	5060	276	57.04	SPFH domain-containing protein
NV1_p18	9913	384	54.46	Amidoligase
NV1_p19	11098	489	53.5	Glutamine amidotransferase
NV1_p21	12926	291	54.45	ATP-grasp enzyme
NV1_p24	14445	584	51.51	DNA primase/helicase
NV1_p25	16138	183	49.28	DNA polymerase part I
NV1_p29	17588	85	51.4	Holin
NV1_p33	18725	546	53.12	DNA polymerase Family A
NV1_p38	22125	294	52.84	5'-3' exonuclease
NV1_p40	23888	120	52.06	Endonuclease
NV1_p44	25881	295	53.48	Phage structural protein
NV1_p45	26779	1057	53.64	Phage structural protein
NV1_p46	29959	571	54.21	Phage structural protein
NV1_p47	31676	134	52.08	Phage structural protein
NV1_p48	32077	366	53.17	Phage particle protein
NV1_p49	32699	93	54.6	Phage particle protein
NV1_p52	34147	514	50.29	Phage particle protein
NV1_p53	35699	211	51.41	Tail fiber protein
NV1_p56	36755	209	50.32	Phage structural protein
NV1_p58	37759	317	52.2	Major capsid protein
NV1_p59	38730	333	51.7	Scaffolding protein
NV1_p61	39969	712	50.16	Phage portal protein
NV1_p62	42086	481	50.76	Terminase large subunit
NV1_p63	43535	179	50.56	Lysozyme
NV1_p64	43962	157	47.05	Phage structural protein

A common feature of the *Luz24likeviruses* is the reported presence of localized single-stranded breaks associated with the consensus sequence 5'-TACT/RTGMC-3', for example the genome of *Pseudomonas putida* Phage tf contains 14 single-strand interruptions in the positive strand [307] and the genome of phage UFV-P2 contains 15 on the positive strand [108].

In total, 13 potential single-strand nick sequences were identified in NV1, all in positive strand, with consensus sequence 5'-TACTRTGAC-3'; therefore, it is likely that the genome of NV1 contains single-stranded nicks similar to other *Luz24likeviruses*.

4.2.3 Phylogeny

To analyse the phylogenetic clustering of NV1 within the *LUZ24likevirus* genus, the entire genomic nucleotide sequence was aligned by MUSCLE [260] and all gaps were removed from the sequence. A maximum parsimony analysis was chosen to predict evolutionary history due to the small number of sequences and likelihood of nucleotide conservation between them and run using Mega6.06. The resulting phylogenetic tree is shown in Figure 4.4.

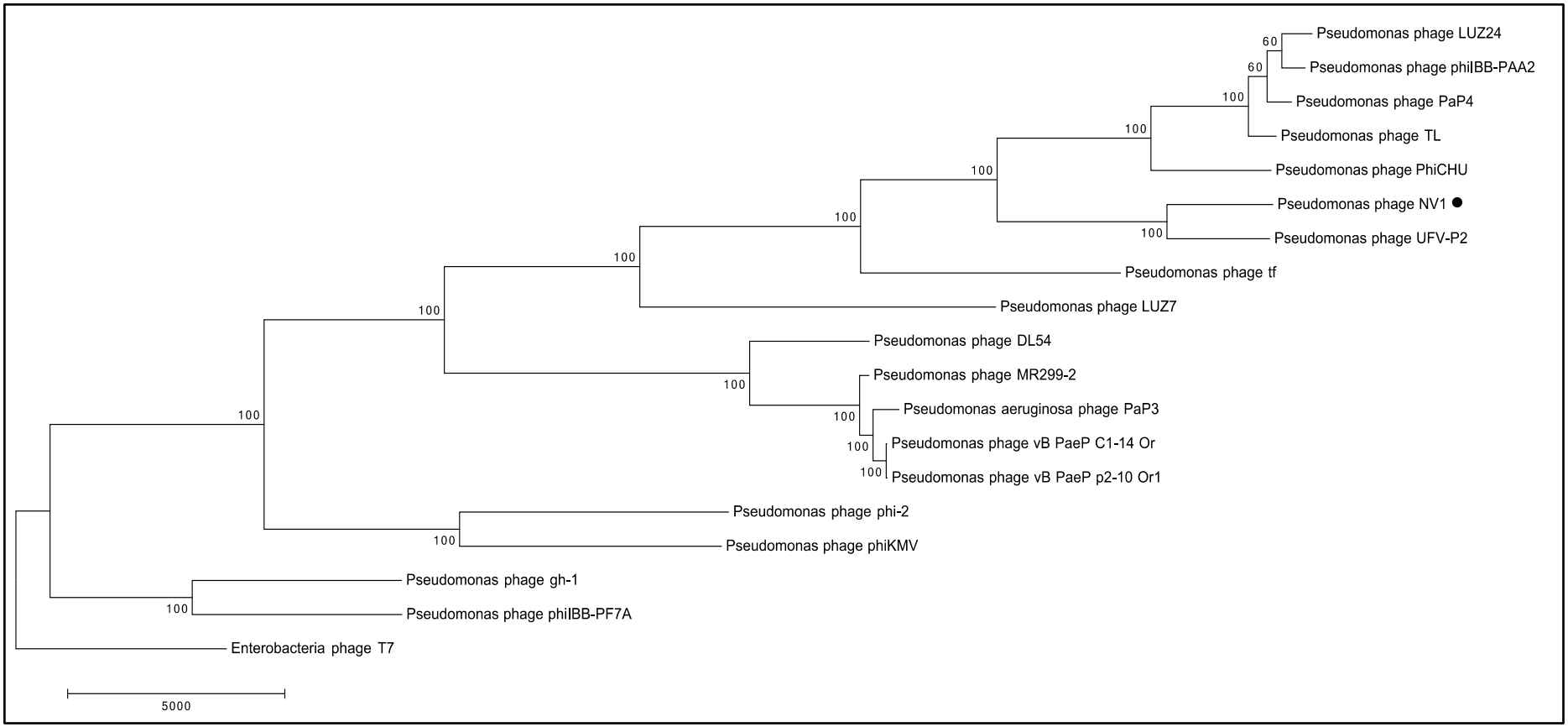

Figure 4.4. Maximum parsimony analysis of *Pseudomonas* phage NV1 (●). The most parsimonious tree with length = 105449 is shown. The consistency index is (0.561417) the retention index is (0.693657), and the composite index is 0.392642 (0.389431) for all sites and parsimony-informative (in parentheses). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches [257]. The MP tree was obtained using the Subtree-Pruning-Regrafting (SPR) algorithm [258] with search level 1 in which the initial trees were obtained by the random addition of sequences (10 replicates). The tree is drawn to scale, with branch lengths calculated using the average pathway method [258]and are in the units of the number of changes over the whole sequence. The analysis involved 19 nucleotide sequences. Codon positions included were $1^{st}+2^{nd}+3^{rd}$+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 24899 positions in the final dataset. Evolutionary analyses were conducted in MEGA6.06 [256].

The phylogenetic tree in Figure 4.4 confirms that NV1 and UFV-P2 are closely related and possibly of the same species, clustered together on an off-shoot of the branch most closely related to other phage of the *Luz24likevirus* genus of the *Podoviridae* family. Two *phiKMVvirus* phage were included in the analysis, *Pseudomonas* phage phi-2 and *Pseudomonas* phage phiKMV, which are located in a branch separate from the *Luz24viruses* towards the lower end of the tree. Likewise, the *T7virus* phages, *Pseudomonas* phage gh-1, *Pseudomonas* phage philBB-PF7A are located in a single branch closest to *Enterobacteria* phage T7 which was used to root the phylogenetic tree. A further phylogenetic analysis utilising only the large terminase subunit gene is available in Appendix 4, Figure Ap4.6, which confirms the results shown in Figure 4.4, with NV1 clustering closely with UFV-P2 and other *Luz24likevirus* members.

4.2.4 Growth characteristics

To characterise the ability of phage NV1 to lyse *P. tolaasii* a variety of kill curves were performed at different starting MOIs and over either 150 minutes, as shown in Figure 4.5, or 700 minutes, as shown in Figure 4.6.

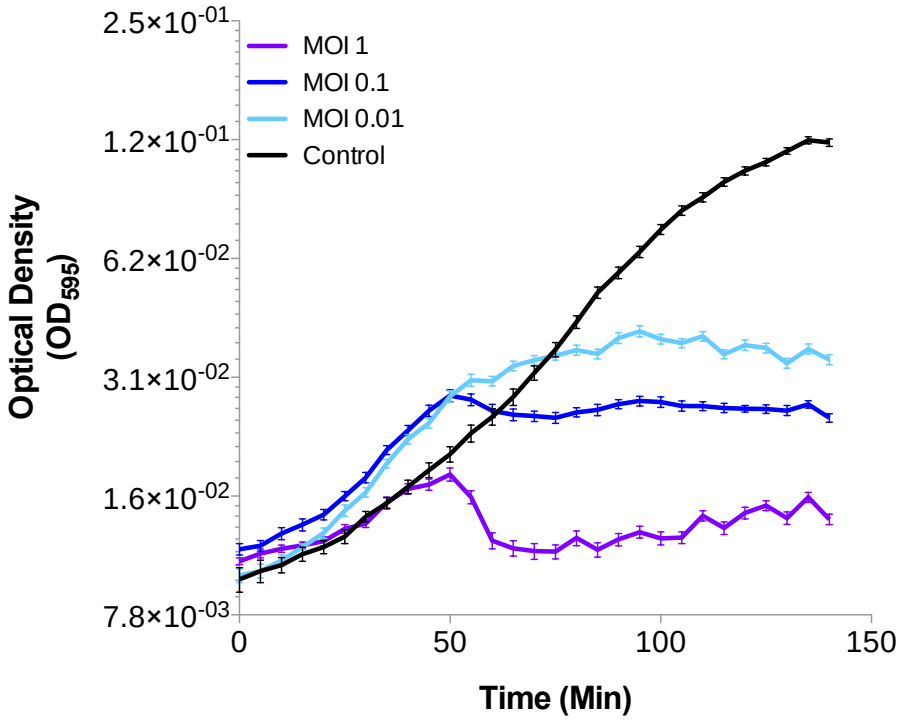

Figure 4.5. Growth curve of *P. tolaasii* infected with NV1 at a starting bacterial concentration of ~1x10^8 cfu/ml, incubated at 28°C. Phage NV1 was added to the cultures at T=0 at an MOI of 1, 0.1, 0.01. Mean values of 10 replicates, error bars = +/- SEM.

At an MOI of 1 after 50 minutes the OD dropped from 0.018 Abs to 0.011 Abs at 70-75 minutes, a difference of 0.007 Abs, in comparison across the same time points at a MOI of 0.1 there is a drop in OD of only 0.003 Abs. However, at a MOI of 0.01 the OD increases by a total of 0.007 Abs between 50 and 75 minutes.

By 150 minutes the OD data in Figure 4.5 shows little change across all MOIs, however the final ODs vary greatly, at an MOI of 1 the final OD is 0.014 Abs, at an MOI of 0.1 the final OD is 0.025 Abs and at a MOI of 0.01 the final OD is 0.035 Abs; in comparison, the control OD at the final time point is significantly higher at 0.123 Abs. As OD is correlated with total bacterial cell count it is clear that the largest drop in bacterial cell count occurs at the highest MOI of 1 and that initial MOI correlates with the cell count at the final time point, with the largest MOIs demonstrating the lowest final ODs.

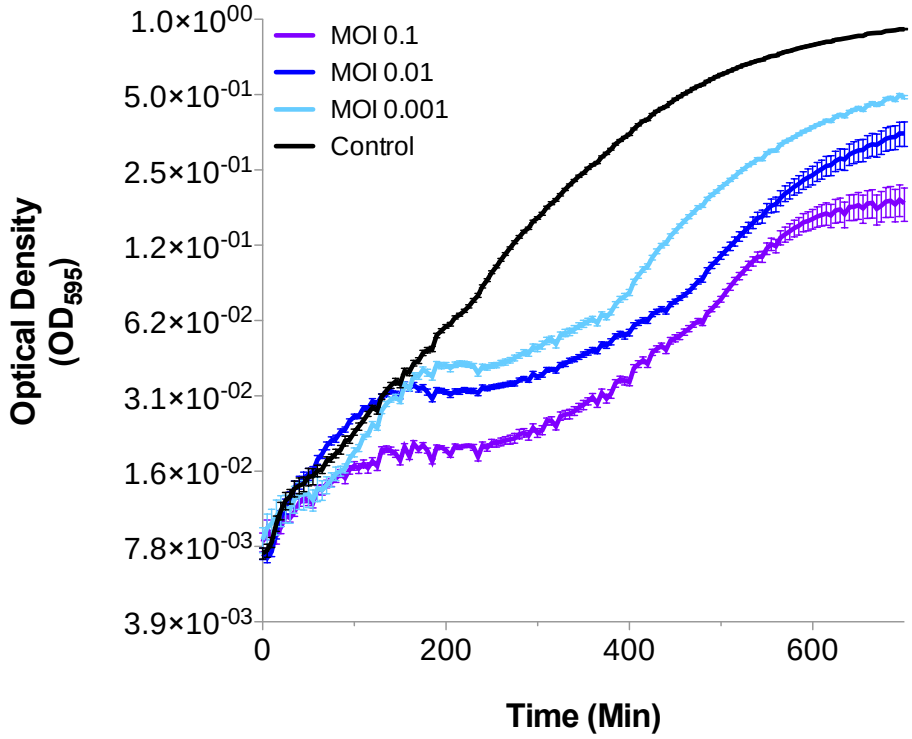

Figure 4.6. Growth curve of *P. tolaasii* infected with NV1 at a starting bacterial concentration of ~1x10^8 cfu/ml, incubated at 28°C for 2 hours with a MOI of either 0.1, 0.001 or 0.001. OD at 595 nm was read every 5 minutes. Mean values of 10 replicates, error bars = +/- SEM.

In order to determine the effectiveness of phage at lower MOI's that would more accurately represent the conditions likely to occur on the surface of *A. bisporus* or within the casing soil layer a further set of kill curve experiments were performed as shown in Figure 4.6.

At t=90 minutes the OD data for the MOI of 0.1 begins to deviate from the MOI of 0.001 which more closely matches the uninfected control in rate of OD increase. The rate of increase in OD of MOI of 0.1 begins to flatten at t=140 minutes at an OD of approximately 0.02 Abs, indicating that the rate of cell lysis is matching the rate of cell replication; the OD begins to once again rise beyond the t=250 minutes point and continues to rise before stabilising once again at around t=600 minutes, ending at an OD of approximately 0.18 Abs.

The MOI of 0.01 roughly follows the rate of increase in OD of the control until t=160 minutes where it begins to flatten at an OD of approximately 0.03 Abs; similar to the MOI of 0.1, past t=250 minutes the OD begins to rise again, although it does not plateau towards the final time

point, it does however show a slowing in the rate of OD increase, ending at a final OD of approximately 0.35Abs.

The lowest MOI of 0.001 only begins to flatten at t=195 minutes at an OD of approximately 0.04 Abs, the largest OD of all MOI's tested. However, past the t=250 time point, similar to both MOIs of 0.1 and 0.01 the OD begins to increase again across the remainder of the experiment, ending at a final OD of approximately 0.48 Abs.

The uninfected control in contrast shows a steady rate of increase in OD until t=400 minutes where the rate of increase in OD begins to slow, ending at a final OD of 0.91 Abs.

No single step growth curves or adsorption assays were performed with phage NV1 because of difficulties in getting accurate plaque counts due to the small size and hazy appearance of NV1 plaques.

4.3 Isolation and characterisation of *Pseudomonas* phage ϕNV3

4.3.1 Phage morphology

Pseudomonas phage ϕNV3 was initially identified from samples of untreated sewage where it was identified as causing plaques on plates of *P. agarici* NCPPB 2472 and was then imaged by TEM. TEM imaging of phage virion particles allows for a broad classification of the phage based on morphology. TEM was performed on clarified phage lysates with negative staining using uranyl acetate. From the presence of an extremely short tail in the electron micrograph in Figure 4.7, we can tentatively identify phage ϕNV3 as a member of the *Caudovirales*, similar to phage NV1.

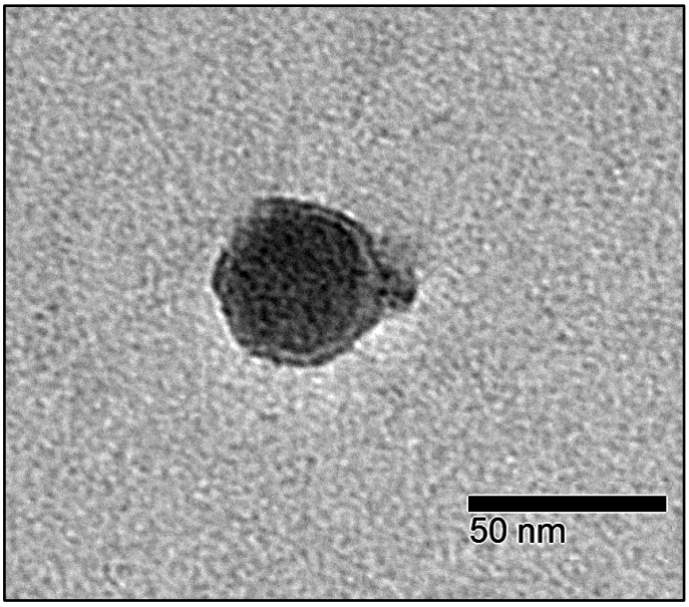

Figure 4.7. Electron micrograph at 80 kV, illustrating the virion morphology of ϕNV3, clearly visible are the icosahedral head and short tail characteristic of the *Podoviridae*.

The head of ϕNV3 is approximately 43nm (1.5 SD) in diameter and an extremely short tail that is approximately 10 nm (3 SD) in length, with no tail fibers visible on any TEM images, similar to phage NV1. The presence of the short stubby tail and icosahedral head suggest that phage ϕNV3 is a member of the *Podoviridae*.

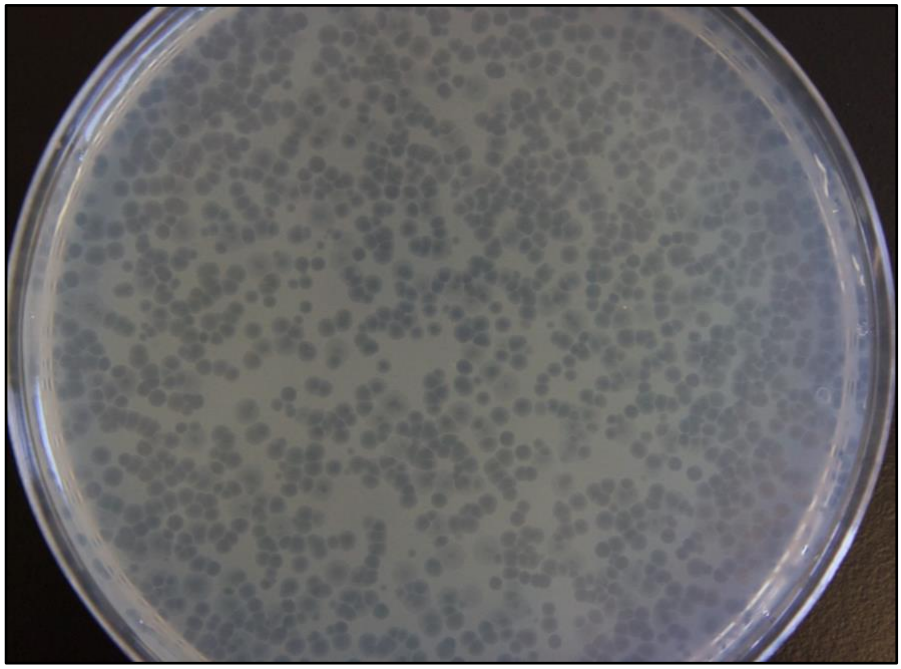

Figure 4.8. Plaque assay plate of phage ϕNV3 and *P. agarici* NCPPB 247 showing near-confluent lysis.

Phage ϕNV3 has a narrow host range, only infecting *P. agarici* NCPPB 2472, of all the *Pseudomonas* species tested, as shown in Table 4.4; on plaque assays of *P. agarici*, ϕNV3 plaques show large clear plaques of approximately 3-5 mm in diameter, illustrated in Figure 4.8.

Table 4.4. Host range assay results of phage ϕNV3.

Strain:	**ΦNV3 Infection**
P. tolaasii NCPPB 2192T	-
P. agarici NCPPB 2472	+
P. sp NS1	-
P. syringae pv. morsprunorum R1 5244	-
P. syringae pv. syringae 9097	-
P. syringae pv. morsprunorum R2 5255	-
P. fluorescens Pfo-1	-
P. aeruginosa 14207	-

4.3.2 Sequencing and genome organisation

The results of the TEM imaging of phage ϕNV3 indicated that it belonged to the *Podoviridae*, a family of phage within Group I of the Baltimore classification scheme, and therefore most likely possessed a dsDNA genome. The genetic material of phage ϕNV3 was extracted using a Norgen Biotek Phage DNA Isolation Kit, returning 75 µl of DNA at a concentration 126.4 ng/µl as determined by NanoDrop™ 2000. The genomic DNA was sent for sequencing by Source Bioscience on an Illumina MiSeq®, with paired-end reads of 300 bp. Full genome sequencing of ϕNV3 returned a single contig of 43,184 bp in length with a coverage of 9.2x and an average GC content of 58.29%, which was confirmed with extensive primer walking combined with Sanger Sequencing. The complete genome of ϕNV3 was analysed by PhageTerm [270] for the presence of direct terminal repeats (DTRs), which identified a T7-like packaging scheme with DTRs of 693bp in length. The completed sequence was uploaded to GenBank (Accession code: MG845683).

ORFs over 100 bp in length were considered as potential genes and 49 ORFs were identified and predicted to encode proteins, all of the ORFs reside on the positive strand and the organisation of the genes shows significant similarities to that reported for ϕKMV [100] and is illustrated graphically in Figure 4.9; however the total nucleotide identity to any known phage is extremely low, with the largest as identified by BLASTN [253] to be 77% identity with only 3% cover of the genome to phage ϕKMV. Of the 49 identified ORFs, 25 showed amino acid identity to phage proteins of known or predicted function as identified by BLASTP [253], these are shown in Table 4.5. A full table of all ORFs identified is available in Appendix 3, Table Ap3.2.

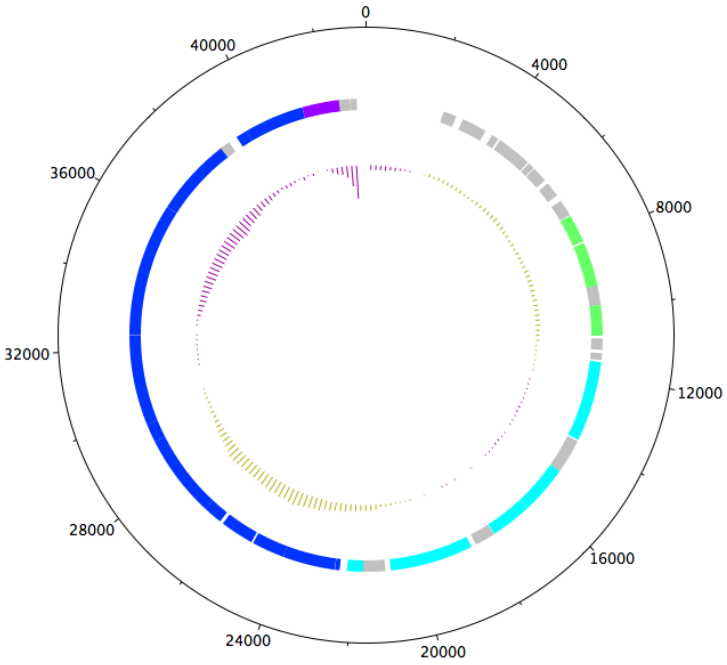

Figure 4.9. ϕNV3 genome organisation, green early host preparation genes, light blue is mid, dark blue late and structural, purple is lysis cassette, hypothetical proteins in grey, inner circle denotes GC usage with yellow being above average and purple being below average, genome is shown circular for illustrative purposes only. Figure generated using DNAPlotter function of the Artemis v16.0.0.1 genome browser.

Similar to phage NV1 the majority of genes with identified functions in Table 4.5 are phage structural proteins and those involved in DNA replication. Notably, ϕNV3 encodes a single subunit RNA polymerase upstream of DNA replication genes, adjacent to the structural protein region of the genome.

Table 4.5. ORFS of phage ϕNV3 with BLASTP [253] similarity to proteins of known or predicted function.

Locus Tag	Start Position	Size (aa)	Total GC (%)	Predicted Function (putative):
phiNV3_p13	7159	278	57.46	DNA primase
phiNV3_p14	8047	436	58.81	DNA helicase
phiNV3_p16	9913	299	60.11	ATP-dependent DNA ligase
phiNV3_p19	11583	786	58.83	DNA polymerase I
phiNV3_p21	15008	319	58.64	Integrase
phiNV3_p22	15948	158	54.92	DNA endonuclease VII
phiNV3_p23	16432	406	59.04	DNA exonuclease
phiNV3_p26	18414	818	58.4	RNA polymerase
phiNV3_p29	21679	156	57.53	DNA endonuclease
phiNV3_p30	22363	45	60.86	Virion structural protein
phiNV3_p31	22510	512	57.24	Head-tail connector protein
phiNV3_p32	24048	321	61.9	Scaffolding protein
phiNV3_p33	25069	330	60.12	Capsid protein
phiNV3_p34	26161	184	56.75	Tail tubular protein A
phiNV3_p35	26712	823	59.54	Tail tubular protein B
phiNV3_p36	29183	181	59.7	Internal virion protein A
phiNV3_p37	29721	886	59.9	Baseplate hub subunit and tail lysozyme
phiNV3_p38	32390	1310	60.56	Internal core protein
phiNV3_p39	36323	788	54.79	T7-like tail protein
phiNV3_p41	39215	110	57.7	DNA maturase A
phiNV3_p42	39529	590	57.13	DNA maturase B
phiNV3_p43	41303	79	60.41	Pinholin
phiNV3_p44	41526	166	58.48	Lysozyme
phiNV3_p45	41975	110	60.96	Rz-like protein
phiNV3_p46	42165	81	62.6	Rzl-like protein

phiNV3_p37, a protein composed of 886 amino acids, is predicted to encode the baseplate hub subunit and tail lysozyme. A 122 aa T4-like lysozyme conserved domain is located at the C-terminus of the protein between residues 738-860. This C-terminus lysozyme domain is characteristic of the *Autographivirinae* including phage ϕKMV and SP6, the T7-like phage of the *Podoviridae* instead have this domain at the N-terminus [100, 308].

4.3.3 Phylogeny

To analyse the phylogenetic clustering of ϕNV3 the entire genomic sequence was aligned by MUSCLE [260] to other members of the *Autographivirinae* and all gaps were removed from the sequence. A maximum parsimony analysis was chosen to predict and evolutionary history due to the small number of sequences and likelihood of nucleotide conservation between them and run using Mega6.06. The resulting phylogenetic tree is shown in Figure 4.10.

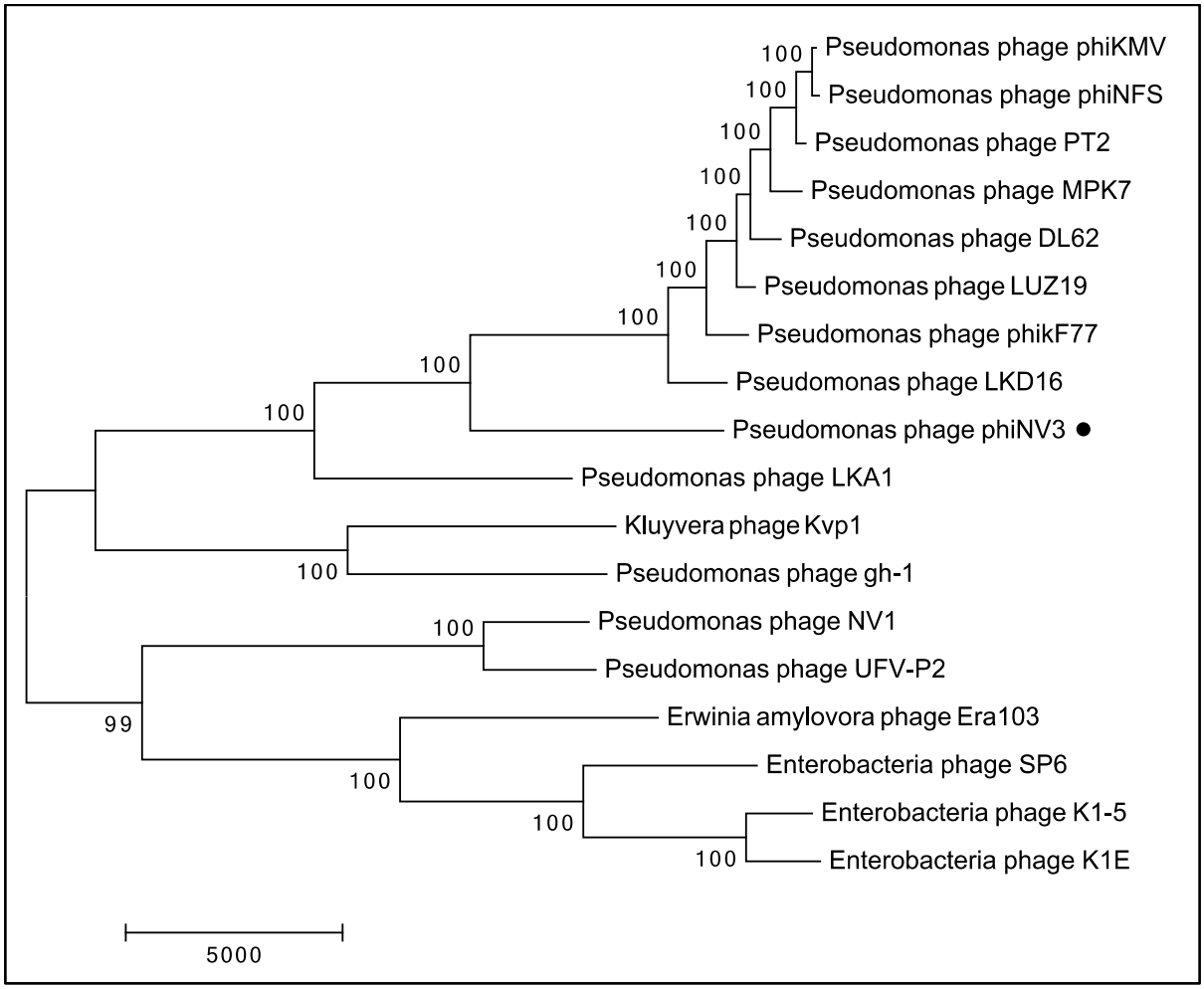

Figure 4.10: Maximum Parsimony analysis of taxa: The evolutionary history of *Pseudomonas* phage ϕNV3 (•), was inferred using the Maximum Parsimony method. The most parsimonious tree with length = 95095 is shown. The consistency index is (0.613301), the retention index is 0.422029 (0.417186) for all sites and parsimony-informative sites (in parentheses). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches [257]. The MP tree was obtained using the Subtree-Pruning-Regrafting algorithm [258] with search level 1 in which the initial trees were obtained by the random addition of sequences (10 replicates). The tree is drawn to scale, with branch lengths calculated using the average pathway method [258] and are in the units of the number of changes over the whole sequence. The analysis involved 18

nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 26547 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 [256].

The results in Figure 4.10 show that phage ϕNV3 clusters with other members of the *phiKMVviruses*, although on a separate branch to the majority, indicating that it may have diverged earlier than other members, similar to phage LKA1. A further phylogenetic analysis utilising only the large terminase subunit gene is available in Appendix 4, Figure Ap4.6, which confirms the results shown in Figure 4.10, with ϕNV3 clustering with other members of the *phiKMVviruses* and other members of the *Autographivirinae*.

4.3.4 Growth characteristics

In order to further characterise the ϕNV3 life cycle several studies were performed. As the first stage of a phage life cycle is finding and adhering to its host, an adsorption assay was performed for ϕNV3, corresponding to typical growth parameters *in vitro*. The results of this adsorption (Figure 4.11), were used to calculate the adsorption constant of ϕNV3, which was found to be approximately k_a=3.5x10^{-9} ml min^{-1} demonstrating ϕNV3 efficiently adsorbs to *P. agarici* cells.

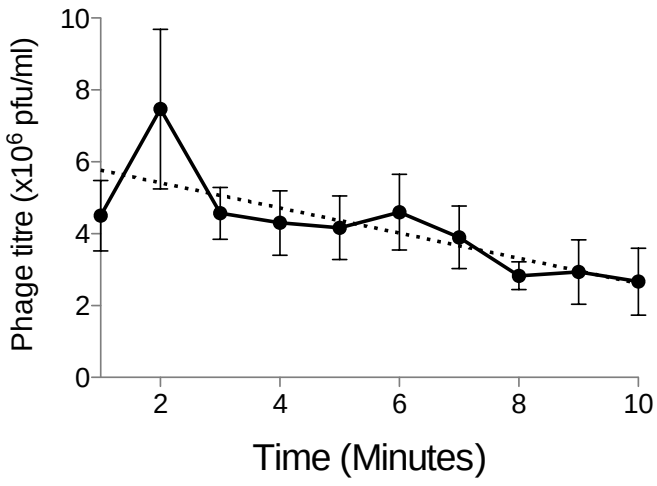

Figure 4.11. Adsorption curve at 28°C with 200 rpm shaking (dotted line is linear regression) Adsorption constant: 3.501x10^{-9} phage/cell/ml/min, error bars = +/- SEM.

