University of Alberta

GENOME SEQUENCE AND CHARACTERIZATION OF *BURKHOLDERIA CEPACIA* COMPLEX BACTERIOPHAGE, KS10

by

Amanda Dawn Goudie 

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of Master of Science

in

Microbiology and Biotechnology

Department of Biological Sciences

Edmonton, Alberta
Fall 2008

Your file Votre référence
ISBN: 978-0-494-47254-5
Our file Notre référence
ISBN: 978-0-494-47254-5

ABSTRACT

The *Burkholderia cepacia* complex is a group of Gram negative organisms that can be found throughout the environment in sources such as soil, water, and plants, making these ideal locations to isolate bacteriophages specific to the complex. While *B. cepacia* complex bacteria can be involved in beneficial interactions with plants, they are also considered opportunistic pathogens, specifically in patients with cystic fibrosis and chronic granulomatous disease. The objective of this study was to sequence and characterize phage KS10, a previously isolated lysogenic bacteriophage of *B. cenocepacia* K56-2. Genome sequence analysis and annotation of this phage revealed that KS10 shows the closest sequence homology to Mu and BcepMu. Assembly of the KS10 genome shows that the organization of this genome differs from that of most phages in that the capsid module is interrupted. KS10 is found to be a prophage in three different strains of *B. cenocepacia*, including strains K56-2, J2315, and C5424, and seven tested clinical isolates of *B. cenocepacia*, but no other *Burkholderia cepacia* complex (BCC) species. A survey of 23 strains and 20 clinical isolates of the BCC revealed that KS10 is able to form plaques on lawns of *B. ambifaria* LMG 19467, *B. cenocepacia* PC184, and *B. stabilis* LMG 18870. There were no potential virulence factors identified in KS10.

TABLE OF CONTENTS

Page

Chapter 1: Introduction 1

1.1 The Burkholderia cepacia complex 2

1.2 Bacteriophages 6
1.2.1 Tailed bacteriophages 8
1.2.2 Mechanisms of bacteriophage bacteriolysis or lysogeny 12
1.2.3 Transposable bacteriophages 16
1.2.4 Lysogenic conversion 23
1.2.5 Restriction modification systems 24

1.3 Phage therapy 26
1.3.1 Infection models with the *B. cepacia* complex 28
1.3.2 Uses for phages 30

1.4 *Burkholderia* phages 32
1.4.1 *B. cepacia* complex specific phages 33

1.5 Research Objectives 38

Chapter 2: Materials and Methods 40

2.1 Bacterial strains, phages, plasmids, and growth conditions 41
2.1.1 *B. cepacia* complex and *E. coli* strains and plasmids 41
2.1.2 Growth conditions 41
2.1.3 Preparations of frozen bacterial stocks 44

2.2 Isolations of phages and host range testing 44
2.2.1 Isolation of phages from environmental sources 44
2.2.2 Preparation of high titer phage stock 45
2.2.3 Determination of host range 45
2.2.4 KS10 production, host range, and distribution 45
2.2.5 Isolation of PC184 and LMG 19467 lysogens 46

2.3 DNA manipulation 46
2.3.1 Phage DNA isolation 46
2.3.2 Isolation of chromosomal DNA 48

2.3.3 Gel electrophoresis of phage DNA 48
2.3.4 Purification of DNA fragments from agarose gel 48
2.3.5 Library preparation 49
2.3.6 Identification of KS10 integration sites 49

2.4 Sequence assembly and annotation 50
2.4.1 DNA sequence analysis 50
2.4.2 Polymerase chain reaction 50
2.4.3 Computer-assisted sequence analysis 51

2.5 Transmission electron microscopy (TEM) 52

2.6 Virulence of *B. ambifaria* LMG 19467 lysogens 52

Chapter 3: Results 54

3.1 Identification of four B. cepacia complex specific phages 55

3.2 Characterization of phage KS10 56

3.3 Host range of KS10 58

3.4 Determination of KS10 genome sequence and annotation 59

3.5 Organization of the KS10 genome 68

3.6 KS10 genome analysis 88

3.7 Integration sites of KS10 90

3.8 Lysogenic conversion of *B. ambifaria* LMG 19467 lysogens 91

Chapter 4: Discussion 94

4.1 Architecture of the KS10 genome 96

4.2 Head and tail assembly 101

4.3 Host cell lysis and transposition 105

4.4 Conclusions 108

Chapter 5: Literature Cited **112**

Chapter 6: Appendix: Construction of an integrative plasmid **130**

6.1 Introduction **131**

6.2 Materials and Methods **132**
6.2.1 Bacterial strains and plasmids 132
6.2.2 Growth conditions 132
6.2.3 Isolation of DNA 133
6.2.4 Polymerase chain reactions 133
6.2.5 Ligation of PCR products and transformations 133
6.2.6 SOEing PCR 134

6.3 Results and Discussion **140**
6.3.1 Construction of the integrative vector 140
6.3.2 Construction of the integrative vector using ligation assisted by restriction enzymes 141
6.3.3 Construction using SOEing PCR 143

LIST OF TABLES

Page

Chapter 2: Materials and Methods

Table 2.1 Bacterial strains and isolates used in this study 42

Chapter 3: Results

Table 3.1 KS10 putative genes and homologues 70
Table 3.2 Additional properties of KS10 proteins gp1-gp49 73
Table 3.3 Putative KS10 transposase binding sites 92
Table 3.4 *B. cepacia* complex strains/isolates testing positive for a KS10 prophage 93

Chapter 6: Appendix

Table 6.1 Plasmids, chromosomes, and plasposons 135
Table 6.2 Primers used for amplification of relevant genes/fragments 136

LIST OF FIGURES

Page

Chapter 1: Introduction

Figure 1.1 Bacteriophage life cycle 13
Figure 1.2 Transposable phage Mu life cycle 20
Figure 1.3 Mu lytic/lysogenic operators & transposase binding sites 21

Chapter 3: Results

Figure 3.1 RFLP comparison of phages KS10, MG1, MG3, and MG4 57
Figure 3.2 Electron micrograph of KS10 60
Figure 3.3 PCR showing KS10 prophage in *B. cenocepacia* strains and isolates 61
Figure 3.4 RFLP analysis of the KS10 genome 63
Figure 3.5 KS10/B. cenocepacia J2315 DNA junction 64
Figure 3.6 Genome map of KS10 and related phages BcepMu and Mu 86
Figure 3.7 Comparing genome organization of KS10 to Mu and BcepMu 87

Chapter 6: Appendix

Figure 6.1 Vector construction using a three-way ligation 138
Figure 6.2 Vector construction using SOEing PCR 139

LIST OF ABBREVIATIONS

α	Alpha
β	Beta
λ	Lambda
ϕ	Phi
ABC	ATP-binding cassette
ACT	Artemis comparison tool
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
BLAST	Basic local alignment search tool
BLASTX	Variation of BLAST for comparing a translated nucleotide query to a protein database
BLASTP	Variation of BLAST for comparing a protein query to a protein database
bp	Base pair
CGD	Chronic granulomatous disease
CF	Cystic fibrosis
CFTR	Cystic fibrosis transmembrane regulator
cfu	Colony forming unit
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate

EM	Electron micrograph
gp	Gene product
HIV	Human immunodeficiency virus
IAS	Internal activation sequence
IHF	Integration host factor
kb	Kilobase
MOI	Multiplicity of infection
NCBI	National Center for Biotechnology Information
OD	Optical density
ORF	Open reading frame
PCR	Polymerase chain reaction
pfu	Plaque forming unit
pN	picoNewton
RBS	Ribosome binding site
RFLP	Restriction fragment length polymorphism
ROR	Reactive oxygen radical

TEM Transmission electron microscopy

Chapter 1: Introduction

1. Introduction

1.1 The Burkholderia cepacia complex

The *Burkholderia cepacia* complex is a group of at least nine closely related species of Gram-negative, motile, bacilli that were first described in 1950 by W.H. Burkholder as an onion pathogen, causing soft rot (Burkholder, 1950). *B. cepacia* was previously known as *Pseudomonas cepacia*, *Pseudomonas multivorans*, and *Pseudomonoas kingii* (Govan *et al.*, 1996), but was later classified into nine different genomovars, collectively known as the *Burkholderia cepacia* complex. These genomovars are now classified as individual species as they are phenotypically similar, but genetically distinct: *B. cepacia* (formerly genomovar I) (Vandamme *et al.*, 1997), *B. multivorans* (formerly genomovar II) (Vandamme *et al.*, 1997), *B. cenocepacia* (formerly genomovar III) (Vandamme *et al.*, 1997; Vandamme *et al.*, 2000), *B. stablilis* (formerly genomovar IV) (Vandamme *et al.*, 1997; Vandamme *et al.*, 2000), *B. vietnamiensis* (formerly genomovar V) (Vandamme *et al.*, 1997), *B. dolosa* (formerly genomovar VI) (Coenye *et al.*, 2001a; Vermis *et al.*, 2004), *B. ambifaria* (formerly genomovar VII) (Coenye *et al.*, 2001b), *B. anthina* (formerly genomovar VIII) (Vandamme *et al.*, 2002), and *B. pyrrocinia* (formerly genomovar IX) (Vandamme *et al.*, 2002). These species can be identified based on a number of biochemical and molecular tests.

The *B. cepacia* complex has been isolated from many environmental sources including soil, water, and plants (Ramette *et al.*, 2005) and is considered to be quite metabolically versatile. Although discovered as a plant pathogen, some

strains of the *B. cepacia* complex are actually beneficial in the environment as bioremediation and biocontrol agents. Certain strains within the complex have been found to enhance plant growth, while others are capable of degrading toxic man-made compounds including those found in pesticides. For example, *B. cepacia* AC1100 has biodegradation abilities against the persistent herbicide 2,4,5-trichlorophenoxyacetate (Kilbane *et al.*, 1982) which is used for weed control, as well as an agent to slow the decoloration of lemons, and increase the size of citrus fruits (Grant, 1979). Strain AC1100 was able to grow with 2,4,5-T as its sole carbon source, but this capability is unstable, and lost spontaneously at high frequency (Kilbane *et al.*, 1982). Bcc strains have also been found effective in reducing sunflower wilt, crown rot in wheat, and mold on apples (McLoughlin *et al.*, 1992; Janisiewicz and Roitman, 1986; Huang and Wong, 1998).

More recently, however, members of the *B. cepacia* complex have become known as important opportunistic pathogens in people who are immunocompromised, cystic fibrosis (CF) patients, and patients with chronic granulomatous disease (CGD). The agriculturally useful features of the *B. cepacia* complex have led to its commercial use in the United States. However, this commercial application has been held back by the risk *B. cepacia* complex bacteria may pose to human health, as all species of the *B. cepacia* complex have been isolated from cystic fibrosis patients. Therefore, the use of *B. cepacia* complex bacteria in commercial applications should be treated with caution since it is unclear whether environmental strains have the potential to become pathogenic (Butler *et al.*, 1995).

As opportunistic pathogens, *B. cepacia* complex species do not normally infect healthy individuals but only those who are immunocompromised. Infection with *B. cepacia* complex bacteria is usually acquired nosocomially from other patients, from contaminated medical devices, or from the environment. The transmission of the *B. cepacia* complex is thought to occur by aerosol droplet or direct contact with other infected individuals or contaminated surfaces (Gibson *et al.*, 2003). A study by Oie and Kamiya (1996) demonstrated the potential that home-use nebulizers may have in the acquisition of *B. cepacia* complex bacteria. Patients with CF and CGD are particularly susceptible to infection with the *B. cepacia* complex.

Chronic granulomatous disease is an inherited disease that normally presents itself as pneumonia, infectious dermatitis, and abscess formation beneath the skin (Johnston, 2001). Phagocytic cells of patients with CGD cannot produce reactive oxygen radicals (ROR) to kill bacterial cells. Since *B. cepacia* complex bacteria are resistant to non-oxidative killing, patients with CGD are susceptible to infection with this bacterium, making it the leading bacterial cause of death for these patients (Winkelstein *et al.*, 2000). Bylund *et al.* (2007) observed that in a healthy human, a *B. cenocepacia* isolate is ingested by neutrophils followed by normal clearing by macrophages. After ingestion of the same isolate, neutrophils from patients with CGD became necrotic, giving further evidence that the virulence of the *B. cepacia* complex in these patients is due to their inability to produce ROR. Infection with *B. cepacia* complex bacteria is also the second most common cause of death (bacterial or other) in patients with CGD (Johnston,

2001).

Cystic fibrosis is a genetic disease in which the gene encoding the CFTR (cystic fibrosis transmembrane regulator) is defective, causing the mucous which lines the digestive tract and lungs to be abnormally thick, making it difficult for cells normally involved in the immune response to act effectively against infection in this area (Gibson *et al.*, 2003). The occurrence of infection with *B. cepacia* complex bacteria in a population of 500 CF patients in Toronto, Canada increased from 10% in 1971 to 18% in 1981, suggesting that infection with the *B. cepacia* complex was an emerging problem (Isles *et al.*, 1984). More recently, a 3-year study of 300 American CF patients showed that while 40% were infected with *Pseudomonas aeruginosa*, the most common pathogen associated with CF, 7% were infected with members of the *B. cepacia* complex (Lambiase *et al.* 2006). However, between 1992 and 1997 in Canada an estimated 16% of CF patients were infected with strains of the *B. cepacia* complex (Speert *et al.* 2002). Infection of CF patients with *B. cepacia* complex bacteria, which decreases life expectancy by 50%, causes accelerated decline in the function of the lungs and can lead to a fatal acute necrotizing pneumonia and bacteremia known as "cepacia syndrome" in approximately 20% of infected patients (Dobbin *et al.*, 2000).

An epidemic strain of the *B. cepacia* complex spread between Toronto, Canada and the United Kingdom between 1986 and 1992. Epidemiological analysis confirmed that affected patients were colonized by a highly transmissible, virulent, Edinburgh/Toronto strain, now known as the ET12 lineage (Govan *et al.*, 1993), which is a *B. cenocepacia* strain. *B. cenocepacia* is most commonly

isolated from clinical material (Mahenthiralingam, 2001), yet it is still unknown if "cepacia syndrome" can be linked exclusively to specific strains or species of the *B. cepacia* complex. In a study of 62 CF patients infected with the *B. cepacia* complex, 46 were infected by *B. cenocepacia*, while 19 were infected with *B. multivorans*. The highest mortality was associated with those infected with *B. cenocepacia*, which was also found capable of spreading to, and replacing infection with *B. multivorans* in other patients, leading to death or chronic infection in most cases (Mahenthiralingam, 2001).

To date only one strain within the *B. cepacia* complex, *B. cenocepacia* J2315, has been completely sequenced and is awaiting formal annotation, while other species remain partly sequenced. All members of the *B. cepacia* complex have large genomes with multiple replicons, and the genome size varies even within a species (Cheng and Lessie, 1994). The genomes of *B. cepacia* complex species are also known to contain many insertion sequences, which promote genetic rearrangement and are capable of activating gene expression (Gaffney and Lessie, 1987). A study by Scrodilis *et al.* (1987) identified five transposable elements that could promote an increase in resistance to penicillin by insertional activation of a poorly expressed β-lactamase gene. The unique genome of *B. cepacia* complex species undoubtedly contributes to their diversity, and can explain the versatile metabolic capabilities of the group.

1.2 Bacteriophages

Bacteriophages, or phages, were discovered independently by Fredrick

Twort in 1915, and Felix d'Herelle in 1917 (D'Herelle, 1922). Phages are simple viruses that can infect bacteria. Most phage genomes contain early, middle, and late genes. The early genes are transcribed first, as they encode proteins used to take over the replication machinery of the bacterial host (Guttman *et al.*, 2005). The middle genes, transcribed after the early genes, encode proteins involved in the replication of the phage genome (Guttman *et al.*, 2005). The late genes are transcribed last, and encode proteins needed for phage assembly, such as the capsid and tail genes, and the release of progeny phage (Guttman *et al.* 2005).

Phages have been evolving, presumably, over billions of years, though a single theory of phage evolution has yet to be agreed upon. One widely accepted theory is the modular theory of phage evolution (Susskind and Botstein, 1978). Usually phages that are closely related will have genomes arranged in a similar way. For example, they will have genes for tail formation, capsid formation, lysis, integration, etc. arranged in a similar order. These functional genetic units are known as modules. Through complete genome sequencing of known phages, it is apparent that phages are mosaics with respect to each other. Phages from the same family have much DNA sequence similarity, but this is interspersed with areas showing no sequence identity (Hendrix, 2002). The theory suggests that phage evolution occurs at the modular level, so the genes within a module can be assorted, while the actual modules will be conserved. For example, the order of the genes involved in lysis/lysogeny may be rearranged, but they will stay within the module. While studying phage P22, Susskind and Botstein (1978) suggested that there were special sequences, called "linkers", between the modules that

promote module recombination, though there is now much evidence suggesting this is not the case. For example, the areas between the modules often do not show sequence identity to the boundaries in other phages. Also, the boundaries are often found in different positions (Hendrix, 2002). It is when illegitimate recombination events occur that a rearrangement of the phage genome occurs and can give rise to these areas showing no sequence homology to related phages. These recombination events can take place essentially anywhere in the phage genome, but it is thought that only rearrangements that do not disrupt the modules will lead to viable phage, and so these areas show up between the modules in most cases (Hendrix, 2002; Brussow *et al.*, 2004). Hendrix *et al.* (1998) suggested, based on the similarity between phages, that most phage genes are probably derived from a common ancestral gene pool and are exchanged by horizontal transfer.

1.2.1 Tailed bacteriophages

Phages are classified based on morphology, type of nucleic acid, and presence of an envelope or lipid. The majority of known phages are "tailed phages", all of which have a double-stranded DNA genome, an icosahedral head, and a tail (Ackermann, 2001). Tailed phages, which belong to the order *Caudovirales*, are then further classified based on the morphological features of the tail. Phages in the *Myoviridae* family have contractile tails, those belonging to *Siphoviridae* have long non-contractile tails, and phages in the *Podoviridae* family have very short tails (Maniloff and Ackermann, 1998). Four percent of phages

are not tailed phages and can be cubical, filamentous, or pleomorphic, and can contain double-stranded or single-stranded DNA or RNA as the genome (Ackermann, 2001).

The *Siphoviridae* family includes λ and λ-like viruses, T1 and T1-like viruses, and T5 and T5-like viruses. This family of phages makes up 61% of all known tailed phages. 25% of known phages belong to the *Myoviridae* family (Ackermann, 2007), which differs from other phage families in that they have a tail sheath involved in contraction. Within this family are T4 and T4-like viruses, P1 and P1-like viruses, and Mu and Mu-like viruses. Regardless of family classification, infection by all tailed phages is initiated by the binding of the tail fiber to the host cell surface. All tailed phages also make a tail tube structure that is involved in moving the phage DNA from the capsid into the host cell, leaving the capsid and tail outside the cell. The main difference between the two dominant phage types is the mechanism by which DNA is inserted into its host, though it is not fully understood how *Siphoviridae* do this. There has been evidence suggesting that the T5 phage tail proteins form a channel through the cell wall. The mechanism by which T5 phages transfer their genome is unique. Approximately 8% of the genome enters the host and is transcribed and translated. The proteins encoded by this initial fragment are involved in degradation of host DNA as well as the transfer of the remaining phage DNA (McCorquodale *et al.*, 1977; Snyder and Benzinger, 1981). However, the driving force behind the genome transfer is not clear.

One hypothesis regarding the *Siphoviridae* mechanism of DNA transport relies on the pressure inside the capsid being very high relative to the pressure outside. Using force-measuring optic tweezers, Smith *et al.* (2001) determined the force needed to package a strand of DNA into the capsid of ϕ29. They observed that the pressure inside the capsid begins to build up once approximately half of the phage genome had been packaged, and builds up to approximately 50 picoNewtons (pN). They predict that this force may be responsible for at least the initial ejection of phage DNA.

In *Myoviridae*, the sheath surrounding the tail tube contracts to push the viral DNA into the host cytoplasm. The method of injection used by the T4 phage is known in more detail than methods used by siphoviruses. T4 proteins gp5 and gp27, found at the end of the tail tube (gp19), are capable of forming a heterohexameric complex. Gp5 forms a trimer that acts as a needle, while the gp27 protein forms a hollow cylinder into which the gp5 needle is attached (Kanamaru *et al.*, 2002). The trimer has three domains, including a domain inserted into the gp27 cylinder, and a lysozyme domain that degrades the peptidoglycan of the bacterial host (Kanamaru *et al.*, 2002). Following phage adsorption and contraction of the tail sheath, the tail tube is moved closer to the cell membrane, and the lysozyme domain of the trimer complex is involved in puncturing the host cell membrane to allow DNA to pass through (Kanamaru *et al.*, 2002). The contraction that drives the tail tube through the cell wall involves the tail sheath protein, gp18 (Kanamaru *et al.*, 2002). The tail sheath, which surrounds the tail tube, is made up of stacks of protein rings of 6 subunits each

(Miller *et al.*, 2003). The tail sheath protein has three domains: the N-terminal domain, the middle domain, and the C-terminal domain. During different states of sheath contraction the domains have been observed to have different conformations (Kostyuchenko *et al.*, 2005). During contraction, the sheath becomes wider and shorter as the subunits are vertically closer together, the middle domains do not interact with neighboring gp18 middle domains, and the gp18 proteins are pulled away from the tail tube. When extended, neighboring gp18 middle domains are in close contact with each other, and the C-terminal domain is moved closer to the tail tube (Kostyuchenko *et al.*, 2005). How the entire genome is transported through the membrane is still not completely understood, and it is seems unlikely that the tail contraction by itself would be sufficient to transport the relatively long DNA molecule through the host membrane.

Phage T7, a member of the *Podoviridae* family, has DNA transport coupled to transcription. After the release of the gp16 protein, which is thought to locally hydrolyze the peptidoglycan, a portion of the genome, containing three *E. coli* RNA polymerase promoters and a T7 promoter is passively transported into the host cell. Once inside, host RNA polymerase will bind, initiate transcription, and pull more (approximately 7 kb) of the genome into the cell (Garcia and Molineux, 1996). The second portion of the genome that enters the host encodes a T7 polymerase, which will transcribe the rest of the phage genome. This transport mechanism is not, however, characteristic of all *Podoviridae*.

In both *Siphoviridae* and *Myoviridae*, the tail tape measure protein has been shown to be a molecular ruler involved in determining the length of the tail (Cornelis *et al.*, 2006). It is thought that the tail tube is assembled around the tape measure protein and the length of the tape measure protein somehow controls the tail tube length (Katsura and Hendrix, 1984; Katsura, 1984). One hypothesis is that the N-terminal domain of the tape measure protein is attached to the growing end of the tail tube, and as the tail tube grows the tape measure protein will be stretched (Journet *et al.*, 2003). In phage T4, changing the length of the tape measure gene, gene 29, caused a proportional change in tail length (Abuladze *et al.*, 1994). In phage λ the tape measure protein is encoded by gene H. Katsura (1984) reported that by changing the length of the tape measure protein, by deleting or duplicating a middle region of the gene H, the tail tube length changes proportionally.

1.2.2 Mechanisms of phage bacteriolysis and lysogeny.

Most phages can be classified as either lytic or temperate (Matsuzaki *et al.*, 2005). Lytic phages repeat a life cycle that results in self-proliferation and lysis of the host bacteria. Temperate phages can undergo a lysogenic cycle (Figure 1.1) in addition to a lytic cycle. In the lysogenic cycle the phage genome is integrated into the host cell genome and is replicated along with host DNA without destroying the host bacteria. In this state the phage is known as a prophage. Eventually, spontaneously or due to an induction event, the lysogenized phage can de-integrate and undergo a lytic life cycle.

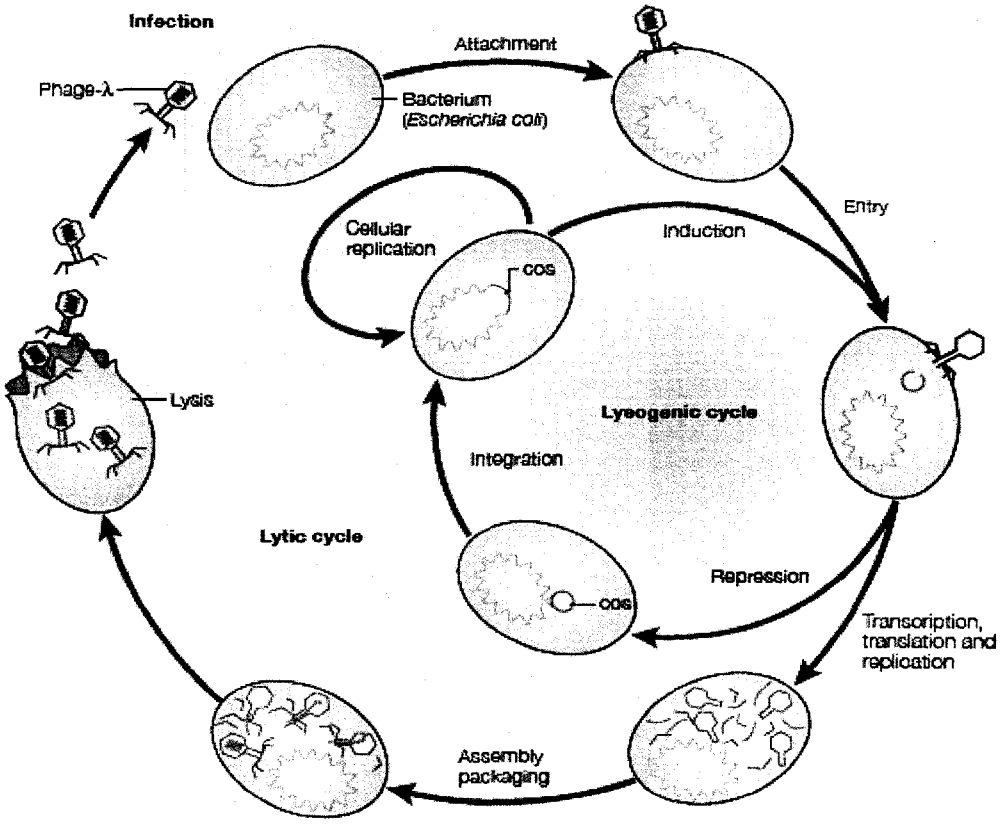


Figure 1.1 Bacteriophage lytic and lysogenic life cycles. Taken, with permission, from Campbell, 2003.

While a lytic phage can only undergo the lytic life cycle, the temperate phage can undergo either, depending on the environment. The lysogenic cycle is favored in starved cells, or at high multiplicity of infection (MOI). The temperate phage genome, excluding transposable phages, typically contains an *int* gene, encoding an integrase that promotes site-specific recombination between the phage and bacterial insertion sites, as well as a *xis* gene encoding a protein for excision of the phage genome from the host genome. Phage integrases carry out recombination between attachment sites on the phage and bacterial genomes, known as *attP* and *attB*, respectively.

The system used by phage λ for controlling the switch between the lysogenic and lytic life cycle has been characterized better than that of other known phages and is shown in figure 1.2. For λ, the main proteins involved in determining the type of growth are CI CII CIII and Cro (Herskhowitz and Hagen, 1980). CI is a repressor of genes involved in the lytic cycle, while the Cro protein represses the genes required for lysogeny. CII promotes the lysogenic cycle by stimulating the synthesis of CI and Int, and by inhibiting lytic growth. A bacterial protein called Hfl degrades the phage CII protein, but Hf1 is inhibited by CIII. The CIII protein increases as MOI increases, therefore Hfl is inhibited at high MOIs and CII can exist at high levels to promote lysogenic growth since it is not being degraded by Hfl. Conversely, at low a low MOI, CIII will be low, Hfl is not inhibited, and CII activity is reduced. As a result CI is not made, Cro concentrations increase and lytic growth is promoted (Herskowitz and Hagen, 1980). The λ lytic genes are expressed from promoters *pL* and *pR*. Expression

from these promoters is controlled by an operator region made up of three segments, *oL1/oL2/oL3* and *oR1/oR2/oR3*, each of which can bind CI (Snyder and Champness, 2003). During lysogeny, CI binds to *oL1/oL2* and *oR1/oR2*, causing repression of the lytic genes (Snyder and Champness, 2003). CI, when bound to *oR2*, activates transcription from its own promoter, *pRM*, and thus controls its own expression. The lytic life cycle begins when CI is cleaved during the SOS response (Snyder and Champness, 2003). The Cro protein, which binds to *oR3*, represses transcription from *pRM*, preventing the expression of CI, therefore removing repression of transcription from *pL* and *pR*, allowing the transcription of lytic genes (Snyder and Champness 2003).

Regardless of the life cycle, the first stage of infection involves the phage binding to its receptor on a specific host. Phage receptors can include pili, outer membrane proteins, or lipopolysaccharides (Datta *et al.*, 1977). The phage will have a specific host range to which it can bind and infect. As shown in Figure 1.1, once the phage binds to its specific receptor (known as adsorption), the phage can inject phage DNA into the host cell cytoplasm by releasing an enzyme to break down a portion of the cell wall (Tsugita *et al.*, 1968). If the phage has the genes for lysogeny and the conditions are fitting, it can then integrate into the host genome. Otherwise, phage proteins will degrade host DNA and phage DNA will be replicated using host cell machinery. Phage proteins will be transcribed and translated, and newly synthesized DNA will be packaged into new phage heads, followed by attachment of a tail to the DNA-filled head (Matsuzaki *et al.*, 2005).

