University of Alberta

Genomic analysis and modification of *Burkholderia cepacia* complex bacteriophages

by

Karlene Heather Lynch

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

Doctor of Philosophy
in
Microbiology and Biotechnology

Department of Biological Sciences

Spring 2012
Edmonton, Alberta

Your file Votre référence
ISBN: 978-0-494-87891-0

Our file Notre référence
ISBN: 978-0-494-87891-0

For Mom and Dad

The *Burkholderia cepacia* complex (Bcc) is a group of seventeen Gram-negative predominantly environmental bacterial species that cause potentially fatal infections in cystic fibrosis (CF) patients. Although its prevalence in these individuals is lower than that of *Staphylococcus aureus* and *Pseudomonas aeruginosa*, the Bcc remains a serious problem in the CF community because of the transmissibility, pathogenicity, and inherent antibiotic resistance of these organisms. An alternative treatment for Bcc infections that is currently being developed is phage therapy, the clinical use of viruses that infect bacteria. In order to assess the suitability of individual phage isolates for therapeutic use, we have determined and analyzed the complete genome sequences of a panel of ten Bcc-specific phages. These sequences range from 32 to 62 kilobase pairs in length and encode a broad range of proteins with a gradient of relatedness to phage and bacterial gene products from *Burkholderia* and other genera. Although the majority of these phages were found not to encode virulence factors, their temperate nature may be considered a drawback with respect to their potential for use in a phage therapy protocol. To circumvent this problem, we engineered a lytic mutant of a *Burkholderia pyrrocinia* prophage by knocking out its putative repressor gene. The resulting phage did not form stable lysogens and was active against a CF epidemic strain in an invertebrate infection model, thus providing a proof-of-principle that temperate phages can be engineered to become lytic and that these constructs are active *in vivo*. Both the genomic characterization and subsequent engineering and modification of Bcc-specific phages are fundamental to the development of an effective phage therapy strategy for these bacteria.

Acknowledgements

The work presented here would not have been possible without the guidance and encouragement of an extraordinary group of people.

Thanks to Dr. Jonathan Dennis, who was willing to take on a shaky-handed eighteen year old back in the summer of 2004. Thank you for your tutelage and your friendship (and thanks to Dr. Kimberley Seed for teaching me what tutelage means).

Thanks to all of my friends in the Dennis lab for your support, your insightfulness, and for making the lab a great place to be. A special thank you to Dr. Seed, Amanda Goudie, and Diana Semler, my sisters in the Dennis lab family.

Thanks to Dr. Julia Foght, Dr. Susan Jensen, Dr. Brenda Leskiw, Dr. Stefan Pukatzki, and Dr. Tracy Raivio for sharing your infinite wisdom and for kindly tolerating my umpteen requests for reference letters.

Thanks to our collaborators, particularly Dr. Paul Stothard, Dr. David Wishart, Dr. Warren Finlay, Dr. Reinhard Vehring, and Helena Orszanska and to the talented staff of the Molecular Biology Service Unit and the Advanced Microscopy Facility.

Thanks to NSERC, Alberta Innovates – Health Solutions, CF Canada, and the Killam Trusts for allowing me to do what I love without having to subsist on ramen noodles.

Thanks to Miles Peterson for your love and encouragement, for publication-quality genome maps, for your uncanny proofreading abilities, and for wholeheartedly embracing my inherent dorkiness.

Finally, thanks to Mom and Dad for taking me to grad school *in utero* the first time around, for buying me the little red microscope from the Toys "R" Us, for reminding me to "stay in school" every time we received poor service at a retail establishment, for driving me to the lab through countless snowstorms to start overnight cultures, for being proud of me no matter how many times my PCRs failed, and for always reminding me to keep going.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

aa	amino acid
AMP	adenosine monophosphate
Bcc	*Burkholderia cepacia* complex
BCSA	*Burkholderia cepacia* selective agar
BLAST	Basic Local Alignment Search Tool
bp or kbp	base pair or kilobase pair
BRED	bacteriophage recombineering of electroporated DNA
°C	degrees Celsius
CAMP	cationic antimicrobial peptide
CF	cystic fibrosis
CFTR	cystic fibrosis transmembrane conductance regulator
CFU	colony forming units
CGD	chronic granulomatous disease
DNA	deoxyribonucleic acid
EOP	efficiency of plating
EPA	Environmental Protection Agency
EPS	exopolysaccharide
g or mg or μg	gram or milligram or microgram
GCHII	GTP cyclohydrolase II
GNAT	GCN5-related *N*-acetyltransferase
h	hour
HEG	homing endonuclease gene

ID	identity
IS	insertion sequence
IVET	*in vivo* expression technology
l or ml or µl	liter or milliliter or microliter
LB	Luria-Bertani
LD_{50}	50% lethal dose
LES	Liverpool Epidemic Strain
LPS	lipopolysaccharide
M or mM	molar or millimolar
min	minute
MLST	multilocus sequence typing
mm or µm	millimeter or micrometer
MOI	multiplicity of infection
MRSA	methicillin-resistant *Staphylococcus aureus*
N/A	not applicable
NCBI	National Center for Biotechnology Information
OD	optical density
ORF	open reading frame
PCR	polymerase chain reaction
PFU	plaque forming units
RBS	ribosome binding site
RFLP	restriction fragment length polymorphism
RND	resistance-nodulation-division

RT	reverse transcriptase
SM	suspension medium
SNP	single nucleotide polymorphism
Tc	tetracycline
Tp	trimethoprim
TR	template repeat
tRNA	transfer ribonucleic acid
vol	volume
VR	variable repeat
VSP	very short patch
Vsr	very short patch repair
wt	weight

Chapter 1

Introduction

Portions of this chapter have been published as:

Lynch, K. H., and J. J. Dennis. 2011. *Burkholderia*, p. 723-736. *In* D. Liu (ed.), Molecular Detection of Human Bacterial Pathogens. CRC Press, Boca Raton, FL.

Lynch, K. H., and J. J. Dennis. In press. Genomic analysis and modification of *Burkholderia cepacia* complex bacteriophages: Cangene Gold Medal Lecture. Can. J. Microbiol.

Semler, D. D., K. H. Lynch, and J. J. Dennis. In press. The promise of bacteriophage therapy for *Burkholderia cepacia* complex respiratory infections. Front. Cell. Inf. Microbio.

BURKHOLDERIA CEPACIA COMPLEX (Bcc)

Isolation and taxonomy. Originally classified as phytopathogens, the first isolates of the *Burkholderia cepacia* complex (Bcc) were identified as the causative agent of sour skin onion rot by W. H. Burkholder (34). The original description of these bacteria noted that the cells were Gram-negative rods with polar flagella and an average length of 1.9 μm (34). Although the taxonomy of the Bcc is constantly evolving, this group is currently comprised of seventeen species (299). As these bacteria were first isolated from onions, they were originally named *Pseudomonas cepacia* (cepa meaning 'onion' in Latin) (34). In 1992, the novel genus *Burkholderia* was created following assessment of molecular and phenotypic characteristics and *Pseudomonas cepacia* was renamed *Burkholderia cepacia* (326).

The first description of the Bcc as a group was published in 1997. Following DNA-DNA hybridization analysis, *B. cepacia* isolates could be divided into five distinct groups (302). Each of these groups was designated as a genomic species or genomovar: I, II (identified as *Burkholderia multivorans*), III, IV, and V (previously named *Burkholderia vietnamiensis*) (90, 302). In subsequent years, the genomovars were redesignated as species: I as *B. cepacia*, III as *Burkholderia cenocepacia*, and IV as *Burkholderia stabilis* (301, 303). Four new genomovars/species were added between 2001 and 2004 (VI/*Burkholderia dolosa*, VII/*Burkholderia ambifaria*, VIII/*Burkholderia anthina*, and IX/*Burkholderia pyrrocinia*) such that, until 2008, the complex consisted of nine different species (56, 57, 300, 308). Based on *recA* sequence analysis (discussed

in the Appendix) and other tests, an additional eight species were added to the Bcc in 2008 and 2009: *Burkholderia latens*, *Burkholderia diffusa*, *Burkholderia arboris*, *Burkholderia seminalis, Burkholderia metallica*, *Burkholderia ubonensis*, *Burkholderia contaminans*, and *Burkholderia lata* (304, 305). As isolates have been found that occupy novel *recA* and multilocus sequence typing (MLST) clusters not included in the current taxonomy, it is likely that the number of species present in the Bcc will increase in the future (304).

Environmental significance. Although this thesis focuses on the Bcc in the context of clinical infection, it is important to note that these species have important roles in both environmental and industrial settings as well. Bcc bacteria can be isolated from many locations in the environment, including surface water (307), sewage (200), plants, and soil/rhizospheres (34, 57, 90, 300, 302, 304, 305). These organisms can be found as contaminants in industrial processes and have caused significant clinical problems when present in medical products such as trypan blue, chlorhexidine, heparin, and povidone-iodine (121, 133, 226, 289, 327).

Bcc bacteria have three characteristics that had historically made them promising candidates for both industrial and agricultural use. First, Bcc species are able to degrade a wide variety of toxic organic compounds, including phenol, trichloroethylene, and diesel fuel (84, 207). Second, these bacteria produce antimicrobial, antifungal, and antiparasitic agents that have been found to be active against a wide variety of plant pathogens (77, 159). Third, Bcc strains have been shown to have nitrogenase activity and thus the ability to fix N_2 (90).

However, due to concerns that environmental application of these bacteria may increase the risk of transmission to susceptible individuals, the Environmental Protection Agency (EPA) has restricted the commercial production and use of these strains (78). These regulations are particularly important in light of studies that have shown that a *B. cenocepacia* epidemic strain can be found in soil and that over 20% of clinical isolates are of the same MLST sequence type as environmental isolates (14, 165).

Clinical significance. In the past 25 years, it has become clear that Bcc species are not only phytopathogens and environmental species, but also opportunistic pathogens. In most cases, immunocompetent individuals are not susceptible to Bcc infection, although in rare cases community-acquired pneumonia, skin infections, endocarditis, and infections resulting from exposure to contaminated medical solutions have been reported for these individuals (88, 106, 238, 291). Infections often occur in patients with chronic granulomatous disease (CGD), a genetic condition caused by mutations in the genes encoding phagocytic NADPH oxidase components (268). Because this enzyme complex is crucial for oxidative burst activity, patients become susceptible to infection by a variety of different pathogens, including *Burkholderia* (268).

As opportunistic pathogens, Bcc bacteria are most closely associated with cystic fibrosis (CF) patients. Approximately 4% of the population carries the mutated gene causing this autosomal recessive disease, resulting in an incidence of 1 in 3,600 Canadian births (64). CF is caused by mutations in the cystic fibrosis transmembrane conductance regulator (CFTR), an epithelial cell chloride channel

(247). Due to this defect in chloride transport, thick mucus accumulates in the lungs that cannot be removed by the mucociliary escalator (249). Bacteria and fungi – typically (in order of prevalence) *Staphylococcus aureus* (including methicillin-resistant *S. aureus* [MRSA]), *Pseudomonas aeruginosa*, *Aspergillus fumigatus*, *Haemophilus influenzae*, *Stenotrophomonas maltophilia*, and the Bcc – can then colonize the lungs (63). With a median survival age of 46.7 years in 2009 (compared to 25.2 years in 1984), Canadian CF patients have a longer life expectancy than ever before (63). However, respiratory infections continue to be a significant cause of early mortality in this population (63).

In Canadian CF patients, the majority of infections are caused by *S. aureus* (45.7% prevalence and 3.8% MRSA prevalence) and *P. aeruginosa* (42.4% prevalence) (63). In contrast, the Bcc has a much lower national prevalence of 4.3% (with some regional variation) (63, 280). Despite this relatively low prevalence, there are three reasons why Bcc infections are a serious concern for the CF community (94): transmissibility, pathogenicity, and resistance.

Transmissibility: Person-to-person transmission of Bcc strains has thus far been observed for *B. cepacia*, *B. multivorans*, *B. cenocepacia* (most commonly), *B. dolosa*, and *B. contaminans* (23, 304). Although previous studies provided evidence that epidemic transmission may be occurring among CF patients, the first conclusive study was published by LiPuma et al. (163). Ribotyping was used to show both that transmission occurred from a Bcc-positive patient to a Bcc-negative patient at a CF summer camp and that this strain was prevalent in the CF centre attended by the Bcc-positive individual (163). This type of spread has

necessitated the implementation of strict infection control and isolation measures for CF patients (80). Revised measures recently released by CF Canada state that:

> "...persons with CF who have tested positive in the last 12 months for *B. cepacia* complex and/or Methicillin-resistant *Staphylococcus aureus* (MRSA) should not attend, in person, any Cystic Fibrosis Canada-sponsored meeting and event where persons with CF may be in attendance" and that "...persons with CF should not attend Cystic Fibrosis Canada-sponsored meetings and events held indoors since the risk of close or prolonged contact between persons with CF is increased. Participation is encouraged through alternative methods, such as teleconferencing, web conferencing, or other remote applications" (65).

Pathogenicity: Bcc species are capable of causing serious and potentially fatal disease in CF patients. Isles et al. (130) published the key study identifying '*P. cepacia*' as a CF pathogen. This study describes the experience of clinicians at The Hospital for Sick Children (now SickKids) CF Clinic in Toronto in the 1970s and early 1980s. Over this period, '*P. cepacia*' prevalence had risen to 18.1% from 9.6% (130). The infections in these patients were observed to follow one of three courses: relatively asymptomatic infection, slow decline over a period of months, and what later became known as 'cepacia syndrome,' a fatal condition characterized by the rapid development of lung abscesses and bacteremia in up to 20% of patients (130, 253). This type of decline was not reported in patients with

only *P. aeruginosa* infections and has since been observed to be caused by *B. multivorans* or *B. cenocepacia* (130, 135). There are only three published reports in which CF patients have recovered from cepacia syndrome (97). In each case, extreme treatment measures were required, including intravenous administration of at least three of the following antibiotics: trimethoprim-sulfamethoxazole, ceftazidime, tobramycin, meropenem, ciprofloxacin, chloramphenicol, and temocillin, in addition to an intravenous anti-inflammatory or nebulised antibiotic(s) (with or without DNase) (97). The cause of this syndrome is not yet known, although it is hypothesized that characteristics of both the strain and the patient may play a role (253).

Although all seventeen Bcc species have been isolated clinically (299), the majority of infections are caused by *B. multivorans* and *B. cenocepacia*. In the most recent epidemiological study from the United States (published in 2005), *B. multivorans* accounted for 38.7% of infections in CF patients, while *B. cenocepacia* accounted for 45.6% (242). Similarly, in a Canadian study from 2002, 9.6% of patients had *B. multivorans* infections while 82.6% had *B. cenocepacia* infections (280). While *B. cenocepacia* currently accounts for the greatest proportion of infections, the epidemiology of Bcc infection has been changing in recent years. Between 1997 and 2004, the proportion of new infections caused by *B. cenocepacia* decreased by ~30% while the proportion caused by *B. multivorans* increased by ~20% (242). It has been suggested that this shift is due to continued transmission of *B. multivorans* from environmental sources, while infection control protocols have led to a decrease in person-to-

person transmission of *B. cenocepacia* strains (13, 242). However, as epidemic *B. cenocepacia* strains can be found in the environment and *B. multivorans* strains can be transmitted from person-to-person, this answer may oversimplify the factors involved (23, 165).

When comparing the two Bcc species most commonly isolated from patients, *B. cenocepacia* is considered to be more virulent than *B. multivorans*. An analysis of patient records from a CF clinic in Manchester, UK showed that following infection by *B. cenocepacia*, almost all patients developed a chronic infection, compared to only half of those infected by *B. multivorans* (135). In addition, mortality of those with a chronic *B. cenocepacia* infection was significantly higher than of those infected with *P. aeruginosa*, while there was no significant difference in mortality following *B. multivorans* and *P. aeruginosa* infection (135). It is thought that differences in gene distribution among members of the Bcc may play an important role in the enhanced pathogenicity of *B. cenocepacia* (19).

Although Bcc species (particularly *B. cenocepacia*) express a wide range of putative virulence factors, there is no clear consensus as to which of these are the most important during infection (185). The characteristics that influence Bcc pathogenicity have been grouped here into two categories: surface structures and enzymes/extracellular products (antibiotic resistance is discussed below). In the following section, the Bcc is discussed in general terms: it should be noted that not all of these characteristics are shared by all seventeen species and that this list is not exhaustive (185).

Bcc bacteria encode multiple surface structures that allow them to carry out a variety of different functions. They may express type III, type IV and type VI secretion systems, all of which have been implicated in virulence (11, 257, 294). The lipopolysaccharide (LPS) not only contributes to resistance (discussed below), but also promotes an inflammatory response (168, 272). Exopolysaccharide (EPS) is thought to inhibit phagocytosis and promote persistence in the lungs (60). *B. cenocepacia* may encode cable pili and an associated adhesin that can bind to both mucin and cytokeratin 13 on host cells and facilitate attachment (258, 259). Studies have suggested that Bcc bacteria can invade and survive inside host cells, including epithelial cells and macrophages, thus avoiding the host immune response and allowing dissemination throughout the body (35, 255). *B. multivorans* and *B. cenocepacia* have been found to be more invasive than other Bcc species (139) and flagella have been shown to be an important factor in this process (295).

A variety of enzymes and extracellular products are also produced by Bcc species. Three quorum-sensing systems exist in these bacteria, *cepIR*, *cciIR* and *bviIR* (found in *B. vietnamiensis*) (158, 173, 188). The synthesis of several products, including zinc metalloproteases, lipase, and the siderophore ornibactin, is quorum sensing regulated in the Bcc (221). To protect cells from the oxidative bursts of phagocytes, these bacteria may produce a melanin-like pigment, catalase, catalase-peroxidase, and/or superoxide dismutase (156, 334). Other extracellular Bcc products include hemolysin and phospholipase C (44).

Resistance: The most serious concern regarding Bcc species is that they are innately antibiotic resistant. In the study by Isles et al. (130) (discussed above), isolates from the CF Clinic in Toronto were resistant to tetracycline (98%), ampicillin (97%), gentamicin (97%), amikacin (94%), tobramycin (94%), trimethoprim (92% [dependent on concentration]), sulfamethoxazole (90%), ticarcillin (84%), trimethoprim-sulfamethoxazole (81%), and chloramphenicol (45%). Further surveys of Bcc species also observed this type of broad resistance (219, 231, 332). Although current treatment protocols include tobramycin, trimethoprim/sulfamethoxazole, ceftazidime, ciprofloxacin, meropenem, and minocycline, resistance to these antibiotics is prevalent within the Bcc (1, 231). Some factors that contribute to this resistance are the expression of β-lactamases (125), low numbers of small porins (10, 227), altered dihydrofolate reductase (36), efflux pumps (37), hopanoids (189), LPS (168), and biofilm formation (42).

The description of a recent outbreak in Ontario CF patients serves to illustrate the interplay of transmissibility, pathogenicity, and resistance in Bcc infections. In 2008, 5 CF patients in Toronto became infected with *B. cenocepacia* ET-12, an epidemic Bcc lineage (185, 198). These patients had attended a CF centre in August 2008 and were diagnosed with Bcc infections between mid-September and early October. The *B. cenocepacia* strain infecting these 5 patients was found to be broadly antibiotic resistant (198). This strain may have been transmitted among these patients through social contact or from some other source in the clinic or environment. Of these five individuals, four died within 1-3 months of becoming infected (198).

PHAGE THERAPY

Principles. Due to the extent of Bcc antibiotic resistance, an alternative treatment protocol called phage therapy is being developed for these bacteria. This type of therapy uses bacteriophages (also referred to as phages), which are viruses that specifically infect bacteria, as the active agent to prevent or to treat infections. In this procedure, a high-titre phage preparation, ideally containing a cocktail of phages with multiple host and receptor targets, is administered to the patient (146). Following administration, these phages infect and replicate within the bacteria, lysing the cells and releasing new phages that can go on to infect other hosts and eventually clear the pathogen (293). There are several key advantages of phage therapy over antibiotic treatment (114):

1) The most important and most relevant advantage with respect to the Bcc is that phage therapy is active against antibiotic resistant pathogens. Although resistance to phages may occur (for example, by mutation of a bacterial cell surface receptor), these mutations can make the pathogen less virulent. A classic example is that of the *E. coli* K1 strain MW: phage-resistant mutants isolated from MW-infected, phage-treated mice had lost the K1 capsule, an important virulence factor (274). Furthermore, these cells remain susceptible to other phages that bind to different receptors, as would be found in a phage cocktail or that may arise via mutation (140, 146).

2) Because phages tend to infect a relatively narrow range of hosts (i.e. a limited number of genera or species), they do not compromise the normal flora.

3) In comparison with antibiotics, which take many years and upwards of half a billion dollars to develop (72) – after which resistance almost inevitably occurs – isolation and screening of phages is both rapid and economical. When phage resistance occurs, new phages can be isolated or previously characterized phages can be modified.

4) While antibiotics can cause numerous side effects, phages are an innocuous component of the normal flora and do not cause adverse reactions.

5) As mentioned above, replication can increase the phage titre upon infection, thus increasing the level of active agent within the body.

Although phage therapy enjoyed significant use in historical times, this form of treatment became relatively obsolete in western countries following the introduction of antibiotics in the 1940s (203). It is only now, with the emergence of multidrug resistant bacterial pathogens, that these methods are being reevaluated for western use (203).

Clinical trials. Results from three key trials published in 2009 have shown that phages are both safe and effective for human administration. Merabishvili et al. (202) developed a cocktail containing lytic phages specific for *P. aeruginosa* (myovirus 14/1 and podovirus PNM) and *S. aureus* (myovirus ISP). This cocktail was used on eight patients in a burn centre in Brussels, Belgium and was found to

be safe. In a Phase I trial, Rhoads et al. (244) tested a cocktail containing eight lytic podoviruses and myoviruses specific for *P. aeruginosa*, *S. aureus*, and *E. coli* in venous leg ulcers in 18 patients (with another 21 patients in a control group) in a wound centre in Lubbock, Texas. Phage administration was found to be safe in this study as well.

Wright et al. (323) completed the sole Phase I/II trial to date where phages have been rigorously tested in humans and evaluated for both safety and efficacy. Biophage-PA cocktail, containing six lytic phages active against *P. aeruginosa*, was used in 24 chronic otitis patients in the UK. Prior to treatment, it was verified that at least one of the Biophage-PA phages lysed the *P. aeruginosa* strains causing the otitis. Following a single administration of 200 µl of the cocktail (containing 6×10^5 plaque forming units [PFU]), the treatment was determined to be safe and both symptoms and bacterial counts were found to decrease (with *P. aeruginosa* completely cleared in three patients by the end of the trial) (323). Following treatment, 1.27×10^8 phages were isolated from the ears (indicating that the phages had replicated) and phage titres dropped following *P. aeruginosa* clearance. Two of the most impressive aspects of this study are that efficacy was observed after only one phage application and that these patients – whose ear conditions had persisted between 2 and 58 years – saw improvement in six weeks or less (323). Thanks to the success of this trial, AmpliPhi Biosciences (the developers of Biophage-PA) are preparing for Phase III trials with Biophage-PA and are in the preclinical phase for Biophage-PR, targeting *P. aeruginosa* infections in CF patients (8).

Genomics and phage therapy. In the clinical trials discussed in the previous section, complete genome sequencing was performed for each phage prior to human administration (111, 202, 244, 278). The first and most important reason for genomic characterization of phages for clinical use is to assess safety. It is well-established that certain temperate phages encode proteins that can increase host pathogenicity via lysogenic conversion (26). The significance of these genes is especially obvious in cases such as *Vibrio cholerae* and *Corynebacterium diphtheriae*, pathogens whose characteristic disease symptoms are caused by phage-encoded toxins (87, 312). Phages chosen for therapeutic use must be carefully screened to ensure that they do not encode genes that could increase host virulence, potentially worsening an infection rather than clearing it.

With respect to Bcc phages, identification of putative lysogenic conversion genes is especially important in light of several studies that have linked phages and phage genes to the evolution and virulence of CF pathogens. In *S. aureus* and *P. aeruginosa*, temperate or lytic phages can promote strain evolution, particularly in CF patients (27, 92). Phage-encoded toxins have also been identified in both of these pathogens (22, 214). In *P. aeruginosa* PAO1, the filamentous prophage Pf4 is required for the formation of both mature biofilms and small colony variants, as well as for full virulence in a mouse model (246). In the *P. aeruginosa* Liverpool Epidemic Strain (LES), in which six prophages and prophage remnants have been identified, mutations in three of these prophages were found to decrease host fitness as determined using *in vivo* competition experiments in a rat model (321). Most recently, a metagenomic study found that

phages present in the CF lung encode proteins linked to antibiotic resistance (320). In the virome of explant and post-mortem CF lungs, genes for aminoglycoside, β-lactam, and macrolide-lincosamide-streptogramin resistance were identified, in addition to genes encoding resistance-nodulation-division (RND) efflux pump components (320).

In addition to virulence genes, a second issue with respect to phage clinical safety is lysogeny. The rationale behind the current dogma that only strictly lytic phages should be used therapeutically is as follows (176):

1) If a temperate phage lysogenizes its host, it neither kills the cell nor releases new virions (146).

2) Superinfection immunity prevents lysis by phages with operators that can be bound by the prophage repressor protein (223).

3) Prophage gene expression can alter the host phenotype (including its pathogenicity) via lysogenic conversion (26).

4) Although generalized transduction can occur with either lytic or temperate phages, specialized transduction occurs only in temperate phages and can result in the transfer of virulence genes between hosts upon induction (96).

Again, such considerations are especially important in the CF lung as it has been shown that antibiotics used in CF patients can stimulate phage induction from clinical strains. Both ciprofloxacin and norfloxacin have been shown to significantly increase phage release from multiple *P. aeruginosa* LES isolates, which could potentially promote the spread of virulence genes among strains in

the lung (86). Similarly, CF-Marseille (an MRSA strain isolated from a French CF patient) was found to contain a prophage inducible by ciprofloxacin, erythromycin, fusidic acid, imipenem, rifampin, tobramycin, and trimethoprim-sulfamethoxazole (250).

Aside from the identification of virulence and lysogeny genes to assess phage safety (111), another reason to characterize phages via genome sequencing prior to therapeutic use is to identify targets for genetic modification that may enhance the activity of a phage *in vivo*. In previous studies, several different approaches have been taken with regards to the manipulation of phages for clinical use (210). Westwater et al. (316) engineered an *E. coli* M13 phagemid to encode two toxins, Gef and ChpBK. These toxins were expressed upon infection and host killing was observed both *in vitro* and in a mouse model (316). Hagens and Bläsi (103) also showed activity against *E. coli* with engineered M13, using a variant that encoded either a restriction enzyme or a holin. This type of modification was also effective for *P. aeruginosa*, as a filamentous *Pseudomonas* phage encoding a restriction enzyme was active both *in vitro* and in a mouse model (104). Both holin mutants (of *E. coli* phage T4) and lysin mutants (of *S. aureus* phage P954) have been shown to be active in mouse models (195, 228). For the majority of these studies, the rationale behind using mutant phages was to limit the release of endotoxin or other inflammatory components upon host cell lysis. In two studies, Lu and Collins (169, 170) used engineered phages to address the issues of biofilms and antibiotic resistance. In the first study, an *E. coli* T7 variant was modified to express dispersin B (a β-*N*-acetylglucosaminidase that

degrades a biofilm adhesin) and was shown to be active against biofilms *in vitro* (169, 239). In the second study, an M13 variant was modified to express proteins that could increase host antibiotic susceptibility (170). One of these proteins, CsrA, will be discussed in Chapter 5. These phages were found to be effective with ofloxacin, gentamicin, and ampicillin *in vitro* and with ofloxacin in a mouse model (170). Biotechnology companies are also developing engineered phages, including *S. aureus* lysin mutants from Gangagen (228) and Gram-negative host range mutants from Pherecydes Pharma (237). Phage engineering and *in vivo* testing for the Bcc will be discussed in Chapter 2.

Finally, phage genome sequencing allows for comparative genomic analyses that are informative from the standpoints of both applied and basic science. From a clinically-focused perspective, comparative genomics allows one to select phages for therapeutic use that show key genetic similarities to phages that have previously been shown to be effective *in vivo* (178). From a basic science perspective, one can survey the genetic content of both environmental and lysogen-derived phages, examine the distribution of these phages and their genes, and assess relatedness between these phages and other phage groups to analyze evolutionary relationships among these viruses (113).

Bcc BACTERIOPHAGES

Early Bcc phage research. Research into phages infecting and lysogenizing Bcc species is a relatively recent endeavor, particularly when compared to nearly a century of work related to enterobacteria phages. The goal

of early Bcc phage studies was not to find phage therapy candidates, but instead to identify generalized transducing phages that could be used for the genetic manipulation of Bcc strains (54, 196, 218). Cihlar et al. (54) were the first to publish a description of a Bcc phage in 1978: CP1, a temperate myovirus of *P. cepacia* 249 (now *B. multivorans* ATCC 17616). Bcep781 was also isolated in 1978, but a description of this phage was not published until 2006 (287). CP1 could infect *P. cepacia* 104, 382, and 383 (now *B. cepacia* ATCC 17478, *B. cepacia* ATCC 17759, and *B. lata* ATCC 17760, respectively), but not 249. Host range mutants of CP1 were able to infect *Pseudomonas pickettii* 472 (now *Ralstonia pickettii*), but not *P. aeruginosa*, *Pseudomonas fluorescens*, or *Pseudomonas phaseolicola* (now *Pseudomonas syringae* pathovar phaseolicola). Matsumoto et al. (196) were the second group to characterize a Bcc phage and the first to identify one capable of generalized transduction. CP75 is a temperate phage of *P. cepacia* PCT1, an uncharacterized clinical isolate. This myovirus has a ~52 kbp genome and a broad host range, infecting 46 out of 105 *P. cepacia* strains screened (the identity and relatedness of these strains was not discussed).

More recently, Nzula et al. (218) identified two generalized transducing phages, NS1 and NS2. These are ~48 kbp temperate phages of *B. vietnamiensis* ATCC 29424 (DBO1) and *B. multivorans* ATCC 17616, respectively. Interestingly, both CP1 and NS2 (in addition to KS5 [discussed in Chapter 3]) are temperate myoviruses of ATCC 17616, but they were found to have differing generalized transduction efficiencies and sensitivity to chloroform (54, 218). Similar to KS5, both NS1 and NS2 are predicted to use LPS as a receptor (176,

218). NS1 and NS2 can infect strains with different serotypes, which may be a result of minimal phage interaction with the O-antigen or changes in the strains that were screened (141). These phages also have a broad host range, particularly NS2: each could infect strains of *B. cepacia*, *B. multivorans*, *B. cenocepacia*, *B. vietnamiensis*, and *P. aeruginosa* (218). Certain Bcc strains were also found to be susceptible to infection by *P. aeruginosa* phages, specifically E79, B3, F116L, and G101. While phages that infect both Bcc and *Pseudomonas* strains may be extremely useful with respect to phage therapy development, it remains a concern that generalized transducing phages with such a broad host range could facilitate the exchange of virulence genes between *Burkholderia* and *Pseudomonas* (218).

One Bcc phage that has been putatively shown to transfer virulence genes between distinct species is BcP15, a ~12 kbp temperate siphovirus of the environmental isolate *B. cepacia* DR11 (119, 120). BcP15 was found to infect only *Shigella flexneri* Pl−35, a plasmid-cured isolate of *S. flexneri* NK1925 (119, 120). Putative Pl−35 lysogens were resistant to trimethoprim/sulfamethoxazole, trimethoprim, and erythromycin and it was suggested that the integrated BcP15 DNA was responsible for this phenotype (119). Further characterization of BcP15, its hosts, and the mechanisms of gene transfer and resistance are needed in order to assess the possible clinical significance of these results.

Bcc phage isolation. In order for phage therapy to be a viable alternative treatment for Bcc infections, Bcc-specific phages appropriate for clinical use must first be isolated (Figure 1-1). Selection criteria for such phages include a broad host range (particularly clinically prevalent species and/or strains), the absence of

genes encoding potential virulence factors, and an obligately lytic lifestyle. To date, one of the major challenges in Bcc phage therapy development has been finding phages that satisfy all of the above criteria. Analysis of Bcc phages with only some of these characteristics remains an important exercise because it provides essential data for studying Bcc comparative genomics, potential clues as to the development of Bcc virulence, and candidates for genetic modification to develop phages that fit these criteria more closely.

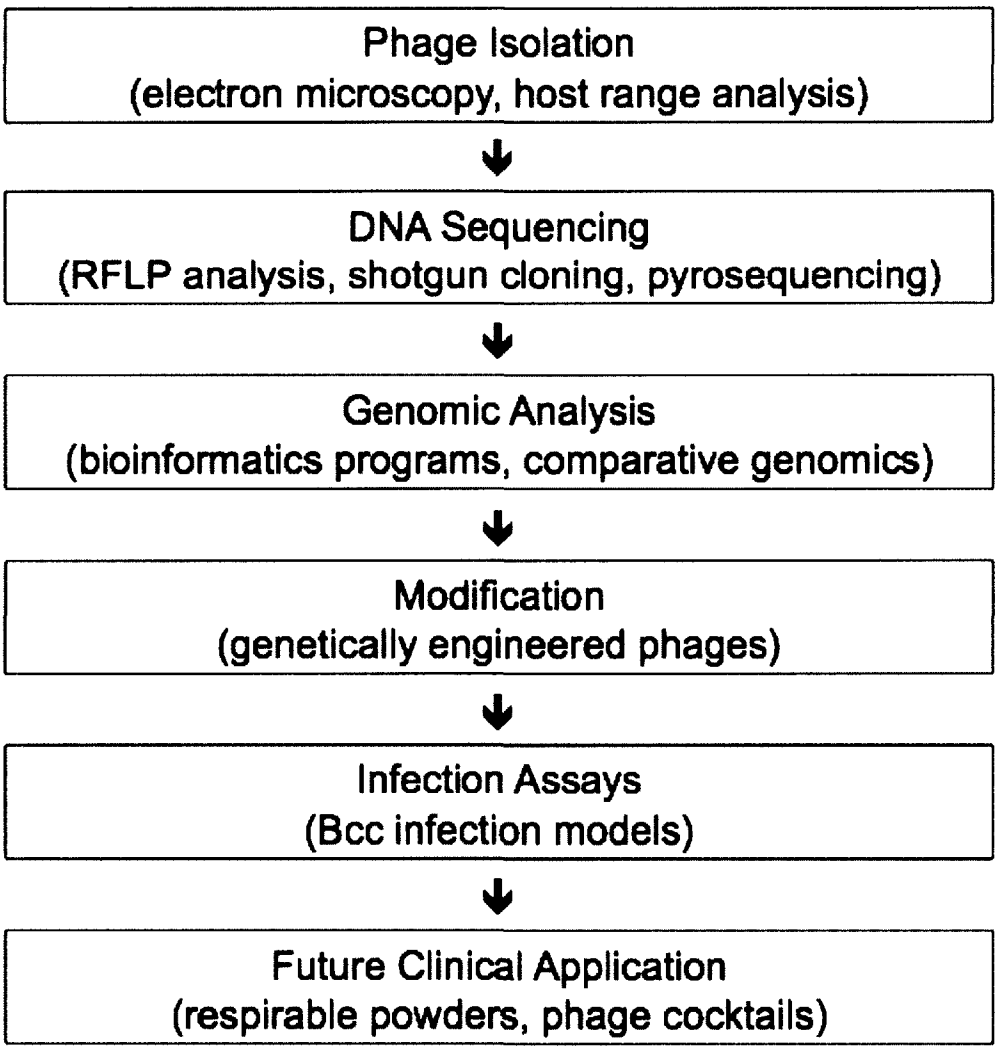


Figure 1-1: Preliminary steps in Bcc phage therapy development.

As discussed above, outside of their role as opportunistic pathogens, Bcc members are environmental organisms. As these species can be isolated from both water and soil, phage screening has focused on these sources. The following Bcc phages have been isolated from environmental samples: JB1, JB3, JB5, RL1, and RL2 (152); KS1, KS2, KS5 (discussed in Chapter 3), and KS6 (265); BcepB1A, Bcep43, Bcep1, and Bcep781 (287); BcepC6B, BcepF1, BcepGomr, and BcepNazgul (286); KS14, DC1, and KS12 (discussed in Chapters 3, 5, and 6, respectively) (267); Bcep22 and BcepIL02 (89); and KL1 and AH2 (discussed in Chapter 4) (179). Three of these phages have been shown to be effective against *B. cenocepacia* in Bcc infection models: KS14 and KS12 in *Galleria mellonella* larvae (267) and BcepIL02 in mice (43). Except for KL1, which was isolated from a sewage sample (179), all of these phages were isolated from soils, rhizospheres, and sediments, indicating that these environmental sources contain many potentially novel Bcc phages. Although these phages may be either lytic or temperate, the majority of Bcc phages characterized to date are temperate.

A second source of Bcc phages is lysogens harboring complete prophages that can be induced into the lytic cycle. The following temperate phages have been isolated from Bcc strains: CP1 (54); CP75 (196); NS1 and NS2 (218); DK1-DK4 and MM1-MM10 (152); BcepMu (288); BcP15 (120); KS4 (discussed in Chapter 6), KS8, KS9 (discussed in Chapter 2), KS10, and KS11 (265); Bcep176 (286); KL3 (discussed in Chapter 3) (177); and KL4 (discussed in Chapter 6). Several of these phages have been shown to be effective *in vivo* against *B. cenocepacia* in the *G. mellonella* model: KS4, KS4-M (a liquid clearing variant of

KS4 [discussed in Chapter 6]) (267), KS9, and KS9c (a repressor mutant of KS9 [discussed in Chapter 2]) (176). Numerous putative prophages have been identified in fully sequenced Bcc strains, some of which may eventually be isolated as temperate phages (251). It is important to note that, in certain cases, it is difficult to discern the exact origin of a newly isolated phage. Although KS5 (discussed in Chapter 3) was isolated from soil planted to onions, it was identified as a prophage in *B. multivorans* ATCC 17616 and was likely released into the soil from a lysogen (177, 265). Similarly, myovirus SR1 (discussed in Chapter 6) could have originated from either an environmental sample or a Bcc strain, as it was isolated from an extraction containing both soil and a characterized strain carrying a nearly-identical prophage (252).

Following isolation, preliminary phage characterization methods include electron microscopy and host range analysis (Figure 1-1). All Bcc phages with known virion morphology have tails and icosahedral capsids, thus belonging to the order *Caudovirales*. These phages contain double-stranded DNA and may belong to one of three families: *Myoviridae* (with long contractile tails), *Siphoviridae* (with long non-contractile tails), or *Podoviridae* (with short non-contractile tails) (5). For host range determination, high-titre phage stocks are tested for their ability to lyse strains of various Bcc species and occasionally other clinically relevant species such as *P. aeruginosa* and *B. gladioli* (265). Stronger emphasis is generally placed on the testing of *B. multivorans* and *B. cenocepacia* strains because of their clinical importance. Substantial variation is observed among Bcc phage host ranges (265). Broad host range phages are preferable for

clinical use as they are active against many strains that may be present in a patient's lungs, particularly in cases of co-colonization with Bcc, *P. aeruginosa*, and/or *B. gladioli* (130, 142). A narrow host range phage may still have utility if it is active against a clinical strain, particularly if used as part of a cocktail. Aside from the tropism itself, another characteristic that is generally noted during host range analysis is plaque morphology. Although clear plaques tend to indicate virulent phages and turbid plaques tend to indicate temperate phages, plaque morphology is not always a reliable indicator of Bcc phage lifestyle. In the case of KS5 and KS14, these phages form clear plaques but are temperate (177). Similarly, Bcep22-like phages form clear plaques but have an indeterminate lifestyle (89, 178).

Two key studies describing the isolation and preliminary characterization of novel Bcc phages were published by Langley et al. (152) and Seed and Dennis (265). Langley et al. (152) isolated phages from both Bcc strains (*B. cepacia*, *B. multivorans*, *B. cenocepacia*, and *B. stabilis*) and the environment (soil, rhizosphere, and sediment samples). Many of these phages have broad host ranges, infecting multiple Bcc species and even *P. aeruginosa* or *B. gladioli* in some cases. These phages belong to either the *Siphoviridae* or the *Myoviridae*. Similarly, Seed and Dennis (265) isolated putatively lytic phages from onion soil and temperate phages from *B. multivorans*, *B. cenocepacia*, *B. ambifaria*, and *B. pyrrocinia*. These phages had narrow to broad host ranges within the Bcc and in some cases could infect *B. gladioli* (but not *P. aeruginosa*). These phages also belong to either the *Siphoviridae* or the *Myoviridae*.

A common occurrence in Bcc phage studies (including those of Langley et al. [152] and Seed and Dennis [265]) is the identification and characterization of identical or nearly identical temperate phages by multiple groups. This issue is best illustrated by BcepMu, DK3, and KS4/KS4-M. BcepMu was isolated and fully sequenced by Summer et al. (288). This phage is a 36,748 bp transposable myovirus found as a prophage in *B. cenocepacia* J2315, BC7, C5424 and other *B. cenocepacia* ET-12 strains (excluding K56-2). BcepMu shows relatedness to transposable phages of *Salmonella typhi* (SalMu), *Photorhabdus luminescens* (PhotoMu), and *Chromobacterium violaceum* (ChromoMu) and encodes a putative O-antigen acyltransferase that may be involved in virulence (288). DK4, isolated and described by Langley et al. (152) as a temperate phage of J2315, was later identified as BcepMu based on sequencing of PCR amplicons and RFLP analysis (153). Like BcepMu, KS4 was isolated from a culture of J2315 and propagated on K56-2 (265). Repeated passage in broth culture yielded a liquid-clearing variant of this phage named KS4-M (267). As phage characteristics such as host range and apparent virion dimensions may be recorded differently among various research groups, the determination of at least partial genome sequences remains a crucial first step to ensure that newly isolated Bcc phages are novel.

Bcc phage genomics.

Overview: Following phage isolation and preliminary characterization, the next steps in phage therapy development are complete genome sequencing and bioinformatics analysis (Figure 1-1). Almost all Bcc phages sequenced to date have genome sizes between 35 and 65 kbp. KS14 and BcepF1 are outliers in this

regard, with 32.3 and 72.4 kbp genomes, respectively (177). In general, the genome of each of these phages shows relatedness to other Bcc phage sequences. Related groups of phages include the lytic myoviruses Bcep1, Bcep43, Bcep781, and to a lesser extent BcepB1A (287), the transposable myoviruses BcepMu and KS10 and their phenotypic variants KS4-M and SR1 (93, 252, 267, 288), the P2-like myoviruses KS5, KS14, and KL3 (177), the Bcep22-like podoviruses Bcep22, BcepIL02, and DC1 (89, 178), and the siphoviruses AH2 and BcepNazgul (179). Such relatedness is not observed for all Bcc phages. For example, KL1 is closely related to *Pseudomonas* phage 73 (179). In addition to size and relatedness, a third commonality among these genomes is that, despite thorough bioinformatics analysis, many of the predicted Bcc phage genes encode hypothetical proteins with no assigned function. This issue is common to the majority of phage genome sequencing projects and is not limited to the Bcc (113). Notably, with respect to phage therapy development, most of these phages are temperate but do not encode virulence genes (286).

As all Bcc phages characterized to date are members of the order *Caudovirales* and the majority of these viruses are temperate, Bcc proteins with putative functions can typically be assigned to one of four broad categories of proteins found in tailed temperate phages: lysogeny, virion morphogenesis, lysis, and DNA binding and/or modification. Note that there is some overlap among these categories and that the following is not an exhaustive listing.

1) Lysogeny: Proteins typically encoded by temperate Bcc phages include a tyrosine integrase (or serine recombinase in the case of Bcep22) and a repressor

protein that together allow for prophage integration and maintenance (89). As discussed in Chapter 2, KS9 gp41 has been examined by mutational analysis and is thus far the best characterized Bcc repressor protein (176). Some Bcc phages encode other lysogeny proteins similar to those of the λ system, such as AH2 Cro (179).

2) Virion morphogenesis: This category of proteins can be subdivided into two functions: tail morphogenesis and capsid morphogenesis/DNA packaging. The set of tail proteins encoded by a phage will differ somewhat based on family. *Myoviridae* such as the P2-like Bcc phages discussed in Chapter 3 encode tail proteins, tail tape measure proteins, frameshifted tail proteins, tail tube proteins, tail sheath proteins, tail fiber assembly proteins, tail fiber proteins, baseplate assembly proteins, and tail completion proteins (177). Bcc siphovirus tails have proteins in common with myovirus tails, but they do not contain tail sheath proteins and form less complex structures as a result (176). To date, Bcc podoviruses have poorly characterized tails, with only the genes encoding multiple sets of tail fiber proteins identified in each phage (89, 178). Similar types of capsid morphogenesis and DNA packaging proteins are found in all three phage families. Examples in the siphovirus AH2 (discussed in Chapter 4) include the major capsid protein, decorator protein, prohead protease, portal protein, head-tail joining protein, and terminase large and small subunits (179).

3) Lysis: Four to five lysis proteins are typically encoded by Bcc phages. These include a holin and antiholin (both of which contain transmembrane domains), a lysin (which contains a peptidoglycan degradation domain), an Rz-

like protein (which contains an N-terminal transmembrane domain), and an Rz1-like protein (which has a signal peptidase II cleavage site and is usually encoded within the Rz ORF) (285, 328).

4) DNA binding and/or modification: This category is extremely broad and includes proteins involved in such processes as regulation, replication, and repair. Specific examples, such as the putative DNA repair systems of KL3 and AH2, will be discussed in Chapters 3 and 4, respectively. A single Bcc phage may encode several proteins belonging to this category. For example, AH2 is predicted to encode a DNA polymerase and DNA polymerase III β subunit, exonuclease, transcriptional regulators, primase, Vsr endonuclease, excinuclease, restriction endonuclease, cytosine methyltransferase, helicase, resolvase, and single-stranded DNA binding protein (179).

Other features identified in Bcc phage genomes include tRNA genes (in the Bcep22-like phages), insertion sequences (in KS5 and Bcep22), and other genes encoding proteins that are not assigned to one of the above categories (such as MazG in KL1 and AH2) (89, 177-179, 260). In phage genomes, genes encoding proteins with a common function are often clustered together in groups known as modules (31). It has been proposed that phage genome evolution occurs at the level of the module as these sequence blocks may be readily transferred among different phages (31). Such modules have been consistently identified in Bcc phage genomes.

Methods: To isolate genomic DNA, phages are propagated to a high titre on agarose plates and the capsids are broken open using protocols such as

SDS/proteinase K or guanidine thiocyanate lysis (177, 179). Preliminary characterization generally includes restriction fragment length polymorphism (RFLP) analysis to identify closely related genomes, particularly for phages with similar virion morphologies and host ranges (179). Prior to sequencing of the complete genome, a small number of shotgun clones are sequenced to assess if the phage is novel, if it is found as a prophage in a sequenced bacterial genome, and if it shows relatedness to other phages and prophages (177). Although shotgun cloning was used to determine full Bcc phage genome sequences for many years, next-generation sequencing analysis is now the preferred method because of its efficiency and depth of coverage (116, 284).

Following collection and assembly of sequence data, the next step is to annotate and analyze the genome. A common starting point is to use a program such as GeneMark to assign open reading frames (ORFs) throughout the phage genome (172). Additional manual annotation is required because this program will not identify all genes, particularly frameshifted tail genes, Rz1-like genes, and tRNA genes. Frameshifted tail genes are located proximal to the tail tape measure gene and contain a XXXYYYZ (where Y is A or T) sequence that can be identified manually or by using FSFinder (208, 324). Rz1-like genes may be encoded out of frame within the Rz gene (as in λ) or may have a start codon within the Rz gene and a stop codon within a downstream gene (as in P2) (285). In either case, this protein will have a signal peptidase II cleavage site that can be identified using LipoP (136). Phage tRNA genes, such as those found in Bcep22-like phages, can be identified using tRNAscan-SE (89, 178, 262). Following gene

assignment, predicted proteins undergo preliminary annotation using BLASTP and Conserved Domain Search to assign putative functions and to assess relatedness to other phage and bacterial proteins (7, 191). Functional assignments can then be refined using tools such as HHpred (277). If the novel phage appears to be closely related to a previously characterized phage, CoreGenes analysis can be used to assess if these viruses belong to the same subfamily or genus (154, 155, 330).

Although these techniques are effective in most cases, not all Bcc phages have been amenable to analysis using this protocol. KS12 is a myovirus isolated from *Dietes grandiflora* (iris) soil that infects *B. multivorans* C5274 and *B. cenocepacia* K56-2 (267). Seed and Dennis (267) showed that this phage is active against K56-2 in the *G. mellonella* model and that the hemolymph of many treated larvae no longer contained K56-2. Because lysogens were absent, these results suggest that KS12 may be lytic. Unfortunately, attempts to sequence this genome have been unsuccessful to date. As discussed in Chapter 6, KS12 DNA appears to be resistant to both shearing and restriction by most enzymes, thus precluding the collection of sequence data by either shotgun cloning or pyrosequencing. As this phage may be an excellent candidate for clinical use, efforts are ongoing to collect sequence data for further characterization.

Virion sequence analysis: Although there are many Bcc phages with sequences that have been deposited into GenBank, this section will focus on those that are both in the database and either published or submitted.

1) *Siphoviridae*: To date, only three Bcc siphovirus sequences have been published or submitted: a prophage of *B. pyrrocinia* LMG 21824 named KS9 (176, 265) and two putatively temperate phages isolated from environmental samples named KL1 and AH2 (179). These phages will be discussed in Chapters 2 and 4, respectively.

2) *Myoviridae*: Summer et al. (287) published the first and thus far only genomic characterization of obligately lytic Bcc phages isolated from the environment: myoviruses BcepB1A (47,399 bp), Bcep43 (48,024 bp), Bcep1 (48,177 bp), and Bcep781 (48,247 bp). These phages have narrow host ranges, only infecting one or two strains each. The latter three phages are very closely related, with percent identities between 87.4% and 97.6% for pairwise comparisons.

Similar to BcepMu (discussed above), Goudie et al. (93) characterized a second transposable temperate myovirus of *B. cenocepacia* J2315. KS10 is 37,635 bp in length and infects *B. cenocepacia* PC184, *B. stabilis* 18870, and *B. ambifaria* LMG 19467. Like BcepMu, it has a relatively wide distribution in *B. cenocepacia*, lysogenizing K56-2, J2315, C5424, and seven clinical isolates. Although this distribution suggests that KS10 may be selected for in *B. cenocepacia*, it was not found to encode any proteins that could be considered virulence factors.

The sequences of three Bcc-specific P2-like myoviruses have been characterized: KS5 (isolated from onion soil), KS14 (isolated from *Dracaena* sp.

soil), and KL3 (isolated from *B. cenocepacia* CEP511) (177). These phages will be discussed in Chapter 3.

3) *Podoviridae*: The only Bcc podoviruses characterized to date, Bcep22 (63,882 bp), BcepIL02 (62,714 bp), and DC1 (61,847 bp; discussed in Chapter 5), are closely related and belong to a single recently identified phage genus: the Bcep22-like phages (89, 178). These phages encode multiple copies of their tail fiber proteins (3 for Bcep22 and 4 for BcepIL02 and DC1) and a large protein (4,602 – 4,667 amino acids) with soluble lytic transglycosylase, helicase, and methylase domains. Although BcepIL02 – which has been shown to be active *in vivo* – was initially thought to be lytic, it was found that these phages encode proteins for integration but cannot form stable lysogens (43, 89, 178). Both BcepIL02 and DC1 carry a putative lipid A modification gene which could modulate host virulence (89, 178).

Prophage sequence analysis: Although most of the studies discussed in the following section are only peripherally related to phage genomics, they provide a wealth of information regarding the distribution of prophages in Bcc genomes and their patterns of expression under different environmental conditions.

1) *B. multivorans*: Using *in vivo* expression technology (IVET), Nishiyama et al. (216) examined the expression of *B. multivorans* ATCC 17616 genes in soil. They found that several phage genes were induced under these conditions, including BMULJ_03661-03667 (KS5 genes *21-27*, encoding a translational regulator and tail morphogenesis proteins). In that same year,

Ronning et al. (251) identified putative prophage regions in several *Burkholderia* species, including *B. multivorans*. These included 2 regions in *B. multivorans* CGD1, 3 regions in *B. multivorans* CGD2, and 3 regions in ATCC 17616 (including those containing Bcep176 [PI 17616-3] and KS5 [PI 17616-4]).

2) *B. cenocepacia*: To identify genomic differences between the highly pathogenic *B. cenocepacia* and other less pathogenic Bcc species, Bernier and Sokol (19) used suppression-subtractive hybridization with *B. multivorans* C5393, *B. cenocepacia* K56-2, and *B. stabilis* LMG 14294. Several putative phage or phage-related genes were identified that were unique to *B. cenocepacia*. When the complete genome sequence of J2315 was published in 2009, at least 5 prophages were identified: BcenGI1 (25.0 kbp) and BcenGI7 (37.7 kbp; KS10) on chromosome 1, BcenGI12 (46.8 kbp) and BcenGI13 (46.2 kbp) on chromosome 2, and BcenGI14 (36.7 kbp; BcepMu) on chromosome 3 (126). BcenGI7 and BcenGI14 have been sequenced independently of J2315 as KS10 and BcepMu (93, 288).

Two studies profiling the transcription patterns of *B. cenocepacia* strains under different conditions have identified prophage genes as part of their analysis. O'Grady et al. (221) collected microarray data for stationary phase *B. cenocepacia* K56-2 quorum sensing mutants (*cepR*, *cciR*, and *cepRcciIR*) and found that 24, 5, and 20 phage genes, respectively, had different expression in the mutants compared to wild-type K56-2. Many of these genes corresponded to BcenGI12. When Peeters et al. (232) exposed J2315 biofilms to hydrogen peroxide for extended periods (30-60 min) and then performed microarray

analysis, they observed increased gene expression from BcenGI14 (specifically the region containing BcepMu genes *1-16*), with some genes showing an increase in expression greater than 10-fold compared to biofims not treated with hydrogen peroxide.

SUMMARY

The Bcc is a group of transmissible, virulent, and antibiotic resistant opportunistic pathogens for which novel therapeutic agents are urgently needed. Phage therapy is a promising alternative that has already been shown to be active against *B. cenocepacia in vivo* (43, 176, 267). In order to develop safe and effective therapeutic regimens for patients with these infections, Bcc phages must be sequenced and characterized to identify virulence genes, determine lifestyle, assess relatedness, and allow for genetic manipulation. The following chapters, summarized in Table 1-1, will describe the genomic characterization of ten novel candidates for Bcc phage therapy.

Table 1-1: Thesis overview

Phage	Source	Host range	Length (bp)	GC (%)	Family	Relatedness	Unique features identified
Chapter 2							
KS9	*B. pyrrocinia* LMG 21824	*B. cenocepacia* K56-2	39,896	60.7	*Siphoviridae*	*B. thailandensis* phage ϕE125, *B. pseudomallei* phage ϕ1026b, *B. cenocepacia* PC184	-GTP cyclohydrolase II integration -no effect on host virulence -KS9c (repressor knockout) has lytic phenotype -KS9 and KS9c active *in vivo*
Chapter 3							
KS5	onion soil, *B. multivorans* ATCC 17616	*B. multivorans* C5393 *B. cenocepacia* 715J *B. cenocepacia* J2315 *B. cenocepacia* K56-2 *B. cenocepacia* C6433 *B. cenocepacia* C5424	37,236	63.7	*Myoviridae*	*B. multivorans* ATCC 17616 (100%), enterobacteria phage P2	-AMP nucleosidase integration -insertion sequence -reverse transcriptase
KS14	*Dracaena* sp. soil	*B. multivorans* C5393 *B. multivorans* C5274 *B. cenocepacia* 715J *B. cenocepacia* C6433 *B. cenocepacia* C5424 *B. dolosa* LMG 21443 *B. ambifaria* LMG 17828	32,317	62.3	*Myoviridae*	enterobacteria phage P2	-plasmid prophage
KL3	*B. cenocepacia* CEP511	*B. ambifaria* LMG 17828	40,555	63.2	*Myoviridae*	enterobacteria phage P2	-threonine tRNA integration -DNA protection and repair module
Chapter 4							
KL1	sewage	*B. cenocepacia* K56-2 *B. cenocepacia* C6433	42,832	54.6	*Siphoviridae*	*Pseudomonas* phage 73	-delayed lysis phenotype -convergent evolution -RecET pair -MazG
AH2	*Nandina* sp. soil	*B. cenocepacia* K56-2 *B. cenocepacia* C6433	58,065	61.3	*Siphoviridae*	Bcc phage BcepNazgul	-delayed lysis phenotype -convergent evolution -DNA protection and repair module -MazG
Chapter 5							
DC1	*Dracaena* sp. soil	*B. cepacia* LMG 18821 *B. cenocepacia* C6433 *B. cenocepacia* PC184 *B. cenocepacia* CEP511 *B. stabilis* LMG 18870	61,847	66.2	*Podoviridae*	*B. cenocepacia* phage Bcep22, *B. cenocepacia* phage BcepIL02	-RecET pair -serine tRNA -lipid A palmitoyltransferase -carbon storage regulator CsrA -4 tail fiber genes -4650 residue soluble lytic

							transglycosylase/helicase/methylase protein
Chapter 6							
KS4-M	KS4 liquid propagation	*B. cenocepacia* K56-2	36,748	62.9	*Myoviridae*	*B. cenocepacia* phage KS4, *B. cenocepacia* phage BcepMu, *B. cenocepacia* J2315 BcenGI14	-single base substitution in baseplate assembly gene *50* (predicted missense mutation)
SR1	onion/carrot/chive soil extraction containing *B. cenocepacia* K56-2	*B. multivorans* ATCC 17616 *B. cenocepacia* PC184 *B. stabilis* LMG 18870 *B. ambifaria* LMG 19467	37,635	62.9	*Myoviridae*	Bcc phage KS10, *B. cenocepacia* J2315 BcenGI7	-single base insertion in tail fiber gene *48* (predicted frameshift mutation)
KL4	*B. multivorans* C5274	unknown	42,250	63.2	unknown	*B. multivorans* CGD1/2/2M	-LacI family transcriptional regulator integration -RecET pair
KS12	*Dietes grandiflora* soil	*B. multivorans* C5274 *B. cenocepacia* K56-2	unknown	unknown	*Myoviridae*	unknown	unknown

Note: The Appendix (discussing the development of a Bcc diagnostic test) is unrelated to phage characterization and has not been included here.

Chapter 2

Inactivation of *Burkholderia cepacia* complex phage KS9 gp41 identifies the phage repressor and generates lytic virions

A version of this chapter has been published as:

Lynch, K. H., K. D. Seed, P. Stothard, and J. J. Dennis. 2010. Inactivation of *Burkholderia cepacia* complex phage KS9 gp41 identifies the phage repressor and generates lytic virions. J. Virol. **84:**1276-1288.

OBJECTIVES

The objectives of this project were fourfold: a) to sequence and characterize the genome of KS9 (vB_BceS_KS9), a prophage of *Burkholderia pyrrocinia* LMG 21824, b) to assess the contribution of the KS9 prophage to Bcc host virulence, c) to convert KS9 into a lytic phage through specific molecular modification of its putative repressor gene, and d) to assess if this genetically modified phage is active against *B. cenocepacia in vivo*.

MATERIALS AND METHODS

Bacterial strains and growth conditions. *Burkholderia cenocepacia* K56-2 and *B. pyrrocinia* LMG 21824 are members of the Bcc experimental strain panel (183) and the updated Bcc strain panel (59), obtained from Belgium Coordinated Collection of Microorganisms LMG Bacteria Collection (Ghent, Belgium) and the Canadian *Burkholderia cepacia* Complex Research and Referral Repository (Vancouver, BC). K56-2 lipopolysaccharide (LPS) mutants were kindly provided by Miguel Valvano (University of Western Ontario, London, ON). These strains were grown aerobically overnight at 30°C on half-strength Luria-Bertani (½ LB) solid medium or broth. Transformations were performed with chemically competent DH5α (Invitrogen, Carlsbad, CA). Transformed DH5α were plated on LB solid medium containing 100 μg/ml ampicillin or 100 μg/ml trimethoprim and 15 μg/ml tetracycline and grown overnight at 37°C. Electroporations were performed for Bcc strains using a Bio-Rad MicroPulser (Bio-Rad, Hercules, CA) and plated on ½ LB solid medium containing 300 μg/ml trimethoprim and 300

μg/ml tetracycline. Strains were stored at -80°C in LB broth containing 20% glycerol.

Electron microscopy. Liquid cultures of LMG 21824 were pelleted by centrifugation at 10,000 x *g* for 2 min. The supernatant was filter sterilized using a Millex-HA 0.45 μm syringe driven filter unit (Millipore, Billerica, MA), incubated on a carbon grid for 5 min at room temperature, and stained with 2% phosphotungstic acid. A Philips/FEI (Morgagni) transmission electron microscope with charge-coupled device camera was used with the assistance of the University of Alberta Department of Biological Sciences Advanced Microscopy Facility.

KS9 propagation and DNA isolation. KS9 was isolated by Seed and Dennis (265) from a single plaque on a lawn of LMG 21824. Phage stocks were prepared in 1.5 ml of suspension medium (SM; 50 mM Tris-HCl [pH 7.5], 100 mM NaCl, 10 mM $MgSO_4$, 0.01% gelatin solution) containing KS9 plaques isolated using a sterile Pasteur pipette. After the addition of $CHCl_3$ and incubation for 1 h at room temperature, stocks were stored at 4°C. KS9 propagation and plaque assays were performed as described previously (265). A 100 μl aliquot of phage stock was incubated for 20 min at room temperature with 100 μl of *B. cenocepacia* K56-2 liquid culture. After the addition of 3 ml of soft nutrient agar, this mixture was poured onto ½ LB solid medium and grown overnight at 30°C. For efficiency of plating (EOP) determinations, this procedure was repeated in three separate trials. For DNA isolation, plates of ½ LB solid medium showing confluent lysis were prepared as described above using 100 μl of a KS9 high titre stock, 100 μl of K56-2 liquid culture, and 3 ml of soft nutrient agarose. Phages

were isolated from these plates by overlaying with SM and the DNA was extracted using the Wizard Lambda Preps DNA purification system (Promega, Madison, WI).

Shotgun library construction and sequence analysis. KS9 DNA was digested with EcoRI and SalI (Invitrogen). Restriction fragments were separated on 0.8% (wt/vol) agarose gels in 1x Tris-acetate-EDTA (pH 8.0), purified using the GENECLEAN II kit (Qbiogene, Irvine, CA), and ligated into pUC19. Blue/white selection was performed on LB solid medium containing 100 μg/ml ampicillin after transformation of the constructs into DH5α. Constructs were isolated using a QIAprep miniprep kit (Qiagen, Hilden, Germany). The presence of an insert in pUC19 was verified by restriction digest and agarose gel electrophoresis. The sequencing of plasmid inserts was performed on an ABI 3100 genetic analyzer (Applied Biosystems, Foster City, CA) by the University of Alberta Department of Biological Sciences Molecular Biology Service Unit.