Following adsorption, the phage genome must enter the host cell and undergo replication and transcription/translation to produce progeny phages. A single step growth curve was performed to determine duration of the phases of phage replication and burst size (defined as the number of progeny phage release after a single phage has infected a single bacterium) for ϕNV3. The results (Figure 4.12) show the latent period of phage ϕNV3 is approximately 45 minutes in length, followed by a steep rise period of approximately 10 minutes due to the lysis of infected cells and release of progeny phage, the calculated burst size was approximately 55 phage per cell for ϕNV3. This stage was rapidly followed by a sharp decrease in phage titre, which can be attributed to the attachment of progeny phages to new host cells.

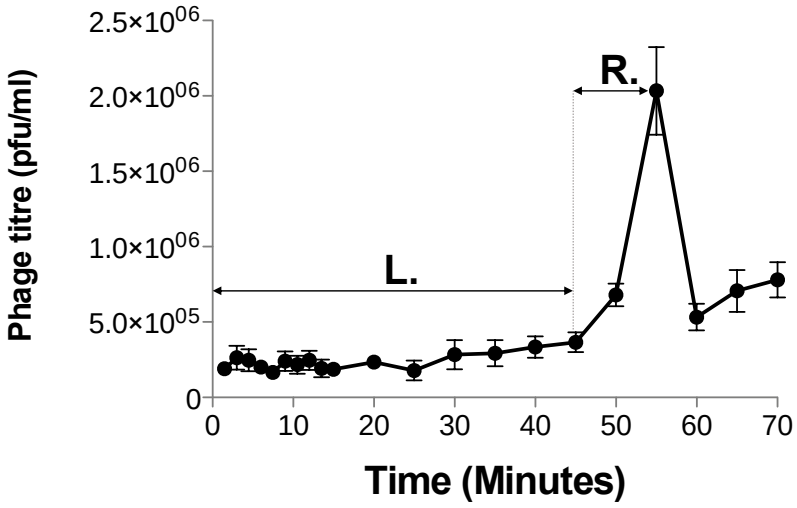

Figure 4.12. Single step growth curve of ϕNV3 with *P. agarici* NCPPB 2472. Latent period and rise periods are indicated with lines labelled L. and R. respectively. Mean values of 3 replicates are plotted, error bars = +/- SEM.

In order to characterise the ability of ϕNV3 to effectively control bacterial populations, kill curve assays were performed at high phage titres with MOIs of between 0.1 and 10, the results are shown in Figure 4.13.

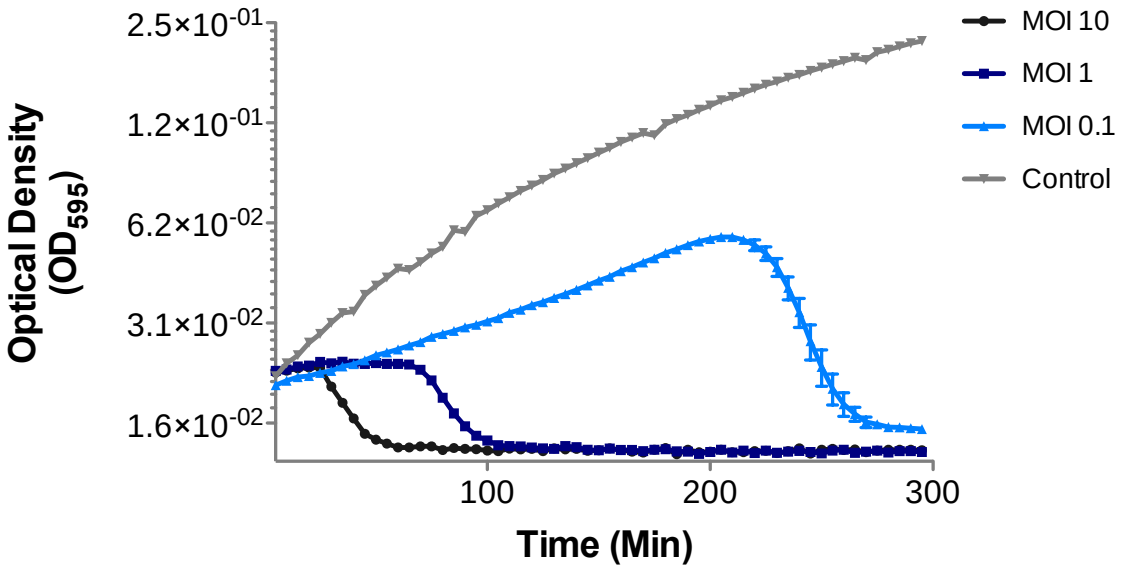

Figure 4.13. High titre kill curves with ϕNV3 and *P. agarici* NCPPB 2472 at a starting bacterial concentration of $1x10^8$cfu/ml at 28°C, mean values of 10 replicates are plotted, error bars = +/- SEM.

With a high MOI of 10 there is a rapid drop in OD commencing at 25 minutes and stabilising at approximately 60 minutes. This would appear to correspond to a single infective cycle where all bacterial cells have been successfully infected with ϕNV3 resulting in total lysis of all bacterial cells within 60 minutes, which matches the infection cycle characterised in the single step growth curve in Figure 4.12. The total lysis of all cells is illustrated by the lack of any rise in the OD of the culture over the remaining duration of the experiment.

However, when the MOI is 1 the time required for a similar drop in OD to commence increases over two times to approximately 65 minutes with the total lysis as indicated by a stabilising of OD at 110 minutes. Once again, the drop in OD appears to correspond to complete lysis of all the bacterial cells as no increase in OD is observed in the culture of the remaining duration of

the experiment after the initial drop in OD between 65-110 minutes. The time required for this drop in OD would be equivalent to two cycles of phage replication.

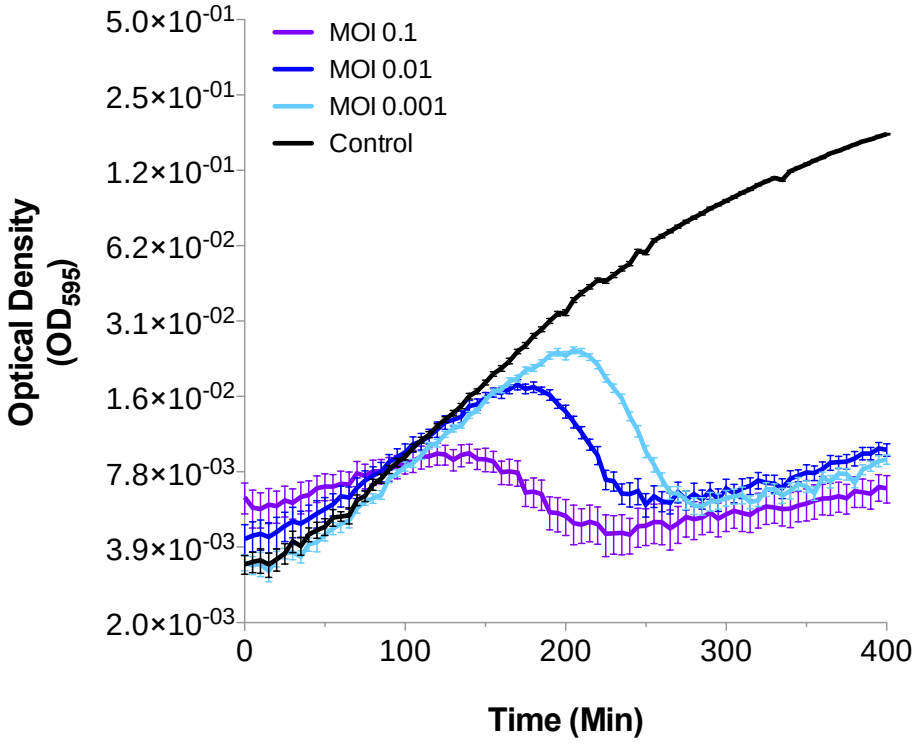

Figure 4.14. Low MOI kill curves kill curves with ϕNV3 at a starting bacterial concentration of 1x10^{8}cfu/ml at 28°C. Mean values of 10 simultaneous replicates are plotted, error bars= +/- SEM.

In order to determine the effectiveness of phage at lower MOI's that would more accurately represent the conditions likely to occur on the surface of *A. bisporus* or within the casing soil layer a further set of kill curve experiments were performed as shown in Figure 4.14.

At higher MOIs the drop in OD occurs more rapidly than at lower MOIs, for example the drop in OD at an MOI of 0.1 occurs at t=140 (0.009 Abs) and reaches a minimum at t=240 (0.004 Abs), whereas at an MOI of 0.01 the drop does not occur until t=185 (0.016 Abs) and reaches a minimum at t=250 (0.006 Abs). Likewise, the drop in OD for the lowest MOI of 0.001 does not occur until t=210 (0.023 Abs) and reaches a minimum at t=285 (0.006 Abs).

Interestingly the drop in OD expressed as a percentage of the maximum peak in OD shows that the largest percentage drop is at the lowest MOIs, at an MOI of 0.1 the drop in OD corresponds to a 55.6% drop, at an MOI of 0.01 it corresponds to a 62.5% drop and at the lowest MOI of 0.001 the drop corresponds to 74% of the peak OD value.

4.3.5 Phage/host co-evolution

In order to investigate the effects of antagonistic co-evolution on the bacterial host a co-evolution study was performed. This was performed using sequential transfers of 100 µl of phage/bacteria in 6 ml of KB followed by overnight incubation at 27°C. From this, samples of purified phage and bacteria were taken, before phage infection capacity was measured by streaking on plates of bacterial colonies of each transfer. The graphical results are shown in Figure 4.14 and full results are available in Appendix 4, Table Ap4.4.

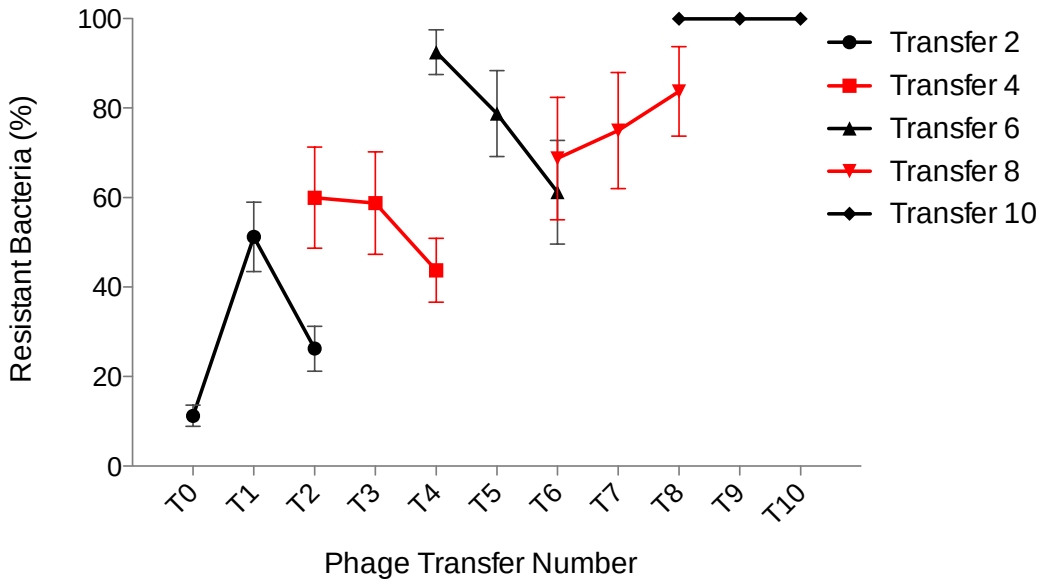

Figure 4.15. Results of co-evolution study between phage ϕNV3 and *P. agarici* NCPPB 2472. Percentage resistant bacteria was determined using 16 arbitrarily sampled bacterial colonies from streak plates of each bacterial transfer to which 5µl of either past/present/future phage was added, each point is the mean value of 5 replicates, error bars= +/- SEM. x axis labels indicate phage transfer number, each bacterial transfer number is illustrated by data point shape and outlined in the key.

From the results shown in Figure 4.15 there is evidence of a high degree of co-evolution between the phage and host. For bacterial Transfers 1, 3, 5 it is evident that ‘future phage’ from the next sequential transfer is more effective; however, at bacterial transfer 8 the ‘future phage’ is less effective and by bacterial Transfer 9 all bacteria are resistant to all phage tested.

Over the course of the experiment escalatory evolution of bacterial resistance is evident in the gradual increase in percentage bacteria resistant across all past, present and future phage, resulting in the final total resistance of the bacteria to all phage and eventual phage extinction. The initial bacterial transfer (Transfer 1) shows that the ‘future phage’ (T2) are capable of overcoming the initial bacterial resistance mechanisms, a similar result is shown in transfers 4 and 6.

4.3.6 Lysis cassette

In order to characterise the process by which phage ϕNV3 is able to cause lysis of *P. agarici* NCPPB 2472, the lysis cassette was identified and investigated.

The bacteriophage lysis cassette, which is comprised of the phage encoded proteins involved in lysing the host cell to allow progeny phage escape at the end of the phage replication cycle, such as the lysin and holin (S) [109], or in the case of SAR-endolysins, the pinholin, lysin and Rzl/Rz1-like proteins [102].

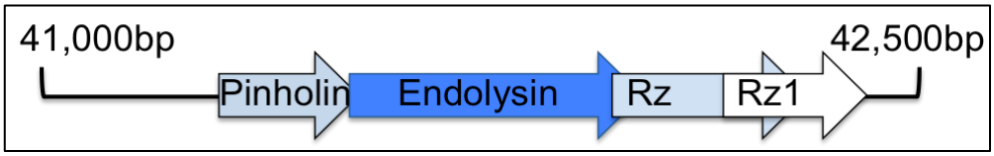

Figure 4.16. Diagram representing the lysis cassette of ϕNV3, showing the ORF's encoding the putative pinholin, endolysin, Rz and Rz1-like proteins and the high degree of overlap between each ORF's.

The lysis cassette of ϕNV3 comprises of four proteins, as shown in Figure 4.16. The presence of a predicted pinholin and Rz-like and Rz1-like proteins as well as the order of these genes indicate that ϕNV3 possesses a SAR-endolysin system conserved in all *phiKMVviruses* [102], rather than a canonical system such as that found in NV1 that is comprised of solely a holin and endolysin proteins.

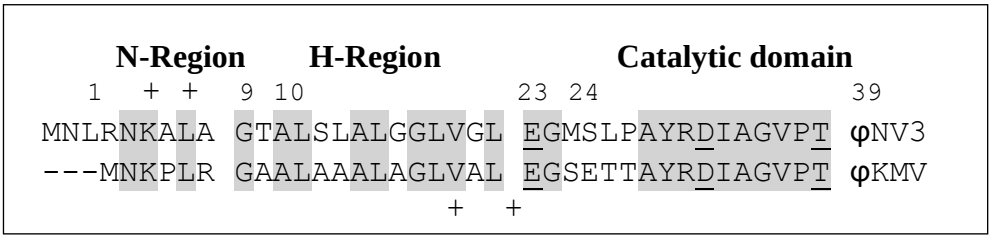

Figure 4.17. Direct comparison of the N-terminal *signal-arrest-release* domains of ϕNV3 and ϕKMV. The positively charged N-region, the hydrophobic H-region and the catalytic domain are indicated with the conserved catalytic domains underlined, based on work published by Briers *et al.* [102].

Analysis of the amino acid terminal sequence of phiNV3_p44, a putative endolysin, shows that it is almost identical to that of ϕKMV. That is, the ϕNV3 N-region of 9 amino acid residues,

which is 3 residues longer than that of ϕKMV, has a net positive charge (identified in Figure 4.17 by the + symbol), and thus acts as a positive anchor to the negatively charged inner site of the cytoplasmic membrane.

The H-region shows a large degree of conservation with that of ϕKMV and is composed of hydrophobic residues, which tend to form an alpha α-helix, as tentatively confirmed by PSIPRED Figure 4.18).

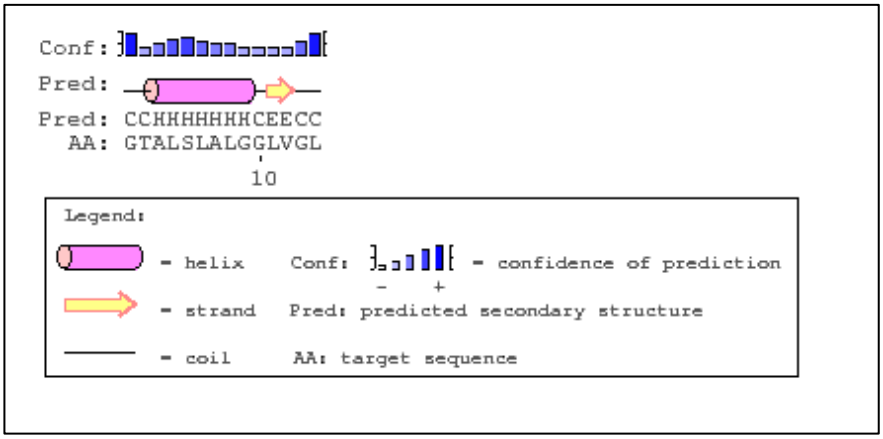

Figure 4.18. PSIPRED results of the H-region of the ϕNV3 signal-arrest-release domain illustrating the predictedα-helix.

This α-helix enhances insertion of the signal peptide into the phospholipid double layer. The catalytic domain shows again a high degree of amino acid sequence conservation between ϕNV3 and ϕKMV with the catalytic residues (E24, D33, T39) being identical [102].

4.3.7 Lysis cassette expression

To confirm the function and lytic activity of the predicted lysis cassette, the three genes that encode pinholin, endolysin and Rzl/Rz1-like of ϕNV3 were cloned individually in to the pEXP5-CT/TOPO® vector containing a T7 promotor and transformed in to BL21-AI™ competent cells, which carry the T7 polymerase gene under control of an arabinose induced promotor (*araBAD*) allowing for control over T7 RNA polymerase expression. The effect of expression was tested by inducing expression in exponential phase cells at a starting OD of 0.4 Abs by adding L-arabinose to a final concentration of 0.2% at T=0 in Figure 4.19.

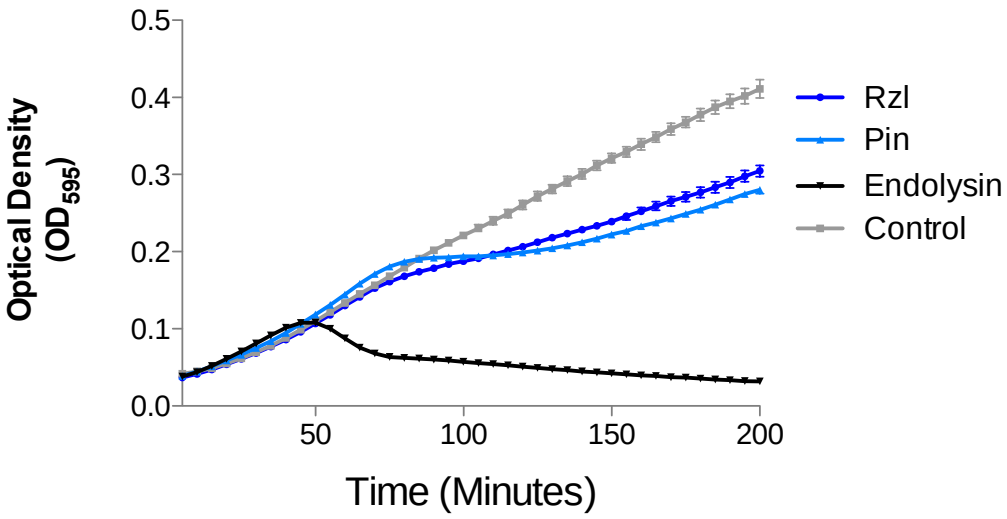

Figure 4.19. Effect of expression of individual lysis protein constructs containing either the ϕNV3 endolysin, pinholin or Rz/Rz1-like encoding genes on optical density of BL21-AI™ *E. coli* following induction with 0.2% L-arabinose at T=0. The control comprised of an uninduced BL21-AI™ containing the Rzl/Rz1-like construct. Mean values of 6 replicates are plotted, error bars= +/- SEM.

The results of the expression assays in Figure 4.19 show the strongest decrease in optical density, corresponding to a drop in bacterial cell numbers, is with expression of the BL21-AI™ *E. coli* containing the endolysin construct. The OD of the BL21-AI™ containing the endolysin construct begins to decrease rapidly at t=50 minutes from a peak of 0.108 Abs to 0.062 Abs at

t=80 minutes, a 43% reduction; after t=80 minutes the OD continues to gradually decrease over the remaining 120 minutes at a reduced rate, ending on a final OD of 0.032 Abs, a total decrease of approximately 70% from the peak OD at 50 minutes.

The OD of the pinholin expression assay continues to rise at a similar rate to the control until 80 minutes, at an OD of 0.187 Abs, where the increase in OD stalls until 115 minutes post induction at an OD of 0.197 Abs (an increase of 0.01 Abs, compared to a difference of 0.08 between 45 minutes and 80 minutes), it begins to rise after 115 minutes again, however at a lower rate; the final OD at 200 minutes is significantly higher than the endolysin expression, at 0.28 Abs.

The Rz/Rzl-like expression OD follows the control almost exactly until 75 minutes, where it begins to deviate, rising at a reduced rate in comparison to the control ending at a final OD of 0.305 Abs, higher than the 0.28 Abs of the pinholin expression, although significantly (P(T<=t) one-tail=2.9E-05) lower than the control which has a final OD at 200 minutes of 0.411 Abs.

To determine the effects of pinholin and Rz/Rz1-like protein complementation on the ability of the endolysin protein to lyse the host cells, a further two constructs were created containing either the pinholin and endolysin or the full lysis cassette (pinholin, endolysin and Rzl/Rzl-like protein) and expressed in BL21-AI™, the results are shown in Figure 4.20.

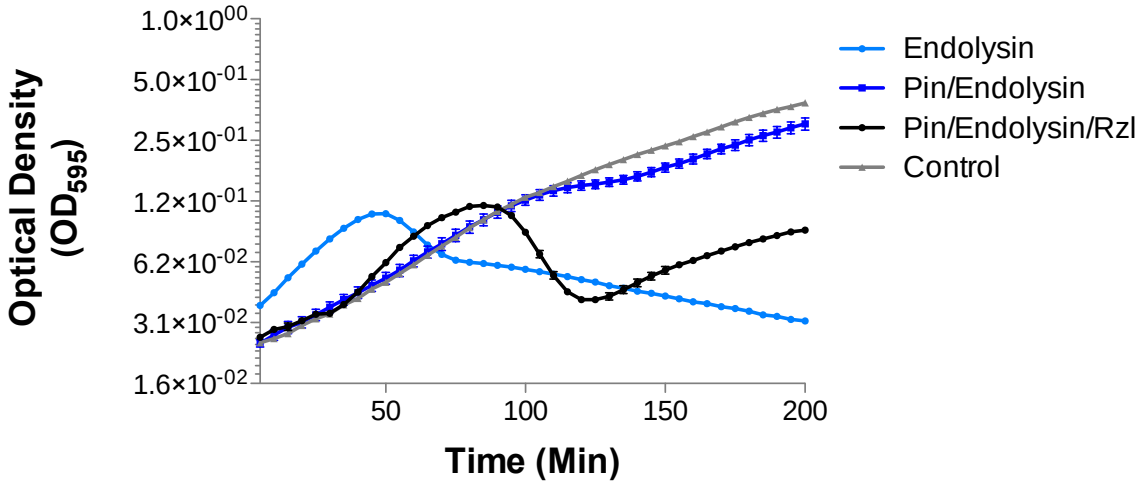

Figure 4.20. Complementation assay of lysis cassette proteins of ϕNV3 by expression of constructs containing either the ϕNV3 endolysin gene, pinholin and endolysin genes or the complete lysis cassette in BL21-AI™ *E. coli* following induction with 0.2% L-arabinose at T=0. The control comprised of an uninduced BL21-AI™ containing the Pin+Endolysin construct. Mean values of 6 replicates are plotted, error bars= +/- SEM.

The endolysin alone expression shows the most rapid increase in OD across all induced constructs as illustrated in Figure 4.20. The increase in OD continues to a peak at t=50 minutes of 0.108 Abs followed by a rapid drop, indicative of cell lysis, that continues to t=75 minutes when it is at an OD of 0.064 Abs, a decrease of 41%; after this time point, the decrease in OD continues over the time course of the experiment, albeit at a reduced rate similar to the results in Figure 4.19.

In contrast, the complete lysis cassette construct expression OD matches the rate of increase as the uninduced control until t=40 minutes, where the increase in OD is markedly more rapid than the control. The rapid increase continues to a peak at 85 minutes at an OD of 0.119 Abs, where it begins a rapid decrease to a minimum of 0.041 Abs at t=120 minutes, which corresponds to a decrease of ~66% compared to the maximum peak at 85 minutes. After this rapid decrease, the OD begins to rise slowly over the remaining time course of the experiment.

The construct containing both the pinholin and endolysin however matches the rate of increase in OD of the control up until t=105 minutes, where it begins to deviate with the OD showing no decrease, but instead a slower rate of increase, when compared to the uninduced control rate. By t=135 minutes the control OD is at 0.2005 Abs compared to 0.1598 Abs in the pinholin/endolysin expression. The pinholin/endolysin construct expression shows the largest standard error of the mean (SEM) of all constructs as indicated by the size of the SEM bars in Figure 4.20.

4.3.8 *P. agarici* Type IV pili knockout

Previous work by Chibeu *et al.* with the bacteriophage PhiKMV and its host *Pseudomonas aeruginosa* has demonstrated that the adsorption of PhiKMV requires Type IV pili [101]. Therefore, it was hypothesised that, as a member of the *phiKMVlikeviruses* it was possible phage ϕNV3 utilised Type IV pili as a binding receptor and to test this hypothesis it would be necessary to create a knockout of *P. agarici* NCPPB 2472 that was incapable of expressing the Type IV pili on its surface. It has been reported that mutations in the PilB gene results in nonpiliated cells [309], therefore it was chosen as the knockout gene to prevent Type IV pilus assembly in *P. agarici*. Two methods of generating competent *P. agarici* cells were attempted, chemical competency and electrocompetency.

For chemically competent *P. agarici* reactions no colonies were observed on any plates, including on both the positive and negative controls.

For the electrocompetent *P. agarici* cells no colonies were observed on any of the pCR2.1™-TOPO® construct transformation plates, however multiple colonies were visible on the pBBR1MCS-2 positive control transformation plates. Six of these colonies were chosen to be

analysed using PCR for the presence of the M13-F and M13-R regions, all 6 colonies returned positive results indicating the presence of the pBBR1MCS-2 vector.

4.4 Phage NV6

4.4.1 Host range

Phage NV6 was isolated simultaneously with ϕNV3 from the same untreated sewage sample and was identified as being able to form plaques on plates of both *P. tolaasii* NCPPB 2192T and *P. agarici* NCPPB 2472, however the plaques formed on *P. tolaasii* plates were significantly smaller (~1mm) and cloudy compared to the larger and clear plaques it formed on *P. agarici*. Phage NV6 did not infect any other *Pseudomonas* strain tested, as shown in Table 4.7.

Table 4.7. Host range assay of phage NV6.

Strain:	**NV6 Infection**
P. tolaasii NCPPB 2192T	+
P. agarici NCPPB 2472	+
P. sp NS1	-
P. syringae pv. morsprunorum R1 5244	-
P. syringae pv. syringae 9097	-
P. syringae pv. morsprunorum R2 5255	-
P. fluorescens Pfo-1	-
P. aeruginosa 14207	-

4.4.2 Growth characteristics

As shown in Figure 4.21, the kill curve characteristics of phage NV6 are almost indistinguishable from those of ϕNV3 under similar conditions (See Figure 4.14 in Section 4.2.4). At higher MOIs the drop in OD occurs faster, at an MOI of 0.1 the drop in OD begins at approximately t=150 minutes (approximately 3 standard phage replication cycles), decreasing from 0.007Abs to 0.003Abs by t=200 minutes. The corresponding drop in OD at the lower MOI's occurs at 180 minutes (0.017 Abs at t=180 to 0.006 Abs at t=240) at an MOI of 0.01 and 210 minutes (0.02 Abs at t=210 to 0.005 Abs at t=275 minutes) at an MOI of 0.001.

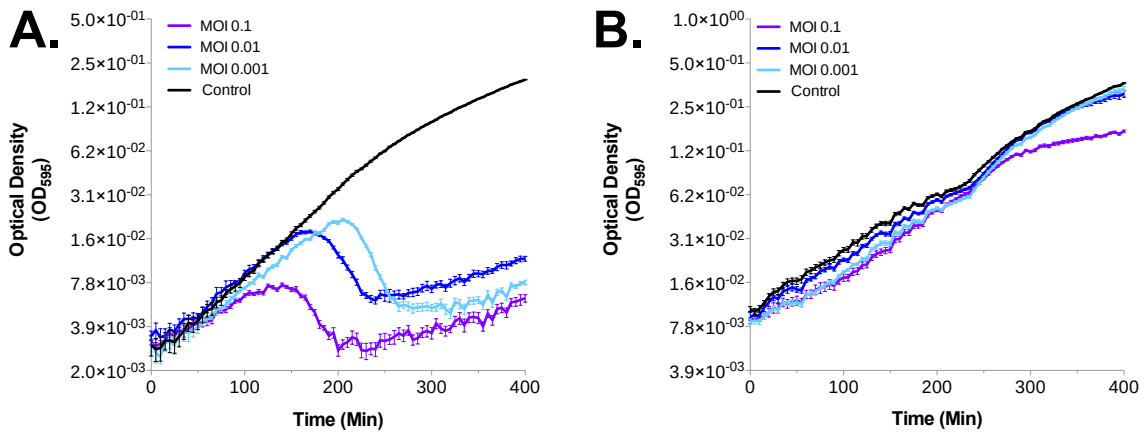

Figure 4.21. Kill curves of NV6 with A., *P. agarici* 2472 and B. *P. tolaasii* 2192T, at a starting bacterial concentration of ~1x10^8 cfu/ml, incubated at 28°C for 12 hours with a MOI of either 0.1, 0.01 or 0.001. OD at 595nm was read every 5 minutes. Mean values of 10 simultaneous replicates are plotted, error bars= +/-SEM.

However, the kill curve of NV6 with *P. tolaasii* is considerably different from that of NV6 with *P. agarici*, with only a small drop or stall in OD across all MOIs. At an MOI of 0.1there is a slight drop in OD between t=30 and t=45 minutes (0.012 Abs to 0.011 Abs), likewise at an MOI of 0.01 at t=50 to t=55 minutes, approximately equal to a single cycle of phage replication, there is a small drop in OD (0.014 Abs to 0.0138 Abs). However, the OD of all phage infected samples, independent of MOI are consistently lower across the entirety of the experiment than the uninfected control; with the final ODs correlating to the MOIs, with MOI of 0.1 final OD of 0.419 Abs, MOI of 0.01 final OD of 0.576 Abs, MOI of 0.001 final OD of 0.724 Abs and the uninfected control final OD of 0.805 Abs.

4.4.3 Sequencing

Initial PCR amplification identified NV6 to show striking resemblance genetically to ϕNV3 and it was hypothesised that NV6 was either closely related to ϕNV3 or had acquired a mutation that allowed it to infect a broader host range. To identify the any nucleotide changes between ϕNV3 and NV6 the genomic DNA was extracted and purified, returning a concentration of 201 ng/µl which was then sent for paired-end sequencing performed by Source Bioscience.

The sequencing results returned a single contig of 43,217 bp for NV6, with a coverage of 54.4x, when compared to the figure of 43,184 bp for ϕNV3, there was an addition of 33 bases. Alignment of the returned NV6 sequence to the genome of ϕNV3 by BLASTn [253] indicated that these 33 bases were in a single large addition. This 33 bp addition was identified to be within phiNV3_p40, a 2,367 bp gene predicted to encode a putative T7-like tail protein, an alignment of the nucleotide sequence is shown in Figure 4.22.

```
1635 TAAAAACTTGCA--------------------------G-----GGGCATGCCTTTGGGCC 1664
1635 TAAAAACTTGCATGGGCATGCCTTTGGGCCGGGCGTACGTGATATCGGGCATGCCTTTGGGCC 1697
```

Figure 4.22. DNA sequence comparison of a section of the T7-like Tail Fiber Protein gene of top, ϕNV3 and below, NV6, showing the base additions highlighted in grey.

The 33 bp addition maintained the single reading frame of the phiNV3_p40 tail protein and when the amino acid sequence was analysed in Jalview 2.9.0b2 it was shown to have added a sequence of 11 aa (-PGVRDIGHAFG-) towards the C-terminal, and two amino acid substitutions upstream and downstream of the addition region, which is illustrated in the amino acid alignment in Figure 4.23.

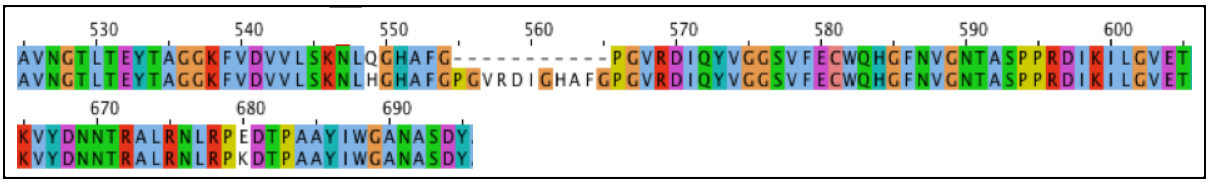

Figure 4.23. Amino acid sequence comparison of a section of the T7-like tail fiber protein gene of top, ϕNV3 and below, NV6.