Two phages proteins, holin and lysin, are responsible for the disruption of the host cell wall, which leads to the release of newly formed phage particles from the cell. The holin protein is a small protein that accumulates in the bacterial cell membrane. This accumulation is associated with the collapse of the membrane potential and permeabilization of the host cell membrane (Wang *et al.*, 2000), essentially forming holes throughout the membrane. A holin inhibitor is normally encoded within the same gene as the holin and is involved in controlling lysis time. This indicates that holin proteins are also responsible for the length of time it takes for the lytic cycle to occur (Wang *et al.*, 2000). As a broad classification, lysins are known as peptidoglycan hydrolases. More specifically, lysins can be endo-β-N-acetylglucosaminidases, N-acetylmuramidases, or N-acetylmuramoyl-L-alanine amidases, depending on where they cleave the peptidoglycan (Young, 1992). Lysin moves through the holes in the membrane formed by the holin and degrade the peptidoglycan layers (Matsuzaki *et al.*, 2005), allowing phage particles to be released and infect neighboring bacteria.

1.2.3 Transposable bacteriophages

The life cycle of transposable phages differs slightly from that of the strictly lytic or temperate phage, as outlined in Figure 1.2. The best example of this life cycle is found in phage Mu, an *Escherichia coli* phage, which combines properties of a transposon and a temperate phage. There have been many other Mu-like phages isolated, and they have been found in both Gram positive species, such as *Deinococcus radiodurans* (Morgen *et al.*, 2002) and Gram negative

species including *E. coli*, *P. aeruginosa*, *Haemophilis influenzae*, *Neisseria meningitidis*, and *B. cenocepacia* (Hull *et al.*, 1978; Rehmat and Sapiro, 1983; Fleischmann *et al.*, 1995; Klee *et al.*, 2000; Summer *et al.*, 2004).

Mu uses non-replicative transposition to integrate into the host genome, and replicative transposition during lytic growth. Transposition in Mu begins with a single-strand break at the 3′-ends of the genome. These ends are then transferred to the free 5′-ends of the target site. During nonreplicative transposition the DNA is simply inserted. During replicative transposition the whole phage genome is replicated, initiated at the 3′-ends. Replication joins the donor and target sites, creating the cointegrate structure. Once resolved, this results in another copy of the Mu genome in a new location, as well as in the initial location (Shapiro, 1979). This replicative transposition is the mechanism Mu uses for lysis, during which approximately 100 copies of the DNA will be produced in 40 minutes.

As is the case with all temperate phages, repressor and regulatory proteins (Rep and Ner, respectively, in the case of Mu) are involved in the switch from a lysogenic to a lytic life cycle (Figure 1.3). These proteins compete with each other and with the transposase protein for binding sites on the operator region. There are three sites in the operator region that control transcription from two different promoters, *Pe* and *Pcm* (Krause and Higgins, 1986). The Ner protein negatively regulates Rep, and at high enough concentration, initiates the lytic cycle due to inhibition of the repressor protein (Goosen and van de Putte, 1987). At low concentration, Rep will bind to the operator site (*o1* and *o2*) and block

transcription of genes from the *Pe* promotor, including genes needed for lytic growth, such as *ner*, and therefore the prophage remains repressed (Krause *et al.*, 1983). When bound to these operator sites the repressor also prevents the transposase from binding. If there is a high concentration of Rep, however, it binds to the *o3* site of the operator region and transcription is blocked from the *Pcm* promoter (Krause and Higgins, 1986). This reduction in *Pcm* activity blocks transcription of Rep due to synthesis of the Ner repressor. Three host proteins, Integration host factor (IHF), H-NS, and ClpXP protease, are also involved in Rep activity. IHF and H-NS are thought to help stabilize the binding of the repressor to 01 and 02 sites, during which the phage remains lysogenized (Krause and Higgins, 1986). ClpXP is able to degrade the repressor, thus causing a switch to the lytic cycle under certain conditions (Geuskens *et al.*, 1992), such as high temperature and stationary phase. At a higher temperature the repressor protein, Rep, changes conformation, lowering its DNA binding affinity, but also exposing a ClpXP recognition site. During stationary phase it is thought that a similar conformation change occurs, also exposing a recognition site for the protease (Ranquet *et al.*, 2005). The actual replicative transposition event requires MuA, a transposase, MuB, a transposition protein, and an enhancer element found approximately 950 bp from the left end of the genome between *o1* and *o2* (Leung *et al.*, 1989).

There are three MuA binding sites at each end of the Mu genome, *L1, L2, L3*, and *R1, R2*, and *R3*. During initial transposition, MuA is bound to *L1, L2, R1*, and *R2* (Figure 1.3).

The MuA proteins form a tetramer and interact with the enhancer element to form the LER complex (Left end, Enhancer, Right end). This enhancer site is known as the IAS (internal activation sequence), and is recognized by the N-terminal domain of MuA. The IAS overlaps with the operator region where the repressor protein binds to promote lytic growth (*o1* and *o2*) (Mizuuchi and Mizzuchi, 1989; Leung *et al.*, 1989). This overlap allows the repressor protein to directly inhibit transposition from occurring (and therefore lytic growth). The enhancer has been shown to be necessary for the initial formation of the transposome (the protein-DNA complex), but is not involved in the cleaving of the DNA, as cleavage can occur in the absence of this IAS (Surette and Charconas, 1992). The MuA tetramer is able to catalzye the single strand break at the 3'-ends of Mu, as well as at the 5' end of the target sites and catalyzes the joining of the ends of the phage DNA to the target DNA. Baker and Luo (1994) isolated MuA mutants with single amino acid substitutions that were unable to cleave and transfer the Mu genome. They found that substitutions in Asp-269 and Glu-392 residues were responsible for the defective transposition. The mutants were able to form the tetramer, but could not cleave/transfer DNA. These residues were found to be analogous to residues required for transposition by the retroviral integrases, such as those found in HIV and Rous sarcoma virus (Baker and Luo, 1994).

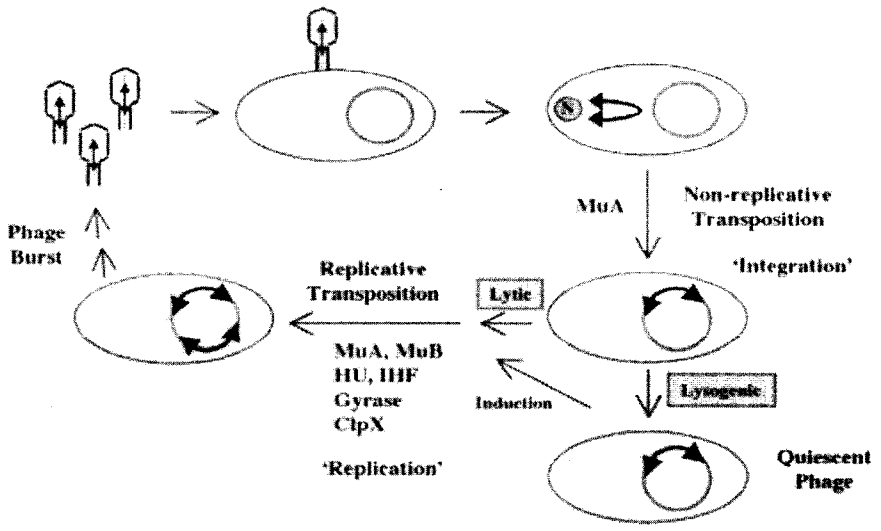


Figure 1.2 Lytic and lysogenic lifecycles of transposable bacteriophage Mu. Integrative transposition leads to an integrated phage. Following induction, replicative transposition, requiring both phage and host encoded proteins, leads to host cell lysis. Taken, with permission, from Sokolsky and Baker, 2003.

MuB, the transposition protein, is involved in the selection of the target DNA and activating strand transfer through the activation of the transposome (Baker *et al.*, 1991). MuB has been shown to bind DNA with high affinity when bound to ATP, but not when bound to ADP. MuB preferentially binds to areas that are not Mu sequence. Mu DNA sequence somehow mediates the stimulation of MuB, by MuA, to hydrolyze ATP, which results in MuB-ADP, causing a release from the DNA. Thus MuB, along with MuA, is involved in choosing the target site. In this way, MuB can be prevented from binding to an incorrect target site, such as the integrated Mu genome (Adzuma and Mizuuchi, 1988; 1989). MuB is also involved in stabilizing the whole DNA complex during transposition by interacting directly with any of the MuA monomers at each end and interacts with the transposome from the very beginning (Naigamwalla and Chaconas, 1997). Because MuB can interact with either of the MuA monomers at each end it does not interfere with the cleavage activity of MuA. MuB has also been found to prolong the amount of time in which a single end of Mu donor DNA is bound to the end of the target. This may be to ensure that the other end of the Mu DNA has time to join the target DNA as well (Lemberg *et al.*, 2007). Another role of MuB is to protect the MuA-DNA complex from ClpX until it is needed. The region of MuA that recognizes MuB overlaps with the ClpX recognition site (Levenchko *et al.*, 1997). ClpX is an ATPase involved in disassembly of the LER complex to allow replication of the phage DNA to begin (Levenchko *et al.*, 1995). By blocking this site, MuB ensures that ClpX cannot interfer with the complex until strand transfer is complete.

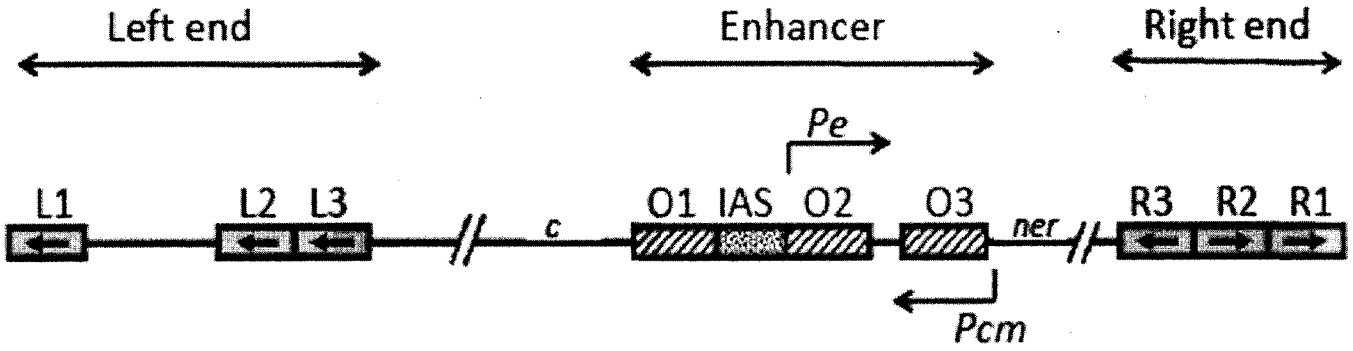


Figure 1.3 Indicator system for Mu lysogeny/lysis and binding sites of transposase proteins. The six transposase binding sites and three operators (O1-O3) are shown, as well as the internal activation site (IAS). Arrows indicate the direction of transcription from the early lytic promoter (*Pe*) and the lysogenic promoter (*Pcm*).

1.2.4 Lysogenic conversion

Prophages are often found in bacterial genomes. In *E. coli* for example, there have been 51 different functional prophages identified from 27 O157:H7 isolates (Osawa *et al.*, 2000). This number does not even take into account prophage remnants, or prophages that are not easily induced. Another study found 136 functional phages released from 173 *Salmonella typhimurium* isolates (Schicklmaier *et al.*, 1998). Because these prophages are sometimes capable of transferring useful genes to the lysogen, they can be important in the evolution of pathogenic bacteria. Lysogenic conversion is a process by which a lysogenic bacteriophage can convert a nonpathogenic strain of bacteria into a pathogenic one. For example, the toxins for cholera and scarlet fever produced by *Vibrio cholera* (Waldor and Mekalanos, 1996) and *Streptococcus pyogenes* (McCloskey, 1973), respectively, are all encoded by prophage-acquired genes. Another example where lysogenic conversion may have occurred is in the *B. cenocepacia* J2315 lysogen harbouring the Mu-like phage, BcepMu. This phage encodes a possible ExeA homologue, which is involved in secretion of toxins, and a 3-O-acyltransferase homolog, which is responsible for resistance to certain antibiotics (Summer *et al.*, 2004).

There is also an abundance of prophage-encoded genes that can increase the fitness of the lysogen, besides genes encoding extracellular toxins (Brussow *et al.*, 2004). Prophages in *Salmonella enterica* and *E. coli* O157, for example, encode superoxide dismutase, which can protect the lysogen from reactive oxygen species. Other proteins encoded by prophages include those that alter

antigenicity, such as O-antigen acetylase and glucosyl transferase proteins, and type III effector proteins (found in *S. enterica* serovar typhimurium prophages) (Brussow *et al.*, 2004). Therefore, phages being considered for use as therapeutic agents should be sequenced to ensure that they do not harbor potential virulence determinants.

Transfer of phage genes to its host is not the only mechanism by which phages can change the virulence of a bacterial strain. Phage integration events can also cause virulence by disrupting certain bacterial genes, or can sometimes attenuate a strain if integrated into a gene involved in virulence. Lysogenization of *Staphylococcus auerus* with bacteriophage ϕ13, for example, results in an interruption of the β-toxin gene, and therefore causes a loss in expression of this toxin (Coleman *et al.*, 1991). Prophage integration can cause a change in the expression of bacterial genes, and bacterial virulence/fitness through a variety of mechanisms. As more and more prophages are sequenced it will provide more insight into where these phage virulence factors originate and how they are involved in the evolution of pathogenic bacteria. Perhaps they originate from bacterial hosts and the phages have incorporated them over time, or they may be original phage genes.

1.2.5 Restriction modification systems

Phages are thought to be one of the most abundant particles in the biosphere (Hendrix *et* al., 1999), much more abundant than their bacterial hosts.

Because they are constantly in contact with phages, some bacteria have developed mechanisms to protect themselves against phage infection. One of these mechanisms is the host restriction modification system. By this mechanism, when phage DNA enters the host cell the phage DNA is degraded by host restriction enzymes, but the host DNA is protected, usually by a specific pattern of methylation. This is one reason why a phage may be capable of infecting some strains better than others. If a host uses this mechanism to protect against phage infection, only those phage particles that have become resistant to this mechanism will be capable of infecting that strain. Because it is possible for phages to become resistant to these mechanisms of defense, they force the bacteria to evolve new defense mechanisms. This is another way phages aid in driving bacterial evolution.

Although some phages propagate more readily in some strains than others, the presence of a host-encoded restriction modification system could cause a phage to propagate poorly in a host it is capable of adsorbing to. Burkholderia phage CP1 was found to propagate more rapidly in *B. cepacia* strain 383 than strains 104 and 382, and it was suggested that strain 383 lacks a restriction modification system for protecting itself from foreign DNA (Cihlar *et al.*, 1978). However, the authors provide no evidence to show that the poor propagation is actually due to a restriction modification system and not due to some other factor.

Some phages encode their own restriction modification systems to degrade host DNA. Bacteriophage T4, for example, has its own restriction system that can degrade host DNA but methylates T4 DNA, protecting it from other restriction

systems (Carlson *et al.*, 1994; Evdokimov *et al.*, 2002). T4 also encodes an α- and a β-glucosyltransferase that also protects the T4 DNA from host restriction systems (Carlson *et al.*, 1994). *B. mallei*-specific phage ϕE125, a prophage in *B. thailandensis* E125, was found to encode two DNA methyltransferases, gp27 and gp56, in its genome (Smith and Jeddeloh, 2004). Once ϕE125 is induced from the genome of *B. thailandensis,* the phage DNA but not host DNA, is methylated at cytosine bases thus protecting phage DNA from any restriction enzymes produced by its host (Smith and Jeddeloh, 2004).

1.3 Phage therapy

Phage therapy involves the use of lytic phages to kill infecting bacteria. However, modified temperate phages or phage products may also be considered for use as antimicrobial agents (Hagens *et al.*, 2004). Since their discovery, phages have been studied worldwide and used in a number of practical applications as antimicrobial agents (Sharp, 2001). Phages were first used as therapeutic agents in 1921 when *Staphylococcus* specific phages were injected subcutaneously to treat boils (d'Herrell, 1922). Phages were also used extensively to control cholera outbreaks in 1932 (Sharp, 2000). With the discovery of chemical antibiotics, however, interest in phage diminished in the Western world. While many countries including Russia and Georgia still use phages to treat antibiotic- resistant bacterial infections, much more research must be carried out before phage therapy will be deemed acceptable in countries such as Canada and the United States.

In order for phage therapy to be effective, selected phages cannot simply be added to the bacterial infection without further preparation. The early use of phages as therapeutic agents was fraught with many problems, such as the use of impure phage preparations that resulted in deleterious effects, and poor commercial phage preparations containing low activity titers or nonviable phages (Straub *et al.*, 1933). In 1929, Larkum observed that some patients receiving phage injections against staphylococcal infections experienced mild or severe side effects. Merril *et al.* (1996) suggest that early problems with phage therapy involved phage preparations that still contained toxins from lysed host cell debris. Thus it is extremely important to prepare phages properly in order to develop reliable therapeutics. Using phage preparations that have been contaminated can lead to morbidity or even mortality. Also, because phages are isolated after they have infected their bacterial hosts, removal of any bacterial endotoxin from the preparation will also be important.

Biswas *et al.* (2002) showed the importance of testing the phage preparations on animals under stress, as opposed to limiting safety testing to healthy individuals, in their study using phages to rescue mice infected with *Enterococcus faecium.* If therapeutic phages are only tested on healthy individuals, possible negative effects may not become evident. In most cases phage therapy will be used to treat existing infections, so animals will already be sick when the therapy is employed. As a control, a phage known to be inactive against *E. faecium* was given to both infected and healthy mice. Biswas *et al.* (2002) found the control phage to increase mortality in the infected mice, but not

in healthy mice. The authors suggest that stressed animals are more sensitive to the phages, or to trace amounts of endo- and exotoxins in the phage preparation. This emphasizes the importance of developing purified phage preparations.

Although animal infection models outlining the use of phages to treat infection with the *B. cepacia* complex have not yet been published, the success in isolating and characterizing new phages specific to the *B. cepacia* complex, and the success of phages in other models suggest that a treatment strategy involving phage therapy is promising.

1.3.1 Infection models with the B. cepacia complex

As described previously, *B. cepacia* complex bacteria are resistant to many commercially available antibiotics, so an alternative treatment method such as phage therapy may be more effective in treating infections with the *B. cepacia* complex. In order to determine the efficacy of phage therapy against *B. cepacia* complex infections, an animal infection model must be employed to determine which strains of the *B. cepacia* complex are most virulent, and which would be best to use in phage rescue experiments. Infection models for other *Burkholderia* species, such as *Burkholderia mallei* and *Burkholderia pseudomallei,* have involved Syrian hamsters or mice (Moore *et al.*, 2004; Ulrich *et al.*, 2004). A number of non-mammalian models including alfalfa seedlings (Bernier *et al.*, 2003), *Caenorhabditis elegans* (Cardona *et al.*, 2005; Kothe *et al.*, 2003; Markey *et al.*, 2006) and *Acanthamoeba* species (Lamothe *et al.*, 2004; Marolda *et al.*, 1999) have been used in *B. cepacia* complex infection models, but the correlation

between virulence of *B. cepacia* complex strains in these models and in mice and rats is fairly poor. Using the *C. elegans* model, for example, Cardona *et al.* (2005) found that *B. cenocepacia* strains that are known to be clonal (Mahenthiralingam *et al.*, 2000) showed different pathogenicity phenotypes. *B. cenocepacia* J215, for example, only received a pathogenicity score of 1 (on a scale of 0-3), while strain K56-2 received a score of 3 (Cardona *et al.*, 2005). Using alfalfa seedlings, while more ethically acceptable than mice or rats, is not ideal, as symptoms and severity of infection are difficult to measure accurately because they rely on visual observations such as leaf color (Bernier *et al.*, 2003).

In 2008, Seed and Dennis developed an infection model for the *B. cepacia* complex using *Galleria mellonella,* or the "wax worm". This model uses worm death as an endpoint, which is easily assessed. When the wax worms become infected with virulent strains of the *B. cepacia* complex and subsequently die, they turn black due to melanization, a part of the worm's innate immune system (Seed and Dennis, 2008). This model shows that the most virulent strains against the wax worm belong to *B. cenocepacia* and *B. cepacia*, while strains belonging to *B. multivorans* were found to persist the longest (Seed and Dennis, 2008). Most strains isolated in a clinical setting belong to *B. cenocepacia* and *B. multivorans* (Speert *et al.*, 2002), which would suggest that perhaps these two strains would be most virulent. Although this was not that case in the wax worm model, the results obtained using this model seem to correlate well with what has been shown using other models, suggesting that is a reliable model (Seed and Dennis, 2008).

1.3.2 Uses for phage

Natural phages attack and kill only certain hosts, and so, unlike antibiotics, they should be ineffective against most normal flora. There have been a number of successful studies using phages to treat infections with pathogens including *E. coli* and *E. faecium*. In 1982, Smith *et al.* reported that phages could effectively treat *E. coli* infections in mice. One intramuscular injection of the *E. coli* strain 018:K1:H7 ColV+ specific phage in mice was more effective than multiple intramuscular injections of tetracycline, ampicillin, chloramphenicol, and trimethoprim, and just as effective as multiple injections of streptomycin. In 1987 Smith *et al.* also used phages to successfully treat calves with gastrointestinal infections caused by enteropathogenic strains of *E. coli*. Another study showing the potential of phage therapy used mice infected with an antibiotic resistant strain of *E. faecium* (Biswas *et al.*, 2002). A 200-fold decrease in blood bacterial titers was observed when ENB6, an *E. faecium*-specific phage, was administered to infected mice when compared with blood bacterial titers from control mice 20 hours after initiation of the experiment. A single dose of ENB6 was also shown to rescue 100% of mice when administered 45 minutes after infection (Biswas *et al.*, 2002).

More recently, phages were shown to be effective in eliminating pathogens such as *Listeria monocytogenes*, *Campylobacter jejuni*, and *Salmonella* spp. (Greer, 2005), and in 2006 the Food and Drug Administration approved the use of a topical phage preparation for elimination of strains of *L. monocytogenes* from meat and poultry products (Lang, 2006). Although the majority of studies

using phage therapy involve animal trials, natural phages are also being tested for their safety in human trials. In 2005, Bruttin and Brussow employed 15 healthy volunteers to receive varying doses of the *E.coli* phage T4 in their drinking water. As expected, no phages were detected in stool samples when volunteers consumed drinking water with the placebo. The phage was detected in all stool samples when volunteers consumed the high dose (10^5 pfu/ml) and was detected in 13 or the 15 when volunteers consumed the low dose of phage (10^3 pfu/ml). There were no side effects attributed to the phages in any of the volunteers.

Phage lysins are also being considered for use as antimicrobial agents against Gram positive bacteria. As described earlier, lysins are enzymes that can directly degrade the peptidoglycan layer of the bacterial cell wall. Lysins are able to degrade the cell wall of non-growing cells, unlike some antibiotics that rely on destroying newly forming peptidoglycan. Although the normal phage cycle involves lysins acting from within the bacterial cell wall, these enzymes can also degrade the peptidoglycan from outside the cell (Matsuzaki *et al.*, 2005). There have been numerous studies using phage lysins to treat antibiotics resistant pathogens, such as *Streptococcus pneumoniae*. When mice were injected intraperitoneally with phage lytic enzymes one hour after infection with antibiotic resistant *S. pneumoniae* they were rescued from the infection. Mice challenged with the strain alone died within 72 hours (Jado *et al.*, 2003). When rats with experimental endocarditis caused by *S. pneumoniae* were treated with a lytic enzyme from pneumococcal phage, Cp1-1, pneumococci were cleared from blood quite rapidly (within 30 minutes). For the enzyme to work efficiently the rats had

to receive a high continuous dose (250 mg/kg/h for 6 hours). When rats were treated with vancomycin, however, pneumococci were cleared from the blood in only one of five rats. Although there was an increase in cytokine secretion in those animals treated with Cpl-1 compared with those treated with vancomycin (presumably due to the quick release of bacterial cell wall fragments), this data suggests lytic enzymes may be a good alternative approach to treat antibiotic resistant pathogens (Entenza *et al.*, 2005).

1.4 *Burkholderia* phages

Burkholderia phages are commonly found in environmental sources, such as agricultural soils (Seed and Dennis, 2005; Summers *et al.*, 2006), but many are also isolated as prophages (Langley *et al.*, 2003; Seed and Dennis, 2005; Summers *et al.*, 2007). Two early isolated *Burkholderia* phages, CP1 and CP75, were isolated prior to the *Burkholderia* genus classification (Cihlar *et al.*, 1978; Matsumoto *et al.*, 1986). CP1 was discovered as a prophage in *Pseudomonas cepacia* strain 249 and had a fairly small host range on the strains tested (Cihlar *et al.*, 1978). CP75, however, was reported as a transducing phage capable of infecting approximately 50 different strains of *P. cepacia* (Matsumoto *et al.*, 1986). Typically, phage host ranges are not that broad, and it is possible that due to difficulty in distinguishing between species/strains in the past, the CP75 host range is much smaller than initially suggested. More recently, two transducing phages, NS1 and NS2, were isolated from *B. cepacia* strains ATCC 29424 and ATCC 17616, respectively (Nzula *et al.*, 2000). Both phages were reported to be

T-even-like phages and similar to phage CP75. Host ranges of these phages included both clinical isolates and environmental isolates of the *B. cepacia* complex, as well as some strains of *P. aeruginosa* (Nzula *et al.*, 2000).

Burkholderia species can cause a variety of diseases in both animals and plants, and, as mentioned, many strains are known to harbor prophages (Langley *et al.*, 2003). Two additional strains of *Burkholderia*, *B. pseudomallei* and *B. mallei*, although not in the *B. cepacia* complex, are also capable of causing infection in humans (melioidosis and glanders, respectively) (Wiersinga *et al.*, 2006). Temperate phages have been shown to make up over 10% of the *B. pseudomallei* genome, while *B. mallei* is known to harbor no temperate phages (Nierman *et al.*, 2003). There have been a number of *Burkholderia* phages isolated from pathogenic strains of *Burkholderia,* and some of these phages are related to phages from other genera that are known to transfer virulence traits to their hosts (Summer *et al.*, 2007). While those *Burkholderia* phages that have been sequenced do not appear to encode definite virulence factors, more research is necessary to determine if this is actually the case (Summer *et al.*, 2007). It seems unlikely that none of these prophages would play a role in this pathogenicity, though the published literature would suggest otherwise.

1.4.1 B. cepacia complex specific phages

Recently there have been a number of phages characterized and sequenced that are specific to the *B. cepacia* complex. In 2005 Seed and Dennis isolated four lytic phages from onion rhizosphere as well as five temperate phages from

five different *B. cepacia* complex strains. These phages were found to be specific to the *B. cepacia* complex, and were not lytic against other bacterial species tested, including *P. aeruginosa*, *P. putida*, and *B. pickettii*. The lytic phages named KS1, KS5, and KS6 exhibited a fairly broad host range, lysing up to 10 of the 24 *B. cepacia* complex strains tested (Seed and Dennis, 2005). Of these nine phages isolated by Seed and Dennis (2005), only temperate phages KS9, isolated from a *B. pyrrocinia* 21824 lysogen, and KS10, isolated from a *B. cenocepacia* K56-2 lysogen, have been sequenced and annotated, revealing a phage similar to Phi E125 and a transposable phage showing homology to Mu. In 2004 Summer *et al.* discovered a transposable phage, BcepMu, as a temperate phage of *B. cenocepacia* J2315, and in 2005 characterized three more *B. cepacia* complex specific phages isolated from soil. Most *B. cenocepacia* strains similar to K56-2 harbor BcepMu, though K56-2 does not. Through searching the sequence of *B. cenocepacia* strain J2315, as well as in a more extensive search for lysogeny, a previous study found BcepMu to be the only known lysogen of J2315 (Langley *et al.*, 2003). However, PCR results, as well as searching the genome of recently sequenced J2315 for KS10 specific sequence, show that, like BcepMu, KS10 had been integrated into the genome of J2315.

Summer *et al.*, 2004

BcepMu was isolated as a 36,748 bp prophage in *B. cenocepacia* J2315 and is capable of infecting *B. cenocepacia* K56-2. Because BcepMu has characteristics similar to Mu, it can integrate into near random locations in the

genome of its host. PCR revealed that BcepMu is also a prophage within many other ET12 lineage isolates, including C5424 and BC7, but not capable of infecting any other strains/isolates tested. In *B. cenocepacia* J2315 the prophage was found in chromosome 3, inserted in a gene encoding a putative transcriptional regulator with an AAA domain. The BcepMu genome is organized fairly similarly to Mu. The genome is divided into four modules: genes for replication/regulation/pathogenesis, genes involved in host cell lysis, genes required for capsid formation, and tail/tail fiber formation. The first three modules contain genes that are closely related to the analogous genes in Mu, though the first module appears to be inverted with respect to the Mu genome. The BcepMu tail and tail fiber genes are more closely related to the analogous genes in bacteriophages P2, and the genome lacks the invertible tail region found in Mu.