Sequences were edited with EditView (Perkin-Elmer, Waltham, MA) and aligned into a single contig using AutoAssembler (Perkin-Elmer). A total of 327 sequences were used, with an assembled length of 39,896 base pairs (bp). Primers (Sigma-Genosys, Oakville, ON) designed to amplify internal regions of the plasmid inserts were used for primer walking. Gaps between sequences were filled by PCR amplification and subsequent cloning and sequencing of the amplicons using the TOPO TA kit (Invitrogen). This protocol was also used to sequence the KS9 insertion sites (LMG 21824 forward primer, 21824F [CCCACGCGCTACGGTACG]; KS9 reverse primer, KS9R

[CCGATGTAGTCCAGGCACACC]; KS9 forward primer, KS9F [CACTGGGCGCCCGTTGAG]; LMG 21824 reverse primer, 21824R [AGGTTGGTGTCGGCCGTCC]). PCR was performed with an Eppendorf MasterCycler gradient DNA thermal cycler (Eppendorf, Hamburg, Germany) using TopTaq DNA polymerase (Qiagen) and the recommended reaction composition and cycling conditions.

To identify the KS9 integration site, the sequence between gene *30* and the integrase gene *31* was analyzed using BLASTN. This region showed 71% identity with a 207 bp region (positions 1,155,342 to 1,155,547) of *Burkholderia lata* 383 chromosome 2. The *B. lata* sequence flanking the similar region was used to design primers for amplification of the junction between the LMG 21824 genome sequence and the 5' end of the KS9 prophage. The site of integration was verified by sequencing this amplicon and a second amplicon from the 3' KS9/LMG 21824 junction. An identical region of sequence in both of these amplicons was identified as the *attP* overlap region. The K56-2/5' KS9 junction in K56-2::KS9 was also identified using this procedure.

The assembled sequences were analyzed using the NCBI suite of programs. Annotation was performed using this suite, including BLASTX (http://blast.ncbi.nlm.nih.gov) (7) and ORF Finder (http://www.ncbi.nlm.nih.gov/projects/gorf). Manual annotations were verified using GeneMark.hmm-P (http://exon.biology.gatech.edu) (172). In all cases (except genes *12* [expressed by a translational frameshift] and *24'* [a gene embedded in *24*]), the manual assignments matched the GeneMark results. Gene

and gene product numbers were assigned based on the numbering system of ϕE125 and ϕ1026b: the predicted KS9 protein most similar to gp2 of these phages was assigned the name gp2, while subsequent proteins were named gp3, etc. Protein transmembrane domains were identified using OCTOPUS (http://octopus.cbr.su.se) (310). Stem-loop structures were identified using mfold (http://mfold.bioinfo.rpi.edu) (335). Signal peptidase II cleavage sites were identified using LipoP (http://www.cbs.dtu.dk/services/LipoP) (136). The sequences of KS9, ϕE125, ϕ1026b, and *B. cenocepacia* PC184 were compared using the Artemis Comparison Tool (45). Plasmid and genome maps were prepared using Savvy (http://bioinformatics.org/savvy) and GenVision (DNASTAR, Madison, WI), respectively.

Gene *41* mutagenesis and analysis. Plasmid pKL1 was created by ligating tetracycline-resistant pALTER-1 (Promega) to a) an EcoRI/KpnI PCR amplicon of KS9 bp 5054 to 6078 (upstream of KS9 gene *41* [KS9-US in Figure 2-1]; EcoRIF, AAGAATTCCAGCGCGGCATCG; KpnIR, TTGGTACCCGCCGTGTGCTTG); b) the 630 bp trimethoprim-resistant KpnI fragment of p34S-Tp2 (p34S-Tp2 in Figure 2-1) (69, 70), and c) a KpnI/BamHI PCR amplicon of KS9 bp 6481 to 7580 (downstream of KS9 gene *41* [KS9-DS in Figure 2-1]; KpnIF, AAGGTACCGTCTGCAATTCAATAGC; BamHIR, TTGGATCCTTGGTGCTTTCTCG).

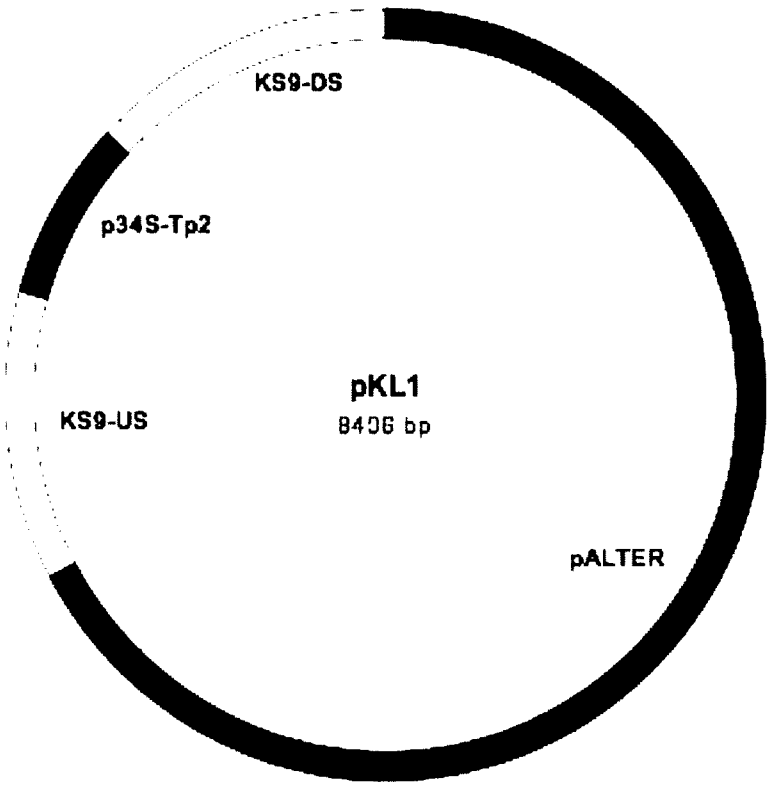


Figure 2-1: Structure of pKL1. Shown are the tetracycline-resistant pALTER-1 backbone (gray), upstream (US) and downstream (DS) sequences flanking KS9 gene *41* (yellow), and the trimethoprim-resistance cassette of p34S-Tp2 (blue).

LMG 21824 was transformed by electroporation with pKL1 and mutants with a single crossover (LMG 21824::pKL1) were selected for on LB solid medium containing 300 μg/ml trimethoprim and 300 μg/ml tetracycline. Construction of LMG 21824 (KS9 *32*) has been described previously (174). Briefly, an internal gene *32* segment was amplified and ligated to the *oriR*/trimethoprim-resistance cassette of pTn*Mod*-OTp' (70). This suicide vector was transformed into LMG 21824 by electroporation and transformants were selected on LB solid medium containing 300 μg/ml trimethoprim. To isolate phages from LMG 21824::pKL1, liquid cultures of these mutants were pelleted by centrifugation at 10,000 x *g* for 2 min and the supernatant was filter sterilized

using a 0.45 μm filter unit. Plaque assays were performed as described above and single plaques to be screened for the mutation were added to 500 μl SM. PCR screening was performed using the primers 40F (TCGTGACTGGCTGTTTTCGGAC) and 42R (GCGGCCAATTTCACGAGTCG). The reaction composition and cycling protocol for TopTaq DNA polymerase were used with the replacement of template DNA by 1 μl of phage suspension. A single representative plaque containing the gene *41* mutation was isolated and propagated as KS9c.

To isolate KS9- and KS9c-insensitive K56-2 (including K56-2::KS9), plates of K56-2 and KS9 or K56-2 and KS9c exhibiting confluent lysis were prepared as described above and incubated overnight at 30°C. Plates were overlaid with 3 ml of water and incubated on a platform rocker at 4°C for 4 h. Following recovery of the solution, cells were pelleted by centrifugation at 10,000 x *g* for 2 min. The supernatant was removed and the cells were resuspended in water and plated on ½ LB solid medium. Single colonies were isolated and colony PCR was used to screen for lysogeny using primers 21824F and KS9R. In this procedure, prior to the addition of *Taq* in the TopTaq protocol, a colony is added to the reaction mixture, incubated at 99.9°C for 5 min, and cooled on ice.

Galleria mellonella infection. We followed the procedure for *G. mellonella* infection and treatment outlined by Seed and Dennis (266, 267). Larvae were stored at 4°C and warmed to room temperature prior to infection. Infections were performed using a 250 μl syringe fitted with a reproducibility adapter (Hamilton, Reno, NV). For all experiments, 1 ml of a liquid culture of

K56-2 or K56-2::KS9 was pelleted by centrifugation at 10,000 x *g* for 2 min and resuspended in 1 ml of 10 mM $MgSO_4$ supplemented with 1.2 mg/ml ampicillin. To compare the virulence of K56-2 and K56-2::KS9, serial dilutions from 10^1 to 10^7 (bacterial concentrations of 10^6 to 10^0 cells/5 µl, respectively) were made in $MgSO_4$-ampicillin solution. For each larva, 5 µl of serially diluted bacteria was injected into the hindmost left proleg. This procedure was repeated with 10 larvae for each dilution. Ten control larvae were injected with $MgSO_4$-ampicillin solution. Larvae were incubated at 30°C and mortality was recorded 48 h post-infection. This protocol was repeated three times for each strain. Means and standard deviations were calculated using Microsoft Excel.

To assess the activity of KS9 and KS9c in *G. mellonella*, phage lysates were first passaged through Pierce Detoxi-Gel endotoxin removal gel (Thermo Scientific, Rockford, IL) and dilutions were made in $MgSO_4$-ampicillin solution. Larvae were injected with 5 µl of a $1:10^4$ dilution of K56-2 into the hindmost left proleg and 5 µl of a $1:10^0$, $1:10^1$, or $1:10^2$ dilution of phage into the second hindmost left proleg. Ten control larvae were injected with K56-2 and $MgSO_4$-ampicillin solution and another ten control larvae were injected with undiluted phage and $MgSO_4$-ampicillin solution. Larvae were incubated at 30°C and survival was recorded 48 h post-infection. This protocol was repeated three times for each multiplicity of infection (MOI) tested. To isolate K56-2 from KS9-treated larvae, hemolymph was extracted from larvae surviving 48 h post-infection using a 20 gauge needle. Serial dilutions were made in $MgSO_4$-ampicillin solution and cells were plated on ½ LB solid medium containing 100

μg/ml ampicillin.

RESULTS AND DISCUSSION

Plaque and virion morphology. KS9 was originally isolated from a culture of *B. pyrrocinia* LMG 21824 (265). Induction with UV light or mitomycin C was not necessary. When propagated on *B. cenocepacia* K56-2, KS9 forms small clear plaques, 0.3 to 1.0 mm in diameter. Preliminary characterization of short KS9 sequences by Seed and Dennis (265) indicated that its genome shows similarity to the genomes of two non-Bcc *Burkholderia* phages: ϕE125, a prophage of *Burkholderia thailandensis* E125 (322) and ϕ1026b, a prophage of *Burkholderia pseudomallei* 1026b (71). However, KS9 is unable to form plaques on *B. mallei* (10 strains tested), *B. pseudomallei* (10 strains tested), or *B. thailandensis* (4 strains tested).

Transmission electron microscopy indicates that KS9 has the B1 morphotype of the family *Siphoviridae* in the order *Caudovirales* (Figure 2-2) (4). The virion is comprised of an icosahedral head (with a diameter of 75 nm) and a long, noncontractile tail (with a length of 250 nm). The phage particles were highly fragile, as a large number of intact capsids were visible in which the tail had broken off close to or at the head/tail adaptor. The KS9 virion is larger than that of both ϕE125 (63 nm head diameter, 203 nm tail length) and ϕ1026b (56 nm head diameter, 200 nm tail length) (71, 322).

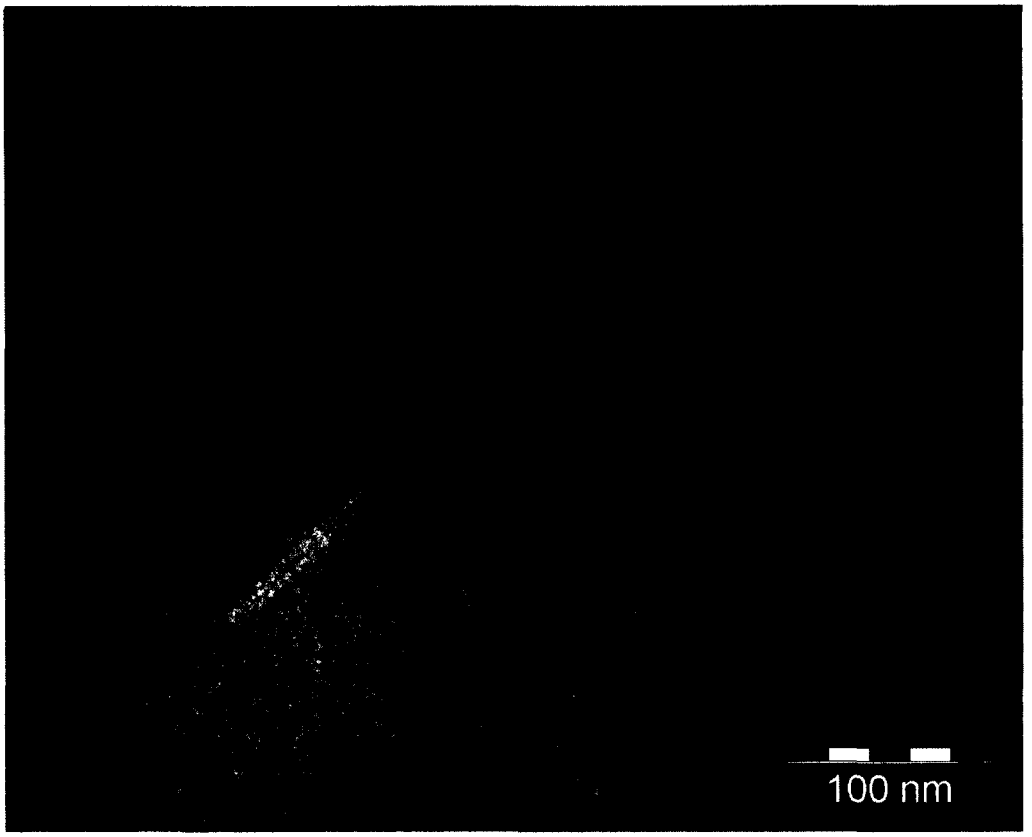


Figure 2-2: Transmission electron micrograph of KS9. The sample was stained with 2% phosphotungstic acid and viewed at 180,000-fold magnification with a Philips/FEI (Morgagni) transmission electron microscope. Reproduced with permission from the American Society for Microbiology.

Receptor binding. To determine if the LPS is a KS9 receptor, K56-2 LPS mutant strains were tested in a plaque assay with high-titre stocks of KS9 and KS5 (a second Bcc-specific phage isolated by Seed and Dennis [265], discussed in Chapter 3). Although KS5 produced confluent lysis on strains XOA7 (*waaL*::pGPΩTp, EOP = 0.8), XOA15 (*wabR*::pGPΩTp, EOP = 1.3), XOA17 (*wabS*::pGPApTp, EOP = 1.1), RSF19 (*wbxE*::pRF201, EOP = 0.5), and K56-2 (EOP = 1), KS9 was unable to form plaques on any of these strains excluding the wild-type (168, 224). Both phages are likely to use the LPS as a receptor because neither phage was able to lyse the LPS mutants XOA8 (*wabO*::pGPΩTp) or

CCB1 (*waaC*::pGPΩTp) (224). These results suggest that KS9 requires a relatively complete LPS structure in order to infect K56-2, likely binding to an LPS component located distal to lipid A (i.e. O-antigen), whereas KS5 likely binds to a receptor located deeper within the LPS, proximal to lipid A (i.e. inner core). These results should be interpreted cautiously because, since the mutations have caused significant deficits in LPS structure, the overall organization of the outer membrane may have been altered as well (168). Although these results are consistent with both KS9 and KS5 using LPS components as receptors, further experiments are required in order to exclude the possibility of adhesion to other outer membrane structures.

KS9 genome. The KS9 genome is 39,896 bp in length and encodes 50 proteins (Table 2-1). The GC content of the genome is 60.7%, which is identical to that of ϕ1026b and slightly lower than that of ϕE125 (61.2%) (71, 322). The majority of the start codons are ATG (42 of 50), with fewer GTG (8) present (Table 2-1). Each of the predicted KS9 proteins has some degree of similarity to other proteins as determined by a BLASTX search, except for gp36, gp38, and gp39 (Table 2-1). The protein with the lowest detectable similarity to others is gp46, putatively involved in replication, which has 23% identity with a phage O protein of *Salmonella enterica* subsp. *enterica* serovar Javiana strain GA_MM04042433 (Table 2-1). The predicted proteins most similar to database entries are gp13, the tail tape measure, and gp18, a tail component protein. Both of these proteins have 97% identity with proteins encoded by *B. cenocepacia* PC184 (Table 2-1). The complete KS9 genome sequence was deposited in

GenBank with the accession number NC_013055.1.

Table 2-1: KS9 genome annotation

Gene	Coding region	Putative function	Strand	Predicted RBS and start codon	Length (aa)	Closest relative	Alignment region in closest relative (aa)	% ID	Source	GenBank accession number
31	241-1515	integrase	+	TAGTtttgcATG	424	phage integrase	1-424/424	91	*Burkholderia cenocepacia* MC0-3	YP_001779169.1
32	1658-2962	unknown	+	ATTTGGGcatctcatGTG	434	hypothetical protein PROVALCAL_01964	3-396/396	28	*Providencia alcalifaciens* DSM 30120	ZP_03319024.1
33	2955-3710	unknown	+	GGAGGccttggccaATG	251	hypothetical protein CPS_0494	1-203/247	33	*Colwellia psychrerythraea* 34H	YP_267252.1
34	3823-4098	transcriptional regulator	-	GGGAgaatcaATG	91	Prophage CP4-57 regulatory protein (AlpA) family	1-76/80	59	*Burkholderia thailandensis* E264	YP_442445.1
35	4123-4947	chromosome partitioning	-	AACAAAAtataaccATG	274	hypothetical protein BB1674	1-256/274	65	*Bordetella bronchiseptica* RB50	NP_888219.1
36	4999-5064	unknown	-	GAAGGAATgatcgATG	21	N/A				
37	5068-5274	unknown	-	GTGAGcacgccATG	68	hypothetical protein Bmul_4870	10-67/67	53	*Burkholderia multivorans* ATCC 17616	YP_001584833.1
38	5282-5407	unknown	-	ATTTGAGGtgcacATG	41	N/A				
39	5415-5555	unknown	-	AGACGAGGccctgATG	46	N/A				
40	5824-6111	unknown	+	GAGGGTggggctgttgATG	95	hypothetical protein BthaB_33341	1-77/77	71	*Burkholderia thailandensis* Bt4	ZP_02389868.1

41	6079-6480	repressor	-	TGAAttgcagacATG	133	gp52	1-133/133	75	*Burkholderia thailandensis* Bt4	ZP_02389867.1
42	6566-6805	unknown	+	AAAGttgcggcATG	79	hypothetical protein BthaT_11693	1-77/80	71	*Burkholderia thailandensis* TXDOH	ZP_02371673.1
43	6986-7186	unknown	-	GGAGccggatttATG	66	hypothetical protein OsJ_023465	384-438/613	40	*Oryza sativa* (japonica cultivar-group)	EAZ39982.1
44	7276-7782	unknown	+	AAGGGtaaaaATG	168	hypothetical protein Bpse9_06757	5-175/176	46	*Burkholderia pseudomallei* 91	ZP_02446505.1
45	7810-9177	helicase	+	GAAGAATctggggaaacaATG	455	replicative DNA helicase	1-455/455	60	*Burkholderia cenocepacia* PC184	YP_002091324.1
46	9223-10263	replication protein	+	TGGCGccgcgcaATG	346	phage O protein	16-289/319	23	*Salmonella enterica* subsp. *enterica* serovar Javiana str. GA_MM040424 33	EDN74656
47	10274-10672	unknown	+	GAAGAATGAcggggaaattATG	132	gp66, conserved hypothetical protein	5-133/139	41	*Burkholderia* phage ϕ644-2	YP_001111145.1
48	10844-11419	unknown	+	GGGGAGttcgttaATG	191	hypothetical protein syc0482_c	48-180/182	27	*Synechococcus elongatus* PCC 6301	YP_171192.1
49	11449-11817	homing endonuclease	+	GGAGAgcgtaATG	122	gp82 phage protein	38-148/159	92	*Burkholderia cenocepacia* PC184	YP_002091325.1
1	12013-12516	unknown	+	GGAAGcctcacATG	167	hypothetical protein BCPG_00011	1-167/167	91	*Burkholderia cenocepacia* PC184	YP_002091326.1
2	12561-14237	terminase	+	GGGTATGGcctgcccGTG	558	phage terminase-like protein	62-615/615	89	*Burkholderia cenocepacia* PC184	YP_002091327.1

3	14234-15523	portal protein	+	GGAAATctacatcctATG	429	phage-related protein	1-429/429	85	*Burkholderia cenocepacia* PC184	YP_002091328.1
4	15552-16307	protease/ scaffold protein	+	GTGccttcagcGTG	251	protease subunit of ATP-dependent Clp proteases	1-251/251	93	*Burkholderia cenocepacia* PC184	YP_002091329.1
5	16376-17641	major capsid protein	+	GATGGAGactgtATG	421	phage ϕ-c31 gp36-like protein	1-418/418	58	*Ralstonia solanacearum* MolK2	YP_002254489.1
6	17692-18276	unknown	+	GTGAGGaataaATG	194	hypothetical protein RSMK01647	4-176/184	40	*Ralstonia solanacearum* MolK2	YP_002254490.1
7	18287-18613	head-tail adaptor	+	GAGGaccgtATG	108	phage head-tail adaptor, putative	1-108/108	84	*Burkholderia oklahomensis* C6786	ZP_02367135.1
8	18606-19028	unknown	+	GAAGGTGggggagaagtATG	140	gp9	1-140/140	85	*Burkholderia* phage ϕ1026b	NP_945039.1
9	19025-19369	unknown	+	GGGGGtggccgGTG	114	hypothetical protein BCPG_00017	1-114/114	86	*Burkholderia cenocepacia* PC184	YP_002091332.1
10	19429-19893	major tail subunit	+	ATGAGGggcttATG	154	gp70	1-154/154	83	*Burkholderia* phage Bcep176	YP_355406.1
11	19922-20389	tail assembly chaperone	+	GAAAGGAAAgagtgATG	155	gp69'	8-133, 142-155/246	76, 78	*Burkholderia* phage Bcep176	YP_355404.1
12	19922-20664	minor tail protein	+	GAAAGGAAAgagtgATG	247	gp69'	8-246/246	78	*Burkholderia* phage Bcep176	YP_355404.1
13	20678-24778	tail tape measure	+	AGGTGGAtagaaaatATG	1366	phage-related minor tail protein	1-1366/1366	97	*Burkholderia cenocepacia* PC184	YP_002091337.1
14	24778-25116	minor tail protein	+	GGGATGgggtgATG	112	gp67	1-112/112	88	*Burkholderia* phage Bcep176	YP_355402.1

15	25126-26934	unknown	+	ATAGGGttagaaaATG	602	gp66	1-193/335	52	*Burkholderia* phage Bcep176	YP_355401.1
16	26939-27622	minor tail protein	+	GAGTTAgccatATG	227	gp65 phage protein	1-227/227	92	*Burkholderia cenocepacia* PC184	YP_002091339.1
17	27672-28424	tail component protein	+	GGGTTttttATG	250	hypothetical protein BCPG_00025	1-250/250	95	*Burkholderia cenocepacia* PC184	YP_002091340.1
18	28421-28984	tail component protein	+	GGGGGTgcgaaGTG	187	gp63 phage protein	1-187/187	97	*Burkholderia cenocepacia* PC184	YP_002091341.1
19	28981-32889	tail tip fiber protein	+	AGAGGAtcaagtGTG	1302	phage-related protein, tail component	1-1300/1301	92	*Burkholderia cenocepacia* PC184	YP_002091342.1
20	32886-33194	unknown	+	GGGGGAGGtgggcGTG	102	hypothetical protein BTH_II1063	3-104/104	95	*Burkholderia thailandensis* E264	YP_439260.1
21	33194-33910	unknown	+	GGAGTGtattaATG	238	hypothetical protein BTH_II1064	9-246/246	93	*Burkholderia thailandensis* E264	YP_439261.1
22	33910-34248	holin	+	GGGAATTgtgtgATG	112	holin	36-147/147	94	*Burkholderia thailandensis* E264	YP_439262.1
23	34251-34700	endolysin	+	GGAGAAAAAtaaacATG	149	hypothetical protein BTH_II1066	1-149/149	92	*Burkholderia thailandensis* E264	YP_439263.1
24	34697-35179	Rz	+	CGGAGgggctATG	160	gp23 phage protein	1-160/160	81	*Burkholderia cenocepacia* PC184	YP_002091345.1
24'	34896-35159	Rz1	+	AGGGGGAAcgctgaagATG	87	hypothetical protein Bamb_1851	6-92/180	65	*Burkholderia ambifaria* AMMD	YP_773741.1

25	35318-36109	DNA adenine methylase	+	AAATACGGctttcacaATG	263	DNA adenine methylase	1-262/262	87	*Burkholderia pseudomallei* 1106b	ZP_02107289.1
26	36211-37107	unknown	+	TGAGAacacaATG	298	hypothetical protein BTH_II1069	1-298/298	93	*Burkholderia thailandensis* E264	YP_439266.1
27	37131-37622	unknown	-	GGGAtacactGTG	163	hypothetical protein BCPG_00032	1-162/163	75	*Burkholderia cenocepacia* PC184	YP_002091347.1
28	37677-38312	unknown	+	AGGGGAGcggtctATG	211	hypothetical protein BthaT_29201	1-210/214	89	*Burkholderia thailandensis* TXDOH	ZP_02375127.1
29	38440-39201	unknown	-	AGAGGCTcattcccATG	253	hypothetical protein SCP1.182	118-238/251	49	*Streptomyces coelicolor* A3(2)	NP_639763.1
30	39194-39565	unknown	-	TGCGGAAcgaacATG	123	ribonuclease G/E	296-393/611	28	*Prochlorococcus marinus* subsp. *marinus* str. CCMP1375	NP_876048.1

Abbreviations: RBS, ribosome binding site; aa, amino acid residues; ID, identity; N/A, not applicable. Putative functions were assigned based on BLASTX results. Reproduced with permission from the American Society for Microbiology.

The KS9 genome exhibits a modular organization (Figure 2-3). Module boundaries were assigned based on BLASTX predictions of gene function. KS9 genes encoding hypothetical proteins were not grouped as part of a module unless the flanking genes encoded two proteins with similar functions (as is the case for genes *6* and *15*). The smallest multigene module, which encodes proteins involved in replication, is made up of genes *45* and *46*. The DNA packaging/head morphogenesis module is made up of genes *2* to *7*. Although the product of gene *6* has not been assigned a putative function, it is included as part of this module because of its location. Similarly, genes *10* to *19* comprise the tail morphogenesis module. Gene *15*, encoding a hypothetical protein, has been included here because of its position. The fourth module in the genome, involved in lysis, includes genes *22* to *24'*, which encode the putative holin, endolysin, Rz and Rz1 proteins.

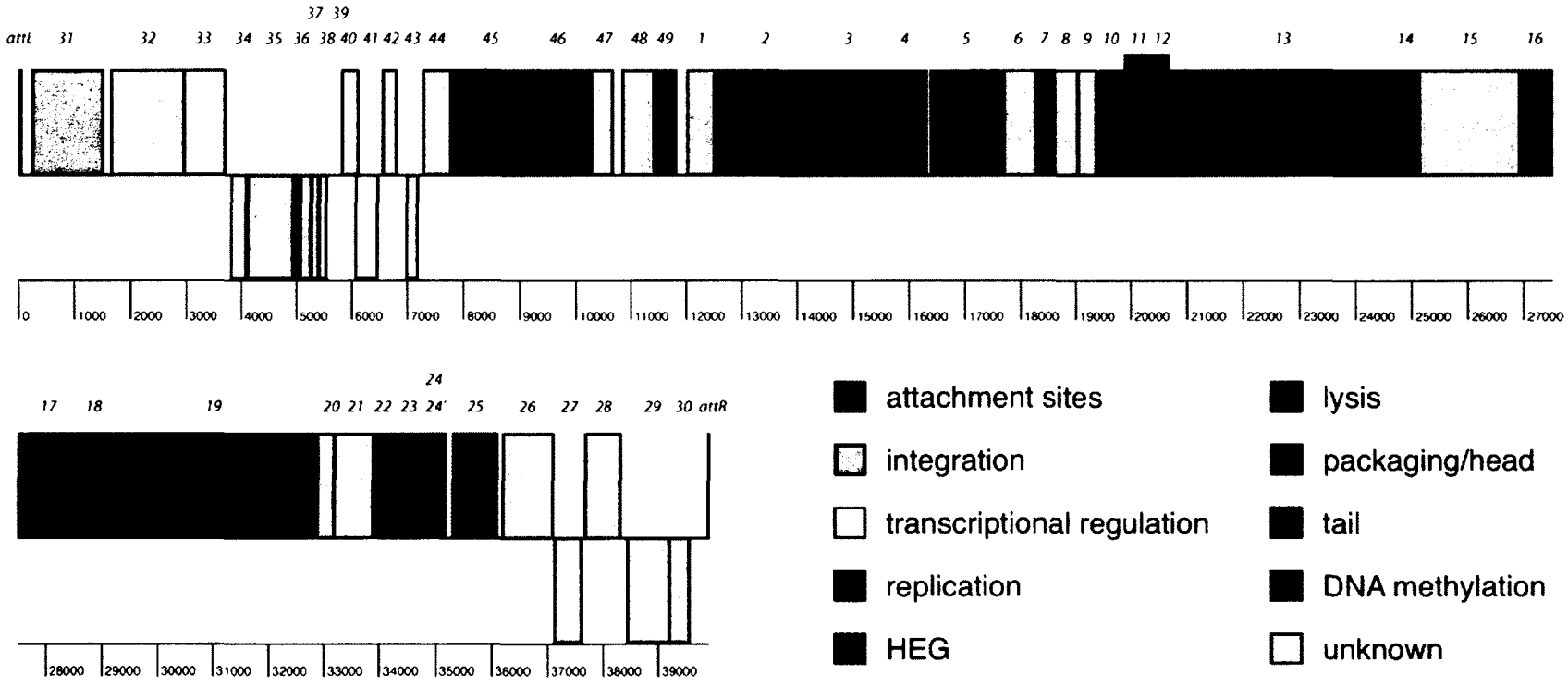


Figure 2-3: Map of the KS9 prophage. Genes transcribed in the forward direction are displayed above those transcribed in the reverse direction. Gene names are listed above and the scale (in base pairs) is shown below. The vertical extension of gene *12* indicates the ORF following a translational frameshift. Gene *24'* is shown embedded in gene *24*. HEG, homing endonuclease gene. Reproduced with permission from the American Society for Microbiology.

Similarity to *B. cenocepacia* PC184. The predicted gene products of KS9 are most similar to those of a *B. cenocepacia* PC184 prophage element (vB_BceZ_PC184; Table 2-2 and Figure 2-4). KS9 gp1-4, gp7-19, gp22-24', gp26-28, gp45, and gp49 are similar to proteins encoded in a single locus in PC184, spanning from BCPG_00009 to BCPG_00033 (Table 2-2). An exception is KS9 gene *15*, which encodes a protein similar to PC184 BCPG_01060 (Table 2-2). For instances in which KS9 proteins are most closely related to PC184 proteins, percent identities range from 60% (gp45) to 97% (gp13 and gp18) (Table 2-1). Such proteins are involved in DNA packaging/head assembly, tail assembly, lysis, replication, and homing endonuclease activity.

Casjens (46) outlines several standards by which one can predict if a bacterial genome sequence contains a prophage. These include: a) the presence of phage-related genes (especially those required for morphogenesis), b) continuous organization undisturbed by non-phage related genes, c) characteristic prophage gene order, and d) the presence of genes encoding hypothetical proteins (especially if phage related). Based on these criteria, this locus in PC184 is predicted to contain an uncharacterized prophage or prophage element: it contains multiple phage genes (including morphogenesis genes), the organization is continuous, genes are present in the same order as those in the KS9 prophage, and genes encoding proteins similar to phage-related hypothetical proteins (such as KS9 gp1) are present.

Table 2-2: Proteins similar to KS9 proteins in *Burkholderia cenocepacia* PC184 and phages ϕ1026b and ϕE125

KS9 protein	*B. cenocepacia* PC184 protein (GenBank accession number, percent identity)	ϕ1026b protein (GenBank accession number, percent identity)	ϕE125 protein (GenBank accession number, percent identity)
gp31	none	none	none
gp32	none	none	none
gp33	none	none	none
gp34	none	none	none
gp35	none	gp68 (NP_945099.1, 46%)	gp58 (NP_536415.1, 47%)
gp36	none	none	none
gp37	none	none	none
gp38	none	none	none
gp39	none	none	none
gp40	none	none	none
gp41	none	none	gp52 (NP_536409.1, 55%)
gp42	none	none	none
gp43	none	none	none
gp44	none	none	none
gp45	BCPG_00009 (YP_002091324.1, 60%)	none	none
gp46	none	none	none
gp47	none	none	none
gp48	none	none	none
gp49	BCPG_00010 (YP_002091325.1, 92%)	gp82 (NP_945113.1, 56%)	gp70 (NP_536427.1, 56%)

gp1	BCPG_00011 (YP_002091326.1, 91%), BCPG_00012 (YP_002091327.1, 69%)	none	none
gp2	BCPG_00012 (YP_002091327.1, 89%)	gp2 (NP_945032.1, 35%)	gp2 (NP_536358.1, 34%)
gp3	BCPG_00013 (YP_002091328.1, 85%)	gp3 (NP_945033.1, 29%)	none
gp4	BCPG_00014 (YP_002091329.1, 93%)	none	none
gp5	none	none	none
gp6	none	none	gp8 (NP_536364.1, 25%)
gp7	BCPG_00015 (YP_002091330.1, 95%)	gp8 (NP_945038.1, 66%)	gp9 (NP_536365.1, 81%)
gp8	BCPG_00016 (YP_002091331.1, 93%), BCPG_05088 (YP_002096214.1, 28%)	gp9 (NP_945039.1, 85%)	gp10 (NP_536366.1, 84%)
gp9	BCPG_00017 (YP_002091332.1, 86%)	gp10 (NP_945040.1, 59%)	gp11 (NP_536367.1, 59%)
gp10	BCPG_00018 (YP_002091333.1, 92%)	gp11 (NP_945041.1, 75%)	gp12 (NP_536368.1, 75%)
gp11	BCPG_00019 (YP_002091334.1, 58%), BCPG_00020 (YP_002091335.1, 91%)	gp12 (NP_945042.1, 56%)	gp13 (NP_536369.1, 56%)
gp12	BCPG_00019 (YP_002091334.1, 58%), BCPG_00020 (YP_002091335.1, 85%), BCPG_00021 (YP_002091336.1, 92%)	gp12 (NP_945042.1, 56%), gp13 (NP_945043.1, 63%)	gp13 (NP_536369.1, 55%), gp14 (NP_536370.1, 64%)

gp13	BCPG_00022 (YP_002091337.1, 97%)	gp14 (NP_945044.1, 73%)	gp15 (NP_536371.1, 73%)
gp14	BCPG_00023 (YP_002091338.1, 94%)	gp15 (NP_945045.1, 73%)	gp16 (NP_536372.1, 73%)
gp15	BCPG_01060 (YP_002092331.1, 26%)	gp16 (NP_945046.1, 41%)	gp17 (NP_536373.1, 41%)
gp16	BCPG_00024 (YP_002091339.1, 92%)	gp17 (NP_945047.1, 81%)	gp18 (NP_536374.1, 79%)
gp17	BCPG_00025 (YP_002091340.1, 95%)	gp18 (NP_945048.1, 74%)	gp19 (NP_536375.1, 75%)
gp18	BCPG_00026 (YP_002091341.1, 97%)	gp19 (NP_945049.1, 77%)	gp20 (NP_536376.1, 76%)
gp19	BCPG_00027 (YP_002091342.1, 92%)	gp20 (NP_945050.1, 73%)	gp21 (NP_536377.1, 73%)
gp20	none	gp21 (NP_945051.1, 47%)	gp22 (NP_536378.1, 47%)
gp21	none	gp22 (NP_945052.1, 47%)	gp23 (NP_536379.1, 47%)
gp22	BCPG_00028 (YP_002091343.1, 95%)	none	none
gp23	BCPG_00029 (YP_002091344.1, 97%)	none	none
gp24	BCPG_00030 (YP_002091345.1, 81%)	none	none
gp24'	BCPG_00030 (YP_002091345.1, 81%)	none	none
gp25	none	gp26 (NP_945057.1, 86%)	gp27 (NP_536383.1, 86%)
gp26	BCPG_00031 (YP_002091346.1, 89%)	gp27 (NP_945058.1, 52%)	gp28 (NP_536384.1, 52%)

gp27	BCPG_00032 (YP_002091347.1, 75%)	none	none
gp28	BCPG_00033 (YP_002091348.1, 59%)	gp32 (NP_945063.1, 55%)	gp33 (NP_536389.1, 55%)
gp29	none	none	none
gp30	none	none	none

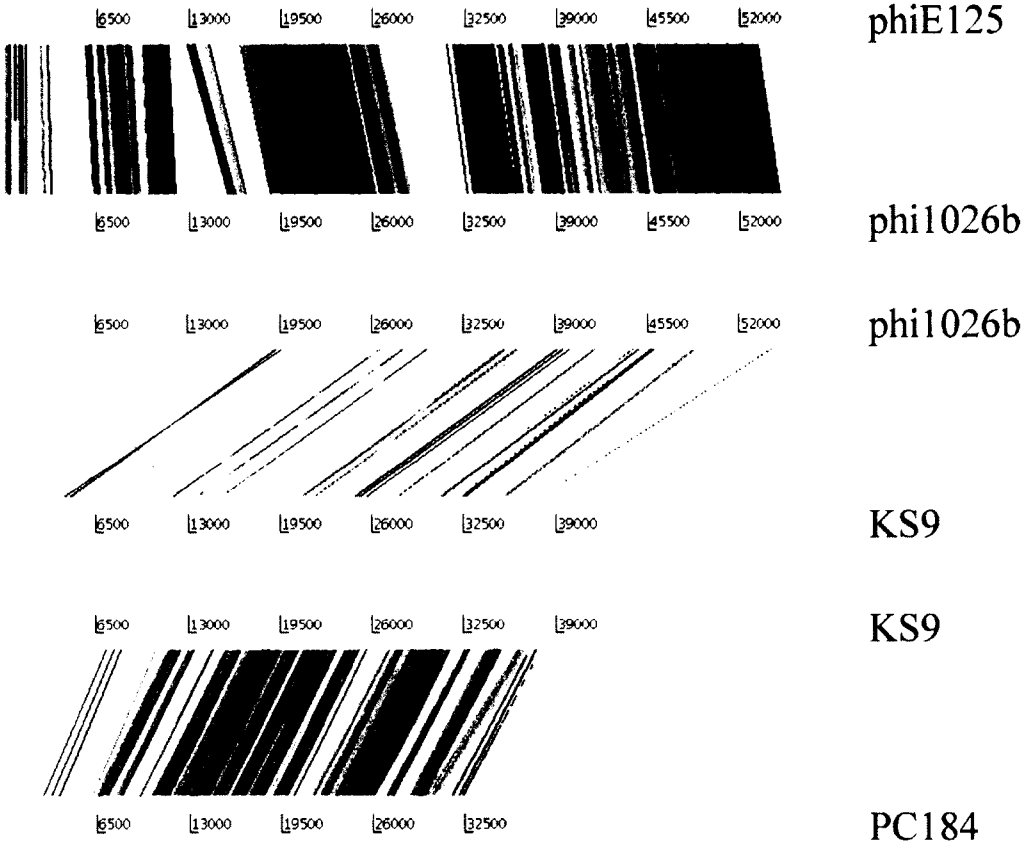


Figure 2-4: Artemis Comparison Tool analysis of ϕE125, ϕ1026b, KS9, and a similar locus of *Burkholderia cenocepacia* PC184. From top to bottom: comparison of ϕE125 (above) and ϕ1026b (below); comparison of ϕ1026b (above) and KS9 (below); comparison of KS9 (above) and PC184 (below). Reproduced with permission from the American Society for Microbiology.

KS9 integration site. Because of the similarities between KS9, ϕE125, and ϕ1026b, it was predicted that the overlap region of the KS9 *attP* site would be located at a similar position in the phage genome (upstream of the integrase gene) and that it would facilitate integration into a similar locus (the 3' end of a $tRNA^{Pro-3}$ coding region) (71, 322). Although the *attP* position was found to be similar, in KS9 it allows for recombination into a different host gene. Using a 20 bp *attP* overlap region (Figure 2-5), KS9 integrates into the 3' end of a gene encoding the LMG 21824 or K56-2 GTP cyclohydrolase II (GCHII). GCHII is responsible for

the production of 2,5-diamino-6-β-ribosyl-4(3*H*)-pyrimidinone 5'-phosphate from GTP. This product is the first intermediate in the synthesis of riboflavin, which is necessary for metabolism (243). The GCHII ORF lies immediately upstream of an ORF encoding a GCN5-related *N*-acetyltransferase (GNAT). Since phage integration does not disrupt the sequence of the predicted GCHII ORF (with the *attP* overlap region replacing the last 19 bp of the gene, including the stop codon), there is likely no loss of GCHII function in a KS9 lysogen.

attL	TCGGGCATTA**CTTCGAGGAAAACGAGTAAC**GTAATCAAGG
virion	AATGTTGTTT**CTTCGAGGAAAACGAGTAAC**GTAATCAAGG
attR	AATGTTGTTT**CTTCGAGGAAAACGAGTAAC**CGCCACGAGC

Figure 2-5: Sequence of the KS9 *attP* overlap region in *Burkholderia pyrrocinia* LMG 21824. The 20 bp overlapping sequence present in *attL* and *attR* of the KS9 prophage and in the chromosome of the vegetative phage (virion) is underlined. Reproduced with permission from the American Society for Microbiology.

To the best of our knowledge, there are no published reports of prophage integration into a GCHII gene. However, we have identified a sequence in GenBank that strongly suggests that prophage insertion can occur at this site in other Bcc genomes. In chromosome 2 of *B. cenocepacia* AU 1054, the gene for Bcen_3636 (GCHII) is located 5' to the gene for Bcen_3637 (GNAT), separated by 125 bp. In contrast, the distance separating these genes in *B. lata* 383 (for Bcep18194_B1027 GCHII and Bcep18194_B1047 GNAT) is 18,957 bp. Between these two genes in *B. lata* 383, the annotated sequence contains an additional 19 genes encoding a phage integrase (Bcep18194_B1028), a terminase (Bcep18194_B1042), a lytic transglycosylase (Bcep18194_B1044), and 16 other proteins without assigned functions.

In order to characterize this region further, we analyzed the sequence beginning with the GCHII start codon and ending with the GNAT stop codon using GeneMark (30) and BLASTX. Thirty-six proteins were identified, including the GCHII (383-1) and GNAT (383-36). According to BLASTX analysis, 16 of these are similar to phage-related proteins (Table 2-3). Interestingly, four of these proteins (383-3, 383-17, 383-19, and 383-32) are similar to ϕE125 and ϕ1026b proteins, respectively: gp33/gp32 (383-32), gp36 of ϕ1026b (383-3), gp60/gp71 (383-17), and gp65/gp76 (383-19) (Table 2-3). Only one of these sets, gp33/gp32, is similar to a KS9 protein, gp28. These data suggest that an uncharacterized prophage has integrated into *B. lata* 383 at a similar site as KS9 and that these two phages are at least partially related. Because of its small size, it is not known if this prophage is defective or if it has retained all of the elements necessary for a productive lytic cycle. Using Casjens' (46) parameters for prophage identification, the region that we have identified in *B. lata* 383 meets these requirements: a) genes encoding phage proteins such as integrase, terminase, and virion morphogenesis factors are present; b) there are no obvious nonphage genes in the almost 19 kbp sequence; c) the gene order is consistent with the ϕ1026b prophage; and d) genes encoding proteins similar to phage-related hypothetical proteins are present.

Table 2-3: Phage-associated proteins similar to GeneMark-predicted *Burkholderia lata* 383 proteins between Bcep18194_B1027 and Bcep18194_B1047

Predicted 383 protein	Similar protein (source)	% identity	GenBank accession number
383-2	phage integrase (*Burkholderia lata* 383)	100	YP_371786.1
	integrase family protein (*Burkholderia cenocepacia* MC0-3)	93	YP_001779169.1
	phage integrase (*Burkholderia dolosa* AU0158)	92	YP_002100279.1
383-3	gp36-related protein (*Burkholderia thailandensis* TXDOH)	57	ZP_02371687.1
	gp36-related protein (*Burkholderia pseudomallei* 1710b)	55	YP_333059.1
	gp36 (ϕ1026b)	55	NP_945067.1
383-6	Bcep22p48 (*Burkholderia cepacia* phage Bcep22)	63	NP_944277.1
	Orf63 (*Pseudomonas* phage D3)	61	NP_061559.1
383-11	bacteriophage protein (*Burkholderia oklahomensis* C6786)	24	ZP_02367173.1
	hypothetical protein Tuc2009_27 (*Lactococcus* phage Tuc2009)	20	NP_108706.1
383-14	hypothetical protein BcepC6B_gp42 (*Burkholderia* phage BcepC6B)	54	YP_024962.1
	Bacteriophage protein Gp48 (*Burkholderia pseudomallei* NCTC 13177)	55	ZP_02494570.1
	Bacteriophage protein Gp48 (*Burkholderia pseudomallei* K96243)	46	YP_110419.1
383-17	GP60 (*Burkholderia phymatum* STM815)	69	YP_001858131.1
	hypothetical protein ϕE125p61 (bacteriophage ϕE125)	67	NP_536417.1
	gp60 (*Burkholderia thailandensis* E264)	67	YP_439230.1
	gp71 (Bacteriophage ϕ1026b)	64	NP_945102.1
383-18	gp8 (*Burkholderia* phage Bcep176)	78	YP_355343.1
	bacteriophage protein (*Burkholderia multivorans* ATCC 17616)	78	YP_001949033.1
	gp72 (*Burkholderia oklahomensis* EO147)	65	ZP_02359388.1
383-19	gp65 (*Burkholderia pseudomallei* 1710a)	43	ZP_02109365.1
	gp65 (*Burkholderia* phage ϕ644-2)	49	YP_001111144.1

	hypothetical protein ϕE125p66 (bacteriophage ϕE125)	44	NP_536422.1
	gp76 (Bacteriophage ϕ1026b)	43	NP_945107.1
383-20	Bcep22gp19 (*Burkholderia cepacia* phage Bcep22)	51	NP_944247.1
	phage-related hypothetical protein (*Bordetella bronchiseptica* RB50)	51	NP_888225.1
	phage-related protein (*Bordetella avium* 197N)	47	YP_785969.1
383-25	hypothetical protein PPF10_gp063 (*Pseudomonas* phage F10)	25	YP_001293407.1
	hypothetical protein GOX2337 (*Gluconobacter oxydans* 621H)	33	YP_192725.1
383-28	Phage DNA packaging protein Nu1 subunit of terminase-like protein (*Rhodobacter sphaeroides* ATCC 17025)	32	YP_001166346.1
383-29	Phage terminase GpA (*Burkholderia lata* 383)	100	YP_371800.1
	COG5525: Bacteriophage tail assembly protein (*Yersinia bercovieri* ATCC 43970)	69	ZP_00822438.1
	terminase GpA (*Burkholderia phymatum* STM815)	47	YP_001858114.1
383-30	hypothetical protein GOX2338 (*Gluconobacter oxydans* 621H)	48	YP_192726.1
383-31	Lytic transglycosylase, catalytic (*Burkholderia lata* 383)	100	YP_371802.1
	λ family phage portal protein (*Escherichia coli* HS)	54	YP_001458777.1
	phage portal protein, λ family (*Escherichia coli* E22)	54	ZP_03045897.1
383-32	gp33 (*Burkholderia pseudomallei* DM98)	80	ZP_02402318.1
	gp33 (*Burkholderia pseudomallei* 1710a)	80	ZP_02109328.1
	hypothetical protein ϕE125p33 (*Burkholderia* phage ϕE125)	80	NP_536389.1
	gp32 (*Burkholderia* phage ϕ1026b)	80	NP_945063.1
383-34	putative bacteriophage protein (*Salmonella enterica* subsp. *enterica* serovar Typhi str. CT18)	28	NP_455545.1
	putative bacteriophage protein (*Salmonella enterica* subsp. *enterica* serovar Typhi str. CT18)	32	NP_456380.1
	putative bacteriophage protein (*Salmonella enterica* subsp. *enterica* serovar Typhi Ty2)	32	NP_805642.1

The three phage-associated proteins most closely related to the predicted *B. lata* 383 proteins (based on BLASTX analysis) are shown. In order to include proteins from ϕE125 and ϕ1026b (if present), four entries were included for 383-17, 383-19 and 383-32.

KS9 morphogenesis genes. The predicted DNA packaging/head assembly proteins of KS9 are similar to those encoded by PC184, ϕE125, and ϕ1026b, but the major capsid protein itself is dissimilar (Table 2-1 and 2-2). Each putative KS9 tail protein has a similar protein encoded by PC184, ϕE125, and ϕ1026b (Table 2-2). As discussed above, although KS9 gene *15* is assumed to be part of the tail morphogenesis module because of its location, its product does not yet have an assigned function.

In many phage genomes, two proteins are encoded between the major tail gene and tail tape measure gene by way of a -1 translational frameshift (324). It is thought that frameshifting allows phages to control the relative expression levels of two proteins during infection (324). In KS9, these proteins are gp11 (a tail assembly chaperone) and gp12 (a minor tail protein). These proteins are predicted to have the same N-terminal sequence, but gp12 (expressed via the frameshift) is predicted to be longer and have a different C-terminus. We have identified a putative frameshift site between bases 20,347 and 20,353. This site contains a stretch of seven adenine residues that can cause the ribosome to slip backwards by one position, resulting in a -1 frameshift. Although AAAAAAA is not the canonical frameshift sequence (XXXYYYZ, where Y is A or T), it is the same sequence identified for phages such as c2 of *Lactococcus* and PSA of *Listeria monocytogenes* (324). Providing further support that this is the correct frameshift sequence, we have used the mfold program (335) to identify a potential stem-loop structure formed by 65 downstream bases (Figure 2-6). Although such structures are not necessarily found at frameshift sites, their presence suggests a mechanism

by which the ribosome may stall and subsequently change its reading frame (324).

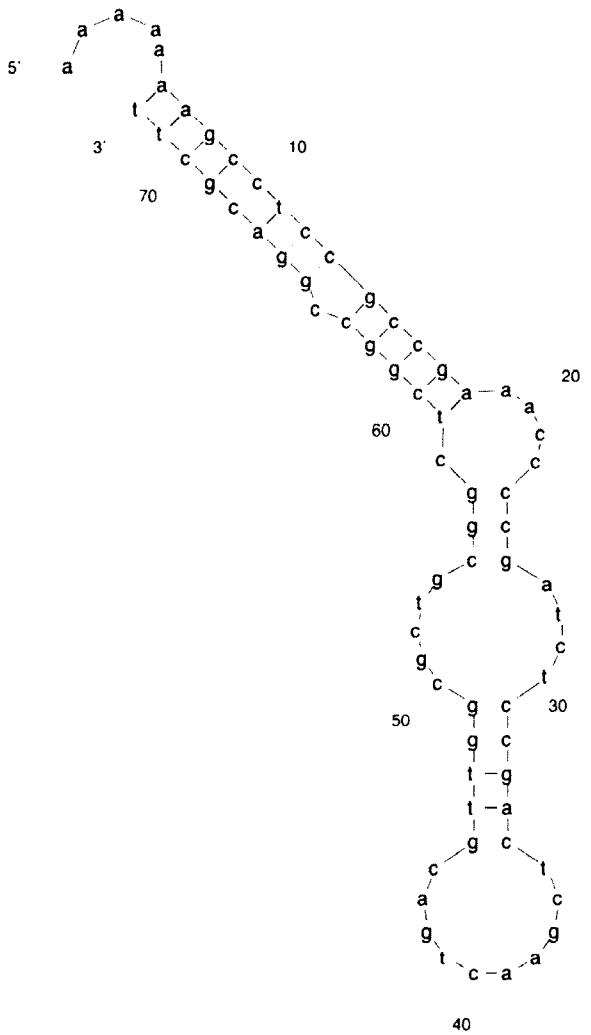


Figure 2-6: Predicted stem-loop structure downstream of the KS9 gp11 translational frameshift sequence. Structure was determined using mfold analysis of the 65 bases downstream of the AAAAAAA site (335). Reproduced with permission from the American Society for Microbiology.

KS9 lysis genes. KS9 encodes four proteins putatively involved in lysis, all of which are similar to proteins from PC184, but not ϕE125 or ϕ1026b (Table 2-2). The genes encoding these proteins are part of a single module (shown in blue in Figure 2-3). The first gene in the module, *22*, encodes a putative holin.

Class I holins have three transmembrane domains, whereas class II holins only have two (328). Because OCTOPUS analysis (310) indicates that gp22 has three transmembrane domains, we predict that this protein is a class I holin. Phage endolysins may be one of three major types: endo-β-*N*-acetylglucosaminidases or *N*-acetylmuramidases, endopeptidases, or *N*-acetylmuramyl-L-alanine amidases (82). We predict that the KS9 endolysin is an endopeptidase: gp23 is similar to M15A peptidases and is predicted to belong to the Peptidase_M15_3 superfamily (E-value = $2e^{-23}$). The KS9 Rz/Rz1 pair is the last gene segment in the lysis module. Rz/Rz1 proteins function in lysis by joining the inner and outer membranes following holin insertion and lysin activity (20). The Rz1 gene, *24'*, is positioned out of frame in the Rz gene, *24*. Using the LipoP program (136), it is predicted that a signal peptidase II cleavage site is present between residues 22 (alanine) and 23 (cysteine) of gp24'. Cleavage at this site would produce a 65 residue lipoprotein. Of these 65 amino acids, 7 are proline (10.8%), a value consistent with the proportion of proline found in the processed BcepMu (23.1%), SalMu (15.5%), and PhotoMu (13.4%) Rz1 lipoproteins (288).

Depending on the organism, the proteins similar to KS9 gp49 in the GenBank database have been annotated as either putative class I holins or as HNH homing endonucleases. If gp49 were to act as a class I holin, it would likely function with gp22 as part of a holin-antiholin pair. These systems are used to control when the onset of lysis occurs (150). However, gene *49* is not part of the lysis module and OCTOPUS analysis of the gp49 sequence suggests that it has no transmembrane domains. We predict that this protein is not a holin and that some

other protein encoded by KS9 may be involved in controlling lysis timing.

A Conserved Domain Search of the gp49 sequence indicates that it belongs to the HNHc superfamily (E-value = $8e^{-6}$) found in HNH homing endonucleases. Homing endonuclease genes (HEGs) are common in phage genomes: for example, T4 encodes 15 homing endonucleases, 6 of which (I-TevIII and MobA-E) belong to this family (204). HEGs are referred to as selfish genetic elements and use a mechanism called homing to copy their sequence from one place in one gene to the same place in the same gene at a different locus (38). In order to complete this process, endonucleases (such as those in the HNH family) are used to cut the DNA at a specific 15-30 bp site which is then fixed using double-strand break repair. Further experiments are required to determine if gp49 has homing endonuclease activity or if it performs some other function during infection.

Contribution of the KS9 prophage to Bcc virulence. In many pathogenic bacterial species, prophage genes make significant contributions to host virulence. Although examples have been documented in a wide variety of Gram-negative organisms (including *Vibrio cholerae*, *Escherichia coli*, *Pseudomonas aeruginosa*, *Neisseria meningitidis*, *Salmonella enterica*, and *Shigella flexneri*), there is limited evidence that prophage genes contribute to virulence in *Burkholderia* species (26, 286). It has been suggested that because *Burkholderia* are not strictly pathogenic and can instead survive in a variety of both terrestrial and aquatic environments, classical toxin genes may not have provided a strong evolutionary advantage to these species (286). Instead, lysogenic conversion

genes of *Burkholderia* prophages would be more likely to encode proteins that would increase the viability of the cell both in the environment and *in vivo* (286).

In the KS9 genome, we identified one gene whose product may have an effect on the pathogenicity of LMG 21824 or a K56-2 KS9 lysogen – gene *32*. When the gp32 sequence was analyzed using a Conserved Domain Search, the domain "COG3950: predicted ATP-binding protein involved in virulence" was identified (E-value = $6e^{-16}$). To determine if KS9 integration increases the pathogenicity of *B. cenocepacia* K56-2, we compared the virulence of wild-type and KS9-lysogenized K56-2 (K56-2::KS9) in the *Galleria mellonella* model. In this invertebrate infection model, the 50% lethal dose (LD_{50}) for K56-2 is 900 colony forming units (CFU) (266). For both K56-2 and K56-2::KS9, infection with between 10^4 and 10^6 CFU resulted in 100% mortality (Figure 2-7). There was no significant difference between the mortality of larvae infected with either K56-2 or K56-2::KS9 at any of the doses tested (Figure 2-7). Although gene *32* contains a domain putatively involved in virulence, these data suggest that it has no effect on the pathogenicity of KS9 lysogens.

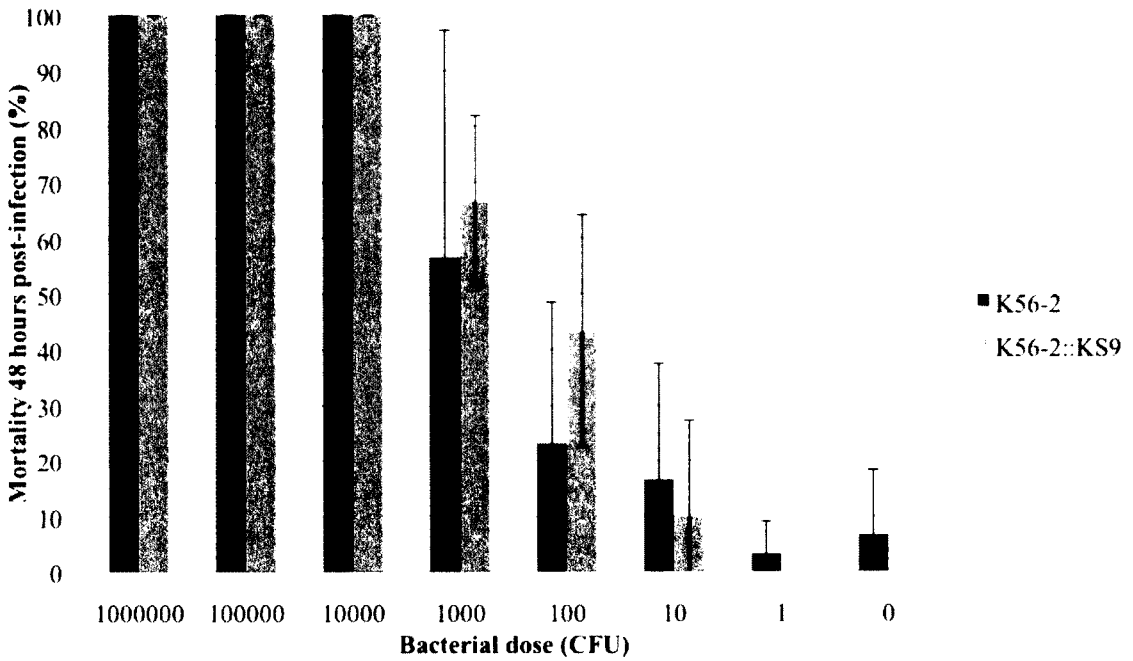


Figure 2-7: Mortality of *Galleria mellonella* infected with K56-2 and KS9-lysogenized K56-2 (K56-2::KS9) 48 h post-infection. For each bacterial dose, ten larvae were infected in three separate experiments.

Construction and analysis of a KS9 lytic variant. To convert KS9 into a lytic phage, we insertionally-inactivated the putative KS9 repressor gene, *41*. It is well-established that repressor mutant phages cannot stably lysogenize (290). Platt et al. (235) constructed a strain of *E. coli* expressing λ CI from the chromosome and lysogenized this strain with a λ *c*I mutant, W30 (18). When this strain was co-cultured with λ-sensitive *E. coli*, the lytic phages released from the lysogens were able to kill the sensitive cells. The authors proposed that this system could be used to deliver lytic phages *in vivo* (235). We wanted to expand on this concept in two ways: by showing that a lytic mutant could be constructed using genetic engineering and by showing that this mutant phage could be active *in vivo*.

Gene *41* encodes a 133 amino acid protein. According to BLASTP analysis, this protein shows 55% identity with gp52, the putative repressor protein of ϕE125 (E-value = $4e^{-30}$), as well as similarity to CI-like proteins of phages BP-4795, H-19B, SfV, cdtI, and a putative transcriptional repressor of phage V10. A suicide vector pKL1 (Figure 2-1) – containing DNA flanking gene *41* interrupted by a trimethoprim-resistance (Tp^R) cassette – was constructed in a tetracycline-resistant (Tc^R) pALTER-1 backbone. After transformation of this construct into *B. pyrrocinia* LMG 21824, a Tc^RTp^R mutant (LMG 21824::pKL1, generated by a single-crossover) was isolated.

If a double-crossover event occurs that replaces gene *41* with a Tp^R cassette, the cells should no longer be viable because the prophage will be induced in the absence of the repressor protein, resulting in lysis. We used two different assays to assess if this induction was occurring in LMG 21824::pKL1. First, we screened 1800 Tc^RTp^R LMG 21824::pKL1 colonies for loss of tetracycline-resistance and maintenance of trimethoprim-resistance (i.e. a double-crossover without loss of viability). No Tc^STp^R colonies were found, suggesting that the double-crossover event was lethal due to prophage induction. Second, we compared the release of K56-2-infecting phages from LMG 21824, LMG 21824::pKL1, and LMG 21824 with a mutation in KS9 gene *32* (LMG 21824 [KS9 *32*]). The number of phages released from LMG 21824::pKL1 was more than 10 times greater than the number released from the other two strains (Figure 2-8), a finding consistent with phage induction occurring in LMG 21824::pKL1 following a double-crossover event. These two experiments provide further

evidence that gene *41* is the repressor, as it is required to maintain stable integration of the KS9 prophage in LMG 21824.