4.5 Discussion

The phage *Pseudomonas* phage NV1 was initially isolated from water samples taken from the River Thames and was identified as forming very small hazy plaques on lawns of lawns of *P. tolaasii* 2192T. Hazy plaque morphology is often associated either with temperate phage or with inefficient phage adsorption to the host[310], however it is worth noting that no genes associated with lysogeny were identified within the NV1 genome. NV1 was identified as

having a narrow host range after failing to show evidence of plaques on any further *Pseudomonas* strains tested. From the virion morphology, shown in Figure 4.1 I was able to identify NV1 as belonging to the *Podoviridae* family of the *Caudovirales* order, a large order to which ~96% of all phage examined by electron microscopy occupy [311].

The genome of NV1 was isolated and sequenced by Source Bioscience, which comprised of 45,059 bp containing 64 identified ORFS, of which only 25 (~40%) showed amino acid identity to proteins of known or predicted function. Analysis of the returned contig revealed a high degree of nucleotide identity with the *Pseudomonas fluorescens* phage UFV-P2 (84% Ident over 86% of the genome). The genome of NV1 was then oriented to be collinear with that of phage UFV-P2, annotated and submitted to GenBank (Accession code pending). The International Committee on Taxonomy of Viruses (ICTV) guidelines recommend DNA sequence identity of 95% as a threshold for species delineation, with phage NV1 showing a maximum nucleotide identity of 84%, it is below the species delineation threshold[312].

The differing host specificity, GC content (52.87% NV1 to 51.5% UFV-P2) and genome size would indicate that phage NV1 is a new species, albeit with significant similarities in gene synteny and significant amino acid and nucleotide similarity of several predicted genes.

Likewise, with full genome phylogenetic analysis as shown in Figure 4.4, phage NV1 shows tight clustering with the closely related UFV-P2, therefore I suggest that NV1 is a new species of "UFV-P2" like virus and that the genome and results of this study provide evidence for the potential creation of a "*UFV-P2virus*" genus, distinct from the *Luz24likevirus* genus to which UFV-P2 has previously been attributed [108].

The results of the kill curve experiments in Figure 4.5 and 4.6 show that phage NV1 is capable of lysis of *P. tolaasii* 2192T *in vitro*, however the minimum MOI needed to cause a decrease in OD that would correspond to a decrease in total bacterial numbers *in vitro* is 1, which would

be difficult to achieve in a real-world setting. While at lower MOIs the time required to show a slowing in bacterial growth is increased and a drop in OD that would correspond to a drop in total bacterial cell counts is never observed, illustrating that lower MOIs of phage treatment may not suitable for use in phage therapy.

As is also shown in the low MOI kill curve results, illustrated in Figure 4.6, *P. tolaasii* appears to be able to rapidly develop resistance to phage NV1 and it is possible this ability of *P. tolaasii* to gain resistance is related to a shift in phenotype (smooth/rough) [14]. Many phages, such as the phage LKA1 of *P. aeruginosa*, are dependent on the expression of particular LPS receptors that are related to phenotype [155] and it is possible that phage NV1 infection is dependent on a specific LPS phenotype. However, this change in phenotype can result in *P. tolaasii* shifting from pathogenic to non-pathogenic [21] and therefore the pressure of developing phage resistance may reduce its infective potential.

Further tests *in vivo* would be required to determine if phage infection and subsequent *P. tolaasii* resistance development would reduce or prevent disease symptoms on *Agaricus bisporus*; similar to those performed by Kim *et al.* with phage hb1a, where they were able to suppress blotch formation on *P. ostreatus*, with the addition of phage simultaneously with *P. tolaasii* [74].

Pseudomonas phage ϕNV3 was initially isolated from untreated sewage water samples and identified as capable of forming large (5 mm), clear plaques on lawns of *P. agarici* NCPPB 2472. It was identified as having a narrow host range after failing to show evidence of plaques on any further *Pseudomonas* strains tested. TEM imaging of the virion morphology, shown in Figure 4.7 allowed us to identify ϕNV3 as belonging to the *Podoviridae* family of the *Caudovirales* order, similar to phage NV1.

The genome of ϕNV3 was isolated and sequenced by Source Bioscience, and identified to be 43,184bp in length; the genome of ϕNV3 was also identified as showing genome organisation similarities with the reported organisation of *Pseudomonas* phage ϕKMV [100], which was then used to orient the genome of ϕNV3 correctly. The genome sequence and orientation was then confirmed with extensive primer walking and Sanger Sequencing. The genome of ϕNV3 has an average GC content of 58.29% which is significantly lower than ϕKMV (62.3%) [100], but closer to Pseudomonas phage LKA1 (60.9%) [308] and close to the host *P. agarici* NCPPB 2472 GC content (60.55%).

In total 50 potential gene encoding ORFs were identified in the genome of ϕNV3, of these only 25 showed amino acid identity to phage proteins of known or predicted functions as identified by BLASTP [253], shown in Table 4.5, consequently 50% of the ORFs have no known or predicted function, 29 ORFs (58%) have orthologs found in phage ϕKMV; however, the genome of ϕNV3 shows very little nucleotide similarity to any known phage with a very low similarity to ϕKMV with a nucleotide identity of 77% over only 3% of the entire genome.

Similar to phage NV1 the majority of genes with identified functions are phage structural proteins and those involved in DNA replication, such as the single subunit RNA polymerase upstream of DNA replication genes, adjacent to the structural protein region of the genome, which is characteristic of the *Autographivirinae* subfamily of bacteriophage and similar to phage ϕKMV and other ϕKMV-like phages [100, 308].

The results of the full genome phylogenetic analysis likewise show that ϕNV3 clusters with other members of other members of the *phiKMVvirus* genus, although on a separate branch to

the majority, indicating that it may have diverged earlier than other members, similar to phage LKA1.

Current phage classification guidelines based on nucleotide identity state that the species demarcation is set at 95% identity and genus at >50% identity [312]. From the genome sequencing results including the apparent lack of nucleotide identity (77% identity over 3% of the genome to phage ϕKMV), the similarities of genome organisation to ϕKMV [100] and the results of the phylogenetic analysis I propose that ϕNV3 is most likely a new species of the *phiKMVvirus* genus, however it is possible due to the extremely low nucleotide identity that ϕNV3 represents a new bacteriophage genus within the *Autographivirinae*, closely related to the *phiKMVviruses*.

The life cycle of phage ϕNV3 was characterised in order to identify whether it would represent a possible candidate for biocontrol applications and to understand its interactions with the host *P. agarici* NCPPB 2472. It was identified via adsorption assay that ϕNV3 is capable of efficient adsorption to host *P. agarici* NCPPB 2472 cells with an adsorption constant of k_a=3.5x10^{-9} ml min^{-1}. This adsorption rate is similar to that of the *P. aeruginosa* phage LKA1 (k_a=3.9x10^{-9} ml min^{-1}) although less than half of the value for *Pantoea virus* LIMEzero (8.29x10^{-9} ml min^{-1}), while it has been reported that both phage LKD16 and ϕKMV failed to produce clear adsorption curves due to their inefficient adsorption [103, 308].

Likewise, it was identified via single step growth profile that the latent period of ϕNV3 is approximately 45 minutes, with a rise period of 10 minutes. The lysis of host cells after 45 minutes post infection is longer than the time required for cell lysis of other members of the ϕKMV-like phage such as LKD16, ϕKMV and LKA1 (25, 35 and 40 minutes respectively

[308]). The calculated burst size of approximately 55 phage per cell for ϕNV3, is identical to that of ϕKMV although lower than LKD16 and LKA1 (120 and 255 respectively) [308].

The results of the kill curve assays in Figure 4.12 and 4.13 show that at high MOIs such as a MOI of 10 the total lysis of all host cells is rapid, occurring between 25 and 60 minutes, and that at MOI of 1 the time required for total cell lysis increases to 110 minutes. However, at MOIs lower than 1 total bacterial cell lysis is not seen and the emergence of resistance to ϕNV3 is rapid indicating that for the use of ϕNV3 in biocontrol applications, a high phage titre would need to be used.

This emergence of resistance is similarly displayed in the results of the co-evolution study illustrated in Figure 4.15, with up to 51% of bacteria resistant to the 'current' phage transfer within one 24-hour cycle. However, they appear to be less resistant to 'future' phage for bacterial transfer numbers 1, 3 and 5 which may indicate that the development of phage resistance by *P. agarici* at this stage comes at cost and may be due to phase variation, similar to *C. jejuni* which can alter the expression of the polysaccharide receptor required by phage F336, or a change in bacterial phenotype similar to that of *P. tolaasii* [14]. The resistance to 'present' phage increases to 79% by Transfer 5 and by Transfer 9 the bacteria are resistant to all 'past' 'present' and 'future' phage; this complete resistance to all phage assayed in transfer 9 may be due to the CRISPR/Cas system of *P. agarici* NCPPB 2472 identified in Section 3.3.6 and the acquisition of a spacer region associated with ϕNV3. This hypothesis could be tested by PCR amplification of the CRISPR repeat region and sequencing of the resulting product to identify any newly acquired spacer regions.

While whole phage treatments of bacterial infections have been well documented in literature, phage lysis proteins have also been studied intensively, as potential enzybiotics treatments of bacterial infections [111, 112, 134] and have shown promise *in vivo* [135]. For this reason, the lysis cassette of ϕNV3 was characterised in detail.

ϕNV3 possesses a SAR-endolysin system comprised of 4 proteins, the pinholin, endolysin, Rz and Rzl-like proteins, a system that is conserved in all *phiKMVviruses* [102] and differing from the canonical system putatively identified in phage NV1. As outlined in Section 4.3.6 the SAR-endolysin system of ϕNV3 shows a great deal of conservation with that of ϕKMV, especially within the N-terminal *signal-arrest-release* domain of the endolysin protein as illustrated in Figure 4.16, as well as sharing conserved catalytic residues within the catalytic domain. This would indicate that the lysis cassette of ϕNV3 likely operates with a similar mechanism as that reported for ϕKMV by Briers *et al.* [102].

In order to confirm the predicted function of the endolysin proteins and associated proteins of the lysis cassette the proteins were cloned into the pEXP5-CT/TOPO® vector and transformed in to BL21-AI™ cells for expression. The results of these expression assays are shown in Figures 4.18 and 4.19 and confirm the lytic ability of the ϕNV3 endolysin protein, even without the presence of the pinholin and Rz/Rzl-like proteins, although the greatest drop in bacterial cell numbers is correlated with the full lysis cassette protein expression although the time required for this drop is longer in comparison to the endolysin protein expressed alone, however this may be attributed to the metabolic pressure of expressing multiple proteins simultaneously compared to a single endolysin protein.

Expression of the lysis proteins of NV1 were also performed, the results of which are available in Appendix 4, Figures Ap4.7 and Ap4.8; however, the results demonstrated limited differences in optical density between the induced and uninduced controls. This limited differences in induced versus uninduced in comparison to the distinct differences seen in the expression of the ϕNV3 lysis proteins may be due to the different lysis systems involved (canonical NV1 system versus the SAR-endolysin system of ϕNV3) where we would expect the lysozyme of NV1 to accumulate in the cytosol as it would lack access to the cell wall, although noticeable differences would be expected with the holin protein expression due to the large pores formed in the cell membrane. The lack of OD reduction, especially in Figure Ap4.7 would appear to indicate that the proteins had not been expressed at all, as a small reduction due to the metabolic stress of protein production would be expected, which may account for the slight differences in OD visible towards the end of the experiment in Figure Ap4.8. However, further experiments would be required to confirm protein expression or the lack thereof in both construct expressions.

The final phage characterised, NV6, was initially identified on mixed culture plates of *P. tolaasii* 2192T from untreated sewage and was originally thought to be a unique phage; however, it was swiftly noted to possess the ability to cause plaques on both *P. tolaasii* NCPPB 2192T and *P. agarici* NCPPB 2472 and to bear striking similarities to phage ϕNV3. Kill curve assays, as illustrated in Figure 4.21 demonstrated that NV6 possessed a significantly greater ability to lyse host cells of *P. agarici* NCPPB 2472 than *P. tolaasii* 2192T, with only minimal changes in growth demonstrated with *P. tolaasii*.

Consequently, the full genome of phage NV6 was isolated and sent for sequencing by Source Bioscience in order to determine if phage NV6 was a mutant of ϕNV3 and if so, whether there

was a genetic difference that may account for the broader host range. It was identified that the only difference between the genomes of NV6 and phiNV3 is a single addition of 33 bases, within phiNV3_p40, which encodes a putative T7 tail protein. It is therefore likely that this region within the T7-like tail protein of ϕNV3 is a determinant of host specificity potentially involved in receptor binding, as demonstrated by the broader host range of phiNV6.

Interestingly while the full genome molecular phylogenetic analysis of phylogeny of both phage appears to correctly place both NV1 and ϕNV3 with members of the same phage families, similar to the further analysis performed using only the large terminase subunit gene, available in Appendix 4 Figure Ap4.6; the analysis only utilising the large terminase places members of the *Myoviridae* and *Siphoviridae* on branches alongside members of the *Podoviridae*. This may indicate that in future studies the approach of full genome phylogenetic analysis, while computationally heavy may yield significant increases in accuracy due to the much larger number of sites used for the analysis, 26,547 and 24,899 sites used to produce Figures 4.10 and 4.4 respectively in comparison to 804 sites used in Figure Ap4.6.

Finally, while the attempts to create a *P. agarici* NCPPB 2472 Type-IV pili knockout using the pCR2.1™-TOPO® vector were unsuccessful, possibly be due to degradation by the identified Type 1 CRISPR system previously identified. It was however demonstrated that the method utilised in creating the electrocompetent cells was successful using the pBBR1MCS-2 vector, and so this cell preparation method could be used in further studies utilising other vectors.

Chapter 5:
RNA Sequencing of *P. agarici* and ϕNV3 infected *P. agarici*

5.1 Introduction

Having successfully sequenced and assembled both the phage, ϕNV3 and the host, *P. agarici* NCPPB 2472 genomes, the next step was to analyse their interactions on a transcriptional level with the aim of potentially elucidating the mechanisms involved in developing phage resistance. In order to do this RNA sequencing and differential gene expression profiling was chosen in order to maximise the number of genes that could be analysed simultaneously.

P. agarici NCPPB 2472 cells infected with phage ϕNV3 at an initial MOI of 1 to ensure that the majority of cells were infected to allow for the largest change in RNA transcription, while avoiding potential lysis from without that can occur at higher MOIs.

5.2 RNA isolation and purification

In total 6 total RNA samples were prepared, 3 non-phage infected controls (P4-P6) and 3 phage infected samples (PN4-PN6). The RNA was extracted from homogenised cell cultures at 45 minutes post phage infection, to correspond with mid to late log-phase phage growth (see single step growth curve, Section 4.3.4, Figure 4.12), which would allow the maximum bacterial response to infection at a transcriptional level.

The RNA was then concentrated to a level appropriate for RNA sequencing and the resulting concentrations varied from a low of 96 ng/µl (P6) to a high of 358 ng/µl (P4) as ascertained by an Agilent 2100 Bioanalyzer, the full results including RNA Integrity Scores (RIN) are shown in Table 5.1. In addition to the RNA integrity scores, the bioanalyzer returned gel electrophoresis images of the RNA samples, which allowed visual interpretation of the RNA integrity as determined by the clarity of the 16S and 23S rRNA bands in the resulting gel image. The gel images are shown below in Figure 5.1.

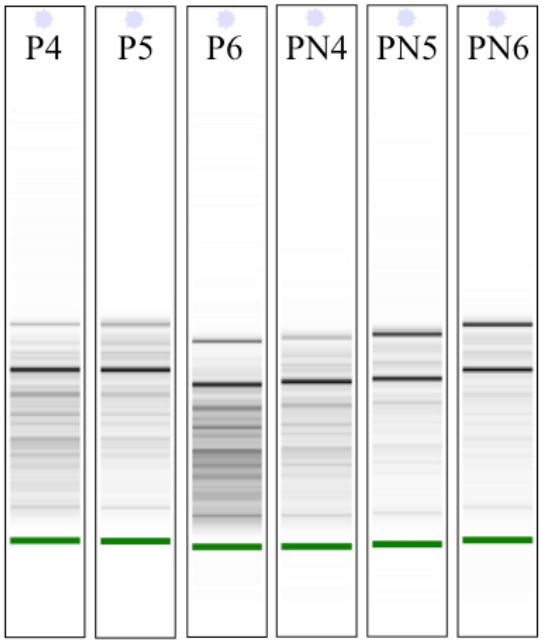

Figure 5.1. Gel images of RNA samples used for RNA sequencing.

As is shown in Figure 5.1 above, the samples with the darkest and most clear 16S and 23S bands, the dark bars located towards the top of the gel image, are samples PN5 and PN6. Both these samples also show minimal low molecular weight smearing which is most evident in sample P6. This is further confirmed by the RIN scores or 7.2 and 7 for PN5 and PN6 respectively, in Table 5.1.

Table 5.1: RNA sample concentration, RIN and total mapped reads for uninfected *P. agarici* NCPPB 2472 (P4,P5,P6) and ϕNV3 infected (PN4, PN5, PN6).

Sample	Concentration (ng/ul)	RIN	Aligned Reads
P4	358	5.4	68734869
P5	296	6.2	51438405
P6	235	5.1	61971656
PN4	242	5.9	60307697
PN5	96	7.2	50277057
PN6	194	7	70485550

All passed quality control with FastQC, however both P5 and PN5 had a 50 bp sequence of 'G' which accounted for 0.172% and 0.137% of the total reads respectively. The FastQ sequencing read files returned by Source Bioscience were aligned to a Fasta sequence file containing the concatenated genomes of *P. agarici* NCPPB 2472 and phage ϕNV3, the total number of aligned reads per sample is also shown in Table 5.1.

The aligned read file was then used to create counts for each gene of *P. agarici* NCPPB 2472 identified via the NCBI Prokaryotic Genome Annotation Pipeline, the total counts are illustrated in Table 5.2.

Table 5.2: Mean read counts for uninfected *P. agarici* NCPPB 2472 (P4,P5,P6) and ϕNV3 infected (PN4, PN5, PN6).

Sample	Read Count
P4	30264541
P5	21693870
P6	27264739
PN4	11884370
PN5	10384432
PN6	11550021

The mean counts for the uninfected samples (P4-P6) in Table 5.2 is 26,407,717 however, for the phage infected (PN4-PN6) the mean counts is 57% lower at 11,272,941.

A principal components analysis (PCA) plot was constructed from the regularized-logarithm transformed (rld) results in order to identify any clustering based on phage infection or bacterial cell line, the results are shown in Figure 5.2.

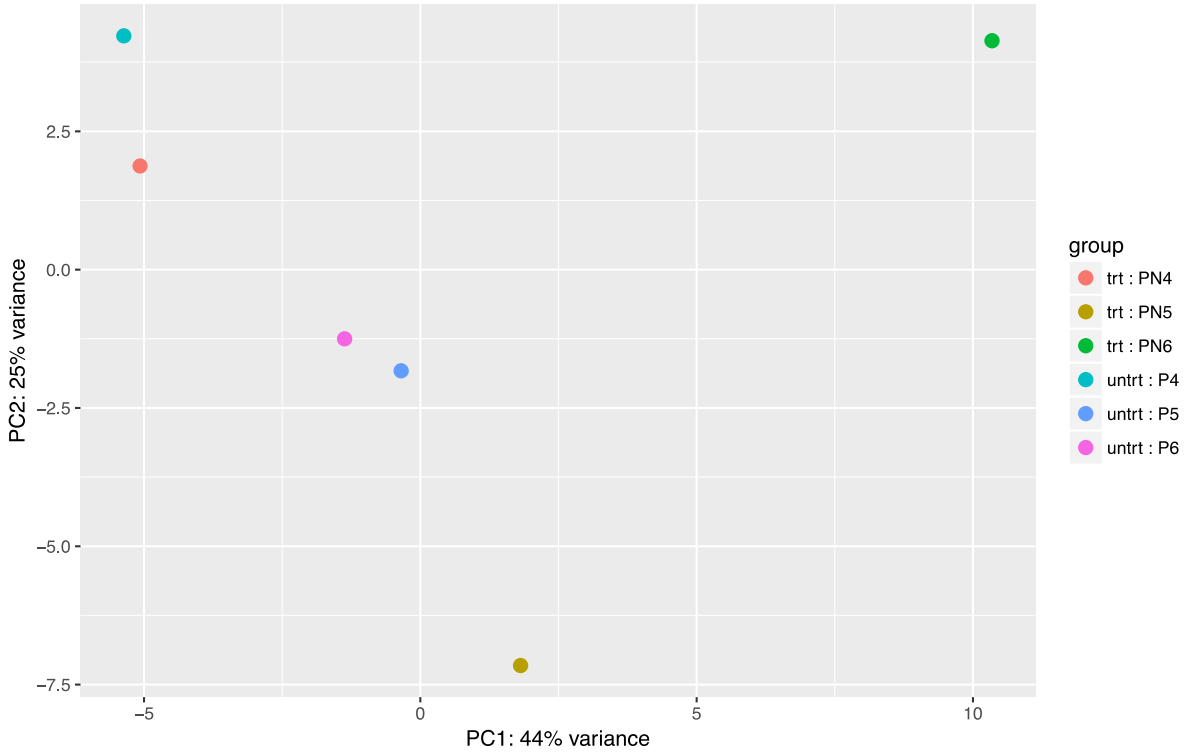

Figure 5.2. PCA plot generated for the *rlog*-transformed values of mapped reads, using *plotPCA* function in the DESeq2 [267] package and plotted using the *ggplot2* package [269].

From the plot in Figure 5.2 we can see that while there is significant clustering of the untreated (uninfected) samples (P4-P6) in both the x and y axis; there is no apparent clustering of the infected samples which are spread across both axes, with both samples PN5 and PN4 located close to the untreated in the y axis and PN4 clustering with the uninfected samples both in the x and y axes.

5.3 Phage gene expression

To determine the level of phage infection per sample, as determined by the highest total level of phage gene transcription, the RPKM values were obtained for each phage gene per sample then a total calculated, as shown in Table 5.2.

Table 5.2. Total RPKM values of phage infected *P. agarici* NCPPB 2472 samples.

Sample:	Total (RPKM):
PN4	747046.571
PN5	684582.436
PN6	859655.397

Total RPKM value for each sample, shown in Table 5.2 shows that the highest levels of phage gene transcription are in sample PN6 followed by sample PN4, then PN5.

To analyse the phage genes with the highest levels of transcription a bar graph of the mean RPKM values per gene and the standard deviation was plotted, shown in Figure 5.3.

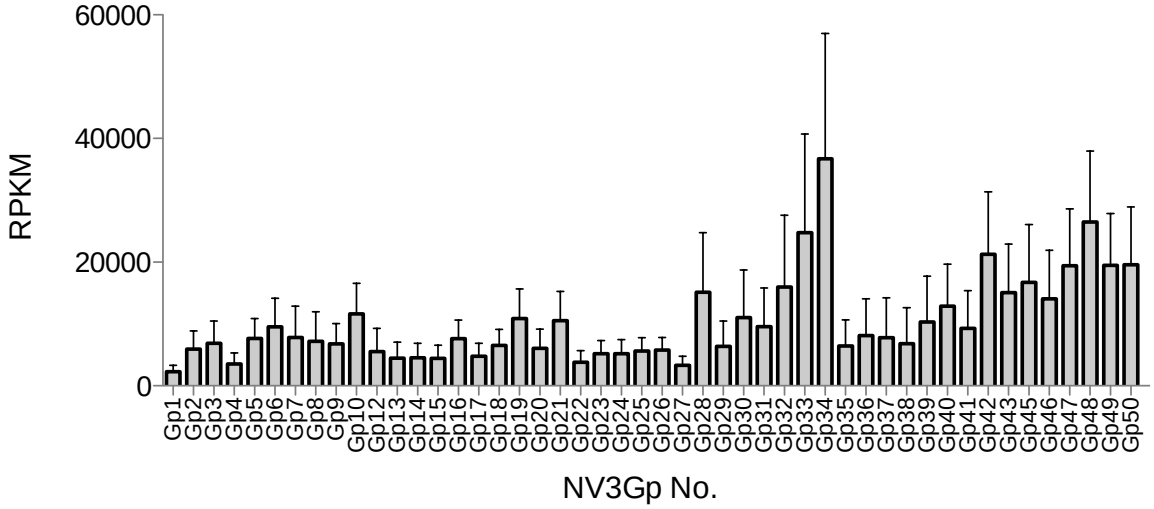

Figure 5.3. Mean RPKM values and STDev of reads per putative gene encoding ORF of phage ϕNV3 as calculated using Artemis v16.0.0.1 genome browser.

As shown in Figure 5.3, the genes with the largest levels of transcription (Gp28-34 and Gp42-50) are clustered towards the right of the graph, corresponding to the mid to late genes of the phage genome. The two genes with the highest level of expression are Gp33 and 34, encoding a putative capsid protein and scaffolding protein respectively.

5.4 Differential gene expression

In total, the expression of 4895 *P. agarici* NCPPB 2472 genes were analysed, 9 returned padj of below 0.01 (0.18%), 24 with a padj of less than 0.05 (0.49%) and 39 with a padj of less than 0.1. Of the 39 genes with a padj of less than 0.1, 26 were up-regulated and 13 were down regulated, the full differential gene expression results are visualised in an MA-plot in Figure 5.4. Full results of all genes with a padj of <0.1 is available in Appendix 4, Table Ap4.1

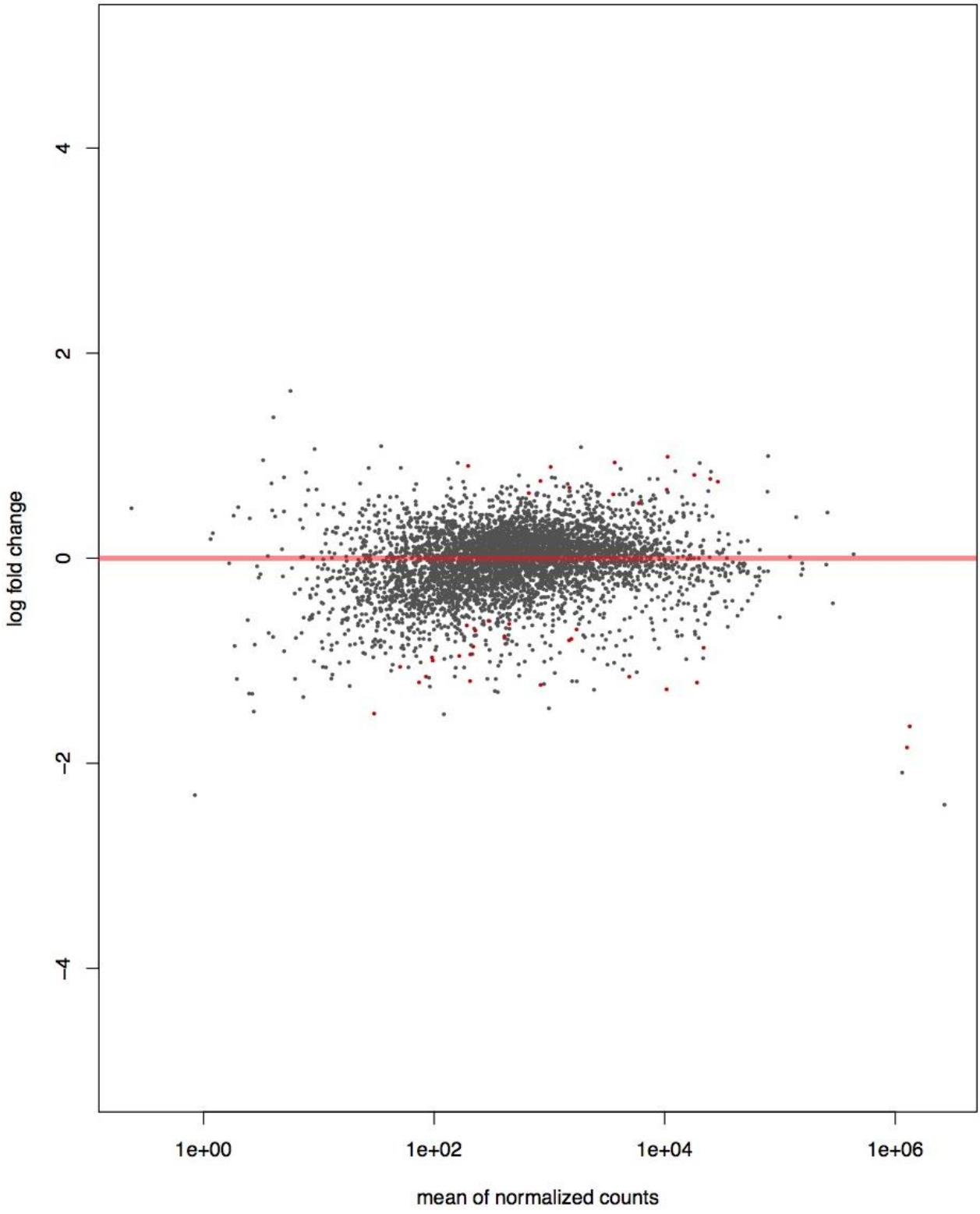

Figure 5.4. MA-plot of DEseq2 [267] comparison results of phage treated vs. untreated samples, genes with an padj value below the threshold of 0.1 are indicated in red.

Further analysis of the differential gene expression data revealed an operon of four genes that showed a mean log2 fold increase of 1.15 between the control and infected samples, with padj values of below 0.05. The genes in this operon and their precise log2 fold changes are outlined below in Table 5.3.

Table 5.3. Differential gene expression data of an unidentified operon of *P. agarici* NCPPB 2472.

Locus Tag	Predicted Function	Log2Fold Change	Padj
RS16090	DUF3050 domain-containing protein	1.159423991	0.0216997
RS16095	P-aminobenzoate N-oxygenase AurF	1.002180797	0.0280734
RS16100	peptide transporter	1.239879524	0.0003224
RS16105	SRPBCC family protein	1.202543288	8.38E-06

The normalised count data for the differentially expressed operon outlined in Table 5.3 is visualised graphically in Figure 5.5.

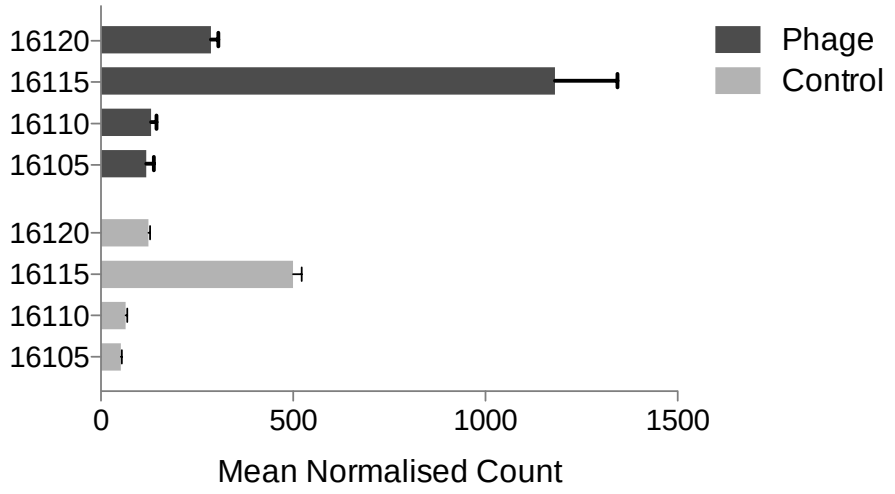

Figure 5.5. Mean normalized count data and SEM+ for gene cluster.

A large number of the differentially expressed genes were identified as hypothetical proteins, however the host DNA polymerase II (RS18970) showed a log2fold increase of 0.694 in expression in phage treated samples.

The tmRNA encoding gene *ssrA* shows significantly increased expression in phage treated samples with a log2fold increase of 1.64 and a padj of 8.67E-15. However, the gene encoding the associated ATP-dependent zinc protease HflB is under-expressed with a log2fold decrease of 0.77 and a padj of 0.019. A Clp protease ATP-binding subunit (RS14345) also shows similar levels of under-expression with log2fold decrease of 0.75 and a padj of 0.02.

5.5 Discussion

The RNA of phage infected and non-infected *P. agarici* NCPPB 2472 was isolated with a wide variation of final concentrations and RIN scores. However, all samples were successfully sequenced by Source Bioscience, with the resulting FastQ files being aligned to a concatenated sequence of the phage ϕNV3 and *P. agarici*, with the number of aligned reads varying between 50.2-70.4 million, shown in Figure 5.1.

These aligned reads were then counted by gene for *P. agarici* NCPPB 2472 as previously identified by the NCBI Prokaryotic Genome Annotation Pipeline (as described in Section 3.3.3), using the *summarizeOverlaps* function of the *GenomicAlignments* package of Bioconductor. The number of reads counted are shown in Table 5.2, with a significantly ($p=0.018$) lower number of counts for *P. agarici* genes in the phage infected than in the uninfected samples, with a 57% reduction in mean reads mapped to the bacterial genome. This reduction is likely due to the high levels of phage gene transcripts present in the infected samples and host RNA degradation at this particular timepoint, as has been previously reported by Lavigne *et al.* [313] with phage LUZ19 and *P. aeruginosa* PAO1 where it was observed that the total proportion of phage RNA to bacterial RNA increased from 23.6% to 60.2%

between 5 to 15 minutes after infection, which corresponds to early to mid/late stages of infection of the LUZ19 replication cycle.