Two BcepMu genes have been identified as possible virulence factors. BcepMu gp8 encodes a homologue of ExeA, which is from *Aeromonas hydrophilia*. This gene is involved in secreting a group of toxins. Three other phages related to BcepMu, SalMu Ty2, ChromoMu, and PhotoMu, also encode this gene, although they are prophages in *Salmonella typhi, Chromobacterium violaceum*, and *Photorhabdus luminescens*, respectively. Another possible virulence factor identified in BcepMu, but not in the three related phages, is an MdmB and OafA homologue, encoded by gp53. MdmB is a 3-*O*-acyltransferase from *Streptomyces mycarofaciens*, while OafA is a *S. typhimurium* gene. MdmB is involved in resistance against macrolide antibiotics, such as erythromycin,

while OafA is responsible for acetylation of the *O*-antigen (Hara and Hutchinson, 1992; Slauch *et al.*, 1996)

Summer *et al.*, 2006

Using a soil enrichment procedure with *B. cepacia* strain 74-34NE, four lytic myophages, Bcep781, Bcep1, Bcep43, and BcepB1A, were identified and subsequently sequenced. Phage genomes were 48,247 bp, 48,177 bp, 48,024 bp, and 47,399 bp in length, respectively. The final assemblies of all phage genomes were found to overlap, with no variable host DNA at either end, suggesting these phages are not transposable. Comparison of the genomes the four phages revealed Bcep781, Bcep43, and Bcep1 are closely related, while BcepB1A had 12 genes sharing homology with the other three. Bcep781 exhibited 97.6% sequence identity to Bcep43 and 87.4% sequence identity to Bcep1. Bcep43 exhibited 88.7% sequence identity to Bcep1. Differences between the three were mostly due to single nucleotide polymorphisms, insertions, and/or deletions. The three phages closely related were found to have a GC content of 63%, while BcepB1A had a GC content of only 53%. Bcep43, Bcep1, and Bcep74 phage genomes were organized into four main transcription units, and arranged in the usual way, with genes for head assembly/DNA packaging first, followed by genes involved in lysis, and genes required for tail/tail fiber formation at the far right end. The BcepB1A genome is arranged differently, with most of its genes encoded on one strand, with transcription orientation from right to left. It is arranged into modules as well, but the lysis module is followed by genes involved in tail assembly, with

genes encoding proteins involved in head assembly to the right of the tail assembly module. However, due to transcription orientation, genes involved in capsid formation will still be transcribed before tail genes.

Hens *et al.*, 2005; 2006

BcP15 is a small temperate phage of *B. cepacia* strain DR11, an environmental isolate from a delta in the Sunderbans, a vast area of forest and saltwater swamp in Bangladesh and India. This phage is from the *Siphovirdae* family of phages and has been characterized as having a head of approximately 65 nm in diameter and a tail length of 200nm. Interestingly, EMs of this phage shows wavy tail fibers that range from 400-700 nm. The genome of this phage only 11.9 kb in length, which is much smaller than other reported *B. cepacia* complex phages (Summers *et al.*, 2004; 2006), and seems quite small to still encode all necessary phage genes. This length was determined by agarose gel electrophoresis in addition to electron micrography, since BcP15 DNA resisted digestion by 30 different restriction enzymes. This phage, though characterized as a *B. cepacia* complex phage, is only known to have one host: *Shigella flexneri* NK1925 P1-35, a plasmid-less NK1925 strain. Perhaps these two strains (DR11 and P1-35) coexist in the same area of the Sunderbans, and so this phage is able to infect this strain of a different species. This phage is thought to be responsible for the resistance of *B. cepacia* DR11 to co-trimoxazole, trimethoprim, and erythromycin. BcP15, when lysogenizing *S. flexneri* NK1925 cured of its

antibiotic resistance plasmids (P1-35R), appears to transmit this resistance to the NK1925, suggesting it is the prophage in DR11 responsible for the resistance.

1.5 Research objectives

As mentioned, there have been numerous phages characterized that are specific to the *B. cepacia* complex, and presumably there are even more that have been isolated and not yet characterized. Because using phage therapy as an alternative to antibiotics in the treatment of patients infected with strains of the *B. cepacia* complex is being considered, discovery of broad host range phages or extremely virulent phages is important. Initially, my research focused on isolating phages from environmental sources such as yellow and sweet corn soil, rhubarb soil, alfalfa soil, swamp mud, sea water, and rotting onions. As members of the *B. cepacia* complex have been isolated from soil, we would expect to find phages specific for the complex in soil as well. Our goal was to isolate and characterize phages able to infect a broad range of *B. cepacia* complex strains, specifically strains of the most clinically relevant species, *B. cenocepacia* and *B. multivorans*. Following isolation of a phage capable of infecting a range of clinically important strains, sequencing and annotation of the genome would allow for further characterization and determine utility for therapy use. As mentioned, a phage encoding virulence genes would not be a good candidate for use in phage therapy. Using the soft agar overlay method following an extraction step, only four phages were identified, MG1, MG3, MG4, and MG5, none of which were capable of infecting a broad range of *B. cepacia* complex strains.

Because an appropriate phage was not identified by initial research, the second objective of this research was to characterize phage KS10, isolated by Seed and Dennis (2005). To do this, the host range, genome sequence and organization, and putative gene functions of the transposable phage KS10 were determined and compared to known phages BcepMu and Mu. Analysis of the KS10 genome also allowed us to identify any potential virulence determinants encoded by this phage.

Chapter 2: Materials and Methods

2. MATERIALS AND METHODS

2.1 Bacterial strains, phages, plasmids, and growth conditions.

2.1.1 Burkholderia cepacia complex and Escherichia coli strains and plasmids

Strains of each of the *B. cepacia* complex species were used in this research and are from the *B. cepacia* complex experimental strain panel (Mahenthiralingam *et al.*, 2000; Coenye *et al.*, 2001). Additional isolates were obtained from the University of Alberta Hospitals (Pediatric/Adult) Cystic Fibrosis Clinic. All strains and isolates are shown in Table 2.1. Chemically competent *E. coli* DH5α cells were commercially obtained (Invitrogen, Carlsbad, CA) and used for phage DNA cloning experiments. pUC19 or pGEM7Z cloning vectors maintained in DH5α were used for sequencing experiments.

2.1.2 Growth Conditions

B. cepacia complex bacteria were grown in half-strength Luria-Bertani (LB) broth or solid media (Sambrook *et al.*, 1989). Growth of *B. cepacia* complex strains was carried out aerobically 30°C overnight, with shaking at approximately 250 rpm when grown in broth. LB solid medium supplemented with ampicillin (0.1g/L) for selection was used to grow competent *E. coli* DH5α. *E. coli* cells were grown aerobically overnight at 37°C with shaking at approximately 250 rpm. Phages were stored at 4°C in suspension media (SM) (50 mM Tris/HCl, pH 7.5, 100 mM NaCl, 10 mM $MgSO_4$, and 0.01% gelatin solution) (Sambrook *et al.*, 1989).

Table 2.1 Bacterial strains and isolates used in this study

Species	Strain	Source/Location	Reference or strain source
B. cepacia	LMG 18821	CF, Australia	Mahenthiralingam *et al.* (2000)
	ATCC 17759	Soil, Trinidad	Mahenthiralingam *et al.* (2000)
B. multivorans	C53393	CF, Canada	Mahenthiralingam *et al.* (2000)
	C3430	CF, Canada	CRRR
	C5274	CF, Canada	CRRR
	C5568	CF, Canada	CRRR
	ATCC 17616	Soil, United States	Mahenthiralingam *et al.* (2000)
	M1512	CF, Canada	UAHCFC
	M1865	CF, Canada	UAHCFC
	R810	CF, Canada	UAHCFC
	R1159	CF, Canada	UAHCFC
B. cenocepacia	715J	CF, Canada	McKevitt et al. (1989)
	J2315	CF-e, United Kingdom	Mahenthiralingam *et al.* (2000)
	K56-2	CF-e, Canada	Mahenthiralingam *et al.* (2000)
	C6433	CF, Canada	Mahenthiralingam *et al.* (2000)
	C1257	CF, Canada	CRRR[β]
	C5424	CF, Canada	Mahenthiralingam *et al.* (2000)
	C4455	CF, Canada	CRRR[β]
	Cep511	CF, Australia	Mahenthiralingam *et al.* (2000)
	PC184	CF-e, United States	Mahenthiralingam *et al.* (2000)
	R161	CF, Canada	UAHCFC
	R452	CF, Canada	UAHCFC
	R750	CF, Canada	UAHCFC
	R1284	CF, Canada	UAHCFC
	R1314	CF, Canada	UAHCFC
	R1434	CF, Canada	UAHCFC
	R1619	CF, Canada	UAHCFC
	R1882	CF, Canada	UAHCFC
	R1883	CF, Canada	UAHCFC
	R1884	CF, Canada	UAHCFC
	R2314	CF, Canada	UAHCFC
	S11528	CF, Canada	UAHCFC

(continued)

Table 2.1- *Continued*

Species	Strain/isolate	Source/location	Reference or strain source
B. stabilis	LMG 14294	CF, Belgium	Mahenthiralingam *et al.* (2000)
	LMG 18870	CF, Canada	Mahenthiralingam *et al.* (2000)
	R450	CF, Canada	UAHCFC
	R2140	CF, Canada	UAHCFC
	R2339	CF, Canada	UAHCFC
B. vietnamiensis	DB01	Soil, United States	Walsh and Ballou (1983)
	LMG 10929	Soil, Vietman	Mahenthiralingam *et al.* (2000)
	LMG 18835	CF, United States	Mahenthiralingam *et al.* (2000)
B. dolosa	LMG 18943	CF, United States	Coenye *et al.* (2001)
	LMG 21443	Rhizosphere, Senegal	Coenye *et al.* (2003)
B. ambifaria	LMG 19467	CF, Australia	Coenye *et al.* (2003)
	LMG 17828	Soil, United States	Mahenthiralingam *et al.* (2000)
	LMG 19182	Soil, United States	Mahenthiralingam *et al.* (2000)
B. anthina	LMG 16670	Rhizosphere, United Kingdom	Coenye *et al.* (2003)
B. pyrrocinia	LMG 14191	Soil, Japan	Coenye *et al.* (2003)
B. gladioli	R406	CF, Canada	UAHCFC
	R1879	CF, Canada	UAHCFC
Pseudomonas aeruginosa	R285	CF, Canada	UAHCFC
Herbaspirillum sp.	R740	CF, Canada	UAHCFC
Listeria monocytogenes	R1653	CF, Canada	UAHCFC
Pandoraea sp.	R1717	CF, Canada	UAHCFC

[α]Abbreviations: CF, cystic fibrosis; CF-e, strain that has spread epidemically among patients with CF; CRRR, from the Canadian *Burkholderia cepacia* complex Research and Referral Repository; UAHCFC, University of Alberta Hospital (Pediatrics/Adults) Cystic Fibrosis Clinic

2.1.3 Preparation of frozen bacterial stocks

E. coli DH5 clones were grown overnight on LB agar supplemented with 100 ug/mL ampicillin, then suspended in 1-2 mL of LB + 20% glycerol (vol/vol) using a 1 mL pipette tip and stored at -80°C.

2.2. Isolation of bacteriophages and host range testing

2.2.1 Isolation of phages from environmental sources

Soil samples were collected in Edmonton, Alberta and tested for the presence of phages specific to the *B. cepacia* complex using a modification of a phage enrichment technique. Rotting onions were also collected from a grocery store in Edmonton, Alberta and tested for the presence of phages using the same technique.

Approximately 10 g of soil or onion was measured into 50 mL conical tubes or 150 mL flasks. 10 mL of half-strength LB broth, 1 mL of *B. cepacia* complex culture, and an optional 1 mL of SM were added to the soil and placed on a shaker at 30°C overnight. Following centrifugation at 10,000 rcf for 10 min at 4°C, supernatant was filter sterilized (0.45 μm filter) into 15 mL conical tubes and a drop of chloroform ($CHCl_3$) was added. Four hundred microliters of extract was added to 100 μL of bacterial culture in a glass culture tube and incubated at room temperature for 20 minutes. Following incubation, 3 mL of molten top agar was added and the mixture was poured over a half-strength LB agar plate. After incubation overnight, plates were checked for the presence of plaques. In cases where no plaques were identified, 1 mL of bacterial culture was added to the extract and incubated again on the shaker overnight at 30°C. Following centrifugation at 4000 rcf for 15 min at 4°C,

supernatant was filter sterilized (0.5 μm filter) into a 15 mL conical tube and tested again for the presence of phages.

2.2.2 Preparation of high titer phage stock

Phages MG1, MG2, and MG4 plaques were picked from plates of infected LMG 19467 using a sterile glass Pasteur pipette and stored at 4°C. MG5 was picked from plates of infected K56-2, and KS10 was picked from uninduced lawns of K56-2 in the same way. High titer phage stocks were prepared by rocking plates containing confluently lysed bacteria with 3-4 mL of SM at 4°C for 2-6 hours. SM containing the phage was then filter sterilized and stored at 4°C.

2.2.3 Determination of phage host range

To determine the host range of isolated phages a high titer stock was tested against the 24 strains from the *B. cepacia* complex experimental strain panels (Mahenthiralingham *et al.*, 2000; Coenye *et al.*, 2003). One hundred milliliters of the high titer stock was added to 100 mL of each of the bacterial strains in triplicate and incubated at room temperature for 20 min as previously described. After incubation at 30°C overnight the plates were checked for plaque formation. Lytic activity was recorded as: (-) no plaques, (+) ≤20 plaques, (++) ≥20 plaques/plate, or (+++) confluent lysis.

2.2.4 KS10 production, host range, and distribution

To determine the titer of just one KS10 plaque in 1 ml of SM, the lysate was serially diluted in SM, and the number of plaque forming units (pfu) was determined using the soft agar overlay method. A high titer stock of KS10 was used to determine the host range as described previously. KS10 distribution within the *B. cepacia* complex was determined using a PCR assay and products were analyzed by agarose gel electrophoresis and purified for sequencing to ensure product was KS10 (methods explained in section 2.4.2).

2.2.5 Isolation of PC184 and LMG 19467 lysogens

KS10 was propagated on BCC strains PC184 and LMG 19467 using the soft agar overlay method. Turbid plaques were identified, picked using a 20-gauge needle, and placed in 1 mL of half-strength LB in an incubating shaker overnight at 30°C. Cultures were streaked for individual colonies that were then tested for their inability to support plaque formation by KS10. PCR using KS10-specific primers further confirmed the presence of a KS10 prophage (described in section 2.4.2).

2.3. DNA manipulation

2.3.1 Phage DNA Isolation

KS10 was propagated on host *B. ambifaria* LMG 19467 for DNA extraction. To obtain a large volume of phage DNA for cloning, 12 large (150 x 15 mL) half-strength LB agarose plates (1.3% half-strength LB agarose) of LMG 19467 confluently lysed by KS10 were covered with 10 mL SM. Plates were prepared by the soft agar overlay method using 400 µL phage stock + 400 µL bacterial culture

and 7 mL 0.6% ½ LB top agarose. Plates were placed on a rocker at 4°C for 4-8 hours. SM containing phage was then transferred to 50 mL conical tubes with 2% v/v $CHCl_3$ added and centrifuged at 10,000 rcf for 10 minutes at 4°C. Ten microliters of DNase I (Fermentas), 10 μL DNase I buffer (Fermentas), and 6 μL RNase A/T1 mix (Fermentas) was added to the supernatant and incubated at 37°C for 40 min. Following incubation lysate was centrifuged at 4,000 rcf for 10 min at 4°C. Supernatant was collected into a sterile 250 mL flask. NaCl was added to a final concentration of 1 M and stirred continuously at 4°C for 1 h. Polyethylene glycol (PEG) 8000 was slowly added to obtain a 10% w/v solution and stirred at 4°C for approximately 5 min then transferred to 50 mL conical tubes and allowed to stand at 4°C overnight (an additional overnight incubation may be required to increase DNA yield with certain phages). After incubation tubes were centrifuged at 10,000 rcf for 20 min at 4°C. Supernatant was removed and 1.6 mL SM for each 100 mL of supernatant removed was added to the pellet. Pellet + SM was incubated at 4°C for 2-4 h to soften pellet before mixing, and then transferred to a 15 mL conical tube. Protease (20 μL of 20 mg/mL proteinase K added to 800 mL resuspended pellet) was added and incubated at 37°C for 10 min. PEG was extracted by adding an equal volume of $CHCl_3$ and vortexing for 30 sec. Organic and aqueous phases were separated by centrifugation at 3,000 rcf for 15 min at 4°C. The aqueous phase was collected and EDTA added to a final concentration of 100 mM. Half volume of 6 M guanidine thiocyanate was added and phage DNA was recovered by using the Geneclean Turbo Kit (Q-BIOgene, Irvine, CA).

2.3.3 Isolation of chromosomal DNA

To test for the presence of a lysogenized KS10, genomic DNA was isolated from the *B. cepacia* complex using a previously described method (Ausubel *et al.*, 1999).

2.3.2 Gel electrophoresis of phage DNA

Purified KS10 DNA was digested using the restriction enzymes *Sph*I, *Eco*RI, and *Xho*I (Invitrogen Corp., Carlsbad, CA) in a 10 μL reaction. Following incubation at 37°C for 1-4 h, 1μL of loading buffer containing 0.25% bromophenol blue and 40% sucrose (Sambrooke *et al.*, 1989) was added to each sample. DNA samples were subjected to electrophoresis in 0.75% agarose, 1X TAE gels at 120 V. To estimate fragment size a 1 kb Plus DNA ladder (Invitrogen) was used as a molecular weight marker. Following gel eletrophoresis, gels were stained in ethidium bromide and visualized with ultraviolet light.

In host range experiments explained previously, PCR reactions were similarly suspended in loading buffer and subjected to electrophoresis as described above.

2.3.3 Purification of DNA fragments from agarose gel

The Geneclean II Kit (Q-BIOgene, Irvine, CA) was used to purify bands excised from agarose gels. Gel fragments were melted in NaI solution at 55°C and a DNA-binding silica matrix was added to bind the DNA from the NaI. The solution was centrifuged briefly in a microcentrifuge at maximum speed to pellet the silica-

DNA complex. The pellet was washed twice with ethanol and dried under a vacuum to remove ethanol. The silica-DNA complex was then suspended in TE (pH 8) to elute DNA. Following centrifugation the supernatant containing the DNA was removed and stored at 4°C until used.

2.3.4 Library Preparation

To create a chromosomal library, purified fragments of KS10 DNA were ligated into pUC19 or pGEM7Z. Ligation reactions were carried out in 15 μL volumes using T4 DNA ligase (Promega, Madison, WI) at 16°C (non-directional cloning) for 4-8 h, or at room temperature (directional cloning) for 2-4 h. Ligated plasmids were transformed into chemically competent *E. coli* DH5α subcloning efficiency (Invitrogen) according to product specifications. Plasmid DNA from *E. coli* was purified for sequencing using the QIAprep miniprep kit (Qiagen Inc., Mississauga, Ont.) and was used as specified by the manufacturer. When PCR products were used for sequencing, products were first cloned into the pCR2.1-TOPO vector (Invitrogen) and subsequently transformed into One Shot TOP10 chemically competent cells (Invitrogen).

2.3.5 Identification of KS10 integration sites

In an attempt to determine the integration sites of KS10 in the three isolated lysogens, *B. cenocepacia* strain C5424 and the seven *B. cenocepacia* clinical isolates, PCR using arbitrary primers (ARB6 and ARB2) and a specific KS10 primer

were used. In addition, the APA gene Gold genome walker kit (Bio S&T, Montreal) was used according to the manufacturer's directions.

2.4 Sequence assembly and annotation

2.4.1 DNA sequence analysis

Sequencing of the KS10 inserts was carried out using the DYEnamic ET kit (Amersham Biosciences, Piscataway, NJ). Sequencing reactions were carried out using 6 μL plasmid DNA (approximately 250 nanograms), 1 μL of primer (approximately 50 picograms), 2 μL of ET reagent, and 11 μL of sequencing buffer (20 mM Tris, pH 9.0, 5 mM MgCl2). The standard sequencing reaction program consisted of 35 cycles of a 20 sec denaturation step at 95°C, a 15 sec primer annealing step at 50°C, and an extension at 60°C for 1 min, followed by a 4°C hold. Reactions were stopped by adding 2 μL of sodium acetate/EDTA, and, following addition of 80 mL of 95% ethanol, tubes were placed at -20°C for 15 min to precipitate DNA. After centrifugation at 13,200 rpm for 10 min, ethanol was removed and tubes were dried and stored at -80°C. DNA was sequenced using the ABI 377 Gene Analyzer with the assistance of the Biological Sciences department's Molecular Biology Service Unit.

2.4.2 Polymerase chain reaction

The sequence of the left end of KS10, as it is in its host *B. cenocepacia* K56-2 was determined using PCR with a non-specific primer cep3 (5'-CGG CAA ACG GCG TGG GGC AG), that would anneal to the bacterial chromosome, and a primer

that would anneal to the end of the known KS10 left end, L1.1 (5'-CTG CCT GCC GTC ATC GCC CA). PCR was also used to fill in gaps in the KS10 sequence. Reactions were performed in 50 μL volumes in 0.2 mL tubes using the Eppendorf Mastercycler gradient thermocycler (Wesbury, NY). Reactions contained 5 μL of 5X PCR reaction buffer (Invitrogen, Carlsbad, CA), 1.5 μL of MgCl2, 1 μL of 10 mM dNTPs, 0.5 μL of forward and reverse primers, 0.5 μL of a chromosomal DNA preparation, and *Taq* polymerase. Milli-Q H_20 was added to a final volume of 50 μL. PCR programs typically consisted of a 2 min denaturation step at 94°C followed by 30 cycles of a 45 sec denaturation step at 94°C, a 30 sec annealing step at 50-65°C (usually 10-15 degrees lower than the melting temperature of the primers), and an extension period at 72°C for 1 min. An additional 2 min extension time at 72°C was used at the end of the final cycle.

In order for the KS10 distribution within the *B. cepacia* complex to be determined, a PCR reaction with oligonucleotide primers F3 (5'CCGATTCCCACATCACGATCC) and R3 (5'-TGCGGGCATTTCAGCTTTCG) was carried out. PCR was performed in 50 μl reactions containing 1 μl of each primer and 0.5 μl template DNA using *Taq*PCRx DNA Polymerase, Recombinant (Invitrogen). PCR was carried out as specified by the manufacturer's recommendations.

2.4.3 Computer-assisted sequence analysis

Sequencing data were edited using EditView 1.0.1 (Perkin Elmer). Clean sequences were assembled using AutoAssembler (Perkin Elmer). Once the complete

sequence was assembled, GeneMark and NCBI's ORF Finder programs were used to detect possible open reading frames (ORFs). When multiple start codons or OFRs were suggested, the presence of a potential ribosomal binding site (RBS) was used to help identify the most likely ORF. RBS finder (http://nbc11.biologie.uni-kl.de/framed/left/menu/auto/right/glimmer2.02) (Delcher *et al.*, 1999) was also used to identify possible ribosomal binding sites. Each identified ORF was used in conjunction with BLASTX (http://www.ncbi.nlm.nih.gov/BLAST) analysis to assign a putative function to each. When BLASTX revealed no significant matches a PSI-BLAST (NCBI) was also used. Genome maps were constructed using GenVision software (DNASTAR, Inc., Madison, Wisconsin).

2.5 Transmission electron microscopy (TEM)

KS10 was obtained from an overnight culture of K56-2 (OD_{600} approximately 2.00). The culture was centrifuged at 10,000 rcf for 2 minutes and filter sterilized using 0.45 μm filters. Filtrate was spotted onto copper grids and stained with 2% phosphotungstic acid. Micrographs were obtained using a Philips/FEI (Morgagni) Transmission Electron Microscope with CCD camera in the Biological Science Department's Microscopy Unit.

2.6 Testing virulence of *B. ambifaria* LMG 19467 lysogens

To test the virulence of the newly created LMG 19467 lysogens, we followed the procedure outlined by Seed and Dennis (2008) involving the use of *Galleria mellonella* larvae, or wax worms. Briefly, overnight cultures of wild type LMG

19467, LMG 19467 lysogen 1 (lys1) and lysogen 2 (lys2) were obtained by shaking overnight at 30°C. One milliliter of each culture was centrifuged at 10,000 rcf for 2 min and pellet was resuspended in 1 mL of 10 mM $MgSO_4$ supplemented with 1.2 mg/mL of ampicillin. Ten worms for each group (wild-type, lys1, and lys2) were injected in the leftmost hind leg with 5 μL culture diluted 1:5 with MgSo4. Another 10 control worms were injected with 5 μL 10 mM $MgSO_4$ + amp. Worms were placed at 30°C and mortality was checked at 24 and 48 h post infection. Injections were performed with a 10 μL Hamilton syringe. This experiment was repeated again with 20 worms in each group, using 1:10 and 1:100 dilutions for each group. Again, mortality was recorded after 24 and 48 h.

Chapter 3: Results

3. RESULTS

3.1 Identification of Four *B. cepacia* complex specific phages

Using various types of soil/plant rhizosphere, we wanted to isolate numerous phages capable of infecting strains of the *B. cepacia* complex. Ideally, phages with the broadest host ranges would be more useful in therapy and could be used in an animal model to rescue animals from infection with these bacterial strains. Four lytic phages, MG1, MG3, MG4, and MG5 were isolated from soil as described above. Phages were isolated from corn rhizosphere, rotting yellow onion, a swamp area in Nova Scotia, and *Taxus sp.* rhizosphere, respectively. The host range was determined for all phages by testing for plaque forming ability on an experimental strain panel of 24 *B. cepacia* complex strains. MG1, MG3, and MG4 phages were capable of infecting LMG 19467, while MG3 and MG4 were also capable of infecting *B. cenocepacia* PC184 and *B. stabilis* LMG 18870. MG5 was only capable of infecting *B. cenocepacia* K56-2. However, restriction fragment length polymorphism (RFLP) analysis of MG1, MG3, and MG4 genomes, along with partial sequencing, identified these phages as previously identified KS10 (Figure 3.1). Although MG1 appears to have a slightly different host range, it is likely a host range mutant of KS10. RFLP analysis of the MG5 genome revealed this phage to be the previously identified BcepMu (data not shown). Therefore, no new *B. cepacia* complex specific phages were isolated in this research. It is unknown if MG1, MG3, and MG4 were isolated from these environmental samples, or if phage stocks or media was contaminated with phage KS10. Because plaques were sometimes

identified on the top layer of LMG 19467 controls, it is possible that these isolated phages were actually contaminants.

3.2 Characterization of phage KS10

KS10 was originally identified as small pinpoint plaques on lawns of uninduced *B. cenocepacia* strain K56-2. It was not necessary to use mitomycin C or exposure to UV light to initiate induction because KS10 in K56-2 appears to spontaneously switch to a lytic lifecycle at a high frequency. When KS10 lysate is plated with another strain, *B. ambifaria* strain LMG 19467, plaques are slightly larger in size. KS10 is quite virulent against this strain; one plaque of KS10 in 1mL of SM will produce approximately 2.74×10^5 pfu when plated with strain LMG 19467. KS10 particles are very stable over time, with titers staying quite high even after storage in SM at 4°C for a year.

An electron micrograph (EM) of KS10 virions negatively stained with 2% phosphotungstic acid shows an icosahedral head and long tail that is expected to be contractile although contraction is not evident from this micrograph (Figure 3.2). These characteristics are typical of the *Myoviridae* family of phages. The EM also revealed the average head size to be approximately 80 nm with a tail length of approximately 140 nm.

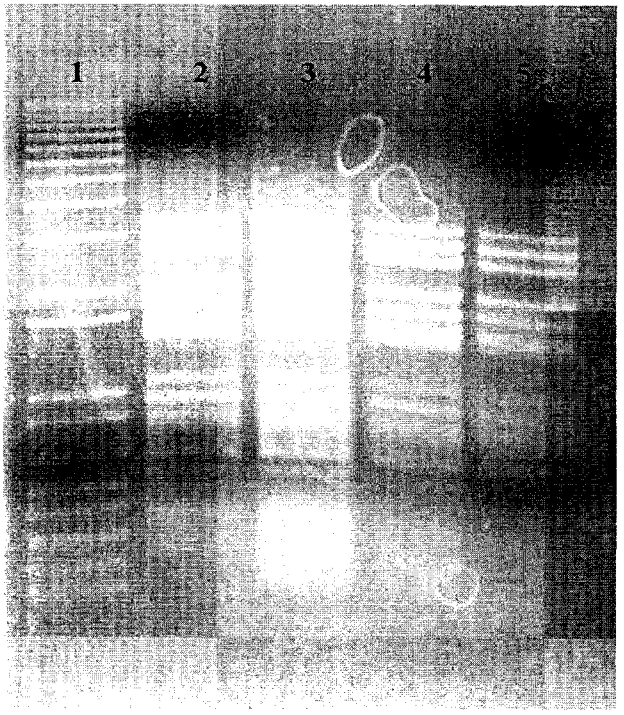


Figure 3.1 Restriction fragment length polymorphisms of the genomes of bacteriophages KS10, MG1, MG3, and MG4. The 0.8% agarose electrophoretic gel shows a 1 kb+ DNA ladder in lane 1 with the *Pst*I-digested genomes of KS10, MG1, MG3, and MG4 lanes 2–5, respectively. This RFLP pattern indicates that MG1, MG3, and MG4 re-isolations of KS10.

3.3 Host range of KS10

Using the plaque assay, 23 *B. cepacia* complex strains were tested for their sensitivity to KS10. Only three of the tested strains (representing three different species) were found to support plaque formation with KS10: *B. ambifaria* LMG 19467, *B. cenocepacia* PC184, and *B. stabilis* LMG 18870. Lysate obtained from filter sterilized uninduced overnight cultures of both K56-2 and J2315 was able to form plaques on LMG 19467, PC184, and LMG 18870. Since LMG 19467 (but not PC184) is also a host for BcepMu, lysate obtained from an overnight culture of J2315 will also form BcepMu plaques on a lawn of LMG 19467, so there is a noticeable increase in the number of plaques on those plates, suggesting that both KS10 and BcepMu are present on the plate.

To determine the distribution of KS10 within the genome of other *B. cepacia* complex species/strains, PCR using KS10 specific primers was carried out. *B. cenocepacia* K56-2 and J2315 chromosomes were used as positive controls since they are known to be lysogenized by KS10. *B. ambifaria* LMG 19467 chromosome was used as a negative control due to the ability of KS10 to form plaques on this strain, and therefore the majority of the cells should not be lysogenized. Seven strains of *B. cenocepacia* and one strain from each of the other species in the *B. cepacia* complex experimental strain panel (Mahenthiralingham *et al.*, 2002; Coenye *et al.*, 2003) were tested. We found only *B. cenocepacia* C5424 to also be lysogenized with KS10. This strain was previously shown to harbour BcepMu as well (Summer *et al.*, 2004). Additionally, 27 clinical isolates obtained from University of Alberta Hospitals

(Pediatric/Adult) Cystic Fibrosis Clinic were also tested, revealing that KS10 is a prophage in seven of these isolates. All seven isolates were characterized as *B. cenocepacia* based on their *fur* gene sequence (Lynch and Dennis, 2007), suggesting that KS10 is a lysogenic phage specific to *B. cenocepacia* (Figure 3.4; Table 3.5).