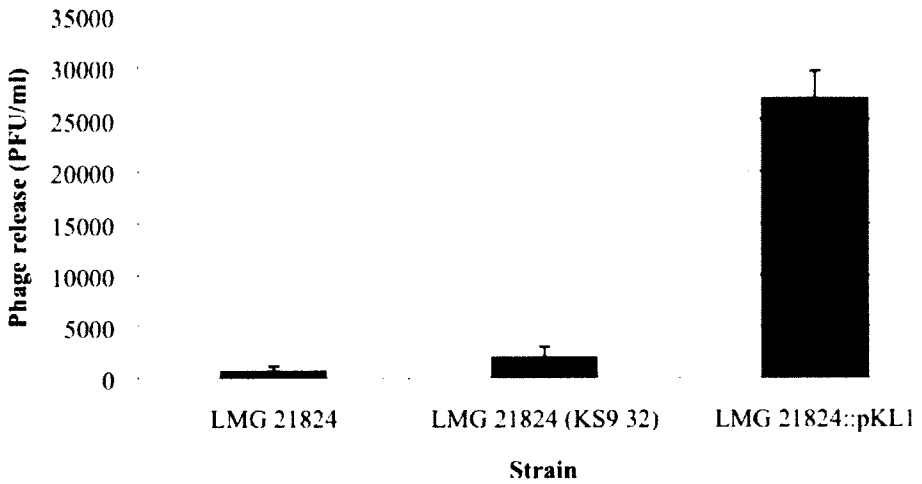


Figure 2-8: Enumeration of K56-2-infecting phages released into the supernatant from 16 h overnight cultures of LMG 21824, LMG 21824 (KS9 *32*), and LMG 21824::pKL1. Cultures were pelleted by centrifugation at 10,000 x *g* for 2 min and the supernatant was filter sterilized using a 0.45 μm filter unit and tested in a plaque assay with K56-2. This procedure was repeated three times for each strain. Reproduced with permission from the American Society for Microbiology.

Phages isolated from culture supernatants of LMG 21824::pKL1 were screened for the presence of the Tp^R cassette. Using primers 40F and 42R (flanking gene *41*), a PCR product of 603 bp was amplified from wild-type KS9, whereas a product of 832 bp was amplified from phages with the integrated Tp^R cassette (Figure 2-9). A representative phage isolate that tested positive during PCR screening was named KS9c. This isolate showed no defects in activity and was able to produce confluent lysis of *B. cenocepacia* K56-2 in an agar overlay at titres of 10^7/ml, similar to wild-type KS9. The host range of KS9c was found to be identical to that of wild-type KS9, suggesting that there was no change in the susceptibility of Bcc strains to this phage.

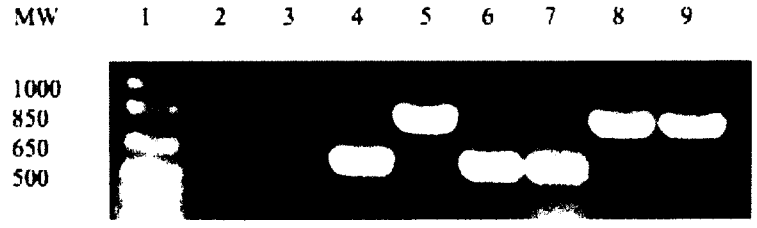


Figure 2-9: PCR amplicons for KS9 and KS9c with primers 40F and 42R. PCR products were separated by electrophoresis on a 0.8% (wt/vol) agarose gel and visualized with ethidium bromide. Sizes of the 1 Kb Plus Ladder (Invitrogen) markers (lane 1) are shown on the left of the gel (MW, molecular weight). Lane 2: no DNA; lane 3: K56-2 DNA; lane 4: LMG 21824 DNA; lane 5: pKL1; lane 6: KS9 lysate; lane 7: KS9 DNA; lane 8: KS9c lysate; lane 9: KS9c DNA.

To assess if KS9c is able to stably integrate into K56-2, plates of K56-2 and KS9 or K56-2 and KS9c exhibiting confluent lysis were overlaid with water and the surviving cells were isolated. One hundred and fifty isolates from KS9c lysates were replica plated onto solid medium with or without trimethoprim. All of these cells were Tp^S, suggesting that KS9c had not stably integrated into any of these bacteria. A total of 50 KS9-insensitive and 50 KS9c-insensitive K56-2 isolates were recovered from the phage lysates and were assayed for a) growth phenotype, b) trimethoprim-resistance, c) phage sensitivity, and d) PCR-positivity with primers designed to the K56-2/5' KS9 prophage junction. The majority of KS9-insensitive isolates (41 isolates, 82%) were non-lysogenized (Figure 2-10). These bacteria grew normally, were Tp^S, insensitive to KS9 and KS9c, and were PCR-negative for KS9 integration. Eighteen percent (9 isolates) were stably lysogenized (Figure 2-10). Like the non-lysogenized isolates, these bacteria grew normally, were Tp^S, and insensitive to KS9 and KS9c, but they were PCR-positive for integration. A different distribution was observed for the KS9c-insensitive isolates. In this case, all 50 isolates (100%) were non-lysogenized

(Figure 2-10). Like the non-lysogenized KS9-insensitive isolates, these bacteria grew normally, were Tp^S, insensitive to KS9 and KS9c, and were PCR-negative for integration. While almost one-fifth of KS9-insensitive isolates were stably lysogenized, none of the KS9c-insensitive isolates showed this phenotype (Figure 2-10). Consequently, we conclude that KS9 gene *41* encodes the repressor protein and that KS9 is converted from a temperate to a functionally lytic phage following the loss of this gene.

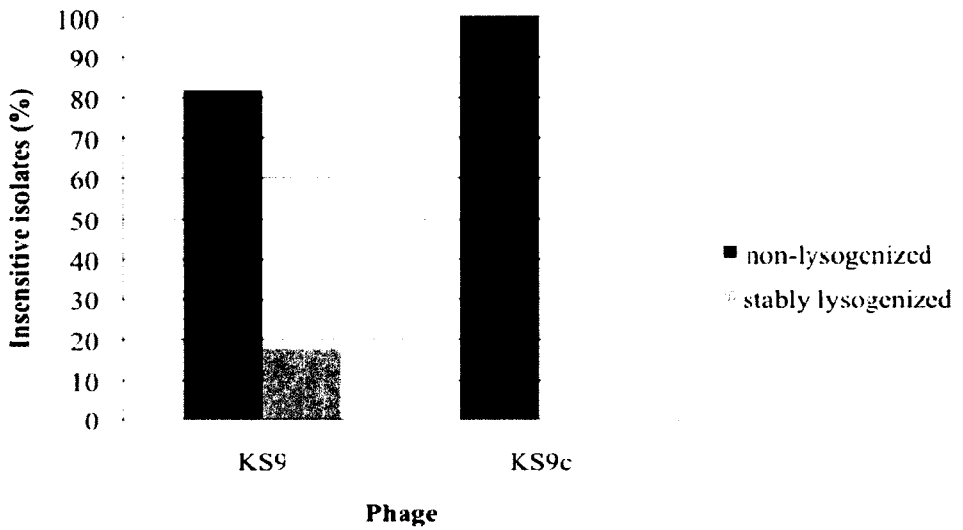


Figure 2-10: Prevalence of stable lysogeny in KS9-insensitive and KS9c-insensitive K56-2 isolates. Plates of K56-2 and KS9 or K56-2 and KS9c exhibiting confluent lysis were overlaid with water and surviving cells were isolated and assayed for phage-insensitivity. Lysogeny was assayed in phage-insensitive cells using PCR primers to the K56-2/5' KS9 junction.

We used the *G. mellonella* infection model to assess the *in vivo* efficacy of KS9c. Seed and Dennis (267) showed that *G. mellonella* larvae infected with a lethal dose of K56-2 could be rescued after administration of phages KS4-M or KS12. We tested the efficacy of treatment with endotoxin-removed KS9c lysates in K56-2-infected larvae at MOIs between 0.5 and 100. Treatment with KS9c at

MOIs of 50 or 100 increased the survival of infected larvae to that of uninfected controls (Figure 2-11). This result is in sharp contrast to untreated infected larvae that had survival rates near zero (Figure 2-11). Phage treatment had little or no effect at MOIs lower than 50: the survival of larvae after treatment was relatively low (Figure 2-11) and those that remained alive 48 h post-infection exhibited both reduced movement and signs of melanization, suggesting that K56-2 infection had progressed but was not yet fatal. This result is expected, as higher MOIs of lytic phages are generally more effective for therapeutic purposes (24). Our results indicate that KS9c is active in an *in vivo* model and can effectively treat experimental *B. cenocepacia* K56-2 infections when administered at an MOI of 50 or greater.

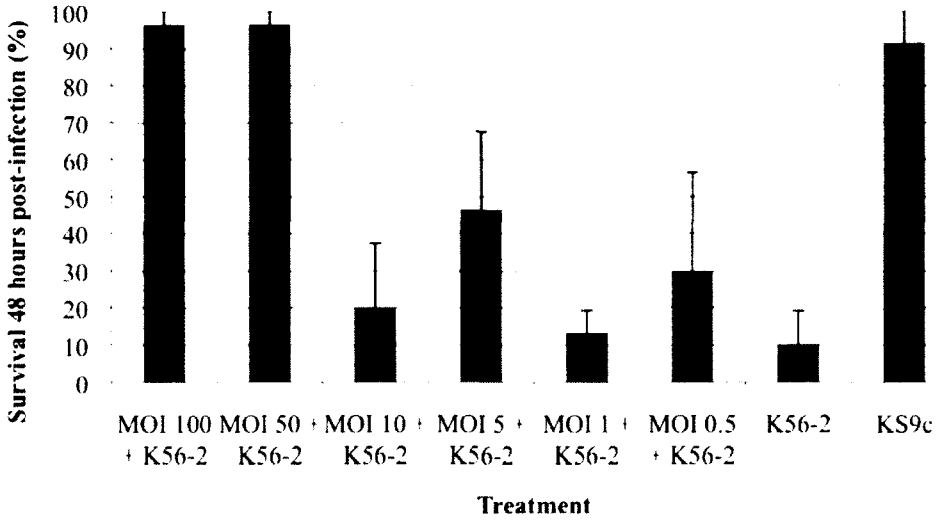


Figure 2-11: Survival of K56-2-infected and KS9c-treated *Galleria mellonella* 48 h post-infection. Larvae were infected with twice the LD_{50} of K56-2 and treated with KS9c at MOIs between 0.5 and 100. Control larvae were injected with either twice the LD_{50} of K56-2 and $MgSO_4$-ampicillin solution or undiluted KS9c and $MgSO_4$-ampicillin solution. For each MOI, 10 larvae were infected and treated in three separate experiments. Reproduced with permission from the American Society for Microbiology.

Because KS9c is unable to stably lysogenize K56-2 and facilitate

superinfection immunity, we predicted that KS9c might be a more effective therapeutic agent than KS9. We treated K56-2-infected larvae with either KS9 or KS9c at MOIs of 50, 5, and 0.5 and monitored larvae survival. In contrast to our prediction, we found no significant difference between the survival of larvae treated with the two different phages (Figure 2-12). In order to assess why no difference was observed between the two treatments, we isolated K56-2 from the hemolymph of infected, KS9-treated larvae and screened these isolates for lysogeny. In contrast to the *in vitro* screening presented above (where nearly 20% of insensitive isolates were stably lysogenized), 0 of 50 KS9-insensitive K56-2 hemolymph isolates were lysogenized based on PCR screening.

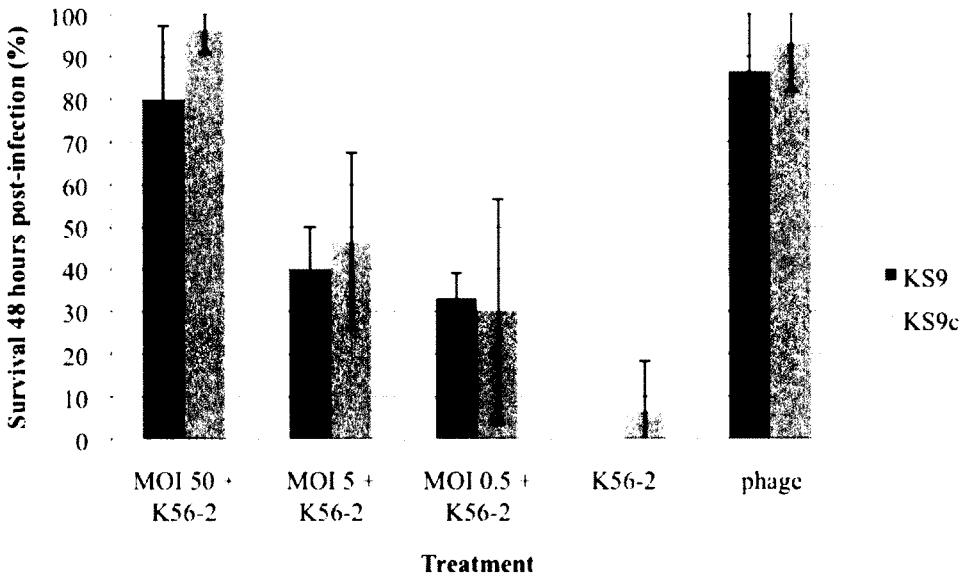


Figure 2-12: Survival of K56-2-infected and KS9- or KS9c-treated *Galleria mellonella* 48 h post-infection. Larvae were infected with twice the LD_{50} of K56-2 and treated with KS9 or KS9c at an MOI of 50, 5, or 0.5. Control larvae were injected with either twice the LD_{50} of K56-2 and $MgSO_4$-ampicillin solution or undiluted KS9 or KS9c and $MgSO_4$-ampicillin solution. For each MOI, 10 larvae were infected and treated in three separate experiments. Reproduced with permission from the American Society for Microbiology.

Because lysogeny does not appear to make a significant contribution to

KS9 resistance in infected and treated larvae, this result explains why there was no significant difference in the *in vivo* efficacy of KS9 and KS9c. Differing phage MOIs most likely caused the discrepancy observed in the proportion of lysogenized bacterial isolates obtained *in vitro* and *in vivo*. MOIs were initially higher for the *in vivo* screening (100 *in vivo* versus approximately 0.1-0.01 *in vitro*), which would tend to favor lysogeny. However, rapid infection, lysis and release of progeny phage would be predicted to increase the effective MOI *in vitro* more so than *in vivo*, leading to a greater proportion of lysogenized cells in the lysate than in the larvae (230).

Although KS9c has been shown to be effective *in vivo* in this proof-of-principle study, there are two reasons why this particular phage could not be used clinically. First, KS9c retains other sequences required for lysogeny, including operators and an integrase gene. As a result, KS9c can integrate into the host chromosome but the prophage cannot be maintained. We observed evidence of this phenotype, as some KS9c lysate isolates were PCR-positive, but a) either grew normally, poorly, or produced plaques when struck out, b) were Tp^S, and c) exhibited confluent lysis when plated in an agar overlay with KS9c, while producing plaques or confluent lysis on negative control plates. In addition, recombination with prophage sequences could potentially restore KS9c to a temperate phenotype if a compatible repressor gene were to integrate into the phage chromosome. In future experiments, all integration-related sites and genes should be knocked out in order to prevent transient prophage integration and minimize the effects of recombination. The second issue with KS9c in its current

form is that it carries a trimethoprim-resistance cassette that could potentially be transferred to the host upon infection. We did not see evidence of recombination in this study as all KS9c-insensitive cells remained trimethoprim-sensitive. However, to ensure that antibiotic resistance genes are not transferred from the phage genome to the host, similar deletions constructed in future experiments will be unmarked.

CONCLUSIONS

We have determined the genome sequence of the siphovirus KS9, a prophage of *B. pyrrocinia* LMG 21824. The KS9 prophage integrates into both LMG 21824 and *B. cenocepacia* K56-2 in the 3' end of the GTP cyclohydrolase II gene. It is 39,896 bp in length and encodes 50 proteins, many of which show similarity to proteins of the *Burkholderia* prophages ϕE125 and ϕ1026b and an uncharacterized prophage element of *B. cenocepacia* PC184. Although one gene product was identified with a putative role in pathogenicity, KS9 integration was shown to have no effect on the virulence of *B. cenocepacia* K56-2 in the *G. mellonella* model. We were able to show that gp41 encodes the phage repressor by developing a functionally lytic mutant of KS9 named KS9c. Unlike KS9, KS9c is unable to stably lysogenize LMG 21824 or K56-2. Treatment of K56-2-infected *G. mellonella* larvae with KS9c at MOIs of 50 or greater resulted in larvae survival comparable to uninfected controls, indicating that KS9c is an effective antibacterial agent in a model system. As a proof-of-principle, we have shown that temperate phages can be genetically engineered to a lytic form and that these

engineered phages are active *in vivo*.

ACKNOWLEDGMENTS

We thank David DeShazer (U.S. Army Medical Research Institute for Infectious Diseases) for testing the *B. pseudomallei* and *B. mallei* host ranges and Miguel Valvano (University of Western Ontario) for providing the *B. cenocepacia* LPS mutants. We also thank Randy Mandryk (University of Alberta Advanced Microscopy Facility) and Pat Murray and Lisa Ostafichuk (University of Alberta Molecular Biology Service Unit) for their assistance.

Chapter 3

Genomic analysis and relatedness of P2-like phages of the *Burkholderia cepacia* complex

A version of this chapter has been published as:

Lynch, K. H., P. Stothard, and J. J. Dennis. 2010. Genomic analysis and relatedness of P2-like phages of the *Burkholderia cepacia* complex. BMC Genomics **11**:599.

OBJECTIVES

The objectives of this project were to sequence and characterize the genomes of three novel Bcc-specific phages: KS5 (vB_BceM_KS5 or vB_BmuZ_ATCC 17616), KS14 (vB_BceM_KS14), and KL3 (vB_BamM_KL3 or vB_BceZ_CEP511).

MATERIALS AND METHODS

Bacterial strains and growth conditions. Bcc strains used for phage isolation and propagation were obtained from the Belgium Coordinated Collection of Microorganisms LMG Bacteria Collection (Ghent, Belgium) and the Canadian *Burkholderia cepacia* complex Research and Referral Repository (Vancouver, BC). Many of the strains used are from the *Burkholderia cepacia* complex experimental strain panel and updated experimental strain panel (59, 183). Strains were grown aerobically overnight at 30°C on half-strength Luria-Bertani (½ LB) solid medium or in ½ LB broth with shaking. Transformations were performed with chemically-competent DH5α (Invitrogen, Carlsbad, CA), plated on LB solid medium containing 100 μg/ml ampicillin, and grown aerobically overnight at 37°C. Strains were stored in LB broth containing 20% glycerol at -80°C.

Electron microscopy. To prepare samples for transmission electron microscopy, phage lysates were filter sterilized using a Millex-HA 0.45 μm syringe driven filter unit (Millipore, Billerica, MA), incubated on a carbon-coated copper grid 5 min at room temperature, and stained with 2% phosphotungstic acid for 2 min. Micrographs were taken with the assistance of the University of

Alberta Department of Biological Sciences Advanced Microscopy Facility using a Philips/FEI (Morgagni) transmission electron microscope with charge-coupled device camera at 140,000-fold magnification.

Phage isolation, propagation, and DNA isolation. Isolation of KS5 from onion soil and KS14 from *Dracaena* sp. soil has been described previously (265, 267). KL3 was isolated from a single plaque on a lawn of *B. cenocepacia* CEP511. The plaque was isolated using a sterile Pasteur pipette, suspended in 1 ml of suspension medium (SM; 50 mM Tris-HCl [pH 7.5], 100 mM NaCl, 10 mM $MgSO_4$, 0.01% gelatin solution) with 20 µl $CHCl_3$ and incubated 1 h at room temperature to generate a KL3 stock. KL3 was propagated on *Burkholderia ambifaria* LMG 17828 in soft agar overlays: 100 µl of phage stock and 100 µl of liquid culture were incubated 20 min at room temperature and 3 ml of soft nutrient agar was added to this mixture, poured onto ½ LB solid medium, and incubated overnight at 30°C.

Phage genomic DNA was isolated using a modified version of a λ proteinase K/SDS lysis protocol (261). ½ LB agarose plates (prepared with soft nutrient agarose) showing confluent phage lysis were overlaid with 3 ml of SM and incubated for 6 h at 4°C on a platform rocker. The lysate was pelleted by centrifugation at 10 000 x *g* for 2 min and filter-sterilized using a 0.45 µm filter unit. 10 ml of lysate was treated with 10 µl DNase I/10 µl DNase I buffer and 6 µl RNase I (Fermentas, Burlington, ON) and incubated 1 h at 37°C. Following addition of 0.5 M EDTA (pH 8.0) to 20 mM, proteinase K to 50 µg/ml, and SDS to 0.5%, the solution was mixed and incubated 1 h at 37°C. Standard

phenol:chloroform extraction and ethanol precipitation were then used to purify the phage DNA. Samples were resuspended in TE (pH 8.0) and quantified using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific, Waltham, MA).

KS14 plasmid prophage DNA was isolated from five putatively lysogenized KS14-resistant C6433 isolates (267) using a QIAprep Spin Miniprep kit (Qiagen, Hilden, Germany). Lysogeny was predicted using PCR with KS14-specifc primers (KS14F: GCAGCTAACCGAGTCGCACG, KS14R: CTCTGAAAAGGTGGGCGGTGG) (Sigma-Genosys, Oakville, ON) and TopTaq DNA polymerase and buffers (Qiagen). *Burkholderia multivorans* ATCC 17616 and *Burkholderia cenocepacia* C6433, CEP511 and K56-2 were used as negative controls. Two milliliter aliquots of 16 h liquid cultures (OD_{600}: 2.0-2.2) were pelleted, washed 3x with sterile H_2O to remove exogenous phages, and treated using the standard kit protocol. For each sample, two 20 µl EcoRI (Invitrogen) reactions each containing 17 µl of plasmid DNA were digested overnight, pooled and separated on 0.8% (wt/vol) agarose gels in 1x Tris-acetate-EDTA (pH 8.0).

Sequencing and bioinformatics analysis. Preliminary sequence analysis was performed using a shotgun cloning protocol. Phage DNA was digested using EcoRI (Invitrogen), separated on 0.8% (wt/vol) agarose gels, purified using the GENECLEAN II kit (Qbiogene, Irvine, CA), ligated into pUC19 or pGEM-7Z, and transformed into DH5α (Invitrogen). Following blue-white selection on LB solid medium containing 100 µg/ml ampicillin, constructs with phage DNA inserts were isolated using a QIAprep Spin Miniprep kit (Qiagen), digested using

EcoRI, and viewed using gel electrophoresis. Inserts were sequenced with the assistance of the University of Alberta Department of Biological Sciences Molecular Biology Service Unit using an ABI 3730 DNA analyzer (Applied Biosystems, Foster City, CA). Sequences were edited using EditView and aligned using AutoAssembler (Perkin-Elmer, Waltham, MA). For completion of the three genomes, DNA samples were submitted for pyrosequencing analysis (454 Life Sciences, Branford, CT). Gaps between the assembled sequences were filled following PCR amplification and cloning using primers (Sigma-Genosys) designed to amplify across the gaps, TopTaq DNA polymerase and buffers (Qiagen), and the CloneJET PCR cloning kit (Fermentas). The complete genome sequences of KS5, KS14, and KL3 were deposited in GenBank with the accession numbers GU911303, HM461982, and GU911304, respectively.

Annotation of the assembled sequences was performed using GeneMark.hmm-P (http://exon.biology.gatech.edu) (172). For KS5, annotations were based on those of the ATCC 17616 chromosome 2 sequence (GenBank: NC_010805.1; BMULJ_03640-BMULJ_03684, bp 477,496-514,731). Manual annotations were performed for the *E+E'* and *lysC/Rz1* genes. Proteins were numbered based on the order of the genes in the prophage (i.e. the integrase gene was named *1* and the integrase was named gp1). Relatedness of the predicted proteins was assessed using BLASTP (http://blast.ncbi.nlm.nih.gov) (7). Protein transmembrane domains, stem-loop structures, and signal peptidase II cleavage sites were identified using OCTOPUS (http://octopus.cbr.su.se), mfold (http://mfold.rna.albany.edu), and LipoP (http://www.cbs.dtu.dk/services/LipoP),

respectively (136, 310, 335). Repeat sequences in the DNA were identified using REPuter (http://bibiserv.techfak.uni-bielefeld.de/reputer) (149). Restriction sites were predicted using NEBcutter (http://tools.neb.com/NEBcutter2) (311). Whole genome sequence comparisons were performed using CoreGenes with a stringency setting of 75 (http://www.binf.gmu.edu/genometools.html) (154, 330). Comparison figures were constructed using PROmer (http://mummer.sourceforge.net) and Circos (http://mkweb.bcgsc.ca/circos) (68, 147).

To identify the KS5 prophage insertion site in ATCC 17616, the assembled KS5 sequence was compared to the vB_BmuZ_ATCC 17616 sequence in a BLASTN search and the left prophage junction was determined. Primers designed to this region (KS5*attL*F: TGCACGGCGAGCTGAAACTG, KS5*attL*R: GAAGGCACGCGAGGTAGAACG) were used to amplify the C6433/KS5 prophage junction in C6433 lysogens. To identify the KL3 insertion site in CEP511, the region proximal to the KL3 integrase gene *1* was analyzed using BLASTN and found to be similar to a region containing a tRNA-Thr gene in several *Burkholderia* strains including *B. ambifaria* AMMD (Bamb_R0016; chromosome 1, bp 403,358-403,433). Primers designed to this region (KL3*attL*F: AGCTGCAGATGGGTAACGAGTGG, KL3*attL*R: CCACTCACGAAGGGCAAGCTG) were used to amplify the CEP511/KL3 prophage junction.

RESULTS AND DISCUSSION

Isolation, host range, and morphology. Seed and Dennis (265) isolated KS5 from an extract of onion soil plated on *B. cenocepacia* K56-2. This phage produces clear plaques on K56-2 with a diameter of 0.5-1.0 mm (265). As discussed in Chapter 2, KS5 was tested for its ability to form plaques on K56-2 LPS mutants and it was determined that it could efficiently lyse wild-type K56-2 (EOP = 1), XOA7 (*waaL*::pGPΩTp, EOP = 0.8), XOA15 (*wabR*::pGPΩTp, EOP = 1.3), XOA17 (*wabS*::pGPApTp, EOP = 1.1) and RSF19 (*wbxE*::pRF201, EOP = 0.5), but not XOA8 (*wabO*::pGPΩTp) or CCB1 (*waaC*::pGPΩTp) (168, 176, 224). Based on these results, it was predicted that KS5 uses the K56-2 LPS as a receptor and that it binds within the inner core region (176). KS5 has a relatively wide host range compared to many Bcc phages, infecting *B. multivorans* C5393 and *B. cenocepacia* 715J, J2315, K56-2, C6433, and C5424 (265).

KS14 was isolated from an extract of *Dracaena* sp. soil plated on *B. multivorans* C5393 (267). Both Bcc phages and bacteria have been recovered from the *Dracaena* rhizosphere (267, 300). On C5393, KS14 forms small clear plaques 0.5-1.0 mm in diameter, similar to the morphology of KS5 on K56-2. The host range of KS14 includes *B. multivorans* C5393 and C5274, *B. cenocepacia* 715J, C6433, C5424 and PC184, *Burkholderia dolosa* LMG 21443, and *B. ambifaria* LMG 17828 (267).

KL3 was isolated from a single plaque on a lawn of *B. cenocepacia* CEP511, an Australian CF epidemic isolate (183). Phage induction from CEP511 was stochastic, as treatment with inducing agents such as UV or mitomycin C was

not necessary. On LMG 17828, KL3 forms small turbid plaques 0.5-1.0 mm in diameter. KL3 has a narrow host range, infecting *B. ambifaria* LMG 17828.

Electron microscopy of KS5, KS14, and KL3 indicates that these phages belong to the family *Myoviridae* (Figure 3-1). These three phages exhibit the A1 morphotype with icosahedral capsids and contractile tails (4). KS5, KS14, and KL3 have similarly sized capsids, each 65 nm in diameter (Figure 3-1). In contrast, their tails vary in length: 140 nm for KS14, 150 nm for KS5, and 160 nm for KL3 (Figure 3-1). These sizes correspond to the length of the tail tape measure protein for each of these three phages: KS14 gp12 is 842 amino acids (aa) in length, KS5 gp15 is 920 aa, and KL3 gp17 is 1075 aa (Tables 3-1, 3-2, and 3-3).

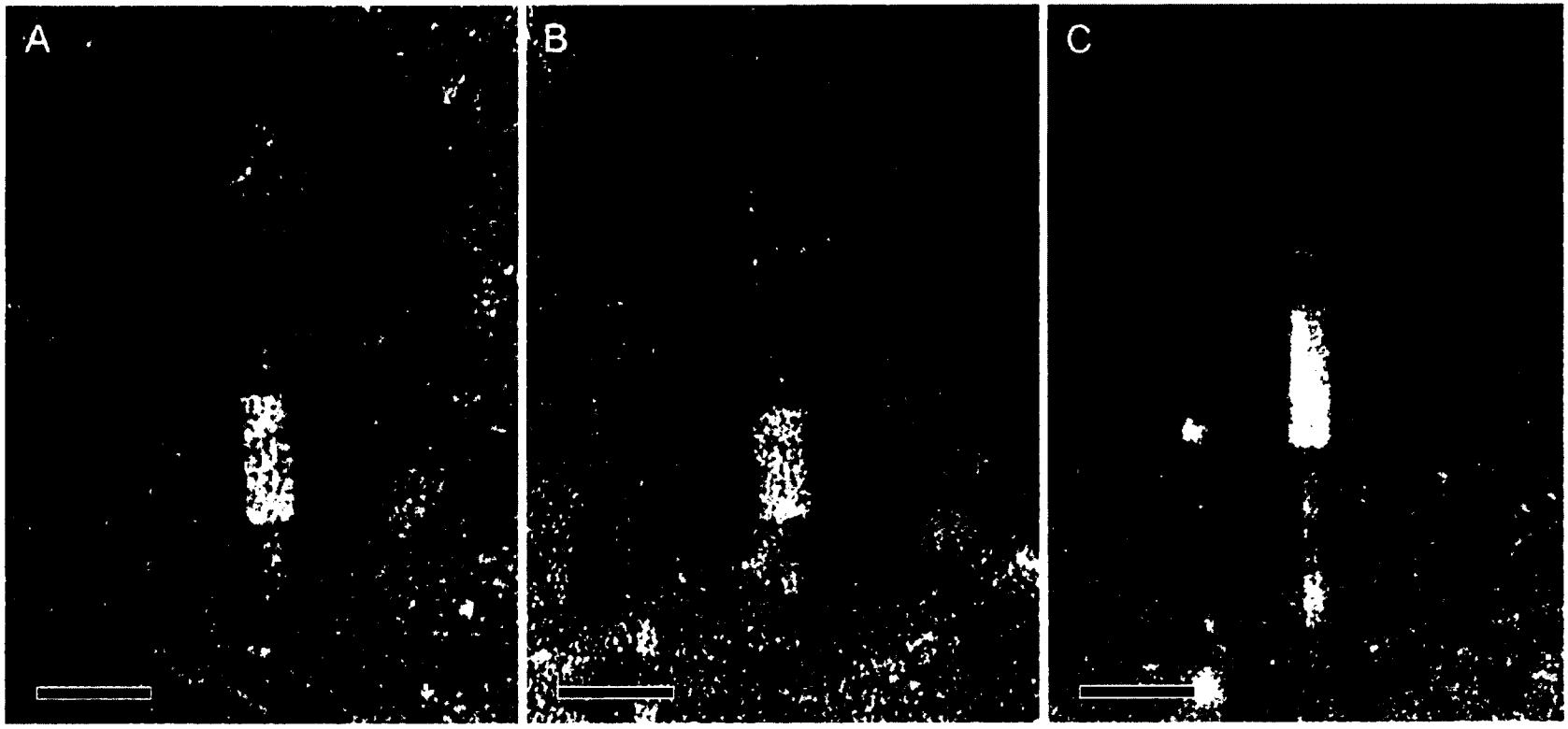


Figure 3-1: Transmission electron micrographs of KS5 (A), KS14 (B) and KL3 (C). Phages were stained with 2% phosphotungstic acid and viewed at 140,000-fold magnification. Scale bars represent 50 nm.

Table 3-1: KS5 genome annotation

Gene	Start	End	Putative function	Strand	Predicted RBS and start codon	Length (aa)	Closest relative (excluding ATCC 17616)	Alignment region in closest relative (aa)	% ID	Source	GenBank accession no.	ATCC 17616 locus tag	ATCC 17616 GenBank accession number
1	108	815	integrase	-	AGCAACAAG cacaaggcaTTG	235	integrase family protein	134-368/368	87	*Burkholderia thailandensis* MSMB43	ZP_02468407.1	BMULJ_03640	YP_001948048.1
2	2142	4937	zinc finger CHC2-family protein	-	GAGCAACAG caataacgATG	931	conserved hypothetical protein	1-931/931	96	*Burkholderia multivorans* CGD1	ZP_03587581.1	BMULJ_03641	YP_001948049.1
3	4940	5200	unknown	-	GGGGGAAGc cgcATG	86	conserved hypothetical protein	1-86/86	91	*Burkholderia multivorans* CGD1	ZP_03587582.1	BMULJ_03642	YP_001948050.1
4	5197	5556	unknown	-	GGGGGTGAtg tgATG	119	conserved hypothetical protein	1-119/119	97	*Burkholderia multivorans* CGD1	ZP_03587583.1	BMULJ_03643	YP_001948051.1
5	5561	5755	membrane protein	-	GGAGccaaacc ATG	64	putative phage-encoded membrane protein	1-64/64	78	*Burkholderia ambifaria* MEX-5	ZP_02905725.1	BMULJ_03644	YP_001948052.1
6	5798	6001	unknown	-	GGATGcactgac cgATG	67	conserved hypothetical protein	1-67/67	92	*Burkholderia multivorans* CGD1	ZP_03587585.1	BMULJ_03645	YP_001948053.1
7	6005	6199	unknown	-	GGAGAGActc ATG	64	conserved hypothetical protein	1-64/64	98	*Burkholderia multivorans* CGD1	ZP_03587586.1	BMULJ_03646	YP_001948054.1
8	6289	6537	transcriptional activator (Ogr)	-	GTAGGAGccc cgaATG	82	transcriptional activator Ogr/delta	1-82/82	91	*Burkholderia cenocepacia* MC0-3	YP_001763475.1	BMULJ_03647	YP_001948055.1
9	6547	6825	DNA binding protein	-	GGGCGttgagtc ATG	92	putative phage DNA-binding protein	1-92/92	98	*Burkholderia ambifaria* MEX-5	ZP_02905729.1	BMULJ_03648	YP_001948056.1
10	6829	7065	unknown	-	GAAGGGAAG tatacgtcATG	78	putative bacteriophage protein	1-77/78	80	*Burkholderia* sp. CCGE1001	ZP_06292840.1	BMULJ_03649	YP_001948057.1
11	7121	7603	repressor	+	GATAATACA cacgatcgGTG	160	putative phage DNA-binding protein	12-166/167	79	*Burkholderia pseudomallei* K96243	YP_106769.1	BMULJ_03650	YP_001948058.1
12	7660	8406	membrane	+	AGGGAAttcaA	248	putative phage-	1-241/249	43	*Burkholderia*	YP_1067	BMULJ_0	YP_00194

			protein		TG		encoded membrane protein			*pseudomallei* K96243	70.1	3651	8059.1
	8971	8975	direct repeat flanking IS*Bmu*23										
IS*Bmu* 23	8976	10185	IS*Bmu*23 insertion sequence										
	8976	8991	IS*Bmu*23 inverted repeat										
	9063	10055	IS*Bmu*23 transposase	+	GGAACGGAcc cacgacgATG	330	transposase IS*4* family protein	1-330/330	100	*Burkholderia* sp. Ch1-1	ZP_0684 6513.1	BMULJ_0 3652	YP_00194 8060.1
	10170	10185	IS*Bmu*23 inverted repeat										
	10185	10189	direct repeat flanking IS*Bmu*23										
13	10359	11510	tail protein (D)	-	AAGGAGGcga tctcgctATG	383	phage late control gene D protein	1-379/382	96	*Burkholderia multivorans* CGD1	ZP_0358 7594.1	BMULJ_0 3653	YP_00194 8061.1
14	11507	11935	tail protein (U)	-	GAAGGAGGG AttgtcATG	142	bacteriophage gpU	1-142/142	95	*Burkholderia multivorans* CGD1	ZP_0358 7595.1	BMULJ_0 3654	YP_00194 8062.1
15	11949	14711	tail tape measure protein (T)	-	GAGCGAGGc gacgaATG	920	putative phage-related tail transmembrane protein	1-919/919	91	*Burkholderia cenocepacia* MC0-3	YP_0017 63483.1	BMULJ_0 3655	YP_00194 8063.1
16	14827	15138	tail protein (E)	-	AGAGGAAcca tacgATG	103	phage tail protein E	1-103/103	97	*Burkholderia multivorans* CGD1	ZP_0358 7598.1	BMULJ_0 3657	YP_00194 8065.1
17	14708	15138	tail protein (E+E')	-	AGAGGAAcca tacgATG	143	phage tail protein E	1-87/103	97	*Burkholderia multivorans* CGD1	ZP_0358 7598.1	BMULJ_0 3656 BMULJ_0 3657	YP_00194 8064.1 YP_00194 8065.1
18	15171	15680	tail tube protein (FII)	-	AGGGAAAcgc aATG	169	phage major tail tube protein	1-169/169	94	*Burkholderia multivorans* CGD1	ZP_0358 7599.1	BMULJ_0 3658	YP_00194 8066.1
19	15710	16882	tail sheath protein (FI)	-	GGGAGAttgcA TG	390	tail sheath protein	1-390/390	94	*Burkholderia cenocepacia* MC0-3	YP_0017 63487.1	BMULJ_0 3659	YP_00194 8067.1
20	16993	17742	N-4/N-6 DNA methylase	-	GAGGGAAtcg ccccATG	249	DNA methylase N-4/N-6 domain protein	1-249/249	89	*Burkholderia ambifaria* MEX-5	ZP_0290 5740.1	BMULJ_0 3660	YP_00194 8068.1
21	17720	17902	Com translational	-	AAGCAGGAA tcaccgATG	60	hypothetical protein	1-60/60	85	*Burkholderia cenocepacia* MC0-3	YP_0017 63489.1	BMULJ_0 3661	YP_00194 8069.1

			regulator				Bcenmc03_0187						
22	18049	18927	tail fiber assembly protein	-	GAGACACAcctATG	292	gp31, bacteriophage-acquired protein	1-272/278	89	*Burkholderia multivorans* CGD1	ZP_03587603.1	BMULJ_03662	YP_001948070.1
23	18937	20547	tail fiber protein	-	GGATAcctgaaacATG	536	bacteriophage protein	1-536/536	99	*Burkholderia multivorans* CGD1	ZP_03587604.1	BMULJ_03663	YP_001948071.1
24	20550	21104	baseplate assembly protein (I)	-	GGGGTGGccgATG	184	phage tail protein I	1-184/184	92	*Burkholderia multivorans* CGD1	ZP_03587605.1	BMULJ_03664	YP_001948072.1
25	21097	22002	baseplate assembly protein (J)	-	GAGGCAcggcATG	301	baseplate assembly protein J (GpJ)	1-301/301	94	*Burkholderia multivorans* CGD1	ZP_03587606.1	BMULJ_03665	YP_001948073.1
26	21999	22376	baseplate assembly protein (W)	-	GAAGGGGcacggATG	125	baseplate assembly protein W (GpW)	1-125/125	89	*Burkholderia multivorans* CGD1	ZP_03587607.1	BMULJ_03666	YP_001948074.1
27	22373	23005	baseplate assembly protein (V)	-	GCGGCAtccttgccgcATG	210	phage baseplate assembly protein V	1-137/234	78	*Burkholderia cenocepacia* MC0-3	YP_001763496.1	BMULJ_03667	YP_001948075.1
28	23206	25086	exonuclease (Old)	-	AAGTGGGGAccaactATG	626	ATP-dependent endonuclease	1-625/626	72	*Cupriavidus metallidurans* CH34	YP_586772.1	BMULJ_03668	YP_001948076.1
29	25269	25718	tail completion protein (S)	-	GGGGAcgtgATG	149	phage virion morphogenesis protein	1-148/149	89	*Burkholderia multivorans* CGD1	ZP_03587610.1	BMULJ_03669	YP_001948077.1
30	25718	26128	tail completion protein (R)	-	AGGAGGcgccGTG	136	P2 phage tail completion protein R (GpR)	1-136/136	96	*Burkholderia multivorans* CGD1	ZP_03587611.1	BMULJ_03670	YP_001948078.1
31	26172	26366	Rz1	-	AAGGAGGttccggtttATG	64	Ribonuclease, Rnc/Rng family	15-48/928	47	*Propionibacterium freudenreichii* subsp. *shermanii* CIRM-BIA1	YP_003687809.1	none	
32	26125	26616	Rz	-	GGGTGGccgcATG	163	conserved hypothetical protein	1-163/163	85	*Burkholderia ambifaria* MEX-5	ZP_02905751.1	BMULJ_03671	YP_001948079.1
33	26613	27413	endolysin	-	GGGGGcgccATG	266	peptidoglycan binding domain-containing protein	1-266/266	90	*Burkholderia cenocepacia* MC0-3	YP_001763501.1	BMULJ_03672	YP_001948080.1
34	27406	27726	holin	-	AAGGGGAGGGAcaagtgATG	106	protein of unknown	1-106/106	88	*Burkholderia ambifaria* MEX-5	ZP_02905753.1	BMULJ_03673	YP_001948081.1

							function DUF754						
35	27726	28100	putative antiholin	-	ATGGGActgag aATG	124	phage-related transmembrane protein	1-124/124	96	*Burkholderia multivorans* CGD1	ZP_0358 7615.1	BMULJ_0 3674	YP_00194 8082.1
36	28103	28315	tail protein (X)	-	AGGGAGctgtc ctgATG	70	tail X family protein	1-70/70	94	*Burkholderia cenocepacia* MC0-3	YP_0017 63504.1	BMULJ_0 3675	YP_00194 8083.1
37	28315	28557	unknown	-	GTGGAGctcatc tgATG	80	conserved hypothetical protein	1-80/80	72	*Burkholderia multivorans* CGD1	ZP_0358 7617.1	BMULJ_0 3676	YP_00194 8084.1
38	28557	29033	capsid completion protein (L)	-	AACGTGACG AAcccgaccAT G	158	head completion protein	1-160/160	85	*Burkholderia ambifaria* MEX-5	ZP_0290 5755.1	BMULJ_0 3677	YP_00194 8085.1
39	29138	29824	terminase small subunit (M)	-	GGGTGGcgcA TG	228	terminase	1-228/228	93	*Burkholderia multivorans* CGD1	ZP_0358 7619.1	BMULJ_0 3678	YP_00194 8086.1
40	29821	30846	capsid protein (N)	-	AAACGGAGA AtccATG	341	phage major capsid protein, P2 family	1-339/339	77	*Burkholderia ambifaria* MEX-5	ZP_0290 5757.1	BMULJ_0 3679	YP_00194 8087.1
41	30884	31705	capsid scaffolding protein (O)	-	AGAGGtttcgca cATG	273	phage capsid scaffolding protein GpO	1-273/273	95	*Burkholderia multivorans* CGD1	ZP_0358 7621.1	BMULJ_0 3680	YP_00194 8088.1
42	31855	33621	terminase large subunit (P)	+	GGTAGccttgct gcATG	588	putative ATPase subunit of terminase (gpP-like)	1-583/583	92	*Burkholderia multivorans* CGD1	ZP_0358 7622.1	BMULJ_0 3681	YP_00194 8089.1
43	33621	34673	portal vertex protein (Q)	+	ATGGAGAtttttc tgATG	350	phage portal protein, pbsx family	1-348/349	92	*Burkholderia multivorans* CGD1	ZP_0358 7623.1	BMULJ_0 3682	YP_00194 8090.1
44	35144	36163	reverse transcriptase	-	GAATGGAtttc cgaaaATG	339	putative reverse transcriptase	2-285/292	42	*Sideroxydans lithotrophicus* ES-1	YP_0035 22714.1	BMULJ_0 3683	YP_00194 8091.1
45	36120	36443	transcriptional regulator	-	GAAGGAGttgc atATG	107	transcriptional regulator	1-97/97	52	*Acinetobacter baumannii* ACICU	YP_0018 40883.1	BMULJ_0 3684	YP_00194 8092.1

Abbreviations: RBS, ribosome binding site; aa, amino acid; % ID, percent identity. The P2 proteins that are similar to KS5 proteins based on CoreGenes analysis are shown in brackets in the putative function column. Excluding genes *17* and *31*, annotations were based on those of the *B. multivorans* ATCC 17616 chromosome 2 sequence (GenBank: NC_010805.1; BMULJ_03640-BMULJ_03684, bp 477,496-514,731).

Table 3-2: KS14 genome annotation

Gene	Start	End	Putative function	Strand	Predicted RBS and start codon	Length (aa)	Closest relative	Alignment region in closest relative (aa)	% ID	Source	GenBank accession number
1	1	261	unknown	-	GAGGCGAggc ATG	86	hypothetical protein BB1680	18-96/101	35	*Bordetella bronchiseptica* RB50	NP_8882 25.1
2	270	3041	zinc finger CHC2-family protein	-	GCGATTCTGA aaaATG	923	hypothetical protein RPRSA1_gp47	1-933/934	65	*Ralstonia* phage ϕRSA1	YP_0011 65296.1
3	3122	3511	unknown	-	GAGGGAccgaa ccATG	129	hypothetical protein Csal_1360	9-103/130	33	*Chromohalobacter salexigens* DSM 3043	YP_5734 14.1
4	3857	4036	unknown	-	GAAAAcaccAT G	59	hypothetical protein PC1_2629	21-62/63	45	*Pectobacterium carotovorum* subsp. carotovorum PC1	YP_0030 18195.1
5	4142	4849	repressor	+	AAGGcccaatAT G	235	hypothetical protein GCWU000324_ 01220	13-220/226	33	*Kingella oralis* ATCC 51147	ZP_0460 1747.1
6	4809	5486	serine recombinase	-	GAAGGCGAtac aagaaaATG	225	resolvase domain-containing protein	4-194/195	57	*Shewanella* sp. W3-18-1	YP_9654 29.1
7	5479	5790	unknown	-	GAGGCGGGcg agctATG	103	hypothetical protein BuboB_03089	9-106/114	45	*Burkholderia ubonensis* Bu	ZP_0237 6681.1
8	5787	6368	unknown	-	GACTACAGGc gaccATG	193	hypothetical protein pRALTA_0144	25-180/255	48	*Cupriavidus taiwanensis*	YP_0017 96036.1
9	6506	7054	unknown	+	AGGAAAGAA	182	hypothetical	20-	31	*Acidovorax* sp.	YP_9875

					AAcggtcgtTTG		protein Ajs_3318	138/138		JS42	16.1
10	7093	8130	tail protein (D)	-	AAAAAagaAT G	345	fels-2 prophage protein	25-366/366	65	*Burkholderia oklahomensis* EO147	ZP_0235 3972.1
11	8127	8558	tail protein (U)	-	GGAGGAAAG AAAcgaTGA	143	bacteriophage tail-related protein	1-133/141	64	*Burkholderia oklahomensis* EO147	ZP_0235 3973.1
12	8574	11102	tail tape measure protein (T)	-	GTATGGAAGc gaATG	842	phage tail tape measure protein, TP901 family	5-918/924	39	*Pantoea* sp. At-9b	ZP_0573 0476.1
13	11251	11535	tail protein (E)	-	AGAGAAAgaa ATG	94	hypothetical protein BPSL0148	12-103/114	72	*Burkholderia pseudomallei* K96243	YP_1067 76.1
14	11099	11535	tail protein (E+E')	-	AGAGAAAgaa ATG	145	gpE+E'	3-142/142	49	Enterobacteria phage P2	NP_0467 80.1
15	11623	12138	tail tube protein (FII)	-	AGGGAGtaaAT G	171	phage major tail tube protein	1-169/169	66	*Burkholderia* sp. CCGE1001	ZP_0629 2830.1
16	12150	13325	tail sheath protein (FI)	-	AAACAGGAAtt cagATG	391	putative phage major tail sheath protein	1-390/390	72	*Burkholderia cenocepacia* J2315	YP_0022 29261.1
17	13387	14007	tail fiber assembly protein (G)	-	GAAGGGAAcc gaccATG	206	gp31, bacteriophage-acquired protein	1-189, 196-266/278	66, 54	*Burkholderia multivorans* CGD1	ZP_0358 7603.1
18	14026	15534	tail fiber protein	-	ACGGATAtctga cctATG	502	hypothetical protein BuboB_27067	3-534/534	55	*Burkholderia ubonensis* Bu	ZP_0238 1413.1
19	15538	16080	baseplate assembly protein (I)	-	GGGGGCAtttac gaaaATG	180	phage-related tail protein	1-179/180	73	*Burkholderia ubonensis* Bu	ZP_0238 1412.1
20	16073	16987	baseplate assembly protein (J)	-	AACGGGGGT GcggcATG	304	phage baseplate assembly protein	1-303/304	74	*Burkholderia ubonensis* Bu	ZP_0238 1411.1

21	16984	17346	baseplate assembly protein (W)	-	GTGAGCGCAc cgcaATG	120	phage baseplate assembly protein	1-117/117	66	*Burkholderia thailandensis* MSMB43	ZP_0246 6378.1
22	17343	18002	baseplate assembly protein (V)	-	AACATGGAgg cATG	219	bacteriophage baseplate assembly protein V	1-227/227	57	*Burkholderia ubonensis* Bu	ZP_0238 1409.1
23	18087	18545	tail completion protein (S)	-	GATCCGGCGG cgcaATG	152	phage virion morphogenesis protein	1-153/155	66	*Burkholderia* sp. CCGE1001	ZP_0629 2822.1
24	18533	18985	tail completion protein (R)	-	ACCGcccccgacc ATG	150	P2 phage tail completion R family protein	1-136/140	54	*Burkholderia* sp. CCGE1001	ZP_0629 2821.1
25	18978	19256	Rz1 (LysC)	-	GAGGCGttgaaa cATG	92	hypothetical phage protein	1-83/91	59	*Burkholderia pseudomallei* 1655	ZP_0489 0536.1
26	19102	19554	Rz (LysB)	-	GAGAAGGcggc cgcATG	150	putative phage-encoded lipoprotein	19-142/142	44	*Burkholderia glumae* BGR1	YP_0029 10045.1
27	19551	20357	endolysin	-	GCGGAGtgaAT G	268	putative phage-encoded peptidoglycan binding protein	5-268/268	66	*Burkholderia ubonensis* Bu	ZP_0237 6668.1
28	20354	20620	holin	-	GAAAGGGctga cccATG	88	protein of unknown function DUF754	1-88/88	61	*Burkholderia* sp. CCGE1001	ZP_0629 2819.1
29	20624	20965	putative antiholin	-	GGAGtcgccaac ATG	113	hypothetical protein BuboB_26997	1-113/114	74	*Burkholderia ubonensis* Bu	ZP_0238 1399.1
30	20980	21186	tail protein (X)	-	GATCGAGctgat ctgATG	68	putative phage tail protein	1-67/67	61	*Erwinia tasmaniensis* Et1/99	YP_0019 06519.1
31	21186	21677	capsid completion	-	AGAGctgaaacc ATG	163	fels-2 prophage protein	31-172/172	55	*Burkholderia thailandensis* E264	YP_4395 44.1

			protein (L)								
32	21779	22465	terminase small subunit (M)	-	AACGGAGGcat gacgcgATG	228	bacteriophage terminase, endonuclease subunit	3-220/229	59	*Burkholderia oklahomensis* EO147	ZP_0236 0025.1
33	22481	23497	capsid protein (N)	-	GGAGAAcacacc acATG	338	bacteriophage protein	1-338/338	68	*Ralstonia solanacearum* GMI1000	NP_5200 58.1
34	23540	24577	capsid scaffolding protein (O)	-	GGAGAcctaaca ATG	345	capsid scaffolding	4-349/349	50	*Burkholderia* sp. CCGE1001	ZP_0629 2813.1
35	24693	26507	terminase large subunit (P)	+	GGGTACAcatag gcgggcGTG	604	protein of unknown function DUF264	13-586/588	75	*Burkholderia* sp. CCGE1001	ZP_0629 2812.1
36	26507	27559	portal vertex protein (Q)	+	ATGGAGttctctta ATG	350	putative phage portal vertex protein	1-347/351	70	*Burkholderia pseudomallei* 7894	ZP_0248 7524.1
37	27740	28495	replication initiation	-	AGGGGAAGcgt cccaATG	251	initiator RepB protein	16-251/251	73	*Ralstonia pickettii* 12J	YP_0019 01323.1
38	28834	29010	unknown	+	GTGAGGGGcaa caaGTG	58	N/A				
39	29150	29695	unknown	+	GTGATGCACG AccgcccgaATG	181	flagellar hook-associated protein FlgK	257-334/672	29	*Acidovorax ebreus* TPSY	YP_0025 54543.1
40	30436	31086	DNA partitioning	+	GGAGCATGcga aATG	216	ParA family protein, putative	1-211/217	69	*Burkholderia thailandensis* E264	YP_4395 56.1
41	31127	31438	unknown	+	AGCGAGGtaata gcaaaATG	103	hypothetical protein BuboB_03094	1-73/88	49	*Burkholderia ubonensis* Bu	ZP_0237 6682.1
42	31435	31581	unknown	+	AAAGAGGGgg cATG	48	N/A				
43	31603	31821	unknown	+	GAAAAGGGG	72	hypothetical	10-63/63	57	*Serratia*	NP_9411

					AAttgaATG		protein SMR0083			*marcescens*	57.1
44	31909	32217	unknown	-	GGAGTGAtgttt ATG	102	hypothetical protein BuboB_03104	1-69/84	73	*Burkholderia ubonensis* Bu	ZP_0237 6684.1

Abbreviations: RBS, ribosome binding site; aa, amino acid; % ID, percent identity; N/A, not applicable. The P2 proteins that are similar to KS14 proteins based on CoreGenes analysis are shown in brackets in the putative function column.

Table 3-3: KL3 genome annotation

Gene	Start	End	Putative function	Strand	Predicted RBS and start codon	Length (aa)	Closest relative	Alignment region in closest relative (aa)	% ID	Source	GenBank accession number
1	122	1150	integrase	-	GGCGCAGtgtg ATG	342	integrase family protein	1-342/342	98	*Burkholderia ambifaria* MEX-5	ZP_02905 720.1
2	1150	1416	unknown	-	GAAAAtcaccA TG	88	hypothetical protein BamMEX5DRAF T_1075	1-88/88	97	*Burkholderia ambifaria* MEX-5	ZP_02905 721.1
3	2096	4891	zinc finger CHC2-family protein	-	AACAGCAAta acgATG	931	zinc finger CHC2-family protein	1-931/931	95	*Burkholderia ambifaria* MEX-5	ZP_02905 722.1
4	4894	5142	unknown	-	GGAGGcgcagc agcATG	82	conserved hypothetical protein	1-82/84	91	*Burkholderia ambifaria* MEX-5	ZP_02905 723.1
5	5139	5501	unknown	-	GCGGGGctgac acgATG	120	conserved hypothetical protein	1-120/120	97	*Burkholderia ambifaria* MEX-5	ZP_02905 724.1
6	5506	5700	membrane protein	-	GGAAccacaccA TG	64	putative phage-encoded membrane protein	1-64/64	93	*Burkholderia ambifaria* MEX-5	ZP_02905 725.1
7	5744	5938	unknown	-	GCACTGAtccg ATG	64	hypothetical protein bglu_1g03740	1-64/64	95	*Burkholderia glumae* BGR1	YP_00291 0278.1
8	5943	6155	unknown	-	GAAAAAAGG AGAtcagcATG	70	conserved hypothetical protein	1-70/70	72	*Burkholderia* sp. CCGE1001	ZP_06292 843.1
9	6243	6491	transcriptional activator (Ogr)	-	GGAGTAAGcc gaaATG	82	putative phage transcriptional	1-82/82	92	*Burkholderia glumae* BGR1	YP_00291 0024.1

							activator Ogr/Delta				
10	6479	6682	unknown	-	AATGAGTAG ctcctacgATG	67	hypothetical protein BoklE_00724	1-67/67	85	*Burkholderia oklahomensis* EO147	ZP_02353 966.1
11	6722	6919	unknown	-	GAGGAGcccgc ATG	65	hypothetical protein BoklE_00729	1-65/65	87	*Burkholderia oklahomensis* EO147	ZP_02353 967.1
12	6934	7143	unknown	-	AAAGTATAcc gaccATG	69	hypothetical protein BoklE_00734	1-62/71	87	*Burkholderia oklahomensis* EO147	ZP_02353 968.1
13	7193	7732	repressor	+	GGTAAGGctag tgtaATG	179	hypothetical protein BCAL0086	1-163/163	62	*Burkholderia cenocepacia* J2315	YP_00222 9252.1
14	7904	8179	unknown	+	GAGGGAccaga agaATG	91	hypothetical protein BuboB_27112	1-91/98	50	*Burkholderia ubonensis* Bu	ZP_02381 422.1
15	8238	9329	tail protein (D)	-	GGACGCGGA GccgaaggcATG	363	fels-2 prophage protein	19-366/366	81	*Burkholderia oklahomensis* EO147	ZP_02353 972.1
16	9326	9784	tail protein (U)	-	ACGGAGGAtc tgtcccATG	152	bacteriophage tail-related protein	1-133/141	66	*Burkholderia oklahomensis* EO147	ZP_02353 973.1
17	9806	13033	tail tape measure protein (T)	-	GAAGCGGAca cgagtaacgATG	1075	hypothetical protein bglu_1g01240	1-1079/1079	59	*Burkholderia glumae* BGR1	YP_00291 0030.1
18	13158	13508	tail protein (E)	-	AGGACACGca acatATG	116	gpE+E'	1-114/114	76	*Burkholderia pseudomallei* 112	ZP_02501 899.1
19	13036	13508	tail protein (E+E')	-	AGGACACGca acatATG	157	gpE+E'	1-100/114	74	*Burkholderia pseudomallei* 112	ZP_02501 899.1
20	13578	14087	tail tube	-	AGGAGtcacaca	169	phage major tail	1-169/169	74	*Burkholderia*	YP_00222

			protein (FII)		cATG		tube protein			*cenocepacia* J2315	9260.1
21	14103	15275	tail sheath protein (FI)	-	AGGAGctgcaca ccATG	390	phage tail sheath protein	1-390/390	84	*Burkholderia pseudomallei* 1655	ZP_04890 547.1
22	15328	15951	tail fiber assembly protein	-	ACGGAcctcgaa acATG	207	tail fiber assembly protein from lambdoid prophage e14	1-190/209	85	*Burkholderia ubonensis* Bu	ZP_02381 414.1
23	15969	18632	tail fiber protein	-	GGATAcctgaac ATG	887	putative phage tail protein	1-883/883	71	*Burkholderia cenocepacia* J2315	YP_00222 9263.1
24	18635	19189	baseplate assembly protein (I)	-	GATGGCGGG GtcgcggATG	184	phage-related tail protein	1-183/184	84	*Burkholderia pseudomallei* 7894	ZP_02487 653.1
25	19182	20087	baseplate assembly protein (J)	-	GAACGGAGtc ggcgcATG	301	baseplate J-like protein	1-301/301	90	*Burkholderia thailandensis* E264	YP_43953 1.1
26	20084	20446	baseplate assembly protein (W)	-	GGAGCGGtgc ATG	120	phage baseplate assembly protein	1-117/120	78	*Burkholderia pseudomallei* 7894	ZP_02487 655.1
27	20443	21138	baseplate assembly protein (V)	-	GAGGGCGGcc ggcaacATG	231	phage baseplate assembly protein	33-261/261	72	*Burkholderia* phage ϕ52237	YP_29373 5.1
28	21304	22080	N-4/N-6 DNA methylase	+	ACGTTGcctcag aaccATG	258	site-specific DNA methyltransferase	34-290/291	78	*Burkholderia pseudomallei* K96243	YP_11108 9.1
29	22060	22527	tail completion protein (S)	-	GAGCAATGG GtggcgtgATG	155	phage virion morphogenesis protein	1-155/155	87	*Burkholderia thailandensis* MSMB43	ZP_02466 375.1
30	22527	22943	tail completion protein (R)	-	AGACGGccgcc cATG	138	bacteriophage tail completion protein R	1-138/138	73	*Burkholderia pseudomallei* K96243	YP_11108 6.1

31	22936	23220	Rz1 (LysC)	-	GGAGActcatcg ATG	94	hypothetical phage protein	1-83/91	75	*Burkholderia pseudomallei* 1655	ZP_04890 536.1
32	23057	23497	Rz (LysB)	-	GAAGGcggccg cGTG	146	protein lysB	1-146/146	62	*Burkholderia thailandensis* E264	YP_43953 8.1
33	23494	24303	endolysin	-	GGAGCAccgaa tcATG	269	putative phage-encoded peptidoglycan binding protein	1-269/270	73	*Burkholderia pseudomallei* K96243	YP_10679 1.1
34	24300	24572	holin	-	AGGGGGAAA tgacATG	90	protein of unknown function DUF754	1-88/88	67	*Burkholderia* sp. CCGE1001	ZP_06292 819.1
35	24574	24918	putative antiholin	-	GGAAttgtccga ATG	114	hypothetical protein Bpse38_23639	1-113/114	83	*Burkholderia thailandensis* MSMB43	ZP_02466 369.1
36	24934	25140	tail protein (X)	-	GGTTGAActga tctgATG	68	phage tail protein X	1-68/68	88	*Burkholderia pseudomallei* 7894	ZP_02487 665.1
37	25140	25619	capsid completion protein (L)	-	GAATCGaccA TG	159	fels-2 prophage protein	1-159/159	86	*Burkholderia thailandensis* E264	ZP_05590 935.1
38	25719	26408	terminase small subunit (M)	-	GAGCTGGtggc ggcATG	229	hypothetical protein bglu_1g01450	1-228/229	95	*Burkholderia glumae* BGR1	YP_00291 0051.1
39	26405	27418	capsid protein (N)	-	GGAGAAcccaa ctcATG	337	Gp2, phage major capsid protein, P2 family protein	1-337/337	98	*Burkholderia glumae* BGR1	YP_00291 0052.1
40	27454	28266	capsid scaffolding protein (O)	-	GGTTCGAcctct ctctATG	270	phage capsid scaffolding protein (GPO)	1-270/270	95	*Burkholderia glumae* BGR1	YP_00291 0053.1
41	28345	30180	terminase large subunit	+	AACGAGcggcg tATG	611	phage terminase, ATPase subunit	1-589/589	99	*Burkholderia glumae* BGR1	YP_00291 0054.1

			(P)								
42	30177	31226	portal vertex protein (Q)	+	GGAGttctattcA TG	349	Gp5, phage portal protein, pbsx family protein	1-347/351	99	*Burkholderia glumae* BGR1	YP_00291 0055.1
43	31595	31732	unknown	+	GATGcgcgAT G	45	N/A				
44	31722	33290	unknown	-	GGGGAAAGca acatATG	522	hypothetical protein ECO103_1901	1-526/527	54	*E. coli* O103:H2 str. 12009	YP_00322 1840.1
45	33455	34699	EcoRII-C endonuclease	-	AACGGAGcttc ggggATG	414	type II restriction endonuclease, EcoRII-C domain protein	1-401/401	77	*Candidatus* Hamiltonella defensa 5AT (*Acyrthosiphon pisum*)	
46	34696	35142	Vsr endonuclease	-	TCGCctgATG	148	DNA mismatch endonuclease Vsr	1-148/148	77	*Burkholderia graminis* C4D1M	ZP_02883 050.1
47	35142	36410	EcoRII DNA cytosine methylase	-	AGCGAGAGca aatATG	422	DNA-cytosine methyltransferase	1-426/426	81	*Burkholderia phytofirmans* PsJN	YP_00189 4783.1
48	36570	37685	unknown	-	AAGCTGAcgct ATG	372	conserved hypothetical protein	1-371/371	95	*Burkholderia ambifaria* MEX-5	ZP_02905 764.1
49	37728	38816	unknown	-	AGTTctctaattga cATG	362	GP30 family protein	1-362/362	96	*Burkholderia ambifaria* MEX-5	ZP_02905 765.1
50	38822	39526	unknown	-	GGGAGAAGcc tgaATG	234	VRR-NUC domain protein	1-234/234	99	*Burkholderia ambifaria* MEX-5	ZP_02905 766.1
51	39523	40032	unknown	-	AGGAGttcagc ATG	169	PAAR repeat-containing protein	1-169/169	97	*Burkholderia ambifaria* MEX-5	ZP_02905 767.1
52	40202	40408	transcriptional	+	AAGGAGAAA	68	phage	1-68/72	94	*Burkholderia*	ZP_02905

			regulator		tagcATG		transcriptional regulator, AlpA			*ambifaria* MEX-5	768.1

Abbreviations: RBS, ribosome binding site; aa, amino acid; % ID, percent identity; N/A, not applicable. The P2 proteins that are similar to KL3 proteins based on CoreGenes analysis are shown in brackets in the putative function column.