A PCA plot was constructed from the rld results in order to results in order to identify any initial clustering based on phage infection or bacterial cell line, shown in Figure 5.2, from which it is apparent that the phage infected sample PN4 cluster heavily with the non-infected samples, likewise PN5 is remarkably close to the non-infected cluster. Also evident is that the phage infected sample PN6 is the furthest outlier of the infected samples in both the x and y axes. This may broadly indicate that phage treatment does not have a significant effect on bacterial transcription levels at this time point of the phage lifecycle, however it is also likely due to the MOI of phage used. At a MOI of 1, which was the MOI used in this study, the total percentage of infected cells would be approximately 63.2% and the percentage uninfected approximately 36.8%, this would indicate that over a third of the returned RNA would be from uninfected *P. agarici* which when combined with the high levels of phage transcription and normalisation process could obscure or reduce differences in gene expression in the infected samples; in future studies a higher MOI of between 5 (99.3% infected) to 10 (100% infected) would prevent this source of error.

Phage gene expression was analysed using the RPKM values calculated using Artemis v16.0.0.1 genome browser, due to difficulties in constructing a functioning feature file containing both bacterial and phage genes. Total RPKM values of all putative genes identified in Section 4.2.2, illustrated in Table 5.2, which shows that the highest levels of phage gene transcription are in sample PN6 followed by sample PN4, then PN5.

This appears to contradict the data shown in the PCA plot in Figure 5.2, which shows that sample PN4 clusters more closely with the uninfected samples. If the clustering was due to

level of phage gene transcription and thereby the level of phage infection we would instead expect to see sample PN5 to be clustered more with the uninfected than PN4.

The mean RPKM values for each putative phage gene of ϕNV3 are shown graphically in Figure 5.3 which shows that the genes with the highest level of transcription at this time point of phage infection are between Gp28-34 and Gp42-50, which correspond to the mid to late genes of the ϕNV3 genome. The genes with the highest levels of transcription are Gp34 (36,736), Gp48 (26,452), Gp33 (24,757) and Gp42 (21,263) which correspond to the capsid protein, Rzl-like protein, scaffolding protein and a novel hypothetical protein respectively. The sampling time point chosen of 45 minutes for our samples corresponds to the beginning of the rise period of phage ϕNV3, as demonstrated in the single step growth assay previously performed (Figure 4.12) in Section 4.2.4, therefore we would expect to see the highest level of transcription in genes involved with progeny virion production and host cell lysis as observed. Likewise, the lowest levels of transcription observed were in the early, host preparation, and mid genes (ORFs Gp 1-30).

These results are similar to those reported by Zhao *et al.* [314] in their transcriptomic analysis of the interaction between *P. aeruginosa* and the temperate phage PaP3, where they demonstrated that the late genes of PaP3 were expressed beginning at 30 minutes, the beginning of the rise period in the single step growth curve of PaP3 and the 'early' and 'middle' genes peaking in expression between 5-10 minutes and 10-30 minutes respectively. Likewise, these phage expression results are similar to those also reported by Lavigne *et al.* [313] with phage LUZ19, where at the mid/late timepoint the highest levels of gene transcription was observed in ORFs corresponding to capsid and head decoration proteins.

Differential gene expression analysis of phage infected *P. agarici* and non-infected *P. agarici* samples was performed using DESeq2, in order to potentially identify expression changes in genes previously identified in the *P. agarici* genome associated with phage resistance. Of the 4,895 genes analysed only 39 (0.79%) differentially expressed genes had a padj of <0.1. Of these 39 genes, 26 were upregulated and 13 downregulated in the infected verses non-infected samples. This is in sharp contrast to the previously mentioned study by Zhao *et al.* [314], who reported a figure of 38% of the 5633 assayed genes of *P. aeruginosa* being differentially expressed in PaP3 infected versus non-infected cells with 98% of the differentially expressed genes being down regulated in the infected cells; as well as the previously mentioned study by Lavigne *et al.* [313] who identified 220 genes in *P. aeruginosa* PAO1 that were up-regulated and 13 operons that were downregulated during LUZ19 infection.

Analysis of the differential gene expression data revealed the presence of a single four gene operon, the precise function of which is unknown, which showed a log2 fold increase in expression of approximately 1.15 between the control and the infected samples, as shown in Table 5.3. While the majority of genes identified to be differentially expressed were predicted to be hypothetical proteins the host DNA polymerase II (RS18970) showed a small log2fold increase (0.694) with a padj of 0.017.

The tmRNA encoding gene *ssrA* shows significantly increased expression in phage treated samples with a log2fold increase of 1.64 and a padj of 8.67E-15. tmRNA is involved in trans-translation and the release of all the components of a translation complex which has stalled during protein synthesis and tagging the unfinished peptide for degradation and facilitates the degradation of the malfunctioning mRNA [315]. However, the gene encoding the associated

ATP-dependent zinc protease HflB is under-expressed with a log2fold decrease of 0.77 and a padj of 0.019.

No statistically significant gene expression was identified in any of the phage resistance systems previously identified in the *P. agarici* NCPPB 2472 genome in Section 3.2.6.
This included in any of the identified CRISPR-associated proteins, with padj values varying between 0.99 and 0.97; it has been reported that *cas1* and *cas2* expression is required for new spacer region acquisition [244], we would therefore expect to see an increase in the expression of the *cas1* and *cas3* genes (RS05180, RS05185) if the cells were undergoing the acquisition of phage spacers at this time point.

It is likely that the differential gene expression data is not statistically valid when taking into consideration the extremely low percentage of differentially expressed genes (0.79%) and that this was due to the large difference in reads attributed to bacterial genes compared to bacteriophage genes caused by the extremely high levels of phage gene transcription and possible host RNA degradation. The consequent high degree of normalisation required would then most likely mask any true differentially expressed genes.

In any future work, as previously mentioned, a significantly higher MOI of between 5-10 would be needed in order to ensure all cells were infected with phage, as well as using a timepoint earlier in the phage replication cycle in order to reduce the imbalance in phage to bacterial transcripts; furthermore using RT-qPCR for identifying differential transcription in bacterial genes would by-pass the issue of large numbers of phage transcripts, however this approach would be impractical for large numbers of bacterial genes, therefore sequencing to a higher depth may increase the likelihood of changes in bacterial transcription being observed.

Chapter 6:
Discussion and Proposed Further Work

The purpose of this study has been to elucidate the genetics of the phage-bacterium-host interactions of mushroom pathogenic *Pseudomonas* species.

I have successfully isolated and sequenced the genome of the mushroom pathogen *P. tolaasii* 2192T, which is available under the accession code NZ_CP020369, including the identification of a total of 6,286 putative gene encoding ORFs. Six NRPS proteins were identified within the *P. tolaasii* 2192T genome that are orthologous to those required for tolaasin biosynthesis, although further work would be needed to confirm these predicted functions.

A putative pyocin encoding gene cluster was likewise identified within the genome that is likely of phage origin, that will also need further work to confirm.

No CRISPR elements were identified in the genome of *P. tolaasii* 2192T. However, I have identified a biosynthetic gene cluster with predicted proteins orthologous to those required for alginate biosynthesis as well as multiple restriction endonuclease enzymes that can potentially act as phage resistance systems.

I have likewise successfully isolated and sequenced the genome of the mushroom pathogen *P. agarici* NCPPB 2472, which is available under the accession code NZ_CP014135, including the construction of a scaffold genome and made significant progress in finalising the genome via stitching of the remaining contig sequences. However, future work would involve completing the sequencing of the remaining junctions and assembly of the sequences into a complete genome sequence. I have identified 4,901 putative gene encoding ORFs within the *P. agarici* 2472 genome including 10 NRPS/polyketide synthase proteins, although it has not been possible to identify their products and further work could address this shortcoming.

Similar to the work performed on *P. tolaasii* 2192T I have identified a putative pyocin encoding biosynthetic gene cluster within the *P. agarici* 2472 genome. I have also identified a

biosynthetic gene cluster with significant similarities in amino acid identify and gene organisation to the achromobactin biosynthesis cluster of *P. syringae* pv. syringae B728a, which I hypothesise to be involved in the production of achromobactin in *P. agarici* NCPPB 2472, further work would be needed to confirm this prediction however. Most notably I have identified a complete Type I-F CRISPR/Cas system within the genome of *P. agarici* NCPPB 2472, including the identification of 34 spacer regions, of which one showed similarity to a known plasmid sequence involved in the production of naphthalene degrading enzymes. I have also identified multiple restriction endonuclease enzymes and their associated methyltransferase enzymes as well as a complete biosynthetic cluster with proteins orthologous to those required for alginate production, however further work would be needed to confirm the function of this cluster.

During my investigations of *P. tolaasii* 2192T and *P. agarici* NCPPB 2472 I have identified a potentially new species, *Pseudomonas* sp. NS1, which shows significant genomic similarities with both the *P. azotoformans* species and *P. fluorescens* species. I have isolated and sequenced the full genome of *P.* sp. NS1 and made it publicly available under the accession code CP022960. I have identified a large biosynthetic gene cluster containing multiple NRPSs that shows significant nucleotide and amino acid similarity as well as gene organisation to the known WLIP production cluster 2 of *Pseudomonas fluorescens* strain LMG 5329; I have also demonstrated that *P.* sp. NS1 is capable of WLIP production and therefore provided strong evidence that this biosynthetic cluster does produce the WLIP.

Similar to both *P. agarici* NCPPB 2472 and *P. tolaasii* 2192T I have also identified a biosynthetic gene cluster predicted to produce alginate and multiple putative restriction endonuclease enzymes, although both would require further work to confirm these predictions.

I have identified two new potential new species of bacteriophage capable of infecting either *P. tolaasii* 2192T or *P. agarici* NCPPB 2472.

The first, *Pseudomonas* phage NV1, which is a narrow host range phage having only been demonstrated to be capable of infecting and lysis of *P. tolaasii* 2192T cells. I have isolated and sequenced the full genome of this phage, identifying that it shares significant nucleotide identity and genome organisation to the *Pseudomonas fluorescens* phage UFV-P2. However, the identity is below the DNA sequence identity threshold of 95% guideline set by the ICTV for delineation of a new phage species, and thus with the full genome sequence differences and differing host range I suggest that Pseudomonas phage NV1 is a new species of *Luz24likevirus*. Further work involving phage NV1 would be needed to further characterise the growth and lifecycle, as well as identifying the host cell surface receptor and identifying whether resistance of *P. tolaasii* NCPPB 2192T to phage NV1 is acquired through a change in *P. tolaasii* phenotype.

The second phage identified, *Pseudomonas* phage ϕNV3, is likewise a narrow host range phage, having only been demonstrated to be capable of infecting and lysing *P. agarici* NCPPB 2472 cells. I have extensively characterised the growth and replication of phage ϕNV3 including the adsorption constant and single step growth profile and co-evolution with the host *P. agarici* NCPPB 2472.

I have also isolated and sequenced the full genome of phage ϕNV3, including confirmation via extensive primer waling. I have identified that phage ϕNV3 shares significant genome organisation similarities with *Pseudomonas* phage ϕKMV as well as identifying a single subunit RNA polymerase which is characteristic of the *Autographivirinae* subfamily of bacteriophage. Likewise, I have also identified conserved features between the lysis cassette of phage ϕNV3 and ϕKMV, indicating that they share a conserved mechanism of action. I have

also confirmed the lytic ability of the endolysin protein of ϕNV3 as well as the complementary actions of the associated lysis cassette proteins and via the identification and characterisation of a spontaneous mutant of phage ϕNV3, phage NV6, identified the T7-like tail protein as the host-specificity determinant of phage ϕNV3. On the basis of the results presented in this study I conclude that ϕNV3 is a new species of *phiKMVlikevirus* within the *Autographivirinae* subfamily.

While I have identified that *P. agarici* NCPPB 2472 rapidly develops resistance to phage ϕNV3 I have not been able to identify the exact mechanism underlying this resistance. I believe that it is likely to be due to the acquisition of a phage DNA spacer within the CRISPR repeat region, which could be tested easily in future work by PCR of the CRISPR repeat region and sequencing of the resulting product, followed by screening for the acquisition of an additional spacer region with nucleotide identity to the genomic sequence of ϕNV3.

While the attempts to create a Type IV knockout of *P. agarici* NCPPB 2472 to identify the binding receptor of ϕNV3 were unsuccessful, I have demonstrated that the technique used for electrocompetent cells was successful and this information could be used in further studies.

The attempts to elucidate the changes in transcription associated with phage ϕNV3 infection of *P. agarici* NCPPB 2472 via RNA sequencing did not yield statistically significant results. This is most likely due to the significantly reduced read counts associated with bacterial genes in the phage infected samples caused by the high levels of phage RNA transcription and host RNA degradation. Future work on this would include the use of higher MOIs to ensure all host bacteria are infected with phage, an earlier timepoint in the phage infective cycle to ensure a

more even ratio of phage to bacterial RNA and potentially sequencing to a higher depth in order to identify changes in gene expression.

Chapter 7:
References

1. Soler-Rivas, C., et al., Biochemical and physiological aspects of brown blotch disease of *Agaricus bisporus.* FEMS Microbiol Rev, 1999. **23**(5): p. 591-614.
2. Beyer, D.M., Basic procedures for agaricus mushroom growing. 2003.
3. Cole, W.M.G.a.A.L.J., Cavity disease of *Agaricus bitorquis* caused by *Pseudomonas cepacia.* Canadian Journal of Microbiology, 1992. **38**(5): p. 394-397.
4. Tsuneda, W.M.G.a.A., The interaction of the soft rot bacterium *Pseudomonas gladioli* pv. agaricicola with Japanese cultivated mushrooms. Canadian Journal of Microbiology, 1997. **43**(7): p. 639-648.
5. Roy Chowdhury, P. and J.A. Heinemann, The general secretory pathway of *Burkholderia gladioli* pv. agaricicola BG164R is necessary for cavity disease in white button mushrooms. Appl Environ Microbiol, 2006. **72**(5): p. 3558-65.
6. Steven P. Lincoln, T.R.F.a.B.J.T., *Janthinobacterium agaricidamnosum* s p. n ov., a soft rot pathogen of *Agaricus bisporus.* International Journal of Systematic Bacteriology, 1999. **49**: p. 1577-1589.
7. Katharina Graupner, K.S., Tom Bretschneider, Gerald Lackner, Martin Roth, Harald Gross and Christian Hertweck Imaging Mass Spectrometry and Genome Mining Reveal Highly Antifungal Virulence Factor of Mushroom Soft Rot Pathogen. Angewandte Chemie International Edition 2012. **51**: p. 13173-13177.
8. Paine, S.G., Studies in Bacteriosis Ii1.: A Brown Blotch Disease of Cultivated Mushrooms. Annals of Applied Biology, 1919. **5**(3-4): p. 206-219.
9. Godfrey, S.A., et al., Characterization by 16S rRNA sequence analysis of pseudomonads causing blotch disease of cultivated *Agaricus bisporus.* Appl Environ Microbiol, 2001. **67**(9): p. 4316-23.
10. Houdeau, G.a.O.J.M., Pathologie des *Pleurotus.* Champignon, 1989. **337**: p. 10-14.
11. Lynch, T.R.F.a.J.M., Bacterial blotch disease of the cultivated mushroom *Agaricus bisporus:* screening, isolation and characterization of bacteria antagonistic to the pathogen (*Pseudomonas tolaasii*). Journal of Applied Bacteriology, 1988. **65**: p. 179-187.
12. Wong, W.C. and T.F. Preece, *Pseudomonas tolaasi* in Mushroom Crops: A Note on Primary and Secondary Sources of the Bacterium on a Commercial Farm in England. Journal of Applied Bacteriology, 1980. **49**(2): p. 305-314.
13. Anzai, Y., et al., Phylogenetic affiliation of the pseudomonads based on 16S rRNA sequence. Int J Syst Evol Microbiol, 2000. **50 Pt 4**: p. 1563-89.
14. Shiv I. S. Grewal, B.H.a.K.J., Identification and Characterization of a Locus Which Regulates Multiple Functions in *Pseudomonas tolaasii*, the Cause of Brown Blotch Disease of *Agaricus bisporus.* Journal of Bacteriology, 1995. **177**(16): p. 4658-4668.
15. S. S. Cutri, B.J.M.a.W.P.R., Characteristics of pathogenic non-fluorescent (smooth) and non-pathogenic fluorescent (rough) forms of *Pseudomonas tolaasii* and *Pseudomonas gingeri.* Journal of Applied Bacteriology, 1984. **57**: p. 291-298.
16. Rainey, S.I.S.G.a.P.B., Phenotypic variation of *Pseudomonas putida* and *P. tolaasii* affects the chemotactic response to *Agaricus bisporus* mycelial exudate. Journal of General Microbiology, 1991. **137**: p. 2761-2768.
17. Christiane Guillorit-Roneau, L.M.a.R.S., Identification of two serological flagellar types (H1 and H2) in *Pseudomonas syringae* pathovars. European Journal of Plant Pathology 1996. **102**(1): p. 99-104.
18. Willis, E.M.H.a.D.K., The *lemA* Gene Required for Pathogenicity of *Pseudomonas syringae* pv. syringae on Bean Is a Member of a Family of Two-Component Regulators. Journal of Bacteriology, 1992. **174**(9): p. 3011-3020.

19. B. Arico, J.F.M., C. Roy, S. Stibitz, D. Monack, S. Falkow, R. Gross and R. Rappuoli, Sequences required for expression of *Bordetella pertussis* virulence factors share homology with prokaryotic signal transduction proteins. Proc Natl Acad Sci U S A, 1989. **86**: p. 6671-6675.
20. Gottesman, V.S.a.S., *RcsB* and *RcsC:* a Two-Component Regulator of Capsule Synthesis in Escherichia coli. Journal of Bacteriology, 1990. **172**(2): p. 659-669.
21. Bin Han, A.P.a.K.J., Spontaneous duplication of a 661 bp element within a two-component sensor regulator gene causes phenotypic switching in colonies of *Pseudomonas tolaasii*, cause of brown blotch disease of mushrooms. Molecular Microbiology, 1997. **25**(2): p. 211-218.
22. Himanshu Sinha, A.P.a.K.J., Analysis of the Role of *recA* in Phenotypic Switching of *Pseudomonas tolaasii.* Journal of Bacteriology, 2000. **182**(22): p. 6532-6535.
23. Moore, E.R.B., et al., The Determination and Comparison of the 16S rRNA Gene Sequences of Species of the Genus *Pseudomonas* (*sensu stricto* and Estimation of the Natural Intrageneric Relationships. Systematic and Applied Microbiology, 1996. **19**(4): p. 478-492.
24. Young, J.M., Drippy gill: A bacterial disease of cultivated mushrooms caused by *Pseudomonas agaric* in.sp. New Zealand Journal of Agricultural Research, 2012. **13**(4): p. 977-990.
25. Mulet, M., J. Lalucat, and E. Garcia-Valdes, DNA sequence-based analysis of the *Pseudomonas* species. Environ Microbiol, 2010. **12**(6): p. 1513-30.
26. Gill, W., "Drippy Gill": An Ooze Disease of the Cultivated Mushroom *Agaricus Bisporus* Caused By *Pseudomonas Agarici.* . Doctor of Philosophy, 1994. **University of Canterbury**.
27. Cole, W.G.a.T., Aspects of the pathology and etiology of 'drippy gill' disease of the cultivated mushroom *Agaricus bisporus.* Canadian Journal of Microbiology, 2000. **46**(3): p. 246-258.
28. Ho-Seong Lim, Y.-S.K.a.S.-D.K., *Pseudomonas stutzeri* YPL-1 Genetic Transformation and Antifungal Mechanism against *Fusarium solani*, an Agent of Plant Root Rot Applied and Environmental Microbiology, 1991. **57**(2): p. 510-516.
29. Folders, J., et al., Characterization of *Pseudomonas aeruginosa* chitinase, a gradually secreted protein. J Bacteriol, 2001. **183**(24): p. 7044-52.
30. M. N. Nielsen, J.S., J. Fels and H. C. Pedersen, Secondary Metabolite- and Endochitinase-Dependent Antagonism toward Plant-Pathogenic Microfungi of *Pseudomonas fluorescens* Isolates from Sugar Beet Rhizosphere. Applied and Environmental Microbiology, 1998. **64**(10): p. 3563-3569.
31. Megan D lenardon, C.A.M.a.N.A.G., Chitin synthesis and fungal pathogenesis. Curr Opin Microbiol, 2010. **13**(4): p. 416-423.
32. A. Obradović, K.G.a.M.I., Bacterial diseases of *Agaricus bisporus* in Serbia. *Pseudomonas syringae* Pathovars and Related Pathogens – Identification, Epidemiology and Genomics, 2008. **Fatmi M. et al. (eds)** (Springer): p. 427-430.
33. Margraff, T.L.a.R., Secondary Metabolites of the Fluorescent Pseudomonads. Microbiological Reviews, 1979. **43**(3): p. 422-442.
34. Strieker, M., A. Tanovic, and M.A. Marahiel, Nonribosomal peptide synthetases: structures and dynamics. Curr Opin Struct Biol, 2010. **20**(2): p. 234-40.
35. Raaijmakers, J.M., et al., Natural functions of lipopeptides from *Bacillus* and *Pseudomonas*: more than surfactants and antibiotics. FEMS Microbiol Rev, 2010. **34**(6): p. 1037-62.

36. Berti, A.D., et al., Identification of a biosynthetic gene cluster and the six associated lipopeptides involved in swarming motility of *Pseudomonas syringae* pv. tomato *DC3000.* J Bacteriol, 2007. **189**(17): p. 6312-23.
37. Gross, H. and J.E. Loper, Genomics of secondary metabolite production by *Pseudomonas* spp. Nat Prod Rep, 2009. **26**(11): p. 1408-46.
38. Roongsawang, N., K. Washio, and M. Morikawa, Diversity of nonribosomal peptide synthetases involved in the biosynthesis of lipopeptide biosurfactants. Int J Mol Sci, 2010. **12**(1): p. 141-72.
39. Vallet-Gely, I., et al., Association of hemolytic activity of *Pseudomonas entomophila*, a versatile soil bacterium, with cyclic lipopeptide production. Appl Environ Microbiol, 2010. **76**(3): p. 910-21.
40. Fiore, A., et al., Bioactive lipopeptides of ice-nucleating snow bacterium *Pseudomonas syringae* strain 31R1. FEMS Microbiol Lett, 2008. **286**(2): p. 158-65.
41. Li, W., et al., The antimicrobial compound xantholysin defines a new group of *Pseudomonas* cyclic lipopeptides. PLoS One, 2013. **8**(5): p. e62946.
42. Rokni-Zadeh, H., et al., Genetic and functional characterization of cyclic lipopeptide white-line-inducing principle (WLIP) production by rice rhizosphere isolate *Pseudomonas putida* RW10S2. Appl Environ Microbiol, 2012. **78**(14): p. 4826-34.
43. Nair N. G., F.P.C., Toxin Production by *Pseudomonas tolaasii* Australian Journal of Biological Sciences, 1973. **26**: p. 509-512.
44. Peng, J.T., Resistance to disease in *Agaricus bisporus* (Lange) Imbac. Thesis (Ph.D.), 1989. University of Leeds.
45. Jennifer C. Nutkins, R.J.M.-S., Leonard C. Packman, Catherine L. Brodey, Paul B. Rainey, Keith Johnstone and Dudley H. Williams, Structure Determination of Tolaasin, an Extracellular Lipodepsipeptide Produced by the Mushroom Pathogen *Pseudomonas tolaasii Paine.* Journal of the American Chemical Society, 1991. **113**: p. 2621-2627.
46. A. Shirata, K.S., M. Takasugi and K. Monde, 栽培ヒラタケに腐敗病を起こす日本産*Pseudomonas tolaasii*の産生する毒素とその生物活性. 日植病報61, 1995. **61**(493-502).
47. Bassarello, C., et al., Tolaasins A--E, five new lipodepsipeptides produced by *Pseudomonas tolaasii.* J Nat Prod, 2004. **67**(5): p. 811-6.
48. C. L. Brodey, P.B.R., M. Tester and K. Johnstone, Bacterial Blotch Disease of the Cultivated Mushroom is Caused by an Ion Channel Formin Lipodepsipeptide Toxin. Molecular Plant-Microbe Interactions, 1991. **4**(4): p. 407-411.
49. Johnstone, M.L.H.a.K., Evidence for the involvment of the surface active properties of the extracellular toxin tolaasin in the manifestation of brown blotch disease symptoms by *Pseudomonas tolaasii* on *Agaricus bisporus.* Physiological and Molecular Plant Pathology, 1993. **42**: p. 373-384.
50. Scherlach, K., et al., Biosynthesis and mass spectrometric imaging of tolaasin, the virulence factor of brown blotch mushroom disease. Chembiochem, 2013. **14**(18): p. 2439-43.
51. Gessard, M.C., Sur la fonction fluorescigenè des microbes. Annales de l'Institut Pasteur, 1892. **6**: p. 801-823.
52. Meyer, J.M., The Fluorescent Pigment of *Pseudomonas fluorescens*: Biosynthesis, Purification and Physicochemical Properties. Journal of General Microbiology, 1978. **107**: p. 319-328.
53. Meyer, J.M., Pyoverdines: pigments, siderophores and potential taxonomic markers of fluorescent *Pseudomonas* species. Arch Microbiol, 2000. **174**(3): p. 135-42.

54. Parker, D.L., et al., Pyoverdine synthesis by the Mn(II)-oxidizing bacterium *Pseudomonas putida GB-1.* Front Microbiol, 2014. **5**: p. 202.
55. Merriman, T.R., M.E. Merriman, and I.L. Lamont, Nucleotide sequence of *pvdD,* a pyoverdine biosynthetic gene from *Pseudomonas aeruginosa:* PvdD has similarity to peptide synthetases. Journal of Bacteriology, 1995. **177**(1): p. 252-258.
56. Lehoux, D.E., F. Sanschagrin, and R.C. Levesque, Genomics of the 35-kb *pvd* locus and analysis of novel *pvdIJK* genes implicated in pyoverdine biosynthesis in *Pseudomonas aeruginosa.* FEMS Microbiol Lett, 2000. **190**(1): p. 141-6.
57. Mossialos, D., et al., Identification of new, conserved, non-ribosomal peptide synthetases from fluorescent pseudomonads involved in the biosynthesis of the siderophore pyoverdine. Mol Microbiol, 2002. **45**(6): p. 1673-85.
58. Quadri, L.E.N., et al., Assembly of the *Pseudomonas aeruginosa* Nonribosomal Peptide Siderophore Pyochelin: In Vitro Reconstitution of Aryl-4,2-bisthiazoline Synthetase Activity from PchD, PchE, and PchF†. Biochemistry, 1999. **38**(45): p. 14941-14954.
59. Reimmann, C., et al., Dihydroaeruginoic acid synthetase and pyochelin synthetase, products of the *pchEF* genes, are induced by extracellular pyochelin in *Pseudomonas aeruginosa.* Microbiology, 1998. **144 (Pt 11)**(11): p. 3135-48.
60. Serino, L., et al., Biosynthesis of pyochelin and dihydroaeruginoic acid requires the iron-regulated *pchDCBA* operon in *Pseudomonas aeruginosa.* J Bacteriol, 1997. **179**(1): p. 248-57.
61. Berti, A.D. and M.G. Thomas, Analysis of achromobactin biosynthesis by *Pseudomonas syringae* pv. syringae B728a. J Bacteriol, 2009. **191**(14): p. 4594-604.
62. Kohler, T., V. Donner, and C. van Delden, Lipopolysaccharide as shield and receptor for R-pyocin-mediated killing in *Pseudomonas aeruginosa.* J Bacteriol, 2010. **192**(7): p. 1921-8.
63. Ghequire, M.G. and R. De Mot, Ribosomally encoded antibacterial proteins and peptides from *Pseudomonas.* FEMS Microbiol Rev, 2014. **38**(4): p. 523-68.
64. Riley, M.A. and J.E. Wertz, Bacteriocins: evolution, ecology, and application. Annu Rev Microbiol, 2002. **56**: p. 117-37.
65. Bernbom, N., et al., Effects of *Lactococcus lactis* on composition of intestinal microbiota: role of nisin. Appl Environ Microbiol, 2006. **72**(1): p. 239-44.
66. Hansen, J.N. and W.E. Sandine, Nisin as a model food preservative. Critical Reviews in Food Science and Nutrition, 1994. **34**(1): p. 69-93.
67. Scholl, D., Phage Tail-Like Bacteriocins. Annual Review of Virology, 2017. **4**: p. 453-467.
68. Michel-Briand, Y. and C. Baysse, The pyocins of *Pseudomonas aeruginosa.* Biochimie, 2002. **84**(5-6): p. 499-510.
69. Nakayama, K., et al., The R-type pyocin of *Pseudomonas aeruginosa* is related to P2 phage, and the F-type is related to lambda phage. Molecular Microbiology, 2002. **38**(2): p. 213-231.
70. Fischer, S., et al., Characterization of a phage-like pyocin from the plant growth-promoting rhizobacterium *Pseudomonas fluorescens* SF4c. Microbiology, 2012. **158**(Pt 6): p. 1493-503.
71. Matsui, H., et al., Regulation of pyocin genes in *Pseudomonas aeruginosa* by positive *(prtN)* and negative *(prtR)* regulatory genes. J Bacteriol, 1993. **175**(5): p. 1257-63.
72. Duckworth, D.H., "Who discovered bacteriophage?". Bacteriol Rev, 1976. **40**(4): p. 793-802.

73. Zarenejad, F., B. Yakhchali, and I. Rasooli, Evaluation of indigenous potent mushroom growth promoting bacteria (MGPB) on *Agaricus bisporus* production. World J Microbiol Biotechnol, 2012. **28**(1): p. 99-104.
74. Kim, M.-H., Bacteriophages of *Pseudomonas tolaasii* for the Biological Control of Brown Blotch Disease. Journal of the Korean Society for Applied Biological Chemistry, 2011. **54**(1).
75. Hendrix, R.W., Bacteriophage genomics. Curr Opin Microbiol, 2003. **6**(5): p. 506-11.
76. Grath, S.M. and D. van Sinderen, Bacteriophage: Genetics and Molecular Biology. 2007: Caister Academic Press.
77. Lodish H, B.A., Zipursky SL, *et al.*, Section 6.3, Viruses: Structure, Function and Uses. Molecular Cell Biology. **4th edition.** New York: W. H. Freeman.
78. Ackermann, H.W., Classification of bacteriophages. . In The Bacteriophages, 2006. Ed. Calendar R(Oxford University Press).
79. Evelien M. Adriaenssens, M.K., Petar Knezevic, Hans-Wolfgang Ackermann, Jakub Barylski, J. Rodney Brister, Martha R. C. Clokie, Siobain Duffy, Bas E. Dutilh, Robert A. Edwards, Francois Enault, Ho Bin Jang, Jochen Klumpp, Andrew M. Kropinski, Rob Lavigne, Minna M. Poranen, David Prangishvili, Janis Rumnieks, Matthew B. Sullivan, Johannes Wittmann, Hanna M. Oksanen, Annika Gillis, Jens H. Kuhn, Taxonomy of prokaryotic viruses: 2016 update from the ICTV bacterial and archaeal viruses subcommittee. Archives of Virology, 2017. **162**(4): p. 1153-1157.
80. Fokine, A. and M.G. Rossmann, Molecular architecture of tailed double-stranded DNA phages. Bacteriophage, 2014. **4**(1): p. e28281.
81. Virus Taxonomy: 2016 Release.
82. Demerec, M. and U. Fano, Bacteriophage-Resistant Mutants in *Escherichia Coli.* Genetics, 1945. **30**(2): p. 119-36.
83. Studier, F.W., The genetics and physiology of bacteriophage T7. Virology, 1969. **39**(3): p. 562-74.
84. Studier, F.W. and J.V. Maizel, Jr., *T7-directed protein synthesis.* Virology, 1969. **39**(3): p. 575-86.
85. Dunn, J.J. and F.W. Studier, Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. J Mol Biol, 1983. **166**(4): p. 477-535.
86. Huber, H.E., B.B. Beauchamp, and C.C. Richardson, *Escherichia coli* dGTP triphosphohydrolase is inhibited by gene 1.2 protein of bacteriophage T7. J Biol Chem, 1988. **263**(27): p. 13549-56.
87. Moffatt, B.A. and F.W. Studier, T7 lysozyme inhibits transcription by T7 RNA polymerase. Cell, 1987. **49**(2): p. 221-7.
88. Prosen, D.E. and C.L. Cech, Bacteriophage T7 E promoter: identification and measurement of kinetics of association with *Escherichia coli* RNA polymerase. Biochemistry, 1985. **24**(9): p. 2219-27.
89. Molineux, I.J., The T7 Group, in The Bacteriophages, C. R.L, Editor. 2005, Oxford University press.
90. Studier, F.W., Bacteriophage T7. Science, 1972. **176**(4033): p. 367-76.
91. Kruger, D.H. and C. Schroeder, Bacteriophage T3 and bacteriophage T7 virus-host cell interactions. Microbiol Rev, 1981. **45**(1): p. 9-51.
92. Rontó, G., et al., Symmetry and structure of bacteriophage T7. Computers & Mathematics with Applications, 1988. **16**(5-8): p. 617-628.
93. Steven, A.C., et al., Molecular substructure of a viral receptor-recognition protein. The gp17 tail-fiber of bacteriophage T7. J Mol Biol, 1988. **200**(2): p. 351-65.