3.4 Determination of KS10 genome sequence and annotation

KS10 was sequenced using a shotgun cloning and sequencing approach. Phage DNA was digested with restriction enzymes, ligated into pUC19 or pGEM7Z, and transformed into chemically competent DH5α cells. Inserts larger than 10 kb were subcloned and sequenced. All inserts were sequenced at least twice, and PCR or primer walking was used to fill any gaps in the sequence. Initial restriction digests suggested an estimated genome size of about 35 kb (Figure 3.4). Approximately 249 runs with an average read-length of 680 bp each were assembled to give greater than 4-fold genome coverage, resulting in a single contig that was 37,635 bp in length. This DNA sequence was approximately 63% GC content, which is slightly higher than that of phage Mu, but similar to BcepMu. To confirm that the recently sequenced *B. cenocepacia* J2315 was lysogenized by KS10, a BLAST analysis was carried out at the Wellcome Trust Sanger Institute site (http://www.sanger.ac.uk/Projects/B_cepacia/blast_server.shtml) to compare the KS10 sequence with the deposited J2315 sequence.

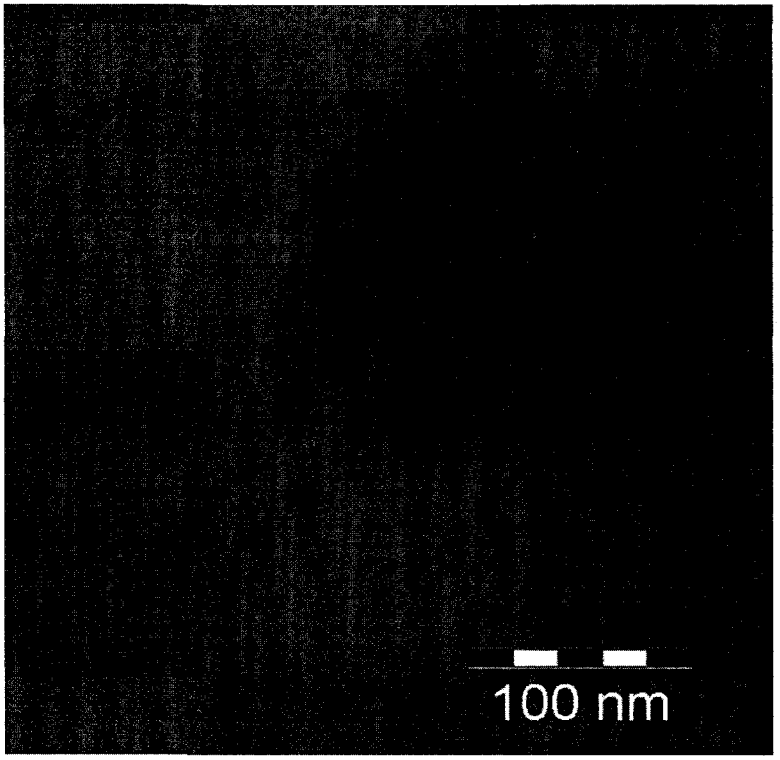


Figure 3.2 Electron micrograph showing morphology of KS10 particles in lysate stored at 4°C for approximately one week prior to imaging. Phage lysate was negatively stained with 2% phosphotungstic acid. Image was viewed at 110,000-fold magnification using a Philips/FEI (Morgagni) Transmission Electron Microscope with CCD camera in the University of Alberta Biological Science Department's Microscopy Unit.

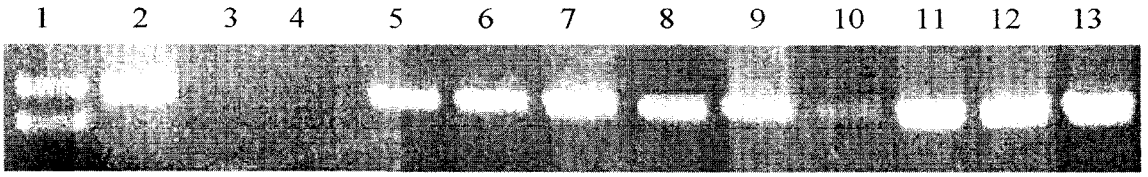


Figure 3.3 KS10 lysogens as determined by PCR of B. cepacia complex DNA with primers F3 and R3. The 0.8% agarose electrophoretic gel shows the 1.65 kb (bottom) and 2 kb (top) fragments from a 1 kb + DNA ladder in lane 1. Lane 2 contains the positive control, *B. cenocepacia* K56-2, lanes 3 and 4 contain *B. cenocepacia* PC184 and *B. ambifaria* LMG 19467, respectively, as negative controls. Shown in lanes 5-13 are *B. cenocepacia* strains/isolates J2315, C5424, R1882, R1883, R1884, S11528, R1434, R2314, and R750.

This BLAST analysis revealed that KS10 is a prophage in J2315. Analysis of the host DNA flanking the phage sequence from the Sanger site implies that in strain J2315 the prophage is located in chromosome 1 (bp 1,766,551 to bp 1,728,918), on the complementary strand, and has inserted in an oxidoreductase gene in the Gfo/Idh/MocA around amino acid 235 (Figure 3.5). The final KS10 sequence was also compared with the sequence from J2315 on the Sanger site using AutoAssembler (Perkin Elmer), showing a perfect overlap in sequences.

The far right end of the KS10 genome was determined from the sequencing of three clones that contained the phage/*B. cenocepacia* K56-2 DNA junction at two different insertion sites. One of these insertion sites was a gene encoding a transcriptional regulator in the GntR family, while the other insertion site was a region showing no homology to gene in the Blast database. This shows random insertion within the host genome, a feature characteristic of transposable phages. The left end was obtained using PCR of K56-2 genomic DNA since no shotgun clones containing a phage/host insert were isolated for this end. Forward primers were designed to the chromosome for the region upstream of the gene interrupted by KS10 prophage at the right end insertion site, and reverse primers were designed from the left end of completed KS10 sequence. Regions upstream of the ORFs were examined to determine potential ribosome binding sites found between 4 and 14 bases upstream of the start codon. RBS candidates were those closest to the consensus sequence TATAAT (Harley and Reynolds, 1987), though it is unlikely that every gene will have a RBS.

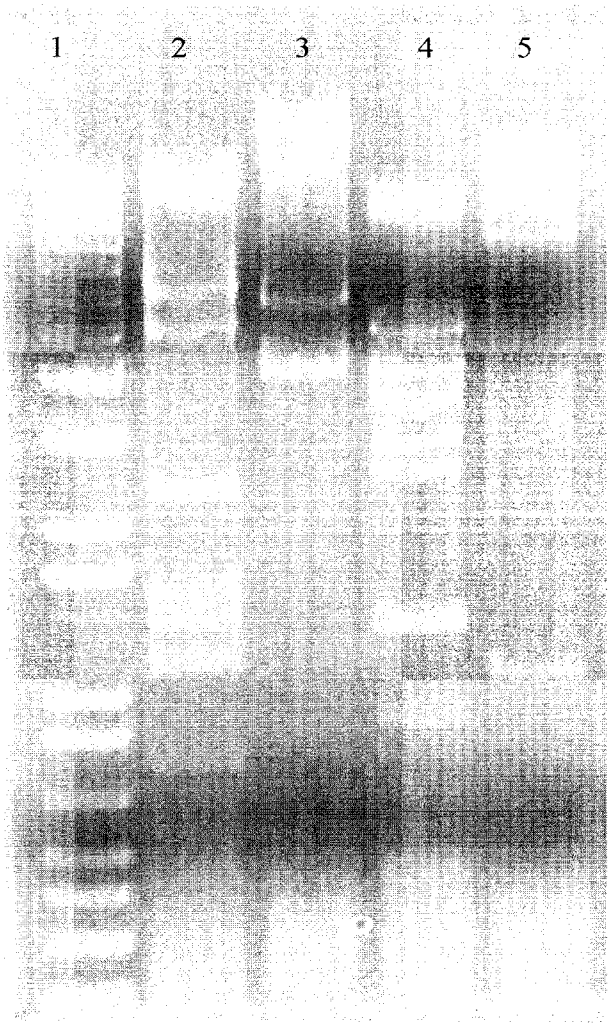
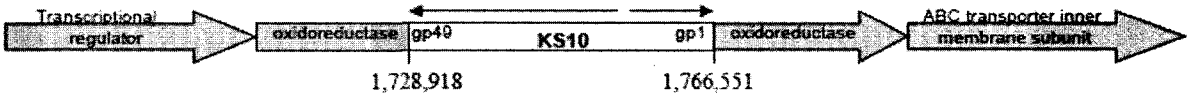


Figure 3.4. Restriction fragment length polymorphism analysis of the genome of bacteriophage KS10. The 0.8% agarose electrophoretic gel shows a 1 kb+ ladder in lane 1 and KS10 DNA digested with restriction enzymes *Eco*RI, *Hind*III, *Sph*I, and *Xho*I in lanes 2-4, respectively.

(a)



(b)

Host/Phage left end DNA junction in J2315
1,766,571-aagagctgcgcgatcccgccGAGAGGGGGGACGTTTAGTT-1,766,532

Phage right end/Host DNA junction in J2315
1,728,937-AAATTAAACGGCGCGCTACTcgccgttcgcatgcacgctc-1,728,897

Figure 3.5 Phage/Host DNA junction and putative transposase binding sites. (a) KS10 has integrated into an oxidoreductase gene of *B. cenocepacia* J2315. Within the J2315 genome, bases 1,728,819 through 1,766,551 are KS10 prophage sequence. Direction of transcription of surrounding genes is indicated with thick arrows, while the direction of transcription of KS10 genes is indicated with thin arrows. The diagram is not drawn to scale. (b) Uppercase letters indicate KS10 sequence while lowercase represent *B. cenocepacia* J2315 host DNA. Phage right end is on the left in the diagram due to its orientation in the J2315 genome, while the left end of KS10 is on the right.

RBS finder (Delcher *et al.*, 1999) was employed to predict RBS within the KS10 genome. However, only three RBS were predicted which corresponded to the ORFs predicted for KS10. A total of 49 putative genes were identified, with just under half encoding hypothetical proteins, having no known function. KS10 genes 11 and 12 have significantly less GC content than the overall genome (Table 3.1), suggesting that these genes may have been acquired by KS10 more recently. All putative genes utilized AUG as their translational start codon, except for two that possessed UUG or GUG start codon.

BLAST analysis (NCBI) (Altschul *et al.*, 1997) revealed that approximately 18% of KS10 proteins are homolgous to proteins of BcepMu, and another 18% show homology to Mu proteins. Twelve percent of KS10 proteins show homology to proteins from *Ralstonia solanacearum* UW551. Although they have not been annotated as such in the UW551 GenBank entry, these are likely proteins of a prophage within this strain of *R. solanacearum*, not bacterial proteins. The highest percent identity that KS10 protein sequences show to an orthologous protein is 67%, while the majority of KS10 proteins show only moderate or low identity and similarity to other phage proteins (Table 3.2).

The genes identified as hypothetical proteins using the standard BLAST searches were subjected to analysis with PSI-BLAST. This program uses the predicted amino acid sequence of each gene and detects putative conserved domains within the sequence. Using this program KS10 gp4 and gp5 were identified as being the large and small terminase subunits, respectively. These terminases are ATP-binding proteins that are involved in packaging phage DNA

into the procapsid (Morgan *et al.*, 2002). Although these proteins were not identified in a standard BLAST search, it is expected that this identification is correct, as they are necessary for DNA packaging, and no other terminase homologues have been identified in the KS10 genome. In addition, these genes are encoded after the virion morphogenesis and portal genes, an order that is conserved in many phage genomes (Summer *et al.*, 2004; Morgan *et al.*, 2002). Also using a PSI-BLAST analysis, KS10 gp39 was found to show homology to FluMu41. This gene is thought to encode a protein analogous to λ G, a tail assembly chaperone (Levin *et al.*, 1992).

Using the GTOP sequence homology search (http://spock.genes.nig.ac.jp/~genome/adseqsch.html) to compare KS10 gp48 to all viruses in the database, the amino acid sequence of KS10 gp48 showed homology to other phage proteins annotated as being involved in host specificity and putative tail-host specificity. Though this homology was fairly low (25%), it suggests that this 742 amino acid protein is possibly a tail fiber protein involved in recognizing the phage receptor on the host cell, and not part of an ABC-type phosphate transport system.

Higher order bioinformatic analyses were performed on the KS10 amino acid sequence using programs for protein domain identification and functional prediction. These analyses determined the presence (if any) of signal peptides, transmembrane helices or barrels, leucine zipper domains, as well as identifying the isoelectric point, amino acid frequency, and polarity of the proteins. In this case, these analyses confirmed the results obtained using BLAST analyses, but

provided little additional functional KS10 protein information (Table 3.2). Such information may be useful to confirm the identity of putative proteins, but because there are frequently many unknown genes within all phage genomes, it is difficult to make an assumption regarding what these unknown genes may encode. Proteins containing a signal peptide will be released from the host cell, while proteins with a transmembrane domain will span the membrane in this region. Leucine zipper proteins are a category of DNA binding proteins. A protein capable of forming a leucine zipper will contain a pattern with leucine repeated every seven amino acids (Landschulz *et al.*, 1988). A protein with a high percentage of cysteine residues may contain multiple disulfide bonds. It has been shown previously that the transcriptional regulator, OxyR, as well as the heat shock protein, Hsp33, are activated during oxidative stress through the state of disulphide bonding (Zheng *et al.*, 1998; Jakob *et al.*, 1999) So, it is possible that unknown proteins with a higher percentage of cysteine residues may be chaperone proteins or transcriptional regulators. Another way to identify certain proteins is based on amino acid frequency. The amino acid sequence of the Rz1 phage protein, for example, is proline rich. Amino acid frequency may also be useful if experiments are performed in which a protease will be used. Thus, knowing additional information about each putative protein may eventually help to identify a putative function.

Potential virulence factor genes were not identified within the genome of KS10. However, there remain several genes encoding hypothetical proteins

within the KS10 genome that cannot be excluded as potential virulence factors until their functions are determined.

3.5 Organization of KS10 genome

The organization of KS10 appears to be somewhat scrambled when compared to other phages. Mu and Mu-like phages, as is the case with most related phages, are usually genetic mosaics of each other and are often arranged in modules so that genes encoding proteins that interact, such as the phage capsid and tail, will not be separated by nonhomologous events (Hendrix, 2003). The organization of and direction of transcription in KS10, however, allows the genes responsible for integration and transcription regulation to interrupt the head assembly module. This is uncommon for a Mu-like phage, as Mu phages generally have genes encoding proteins involved in head assembly in the middle of the genome, in the late region, which is usually more conserved than the early and middle regions (Morgan *et al.*, 2002). Unlike other sequenced phages, the first five gene products of KS10 are thought to be involved in head assembly. After these genes in the KS10 genome are a number of genes involved in host cell lysis, followed by genes for transposition/integration, followed by more genes involved in head assembly in the middle of the genome, as expected, and finally genes involved in tail formation at the right end. To the best of our knowledge this organization (separation of the capsid module) is unique to KS10.

When the KS10 genome is compared with the genome of Mu, the first approximately 17 kb of KS10 genome is the most varied, and appears to be

flipped (Figure 3.6), with genes responsible for host cell lysis remaining between integration and head assembly genes. It is unknown why this phage, found in multiple *B. cenocepacia* genomes, would have its genome arranged this way. This rearrangement is also evident using the ACT tool to compare the DNA sequence of KS10 to *E.coli* Mu and BcepMu (Figure 3.7). This comparison demonstrates that KS10 shows more identity to BcepMu in the left half of its genome, although largely rearranged, but the right end of the genome shows more identity to Mu.

Table 3.1 Phage KS10 putative genes and homologues

Gene product	Coding	Strand	Possible RBS* and Start Codon	AA	Putative Function	Alignment Region	% Identity	% GC	Significant Matches to proteins in NCBI's GenBank
1	205-1437	-	AGGCGcgtaaATG	410	virion morphogenesis	248(16-263)	33%	63.26	BcepMu gp30/ MuF
2	1438-1698	-	AAAGGAGccaacaATG	86	hyp. protein			61.69	no sig. match
3	1743-3278	-	GGAAG//36bp//cgtttATG	511	portal	354(35-374)	32%	62.57	BcepMu gp29
4	3268-4884	-	AAGAGGcctgatccacgATG	538	terminase large subunit	520(4-521)	55%	58.94	D3112 p26
5	4891-5388	-	AAGGGttgacgcATG	165	terminase small subunit	164(1-164)	37%	59.24	D3112 p24
6	5392-5682	-	AGGGTATccgcgATG	96	hyp. protein	-	-	59.45	no sig. match
7	5679-5903	-	GGATAacgATG	74	dksA/traR C4-type zinc finger	74(1-72)	45%	60.89	Bacteriophage L-413C orf82
8	5893-6186	-	GAGGAtgcccgtcATG	102	hyp. protein			65.99	no sig. match
9	6466-7137	-	GTGGGcctcgcagcATG	223	endolysin	186(20-203)	54%	65.03	BcepMu gp22
10	7134-7535	-	GGGTGccgccgtcgcgaATG	133	holin	116(8-113)	39%	58.96	BcepMu gp21
11	7650-8258	-	GGGAAGcgcgaaATG	202	hyp. protein	177(53-211)	26%	53.53	*Pseudomonas entomophila* L48
12	8251-8784	-	GGGAATAtgtaaagcATG	177	hyp. protein			54.87	no sig. match
13	8839-9309	-	AAGAGGccgaccATG	156	phage repressor	137(2-127)	36%	56.69	BcepMu gp17
14	9391-9582	+	GGAGcaaatATG	63	DNA binding protein	62(5-65)/ 57(1-57)	33%/52 %	58.85	BcepMu gp16/*Ralstonia solanacearum* UW551
15	9579-10631	+	GGGGAGGtggATG	350	conserved protein	141(49186)/ 127(77-203)	28%	64.71	BcepMu gp10/*Ralstonia solanacearum* UW551
16	10671-12296	+	AGGTGcgatATG	541	transposase	539(24-559)	36%	65.19	*Ralstonia solanacearum* UW551
17	12306-13298	+	GAATAAGGAGtgaccATG	330	transposition protein	309(7-311)	42%	64.15	*Ralstonia solanacearum* UW551
18	13306-13494	+	GTGAGGccatcATG	62	hyp. protein	-	-	66.14	no sig. match
19	13491-13796	+	GAGGTGAcggcATG	101	hyp. protein	-	-	69.28	no sig. match
20	13859-14371	+	TGGccgacgacagcATG	170	hyp. protein	172(55-187)	60%	67.25	*Ralstonia solanacearum* UW551
21	14368-14994	+	AAGGAAcccatcATG	208	conserved protein	200(9-210)/ 199(6-204)	72%/55 %	63	*Ralstonia solanacearum* UW551/BcepMu gp05
22	15005-15397	+	GAGGGGctggccATG	130	hyp. protein	-	-	66.16	no sig. match *(Continued)*

Table 3.1 - *Continued*

Gene product	Coding	Strand	RBS and Start Codon	AA	Putative Function	Alignment Region	% Identity	% GC	Significant Matches to proteins in NCBI's GenBank
23	15460-15732	+	AGGAGAAAcaccctcATG	90	DNA binding protein Hu-beta	90(1-90)	57%	64.47	*Bordetella pertussis*
24	15809-16642	+	GAGGAAccccaaaATG	277	hyp. protein	117(51167)/ 58(344-401)	53%/ 67%	62.83	*Burkholderia thailandensis* E264 gp38
25	16806-17126	+	GGGGAGTGAcactgtgATG	106	hyp. protein	111(9-119)	37%	64.49	*Burkholderia pseudomallei* K96243
26	17128-17568	+	GGAGGcccgctgacATG	146	modulation of host genes?	129(9-135)	30%	64.17	*Escherichia coli* B7A/ Mu gp16
27	17565-17963	+	GAAACGAccgcATG	132	middle operon regulator (Mor)**	106(14-118)	31%	64.91	*Pseudomonas entomophila* L48
28	18120-19322	+	AGAGAAccatTTG	400	protease	330(10-322)	33%	65.59	BcepMu gp32
29	18876-19322	+	AGGAagccATG	148	scaffold	-	-	67.11	BcepMu gp33
30	19368-19724	+	AGAGGATtcacATG	118	conserved hyp. protein	80(32-109)	42%	71.71	BcepMu gp35
31	19775-20722	+	GGAGctatccATG	315	major head subunit	311(4-307)	47%	64.87	Bacteriophage B3 gp34
32	20797-21186	+	AAGAGAGatcATG	129	hyp. protein	-	-	66.92	no sig. match
33	21183-21686	+	GAAGGGccgcaaATG	167	hyp. protein	-	-	65.48	no sig. match
34	21683-22114	+	GGATAGGTAcgggaaATG	143	virion morphogenesis	121(7-124)	36%	60.19	*Escherichia coli* 53638/ MuG (gp31)
35	22114-22716	+	GGTGAGGATGAtgtgATG	200	phage-related conserved hyp. protein	165(1-164)	24%	61.36	*Burkholderia vietnamiensis* G4
36	22700-22978	+	ACGGG//25bp//gatATG	92	hyp. protein	-	-	64.52	no sig. match
37	23022-24500	+	AAGGGAcattcgacATG	492	tail sheath protein	474(12-481)	34%	63.15	MuSo2/ MuL (gp39)
38	24546-24917	+	AAGGGAGTGaaacATG	123	hyp. protein	-	-	62.9	no sig. match
39	25000-25554	+	TGAGATTcccaccATG	184	tail assembly chaperone	-	-	62.88	FluMu gp41
40	25603-28038	+	GAGGAAGAgacgATG	811	tail tape measure (TP109 fam.)	484(3-458)	38%	64.2	*Burkholderia vietnamiensis* G4

Table 3.1 - *Continued*

Gene product	Coding	Strand	RBS and Start Codon	AA	Putative Function	Alignment Region	% Identity	% GC	Significant Matches to proteins in NCBI's GenBank
41	28038-29408	+	GGAGGAAcgaactgATG	456	DNA circulation protein	453(3-433)	24%	65.43	*Escherichia coli* B7A/ MuN (gp43)
42	29414-30571	+	TGAcccctATG	385	tail protein	352(96-337)	33%	63.04	*Polaromonas* sp. JS666/ MuP (gp44)
43	30571-31092	+	GGAGcaaactgATG	173	baseplate assembly	137(38-167)	29%	61.88	*Escherichia.coli* 53638/ MuQ (gp45)
44	31177-31758	+	AGGccatcATG	193	tail protein	99(25-112)	47%	64.6	MuV (gp46)
45	31755-32876	+	GGAAAAcatcATG	476	tail protein	306(24-325)	36%	65.86	*Pseudomonas entomophila* L48/ MuW (gp47)
46	32879-33478	+	AGGAGTgaccGTG	173	tail protein	197(5-192)	26%	65.17	*Desulfovibrio vulgaris* subsp. vulgaris str. Hildenborough
47	33478-34455	+	GGAcatcgactgATG	325	tail collar protein	229(310-534)	42%	62.47	*Burkholderia multivorans* ATCC 17616
48	34463-36691	+	TGAGGcacgcATG	742	ABC-type phosphate transport sys.	459(40-493)	46%	60	*Pseudomonas stutzeri* A1501
49	36711-37475	+	GAGGTAcaaATG	254	hyp. protein	-	-	59.48	no sig. match

Abbreviations: RBS, ribosome-binding site; AA, amino acid; hyp, hypothetical; sig,, significant; subsp, subspecies;
*Putative RBS binding sites for gp1, gp4, and gp37 were identified using RBS finder.
**Function determined using PHYRE database

Table 3.2 Additional properties of KS10 proteins gp1-gp49

GP#	Folding rate	SP	TM helix	TM Barrel	Leucine Zipper	MW (Da)	Cysteine (%)	Theoretical pI	Negatively charged residues (Asp + Glu)	Positively charged residues (Arg + Lys)	Protein stability (II)	AA Frequency	
												A-10.49	M-3.17
												C-0.73	N-3.41
												D-5.85	P-6.83
												E-5.12	Q-4.39
												F-1.71	R-7.80
												G-7.56	S-5.37
												H-1.46	T-6.34
												I-2.20	V-10.49
											Unstable	K-2.68	W-3.41
1	28.2/sec	None	None	None	None	45441	0.7	6.35	45	43	(42.71)	L-7.56	Y-3.41
												A-13.95	M-3.49
												C-1.16	N-1.16
												D-9.30	P-4.65
												E-4.65	Q-3.49
												F-1.16	R-4.65
												G-4.65	S-8.14
												H-2.33	T-6.98
												I-4.65	V-4.65
											Unstable	K-2.33	W-0.00
2	-1.9/sec	None	None	None	None	9099	1.2	4.5	12	6	(45.32)	L-18.60	Y-0.00
												A-14.68	M-3.91
												C-0.78	N-2.94
												D-7.63	P-4.70
												E-6.85	Q-5.09
												F-4.11	R-6.85
												G-6.07	S-5.28
												H-0.98	T-3.72
												I-5.09	V-7.63
											Unstable	K-2.35	W-2.35
3	5.07/sec	None	None	None	None	56387	0.8	4.64	74	47	(41.13)	L-6.85	Y-2.15
												A-7.81	M-3.35
												C-1.67	N-3.53
												D-7.62	P-5.20
												E-5.76	Q-4.46
												F-4.83	R-6.88
												G-5.76	S-5.02
												H-3.16	T-2.97
												I-5.58	V-7.06
											Unstable	K-6.13	W-2.04
4	-6.53/sec	None	None	None	None	61531	1.7	6.8	72	70	(42.15)	L-8.74	Y-2.42

Table 3.2 - *Continued*

GP #	Folding rate	SP	TM helix	TM barrel	Leucine zipper	MW (Da)	Cysteine (%)	Theoretical pI	Negatively charged residues (Asp + Glu)	Positively charged residues (Arg + Lys)	Protein stability (II)	AA Frequency	
												A-7.81	M-3.35
												C-1.67	N-3.53
												D-7.62	P-5.20
												E-5.76	Q-4.46
												F-4.83	R-6.88
												G-5.76	S-5.02
												H-3.16	T-2.97
												I-5.58	V-7.06
											Unstable	K-6.13	W-2.04
5	-6.53/sec	None	None	None	None	61531	1.7	6.8	72	70	(42.15)	L- 8.74	Y-2.42
												A-11.52	M-4.24
												C-1.21	N-4.85
												D-6.06	P-2.42
												E-5.45	Q-7.27
												F-3.64	R-6.67
												G-2.42	S-7.27
												H-1.21	T-4.85
												I-3.64	V-6.67
											Unstable	K-8.48	W-1.21
6	-9.43/sec	None	None	None	None	18826	1.2	9.37	19	25	(43.98)	L-7.88	Y-3.03
												A-8.33	M-4.17
												C-0.00	N-4.17
												D-8.33	P-0.00
												E-5.21	Q-6.25
												F-1.04	R-6.25
												G-14.58	S-1.04
												H-2.08	T-7.29
												I-4.17	V-7.29
											Stable	K-6.25	W-0.00
7	26.1/sec	None	None	None	1 (39-70)	10292	3	6.06	13	12	(7.65)	L-12.50	Y-1.04
												A-13.51	M-4.05
												C-8.11	N-1.35
												D-1.35	P-5.41
												E-14.86	Q-9.46
												F-2.70	R-13.51
												G-6.76	S-0.00
												H-1.35	T-4.05
												I-4.05	V-4.05
											Unstable	K-0.00	W-0.00
8	-33.8/sec	None	None	None	None	8448	0.6	5.38	12	10	(83.37)	L-2.70	Y-2.70

Table 3.2 - *Continued*

GP #	Folding rate	SP	TM helix	TM barrel	Leucine zipper	MW (Da)	Cysteine (%)	Theoretical pI	Negatively charged residues (Asp + Glu)	Positively charged residues (Arg + Lys)	Protein stability (II)	AA Frequency	
												A-14.35	M-1.79
												C-1.79	N-3.59
												D-4.93	P-6.28
												E-2.69	Q-4.48
												F-1.35	R-8.52
												G-7.62	S-5.38
												H-1.79	T-3.59
												I-5.38	V-5.83
											Unstable	K-3.14	W-2.69
9	-7.79/sec	1-45	None	None	None	24501	1.8	9.59	17	26	(44.50)	L-10.31	Y-4.48
												A-12.03	M-3.01
												C-1.50	N-0.00
												D-3.76	P-5.26
												E-5.26	Q-1.50
												F-3.76	R-6.02
												G-5.26	S-6.77
												H-3.01	T-3.76
												I-6.02	V-9.77
											Unstable	K-3.01	W-2.26
10	-6.76/sec	None	20-37	None	None	14706	1.5	6.9	12	12	(47.80)	L-15.04	Y-3.01
												A-9.90	M-3.47
												C-2.97	N-4.95
												D-6.44	P-5.45
												E-6.44	Q-6.93
												F-2.48	R-6.44
												G-3.96	S-10.40
												H-0.50	T-4.46
												I-4.95	V-2.48
											Unstable	K-5.94	W-1.49
11	4.33/sec	None	None	None	None	22504	3	5.95	26	25	(57.76)	L-8.91	Y-1.49
												A-14.69	M-5.08
												C-0.56	N-5.08
												D-2.82	P-2.26
												E-7.34	Q-3.39
												F-3.39	R-7.34
												G-5.08	S-6.78
												H-1.13	T-6.78
												I-3.95	V-7.91
											Stable	K-3.39	W-3.39
12	-10.9/sec	1-42	None	None	None	19638	0.6	7.91	18	19	(22.40)	L-8.47	Y-1.13