Genome characterization.

KS5: The KS5 genome is 37,236 base pairs (bp) in length and encodes 46 proteins (including the transposase of a predicted insertion sequence, discussed below) (Table 3-1). This genome has a 63.7% GC content. Forty-three of the start codons are ATG, 2 are GTG and 1 is TTG (Table 3-1). As KS5 was isolated from an environmental sample, it was predicted that this phage might be obligately lytic (265). However, KS5 encodes an integrase and a repressor and is found as a prophage in chromosome 2 of the fully sequenced *B. multivorans* strain ATCC 17616 (GenBank: NC_010805.1; BMULJ_03640-BMULJ_03684, bp 477,496-514,731) (Table 3-1). Because of this similarity, the possibility exists that KS5 originated from ATCC 17616 or a closely related strain found in the soil enrichment. Excluding the ATCC 17616 prophage, KS5 is most similar to a putative prophage element in *Burkholderia multivorans* CGD1. Twenty-three of 46 KS5 proteins are most closely related to a protein from CGD1, with percent identities ranging from 72-99% (Table 3-1).

KS14: The KS14 genome is 32,317 bp in length and encodes 44 proteins (Table 3-2). This genome has a 62.3% GC content. Forty-one of the start codons are ATG, 2 are GTG and 1 is TTG (Table 3-2). All predicted KS14 proteins show similarity to at least one protein in the database (as determined by a BLASTP search) except for gp38 and gp42. The protein with the most similarity to others in the database is the terminase large subunit, gp35, which has 75% identity with a protein of unknown function DUF264 of *Burkholderia* sp. CCGE1001. Aside from gp38 and gp42, the least similar protein is the hypothetical protein gp39,

which has 29% identity with the flagellar hook-associated protein FlgK of *Acidovorax ebreus* TPSY (Table 3-2).

KL3: The KL3 genome is 40,555 bp in length and encodes 52 proteins (Table 3-3). This genome has a 63.2% GC content. Fifty-one of the start codons are ATG and one is GTG (Table 3-3). Similarly to KS14, all predicted KL3 proteins show similarity to at least one protein in the database except for gp43. The proteins with the most similarity to others in the database are the terminase large subunit (gp41) and the portal protein (gp42) that have 99% identity with *Burkholderia glumae* BGR1 proteins and the hypothetical protein gp50 which has 99% identity with a *B. ambifaria* MEX-5 protein. Aside from gp43, the least similar protein is the hypothetical protein gp14, which has 50% identity with the hypothetical protein BuboB_27112 of *Burkholderia ubonensis* Bu (Table 3-3).

Modular organization. The genome maps of KS5, KS14, and KL3 are shown in Figure 3-2. Each of these phages has a modular organization, with genes for tail formation (shown in purple), lysis (shown in light blue), and head formation (shown in dark blue) clustered in each phage (Figure 3-2). In KS5, genes *13-19*, *22-27*, *29-30*, and *36* encode tail proteins, genes *31-35* encode lysis proteins, and genes *38-43* encode capsid proteins (Table 3-1, Figure 3-2). In KS14, genes *10-24* and *30* encode tail proteins, genes *25-29* encode lysis proteins, and genes *31-36* encode capsid proteins (Table 3-2, Figure 3-2). In KL3, genes *15-27*, *29-30*, and *36* encode tail proteins, genes *31-35* encode lysis proteins, and genes *37-42* encode capsid proteins (Table 3-3, Figure 3-2). In addition, KL3 has a three gene DNA modification module: gene *45*, encoding an EcoRII-C

restriction endonuclease, gene *46*, encoding a Vsr endonuclease, and gene *47*, encoding an EcoRII methylase (Table 3-3, Figure 3-2).

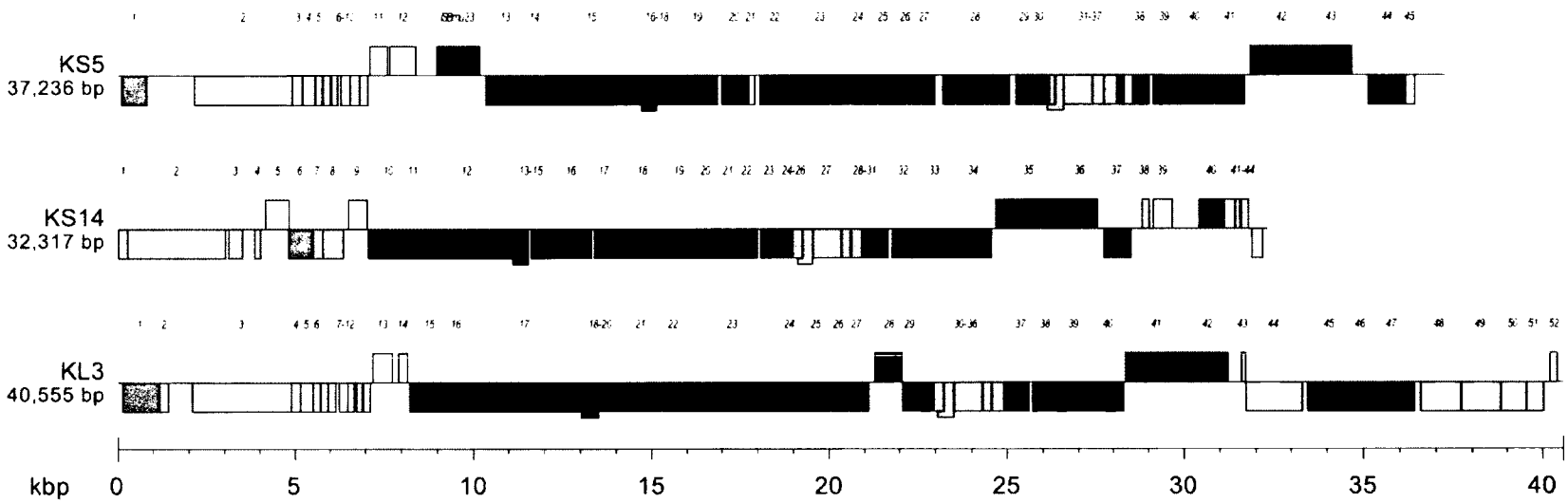


Figure 3-2: Genome maps of KS5, KS14, and KL3. Genes transcribed in the forward direction are shown above and genes transcribed in the reverse direction are shown below. The scale (in kbp) is shown on the bottom. The prophage gene order is shown for KS5 and KL3. The gene order for KS14 (which is maintained as a plasmid prophage) was chosen based on alignment with the other two sequences. Legend: orange, recombinase; yellow, transcriptional or translational regulation; black, insertion sequence; purple, tail morphogenesis; red, DNA modification; light blue, lysis; dark blue, capsid morphogenesis and DNA packaging; green, reverse transcription; brown, replication and partitioning; gray, unknown function.

Similarity to P2. KS5, KS14, and KL3 all show similarity to enterobacteria phage P2 (GenBank: NC_001895.1). P2 is a temperate myovirus that was isolated from *E. coli* strain Li by Bertani in 1951 (21). P2 has recently been classified as part of a novel subfamily, placing it in the order *Caudovirales*, family *Myoviridae*, subfamily *Peduovirinae*, and genus "P2-like viruses" (154). This genus includes phages P2, Wϕ, 186, and PsP3 of enterobacteria, L-413C of *Yersinia*, Fels-2 and SopEϕ of *Salmonella*, ϕ-MhaA1-PHL101 of *Mannheimia*, ϕCTX of *Pseudomonas*, RSA1 of *Ralstonia*, ϕE202 of *Burkholderia thailandensis*, and ϕ52237 and ϕE12-2 of *Burkholderia pseudomallei* (154).

A four-way comparison of the P2, KS5, KS14, and KL3 genomes prepared using PROmer/Circos is shown in Figure 3-3. In this comparison, regions of similarity on the same strand are shown in green, while regions of similarity on the opposite strand are shown in red. The majority of similar regions among these phages are on the same strand, except for a short conserved region in KS5 and KL3 containing DNA methylase genes (KS5 *20* and KL3 *28*) on the minus strand in KS5 and on the plus strand in KL3 (Tables 3-1 and 3-3). KS5, KS14, and KL3 all encode proteins similar to P2 D, U, T, E, E+E', FII, FI, I, J, W, V, S, R, and X (involved in tail formation) and L, M, N, O, P, and Q (involved in capsid formation) (Table 3-4). In addition, KS5 gp8 and KL3 gp9 are similar to Ogr (transcriptional activator), KS5 gp28 is similar to Old (phage immunity protein), KS14 gp17 is similar to G (tail fiber assembly protein), and KS14 gp26/gp25 and KL3 gp32/gp31 are similar to LysBC (Rz/Rz1-like lysis proteins) (Table 3-4). The percent identity of the similar proteins ranges from 25-64% in

KS5, 24-64% in KS14, and 31-62% in KL3 (Table 3-4).

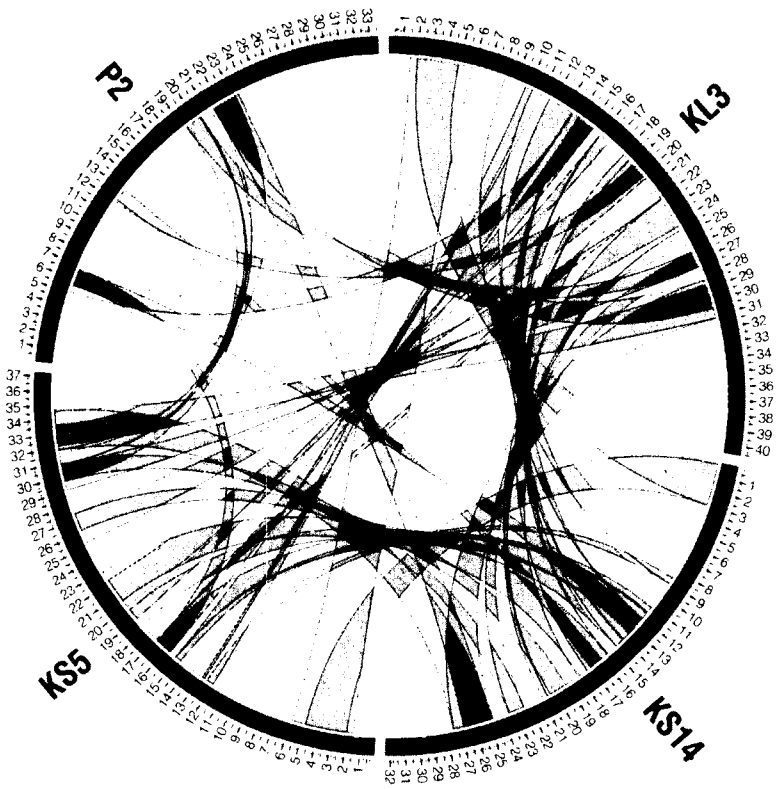


Figure 3-3: PROmer/Circos comparison of the KS5, KS14, KL3, and P2 prophages. Regions of similarity on the same strand are shown in green and regions of similarity on the opposite strand are shown in red. The scale (in kbp) is shown on the outside. The sequence start site for the KS14 prophage (which is maintained as a plasmid) was chosen based on alignment with the other three sequences. PROmer parameters: breaklen = 60, maxgap = 30, mincluster = 20, minmatch = 6.

Table 3-4: CoreGenes comparison of P2, KS5, KS14, and KL3

P2 protein	P2 function	Similar KS5 protein (% ID)	Similar KS14 protein (% ID)	Similar KL3 protein (% ID)
Old	phage exclusion	gp28 (25%)		
Tin	phage exclusion			
Orf91	unknown			
A	DNA replication			
Orf83	unknown			
Orf82	unknown			
Orf81	unknown			
Orf80	unknown			
B	DNA replication			
Orf78	unknown			
Cox	transcriptional repressor; excision			
C	repressor			
Int	integrase			
Ogr	transcriptional activator	gp8 (39%)		gp9 (34%)
D	tail protein	gp13 (50%)	gp10 (38%)	gp15 (41%)
U	tail protein	gp14 (48%)	gp11 (45%)	gp16 (40%)
T	tape measure protein	gp15 (28%, 42%)	gp12 (25%)	gp17 (31%, 33%)
E	tail protein	gp16 (59%)	gp13 (55%)	gp18 (55%)
E+E'	tail protein	gp17 (50%)	gp14 (49%)	gp19 (51%)
FII	tail tube protein	gp18 (48%)	gp15 (48%)	gp20 (48%)
FI	tail sheath protein	gp19 (64%)	gp16 (64%)	gp21 (58%)
Z/Fun	phage exclusion			

G	tail fiber assembly		gp17 (24%)	
H	tail fiber protein			
I	baseplate assembly protein	gp24 (37%)	gp19 (36%)	gp24 (39%)
J	baseplate assembly protein	gp25 (48%)	gp20 (49%)	gp25 (44%)
W	baseplate assembly protein	gp26 (43%)	gp21 (39%)	gp26 (36%)
V	baseplate assembly protein	gp27 (38%)	gp22 (35%)	gp27 (31%)
Orf30	unknown			
S	tail completion protein	gp29 (44%)	gp23 (35%)	gp29 (35%)
R	tail completion protein	gp30 (43%)	gp24 (43%)	gp30 (39%)
LysC	Rz1-like		gp25 (36%)	gp31 (48%)
LysB	Rz-like		gp26 (33%)	gp32 (42%)
LysA	antiholin			
K	endolysin			
Y	holin			
X	tail protein	gp36 (51%)	gp30 (55%)	gp36 (62%)
L	capsid completion protein	gp38 (45%)	gp31 (39%)	gp37 (43%)
M	terminase small/endonuclease subunit	gp39 (49%)	gp32 (47%)	gp38 (46%)
N	major capsid protein	gp40 (51%)	gp33 (54%)	gp39 (55%)
O	capsid scaffolding protein	gp41 (46%)	gp34 (44%)	gp40 (40%)
P	terminase large/ATPase subunit	gp42 (59%)	gp35 (60%)	gp41 (57%)
Q	portal protein	gp43 (57%)	gp36 (54%)	gp42 (55%)

Abbreviations: % ID, percent identity.

The genes in common between P2 and the P2-like Bcc phages are almost exclusively limited to structural genes involved in virion formation (Table 3-4). Other P2 genes, such as those involved in DNA replication, phage immunity, lysogeny, and lysis are dissimilar among these phages. A similar pattern is observed (with some exceptions) following CoreGenes analysis of the P2-like phages ϕE202 of *B. thailandensis* and ϕ52237 and ϕE12-2 of *B. pseudomallei* (154). A likely explanation for this pattern is that, while phage structural components predominantly interact with each other, components from other phage systems may interact with host-specific proteins (such as those involved in transcription and DNA replication) (67, 206). KS5, KS14, and KL3 appear to have retained P2 modules for the closely interacting capsid and tail proteins, while acquiring new modules for carrying out *Burkholderia* host-specific processes. These genes replace P2 genes at the right end of the P2 genome (the TO-region), P2 *Z/fun* (the Z-region), and P2 *orf30* (Table 3-4) (215). As it is very common for genes not found in P2 to be identified in these three regions in other P2-like phages, it is predicted that these loci contain genes that have been acquired via horizontal transfer (215).

Although a phage may show relatedness to a well-characterized phage such as P2, specific guidelines must be used to determine both the degree of relatedness of two phages and if the novel phage can be classified as a "P2-like virus" in a strict taxonomic sense. Lavigne et al. (155) proposed the use of the comparison program CoreGenes to aid in phage taxonomic analysis. This program can be used to compare the proteomes of two or more phages (155). If a phage

shares at least 40% of its proteins (those with a BLASTP score ≥ 75) with a reference phage such as P2, then these two phages can be considered as part of the same genus, while if it shares 20-39% of its proteins with a reference phage, then they can be considered as part of the same subfamily (155). When KS5, KS14, and KL3 were analyzed with CoreGenes using P2 as a reference genome, the percentage of proteins in common with respect to P2 were 51.16%, 53.49%, and 53.49%, respectively. These are similar to the percentages for ϕE202 (55.81%), ϕ52237 (51.16%), and ϕE12-2 (48.84%) (154). Based on these results, KS5, KS14, and KL3 can be classified as members of the *Peduovirinae* subfamily and "P2-like viruses" genus (154).

Integration site characterization. In *E. coli*, P2 is able to integrate at over 10 different loci, but certain sites may be used more commonly than others (16). None of the three P2-like Bcc phages characterized here were found to integrate into a locus similar to that of P2. Only KL3 was found to have a previously characterized integration site. Following PCR amplification and sequencing from the *B. cenocepacia* CEP511 chromosome (where KL3 is carried as a prophage), it was determined that, like many phages, KL3 integrates into a tRNA gene. Specifically, it integrates into the middle of a threonine tRNA gene: bp 1 of the KL3 prophage corresponds to bp 32 of the tRNA based on comparison with a 76 bp threonine tRNA gene of *B. cenocepacia* HI2424 chromosome 1 (Bcen2424_R0015, bp 491,047-491,122). Other phages that integrate into threonine tRNA genes include enterobacteria phage P22, *Shigella flexneri* phage SfV and *Salmonella enterica* serovar Typhimurium phage ST104 (101, 122, 162).

KL3 integration should not affect threonine tRNA synthesis as bp 1-45 of KL3 has an identical sequence to bp 32-76 of the tRNA gene.

In both *B. multivorans* ATCC 17616 and *B. cenocepacia* C6433, KS5 integrates into the 3' end of an AMP nucleosidase gene. AMP nucleosidases convert AMP into adenine and ribose 5-phosphate (331). This gene has not been previously identified as a phage integration site. KS5 bases 1-815 (including the integration site and the integrase gene sequence) show similarity to sequences encoding pairs of adjacent AMP nucleosidase and integrase genes in several *Burkholderia* genomes. For example, in *B. pseudomallei* K96243 chromosome 2, the AMP nucleosidase (BPSS1777) and integrase (BPSS1776) genes are adjacent to genes annotated as encoding a putative phage capsid related protein (fragment) (BPSS1775) and putative phage-related tail protein (fragment) (BPSS1774A). Similarly, in *B. pseudomallei* 1106a chromosome 2, the AMP nucleosidase (BURPS1106A_A2416) and integrase (BURPS1106A_A2415) genes are adjacent to genes annotated as encoding a phage portal domain protein (BURPS1106A_A2414) and phage tail completion protein (BURPS1106A_A2413). The identification of phage related genes at this site in other *Burkholderia* genomes suggests that the AMP nucleosidase gene may be a conserved integration site among some *Burkholderia*-specific temperate phages.

KS14 is different from other P2-like phages in that it does not encode a tyrosine integrase. Most temperate phages use a tyrosine recombinase (or in rare cases a serine recombinase) to facilitate recombination between the phage *attP* site and the host *attB* site (100). KS14 encodes a serine recombinase (gp6), but

this protein is unlikely to mediate prophage integration for three reasons. First, gp6 is more closely related to invertases such as Mu Gin (49% identity, E-value: $8e^{-44}$) and P1 Cin (49% identity, E-value: $7e^{-43}$) than to integrases such as those from *Streptomyces lividans* phage ϕC31 (29% identity, E-value: 1.2) and *Mycobacterium smegmatis* phage Bxb1 (29% identity, E-value: $3e^{-4}$) (123, 143, 148, 234). Second, gp6 lacks the conserved cysteine-rich and leucine/isoleucine/valine/methionine-rich regions found in other serine integrases (275). Third, gp6 is only 225 aa in length, which is substantially smaller than the serine integrases that are typically between 450-600 aa in length (275). We did not believe KS14 to be obligately lytic because it encodes a putative repressor protein (gp5) and because previously collected KS14-resistant C6433 isolates were predicted to be lysogenized based on PCR-positivity with KS14-specific primers (Figure 3-4) (267).

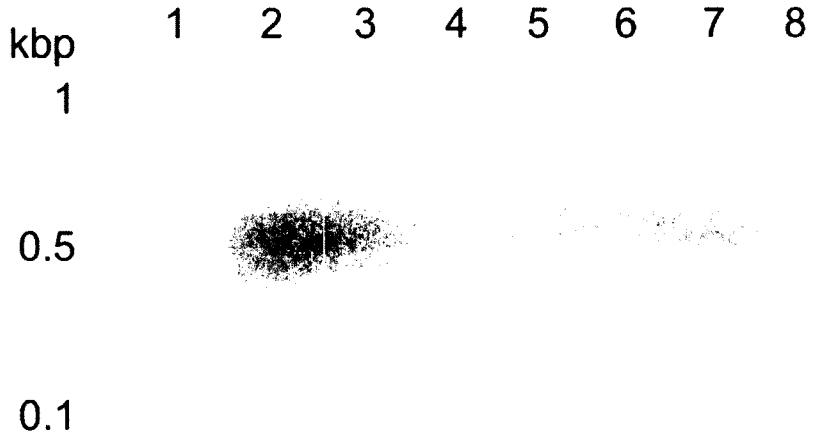


Figure 3-4: Detection of lysogeny in KS14-resistant *Burkholderia cenocepacia* C6433 isolates (267). Bacterial genomic DNA was amplified using KS14-specific primers. Lane 1: 1 Kb Plus DNA ladder (Invitrogen), lane 2: DNA-free control, lane 3: C6433 control, lane 4: KS14-resistant C6433 isolate I, lane 5: KS14-resistant C6433 isolate II, lane 6: KS14-resistant C6433 isolate III, lane 7: KS14-

resistant C6433 isolate IV, lane 8: KS14-resistant C6433 isolate V. The size of the markers (in kbp) is shown on the left.

Phages such as P1, P7, and N15 of enterobacteria, ϕ20 of *Bacillus anthracis*, ϕBB-1 of *Borrelia burgdorferi*, LE1 of *Leptospira biflexa*, pGIL01 of *Bacillus thuringiensis*, and pKO2 of *Klebsiella oxytoca* lysogenize their hosts as plasmids (48, 76, 91, 127, 129, 241, 306, 329). Because KS14 gene *39* encodes a putative ParA protein (involved in partitioning in other plasmid prophages), we predicted that the KS14 prophage might exist as a plasmid (3, 171). To test this hypothesis, we used a standard protocol for the QIAprep Spin Miniprep plasmid isolation kit with cells of C6433 (a KS14 host), ATCC 17616 (a KS5 lysogen), CEP511 (a KL3 lysogen), K56-2 (a lysogen of KS10, a previously characterized Bcc-specific phage) and five putatively lysogenized KS14-resistant C6433 isolates (93, 267). These preparations were then treated with EcoRI and the resulting fragments were separated using agarose gel electrophoresis. For each of the four control strains, no distinct bands were observed (Figure 3-5, left). In contrast, preparations from each of the five putatively lysogenized strains contained identical distinct bands (Figure 3-5, right). Furthermore, these bands were the same size as those predicted and observed for an EcoRI digest of KS14 DNA (with predictions based on a circular genome sequence) (Figure 3-5, far right) and sequences from selected bands matched the KS14 genome sequence. Based on these results, we predict that KS14 is a temperate phage that, in contrast to other P2-like phages, lysogenizes host strains as a plasmid.

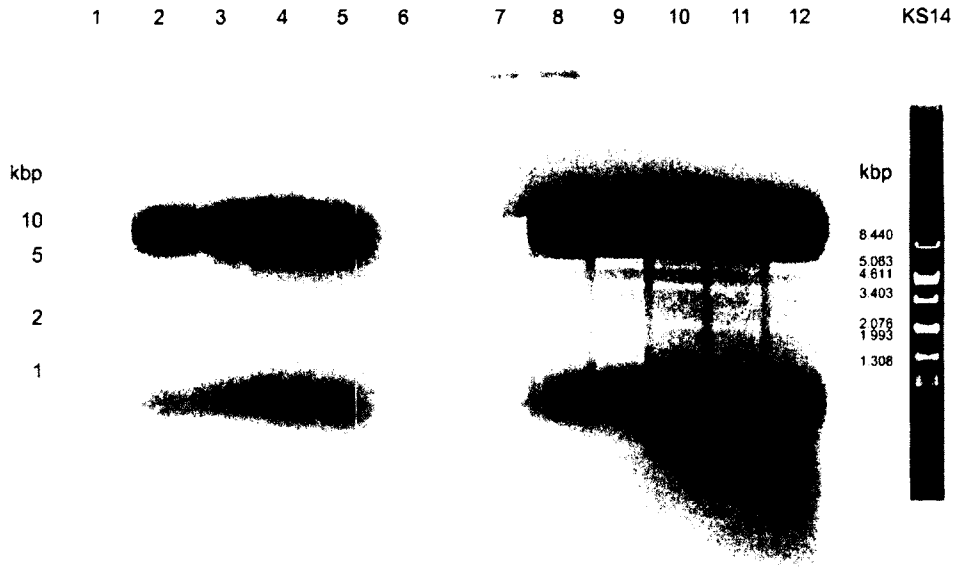


Figure 3-5: Isolation of the KS14 plasmid prophage. DNA was isolated using a QIAprep Spin Miniprep plasmid isolation kit (Qiagen) and digested with EcoRI (Invitrogen). Lane 1: 1 Kb Plus DNA ladder (Invitrogen), lane 2: *B. cenocepacia* C6433, lane 3: *B. multivorans* ATCC 17616, lane 4: *B. cenocepacia* CEP511, lane 5: *B. cenocepacia* K56-2, lane 6: blank, lane 7: 1 Kb Plus DNA ladder, lane 8: KS14-resistant C6433 isolate I, lane 9: KS14-resistant C6433 isolate II, lane 10: KS14-resistant C6433 isolate III, lane 11: KS14-resistant C6433 isolate IV, lane 12: KS14-resistant C6433 isolate V. The size of the markers (in kbp) is shown on the left. A KS14 EcoRI DNA digest and the size of the bands predicted for this digest (>1 kbp in size) are shown on the far right.

Morphogenesis genes. As discussed above, the KS5, KS14, and KL3 structural genes are related to those from P2 and function to construct a P2-like myovirus with a contractile tail. The only virion morphogenesis genes of P2 that these phages lack are *G* (encoding the tail fiber assembly protein, missing in KS5 and KL3) and *H* (encoding the tail fiber protein) (Table 3-4). Because the tail fibers are involved in host recognition, it is expected that these proteins would be dissimilar in phages infecting *E. coli* and those infecting the Bcc.

As discussed in Chapter 2, a commonly identified characteristic in tailed

phages is the expression of two tail proteins from a single start codon via a translational frameshift (324). These proteins (encoded in a region between the genes for the tail tape measure and the major tail protein) share the same N-terminus but have different C-termini due to stop codon readthrough in the -1 frame (324). In P2, this -1 frameshift occurs at a TTTTTTG sequence and produces the 91 aa protein E and the 142 aa protein E+E' from the same translational start site (Figure 3-6) (53, 324). KS5, KS14, and KL3 all encode proteins similar to both E and E+E' with percent identities ranging from 49-59% (Table 3-4). Despite the relatively low degree of similarity, the P2 frameshift site appears to be conserved amongst these phages, suggesting that they likely use a similar frameshifting mechanism (Figure 3-6). In rare cases, RNA secondary structure can be identified downstream of the phage frameshift sequence (176, 324). When the KS5, KS14, and KL3 *E+E'* sequences 60 bp downstream of the TTTTTTG sequence were screened for secondary structure, no predicted hairpins were identified. This result was anticipated based upon the absence of these structures in the P2 *E+E'* gene (324).

Figure 3-6: Conservation of the P2 *E/E+E'* frameshift sequence in KS5, KS14, and KL3. For each phage, the DNA sequence is shown in the first line, the translation in the original frame is shown in the second line, the translation in the -1 frame is shown in the third line and the amino acid sequence of the frameshifted protein is shown in the fourth line. The conserved TTTTTTG frameshift sequence is underlined. The frameshift is predicted to occur after the terminal G in this sequence.

Lysis genes. In P2, the lysis module consists of five genes: *Y* (holin), *K* (endolysin), *lysA* (antiholin), *lysB* (Rz), and *lysC* (Rz1) (285, 333). The P2-like Bcc phages are predicted to encode endolysins, holins and antiholins that are dissimilar to those of P2 (Table 3-4). KS5 gp33, KS14 gp27, and KL3 gp33 are putative endolysins as they all have the conserved domain pfam01471 (PG_binding_1, putative peptidoglycan binding domain; E-values: $3e^{-11}$, $3e^{-10}$, and $9e^{-10}$, respectively) and show similarity to other phage endolysins. P2 Y is a type I holin with three transmembrane domains (314). Although KS5 *34*, KS14 *28*, and KL3 *34* are dissimilar to P2 *Y*, it is predicted that these three genes encode type I holins because they are each immediately upstream of a putative endolysin gene

and they each encode proteins that a) have three transmembrane domains based on OCTOPUS analysis and b) show similarity to other phage holins.

Antiholins such as P2 LysA inhibit holin activity and delay lysis of infected cells in order to optimize the phage burst size (150, 333). Although some phages such as λ express antiholins from a second translational start site two codons upstream of the holin start codon, phages such as P2 and ϕO1205 of *Streptococcus thermophilus* encode an antiholin from a separate gene (95, 270, 333). The location of the putative antiholin genes KS5 *35*, KS14 *29*, and KL3 *35* is similar to that in ϕO1205, in which the holin and antiholin genes are adjacent immediately upstream of the endolysin gene (as opposed to P2, in which gene *K* separates *Y* and *lysA*) (270, 333). Based on OCTOPUS analysis, KS5 gp35 has three transmembrane domains, while KS14 gp29, KL3 gp35, and P2 LysA have four. Based on gene organization and protein transmembrane structure, it is predicted that the P2-like Bcc phages have separate antiholin genes in their lysis modules.

P2 encodes two proteins, LysB and LysC, that are predicted to function similarly to λ Rz and Rz1 (285). Rz is an inner membrane protein with an N-terminal transmembrane domain and Rz1 is a proline-rich outer membrane lipoprotein (20). Rz/Rz1 pairs fuse the inner and outer membranes following holin and endolysin activity and facilitate phage release (20). The P2 *lysC* start codon is in the +1 frame within the *lysB* gene, while the *lysC* stop codon is out of frame in the downstream tail gene *R* (193). In contrast, the *Rz1* gene in λ is entirely contained within the *Rz* gene (109). KS14 and KL3 LysBC pairs (gp26/gp25 and

gp32/gp31, respectively) are similar to that of P2 (Table 3-4). In KS14 and KL3, the *lysC* genes start approximately 160 bp upstream from the *lysB* stop codon and extend into the first 8 bp of *R* (gene *24* in KS14 and *30* in KL3) (Figure 3-7). Both KS14 and KL3 LysC proteins are predicted to have a signal peptidase II cleavage site between positions 20 (alanine) and 21 (cysteine). Signal peptidase II cleavage would produce a 72 aa lipoprotein with 7 prolines (9.7% proline) for KS14 LysC and a 74 amino acid lipoprotein with 7 prolines (9.5% proline) for KL3 LysC.

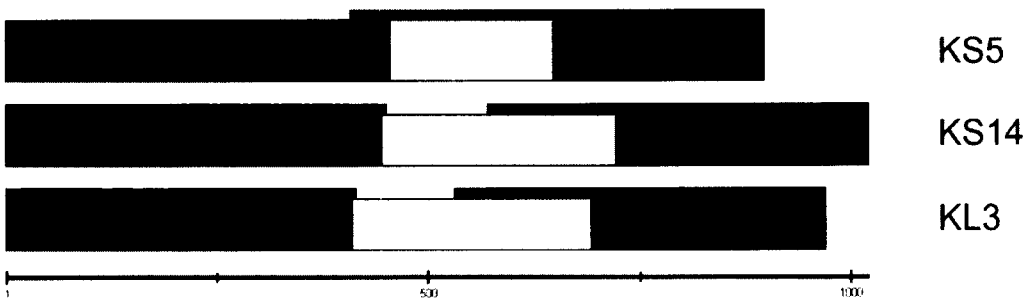


Figure 3-7: Organization of the *Rz/Rz1* and *lysBC* genes in KS5, KS14, and KL3. *R* genes (KS5 *30*, KS14 *24*, and KL3 *30*) are shown in blue, *Rz/lysB* genes (KS5 *32*, KS14 *26*, and KL3 *32*) are shown in green, and *Rz1/lysC* genes (KS5 *31*, KS14 *25*, and KL3 *31*) are shown in gray. The scale (in bp) is shown below.

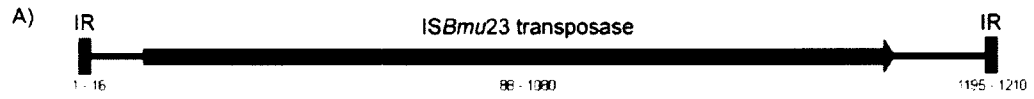
In contrast to the P2-like *lysBC* gene organization found in KS14 and KL3, the KS5 genes *32/31* have a similar organization to λ *Rz/Rz1*. KS5 Rz1 is encoded in the +1 frame within the *Rz* gene (Figure 3-7). It is predicted to have a signal peptidase II cleavage site between positions 18 (alanine) and 19 (cysteine), which would produce a 46 amino acid lipoprotein with 12 prolines (26.1%). The differences in both gene organization and proline content between the P2-like KS14 and KL3 LysC proteins and the λ-like KS5 Rz1 protein suggest that KS5 may have acquired genes *31* and *32* – and potentially the entire lysis module – through horizontal transfer from a phage similar to λ.

Sequence elements unique to KS5 and/or KL3.

Insertion sequences: Insertion sequences (ISs) are short genetic elements that can insert into nonhomologous regions of DNA (186). These elements, comprised of a transposase gene and inverted repeats, create flanking direct repeats following insertion (186). Many mutants of well-characterized phages have been found to carry ISs, including λ and Mu (52, 233). However, it is relatively rare for wild-type phages to carry ISs because they can interfere with gene expression (260). Sakaguchi et al. (260) determined the genome sequence of the *Clostridium botulinum* phage c-st and determined that it carries 12 ISs (5 of which are incomplete). Of the 284 genomes sequenced at the time, one IS was found in each of eight phages: *Burkholderia* phages ϕE125 and Bcep22, enterobacteria phages P1 and HK022, *Lactobacillus* phages ϕAT3 and LP65, *Rhodothermus* phage RM378, and *Shigella* phage Sf6 (260).

A novel insertion sequence (named IS*Bmu*23 in vB_BmuZ_ATCC 17616) is found in the KS5 genome between gene *12*, encoding a membrane protein and gene *13*, encoding the tail protein D (Table 3-1). This IS does not appear to disrupt any putative ORFs and so may not have any significant effect on phage gene expression. IS*Bmu*23 is 1210 bp in length and contains two imperfect 16 bp inverted repeats (Table 3-1, Figure 3-8). In KS5, it is flanked by two copies of a 5 bp direct repeat, CCTAA. IS*Bmu*23 encodes a 330 aa transposase that has the conserved domain COG3039 (transposase and inactivated derivatives, IS*5* family; E-value: $8e^{-29}$). Excluding IS*Bmu*23, this protein is most similar to the transposase of IS*Bmu*2 (85% identity), an IS*5*-like IS present in nine copies in ATCC 17616

(222). IS*Bmu*2 and IS*Bmu*23 are very similar as they a) are present in the same genome, b) are both 1210 bp in length, c) encode similar 330 aa transposases, d) have similar 16 bp inverted repeats (the right inverted repeats of IS*Bmu*2 and IS*Bmu*23 are identical, while the left repeats differ by 3 bp), and e) preferentially integrate into CTAA sequences (Figure 3-8) (222). Ohtsubo et al. (222) found that the transposition of ISs in ATCC 17616 increased when the cells were grown at high temperatures. Because these temperatures are similar to what the cell may encounter during infection of an animal or human, it is suggested that this change may provide a selective advantage to ATCC 17616 by modifying its genome under *in vivo* conditions (222). Further experiments are required to determine if IS*Bmu*23 transposition is affected by temperature and if this IS may provide a selective advantage to KS5 lysogens *in vivo*.



B)

IS*Bmu*2 L	G G A A A C G C T G A T T T A T
IS*Bmu*2 Ri	G G A A A C G C T G A T C A A T
IS*Bmu*23 L	G G A A A C A C T G A T C A A T
IS*Bmu*23 Ri	G G A A A C G C T G A T C A A T

Figure 3-8: Comparison of the IS*Bmu*23 and IS*Bmu*2 insertion sequences. A) Structure of IS*Bmu*23. IR, inverted repeats. Relative positions of the inverted repeats and transposase gene (in bp) are shown below. B) Alignment of the IS*Bmu*2 and IS*Bmu*23 inverted repeats. Non-consensus bases are underlined. IS*Bmu*2 sequences are from Ohtsubo et al. (222). L, left repeat; Ri, right repeat inverted.

Reverse transcriptases: Reverse transcriptases (RTs) are RNA-dependent DNA polymerases most commonly associated with retroviruses and retrotransposons (81). RTs have also been identified in several phage genomes, including those of P2-like phages (74, 201, 220). One function of these proteins

was extensively characterized in *Bordetella bronchiseptica* phage BPP-1. This phage has the ability to change its host range by making amino acid substitutions in its tail fiber protein, Mtd (major tropism determinant) (166). This switch requires the phage-encoded RT Brt (*Bordetella* RT) that synthesizes a DNA copy of a 134 bp locus (the template repeat, TR) that has 90% identity with a 134 bp region of the *mtd* gene (the variable repeat, VR) (74, 166). Adenines in the reverse transcribed copy of TR are mutagenized and the altered DNA integrates or recombines at VR by an unknown mechanism, generating a tail fiber gene with multiple base substitutions (74, 201).

A second function associated with phage RTs is phage exclusion. In *Lactococcus lactis*, expression of the RT-like protein AbiK significantly lowers the efficiency of plating of infecting phages (85, 313). Expression of Orf570, an RT identified in the P2-like enterobacteria prophage P2-EC30, was found to inhibit T5 infection of *E. coli* (220). When a region of Orf570 containing an RT conserved motif was deleted, T5 infection was no longer inhibited (220).

KS5 encodes a putative RT, gp44. This protein has the conserved domain cd03487 (RT_Bac_retron_II, reverse transcriptases in bacterial retrotransposons or retrons; E-value: $2e^{-45}$). It is unlikely that gp44 and Brt have the same function: the two proteins show minimal similarity (21% identity, E-value: $7e^{-4}$), gene *44* is located distal to the tail fiber gene (in contrast to *brt* and *mtd*), neither nucleotide substitutions in the tail fiber gene nor variations in KS5 tropism were observed, and no repeated sequences were identified in the KS5 genome longer than 28 bp (166). When compared to Orf570, gp44 shows almost no relatedness (41% over

12/546 amino acids; E-value: 2.7) but is found at the same locus (in the prophage, both *orf570* and *44* would be located proximal to the portal protein gene *Q*). Further experiments are required to determine if the KS5 RT is involved in tropism modification, phage exclusion, or some uncharacterized function.

DNA methylation, restriction and repair: DNA methylase and endonuclease genes are commonly found in phage genomes. Methylases modify the DNA such that it becomes resistant to bacterial restriction systems (276). Although P2 does not encode any putative methylases, such proteins are encoded by both KS5 and KL3 (KS5 gp20 and KL3 gp28 and gp47) (Tables 3-1 and 3-3). All three methylases are predicted to belong to the AdoMet_MTase superfamily (cl12011; S-adenosylmethionine-dependent methyltransferases). KS5 gp20 is most similar to a DNA methylase N-4/N-6 domain protein of *B. ambifaria* MEX-5 (89% identity). KL3 gp28 is most similar to a site-specific DNA methyltransferase of *B. pseudomallei* K96243 (78% identity). Both of these proteins have the conserved domain pfam01555 (N6_N4_Mtase, DNA methylase; KS5 gp20 E-value: $5e^{-22}$, KL3 gp28 E-value: $4e^{-25}$). Because this domain is associated with both N-4 cytosine and N-6 adenine methylases, these proteins may have either cytosine or adenine methylase activity. KL3 gp47 shows similarity to a modification methylase EcoRII from several bacterial species, with E-values as low as $4e^{-114}$. This protein has the conserved domain cd00315 (Cyt_C5_DNA_methylase, Cytosine-C5 specific DNA methylases; E-value: $6e^{-68}$) and so can be classified as a cytosine-C5 methylase. KS5 gp20 and KL3 gp28 are likely involved in protecting the phage DNA from Bcc restriction systems. As

discussed below, the function of KL3 gp47 is likely to protect the phage DNA from a phage-encoded restriction enzyme.

Phage nucleases have a number of functions, including degradation of the bacterial DNA (to both inhibit the host and provide nucleotides for the phage), phage exclusion, and DNA processing (204). KL3 encodes two endonucleases, gp45 and gp46. Gp45 is most similar to a type II restriction endonuclease, EcoRII-C domain protein of *Candidatus* Hamiltonella defensa 5AT (*Acyrthosiphon pisum*) (77% identity). This protein has the conserved domain pfam09019 (EcoRII-C, EcoRII C terminal; E-value: $6e^{-65}$). Gp46 is most similar to a DNA mismatch endonuclease Vsr of *Burkholderia graminis* C4D1M (77% identity). This protein has the conserved domain cd00221 (Vsr, Very Short Patch Repair [Vsr] endonuclease; E-value: $9e^{-38}$).

The organization of genes *45-47* (encoding an EcoRII-C endonuclease, Vsr endonuclease, and EcoRII methylase, respectively) in a single module suggests that the proteins that they encode are functionally related. The EcoRII-C endonuclease (which has a CCWGG recognition sequence where W = A or T) is likely to degrade either bacterial DNA to inhibit the host during the KL3 lytic cycle or superinfecting phage DNA (204, 211). KL3 DNA would be protected from this cleavage by EcoRII methylation at the second position in the EcoRII-C recognition sequence (forming $CC^{m}WGG$ where C^{m} = 5-methylcytosine) (264). Expression of the Dcm methylase, which has an identical recognition sequence and methylation site as EcoRII methylase, is mutagenic in *E. coli* because 5-methylcytosines are deaminated to thymines, causing T/G mismatches (61, 160).

EcoRII methylase expression would presumably cause mismatched sites in KL3 with the sequence C(T/G)WGG. In *E. coli*, these mismatches are repaired by very short patch (VSP) repair which starts with the recognition and nicking of the sequence C(T/G)WGG by a Vsr endonuclease (117). As KL3 expresses a Vsr endonuclease, it could repair post-methylation T/G mismatches using the same mechanism.

The proposed model for methylase and endonuclease interaction during the KL3 lytic cycle is as follows. Unmethylated host DNA (or, alternatively, superinfecting phage DNA) is degraded by gp45. KL3 DNA is protected from gp45 degradation by gp47-mediated conversion of cytosine to 5-methylcytosine. These 5-methylcytosine bases are deaminated to thymine, but the resulting T/G mismatches are cleaved by gp46 and fixed using VSP repair. Although further experiments are required to test the validity of this model, KL3 appears to encode an elegant system for degradation of bacterial or superinfecting phage DNA, protection of the phage genome, and repair of resulting mutations.

CONCLUSIONS

This study is the first to identify and characterize P2-like phages of the Bcc. Like other previously characterized P2-like *Burkholderia* phages, KS5, KS14, and KL3 share structural genes with P2 but encode dissimilar accessory proteins. KS5, a 37,236 bp prophage of *B. multivorans* ATCC 17616, integrates into an AMP nucleosidase gene, has a λ-like *Rz/Rz1* cassette, carries an IS*Bmu*2-like insertion sequence, and encodes a reverse transcriptase. KS14, a 32,317 bp

soil phage, encodes a serine recombinase but is maintained as a plasmid prophage. KL3, a 40,555 bp prophage of *B. cenocepacia* CEP511, integrates into a threonine tRNA gene and encodes a series of proteins capable of degrading bacterial or superinfecting phage DNA, methylating the phage genome, and repairing methylation-induced mismatches. As KS14 has already been shown to be active *in vivo* (267), characterization of these three related phages is an important preliminary step in the development of a phage therapy protocol for the Bcc.

ACKNOWLEDGEMENTS

The authors thank Randy Mandryk in the University of Alberta Department of Biological Sciences Advanced Microscopy Facility for assistance with electron micrographs, the University of Alberta Department of Biological Sciences Molecular Biology Service Unit for sequence data collection, Miles Peterson for assistance with figure construction, and Kimberley Seed for preliminary sequence collection.

Chapter 4

Comparative analysis of two convergently evolved *Burkholderia cenocepacia*-specific bacteriophages

A version of this chapter has been submitted as:

Lynch, K. H., P. Stothard, and J. J. Dennis. Submitted. Comparative analysis of two convergently evolved *Burkholderia cenocepacia*-specific bacteriophages. BMC Genomics.

OBJECTIVES

The objectives of this project were to a) sequence and characterize the genomes of KL1 (vB_BceS_KL1) and AH2 (vB_BceS_AH2), two *Burkholderia cenocepacia*-specific phages isolated from environmental samples and b) determine if the similar and unique phenotype of these phages can be attributed to either genomic relatedness or convergent evolution.

MATERIALS AND METHODS

Bacterial strains and growth conditions. *B. cenocepacia* strains K56-2 and C6433, part of the Bcc experimental strain panel (59, 183), were used for phage isolation and propagation. Strains used for host range analysis (also part of the panel) were acquired from the Belgium Coordinated Collection of Microorganisms LMG Bacteria Collection (Ghent, Belgium) and the Canadian *Burkholderia cepacia* complex Research and Referral Repository (Vancouver, BC). Strains were grown aerobically overnight at 30°C on half-strength Luria-Bertani (½ LB) solid medium or in ½ LB broth with shaking. Lysates for DNA isolation were prepared from soft agar overlays made with ½ LB medium containing agarose instead of agar.

Phage isolation and propagation. KL1 and AH2 were isolated from sewage and *Nandina* sp. soil, respectively, using standard extraction protocols (265). Environmental samples were incubated with shaking at 30°C in a slurry of ½ LB broth, suspension medium (SM; 50 mM Tris-HCl [pH 7.5], 100 mM NaCl, 10 mM $MgSO_4$, 0.01% gelatin solution), and Bcc liquid culture (K56-2 for KL1

isolation and C6433 for AH2 isolation). Solids were pelleted by centrifugation and the supernatant was filter sterilized, plated in soft agar overlays with the Bcc strain used in the extraction, and incubated overnight at 30°C and >24 h at room temperature. Plaques were picked using a sterile Pasteur pipette and transferred into 1 ml SM. Phage propagation and host range analysis were performed using soft agar overlays: 100 µl liquid culture and 100 µl phage stock (diluted in SM if necessary) were incubated 20 min at room temperature, mixed with 3 ml 0.7% ½ LB top agar, overlaid on a plate of ½ LB solid medium, and incubated at 30°C and room temperature until plaque formation was complete. High titre stocks were made by transferring multiple plaques into SM or by overlaying plates with SM and incubating 4-8 h at 4°C on a platform rocker.

Electron microscopy. Filter-sterilized high titre stocks of KL1 and AH2 were used for electron microscopy. 5-10 µl of phage lysate was deposited onto a carbon-coated copper grid and incubated 5 min at room temperature. Following adsorption of excess lysate onto a filter paper, the grids were stained with 2% phosphotungstic acid for 2 min. Grids were viewed using a Philips/FEI (Morgagni) transmission electron microscope with charge-coupled device camera (University of Alberta Department of Biological Sciences Advanced Microscopy Facility).

DNA isolation and sequencing. Phage DNA was isolated using polyethylene glycol precipitation and guanidine thiocyanate lysis. One hundred ml of phage lysate (propagated on C6433) was collected by overlaying turbid-clear or mottled ½ LB agarose plates with SM and incubating at 4°C 4-8 h on a

platform rocker. Following the addition of chloroform, debris in the lysate was pelleted by centrifugation for 10 min at 10,000 x *g* and 4°C and the supernatant was filter sterilized with a Millex-HA 0.45 µm syringe driven filter unit (Millipore, Billerica, MA). Fifty ml aliquots of the supernatant were incubated at 37°C ≥40 min with 10 µl DNase I, 10 µl DNase I buffer, and 6 µl RNase (Fermentas, Burlington, ON) to degrade contaminating bacterial nucleic acids. Following centrifugation for 10 min at 4000 x *g* and 4°C, phages in the supernatant were precipitated in 1 M NaCl and 10% w/v PEG 8000 at 4°C. The precipitated phages were pelleted by centrifugation for 20 min at 10,000 x *g* and 4°C and resuspended in 1.6 ml SM. To eliminate residual DNase I activity, the phage suspension was incubated at 37°C 10 min with 40 µl 20 mg/ml proteinase K. Following extraction of the phages with an equal volume of chloroform and the addition of EDTA to 100 mM, ½ volume of 6 M guanidine thiocyanate was added to disrupt the capsids and release the phage DNA. DNA was then purified using the GENECLEAN Turbo Kit (Qbiogene, Irvine, CA). Phage DNA was quantified using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific, Waltham, MA). RFLP analysis was performed using 5 µg of phage DNA digested overnight at 37°C with EcoRI (Invitrogen, Carlsbad, CA). For *cos* site screening, 5 µg EcoRI digests were incubated 20 min at 80°C, cooled on ice, and separated on 0.8% agarose gels in 1x Tris-acetate-EDTA (pH 8.0). Bands present only in the heated sample were excised from the gel, purified using a GENECLEAN III kit (Qbiogene), cloned into pJET1.2 (Fermentas), and sequenced to identify the *cos* site. Preliminary sequencing of EcoRI phage DNA fragments cloned into

pUC19 was performed as described previously (176, 177). For complete genome sequencing, phage DNA was submitted to 454 Life Sciences (Branford, CT) for pyrosequencing analysis. The genome sequences of KL1 and AH2 have been deposited in GenBank with the accession numbers JF939047 and JN564907. Sequence start sites for these files were chosen based on alignment with *Pseudomonas* phage 73 (NC_007806.1) for KL1 and at the *cos* site for AH2.

Bioinformatics analysis. Annotation of the genome sequences and determination of GC contents were performed using GeneMark.hmm-P (http://exon.biology.gatech.edu/gmhmm2_prok.cgi) (172). Manual annotations were performed for KL1 *5* (encoding Rz1) and KL1 *19*/AH2 *55* (encoding translationally-frameshifted tail proteins). Homology searches and conserved domain searches were performed using HHpred (http://toolkit.tuebingen.mpg.de/hhpred) (277) and NCBI's BLASTN/BLASTP (for full genomes and individual proteins, respectively) (http://blast.ncbi.nlm.nih.gov) (7) and Conserved Domain Search (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) (191). Sequence alignments were constructed using LALIGN (http://www.ch.embnet.org/software/LALIGN_form.html). FSFinder was used for translational frameshift identification (http://wilab.inha.ac.kr/fsfinder) (208). Sequence comparisons were visualized using Circos (http://mkweb.bcgsc.ca/circos) (147) and PROmer (http://mummer.sourceforge.net) (68) with the following parameters: breaklen = 60, maxgap = 30, mincluster = 20, minmatch = 6. Lysis protein analysis was

performed using TMHMM for transmembrane region identification (http://www.cbs.dtu.dk/services/TMHMM) (145) and LipoP for signal peptidase II cleavage site identification (http://www.cbs.dtu.dk/services/LipoP) (136).

RESULTS AND DISCUSSION

Isolation, host range, and morphology. KL1 was isolated from sewage using *B. cenocepacia* K56-2 as a host. In contrast to enterobacteria phages, which are commonly found in sewage (128), this is the first report of Bcc phage isolation from this source. AH2 was isolated from *Nandina* sp. (also known as heavenly bamboo) soil using *B. cenocepacia* C6433. Bcc phages have commonly been isolated from both rhizospheres and soil samples, including that of onion and *Dracaena* sp. (89, 152, 265, 267, 287).

KL1 and AH2 are very similar with respect to both host range and lysis characteristics. These phages have a relatively narrow tropism, infecting only *B. cenocepacia* K56-2 and C6433 out of over fifty Bcc strains tested. Both KL1 and AH2 exhibit a pattern of lysis that is unique in our collection of Bcc-specific phages: although high titre stocks of these phages are very concentrated (up to 10^{11} PFU/ml), these phages do not produce clear lysis in agar overlays after 16 h incubation like other phages that we have previously isolated (93, 176, 177, 265, 267). Instead, turbid clearing (on C6433) or no clearing (on K56-2) is observed at high titres, with mottling or individual plaques observed at lower titres (approximately 10^7 PFU/ml or less). Incubation at 30°C for greater than 16 h is

required for plaque formation at low titres (Figure 4-1). Individual plaques are turbid (more so on K56-2 than C6433) with a diameter of 0.5-2 mm.

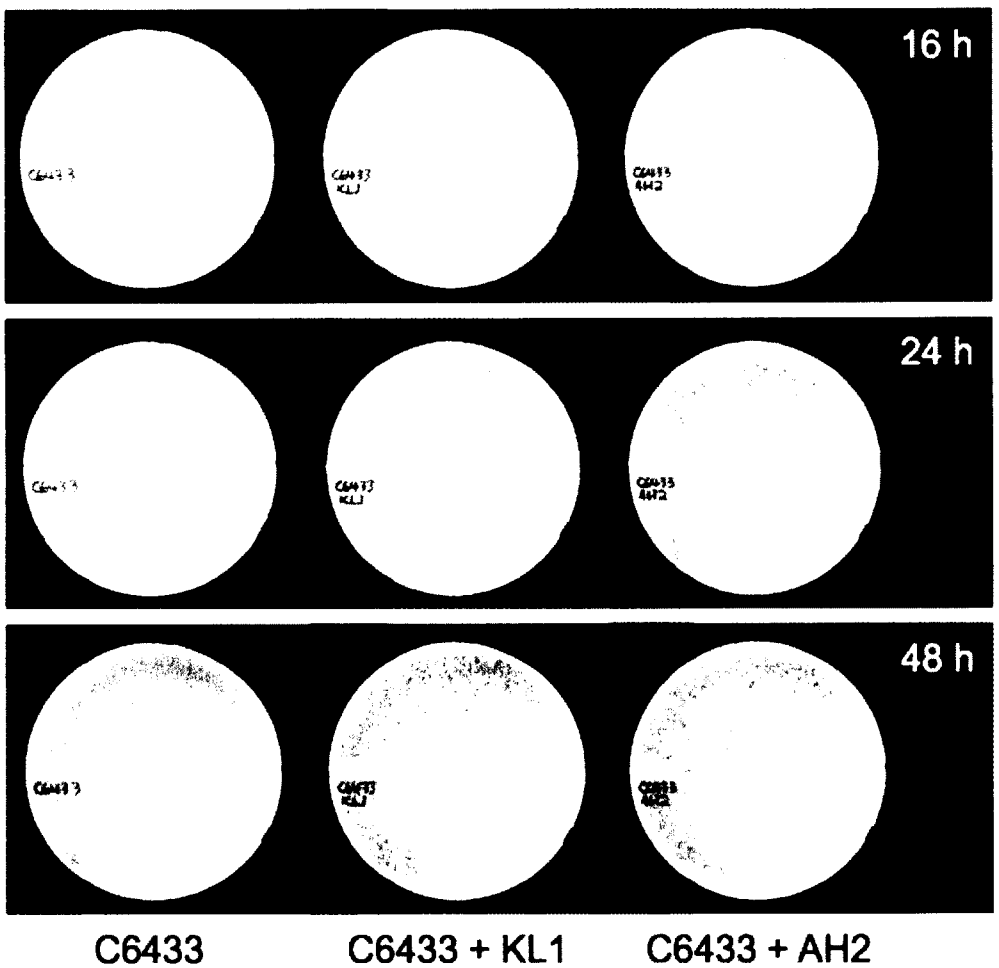


Figure 4-1: Development and morphology of KL1 and AH2 plaques. Phages were plated in agar overlays with *Burkholderia cenocepacia* C6433 at an OD_{600} of 2.0. Plates were incubated at 30°C and photographed after 16, 24, and 48 h.

Both KL1 and AH2 belong to the order *Caudovirales* and family *Siphoviridae* as determined by electron microscopy. The KL1 virion has a non-contractile tail approximately 160 nm in length and a capsid approximately 55 nm in diameter (Figure 4-2A). The AH2 virion is slightly larger, with a non-contractile tail approximately 220 nm in length and a capsid approximately 60 nm

in diameter (Figure 4-2B). The stacked rings comprising the tail structure are visible in the AH2 micrograph (Figure 4-2B).

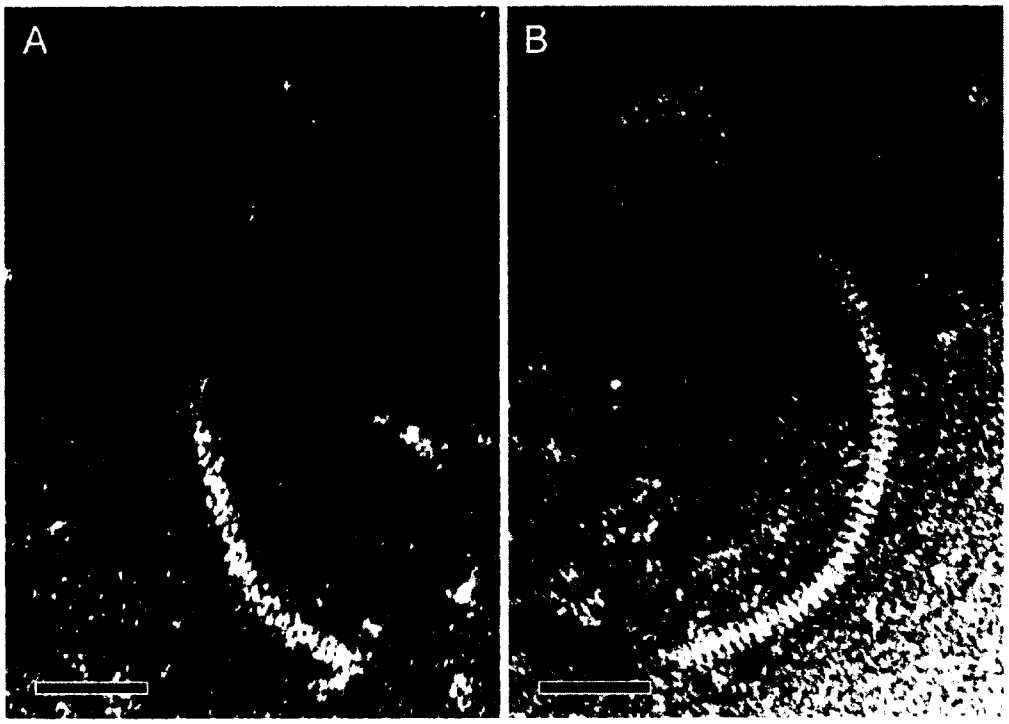


Figure 4-2: KL1 (A) and AH2 (B) virion morphology. Phages were stained with 2% phosphotungstic acid and visualized at 180,000-fold magnification by transmission electron microscopy. Scale bars represent 50 nm.

Genome characterization. Despite the similarities in phenotype between KL1 and AH2 with respect to host range and lysis characteristics, the genomes of these two phages are dissimilar. RFLP analysis shows distinct banding patterns of EcoRI-digested KL1 and AH2 genomic DNA, suggesting that their sequences are substantially different (Figure 4-3). This prediction is confirmed by the results of whole genome pyrosequencing (discussed below).

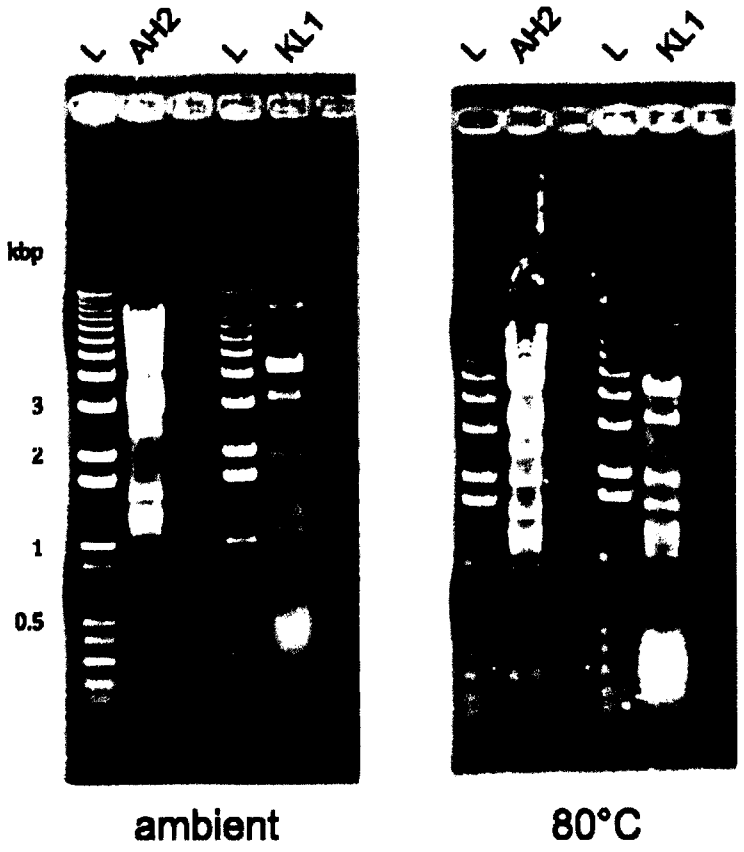


Figure 4-3: RFLP analysis of KL1 and AH2 genomic DNA. Five micrograms of genomic DNA were digested overnight with EcoRI and separated on a 0.8% agarose gel. The DNA in the ambient gel (left) was not heated, whereas the DNA in the 80°C gel (right) was incubated 20 min at 80°C and chilled on ice prior to loading. Arrows indicate bands containing *cos* site DNA. L: 1 Kb Plus DNA Ladder (Invitrogen).