94. Garcia-Doval, C. and M.J. van Raaij, Structure of the receptor-binding carboxy-terminal domain of bacteriophage T7 tail fibers. Proc Natl Acad Sci U S A, 2012. **109**(24): p. 9390-5.
95. Serwer, P., et al., Evidence for bacteriophage T7 tail extension during DNA injection. BMC Res Notes, 2008. **1**(1): p. 36.
96. Kemp, P., M. Gupta, and I.J. Molineux, Bacteriophage T7 DNA ejection into cells is initiated by an enzyme-like mechanism. Mol Microbiol, 2004. **53**(4): p. 1251-65.
97. Garcia, L.R. and I.J. Molineux, Rate of translocation of bacteriophage T7 DNA across the membranes of *Escherichia coli.* J Bacteriol, 1995. **177**(14): p. 4066-76.
98. Savalia, D., et al., The role of the T7 Gp2 inhibitor of host RNA polymerase in phage development. J Mol Biol, 2010. **402**(1): p. 118-26.
99. Lavigne, R., et al., Unifying classical and molecular taxonomic classification: analysis of the *Podoviridae* using BLASTP-based tools. Res Microbiol, 2008. **159**(5): p. 406-14.
100. Lavigne, R., et al., The genome of bacteriophage *phiKMV,* a T7-like virus infecting *Pseudomonas aeruginosa.* Virology, 2003. **312**(1): p. 49-59.
101. Chibeu, A., et al., The adsorption of *Pseudomonas aeruginosa* bacteriophage *phiKMV* is dependent on expression regulation of type IV pili genes. FEMS Microbiol Lett, 2009. **296**(2): p. 210-8.
102. Briers, Y., et al., The lysis cassette of bacteriophage *phiKMV* encodes a signal-arrest-release endolysin and a pinholin. Bacteriophage, 2011. **1**(1): p. 25-30.
103. Adriaenssens, E.M., et al., Bacteriophages *LIMElight* and *LIMEzero* of *Pantoea agglomerans,* belonging to the *"phiKMV-like viruses".* Appl Environ Microbiol, 2011. **77**(10): p. 3443-50.
104. Monier, J.M. and S.E. Lindow, Aggregates of resident bacteria facilitate survival of immigrant bacteria on leaf surfaces. Microb Ecol, 2005. **49**(3): p. 343-52.
105. Andersson, A.M., et al., Dust-borne bacteria in animal sheds, schools and children's day care centres. Journal of Applied Microbiology, 1999. **86**(4): p. 622-634.
106. Ceyssens, P.J., et al., The intron-containing genome of the lytic *Pseudomonas phage LUZ24* resembles the temperate phage *PaP3.* Virology, 2008. **377**(2): p. 233-8.
107. Eller, M.R., et al., Complete Genome Sequence of the *Pseudomonas fluorescens* Bacteriophage *UFV-P2.* Genome Announc, 2013. **1**(1).
108. Eller, M.R., et al., *UFV-P2* as a member of the *Luz24likevirus genus:* a new overview on comparative functional genome analyses of the *LUZ24-like phages.* BMC Genomics, 2014. **15**: p. 7.
109. Young, R., Bacteriophage holins: deadly diversity. J Mol Microbiol Biotechnol, 2002. **4**(1): p. 21-36.
110. Fischetti, V.A., Bacteriophage lytic enzymes: novel anti-infectives. Trends Microbiol, 2005. **13**(10): p. 491-6.
111. Hermoso, J.A., J.L. Garcia, and P. Garcia, Taking aim on bacterial pathogens: from phage therapy to enzybiotics. Curr Opin Microbiol, 2007. **10**(5): p. 461-72.
112. Nelson, D.C., et al., Endolysins as antimicrobials. Adv Virus Res, 2012. **83**: p. 299-365.
113. Oliveira, H., et al., Bacteriophage endolysins as a response to emerging foodborne pathogens. Trends in Food Science & Technology, 2012. **28**(2): p. 103-115.
114. Young, R., Phage lysis: three steps, three choices, one outcome. J Microbiol, 2014. **52**(3): p. 243-58.
115. Fischetti, V.A., Bacteriophage endolysins: a novel anti-infective to control Gram-positive pathogens. Int J Med Microbiol, 2010. **300**(6): p. 357-62.

116. Savva, C.G., et al., Stable micron-scale holes are a general feature of canonical holins. Mol Microbiol, 2014. **91**(1): p. 57-65.
117. Oliveira, H., et al., Molecular aspects and comparative genomics of bacteriophage endolysins. J Virol, 2013. **87**(8): p. 4558-70.
118. Xu, M., et al., Disulfide isomerization after membrane release of its SAR domain activates P1 lysozyme. Science, 2005. **307**(5706): p. 113-7.
119. Xu, M., et al., A signal-arrest-release sequence mediates export and control of the phage P1 endolysin. Proc Natl Acad Sci U S A, 2004. **101**(17): p. 6415-20.
120. Park, T., et al., The pinholin of lambdoid phage 21: control of lysis by membrane depolarization. J Bacteriol, 2007. **189**(24): p. 9135-9.
121. Young, R., Phage lysis: do we have the hole story yet? Curr Opin Microbiol, 2013. **16**(6): p. 790-7.
122. Rajaure, M., et al., Membrane fusion during phage lysis. Proc Natl Acad Sci U S A, 2015. **112**(17): p. 5497-502.
123. Wang, I.N., D.L. Smith, and R. Young, Holins: the protein clocks of bacteriophage infections. Annu Rev Microbiol, 2000. **54**: p. 799-825.
124. Ramanculov, E. and R. Young, Functional analysis of the phage T4 holin in a lambda context. Mol Genet Genomics, 2001. **265**(2): p. 345-53.
125. Wang, I.N., J. Deaton, and R. Young, Sizing the holin lesion with an endolysin-beta-galactosidase fusion. J Bacteriol, 2003. **185**(3): p. 779-87.
126. White, R., et al., Holin triggering in real time. Proc Natl Acad Sci U S A, 2011. **108**(2): p. 798-803.
127. Kumar, S., Textbook of Microbiology. 2012: Jaypee Brothers,Medical Publishers Pvt. Limited.
128. Briers, Y., et al., Muralytic activity and modular structure of the endolysins of *Pseudomonas aeruginosa* bacteriophages *phiKZ* and *EL*. Mol Microbiol, 2007. **65**(5): p. 1334-44.
129. Briers, Y., et al., The high-affinity peptidoglycan binding domain of *Pseudomonas* phage endolysin *KZ144*. Biochem Biophys Res Commun, 2009. **383**(2): p. 187-91.
130. Summer, E.J., et al., *Rz/Rz1* lysis gene equivalents in phages of Gram-negative hosts. J Mol Biol, 2007. **373**(5): p. 1098-112.
131. Zhang, N. and R. Young, Complementation and characterization of the nested *Rz* and *Rz1* reading frames in the genome of bacteriophage lambda. Mol Gen Genet, 1999. **262**(4-5): p. 659-67.
132. Berry, J., et al., The final step in the phage infection cycle: the Rz and Rz1 lysis proteins link the inner and outer membranes. Mol Microbiol, 2008. **70**(2): p. 341-51.
133. Berry, J., et al., The spanin complex is essential for lambda lysis. J Bacteriol, 2012. **194**(20): p. 5667-74.
134. Nelson, D., L. Loomis, and V.A. Fischetti, Prevention and elimination of upper respiratory colonization of mice by group A streptococci by using a bacteriophage lytic enzyme. Proc Natl Acad Sci U S A, 2001. **98**(7): p. 4107-12.
135. Loeffler, J.M., S. Djurkovic, and V.A. Fischetti, Phage Lytic Enzyme Cpl-1 as a Novel Antimicrobial for Pneumococcal Bacteremia. Infection and Immunity, 2003. **71**(11): p. 6199-6204.
136. Briers, Y., M. Walmagh, and R. Lavigne, Use of bacteriophage endolysin EL188 and outer membrane permeabilizers against *Pseudomonas aeruginosa*. J Appl Microbiol, 2011. **110**(3): p. 778-85.
137. Moak, M. and I.J. Molineux, Peptidoglycan hydrolytic activities associated with bacteriophage virions. Mol Microbiol, 2004. **51**(4): p. 1169-83.

138. Masschalck, B., D. Deckers, and C.W. Michiels, Lytic and nonlytic mechanism of inactivation of gram-positive bacteria by lysozyme under atmospheric and high hydrostatic pressure. J Food Prot, 2002. **65**(12): p. 1916-23.
139. Briers, Y., et al., Analysis of outer membrane permeability of *Pseudomonas aeruginosa* and bactericidal activity of endolysins KZ144 and EL188 under high hydrostatic pressure. FEMS Microbiol Lett, 2008. **280**(1): p. 113-9.
140. Lindberg, A.A., Bacteriophage receptors. Annu Rev Microbiol, 1973. **27**: p. 205-41.
141. Rakhuba, D.V., *et al*., Bacteriophage Receptors, Mechanisms of Phage Adsorption and Penetration into Host Cell. Polish Journal of Microbiology, 2010. **59**(3): p. 145-155.
142. Loek van Alphen, L.H.a.B.L., Major outer membrane protein d of *Escherichia coli* K 12. FEBS Letters, 1977. **75**(1-2): p. 285-290.
143. Ronal A. Skurray, R.E.W.H.a.P.R., Con- Mutants: Class of Mutants in *Escherichia coli* K-12 Lacking a Major Cell Wall Protein and Defective in Conjugation and Adsorption of a Bacteriophage. Journal of Bacteriology, 1974. **119**(3): p. 726-735.
144. Koebnik, R., Structural and functional roles of the surface-exposed loops of the beta-barrel membrane protein OmpA from *Escherichia coli.* J Bacteriol, 1999. **181**(12): p. 3688-94.
145. JP, M.C.a.R., Conjugation-deficient mutants of *Escherichia coli* distinguish classes of functions of the outer membrane OmpA protein. Molecular Genetics and Genomics, 1982. **187**(1): p. 148-156.
146. Mizushima, F.Y.a.S., Roles of Lipopolysaccharide and Outer Membrane Protein OmpC of *Escherichia coli* K-12 in the Receptor Function of Bacteriophage T4. Journal of Bacteriology, 1982. **151**(2): p. 718-722.
147. Igo, C.-Y.L.a.M.M., Differential Expression of the OmpF and OmpC Porin Proteins in *Escherichia coli* K-12 Depends upon the Level of Active OmpR. Journal of Bacteriology, 1998. **180**(1): p. 171-174.
148. Braun, V., FhuA (TonA), the career of a protein. J Bacteriol, 2009. **191**(11): p. 3431-6.
149. Clett Erridge, E.B.-G.a.I.R.P., Structure and function of lipopolysaccharides. Microbes and Infection, 2002. **4**: p. 837-851.
150. Wilkinson, S.G., Bacterial lipopolysaccharides--themes and variations. Prog Lipid Res, 1996. **35**(3): p. 283-343.
151. Delcour, A.H., Outer membrane permeability and antibiotic resistance. Biochim Biophys Acta, 2009. **1794**(5): p. 808-16.
152. Auer, G.K. and D.B. Weibel, Bacterial Cell Mechanics. Biochemistry, 2017. **56**(29): p. 3710-3724.
153. A. A. Lindberg, R.W., P. Gemski and J. A. Wohlhieter, Interaction Between Bacteriophage Sf6 and *Shigella flexneri.* Journal of Virology, 1978. **27**(1): p. 38-44.
154. Ceyssens, P.J., et al., Phenotypic and genotypic variations within a single bacteriophage species. Virol J, 2011. **8**(8): p. 134.
155. M J coyne, J., K S Russell, C L Coyle and J B Goldberg, The *Pseudomonas aeruginosa algC* Gene Encodes Phosphoglucomutase, Required for the Synthesis of a Complete Lipopolysaccharide Core. Journal of Bacteriology, 1994. **176**(12): p. 3500-3507.
156. Cross, A.S., The Biologic Significance of Bacterial Encapsulation, in Bacterial Capsules, K. Jann and B. Jann, Editors. 1990, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 87-95.

157. Boyce, J.D. and B. Adler, The Capsule Is a Virulence Determinant in the Pathogenesis of *Pasteurella multocida* M1404 *(B:2).* Infection and Immunity, 2000. **68**(6): p. 3463-3468.
158. Roberts, I.S., The biochemistry and genetics of capsular polysaccharide production in bacteria. Annu Rev Microbiol, 1996. **50**: p. 285-315.
159. Willis, L.M. and C. Whitfield, Structure, biosynthesis, and function of bacterial capsular polysaccharides synthesized by ABC transporter-dependent pathways. Carbohydr Res, 2013. **378**: p. 35-44.
160. Stirm, S. and E. Freund-Molbert, *Escherichia coli* capsule bacteriophages. II. Morphology. J Virol, 1971. **8**(3): p. 330-42.
161. Jakobsson, E., et al., Identification of amino acid residues at the active site of endosialidase that dissociate the polysialic acid binding and cleaving activities in *Escherichia coli K1* bacteriophages. Biochem J, 2007. **405**(3): p. 465-72.
162. Gross, R.J., T. Cheasty, and B. Rowe, Isolation of bacteriophages specific for the K1 polysaccharide antigen of *Escherichia coli.* J Clin Microbiol, 1977. **6**(6): p. 548-50.
163. Muhlenhoff, M., et al., Proteolytic processing and oligomerization of bacteriophage-derived endosialidases. J Biol Chem, 2003. **278**(15): p. 12634-44.
164. Scholl, D., S. Adhya, and C. Merril, *Escherichia coli* K1's capsule is a barrier to bacteriophage T7. Appl Environ Microbiol, 2005. **71**(8): p. 4872-4.
165. Holst Sorensen, M.C., et al., Phase variable expression of capsular polysaccharide modifications allows *Campylobacter jejuni* to avoid bacteriophage infection in chickens. Front Cell Infect Microbiol, 2012. **2**: p. 11.
166. The Nature of the Bacterial Surface. The Yale Journal of Biology and Medicine, 1950. **23**(2): p. 158-158.
167. Proft, T. and E.N. Baker, Pili in Gram-negative and Gram-positive bacteria - structure, assembly and their role in disease. Cell Mol Life Sci, 2009. **66**(4): p. 613-35.
168. Connell, I., et al., Type 1 fimbrial expression enhances *Escherichia coli* virulence for the urinary tract. Proc Natl Acad Sci U S A, 1996. **93**(18): p. 9827-32.
169. Zhou, G., et al., Uroplakin Ia is the urothelial receptor for uropathogenic *Escherichia coli:* evidence from in vitro FimH binding. J Cell Sci, 2001. **114**(Pt 22): p. 4095-103.
170. Shi, W. and H. Sun, Type IV pilus-dependent motility and its possible role in bacterial pathogenesis. Infect Immun, 2002. **70**(1): p. 1-4.
171. Craig, L. and J. Li, Type IV pili: paradoxes in form and function. Curr Opin Struct Biol, 2008. **18**(2): p. 267-77.
172. McBride, M.J., Bacterial gliding motility: multiple mechanisms for cell movement over surfaces. Annu Rev Microbiol, 2001. **55**(1): p. 49-75.
173. Kim, S., et al., *Pseudomonas aeruginosa* bacteriophage PA1O requires type IV pili for infection and shows broad bactericidal and biofilm removal activities. Appl Environ Microbiol, 2012. **78**(17): p. 6380-5.
174. Olsen, A., A. Jonsson, and S. Normark, Fibronectin binding mediated by a novel class of surface organelles on *Escherichia coll.* Nature, 1989. **338**(6217): p. 652-655.
175. Kikuchi, T., et al., Curli fibers are required for development of biofilm architecture in *Escherichia coli* K-12 and enhance bacterial adherence to human uroepithelial cells. Microbiol Immunol, 2005. **49**(9): p. 875-84.
176. Barnhart, M.M. and M.R. Chapman, Curli biogenesis and function. Annu Rev Microbiol, 2006. **60**: p. 131-47.
177. Sakellaris, H., G.P. Munson, and J.R. Scott, A conserved residue in the tip proteins of CS1 and CFA/I pili of enterotoxigenic *Escherichia coli* that is essential for adherence.

Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(22): p. 12828-12832.
178. Rossez, Y., et al., Bacterial flagella: twist and stick, or dodge across the kingdoms. PLoS Pathog, 2015. **11**(1): p. e1004483.
179. Sourjik, V. and N.S. Wingreen, Responding to chemical gradients: bacterial chemotaxis. Curr Opin Cell Biol, 2012. **24**(2): p. 262-8.
180. Choi, Y., et al., Identification and characterization of a novel flagellum-dependent *Salmonella*-infecting bacteriophage, iEPS5. Appl Environ Microbiol, 2013. **79**(16): p. 4829-37.
181. Evans, T.J., et al., Exploitation of a new flagellatropic phage of *Erwinia* for positive selection of bacterial mutants attenuated in plant virulence: towards phage therapy. J Appl Microbiol, 2010. **108**(2): p. 676-85.
182. Andres, D., et al., Tail morphology controls DNA release in two *Salmonella* phages with one lipopolysaccharide receptor recognition system. Mol Microbiol, 2012. **83**(6): p. 1244-53.
183. Molineux, I.J., No syringes please, ejection of phage T7 DNA from the virion is enzyme driven. Mol Microbiol, 2001. **40**(1): p. 1-8.
184. Israel, V., A model for the adsorption of phage P22 to *Salmonella typhimurium.* J Gen Virol, 1978. **40**(3): p. 669-73.
185. Nordstrom, K. and A. Forsgren, Effect of protein A on adsorption of bacteriophages to *Staphylococcus aureus.* J Virol, 1974. **14**(2): p. 198-202.
186. Riede, I. and M.L. Eschbach, Evidence that TraT interacts with OmpA of *Escherichia coli.* FEBS Lett, 1986. **205**(2): p. 241-5.
187. Byl, C.V. and A.M. Kropinski, Sequence of the Genome of *Salmonella* Bacteriophage *P22.* Journal of Bacteriology, 2000. **182**(22): p. 6472-6481.
188. Perry, L.L., et al., Sequence analysis of *Escherichia coli* O157:H7 bacteriophage *PhiV10* and identification of a phage-encoded immunity protein that modifies the O157 antigen. FEMS Microbiol Lett, 2009. **292**(2): p. 182-6.
189. Sarkar, S., et al., Role of capsule and O antigen in the virulence of uropathogenic *Escherichia coli.* PLoS One, 2014. **9**(4): p. e94786.
190. Henderson, I.R., P. Owen, and J.P. Nataro, Molecular switches--the ON and OFF of bacterial phase variation. Mol Microbiol, 1999. **33**(5): p. 919-32.
191. van der Woude, M.W., Re-examining the role and random nature of phase variation. FEMS Microbiol Lett, 2006. **254**(2): p. 190-7.
192. Kim, M. and S. Ryu, Spontaneous and transient defence against bacteriophage by phase-variable glucosylation of O-antigen in *Salmonella enterica* serovar *Typhimurium.* Mol Microbiol, 2012. **86**(2): p. 411-25.
193. Geisinger, E. and R.R. Isberg, Antibiotic modulation of capsular exopolysaccharide and virulence in *Acinetobacter baumannii.* PLoS Pathog, 2015. **11**(2): p. e1004691.
194. Kachlany, S.C., et al., Structure and carbohydrate analysis of the exopolysaccharide capsule of *Pseudomonas putida G7.* Environ Microbiol, 2001. **3**(12): p. 774-84.
195. Li, P., S.E. Harding, and Z. Liu, Cyanobacterial exopolysaccharides: their nature and potential biotechnological applications. Biotechnol Genet Eng Rev, 2001. **18**: p. 375-404.
196. Nivens, D.E., et al., Role of alginate and its O acetylation in formation of *Pseudomonas aeruginosa* microcolonies and biofilms. J Bacteriol, 2001. **183**(3): p. 1047-57.
197. Cerning, J., Production of exopolysaccharides by lactic acid bacteria and dairy propionibacteria. Lait, 1995. **75**(4-5): p. 463-472.

198. O'Toole, G., H.B. Kaplan, and R. Kolter, Biofilm formation as microbial development. Annu Rev Microbiol, 2000. **54**(1): p. 49-79.
199. Puranik, P.R. and K.M. Paknikar, Biosorption of lead, cadmium, and zinc by *Citrobacter* strain *MCM B-181:* characterization studies. Biotechnol Prog, 1999. **15**(2): p. 228-37.
200. Glonti, T., N. Chanishvili, and P.W. Taylor, Bacteriophage-derived enzyme that depolymerizes the alginic acid capsule associated with cystic fibrosis isolates of *Pseudomonas aeruginosa.* J Appl Microbiol, 2010. **108**(2): p. 695-702.
201. Chung, I.Y., et al., A phage protein that inhibits the bacterial ATPase required for type IV pilus assembly. Proc Natl Acad Sci U S A, 2014. **111**(31): p. 11503-8.
202. Heo, Y.J., et al., Genome sequence comparison and superinfection between two related *Pseudomonas aeruginosa* phages, *D3112 and MP22.* Microbiology, 2007. **153**(Pt 9): p. 2885-95.
203. Pingoud, A. and A. Jeltsch, Structure and function of type II restriction endonucleases. Nucleic Acids Res, 2001. **29**(18): p. 3705-27.
204. Tock, M.R. and D.T. Dryden, The biology of restriction and anti-restriction. Curr Opin Microbiol, 2005. **8**(4): p. 466-72.
205. Roberts, R.J., A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. Nucleic Acids Research, 2003. **31**(7): p. 1805-1812.
206. Murray, N.E., Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). Microbiol Mol Biol Rev, 2000. **64**(2): p. 412-34.
207. Webb, J.L., et al., Restriction by EcoKI is enhanced by co-operative interactions between target sequences and is dependent on DEAD box motifs. The EMBO Journal, 1996. **15**(8): p. 2003-2009.
208. Hedgpeth, J., H.M. Goodman, and H.W. Boyer, DNA nucleotide sequence restricted by the RI endonuclease. Proc Natl Acad Sci U S A, 1972. **69**(11): p. 3448-52.
209. Dryden, D.T., N.E. Murray, and D.N. Rao, Nucleoside triphosphate-dependent restriction enzymes. Nucleic Acids Res, 2001. **29**(18): p. 3728-41.
210. Janscak, P., et al., Subunit assembly and mode of DNA cleavage of the type III restriction endonucleases EcoP1I and EcoP15I. J Mol Biol, 2001. **306**(3): p. 417-31.
211. Iida, S., et al., DNA restriction--modification genes of phage P1 and plasmid p15B. Structure and in vitro transcription. J Mol Biol, 1983. **165**(1): p. 1-18.
212. Lepikhov, K., et al., Characterization of the type IV restriction modification system BspLU11III from *Bacillus sp.* LU11. Nucleic Acids Res, 2001. **29**(22): p. 4691-8.
213. Stewart, F.J., et al., Methyl-specific DNA binding by McrBC, a modification-dependent restriction enzyme. J Mol Biol, 2000. **298**(4): p. 611-22.
214. Samson, J.E., et al., Revenge of the phages: defeating bacterial defences. Nat Rev Microbiol, 2013. **11**(10): p. 675-87.
215. Meisel, A., et al., Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage. Nature, 1992. **355**(6359): p. 467-9.
216. Weigele, P. and E.A. Raleigh, Biosynthesis and Function of Modified Bases in Bacteria and Their Viruses. Chem Rev, 2016. **116**(20): p. 12655-12687.
217. Khatoon, H. and A.I. Bukhari, Bacteriophage Mu-induced modification of DNA is dependent upon a host function. J Bacteriol, 1978. **136**(1): p. 423-8.
218. Nikolskaya, I., N. Lopatina, and S. Debov, Methylated guanine derivative as a minor base in the DNA of phage DDVI *Shigella disenteriae.* Biochimica et Biophysica Acta (BBA)-Nucleic Acids and Protein Synthesis, 1976. **435**(2): p. 206-210.

219. Zabeau, M., et al., The *ral* gene of phage *lambda. I.* Identification of a non-essential gene that modulates restriction and modification in *E. coli.* Mol Gen Genet, 1980. **179**(1): p. 63-73.
220. Iida, S., et al., Two DNA antirestriction systems of bacteriophage *P1, darA,* and *darB: c*haracterization of *darA*- phages. Virology, 1987. **157**(1): p. 156-66.
221. Atanasiu, C., et al., Interaction of the *ocr gene 0.3* protein of bacteriophage *T7* with EcoKI restriction/modification enzyme. Nucleic Acids Res, 2002. **30**(18): p. 3936-44.
222. Dy, R.L., et al., A widespread bacteriophage abortive infection system functions through a Type IV toxin-antitoxin mechanism. Nucleic Acids Res, 2014. **42**(7): p. 4590-605.
223. Fineran, P.C., et al., The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. Proc Natl Acad Sci U S A, 2009. **106**(3): p. 894-9.
224. Chopin, M.C., A. Chopin, and E. Bidnenko, Phage abortive infection in lactococci: variations on a theme. Curr Opin Microbiol, 2005. **8**(4): p. 473-9.
225. Snyder, L., Phage-exclusion enzymes: a bonanza of biochemical and cell biology reagents? Mol Microbiol, 1995. **15**(3): p. 415-20.
226. Smith, H.S., et al., Abortive infection of *Shigella dysenteriae P2* by T2 bacteriophage. J Virol, 1969. **4**(2): p. 162-8.
227. Ishino, Y., et al., Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli,* and identification of the gene product. J Bacteriol, 1987. **169**(12): p. 5429-33.
228. Jansen, R., et al., Identification of genes that are associated with DNA repeats in prokaryotes. Mol Microbiol, 2002. **43**(6): p. 1565-75.
229. Bolotin, A., et al., Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology, 2005. **151**(Pt 8): p. 2551-61.
230. Barrangou, R., et al., CRISPR provides acquired resistance against viruses in prokaryotes. Science, 2007. **315**(5819): p. 1709-12.
231. Komor, A.C., A.H. Badran, and D.R. Liu, CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. Cell, 2017. **168**(1-2): p. 20-36.
232. Makarova, K.S., et al., An updated evolutionary classification of CRISPR-Cas systems. Nat Rev Microbiol, 2015. **13**(11): p. 722-36.
233. Wright, A.V., J.K. Nunez, and J.A. Doudna, Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. Cell, 2016. **164**(1-2): p. 29-44.
234. Gong, B., et al., Molecular insights into DNA interference by CRISPR-associated nuclease-helicase Cas3. Proc Natl Acad Sci U S A, 2014. **111**(46): p. 16359-64.
235. Jore, M.M., et al., Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nat Struct Mol Biol, 2011. **18**(5): p. 529-36.
236. Rath, D., et al., The CRISPR-Cas immune system: biology, mechanisms and applications. Biochimie, 2015. **117**: p. 119-28.
237. Fagerlund, R.D., et al., Spacer capture and integration by a type I-F Cas1-Cas2-3 CRISPR adaptation complex. Proc Natl Acad Sci U S A, 2017. **114**(26): p. E5122-E5128.
238. Xiao, Y., et al., How type II CRISPR-Cas establish immunity through Cas1-Cas2-mediated spacer integration. Nature, 2017. **550**(7674): p. 137-141.
239. Rollins, M.F., et al., Cas1 and the Csy complex are opposing regulators of Cas2/3 nuclease activity. Proc Natl Acad Sci U S A, 2017. **114**(26): p. E5113-E5121.
240. Mojica, F.J., et al., Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology, 2009. **155**(Pt 3): p. 733-40.

241. Deltcheva, E., et al., CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature, 2011. **471**(7340): p. 602-7.
242. Deveau, H., et al., Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus.* J Bacteriol, 2008. **190**(4): p. 1390-400.
243. Sapranauskas, R., et al., The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli.* Nucleic Acids Res, 2011. **39**(21): p. 9275-82.
244. Datsenko, K.A., et al., Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. Nat Commun, 2012. **3**: p. 945.
245. Bondy-Denomy, J., et al., Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. Nature, 2013. **493**(7432): p. 429-32.
246. Betts, A., et al., Back to the future: evolving bacteriophages to increase their effectiveness against the pathogen *Pseudomonas aeruginosa* PAO1. Evol Appl, 2013. **6**(7): p. 1054-63.
247. Jiang, H., et al., Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinformatics, 2014. **15**(1): p. 182.
248. Bankevich, A., et al., SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol, 2012. **19**(5): p. 455-77.
249. Choi, K.H., A. Kumar, and H.P. Schweizer, A 10-min method for preparation of highly electrocompetent *Pseudomonas aeruginosa cells:* application for DNA fragment transfer between chromosomes and plasmid transformation. J Microbiol Methods, 2006. **64**(3): p. 391-7.
250. Chuanchuen, R., C.T. Narasaki, and H.P. Schweizer, Benchtop and microcentrifuge preparation of *Pseudomonas aeruginosa* competent cells. Biotechniques, 2002. **33**(4): p. 760, 762-3.
251. Marco Galardini, E.G.B., Marco Bazzicalupo and Alessio Mengoni, CONTIGuator: a bacterial genomes finishing tool for structural insights on draft genomes. Source Code for Biology and Medicine, 2011. **6:11**.
252. Bosi, E., et al., MeDuSa: a multi-draft based scaffolder. Bioinformatics, 2015. **31**(15): p. 2443-51.
253. Altschul, S.F., et al., Basic local alignment search tool. J Mol Biol, 1990. **215**(3): p. 403-10.
254. Mount, D.W., Maximum parsimony method for phylogenetic prediction. CSH Protoc, 2008. **2008**: p. pdb top32.
255. Julie D.Thompson, D.G.H.a.T.J.G., CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res, 1994. **22**(22): p. 4673-4680.
256. Tamura, K., et al., MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol, 2013. **30**(12): p. 2725-9.
257. Felsenstein, J., Confidence Limits on Phylogenies: An Approach Using the Bootstrap. Evolution, 1985. **39**(4): p. 783-791.
258. Kumar, M.N.a.S., Molecular Evolution and Phylogenetics Oxford University Press, 2000.
259. Wagner, J., et al., Identification and characterisation of *Pseudomonas* 16S ribosomal DNA from ileal biopsies of children with Crohn's disease. PLoS One, 2008. **3**(10): p. e3578.
260. McWilliam, H., et al., Analysis Tool Web Services from the EMBL-EBI. Nucleic Acids Res, 2013. **41**(Web Server issue): p. W597-600.

261. M, T.K.a.N., Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. . Molecular Biology and Evolution, 1993. **10**: p. 512-526.
262. Weber, T., et al., antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res, 2015. **43**(W1): p. W237-43.
263. Finn, R.D., et al., InterPro in 2017-beyond protein family and domain annotations. Nucleic Acids Res, 2017. **45**(D1): p. D190-D199.
264. Huber, W., et al., Orchestrating high-throughput genomic analysis with Bioconductor. Nat Methods, 2015. **12**(2): p. 115-21.
265. Gentleman, R.C., et al., Bioconductor: open software development for computational biology and bioinformatics. Genome Biol, 2004. **5**(10): p. R80.
266. Lawrence, M., et al., Software for computing and annotating genomic ranges. PLoS Comput Biol, 2013. **9**(8): p. e1003118.
267. Love, M.I., W. Huber, and S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol, 2014. **15**(12): p. 550.
268. Witten, D.M., Classification and clustering of sequencing data using a Poisson model. The Annals of Applied Statistics, 2011. **5**(4): p. 2493-2518.
269. Wickham, H., ggplot2. Springer-Verlag New York, 2009. **1**.
270. Garneau, J.R., et al., PhageTerm: a Fast and User-friendly Software to Determine Bacteriophage Termini and Packaging Mode using randomly fragmented NGS data. bioRxiv, 2017.
271. Afgan, E., et al., The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Research, 2016. **44**(W1): p. W3-W10.
272. Yamamoto, S., et al., Phylogeny of the genus *Pseudomonas:* intrageneric structure reconstructed from the nucleotide sequences of *gyrB* and *rpoD* genes. Microbiology, 2000. **146 (Pt 10)**: p. 2385-94.
273. Gast, F.U., et al., The recognition of methylated DNA by the GTP-dependent restriction endonuclease McrBC resides in the N-terminal domain of McrB. Biol Chem, 1997. **378**(9): p. 975-82.
274. Steczkiewicz, K., et al., Sequence, structure and functional diversity of PD-(D/E)XK phosphodiesterase superfamily. Nucleic Acids Res, 2012. **40**(15): p. 7016-45.
275. Remminghorst, U. and B.H. Rehm, Alg44, a unique protein required for alginate biosynthesis in *Pseudomonas aeruginosa.* FEBS Lett, 2006. **580**(16): p. 3883-8.
276. Maleki, S., et al., Alginate Biosynthesis Factories in *Pseudomonas fluorescens:* Localization and Correlation with Alginate Production Level. Appl Environ Microbiol, 2016. **82**(4): p. 1227-1236.
277. Kang Yoon-Suk, Y.J.K., Che Ok Jeon and Woojun Park, Characterization of Napthalene-Degrading *Pseudomonas* Species Isolated from Pollutant-Contaminated Sites: Oxidative Stress During their. Journal of Microbiology and Biotechnology, 2006. **16**(11): p. 1819-1825.
278. Richter, M., et al., JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. Bioinformatics, 2016. **32**(6): p. 929-31.
279. Rokni-Zadeh, H., et al., Distinct lipopeptide production systems for WLIP (white line-inducing principle) in *Pseudomonas fluorescens* and *Pseudomonas putida.* Environ Microbiol Rep, 2013. **5**(1): p. 160-9.
280. Bickle, T.A. and D.H. Kruger, Biology of DNA restriction. Microbiol Rev, 1993. **57**(2): p. 434-50.
281. Parte, A.C., LPSN--list of prokaryotic names with standing in nomenclature. Nucleic Acids Res, 2014. **42**(Database issue): p. D613-6.