Table 3.2 - *Continued*

GP #	Folding rate	SP	TM helix	TM barrel	Leucine zipper	MW (Da)	Cysteine (%)	Theoretical pI	Negatively charged residues (Asp + Glu)	Positively charged residues (Arg + Lys)	Protein stability (II)	AA Frequency	
												A-13.46	M-3.21
												C-0.00	N-4.49
												D-4.49	P-3.21
												E-3.85	Q-5.77
												F-2.56	R-6.41
												G-10.26	S-7.69
												H-1.92	T-8.33
												I-3.21	V-11.54
											Stable	K-3.85	W-0.00
13	6.01/sec	1-25	None	None	None	16306	0	9.45	13	16	(19.54)	L-3.85	Y-1.92
												A-9.52	M-3.17
												C-0.00	N-1.59
												D-4.76	P-1.59
												E-7.94	Q-4.76
												F-0.00	R-6.35
												G-12.70	S-1.59
												H-3.17	T-3.17
												I-3.17	V-12.70
											Stable	K-12.70	W-1.59
14	12.6/sec	None	None	None	None	6998	0	9.57	8	12	(4.37)	L-4.76	Y-4.76
												A-13.14	M-1.71
												C-1.14	N-2.57
												D-7.14	P-4.29
												E-8.86	Q-2.86
												F-2.86	R-10.29
												G-5.71	S-5.14
												H-2.29	T-4.86
												I-2.86	V-7.14
					1 (185-						Stable	K-4.29	W-0.57
15	10.4/sec	1-25	None	None	206)	38672	1.1	5.99	56	51	(39.92)	L-11.43	Y-0.86
												A-9.80	M-1.66
												C-0.55	N-2.77
												D-6.28	P-4.44
												E-5.73	Q-4.62
												F-2.59	R-9.43
												G-6.47	S-5.55
												H-2.96	T-4.99
												I-3.51	V-7.21
											Unstable	K-4.62	W-1.48
16	22.9/sec	None	None	None	None	61199	0.6	9.31	65	76	(42.97)	L-11.28	Y-4.07

Table 3.2 - *Continued*

GP #	Folding rate	SP	TM helix	TM barrel	Leucine zipper	MW (Da)	Cysteine (%)	Theoretical pI	Negatively charged residues (Asp + Glu)	Positively charged residues (Arg + Lys)	Protein stability (II)	AA Frequency	
												A-11.21	M-3.33
												C-0.91	N-4.55
												D-6.06	P-3.33
												E-4.85	Q-3.94
												F-1.82	R-7.27
												G-6.06	S-6.67
												H-2.12	T-6.97
												I-4.85	V-7.88
											Stable	K-3.64	W-1.52
17	20.2/sec	None	None	None	None	36411	0.9	7.06	36	36	(37.01)	L-10.00	Y-3.03
												A-19.35	M-1.61
												C-0.00	N-3.23
												D-3.23	P-4.84
												E-3.23	Q-3.23
												F-4.84	R-9.68
												G-6.45	S-6.45
												H-3.23	T-16.13
												I-3.23	V-3.23
											Stable	K-0.00	W-1.61
18	11.1/sec	None	None	None	None	6620	0	10.57	4	6	(34.08)	L-4.84	Y-1.61
												A-11.88	M-1.98
												C-0.99	N-3.96
												D-5.94	P-1.98
												E-4.95	Q-5.94
												F-0.99	R-21.78
												G-4.95	S-2.97
												H-0.00	T-5.94
												I-0.99	V-5.94
											Unstable	K-0.99	W-0.99
19	20.6/sec	1-15	None	None	None	11775	1	11.67	11	23	(63.59)	L-14.85	Y-1.98
												A-15.29	M-4.71
												C-0.00	N-1.18
												D-4.12	P-8.82
												E-5.29	Q-3.53
												F-0.59	R-9.41
												G-7.06	S-4.71
												H-2.94	T-4.71
												I-2.94	V-6.47
											Unstable	K-4.12	W-2.35
20	-10.9/sec	None	None	None	None	18571	0	10.14	16	23	(40.22)	L-10.00	Y-1.76

Table 3.2 - *Continued*

GP #	Folding rate	SP	TM helix	TM barrel	Leucine zipper	MW (Da)	Cysteine (%)	Theoretical pI	Negatively charged residues (Asp + Glu)	Positively charged residues (Arg + Lys)	Protein stability (II)	AA Frequency	
												A-10.58	M-1.92
												C-0.48	N-2.88
												D-6.73	P-1.92
												E-8.17	Q-5.77
												F-3.37	R-7.69
												G-6.73	S-4.33
												H-0.96	T-5.29
												I-7.21	V-7.69
											Stable	K-6.73	W-0.96
21	10.1/sec	None	None	None	None	23315	0.5	6.2	31	30	(23.24)	L-8.17	Y-2.40
												A-8.46	M-2.31
												C-0.77	N-2.31
												D-8.46	P-7.69
												E-5.38	Q-4.62
												F-4.62	R-10.00
												G-8.46	S-2.31
												H-1.54	T-3.08
												I-5.38	V-6.92
											Unstable	K-1.54	W-1.54
22	19.2/sec	None	None	None	None	14793	0.8	5.33	18	15	(40.94)	L-10.00	Y-4.62
												A-21.11	M-1.11
												C-0.00	N-5.56
												D-4.44	P-2.22
												E-4.44	Q-6.67
												F-3.33	R-3.33
												G-7.78	S-5.56
												H-0.00	T-4.44
												I-4.44	V-7.78
											Stable	K-10.00	W-0.00
23	10.6/sec	None	None	None	None	9252	0	9.7	8	12	(25.76)	L-7.78	Y-0.00
												A-14.08	M-2.17
												C-3.97	N-2.89
												D-6.14	P-6.14
												E-8.30	Q-4.33
												F-2.89	R-6.50
												G-5.42	S-7.22
												H-2.53	T-4.69
												I-3.25	V-6.86
											Unstable	K-1.81	W-1.08
24	7.19/sec	1-19	None	None	None	30057	4	4.76	40	23	(52.80)	L-7.58	Y-2.17

Table 3.2 - *Continued*

GP #	Folding rate	SP	TM helix	TM barrel	Leucine zipper	MW (Da)	Cysteine (%)	Theoretical pI	Negatively charged residues (Asp + Glu)	Positively charged residues (Arg + Lys)	Protein stability (II)	AA Frequency	
												A-15.09	M-2.83
												C-0.94	N-1.89
												D-6.60	P-2.83
												E-6.60	Q-2.83
												F-3.77	R-7.55
												G-7.55	S-3.77
												H-3.77	T-6.60
												I-2.83	V-10.38
											Stable	K-2.83	W-1.89
25	11/sec	None	None	None	None	11633	0.9	5.73	14	11	(21.06)	L-6.60	Y-2.83
												A-18.49	M-4.11
												C-0.68	N-1.37
												D-4.79	P-3.42
												E-5.48	Q-6.16
												F-1.37	R-8.22
												G-3.42	S-6.16
												H-2.05	T-4.79
												I-1.37	V-3.42
											Unstable	K-6.85	W-3.42
26	0.852/sec	None	None	None	None	16301	0.7	9.86	15	22	(44.54)	L-13.01	Y-1.37
												A-9.09	M-3.79
												C-0.76	N-1.52
												D-4.55	P-6.82
												E-7.58	Q-5.30
												F-3.79	R-9.85
												G-6.82	S-2.27
												H-3.03	T-4.55
												I-6.06	V-9.09
											Stable	K-2.27	W-0.76
27	3.04/sec	None	None	None	None	14884	0.8	6.92	16	16	(35.77)	L-10.61	Y-1.52
												A-16.00	M-1.75
												C-0.25	N-3.00
												D-6.50	P-4.25
												E-7.00	Q-5.50
												F-5.25	R-7.75
												G-8.00	S-5.50
												H-1.75	T-4.75
												I-4.25	V-7.25
											Unstable	K-3.00	W-2.50
28	3.47/sec	None	None	None	None	43542	0.2	5.27	54	43	(40.30)	L-5.00	Y-0.75

Table 3.2 - *Continued*

GP #	Folding rate	SP	TM helix	TM barrel	Leucine zipper	MW (Da)	Cysteine (%)	Theoretical pI	Negatively charged residues (Asp + Glu)	Positively charged residues (Arg + Lys)	Protein stability (II)	AA Frequency	
												A-18.18	M-3.03
												C-0.00	N-1.52
												D-9.09	P-4.55
												E-3.03	Q-7.58
												F-4.55	R-3.03
												G-6.06	S-7.58
												H-1.52	T-6.06
												I-4.55	V-7.58
											Stable	K-6.06	W-1.52
29	7.84/sec	None	None	None	None	7011	0	4.99	8	6	(36.11)	L-3.03	Y-1.52
												A-22.03	M-0.85
												C-0.00	N-1.69
												D-7.63	P-3.39
												E-2.54	Q-0.85
												F-0.00	R-5.93
												G-14.41	S-3.39
												H-0.85	T-5.93
												I-5.93	V-9.32
	0.0612/se										Stable	K-2.54	W-0.00
30	c	None	None	None	None	11536	0	5.17	10	12	(4.99)	L-10.17	Y-2.54
												A-11.75	M-1.59
												C-0.32	N-4.44
												D-6.67	P-6.98
												E-5.40	Q-5.71
												F-3.81	R-6.98
												G-7.30	S-3.49
												H-0.95	T-6.35
												I-5.08	V-8.57
											Stable	K-3.17	W-0.63
31	10.6/sec	None	None	None	None	34304	0.3	5.26	38	32	(35.13)	L-8.57	Y-2.22
												A-15.50	M-1.55
												C-0.00	N-0.78
												D-4.65	P-6.20
												E-6.98	Q-3.10
												F-0.78	R-6.98
												G-12.40	S-11.63
												H-2.33	T-7.75
												I-0.78	V-7.75
											Unstable	K-6.98	W-0.00
32	8.27/sec	None	None	None	None	13026	0	9.39	15	18	(40.38)	L-3.10	Y-0.78

Table 3.2 - *Continued*

GP #	Folding rate	SP	TM helix	TM barrel	Leucine zipper	MW (Da)	Cysteine (%)	Theoretical pI	Negatively charged residues (Asp + Glu)	Positively charged residues (Arg + Lys)	Protein stability (II)	AA Frequency	
												A-8.38	M-2.40
												C-2.99	N-2.40
												D-8.98	P-5.39
												E-7.78	Q-5.39
												F-1.20	R-8.98
												G-4.79	S-6.59
												H-0.00	T-7.19
												I-0.60	V-4.19
											Unstable	K-3.59	W-2.40
33	7.69/sec	None	None	None	None	19046	3	4.74	28	21	(54.28)	L-12.57	Y-4.19
												A-9.09	M-4.20
												C-0.00	N-2.80
												D-9.79	P-1.40
												E-2.80	Q-5.59
												F-2.10	R-9.79
												G-11.19	S-6.29
												H-2.80	T-3.50
												I-6.29	V-8.39
											Unstable	K-1.40	W-1.40
34	9.79/sec	None	None	None	None	15859	0	6.04	18	16	(43.65)	L-6.99	Y-4.20
												A-10.50	M-4.00
												C-1.00	N-1.00
												D-5.00	P-5.50
												E-9.00	Q-3.50
												F-4.00	R-6.00
												G-6.50	S-6.50
												H-1.00	T-7.00
												I-3.50	V-8.50
											Stable	K-5.00	W-3.00
35	1.65/sec	None	None	None	None	22377	1	5.09	28	22	(39.36)	L-6.50	Y-3.00
												A-13.04	M-4.35
												C-0.00	N-2.17
												D-9.78	P-6.52
												E-3.26	Q-3.26
												F-1.09	R-6.52
												G-8.70	S-4.35
												H-1.09	T-9.78
												I-3.26	V-11.96
											Stable	K-3.26	W-0.00
36	-13.6/sec	None	None	None	None	9657	0	4.86	12	9	(36.19)	L-5.43	Y-2.17

Table 3.2 - *Continued*

GP #	Folding rate	SP	TM helix	TM barrel	Leucine zipper	MW (Da)	Cysteine (%)	Theoretical pI	Negatively charged residues (Asp + Glu)	Positively charged residues (Arg + Lys)	Protein stability (II)	AA Frequency	
												A-11.59	M-3.86
												C-0.81	N-5.28
												D-5.28	P-4.88
												E-5.08	Q-2.64
												F-2.44	R-4.27
												G-8.13	S-4.88
												H-1.42	T-8.54
												I-6.50	V-8.33
											Stable	K-3.46	W-1.83
37	-11.3/sec	None	None	None	None	53071	0.8	5.09	51	38	(30.02)	L-7.72	Y-3.05
												A-6.50	M-4.07
												C-2.44	N-1.63
												D-8.13	P-4.07
												E-5.69	Q-4.88
												F-6.50	R-6.50
												G-8.13	S-7.32
												H-0.81	T-8.94
												I-4.07	V-8.13
											Unstable	K-4.88	W-1.63
38	4.27/sec	None	None	None	None	13661	2.4	5.1	17	14	(40.74)	L-4.88	Y-0.81
												A-12.50	M-3.26
												C-1.63	N-1.09
												D-9.78	P-2.17
												E-6.52	Q-4.89
												F-2.17	R-8.70
												G-7.07	S-5.98
												H-1.63	T-5.43
												I-3.26	V-7.07
											Stable	K-2.72	W-0.00
39	12.9/sec	1-21	None		None	20101	1.6	4.82	30	21	(24.85)	L-11.96	Y-2.17
												A-17.51	M-3.58
												C-0.00	N-4.32
												D-5.18	P-3.45
												E-2.96	Q-4.56
												F-1.97	R-6.41
												G-12.33	S-6.17
												H-0.25	T-6.66
												I-3.70	V-6.04
											Stable	K-4.19	W-0.62
40	3.78/sec	None	None	None	None	82943	0	9.85	66	86	(32.81)	L-8.75	Y-1.36

Table 3.2 - *Continued*

GP #	Folding rate	SP	TM helix	TM barrel	Leucine zipper	MW (Da)	Cysteine (%)	Theoretical pI	Negatively charged residues (Asp + Glu)	Positively charged residues (Arg + Lys)	Protein stability (II)	AA Frequency	
												A-13.82	M-1.75
												C-0.22	N-3.29
												D-7.46	P-5.04
												E-3.95	Q-4.17
												F-3.07	R-5.48
												G-5.70	S-8.33
												H-1.32	T-6.80
												I-5.04	V-7.02
											Stable	K-2.19	W-2.19
41	1.13/sec	None	None	None	None	49389	0.2	4.78	52	35	(33.74)	L-10.09	Y-3.07
												A-7.79	M-1.82
												C-1.30	N-3.12
												D-7.53	P-3.64
												E-5.71	Q-2.60
												F-2.34	R-7.79
												G-8.83	S-5.45
												H-2.34	T-5.71
												I-5.97	V-10.65
											Stable	K-6.75	W-1.56
42	13.9/sec	None	None	None	None	42538	1.3	8.8	51	56	(29.00)	L-6.75	Y-2.34
												A-7.51	M-2.31
												C-0.58	N-5.78
												D-9.25	P-2.31
												E-5.20	Q-4.05
												F-2.89	R-7.51
												G-10.40	S-4.62
												H-3.47	T-6.36
												I-4.62	V-10.40
											Stable	K-3.47	W-0.58
43	0.365/sec	None	None	None	None	18867	0.6	5.46	25	19	(32.33)	L-7.51	Y-1.16
												A-12.44	M-1.55
												C-2.07	N-2.59
												D-9.84	P-3.63
												E-4.66	Q-2.07
												F-3.11	R-8.81
												G-4.15	S-7.77
												H-2.07	T-8.29
												I-4.15	V-6.74
											Stable	K-1.04	W-4.66
44	1.84/sec	1-25	None	None	None	21648	2.1	4.84	28	19	(34.62)	L-8.81	Y-1.55

Table 3.2 - *Continued*

GP #	Folding rate	SP	TM helix	TM. barrel	Leucine zipper	MW (Da)	Cysteine (%)	Theoretical pI	Negatively charged residues (Asp + Glu)	Positively charged residues (Arg + Lys)	Protein stability (II)	AA Frequency	
												A-14.21	M-0.80
												C-0.27	N-2.14
												D-6.70	P-5.90
												E-4.83	Q-4.29
												F-3.22	R-5.90
												G-8.85	S-5.36
												H-1.07	T-7.24
												I-5.09	V-10.19
											Stable	K-2.14	W-0.80
45	1.42/sec	1-16	None	None	None	39279	0.3	4.83	43	30	(28.22)	L-8.58	Y-2.41
												A-10.61	M-2.02
												C-3.03	N-2.53
												D-7.58	P-5.05
												E-7.07	Q-2.02
												F-3.54	R-12.1
												G-6.57	S-3.54
												H-2.02	T-2.53
												I-3.54	V-7.58
											Unstable	K-1.52	W-3.03
46	-13.6/sec	1-34	None	None	None	22676	3	6.14	29	27	(49.01)	L-11.11	Y-3.03
												A-12.62	M-2.46
												C-0.62	N-5.23
												D-8.31	P-3.69
												E-3.69	Q-2.46
												F-2.15	R-6.77
												G-12.00	S-6.77
												H-2.77	T-4.62
												I-4.62	V-6.77
											Stable(26	K-2.46	W-1.85
47	7.72/sec	None	None	None	None	34639	0.6	5.46	39	30	.63)	L-6.77	Y-3.38
												A-11.46	M-1.75
												C-0.94	N-4.31
												D-7.82	P-5.26
												E-4.31	Q-3.77
												F-3.64	R-6.33
												G-9.30	S-5.93
												H-1.08	T-5.80
												I-3.64	V-8.63
											Stable	K-2.56	W-1.89
48	23.7/sec	None	None	None	None	80473	0.9	4.87	90	66	(31.46)	L-7.68	Y-3.91

Table 3.2 - *Continued*

GP #	Folding rate	SP	TM helix	TM barrel	Leucine zipper	MW (Da)	Cysteine (%)	Theoretical pI	Negatively charged residues (Asp + Glu)	Positively charged residues (Arg + Lys)	Protein stability (II)	AA Frequency	
												A-9.84	M-1.97
												C-1.18	N-3.94
												D-5.51	P-3.94
												E-4.72	Q-1.97
												F-5.51	R-5.12
												G-11.42	S-9.45
												H-1.97	T-6.69
												I-5.91	V-6.69
											Stable	K-2.36	W-0.79
49	11.5/sec	None	None	None	None	26957	1.2	5.2	26	19	(28.72)	L-8.27	Y-2.76

Abbreviations: GP-gene product; SP-signal peptide; TM- transmembrane; MW- molecular weight; pI- isoelectric point; AA- amino acid

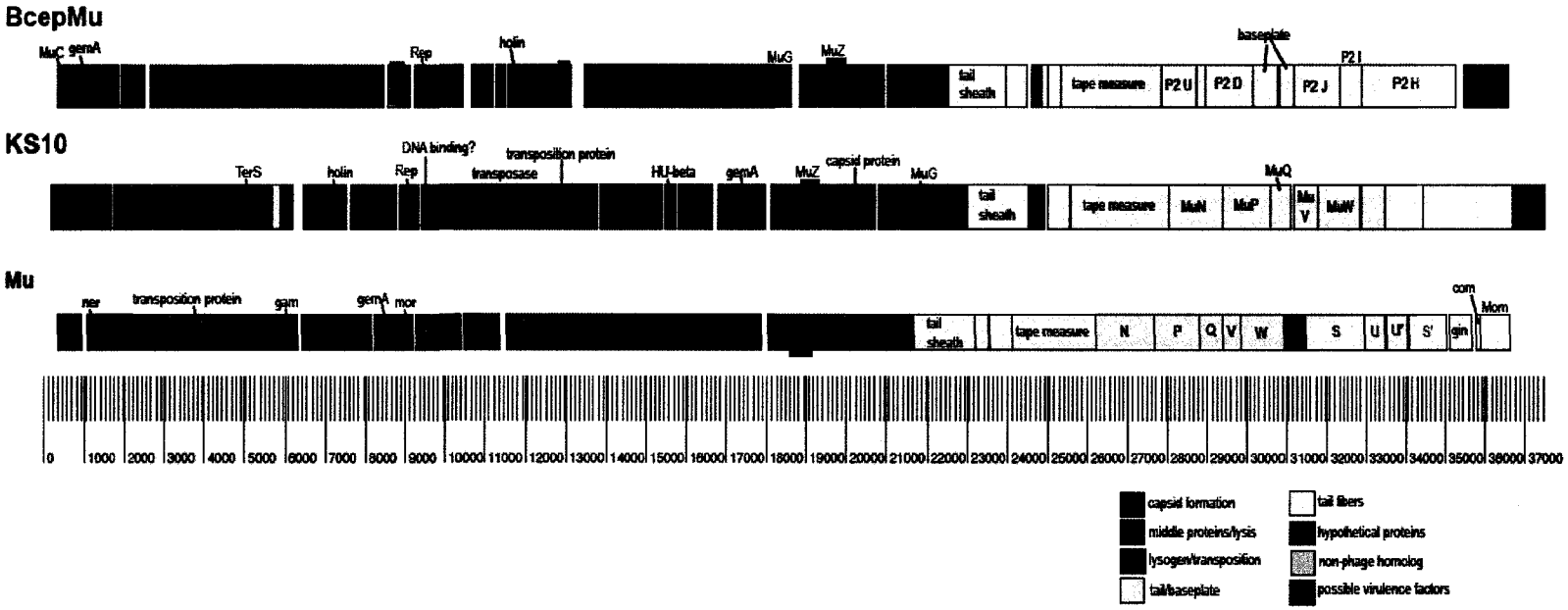


Figure 3.6 Genome maps of KS10 and related phages BcepMu and Mu (derived from NC_005882 and NC_000929, respectively). Each box represents a predicted gene drawn to scale using GenVision program (DNASTAR). Homologues and known phage proteins are indicated (Table 3.1). Different colors represent different modules. Dark grey boxes indicate genes with no known phage homologues and are annotated as hypothetical proteins.

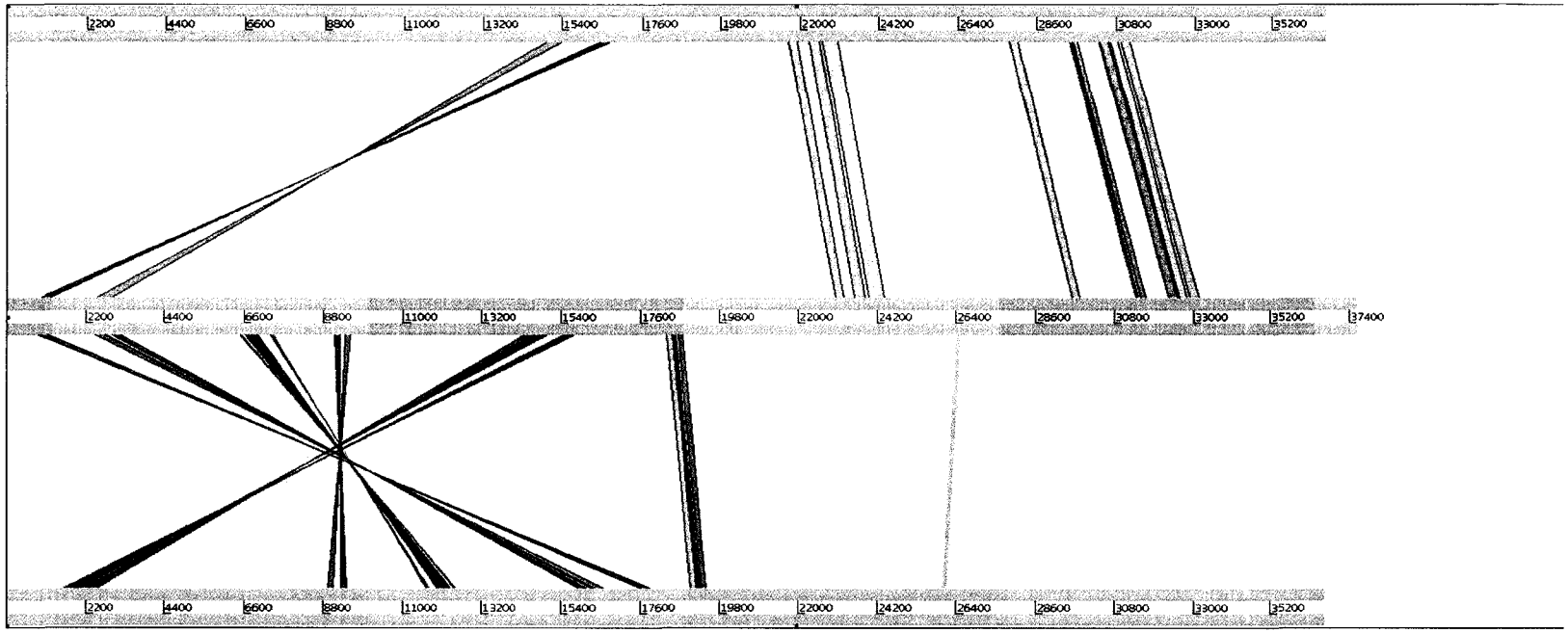


Figure 3.7 Comparison Mu, KS10, and BcepMu genomes (from top to bottom) using the Artemis Comparison Tool (ACT) program. Translated BLAST (tblastx) was used to align translated genomic sequences of each phage. An E-value cutoff of 10 and a score cutoff of 40 were used in this comparison. Nucleotide basepairs are indicated between grey lines for each phage genome. The blue and red lines represent the reverse and forward matches, respectively, and color intensity is proportional to the sequence homology.

3.6 KS10 genome analysis

The KS10 prophage, in both *B. cenocepacia* K56-2 and J2315, lacks the 5'- TG...CA-3' dinucleotides at its termini that are characteristic of Mu-like prophage elements related to BcepMu carrying an Rve integrase catalytic core domain (Summer *et al.*, 2004). Interestingly, where the KS10 prophage is integrated, there is no evidence of direct repeats in the flanking *B. cenocepacia* DNA. These findings, along with the organization of the genome, suggest that the genome of KS10 may have been rearranged after entry into the J2315 host genome or that when it leaves/enters the genome it does so with blunt ends. Because only one end of the KS10 insertion site in K56-2 was identified, we cannot determine if direct repeats are present in this location. However, the DNA flanking the KS10 genome in J2315 is different from that of the DNA flanking the right end of the KS10 genome in K56-2.

Four putative transposase binding sites were identified at the terminal ends of KS10. The position and sequence of the four 15 nt direct repeats which are predicted to be TnpA binding sites in KS10 are indicated in Table 3.4. The *L1* and *L2* transposase binding sites, at the left end, are inverted relative to the *R1* and *R1* binding sites at the right end. The proteins involved in transposition of KS10, mainly gp16 and gp17, show the most homology to the transposase and transposition protein of *R. solanacearum* strain UW551. The suggested transposase binding sites for KS10 are imperfect direct repeats of 15 nucleotides. They do not appear to be similar to those found in BcepMu or Mu, which may be due to the transposase of KS10 not showing significant homology to the

transposase A of Mu or BcepMu. However, the putative KS10 transposase binding sites do contain the characteristic repeated A nucleotides within the sequence (Table 3.4) Unlike the phages Mu, FluMu, Pnm1, and Sp18, which have six putative transposase binding sites within their genomes (Morgan *et al.*, 2002) there are only four identified in KS10. These putative sites in KS10 were identified after a search for the consensus transposase binding sites in both Mu and BcepMu were not detected in the KS10 sequence.

Gyrase binding sites have been identified in Mu and some other Mu-like phages where they are thought to promote the replicative transposition process that occurs during Mu lytic growth. In Mu and FluMu this binding site is located between MuG and MuI (Morgan *et al.*, 2002). A search for this site in KS10 revealed no similar sequence, though we predict that this binding site is present, as a previous study by Sokolsky and Baker (2003) revealed that gyrase is necessary in Mu replicative transposition. By using a drug that inhibits gyrase, they concluded that gyrase activity is important for the lytic life cycle of phage Mu, but it is not necessary for the initial integration into the host chromosome.

Phage Mu has three operator binding sites that are putative binding sites for the phage repressor as well as the transposase during transposition (Mizuuchi and Mizuuchi, 1989). The operator sites in phage Mu contain areas of imperfect repeats of 11 bp long with a consensus of 5'-CTTTTNNN(A/T)(A/T)(A/T)-3' and are located flanking the Integration Host Factor (IHF) binding site. Using DNaseI footprinting, the Mu-like phage D108 operator sites were identified as having the consensus sequence of 5'-AAATC-3' (van Rijn *et al.*, 1988). These sequences

were not identified in KS10, though it is possible that the consensus sequence of Mu binding site is not conserved in KS10 just as the consensus in D108 appears to be different from that in Mu.

3.7 Integration sites of KS10

Normally KS10 will form lytic plaques on lawns of *B. ambifaria* LMG 19467 and *B. cenocepacia* PC184, but at a lower frequency KS10 will also enter into its lysogenic life cycle. To demonstrate that KS10 can insert into the chromosomes of these hosts, colonies of these strains that had been lysogenized with KS10 were selected. In order to obtain these lysogens, KS10 was plated with LMG 19467 and PC184 and turbid plaques indicative of lysogeny were selected using a needle tip. Two LMG 19467 lysogens and one PC184 lysogen were collected and demonstrated by PCR to harbour KS10.