The KL1 genome is 42,832 base pairs (bp) in length and has a 54.6% GC content. This percentage is lower than that for most *Burkholderia*-specific phages, which tend to have GC contents between 60–65% (excluding phages such as BcepB1A [54.5%], BcepF1 [55.9%], and BcepGomr [56.3%]). We were unable to identify a KL1 *cos* site following incubation of the DNA at 80°C, as the RFLP profiles appeared identical both before and after heating (Figure 4-3). KL1 is predicted to encode 55 proteins, all of which have an ATG start codon, except for

gp2 which has a GTG codon (Figure 4-4, Table 4-1). The proteins with the most and least similarity to others in the database are gp33 (which has 91% identity with *Pseudomonas* phage 73 [PA73] hypothetical protein ORF032) and gp23 (which has 25% identity with *Pseudomonas* phage M6 hypothetical protein ORF033), respectively (Table 4-1).

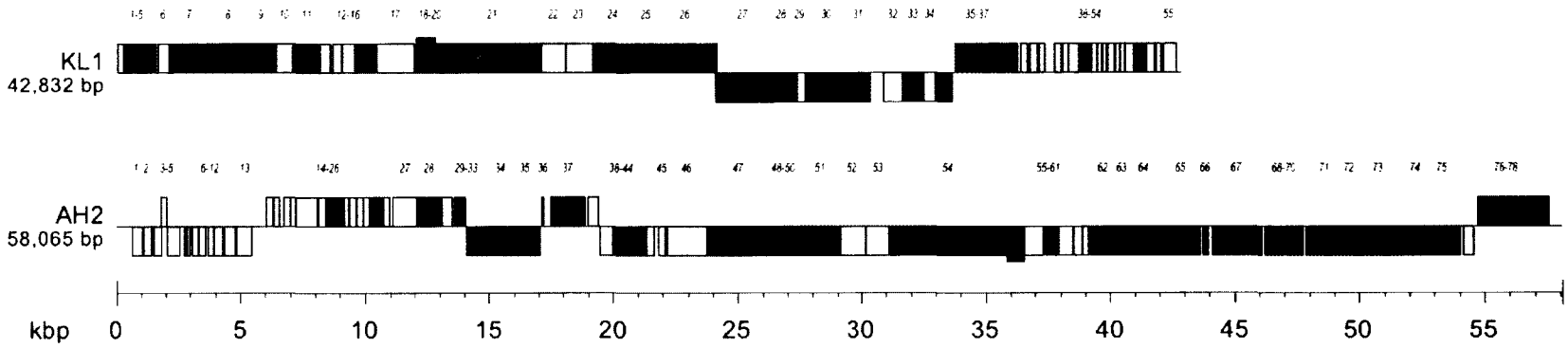


Figure 4-4: Genome maps of KL1 and AH2. Genes transcribed in the forward direction are shown above and those transcribed in the reverse direction are shown below. The scale (in kbp) is shown below the maps. Legend: light blue, lysis; purple, capsid morphogenesis and DNA packaging; pink, tail morphogenesis; red, DNA binding; green, MazG; gray, unknown function.

Table 4-1: KL1 genome annotation

Gene	Start	End	Putative function	Strand	Predicted RBS and start codon	Length (aa)	Closest relative	Alignment region in closest relative (aa)	% ID	Source	GenBank accession number
1	1	267	unknown	+	AGGGGCGAActtcgtATG	88	hypothetical protein ORF001	1-84/84	77	*Pseudomonas* phage 73	YP_001293408.1
2	264	560	holin	+	AAAGGGGCGGtaacGTG	98	hypothetical protein ORF002	3-88/88	42	*Pseudomonas* phage 73	YP_001293409.1
3	514	1080	lysozyme	+	AAAAGGGGttatcgaATG	188	hypothetical protein bglu_1g27070	2-181/188	47	*Burkholderia glumae* BGR1	YP_002912484.1
4	1091	1408	Rz	+	AAGTAAGGGGttcgaaATG	105	hypothetical protein ORF004	1-101/101	37	*Pseudomonas* phage 73	YP_001293411.1
5	1329	1592	Rz1	+	GAAAGGtgccgccgATG	87	conserved hypothetical protein	1-79/86	40	*Burkholderia* sp. Ch1-1	ZP_06842908.1
6	1647	2138	unknown	+	ACTAGGccgcgattATG	163	hypothetical protein ORF005	1-162/162	59	*Pseudomonas* phage 73	YP_001293412.1
7	2116	3756	terminase large subunit	+	AACAGGAAttgcttaATG	546	hypothetical protein ORF006	10-531/531	84	*Pseudomonas* phage 73	YP_001293413.1
8	3770	5266	portal protein	+	AAAGGAAAcgaaatcATG	498	hypothetical protein ORF007	3-494/501	85	*Pseudomonas* phage 73	YP_001293414.1
9	5269	6384	head morphogenesis protein	+	GGGGCGTAatcATG	371	hypothetical protein ORF008	1-364/364	73	*Pseudomonas* phage 73	YP_001293415.1
10	6403	7110	unknown	+	AAGGAGtccttgaaATG	235	hypothetical protein ORF009	1-235/239	82	*Pseudomonas* phage 73	YP_001293416.1
11	7123	8097	major capsid protein	+	AAGGAcactttatcATG	324	hypothetical protein ORF010	1-325/325	90	*Pseudomonas* phage 73	YP_001293417.1
12	8171	8587	unknown	+	AAGGAGtttcgaacATG	138	hypothetical protein ORF011	1-134/134	69	*Pseudomonas* phage 73	YP_001293418.1
13	8656	9033	unknown	+	AAAGGAGcgtcgaacATG	125	hypothetical protein ORF012	1-123/123	70	*Pseudomonas* phage 73	YP_001293419.1
14	9047	9565	unknown	+	AAGGGGcgcggcatcATG	172	hypothetical protein ORF013	1-172/172	83	*Pseudomonas* phage 73	YP_001293420.1
15	9570	9944	head-tail joining protein	+	GATAAGGGtctaacgctATG	124	hypothetical protein ORF014	1-124/126	59	*Pseudomonas* phage 73	YP_001293421.1
16	9941	10399	minor tail protein	+	ATACGGTAttgttcgcacaATG	152	hypothetical protein ORF015	5-151/151	68	*Pseudomonas* phage 73	YP_001293422.1
17	10412	11965	unknown	+	AAGGAGttacgaaaATG	517	hypothetical protein ORF016	3-511/511	78	*Pseudomonas* phage 73	YP_001293423.1
18	12030	12458	tail protein	+	GGAGTAAAccaa	142	hypothetical protein	1-142/142	79	*Pseudomonas* phage 73	YP_001293424.1

					ATG		ORF017				
19	12030	12823	tail protein	+	GGAGTAAAccaa ATG	264	hypothetical protein ORF017	1-142/142	79	*Pseudomonas* phage 73	YP_001293424.1
							hypothetical protein ORF018	1-118/118	78	*Pseudomonas* phage 73	YP_001293425.1
20	12792	13226	tail protein	+	AAAAGGCGGcg caacagaATG	144	hypothetical protein ORF019	1-144/144	80	*Pseudomonas* phage 73	YP_001293426.1
21	13232	17050	tail tape measure	+	AAGGAttagcagaa ATG	1272	hypothetical protein ORF020	1-78, 131-1202/1204	61, 57	*Pseudomonas* phage 73	YP_001293427.1
22	17069	18067	unknown	+	AGGAAtacgaattA TG	332	hypothetical protein XAlc_0225	1-295/307	30	*Xanthomonas albilineans* GPE PC73	YP_003374757.1
23	18070	19179	unknown	+	GAGGAAAActaa tcATG	369	hypothetical protein ORF033	1-332/333	25	*Pseudomonas* phage M6	YP_001294541.1
24	19179	20870	tail assembly protein	+	AAGAAGAtcgcat aATG	563	hypothetical protein ORF023	63-565/568	36	*Pseudomonas* phage 73	YP_001293430.1
25	20867	21688	tail assembly protein	+	AAGGAcgattccag aATG	273	hypothetical protein ORF024	1-273/274	49	*Pseudomonas* phage 73	YP_001293431.1
26	21689	24100	tail assembly protein	+	AAGATGGGGtc ggttaaATG	803	hypothetical protein ORF025	1-755/813	49	*Pseudomonas* phage 73	YP_001293432.1
27	24097	26166	DNA polymerase	-	AAGGAAtttgcccg ATG	689	hypothetical protein ORF026	1-682/683	83	*Pseudomonas* phage 73	YP_001293433.1
28	26179	27339	DNA polymerase III β subunit	-	AAGGGGttaaaaA TG	386	hypothetical protein ORF027	2-380/380	74	*Pseudomonas* phage 73	YP_001293434.1
29	27323	27691	unknown	-	GAATGGtgaaatt ATG	122	hypothetical protein Dole_2913	5-84/87	33	*Desulfococcus oleovorans* Hxd3	YP_001530793.1
30	27696	29351	superfamily II helicase/ restriction enzyme	-	AAGGGttacgaAT G	551	hypothetical protein ORF029	1-551/551	90	*Pseudomonas* phage 73	YP_001293436.1
31	29344	30342	exonuclease	-	GGAAGGcgaaga acgATG	332	hypothetical protein ORF030	1-365/365	65	*Pseudomonas* phage 73	YP_001293437.1
32	30852	31637	unknown	-	GAAAGGtgaaacg aacATG	261	hypothetical protein Isop_2441	1-118/151	37	*Isosphaera pallida* ATCC 43644	YP_004179564.1
33	31696	32412	recombinase	-	AGGTGAAcgtAT G	238	hypothetical protein ORF032	1-238/238	91	*Pseudomonas* phage 73	YP_001293439.1
34	32471	32980	unknown	-	AAGGAAccccaaa ATG	169	hypothetical protein ORF033	7-146/146	49	*Pseudomonas* phage 73	YP_001293440.1
35	33059	33598	pyrophospho-hydrolase	-	AGGGGcatcgtAT G	179	hypothetical protein ORF034	8-185/185	69	*Pseudomonas* phage 73	YP_001293441.1
36	33746	33934	transcriptional	+	GGGGcaagcATG	62	hypothetical protein	1-61/62	51	*Pseudomonas* phage 73	YP_001293442.1

			regulator				ORF035				
37	33924	36233	primase	+	GAAGGcttgcgcaa atATG	769	hypothetical protein ORF036	1-773/773	85	*Pseudomonas* phage 73	YP_001293443.1
38	36366	36668	unknown	+	GAAGGAgttacga acATG	100	hypothetical protein	132-217/217	44	*Deftia* phage ϕW-14	YP_003359005.1
39	36735	37091	unknown	+	GAAGGAGtacac gccATG	118	hypothetical protein BATDEDRAFT_85840	84-158/952	30	*Batrachochytrium dendrobatidis* JAM81	EGF83277.1
40	37097	37360	unknown	+	AGAAGAAGGA GtaagcgccATG	87	PREDICTED: hypothetical protein	22-86/179	32	*Vitis vinifera*	XP_002271666.1
41	37728	38024	unknown	+	AAAGGAGcgcca gccATG	98	hypothetical protein ORF039	1-97/98	70	*Pseudomonas* phage 73	YP_001293446.1
42	38060	38296	unknown	+	AAGGAAccccgat cATG	78	hypothetical protein ORF040	1-80/80	50	*Pseudomonas* phage 73	YP_001293447.1
43	38302	38703	unknown	+	AAAGGGGtaatta ctATG	133	hypothetical protein ORF042	1-120/124	40	*Pseudomonas* phage 73	YP_001293449.1
44	38707	39195	Vsr endonuclease	+	GACGAAGttgcatt aagccATG	162	hypothetical protein ORF043	1-176/179	61	*Pseudomonas* phage 73	YP_001293450.1
45	39201	39458	unknown	+	GGAAGGAGtaac ccaaATG	85	hypothetical protein Astex_0306	3-81/183	44	*Asticcacaulis excentricus* CB 48	YP_004086155.1
46	39455	39655	unknown	+	GGCGAAGtcgtcg aATG	66	monooxygenase, FAD-binding	385-445/546	38	*Streptomyces griseoflavus* Tu4000	ZP_07309792.1
47	39652	39840	unknown	+	AAGGAGtacgcac cATG	62	hypothetical protein METUNv1_00516	11-65/68	39	*Methyloversatilis universalis* FAM5	ZP_08503515.1
48	39882	40154	unknown	+	AAAAGGAGtaac gaacATG	90	hypothetical protein Cflav_PD2164	58-133/172	30	bacterium Ellin514	ZP_03630603.1
49	40138	40374	unknown	+	GAACCGGAttac gattATG	78	hypothetical protein ORF047	2-77/77	67	*Pseudomonas* phage 73	YP_001293454.1
50	40374	40550	unknown	+	GGGTTAcgaataA TG	58	SNARE associated Golgi protein-like protein	90-140/227	29	*Glaciecola agarilytica* 4H-3-7+YE-5	YP_004435864.1
51	40562	40933	unknown	+	GAAAGGtgaaatc ATG	123	hypothetical protein BURMUCGD2M_4586	8-67/70	34	*Burkholderia multivorans* CGD2M	ZP_03569237.1
52	40930	41415	dCMP deaminase	+	GGAACGtccggc ATG	161	hypothetical protein ORF049	2-153/155	75	*Pseudomonas* phage 73	YP_001293456.1
53	41412	41786	unknown	+	AAAGGctgaatcA TG	124	hypothetical protein ORF050	4-125/127	43	*Pseudomonas* phage 73	YP_001293457.1
54	41826	42032	unknown	+	GGGGAtgcccacat tATG	68	hypothetical protein ORF051	37-94/94	45	*Pseudomonas* phage 73	YP_001293458.1
55	42120	42674	unknown	+	AAGGAGttttacaa ATG	184	hypothetical protein ORF052	9-190/190	66	*Pseudomonas* phage 73	YP_001293459.1

Abbreviations: RBS, ribosome binding site; aa, amino acid; % ID, percent identity.

The AH2 genome is 58,065 bp in length and has a 61.3% GC content. Incubation of the DNA at 80°C caused a shift in the RFLP profile (Figure 4-3), suggesting the presence of a *cos* site. Sequencing of the shifted fragments indicates that AH2 has a 12 bp *cos* site with a sequence almost identical (1 bp difference) to that of *Burkholderia* phage BcepNazgul. AH2 is predicted to encode 78 proteins (Figure 4-4, Table 4-2). The majority of the start codons (70) are ATG, six are GTG, and two are TTG (Table 4-2). The proteins with the most and least similarity to others in the database are gp12 (which has 74% identity with *Burkholderia* phage BcepNazgul hypothetical protein Nazgul10) and gp52 (which has 23% identity with *Aggregatibacter actinomycetemcomitans* D11S-1 hypothetical protein D11S_2171), respectively (Table 4-2).

Table 4-2: AH2 genome annotation

Gene	Start	End	Putative function	Strand	Predicted RBS and start codon	Length (aa)	Closest relative	Alignment region in closest relative (aa)	% ID	Source	GenBank accession number
1	619	1035	unknown	-	AAGGAAAcgac ATG	138	hypothetical protein Nazgul32	12-130/130	29	*Burkholderia* phage BcepNazgul	NP_918966.1
2	1073	1423	unknown	-	AGGGGGGAAcg gccATG	116	conserved hypothetical protein	1-116/116	72	*Burkholderia multivorans* CGD1	ZP_03586942.1
3	1501	1818	unknown	-	GGATTActgaccA TG	105	glycosyl transferase family 2	292-387/387	32	*Haloterrigena turkmenica* DSM 5511	YP_003404522.1
4	1809	2024	unknown	+	GAGAAAtagagA TG	71	mobilization protein mbeA	190-237/325	37	*Escherichia coli* E128010	EFZ49597.1
5	2021	2578	unknown	-	AGGGGttacatcA TG	185	hypothetical protein Nazgul06	88-158/330	44	*Burkholderia* phage BcepNazgul	NP_919015.1
6	2728	2877	unknown	-	AGGTGcaaaaAT G	49	hypothetical protein BoklE_20935	6-38/38	48	*Burkholderia oklahomensis* EO147	ZP_02357945.1
7	2874	3002	unknown	-	AGGGGcgatcAT G	42	polysaccharide deacetylase	21-60/287	35	*Bacillus mycoides* Rock3-17	ZP_04156726.1
8	3071	3325	unknown	-	AAAGAgctATG	84	major facilitator superfamily MFS_1	131-209/467	37	*Burkholderia gladioli* BSR3	YP_004349464.1
9	3322	3579	unknown	-	GGAGTAtccgccA TG	85	hypothetical protein Plabr_1809	308-361/603	31	*Planctomyces brasiliensis* DSM 5305	YP_004269441.1
10	3663	3911	unknown	-	GGGGGTAtgacA TG	82	HAD-superfamily subfamily IIA hydrolase like protein	70-119/268	38	*Methanosphaerula palustris* E1-9c	YP_002465429.1
11	3913	4314	unknown	-	AGGGGGAGtaac ggccATG	133	hypothetical protein Nazgul09	1-129/141	59	*Burkholderia* phage BcepNazgul	NP_919018.1
12	4320	4805	unknown	-	AGGGGttacatcA TG	161	hypothetical protein Nazgul10	1-151/160	74	*Burkholderia* phage BcepNazgul	NP_919019.2
13	4846	5454	unknown	-	AAAAAGGGGttt ttgacATG	202	hypothetical protein	101-187/188	43	*Salmonella* phage PVP-SE1	ADP02590.1
14	6021	6302	unknown	+	AAGGAGcaatcA TG	93	hypothetical protein Nazgul13	3-93/93	41	*Burkholderia* phage BcepNazgul	NP_919022.1
15	6311	6550	unknown	+	AGGCGGtcgtAT G	79	hypothetical protein BDB_mp60418	1-67/67	45	blood disease bacterium R229	CCA83252.1
16	6707	7015	unknown	+	ACACGAcaccAT G	102	hypothetical protein MC7420_4162	43-84/88	45	*Microcoleus chthonoplastes* PCC 7420	ZP_05027813.1
17	7012	7218	unknown	+	GAAGGtgccggcA TG	68	hypothetical protein ccc_5268	62-90/161	45	*Cyanothece* sp. ATCC 51142	YP_001798668.1
18	7215	8069	unknown	+	AGGAAAGgaaA	284	hypothetical protein	5-175/177	45	*Thioalkalivibrio* sp. K90mix	YP_003494636.1

					TG		TK90_2682				
19	8123	8407	unknown	+	GAGAAGGcacac acATG	94	GTP-binding protein	150-232/1016	29	*Gemmata* sp. Wa1-1	AAX07516.1
20	8499	9128	DNA polymerase III β subunit	+	GAACGGTGAGc ttATG	209	hypothetical protein Nazgul21	24-216/237	24	*Burkholderia* phage BcepNazgul	NP_918955.1
21	9149	9343	unknown	+	AGGAGAAAGga gATG	64	beta-hexosaminidase	313-360/539	40	*Streptomyces lividans* TK24	ZP_06533176.1
22	9346	9645	unknown	+	GGGGGTAtctgac cATG	99	hypothetical protein PFL_2108	3-63/70	33	*Pseudomonas fluorescens* Pf-5	YP_259216.1
23	9642	9938	unknown	+	GGAGGGtcaTTG	98	aspartoacylase, putative	38-122/317	32	*Rhodospirillum centenum* SW	YP_002297975.1
24	9935	10171	unknown	+	GGGGcttggcgtAT G	78	hypothetical protein Nazgul19	18-97/97	39	*Burkholderia* phage BcepNazgul	NP_919028.2
25	10256	10711	pyrophospho-hydrolase	+	AAGGAAAggac ATG	151	hypothetical protein BCAS0549	15-139/140	60	*Burkholderia cenocepacia* J2315	YP_002153936.1
26	10720	10977	unknown	+	GAGGccggccAT G	85	hypothetical protein AGRO_3677	208-273/300	41	*Agrobacterium* sp. ATCC 31749	ZP_08529674.1
27	11082	12074	unknown	+	AGGAGAAatcG TG	330	hypothetical protein	8-95/113	48	Enterobacteria phage vB_EcoM_ECO1230-10	ADE87960.1
28	12101	13075	transcriptional regulator	+	AAGGAAccgacA TG	324	hypothetical protein Pnap_4317	25-252/342	45	*Polaromonas naphthalenivorans* CJ2	YP_973341.1
29	13078	13497	unknown	+	GCTGACGAtctct gaccATG	139	hypothetical protein SCHCODRAFT_69044	549-631/848	33	*Schizophyllum commune* H4-8	XP_003030158.1
30	13574	13768	transcriptional regulator	+	AGGGAtttttcATG	64	hypothetical protein APA01_20380	8-64/74	53	*Acetobacter pasteurianus* IFO 3283-01	YP_003188539.1
31	13768	14031	transcriptional regulator	+	AAGCGGAGccgt cctgATG	87	hypothetical protein Bcep1808_2468	2-85/86	73	*Burkholderia vietnamiensis* G4	YP_001120302.1
32	14064	14450	Vsr endonuclease	-	GGAGGAatgAT G	128	DNA mismatch endonuclease Vsr	15-141/141	65	*Methylocella silvestris* BL2	YP_002360880.1
33	14450	15025	excinuclease	-	AACAGAGttgcag cGTG	191	Excinuclease ABC C subunit domain protein	3-183/192	58	*Pseudomonas syringae* pv. *lachrymans* str. M301315	EGH83133.1
34	15038	15892	restriction endonuclease	-	GGCAAAGGtcgc cgcATG	284	conserved hypothetical protein	1-285/285	70	*Ralstonia solanacearum* CMR15	CBJ36134.1
35	15889	17031	cytosine methylase	-	AGGGGGttcgcG TG	380	DNA-cytosine methyltransferase	1-385/385	66	*Ralstonia solanacearum* CMR15	CBJ36133.1
36	17107	17199	unknown	+	ACGAAGccttgctt aATG	30	resistance-nodulation-cell division acriflavin:proton (H+) antiporter	850-868/1014	68	*Bacillus pumilus* SAFR-032	YP_001486844.1
37	17511	18842	integrase	+	GAAGGAGGtctt gtagcactgATG	443	chorismate mutase family protein	1-362/386	62	*Phaeobacter gallaeciensis* BS107	ZP_02147383.1

38	18990	19412	unknown	+	AAGGAGGAatc ATG	140	hypothetical protein Dda3937_00584	60-163/163	40	*Dickeya dadantii* 3937	YP_003882998.1
39	19462	20001	unknown	-	GGAGAtttteATG	179	hypothetical protein PcarcW_20243	68-197/198	67	*Pectobacterium carotovorum* subsp. carotovorum WPP14	ZP_03833564.1
40	20034	20264	Rz1	-	GGAGGAcgccAT G	76	hypothetical protein BURPS668_A2333	27-81/81	62	*Burkholderia pseudomallei* 668	YP_001063327.1
41	20277	20588	Rz	-	AGGGGGccgtAT G	103	hypothetical protein ORF004	2-101/101	35	*Pseudomonas* phage 73	YP_001293411.1
42	20585	21091	lysozyme	-	AAGGAGAAGA acaGTG	168	hypothetical protein HMPREF0005_02034	1-161/163	60	*Achromobacter xylosoxidans* C54	EFV83908.1
43	21088	21339	holin	-	GAAGGGGtggac ccgaccATG	83	conserved exported hypothetical protein	1-83/85	35	blood disease bacterium R229	CCA83792.1
44	21336	21665	unknown	-	AAGGGGccagaa gATG	109	hypothetical protein HDEF_1702	3-87/92	31	Candidatus *Hamiltonella defensa* 5AT (*Acyrthosiphon pisum*)	YP_002924457.1
45	21807	22121	unknown	-	AAGGAGAAAtc acATG	104	hypothetical protein M446_5978	1-72/115	35	*Methylobacterium* sp. 4-46	YP_001772689.1
46	22133	23731	unknown	-	GGAACGtggacA TG	532	hypothetical protein Bpse112_32291	69-240/282	45	*Burkholderia pseudomallei* 112	ZP_02502292.1
47	23809	26178	tail assembly protein	-	AGAGGAAGAca aATG	789	hypothetical protein HCH_05649	2-727/728	34	*Hahella chejuensis* KCTC 2396	YP_436732.1
48	26175	26375	tail assembly protein	-	GGGGGCAAgaa ATG	66	hypothetical protein HCH_05650	4-67/71	50	*Hahella chejuensis* KCTC 2396	YP_436733.1
49	26372	26608	tail assembly protein	-	GAGGActgatcAT G	78	putative transmembrane protein	7-82/82	47	*Rhodobacter* sp. SW2	ZP_05845047.1
50	26618	27418	tail assembly protein	-	AGGGGGAtcaaa caATG	266	hypothetical protein HCH_05652	1-268/269	39	*Hahella chejuensis* KCTC 2396	YP_436735.1
51	27415	29100	tail assembly protein	-	AAGAAGAtcacT TG	561	hypothetical protein HCH_05654	35-560/563	32	*Hahella chejuensis* KCTC 2396	YP_436736.1
52	29097	30158	unknown	-	GACGAGGtttgaa ATG	353	hypothetical protein D11S_2171	1-326/327	23	*Aggregatibacter actinomycetemcomitans* D11S-1	YP_003256741.1
53	30160	31122	unknown	-	GAGCGAGGcata acGTG	320	hypothetical protein XALc_0225	1-194/307	35	*Xanthomonas albilineans* GPE PC73	YP_003374757.1
54	31124	35860	tail tape measure	-	GGACTGAAcgga aATG	1578	phage tape measure protein	1-109, 452-1680/1683	33	*Sinorhizobium meliloti* AK83	YP_004548730.1
55	35853	36538	tail protein	-	AAGGGGGCGag cATG	228	pre-tape measure frameshift protein G-T	1-242/243	34	*Burkholderia* phage BcepNazgul	NP_918998.2
56	36098	36538	tail protein	-	AAGGGGGCGag cATG	146	hypothetical protein Sinme_1368	4-126/142	34	*Sinorhizobium meliloti* AK83	YP_004548729.1
57	36549	37337	unknown	-	GAGGAAtcaatcA	262	hypothetical protein	1-257/262	45	*Sinorhizobium meliloti* AK83	YP_004548728.1

					TG		Sinme_1367				
58	37385	37897	minor tail protein	-	GAGGAAAGtata ATG	170	hypothetical protein Sinme_1366	7-177/177	50	*Sinorhizobium meliloti* AK83	YP_004548727.1
59	37897	38517	unknown	-	GACGCAGGtttgc cgacATG	206	hypothetical protein Nazgul55	5-198/205	49	*Burkholderia* phage BcepNazgul	NP_918988.2
60	38514	38873	unknown	-	GAGGCGcgtgAT G	119	hypothetical protein Sinme_1364	3-120/125	38	*Sinorhizobium meliloti* AK83	YP_004548725.1
61	38886	39134	unknown	-	AAAGGAAccatc ATG	82	hypothetical protein Nazgul57	1-38/85	47	*Burkholderia* phage BcepNazgul	NP_918990.1
62	39205	40233	major capsid protein	-	AAGGAGAAAG caaaATG	342	capsid protein E	2-343/346	50	*Burkholderia* phage BcepNazgul	NP_918991.1
63	40290	40688	decorator protein	-	AGGAGAAccatc ATG	132	decorator protein D	4-123/131	49	*Burkholderia* phage BcepNazgul	NP_918992.1
64	40743	42071	prohead protease	-	AGGACCAGAA ccaATG	442	prohead protease ClpP	4-427/434	53	*Burkholderia* phage BcepNazgul	NP_918994.2
65	42068	43591	portal protein	-	GGAAccegtcgAT G	507	phage portal protein	57-554/559	59	*Staphylococcus* phage SA1	ACZ55505.1
66	43736	43960	head-tail joining protein	-	GGACAAcactAT G	74	head-tail joining protein Lambda W	13-76/76	56	*Burkholderia* phage BcepNazgul	NP_918996.1
67	44097	46076	terminase large subunit	-	AAGAcctcgATG	659	terminase large subunit TerL	44-677/677	58	*Burkholderia* phage BcepNazgul	NP_918997.2
68	46210	46803	terminase small subunit	-	GAAGGTGAtagc gATG	197	TerS	9-179/222	49	*Burkholderia* phage BcepNazgul	NP_918999.1
69	46796	46990	transcriptional regulator	-	AGGAGTAcggtA TG	64	aminoglycoside phosphotransferase	423-473/487	29	*Frankia* sp. EUN1f	ZP_06416368.1
70	47047	47736	repressor	-	GAAAGGCAAG GcagcagcATG	229	hypothetical protein Rvan_1213	14-180/242	36	*Rhodomicrobium vannielii* ATCC 17100	YP_004011581.1
71	47833	49446	helicase	-	ACGAcctcctgcgA TG	537	helicase	11-507/522	52	*Burkholderia* phage BcepNazgul	NP_919000.2
72	49443	49745	resolvase	-	GAAAGGAGGAt tcactGTG	100	conserved phage protein	15-103/108	55	*Burkholderia* phage BcepNazgul	NP_919001.2
73	49742	51796	DNA polymerase	-	ACGTcaccATG	684	hypothetical protein ORF026	48-670/683	45	*Pseudomonas* phage 73	YP_001293433.1
74	51875	52609	single-stranded DNA binding protein	-	AAAGGTGAcaaa aATG	244	conserved phage protein	4-186/198	35	*Staphylococcus* phage SA1	ACZ55548.1
75	52655	53995	Cas4 superfamily exonuclease	-	GATCctctcgaccce ATG	446	conserved phage protein	8-448/454	48	*Burkholderia* phage BcepNazgul	NP_919005.2
76	54140	54538	unknown	-	GGAGAAatcAT G	132	hypothetical protein RUMHYD_01446	1-120/122	26	*Blautia hydrogenotrophica* DSM 10507	ZP_03782010.1

77	54718	55017	Cro	+	AACGGAGAtcac aATG	99	hypothetical protein Nazgul73	5-90/97	31	*Burkholderia* phage BeepNazgul	NP_919007.1
78	55054	57534	primase	+	GGAGGGgcaAT G	826	DR0530-like primase	1-843/843	49	*Burkholderia* phage BeepNazgul	NP_919008.2

Abbreviations: RBS, ribosome binding site; aa, amino acid; % ID, percent identity.

Relatedness to other phage genomes. Despite the similarities in phenotype between KL1 and AH2, there is minimal relatedness between these two genomes. These phages share only three proteins with the same closest relative in the database: KL1 gp4 and AH2 gp41 are most similar to PA73 hypothetical protein ORF004, KL1 gp22 and AH2 gp53 are most similar to *Xanthomonas albilineans* GPE PC73 hypothetical protein XALc_0225, and KL1 gp27 and AH2 gp73 are most similar to PA73 hypothetical protein ORF026 (Tables 4-1 and 4-2). However, even these pairs of proteins are relatively dissimilar: in LALIGN alignments, KL1 gp4/AH2 gp41 can be aligned three ways, with between 20.0–29.9% identity, KL1 gp22/AH2 gp53 are aligned with 42.0% identity, and KL1 gp27/AH2 gp73 are aligned with 44.1% identity. In a Circos plot of a PROmer comparison of these phages, no regions of similarity at the protein level are observed under the parameters used (Figure 4-5A). These comparisons confirm that, regardless of the phenotypic similarities of these two phages, they are dissimilar with respect to their genetic makeup.

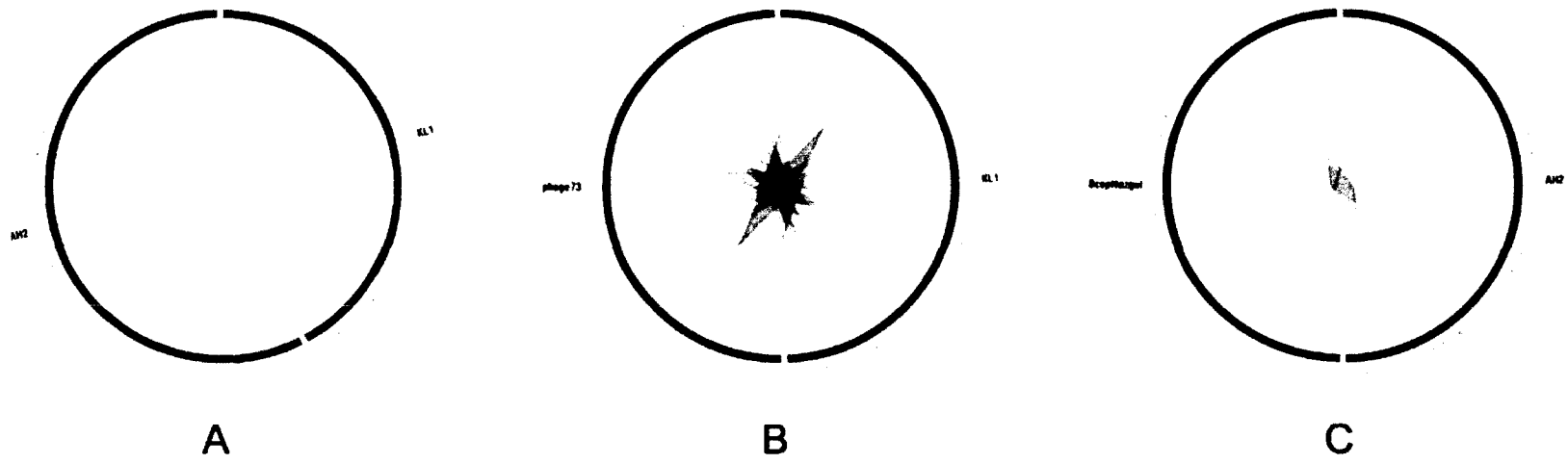


Figure 4-5: Circos plots of KL1 and AH2 PROmer comparisons. Green ribbons indicate regions of similarity between two genomes at the protein level. Each region is on the same strand in both genomes. The scale (in kbp) is shown on the periphery of the plots. PROmer parameters: breaklen = 60, maxgap = 30, mincluster = 20, minmatch = 6. A) KL1/AH2 comparison; B) KL1/*Pseudomonas* phage 73 (PA73) comparison; C) AH2/*Burkholderia* phage BcepNazgul comparison.

KL1 is most similar to PA73, a siphovirus that infects *Pseudomonas aeruginosa* (151). These phages are similar with respect to length (42,999 bp for PA73 and 42,832 bp for KL1), GC content (53.6% for PA73 and 54.6% for KL1), and predicted number of proteins (52 for PA73 and 55 for KL1). BLASTN comparison of KL1 and PA73 indicates that these sequences are similar over 69% of the KL1 genome. KL1 encodes a protein most similar to each PA73 protein from ORF001-ORF052 (excluding 12 proteins) (Table 4-1). Most PA73 proteins show limited similarity to others in the NCBI database and have not been assigned a putative function (151). Of the nine PA73 proteins with predicted functions, all but one (peptidyl-tRNA hydrolase [peptide chain release factor]) is similar to a KL1 protein: holin, terminase large subunit, head morphogenesis protein, tail tape measure protein, DNA polymerase, superfamily II helicase/restriction enzyme, helicase (annotated here as recombinase), and dCMP deaminase (KL1 gp2, gp7, gp9, gp21, gp27, gp30, gp33, and gp52, respectively) (Table 4-1). Of the KL1 proteins most similar to a PA73 protein, the most similar is gp33 (91% identity with ORF032) and the least similar is gp24 (36% identity with ORF023) (Table 4-1). In a Circos plot of a PROmer comparison of these phages, the majority of the two genomes are similar at the protein level (Figure 4-5B).

AH2 is most similar to BcepNazgul, a siphovirus isolated from soil that infects *Burkholderia ambifaria*. Like PA73 and KL1, these phages are similar with respect to length (57,455 bp for BcepNazgul and 58,065 bp for AH2), GC content (60.6% for BcepNazgul and 61.3% for AH2), and predicted number of proteins (73 for BcepNazgul and 78 for AH2). In contrast to KL1 (which is

closely related to a single phage), AH2 has a mosaic structure and encodes proteins similar to those from a variety of bacteria and phages (Table 4-2). Therefore, AH2 is less closely related to BcepNazgul than KL1 is to PA73. BLASTN comparison of AH2 and BcepNazgul indicates that these sequences are similar over 16% of the AH2 genome. Twenty-one AH2 proteins are most similar to a BcepNazgul protein and 39 show some similarity based on BLASTP analysis (Table 4-3). Of the AH2 proteins most similar to a BcepNazgul protein, the most similar is gp12 (74% identity with Nazgul10) and the least similar is gp20 (24% identity with Nazgul21) (Table 4-2). In a Circos plot of a PROmer comparison of these phages, the most similar regions at the protein level correspond to AH2 gp12, gp71, gp78 (similar to BcepNazgul Nazgul10, helicase, and DR0530-like primase, respectively) and a portion of the putative capsid morphogenesis and DNA packaging module (Figure 4-5C).

Table 4-3: AH2 BcepNazgul BLASTP homologs

AH2 protein	BcepNazgul homolog	Alignment region in BcepNazgul homolog	Percent identity	Accession number	E-value
gp1	hypothetical protein Nazgul32	12-130/130	29	NP_918966.1	0.004
gp2	none				
gp3	none				
gp4	none				
gp5	hypothetical protein Nazgul06	88-158/330	44	NP_919015.1	$4e^{-4}$
gp6	none				
gp7	none				
gp8	none				
gp9	none				
gp10	none				
gp11	hypothetical protein Nazgul09	1-129/141	59	NP_919018.1	$3e^{-48}$
gp12	hypothetical protein Nazgul10	1-151/160	74	NP_919019.2	$6e^{-75}$
gp13	none				
gp14	hypothetical protein Nazgul13	3-93/93	41	NP_919022.1	$1e^{-11}$
gp15	none				
gp16	none				
gp17	none				
gp18	hypothetical protein Nazgul34	38-179/265	39	NP_918968.1	$9e^{-25}$
gp19	none				
gp20	hypothetical protein Nazgul21	24-216/237	24	NP_918955.1	$1e^{-5}$
gp21	none				
gp22	hypothetical protein Nazgul33	216-239/244	58	NP_918967.1	1.2
gp23	none				
gp24	hypothetical protein Nazgul19	18-97/97	39	NP_919028.2	$7e^{-11}$
gp25	none				
gp26	none				
gp27	none				

gp28	possible transcription regulator	1-355/388	30	NP_918963.1	$1e^{-22}$
gp29	none				
gp30	none				
gp31	none				
gp32	none				
gp33	none				
gp34	none				
gp35	none				
gp36	none				
gp37	none				
gp38	none				
gp39	none				
gp40	hypothetical protein Rz1	7-89/98	34	YP_544112.1	$2e^{-9}$
gp41	none				
gp42	none				
gp43	holin	33-119/127	32	NP_918972.1	$2e^{-6}$
gp44	hypothetical protein Nazgul40	57-146/146	32	NP_918973.1	$4e^{-4}$
gp45	none				
gp46	none				
gp47	conserved tail assembly protein	1-740/741	31	NP_918976.1	$5e^{-110}$
gp48	conserved tail assembly protein	6-44/70	36	NP_918977.1	0.83
gp49	conserved tail assembly protein	1-78/78	38	NP_918978.1	$5e^{-9}$
gp50	conserved tail assembly protein	1-272/273	27	NP_918979.1	$1e^{-16}$
gp51	conserved tail assembly protein	9-385/838	26	NP_918980.1	$6e^{-27}$
gp52	none				
gp53	none				
gp54	tape measure protein	285-599/830	37	NP_918983.1	$1e^{-46}$
gp55	pre-tape measure frameshift protein G-T	1-242/243	34	NP_918998.2	$4e^{-29}$
gp56	pre-tape measure frameshift protein G-T	1-139/243	31	NP_918998.2	$1e^{-7}$
gp57	hypothetical protein Nazgul53	7-257/258	40	NP_918986.1	$4e^{-51}$
gp58	hypothetical protein Nazgul54	5-172/179	32	NP_918987.2	$4e^{-20}$
gp59	hypothetical protein Nazgul55	5-198/205	49	NP_918988.2	$1e^{-47}$
gp60	hypothetical protein Nazgul56	4-118/118	31	NP_918989.2	$1e^{-6}$

gp61	hypothetical protein Nazgul57	1-38/85	47	NP_918990.1	$6e^{-4}$
gp62	capsid protein E	2-343/346	50	NP_918991.1	$1e^{-110}$
gp63	decorator protein D	4-123/131	49	NP_918992.1	$1e^{-27}$
gp64	prohead protease ClpP	4-427/434	53	NP_918994.2	$9e^{-139}$
gp65	portal protein Lambda B	45-549/553	56	NP_918995.1	0.0
gp66	head-tail joining protein Lambda W	13-76/76	56	NP_918996.1	$2e^{-15}$
gp67	terminase large subunit TerL	44-677/677	58	NP_918997.2	0.0
gp68	TerS	9-179/222	49	NP_918999.1	$7e^{-56}$
gp69	none				
gp70	none				
gp71	helicase	11-507/522	52	NP_919000.2	$2e^{-176}$
gp72	conserved phage protein	15-103/108	55	NP_919001.2	$3e^{-25}$
gp73	none				
gp74	conserved phage protein	1-182/206	32	NP_919004.1	$2e^{-23}$
gp75	conserved phage protein	8-448/454	48	NP_919005.2	$4e^{-139}$
gp76	none				
gp77	hypothetical protein Nazgul73	5-90/97	31	NP_919007.1	$4e^{-8}$
gp78	DR0530-like primase	1-843/843	49	NP_919008.2	0.0

Module analysis.

Overview: We have identified the proteins encoded by KL1 and AH2 as belonging to four different functional categories: virion morphogenesis (including capsid morphogenesis/DNA packaging and tail morphogenesis), lysis, DNA binding (the largest and broadest category), and MazG (Table 4-4). For both phages, putative functions exist for which we have not yet identified the corresponding protein. For example, although both phages are siphoviruses that are predicted to encode a major tail protein and likely a tail fiber protein, the genes for these products have not yet been identified (Table 4-4). These genes are presumably found among the many encoding hypothetical proteins for which putative functions have not been assigned (Tables 4-1 and 4-2). Although the proteins encoded by each phage perform many of the same functions (e.g. both KL1 gp11 and AH2 gp62 are predicted to be major capsid proteins [Table 4-4]), the proteins themselves are dissimilar. As we discuss below, the finding that KL1 and AH2 can maintain nearly identical phenotypes with two dissimilar sets of proteins is compelling evidence for convergent evolution occurring in these Bcc-specific phages.

Table 4-4: Functional categories of KL1 and AH2 proteins

Functional category	Protein function	Putative KL1 protein	Putative AH2 protein
Virion morphogenesis: capsid morphogenesis and DNA packaging	major capsid protein	gp11	gp62
	decorator protein		gp63
	prohead protease		gp64
	portal protein	gp8	gp65
	head-tail joining protein	gp15	gp66
	terminase large subunit	gp7	gp67
	terminase small subunit		gp68
	head morphogenesis protein	gp9	
Virion morphogenesis: tail morphogenesis	minor tail protein	gp16	gp58
	tail tape measure	gp21	gp54
	frameshifted tail protein	gp18/gp19	gp55/gp56
	tail assembly protein	gp24, gp25, gp26	gp47, gp48, gp49, gp50, gp51
	tail protein	gp20	
Lysis	holin	gp2	gp43
	endolysin	gp3	gp42
	Rz	gp4	gp41
	Rz1	gp5	gp40
DNA binding	DNA polymerase	gp27	gp73
	DNA polymerase III B subunit	gp28	gp20
	superfamily II helicase/restriction enzyme	gp30	
	exonuclease	gp31	gp75
	recombinase	gp33	
	transcriptional regulator	gp36	gp28, gp30, gp31, gp69
	primase	gp37	gp78
	Vsr endonuclease	gp44	gp32
	dCMP deaminase	gp52	
	excinuclease		gp33
	restriction endonuclease		gp34
	cytosine methyltransferase		gp35

	integrase		gp37
	repressor		gp70
	helicase		gp71
	resolvase		gp72
	single-stranded DNA binding protein		gp74
	Cro		gp77
MazG	pyrophosphohydrolase	gp35	gp25

Virion morphogenesis: Although we have determined that KL1 is a siphovirus (Figure 4-2A), the identity of many of the structural genes remains unknown. As discussed above, KL1 is most closely related to PA73, a phage whose proteins have largely uncharacterized functions. Based on BLASTP analysis, we have been able to predict the identity of only eight KL1 structural proteins: three involved in capsid morphogenesis and DNA packaging and five involved in tail morphogenesis. Gp7 (terminase large subunit) and gp9 (head morphogenesis protein) are similar to PA73 ORF006 and ORF008, respectively, both of which have been assigned putative functions in the PA73 annotation (Table 4-1). Gp11 (major capsid protein) is similar to the major capsid proteins of *Escherichia* phage K1H and *Listonella* phage ϕHSIC. Gp20 is similar to tail proteins from multiple *Escherichia* phages including K1G, K1H, and K1ind1-K1ind3. Gp21 is predicted to be the tail tape measure as it is the largest protein encoded by KL1 (1272 amino acids [aa]) and it is similar to the predicted PA73 tape measure protein ORF020 (Table 4-1). Finally, gp24-gp26 are similar to BcepNazgul tail assembly proteins. Using HHpred analysis, we were able to identify an additional three proteins at a probability threshold of 75%. Gp8 is similar to bacteriophage SPP1 portal protein (99.44% probability), gp15 is similar to λ gpFII head-tail joining protein (82.86% probability), and gp16 is similar to λ gpU minor tail protein (77.70% probability) (Table 4-5).

Table 4-5: KL1 HHpred predictions

Protein	Closest relative	Probability (%)	Motif	Source
gp1	Fibrinogen alpha chain	84.48	3ghg_A	*Homo sapiens*
gp2	CE-FAR-7, fatty acid/retinol binding protein protein 7, isoform A	26.14	2w9y_A	*Caenorhabditis elegans*
gp3	Hypothetical protein NMB1012	100.00	2ikb_A	*Neisseria meningitidis* MC58
gp4	Hypothetical protein	47.27	2o2z_A	*Bacillus halodurans*
gp5	AXE2A, CJCE2B, putative acetyl xylan esterase	54.01	2w9x_A	*Cellvibrio japonicus*
gp6	Phikz144, lytic transglycosylase	60.54	3bkh_A	*Pseudomonas* phage ϕKZ
gp7	G2P, terminase large subunit	99.93	2wbn_A	*Bacillus* phage SPP1
gp8	Portal protein	99.44	2jes_A	Bacteriophage SPP1
gp9	Cytochrome C-type biogenesis protein CCMH	27.60	2hl7_A	*Pseudomonas aeruginosa*
gp10	AGI/II, PA	42.97	3iox_A	*Streptococcus mutans*
gp11	Major capsid protein	97.56	1ohg_A	Bacteriophage HK97
gp12	Imidazolonepropionase	54.92	2puz_A	*Agrobacterium tumefaciens* str
gp13	S-SEC1	20.83	1epu_A	*Loligo pealei*
gp14	Hypothetical protein YQBG	92.97	1xn8_A	*Bacillus subtilis*
gp15	Gpfii	82.86	1k0h_A	Enterobacteria phage lambda
gp16	Uncharacterized protein	99.96	2l25_A	*Bordetella bronchiseptica*
gp17	Transthyretin-like protein PUCM	8.76	2h0e_A	*Bacillus subtilis*
gp18	Uncharacterized protein YQBN	61.50	3klu_A	*Bacillus subtilis*
gp19	Uncharacterized protein YQBN	67.45	3klu_A	*Bacillus subtilis*
gp20	D-tyrosyl-tRNA(Tyr) deacylase	22.47	1j7g_A	*Haemophilus influenzae* rd KW20
gp21	Methyl-accepting chemotaxis protein	96.85	2ch7_A	*Thermotoga maritima*
gp22	Chitinase A1, (CHBD-CHIA1)	29.49	1ed7_A	*Bacillus circulans*
gp23	ZF-HD homeobox family protein	55.37	1wh7_A	*Arabidopsis thaliana*
gp24	Hydin protein	81.28	2e6j_A	*Homo sapiens*
gp25	Collagen type-I A1 chain	52.78	3ejh_E	*Homo sapiens*

gp26	Tail protein, 43 kDa	97.01	3d37_A	*Neisseria meningitidis* MC58
gp27	DNA polymerase	100.00	2gv9_A	Human herpesvirus 1 strain kos
gp28	DNA polymerase III subunit beta	98.71	3d1g_A	*Escherichia coli*
gp29	DNA-directed RNA polymerase RPO12 subunit	58.55	2waq_P	*Sulfolobus shibatae*
gp30	RECQ helicase, ATP-dependent DNA helicase	100.00	1oyw_A	*Escherichia coli*
gp31	RED alpha, lambda exonuclease	99.21	1avq_A	Enterobacteria phage lambda
gp32	SM B	85.48	3pgw_B	*Homo sapiens*
gp33	DNA repair and recombination protein RADB	98.66	2cvh_A	*Thermococcus kodakarensis*
gp34	ALDH, aldehyde dehydrogenase	44.15	1ez0_A	*Vibrio harveyi*
gp35	Protein MAZG	100.00	3cra_A	*Escherichia coli*
gp36	Transcriptional regulator, LACI family	95.68	3kjx_A	*Silicibacter pomeroyi*
gp37	ORF904	96.05	3m1m_A	*Sulfolobus islandicus*
gp38	Hypoxanthine-guanine phosphoribosyltransferase	23.48	3acd_A	*Thermus thermophilus*
gp39	Exosome component 10	48.54	2cpr_A	*Homo sapiens*
gp40	D-ornithine aminomutase S component	27.50	3kp1_E	*Clostridium sticklandii*
gp41	Urease beta subunit	26.38	1ejx_B	*Klebsiella aerogenes*
gp42	COP associated protein	72.76	1yg0_A	*Helicobacter pylori* 26695
gp43	GDP-mannose mannosyl hydrolase	52.18	1rya_A	*Escherichia coli*
gp44	Protein (VSR endonuclease)	78.74	1vsr_A	*Escherichia coli* K12
gp45	NADH-dependent nitrate reductase	43.63	2cnd_A	*Zea mays*
gp46	Repressor protein CI	47.83	1kca_A	Enterobacteria phage lambda
gp47	Ribosome-associated factor Y	87.21	3lyv_A	*Streptococcus pyogenes*
gp48	Crustacyanin A2 subunit	54.97	1gka_B	*Homarus gammarus*
gp49	Hypothetical protein	47.38	1zso_A	*Plasmodium falciparum*
gp50	LEMA protein	57.81	2etd_A	*Thermotoga maritima* MSB8
gp51	Putative polysaccharide binding proteins (DUF1812)	33.32	3gf8_A	*Bacteroides thetaiotaomicron* vpi-5482

gp52	Deoxycytidylate deaminase	100.00	2hvw_A	*Streptococcus mutans*
gp53	Hypothetical protein PA0269	69.12	2o4d_A	*Pseudomonas aeruginosa*
gp54	Zinc resistance-associated protein	86.95	3lay_A	*Salmonella enterica* subsp
gp55	E3 ubiquitin-protein ligase UHRF1	44.24	3db3_A	*Homo sapiens*

In comparison with KL1, the structural proteins of AH2 are well-defined. Genes *62-68* make up the capsid morphogenesis and DNA packaging module, containing genes encoding the major capsid protein, decorator protein, prohead protease, portal protein, head-tail joining protein, and terminase subunits (large and small) (Table 4-2). Each of these proteins is similar to a BcepNazgul protein, with percent identities between 49-58% (Table 4-3). Several genes between *47* and *56* are similar to genes encoding BcepNazgul conserved tail assembly proteins, tape measure protein, and pre-tape measure frameshift protein G-T (with percent identities between 26-38%) (Table 4-3). An additional AH2 tail protein was identified using HHpred analysis (gp58, similar to λ gpU minor tail protein [88.10% probability]) (Table 4-6). Hypothetical proteins encoded in this region are likely to be involved in tail morphogenesis based on the proximity of their genes to this module.

Table 4-6: AH2 HHpred predictions

Protein	Closest relative	Probability (%)	Motif	Source
gp1	Hydroquinone glucosyltransferase	35.45	2vch_A	*Arabidopsis thaliana*
gp2	N-acetylmuramoyl-L-alanine amidase XLYA	40.29	3hma_A	*Bacillus subtilis*
gp3	Colicin IA	50.72	1cii_A	*Escherichia coli*
gp4	Aerotaxis transducer AER2	37.48	3lnr_A	*Pseudomonas aeruginosa*
gp5	Alpha-hemoglobin stabilizing protein	17.95	1z8u_A	*Homo sapiens*
gp6	STAL	30.13	2ov8_A	*Streptomyces toyocaensis*
gp7	Sirohydrochlorin cobaltochelatase	29.29	2xwp_A	*Salmonella enterica*
gp8	Uncharacterized protein	43.25	3fvv_A	*Bordetella pertussis*
gp9	REBM, methyltransferase	70.81	3bus_A	*Lechevalieria aerocolonigenes*
gp10	Dihydrodipicolinate reductase	28.60	1p9l_A	*Mycobacterium tuberculosis*
gp11	Single domain haemoglobin	51.19	2wy4_A	*Campylobacter jejuni*
gp12	Gamma-glutamyl cyclotransferase	31.17	3cry_A	*Homo sapiens*
gp13	Uncharacterized protein	60.14	2k5e_A	*Methanococcus jannaschii*
gp14	Hypothetical protein AGR_PAT_315	24.40	1vqy_A	*Agrobacterium tumefaciens* str
gp15	DNA-directed RNA polymerase III 39 kDa polypeptide	30.05	2dk5_A	*Homo sapiens*
gp16	Protein kinase C, NU type	53.00	2d9z_A	*Homo sapiens*
gp17	PUUE allantoinase	40.51	3cl6_A	*Pseudomonas fluorescens*
gp18	Putative cystein deoxygenase	92.50	3eqe_A	*Bacillus subtilis*
gp19	Spectrin alpha chain, brain	27.41	1u5p_A	*Gallus gallus*
gp20	DNA polymerase III, beta subunit	90.54	1vpk_A	*Thermotoga maritima* MSB8
gp21	Heterochromatin protein 1-binding protein 3	45.74	2rqp_A	*Homo sapiens*
gp22	Glycogen synthesis protein GLGS	60.01	1rrz_A	*Escherichia coli*
gp23	Uncharacterized protein SPOA0173	64.51	3eyt_A	*Silicibacter pomeroyi*
gp24	3-methyl-2-oxobutanoate hydroxymethyltransferase	76.36	1o66_A	*Neisseria meningitidis* serogroup B
gp25	Protein MAZG	99.76	3cra_A	*Escherichia coli*
gp26	ATP-dependent DNA ligase, N-terminal domain prote	57.99	3p4h_A	Candidatus korarchaeum cryptofilum opforganism_taxid

gp27	Thiol:disulfide interchange protein DSBA	41.73	3hd5_A	*Bordetella parapertussis*
gp28	PARB, chromosome partitioning protein PARB	100.00	1vz0_A	*Thermus thermophilus*
gp29	Oxidoreductase, GFO/IDH/MOCA family	67.39	3i23_A	*Enterococcus faecalis*
gp30	TORI, TOR inhibition protein	98.09	1z4h_A	*Escherichia coli*
gp31	XIS, excisionase from transposon TN916	95.91	1y6u_A	*Enterococcus faecalis*
gp32	Protein (VSR endonuclease)	100.00	1vsr_A	*Escherichia coli* K12
gp33	Uvrabc system protein C	98.35	1yd0_A	*Thermotoga maritima*
gp34	Putative lipase/esterase	55.34	3bxp_A	*Lactobacillus plantarum* WCFS1
gp35	Cytosine-specific methyltransferase	100.00	3g7u_A	*Escherichia coli* O157
gp36	Endo-1,4-B-D-mannanase	60.93	1qnr_A	*Trichoderma reesei*
gp37	Integrase	100.00	1z1b_A	Enterobacteria phage lambda
gp38	Microtubule-associated protein RP/EB family member 1	83.86	2qjz_A	*Homo sapiens*
gp39	Lectin	52.04	2bt9_A	*Ralstonia solanacearum*
gp40	RLF, RLF-RBD	35.11	1rlf_A	*Mus musculus*
gp41	Cucurmosin	31.60	3bwh_A	*Cucurbita moschata*
gp42	Australian black SWAN egg white lysozyme	90.90	1gbs_A	*Cygnus atratus*
gp43	CE-FAR-7, fatty acid/retinol binding protein protein 7, isoform A, confirmed by transcript...	46.11	2w9y_A	*Caenorhabditis elegans*
gp44	Homer protein homolog 1	75.30	3cve_A	*Rattus norvegicus*
gp45	Geminin	77.96	1uii_A	*Homo sapiens*
gp46	SPRY domain-containing protein 3	99.85	2yyo_A	*Homo sapiens*
gp47	Prophage MUSO2, 43 kDa tail protein	97.10	3cdd_A	*Shewanella oneidensis* mr-1
gp48	CCR4-NOT transcription complex subunit 7	85.46	2d5r_A	*Homo sapiens*
gp49	Glutathione peroxidase-like protein	50.86	2vup_A	*Trypanosoma brucei*
gp50	Tail protein, 43 kDa	80.25	3d37_A	*Neisseria meningitidis* MC58
gp51	Hydrocephalus-inducing protein homolog	57.73	2ys4_A	*Homo sapiens*
gp52	ZF-HD homeobox family protein	33.02	1wh7_A	*Arabidopsis thaliana*
gp53	Chitinase A1, (CHBD-CHIA1)	28.59	1ed7_A	*Bacillus circulans*
gp54	Methyl-accepting chemotaxis protein	97.16	2ch7_A	*Thermotoga maritima*
gp55	PAS factor	31.46	2b8i_A	*Vibrio vulnificus*
gp56	PAS factor	61.55	2b8i_A	*Vibrio vulnificus*
gp57	MSRB, peptide methionine sulfoxide reductase	18.36	3hcj_A	*Xanthomonas campestris* PV

gp58	Minor tail protein U	88.10	3fz2_A	Enterobacteria phage lambda
gp59	General control protein GCN4	40.60	2wq1_A	*Saccharomyces cerevisiae*
gp60	Conserved protein	53.65	2kbn_A	*Methanosarcina mazei*
gp61	Cytochrome C5	59.68	1cc5_A	*Azotobacter vinelandii*
gp62	Putative capsid protein of prophage	100.00	3bqw_A	*Escherichia coli* CFT073
gp63	Head decoration protein	99.94	1td4_A	Enterobacteria phage P21
gp64	Protease 4	100.00	3bf0_A	*Escherichia coli*
gp65	Portal protein	98.33	2jes_A	Bacteriophage SPP1
gp66	GPW, head-TO-tail joining protein W	99.57	1hyw_A	Enterobacteria phage lambda
gp67	Terminase, DNA packaging protein GP17	98.70	2o0j_A	Enterobacteria phage T4
gp68	GPNU1 DBD	97.58	1j9i_A	Enterobacteria phage lambda
gp69	TORI, TOR inhibition protein	84.08	1z4h_A	*Escherichia coli*
gp70	Repressor protein CI	99.87	2fjr_A	Enterobacteria phage 186
gp71	Similar to RAD54-like	100.00	1z3i_X	*Danio rerio*
gp72	HJC	97.87	2wcw_A	*Archaeoglobus fulgidus*
gp73	DNA polymerase delta catalytic subunit	100.00	3iay_A	*Saccharomyces cerevisiae*
gp74	Helix-destabilizing protein	99.83	1je5_A	Enterobacteria phage T7
gp75	Putative exonuclease	99.79	3l0a_A	*Eubacterium rectale*
gp76	50S ribosomal protein L20	25.67	2ghj_A	*Aquifex aeolicus*
gp77	CRO protein	96.60	3bd1_A	*Xylella fastidiosa* ann-1
gp78	Hypothetical protein ORF904	99.83	1ro2_A	*Sulfolobus islandicus*

As discussed in Chapters 2 and 3, most tailed phages encode two tail proteins proximal to the tail tape measure gene by way of a -1 translational frameshift (324). We have previously identified these frameshifted genes in the Bcc-specific phages KS9, KS5, KS14, and KL3 (176, 177). Using FSFinder and manual scanning for XXXYYYZ motifs, we predict that KL1 gp18/gp19 and AH2 gp55/gp56 are expressed using this mechanism. The predicted frameshift site in KL1 is GGGAAAC, immediately upstream of the gp18 TGA stop codon (Figure 4-6). A -1 ribosomal shift following the terminal C will allow for expression of the 264 aa gp19 and the 142 aa gp18 from the same start codon (Figure 4-6). Although most phages encode their frameshifted proteins immediately upstream of the tail tape measure gene, KL1 encodes an intervening tail protein, gp20 (Table 4-1, Figure 4-4). This organization is similar to that of *Escherichia coli* phage HK97, *Bacillus subtilis* phage SPP1, *Methanobacterium thermoautotrophicum* phage ψM2, *Methanothermobacter wolfei* phage ψM100, *Lactococcus* phages c2 and BIL67, and *Natrialba magadii* phage ϕch1 (324). The predicted frameshift site in AH2 is AAAAAAG, the same sequence used by *E. coli* phage VT1-Sakai, *M. thermoautotrophicum* phage ψM2, *Staphylococcus aureus* phages PVL and PV83, *Lactococcus lactis* phage ul36, and *Borrelia burgdorferi* prophage Borreliapro (Figure 4-6) (324). In the case of AH2, a -1 shift of the ribosome following the G in this sequence will allow for the 228 aa gp55 to be expressed instead of the 146 aa gp56 (Figure 4-6). Although we were unable to identify the KL1 or AH2 major tail genes by either BLASTP or HHpred searches, as these genes are generally positioned upstream of the frameshifted

protein genes, we predict that the major tail proteins may be gp17 in KL1 and gp57 in AH2 (324).