282. Gomila, M., et al., Phylogenomics and systematics in *Pseudomonas.* Front Microbiol, 2015. **6**: p. 214.
283. Migula, W., System der Bakterien. Handbuck der Morphologie, Entwickelung-Geschichte und Systematik der Bakterien. Jena: Verlag von Gustav Fischer, 1900. **2**.
284. Palleroni, N.J., et al., Nucleic Acid Homologies in the Genus *Pseudomonas.* International Journal of Systematic Bacteriology, 1973. **23**(4): p. 333-339.
285. Dahllof, I., H. Baillie, and S. Kjelleberg, *rpoB*-Based Microbial Community Analysis Avoids Limitations Inherent in 16S rRNA Gene Intraspecies Heterogeneity. Applied and Environmental Microbiology, 2000. **66**(8): p. 3376-3380.
286. Fox, G.E., J.D. Wisotzkey, and P. Jurtshuk, Jr., How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. Int J Syst Bacteriol, 1992. **42**(1): p. 166-70.
287. Ochman, H. and A. C. Wilson, Evolution in bacteria: Evidence for a universal substitution rate in cellular genomes. Vol. 26. 1987. 74-86.
288. Mulet, M., et al., Phylogenetic analysis and siderotyping as useful tools in the taxonomy of *Pseudomonas stutzeri:* description of a novel genomovar. Int J Syst Evol Microbiol, 2008. **58**(Pt 10): p. 2309-15.
289. Rajwar, A. and M. Sahgal, Phylogenetic relationships of fluorescent pseudomonads deduced from the sequence analysis of 16S rRNA, *Pseudomonas*-specific and *rpoD* genes. 3 Biotech, 2016. **6**(1): p. 80.
290. Hickson, P.M.W.a.I.D., Structure and function of type II DNA topoisomerases. Biochemical Journal 1994. **303**(3): p. 681-695.
291. Michael Lonetto, M.G.a.C.A.G., The σ70 Family: Sequence Conservation and Evolutionary Relationships. Journal of Bacteriology 1992. **174**(12): p. 3843-3849.
292. Sintchenko, V., et al., *Mycobacterium tuberculosis rpoB* gene DNA sequencing: implications for detection of rifamycin resistance. J Antimicrob Chemother, 1999. **44**(2): p. 294-5.
293. Punekar, A.S., et al., Structural and functional insights into the molecular mechanism of rRNA m6A methyltransferase RlmJ. Nucleic Acids Res, 2013. **41**(20): p. 9537-48.
294. Schalk, I.J. and L. Guillon, Pyoverdine biosynthesis and secretion in *Pseudomonas aeruginosa:* implications for metal homeostasis. Environ Microbiol, 2013. **15**(6): p. 1661-73.
295. Munzinger, M., et al., Achromobactin, a new citrate siderophore of *Erwinia chrysanthemi.* Z Naturforsch C, 2000. **55**(5-6): p. 328-32.
296. Samson, R., et al., Transfer of *Pectobacterium chrysanthemi* (Burkholder et al. 1953) Brenner *et al.* 1973 and *Brenneria paradisiaca* to the genus *Dickeya gen. nov. as Dickeya chrysanthemi comb. nov. and Dickeya paradisiaca comb. nov.* and delineation of four novel species, *Dickeya dadantii sp. nov., Dickeya dianthicola sp. nov., Dickeya dieffenbachiae sp. nov. and Dickeya zeae sp. nov.* Int J Syst Evol Microbiol, 2005. **55**(Pt 4): p. 1415-27.
297. Tomisic, V., et al., Iron(III) uptake and release by chrysobactin, a siderophore of the phytophatogenic bacterium *Erwinia chrysanthemi.* Inorg Chem, 2008. **47**(20): p. 9419-30.
298. Loper, J., Lack of Evidence for In Situ Fluorescent Pigment Production by *Pseudomonas syringae* pv. syringae on Bean Leaf Surfaces. Vol. 77. 1987.
299. Franza, T., B. Mahe, and D. Expert, *Erwinia chrysanthemi* requires a second iron transport route dependent of the siderophore achromobactin for extracellular growth and plant infection. Mol Microbiol, 2005. **55**(1): p. 261-75.
300. Ishii, S.I., Y. Nishi, and F. Egami, The fine structure of a pyocin. J Mol Biol, 1965. **13**(2): p. 428-31.

301. Yui-Furihata, C., Structure of Pyocin R. The Journal of Biochemistry, 1972. **72**(1): p. 1-10.
302. Kuroda, K. and M. Kageyama, Biochemical properties of a new flexuous bacteriocin, pyocin F1, produced by *Pseudomonas aeruginosa.* J Biochem, 1979. **85**(1): p. 7-19.
303. Nakayama, K., et al., The R-type pyocin of *Pseudomonas aeruginosa* is related to P2 phage, and the F-type is related to lambda phage. Mol Microbiol, 2000. **38**(2): p. 213-31.
304. Smith, A.W., et al., The pyocin Sa receptor of *Pseudomonas aeruginosa* is associated with ferripyoverdin uptake. Journal of Bacteriology, 1992. **174**(14): p. 4847-4849.
305. Heussler, G.E., et al., Requirements for *Pseudomonas aeruginosa* Type I-F CRISPR-Cas Adaptation Determined Using a Biofilm Enrichment Assay. J Bacteriol, 2016. **198**(22): p. 3080-3090.
306. Aydin, S., et al., Presence of Type I-F CRISPR/Cas systems is associated with antimicrobial susceptibility in *Escherichia coli.* J Antimicrob Chemother, 2017. **72**(8): p. 2213-2218.
307. Glukhov, A.S., et al., Genomic analysis of *Pseudomonas putida* phage *tf* with localized single-strand DNA interruptions. PLoS One, 2012. **7**(12): p. e51163.
308. Ceyssens, P.J., et al., Genomic analysis of *Pseudomonas aeruginosa* phages *LKD16* and *LKA1:* establishment of the *phiKMV* subgroup within the *T7* supergroup. J Bacteriol, 2006. **188**(19): p. 6924-31.
309. Bardy, S.L., S.Y. Ng, and K.F. Jarrell, Prokaryotic motility structures. Microbiology, 2003. **149**(Pt 2): p. 295-304.
310. Abedon, S.T., Bacteriophage Ecology: Population Growth, Evolution, and Impact of Bacterial Viruses. 2008. **1**: p. 101-111.
311. Ackermann, H.W., 5500 Phages examined in the electron microscope. Arch Virol, 2007. **152**(2): p. 227-43.
312. Adriaenssens, E. and J.R. Brister, How to Name and Classify Your Phage: An Informal Guide. Viruses, 2017. **9**(4).
313. Lavigne, R., et al., A multifaceted study of *Pseudomonas aeruginosa* shutdown by virulent podovirus *LUZ19.* MBio, 2013. **4**(2): p. e00061-13.
314. Zhao, X., et al., Global Transcriptomic Analysis of Interactions between *Pseudomonas aeruginosa* and Bacteriophage *PaP3.* Sci Rep, 2016. **6**: p. 19237.
315. Keiler, K.C., Biology of trans-translation. Annu Rev Microbiol, 2008. **62**: p. 133-51.
316. Letunic, I. and P. Bork, Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res, 2016. **44**(W1): p. W242-5.

Appendix 1: Additional *P. agarici* NCPPB 2472 Contig Sequencing Data

Reaction 1 Sequence:
Contig 3 End -
TGTAACGCTGTAGAATTCGCCTCCCGCTGTCGATAGGTCGACAGCGCAAGTGGTT
GAAGTTGAAGCGTTTTTTGAACGAAAAGCCTTGAAAACTTCTCAAATAATCACTT
GACAGCAACTGGCGCTGCTGTAGAATGCGCGCCTCGGTTCAGCGAACAGCTCAA
CCCACCGCTCTTTAACAACTGAATCAAGCAATTCGTGTGGGTGCTTGTGGAGTCA
GACTGAATAGTCACTAGATTATCAGCATCACAAGTTACTCCGCGAGAAATCAAA
GATGTAACCAACGATTGCTGAGCCAAGTTTAGGGTTTTCTCAAAACCCAAAGATG
TTTGAACTGAAGAGTTTGATCATGGCTCAGATTGAACGCTGGCGGCAGGCCTAAC
ACATGCAAGTCGAGCGGATGAAGAGAGCTTGCTCCCGGATTCAGCGGCGGACGG
GTGAGTAATGCCTAGGAATCTGCCTGGTAGTGGGGGACAACGTTTCGAAAGGAA
CGCTAATACCGCATACGTCCTACGGGAGAAAGCAGGGGACCTTCGGGCCTTGCG
CTATCAGATGAGCCTAGGTCGGATTAGCTAGTTGGTGAGGTAAAGGCTCACCAA
GGCGACGATCCGTAACTGGTCTGAGAGGATGATCAGTCACACTGGAACTGAGAC
ACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGCGAA
AGCCTGATCCAGCCATGCCGCGTGTGTGAAGAAGGTCTTCGGATTGTAAAGCACT
TTAAGTTGGGAGGAAGGGCATTAACCTAATACGTTAGTGTTTTGACGTTACCGAC
AGAATAAGCACCGGCTAACTCTGTGCCAGCAGCCGCGGTAATACAGAGGGTGCA
AGCGTTAATCGGAATTACTGGGCGTAAAGCGCGCGTAGGTGGTTGGTTAAGTTG
GATGTGAAATCCCCGGGCTCAACCTGGGAACTGCATCCAAAACTGGCCAGCTAG
AGTAGGGTAGAGGGTGGTGGAATTTCCTGTGTAGCGGTGAAATGCGTAGATATA
GGAAGGAACACCAGTGGCGAAGGCGACCACCTGGACTCATACTGACACTGAGGT
GCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAA
CGATGTCAACTAGCCGTTGGGAACCTTGAGTTCTTAGTGGCGCAGCTAACGCATT
AAGTTGACCGCCTGGGGAGTACGGCCGCAAGGTTAAAACTCAAATGAATTGACG
GGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAAC
CTTACCAGGCCTTGACATCCAATGAATCTTCCAGAGATGGAGGAGTGCCTTCGGG
AACATTGAGACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGG
TTAAGTCCCGTAACGAGCGCAACCCTTGTCCTTAGTTACCAGCACGTGATGGTGG
GCACTCTAAGGAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCA
AGTCATCATGGCCCTTACGGCCTGGGCTACACACGTGCTACAATGGTCGGTACAA
AGGGTTGCCAAGCCGCGAGGTGGAGCTAATCCCATAAAACCGATCGTAGTCCGG
ATCGCAGTCTGCAACTCGACTGCGTGAAGTCGGAATCGCTAGTAATCGCGAATC
AGAATGTCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCAT
GGGAGTGGGTTGCACCAGAAGTAGCTAGTCTAACCTTCGGGAGGACGGTTACCA
CGGTGTGATTCATGACTGGGGTGAAGTCGTAACAAGGTAGCCGTAGGGGAACCT
GCGGCTGGATCACCTCCTTAATCGACGACATCAGCTGCTGCATAAGTACCCACAC
GAATTGCTTGATTCATTGAAGAAGACGATAGAAGCAGCTCCAGGCTCTGTGATA
AAGAGCAAGGTAAAGTCCGCTTGTTACACCCAGATTCTGGGTCTGTAGCTCAGTT
GGTTAGAGCGCACCCCTGATAAGGGTGAGGTCGGCAGTTCGAATCTGCCCAGAC
CCACCAGTTACTTGGTGAGGGACTTGGCAAGAGCGCACCCCCGCCTCTGAGGTA
AAGAGCGGTGAGGTCAGGGTTTTTTGCAGTTCGAATCTGCCCAGACCCACCAGTT
TGTGTGGGAAGTGTGGTAGGAACCTGTAGATATACGGGGCCATAGCTCAGCTGG
GAGAGCGCCTGCCTTGCACGCAGGAGGTCAACGGTTCGATCCCGTTTGGCTCCAC
CATTAACTGCTTCTGATGTTAGAGCTTAGAAATGAATATTCACACACGAATATTG
ATTTCTAGTCTTTGATTAGATCGTTCTTTAAAAATTTGGGTATGTGATAGAAAGAT
AGACTGAACGTTACTTTCACTGGTAACGGATCAGGCTAAGGTAAAATTTGTGAGT
TATCTTACAGATTTTCGGCGAATGTCGTCTTCATACTGTAACCAGATTGCTTGGG

GTTATAGGGTCAAGTGAAGAAGCGCATACGGTGGATGCCTTGGCAGTCAGAGGC
GATGAAAGACGTGGTAGCCTGCGAAAAGCTTCGGGGAGTCGGCAAACAGACTGT
GATCCGGAGATGTCTGAATGGGGGAACCCAGCCATCATAAGATGGTTATCTTGT
ACTGAATACATAGGTGCAAGAGGCGAACCAGGGGAACTGAAACATCTAAGTACC
CTGAGGAAAAGAAATCAACCGAGATTCCCTTAGTAGTGGCGAGCGAACGGGGAC
TAGCCCTTAAGTGGCTTTGAGATTAGCGGAACGCTCTGGAAAGTGCGGCCATAGT
GGGTGATAGCCCTGTACGCGAAAATCTCTTAGTCATGAAATCGAGTAGGACGGG
GCACGAGAAACCTTGTCTGAATATGGGGGGACCATCCTCCAAGGCTAAATACTA
CTGACTGACCGATAGTGAACTAGTACCGTGAGGGAAAGGCGAAAAGAACCCCGG
AGAGGGGAGTGAAATAGATCCTGAAACCGTCTGCGTACAAGCAGTGGGAGCCCA
CTTGTTGGGTGACTGCGTACCTTTTGTATAATGGGTCAGCGACTTATTTTCAGTGG
CAAGCTTAACCGAATAGGGGAGGCGTAGCGAAAGCGAGTCTTAATAGGGCGTCT
AGTCGCTGGGAATAGACCCGAAACCGGGCGATCTATCCATGGGCAGGTTGAAGG
TTGGGTAACACTAACTGGAGGACCGAACCGACTACCGTTGAAAAGTTAGCGGAT
GACCTGTGGATCGGAGTGAAAGGCTAATCAAGCTCGGAGATAGCTGGTTCTCCT
CGAAAGCTATTTAGGTAGCGCCTCATGTATCACTGTAGGGGGTAGAGCACTGTTT
CGGCTAGGGGGTCATCCCGACTTACCAAACCGATGCAAACTCCGAATACCTACA
AGTGCCGAGCATGGGAGACACACGGCGGGTGCTAACGTCCGTCGTGAAAAGGGA
AACAACCCAGACCGTCAGCTAAGGTCCCAAAGTCATGGTTAAGTGGGAAACGAT
GTGGGAAGGCTTAGACAGCTAGGAGGTTGGCTTAGAAGCAGCCACCCTTTAAAG
AAAGCGTAATAGCTCACTAGTCGAGTCGGCCTGCGCGGAAGATGTAACGGGGCT
CAAACCATGCACCGAAGCTACGGGTATCACTTAGGTGATGCGGTAGAGGAGCGT
TCTGTAAGCCTGTGAAGGTGAGTTGAGAAGCTTGCTGGAGGTATCAGAAGTGCG
AATGCTGACATGAGTAACGACAATGGGTGTGAAAAACACCCACGCCGAAAGACC
AAGGTTTCCTGCGCAACGTTAATCGACGCAGGGTGAGTCGGTCCCTAAGGCGAG
GCTGAAAAGCGTAGTCGATGGAAAACAGGTTAATATTCCTGTACTTCTGGTTATT
GCGATGGAGGGACGGAGAAGGCTAGGCCAGCTTGGCGTTGGTTGTCCAAGTTTA
AGGTGGTAGGCTGGAATCTTAGGTAAATCCGGGGTTTCAAGGCCGAGAGCTGAT
GACGAGTTACCCTTTAGGGTGACGAAGTGGTTGATGCCATGCTTCCAAGAAAAG
CTTCTAAGCTTCAGATAACCAGGAACCGTACCCCAAACCGACACAGGTGGTTGG
GTAGAGAATACCAAGGCGCTTGAGAGAACTCGGGTGAAGGAACTAGGCAAAAT
GGCACCGTAACTTCGGGAGAAGGTGCGCCGGTGAGGGTGAAGGACTTGCTCCGT
AAGCTCATGCCGGTCGAAGATACCAGGCCGCTGCGACTGTTTATTAAAAACACA
GCACTCTGCAAACACGAAAGTGGACGTATAGGGTGTGACGCCTGCCCGGTGCCG
GAAGGTTAATTGATGGGGTTAGCTCACGCGAAGCTCTTGATCGAAGCCCCGGTA
AACGGCGGCCGTAACTATAACGGTCCTAAGGTAGCGAAATTCCTTGTCGGGTAA
GTTCCGACCTGCACGAATGGCGTAACGATGGCGGCGCTGTCTCCACCCGAGACTC
AGTGAAATTGAAATCGCTGTGAAGATGCAGTGTATCCGCGGCTAGACGGAAAGA
CCCCGTGAACCTTTACTATAGCTTTGCACTGGACTTTGAATTTGCTTGTGTAGGAT
AGGTGGGAGGCTTTGAAGCGTGGACGCCAGTTCGCGTGGAGCCAACCTTGAAAT
ACCACCCTGGCAACTTTGAGGTTCTAACTCAGGTCCGTTATCCGGATCGAGGACA
GTGTATGGTGGGTAGTTTGACTGGGGCGGTCTCCTCCTAAAGAGTAACGGAGGA
GTACGAAGGTGCGCTCAGACCGGTCGGAAATCGGTCGTAGAGTATAAAGGCAAA
AGCGCGCTTGACTGCGAGACAGACACGTCGAGCAGGTACGAAAGTAGGTCTTAG
TGATCCGGTGGTTCTGTATGGAAGGGCCATCGCTCAACGGATAAAAGGTACTCC
GGGGATAACAGGCTGATACCGCCCAAGAGTTCATATCGACGGCGGTGTTTGGCA
CCTCGATGTCGGCTCATCACATCCTGGGGCTGAAGCCGGTCCCAAGGGTATGGCT
GTTCGCCATTTAAAGTGGTACGCGAGCTGGGTTTAGAACGTCGTGAGACAGTTCG
GTCCCTATCTGCCGTGGACGTTTGAGATTTGAGAGGGGCTGCTCCTAGTACGAGA
GGACCGGAGTGGACGAACCTCTGGTGTTCCGGTTGTCACGCCAGTGGCATTGCCG

GGTAGCTATGTTCGGGAAAGATAACCGCTGAAAGCATCTAAGCGGGAAACTTGC
CTCAAGATGAGATCTCACTGGAACCTTGAGTTCCCTGAAGGGCCGTCGAAGACT
ACGACGTTGATAGGTGGGGTGTGTAAGCGCTGTGAGGCGTTGAGCTAACCCATA
CTAATTGCCCGTGAGGCTTGACCCTATAACACCCAAGCAATTTGTGAACTTGAGC
CTGAAGGCCAGAGAGCAGATTGCGGGGTGTGAAGACGAAACGAACCGAAAGTT
TGTGCTGAC - Contig 21 Start

Reaction 2 Sequence:
Contig 21 End -
ATACCCACCGAGTCTCTTCTGAGCTTCTCCACACTTTGGGCAAATGGCCTTTATCT
CGGGCAGCCGCTTGCCGATTGCGAACAGCCTGCCTGACAGCTTTTCACTAATAGC
GGGCTTAGCTTAAAGCTCACTCGCCCATCGCTATGTATCGCGTGATCGTCGAAGT
GGAGAAAAACTAAATGAGTGATGGTTATTTTATCGGCCTGGGTGACAAAACCAC
CTGCGGCGGGGAAGTTCTGGATGGTGACGAGAGAATCAATATGTTTGGCGTCCT
GCATGCCTGAGAAGGTGATCGGGTCTCATGTGGAAGGACGGAAAGACTTATCGG
ATCGTCGGTGGCGTTTCCCATATGAATAGCCATGATCGGCTGATGGCTGGCACGC
TGGATAGCCACAGCGATTGCCCCTGCAAAGCCAAACTGGTCCCCTCGGTTCTGAC
GGCTTGCTATCGTAATGGCCCCGCAACCAGCCGAGTCGCCGAGCAACCGGCCTCT
TCAGCAGCTACCAGCCGCTCGCCAGCGCCGCTTAAATCTGTTTTCTCTCCTCCAA
GCCCTCCAGCCCCAGGGATATTCAGCCGCGTGGAGCCCCAGGAGCCGGGTTTCC
ATGTAGTGCCCAAAAGCATGACCCGCGAGGCACTGGAGGCCACGCTGTTCCCCA
CGCCCGACTCGGCGGTGATGCGCAAGTTCCGGGCGCTCAACCCTTATCGTGGCGA
CGTCAAGGCCGGATCGCTGATCGTCCTTGGCGATCCGAACAACCTGCTCTGTACC
CGCGAAGAGGCGCAGTTGATGGCGGCGGCGCAAGCGGTGCATGCCGAGCTGGA
ACCCCTGACACCGGAACAAGCCGACTTCATGCAGCGCCACAGTGTCGAAATTGC
CAGTTTTGCCGGCCATGCTTNCAATCT - Contig 38 Start

Reaction 18 Sequence
Contig 11 End -
CAAAGAAGAAGGCGAAGTAACGGTCAAGGGCGCCAACGCCGGCCATGCCACCG
AAGCCAACATGGACCGCCTGGACAACGCCGGCGGTGCCGCCGCCCTCGAAGCCC
GTGCCGTGACCGCCAACGACTCCGCCGCGATCATCCGCGCCAAGGCCGCTCTCG
ACGCCCTCGATGTCGCCGAAGGCCTGGCCGAACTCGAAGGCGCCTCGGCTCGGG
TCGCCGTCGATGAAAAGCGCATGATCAACTGCCGCGCCGACCTCAACCAGCTCG
TCCCCTTCAAGTACGACTGGGCCTGGCAGAAGTACCTCGACGGCTGCGCCAACC
ACTGGATGCCGCAAGAGGTCAACATGACCGCCGACATCGCCCTGTGGAAAAACC
CCGAAGGCCTGACCGACGACGAGCGCCGCATCGTCATGCGCAACCTCGGCTTCTT
CTCCACCGCCGACTCCCTGGTGGCCAACAACCTGGTACTGGCCGTCTACCGCCTG
ATCACCAACCCCGAGTGCCGCCAGTACATCCTGCGCCAGGCCTTCGAAGAGGCC
ATCCACACCCACGCCTATCAGTACTGCATCGAATCGCTGGCCATGGATGAAGGC
GAAATCTTCAACATGTACCACGAGATTCCTTCGGTGGCGAAAAGGCCGCCTGG
GGCCTGAAGTACACCCGTTCGATCTCCGATCCGAAGTTCGAGACCGGTACCGTCG
ACACCGACAAGGAACTGCTGCGCAACCTGATCGCCTACTACTGCGTGCTCGAAG
GTATCTTCTTCTACTGTGGCTTCACCCAGATCCTGTCCATGGGCCGGCGCAACAA
GATGACC - Contig 9 Start

Reaction 19 Sequence
Contig 9 End -
CATCGTAGGCGAAGGTTTCGCTATGGCCCCGTGGGTTCTCCTGGGTGGCGATGAT
TCGGTCGGTGGCGTCGTAGTGCAGGTGTTGACGCTGGTCGCGGCGGGCGGGGTT

ATCGCGTTCGTCACGCTCGATCAGGTTGTCGAGGACATCGTATTCGAAGTATTTA
TTCCGGGCGGCAGATAGCATTTCCGGCTGGTTGTTCGGGCGGCGCACCCGTGAGC
GCAGGCGTCCGCACCGGTCATACTCGCTGCGGGTGTTGAGTTGTCCCTGGGTGCG
CGAGAGTTCGCGGTGCAGGCGGTCGCGTTCGAAGTCGCTGATGACCTCGCCGTC
GAGGTTGAGCTGATGCAGGTGGCCACTGCCGTAATACAGCCGATTGAGCCAGCG
CCCATCGGGTAGCTGCGTCTGGCTCAGGTTGCCCAGTTCGTCGTAGTGATGCTGC
AAGCTGCCGGCTGCGCTCTGTTCTTCCAGCAACCGGCCGAGGGCGTCATAGGCG
AAGTTGAGTGTTTGTTCATTACCGTCATGGTCGGTAAAGGTGATGGCGGTCACTT
GGTCCAGCCGGTCGTAGCTGTAGGTGGTGCGGCCATCGGCGGTGATCTTGGCGAT
CAGGCGACCCAGTGCATCGCGTTCCAATTGATGGACGATAGGAGCCGGAGTCGG
GCTTCCATCCTCTTGAGGCGCTGGCAAGTGTTCAATGGCGACTACGTTGTCGTGC
GTGTCATAGCTGTACTGCCGGGCGCTGCCGTCCAGATCCTGTTGGCGGATCCGGC
GGTCGTTGGCATCCCAGGCAAAGCGGTAACTTTCCCCGTTCTCGTTGGTCAGTGC
CTGCAAGCGGCCATAGGCGTCGTAGTGGTATTCGATCTGTCGGCCAAGGGCGTC
GGTGCGTTGGCGTACCTGGCCGCGTCGGTTGTACTGGTAGCGCGTGCTGCCGCCA
GCGGCGTCGATATGGGCGATCAGTTGGCCGTGGGCGTCGCGCTGGTAGTGGTCG
AAGCGGTCCACGGGCCGTTCGGTTTGCAGCAGTCGGCCTTGAGCATCGCGTTGGT
AGTGCACCCGTTCGCCAAAGGCATCGGCCACCGTACACAGGTGGCCGCGTCGGT
CGTAGTAAAAGTCCGTGGGATAGTCGGAGCAATCGGCGTGGTAGGTCAGTTGCC
CGCATTCGTTCCAGCGCAAGGTTCGGCTTTTGCCCGTGGCATCGATGATTTCCAC
CACCTGGCCGAAGGCATCGTAGCGATAGGTGGTGGTGTGCCCCAGCGGGTCGGT
TTCGCTGAGGCAGTTGCCGCGTTGGTCGTAGCGGTACTGCCAGGCCTGGCCGGCG
GCATCGGTTTGCACTTGGGGCAGCGACCAGTGTTCCAGCCACAGGGTGGAGTCG
ATACGGCCCAATGGGTCTTCGCTGGTGGACAGGTTGCCGGAGGCGTCGTAGCTG
AAGCGCCAGTGTCCGTCCTGCGGATCAATGGCGCCGAGTAGCTGGCGTTCGTCGT
TCCACTGAAATTGCCAGGTATGGCCGAGGGGGTCGGTGTAGGCGGTGATCTGGT
ACTGGCGGTTCCACTGGAGGGTGCTGACACGCCGTAAACCATCGGTAATACGTG
TGATGCCAGCCTCCAGGTCGTAGTCGAACTGGTAGTCATCGCCGTCATCGGTCCA
GTGCCGGGTGACGCGCCATTCCTGTGTTTTTTGTGTTTTCTGTGTTTCTTGTGTTTC
CTGTGTTTCGATCAGCGCCCATTGATAGAAACAACGCAGGCCGCTGGGCAATTG
GTGTTCGACCAGGCGCCGGTGTTCGTCATAGGCAAAACGCCGCTGAACTTGCCCT
GTGGCATCGCGAACTTCACTCAGGTCGCCCGCCGCATCGTAGTCATAGCTGACCA
GCACTTCGCGGGTTTGATCGGGGTACAGGCGTTCGATCTGGCCGATACGCCGAG
GCCCTTGGGGGCTGTGGTAGATCAGTTCCACCTGGACTTGGTCGAAGGTGTCACG
CAAGCGAATCAGCGTGCCCGACTTGCTGTAGTCGAGGTAGAGCCGGTTGTCGTTG
CGGTCGCCCAACTGGCTCAGGCGCAGTCGCGAGGGATCGGCGAGGGTCGGTTCA
AACAGCCGGTACAGCCCGTCCTCGCTCTCGATCAGCAACTGCCCGTTGTCGTGGC
GCCGCACGCTCAAGCCTTCGCCGGCACTGAACACCGCATGGCCCAGCGGGATTG
AGCCCATATCGATCTGTCGGGCCTGCTCGTCGGTGTAGATCAGCCGTTCGCCGCC
TTCGGGACGGGCTTCGAGGCGCACGTTGATTTCATACGGCAGGCTCCAGCCCTGG
CCGAGCAAACTGTCGTGGCGTTCATCGCGGCTGTTGTAGAAGCGTTGCCAGTCGA
GGGGCAGCAAACCGGGCAAGGCGAAATCGAGCTCGTCCTCGCCGCCCAGCACCT
TGGCACCAGTGGCGGCGTGGACCGGGTTAGGCGAGCCGGACATCGCCGCGCTCA
GGGCGTTGCTGACCTGGCCGGTGCCCCATGACGTCAGTCCGCCGACCAGCATGC
AGGGCAATTGGCTGAAAAACTTCGCGCCGCCACCGCGCAACATCAACAAGGTGC
CCACCGCCAGGCCCACGCCGGGGGGCTTGCCGCTGCGGATTTCGCGTATCACCAC
GGGCGCGCCACCGATGCGCACATTGGGTGAAATCAGCCCGTTGACCGCACCGAC
CCGAGCCTCGCAGGTGCTGCGATCACCCGTGCGCACGGCGGGCTGGCCGTTGAT
GGTGACTTTGCTGGAGCCTTCGGCGAGATAGACCTGGGGCATCGGCGGGTGTTTG
TCGCAGGCCACGCGATCATAAGGGCTGGGAATGGCGCCGGGCGCGGGGGTGGCG

ACGGTGGGCCGCCATAGCTGGGAAAAGAAGCCCTGGGCGATATCCAGGTAGCCA
GCGGACTCGCCGGCCAGTACCGGGCCGATCCTGCCAGCGGCGCGGGCGGCGGGC
TTGCCGTTGATAAAGGTGTCGGGCGAGCCGGTGGTGATATGGGCCTGGACCGTG
GGCGGAAACAGCGCATTGCTTAGCCAGTCGCAGAGCTGTTTCACCTCCTCGTCGG
TGCCGGTCTGGTTCATGCCGACGCCGACCGCAATGCCCACCGCGCTCCCGAGTAC
GAGGCAGCCCAGCCCACCGGTGGCGACGGTGATACCCGTTGCCGCGACAACCGC
CGTGGTGGCCAGTGCGCCGACCGCGACAGAGGCGGCGATTTCCAGGACTCCGCC
CAGGAGGTCGGCCATCGGCGCGGTGTGCAGCAGGGCATCGCCTTCGC - Contig 42 Start

Reaction 23 Sequence
Contig 29 End -
ACCACCTACAGCTACGACCGGCTGGACCAAGTGACCGCCATCACCTTTACCGACC
ATGACGGTAATGAACAAACACTCAACTTCGCCTATGACGCCCTCGGCCGGTTGCT
GGAAGAACAGAGTGCAGCGGGTAGCTTGCAGCATCACTACGACGAACTGGGCAA
CCTGAGCCAGACGCAGCTACCCGATGGTCGCTGGCTCAATCGGCTGTATTACGGC
AGTGGCCACCTGCACCAGCTCAACCTCGACGGCGAGGTCATCAGCGACTTCGAA
CGCGACCGCCTGCACCGCGAACTCTCGCGCACCCAGGGACAACTCAACACCCGC
AGCGAGTATGACCGGTGCGGACGCCTGCGCTCACGGGTGCGCCGCCCGAACAAC
CAGCCGGAAATGCTATCTGCCGCCCGGAATAAATACTTCGAATACGATGTCCTCG
ACAACCTGATCGAGCGTGACGAACGCGATAACCCCGCCCGCCGCGACCAGCGTC
AACACCTGCACTACGACGCCACNGACCGAATCATCGCCACCCAGGAGAACCCAC
GGGGCCATAGCGAAACCTTCGCCTACGATGCCGCCGCCAACCTGCTCAACGGCC
AGCAACAAAACGCGGNACGGGTGCTGCACAACAAGCTGCTGACCTATCAGGACA
AACGCTATCGCTACGATGGCTTTGGCCGCATGATCGAAAAACGCAGCGCCAGCC
ATCGGGTACAACGTTTTGCCTATGACGCCGAACACCGCCTGATCGAAGTCCACAA
CCAGGACGGCGTCCGCGAAACCGTGGTGCGCATGACCTACGACCCGCTGGGCCG
GCGCATCGGCAAAACCGAACACAACCACAACGGCTATTTGCTGGGCGAAACCCG
TTTTACCTGGGACGGTTTGCGGCTGTTGCAGGAACACAAGAACACCCAAACCAG
CCTCTACCTGTATGTCGACGANAGCTATGAACCGTTGGCGCGGGTCGATGGCCTC
GGCGACGTGCAGAAAATCCGCTACTACCACACCGACCCCAACGGCCTGCCCGAG
CAACTGACGGAGGCCGAT Contig 6 Start

Reaction 26 Sequence
Contig 24 End -
TGTATACCCACCGAGTCTCTTCTGAGCTTCTCCACACTTTGGGCAAATGGCCTTTA
TCTCGGGCAGCCGCTTGCCGATTGCGAACAGCCTGCCTGACAGCTTTTCACTAAT
AGCGGGCTTAGCTTAAAGCTCACTCGCCCATCGCTATGTATCGCGTGATCGTCGA
AGTGGAGAAAAACTAAATGAGTGATGGTTATTTTATCGGCCTGGGTGACAAAAC
CACCTGCGGCGGGGAAGTTCTGGATGGTGACGAGAGAATCAATATGTTTGGCGT
CCTGCATGCCTGCGAAGGTGATCGGGTCTCATGTGGAAAAGGACGGAAAGACTT
ATCGGATCGTCGGTGGCGTTTCTCATATGAATAGCCATGGTCGGCTGATGGCTGG
CACGCTGGATAGCCACAGCGATTGCCCCTGCAAAGCCAAACTGGTCCCCTCGGTT
CTGACGGCTTGCTATCGTAATGGACCCGCAACCAGCCGAGTCGCCGAGCAACCG
GCCTCTTCAGCAGCTACCAGCCGCTCGCCAGCGCCGCTTAAATCTGTTTTCTCTCC
TCCAAGCCCTCCAGCCCCAGGGATATTCAGCCGCGTGGAGCCCCAGGAGCCGGG
TTTCCATGTAGTGCCCAAAAGCATGACCCGCGAGGCACTGGAGGCCACGCTGTTC
CCCACGCCCGACTCGGCGGTGATGCGCAAGTTCCGGGCGCTCAACCCTTATCGTG
GCGACGTCAAGGCCGGATCGCTGATCGTCCTTGGCGATCCGAACAACCTGCTCTG
TACCCGCGAAGAGGCGCAGTTGATGGCGGCGGCGCAAGCGGTGCATGCCGAGCT