Chromosomes from these lysogens, the seven *B. cenocepacia* clinical isolates, and *B. cenocepacia* C5424 were used to try to determine the exact location of KS10 integration in these hosts. Arbitrary primers (ARB6 and ARB2) and a specific KS10 primer, in addition to the APA gene Gold genome walker kit (Bio S&T, Montreal), were used in PCR experiments to determine the precise insertion site of the integrated prophage. Unfortunately, the only PCR products obtained were found by sequence analysis to be the result of mispriming. This lack of result limited our ability to identify an exact site of insertion for KS10 in each *B. cenocepacia* genome, even though PCR tests identified its presence in the genomes of these strains (Table 3.4).

3.8 Lysogenic conversion of *B. ambifaria* LMG 19467 lysogens

To determine whether KS10 caused a change in the virulence of *B. ambifaria* LMG 19467 once lysogenized, the *G. mellonella* infection model was employed (Seed and Dennis, 2007). Initial experiments employed 40 worms, 10 injected with a 1:5 dilution of an overnight culture of unlysogenized *B. ambifaria* LMG 19467, 10 injected with LMG 19467 lysogen 1 (lys1), 10 injected with LMG 19467 lysogen 2 (lys2), and 10 control worms injected with 10 mM $MgSO_4$ supplemented with 1.2 mg/mL of ampicillin. After 48 hours, survival of worms infected with the wild type was similar to survival seen in those worms infected with the lysogens. All control worms survived the injection, 8 of the 10 worms injected with unlysogenized LMG 19467 and 7 of the 10 worms injected with lys2 were dead after 48 hours, and none of the worms injected with lys1 survived. I To see if changing the dilution would show a difference in killing rate, this experiment was then repeated using 1:10 and 1:100 dilutions of an overnight culture of LMG 19467, lys1, and lys2. This time twenty worms were used in each group, and, as we found in the initial experiment, there appears to be no significant difference in worm survival after infection with un-lysogenized LMG 19467 and infection with lysogens.

Table 3.3. Putative transposase binding sites of KS10

Position	Name	KS10 sequence*
9	L1	GACGTTTAGTTTGAG
113	L2	GACGTTTATTTGAC
37, 592	R2	GTCAAATTAAACGTC
37, 614	R1	GCAAATTAAACGGC
Consensus sequence		gTCAAAtTAAACGtC

* All sequences are written 5'-3'. Consensus sequence is written in the direction of the R1 and R2 binding sites.

Table 3.4 *B. cepacia* complex strains/isolates testing positive for KS10 prophage

Host (*B. cenocepacia* unless noted otherwise)	Source	Integration Site
K56-2	BCC experimental strain panel	Transcriptional regulator; GntR family with aminotransferase
J2315*	BCC experimental strain panel	Oxidoreductase on Chromosome 1
C5424*	BCC experimental strain panel	ND
R1882	Clinical Isolate	ND
R1883	Clinical Isolate	ND
R1884	Clinical Isolate	ND
S11528	Clinical Isolate	ND
R1434	Clinical Isolate	ND
R750	Clinical Isolate	ND
R2314	Clinical Isolate	ND
B. ambifaria LMG 19467 (lys1)	BCC experimental strain panel (modified, this study)	ND
B. ambifaria LMG 19467 (lys2)	BCC experimental strain panel (modified, this study)	ND
PC184 (lys3)	BCC experimental strain panel (modified, this study)	ND

* - Also known to contain BcepMu
ND – Not determined

Chapter 4: Discussion

4. DISCUSSION

Phage and bacteria evolve together, and so for most known bacterial species there are phages that can infect and potentially kill them. Even if a bacterial strain becomes resistant to a phage, the phage may be able to mutate to overcome this resistance. Due to this, phage therapy may prove to be more effective and efficient than antibiotics, especially since it is much easier to modify a phage or isolate a new phage than it is to develop a new antibiotic when bacterial strains become resistant. In order for phage therapy to be used safely, especially when involving lysogenic phages, it is important to determine if phages encode toxins or genes that could be harmful if acquired by the bacteria. Isolating and sequencing many *B. cepacia* complex phages will help identify potential virulence factors originating from phages. *B. cenocepacia* K56-2 and PC184 appear to be the best strains for isolating phages from environmental sources, perhaps due to a lack of a known restriction modification systems in these strains. Strains of bacteria that appear to lack these restriction modification systems would be useful in isolating new bacteriophages, since phages would be detected at a much frequency on these strains. High titer stocks of these phages could then be tested for their ability to infect other strains that may have restriction systems.

Following unsuccessful isolation of broad host range, novel lytic phages, the goal of my research was to sequence and analyze the genome of a temperate phage specific to the *B. cepacia* complex, specifically *B. cenocepacia*. Because KS10 was known to be a prophage in a pathogenic strain of *B. cenocepacia*,

sequencing this phage allowed me to identify potential phage encoded virulence determinants.

4.1 Architecture of the KS10 genome

The KS10 genome is comprised of areas showing similarity to known phage proteins, dispersed among areas with similarity to proteins of unknown function. Even after higher order bioinformatic analyses were carried out on the KS10 amino acid sequence, no functions could be determined for these unknown proteins. Of the proteins with known function, KS10 shows similarity to a number of phages, including the *B. cenocepacia* phage BcepMu, *E. coli* phage Mu, *P. aeruginosa* phage D3112, and a putative prophage in *R. solanacearum* UW551. The mosaicism within the KS10 genome, like most phage genomes, is evident when observing the sequence homology. Genes encoding proteins for capsid formation, tail formation, and lysis/lysogeny are present in all phage genomes, but each gene/module can show homology to different phages. KS10 proteins involved in transposition, for example, show very high similarity to those from the putative prophage in *R. solanacearum* UW551, but no other KS10 proteins show a relationship to this putative prophage. Many KS10 genes show homology to BcepMu genes, but the KS10 tail genes show homology to Mu tail proteins, whereas the tail genes in BcepMu show more homology to phage P2. Mosaicism is especially evident in the tail fiber genes in dsDNA phages (Haggard-Ljungquist *et al.*, 1991). The tail fiber genes of myophage P2 provide good evidence for the theory that phages have evolved through a series of

recombination events with each other. Genes H and G of P2 are predicted to encode a tail fiber protein, and a protein involved in assembly of the tail fibers, respectively (Haggard-Ljungquist *et al.*, 1991). Through amino acid sequence analysis, the P2 tail fiber protein was found to contain regions showing similarity to phages P1, Mu, T2, K3, and λ, and the gpG protein showed the most similarity to phages Mu and P1 (Haggard-Ljungquist *et al.*, 1991). There has also been evidence of mosaicism within the coding region of the integrase (*int*) gene of the temperate λ-related phage, HK022 (Yagil *et al.*, 1989). The amino acid sequence of the excision (*xis*) gene in HK022 differs from λ by only one amino acid. The *int* gene, which is almost identical in location and size in the two phages, however, has a number of different amino acids in the C-terminal end, while the first 55 amino acids are identical. This indicates that there has been a recombination event in the integrase gene of HK022, which has caused the bacterial genome integration site of this phage to be different from that of λ (Yagil *et al.*, 1989).

Previous theories of phage evolution imply that evolution by illegitimate recombination usually occurs by recombination events that will not interrupt the individual modules (Hendrix, 2002). However, the KS10 genome is organized such that the phage head assembly module is interrupted. It is more likely that this rearrangement occurred due to nonhomologous recombination, rather than homologous recombination. Nonhomoglous events are thought to create the module boundaries, whereas homologous recombination is responsible for the shuffling of these genes in the phage population after these boundaries are created

(Hendrix, 2003). In KS10, the scaffold and protease genes are found in the middle of the KS10 genome, while other head assembly genes such as the portal and virion morphogenesis genes, are encoded near the left end. Because proteins involved in head assembly interact with each other, it is unlikely that the head assembly proteins encoded earlier in the genome would be used until the later genes have been transcribed and translated. It is possible that KS10 has evolved in such a way that genes involved in head assembly are transcribed earlier than the rest but translated at the same time as the middle head genes. It has been suggested that phages that recombine in ways that will interrupt the modules will be eliminated because they are at a disadvantage (Hendrix, 2002). Therefore, the rearrangement in KS10 interrupting the head module would suggest that, either it is not important for the head genes to be transcribed together, which is unlikely since these genes interact intimately with each other, or that KS10 has some way of controlling the transcription (via a transcriptional suppressor or activator) so that all genes involved in head assembly will be transcribed together.

KS10 gene 7 shows no homology to other known phage proteins. This gene encodes a protein showing relatively high identity to a dksA/traR C4-type zinc finger protein found in bacteriophage L-413C. DksA is a DnaK suppressor protein that suppresses transcription of DnaK, while TraR is a transcriptional activator. The coliphages P2, 186, and phage Phi MhaA1-PHL101 also encode a protein showing homology to this dskA/traR protein, though other phages similar to KS10 do not. In Phi MhaA1-PHL101, the conserved Dsk/TraR region is only over the last 40 amino acids and the authors suggest this could be involved in

transcriptional activation (Highlander *et al.*, 2006). In KS10 gp7, the domain is conserved across the whole protein. This gene is located upstream of the first module of head genes. A possible role for this protein in KS10 could be to repress the transcription of genes 1-5 until the second head module is transcribed so they can be translated together. As previously mentioned, the proteins encoded by the first set of head genes cannot be used until the second set of head genes are transcribed and translated, so this protein may be responsible for controlling the transcription of the two separate head gene modules. Another possibility is that it activates transcription of the second set of head genes.

As previously discussed, phages can often encode proteins that they have acquired from recombination with a bacterial host or another phage. Since DNA recombination can often occur between prophages in the same genome, it is surprising that none of the KS10 genes show higher identity to BcepMu genes, since KS10 was found to be a prophage in several strains that are also lysogenized with BcepMu. However, the integration of one or both of these phages into the genome of these *B. cenocepacia* strains may be a fairly unusual event. Although a study for polylysogeny in the *B. cepacia* complex did not detect polylysogeny (Langley *et al.*, 2003), we now know this is not the case in *B. cenocepacia* J2315. This is not unexpected, as other species, including *P. aeruginosa* (Holloway *et al.*, 1960), are known to be lysogenized by more than one phage. Perhaps polylysogeny occurs in many *B. cepacia* complex strains, but multiple approaches must be employed in order to prove this condition.

In some cases, bacterial infection with a phage can cause induction of dormant prophages, suggesting that in addition to induction by mutagens such as mitomycin C or UV, prophages can also be induced by phage infection. When an El Tor strain of *Vibrio cholerae* O1 was infected with phage ϕP15, two additional phages were detected in the supernatant of the *V. cholerae* culture after extended incubation (2 and 6 days) (Espeland *et al.*, 2004). Espeland *et al.* (2004) found that these prophages were not induced when exposed to mitomycin C or UV radiation. It is not known, however, which ϕP15 phage proteins can cause this induction, or how they interact with the prophages to cause the induction. Another case of polylysogeny is evident in the marine bacterium *Silicibacter* sp. strain TM1040 (Chen *et al.*, 2006). Three prophages and two prophage remnants were detected within the genome of this strain using multiple approaches, including PCR of a prophage specific gene. One of these phages appeared to be spontaneously induced, like KS10, while two others required induction through the use of mitomycin C (Chen *et al.*, 2006). The events causing induction of any particular prophage are still not fully understood. It is unknown why some phages, such as the two *Vibrio* prophages, are not induced until infection with an additional phage, and some phages, such as KS10, will be spontaneously induced at a high frequency in the absence of a mutagen or phage infection. The use of PCR with primers specific for the catalytic core of a phage integrase or another specific prophage gene, may be more effective than induction at identifying prophages or prophage remnants within a bacterial genome, as different phages appear to have different requirements for induction.

4.2 Head and Tail Assembly

In most phages, head assembly progresses through five major stages, each involving a number of phage encoded proteins. The first stage, initiation, involves the initiation of polymerization of phage coat proteins by minor head proteins. Next, during shell formation, the immature procapsid, comprised of coat proteins, is assembled around the scaffold protein, which is subsequently cleaved to form a mature procapsid. Following maturation of the procapsid are the two final stages, DNA packaging and head completion (Grimaud, 1996). In KS10 gp1-gp5 and gp28-gp34 are expected to be involved in head assembly. KS10 gp28 and gp29, identified as the protease and scaffold proteins, are structurally important in this process. This scaffolding protein has been shown to be essential in assembling the coat proteins, encoded by gene 31 in KS10, into a shell (King and Casjens, 1974). Once this shell is formed, it is thought that the protease, KS10 gp28, is involved in removing this scaffold. As is the case in many other phages, the scaffold gene of KS10 is embedded in frame with the protease gene.

The genes mainly involved in the actual packaging of the DNA into the procapsid are the terminases, portal, and major capsid, which in KS10 are genes 4, 5, 3, and 31, respectively. During packaging, the portal protein is located at one vertex of the capsid, and ensures the connection between the capsid and the tail. This protein is also involved in forming the channel through which the DNA is packaged during morphogenesis and exits during infection (Bear *et al.*, 1984). KS10 gp34 shows homology to MuG, which is thought to be a tail protein, though

the function is unclear (Giphart-Gassler *et al.*, 1981). KS10 gp1 is expected to be involved in virion morphogenesis as it shows relatedness to MuF, though its function remains unknown. There are also a number of genes encoding hypothetical proteins interspersed within these known genes. These genes may encode minor proteins involved in forming the mature packaged prohead.

The module encoding proteins for tail assembly is located at the right end of the genome, similar to Mu, though it lacks the invertible G-region of Mu. This invertible region found in Mu and some related phages encodes the genes involved in tail fiber synthesis, and the orientation of the region determines the host range of the phage (van de Putte *et al.*, 1980). BcepMu, a similar phage found in *B. cenocepacia*, has a right end similar to P2 and also lacks this invertible G region (Summer *et al.*, 2004). Although KS10 tail proteins do show homology to Mu tail proteins, the KS10 genome also lacks this invertible G region.

Unlike proteins involved in head assembly, lysis, and transposition, there is little known about the proteins involved in Mu tail assembly. The last approximately 14 kb of the KS10 genome is involved in tail assembly. Like other members of the *Myoviridae* family of phages with contractile tails, Mu-like phages have a contractile tail sheath outside of an inner tail tube. The phage baseplate is located at the end of the tail and is attached to the tail fibers, which are involved in attachment to the host cell. KS10 genes 37 and 41 encode the tail sheath and tail tape measure, respectively. In both Mu and BcepMu the tail tube gene is found between the genes encoding the tail sheath and the tape measure. In

KS10 there is a hypothetical protein of 123 amino acids that, using a BLAST analysis, shows no homology to anything in the database, and no conserved domains could be identified. This gene is the correct length and in the expected location to encode a tail tube protein based on other similar phages, though there is no functional evidence to support this prediction.

Xu *et al.* (2004) suggest that there is -1 (or -2 in Mu) frameshift conserved among dsDNA tailed phages that occurs before the tail tape measure gene and after the major tail gene. This frameshift occurs in many phages including Mu, FluMu, P2, lambda, and D3. We found no evidence of a frameshift region before the tape measure gene in KS10. Using PSI-BLAST analysis, a FluM-like gp41 conserved domain was detected, which has been annotated as a Lambda G analogue. In many dsDNA tailed phages there is a slippery sequence within this "G" gene that causes a frameshift, creating two overlapping ORFs. This sequence is usually a region of repeated A, T, or G nucleotides. In KS10, a sequence capable of causing this frameshift was not identified using both a manual search as well as a Frame Shift Finder program (http://chainmail.bio.pitt.edu/~junxu/cgi-bin/webshift.cgi).

KS10 gp42 to gp45 are predicted to be involved in baseplate assembly since they show relatedness to Mu proteins with this function. How each protein is involved in baseplate assembly in Mu is yet unknown. KS10 gene 7 encodes a 325 amino acid protein showing 42% identity to a tail collar protein of a *B. multivorans* strain prophage and 27% identity to a tail fiber protein of a prophage in *B. thailandensis* strain E264. In many Mu-like phages, the gene or genes

encoding tail fibers is/are relatively long. BcepMu, for example, has a tail fiber gene encoding a 786 amino acid protein, similar to the phage P2, which is much larger than the small gene in KS10 showing poor homology to a tail fiber gene. KS10 gp48, a 742 amino acid protein at the end of the genome shows 46% similarity to the periplasmic component of an ABC-type phosphate transporter system of *Pseudomonas stutzeri* A1501. However, the region showing homology to the protein from *P. stutzeri* shows no conserved domains when analyzed using PSI-BLAST and no homology to any other periplasmic component from an ABC-type phosphate transporter system, suggesting that phosphate transport may not be the function of this protein. When using BLAST to compare the amino acid sequence of gp48 to sequences in the translated nucleotide database, this sequence also has homology with sequences from chromosome 1 of *B. phytofirmans* PsJN and chromosome 1 from *B. xenovorans* LB400. This may indicate that prophages similar to KS10 are also present in these other species of *Burkholderia*. Because KS10 does not appear to encode any tail fiber genes, perhaps this relatively long gene 48 in KS10 encodes a tail fiber protein, but the ancestry has been lost due to extensive mutational changes. Though we would still expect to see some conserved tail fiber domain, it is possible that the phages most related to KS10 have not yet been sequenced or annotated, or that this gene has accumulated many mutations that mask the encoded protein's function.

Interestingly, Mu-like phages Pnm2 and NeisMu1 have both integrated into an ABC-type transporter gene in two different strains of *Neisseria meningitidis*, indicating that they are probably descendants of the same integration

events (Casjens, 2003). If the last gene in KS10 is actually an ABC-type transporter gene, and not a tail fiber gene, it could indicate that KS10 is a descendant of a phage that had initially integrated into this transporter gene, suggesting that ABC-type transporter genes may contain a sequence that is susceptible to integration by transposable phages. Since we know Mu-like phages are capable of packaging a variable amount of host DNA at the right end (Symonds *et al.*, 1987), it is possible that this ABC-type gene was packaged by an ancestor of KS10 and remains in the genome. More experiments are needed to determine the definite function/identity of KS10 gp48.

4.3 Host cell Lysis and Transposition

Phage host cell lysis usually involves a number of proteins, including a holin and a lysin. In KS10 these proteins are encoded by gene 21 and gene 22, respectively. KS10 holin and lysin proteins show homology to the holin and lysin of BcepMu (Summer *et al.*, 2004) but the lysins differ in the N-terminal region. PSI-BLAST revealed the lysin of KS10 to have a soluble lytic transglycosylase (Slt) domain of 117 amino acids, similar to that of BcepMu and other non-Mu phages such as T7. Most enzymes of this nature found in *E. coli* have been shown to catalyze the cleavage between N-acetylmuramic acid and N-acetylglucosamine, resulting in 1,6-anhydromuramic acid. In *E.coli* these enzymes degrade the cell wall murein during bacterial morphogenesis (Hoeltje *et al.*, 1975). In BcepMu this Slt domain is conserved over residues 46-155, while in KS10 the domain is conserved over residues 70-170. Most phage lysins lack their own signal peptide

(SP) sequence, and are under control of the holin for release. In a search for SP sequences within the lysin gene of KS10, however, a potential SP sequence was identified using SignalP server (http://www.cbs.dtu.dk/services/SignalP-2.0), which predicts the location of SP sequences and cleavage sites. The cleavage site in KS10 gp10 was predicted to be between amino acids 44 and 45. The N-terminal region of KS10 putative lysin protein is also fairly hydrophobic, further suggesting that this protein contains an SP, though this hypothesis has not been confirmed experimentally. Although it does not appear to be a common phenomenon, lysins of other bacteriophages, such as the *Oenococcus oeni* phage fOg44, have been found to have a SP sequence. In this phage, overexpresssion of its lysin (Lys44) in *E. coli* was lethal, which is uncommon for phage lytic enzymes. When the SP was deleted, overexpression of Lys44 was no longer toxic (São-José *et al.*, 2000).

Other phage genes that are often involved in lysis are Rz/Rz1, which are secondary proteins involved in lysis, and an antiholin, which inhibits the holin until the phage are ready to be released (Wang *et al.*, 2000). Rz1 is embedded out of frame within the Rz gene and is usually proline rich. These genes are not essential for lysis, and a deletion has no affect on plating efficiency. Previous studies of phage Mu suggest that these proteins are not encoded by this phage. However, a more detailed search by Summer *et al.* (2007) revealed Rz/Rz1 equivalent genes in many phages that had no previous Rz/Rz1 genes identified. By manually searching the KS10 genome using the same rules employed by Summer *et al.* (2007), Rz/Rz1equivalent genes have not been identified in KS10.

Because so many other *Burkholderia* phages (all 16 of the *Burkholderia* phages checked by Summer *et al.*, 2007) encode these Rz/Rz1 equivalents, we would expect that KS10 would as well, since these phages would presumably have similar lysis strategies. It is unknown if KS10, and other phages without Rz/Rz1, would be at a disadvantage by not encoding these genes.

Phage repressors are involved in promoting lysogeny by blocking transcription of early lytic promoters (Mizuuchi and Mizuuchi, 1989). As mentioned previously, the repressor (Rep) in KS10 shows 36% identity to the Rep in BcepMu and unlike BcepMu, KS10 appears to switch to lysis at a fairly high frequency, as small pinpoint plaques are usually visible on control plates containing lawns of *B. cenocepacia* K56-2. Although *B. cenocepacia* J2315 and C5424 are also lysogenized with KS10, these small plaques are not normally visable on lawns of uninduced cultures and so this frequent induction appears to be limited to K56-2. The ability of KS10 to integrate into both J2315 and C5424 is interesting, as it proves multiple lysogeny can occur within these strains of *B. cenocepacia* with two functionally similar phage. A previous study stated that J2315 contained only one prophage, BcepMu (Langley *et al.*, 2003). KS10 was, perhaps, not detected in previous attempts to isolate prophage from strain J2315 due to its relatively low host range for plaque formation, and its overlap with the host range of BcepMu.

A study by Ranquet *et al.* (2005), mentioned in section 1.2.3, showed that Mu Rep can be degraded by proteases ClpXP and Lon of *E.coli.* They suggest that this is caused by a conformational change in Rep during stationary phase that

causes a ClpX recognition motif to be revealed. Since the repressor is responsible for lysogeny, if it is being degraded it will no longer compete with the transposase for binding sites in the operator region, and the phage will go through its lytic life cycle. Interestingly, although the repressor proteins of KS10 and BcepMu show 36% identity to each other and both are found integrated into J2315 and C5424, only KS10 is a prophage in K56-2, while BcepMu will form plaques on this strain. Since the KS10 prophage in K56-2 is identical to the KS10 prophage in J2315, a strain that also harbours BcepMu, the reason for BcepMu not lysogenizing K56-2 is probably not due to a protein expressed by KS10.

It is still unknown why phage KS10 is so readily induced from K56-2 that pinpoint plaques are visible on lawns of K56-2. One explanation could be that the repressor protein of KS10 has some property causing it to be more readily degraded by proteases of K56-2 than the proteases found in other strains of the same species. However, an experiment identifying a protease of *B. cenocepacia* that is capable of degrading phage repressors has not yet been performed.

4.4 Conclusions

Phages are one of the most abundant particles on earth, but only a relatively small number of these phage's genomes have been sequenced, making it difficult to draw definitive conclusions about the relationship of KS10 and its gene products to other phages. Because the highest percent identity KS10 protein sequences show to an orthologous proteins is only 67%, it has likely been a long time since KS10 diverged from its "parent" phage(s). While the KS10 genome

seems to show unusual variability in its genomic organization, especially in the first half, it is possible that there are many other phages with similar organizations that have not yet been sequenced. There also remain a number of unknown proteins within the KS10 genome, but this is common for many phages. Because there are still many phage genomes that are not yet sequenced, and because phage genes, like all genes, can undergo point mutations, which if accumulate enough, can cause a gene to no longer show sequence similarity to its ancestor (Hendrix, 2002), determining the function of encoded phage proteins can be especially difficult. It is also possible that these unknown genes have no function, other than to make the phage genome the proper size for packaging the genome into the capsid (Hendrix, 2002). Alternatively, these extra genes may be in the phage genome to decrease the frequency of essential genes being interrupted/destroyed by a recombination event, mutations, or deletions in the genome.

In 2001, Clark *et al.* attempted to determine the rate at which mutations occur in lambdoid phages. To do this they compared three head assembly proteins from *Salmonella enterica* phage P22, and two similar phages, APSE-1, and *E. coli* phage HK620, with the homologues in the *E.coli* podoviruses 933W and VT2-Sa. They also compared the three proteins from phages P22, APSE-1, and HK620 with each other. From their comparisons, they suggest that phages may mutate at a rate similar to that of bacteria, or that they may accumulate mutations during phage replication, so their mutation rate may be slightly higher than that of the host (Clark *et al.*, 2001). It also seems possible that different phages, or different phage proteins, mutate at faster rates than others. Clark *et al.*

(2001) observed that the mutation rate between the three head proteins of APSE-1 and P22 was different than for the mutation rate between the three proteins in APSE-1 and HK620. Overall, it would appear that phages mutate at a rate that is not significantly higher than that of their bacterial hosts. It is well known that there is a high degree of divergence amongst tailed phages, seemingly higher than that of bacteria. Because the phage mutation rate, at least that of the three lambdoid phages studied by Clark *et al.* (2001), is similar to that of bacteria, the age of the tailed phage population is suggested to be a hundred million to a billion years (Hendrix, 2002).

Seed and Dennis (2005) found *B. cenocepacia* K56-2 to be highly susceptible to phage infection; many phages are capable of forming plaques on this strain. Because KS10 is a prophage in *B. cenocepacia* K56-2 and appears to switch to its lytic cycle at a high frequency, knowing the sequence of this phage is important for future research involving phages propagated on strain K56-2. Complete sequencing of bacterial genomes has provided increasing opportunities for prophage identification, and has also produced incontrovertible evidence of extensive phage-mediated exchange of genetic material between species. By sequencing and characterizing *B. cepacia* complex phages, regardless of whether they are lytic or lysogenic, we can begin to further develop these phages as novel therapeutic agents for use against infections caused by the highly antibiotic resistant *B. cepacia* complex.

Using *B. ambifaria* LMG 19467 lysogens in a wax worm model we have also provided some evidence that KS10 is not responsible for any virulence

associated with the strains of *B. cenocepacia* it lysogenizes. However, due to the high mortality of the larvae 24 hours after infection with un-lysogenized *B. ambifaria* LMG 19467, even if the lysogens were capable of killing the larvae faster, it may be difficult to observe with the dilutions used. Per haps a more extensive experiment including many dilutions and closer time points may give a better insight into the killing rates of these lysogens. We have also shown that polylysogeny does occur in the *B. cepacia* complex, and that it can occur with two different Mu-like *B. cenocepacia* phages. Further studies are required to understand the interactions of multiple active phages within a genome of a strain of the *B. cepacia* complex and their impact, if any, on pathogenicity.

Chapter 5: Literature Cited

Abuladze, N. K., Gingery, M., Tsai, J., & Eiserling, F. A. (1994). Tail length determination in bacteriophage T4. *Virology, 199*(2), 301-310.

Ackermann, H.-W. (2001). Frequency of morphological phage descriptions in the year 2000. *Archives of Virology, 146*(5), 843-857.

Ackermann, H.-W. (2001). 5500 Phages examined in the electron microscope. *Archives of Virology, 152*(2), 227-243.

Adzuma, K., & Mizuuchi, K. (1988). Target immunity of mu transposition reflects a differential distribution of mu B protein. *Cell, 53*(2), 257-266.

Adzuma, K., & Mizuuchi, K. (1989). Interaction of proteins located at a distance along DNA: Mechanism of target immunity in the Mu DNA strand-transfer reaction. *Cell, 57*(1), 41-47.

Altschul S.F., Gish W., Miller W., Myers E.W., & Lipman D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology, 215*(3), 403-410.

Altschul, S. F.,Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research. 25*(17), 3389-3402.

Amann, E., & Brosius, J. (1985). 'ATG vectors' for regulated high-level expression of cloned genes in *Escherichia coli*. *Gene*, 40(2), 183-190.

Amann, E., Ochs, B., & Abel, K.J. (1988) Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. *Gene, 69*(2), 301-315.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. (eds) (1999). *Short protocols in molecular biology*. New York: JohnWiley and Sons, Inc.

Baker, T.A., Mizuuchi, M., and Mizuuchi, K. (1991). MuB protein allosterically activates strand transfer by the transposase of phage Mu. Cell. *65*(6):1003-1013.

Baker, T. A., & Luo, L. (1994). Identification of residues in the mu transposase essential for catalysis. *Proceedings of the National Academy of Sciences of the United States of America, 91*(14), 6654-6658.

Bear, S. E., Court, D. L., & Friedman, D. I. (1984). An accessory role for *Escherichia coli* integration host factor: Characterization of a lambda

mutant dependent upon integration host factor for DNA packaging. *Journal of Virology, 52*(3), 966-972.

Bernier, S. P., Silo-Suh, L., Woods, D. E., Ohman, D. E., & Sokol, P. A. (2003). Comparative analysis of plant and animal models for characterization of burkholderia cepacia virulence. *Infection and Immunity, 71*(9), 5306-5313.

Besemer J. & Borodovsky M. (2005) GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Research, 33,* W451-454.

Biswas, B., Adhya, S., Washart, P., Paul, B., Trostel, A. N., Powell, B., *et al.* (2002). Bacteriophage therapy rescues mice bacteremic from a clinical isolate of vancomycin-resistant *Enterococcus faecium. Infection and Immunity, 70*(1), 204-210.

Bottone, E.J., Douglas, S.D., Rausen, A.R., and Keusch, G.T. (1975). Association of *Pseudomonas cepacia* with chronic granulomatous disease, *Journal of Clinical Microbiol*ogy, *1*(5), 425-428.

Brüssow, H., Canchaya, C., & Hardt, W. (2004). Phages and the evolution of bacterial pathogens: From genomic rearrangements to lysogenic conversion. *Microbiology and Molecular Biology Reviews, 68*(3), 560-602.