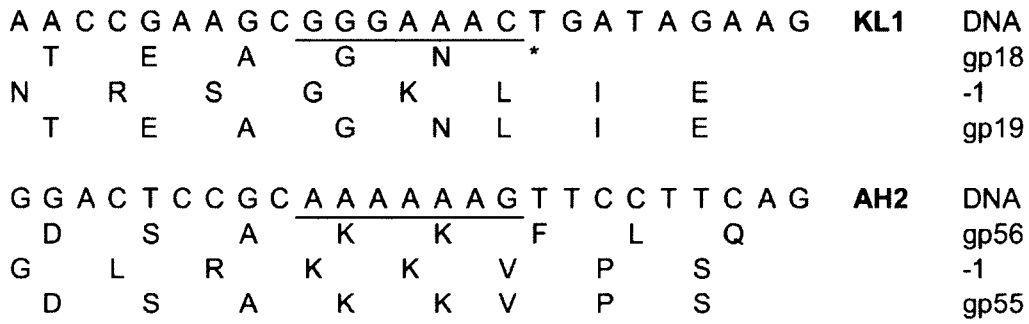


Figure 4-6: Sequence of the KL1 and AH2 predicted translational frameshift sites. For each phage, the first row shows the DNA sequence (with the predicted frameshift site underlined); the second row shows the amino acid sequence in the original frame (the KL1 gp18 stop codon is represented by an asterisk); the third row shows the amino acid sequence in the -1 frame; the fourth row shows the amino acid sequence of the frameshifted protein.

Lysis: In KL1, we have identified the genes putatively encoding the holin, lysin, Rz, and Rz1 lysis proteins. In a BLASTP search, gp2 shows similarity to putative holin proteins of PA73 and BcepNazgul. TMHMM analysis of this protein indicates that it has two transmembrane domains, so gp2 is predicted to be a class II holin (328). Gp3 is similar to the endolysin of *Erwinia* phage vB_EamP_S6 and contains lysozyme and peptidoglycan-binding conserved domains. Although gp4 does not show similarity to any Rz proteins in the NCBI database, it is predicted to contain a single N-terminal transmembrane domain, characteristic of Rz proteins (285). Gp5 is predicted to be the KL1 Rz1 protein as it is similar to BcepNazgul Rz1. LipoP analysis identifies a signal peptidase II cleavage site between positions 17 (serine) and 18 (cysteine), resulting in a 70 aa protein with four proline residues (5.7% proline). The proportion of prolines in the predicted Rz1 lipoprotein is low compared to previously identified Rz1

proteins in Bcc phages (176, 177, 288).

The same lysis proteins were identified in AH2. Like KL1 gp2, the putative AH2 holin gp43 is similar to the BcepNazgul holin, has two transmembrane domains, and is predicted to be a class II holin. Although gp42 shows no similarity to endolysins in a BLASTP search, HHpred analysis reveals similarity to both eukaryotic and prokaryotic lysozyme proteins. Gp41 is predicted to be the AH2 Rz protein as it has a single N-terminal transmembrane domain. Although manual annotation has been required for identification of the Rz1 gene in KL1 and in our previous studies (176, 177), we predict that the GeneMark-assigned gp40 is the AH2 Rz1 protein. Gp40 is similar to BcepNazgul Rz1 and it has a signal peptidase II cleavage site between amino acids 15 (alanine) and 16 (cysteine) as expected. Similar to the predicted KL1 Rz1, the proportion of prolines present in this protein is relatively low (3/61 or 4.9%). It is unclear from this analysis what protein (or proteins) is responsible for the unique lysis phenotype observed in both of these phages. Aside from the low proportion of proline found in the putative Rz1 proteins, KL1 and AH2 appear to have relatively standard lysis modules, suggesting that unique (and as yet unidentified) proteins may be responsible for controlling lysis timing in each phage.

DNA binding: Of the 8 KL1 proteins similar to a PA73 protein with an assigned function, half of these are DNA- or nucleotide-binding proteins: DNA polymerase (gp27), superfamily II helicase/restriction enzyme (gp30), helicase (annotated here as recombinase [gp33]), and dCMP deaminase (gp52) (Table 4-1). In addition, KL1 encodes a putative DNA polymerase III β subunit (gp28),

exonuclease (gp31), transcriptional regulator (gp36), primase (gp37), and Vsr endonuclease (gp44) (Tables 4-1 and 4-5). In a multi-genome analysis performed by Lopes et al. (167), it was determined that PA73 ORF032 is distantly related to *Lactococcus* phage ϕ31 Sak4 recombinase. When this protein was expressed in *E. coli*, it exhibited recombinase activity, but was found to be less efficient than λ Redβ (167). Furthermore, PA73 encodes an exonuclease, as is found in characterized phage recombinase pairs such as Redαβ in λ and RecET in *rac* (167). KL1 gp33 is most closely related to PA73 ORF032 and, with 91% identity, is the KL1 protein most similar to a PA73 protein. In addition, KL1 gp31 has 65% identity with PA73 ORF030 and both of these proteins are similar to λ Redα (99.21% probability for gp31 and 99.17% probability for ORF030) (Tables 4-1 and 4-5). It is interesting to note that, despite the relatively limited similarity between KL1 and other previously sequenced Bcc-specific phages, both gp31 and gp33 are similar to proteins from *Burkholderia* phage BcepGomr (BcepGomrgp43 and BcepGomrgp45, respectively) (167). Although further characterization of these proteins is required in both KL1 and BcepGomr, it is possible that these exonucleases and Sak4-like recombinases represent a conserved recombination system in certain Bcc-specific phages.

AH2 encodes DNA replication, modification, and repair proteins including a putative DNA polymerase III β subunit (gp20), Vsr endonuclease (gp32), excinuclease (gp33), restriction endonuclease/methylase pair (gp34/gp35), integrase (gp37), helicase (gp71), resolvase (gp72), DNA polymerase (gp73), single-stranded DNA binding protein (gp74), Cas4 superfamily exonuclease

(gp75), and primase (gp78) (Table 4-2). Other putative DNA binding proteins are predicted to be involved in transcriptional regulation. Gp28 is similar to partitioning and regulation proteins from *Thermus thermophilus* (100% probability) and *E. coli* (99.86% probability) (Table 4-6). The gp30 and gp31 predicted proteins belong to the helix-turn-helix MerR superfamily and the pyocin activator superfamily, respectively. Both of these proteins, in addition to gp69, also show similarity to excisionase proteins (Table 4-6). Gp70 and gp77 are similar to the lysogeny control proteins CI from enterobacteria phage 186 (99.87% probability) and Cro from *Xylella fastidiosa* Ann-1 (96.60% probability), respectively (Table 4-6).

AH2 gp32-gp35 are predicted to be part of a DNA protection and repair module. Vsr (very short patch repair) endonucleases are involved in the repair of 5-methylcytosine to thymine deamination (117). As discussed in Chapter 3, we identified a Vsr endonuclease in the Bcc-specific phage KL3 that, along with an EcoRII-C endonuclease/methylase pair, was predicted to be part of a novel non-self DNA degradation and self DNA protection/repair module (177). Our model proposed that non-KL3 DNA (i.e. that of the host or a superinfecting phage) would be degraded by the endonuclease (gp45), while KL3 DNA would be protected by the methylase (gp47) (converting cytosine to 5-methylcytosine). Vsr endonuclease (gp46) and very short patch repair would then prevent the accumulation of mutations caused by 5-methylcytosine deamination (177).

The DNA protection and repair system of AH2 is analogous to that of KL3. AH2 gp32 has 51% identity with the KL3 Vsr endonuclease and is similar

to *E. coli* Vsr endonuclease (100% probability) (Table 4-6). AH2 also encodes an endonuclease/methylase pair: gp34 is similar to *Kluyvera ascorbata* KasI (64% identity), while gp35 is similar to *K. ascorbata* M.KasI, *Brevundimonas diminuta* ATCC 11568 cytosine-specific methyltransferase NlaX, and *Acetobacter pomorum* DM001 modification methylase HpaII (63-66% identity). Gp35 also has several methylase conserved domains, including Dcm (an enzyme that produces 5-methylcytosine bases at sites recognized by Vsr endonuclease) (117). Gp33 is similar to *Thermotoga maritima* UvrABC system protein C (98.35% probability) and could function together with UvrAB in nucleotide excision repair (Table 4-6) (161). Although further experiments are required to identify the recognition sites of gp34 and gp35, we predict that this module functions as follows: gp34 cleaves non-self DNA, while self DNA is protected by gp35 methylation and subsequent gp32/gp33 repair. Although the identity and arrangement of genes in this module is different in AH2 than in KL3, the identification of a similar module in an unrelated Bcc-specific phage suggests that these genes may be widely used for DNA protection and repair in this group of phages.

MazG: A notable protein encoded by both KL1 and AH2 is MazG. MazG is a pyrophosphohydrolase that acts on ppGpp, one of the signaling molecules in bacteria produced during the stringent response (99). When bacterial cells are in an amino acid-limited environment, RelA synthesizes pppGpp, the precursor of ppGpp, and the latter activates the expression of genes required for cell survival (such as *rpoS*) and represses genes required for protein and DNA synthesis (181). Recently, there has been a great deal of interest in marine phages (especially

cyanophages) that encode MazG homologs, such as *Prochlorococcus* phages P-SSM2 and P-SSM4, *Synechococcus* phage S-PM2, *Prochlorococcus* and *Synechococcus* phage Syn9, *Roseobacter* phage SIO1, *Pseudoalteromonas* phage H105/1, almost one-fifth of the cyanophages tested by Bryan et al., and all of the cyanophages analyzed by Sullivan et al. (9, 32, 75, 190, 282, 283, 315). It has been suggested that these MazG-encoding phages are better able to infect and propagate within their hosts, which are found in nutrient-limited water. By inactivating ppGpp, these phages can promote the expression of genes that would usually be expressed by an exponential phase cell under nutrient-rich conditions, such as those required for protein and DNA synthesis (55). There are few published reports of the *mazG* gene in non-marine phages, but it has been previously identified in *Myxococcus* phage Mx8 and mycobacteriophage L5 (32).

The putative MazG proteins encoded by KL1 and AH2 are gp35 and gp25, respectively. KL1 gp35 is similar to putative MazG proteins from phages infecting *Synechococcus* (including S-CRM01, S-SM2, and S-ShM2), *Prochlorococcus* (including P-HM1, P-HM2, and P-SSM2), and *Bacillus* (0305ϕ8-36), as well as to PA73 hypothetical protein ORF034 (Table 4-1). AH2 gp25 is similar to putative *Clostridium* MazG proteins and to the *Burkholderia* phage proteins ϕE255 gp37, BcepMu gp06, and BcepB1A gp71. Both gp35 and gp25 are similar to *E. coli* MazG (100% and 99.76% probability, respectively) (Tables 4-5 and 4-6). Because Bcc bacteria found in soil and water are likely to be nutrient-limited (similar to cyanobacteria), MazG proteins in Bcc-specific phages may help to facilitate infection in the environment. This protein may also be

involved in the unique lysis phenotype of these phages, as the appearance of plaques at low titre after >16 h incubation (at which time the bacterial lawn appears intact) (Figure 4-1, top row) suggests that lysis of stationary phase cells may be occurring.

MazG may also have an effect with respect to Bcc pathogenicity. Synthesis of ppGpp has been associated with virulence in species such as *Legionella*, *Listeria*, *Pseudomonas*, *Salmonella*, *Mycobacterium*, and *Vibrio* (although the association in this species has been controversial) (79, 107, 110, 132, 144, 273, 292). For example, in *P. aeruginosa*, *relA* mutants are less virulent than the wild-type when tested in the *Drosophila melanogaster* model (79). Because MazG activity would mimic the effects of a *relA* mutation, it is possible that phage-encoded MazG could inhibit the virulence of a lysogen. Further experiments are required to determine if the putative KL1 and AH2 MazG proteins have pyrophosphohydrolase activity, if these genes are expressed in the lysogen, and if MazG expression has an effect on host cell virulence.

Convergent evolution. Although there have been relatively few papers published on the subject, the occurrence of convergent evolution in phages has been documented previously. Most studies examine the phenomenon at the molecular level by identifying identical base pair and amino acid changes that occur in different phage lineages under the same environmental conditions (25, 33, 317, 318). Structural examples of convergent evolution, such as the *Caudovirales* tail and the tectivirus pseudo-tail, have been reviewed previously (5). Given the ever-increasing number of completed phage genome sequences, it

is expected that many more examples remain to be identified (particularly at the whole genome level). Furthermore, there are likely many examples in the literature of phages with similar phenotypes but dissimilar genomes that have not explicitly been identified as examples of convergent evolution, perhaps because they exhibit what is considered to be a "standard" lysis phenotype.

Here, we have identified KL1 and AH2 as examples of phage convergent evolution at the whole genome level. As discussed above, these two phages exhibit a lysis phenotype that is both similar and unique in comparison to all other Bcc-specific phages that we have characterized previously (with respect to host range and lysis timing). Because of these characteristics, KL1 and AH2 were initially thought to be the same phage (prior to RFLP and genomic analysis). However, KL1 and AH2 appear to have convergently evolved because, as discussed above, their genomes are almost entirely dissimilar (Figure 4-5A).

Given the relative rarity of this phenotype, it is important to consider how it may be favorable in an evolutionary sense. The turbid plaque morphology suggests that these phages are temperate. It is well-established that a temperate lifestyle is beneficial because it protects the phage from conditions that could be damaging to a free phage particle (205) and it increases the effective burst size (as following cell division, each of the daughter cells of the original lysogen can undergo the lytic cycle) (2). Consistent with these advantages, lysogeny is extremely prevalent among Bcc-specific phages (89, 93, 152, 176, 177, 265, 288).

The evolutionary benefits of a narrow host range and delayed lysis phenotype are somewhat less apparent. It has been proposed that phages with a

narrow host range could be at an advantage because they have greater specificity and can better interact with host receptors (2): the tail fibers of a narrow host range phage presumably have optimized binding to a small set of receptors, while a broad host range phage would have less optimized binding to a larger set. As a direct consequence of this narrow host range, these phages would have relatively fewer hosts in their environment and would need to use each one more effectively (2). Because delayed lysis is correlated with a larger burst size (2), these phages can use a combination of lysogeny and delayed lysis to increase the total number of phages produced from a small number of hosts. These predictions are consistent with the high titres obtained during KL1 and AH2 propagation (up to 10^{11} PFU/ml).

Given the potential benefits of these adaptations, it remains puzzling as to why this particular phenotype is not more commonly observed. A likely (albeit simple) explanation is sampling bias. Standard phage isolation protocols most readily identify those phages that have easily visible plaques on a broad range of hosts after overnight incubation. Phages such as KL1 and AH2 may be missed because of poorly visible plaques, incompatible hosts, insufficient incubation times, titres that are too high or too low, overgrowth of bacteria, and/or competition by more rapidly lysing phages. As novel phages continue to be isolated from environmental samples using diverse bacterial hosts, the prevalence, distribution, and genetic basis of this phenotype should become more apparent.

CONCLUSIONS

A recent publication by Ceyssens et al. (49) provides an interesting counterpoint to our study. While we identified KL1 and AH2 as phages that were phenotypically similar but genetically dissimilar, this group analyzed a set of *Pseudomonas* phages that were phenotypically dissimilar but genetically similar. They found that, among ϕKMV-like viruses with between 83-97% nucleotide identity, there were significant differences observed with respect to latent period, host range, and antibody reactivity (49). We have made similar observations with our collection of Bcc-specific phages: as discussed in Chapter 6, two phages can have distinct phenotypes with respect to liquid clearing and host range while at the same time having almost identical genomes (252, 267). Taken together, the observations made by Ceyssens et al. (49) and those discussed in this study provide a) novel examples of both divergent and convergent phage evolution and b) further evidence of the broad diversity of phages that infect Gram-negative opportunistic pathogens.

ACKNOWLEDGEMENTS

The authors would like to thank Amberlie Heaman for isolation and preliminary host range analysis of AH2, Kimberley Seed for development of the DNA isolation protocol, Miles Peterson for assistance with figure construction, and Arlene Oatway (University of Alberta Department of Biological Sciences Advanced Microscopy Facility) for assistance with electron microscopy.

Chapter 5

Characterization of DC1, a broad host range Bcep22-like podovirus

A version of this chapter has been published as:

Lynch, K. H., P. Stothard, and J. J. Dennis. In press. Characterization of DC1, a broad host range Bcep22-like podovirus. Appl. Environ. Microbiol.

OBJECTIVES

The objectives of this project were to sequence and characterize the genome of the Bcep22-like podovirus DC1 (vB_BceP_DC1) and to identify the protein(s) responsible for the putative exopolysaccharide (EPS) depolymerase activity of this phage.

Note: Gill et al. (89) published a full-length manuscript in October 2011 that a) describes the sequencing and characterization of the *Burkholderia cenocepacia*-specific podoviruses Bcep22 and BcepIL02 and b) establishes the "Bcep22-like" phage type, to which both of these viruses belong. As discussed in this chapter, Bcep22 and BcepIL02 are very closely related to DC1. As a result, the manuscript upon which this chapter is based has been formatted as a short-form paper.

MATERIALS AND METHODS

DC1 was propagated in soft agar overlays on half-strength Luria-Bertani (½ LB) solid medium. Electron microscopy was performed as described previously (177). DC1 DNA was isolated using the GENECLEAN Turbo Kit (Qbiogene, Irvine, CA) following guanidine thiocyanate lysis of PEG-precipitated high-titre phage lysates (as described in Chapter 4). The complete genome sequence was determined using pyrosequencing (454 Life Sciences, Branford, CT) with PCR cloning (CloneJET PCR Cloning Kit; Fermentas, Burlington, ON) to fill the contig gaps. Annotation and sequence analysis were performed using GeneMark.hmm-P (172), BLAST (7), EMBOSS matcher (245), TMHMM (145),

LipoP (136), tRNAscan-SE (262), HHpred (277), and CoreGenes (154, 155, 330). Comparison plots were prepared using PROmer and Circos (68, 147). The DC1 sequence has been deposited in GenBank under accession JN662425.

RESULTS AND DISCUSSION

Isolation, host range, and morphology. DC1 was isolated from an extract of *Dracaena* sp. soil in Edmonton, Canada using *Burkholderia cepacia* LMG 18821 as a host (267). When plated with LMG 18821 in an overlay on ½ LB solid medium, DC1 forms mainly clear plaques with a diameter of 1-2 mm. Transmission electron microscopy of DC1 virions indicates that it is a member of the *Podoviridae* family (Figure 5-1).

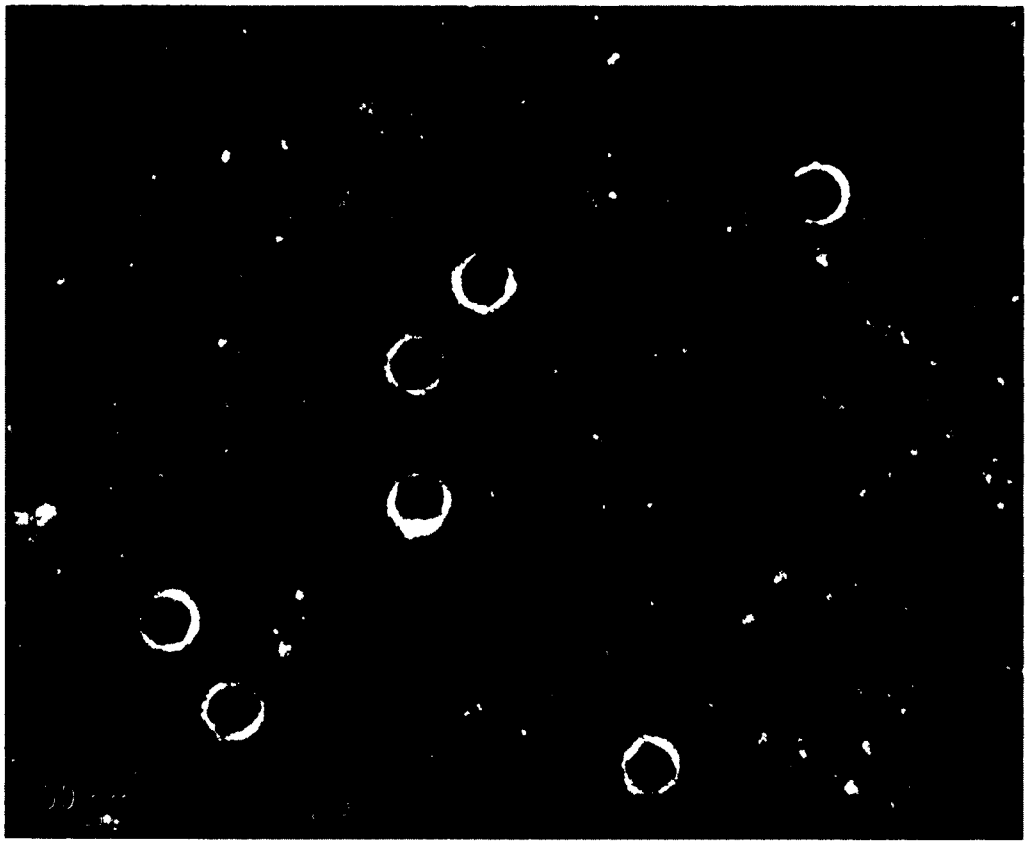


Figure 5-1: Transmission electron micrograph of DC1. Virions were stained with 2% phosphotungstic acid and viewed at 89,000-fold magnification.

While Bcep22 and BcepIL02 – the originally described Bcep22-like phages – were reported to specifically infect *B. cenocepacia* (43, 89), the relatively broader host range of DC1 is a significant advantage with respect to clinical use. In contrast to Bcep22 and BcepIL02, which infect *B. cenocepacia* PC184 (BcepIL02), AU0728 (BcepIL02), and AU1054 (both phages) (43, 89), the DC1 host range includes *B. cepacia* LMG 18821, *B. cenocepacia* C6433, PC184, and CEP511, and *Burkholderia stabilis* LMG 18870 (267). The *B. cepacia* and *B. stabilis* strains are CF isolates, while the *B. cenocepacia* strains are CF epidemic isolates (183). The efficiency of plating for DC1 on each of these strains is similar (within one order of magnitude compared to LMG 18821).

Genome characterization. The DC1 genome is 61,847 base pairs (bp) in length, has a 66.2% GC content, and is predicted to encode 73 proteins and one tRNA (Table 5-1). BLASTN and EMBOSS matcher analysis of the complete genome sequence indicates that it is most closely related to BcepIL02 (79.5% identity) and Bcep22 (73.1% identity). Using CoreGenes analysis to assess phage protein relatedness (154, 155), 52 matches were found between the proteins of DC1 (n=73) and Bcep22 (n=81) (Table 5-2). Although both DC1 gp56 and gp59 (tail fiber proteins) are closely related to Bcep22gp65, the program only tallies gp56 as a match, so the true total is 53 (Table 5-2), resulting in a 65.43% similarity value between these two phages. Based on the recommended CoreGenes genus-level threshold of 40% (154, 155), it is evident that Bcep22-like phages (including DC1) not only comprise a new phage type as previously suggested (89), but that they in fact constitute an entirely novel and distinct podovirus genus.

Table 5-1: DC1 genome annotation

Gene	Start	End	Putative function	Strand	Predicted RBS and start codon	Length (aa)	Closest relative	Alignment region in closest relative (aa)	% identity	Source	GenBank accession no.
1	63	656	hypothetical protein	-	GAGGGAAGGccgtATG	197	gp01	1-205/206	77	*Burkholderia* phage BcepIL02	YP_002922673.1
2	653	847	hypothetical protein	-	AGGGGAGGttgcagcATG	64	gp02	1-64/64	91	*Burkholderia* phage BcepIL02	YP_002922674.1
3	844	1764	hypothetical protein	-	AGGAGAAAggaGTG	306	gp03	1-129, 233-365/394	66, 77	*Burkholderia* phage BcepIL02	YP_002922675.1
4	2170	3429	integrase	+	GAAAGGCAAtgtaccacccA TG	419	gp04	5-422/422	82	*Burkholderia* phage BcepIL02	YP_002922676.1
5	3775	4806	hypothetical protein	-	GGAGGTAGttcATG	343	phage-related protein	1-225/231	79	*Nitrobacter* sp. Nb-311A	ZP_01046190.1
6	4928	5947	RecT (multimer)	-	GAAAAATGGAcgagaagA TG	339	Bcep22gp10	5-337/337	73	*Burkholderia* phage Bcep22	NP_944238.1
7	6013	7044	exonuclease	-	GAGGAGcagatATG	343	gp07	1-343/343	92	*Burkholderia* phage BcepIL02	YP_002922679.1
8	7183	7773	repressor (multimer)	-	AGAAGGGAActgagtggAT G	196	Bcep22gp13	1-196/196	79	*Burkholderia* phage Bcep22	NP_944241.1
9	7898	8254	transcriptional regulator	+	ATGAAATcagccggaaacAT G	118	Bcep22gp14	1-90/118	86	*Burkholderia* phage Bcep22	NP_944242.1
tRNA	8262	8348	tRNA-Ser (UCA codon/TGA anticodon)								
10	8423	8803	hypothetical protein	+	AAGGAGAGAGcatcATG	126	gp10	1-125/126	85	*Burkholderia* phage BcepIL02	YP_002922682.1
11	8862	9671	hypothetical protein	+	AAGACGAcaccaacATG	269	gp11	9-274/280	59	*Burkholderia* phage BcepIL02	YP_002922683.1
12	9668	9757	hypothetical protein	-	ATATCCAGAcgcATG	29	none				
13	9964	10452	single stranded DNA binding protein	+	AAGGAGctcgacATG	162	gp13	1-172/172	81	*Burkholderia* phage BcepIL02	YP_002922685.1
14	10461	10679	hypothetical protein	+	GAGGttcatcATG	72	gp14	1-72/72	75	*Burkholderia* phage BcepIL02	YP_002922686.1
15	10676	11164	hypothetical protein (multimer)	+	GGCAGGTGGGacgATG	162	gp15	1-161/161	82	*Burkholderia* phage BcepIL02	YP_002922687.1
16	11161	11661	recombination protein	+	GGGAGGcgcgGTG	166	gp16	1-166/166	89	*Burkholderia* phage BcepIL02	YP_002922688.1
17	11658	12110	endonuclease	+	GGCGCAGGtagcagcATG	150	PmgM	3-115/130	34	*Clostridium*	ZP_02950170.1

										butyricum 5521	
18	12326	12556	transcriptional regulator	+	GGAGGcagcATG	76	gp19	1-76/78	87	*Burkholderia* phage BcepIL02	YP_002922691.1
19	12553	13359	replication protein (multimer)	+	GAAGGCAGcagcATG	268	gp20	1-268/268	94	*Burkholderia* phage BcepIL02	YP_002922692.1
20	13356	14153	DnaC	+	GGGGGtttgatcgATG	265	gp21	1-265/265	89	*Burkholderia* phage BcepIL02	YP_002922693.1
21	14223	14666	palmitoyltransferase	+	GAGAGGGGtgttATG	147	gp22	1-147/147	97	*Burkholderia* phage BcepIL02	YP_002922694.1
22	14674	14805	hypothetical protein	+	GAGGAcgccATG	43	gp23	1-43/43	93	*Burkholderia* phage BcepIL02	YP_002922695.1
23	14871	15293	hypothetical protein (multimer)	+	AAGGAGAGAacgTTG	140	Bcep22gp30	1-139/139	81	*Burkholderia* phage Bcep22	NP_944259.1
24	15347	15616	hypothetical protein	+	AAGGAGAAAggaGTG	89	gp24	1-93/94	43	*Burkholderia* phage BcepIL02	YP_002922696.1
25	15619	15909	hypothetical protein	+	GAGGAGctgaccATG	96	gp25	1-97/97	74	*Burkholderia* phage BcepIL02	YP_002922697.1
26	15952	17247	hypothetical protein	+	AAAGGAGAGAGAgcaATG	431	gp26	1-410/410	71	*Burkholderia* phage BcepIL02	YP_002922698.1
27	17299	17829	hypothetical protein	-	AGAGGTAAAtaagatcATG	176	gp27	1-150/193	56	*Burkholderia* phage BcepIL02	YP_002922699.1
28	17991	18788	terminase small subunit	+	AGGAAAAAttggcATG	265	gp28	1-268/268	85	*Burkholderia* phage BcepIL02	YP_002922700.1
29	18868	19176	hypothetical protein	+	AAGGAGAAAGgacATG	102	none				
30	19221	19454	hypothetical protein	+	ACCACGaccATG	77	hypothetical protein BURMUCGD1_6655	1-67/67	76	*Burkholderia multivorans* CGD1	ZP_03588378.1
31	19501	19944	hypothetical protein (multimer)	+	AAGGAGAAAGgagATG	147	gp32	1-146/147	59	*Burkholderia* phage BcepIL02	YP_002922704.1
32	19944	20192	hypothetical protein (multimer)	+	GGTGGAGGcatgATG	82	hypothetical protein Bmul_1848	1-77/80	63	*Burkholderia multivorans* ATCC 17616	YP_001580033.1
33	20192	20398	hypothetical protein	+	GAGGGAGGcgtgATG	68	Bcep22gp40	1-67/68	47	*Burkholderia* phage Bcep22	NP_944269.1
34	20555	20797	hypothetical protein	+	GGCCTACActgaGTG	80	Bcep22gp42	1-43/43	95	*Burkholderia* phage Bcep22	NP_944271.1
35	20834	20983	hypothetical protein	+	AGGAGAtcgccGTG	49	Bcep22gp43	1-49/49	98	*Burkholderia* phage Bcep22	NP_944272.1
36	21101	21259	hypothetical protein	+	GAGGAGAAAggaGTG	52	Bcep22gp46	4-53/53	58	*Burkholderia* phage Bcep22	NP_944275.1
37	21256	21615	hypothetical protein	+	GAAGGTGAcggtATG	119	Bcep22gp48	1-111/130	68	*Burkholderia*	NP_944277.1

										phage Bcep22	
38	21690	22622	methyltransferase	+	GGGGAGAGActagcaATG	310	gp41	1-310/310	95	*Burkholderia* phage BcepIL02	YP_002922713.1
39	22627	23238	restriction endonuclease	+	GGGATTGAGgtcGTG	203	gp42	1-203/203	88	*Burkholderia* phage BcepIL02	YP_002922714.1
40	23381	23896	hypothetical protein	+	AGCGCGcaggcATG	171	none				
41	23937	25574	terminase large subunit	+	GGCCGGGctcttctcaGTG	545	gp44	1-531/531	88	*Burkholderia* phage BcepIL02	YP_002922716.1
42	25593	26036	hypothetical protein	+	GAAGGAGAGagcATG	147	gp45	4-148/148	94	*Burkholderia* phage BcepIL02	YP_002922717.1
43	26166	28493	portal protein (virion)	+	GGAGTGAcacTTG	775	Bcep22gp51	1-776/776	87	*Burkholderia* phage Bcep22	NP_944280.1
44	28501	29490	hypothetical protein	+	GAGGAtcgaATG	329	gp47	1-324/325	77	*Burkholderia* phage BcepIL02	YP_002922719.1
45	29500	29859	hypothetical protein	+	AAGGAGccatcATG	119	gp48	1-118/119	98	*Burkholderia* phage BcepIL02	YP_002922720.1
46	29875	30069	carbon storage regulator	+	AGGGGAccgcATG	64	gp49	3-64/64	70	*Burkholderia* phage BcepIL02	YP_002922721.1
47	30070	30159	hypothetical protein	-	AGAAAAGGcaaccagccATG	29	none				
48	30259	31353	major capsid protein (virion)	+	GAGGccctactATG	364	Bcep22gp55	1-364/364	92	*Burkholderia* phage Bcep22	NP_944284.1
49	31422	31868	hypothetical protein (virion)	+	GAACGAGGtatccatcATG	148	gp51	1-148/148	97	*Burkholderia* phage BcepIL02	YP_002922723.1
50	31926	32546	hypothetical protein	+	GAGGAGAGGAGttcATG	206	gp52	1-206/206	92	*Burkholderia* phage BcepIL02	YP_002922724.1
51	32551	33198	hypothetical protein (virion)	+	GAGTAAcgccATG	215	gp53	1-215/215	96	*Burkholderia* phage BcepIL02	YP_002922725.1
52	33195	33824	hypothetical protein (virion)	+	GAACGCGAAccacgtATG	209	gp54	1-209/209	97	*Burkholderia* phage BcepIL02	YP_002922726.1
53	33834	34253	hypothetical protein (virion)	+	AAGGAGGcatcATG	139	Bcep22gp60	1-139/139	83	*Burkholderia* phage Bcep22	NP_944289.1
54	34258	35115	hypothetical protein (virion)	+	AGATGctctaagcgcATG	285	gp56	1-285/285	58	*Burkholderia* phage BcepIL02	YP_002922728.1
55	35115	35393	hypothetical protein	+	GGAGGcataATG	92	gp57	1-92/92	93	*Burkholderia* phage BcepIL02	YP_002922729.1
56	35395	36330	tail fiber protein (virion)	+	GAGTGCGCGActgatATG	311	gp58	1-316/316	74	*Burkholderia* phage BcepIL02	YP_002922730.1
57	36334	37344	tail fiber protein (multimer, virion)	+	GGGGTAActtATG	336	gp59	1-339/339	65	*Burkholderia* phage BcepIL02	YP_002922731.1
58	37341	37844	hypothetical protein (virion)	+	GGAGTCCAAccaATG	167	gp60	1-166/166	92	*Burkholderia*	YP_002922732.1

										phage BcepIL02	
59	37992	38870	tail fiber protein	+	GCGAGGGAttcATG	292	gp61	1-292/292	96	*Burkholderia* phage BcepIL02	YP_002922733.1
60	38873	39781	tail fiber protein	+	GGGAGCCtgactATG	302	gp62	1-302/302	92	*Burkholderia* phage BcepIL02	YP_002922734.1
61	39856	41598	hypothetical protein (virion)	+	GAGGGGtttgcATG	580	gp63	1-580/580	96	*Burkholderia* phage BcepIL02	YP_002922735.1
62	41600	41950	hypothetical protein (multimer, virion)	+	GGCAGGAGtgtgacATG	116	gp64	1-116/116	92	*Burkholderia* phage BcepIL02	YP_002922736.1
63	42010	42441	acyltransferase	+	GCGACGcgctATG	143	gp65	1-145/145	96	*Burkholderia* phage BcepIL02	YP_002922737.1
64	42434	43429	hypothetical protein (virion)	+	GGAGtcccgacATG	331	gp66	1-331/331	94	*Burkholderia* phage BcepIL02	YP_002922738.1
65	43441	44163	phosphoadenosine phosphosulfate reductase	+	AAGGAGAGAGaacGTG	240	gp67	1-239/239	97	*Burkholderia* phage BcepIL02	YP_002922739.1
66	44166	45752	hypothetical protein (virion)	+	GGAAATCGGtgggtaatcATG	528	gp69	1-530/530	85	*Burkholderia* phage BcepIL02	YP_002922741.1
67	45834	59786	multidomain (soluble lytic transglycosylase/helicase/methylase) protein (virion)	+	AGAGGGGAGctcGTG	4650	gp75	1-740, 715-4601/4602	77, 86	*Burkholderia* phage Bcep22	NP_944303.1
68	59827	60069	antiholin	-	GAGGGAAGattATG	80	gp71	1-80/80	84	*Burkholderia* phage BcepIL02	YP_002922743.1
69	60116	60412	hypothetical protein (multimer)	-	GCGGAGAtaccccctATG	98	Bcep22gp77	1-98/98	78	*Burkholderia* phage Bcep22	NP_944305.1
70	60573	60803	holin	+	GGGGccgcgtccacATG	76	gp73	1-75/75	92	*Burkholderia* phage BcepIL02	YP_002922745.1
71	60796	61320	SAR endolysin	+	GGGGAGCGtgacgATG	174	gp79	1-174/174	94	*Burkholderia* phage Bcep22	YP_001531197.1
72	61317	61847	Rz	+	GGGGGcgcaATG	176	Bcep22gp80	1-176/176	93	*Burkholderia* phage Bcep22	NP_944308.1
73	61576	61797	Rz1	+	GAAACCGAcctcaatcATG	73	Bcep22gp81	1-73/73	90	*Burkholderia* phage Bcep22	YP_001531198.1

Abbreviations: RBS, ribosome binding site; aa, amino acid. Multimer and virion protein predictions are based on similarity to proteins analyzed by Gill et al. (89). Reproduced with permission from the American Society for Microbiology.

Table 5-2: BLASTP and CoreGenes comparisons of DC1, BcepIL02, and Bcep22

Gene	BcepIL02 BLASTP homolog and assigned function	Alignment region in BcepIL02 homolog (aa)	% identity	GenBank accession no.	CoreGenes match (total matches=62)	Bcep22 BLASTP homolog and assigned function	Alignment region in Bcep22 homolog (aa)	% identity	GenBank accession no.	CoreGenes match (total matches=52)
1	gp01	1-205/206	77	YP_002922673.1	yes	none				no
2	gp02	1-64/64	91	YP_002922674.1	yes	none				no
3	gp03 conserved phage protein; contains LysM domain	1-129, 233-365/394	66, 77	YP_002922675.1	yes	none				no
4	gp04 putative tyrosine recombinase	5-422/422	82	YP_002922676.1	yes	none				no
5	none				no	none				no
6	gp06 putative RecT-like protein	5-315/342	77	YP_002922678.1	yes	Bcep22gp10 RecT-like protein	5-337/337	73	NP_944238.1	yes
7	gp07 putative nuclease/RecB-like protein	1-343/343	92	YP_002922679.1	yes	Bcep22gp12 putative endonuclease	1-255/303	94	NP_944240.1	yes
8	gp08 putative lambda-like transcriptional repressor	1-177/177	80	YP_002922680.1	yes	Bcep22gp13 HTH-XRE-like protein	1-196/196	79	NP_944241.1	yes
9	gp09 conserved phage protein	12-98/129	84	YP_002922681.1	yes	Bcep22gp14 conserved phage protein	1-90/118	86	NP_944242.1	yes
10	gp10 conserved phage protein	1-125/126	85	YP_002922682.1	yes	none				no
11	gp11 conserved phage protein	9-274/280	59	YP_002922683.1	yes	none				no
12	none				no	none				no
13	gp13 putative single-stranded DNA binding protein	1-172/172	81	YP_002922685.1	yes	Bcep22gp17 single stranded DNA-binding protein	1-169/169	74	NP_944246.1	yes
14	gp14 conserved phage protein	1-72/72	75	YP_002922686.1	yes	Bcep22gp18	1-69/72	72	YP_001531194.1	yes
15	gp15 conserved phage protein	1-161/161	82	YP_002922687.1	yes	Bcep22gp19 conserved phage protein	70-153/153	74	NP_944247.1	yes

16	gp16 conserved phage protein	1-166/166	89	YP_002922688.1	yes	Bcep22gp20 conserved phage protein	1-166/166	85	NP_944248.1	yes
17	none				no	none				no
18	gp19 conserved phage protein	1-76/78	87	YP_002922691.1	yes	Bcep22gp24 conserved phage protein	1-76/78	86	NP_944252.1	yes
19	gp20 putative phage replication protein	1-268/268	94	YP_002922692.1	yes	Bcep22gp27 DNA replication protein; similar to phage LE1 helix-turn-helix motif protein	84-257/257	70	NP_944256.1	yes
20	gp21 putative DnaC-like protein	1-265/265	89	YP_002922693.1	yes	Bcep22gp28 putative DnaC; AAA protein	1-265/265	87	NP_944257.1	yes
21	gp22 conserved phage protein, contains OmpA-like fold; contains putative Sec secretion signal	1-147/147	97	YP_002922694.1	yes	none				no
22	gp23 conserved phage protein	1-43/43	93	YP_002922695.1	yes	Bcep22gp29	19-55/57	54	NP_944258.1	no
23	none				no	Bcep22gp30	1-139/139	81	NP_944259.1	yes
24	gp24	1-93/94	43	YP_002922696.1	yes	none				no
25	gp25	1-97/97	74	YP_002922697.1	yes	none				no
26	gp26 conserved phage protein	1-410/410	71	YP_002922698.1	yes	Bcep22gp31 conserved phage protein	1-409/409	69	NP_944260.2	yes
27	gp27 conserved phage protein	1-150/193	56	YP_002922699.1	yes	Bcep22gp35	1-165/192	53	NP_944264.1	yes
28	gp28 putative terminase small subunit	1-268/268	85	YP_002922700.1	yes	Bcep22gp36 TerS-like protein	1-259/259	48	NP_944265.2	yes
29	none				no	none				no
30	none				no	none				no
31	gp32	1-146/147	59	YP_002922704.1	yes	Bcep22gp37	1-154/154	34	NP_944266.1	yes
32	none				no	Bcep22gp39 Bcep176 gp38-like protein	1-70/70	54	NP_944268.1	yes
33	gp34 conserved	1-67/68	46	YP_002922706.1	yes	Bcep22gp40	1-67/68	47	NP_944269.1	yes

	phage protein									
34	gp37 conserved phage protein	2-27/45	77	YP_002922709.1	yes	Bcep22gp42	1-43/43	95	NP_944271.1	yes
35	none				no	Bcep22gp43 BcepNY3 gp34-like protein	1-49/49	98	NP_944272.1	yes
36	gp40 hypothetical phage protein; contains putative Sec secretion signal	4-53/65	48	YP_002922712.1	yes	Bcep22gp46 predicted signal peptide	4-53/53	58	NP_944275.1	yes
37	none				no	Bcep22gp48 conserved phage protein; predicted to contain two transmembrane domains	1-111/130	68	NP_944277.1	yes
38	gp41 putative methyltransferase	1-310/310	95	YP_002922713.1	yes	none				no
39	gp42 putative type II restriction endonuclease	1-203/203	88	YP_002922714.1	yes	none				no
40	none				no	none				no
41	gp44 putative terminase large subunit	1-531/531	88	YP_002922716.1	yes	Bcep22gp49 TerL-like protein	1-530/530	87	NP_944278.1	yes
42	gp45 conserved phage protein	4-148/148	94	YP_002922717.1	yes	Bcep22gp50 conserved phage protein	4-148/148	90	NP_944279.1	yes
43	gp46 putative portal protein	1-775/775	87	YP_002922718.1	yes	Bcep22gp51 phage structural protein	1-776/776	87	NP_944280.1	yes
44	gp47 conserved phage protein	1-324/325	77	YP_002922719.1	yes	Bcep22gp52 phage structural protein	1-321/322	66	NP_944281.1	yes
45	gp48 conserved phage protein	1-118/119	98	YP_002922720.1	yes	Bcep22gp53	83-105/106	74	NP_944282.1	yes
46	gp49 conserved phage protein	3-64/64	70	YP_002922721.1	yes	Bcep22gp54 CsrA-like protein	3-62/64	67	NP_944283.1	yes
47	none				no	none				no
48	gp50 conserved virion-associated	1-364/364	92	YP_002922722.1	yes	Bcep22gp55 phage structural	1-364/364	92	NP_944284.1	yes

	phage protein					protein				
49	gp51 conserved virion-associated phage protein	1-148/148	97	YP_002922723.1	yes	Beep22gp56 conserved phage protein	1-154/157	60	NP_944285.1	yes
50	gp52 conserved phage protein	1-206/206	92	YP_002922724.1	yes	Beep22gp57	1-211/212	79	NP_944286.2	yes
51	gp53 conserved virion-associated phage protein	1-215/215	96	YP_002922725.1	yes	Beep22gp58 conserved phage protein	1-168/197	88	NP_944287.1	yes
52	gp54 conserved phage protein	1-209/209	97	YP_002922726.1	yes	Beep22gp59 conserved phage protein	1-210/210	87	NP_944288.1	yes
53	gp55 putative major capsid protein	1-140/140	70	YP_002922727.1	yes	Beep22gp60 conserved phage protein	1-139/139	83	NP_944289.1	yes
54	gp56 conserved virion-associated phage protein	1-285/285	58	YP_002922728.1	yes	Beep22gp61	1-150/246	48	NP_944290.1	yes
55	gp57 conserved virion-associated phage protein	1-92/92	93	YP_002922729.1	yes	Beep22gp63 conserved phage protein	1-92/93	45	NP_944292.2	yes
56	gp58 putative tail fiber; contains T4 gp12 binding domain	1-316/316	74	YP_002922730.1	yes	Beep22gp65 phage tail fiber protein	1-290/292	49	NP_944294.2	yes
	gp61 putative tail fiber; contains T4 gp12 binding domain	1-288/292	51	YP_002922733.1		Beep22gp66 phage tail fiber protein	1-290/300	47	NP_944295.1	
	gp62 putative tail fiber; contains T4 gp12 binding domain	1-290/302	49	YP_002922734.1		Beep22gp64 phage tail fiber protein	1-114/425	66	NP_944293.1	
	gp59 putative phage tail fiber	1-98/339	64	YP_002922731.1						
57	gp59 putative phage tail fiber	1-339/339	65	YP_002922731.1	yes	Beep22gp64 phage tail fiber protein	1-120/425	73	NP_944293.1	yes
	gp61 putative tail fiber; contains T4 gp12 binding domain	1-99/292	57	YP_002922733.1		Beep22gp65 phage tail fiber protein	1-99/292	57	NP_944294.2	

	gp58 putative tail fiber; contains T4 gp12 binding domain	1-105/316	52	YP_002922730.1		Bcep22gp66 phage tail fiber protein	1-123/300	50	NP_944295.1	
	gp62 putative tail fiber; contains T4 gp12 binding domain	1-120/302	52	YP_002922734.1						
58	gp60 conserved phage protein; contains mannose-binding lectin domain	1-166/166	92	YP_002922732.1	yes	none				no
59	gp61 putative tail fiber; contains T4 gp12 binding domain	1-292/292	96	YP_002922733.1	yes	Bcep22gp65 phage tail fiber protein	1-292/292	91	NP_944294.2	no
	gp62 putative tail fiber; contains T4 gp12 binding domain	1-293/302	66	YP_002922734.1		Bcep22gp66 phage tail fiber protein	1-293/300	65	NP_944295.1	
	gp58 putative tail fiber; contains T4 gp12 binding domain	1-308/316	47	YP_002922730.1		Bcep22gp64 phage tail fiber protein	1-104/425	54	NP_944293.1	
	gp59 putative phage tail fiber	1-99/339	53	YP_002922731.1						
60	gp62 putative tail fiber; contains T4 gp12 binding domain	1-302/302	92	YP_002922734.1	yes	Bcep22gp66 phage tail fiber protein	1-300/300	83	NP_944295.1	yes
	gp61 putative tail fiber; contains T4 gp12 binding domain	1-291/292	66	YP_002922733.1		Bcep22gp65 phage tail fiber protein	1-291/292	66	NP_944294.2	
	gp58 putative tail fiber; contains T4 gp12 binding domain	1-306/316	46	YP_002922730.1		Bcep22gp64 phage tail fiber protein	1-99/425	56	NP_944293.1	
	gp59 putative phage tail fiber	1-93/339	55	YP_002922731.1						
61	gp63 conserved	1-580/580	96	YP_002922735.1	yes	Bcep22gp67	1-580/580	92	NP_944296.1	yes

	virion-associated phage protein; contains aldehyde dehydrogenase conserved site					conserved phage protein				
62	gp64 conserved phage protein	1-116/116	92	YP_002922736.1	yes	Bcep22gp68	1-115/117	69	NP_944297.1	yes
63	gp65 conserved phage protein; contains acyl-CoA N-acyltransferase domain	1-145/145	96	YP_002922737.1	yes	Bcep22gp69 acetyltransferase; RimL	1-143/143	93	NP_944298.1	yes
64	gp66 conserved virion-associated phage protein	1-331/331	94	YP_002922738.1	yes	Bcep22gp71 conserved phage protein	1-169/171	93	NP_944299.1	yes
						Bcep22gp70 potential signal peptide	1-71/104	87	YP_001531196.1	
						Bcep22gp72 potential signal peptide; conserved phage protein	1-83/83	95	NP_944300.2	
65	gp67 putative PAPS reductase-like protein; contains Rossman-like alpha/beta/alpha fold	1-239/239	97	YP_002922739.1	yes	Bcep22gp73 PAPS reductase	3-242/242	94	NP_944301.1	yes
66	gp69 conserved virion-associated phage protein; contains TPR-like domain	1-530/530	85	YP_002922741.1	yes	Bcep22gp74 conserved phage protein	1-532/532	78	NP_944302.1	yes
67	gp70 putative virion-associated antirestriction protein, DarB-like	1420-4666/4667	93	YP_002922742.1	yes	Bcep22gp75 large multi-functional domain protein; includes SLT, helicase and methylase	1-740, 715-4601/4602	77, 86	NP_944303.1	yes

						domains				
68	gp71 conserved phage protein	1-80/80	84	YP_002922743.1	yes	Bcep22gp76 potential transmembrane domain	1-80/80	84	NP_944304.1	yes
69	gp72 conserved phage protein; contains ribbon-helix-helix domain	1-98/98	78	YP_002922744.1	yes	Bcep22gp77	1-98/98	78	NP_944305.1	yes
70	gp73 putative *holin*	1-75/75	92	YP_002922745.1	yes	Bcep22gp78 *pinholin*; SAR domain; type II holin	1-63/63	92	NP_944306.1	yes
71	gp74 putative lysozyme; contains SAR domain	1-174/174	93	YP_002922746.1	yes	Bcep22gp79 phage SAR endolysin	1-174/174	94	YP_001531197.1	yes
72	gp75 putative Rz-like phage lysis protein	32-171/171	51	YP_002922747.1	yes	Bcep22gp80 N-terminal transmembrane domain; lysis protein; Rz	1-176/176	93	NP_944308.1	yes
73	none				yes	Bcep22gp81 SPII processed outer membrane lipoprotein; lysis protein; Rz1	1-73/73	90	YP_001531198.1	yes

Note: DC1 gp73 was tallied in the CoreGenes output for BcepIL02 but was not identified in the BLASTP search using standard parameters; DC1 gp59 was not tallied in the CoreGenes output for Bcep22 because both DC1 gp56 and DC1 gp59 are most similar to Bcep22gp65. Comparisons are based on BcepIL02 FJ937737.1 (23-MAY-2009) and NC_012743.1 (08-DEC-2010) and Bcep22 AY349011.2 (23-OCT-2007) and NC_005262.2 (02-NOV-2007). Reproduced with permission from the American Society for Microbiology.

Relatedness to BcepIL02 and Bcep22. Predicted DC1 genes show similarity to the majority of both BcepIL02 and Bcep22 genes (Table 5-2), including those encoding the tyrosine integrase (BcepIL02), RecT/nuclease pair, transcriptional repressor, serine tRNA, replication proteins, PagP (BcepIL02), methyltransferase/endonuclease pair (BcepIL02), capsid morphogenesis and DNA packaging proteins, CsrA, multiple tail fiber proteins, acyltransferase, PAPS reductase, large multi-domain protein, and lysis proteins (although, based on TMHMM analysis, we predict that gp68 is the putative antiholin and that the putative holin gp70 contains only one transmembrane domain) (89). Two of these proteins are predicted to be involved in lysogeny: the integrase gp4 and the repressor gp8. Interestingly, with regard to phenotypic similarities between all three phages, we have also observed evidence of unstable lysogeny in DC1 hosts LMG 18821 and C6433 (although the nature of this phenomenon requires further investigation) (89). Three proteins similar to BcepIL02 and Bcep22 conserved proteins have been assigned putative functions based on HHpred analysis (with a 95% probability value cutoff): transcriptional regulators gp9 and gp18 and recombination protein gp16 (Tables 5-1 and 5-3).

Table 5-3: DC1 HHpred predictions

Protein	Closest relative	Probability (%)	Motif	Source
gp1	DHDPS, dihydrodipicolinate synthase	63.80	2yxg_A	*Methanocaldococcus jannaschii* DSM2661
gp2	RAS-related protein RAB-7	53.56	1vg8_A	*Rattus norvegicus*
gp3	Hypothetical protein SB145	81.13	2djp_A	*Homo sapiens*
gp4	Integrase	100.00	1z1b_A	Enterobacteria phage lambda
gp5	Signal transducing adaptor molecule 2	57.98	1x5b_A	*Homo sapiens*
gp6	AGR_C_3712P	14.37	2nys_A	*Agrobacterium tumefaciens* str
gp7	RED alpha, lambda exonuclease	100.00	1avq_A	Enterobacteria phage lambda
gp8	Lambda repressor	99.96	3bdn_A	Enterobacteria phage lambda
gp9	HTH-type transcriptional regulator MQSA (YGIT/B3021)	98.89	3fmy_A	*Escherichia coli* k-12
gp10	Thioesterase	61.21	1q4t_A	*Arthrobacter* SP
gp11	Endothelial differentiation-related factor 1	45.18	1x57_A	*Homo sapiens*
gp12	Cold shock domain-containing protein E1	40.76	2ytx_A	*Homo sapiens*
gp13	Single-stranded DNA-binding protein	100.00	3pgz_A	*Bartonella henselae*
gp14	Uncharacterized protein YBII	94.24	2kgo_A	*Escherichia coli* k-12
gp15	Nuclear mRNA export protein SAC3	26.08	3fwb_B	*Saccharomyces cerevisiae*
gp16	Recombination enhancement function protein	100.00	3plw_A	Enterobacteria phage P1
gp17	SDAI restriction endonuclease	84.63	2ixs_A	*Streptomyces diastaticus*

gp18	BIRA bifunctional protein	96.46	1bia_A	*Escherichia coli*
gp19	DNA replication protein DNAD	96.75	2v79_A	*Bacillus subtilis*
gp20	DNAI, primosome component (helicase loader)	100.00	2w58_A	*Geobacillus kaustophilus* HTA426
gp21	Protein PAGP	100.00	3gp6_A	*Escherichia coli*
gp22	Stannin, AG8_1	86.89	1zza_A	*Homo sapiens*
gp23	NitrIle hydratase beta subunit	59.23	3hht_B	*Geobacillus pallidus*
gp24	39S ribosomal protein L19, mitochondrial	67.13	2ftc_K	*Bos taurus*
gp25	Protein RV2228C/MT2287	49.83	3hst_B	*Mycobacterium tuberculosis*
gp26	Smooth muscle myosin heavy chain	80.18	1i84_S	*Gallus gallus*
gp27	Protein YBL047C	39.61	2g3q_A	*Saccharomyces cerevisiae*
gp28	G1P, terminase small subunit	99.87	2cmp_A	Bacteriophage SF6
gp29	Flavoredoxin	62.74	2d5m_A	*Desulfovibrio vulgaris* str
gp30	UPF0066 protein AF_0241	68.08	2nv4_A	*Archaeoglobus fulgidus*
gp31	Hypothetical protein	46.69	2p97_A	*Anabaena variabilis* atcc 29413
gp32	Proteasome activator protein PA26	73.25	1yar_O	*Trypanosoma brucei*
gp33	Urease beta subunit	56.90	1ejx_B	*Klebsiella aerogenes*
gp34	TAFII28	29.15	1bh9_B	*Homo sapiens*
gp35	DNA polymerase beta-like protein	66.45	1jaj_A	African swine fever virus BA71V
gp36	MU-conotoxin PIIIA	43.88	1r9i_A	*Conus purpurascens*
gp37	ZF-HD homeobox family protein	70.08	1wh5_A	*Arabidopsis thaliana*
gp38	Adenine-specific methyltransferase MBOIIA	100.00	1g60_A	*Moraxella bovis*
gp39	Bstyi	100.00	1sdo_A	*Geobacillus stearothermophilus*

gp40	DNA-directed RNA polymerase RPO6 subunit	19.33	2waq_K	*Sulfolobus shibatae*
gp41	Terminase, DNA packaging protein GP17	100.00	3cpe_A	Bacteriophage T4
gp42	Inositol oxygenase	67.19	2ibn_A	*Homo sapiens*
gp43	Portal protein	98.68	2jes_A	Bacteriophage SPP1
gp44	Poly(beta-D-mannuronate) C5 epimerase 4	47.63	2agm_A	*Azotobacter vinelandii*
gp45	Mitochondrial ribosomal protein L2	60.19	2ftc_B	*Bos taurus*
gp46	Carbon storage regulator homolog	98.46	2bti_A	*Yersinia enterocolitica*
gp47	Chlorophyll A-B binding protein AB80	31.22	2bhw_A	*Pisum sativum*
gp48	GP10, T7-like capsid protein	98.25	2xd8_A	*Prochlorococcus* phage p-ssp7
gp49	PACS, cation-transporting ATPase, P-type	34.38	2hc8_A	*Archaeoglobus fulgidus*
gp50	Protein (potassium channel KV1.1)	34.54	1t1d_A	*Aplysia californica*
gp51	Protein EB3, microtubule-associated protein RP/EB family member 3	34.89	1wyo_A	*Homo sapiens*
gp52	Major ampullate spidroin 1	46.17	3lr2_A	*Euprosthenops australis*
gp53	2-keto-4-pentenoate hydratase	22.67	2wqt_A	*Escherichia coli*
gp54	PBP4, penicillin-binding protein 4	26.30	3a3d_A	*Haemophilus influenzae*
gp55	KIAA1837 protein	90.88	2yrl_A	*Homo sapiens*
gp56	Bacteriophage T4 short tail fibre	99.80	1ocy_A	Bacteriophage T4
gp57	NECK appendage protein	99.83	3gud_A	*Bacillus* phage ga-1
gp58	Lipoprotein	29.92	3lhn_A	*Shewanella oneidensis*

gp59	Bacteriophage T4 short tail fibre	99.88	1ocy_A	Bacteriophage T4
gp60	Bacteriophage T4 short tail fibre	99.84	1ocy_A	Bacteriophage T4
gp61	Putative gluconolactonase	75.89	3e5z_A	*Deinococcus radiodurans*
gp62	Middle operon regulator	47.56	1rr7_A	Enterobacteria phage MU
gp63	Ribosomal-protein-alanine acetyltransferase	99.95	2z10_A	*Thermus thermophilus*
gp64	Colicin E3	80.92	2b5u_A	*Escherichia coli*
gp65	Phosphoadenosine phosphosulfate reductase	100.00	2o8v_A	*Escherichia coli*
gp66	Eukaryotic translation initiation factor 4E- binding protein 1	19.68	3hxi_C	*Homo sapiens*
gp67	RNA polymerase-associated protein RAPA	100.00	3dmq_A	*Escherichia coli* K12
gp68	S19E protein	71.49	3jyv_T	*Thermomyces lanuginosus*
gp69	Nickel-responsive regulator	92.96	2hza_A	*Escherichia coli*
gp70	Mature capsid protein gamma	74.97	1f8v_D	Pariacato virus
gp71	Lysozyme	100.00	2anv_A	Enterobacteria phage P22
gp72	Zinc resistance-associated protein	79.13	3lay_A	*Salmonella enterica* subsp
gp73	'BLR6230 protein	29.16	3m8t_A	*Bradyrhizobium japonicum*

Like many phages, DC1 has a mosaic structure in which regions of strong similarity to BcepIL02 and Bcep22 are interspersed with regions of minimal to no similarity throughout its genome. This mosaicism is evident in the PROmer/Circos plots comparing these three phages (Figure 5-2). Based on BLASTP analysis, DC1 lacks homologs of 15 BcepIL02 proteins and 26 Bcep22 proteins (Table 5-2). The majority of these proteins have no assigned functions. However, DC1 lacks a homolog of a putative transcriptional regulator, DNA ligase, and Rz1-like lysis protein of BcepIL02 and a putative serine recombinase, HNH endonuclease (2 proteins), methyltransferase, transposase, transcriptional regulator, and pectin lyase-like protein of Bcep22 (89). The only DC1 proteins with low E-value BLASTP matches to proteins not found in either BcepIL02 or Bcep22 are gp5, gp17, and gp30. Of these, only gp17 has been assigned a putative function (Table 5-1).

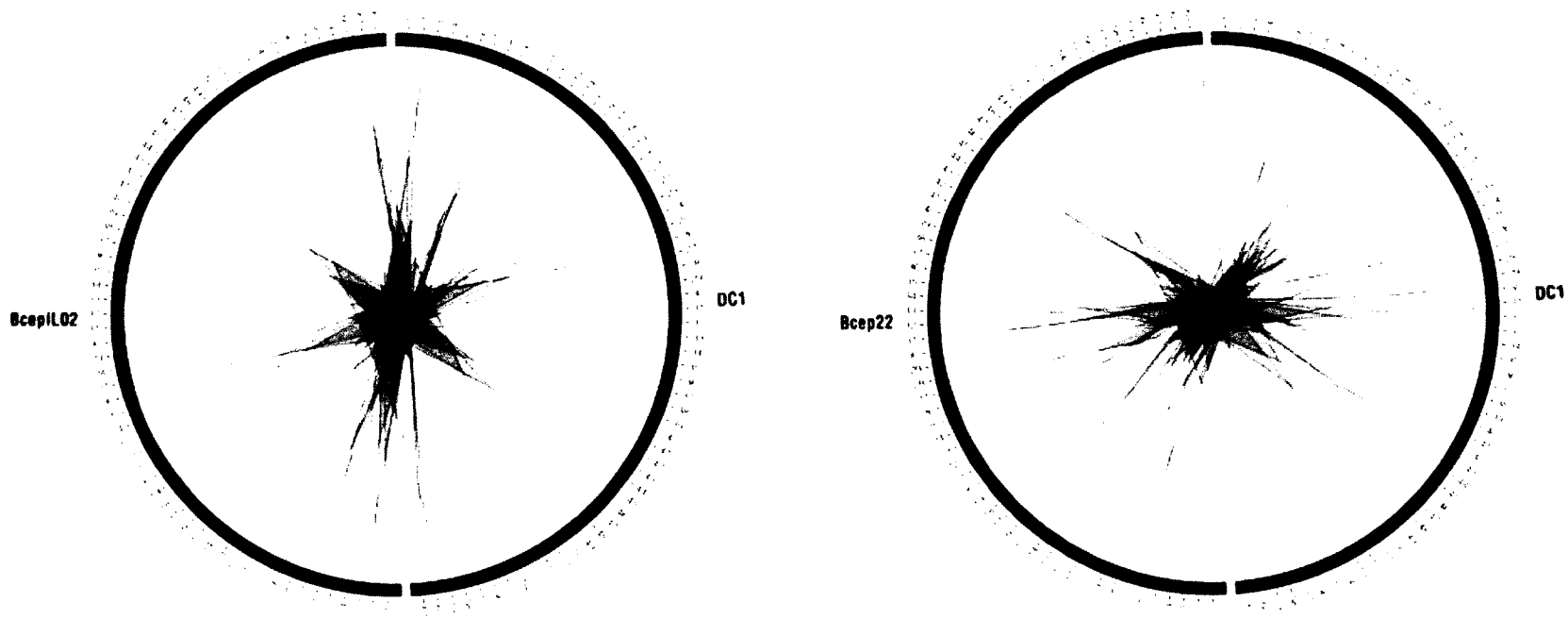


Figure 5-2: PROmer/Circos comparisons of DC1 and BcepIL02 (left) or Bcep22 (right). The scale (in kbp) is shown on the periphery. Green ribbons connect regions of protein-level similarity involving the same strand on both genomes. No matches involving opposite strands were detected. PROmer parameters: breaklen = 60, maxgap = 30, mincluster = 20, minmatch = 6. Reproduced with permission from the American Society for Microbiology.