GGAACCCCTGACACCGGAACAAGCCGACTTCATGCAGCGCCACAGTGTCGAAAT
TGCCAGTTTTGCCGGCCATGCCTCAA - Contig 15 Start

Reaction 28 Sequence

Contig 13 End -

GCTACGACGCCAGCGCCGACGGTACGACCGCCTTCACGAATAGCGAAACGCAGG
CCGTCTTCCATCGCGATGGTCTTGATCAGCGTGACAGTCATCTGAATATTGTCAC
CTGGCATCACCATTTCAACGCCTTCTGGCAGTTCGCAGTTACCGGTCACGTCAGT
TGTACGGAAGTAGAACTGTGGACGGTAGCCTTTGAAGAACGGCGTGTGACGACC
GCCTTCTTCCTTGCTCAGAACATAGACTTCTGCAGTGAACTTGGTGTGCGGCTTG
ACCGAACCTGGCTTGACCAGAACCTGGCCACGCTCCACGTCGTCACGCTTGGTAC
CACGCAGCAGCACGCCGCAGTTCTCGCCGGCACGACCTTCGTCGAGCAGCTTGC
GGAACATTTCAACACCGGTGCAGGTGGTGGTGGTGGTATCACGCAGACCAACGA
TTTCCAGTGGATCCTGAACACGAACGATACCACGCTCGATACGACCGGTTACCAC
AGTGCCGCGACCGGAGATCGAGAACACGTCTTCGATTGGCATCAGGAACGGCTT
GTCGATAGCGCGCTCTGGCTCTGGAATGTAGCTGTCCAGAGTTTCCACCAACTTC
TTGACGGCGGTGGTGCCCATCTCGTTGTCGTCTTTACCCTCCAGCGCCATACGGG
CCGAACCGATGATGATCGGAGTGTCATCACCTGGGAAGTCGTAGGTGCTCAGCA
GGTCGCGAACTTCCATCTCAACCAGTTCCAGCAGCTCTGCGTCGTCTACCAGGTC
AGCCTTGTTCAGGAAAACCACGATGTACGGAACGCCTACCTGACGGGACAGCAG
GATGTGCTCACGGGTTTGCGGCATCGGACCATCGGCGGCCGAGCAAACCAGGAT
CGCGCCGTCCATCTGCGCAGCACCGGTGATCATGTTCTTCACGTAGTCGGCGTG -
Contig 46 Start

Reaction 30 Sequence

Contig 12 End -

CCCGATCACCGCGTCCTGTACGCCGCGGGGATAAGCCCAATCAGTGACCCGGAA
GGGAAAACCTCATGACCGACCCTGACATCGTTGTCCCCGTTGCCCTGAAACACTC
CCAACAAGTGACCTGCAGCGCGCCCTGGTACGTGCAGAGGAGTGAATATCACCC
CATGGAGGCCACCTATCAGCCGTTGATCAACGGCGAAGAAACCTTCAAGGCCGT
GCATCTGGCGATTGCCCGAGCCACTAAAACCATCGACATCATCTGTTGGGGCTTT
CAGCCGTCGATGTATTTCATTCGCGACGGTAAAGCCCCGAGTATCGGCGAACTGC
TCAAGGCCAAGGCTCGGGAGGGAGTCAAGGTGCGCGTATTGGGATGGGAAATGC
CGCTCAACCTCGCGGGTTTCGCCGGTGAGGCGAACCTGCCCGGTAAAGGTACGG
TTCGCCTCAAGGACCGAGCCATGCAACGTTCTACTCAGGCGCAATACGACGAGG
ACCGCCGGTGGTTTGCCGAATGTGCGGTCGCGGACGACAAGGCGGCTCAGCAGG
GAGCGAGTGGGCTTCCGGTGTTTGTCAGTCGCGGTTTCGACCTGAATGAAAGAGC
CAAAATTGCCCATGAGGTGAAATACGAAGGCCTCGATCGCGAGATCAGTGACAA
GATGCGTCACACACTGAGGTGGACAGCCACCCATCACCAAAAGAGTGTGCTGGT
CGACTATGGCCTGCCGGATTGCGCCGTCGGTTTCGTCATGGGCCACAACATGCTC
GACGAATACTGGGACACCAACGCGCACTCGGCGCTGAACCGCTCCGAGGATAGT
AAGCCTGCCCCCAATAGCGGCCCGCGTGGCAATACCCCACGCCAGGATATTTCC
AGCCAGATCAGCGGGCCGATTCTGGAACACCTGCACCATAACTTCGCCAGCGCC
TGGCGTAAGGAAACCGGCGAGGACCTGCTCATTTCTCGCCAGGCCAAACAGGTG
GGACCGCACCTGAAATGCACCCCCGGTGCCACTCGCCAACTGGCCCAACTGCTA
CGCACCCAGGCGCAGTCGGGCAAGCGCGATATCGAAAGGGCTTACCTGAAAGCG
GTCAACAACGCGACCCAGTTCATCTACATCGAGAACCAGTACTTTCGCTGGCCAC
CGCTGGCCGAAGCCATCAAAAAGGCCGCGGCCGACCAGACCGGCGCGGGACGC

GATCCCGGCTTGCATGGCGCCTTACACCTGTTCGTGATCACCAATGCCACGGACG
ATGGCATTGGCGCGGGCACGGTGAACACCCAGCGCATGCTCGACAGCCTTGGAC
GCGCCGAGACCATTCCCGAGGTCACCAAGCTGCGGCTGATCAAAAAGATCAAAA
AAGATGCCCCACCCCAACCGCAACCTGATCTGCGTGATCACGCAGGGAAGAGGG
AATTGGCTAACTGGCAGGCCGAACTTGATCGGCAAATCGAGGACGTCCAGAACA
GCACAATCGTGCCGCAGAAGGTACCGGGCCTGAAAATCCACGTGTGCTCGCTGG
TTGCGCCCGATTCACCGGCTGGGAAACCGTGGATGCCGGTCTATATCCACTCCAA
GCTGATGATCGTCAATGACNTGTTCACCACCCACGGTTCGGCCAACATCAACACC
CGCAGCATGCAGGTGG - Contig 23 Start

Reaction Rb 14 Sequence

AAATGCGCCTGCTCTCCTGGCCGTTGATTGCCGCAGGACAAAAAGTCTGGCTGCG
GCTGGAGGGACTGAACGCCAACAACGCGCCGCACAATCACACCCTGTGGGAAGC
CGCCACGGTTTCATCGACGTGGGTCAGCGACGGCCATGCGGAGATATTGGTACC
GGCCAGCTACCTGAGTGGATTGGGCCATGGAACTTCCCTGTCGGCGACGTTCAAG
GCGACGTTTGATCGCAGCAGCGTCGAGGCCAATGCGCTGACGTTCCCGGCGCAC
GGATTGAGGGTGGAAAACAACCCGCGCA - Contig 22/Contig 35 Overlap (Contig 22 in Pink)

Reaction Rb20 Sequence
Contig 15 End -
TTGAGGCATGGCCGGCAAAACTGGCAATTTCGACACTGTGGCGCTGCATGAAGT
CGGCTTGTTCCGGTGTCAGGGGTTCCAGCTCGGCATGCACCGCTTGCGCCGCCGC
CATCAACTGCGCCTCTTCGCGGGTACAGAGCAGGTTGTTCGGATCGCCAAGGAC
GATCAGCGATCCGGCCTTGACGTCGCCACGATAAGGGTTGAGCGCCCGGAACTT
GCGCATCACCGCCGAGTCGGGCGTGGGGAACAGCGTGGCCTCCAGTGCCTCGCG
GGTCATGCTTTTGGGCACTACATGGAAACCCGGCTCCTGGGGCTCCACGCGGCTG
AATATCCCTGGGGCTGGAGGGCTTGGAGGAGAGAAAACAGATTTAAGCGGCGCT
GGCGAGCGGCTGGTAGCTGCTGAAGAGGCCGGTTGCTCGGCGACTCGGCTGGTT
GCGGGTCCATTACGATAGCAAGCCGTCAGAACCGAGGGGACCAGTTTGGCTTTG
CAGGGGCAATCGCTGTGGCTATCCAGCGTGCCAGCCATCAGCCGACCATGGCTA
TTCATATGAGAAACGCCACCGACGATCCGATAAGTCTTTCCGTCCTTTCCACATG
AGACCCGATCACCTTCGCAGGCATGCAGGACGCCAAACATATTGATTCTCTCGTC
ACCATCCAGAACTTCCCCGCCGCAGGTGGTTTTGTCACCCAGGCCGATAAAATAA
CCATCACTCATTTAGTTTTTCTCCACTTCGACGATCACGCGATACATAGCGATGG
GCGAGTGAGCTTTAAGCTAAGCCCGCTATTAGTGAAAAGCTGTCAGGCAGGCTG
TTCGCAATCGGCAAGCGGCTGCCCGAGATAAAGGCCATTTGCCCAAAGTGTGGA
GAAGCTCAGAAGAGACTCGGTGGGTATGC - Contig 24 Start

Reaction Rb42 Sequence

ACGGGTCCGCAGCTTCCCCTGGGTCAAGTTTTTTGACGGGGACTTAGCCTCCNAA
AAGCCAGCGACCACTGTCCAGAAATGTAAGTGTAGTACGTAGTGTTTCTAGCCA
AATGGGCCGCGGAGTGCCGGGGAGGCGGTTTTTTAGGACATCGTTGTCCATCTTG
AGCTGCCGGTTCTCTTGCTCCAACTGCCGAATACGCTGCTGCTCAGGTGTCAACG
GCTTACCGATC Contig61/39 Overlap (61 End in Pink)

Reaction 8 (Incomplete)

Contig 59 end-
TCTGTAAGGGGCCGGGGAACACACCTTCCGAGCCCGAGCAATAAGGACGATTGT
GTCCGCGTATTTTTAAGCGGACGCAATTACCCTTAATGCCAGGATCGTCATGCGC
CAACGTAAGTCATACCCGAAATCCTTCAAGACCCAAGTCGTTCAAGAGTGCGAG
CAGCCCGGTGTTTCCGTGGCAGCTATTGCGATGAGTCACGGGATTAATGCCAATG
TCGTTCGCCGGTGGATACCGCTTTACCGTGATCAGCAGACAGTCGCGCTGCCAGC
TTTCATTCCTTTGAAAGTCGCGCCGGCTGAACCAAAACATAAGACCGAAGCGTCG
GCGATCATTGAGCTGCCGCTTGGCGAGCAATCACTCATCGTGAAATGGCCAACTT
CCGACCCTGACGGGTGCGCCCGCTTTGTCCGAGGGCTTGTCCTTTGATCCGCATC
GATGCCATCTGGCTCGCCACCGAGCCGATGGACATGCGCGCCGGTACCGAGACG
GCATTGGCCAGGGTGATCGCGGTGTTCGGTGCGGCGAGGCCGCACTGCGCTTATC
TGTTCGCCAACCGCCGCGCCACACGCATGAAAGTTTTGGTGCATGACGGCTTCGG
TATCTGGCTGGCGGCTCGCCGATTGAACCAAGGCAAGTTCCACTGGCCAGGTATT
CGCCAAGGCTCTGAATTGGAGTTGGCTCCCGAGCAACTTCAGGCTTTAGTACTGG
GCCTGCCATGGCAACGCGTATGTTCCGGCGGCTCGATCACACTGCTTTAACGGCT
GCCATTAGCCTATCGGTCTATCGCCGCGAACTGCTTGCTCTGGCAAAATCGGCGC
CTTGACTTCGCAACCCAATCTCGATCACCTGACCCCTGAACAACTGCGCGCCTGG
CGGCGCAGTTGATGCAGCGTGTCGAAGACGCTCGACCACCAGGTCGACACGCTG
GGCAAGACGGTCGAAACGATGGGCAAGAAGATCAACCGCGATCAGACGGTGAT
CGAAAAGCTGACCCACGAGATTGCACAGCTCAAGCGTTGAAGTTGCCAAGCGCA
GGAGCAGATGAATCATGAGCAGGCGAGCCTGCTTGAGACCTGATCGATACCGAT
ATCGCGGCGATGAGGCCGAGCTTCAGGCCTGCAAATAGCCCCAGCGGCGACCGA
GAAAAAGCAAACGCCCAAGCGCACGTCGTTGCCGGCAGAGTTTCCACGCACCTT
GATCCATCACGAACCGGACAACACCCACTGCCCGTGCGGCTGCGCGCTCAAGCG
CATCGGTGAAGATGTCAGCGAAAAGCTGGATTACATGCCGGGCGTGTTTACCGTT
GAACGCCATGTTCGTGGCAAGTGGGTTTGCGATAACTGCGAAACGATGATCCAG
GCACCCGTTCCAGCGCAGGTTATTGATAAGGGCATTCCGACTGCGGGCCTACTTG
CCCACGTCATGATCGCCAAGTTTGCCGACCATCTGCCGCTTTACCGTCAGGAATC
GATATTCGGTCGAGCGGGCCTGGCGATTCCACGCTCAACCTTGGCTCAATGGGTT
GGCGTGACTGGGGTTCAGTTGCAGCCGCTGGTCCTGCTTGCGCGAGAACCCCTTC
AAACAAAACCATTT

Appendix 2- Complete list of Primers.

Table Ap2.1: Primers derived from mapping sequencing data of *P. agarici* NCPPB 2472 to *P. fluorescens* A506, used for contig stitching.

Primer:	Length:	Tm:	Sequence:
Seq 1	26	61	GGCTTTTTCCAACAACCGAATCAGAC
Seq 2	21	61	TCCACGCGCCTTAATGTCAGC
Seq 3	28	61	GCTAAAGGTCATTGCCGAACATTCCTAG
Seq 4	21	61	CTGTCGATAGGTCGACAGCGC
Seq 6	21	61	CTGTCGATAGGTCGACAGCGC
Seq 7	27	60	GGATTCGGCGCATACTGATAAAGATTG
Seq 8	35	60	CTGACAATGCCTTTGTAGTTCTTAATCTC AAAATC
Seq 9	25	62	GCCGAAATTCAGGTGCATGGATTCG
Seq 10	25	61	GTCGGTGTGGTAGTAGCGGATTTTC
Seq 11	32	61	GCAAGAGCAATCTATTCAGATGTATGTATGCG
Seq 12	22	62	CGCACTTCTTCCAGGGTGTTGC
Seq 13	26	60	GTGGAGTCAGGCTGAATAGTCACTAG
Seq 14	25	60	GTGTTTTCACACGGACGACAGGC
Seq 15	29	62	GCATATTTCAACCTCTCCAATTCCGATGC
Seq 16	23	62	GCTGGGTGCTACCCATGTTTGTC
Seq 17	25	60	CTCATGTAAGGCATAGGCTTGCAAC
Seq 18	28	60	CGTTAAAAGGAGGCGCAATTCTATAGAC
Seq 19	26	61	GAGGGTGAGAGCATGGAATTGTAAGG
Seq 20	23	61	CTTCCAGCCAGCCTTGTTGAGAG
Seq 21	22	61	GCGCGTTACTGAAGAGCGAGAC
Seq 22	26	62	CTGACGATAGCTTTGCTGATGCAAGC
Seq 23	23	62	GAAAGCAGACCGGTGGTTTCTCG
Seq 24	22	61	CGAAGCCGACATCAGTGTGGTC
Seq 25	24	62	GCATCGATCAACTGGCGAATCAGG
Seq 26	25	61	CCGTTGCTCTTGCCTCTATAAGGTC
Seq 27	25	60	GCTCGAGATCTTCCGACATCAAGAG
Seq 28	27	61	GCTGTATTGCACATGCACAATGATGTC
Seq 33	23	61	CAATCTGCTCTCTGGCCTTCAGG
Seq 34	24	61	CTTGTTGCATACCCACCGAGTCTC
Seq 38	24	61	CGTCACACTCATCCACTCATGAGC
Seq 39	29	62	CTTGTTGTATACCCACCGAGTCTCTTCTG
Seq 40	22	62	CCACTGTCACCATCACCGATGC
Seq 41	28	60	CCAGCAAGTTGTAAGTTGTTGTTCTACG
Seq 42	20	60	GCTGGAAGCCCGAATTGCTC
Seq 43	26	61	CGATGTGGTCTTCAGTTCATCGAAGC
Seq 44	20	60	GCTGGAAGCCCGAATTGCTC
Seq 45	20	60	CGGGCAAAGAAGGCTGTAGC
Seq 46	21	61	CACGTCAGGTTCAGGCCATCG
Seq 47	24	62	GTAACTGGCGCTGCTGTAGAATGC
Seq 48	22	61	CAGGCACTGACCAACGAGAACG
Seq 49	23	61	GCTCACTGTCTACCTGCATGCTG
Seq 50	23	61	GGCCCAGGTCATTCAAGAGTGTG
Seq 53	18	61	CAACCCCGCAACGCCAAG
Seq 54	24	61	CAGTTCTTCAGGGCTGTTGCTCTG

Seq 55	20	60	GCATAAGTGTGGCCGTCTGC
Seq 56	21	60	GCAGGATCGAACAGCGCAATG
Seq 57	22	60	GCAATTTCGCCTCTACCAGCAC
Seq 58	17	60	CGTGCTTGGCGCACGTC
Seq 59	25	61	TCAGCAGATCACACTCATGACAAGC
Seq 60	24	61	CCAGACCATTGTGATCTGGTCCTG
Seq 65	25	61	CGTCAGCATGTTCCGATAATCCGTG
Seq 66	20	62	GGTGGCCGTGCTGATGTAGC
Seq 67	20	60	TCGAGTCCAGGTTGTGGAGC
Seq 68	22	60	GATCGGTAAGCCGTTGACACCT
Seq 73	25	60	GTATCGTCAATCGCTGGAAGCAAAT
Seq 74	22	60	CGGTGCATCAGTCAGCCATTTC
Seq 81	20	61	CAGGTCACGCCGACTACGTG
Seq 82	24	61	GTCGCTCTGCCAATTGAGCTACTG
Seq 97	26	60	GTGTCTCAGAGGCGTACTCATAGATC
Seq 98	22	60	GAAGAAGGCGAAGTAGCGGTCA

Table Ap2.2: Primers derived from mapping sequencing data of *P. agarici* NCPPB 2472 to *P. agarici* NCPPB 2289, used for contig stitching.

Primer:	**Length:**	**Tm:**	**Sequence**
Seq 1	26	61	GGCTTTTTCCAACAACCGAATCAGAC
Seq 2	21	61	TCCACGCGCCTTAATGTCAGC
Seq 3	28	61	GCTAAAGGTCATTGCCGAACATTCCTA G
Seq 4	25	61	CAGCCTCTATGTATTGGACGACAGC
Seq 5	20	62	GATTGAGCCAGCGGCCATCG
Seq 6	21	61	CTGTCGATAGGTCGACAGCGC
Seq 7	27	60	GGATTCGGCGCATACTGATAAAGATTG
Seq 8	35	60	CTGACAATGCCTTTGTAGTTCTTAATCT CAAAATC
Seq 9	25	62	GCCGAAATTCAGGTGCATGGATTCG
Seq 10	25	61	GTCGGTGTGGTAGTAGCGGATTTTC
Seq 11	32	61	GCAAGAGCAATCTATTCAGATGTATGTA TGCG
Seq 12	22	62	CGCACTTCTTCCAGGGTGTTGC
Seq 13	26	60	GTGGAGTCAGGCTGAATAGTCACTAG
Seq 14	25	60	GTGTTTTCACACGGACGACAGGC
Seq 15	29	62	GCATATTTCAACCTCTCCAATTCCGATG C
Seq 16	23	62	GCTGGGTGCTACCCATGTTTGTC
Seq 17	25	60	CTCATGTAAGGCATAGGCTTGCAAC
Seq 18	28	60	CGTTAAAAGGAGGCGCAATTCTATAGA C
Seq 19	26	61	GAGGGTGAGAGCATGGAATTGTAAGG
Seq 20	23	61	CTTCCAGCCAGCCTTGTTGAGAG
Seq 21	22	61	GCGCGTTACTGAAGAGCGAGAC
Seq 22	26	62	CTGACGATAGCTTTGCTGATGCAAGC
Seq 23	23	62	GAAAGCAGACCGGTGGTTTCTCG
Seq 24	22	61	CGAAGCCGACATCAGTGTGGTC
Seq 25	24	62	GCATCGATCAACTGGCGAATCAGG
Seq 25	24	62	GCATCGATCAACTGGCGAATCAGG
Seq 26	25	61	CCGTTGCTCTTGCCTCTATAAGGTC
Seq 26	25	61	CCGTTGCTCTTGCCTCTATAAGGTC
Seq 26	25	61	CCGTTGCTCTTGCCTCTATAAGGTC

Seq 27	25	60	GCTCGAGATCTTCCGACATCAAGAG
Seq 28	27	61	GCTGTATTGCACATGCACAATGATGTC
Seq 29	20	62	GGGCTTTAGGCGGCACTCAG
Seq 30	22	62	CACGTCGATGGGTTTCAGGCAC
Seq 31	22	61	CACTTCCGACGCGGTCTAAACG
Seq 32	23	62	GCAGGCCTATGCCTTACATGAGC
Seq 33	23	61	CAATCTGCTCTCTGGCCTTCAGG
Seq 34	24	61	CTTGTTGCATACCCACCGAGTCTC
Seq 35	24	60	GTCTGCCAATGCCCAAATTCTACG
Seq 36	24	60	CTCTATCTCCGCGGCGTTATAGAC
Seq 37	23	61	GAGGTGAAGCAGTTCTGCGACTG
Seq 38	24	61	CGTCACACTCATCCACTCATGAGC
Seq 39	29	62	CTTGTTGTATACCCACCGAGTCTCTTCT G
Seq 40	22	62	CCACTGTCACCATCACCGATGC
Seq 41	28	60	CCAGCAAGTTGTAAGTTGTTGTTCTAC G
Seq 42	20	60	GCTGGAAGCCCGAATTGCTC
Seq 43	26	61	CGATGTGGTCTTCAGTTCATCGAAGC
Seq 44	23	61	GGCCGCTGTCGTCCAATACATAG
Seq 45	20	60	CGGGCAAAGAAGGCTGTAGC
Seq 46	21	61	CACGTCAGGTTCAGGCCATCG
Seq 47	24	62	GTAACTGGCGCTGCTGTAGAATGC
Seq 48	22	61	CAGGCACTGACCAACGAGAACG
Seq 49	23	61	GCTCACTGTCTACCTGCATGCTG
Seq 50	23	61	GGCCCAGGTCATTCAAGAGTGTG
Seq 52	25	61	GCCATGAATGCTTGATCAAACCCAG
Seq 53	18	61	CAACCCCGCAACGCCAAG
Seq 54	24	61	CAGTTCTTCAGGGCTGTTGCTCTG
Seq 54	24	61	CAGTTCTTCAGGGCTGTTGCTCTG
Seq 55	20	60	GCATAAGTGTGGCCGTCTGC
Seq 56	21	60	GCAGGATCGAACAGCGCAATG
Seq 57	22	60	GCAATTTCGCCTCTACCAGCAC
Seq 58	17	60	CGTGCTTGGCGCACGTC
Seq 59	25	61	TCAGCAGATCACACTCATGACAAGC
Seq 60	24	61	CCAGACCATTGTGATCTGGTCCTG
Seq 62	25	60	CCATCAGCAAAGCACTCGATTACAG
Seq 63	20	60	GTCTTGGCTTTGCGGGACTG
Seq 65	25	61	CGTCAGCATGTTCCGATAATCCGTG
Seq 66	20	62	GGTGGCCGTGCTGATGTAGC
Seq 67	20	60	TCGAGTCCAGGTTGTGGAGC
Seq 68	22	60	GATCGGTAAGCCGTTGACACCT
Seq 73	25	60	GTATCGTCAATCGCTGGAAGCAAAT
Seq 74	22	60	CGGTGCATCAGTCAGCCATTTC
Seq 77	22	61	GCAAGCTCTCTTCATCCGCTCG
Seq 78	20	60	GTGTGTAAGCGCTGTGAGGC
Seq 82	24	61	GTCGCTCTGCCAATTGAGCTACTG
Seq 83	21	60	AATGGCTTGCTCGAACCGTTG
Seq 83	21	60	AATGGCTTGCTCGAACCGTTG
Seq 83	21	60	AATGGCTTGCTCGAACCGTTG
Seq 84	19	60	CGTGCCTGATTCGATGCCG
Seq 84	19	60	CGTGCCTGATTCGATGCCG

Seq 86	23	61	GACTTGCCACGGCTGATTTACGT
Seq 87	20	60	TGCCATTTATGCGGCAGCAA
Seq 91	21	60	GGTGTAGTCACCTACCGTGCC
Seq 92	20	60	ACCGTGGATGTGCTCGATCA
Seq 93	22	60	TCGAACAGGAAGCGGACTTTGT
Seq 94	20	61	AACATGACTGGCCGCTGGTC
Seq 95	22	60	TGCCAGCCATGGAAGATCTGTA
Seq 95	22	60	TGCCAGCCATGGAAGATCTGTA
Seq 97	26	60	GTGTCTCAGAGGCGTACTCATAGATC
Seq 98	22	60	GAAGAAGGCGAAGTAGCGGTCA
Seq 99	24	60	CGGAGCAGACCAGTATGAACAGAG
Seq 100	23	60	GTCATGAAAGCGCCACCACTAAC
Seq 101	21	61	ACGGTTCGGCCAACATCAACA
Seq 102	20	61	GCTTATCACCGCGGCGTACA
Seq 103	21	61	GCTCCTTCAGCAGGTGGTTGT
Seq 104	21	61	AGCTATCTCAAAGCGCGAGGC
Seq 107	22	60	CGAATTGATCCAGCAGCAAGCC
Seq 108	20	61	GCTGACCGCTGTCGTCCAAT
Seq 113	22	60	GCTGGCGAACGACAGCAATAAC
Seq 114	20	60	TGGTGCAAACCACGTTCACC

Table Ap2.3: Bacteriophage ϕNV3 forward primers, used for primer walking.

Name:	**Start (bp) :**	**Length:**	**Tm:**	**Sequence:**
Nv3a-F1	1	19	62	TCCCGACCCCTAGCAGCTC
Nv3a-F2	1420	19	60	CAGCAAGGCAGCACAGGAC
Nv3a-F3	2788	19	59	CACGCAGCGACTAGATGGG
Nv3a-F4	4151	19	60	CGCTACATGCATCGCCAGC
Nv3a-F5	5488	20	59	CCCAAGCTCAAACCGCTACC
Nv3a-F6	6827	18	62	GCGGGTGCCTTCGGTACC
Nv3a-F7	8212	21	60	CCACAGGAGACTGAGGTGGAC
Nv3a-F8	9619	19	59	CGATGGCGGTGATGATCGC
Nv3a-F9	10958	19	61	GCTGGCTGACCGACTCAGC
Nv3a-F10	12,291	20	61	CTGGTGCCCACAGATCTGCC
Nv3a-F11	13,646	19	59	CTGGAACCAAGGTGAGGCC
Nv3a-F12	14,992	20	60	CTGCCGGCTCTGTAACATGC
Nv3a-F13	16,348	20	61	CCCGAGCTAAGGCCAACACC
Nv3a-F14	17,700	20	62	CCGTTCTGAAGGGCACCTGC
Nv3a-F15	19,053	21	59	CGAACGTTCCTGCATGAGTCC
Nv3a-F16	20,401	20	59	CGCTAGTGCGCAAGGTAGAC
Nv3a-F17	21,753	19	60	AGGAGTACTGGTGCTGGGC
Nv3a-F18	23,121	21	61	CATTAAGCGCTCGCAAGGTGC
Nv3a-F19	24,468	20	61	ATTGCGTACCTCGAAGCCGC
Nv3a-F20	25,817	20	61	CTCGTAAGGCGATCACCGGC
Nv3a-F21	27,185	21	60	CGCCAAAGACTTCACCGTGAC
Nv3a-F22	28,528	21	61	CCGGGCAGGTTATCCTGATGC
Nv3a-F23	29,875	20	62	CTAGCGGTATCGTGCAGCGC
Nv3a-F24	31,224	25	61	GACTTCCAGAAGTTCCAGAACAGCC
Nv3a-F25	32,561	22	61	CTTGCTGCAATCAAGCAGGAGC
Nv3a-F26	33,936	21	62	CAGCGCTGTCCAACGTACTGC

Nv3a-F27	35,278	20	61	GGCAACCATTATGCACCGGC
Nv3a-F28	36,618	21	60	CCTTGCAGGAGGGTCTACTCC
Nv3a-F29	37,984	20	61	GCCGGGCGTACGTGATATCC
Nv3a-F30	39,313	21	61	GGGTATACCTGTTCCTGCGGC
Nv3a-F31	40,659	20	61	CGGTAACGGTGGTGACGAGC
Nv3a-F32	42,017	21	61	GCTGGATTAAGCTCCTGGGGC
Nv3a-F33	42,678	19	62	CTAACGGCGCCATCACCGC

Table Ap2.4: Bacteriophage ϕNV3 reverse primers, used for primer walking.

Name:	**Start (bp):**	**Length:**	**Tm:**	**Location:**
Nv3a-R1	687	19	62	GCCAAGAGGTAAGGCGCGC
Nv3a-R2	2025	19	61	GTCCGCCGGATCAGTGTGT
Nv3a-R3	3301	19	60	CGCTTGGAAGTCGAACGCC
Nv3a-R4	4649	20	60	GGATTCAACGTGGTCGCAGG
Nv3a-R5	5965	19	61	CCTGCAATGCACCCGTTGC
Nv3a-R6	7310	20	61	CCGGCAACTCTGGCAGTACG
Nv3a-R7	8644	21	61	GCTCGTCTTACCCGCATCAGG
Nv3a-R8	9996	19	61	GGTCGAAGCGAAGGCGGAT
Nv3a-R9	11344	20	59	GTCCTCCTTGGTGACGTTGC
Nv3a-R10	12689	19	62	CTGCTGGAACTCAGCCCGC
Nv3a-R11	14041	19	60	CGGCAGCTTGGTTAGCTGC
Nv3a-R12	15384	22	61	CAGTAGTGGTGTAGCTCCACGC
Nv3a-R13	16746	20	62	CGTTATCGGTGCAGCCGGAC
Nv3a-R14	18103	19	59	GACCAGCGGAACAGAGTGC
Nv3a-R15	19442	20	62	CCAGTTCTTCTTCGGCGGCC
Nv3a-R16	20793	23	61	GGACTTGACATGGAGCTACTGCC
Nv3a-R17	22132	19	61	CTCACCTCCGGGCATACGC
Nv3a-R18	23490	20	60	GCGATTGGTTCACTGCGACG
Nv3a-R19	24842	19	62	AGTTTGAACCACGCCGCCG
Nv3a-R20	26217	19	62	CGGCTCGTCAAGGGAGGGT
Nv3a-R21	27571	20	61	CGTAGTGGGCAGACTCGCAG
Nv3a-R22	28928	21	61	TACTCGAACTCCCCGGTGTCC
Nv3a-R23	30268	20	62	AGCAAGGAAGGGCTTGGTGC
Nv3a-R24	31620	21	61	CGGGTACTTGTCTGCCAGTGC
Nv3a-R25	32966	20	61	CCGCCAACGAGAGCCATACC
Nv3a-R26	34302	20	61	CCAACCTATACCCCGCTGGC
Nv3a-R27	35655	21	61	CCCAGTCCACATCTTTGCCGT
Nv3a-R28	37010	20	60	CCGAACTCCTCCGGGCATAG
Nv3a-R29	38363	19	61	CAGGGGCGTAGTCGCTAGC
Nv3a-R30	39786	20	61	CCTTGTCACCGGAGCCAGAG
Nv3a-R31	41134	20	61	CCCTCAAGGGCGTCGATACG
Nv3a-R32	42448	20	62	CCTTGGCCAAGTTCAGGCCG
Nv3a-R33	43131	21	61	CACATAGAACGCGCACAGGGA

Appendix 3: Full Phage Genome Tables

Table Ap3.1 All identified ORFs of phage NV1 including their function as predicted by BlastP, including Accession codes of homologous protein in phage UFV-P2 (if applicable) and Query Cover, E-value and Identity scores as identified by BlastP.