Burkholder, WH. (1950). Sour skin, a bacterial rot of onion bulbs. *Phytopathology. 40,* 115-117

Butler, S.L., Doherty, C.J., Hughes, J.E., Nelson, J.W., & Govan, J.R. (1995). *Burkholderia cepacia* and cystic fibrosis: Do natural environments present a potential hazard? *Journal of Clinical Microbiology, 33*(4), 1001-1004.

Bylund, J., Campsall, P. A., Ma, R. C., Conway, B. D., & Speert, D. P. (2005). *Burkholderia cenocepacia* induces neutrophil necrosis in chronic granulomatous disease. *Journal of Immunology, 174*(6), 3562-3569.

Canchaya, C., Fournous, G., & Brüssow, H. (2004). The impact of prophages on bacterial chromosomes. *Molecular Microbiology, 53*(1), 9-18.

Cardona, S. T., Wopperer, J., Eberl, L., & Valvano, M. A. (2005). Diverse pathogenicity of *Burkholderia cepacia* complex strains in the *Caenorhabditis elegans* host model. *FEMS Microbiology Letters, 250*(1), 97-104.

Carlson, K., E. Raleigh, A., and Hattman, S. (1994). Restriction and modification. *In* Molecular biology of bacteriophage T4. *Edited* by J. D. Karam, J. W. Drake, K. N. Kreuzer, G. Mosig, D. H. Hall, F. A. Eiserling, L. W. Black, E. K. Spicer, E. Kutter, K. Carlson, and E. S. Miller. ASM Press,

Washington, D.C. pp. 369–381.

Carver, T.J., Rutherford, K.M., Berriman, M., Rajandream, M.A., Barrell, B.G., and Parkhill, J. (2005). ACT: the Artemis Comparison Tool. *Bioinformatics*, 21(16), 3422-3423.

Chene, F., Wang, K., Stewart, J., & Belas, R. (2006). Induction of multiple prophages from a marine bacterium: A genomic approach. *Applied and Environmental Microbiology, 72*(7), 4995-5001.

Cheng, H. & Lessie, T.G. (1994). Multiple replicons constituting the genome of *Pseudomonas cepacia* 17616. *Journal of Bacteriology, 176*(13), 4034-4042.

Cihlar, R. L., Lessie, T. G., & Holt, S. C. (1978). Characterization of bacteriophage CP1, an organic solvent sensitive phage associated with *Pseudomonas cepacia*. *Canadian Journal of Microbiology, 24*(11), 1404-1412.

Clark, A. J., Inwood, W., Cloutier, T., & Dhillon, T. S. (2001). Nucleotide sequence of coliphage HK620 and the evolution of lambdoid phages. *Journal of Molecular Biology, 311*(4), 657-679.

Coenye, T., LiPuma, J.J., Henry, D., Hoste, B., Vandemeulebroecke, K., Gillis, M., Speert, D.P., and Vandame, P. (2001a). *Burkholderia cepacia* genomovar VI, a new member of the *Burkholderia cepacia* complex isolated from cystic fibrosis patients. *International Journal of Systematic and Evolutionary Microbiology, 51*(2), 271-279.

Coenye, T., Mahenthiralingam, E., Henry, D., LiPuma, J.J., Laevens, S., Gillis, M., Speert, D.P., and Vandamme, P. (2001b). *Burkholderia ambifaria* sp. nov., a novel member of the *Burkholderia cepacia* complex including biocontrol and cystic fibrosis-related isolates. *International Journal of Systematic and Evolutionary Microbiology, 51*(2), 1481-1490.

Coenye T., Vandamme P., LiPuma J.J., Govan J.R., & Mahenthiralingham E. (2003) Updated version of the *Burkholderia cepacia* complex experimental strain panel. *Journal of Clinical Microbiology, 41*(6), 2797-2798.

Coleman, D., Knights, J., Russell, R., Shanley, D., Birkbeck, T. H., Dougan, G., *et al.* (1991). Insertional inactivation of the staphylococcus aureus β-toxin by bacteriophage Φ13 occurs by site- and orientation-specific integration of the Φ13 genome. *Molecular Microbiology, 5*(4), 933-939.

Cornelis, G. R., Agrain, C., & Sorg, I. (2006). Length control of extended protein structures in bacteria and bacteriophages. *Current Opinion in Microbiology, 9*(2), 201-206.

Cost, G. J. (2007). Enzymatic ligation assisted by nucleases: Simultaneous ligation and digestion promote the ordered assembly of DNA. *Nature Protocols, 2*(9), 2198-2202.

Datta, D.B., Arden, B., and Henning, U. (1977). Major proteins of the *Escherichia coli* outer cell envelope membrane as bacteriophage receptors. *Journal of Bacteriology, 131*(3), 821-829.

Defenbaugh, D. A., & Nakai, H. (2003). A context-dependent ClpX recognition determinant located at the C terminus of phage mu repressor. *Journal of Biological Chemistry, 278*(52), 52333-52339.

Delcher, A.L., Harmon, D., Kasif, S.,White, O., and Salzberg, S.L. (1999). Improved microbial gene identification with Glimmer. *Nucleic Acids Research. 27*, 4636-4641.

Dennis, J. J., & Zylstra, G. J. (1998). Plasposons: Modular self-cloning minitransposon derivatives for rapid genetic analysis of gram-negative bacterial genomes. *Applied and Environmental Microbiology, 64*(7), 2710-2715.

DeShazer, D., & Woods, D.E. (1996). Broad-host-range cloning and cassette vectors based on the R388 trimethoprim resistance gene. *BioTechniques, 20*(5), 762-764.

D'Herelle, F. (1922). The bacteriophage: its role in immunity. Williams and Wilkens Co./Waverly Press, Baltimore, USA.

Dobbin, C.J., Soni, R., Jelihovsky, T., and Bye, P.T. (2000). Cepacia syndrome occurring following prolonged colonization with *Burkholderia cepacia*. *Australian and New Zealand Journal of Medicine, 30*(2), 288-289.

Dodson, M., Echols, H., & Wickner, S. (1986). Specialized nucleoprotein structures at the origin of replication of bacteriophage λ: Localized unwinding of duplex DNA by a six-protein reaction. *Proceedings of the National Academy of Sciences of the USA, 83*(20), 7638-7642.

Duda, R. L., Wall, J. S., & Hainfeld, J. F. (1985). Mass distribution of a probable tail-length-determining protein in bacteriophage T4. *Proceedings of the National Academy of Sciences of the USA, 82*(16), 5550-5554.

Entenza, J. M., Loeffler, J. M., Grandgirard, D., Fischetti, V. A., & Moreillon, P. (2005). Therapeutic effects of bacteriophage cpl-1 lysin against *Streptococcus pneumoniae* endocarditis in rats. *Antimicrobial Agents and Chemotherapy, 49*(11), 4789-4792.

Espeland, E. M., Lipp, E. K., Huq, A., & Colwell, R. R. (2004). Polylysogeny and prophage induction by secondary infection in *Vibrio cholerae*.

Environmental Microbiology, 6(7), 760-763.

Evdokimov, A. A., Zinoviev, V. V., Malygin, E. G., Schlagman, S. L., & Hattman, S. (2002). Bacteriophage T4 dam DNA-[N6-adenine]methyltransferase. Kinetic evidence for a catalytically essential conformational change in the ternary complex. *Journal of Biological Chemistry, 277*(1), 279-286.

Fischetti, V. A. (2005). Bacteriophage lytic enzymes: Novel anti-infectives. *Trends in Microbiology, 13*(10), 491-496.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., *et al.* (1995). Whose-genome random sequencing and assembly of *Haemophilus influenzae* rd. *Science, 269*(5223), 496-521.

Gaffney, T.D. & Lessie, T.G. (1987). Insertion-sequence-dependent rearrangments of *Pseudomonas cepacia* plasmid pTGL1. *Journal of Bacteriology, 169*(1), 224-230.

García, R.L., & Molineux, I. J. (1996). Transcription-independent DNA translocation of bacteriophage T7 DNA into *Escherichia coli*. *Journal of Bacteriology, 178*(23), 6921-6929.

Geuskens, V., Mhammedi-Alaoui, A., Desmet, L., & Toussaint, A. (1992). Virulence in bacteriophage Mu: A case of transdominant proteolysis by the *Escherichia coli* Clp serine protease. *EMBO Journal, 11*(13), 5121-5127.

Gibson, R.L., Burns, J.L., and Ramsey, B.W. (2003). Pathophysiology and management of pulmonary infections in cystic fibrosis. *American Journal of Respiratory and Critical Care Medicine, 168*(8), 918-951.

Giphart-Gassler, M. Wijffelman, C., and Reeve, J. (1981). Structural polypeptides and products of late genes of bacteriophage Mu: Characterization and functional
aspects. *Journal of Molecular Biology, 145*(1), 139-163.

Goldhaber-Gordon, I., Early, M. H., & Baker, T. A. (2003). MuA transposase separates DNA sequence recognition from catalysis. *Biochemistry, 42*(49), 14633-14642.

Goosen, N., and van de Putte, P. (1987). Regulation of transcription. *In* Phage Mu. *Edited by* N. Symonds, A. Toussaint, P. Van de Putte, and M.M. Howe. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. pp. 41-52.

Govan, J.R., Hughes, J.E., and Vandamme, P. (1996). *Burkholderia cepacia*: medical, taxonomic and ecological issues. *Journal of Medical Microbiology, 45*(6), 395-407.

Grant, W.F. (1979). The genotoxic effects of 2,4,5-T. *Mutation Research, 65*(2), 83-119.

Greer, G.G. (2005). Bacteriophage control of foodbourne bacteria. *Journal of Food Protection, 114*(3), 370-375.

Grimaud, R. (1996). Bacteriophage Mu head assembly. *Virology, 217*(1), 200-210.

Groat, R. G., Schultz, J. E., & Zychlinsky, E. (1986). Starvation proteins in *Escherichia coli*: Kinetics of synthesis and role in starvation survival. *Journal of Bacteriology, 168*(2), 486-493.

Guttman, B., Raya, R., and Kutter, E. 2005. Basic phage biology. *In* Bacteriophages: biology and applications. *Edited by* E. Kutter and A. Sulakvelidze. CRC Press, Boca Raton, Fl. pp. 29-66.

Hagens, S., Habel, A., von Ahsen, U., von Gabain, A., & Blasi, U. (2004). Therapy of experimental pseudomonas infections with a nonreplicating genetically modified phage. *Antimicrobial Agents and Chemotherapy, 48*(10), 3817-3822.

Haggard-Ljungquist, E., Halling, C., & Calendar, R. (1992). DNA sequences of the tail fiber genes of bacteriophage P2: Evidence for horizontal transfer of tail fiber genes among unrelated bacteriophages. *Journal of Bacteriology, 174*(5), 1462-1477.

Hara, O. & Hutchinson, C.R. (1992). A macrolide 3-O-acyltransferase gene from the midecamycin-producing species *Streptomyces mycarofaciens*. *Journal of Bacteriology*, 174(15), 5141-5144.

Harley, C.B., and Reynolds, R.P. (1987). Analysis of *E. coli* promoter sequences. *Nucleic Acids Research, 15*(5), 2343-2361.

Heckman, K. L., & Pease, L. R. (2007). Gene splicing and mutagenesis by PCR-driven overlap extension. *Nature Protocols, 2*(4), 924-932.

Hendrix, R. W. (1988). Tail length determination in double-stranded DNA bacteriophages. *Current Topics in Microbiology and Immunology, 136*, 21-29.

Hendrix, R. W. (2002). Bacteriophages: Evolution of the majority. *Theoretical Population Biology, 61*(4), 471-480.

Hendrix, R. W. (2003). Bacteriophage genomics. *Current Opinion in Microbiology, 6*(5), 506-511.

Hendrix, R. W., Smith, M. C. M., Burns, R. N., Ford, M. E., & Hatfull, G. F. (1999). Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage. *Proceedings of the National Academy of Sciences of the USA, 96*(5), 2192-2197.

Hens, D. K., Chatterjee, N. C., & Kumar, R. (2006). New temperate DNA phage BcP15 acts as a drug resistance vector. *Archives of Virology, 151*(7), 1345-1353.

Hens, D. K., Ghosh, A. N., & Kumar, R. (2005). A new small temperate DNA phage BcP15 isolated from *Burkholderia cepacia* DR11. *Archives of Virology, 150*(12), 2421-2428.

Higgins, C. F. (1992). ABC transporters: From microorganisms to man. *Annual Review of Cell Biology, 8*, 67-113.

Highlander, S. K., Weissenberger, S., Alvarez, L. E., Weinstock, G. M., & Berget, P. B. (2006). Complete nucleotide sequence of a P2 family lysogenic bacteriophage, ϕMhaA1-PHL101, from *Mannheimia haemolytica* serotype A1. *Virology, 350*(1), 79-89.

Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., & Pease, L. R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene, 77*(1), 51-59.

Hoeltje, J. V., Mirelman, D., Sharon, N., & Schwarz, U. (1975). Novel type of murein transglycosylase in *Escherichia coli*. *Journal of Bacteriology, 124*(3), 1067-1076.

Holloway B. W., Egan, J. B. & Monk, M. (1960). Lysogeny in *Pseudomonas aeruginosa*. *Australian Journal of Experimental Biololgy and Medical Science, 38*, 321-330.

Horton, R. M. (1995). PCR-mediated recombination and mutagenesis. SOEing together tailor-made genes. *Molecular Biotechnology, 3*(2), 93-99.

Horton, R. M., Cai, Z., Ho, S. N., & Pease, L. R. (1990). Gene splicing by overlap extension: Tailor-made genes using the polymerase chain reaction. *BioTechniques, 8*(5), 528-535.

Hoskisson, P. A., & Smith, M. C. (2007). Hypervariation and phase variation in the bacteriophage 'resistome'. *Current Opinion in Microbiology, 10*(4), 396-400.

Huang, Y., & Wong, P. T. W. E. -. (1998). Effect of *Burkholderia* (*Pseudomonas*) *cepacia* and soil type on the control of crown rot in wheat. *Plant and Soil, 203*(1), 103-108.

Hull, R. A., Gill, G. S., & Curtiss III, R. (1978). Genetic characterization of Mu-like bacteriophage D108. *Journal of Virology, 27*(3), 513-518.

Isles, A., Maclusky, I., Corey, M., Gold, R., Prober, C., Fleming, P., & Levison, H. (1984). *Pseudomonas cepacia* infection in cystic fibrosis: an emerging problem. *Journal of Pediatrics, 104*(2), 206-210.

Jado, I., Lopez, R., Garcia, E., Fenoll, A., Casal, J., & Garcia, P. (2003). Phage lytic enzymes as therapy for antibiotic resistant *Streptococcus pneumoniae* infection in a murine sepsis model. *Journal of Antimicrobial Chemotherapy, 52*(6), 967-973.

Jakob, U., Muse, W., Eser, M., & Bardwell, J.C. (1999). Chaperone activity with a redox switch. *Cell. 96*, 341-352.

Janisiewicz, W.J. & Roitman, J. (1988). Biological control of blue mold and gray mold on apple and pear with *Pseudomonas cepacia. Phytopathology, 78*, 1697-1700.

Johnson, K. & Lory, S. (1987). Characterization of *Pseudomonas aeruginosa* mutants with altered pilation. *Journal of Bacteriology. 169*, 5663-5667.

Journet, L., Agrain, C., Broz, P., & Cornelis, G. R. (2003). The needle length of bacterial injectisomes is determined by a molecular ruler. *Science, 302*(5651), 1757-1760.

Kanamaru, S., Leiman, P. G., Kostyuchenko, V. A., Chipman, P. R., Mesyanzhinov, V. V., Arisaka, F., *et al.* (2002). Structure of the cell-puncturing device of bacteriophage T4. *Nature, 415*(6871), 553-557.

Katsura, I. (1987). Determination of bacteriophage λ tail length by a protein ruler. *Nature, 326*(6117), 73-75.

Katsura, I. (1990). Mechanism of length determination in bacteriophage lambda tails. *Advances in Biophysics, 26*, 1-18.

Katsura, I., & Hendrix, R. W. (1984). Length determination in bacteriophage lambda tails. *Cell, 39*(3 II), 691-698.

Kilbane, J.J., Chatterjee, D.K., Karns, J.S., Kellogg, S.T., Chakrabarty, A.M. (1982). Biodegradation of 2,4,5-Trichlorophenoxyacetic acid by a pure culture of *Pseudomonas cepacia. Applied and Environmental Microbiology, 44*(1), 72-78.

King, J. (1971). Bacteriophage T4 tail assembly: Four steps in core formation. *Journal of Molecular Biology, 58*(3), 693-709.

King, J., and Casjens, S. (1974). Catalytic head assembling protein in virus

morphogenesis. *Nature, 251*(5471), 112-119.

Klee, S. R., Nassif, X., Kusecek, B., Merker, P., Beretti, J., Achtman, M., *et al.* (2000). Molecular and biological analysis of eight genetic islands that distinguish *Neisseria meningitidis* from the closely related pathogen *Neisseria gonorrhoeae. Infection and Immunity, 68*(4), 2082-2095.

Kostyuchenko, V. A., Chipman, P. R., Leiman, P. G., Arisaka, F., Mesyanzhinov, V. V., & Rossmann, M. G. (2005). The tail structure of bacteriophage T4 and its mechanism of contraction. *Nature Structural and Molecular Biology, 12*(9), 810-813.

Kostyuchenko, V. A., Leiman, P. G., Chipman, P. R., Kanamaru, S., Van Raaij, M. J., Arisaka, F., *et al.* (2003). Three-dimensional structure of bacteriophage T4 baseplate. *Nature Structural Biology, 10*(9), 688-693.

Köthe, M., Antl, M., Huber, B., Stoecker, K., Ebrecht, D., Steinmetz, I., *et al.* (2003). Killing of *Caenorhabditis elegans* by *Burkholderia cepacia* is controlled by the cep quorum-sensing system. *Cellular Microbiology, 5*(5), 343-351.

Krause, H. M., Rothwell, M.R., and Higgins, N.P. (1983). The early promoter of bacteriophage Mu: definition of the site of transcript initiation. *Nucleic Acids Research. 11*(16), 5483-5495.

Krause, H.M., and Higgens, N.P. (1986). Positive and negative regulation of the Mu operator by Mu repression and *Escerichia coli* integration host factor. *Journal of Biological Chemistry. 261*(8), 3744-3752.

Lambiase, A., Raia, V., del Pezzo, M., Sepe, A., Carnovale, V., and Rossano, F. (2006). Microbiology of airway disease in a cohort of patients with cystic fibrosis. *BMC Infectious Diseases*, 6,4.

Lamothe, J., Thyssen, S., & Valvano, M. A. (2004). *Burkholderia cepacia* complex isolates survive intracellularly without replication within acidic vacuoles of acanthamoeba polyphaga. *Cellular Microbiology, 6*(12), 1127-1138.

Landschulz, W.H., Johnson, P.F., and McKnight, S.L. (1988). The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science. 240*(4860), 1759-1764.

Lang, L.H. (2006). FDA approves use of bacteriophages to be added to meat and poultry products. *Gastroenterology, 131*(5), 1370.

Langley, R., Kenna, D.T., Vandamme, P., Ure, R., & Govan, J.R.W. (2003). Lysogeny and bacteriophage host range within the *Burkholderia cepacia* complex. *Journal of Medical Microbiology, 52*(6), 483-490.

Larkum, N.W. (1929) Bacteriophage from Public Health standpoint. *American Journal of Public Health. 19*, 31-36.

Lemberg, K. M., Schweidenback, C. T. H., & Baker, T. A. (2007). The dynamic mu transpososome: MuB activation prevents disintegration. *Journal of Molecular Biology, 374*(5), 1158-1171.

Leung, P.C., Teplow, D.B., and Harshey, R.M. (1989). Interaction of distinct domains in Mu transposase with Mu DNA ends and an internal transpositional enhancer. *Nature. 338*, 656-658.

Levchenko, I., Yamauchi, M., & Baker, T. A. (1997). ClpX and MuB interact with overlapping regions of mu transposase: Implications for control of the transposition pathway. *Genes and Development, 11*(12), 1561-1572.

Levin, M.E., Hendrix, R.W., and Casjens, S.R. (1992). A programmed translational frameshift is required for the synthesis of a bacteriophage λ tail assembly protein. *Journal of Molecular Biology, 234*(1), 124-139.

Loessner, M. J. (2005). Bacteriophage endolysins - current state of research and applications. *Current Opinion in Microbiology, 8*(4), 480-487.

Lynch, K.H., and Dennis, J.J. (2008). Development of a species-specific *fur* gene-based method for identification of the *Burkholderia cepacia* complex. *Journal of Clinical Microbiology, 46*(2), 447-455.

Mahenthiralingam, E., Simpson, D., & Speert, D. (1997). Identification and characterization of a novel DNA marker associated with epidemic *Burkholderia cepacia* strains recovered from patients with cystic fibrosis. *Journal of Clinical Microbiology, 35*(4), 808-816.

Mahenthiralingam E., Coenye T., Chung J.W., Speert D.P., Govan J.R., Taylor P., & Vandamme P. (2000). Diagnostically and experimentally useful panel of strains from the *Burkholderia cepacia complex. Journal of Clinical Microbiology, 38*(2), 1042-1047.

Mahenthiralingam, E., Vandamme, P., Campbell, M.E., Henry, D.A., Gravelle, A.M., Wong, L.T.K., Davidson, A.G.F., Wilcox, P.G., Nakielna, B., and Speert, D.P. (2001). Infection with *Burkholderia cepacia* complex genomovars in patients with cystic fibrosis: Virulent transmissible strains of genomovar III can replace *Burkholderia multivorans. Clinical Infectious Diseases, 33*(9), 1469-1475.

Maniloff, J. & Ackermann, H.-W. (1988). Taxonomy of bacterial viruses: Establishment of tailed virus genera and the order Caudovirales. *Archives of Virology, 143*(10), 2051-2063.

Markey, K. M., Glendinning, K. J., Morgan, J. A. W., Hart, C. A., & Winstanley, C. (2006). *Caenorhabditis elegans* killing assay as an infection model to study the role of type III secretion in *Burkholderia cenocepacia*. *Journal of Medical Microbiology, 55*(7), 967-969.

Marolda, C. L., Hauröder, B., John, M. A., Michel, R., & Valvano, M. A. (1999). Intracellular survival and saprophytic growth of isolates from the *Burkholderia cepacia* complex in free-living amoebae. *Microbiology, 145*(7), 1509-1517.

Marshall-Batty, K. R., & Nakai, H. (2003). Trans-targeting of the phage mu repressor is promoted by conformational changes that expose its ClpX recognition determinant. *Journal of Biological Chemistry, 278*(3), 1612-1617.

Matsumoto, H., Itoh, Y., Ohta, S., & Terawaki, Y. (1986). A generalized transducing phage of *Pseudomonas cepacia*. *Journal of General Microbiology, 132*(9), 2583-2586.

Matsuzai, S., Rashel, M., Uchiyama, J., Sakurai, S., Ujihara, T., Kuroda, M., Ikeuchi, M., Tani, T., Fujieda, M., Wakiguchi, H., and Imai, S. (2005). Bacteriophage therapy: a revitalized therapy against bacterial infectious disease. *Journal of Infection and Chemotherapy, 11*(5), 211-219.

McCloskey, R. V. (1973). Scarlet fever and necrotizing fascitis caused by coagulase-positive hemolytic *Staphylococcus aureus*, phage type 85. *Annals of Internal Medicine, 78*(1), 85-87.

McCorquodale, D. J., Shaw, A. R., Shaw, P. K., & Chinnadurai, G. (1977). Pre early polypeptides of bacteriophages T5 and BF23. *Journal of Virology, 22*(2), 480-488.

McLoughlin, T.J., Quinn, J.P., Bettermann, A., and Bookland, R. (1992). *Pseudomonas cepacia* suppression of sunflower wilt fungus and role of antifungal compounds in controlling the disease. *Applied and Environmental Microbiology, 58*(5), 1760-1763.

Merril, C.R., Biswas, B., Carlton, R., Jensen, N.C., Creed, G.J., Zullo, S., & Adhya, S. (1996). Long-circulating bacteriophages as antibacterial agents. *Proceedings of the National Acadamy of Science of the USA, 93*(8), 3188-3192.

Meury, J., & Kohiyama, M. (1991). Role of heat shock protein DnaK in osmotic adaptation of *Escherichia coli*. *Journal of Bacteriology, 173*(14), 4404-4410.

Miller, E. S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T., & Rüger, W. (2003). Bacteriophage T4 genome. *Microbiology and Molecular Biology*

Reviews, 67(1), 86-156.

Mizuuchi, M., & Mizuuchi, K. (1989). Efficient Mu transposition requires interaction of transposase with a DNA sequence at the Mu operator: Implications for regulation. *Cell, 58*(2), 399-408.

Mizuuchi, M., Rice, P. A., Wardle, S. J., Haniford, D. B., & Mizuuchi, K. (2007). Control of transposase activity within a transpososome by the configuration of the flanking DNA segment of the transposon. *Proceedings of the National Academy of Sciences of the USA, 104*(37), 14622-14627.

Moak, M., & Molineux, I. J. (2000). Role of the Gp16 lytic transglycosylase motif in bacteriophage T7 virions at the initiation of infection. *Molecular Microbiology, 37*(2), 345-355.

Moore, R. A., Reckseidler-Zenteno, S., Kim, H., Nierman, W., Yu, Y., Tuanyok, A., *et al.* (2004). Contribution of gene loss to the pathogenic evolution of *Burkholderia pseudomallei* and *Burkholderia mallei. Infection and Immunity, 72*(7), 4172-4187.

Moreau, S., Blanco, C., & Trautwetter, A. (1999). Site-specific integration of corynephage ϕ16: Construction of an integration vector. *Microbiology, 145*(3), 539-548.

Morgan, G. J., Hatfull, G. F., Casjens, S., & Hendrix, R. W. (2002). Bacteriophage Mu genome sequence: Analysis and comparison with Mu-like prophages in *Haemophilus, Neisseria* and *Deinococcus. Journal of Molecular Biology, 317*(3), 337-359.

Naigamwalla, D. Z., & Chaconas, G. (1997). A new set of mu DNA transposition intermediates: Alternate pathways of target capture preceding strand transfer. *EMBO Journal, 16*(17), 5227-5234.

Nakagawa, I., Kurokawa, K., Yamashita, A., Nakata, M., Tomiyasu, Y., Okahashi, N., *et al.* (2003). Genome sequence of an M3 strain of *Streptococcus pyogenes* reveals a large-scale genomic rearrangement in invasive strains and new insights into phage evolution. *Genome Research, 13*(6 A), 1042-1055.

Nierman, W. C., DeShazer, D., Kim, H. S., Tettelin, H., Nelson, K. E., Feldblyum, T., *et al.* (2004). Structural flexibility in the *Burkholderia mallei* genome. *Proceedings of the National Academy of Sciences of the USA, 101*(39), 14246-14251.

Nzula, S., Vandamme, P., & Govan, J. R. W. (2000). Sensitivity of the *Burkholderia cepacia* complex and *Pseudomonas aeruginosa* to transducing bacteriophages. *FEMS Immunology and Medical Microbiology, 28*(4), 307-312.

O'Brien, A. D., Newland, J.W., Miller, S.F., Holmes, R.K., Smith, H.W., & Formal, S.B. (1984). Shiga-like toxin converting phage from *Escherichia coli* strains that cause hemorrhagic colitis or infantile diarrhea. *Science, 226*(4675), 694-696.

Oie, S., & Kamiya, A. (1996). Microbial contamination of antiseptics and disinfectants. *American Journal of Infection Control, 24*(5), 389-395.

Osawa, R., Iyoda, S., Nakayama, S., Wada, A., Yamai, S., & Watanabe, H. (2000). Genotypic variations of shiga toxin-converting phages from enterohaemorrhagic *Escherichia coli* O157:H7 isolates. *Journal of Medical Microbiology, 49*(6), 565-574.

Ramette, A., LiPuma, J. J., & Tiedje, J. M. (2005). Species abundance and diversity of *Burkholderia cepacia* complex in the environment. *Applied and Environmental Microbiology, 71*(3), 1193-1201.

Ranquet, C., Toussaint, A., De Jong, H., Maenhaut-Michel, G., & Geiselmann, J. (2005). Control of bacteriophage Mu lysogenic repression. *Journal of Molecular Biology, 353*(1), 186-195.

Rehmat, S., & Shapiro, J. A. (1983). Insertion and replication of the *Pseudomonas aeruginosa* mutator phage D3112. *Molecular and General Genetics, 192*(3), 416-423.

Sambrook, J., Fritsch, E.F., & Maniatis, T. (1989). Molecular cloning: a laboratory manual. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

São-José, C., Parreira, R., Vieira, G., and Santos, M.A. (2000). The N-terminal region of the *Oenococcus oeni* Bacteriophage fOg44 lysin behaves as a bona fide signal peptide in *Escherichia coli* and a cis-inhibitory element, preventing lytic activity on Oenococcal cells. *Journal of Bacteriology*, 182(20), 5823-5831.

Schicklmaier, P., Moser, E., Wieland, T., Rabsch, W., & Schmieger, H. (1998). A and comparative study on the frequency of prophages among natural isolates of *Salmonella Escherichia coli* with emphasis on generalized transducers. *Antonie Van Leeuwenhoek, International Journal of General and Molecular Microbiology. 73*(1), 49-54.

Schmieger, H., & Schicklmaier, P. (1999). Transduction of multiple drug resistance of *Salmonella enterica* serovar Typhimurium DT104. *FEMS Microbiology Letters, 170*(1), 251-256T

Scordilis, G.E., Ree, H., & Lessie, T.G. (1987). Identification of transposable elements which activate gene expression in *Pseudomonas cepacia. Journal*

of Bacteriology, 169(1), 8-13.

Seed, K. D., & Dennis, J. J. (2005). Isolation and characterization of bacteriophages of the *Burkholderia cepacia* complex. *FEMS Microbiology Letters, 251*(2), 273-280.