Exopolysaccharide (EPS) depolymerase. When propagated in dilute agar overlays on EPS-inducing solid medium using LMG 18821 or PC184 as a host, DC1 exhibits a plaque morphology characteristic of phages that express EPS depolymerase enzymes (252). These plaques are surrounded by large zones of clearing (i.e. halos) that increase in size for up to one week, a phenotype not observed for overlays on ½ LB solid medium (252). One of the objectives of this project was to identify the DC1 protein(s) associated with this phenotype. Unfortunately, a putative EPS depolymerase gene was not apparent in the annotation following BLAST analysis (Table 5-1). Based on HHpred predictions (Table 5-3), the most likely candidate is gene *57*, which encodes a protein with similarity to the enterobacteria phage K1F endo-*N*-acetylneuraminidase (99.81% probability), a depolymerase that degrades the *E. coli* K1 polysialic acid capsule (105). This tail protein assembles into trimers (212), consistent with the prediction that gp57 is part of the virion and forms multimers (Table 5-1). HHpred analysis indicates that the region of similarity between these proteins is in the C-terminal chaperone domain and not the catalytic domain. This result would be expected for a Bcc phage depolymerase because of structural differences between the *E. coli* capsule and Bcc EPS. Experiments are in progress to characterize this protein and to determine if it is required for DC1 depolymerase activity.

CONCLUSIONS

Although DC1 has certain drawbacks with respect to its potential for use in phage therapy (i.e. genes for lysogeny and a putative lipid A

palmitoyltransferase [89]), it also has four key advantages. First, the DC1 host range is relatively broad, infecting clinical strains of multiple Bcc species. It remains to be determined if amino acid differences between the tail fiber proteins of DC1 and those of BcepIL02 and Bcep22 are responsible for the expanded host range of DC1 (as DC1 gp56, gp57, gp59, and gp60 exhibit 46-96% identity with the tail fiber proteins of BcepIL02 and Bcep22 [Table 5-2]). Second, DC1 is closely related to the only phage shown to date to be active against the Bcc in a mammalian infection model: in C57BL/6 mice infected with *B. cenocepacia* AU0728, BcepIL02 has been shown to decrease both bacterial density and inflammatory cytokine release (43). Third, as DC1 putatively expresses an EPS depolymerase, it may be more active against biofilms in the CF lung than phages lacking such an enzyme. Finally, all three Bcep22-like phages encode putative CsrA-like proteins. F116, a podovirus active against *P. aeruginosa* biofilms, encodes a similar regulator (39, 108). In *E. coli*, CsrA expression is inhibitory to biofilm development by means of both decreased formation and increased dispersion (131). When Lu and Collins (170) engineered an M13 phage derivative to express CsrA in *E. coli*, host susceptibility to ofloxacin increased. If the action of CsrA in *Burkholderia* is analogous to that in *E. coli*, Bcep22-like phages *in vivo* could potentially induce not only direct killing but also reduced biofilm development and increased antibiotic susceptibility.

Together with the findings of Gill et al. (89), we can conclude that Bcep22-like phages have a wide geographic distribution and a potentially broad range of hosts within the Bcc. Since a member of this group has already been

shown to be active against the Bcc *in vivo* (43), isolation of related phages with expanded host ranges is important for Bcc phage therapy development.

ACKNOWLEDGEMENTS

The authors would like to thank Danielle Carpentier for isolating DC1, Kimberley Seed and Amanda Goudie for preliminary DC1 characterization, and Ashraf Abdu and Sarah Routier for isolating putative unstable lysogens.

Chapter 6

Conclusions and future directions

Portions of this chapter have been published as:

Lynch, K. H., and J. J. Dennis. In press. Genomic analysis and modification of *Burkholderia cepacia* complex bacteriophages: Cangene Gold Medal Lecture. Can. J. Microbiol.

PROJECTS IN PROGRESS

B. cenocepacia J2315 transposable myovirus prophage variants. The *B. cenocepacia* J2315 genome is predicted to contain 14 genomic islands, at least 6 of which are either prophages or prophage remnants (126). Two of these phages have been sequenced independently: the transposable myoviruses BcepMu/KS4 (J2315 BcenGI14) and KS10 (J2315 BcenGI7) (93, 126, 265, 288). We have isolated two phenotypic variants of these phages, KS4-M and SR1 (Figure 6-1), and have determined their complete genome sequences via pyrosequencing.

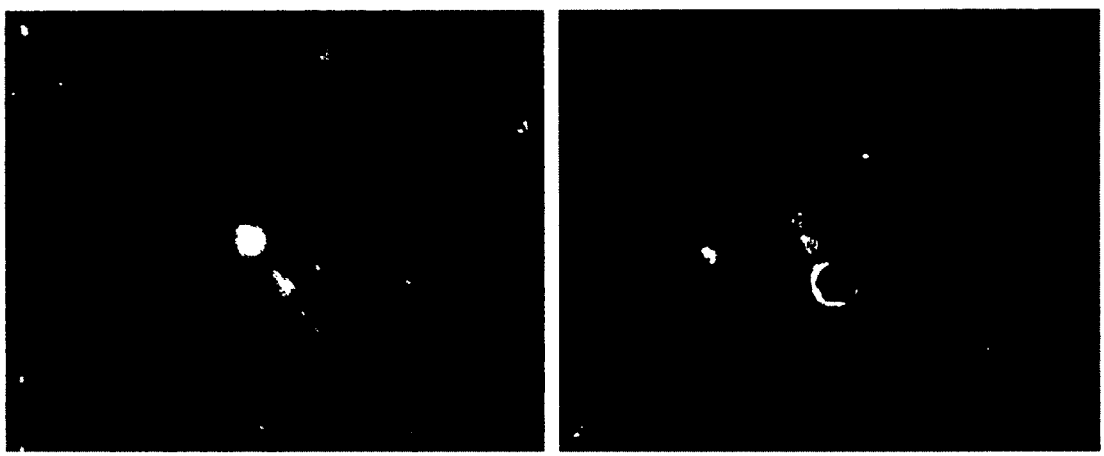


Figure 6-1: Transmission electron micrographs of KS4-M (left) and SR1 (right). Virions were stained with 2% phosphotungstic acid and viewed at 140,000-fold (KS4-M) or 180,000-fold (SR1) magnification.

KS4-M is a liquid clearing variant of KS4 isolated following repeated passage in liquid culture by former graduate student Dr. Kimberley Seed (267). We identified two base pair changes between the sequences of KS4-M, KS4, BcepMu, and J2315 BcenGI14 (Table 6-1). These changes (positions 31,518 and 32,095) result in two nonsynonymous substitutions in gp50, a predicted baseplate assembly protein (Table 6-1) (288). We predict that the threonine to alanine substitution at position 116 of gp50 may be responsible for this phenotype, as it is

the only residue change differentiating the non-liquid clearing KS4 from the liquid clearing KS4-M (Table 6-1). Although preliminary work has suggested that this hypothesis could be accurate, further experiments are required to confirm this prediction.

Table 6-1: Sequence comparison of KS4 gp50 variants

KS4-M base pair	Phage	Base	Amino acid (codon and position in gp50)
31,518	KS4-M	G	Ala/A (GCG 116)
	KS4	A	Thr/T (ACG 116)
	BcepMu	A	Thr/T (ACG 116)
	J2315 BcenGI14	A	Thr/T (ACG 116)
32,095	KS4-M	A	Asp/D (GAC 308)
	KS4	A	Asp/D (GAC 308)
	BcepMu	G	Gly/G (GGC 308)
	J2315 BcenGI14	A	Asp/D (GAC 308)

A second J2315 transposable myovirus prophage variant named SR1 was isolated by former graduate student Sarah Routier (252). This phage is an extended host range variant of KS10 that infects *B. multivorans* ATCC 17616 in addition to the previously identified KS10 hosts *B. cenocepacia* PC184, *B. stabilis* LMG 18870, and *B. ambifaria* LMG 19467 (93, 252). Because this phage was isolated from a soil extraction containing *B. cenocepacia* K56-2 (252), it is unclear if SR1 originated from the environmental sample or from the added Bcc strain. Sequencing of this isolate revealed ten base pair changes between SR1, KS10, and J2315 BcenGI7 (Table 6-2). We predict that the mutation responsible for the tropism shift is at position 35,924 in gene *48*, encoding a tail fiber protein (Table 6-2) (93). The insertion of a G in gene *48* is predicted to cause a frameshift, allowing for the expression of two proteins (536 and 183 residues in length) instead of the single 742 residue gp48. Further experiments are required to

determine if this frameshift is responsible for the expanded host range of SR1.

Table 6-2: Sequence comparison of SR1, KS10, and J2315 BcenGI7

SR1 base pair	gene (function)	SR1 base	KS10 base	J2315 BcenGI7 base
39	noncoding	-	A	A
1,871	*3* (portal protein)	A	A	C
2,488	*3* (portal protein)	G	A	A
7,219	*10* (holin)	C	T	C
8,931	*13* (repressor)	C	-	-
13,828	noncoding	-	A	-
28,630	*41* (DNA circulation protein)	C	C	G
28,650	*41* (DNA circulation protein)	C	T	T
33,666	*47* (tail collar protein)	G	G	A
35,924	*48* (tail fiber protein)	G	-	-

Myovirus KS12. KS12 is a myovirus isolated from *Dietes grandiflora* (iris) soil that infects *B. multivorans* C5274 and *B. cenocepacia* K56-2 (Figure 6-2) (267). As this phage is active against K56-2 in the *Galleria mellonella* model and has not been shown to lysogenize its host (267), genomic characterization of KS12 remains a priority. Although DNA can be isolated from KS12 using standard SDS/proteinase K lysis, it is resistant to both shearing and restriction. As a result, attempts to collect sequence data from this phage have thus far been unsuccessful. Using a standard nebulizer (Invitrogen, Carlsbad, CA) and the nebulization technology utilized by 454 Life Sciences (Branford, CT), neither we nor 454 were able to shear KS12 DNA (a result that, according to personal communications with a 454 representative, they had not previously encountered). With respect to restriction, former undergraduate student Bryan Frobb determined that out of 31 restriction enzymes tested, KS12 DNA is only susceptible to cleavage by HpaII, a four-base cutter with the recognition site CCGG. A third unique property of KS12 DNA is its migration on agarose gels. In Figure 6-3, all

of the Bcc phage genomic DNA samples tested migrate similarly on a 2% agarose gel excluding KS12, which runs at a substantially slower rate. We hypothesize that extensive DNA modification (either protein or chemical) could account for these unique properties, although the nature of this modification remains to be identified.

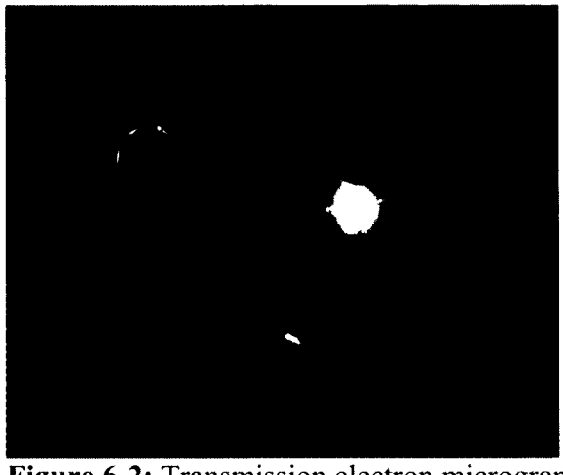


Figure 6-2: Transmission electron micrograph of KS12. Virions were stained with 2% phosphotungstic acid and viewed at 140,000-fold magnification

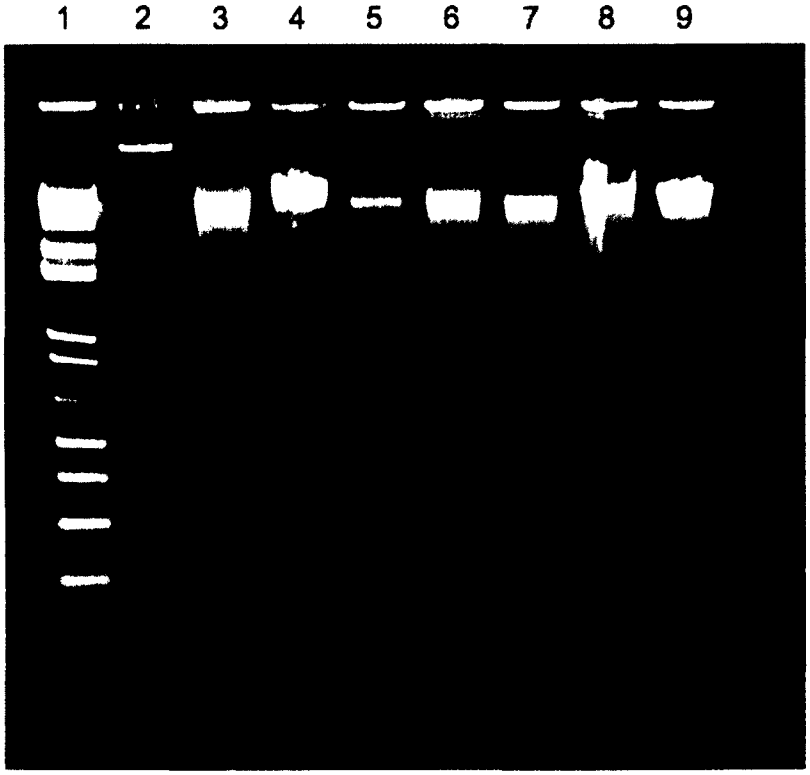


Figure 6-3: 2% agarose gel electrophoresis of Bcc phage genomic DNA. 1: 1 Kb Plus DNA ladder (Invitrogen); 2: KS12+K56-2 DNA; 3: KS12+C5274 (i.e. KL4) DNA; 4: KS14 DNA; 5: KS5 DNA; 6: DC1 DNA; 7: KS1 and KL3 DNA; 8: AH2 and SR1 DNA; 9: KL1 and KS4-M DNA.

B. multivorans C5274 prophage KL4. When DNA was isolated from KS12 propagated on C5274, we occasionally observed two bands following gel electrophoresis: one that migrated slowly, characteristic of KS12 (Figure 6-3, lane 2) and one that migrated at a normal rate (Figure 6-3, lane 3). The latter DNA was susceptible to restriction digest and sequencing of shotgun-cloned fragments showed similarity to phage proteins. PCR primers to these sequences amplified both KS12+C5274 lysates and C5274 chromosomal DNA, indicating that these sequences belong to a C5274 prophage. Although we had no host on which to propagate this phage, subsequently named KL4, sufficient quantities of DNA for pyrosequencing analysis were isolated from KS12+C5274 lysates.

The KL4 genome is 42,250 bp in length, has a 63.2% GC content, and is predicted to encode 60 proteins (Table 6-3). KL4 is most closely related to uncharacterized prophage elements from *B. multivorans* chronic granulomatous disease isolates and to Bcc phage BcepC6B (Table 6-3). In C5274, the KL4 prophage is integrated downstream of a LacI family transcriptional regulator using a 17 bp *attP* overlap region (Figure 6-4). Although this locus has not been previously characterized as a phage insertion site, a putative prophage sequence is found downstream of a LacI regulator in *B. vietnamiensis* G4 chromosome 1, suggesting that it may be used for integration by various Bcc prophages (as was hypothesized for the novel KS9 [Chapter 2] and KS5 [Chapter 3] insertion sites).

Further characterization of KL4 has not yet been performed because we have been unable to identify a suitable host for lytic growth. Host ranges using filter-sterilized C5274 culture supernatant and 26 Bcc strains were negative. In addition, KS12 propagated on C5274 was tested with 64 *Burkholderia* and *Pseudomonas* strains, but lysis was only observed on previously identified KS12 hosts. In order to determine the virion morphology of KL4, we plan to induce C5274 cultures with mitomycin C or UV and analyze the supernatants using transmission electron microscopy. In order to propagate this phage, we will test a new panel of Bcc isolates with C5274 supernatants and we will screen for either a KL4-cured C5274 strain or a KL4 operator mutant that can be propagated on wild-type C5274.

Table 6-3: KL4 genome annotation

Gene	Start	End	Putative function	Strand	Predicted RBS and start codon	Length (aa)	Closest relative	Alignment region in closest relative (aa)	% identity	Source	GenBank accession number
1	183	1220	integrase	+	AAGGGGGcaggATG	345	phage integrase family protein	1-345/345	99	*Burkholderia vietnamiensis* G4	YP_001118990.1
2	1205	1453	unknown	-	GAGGGAGAGcagaaATG	82	conserved hypothetical protein	1-82/82	100	*Burkholderia multivorans* CGD2M	ZP_03574500.1
3	1491	1769	unknown	-	AGGAGAcgcatcGTG	92	hypothetical protein NEUTE1DRAFT_77655	1018-1094/1403	28	*Neurospora tetrasperma* FGSC 2508	EGO61575.1
4	1766	1981	unknown	-	AGGAGGcaagcaATG	71	guanine deaminase	201-252/556	33	*Neosartorya fischeri* NRRL 181	XP_001257280.1
5	1978	2376	unknown	-	AATGGAGccaagcATG	132	hypothetical protein BURMUCGD2M_2600	75-139/142	65	*Burkholderia multivorans* CGD2M	ZP_03574505.1
6	2369	2866	adenine methylase	-	AAGGGAGGtgcgtgATG	165	hypothetical protein BpseN_34450	1-168/169	80	*Burkholderia pseudomallei* NCTC 13177	ZP_02494587.1
7	2866	3174	unknown	-	ATGGGAccaatcATG	102	hypothetical protein BCAM1029	1-104/107	70	*Burkholderia cenocepacia* J2315	YP_002233645.1
8	3274	4620	unknown	-	GAGGAccacaccATG	448	conserved hypothetical protein	96-314/387	67	*Burkholderia multivorans* CGD2M	ZP_03569239.1
9	4639	5739	methylase	-	GAGGAtgcgATG	366	DNA methylase	2-387/390	73	*Burkholderia pseudomallei* Pasteur 52237	ZP_04898485.1
10	5923	6975	RecT	-	GAGGAAtaaaaaccATG	350	RecT protein	1-350/350	99	*Burkholderia multivorans* CGD2M	ZP_03569244.1
11	6981	7793	RecE	-	GGGGGAcgtaATG	270	exonuclease VIII, 5'->3' specific dsdna exonuclease	1-270/270	99	*Burkholderia multivorans* CGD2M	ZP_03569245.1
12	7793	7954	unknown	-	GGGAGtcgctcgATG	53	putative bacteriophage	1-53/53	100	*Burkholderia*	YP_001945863.1

							protein			*multivorans* ATCC 17616	
13	7951	8109	unknown	-	GAAGGAGGcatgATG	52	putative phage membrane protein	1-52/52	92	*Burkholderia cenocepacia* J2315	YP_002233658.1
14	8109	8402	unknown	-	AATGGAGGcgtctgATG	97	Bcep22gp38	1-97/98	63	*Burkholderia* phage Bcep22	NP_944267.1
15	8402	8713	unknown	-	GAGGTGAagcATG	103	gp39	53-105/105	96	*Burkholderia* phage Bcep176	YP_355374.1
16	8710	9027	unknown	-	GAGAGGAccacATG	105	gp38	1-91/91	67	*Burkholderia* phage Bcep176	YP_355373.1
17	9083	9274	unknown	-	AGGAGAAAGaccGTG	63	hypothetical protein BcepF1.037	2-61/65	48	*Burkholderia ambifaria* phage BcepF1	YP_001039721.1
18	9303	9566	nuclease inhibitor	-	AGGAGAAAGaccATG	87	host nuclease inhibitor protein	5-88/88	51	*Escherichia blattae* DSM 4481	AAX12924.1
19	9594	9776	unknown	-	GAAGGGGAAcgacgATG	60	PREDICTED: VPS10 domain-containing receptor SorCS3-like	611-648/1233	37	*Danio rerio*	XP_684106.3
20	9804	9908	unknown	-	GGAGAcgaatcGTG	34	none				
21	9905	10222	unknown	-	AGGAGtgcatATG	105	hypothetical protein Bcep1808_1158	1-101/102	76	*Burkholderia vietnamiensis* G4	YP_001119004.1
22	10884	11084	unknown	-	GGAGAGatcATG	66	conserved hypothetical protein	1-66/66	100	*Burkholderia multivorans* CGD1	ZP_03586938.1
23	11100	11873	repressor	-	GATATTAGcatcctaATG	257	HTH-type transcriptional regulator PrtR (Pyocin repressor protein)	1-257/257	99	*Burkholderia multivorans* CGD1	ZP_03586937.1
24	11894	12100	transcriptional regulator	+	AGGATGCTAAtatcgcctcATG	68	hypothetical protein 111_02140	1-62/69	37	*Neisseria weaveri* LMG 5135	ZP_08826135.1
25	12296	12520	unknown	-	GAAAGGAAGctcctATG	74	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	338-412/419	39	*Streptococcus ictaluri* 707-05	ZP_08728726.1
26	12535	12939	transcriptional	+	AGGAGcttcctttcATG	134	phage-related	3-107/107	92	*Burkholderia*	YP_001119007.1

			regulator				transcriptional transcriptional regulator			*vietnamiensis* G4	
27	12936	13148	unknown	+	AAGGAGGGGcggtcATG	70	hypothetical protein BURMUCGD2M_4608	1-49/59	98	*Burkholderia multivorans* CGD2M	ZP_03569259.1
28	13138	13236	unknown	+	GAGGAtcgtcaacATG	32	none				
29	13238	13615	unknown	+	GGAGcctgacATG	125	hypothetical protein PA13_12715	1-79/132	51	*Pseudomonas aeruginosa* 138244	EGM19366.1
30	13615	14400	replication protein	+	GGCGGGttcgaataATG	261	conserved hypothetical protein	1-93/93	99	*Burkholderia multivorans* CGD2M	ZP_03574528.1
31	14397	15764	helicase	+	AGGGAGGctgcATG	455	replicative DNA helicase	1-455/455	99	*Burkholderia multivorans* CGD2M	ZP_03574529.1
32	15761	16078	Vsr endonuclease	+	AAGGGGGttcgaATG	105	conserved hypothetical protein	1-105/105	100	*Burkholderia multivorans* CGD2M	ZP_03574530.1
33	16078	16446	unknown	+	GAAGGAGGGAGcgtaATG	122	hypothetical protein BURMUCGD2M_2573	1-122/122	100	*Burkholderia multivorans* CGD2M	ZP_03574531.1
34	16549	16818	unknown	+	GGGGAGGcaaagcaATG	89	hypothetical protein BcepC6B_gp01	12-100/100	73	*Burkholderia* phage BcepC6B	YP_024921.1
35	16815	17213	terminase small subunit	+	GGGTGGtgcgggcGTG	132	gp1	31-162/162	99	*Burkholderia multivorans* CGD1	ZP_03586931.1
36	17210	18631	terminase large subunit	+	AAGAGGcaaggcggcgaATG	473	putative TerL	1-473/473	99	*Burkholderia multivorans* CGD1	ZP_03586930.1
37	18693	18920	unknown	+	GGAGAAGcgcATG	75	conserved hypothetical protein	1-75/75	99	*Burkholderia multivorans* CGD2M	ZP_03574535.1
38	18917	20581	head-tail connector	+	AAGGCGAtgctcgggcaATG	554	conserved hypothetical protein	1-549/549	95	*Burkholderia multivorans* CGD1	ZP_03586928.1
39	20581	20877	unknown	+	GAGGGtgtgATG	98	hypothetical protein BURMUCGD2M_2568	1-98/98	93	*Burkholderia multivorans* CGD2M	ZP_03574537.1
40	20874	21218	unknown	+	AAGGACGAGGccaccgaATG	114	hypothetical protein BcepC6B_gp06	79-192/192	94	*Burkholderia* phage BcepC6B	YP_024926.1

41	21231	21605	unknown	+	GACAGGAGcatcccATG	124	gene 66 protein	12-127/136	71	*Burkholderia multivorans* CGD2M	ZP_03572545.1
42	21645	22391	unknown	+	AGGAGGttccATG	248	conserved hypothetical protein	1-241/241	82	*Burkholderia multivorans* CGD2M	ZP_03574539.1
43	22401	23393	major capsid protein	+	AAGGAGcttacATG	330	Bbp17	1-330/330	98	*Burkholderia multivorans* CGD2M	ZP_03574540.1
44	23456	23860	unknown	+	GAAAGGAcatcgcaATG	134	hypothetical protein BcepC6B_gp09	1-139/139	78	*Burkholderia* phage BcepC6B	YP_024929.1
45	23920	24180	unknown	+	AAGGGGAGcagcATG	86	conserved hypothetical protein	1-86/86	99	*Burkholderia multivorans* CGD1	ZP_03586922.1
46	24192	24815	unknown	+	AGTAACGGActgacgaATG	207	conserved hypothetical protein	1-207/207	95	*Burkholderia multivorans* CGD2M	ZP_03574543.1
47	24824	27127	unknown	+	AGGAGcgcgcATG	767	hypothetical protein BcepC6B_gp12	1-768/768	86	*Burkholderia* phage BcepC6B	YP_024932.1
48	27117	27542	GNAT	+	GGATAcgcaagATG	141	acetyltransferase, gnat family	1-141/141	99	*Burkholderia multivorans* CGD2M	ZP_03574545.1
49	27542	28147	unknown	+	GAGGGtgaagtaATG	201	conserved hypothetical protein	1-201/201	99	*Burkholderia multivorans* CGD2	ZP_03580943.1
50	28155	30275	unknown	+	GATAGGActgatATG	706	conserved hypothetical protein	1-706/706	98	*Burkholderia multivorans* CGD1	ZP_03586917.1
51	30275	32524	tail protein	+	GGGAGGtatgtaATG	749	SLT domain-containing tail structural protein	1-749/749	98	*Burkholderia multivorans* CGD1	ZP_03586916.1
52	32525	35119	unknown	+	GGGGGGctaaATG	864	conserved hypothetical protein	1-864/864	98	*Burkholderia multivorans* CGD1	ZP_03586915.1
53	35178	37559	tail fiber	+	AAGGGctgcgccATG	793	putative tail fiber protein	1-283/637	56	*Burkholderia* phage BcepC6B	YP_024938.1
54	37562	38488	unknown	+	GGTTGatcATG	308	hypothetical protein Bcep1808_1190	1-298/299	39	*Burkholderia vietnamiensis* G4	YP_001119036.1

55	38513	38845	antiholin	+	AGGGGGtccatATG	110	putative holin protein	1-110/110	100	*Burkholderia multivorans* CGD2M	ZP_03574551.1
56	38845	39120	holin	+	GGGGGGcaatgATG	91	putative holin protein	2-92/92	99	*Burkholderia multivorans* CGD2	ZP_03580937.1
57	39107	39526	lysozyme	+	AAGAAGGGAGGGcggcATG	139	phage lysozyme	1-139/139	98	*Burkholderia multivorans* CGD2M	ZP_03574553.1
58	39523	40002	Rz	+	GGAGCGGcgcATG	159	gp23	1-154/164	90	*Burkholderia multivorans* CGD2M	ZP_03574554.1
59	39722	39985	Rz1	+	AGGAGAAcgcgaaccATG	87	hypothetical protein BthaA_24648	4-90/90	63	*Burkholderia thailandensis* E264	ZP_05590645.1
60	40081	41727	unknown	+	AATGCGtcgcaaATG	548	hypothetical protein BH160DRAFT_3745	12-540/543	37	*Burkholderia* sp. H160	ZP_03267466.1

Abbreviations: RBS, ribosome binding site; aa, amino acid.

attL	C C G G T T C G A G **T C C G G T C C C C G G C A C C A** T C C C A C T C T G
virion	A A C C C T A G T T **T C C G G T C C C C G G C A C C A** T C C C A C T C T G
attR	A A C C C T A G T T **T C C G G T C C C C G G C A C C A** C G G C C T T T C A

Figure 6-4: Sequence of the KL4 *attP* overlap region in *Burkholderia multivorans* C5274. The identical 17 bp sequence identified in *attL*, the virion, and *attR* is underlined.

CONCLUSIONS AND FUTURE DIRECTIONS

Significance: We have now sequenced and genomically characterized ten phages with potential for use in a Bcc phage therapy protocol. These studies have revealed that each phage has potential advantages and disadvantages with respect to clinical use. In Chapter 2, we showed that KS9 integration does not affect pathogenicity and that this phage can be engineered to a lytic form that is active *in vivo*. However, KS9 has a very narrow host range and our lytic derivative carries a trimethoprim-resistance cassette. In Chapter 3, we analyzed three P2-like phages (KS5, KS14, and KL3) that encoded proteins similar to those of the well-characterized enterobacteria phage P2. Specifically with respect to KS14, this phage lysogenizes as an easily manipulable plasmid, is active *in vivo*, and is stable in a spray-dried powder formulation (194, 267). However, KL3 has a narrow host range and all three of these P2-like phages are temperate. In Chapter 4, we determined that KL1 and AH2 encode MazG proteins that may allow these phages to be active against stationary phase cells in the CF lung. However, both of these phages have a narrow host range, a delayed lysis phenotype *in vitro*, and are predicted to be temperate. In Chapter 6, one of the J2315 transposable myovirus prophages has been shown to be active *in vivo* (267), but KS4-M carries a putative virulence gene and both of these phages are temperate. Chapter 6 also describes two phages with inverse issues: while KS12 is active *in vivo* but cannot be sequenced (267), KL4 has been sequenced but cannot be tested *in vivo*.

One of the Bcc phages with the greatest potential for successful clinical use is DC1, discussed in Chapter 5. This phage has a broad host range (infecting

CF isolates from multiple Bcc species), is closely related to a phage shown to be active in a mammalian model (43), and encodes putative depolymerase and CsrA proteins which may be active against Bcc biofilms. Although DC1 has two disadvantages with respect to clinical use, these issues can be resolved through genetic modification. First, DC1 encodes integrase and repressor proteins and so is putatively temperate. However, similar to observations made by Gill et al. (89), we were unable to find evidence of stable DC1 lysogeny. Therefore, the Bcep22-like phages may not promote some of the negative characteristics associated with temperate phages that form stable lysogens, such as superinfection immunity. Furthermore, the DC1 lysogeny-associated genes could be knocked out, as was performed for KS9c. Second, DC1 encodes a lipid A palmitoyltransferase that could potentially increase host virulence. However, as discussed below, this effect on pathogenicity has not been shown experimentally for either DC1 or BcepIL02 and this gene could also be knocked out prior to therapeutic use (89).

Aside from the assessment of possible phage therapy candidates, the data presented here provide important insights regarding phage evolution and relatedness. Because relatively few Bcc phages have been characterized (particularly in comparison with enterobacteria phages, dairy phages, and mycobacteriophages) (30, 47, 112), each novel phage sequence allows for substantial improvements in our understanding of *Burkholderia* phage distribution, diversity, and relatedness. One of the best such examples is that of the Bcep22-like phages. Bcep22 and BcepIL02, isolated from soil samples in New York and Illinois, respectively, were originally identified as *B. cenocepacia*-

specific phages with sequences that were divergent from any characterized previously (89). As discussed in Chapter 5, the sequencing of one additional phage, DC1, allowed us to conclude that the Bcep22-like phages are part of a novel phage genus with a wide distribution (New York, Illinois, and Alberta), a broad host range (encompassing multiple Bcc species), and putative anti-biofilm activity (89). In Chapters 2 and 3, Bcc phages were found to be closely related to both *B. thailandensis* and *B. pseudomallei* phages, indicating broader relatedness within the *Burkholderia* genus. In Chapter 4, AH2 was found to have a mosaic structure with partial relatedness to BcepNazgul, while KL1 was closely related to a *Pseudomonas* phage and not to other Bcc sequences. The nearly identical lysis phenotypes of these phages suggested that convergent evolution had occurred at the whole genome level. In Chapter 6, we identified single base pair changes in KS4-M and SR1 that were putatively responsible for changes in lysis characteristics, thereby linking sequence evolution to shifts in phenotype. These types of comparative genomic analyses are only feasible following the isolation and sequencing of a broad and diverse panel of *Burkholderia* phages.

Future prospects: There are several reasons why Bcc species are excellent targets for phage therapy. First and foremost, these bacteria are extremely resistant to antibiotic treatment. Aside from the antibiotics commonly used in CF patients – to which Bcc strains are often resistant (231) – there are no broadly effective treatment options for those with Bcc infections. Although some clinical isolates respond to conventional antibiotic treatment, susceptibility is not observed in the majority of cases (332). New classes of drugs active against Bcc

strains, such as enacyloxins, are being explored but remain years from clinical use (184). Because phages are active against antibiotic resistant pathogens, they remain a viable treatment option for the Bcc.

Second, in contrast to many other bacteria, Bcc strains typically lack CRISPR (clustered regularly interspaced short palindromic repeat) regions. CRISPRs are part of defense systems that protect prokaryotes against foreign DNA such as phages and plasmids (187). Although these sequences have been identified in less than half of sequenced bacteria, they are often cited as a potential disadvantage of phage therapy (50, 296). Of the 28 sequenced *Burkholderia* strains included in the CRISPRs database (98) (including Bcc species, *B. gladioli*, *B. glumae*, *B. mallei*, *B. phymatum*, *B. phytofirmans*, *B. pseudomallei*, *B. rhizoxinica*, *B. thailandensis*, and *B. xenovorans*), only one definitive CRISPR sequence was identified. The prevalence of prophages in *Burkholderia* strains (251) is consistent with limited or absent CRISPR immunity. A similar absence of CRISPR loci has been noted in several pathogens (particularly those with an exclusively or partially intracellular lifestyle), including *Chlamydia*, *Borrelia*, and *Brucella* (269). In enterococci, it has been proposed that antibiotic treatment may select against these systems because CRISPR immunity prevents the acquisition of antibiotic resistance genes (225). In a survey of 122 clinical *P. aeruginosa* isolates, CRISPRs were found in only 45 of these strains (although this value may be an underestimation resulting from the methods used) (40). Although further surveys are required, CRISPR immunity may not pose a serious obstacle to phage therapy in the CF lung, particularly for the Bcc.

Third, Bcc phages are dissimilar from those infecting many other pathogenic species in that they do not typically encode virulence genes. Summer et al. (286) have suggested that, because Bcc bacteria live in both the environment and the lung, classical virulence genes may provide less of a benefit than genes that could facilitate survival in both settings. The relative absence of virulence genes in Bcc phages suggests that there may be fewer safety concerns regarding the clinical use of these phages compared to those infecting other hosts. The only Bcc phage genes identified to date with a strong predicted link to pathogenicity and/or resistance encode BcepMu/KS4/KS4-M gp53, a putative O-antigen acyltransferase and BcepIL02 gp22/DC1 gp21, putative lipid A palmitoyltransferases (89, 286, 288). O-antigen acyltransferases may have effects such as serotype conversion (6), while lipid A palmitoyltransferases may increase cationic antimicrobial peptide (CAMP) resistance (102) and decrease TLR4 signaling (138). However, neither one of these proteins has been shown experimentally to produce an effect in *Burkholderia*. Furthermore, proteins that increase CAMP resistance in other species may have a negligible effect as Bcc species are already highly resistant (168). Although both KS4/KS4-M and BcepIL02 encode putative virulence genes, these phages were still shown to be active *in vivo* (43, 267). Prior to clinical use, such genes can be knocked out to ensure that they do not pose a risk to patient safety.

Bcc phage therapy also has certain obstacles. One of the major issues is accessibility within the CF lung. Bcc bacteria reside as part of multi-species biofilms, which are problematic for phage therapy because the cells are both

physically protected from infection and in stationary phase (108, 157, 248, 325). Despite these challenges, many phages have been shown to have activity against biofilms (73), particularly if a depolymerase enzyme is expressed (169). Bcc phages such as DC1, encoding a putative depolymerase enzyme and CsrA (discussed in Chapter 5), KL1/AH2, encoding putative MazG proteins (discussed in Chapter 4), or other Bcc phages engineered to express these proteins may show the greatest activity against stationary phase cells within the biofilm matrix. A second obstacle to phage infection in CF patients is that Bcc bacteria can live intracellularly within epithelial cells and macrophages (35, 255). Phage therapy has been shown to be effective for intracellular pathogens (such as *Mycobacterium tuberculosis* or *Mycobacterium avium*) when phages were released intracellularly by a phage-infected nonpathogenic strain (such as *Mycobacterium smegmatis*) (29). A similar strategy could be used with characterized Bcc strains thought to be nonpathogenic such as *B. vietnamiensis* G4, *B. ambifaria* AMMD, and *B. lata* 383 (263).

A second potential obstacle to Bcc phage therapy is that the vast majority of Bcc phages isolated to date are temperate. This trend is illustrated by the ten phages sequenced and characterized in this thesis, all of which were either temperate or putatively temperate. This result is not entirely surprising, as it is well-established that temperate phages have certain evolutionary benefits (as discussed in Chapter 4). Although temperate or putatively temperate phages have been shown to be active *in vivo* against CF pathogens such as *S. aureus* (41, 197, 228) and the Bcc (43, 176, 267), lytic phages are preferable for clinical use. As

discussed in Chapter 2, an effective way of circumventing this problem is to engineer lytic variants of temperate phages. This solution is relatively complex as the constructed deletions should ideally be unmarked and include all lysogeny-related sequences (particularly the repressor gene, integrase gene, and repressor-binding operators) to prevent reversion to a temperate phenotype by recombination with host prophages. Given the limited number of lytic Bcc phages currently available, genetic modification is a promising method for the development of lytic phages for clinical use.

Although modification of Bcc phages may make them more suitable for clinical use (for example, by removing virulence genes, adding depolymerase, *csrA*, and *mazG* genes, and constructing lytic variants), both the Bcc and its phages are difficult to genetically manipulate. Bcc bacteria are resistant to most antibiotics and compounds commonly used for selection and counterselection and few molecular techniques have been developed specifically for these species (17, 83). These issues are compounded in attempts to genetically manipulate Bcc phages for clinical use, as mutants must be screened instead of directly selected for and as resistance cassettes must not be present. Although we were successful in constructing a KS9 repressor gene knockout (discussed in Chapter 2), the procedure used to make this mutation was inefficient and produced a phage carrying a trimethoprim-resistance cassette.

Recombineering is a technique that uses phage recombination proteins to make marked or unmarked deletions in either bacteria or phages (298). These proteins include the exonuclease/recombinase pairs Exo/Bet from the λ Red

system, RecE/RecT from the Rac prophage, and gp60/gp61 from mycobacteriophage Che9c (297). It is thought that these proteins may be involved in a variety of processes in the phage infection cycle, including replication, chromosome circularization, repair in a DNA-damaging environment or following restriction, and evolution (236). When used in recombineering, these protein pairs promote high levels of recombination using relatively short regions of homology (298). Recently, a recombineering system was developed for mycobacteriophages (that would also be applicable to other phage types) called bacteriophage recombineering of electroporated DNA (BRED) (192). Unmarked deletions, insertions, and point mutations can all be made efficiently in phage genomes using this procedure (192).

Until recently, the λ Red system had not been optimized for use in β-proteobacteria and so could not be used in *Burkholderia* species (137). Recombineering plasmids have now been developed for use in *B. thailandensis* and *B. pseudomallei*, but these have not yet been tested in the Bcc (137). If this system is active in the Bcc, it could then be used to develop a BRED protocol for genetic manipulation of Bcc phages. If this system is not efficient in Bcc strains, an alternative Bcc-optimized recombineering system could be developed using one of the three putative exonuclease/recombinase pairs described here: KL1 gp31/gp33, DC1 gp7/gp6, or KL4 gp11/gp10.

The development of a Bcc BRED system will allow us to pursue three major objectives. The first of these is functional characterization of all Bcc phage genes. Aside from the KS9 repressor, assignment of functions to the proteins

described here has been based solely on bioinformatics. With accompanying experimental data, such as from mutagenesis studies, we would be able to either confirm or refute these annotations. Furthermore, experimental data are needed to discern the function of the numerous hypothetical proteins identified in each genome and to identify genes responsible for certain phenotypes (such as depolymerase activity in DC1). Together with techniques such as RNA-seq (62), BRED will allow us to characterize the expression and function of Bcc phage proteins and to refine our annotations.

As discussed above, the second major objective is the augmentation of phages for therapeutic use. Using BRED, we can delete genes putatively involved in virulence or lysogeny and insert genes with clinical utility. The latter group includes genes encoding depolymerase enzymes, CsrA, or MazG (for activity against biofilms) and tail fiber proteins or methylases (for host range expansion) (203). All of these modifications can be made without resistance cassettes, thus improving phage safety.

The third major objective is the development of a Bcc-specific phage with a minimal genome. Although we had hoped to identify at least one phage that had a broad host range, a strictly lytic lifestyle, and no virulence genes, this ideal therapeutic candidate was not present among the phages characterized here. One of the phages that matches these specifications most closely is KS14: it has a broad tropism, encodes no virulence factors, can be spray-dried into an inhalable powder, is active in the *G. mellonella* model, is closely related to a well-characterized phage and so has few hypothetical proteins, and has a small genome

that is maintained as a plasmid prophage (194, 267). Our goal is to identify the KS14 genes required for the lytic cycle and to construct a KS14 variant containing only those genes. The minimal plasmid genome of this phage could then be used as a scaffold for the addition of other genes such as those encoding depolymerase enzymes. Eventually, using synthetic biology, we can synthesize a KS14 derivative with a modified codon usage and lower GC content so as to minimize the potential for recombination with Bcc prophage sequences.

In a broader context, there are many other goals for future research regarding both Bcc phages and phage therapy development (Figure 1-1). The first is the isolation and characterization of novel phages, ideally those that are lytic with broad host ranges. These phages must be screened for characteristics with clinical utility, such as multi-genus tropism and biofilm degrading activity. New sequencing technologies will allow these novel phages to be sequenced both rapidly and accurately. Similarly, new bioinformatics tools will allow for efficient annotation and comparative analyses that encompass a wide array of phage sequences and provide insights regarding pathogenicity, evolution, and therapeutic development. Following characterization at the sequence level, we can then use this information for further higher-level studies. Phages predicted to be effective for *in vivo* usage can be tested in invertebrate and mammalian models to assess their efficacy both individually and in cocktails. Respirable powders containing these phages can then be developed and tested for both aerosol delivery and long-term storage. Using the results of these higher-level studies, phages can be modified to further improve their activity *in vivo*. Altogether, these

experiments are the foundation for the development of an effective phage therapy strategy for the Bcc.

Appendix

Development of a species-specific *fur* gene-based method for identification of the *Burkholderia cepacia* complex

A version of this chapter has been published as:

Lynch, K. H., and J. J. Dennis. 2008. Development of a species-specific *fur* gene-based method for identification of the *Burkholderia cepacia* complex. J. Clin. Microbiol. **46:**447-455.

INTRODUCTION

In addition to the transmissibility, pathogenicity, and antibiotic resistance of Bcc species, a fourth serious concern regarding these organisms is misdiagnosis. In order to provide optimal antibiotic therapy and to minimize transmission (from infected patients and/or contaminated instruments), Bcc isolates need to be identified by clinical laboratories both quickly and accurately (199). However, commercial systems used for bacterial identification by these labs a) often identify other Gram-negative nonfermenters (such as *Burkholderia gladioli*, *Pseudomonas aeruginosa*, *Stenotrophomonas maltophilia*, and *Alcaligenes xylosoxidans*) as Bcc species and vice versa and b) do not identify Bcc isolates to the species level, even though this classification is correlated with both pathogenicity and prevalence (13, 199, 271). In the United States, false-positive and false-negative rates as high as 11% and 36%, respectively, have been reported for the Bcc (199).

In the last decade, various molecular typing methods have been evaluated to accurately identify and categorize strains belonging to the Bcc (175, 299). Although 16S rDNA sequence analysis is an acceptable means of differentiating many bacteria, it is of limited use in separating Bcc species due to high (>97.7%) sequence identity (28, 58, 164, 182). Techniques using the recombinase A (*recA*) gene locus have proven to be the most successful. Originally designed to be used primarily as an assay to separate Bcc strains at the species level, recent modifications to this procedure permit discrimination between all *Burkholderia* genus members (182, 229, 309). *RecA*-based typing has been used to identify Bcc

strains from both environmental samples and sputa, as well as to identify other *Burkholderia* species (51, 66, 209, 240). However, misidentification of Bcc species has occurred using this approach as a restriction fragment length polymorphism (RFLP)-based strategy (209). Furthermore, the medically important Bcc member *B. cenocepacia* is divided into four different *recA* phylotypes, complicating the identification of this species (301).

Recently, multilocus sequence typing (MLST) schemes have been developed for the precise differentiation of the species and strains of both the Bcc and other *Burkholderia* species (15, 281). In this protocol, seven conserved housekeeping genes are PCR amplified, sequenced, and classified under allelic profiles to identify the sequence type of an isolate (15, 281). Extensive nucleotide sequence diversity was found within all seven Bcc loci, ranging from 13.1% for *atpD* to 37.4% for *gyrB* (15). This MLST scheme differentiated nine Bcc species (those identified at the time of publication) and 114 of 119 Bcc strains (15). Unfortunately, DNA sequencing and comparison of seven gene targets for each Bcc isolate are not within the capabilities of many clinical laboratories. Therefore, simpler, effective strategies for Bcc classification are still needed.

In spite of the numerous methods available for separating Bcc species, a single well-conserved gene locus that would provide a simple method for unambiguous species-specific identification is highly desirable. While studying the potential of using virulence factor genes for the purpose of distinguishing virulent Bcc strains from environmental strains, we discovered that the Bcc ferric uptake regulator (*fur*) gene contained sufficient single nucleotide polymorphisms

(SNPs) to differentiate between Bcc species. The objectives of this project were to a) use *fur* gene SNPs to develop a sequencing- and PCR-based Bcc diagnostic test and b) determine if this test could be used to accurately identify Bcc CF isolates to the species level.

Note: At the time that this study was completed, only nine of the current seventeen Bcc species had been identified and *B. contaminans*/*B. lata* strains were classified as *B. cepacia recA* group K.

MATERIALS AND METHODS

Bacterial strains and growth conditions. The 73 bacterial strains and isolates used in this study are shown in Table A-1. These were primarily members of the Bcc experimental strain panel (183) and the updated Bcc strain panel (59), obtained from the Belgium Coordinated Collection of Microorganisms LMG Bacteria Collection (Ghent, Belgium), the American Type Culture Collection (Manassas, VA), and the Canadian *Burkholderia cepacia* Complex Research and Referral Repository (Vancouver, BC). Clinical isolates putatively identified as being Bcc species (by diagnostic metabolism tests) were provided by the University of Alberta Hospital (pediatric/adult) CF clinic. Growth of these isolates on *Burkholderia cepacia* selective agar (BCSA) (118) was assessed following overnight aerobic incubation at 30°C. For characterized strains, species assignment was based on results from previously published polyphasic analyses (59, 183, 301). Strains and isolates were maintained at -80°C by freezing in Luria-Bertani (LB) medium containing 20% glycerol. Before use, bacteria were plated

on half-strength LB (½ LB) solid medium and grown aerobically overnight at 30°C.

Table A-1: Strains and clinical isolates used for *fur* assay development and testing

Species	Strain or isolate	Source	Location	Growth on BCSA	Amplification with JD490/JD491	Amplification with species-specific primers
Burkholderia cepacia	**CEP509*/LMG 18821***	CF	Australia	N/T	+	+
	ATCC 25416^{T}*/LMG 1222^{T}*	ENV	USA	N/T	+	+
	ATCC 17759*/LMG 2161*	ENV	Trinidad	N/T	+	+
Burkholderia cepacia recA group K	CEP0964	CF	Canada	N/T	+	+
	CEP1056	CF	Canada	N/T	+	+
	R445	CF	Canada	+	+	+
Burkholderia multivorans	**LMG 13010^{T}***	CF	Belgium	N/T	+	+
	ATCC 17616*/LMG 17588*	ENV	USA	N/T	+	+
	C3430	CF	Canada	N/T	+	+
	C5274	CF	Canada	N/T	+	+
	C5393*/LMG 18822*	CF	Canada	N/T	+	+
	C5568	CF	Canada	N/T	+	+
	M1512	CF	Canada	+	+	+
	M1865	CF	Canada	+	+	+
	R810	CF	Canada	+	+	+
	R1159	CF	Canada	+	+	+
Burkholderia cenocepacia	**K56-2*/LMG 18863***	CF-e	Canada	N/T	+	+
	J2315*/LMG 16656*	CF-e	UK	N/T	+	+
	C1257	CF-e	Canada	N/T	+	+
	C4455	CF-e	Canada	N/T	+	+
	C5424*/LMG 18827*	CF-e	Canada	N/T	+	+
	C6433*/LMG 18828*	CF-e	Canada	N/T	+	+
	CEP511*/LMG 18830*	CF-e	Australia	N/T	+	+
	D1	ENV	USA	N/T	+	+

	PC184*/LMG 18829*	CF-e	USA	N/T	+	+
	LMG 19240	ENV	Australia	N/T	+	-
	CEP0868/LMG 21461	CF	Argentina	N/T	+	+
	R161	CF	Canada	+	+	+
	R452	CF	Canada	+	+	+
	R750	CF	Canada	+	+	+
	R1284	CF	Canada	+	+	+
	R1285	CF	Canada	+	+	+
	R1314	CF	Canada	+	+	+
	R1434	CF	Canada	+	+	+
	R1619	CF	Canada	+	+	+
	R1882	CF	Canada	+	+	+
	R1883	CF	Canada	+	+	+
	R1884	CF	Canada	+	+	+
	R2314	CF	Canada	+	+	+
	S11528	CF	Canada	+	+	+
Burkholderia stabilis	**LMG 14294^T***	CF	Belgium	N/T	+	+
	C7322*/LMG 18870*	CF	Canada	N/T	+	+
	R450	CF	Canada	+	+	+
	R2140	CF	Canada	+	+	+
	R2339	CF	Canada	+	+	+
Burkholderia vietnamiensis	**LMG 10929^T***	ENV	Vietnam	N/T	+	+
	PC259*/LMG 18835*	CF	USA	N/T	+	+
	ATCC 29424	ENV	USA	N/T	+	+
	G4	ENV	USA	N/T	+	+
Burkholderia dolosa	**AU0645^T*/LMG 18943^T***	CF	USA	N/T	+	+
	CEP021*/LMG 21819*	CF	USA	N/T	+	+
	E12*/LMG 21820*	CF	UK	N/T	+	+
	L06	CF		N/T	+	+
	STM1441*/LMG 21443*	ENV	Senegal	N/T	+	+

Burkholderia ambifaria	**ATCC 53266*/LMG 17828***	ENV	USA	N/T	+	+
	AMMD^{T}*/LMG 19182^{T}*	ENV	USA	N/T	+	+
	CEP0996*/LMG 19467*	CF	Australia	N/T	+	+
	M53	ENV	USA	N/T	+	+
Burkholderia anthina	**W92^{T}*/LMG 20980^{T}***	ENV	USA	N/T	+	+
	C1765*/LMG 20983*	CF	UK	N/T	+	+
	J2552*/LMG 16670*	ENV	UK	N/T	+	+
	AU1293*/LMG 21821*	CF	USA	N/T	+	+
Burkholderia pyrrocinia	**ATCC 15958^{T}*/LMG 14191^{T}***	ENV	Japan	N/T	+	+
	ATCC 39277*/LMG 21822*	ENV	USA	N/T	+	+
	BC011*/LMG 21823*	ENV	USA	N/T	+	+
	C1469*/LMG 21824*	CF	UK	N/T	+	+
Burkholderia gladioli	*R406*	CF	Canada	+	+	N/A
	R1879	CF	Canada	+	+	N/A
Pseudomonas aeruginosa	*R285*	CF	Canada	-	-	N/A
Herbaspirillum sp.	*R740*	CF	Canada	-	-	N/A
Listeria monocytogenes	*R1653*	CF	Canada	-	-	N/A
Pandoraea sp.	*R1717*	CF	Canada	+	-	N/A
Burkholderia sp.	JS150	ENV	USA	N/T	+	N/A

Abbreviations: BCSA, *Burkholderia cepacia* selective agar; T, type strain; ENV, environmental isolate; CF, cystic fibrosis isolate; CF-e, cystic fibrosis epidemic isolate; N/A, not applicable; N/T, not tested. Strains in italicized type are clinical isolates putatively identified as being Bcc species (provided by the University of Alberta Hospital CF Clinic). Strains in boldface type were used to test each of the nine species-specific primer sets. Asterisks indicate strains of the Bcc experimental strain panel (59, 183). Reproduced with permission from the American Society for Microbiology.

DNA preparation and PCR. Genomic DNA was prepared using a standard protocol (12). Gene amplification was performed in a total volume of 50 µl containing 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 1.5 mM $MgSO_4$, 0.2 mM of each deoxynucleoside triphosphate, 50 pmol of each primer, 1.25 units of *Taq* PCRx DNA polymerase (Invitrogen, Burlington, ON), 5 µl of PCRx enhancer solution (Invitrogen), and 10 ng of genomic DNA. PCR was performed with an Eppendorf Mastercycler gradient DNA thermal cycler (Westbury, NY) under the following conditions: 96°C for 2 min for the first cycle, followed by 30 cycles of 96°C for 1 min, 55°C for 1 min, and 72°C for 1 min, with a final extension step of 72°C for 2 min. PCR products were separated on 0.8% (wt/vol) agarose gels in 1x Tris-acetate-EDTA (pH 8.0).

fur gene sequencing. The *fur* gene (*B. cenocepacia* J2315 chromosome 1, base pairs [bp] 3,702,286-3,702,714) was initially amplified from various Bcc strains using primer set JD490/JD491 (Table A-2). The DNA sequences were determined by direct sequencing from amplicons purified using a GENECLEAN II kit (Qbiogene, Irvine, CA) or cloned into pJET1.2/blunt (Fermentas, Burlington, ON). Nucleotide sequences were determined at least once on each DNA strand using BigDye Terminator reaction mixtures according to the manufacturer's recommendations (Applied Biosystems, Foster City, CA). Products were separated and collected using an ABI Prism 3100 genetic analyzer (Applied Biosystems) using standard sequencing conditions. The sequences were aligned and edited using EditView and AutoAssembler software (Applied Biosystems) and analyzed using BLASTN (NCBI) to verify amplification of the

correct product. When aligned, the *fur* gene sequences from different Bcc species showed heterogeneity at several sites. These sites were used to design PCR primers that would amplify species-specific products. Primer sets were designed to possess similar melting and optimum annealing temperatures and to amplify a *fur* gene product from a specific Bcc species (or set of species) (Table A-2). All primers were purchased from Sigma/Genosys Canada (Oakville, ON). To identify strains that were not amplified by JD490/JD491, 16S rDNA sequences were determined. Primer set 27F/1522R was used to amplify a partial region of the 16S rDNA (134) which was cloned into pJET1.2/blunt (Fermentas) and sequenced as described above.

Table A-2: Species-specific primers used for amplification of the Bcc *fur* gene

Species (genomovar)	Forward primer	Forward primer sequence	Forward primer positions	Reverse primer	Reverse primer sequence	Reverse primer positions	Product length (bp)	Bcc species amplified
B. cepacia (I)/ *B. cepacia recA* group K	F1	GGCNGAAGACGTCTACCGG	102-120	R1	TCGAAGTTGCTGCGCGAC	201-218	117	*B. cepacia*
B. multivorans (II)	F2	AGCAGAGCCCCGTGCGG	77-93	R2	GGTGGGGGCAGTTTTCGGTG	399-418	342	*B. multivorans*
B. cenocepacia (III)	F	TGACCAATCCGACCGATCTCA	2-22	R3	ATCGCCTGCTGGCGGCTC	321-338	337	*B. cenocepacia* IIIA, IIIB, IIID
B. stabilis (IV)	F4	CNACCGTCTATCGCGTGCTC	155-174	R	TCAGTGCTTGCGGTGGGG	412-429	275	*B. multivorans, B. cenocepacia, B. stabilis, B. dolosa*
B. vietnamiensis (V)	F	TGACCAATCCGACCGATCTCA	2-22	R5	CGTGGTGGGAGCCTTCGTTG	243-262	261	*B. vietnamiensis*
B. dolosa (VI)	F6	CTAAAGGCCACCCTACCGCGG	34-54	R	TCAGTGCTTGCGGTGGGG	412-429	396	*B. dolosa*
B. ambifaria (VII)	F7	CCCNGTGCGTCACCTGACT	84-102	R7	CGTGGTGCGAACCTTCATTCAA	241-262	179	*B. ambifaria*
B. anthina (VIII)	F	TGACCAATCCGACCGATCTCA	2-22	R8	CAGGTGACGCACGGGGCTC	81-99	98	*B. multivorans, B. cenocepacia, B. stabilis, B. anthina, B. pyrrocinia* (1 strain)
B. pyrrocinia (IX)	F1	GGCNGAAGACGTCTACCGG	102-120	R9	ATCGCCTGCTGGCGGCC	322-338	237	*B. pyrrocinia*
All (I-IX)	JD490	ATGACCAATCCGACCGATCTCAA	1-23	JD491	TCAGTGCTTGCGITNIGGGCAGTT	406-429	429	All

Phylogenetic analysis. To construct phylogenetic trees of the assembled and edited *fur* gene sequences, the Clustal V algorithm (124) was used as part of the alignment software included with the MegAlign program (DNASTAR, Madison, WI). This algorithm groups sequences into clusters first individually and then collectively to produce an overall alignment. At each alignment stage, a two-sequence alignment method is utilized using algorithms described previously by Wilbur and Lipman for input sequences (319) or by Myers and Miller for ancestral consensus sequences (213). The general approach is to progressively align groups of sequences according to a branching order in a hypothetical phylogenetic tree, with gaps occurring in earlier alignments preserved throughout later alignments. The final phylogenetic relationships are constructed by applying the neighbor-joining method of Saitou and Nei to the distance and alignment data (256). Confidence levels of individual branches within the optimal tree were assessed by pseudoreplicate data set resampling or bootstrap analysis. The resulting rooted phylogenetic tree includes several non-Bcc *fur* sequences as outliers, including *B. pseudomallei* (accession number AF117238), *Burkholderia* sp. strain JS150, and *B. xenovorans* LB400 (accession number NC_007951). EditSeq (DNASTAR) was used to modify the original sequences to files that could be assembled by MegAlign. Nearest-neighbor interchange analysis with these data using the Hein algorithm produced a similarly branched tree (115). By examining sequence pairs, this analysis builds a phylogeny represented by a graph of possible alignments where individual sequences entered into the graph retain a relationship with other sequences that allow the algorithm to ignore the

exploration of distant pair relationships. The phylogeny created by this process is finally reexamined for the best possible arrangement of ancestral branches. The phylogenetic analysis methods providing the most discriminatory results are presented in Figure A-1.

Figure A-1: Phylogenetic tree comparing sequences of the 429 bp *fur* gene from strains and isolates of the Bcc. The phylogenies are rooted due to the assumption of a common ancestor and biological clock. Genetic distance is shown on the scale, with demarcations representing 2% estimated substitutions. The genomovar status (I through IX [Table A-2]) of each Bcc species is shown in parentheses following the strain or isolate name. Abbreviations: Bg, *Burkholderia gladioli*; Bp, *Burkholderia pseudomallei*; Bsp, *Burkholderia* sp.; Bx, *Burkholderia xenovorans*. Reproduced with permission from the American Society for Microbiology.

Nucleotide sequence accession numbers. The *fur* gene and 16S rDNA sequences have been deposited in GenBank under accession numbers EU090823-EU090895, EU273473, and EU273474.

RESULTS

Analysis of *fur* gene sequences. Following chromosome isolation, 73 strains and clinical isolates (Table A-1) were tested in a PCR protocol using primer set JD490/JD491. A region of approximately 400 bp was amplified from 69 of these bacteria (Table A-1). Each strand of these DNA products was sequenced, producing 429 bp contigs. BLASTN analysis showed that each of these contigs was homologous to the *Burkholderia fur* gene, indicating that the PCR protocol had amplified the correct product in each case. These 69 contigs exhibited sufficient sequence similarity to be organized into a single MegAlign alignment. Sequences from strains that had been previously identified to the species level were nearly identical within a single Bcc species. Among different species, however, characteristic SNPs that differentiated groups of species from each other and, in several cases, differentiated one species from all others were present.

The 69 aligned *fur* sequences were used to construct a phylogenetic tree, with sequences from *B. xenovorans, Burkholderia* sp. strain JS150, and *B. pseudomallei* included as outliers (Figure A-1). The clinical isolates R406 and R1879, although putatively identified as being Bcc species, also appeared to be outliers (Figure A-1). To confirm the identities of these two isolates, 16S rDNA

sequencing was performed with primer set 27F/1522R (134). Sequence data showed that these two isolates were not part of the Bcc but were instead *B. gladioli*, a closely related *Burkholderia* species. Both of these isolates were able to grow on BCSA selective medium (Table A-1).

Of the Bcc strains that had been previously characterized, 38 of 40 grouped according to species level on the tree (Figure A-1). The two exceptions were C1469 (a *B. pyrrocinia* strain that appears to be more closely related to *B. cepacia recA* group K isolates than to other *B. pyrrocinia* strains) and LMG 19240 (a *B. cenocepacia* strain that belongs to the *recA* IIIC subgroup) (Figure A-1). The *fur* gene sequence of LMG 19240 is substantially different from the sequences of the other *B. cenocepacia* strains and isolates tested and it forms an isolated *B. cenocepacia* IIIC branch on the tree (Figure A-1). Despite these two exceptions, the strong correlation between the position of any Bcc strain's *fur* sequence on the tree and its species designation suggested that the addition of *fur* gene sequences from unidentified strains would allow unambiguous identification at the species level. As predicted, the 21 uncharacterized clinical isolates clustered on the tree among the previously characterized strains (Figure A-1). The largest cluster of isolates, 13 of 21 (62%), was *B. cenocepacia* based on position in the tree. The next largest cluster was *B. multivorans*, with 4 of 21 isolates (19%). These values are consistent with previous studies in which the majority of Bcc clinical isolates are either *B. cenocepacia* or *B. multivorans* (254). The remainder of the uncharacterized Bcc clinical isolates belonged to *B. stabilis* (three isolates [14%]) and *B. cepacia recA* group K (one isolate [5%]) (Figure A-1).

In contrast to *recA*-based protocols, *fur* gene sequence analysis separates strains and isolates of a single species into a single cluster on the phylogenetic tree (with the exception of C1469 and LMG 19240) (Figure A-1). However, some subgroups are present within these clusters. For example, *B. multivorans* strains and isolates branch into two subgroups, the first including M1512, M1865, R810, C5274, and LMG 13010 and the second including ATCC 17616, C3430, C5393, R1159, and C5568 (Figure A-1). For previously characterized *B. cenocepacia* strains, the subgroups based on the *fur* gene sequence match the groupings assigned by *recA* analysis as IIIA to IIID strains form discrete subgroups on the tree (Figure A-1). The cluster formed by the IIIA strains is the most diffuse: it branches into two subgroups, the first including R1284, R1314, and C1257 and the second including all other IIIA strains and isolates tested (Figure A-1). Based on this analysis, all of the uncharacterized *B. cenocepacia* clinical isolates are most closely related to strains in the IIIA subgroup.