Gp No:	Start:	End:	Strand:	Predicted Function:	Homolog (UFV-P2):	Cover:	Evalue:	Ident:
1	980	1321	+	Hypothetical Protein	YP_006907083.2	80%	5.00E-27	67%
2	1351	1836	+	Hypothetical Protein	YP_006907082.2	98%	4.00E-96	48%
3	1709	2077	+	Hypothetical Protein	YP_007518458.1	45%	2.00E-21	71%
4	2546	2776	+	Hypothetical Protein	YP_007518462.1	97%	5.00E-19	54%
5	3279	3479	+	Hypothetical Protein	YP_007518465.1	100%	1.00E-35	89%
6	3466	3795	+	Hypothetical Protein	YP_007518466.1	98%	9.00E-29	47%
7	3792	4031	+	Hypothetical Protein	N/A	N/A	N/A	N/A
8	4031	4243	+	Hypothetical Protein	YP_007518467.1	85%	7.00E-30	83%
9	4224	4628	+	Hypothetical Protein	YP_006907081.1	94%	3.00E-57	65%
10	4712	4903	+	Novel Hypothetical Protein	N/A	N/A	N/A	N/A
11	5060	5890	+	Putative Transposase	YP_006907080.2	100%	0.00E+00	99%
12	6062	6400	+	Hypothetical Protein	YP_007518468.1	75%	5.00E-30	67%
13	6541	6753	+	Hypothetical Protein	YP_007518470.1	100%	1.00E-31	76%
14	6877	7689	+	Hypothetical Protein	YP_007518471.1	99%	7.00E-80	48%
15	7693	8571	+	Hypothetical Protein	YP_006907079.1	100%	4.00E-109	55%
16	8583	9380	+	Hypothetical Protein	YP_006907078.1	100%	3.00E-175	89%
17	9377	9937	+	Hypothetical Protein	N/A	N/A	N/A	N/A
18	9913	11067	+	Putative Amidoligase	YP_006907076.1	100%	0.00E+00	78%
19	11098	12567	+	Putative Glutamine Amidotransferase	YP_006907074.1	99%	0.00E+00	81%
20	12570	12788	+	Hypothetical Protein	YP_007518472.1	100%	2.00E-42	94%
21	12926	13801	+	Putative ATP-grasp Enzyme	YP_006907073.1	100%	2.00E-159	76%
22	13794	14189	+	Hypothetical Protein	YP_006907072.2	99%	5.00E-78	84%
23	14189	14485	+	Hypothetical Protein	YP_007518474.1	100%	4.00E-56	87%
24	14445	16199	+	Putative DNA Primase/Helicase	YP_006907071.2	100%	0.00E+00	96%

25	16138	16689	+	Putative DNA Polymerase Part I	YP_006907070.2	93%	1.00E-109	88%
26	16646	16957	+	Hypothetical Protein	N/A	N/A	N/A	N/A
27	17121	17351	+	Hypothetical Protein	YP_007518477.1	100%	2.00E-34	75%
28	17364	17567	+	Hypothetical Protein	YP_007518478.1	100%	3.00E-31	84%
29	17588	17845	+	Putative Holin	YP_007518479.1	100%	5.00E-51	94%
30	17827	18060	+	Hypothetical Protein	YP_007518480.1	93%	8.00E-12	40%
31	18099	18527	+	Hypothetical Protein	N/A	N/A	N/A	N/A
32	18496	18717	+	Hypothetical Protein	N/A	N/A	N/A	N/A
33	18725	20365	+	Putative DNA Polymerase Family A	YP_006907069.2	100%	0.00E+00	95%
34	20432	21028	+	Hypothetical Protein	YP_006907068.1	99%	1.00E-116	86%
35	21006	21473	+	Hypothetical Protein	YP_006907067.2	74%	5.00E-66	83%
36	21473	21733	+	Hypothetical Protein	YP_007518485.1	88%	2.00E-19	71%
37	21751	22128	+	Hypothetical Protein	YP_007518486.1	69%	4.00E-18	47%
38	22125	23009	+	Putative 5'-3' Exonuclease	YP_006907065.2	100%	0.00E+00	91%
39	22984	23982	+	Hypothetical Protein	YP_006907064.1	100%	0.00E+00	74%
40	23888	24250	+	Putative Endonuclease	YP_006907063.2	100%	2.00E-68	87%
41	24216	24971	+	Hypothetical Protein	YP_006907062.1	100%	0.00E+00	97%
42	25190	25396	+	Hypothetical Protein	YP_007518490.1	100%	1.00E-41	97%
43	25525	25881	-	Hypothetical Protein	YP_006907061.1	100%	5.00E-60	75%
44	25881	26768	-	Putative Phage Structural Protein	YP_006907060.1	99%	0.00E+00	86%
45	26779	29952	-	Putative Phage Structural Protein	YP_006907059.1	100%	0.00E+00	90%
46	29959	31674	-	Putative Phage Structural Protein	YP_006907058.1	100%	0.00E+00	82%
47	31676	32080	-	Putative Phage Structural Protein	YP_006907057.1	100%	2.00E-84	97%
48	32080	32607	-	Putative Phage Particle Protein	YP_006907056.1	100%	2.00E-113	99%
49	32699	33022	-	Putative Phage Particle Protein	YP_006907056.1	90%	1.00E-41	85%
50	33003	33437	-	Hypothetical Protein	YP_006907055.2	100%	1.00E-90	87%
51	33434	34147	-	Hypothetical Protein	YP_006907054.2	100%	3.00E-129	87%
52	34147	35691	-	Putative Phage Particle Protein	YP_006907053.1	100%	0.00E+00	83%
53	35699	36334	-	Putative Tail Fiber Protein	YP_006907052.1	100%	1.00E-134	88%
54	36324	36575	-	Hypothetical Protein	N/A	N/A	N/A	N/A
55	36557	36751	-	Hypothetical Protein	YP_006907051.2	100%	3.00E-32	91%

56	36755	37384	-	Putative Phage Structural Protein	YP_006907050.1	100%	3.00E-145	96%
57	37388	37570	-	Hypothetical Protein	YP_006907049.2	100%	3.00E-35	93%
58	37759	38712	-	Putative Major Capsid Protein	YP_006907048.1	100%	0.00E+00	96%
59	38730	39728	-	Putative Scaffolding Protein	YP_006907047.1	99%	1.00E-168	75%
60	39718	39969	-	Hypothetical Protein	YP_007641362.1	100%	5.00E-48	88%
61	39969	42107	-	Putative Portal Protein	YP_006907046.2	97%	0.00E+00	95%
62	42086	43531	-	Putatitve Terminase Large Subunit	YP_006907045.1	100%	0.00E+00	91%
63	43535	44074	-	Putative Lysozyme	YP_006907044.2	93%	1.00E-111	93%
64	43962	44435	-	Putative Terminase Small Subunit	YP_006907043.1	100%	4.00E-92	83%

Table Ap3.2 All identified ORFs of phage ϕNV3 including their function as predicted by BlastP, including closest amino acid homolog by species and Query Cover, E-value and Identity scores as identified by BlastP.

Gp No:	Start:	End:	Predicted Function	Closest Similarity:	Cover:	E value:	Ident:
1	2302	2688	Hypothetical protein	Pseudomonas phage VSW-3	81%	2.00E-35	54%
2	2900	3427	Hypothetical protein	Pseudomonas phage phikF77	77%	3.00E-40	49%
3	3424	3639	Novel Hypothetical protein	N/A	N/A	N/A	N/A
4	3867	4079	Hypothetical protein	Pseudomonas phage PAK_P5	98%	8.00E-08	41%
5	4140	4499	Novel Hypothetical protein	N/A	N/A	N/A	N/A
6	4489	4737	Novel Hypothetical protein	N/A	N/A	N/A	N/A
7	4730	4918	Novel Hypothetical protein	N/A	N/A	N/A	N/A
8	4915	5172	Novel Hypothetical protein	N/A	N/A	N/A	N/A
9	5218	5406	Hypothetical protein	Pseudomonas phage phi-2	56%	7.00E-04	57%
10	5436	5891	Hypothetical protein	Pseudomonas phage phikF77	70%	3.00E-18	47%
11	6014	6457	Novel Hypothetical protein	N/A	N/A	N/A	N/A
12	6641	7135	Hypothetical protein	Pseudomonas phage UNO-SLW4	91%	8.00E-25	37%
13	7159	7995	Putative DNA primase	Pseudomonas phage LUZ19	93%	8.00E-81	51%
14	8047	9357	Putative DNA helicase	Pseudomonas phage vB_Pae-TbilisiM32	97%	1.00E-160	55%
15	9350	9913	Hypothetical protein	Pseudomonas phage vB_PaeP_PPA-ABTNL	88%	3.00E-10	29%
16	9913	10812	Putative ATP-dependent DNA ligase	Ralstonia phage RSB3	100%	1.00E-70	44%
17	10897	11223	Novel Hypothetical protein	N/A	N/A	N/A	N/A
18	11322	11518	Novel Hypothetical protein	N/A	N/A	N/A	N/A
19	11583	13943	Putative DNA polymerase I	Pseudomonas phage phiNFS	99%	0	66%
20	13990	15006	Hypothetical protein	Pseudomonas phage phi-2	92%	1.00E-93	51%
21	15008	15967	Putative Integrase	Pseudomonas phage phi-2	92%	5.00E-113	58%
22	15948	16424	Putative DNA Endonuclease VII	Pseudomonas phage phiKMV	69%	1.00E-47	66%
23	16432	17652	Putative DNA Exonuclease	Pseudomonas phage phi-2	81%	3.00E-179	74%

24	17663	18091	Hypothetical protein	Pseudomonas phage Andromeda	86%	1.00E-65	81%
25	18088	18270	Novel Hypothetical protein	N/A	N/A	N/A	N/A
26	18414	20870	Putative RNA Polymerase	Pseudomonas phage vB_PaeP_PPA-ABTNL	99%	0.00E+00	51%
27	21028	21271	Hypothetical protein	Pseudomonas phage phikF77	100%	8.00E-15	48%
28	21270	21701	Hypothetical protein	Pseudomonas phage LKD16	94%	5.00E-26	41%
29	21679	22149	Putative DNA Endonuclease	Pseudomonas phage phi-2	98%	2.00E-34	44%
30	22363	22500	Putative virion structural protein	Pseudomonas phage phikF77	97%	7.00E-12	66%
31	22510	24048	Putative head-tail connector protein	Pseudomonas phage vB_Pae-TbilisiM32	99%	0.00E+00	65%
32	24048	25013	Putative scaffolding protein	Pseudomonas phage LKD16	100%	6.00E-49	42%
33	25069	26061	Putative capsid protein	Pseudomonas phage phikF77	100%	0.00E+00	74%
34	26161	26715	Putative tail tubular protein A	Pseudomonas phage vB_PaeP_PPA-ABTNL	100%	2.00E-57	49%
35	26712	29183	Putative tail tubular protein B	Pseudomonas phage LKD16	99%	0.00E+00	47%
36	29183	29728	Putative internal virion protein A	Pseudomonas phage LKD16	98%	2.00E-49	49%
37	29721	32381	Putative baseplate hub subunit and tail lysozyme	Pseudomonas phage phiKMV	99%	0.00E+00	49%
38	32390	36322	Putative internal core protein	Pseudomonas phage phiKMV	100%	0.00E+00	59%
39	36323	38689	Putative T7-like tail protein	Pseudomonas phage MPK7	17%	3.00E-19	43%
40	38640	38936	Novel Hypothetical protein	N/A	N/A	N/A	N/A
41	39005	39184	Novel Hypothetical protein	N/A	N/A	N/A	N/A
42	39215	39532	Putative DNA Maturase A	Pseudomonas phage LKA1	89%	3.00E-25	54%
43	39529	41301	Putative DNA Maturase B	Pseudomonas phage phiNFS	98%	0.00E+00	79%
44	41303	41542	Putative pinholin	Pseudomonas phage phiKMV	81%	2.00E-19	63%
45	41526	42026	Putative lysozyme	Pseudomonas phage MPK7	83%	5.00E-72	75%
46	41975	42307	Putative Rz-like protein	Pseudomonas phage LUZ19	96%	2.00E-14	42%
47	42165	42410	Putative Rzl-like protein	Pseudomonas phage LKA1	46%	2.00E-06	55%
48	42388	42715	Hypothetical protein	Pseudomonas phage phi-2	51%	3.40E+00	46%
49	42729	42911	Novel Hypothetical protein	N/A	N/A	N/A	N/A

Appendix 4: Supplementary Experimental Data

	baseMean	log2Fold Change	lfcSE	stat	pvalue	padj	P4	P5	P6	PN4	PN5	PN6
RS24005	39706.802	-6.447	0.345	-18.688	6.21E-78	3.04E-74	950.17	661.59	1087.39	44007.70	102839.44	88694.52
ssrA	1334955.215	-1.640	0.189	-8.692	3.55E-18	8.67E-15	598588.72	657415.73	690028.48	1703937.31	2098988.72	2260772.33
RS16105	205.223	-1.198	0.206	-5.813	6.13E-09	9.99E-06	122.00	118.48	131.25	248.05	308.46	303.10
RS04440	1485.348	-0.801	0.153	-5.231	1.69E-07	2.06E-04	1068.54	1122.80	1058.02	1788.77	1921.96	1952.01
RS16100	840.808	-1.236	0.242	-5.110	3.23E-07	3.16E-04	522.35	454.84	523.56	853.63	1346.56	1343.91
RS23260	21763.386	-0.875	0.193	-4.539	5.64E-06	4.60E-03	14391.96	17938.53	13756.57	25676.30	28013.68	30803.27
RS24275	4935.152	-1.156	0.257	-4.491	7.09E-06	4.95E-03	3008.88	3014.54	3145.14	4575.57	7850.95	8015.83
RS24525	19086.990	-1.213	0.279	-4.352	1.35E-05	8.23E-03	11756.36	10626.03	12136.14	16890.83	30608.92	32503.66
RS13830	30.191	-1.514	0.354	-4.273	1.93E-05	9.45E-03	14.54	15.90	17.06	41.92	32.63	59.09
RS17680	1552.339	-0.787	0.183	-4.293	1.77E-05	9.45E-03	1146.91	1118.03	1155.15	2290.70	1815.18	1788.07
RS04410	416.473	-0.836	0.203	-4.121	3.78E-05	1.66E-02	282.39	306.15	310.82	621.88	522.01	455.60
RS18970	1721.115	-0.694	0.169	-4.103	4.07E-05	1.66E-02	1260.43	1276.27	1404.84	1920.37	2289.74	2175.04
RS06350	165.249	-0.953	0.233	-4.083	4.45E-05	1.67E-02	108.67	104.17	124.61	193.32	195.75	264.97
RS16090	84.998	-1.155	0.285	-4.048	5.17E-05	1.73E-02	49.69	47.71	57.80	90.84	157.20	106.75
RS23105	206.453	-0.938	0.233	-4.030	5.57E-05	1.73E-02	137.76	131.21	153.99	217.77	275.84	322.16
RS07595	10603.102	0.990	0.246	4.027	5.65E-05	1.73E-02	14079.68	13790.06	14441.70	8113.52	8260.26	4933.40
RS21855	407.818	-0.762	0.192	-3.965	7.34E-05	1.99E-02	263.40	325.23	320.29	499.60	489.39	549.00
hflB	24881.951	0.773	0.195	3.971	7.17E-05	1.99E-02	29479.17	30576.37	34130.36	19544.87	20385.18	15175.74
RS16095	97.522	-0.998	0.257	-3.880	1.05E-04	2.44E-02	66.66	57.25	70.12	108.30	124.57	158.22
RS22880	836.743	0.753	0.193	3.900	9.63E-05	2.44E-02	962.70	1084.63	1098.76	567.14	729.63	577.60
rnpB	1263911.359	-1.846	0.474	-3.891	1.00E-04	2.44E-02	482360.27	539283.38	629076.74	802993.51	3756891.67	1372862.58
RS14345	28914.912	0.746	0.193	3.868	1.10E-04	2.44E-02	37033.69	35131.17	36517.88	23251.68	23988.85	17566.20
RS15670	95.934	-0.968	0.253	-3.832	1.27E-04	2.70E-02	65.85	69.18	59.23	133.92	148.30	99.13
RS18775	74.087	-1.211	0.329	-3.676	2.37E-04	4.63E-02	50.09	42.14	39.33	64.05	136.44	112.47
RS12730	447.489	-0.641	0.174	-3.684	2.30E-04	4.63E-02	326.82	375.33	346.83	540.36	569.47	526.13
RS22885	3586.449	0.623	0.173	3.612	3.04E-04	5.72E-02	4307.69	4443.48	4290.34	2474.70	3146.91	2855.58
RS13280	3684.993	0.933	0.261	3.583	3.40E-04	6.15E-02	4888.22	4929.34	4694.03	3010.40	2906.66	1681.32
RS07600	1026.488	0.890	0.250	3.556	3.76E-04	6.57E-02	1340.42	1318.41	1344.67	926.99	711.84	516.60
RS19360	663.999	0.635	0.180	3.532	4.12E-04	6.95E-02	765.15	799.16	861.38	569.47	516.08	472.75
RS12095	50.730	-1.059	0.301	-3.513	4.44E-04	7.23E-02	31.51	30.22	38.38	85.01	56.35	62.91
RS18635	227.909	-0.705	0.201	-3.504	4.59E-04	7.23E-02	177.35	174.94	170.10	302.79	237.28	305.00
RS02040	10435.994	0.669	0.191	3.493	4.78E-04	7.30E-02	13663.17	11026.80	13744.73	8260.25	8829.72	7091.29
RS13245	18062.289	0.813	0.233	3.483	4.96E-04	7.35E-02	25366.61	19598.87	24086.08	13292.33	15829.44	10200.40
RS10945	218.514	-0.862	0.249	-3.462	5.36E-04	7.70E-02	153.11	191.64	120.82	279.50	314.39	251.63

RS12065	296.642	-0.610	0.177	-3.443	5.76E-04	7.82E-02	238.35	229.81	235.96	365.67	370.75	339.31
RS03480	1493.979	0.688	0.199	3.450	5.60E-04	7.82E-02	2066.79	1646.82	1816.58	1262.39	1195.29	976.01
RS22640	192.155	-0.655	0.194	-3.383	7.16E-04	8.97E-02	140.99	149.49	156.83	231.75	243.21	230.66
RS22890	5980.916	0.536	0.158	3.386	7.09E-04	8.97E-02	6707.36	7671.13	6857.91	4950.56	5077.76	4620.78
RS14695	10384.225	-1.278	0.377	-3.391	6.97E-04	8.97E-02	5266.75	7520.84	5403.79	11027.25	8957.26	24129.45
RS20420	197.489	0.900	0.267	3.376	7.36E-04	8.99E-02	279.15	302.96	188.10	128.10	151.26	135.34

Table Ap4.1. Complete RNA sequencing results of the differentially expressed genes as calculated by DESeq2 with a padj of <0.1.

Table Ap4.2 Full results of the Jspecies AniB results for analysis of the *Pseudomonas* sp. NS1(2017) genome to closely related strains.

	P. sp NS1	**P. azotoformans S4**	**P. azotoformans F77**	**P. fluorescens PICF7**	**P. fluorescens UK4**	**P. fluorescens Pf0-1**	**P. fluorescens F113**	**P. agarici 2472**	**P. fluorescens LMG 5329**
P. sp NS1	*	93.32 [82.22]	89.72 [77.57]	90.72 [76.13]	83.25 [60.78]	79.91 [58.72]	79.60 [54.86]	78.19 [46.79]	93.43 [81.97]
P. azotoformans S4	93.06 [81.21]	*	89.65 [75.97]	90.47 [73.73]	83.57 [61.20]	79.69 [59.23]	79.63 [54.00]	78.05 [47.28]	95.56 [84.83]
P. azotoformans F77	89.94 [78.95]	90.02 [78.90]	*	89.61 [77.35]	83.43 [61.92]	80.09 [59.55]	79.80 [56.25]	78.47 [47.21]	90.05 [78.22]
P. fluorescens PICF7	91.22 [82.13]	91.08 [81.07]	89.86 [82.11]	*	83.41 [65.95]	79.99 [62.37]	80.04 [60.05]	78.66 [49.01]	91.08 [80.02]
P. fluorescens UK4	83.67 [66.72]	84.07 [67.42]	83.86 [66.70]	83.58 [66.42]	*	80.10 [59.86]	80.36 [56.34]	78.99 [48.23]	83.74 [66.43]
P. fluorescens Pf0-1	80.16 [61.13]	80.14 [62.15]	80.24 [60.29]	79.89 [59.64]	79.95 [56.98]	*	81.53 [62.59]	78.84 [49.16]	80.19 [61.37]
P. fluorescens F113	79.73 [53.93]	79.96 [53.55]	79.77 [54.37]	79.61 [54.51]	79.96 [50.76]	81.51 [58.31]	*	78.97 [44.88]	79.95 [53.21]
P. agarici 2472	78.92 [54.41]	78.89 [55.73]	79.09 [54.05]	78.94 [52.97]	79.29 [51.23]	79.28 [55.03]	79.55 [53.93]	*	78.97 [55.54]
P. fluorescens LMG 5329	93.28 [81.20]	95.59 [85.14]	89.68 [75.89]	90.48 [73.27]	83.11 [59.49]	79.79 [58.12]	79.62 [53.70]	78.15 [47.41]	*

Table Ap4.3. Full results of the Jspecies AniM results for analysis of the *Pseudomonas* sp. NS1(2017) genome to closely related strains.

	P. sp NS1	**P. azotoformans S4**	**P. azotoformans F77**	**P. fluorescens PICF7**	**P. fluorescens UK4**	**P. fluorescens Pf0-1**	**P. fluorescens F113**	**P. agarici 2472**	**P. fluorescens LMG 5329**
P. sp NS1	*	0.99858	0.99663	0.99819	0.97568	0.94809	0.96877	0.95169	0.99892
P. azotoformans S4	0.99858	*	0.99625	0.99737	0.9718	0.94495	0.96403	0.94736	0.99941
P. azotoformans F77	0.99663	0.99625	*	0.99739	0.97997	0.93905	0.97056	0.95052	0.9975
P. fluorescens PICF7	0.99819	0.99737	0.99739	*	0.97987	0.94596	0.97197	0.95542	0.9981
P. fluorescens UK4	0.97568	0.9718	0.97997	0.97987	*	0.93581	0.98022	0.97202	0.97537
P. fluorescens Pf0-1	0.94809	0.94495	0.93905	0.94596	0.93581	*	0.94687	0.93481	0.94523
P. fluorescens F113	0.96877	0.96403	0.97056	0.97197	0.98022	0.94687	*	0.97282	0.96802
P. agarici 2472	0.95169	0.94736	0.95052	0.95542	0.97202	0.93481	0.97282	*	0.951
P. fluorescens LMG 5329	0.99892	0.99941	0.9975	0.9981	0.97537	0.94523	0.96802	0.951	*

Table Ap4.4: Full results of Phage/Host Co-Evolution Study of ϕNV3 and *P. agarici* NCPPB 2472.

Bacteria Transfer No.	Transfer 1			Transfer 3			Transfer 5		
Phage Transfer No.	T0	T1	T2	T2	T3	T4	T4	T5	T6
Repeat 1 (% resistant)	12.50	68.75	37.50	100.00	56.25	50.00	100.00	68.75	100.00
Repeat 2 (% resistant)	6.25	62.50	37.50	43.75	37.50	31.25	75.00	75.00	50.00
Repeat 3 (% resistant)	6.25	25.00	18.75	68.75	62.50	37.50	100.00	100.00	43.75
Repeat 4 (% resistant)	12.50	56.25	25.00	37.50	37.50	31.25	87.50	50.00	37.50
Repeat 5 (% resistant)	18.75	43.75	12.50	50.00	100.00	68.75	100.00	100.00	75.00
Mean (%)	11	51	26	60	59	44	93	79	61

Bacteria Transfer No.	Transfer 7			Transfer 9		
Phage Transfer No.	T6	T7	T8	T8	T9	T10
Repeat 1 (% resistant)	87.50	87.50	100.00	100.00	100.00	100
Repeat 2 (% resistant)	100.00	43.75	100.00	100.00	100.00	100
Repeat 3 (% resistant)	81.25	100.00	62.50	100.00	100.00	100
Repeat 4 (% resistant)	50.00	43.75	56.25	100.00	100.00	100
Repeat 5 (% resistant)	25.00	100.00	100.00	100.00	100.00	100
Mean (%)	69	75	84	100	100	100

Table Ap4.5: Complete table of results of the putative achromobactin biosynthesis cluster of *P. agarici* NCPPB 2472.

***P. agarici* Locus Tag:**	**Predicted Function:**	**Length (bp):**	**Protein ID:**	***P. syringae* Locus Tag:**	**Length (bp):**	**Protein ID:**	**Cover (%):**	**Ident (%):**
RS04660	RNA polymerase sigma factor	509	WP_026013089.1	Psyr_2580	509	YP_235657.1	99	87
RS04665	sugar ABC transporter substrate-binding protein	962	WP_060783937.1	Psyr_2581	965	YP_235658.1	99	73
RS04670	TonB-dependent siderophore receptor	2402	WP_060782254.1	Psyr_2582	2408	YP_235659.1	100	81
RS04675	aspartate aminotransferase family protein	1376	WP_017131159.1	Psyr_2583	1376	YP_235660.1	100	88
RS04680	AcsD protein	1784	WP_060782255.1	Psyr_2584	1784	YP_235661.1	100	81
RS04685	diaminopimelate decarboxylase	1223	WP_060782256.1	Psyr_2585	1208	YP_235662.1	98	80
RS04690	MFS transporter	1394	WP_060783938.1	Psyr_2586	1406	YP_235663.1	100	86
RS04695	AcsC protein	1865	WP_060782257.1	Psyr_2587	1859	YP_235664.1	96	82
RS04700	siderophore biosynthesis protein SbnG	773	WP_060782258.1	Psyr_2588	776	YP_235665.1	100	86
RS04705	AcsA protein	1868	WP_060782259.1	Psyr_2589	1883	YP_235666.1	99	83
RS04710	iron ABC transporter substrate-binding protein	896	WP_017131166.1	Psyr_2590	905	YP_235667.1	98	76
RS04715	siderophore ABC transporter permease	989	WP_017131167.1	Psyr_2591	989	YP_235668.1	100	80
RS04720	iron ABC transporter permease	1031	WP_060782260.1	Psyr_2592	1031	YP_235669.1	100	85
RS04725	iron-enterobactin transporter ATP-binding protein	791	WP_017131169.1	Psyr_2593	791	YP_235670.1	100	91
RS04730	siderophore-iron reductase, Fe-S cluster protein	779	WP_060782261.1	Psyr_2594	779	YP_235671.1	100	75
RS04735	S-adenosylmethionine--2-demethylmenaquinone methyltransferase	632	WP_060782262.1	Psyr_2595	644	YP_235672.1	96	72

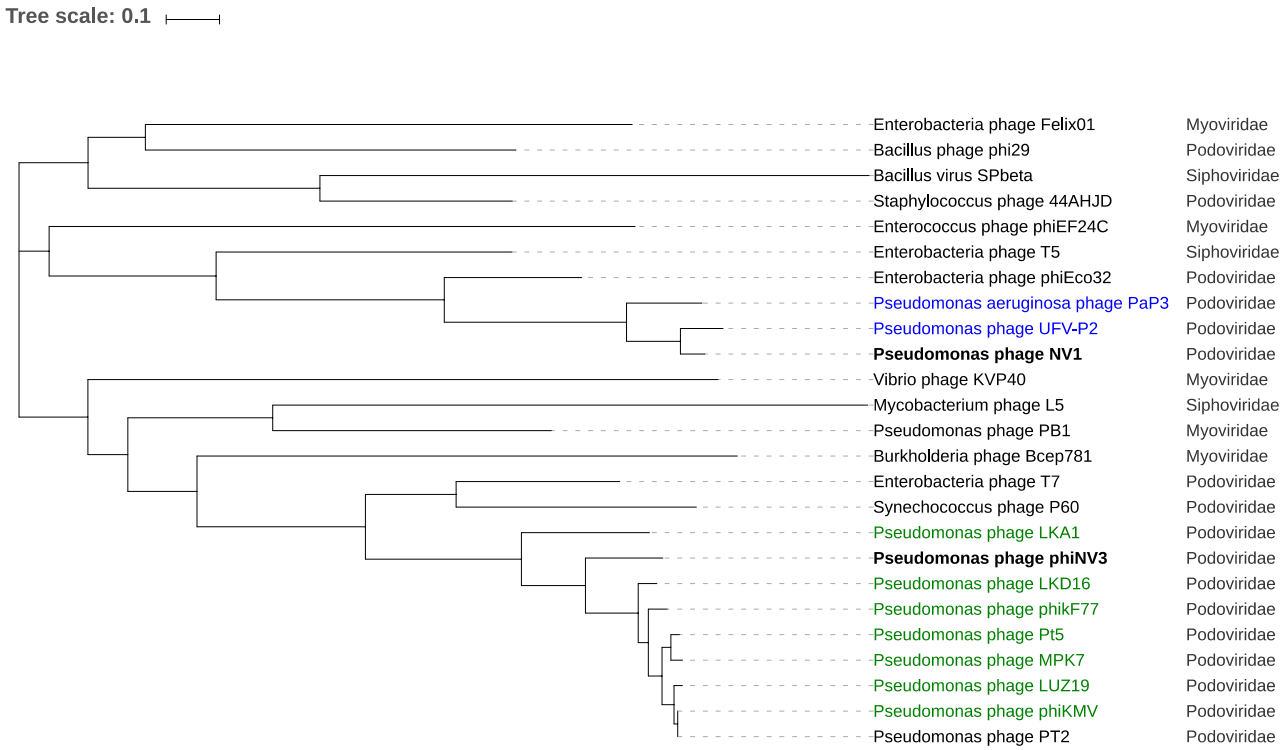

Figure Ap4.6: Molecular phylogenetic analysis of bacteriophage large terminase subunit genes by Maximum Likelihood. The evolutionary history was inferred by using the maximum likelihood method based on the Tamura-Nei model. The tree with the highest log likelihood (-19125.23) is shown (1000 bootstrap replicates) [257]. Initial tree for the heuristic search was obtained automatically by applying the Maximum Parsimony method [258]. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 25 nucleotide sequences. Codon positions included were $1^{st}+2^{nd}+3^{rd}$+Noncoding. All positions containing gaps and missing data were eliminated. There were a total of 804 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 [256] and tree edited using iTOL v4 [316].

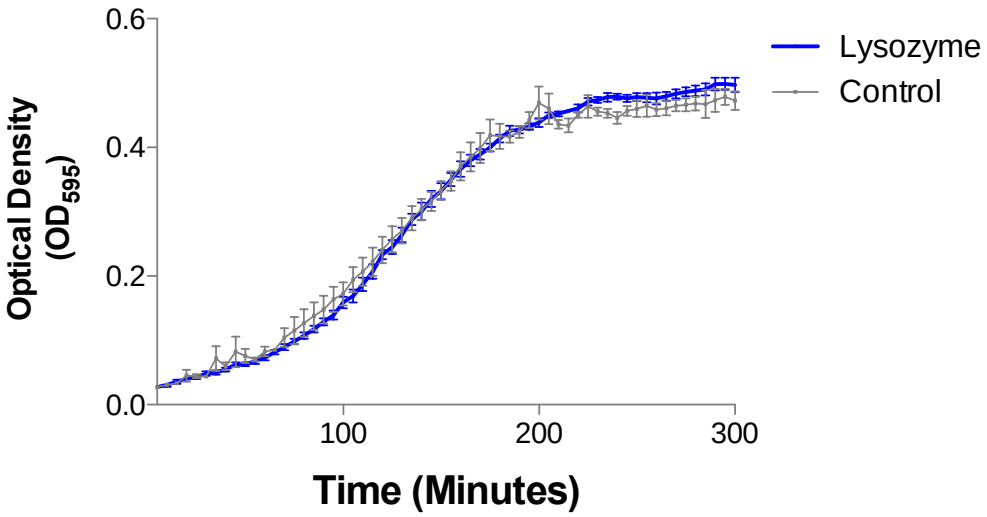

Figure Ap4.7: Effect of expression of a lysis protein construct containing the NV1 putative lysozyme encoding gene on optical density of BL21-AI™ *E. coli* following induction with 0.2% L-arabinose at T=0. The control comprised of an uninduced BL21-AI™ containing the lyozyme construct. Mean values of 6 replicates are plotted, error bars= +/- SEM.

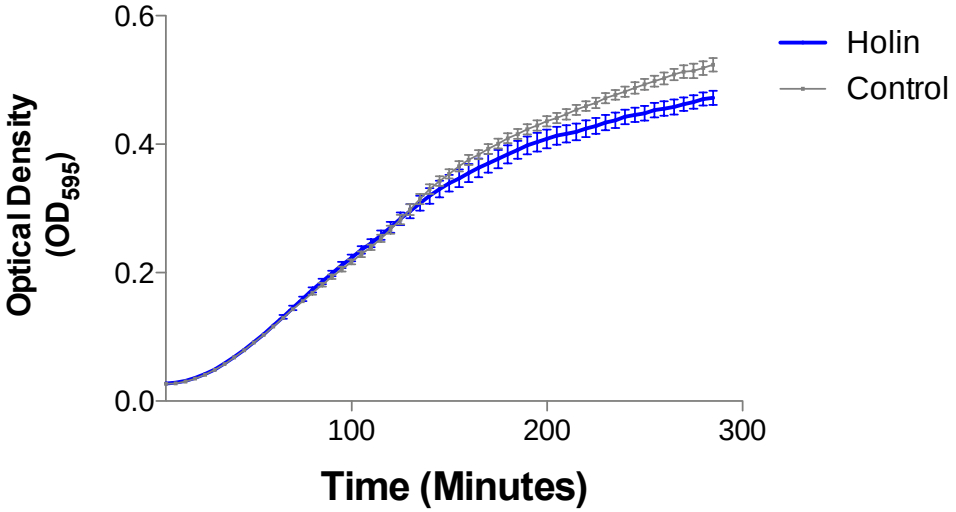

Figure Ap4.8: Effect of expression of a lysis protein construct containing the NV1 putative holin encoding gene on optical density of BL21-AI™ *E. coli* following induction with 0.2% L-arabinose at T=0. The control comprised of an uninduced BL21-AI™ containing the holin construct. Mean values of 6 replicates are plotted, error bars= +/- SEM.