Seed, K. D., & Dennis, J. J. (2008). Development of galleria mellonella as an alternative infection model for the *Burkholderia cepacia* complex. *Infection and Immunity, 76*(3), 1267-1275.

Shanmuganatham, K. K., Ravichandran, M., Howe, M. M., & Park, H. (2007). Crystallization and preliminary X-ray analysis of phage Mu activator protein C in a complex with promoter DNA. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications, 63*(7), 620-623.

Shapiro, J. A. (1979). Molecular model for the transposition and replication of bacteriophage Mu and other transposable elements. *Proceedings of the National Academy of Sciences of the USA, 76*(4), 1933-1937.

Sharp, R. (2001). Bacteriophages: Biology and history. *Journal of Chemical Technology & Biotechnology, 76*(7), 667-672.

Slauch, J.M., Lee, A.A., Mahan, M.J., & Mekalanos, J.J. (1996). Molecular characterization of the *oafA* locus responsible for acetylation of *Salmonella typhimurium* O-antigen: *oafA* is a member of a family of integral membrane trans-acylases. *Journal of Bacteriology*, 178(20), 5904-5909.

Smith, M. J., & Jeddeloh, J. A. (2005). DNA methylation in lysogens of pathogenic *Burkholderia* spp. requires prophage induction and is restricted to excised phage DNA. *Journal of Bacteriology, 187*(3), 1196-1200.

Snyder Jr., C. E., & Benzinger, R. H. (1981). Second-step transfer of bacteriophage T5 DNA: Purification and characterization of the T5 gene A2 protein. *Journal of Virology, 40*(1), 248-257.

Sokolsky, T. D., & Baker, T. A. (2003). DNA gyrase requirements distinguish the alternate pathways of Mu transposition. *Molecular Microbiology, 47*(2), 397-409.

Speert, D. P., D. Henry, P. Vandamme, M. Corey, and E. Mahenthiralingam. (2002). Epidemiology of *Burkholderia cepacia* complex in patients with cystic fibrosis, Canada. *Emerging Infectious Diseases, 8*(2), 181–187.

Straub, M.E. & Applebaum, M. (1933). Studies on commercial bacteriophage products. *Journal of the American Medical Association. 100*, 110-113.

Summer, E. J., Berry, J., Tran, T. A. T., Niu, L., Struck, D. K., & Young, R.

(2007). Rz/Rz1 lysis gene equivalents in phages of Gram-negative hosts. *Journal of Molecular Biology, 373*(5), 1098-1112.

Summer, E. J., Gill, J. J., Upton, C., Gonzalez, C. F., & Young, R. (2007). Role of phages in the pathogenesis of *Burkholderia*, or 'Where are the toxin genes in *Burkholderia* phages?'. *Current Opinion in Microbiology, 10*(4), 410-417.

Summer, E. J., Gonzalez, C. F., Carlisle, T., Mebane, L. M., Cass, A. M., Savva, C. G., et al. (2004). *Burkholderia cenocepacia* phage BcepMu and a family of Mu-like phages encoding potential pathogenesis factors. *Journal of Molecular Biology, 340*(1), 49-65.

Surette, M. G., & Chaconas, G. (1992). The Mu transpositional enhancer can function in trans: Requirement of the enhancer for synapsis but not strand cleavage. *Cell, 68*(6), 1101-1108.

Susskind, M. M., & Botstein, D. (1978). Molecular genetics of bacteriophage P22. *Microbiological Reviews, 42*(2), 385-413.

Symonds, N., Toussaint, A., van de Putte, P., and Howe, M. (1987). Phage Mu. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Takeda, S., Sasaki, T., Ritani, A., Howe, M. M., & Arisaka, F. (1998). Discovery of the tail tube gene of bacteriophage Mu and sequence analysis of the sheath and tube genes. *Biochimica Et Biophysica Acta - Gene Structure and Expression, 1399*(1), 88-92.

Tsugita, A., Inouye, M., Terzaghi, E., & Streisinger, G. (1968). Purification of bacteriophage T4 lysozyme. *Journal of Biological Chemistry, 243*(2), 391-397.

Ulrich, R. L., DeShazer, D., Brueggemann, E. E., Hines, H. B., Oyston, P. C., & Jeddeloh, J. A. (2004). Role of quorum sensing in the pathogenicity of *Burkholderia pseudomallei*. *Journal of Medical Microbiology, 53*(11), 1053-1064.

Vandamme, P., Holmes, B., Vancanneyt, M., Coenye, T., Hoste, B., Coopman R., Revets, H., Lauwers, S., Gillis, M., Kersters, K., & Govan, J.R. (1997). Occurrence of multiple genomovars of Burkholderia cepacia in cystic fibrosis patients and proposal of Burkholderia multivorans sp. nov. *International Journal of Systematic Bacteriology, 47*(4), 1188-1200.

Vandamme, P., Mahenthiralingam, E., Holmes, B., Coenye, T., Hoste, B., De Vos, P., Henry, D., & Speert, D.P. (2000). Identification and population structure of *Burkholderia stabilis* sp. nov. (formerly *Burkholderia cepacia* genomovar IV). *Journal of Clinical Microbiology*, 38(3), 1042-1047.

Vandamme, P., Henry, D., Coenye, T., Nzula, S., Vancanneyt, M., LiPuma, J.J., Speert, D.P., Govan, J.R., & Mahenthiralingam, E. (2002). *Burkholderia anthina* sp. nov. and *Burkholderia pyrrocinia*, two additional *Burkholderia cepacia* complex bacteria, may confound results of new molecular diagnostic tools. *FEMS Immunology and Medical Microbiology*, 33(2), 143-149.

Van De Putte, P., Cramer, S., & Giphart-Gassler, M. (1980). Invertible DNA determines host specificity of bacteriophage Mu. *Nature*, 286(5770), 218-222.

Van Leerdam, E., Karreman, C., & Van De Putte, P. (1982). Ner, a cro-like function of bacteriophage mu. *Virology, 123*(1), 19-28.

Vermis, K., Coenye, T., LiPuman, J.J., Mahenthiralingam, E., Nelis, H.J., & Vandamme, P. (2004). Proposal to accommodate *Burkholderia cepacia* genomovar VI as *Burkholderia dolosa* sp. nov. *International Journal of Systematic and Evolutionary Bacteriology*, 54(3), 689-691.

Waldor, M. K. & Mekalanos, J. J. (1996). Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science, 272*(5270), 1910-1913.

Wang, I., Smith, D. L., & Young, R. (2000). Holins: The protein clocks of bacteriophage infections. *Annual Review of Microbiology, 54*, 799-825.

Wang, P.W., Chu, L., and Guttman, D.S. (2004). Complete sequence and evolutionary genomic analysis of the *Pseudomonas aeruginosa* transposable bacteriophage D3112. *Journal of Bacteriology*, 186(2), 400-410.

Wiersinga, W. J., van der Poll, T., White, N. L., Day, N. P., & Peacock, S. J. (2006). Melioidosis: Insights into the pathogenicity of *Burkholderia pseudomallei*. *Nature Reviews Microbiology, 4*(4), 272-282.

Winkelstein, J.A., Marino, M.C., Johnston, R.B. Jr., Boyle, J., Curnutte, J., Gallin, J.I., Malech, H.L., Holland, S.M., Ochs, H., Quie, P., Buckley, R.H., *et al.* (2000). Chronic granulomatous disease. Report on a national registry of 368 patients. *Medicine, 79*(3), 155-169.

Xu, J., Hendrix, R., and Duda, R. (2004). Conserved translational frameshift in dsDNA bacteriophage tail assembly genes. *Molecular Cell,* 16(1), 11-21.

Yagil, E., Dolev, S., Oberto, J., Kislev, N., Ramaiah, N., & Weisberg, R. A. (1989). Determinants of site-specific recombination in the lambdoid coliphage HK022. an evolutionary change in specificity. *Journal of Molecular Biology, 207*(4), 695-717.

Young, R. (1992). Bacteriophage lysis: Mechanism and regulation.

Microbiological Reviews, 56(3), 430-481.

Zheng M., Åslund F, & Storz G. (1998). Activation of the OxyR transcription factor by reversible disulfide bond formation. *Science. 279*, 1718– 21

Chapter 6: Appendix

6. APPENDIX: Construction of an integrative plasmid for *B. cepacia* complex mutagenesis.

6.1 Introduction

Initially, recombinant DNA technology was limited to the use of *E. coli* plasmids or bacteriophage λ. With the emergence of many antibiotic resistant pathogens, there has been a recent increase in the number of genetic tools available for use in many bacteria. Cloning vectors may be modified for use in a larger number of hosts by including a broad host range replicon, for example. Vectors can be constructed containing a *mob* gene so that bacteria that are difficult to transform can receive plasmids via conjugal transfer. Expression vectors, under the control of an inducible promoter, can allow for the controlled expression of a cloned gene in a chosen host. Using such genetic tools can be important in identifying the function of unknown genes.

The *Burkholderia cepacia* complex, reviewed in section 1.1, are important opportunistic pathogens due to its intrinsic antibiotic resistance. The highly resistant nature of the complex, combined with their genomic plasticity, makes it quite difficult to develop effective treatments against these bacteria. Molecular tools that allow characterization and manipulation of these bacteria are important in understanding which genes are involved in pathogenesis, and possibly developing new antibiotic targets. However, there are relatively few tools available for use in the *B. cepacia* complex (Lefebre and Valvano, 2002; Cardona *et al.*, 2006).

In this study we attempted to construct a plasmid containing an integrase gene from bacteriophage KS9 (unpublished data). KS9 is a prophage of *B. pyrrocinia* LMG 21824 and has a fairly limited host range. This plasmid encoding the KS9 integrase should integrate into the chromosome of species of the *B. cepacia* complex at the phage attachment site. Since the initial adsorption of the phage to the host is not necessary, the host range of the plasmid should not be limited to the host range of phage KS9. We expect that by creating a plasmid that can integrate into the chromosome of certain *B. cepacia* complex strains and successfully express foreign genes we can test whether certain genes are involved directly in pathogenesis by using the *G. mellonella* larvae model (Seed and Dennis, 2008).

6.2 Materials and Methods

6.2.1 Bacterial strains and plasmids

E. coli DH5α was used in this study to maintain constructs. Plasmids and their relevant characteristics used for integrative plasmid construction are listed in Table 6.1

6.2.2 Growth Conditions

E. coli DH5α was grown on LB medium at 37°C. LB was supplemented with 100μg/mL ampicillin or 100μg/mL trimethoprim, depending on plasmid requirements. Transformed *B. cepacia* complex species were grown at 30°C on LB medium supplemented with 300μg/mL trimethoprim.

6.2.3 Isolation of DNA

Plasmid and chromosomal DNA was isolated as described in section 2.3.

6.2.4 Polymerase chain reactions

To construct the integrative plasmid, PCR was used to amplify individual fragments. Primers and annealing temperatures used in each experiment are listed in Table 6.2. Reactions were carried out as described in section 2.4.2.

6.2.5 *Ligation of PCR products and transformations*

Individual fragments were designed to have a restriction enzyme cut site at each end (Table 6.1). Fragments to be ligated together were designed to contain complementary overhanging restriction enzyme cut sites (Figure 6.1). For each ligation, 10 μL of each product was added to 1.5 μL ligation buffer (Invitrogen) and 1 μL T4 DNA ligase (Invitrogen). Ligation was carried out at room temperature and at 16°C for 4 hours. Following ligation, DNA was transformed into chemically competent *E. coli* DH5α according to product specifications and plated on LB media supplemented with 100 mg/mL trimethoprim. Between 30 and 38 colonies were routinely screened. Constructs were isolated using a QIAprep Miniprep Kit (QIAGEN). Presence of the correct construct was checked by digestion with *Eco*RI. Once the correct the construct is identified, the remaining three fragments, *mob*, $lacI^q$, and T_1T_2, can be ligated into the three-piece construct.

6.2.6 SOEing PCR

After a three-way ligation attempt failed, SOEing PCR was used to attempt to fuse OriTp^r, *lacI*^{q}, and KS9 integrase (referred to as P_a fusion), and also to fuse T1T2, MCS, and *mob* (referred to as P_b fusion). PCR was carried out as previously described (Horton *et al.*, 1990; Heckman and Pease, 2007). See Figure 6.2 for overview of SOEing PCR used in this experiment.

Table 6.1 Plasmids, chromosomes, and plasposons

Plasmid/strain/plasposon	**Relevant Characteristics**	**Reference or Source**
pBBR1-TP	Source of *mob* gene	DeShazer and Woods, 1996
pTrc99A	Source of *lacI*^{q} and terminators	Amann and Brosius 1985; Amann *et al.*, 1988
pUC19	Source of multiple cloning site; Source of P_{lac} and *lacZ*α	Invitrogen
B. pyrrocinia strain 21824	Source of integrase gene and *attP* from prophage KS9	Coenye *et al.*, 2003
pTnModO-Tp'	Source of origin of replication; Source of trimethoprim resistance cassette	Dennis and Zylstra, 1998

Table 6.2 Primers used for amplification of relevant genes/fragments

Gene	Primers (written 5′-3′)	Annealing temperature
Ori Tp^{r*}	$P1_a$: ACC TAG GCC GGC CGC GAT CGC	70°C
	$P3_a$: CGA AAG GTT TTG CAC CAT TCA AGA GCT CGC TTA GGC CAC A	
lacIq	$LacI^q{}_a$: CTT CTA GAC AAA ACC TTT CGC	48°C
	$LacI^q{}_{b2}$: ATT CTA GAT TGC GCT CAC TGC	
	$P4_a$: CAA AAC CTT TCG CGG TAT	
	$P5_a$: CGT TGC GCT CAC TGC C	50°C
KS9 integrase	$KS9_a$: TTA CTA GTG GTA TGG CCA TAT GTG	45°C
	$KS9_{b2}$: CTC CTA GGG TCT GTT TAG TTT AG	
	$P6_a$: GCC CGT CAC TCG CGT TGC CTA AAC	68°C
	TAA ACA GAC CGC	
	$P8_a$: TCC GAC GTC CGC GCT TGG CC	
Terminators	$rrnB_{a3}$: CTG CTA GCT CAT GGC GGA TAC AT	55°C
	$rrnB_{b3}$: CAG CTA GCT AGG GAA CTG CCA GG	
	$P1_b$: CAA TAT TAT TGA AGC ATT TAT	
	CAG GGT TAT TGT CTC ATG AGC GG	68°C
	$P3_b$: AAG CAT AAA GTG TAA AGC CTG GAC CGT CAA GGG ATG AGA GCG T	
MCS	MCS_{a3}: GAC CTA GGA GAT CTT CAG GGC GCG TCA GC	59°C
	MCS_b: TCC CTA GGA TTA GGC ACC CCA GGC	
		47°C
	$P4_b$: ATT AGG CAC CCC AGG C	
	$P5_b$: TCA GGG CGC GTC AGC	

(Continued)

Table 6.2 - *Continued*

Gene	Primer (written 5'-3')		Annealing temperature
mob	mob_{a2}:	TCG GAT CCC CTC GTG ATA CGC CT	60°C
	mob_{b2}:	TCG GAT CCG GAC ACC AAA AGG	
	$P6_b$:	GTG ACC GGC AGC AAA ATG TCG AAG	
		AGA CGA AGG G	71°C
	$P8_b$:	GCC TTG CGC TGC CCC CGT CG	

* Ori-Tpr fragment was obtained without PCR from cutting the *Bgl*II-*Spe*I fragment pTnModO-Tp'

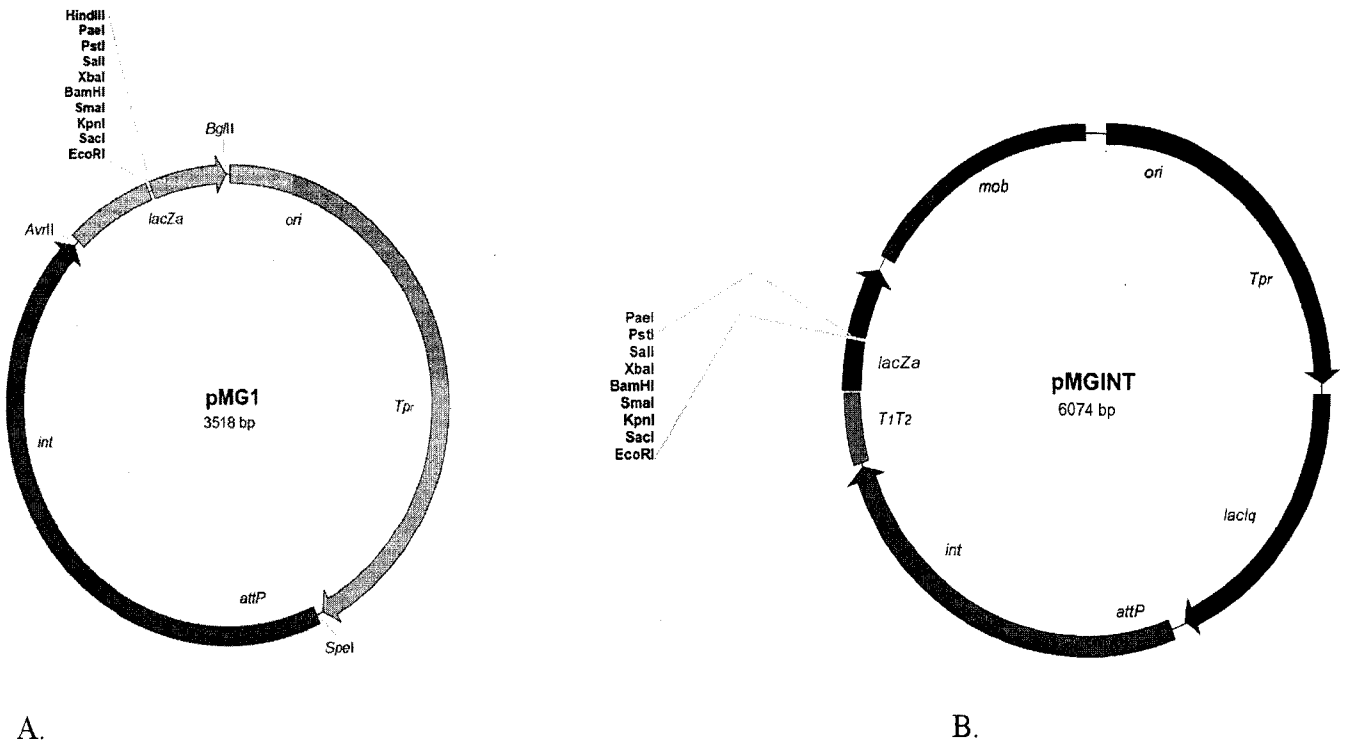


Figure 6.1 Vector construction using a three-way ligation method. A) pMG1. The expected plasmid after a successful three-way ligation. T_1T_2, *mob,* and *lacI*^{q} fragments can be ligated in at *Avr*II, *Bgl*II, and *Spe*I sites, respectively. B) Design of final integrative vector, pMGINT.

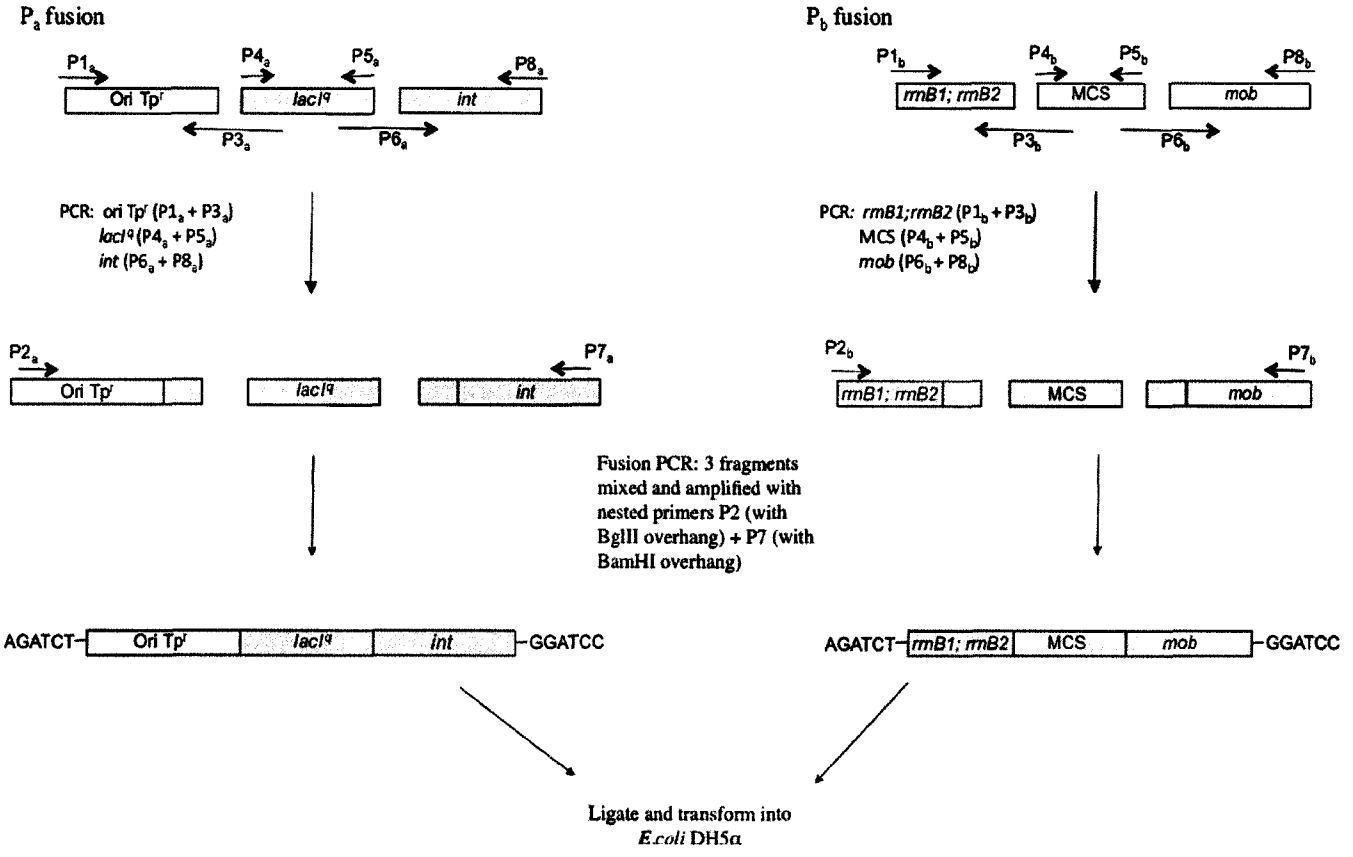


Figure 6.2 Steps for fusion PCR used in this study. Arrows labeled $P1_{a,b}$-$P7_{a,b}$ show location of primer design. Direction of arrows indicate fragments to be amplified. The first half of $P3_{a,b}$ primers are identical to the 5′ ends of *lacI*^{q} and MCS (in *lacZα*), respectively, while the last half of $P3_{a,b}$ primers are identical to the 3′ ends of Ori Tp^r and *rrnB1;rrnB2*, respectively. The first half of $P6_{a,b}$ primers are identical to the 3′ ends of *lacI*^{q} and MCS (in *lacZα*), respectively, while the last half of $P6_{a,b}$ primers are identical to the 5′ ends of *int* and *mob*, respectively.

6.3 Results and Discussion

6.3.1 Construction of the integrative vector

In this study, we attempted to create a vector that would integrate into the chromosome of *B. cepacia* complex bacteria using the *attP* and integrase gene of temperate phage KS9 (unpublished data). Because the integrative plasmid would be present in single copy, and does not require the constant presence of an antibiotic for it to remain in the cell, gene expression from this vector, as compared to a more traditional expression vector, is more representative of natural gene expression. Although KS9 has a very narrow host range, it is expected that many *B. cepacia* complex species may have the *attB* site necessary for integration. Moreau *et al.* (1999) constructed an integrative vector containing the integrase gene of corynephage ϕ16. In this case, the vector was able to integrate into the chromosome of a number of corynebacteria species that were not hosts for the ϕ16 phage (Moreau *et al.*, 1999). In order to facilitate conjugative transfer of the vector into a broad range of *B. cepacia* complex species, we designed the integrative vector to contain the *mob* region of pBBR1-Tp. The vector would also contain the P_{lac} promoter upstream from the multiple cloning site of pUC19. This site contains unique cloning sites and a promoterless *lacZ*. The T_1T_2 terminators from pTrc99A would prevent upstream transcription from entering the *lacZ* gene. The *lacI*^{q} gene would also be present on this vector to control transcription from the P_{lac} promoter; therefore we could control the expression of genes that have been cloned into the MCS. In order to replicate in its *E. coli* host, the vector was also designed to contain a narrow host-range origin

of replication to allow maintenance, but would not be able to replicate in *B. cepacia* complex species. To select for bacteria that received the vector via conjugation, the vector would also contain a trimethoprim resistance cassette. Although some *B. cepacia* complex strains do exhibit some resistance to trimethoprim, using a high concentration (300 mg/mL) should allow for selection.

6.3.2 Construction of the integrative vector using ligation assisted by restriction enzymes

There have been a number of integrative vectors constructed that employ phage integrase enzymes to integrate a vector into the bacterial chromosome. Yang *et al.* (2002) created an integrative vector using the integrase gene from the phage ϕFC1. In this case, a fragment containing the *attP* and integrase gene was inserted into a plasmid derived from pUC19, transformed into *E. faecalis* KBL 707 and was successfully integrated into the chromosome (Yang *et al.*, 2002). For the creation of our integrative vector, however, a backbone plasmid with the desired characteristics was not available, and so each fragment was obtained from several different plasmids (Table 6.1).

Initially, we attempted to construct our vector by carrying out a three-way ligation between the MCS fragment, the *attP/integrase* fragment, and the *ori-Tp*^{r} fragment, which would result in a three piece vector with three restriction sites; *Bgl*II, *Avr*II, and *Spe*I. In these experiments the vector construction would involve the use of enzyme pairs with different but compatible ends (Cost, 2007). *Bgl*II and *Bam*HI, *Avr*II and *Nhe*I, and *Spe*I and *Xba*I were chosen so that,

following a ligation, recognition sites of both precursor enzymes would be eliminated. Following a successful three-way ligation, the *LacI*^{q}, *mob*, and terminator fragments would be added to the construct (pMG1) by first digesting pMG1 at each restriction site and then ligating each fragment to pMG1 (Figure 6.1). After a number of ligation attempts (each attempt screened between 30-38 colonies) and the use of various ligation buffers, the desired ligation product was not obtained. Three-way ligations, and sometimes four-way ligations, are routinely used in experiments, and it is unknown why none of the clones received the desired construct. One problem could be insufficient digestion of the PCR fragments before ligating. Perhaps more bases are needed before the restriction site on the 5′-end of the fragments for successful digestion using each restriction enzyme. In our design, each fragment is amplified using primers containing 2 bp, followed by the desired restriction site, and then between 12 and 15 bp of sequence that will anneal to the template. It is possible that more than two bases were needed on the other end of the cut site in order for the enzymes to cut efficiently. To overcome this problem, PCR products can be cloned, with blunt ends, into a vector such as pJET (Fermentas). Fragments can then be cut out of the vector with the desired restriction enzymes and a three-way ligation may ligate the three fragments together. The remaining fragments, as described previously, can then be inserted into the three-fragment construct.

Another, but unlikely, explanation is that the construct may be integrating into the *E. coli* DH5α genome, and so a plasmid prep will not isolate the correct

construct. In their construction of an integrative vector from ϕFC1 integrase, Yang *et al.* (2002) found that their vector could integrate into both *E. coli* and *E. faecalis*, showing that these plasmids can integrate the chromosomes of species different from the phages bacterial host. However, because *E. coli* and *B. cepacia* complex bacteria are very different, unlike *E. coli* and *E. faecalis*, which are both enterococci, it would seem unlikely that *E. coli* could have the proper *attB* site necessary for a *B. cepacia* complex phage to recombine. Because the integrase may have its own promoter, and the vector is capable of replicating in *E. coli*, the integrase could have been expressed in DH5α, and PCR should have been carried out to determine if this could be happening.

6.3.3 Construction using SOEing PCR

Horton *et al.* (1989) developed a PCR fusion procedure in which two different fragments of DNA can be "spliced" or "fused" together by using overlap extension PCR (Ho *et al.*, 1989). This technique is referred to as SOEing PCR, or Splicing by Overlap Extension. Using this technique, initial PCR products are obtained that contain an end gene segment that overlaps with the end of another PCR product. These fragments are then be used as templates for a final PCR that should fuse the three fragments together to create one product (Heckman and Pease, 2007). Because we were unable to construct an integrative vector using a three-way ligation method, we attempted to use this method to fuse the fragments, as illustrated in Figure 6.2. Normally, this procedure is used for gene splicing and mutagenesis, as the name suggests, but the same method could be employed for

plasmid construction following PCR of individual fragments. However, as with the three-way ligation, the SOEing PCR was unsuccessful. Although individual fragments could be isolated, the fusion PCR continuously gave no product. While the primer design involved in this SOEing PCR method is much more complex than the previous method, if successful, it would have been a useful and rapid way to construct an integrative plasmid, and it does not require the use of restriction sites to fuse genes.

The negative results outlined suggest that the best, or easiest, way to construct an integrative plasmid may be to start with a backbone plasmid already containing some of the desired characteristics, or perhaps to first clone all fragments into a vector. While a backbone plasmid was not available with the desired characteristics for the vector we have designed, perhaps by altering the design we could have found a plasmid that already has some of the required characteristics. Alternatively, subsequent experiments utilizing a vector for cloning PCR products, such as pJET (Fermentas), may prove to be successful. Although initial experiments were unsuccessful, it may be possible to utilize the primers designed in future experiments. We expect that creation of an integrative vector for the *B. cepacia* complex is possible and could prove to be a useful tool for analysis of *B. cepacia* complex genes.