16S rDNA sequences were determined for the four clinical isolates putatively identified as being Bcc species that could not be amplified using primer set JD490/JD491 (Table A-1). Using primer set 27F/1522R, a product of approximately 1,500 bp was obtained for each sample. Sequencing and comparison of these amplified products indicated that these isolates did not belong to the Bcc and were instead the Gram-positive organism *Listeria monocytogenes* and the Gram-negative organisms *P. aeruginosa, Herbaspirillum* sp., and *Pandoraea* sp. (Table A-1). Of these four isolates, only *Pandoraea* sp. was able to grow on BCSA (Table A-1).

Design and application of species-specific primers. While phylogenetic analysis of the *fur* gene sequence allowed Bcc strains and isolates to be easily and accurately identified to the species level (Figure A-1), a complementary PCR protocol was developed to make the classification process more rapid. There are several sites in the *fur* gene where a single base pair difference can be used to differentiate strains of one Bcc species from others. These SNPs were used to design primer sets that would amplify the *fur* gene in a species-specific fashion. These primers, the expected products, and the species from which the products can be amplified are shown in Table A-2. Primer sets were tested against a panel of nine representative Bcc strains (boldface in Table A-1) to verify that a product of the expected size would be amplified for only one (or a select few) of the nine species. Each of these PCR products was sequenced to confirm that the correct region had been amplified.

The primers designed for *B. cepacia*/*B. cepacia recA* group K, *B. multivorans, B. cenocepacia, B. vietnamiensis, B. dolosa, B. ambifaria*, and *B. pyrrocinia* were highly specific. When tested with the nine representative Bcc strains, these primer sets amplified a product of the expected size from only a single species. Although a small number of nonspecific products were amplified in some cases, these amplicons could be distinguished from the expected product by their sizes (Figure A-2). In addition to the nine representative Bcc strains, these primer sets were tested with an additional 50 strains and isolates: 5 *B. cepacia* and *B. cepacia recA* group K strains/isolates for primer set F1/R1, 9 *B. multivorans* strains/isolates for primer set F2/R2, 23 *B. cenocepacia* strains/isolates for primer

set F/R3, 3 *B. vietnamiensis* strains for primer set F/R5, 4 *B. dolosa* strains for primer set F6/R, 3 *B. ambifaria* strains for primer set F7/R7, and 3 *B. pyrrocinia* strains for primer set F1/R9 (Table A-1). In each reaction, a product of the expected size was amplified (except for LMG 19240, which lacks the *fur* gene polymorphism used to design the F/3R primer set) (Table A-1).

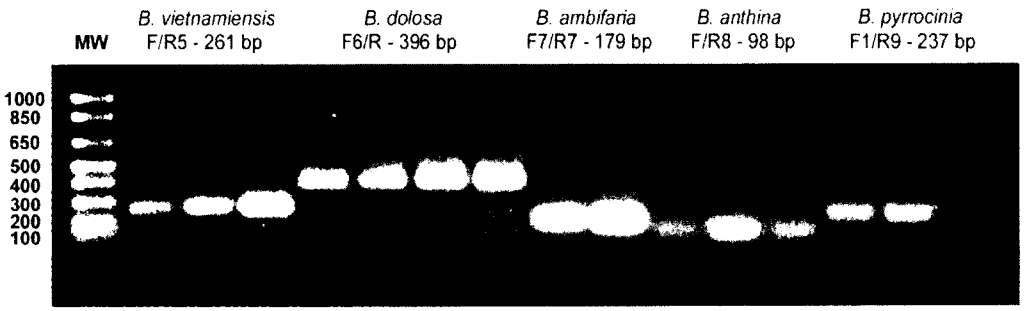


Figure A-2: Gel electrophoresis of *fur* PCR products for *Burkholderia vietnamiensis, Burkholderia dolosa, Burkholderia ambifaria, Burkholderia anthina*, and *Burkholderia pyrrocinia*. Amplicons were separated by electrophoresis on a 0.8% agarose gel prior to visualization with ethidium bromide. Molecular weight standards (MW) are shown in lane 1, with the corresponding base pair sizes shown at the left of the gel. Shown below each Bcc species name is the *fur* primer set used and the size of the expected PCR product with the following template DNAs: lane 2, PC259/LMG 18835; lane 3, G4; lane 4, ATCC 29424; lane 5, STM1441/LMG 21443; lane 6, CEP021/LMG 21819; lane 7, E12/LMG 21820; lane 8, L06; lane 9, CEP0996/LMG 19467; lane 10, M53; lane 11, J2552/LMG 16670; lane 12, C1765/LMG 20983; lane 13, AU1293/LMG 21821; lane 14, ATCC 39277/LMG 21822; lane 15, BC011/LMG 21823; lane 16, C1469/LMG 21824. Reproduced with permission from the American Society for Microbiology.

The primers designed for *B. stabilis* and *B. anthina* were less specific, as the polymorphisms used to design these two primer sets are each present in four species (Table A-2). More specific primer sets could not be designed because of the relatively low number of heterogeneous base pairs in the *fur* gene specific to each of these species. However, discrimination was still possible by designing F4/R (tested with an additional four *B. stabilis* strains/isolates [Table A-1]) to amplify *B. multivorans, B. cenocepacia, B. stabilis*, and *B. dolosa*, while F/R8

(tested with an additional three *B. anthina* strains [Table A-1]) was designed to amplify *B. multivorans, B. cenocepacia, B. stabilis*, and *B. anthina* (Table A-2). In addition, this primer set has the potential to amplify internal *fur* gene products from *B. pyrrocinia* as demonstrated with strain BC011. Although these primer sets are relatively non-specific, they can still be used to differentiate *B. stabilis* and *B. anthina* from other Bcc species. For example, using DNA from a *B. stabilis* isolate will produce an amplicon using primer set F4/R but not F6/R, whereas *B. dolosa* genomic DNA will produce an amplicon in both reactions (Table A-2). Similarly, *B. anthina* DNA will produce an amplicon using primer set F/R8 but not F1/R9, whereas *B. pyrrocinia* DNA will produce an amplicon using F1/R9 and potentially F/R8 (Table A-2). Together, the nine primer sets described here allow for the rapid discrimination of Bcc species from one another.

DISCUSSION

Because of the devastating potential impact of a false-positive or false-negative diagnosis, Bcc infections must be identified both rapidly and accurately in a clinical setting (279). Unfortunately, there are various problems with current diagnostic tests for Bcc species, leading to unacceptably high misidentification rates (199). Therefore, the further development of diagnostic tests that are able to differentiate Bcc species is required.

We have developed a simple, specific, and accurate procedure to identify Bcc isolates to the species level using the sequence of the *fur* gene. As shown in Figure A-1, for 97% of the strains and isolates tested, members of a single species

cluster together on a phylogenetic tree based on these sequences. To identify the species of an uncharacterized strain or isolate with relative certainty, one can use primer set JD490/JD491 to PCR amplify and sequence the *fur* gene for comparison with those of other Bcc species. Thus, one of the main benefits of this procedure is its simplicity. Another benefit of this protocol is that the JD490/JD491 PCR reaction will not produce an amplicon if the isolate is not of the genus *Burkholderia*, thus rapidly discriminating between *Burkholderia* and non-*Burkholderia* isolates (including *Listeria*, *Pseudomonas*, *Herbaspirillum*, and *Pandoraea* [Table A-1]). In our analysis of 27 previously uncharacterized clinical isolates, negative JD490/JD491 PCR amplification results were the first indication that at least four of these samples did not belong to the Bcc (Table A-1). Growth on BCSA was an unreliable measure in this case, as one of these four isolates (*Pandoraea* sp.) was able to grow on this selective medium.

In order to take advantage of SNPs present in the *fur* gene, a PCR protocol that would allow for the rapid determination of Bcc species status was designed. Using this procedure, one can identify the species to which a strain or isolate belongs using only one set of nine PCR reactions. The protocol was designed such that the sequencing data, which are highly accurate but take time to collect, can be used in concert with the PCR, which is less specific but more rapid. It therefore fulfils the need for a diagnostic procedure that is both effective and efficient at identifying clinical and environmental Bcc samples.

In practical use, the procedure was successful in classifying 27 clinical isolates from CF patients putatively identified as being Bcc species. As predicted,

the majority of the isolates were either *B. cenocepacia* or *B. multivorans*, with a lesser number identified as being either *B. stabilis* or *B. cepacia recA* group K. This system was also able to determine that 6 of the 27 isolates had been misidentified and were actually *B. gladioli, P. aeruginosa, Herbaspirillum* sp., *Pandoraea* sp., or *L. monocytogenes* (Table A-1). As shown in Table A-1, three of these six isolates were able to grow on BCSA selective medium. These results further underscore the need for a new, more accurate practical method for examining clinical specimens.

The *fur* gene protocol has a number of advantages over tests developed previously, as a) it does not rely on biochemical measures, b) it can identify strains to the species level (unlike 16S rDNA), c) with few exceptions, it identifies members of a single species as part of a single group (unlike *recA*), and d) it uses a single gene target (unlike MLST). The PCR can be completed rapidly with relatively inexpensive reagents and equipment and does not require special training to complete. This protocol quickly separates the more important clinical isolates – *B. multivorans* and *B. cenocepacia* – from those species that are less able to cause acute disease or to undergo epidemic spread in a susceptible population.

One limitation of this protocol is that it was unable to classify a *recA* IIIC isolate as part of the *B. cenocepacia* cluster. *B. cenocepacia* IIIC may be significantly different at other loci besides *fur* and further testing should be carried out in order to confirm its phylogenetic position among Bcc strains. All IIIC strains characterized to date are environmental isolates, suggesting that this

lineage may not be prevalent clinically (301). If necessary, sufficient *fur* gene polymorphisms exist to allow the design of a IIIC subgroup-specific primer.

A second limitation of this protocol, common to all PCR-based classification systems, is that mutation of a single base pair could prevent amplification by the correct primer set or allow amplification by an incorrect primer set. However, companion DNA sequencing and comparison of the complete *fur* gene would still position the strain correctly in the phylogenetic tree. Furthermore, this gene could alternatively be used as an additional locus in an MLST protocol. Analysis of *fur* sequences from the 66 Bcc strains and isolates examined here indicates that this gene has a nucleotide sequence diversity of 18%, which is intermediate to that of other genetic loci used for MLST (such as *atpD* [13.1%] and *recA* [26.7%]) (15). This result suggests that the Bcc *fur* gene sequence is conserved enough to be accurate in PCR-based diagnostics, but still variable enough to be used as a gene locus candidate in a MLST scheme.

A third limitation of this protocol is that it has not been updated to the current Bcc taxonomy. Subsequent to the publication of this study in February 2008, an additional eight species were added to the Bcc: *B. latens*, *B. diffusa*, *B. arboris*, *B. seminalis*, *B. metallica*, and *B. ubonensis* in July 2008 (305) and *B. contaminans* and *B. lata* in January 2009 (304). Both clinical and environmental strains have been isolated for each of these species, excluding *B. metallica* and *B. latens* (which have thus far been isolated exclusively from CF patients) (299). The *fur* protocol has not yet been updated to reflect the current taxonomy. However, as *B. cepacia recA* group K isolates (now *B. contaminans* and *B. lata*) were found to

cluster on the phylogenetic tree in Figure A-1, we predict that strains of these newly identified species will cluster in a similar fashion and be identifiable using this protocol. Furthermore, because the vast majority of clinical isolates are still either *B. multivorans* or *B. cenocepacia*, the current *fur* protocol remains effective in most cases.

CONCLUSIONS

In summary, we have developed an effective method of differentiating Bcc strains and isolates based on their *fur* gene sequences. Although further large-scale sensitivity and specificity testing is required to validate the clinical utility of this method, it represents a potentially very useful advance in rapid Bcc diagnostics. This protocol can be used in both clinical and research laboratories to quickly and accurately classify Bcc isolates. This Bcc *fur* gene-based method permits unambiguous identification at the species level and, as evidenced by the resulting phylogenetic tree, creates a population structure that coincides with current species assignments for the Bcc. It is able to resolve unknown clinical isolates into specific Bcc species, thereby providing a comprehensive framework for strain classification and epidemiological evaluation.

Since its publication, this protocol has been used for environmental characterization and clinical diagnostics in at least three international laboratories. In 2010, Ma et al. (180) isolated the lipase-expressing strain C737-11 from an oil-contaminated soil sample in China. Preliminary physiological tests suggested that this isolate is a Bcc strain. 16S rDNA sequencing confirmed that C737-11 belongs

to the Bcc, but it could not be identified to the species level because both *B. cepacia* and *B. cenocepacia* sequences clustered with this isolate on a phylogenetic tree. Using *fur* sequencing, C737-11 could clearly be identified as a *B. cepacia* strain (Figure A-3), indicating that this protocol is more effective than 16S rDNA sequencing for species-level identification and that it can be used for both environmental and clinical isolates.

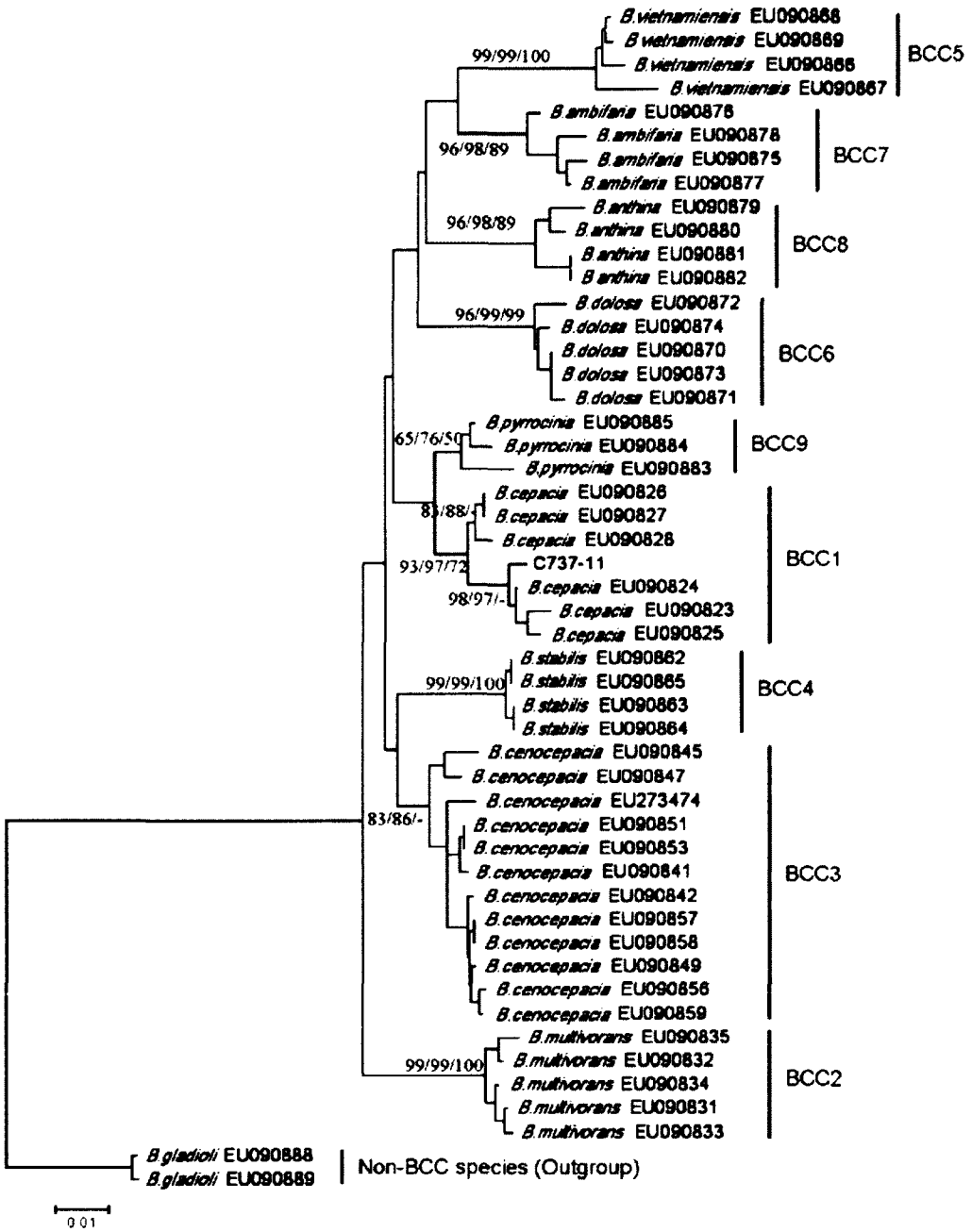


Figure A-3: Phylogenetic analysis of C737-11 based on ferric uptake regulator (*fur*) gene sequences. The tree was rooted with (*fur*) gene sequences from *Burkholderia gladioli* (EU090888, EU090889). The sequences were aligned with ClustalX 2.0, using the Blosum 30 matrix. The alignment result was exported to Mega3.1. Phylogenetic trees were calculated with NJ, ME and MP methods. The topologies of trees generated with different methods were similar. The topology presented in this figure was generated with the NJ method. The bootstrap values provided on the branches were calculated using the NJ, ME, and MP methods with 1,000 replicates, respectively. Figure and legend reproduced from Ma et al. (180) with permission from Springer.

In 2008, Heo et al. (121) used a commercial system, *fur* sequencing, and *recA* RFLP analysis in combination to identify a Bcc outbreak in a South Korean cancer centre. It was determined that a chlorhexidine solution used to disinfect skin prior to venous catheter insertion was contaminated with *B. stabilis*, resulting in bloodstream infections in eight leukemia patients (121). After *B. stabilis* was identified as the infectious agent, the chlorhexidine solutions were discarded and all eight patients recovered following treatment with either trimethoprim/sulfamethoxazole or meropenem (121).

In 2010, Nørskov-Lauritsen et al. (217) used *atpD*, *recA*, and *fur* sequencing to characterize Bcc isolates from Danish CF patients. Notably, *fur* sequencing was used in cases where no PCR product could be obtained using *atpD* or *recA* primers (217), indicating that the *fur* primers designed in this study have (in certain cases) broader utility than two of the primer sets designed for MLST analysis. A unique distribution of species was observed in these patients: 92% of Bcc isolates were *B. multivorans*, 6% were *B. cenocepacia*, and 2% were *B. anthina* (217). As pulsed-field gel electrophoresis analysis indicated that only a minority of patients carried the same strain type, it was suggested that the majority of the *B. multivorans* infections were acquired directly from the environment (a similar conclusion to that made by Baldwin et al. [13]) (217). These three international studies provide independent validation that the *fur* protocol is an effective diagnostic tool for use in both environmental studies and clinical laboratories.

ACKNOWLEDGMENTS

We thank Pamela Sokol, David Speert, and Deb Henry of the Canadian *Burkholderia cepacia* Complex Research and Referral Repository for providing bacterial strains. We also gratefully acknowledge Robert Rennie and LeeAnn Turnbull of the University of Alberta Hospital CF Clinic for providing clinical isolates.

BIBLIOGRAPHY

1. **Aaron, S. D., W. Ferris, D. A. Henry, D. P. Speert, and N. E. MacDonald.** 2000. Multiple combination bactericidal antibiotic testing for patients with cystic fibrosis infected with *Burkholderia cepacia*. Am. J. Respir. Crit. Care Med. **161:**1206-1212.

2. **Abedon, S. T.** 2009. Phage evolution and ecology. Adv. Appl. Microbiol. **67:**1-45.

3. **Abeles, A. L., S. A. Friedman, and S. J. Austin.** 1985. Partition of unit-copy miniplasmids to daughter cells III. The DNA sequence and functional organization of the P1 partition region. J. Mol. Biol. **185:**261-272.

4. **Ackermann, H-W.** 2001. Frequency of morphological phage descriptions in the year 2000. Arch. Virol. **146:**843-857.

5. **Ackermann, H-W.** 2007. 5500 Phages examined in the electron microscope. Arch. Virol. **152:**227-243.

6. **Allison, G. E., and N. K. Verma.** 2000. Serotype-converting bacteriophages and O-antigen modification in *Shigella flexneri*. Trends Microbiol. **8:**17-23.

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

8. **AmpliPhi Biosciences.** 2011. Product Pipeline. http://www.ampliphibio.com/index.php/pipeline/product_pipeline.

9. **Angly, F., M. Youle, B. Nosrat, S. Srinagesh, B. Rodriguez-Brito, P. McNairnie, G. Deyanat-Yazdi, M. Breitbart, and F. Rohwer.** 2009. Genomic analysis of multiple Roseophage SIO1 strains. Environ. Microbiol. **11:**2863-2873.

10. **Aronoff, S. C.** 1988. Outer membrane permeability in *Pseudomonas cepacia*: Diminished porin content in a β-lactam-resistant mutant and in resistant cystic fibrosis isolates. Antimicrob. Agents Chemother. **32:**1636-1639.

11. **Aubert, D. F., R. S. Flannagan, and M. A. Valvano.** 2008. A novel sensor kinase-response regulator hybrid controls biofilm formation and type VI secretion system activity in *Burkholderia cenocepacia.* Infect. Immun. **76:**1979-1991.

12. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seldman, J. A. Smith, and K. Struhl.** 1991. Current protocols in molecular biology. Greene Publishing Associates, New York.

13. **Baldwin, A., E. Mahenthiralingam, P. Drevinek, C. Pope, D. J. Waine, D. A. Henry, D. P. Speert, P. Carter, P. Vandamme, J. J. LiPuma, and C. G. Dowson.** 2008. Elucidating global epidemiology of *Burkholderia multivorans* in cases of cystic fibrosis by multilocus sequence typing. J. Clin. Microbiol. **46:**290-295.

14. **Baldwin, A., E. Mahenthiralingam, P. Drevinek, P. Vandamme, J. R. Govan, D. J. Waine, J. J. LiPuma, L. Chiarini, C. Dalmastri, D. A. Henry, D. P. Speert, D. Honeybourne, M. C. J. Maiden, and C. G.**

Dowson. 2007. Environmental *Burkholderia cepacia* complex isolates in human infections. Emerg. Infect. Dis. **13:**458-461.

15. **Baldwin, A., E. Mahenthiralingam, K. M. Thickett, D. Honeybourne, M. C. J. Maiden, J. R. Govan, D. P. Speert, J. J. LiPuma, P. Vandamme, and C. G. Dowson.** 2005. Multilocus sequence typing scheme that provides both species and strain differentiation for the *Burkholderia cepacia* complex. J. Clin. Microbiol. **43:**4665-4673.
16. **Barreiro, V., and E. Haggard-Ljungquist.** 1992. Attachment sites for bacteriophage P2 on the *Escherichia coli* chromosome: DNA sequences, localization on the physical map, and detection of a P2-like remnant in *E. coli* K-12 derivatives. J. Bacteriol. **174:**4086-4093.
17. **Barrett, A. R., Y. Kang, K. S. Inamasu, M. S. Son, J. M. Vukovich, and T. T. Hoang.** 2008. Genetic tools for allelic replacement in *Burkholderia* species. Appl. Environ. Microbiol. **74:**4498-4508.
18. **Bassford Jr., P. J., C. Bradbeer, R. J. Kadner, and C. A. Schnaitman.** 1976. Transport of vitamin B12 in *tonB* mutants of *Escherichia coli*. J. Bacteriol. **128:**242-247.
19. **Bernier, S. P., and P. A. Sokol.** 2005. Use of suppression-subtractive hybridization to identify genes in the *Burkholderia cepacia* complex that are unique to *Burkholderia cenocepacia*. J. Bacteriol. **187:**5278-5291.
20. **Berry, J., E. J. Summer, D. K. Struck, and R. Young.** 2008. The final step in the phage infection cycle: The Rz and Rz1 lysis proteins link the inner and outer membranes. Mol. Microbiol. **70:**341-351.

21. **Bertani, G.** 1951. Studies on lysogenesis I. The mode of phage liberation by lysogenic *Escherichia coli*. J. Bacteriol. **62:**293-300.

22. **Betley, M. J., and J. J. Mekalanos.** 1985. Staphylococcal enterotoxin A is encoded by phage. Science **229:**185-187.

23. **Biddick, R., T. Spilker, A. Martin, and J. J. LiPuma.** 2003. Evidence of transmission of *Burkholderia cepacia*, *Burkholderia multivorans* and *Burkholderia dolosa* among persons with cystic fibrosis. FEMS Microbiol. Lett. **228:**57-62.

24. **Biswas, B., S. Adhya, P. Washart, B. Paul, A. N. Trostel, B. Powell, R. Carlton, and C. R. Merril.** 2002. Bacteriophage therapy rescues mice bacteremic from a clinical isolate of vancomycin-resistant *Enterococcus faecium*. Infect. Immun. **70:**204-210.

25. **Bollback, J. P., and J. P. Huelsenbeck.** 2009. Parallel genetic evolution within and between bacteriophage species of varying degrees of divergence. Genetics **181:**225-234.

26. **Boyd, E. F., and H. Brüssow.** 2002. Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved. Trends Microbiol. **10:**521-529.

27. **Brockhurst, M. A., A. Buckling, and P. B. Rainey.** 2005. The effect of a bacteriophage on diversification of the opportunistic bacterial pathogen, *Pseudomonas aeruginosa*. Proc. Biol. Sci. **272:**1385-1391.

28. **Brown, A. R., and J. R. W. Govan.** 2007. Assessment of fluorescent in situ hybridization and PCR-based methods for rapid identification of

Burkholderia cepacia complex organisms directly from sputum samples. J. Clin. Microbiol. **45:**1920-1926.

29. **Broxmeyer, L., D. Sosnowska, E. Miltner, O. Chacón, D. Wagner, J. McGarvey, R. G. Barletta, and L. E. Bermudez.** 2002. Killing of *Mycobacterium avium* and *Mycobacterium tuberculosis* by a mycobacteriophage delivered by a nonvirulent *Mycobacterium*: A model for phage therapy of intracellular bacterial pathogens. J. Infect. Dis. **186:**1155-1160.

30. **Brüssow, H.** 2001. Phages of dairy bacteria. Annu. Rev. Microbiol. **55:**283-303.

31. **Brüssow, H., and F. Desiere.** 2001. Comparative phage genomics and the evolution of *Siphoviridae*: Insights from dairy phages. Mol. Microbiol. **39:**213-222.

32. **Bryan, M. J., N. J. Burroughs, E. M. Spence, M. R. J. Clokie, N. H. Mann, and S. J. Bryan.** 2008. Evidence for the intense exchange of MazG in marine cyanophages by horizontal gene transfer. PLoS ONE **3:**e2048.

33. **Bull, J. J., M. R. Badgett, H. A. Wichman, J. P. Huelsenbeck, D. M. Hillis, A. Gulati, C. Ho, and I. J. Molineux.** 1997. Exceptional convergent evolution in a virus. Genetics **147:**1497-1507.

34. **Burkholder, W. H.** 1950. Sour skin, a bacterial rot of onion bulbs. Phytopath. **40:**115-117.

35. **Burns, J. L., M. Jonas, E. Y. Chi, D. K. Clark, A. Berger, and A. Griffith.** 1996. Invasion of respiratory epithelial cells by *Burkholderia* (*Pseudomonas*) *cepacia*. Infect. Immun. **64:**4054-4059.

36. **Burns, J. L., D. M. Lien, and L. A. Hedin.** 1989. Isolation and characterization of dihydrofolate reductase from trimethoprim-susceptible and trimethoprim-resistant *Pseudomonas cepacia*. Antimicrob. Agents Chemother. **33:**1247-1251.

37. **Burns, J. L., C. D. Wadsworth, J. J. Barry, and C. P. Goodall.** 1996. Nucleotide sequence analysis of a gene from *Burkholderia* (*Pseudomonas*) *cepacia* encoding an outer membrane lipoprotein involved in multiple antibiotic resistance. Antimicrob. Agents Chemother. **40:**307-313.

38. **Burt, A., and V. Koufopanou.** 2004. Homing endonuclease genes: The rise and fall and rise again of a selfish element. Curr. Opin. Genet. Dev. **14:**609-615.

39. **Byrne, M., and A. M. Kropinski.** 2005. The genome of the *Pseudomonas aeruginosa* generalized transducing bacteriophage F116. Gene **346:**187-194.

40. **Cady, K. C., A. S. White, J. H. Hammond, M. D. Abendroth, R. S. G. Karthikeyan, P. Lalitha, M. E. Zegans, and G. A. O'Toole.** 2011. Prevalence, conservation and functional analysis of *Yersinia* and *Escherichia* CRISPR regions in clinical *Pseudomonas aeruginosa* isolates. Microbiology **157:**430-437.

41. **Capparelli, R., M. Parlato, G. Borriello, P. Salvatore, and D. Iannelli.** 2007. Experimental phage therapy against *Staphylococcus aureus* in mice. Antimicrob. Agents Chemother. **51:**2765-2773.

42. **Caraher, E., G. Reynolds, P. Murphy, S. McClean, and M. Callaghan.** 2006. Comparison of antibiotic susceptibility of *Burkholderia cepacia* complex organisms when grown planktonically or as biofilm in vitro. Eur. J. Clin. Microbiol. Infect. Dis. **26:**213-216.

43. **Carmody, L. A., J. J. Gill, E. J. Summer, U. S. Sajjan, C. F. Gonzalez, R. F. Young, and J. J. LiPuma.** 2010. Efficacy of bacteriophage therapy in a model of *Burkholderia cenocepacia* pulmonary infection. J. Infect. Dis. **201:**264-271.

44. **Carvalho, A. P. D., G. M. C. Ventura, C. B. Pereira, R. S. Leão, T. W. Folescu, L. Higa, L. M. Teixeira, M. C. M. Plotkowski, V. L. C. Merquior, R. M. Albano, and E. A. Marques.** 2007. *Burkholderia cenocepacia*, *B. multivorans*, *B. ambifaria* and *B. vietnamiensis* isolates from cystic fibrosis patients have different profiles of exoenzyme production. APMIS **115:**311-318.

45. **Carver, T. J., K. M. Rutherford, M. Berriman, M-A. Rajandream, B. G. Barrell, and J. Parkhill.** 2005. ACT: The Artemis comparison tool. Bioinformatics **21:**3422-3423.

46. **Casjens, S.** 2003. Prophages and bacterial genomics: What have we learned so far? Mol. Microbiol. **49:**277-300.

47. **Casjens, S. R.** 2005. Comparative genomics and evolution of the tailed-bacteriophages. Curr. Opin. Microbiol. **8:**451-458.

48. **Casjens, S. R., E. B. Gilcrease, W. M. Huang, K. L. Bunny, M. L. Pedulla, M. E. Ford, J. M. Houtz, G. F. Hatfull, and R. W. Hendrix.** 2004. The pKO2 linear plasmid prophage of *Klebsiella oxytoca*. J. Bacteriol. **186:**1818-1832.

49. **Ceyssens, P., T. Glonti, A. M. Kropinski, R. Lavigne, N. Chanishvili, L. Kulakov, N. Lashkhi, M. Tediashvili, and M. Merabishvili.** 2011. Phenotypic and genotypic variations within a single bacteriophage species. Virol. J. **8:**134.

50. **Chakraborty, S., T. M. Z. Waise, F. Hassan, Y. Kabir, M. A. Smith, and M. Arif.** 2009. Assessment of the evolutionary origin and possibility of CRISPR-Cas (CASS) interference pathway in *Vibrio cholerae* O395. In Silico Biol. **9:**245-254.

51. **Chan, C-H., D. E. Stead, and R. H. A. Coutts.** 2003. Development of a species-specific *recA*-based PCR test for *Burkholderia fungorum*. FEMS Microbiol. Lett. **224:**133-138.

52. **Chow, L. T., and T. R. Broker.** 1978. Adjacent insertion sequences IS*2* and IS*5* in bacteriophage Mu mutants and an IS*5* in a lambda d*arg* bacteriophage. J. Bacteriol. **133:**1427-1436.

53. **Christie, G. E., L. M. Temple, B. A. Bartlett, and T. S. Goodwin.** 2002. Programmed translational frameshift in the bacteriophage P2 FETUD tail gene operon. J. Bacteriol. **184:**6522-6531.

54. **Cihlar, R. L., T. G. Lessie, and S. C. Holt.** 1978. Characterization of bacteriophage CP1, an organic solvent sensitive phage associated with *Pseudomonas cepacia*. Can. J. Microbiol. **24:**1404-1412.

55. **Clokie, M. R. J., and N. H. Mann.** 2006. Marine cyanophages and light. Environ. Microbiol. **8:**2074-2082.

56. **Coenye, T., J. J. LiPuma, D. Henry, B. Hoste, K. Vandemeulebroeke, M. Gillis, D. P. Speert, and P. Vandamme.** 2001. *Burkholderia cepacia* genomovar VI, a new member of the *Burkholderia cepacia* complex isolated from cystic fibrosis patients. Int. J. Syst. Evol. Microbiol. **51:**271-279.

57. **Coenye, T., E. Mahenthiralingam, D. Henry, J. J. LiPuma, S. Laevens, M. Gillis, D. P. Speert, and P. Vandamme.** 2001. *Burkholderia ambifaria* sp. nov., a novel member of the *Burkholderia cepacia* complex including biocontrol and cystic fibrosis-related isolates. Int. J. Syst. Evol. Microbiol. **51:**1481-1490.

58. **Coenye, T., P. Vandamme, J. R. W. Govan, and J. J. LiPuma.** 2001. Taxonomy and identification of the *Burkholderia cepacia* complex. J. Clin. Microbiol. **39:**3427-3436.

59. **Coenye, T., P. Vandamme, J. J. LiPuma, J. R. W. Govan, and E. Mahenthiralingam.** 2003. Updated version of the *Burkholderia cepacia* complex experimental strain panel. J. Clin. Microbiol. **41:**2797-2798.

60. **Conway, B-A. D., K. K. Chu, J. Bylund, E. Altman, and D. P. Speert.** 2004. Production of exopolysaccharide by *Burkholderia cenocepacia* results

in altered cell-surface interactions and altered bacterial clearance in mice. J. Infect. Dis. **190:**957-966.

61. **Coulondre, C., J. H. Miller, P. J. Farabaugh, and W. Gilbert.** 1978. Molecular basis of base substitution hotspots in *Escherichia coli*. Nature **274:**775-780.

62. **Croucher, N. J., and N. R. Thomson.** 2010. Studying bacterial transcriptomes using RNA-seq. Curr. Opin. Microbiol. **13:**619-624.

63. **Cystic Fibrosis Canada.** 2011. Canadian Cystic Fibrosis Patient Data Registry Report 2009. http://www.cysticfibrosis.ca/assets/files/pdf/CPDR_ReportE.pdf.

64. **Cystic Fibrosis Canada.** 2011. Cystic Fibrosis in Canada. http://www.cysticfibrosis.ca/assets/files/pdf/Cystic_fibrosis_in_canadaE.pdf.

65. **Cystic Fibrosis Canada.** 2011. Infection Control and Cystic Fibrosis. http://www.cysticfibrosis.ca/assets/files/pdf/Infection_Control_and_CFE.pdf.

66. **Dalmastri, C., L. Pirone, S. Tabacchioni, A. Bevivino, and L. Chiarini.** 2005. Efficacy of species-specific *recA* PCR tests in the identification of *Burkholderia cepacia* complex environmental isolates. FEMS Microbiol. Lett. **246:**39-45.

67. **Datta, I., S. Sau, A. K. Sil, and N. C. Mandal.** 2005. The bacteriophage lambda DNA replication protein P inhibits the *oriC* DNA- and ATP-binding functions of the DNA replication initiator protein DnaA of *Escherichia coli*. J. Biochem. Mol. Biol. **38:**97-103.

68. **Delcher, A. L., A. Phillippy, J. Carlton, and S. L. Salzberg.** 2002. Fast algorithms for large-scale genome alignment and comparison. Nucleic Acids Res. **30:**2478-2483.

69. **Dennis, J. J., and G. J. Zylstra.** 1998. Improved antibiotic-resistance cassettes through restriction site elimination using Pfu DNA polymerase PCR. BioTechniques **25:**772-776.

70. **Dennis, J. J., and G. J. Zylstra.** 1998. Plasposons: Modular self-cloning minitransposon derivatives for rapid genetic analysis of gram-negative bacterial genomes. Appl. Environ. Microbiol. **64:**2710-2715.

71. **DeShazer, D.** 2004. Genomic diversity of *Burkholderia pseudomallei* clinical isolates: subtractive hybridization reveals a *Burkholderia mallei*-specific prophage in *B. pseudomallei* 1026b. J. Bacteriol. **186:**3938-3950.

72. **DiMasi, J. A., R. W. Hansen, and H. G. Grabowski.** 2003. The price of innovation: New estimates of drug development costs. J. Health Econ. **22:**151-185.

73. **Donlan, R. M.** 2009. Preventing biofilms of clinically relevant organisms using bacteriophage. Trends Microbiol. **17:**66-72.

74. **Doulatov, S., A. Hodes, L. Dal, N. Mandhana, M. Liu, R. Deora, R. W. Simons, S. Zimmerly, and J. F. Miller.** 2004. Tropism switching in *Bordetella* bacteriophage defines a family of diversity-generating retroelements. Nature **431:**476-481.

75. **Duhaime, M. B., A. Wichels, J. Waldmann, H. Teeling, and F. O. Glöckner.** 2011. Ecogenomics and genome landscapes of marine *Pseudoalteromonas* phage H105/1. ISME J. **5:**107-121.

76. **Eggers, C. H., S. Casjens, S. F. Hayes, C. F. Garon, C. J. Damman, D. B. Oliver, and D. S. Samuels.** 2000. Bacteriophages of spirochetes. J. Mol. Microbiol. Biotechnol. **2:**365-373.

77. **El-Banna, N., and G. Winkelmann.** 1998. Pyrrolnitrin from *Burkholderia cepacia*: antibiotic activity against fungi and novel activities against streptomycetes. J. Appl. Microbiol. **85:**69-78.

78. **Environmental Protection Agency.** 2003. *Burkholderia cepacia* complex; significant new use rule. http://www.epa.gov/fedrgstr/EPA-TOX/2003/ June/Day-13/t15010.htm.

79. **Erickson, D. L., J. L. Lines, E. C. Pesci, V. Venturi, and D. G. Storey.** 2004. *Pseudomonas aeruginosa relA* contributes to virulence in *Drosophila melanogaster*. Infect. Immun. **72:**5638-5645.

80. **Festini, F., R. Buzzetti, C. Bassi, C. Braggion, D. Salvatore, G. Taccetti, and G. Mastella.** 2006. Isolation measures for prevention of infection with respiratory pathogens in cystic fibrosis: a systematic review. J. Hosp. Infect. **64:**1-6.

81. **Finnegan, D. J.** 1997. Transposable elements: How non-LTR retrotransposons do it. Curr. Biol. **7:**R245-R248.

82. **Fischetti, V. A.** 2005. The use of phage lytic enzymes to control bacterial infections, p. 321-334. *In* E. Kutter and A. Sulakvelidze (eds.), Bacteriophages: Biology and Applications. CRC Press, Boca Raton, FL.

83. **Flannagan, R. S., T. Linn, and M. A. Valvano.** 2008. A system for the construction of targeted unmarked gene deletions in the genus *Burkholderia*. Environ. Microbiol. **10:**1652-1660.

84. **Folsom, B. R., P. J. Chapman, and P. H. Pritchard.** 1990. Phenol and trichloroethylene degradation by *Pseudomonas cepacia* G4: kinetics and interactions between substrates. Appl. Environ. Microbiol. **56:**1279-1285.

85. **Fortier, L-C., J. D. Bouchard, and S. Moineau.** 2005. Expression and site-directed mutagenesis of the lactococcal abortive phage infection protein AbiK. J. Bacteriol. **187:**3721-3730.

86. **Fothergill, J. L., E. Mowat, M. J. Walshaw, M. J. Ledson, C. E. James, and C. Winstanley.** 2011. Effect of antibiotic treatment on bacteriophage production by a cystic fibrosis epidemic strain of *Pseudomonas aeruginosa*. Antimicrob. Agents Chemother. **55:**426-428.

87. **Freeman, V. J.** 1951. Studies on the virulence of bacteriophage-infected strains of *Corynebacterium diphtheriae*. J. Bacteriol. **61:**675-688.

88. **Ghazal, S. S., K. Al-Mudaimeegh, E. M. Al Fakihi, and A. T. Asery.** 2006. Outbreak of *Burkholderia cepacia* bacteremia in immunocompetent children caused by contaminated nebulized sulbutamol in Saudi Arabia. Am. J. Infect. Control **34:**394-398.

89. **Gill, J. J., E. J. Summer, W. K. Russell, S. M. Cologna, T. M. Carlile, A. C. Fuller, K. Kitsopoulos, L. M. Mebane, B. N. Parkinson, D. Sullivan, L. A. Carmody, C. F. Gonzalez, J. J. LiPuma, and R. Young.** 2011. Genomes and characterization of phages Bcep22 and BcepIL02, founders of a novel phage type in *Burkholderia cenocepacia*. J. Bacteriol. **193:**5300-5313.

90. **Gillis, M., T. Van Van, R. Bardin, M. Goor, P. Hebbar, A. Willems, P. Segers, K. Kersters, T. Heulin, and M. P. Fernandez.** 1995. Polyphasic taxonomy in the genus *Burkholderia* leading to an emended description of the genus and proposition of *Burkholderia vietnamiensis* sp. nov. for N_2-fixing isolates from rice in Vietnam. Int. J. Syst. Bacteriol. **45:**274-289.

91. **Girons, I. S., P. Bourhy, C. Ottone, M. Picardeau, D. Yelton, R. W. Hendrix, P. Glaser, and N. Charon.** 2000. The LE1 bacteriophage replicates as a plasmid within *Leptospira biflexa*: Construction of an *L. biflexa-Escherichia coli* shuttle vector. J. Bacteriol. **182:**5700-5705.

92. **Goerke, C., S. M. Y. Papenberg, S. Dasbach, K. Dietz, R. Ziebach, B. C. Kahl, and C. Wolz.** 2004. Increased frequency of genomic alterations in *Staphylococcus aureus* during chronic infection is in part due to phage mobilization. J. Infect. Dis. **189:**724-734.

93. **Goudie, A. D., K. H. Lynch, K. D. Seed, P. Stothard, S. Shrivastava, D. S. Wishart, and J. J. Dennis.** 2008. Genomic sequence and activity of KS10, a transposable phage of the *Burkholderia cepacia* complex. BMC Genomics **9:**615.

94. **Govan, J. R. W., J. E. Hughes, and P. Vandamme.** 1996. *Burkholderia cepacia*: Medical, taxonomic and ecological issues. J. Med. Microbiol. **45:**395-407.

95. **Graschopf, A., and U. Blăsi.** 1999. Molecular function of the dual-start motif in the λ S holin. Mol. Microbiol. **33:**569-582.

96. **Griffiths, A. J. F., J. H. Miller, D. T. Suzuki, R. C. Lewontin, and W. M. Gelbart.** 2000. An Introduction to Genetic Analysis, Seventh Edition. W. H. Freeman and Company, New York.

97. **Grimwood, K., T. J. Kidd, and M. Tweed.** 2009. Successful treatment of cepacia syndrome. J. Cyst. Fibros. **8:**291-293.

98. **Grissa, I., G. Vergnaud, and C. Pourcel.** 2007. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinformatics **8:**172.

99. **Gross, M., I. Marianovsky, and G. Glaser.** 2006. MazG - A regulator of programmed cell death in *Escherichia coli*. Mol. Microbiol. **59:**590-601.

100. **Groth, A. C., and M. P. Calos.** 2004. Phage integrases: Biology and applications. J. Mol. Biol. **335:**667-678.

101. **Guan, S., and N. K. Verma.** 1998. Serotype conversion of a *Shigella flexneri* candidate vaccine strain via a novel site-specific chromosome-integration system. FEMS Microbiol. Lett. **166:**79-87.

102. **Guo, L., K. B. Lim, C. M. Poduje, M. Daniel, J. S. Gunn, M. Hackett, and S. I. Miller.** 1998. Lipid A acylation and bacterial resistance against vertebrate antimicrobial peptides. Cell **95:**189-198.

103. **Hagens, S., and U. Blăsi.** 2003. Genetically modified filamentous phage as bactericidal agents: A pilot study. Lett. Appl. Microbiol. **37:**318-323.

104. **Hagens, S., A. Habel, U. von Ahsen, A. von Gabain, and U. Bläsi.** 2004. Therapy of experimental *Pseudomonas* infections with a nonreplicating genetically modified phage. Antimicrob. Agents Chemother. **48:**3817-3822.

105. **Hallenbeck, P. C., E. R. Vimr, F. Yu, B. Bassler, and F. A. Troy.** 1987. Purification and properties of a bacteriophage-induced endo-*N*-acetylneuraminidase specific for poly-α-2,8-sialosyl carbohydrate units. J. Biol. Chem. **262:**3553-3561.

106. **Hamilton, J., W. Burch, G. Grimmett, K. Orme, D. Brewer, R. Frost, and C. Fulkerson.** 1973. Successful treatment of *Pseudomonas cepacia* endocarditis with trimethoprim-sulfamethoxazole. Antimicrob. Agents Chemother. **4:**551-554.

107. **Hammer, B. K., and M. S. Swanson.** 1999. Co-ordination of *Legionella pneumophila* virulence with entry into stationary phase by ppGpp. Mol. Microbiol. **33:**721-731.

108. **Hanlon, G. W., S. P. Denyer, C. J. Olliff, and L. J. Ibrahim.** 2001. Reduction in exopolysaccharide viscosity as an aid to bacteriophage penetration through *Pseudomonas aeruginosa* biofilms. Appl. Environ. Microbiol. **67:**2746-2753.

109. **Hanych, B., S. Kedzierska, B. Walderich, B. Uznanski, and A. Taylor.** 1993. Expression of the Rz gene and the overlapping Rz1 reading frame present at the right end of the bacteriophage lambda genome. Gene **129:**1-8.

110. **Haralalka, S., S. Nandi, and R. K. Bhadra.** 2003. Mutation in the *relA* gene of *Vibrio cholerae* affects in vitro and in vivo expression of virulence factors. J. Bacteriol. **185:**4672-4682.

111. **Harper, D. R., and M. C. Enright.** 2011. Bacteriophages for the treatment of *Pseudomonas aeruginosa* infections. J. Appl. Microbiol. **111:**1-7.

112. **Hatfull, G. F.** 2010. Mycobacteriophages: Genes and genomes. Annu. Rev. Microbiol. **64:**331-356.

113. **Hatfull, G. F., and R. W. Hendrix.** 2011. Bacteriophages and their genomes. Curr. Opin. Virol. **1:**298-303.

114. **Hausler, T.** 2006. Viruses vs. Superbugs: A Solution to the Antibiotics Crisis? Macmillan, New York.

115. **Hein, J.** 1990. Unified approach to alignment and phylogenies. Meth. Enzymol. **183:**626-645.

116. **Henn, M. R., M. B. Sullivan, N. Stange-Thomann, M. S. Osburne, A. M. Berlin, L. Kelly, C. Yandava, C. Kodira, Q. Zeng, M. Weiand, T. Sparrow, S. Saif, G. Giannoukos, S. K. Young, C. Nusbaum, B. W. Birren, and S. W. Chisholm.** 2010. Analysis of high-throughput sequencing and annotation strategies for phage genomes. PLoS ONE **5:**e9083.

117. **Hennecke, F., H. Kolmar, K. Brundl, and H-J. Fritz.** 1991. The *vsr* gene product of *E. coli* K-12 is a strand- and sequence-specific DNA mismatch endonuclease. Nature **353:**776-778.

118. **Henry, D. A., M. E. Campbell, J. J. LiPuma, and D. P. Speert.** 1997. Identification of *Burkholderia cepacia* isolates from patients with cystic fibrosis and use of a simple new selective medium. J. Clin. Microbiol. **35:**614-619.

119. **Hens, D. K., N. C. Chatterjee, and R. Kumar.** 2006. New temperate DNA phage BcP15 acts as a drug resistance vector. Arch. Virol. **151:**1345-1353.

120. **Hens, D. K., A. N. Ghosh, and R. Kumar.** 2005. A new small temperate DNA phage BcP15 isolated from *Burkholderia cepacia* DR11. Arch. Virol. **150:**2421-2428.

121. **Heo, S. T., S. J. Kim, Y. G. Jeong, I. G. Bae, J. S. Jin, and J. C. Lee.** 2008. Hospital outbreak of *Burkholderia stabilis* bacteraemia related to contaminated chlorhexidine in haematological malignancy patients with indwelling catheters. J. Hosp. Infect. **70:**241-245.

122. **Hermans, A. P. H. M., A. M. Beuling, A. H. A. M. van Hoek, H. J. M. Aarts, T. Abee, and M. H. Zwietering.** 2006. Distribution of prophages and SGI-1 antibiotic-resistance genes among different *Salmonella enterica* serovar Typhimurium isolates. Microbiology **152:**2137-2147.

123. **Hiestand-Nauer, R., and S. Iida.** 1983. Sequence of the site-specific recombinase gene *cin* and of its substrates serving in the inversion of the C segment of bacteriophage P1. EMBO J. **2:**1733-1740.

124. **Higgins, D. G., and P. M. Sharp.** 1989. Fast and sensitive multiple sequence alignments on a microcomputer. Comput. Appl. Biosci. **5:**151-153.

125. **Hirai, K., S. Iyobe, M. Inoue, and S. Mitsuhashi.** 1980. Purification and properties of a new β-lactamase from *Pseudomonas cepacia*. Antimicrob. Agents Chemother. **17:**355-358.

126. **Holden, M. T. G., H. M. B. Seth-Smith, L. C. Crossman, M. Sebaihia, S. D. Bentley, A. M. Cerdeño-Tárraga, N. R. Thomson, N. Bason, M. A. Quail, S. Sharp, I. Cherevach, C. Churcher, I. Goodhead, H. Hauser, N. Holroyd, K. Mungall, P. Scott, D. Walker, B. White, H. Rose, P. Iversen, D. Mil-Homens, E. P. C. Rocha, A. M. Fialho, A. Baldwin, C. Dowson, B. G. Barrell, J. R. Govan, P. Vandamme, C. A. Hart, E. Mahenthiralingam, and J. Parkhill.** 2009. The genome of *Burkholderia cenocepacia* J2315, an epidemic pathogen of cystic fibrosis patients. J. Bacteriol. **91:**261-277.

127. **Ikeda, H., and J. Tomizawa.** 1968. Prophage P1, an extrachromosomal replication unit. Cold Spring Harb. Symp. Quant. Biol. **33:**791-798.

128. **Imamovic, L., E. Ballesté, J. Jofre, and M. Muniesa.** 2010. Quantification of Shiga toxin-converting bacteriophages in wastewater and in fecal samples by real-time quantitative PCR. Appl. Environ. Microbiol. **76:**5693-5701.

129. **Inal, J. M., and K. V. Karunakaran.** 1996. ϕ20, a temperate bacteriophage isolated from *Bacillus anthracis* exists as a plasmidial prophage. Curr. Microbiol. **32:**171-175.

130. **Isles, A., I. Maclusky, and M. Corey.** 1984. *Pseudomonas cepacia* infection in cystic fibrosis: An emerging problem. J. Pediatr. **104:**206-210.

131. **Jackson, D. W., K. Suzuki, L. Oakford, J. W. Simecka, M. E. Hart, and T. Romeo.** 2002. Biofilm formation and dispersal under the influence of the global regulator CsrA of *Escherichia coli*. J. Bacteriol. **184:**290-301.

132. **Jeong, J-H., M. Song, S-I. Park, K-O. Cho, H. R. Joon, and H. E. Choy.** 2008. *Salmonella enterica* serovar Gallinarum requires ppGpp for internalization and survival in animal cells. J. Bacteriol. **190:**6340-6350.

133. **Jimenez, L., and S. Smalls.** 2000. Molecular detection of *Burkholderia cepacia* in toiletry, cosmetic, and pharmaceutical raw materials and finished products. J. AOAC Int. **83:**963-966.

134. **Johnson, J.** 1994. Similarity analysis of rRNAs, p. 683-700. *In* P. Gerhardt, R. G. E. Murray, W. A. Wood, and N. R. Krieg (eds.), Methods for General and Molecular Bacteriology. ASM Press, Washington, DC.

135. **Jones, A. M., M. E. Dodd, J. R. W. Govan, V. Barcus, C. J. Doherty, J. Morris, and A. K. Webb.** 2004. *Burkholderia cenocepacia* and *Burkholderia multivorans*: Influence on survival in cystic fibrosis. Thorax **59:**948-951.

136. **Juncker, A. S., H. Willenbrock, G. von Heijne, S. Brunak, H. Nielsen, and A. Krogh.** 2003. Prediction of lipoprotein signal peptides in Gram-negative bacteria. Protein Sci. **12:**1652-1662.

137. **Kang, Y., M. H. Norris, B. A. Wilcox, A. Tuanyok, P. S. Keim, and T. T. Hoang.** 2011. Knockout and pullout recombineering for naturally transformable *Burkholderia thailandensis* and *Burkholderia pseudomallei*. Nat. Protoc. **6:**1085-1104.

138. **Kawasaki, K., R. K. Ernst, and S. I. Miller.** 2004. Deacylation and palmitoylation of lipid A by Salmonellae outer membrane enzymes modulate host signaling through Toll-like receptor 4. J. Endotoxin Res. **10:**439-444.

139. **Keig, P. M., E. Ingham, P. A. R. Vandamme, and K. G. Kerr.** 2002. Differential invasion of respiratory epithelial cells by members of the *Burkholderia cepacia* complex. Clin. Microbiol. Infect. **8:**47-49.

140. **Kelly, D., O. McAuliffe, R. P. Ross, J. O'Mahony, and A. Coffey.** 2011. Development of a broad-host-range phage cocktail for biocontrol. Bioeng. Bugs **2:**31-37.

141. **Kenna, D. T., V. A. Barcus, R. J. Langley, P. Vandamme, and J. R. W. Govan.** 2003. Lack of correlation between O-serotype, bacteriophage susceptibility and genomovar status in the *Burkholderia cepacia* complex. FEMS Immunol. Med. Microbiol. **35:**87-92.

142. **Kennedy, M. P., R. D. Coakley, S. H. Donaldson, R. M. Aris, K. Hohneker, J. P. Wedd, M. R. Knowles, P. H. Gilligan, and J. R. Yankaskas.** 2007. *Burkholderia gladioli*: Five year experience in a cystic fibrosis and lung transplantation center. J. Cyst. Fibros. **6:**267-273.

143. **Kim, A. I., P. Ghosh, M. A. Aaron, L. A. Bibb, S. Jain, and G. F. Hatfull.** 2003. Mycobacteriophage Bxb1 integrates into the *Mycobacterium smegmatis groEL1* gene. Mol. Microbiol. **50:**463-473.

144. **Klinkenberg, L. G., J-H. Lee, W. R. Bishai, and P. C. Karakousis.** 2010. The stringent response is required for full virulence of *Mycobacterium tuberculosis* in guinea pigs. J. Infect. Dis. **202:**1397-1404.

145. **Krogh, A., B. Larsson, G. von Heijne, and E. L. L. Sonnhammer.** 2001. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J. Mol. Biol. **305:**567-580.

146. **Kropinski, A. M.** 2006. Phage therapy - Everything old is new again. Can. J. Infect. Dis. Med. Microbiol. **17:**297-306.

147. **Krzywinski, M., J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S. J. Jones, and M. A. Marra.** 2009. Circos: An information aesthetic for comparative genomics. Genome Res. **19:**1639-1645.

148. **Kuhstoss, S., and R. N. Rao.** 1991. Analysis of the integration function of the streptomycete bacteriophage ϕC31. J. Mol. Biol. **222:**897-908.

149. **Kurtz, S., J. V. Choudhuri, E. Ohlebusch, C. Schleiermacher, J. Stoye, and R. Giegerich.** 2001. REPuter: The manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. **29:**4633-4642.

150. **Kutter, E., R. Raya, and K. Carlson.** 2005. Molecular mechanisms of phage infection, p. 165-222. *In* E. Kutter and A. Sulakvelidze (eds.), Bacteriophages: Biology and Applications. CRC Press, Boca Raton, FL.

151. **Kwan, T., J. Liu, M. DuBow, P. Gros, and J. Pelletier.** 2006. Comparative genomic analysis of 18 *Pseudomonas aeruginosa* bacteriophages. J. Bacteriol. **188:**1184-1187.

152. **Langley, R., D. T. Kenna, P. Vandamme, R. Ure, and J. R. W. Govan.** 2003. Lysogeny and bacteriophage host range within the *Burkholderia cepacia* complex. J. Med. Microbiol. **52:**483-490.

153. **Langley, R. J., D. Kenna, J. Bartholdson, D. J. Campopiano, and J. R. W. Govan.** 2005. Temperate bacteriophages DK4 and BcepMu from *Burkholderia cenocepacia* J2315 are identical. FEMS Immunol. Med. Microbiol. **45:**349-350.

154. **Lavigne, R., P. Darius, E. J. Summer, D. Seto, P. Mahadevan, A. S. Nilsson, H-W. Ackermann, and A. M. Kropinski.** 2009. Classification of *Myoviridae* bacteriophages using protein sequence similarity. BMC Microbiol. **9:**224.

155. **Lavigne, R., D. Seto, P. Mahadevan, H-W. Ackermann, and A. M. Kropinski.** 2008. Unifying classical and molecular taxonomic classification: analysis of the *Podoviridae* using BLASTP-based tools. Res. Microbiol. **159:**406-414.

156. **Lefebre, M. D., and M. A. Valvano.** 2001. *In vitro* resistance of *Burkholderia cepacia* complex isolates to reactive oxygen species in relation to catalase and superoxide dismutase production. Microbiology **147:**97-109.

157. **Levin, B. R., and J. J. Bull.** 2004. Population and evolutionary dynamics of phage therapy. Nat. Rev. Microbiol. **2:**166-173.

158. **Lewenza, S., B. Conway, E. P. Greenberg, and P. A. Sokol.** 1999. Quorum sensing in *Burkholderia cepacia*: Identification of the LuxRI homologs CepRI. J. Bacteriol. **181:**748-756.

159. **Li, W., D. P. Roberts, P. D. Dery, S. L. F. Meyer, S. Lohrke, R. D. Lumsden, and K. P. Hebbar.** 2002. Broad spectrum anti-biotic activity and disease suppression by the potential biocontrol agent *Burkholderia ambifaria* BC-F. Crop Prot. **21:**129-135.

160. **Lieb, M.** 1991. Spontaneous mutation at a 5-methylcytosine hotspot is prevented by very short patch (VSP) mismatch repair. Genetics **128:**23-27.

161. **Lin, J-J., and A. Sancar.** 1991. The C-terminal half of UvrC protein is sufficient to reconstitute (A)BC excinuclease. Proc. Natl. Acad. Sci. USA **88:**6824-6828.

162. **Lindsey, D. F., C. Martinez, and J. R. Walker.** 1992. Physical map location of the *Escherichia coli* attachment site for the P22 prophage (*attP22*). J. Bacteriol. **174:**3834-3835.

163. **LiPuma, J. J., S. E. Dasen, D. W. Nielson, R. C. Stern, and T. L. Stull.** 1990. Person-to-person transmission of *Pseudomonas cepacia* between patients with cystic fibrosis. Lancet **336:**1094-1096.

164. **LiPuma, J. J., B. J. Dulaney, J. D. McMenamin, P. W. Whitby, T. L. Stull, T. Coenye, and P. Vandamme.** 1999. Development of rRNA-based PCR assays for identification of *Burkholderia cepacia* complex isolates recovered from cystic fibrosis patients. J. Clin. Microbiol. **37:**3167-3170.

165. **LiPuma, J. J., T. Spilker, T. Coenye, and C. F. Gonzalez.** 2002. An epidemic *Burkholderia cepacia* complex strain identified in soil. Lancet **359:**2002-2003.

166. **Liu, M., R. Deora, S. R. Doulatov, M. Gingery, F. A. Eiserling, A. Preston, D. J. Maskell, R. W. Simons, P. A. Cotter, J. Parkhill, and J. F. Miller.** 2002. Reverse transcriptase-mediated tropism switching in *Bordetella* bacteriophage. Science **295:**2091-2094.

167. **Lopes, A., J. Amarir-Bouhram, G. Faure, M-A. Petit, and R. Guerois.** 2010. Detection of novel recombinases in bacteriophage genomes unveils Rad52, Rad51 and Gp2.5 remote homologs. Nucleic Acids Res. **38:**3952-3962.

168. **Loutet, S. A., R. S. Flannagan, C. Kooi, P. A. Sokol, and M. A. Valvano.** 2006. A complete lipopolysaccharide inner core oligosaccharide is required for resistance of *Burkholderia cenocepacia* to antimicrobial peptides and bacterial survival in vivo. J. Bacteriol. **188:**2073-2080.

169. **Lu, T. K., and J. J. Collins.** 2007. Dispersing biofilms with engineered enzymatic bacteriophage. Proc. Natl. Acad. Sci. USA **104:**11197-11202.

170. **Lu, T. K., and J. J. Collins.** 2009. Engineered bacteriophage targeting gene networks as adjuvants for antibiotic therapy. Proc. Natl. Acad. Sci. USA **106:**4629-4634.

171. **Ludtke, D. N., B. G. Eichorn, and S. J. Austin.** 1989. Plasmid-partition functions of the P7 prophage. J. Mol. Biol. **209:**393-406.

172. **Lukashin, A. V., and M. Borodovsky.** 1998. GeneMark.hmm: New solutions for gene finding. Nucleic Acids Res. **26:**1107-1115.

173. **Lutter, E., S. Lewenza, J. J. Dennis, M. B. Visser, and P. A. Sokol.** 2001. Distribution of quorum-sensing genes in the *Burkholderia cepacia* complex. Infect. Immun. **69:**4661-4666.

174. **Lynch, K. H., and J. J. Dennis.** 2008. Assessment of the contribution of prophage KS9 protein gp32 to the virulence of *Burkholderia pyrrocinia* LMG 21824. Banff Conference on Infectious Diseases, abstract W14.

175. **Lynch, K. H., and J. J. Dennis.** 2011. *Burkholderia*, p. 723-736. *In* D. Liu (ed.), Molecular Detection of Human Bacterial Pathogens. CRC Press, Boca Raton, FL.

176. **Lynch, K. H., K. D. Seed, P. Stothard, and J. J. Dennis.** 2010. Inactivation of *Burkholderia cepacia* complex phage KS9 gp41 identifies the phage repressor and generates lytic virions. J. Virol. **84:**1276-1288.

177. **Lynch, K. H., P. Stothard, and J. J. Dennis.** 2010. Genomic analysis and relatedness of P2-like phages of the *Burkholderia cepacia* complex. BMC Genomics **11:**599.

178. **Lynch, K. H., P. Stothard, and J. J. Dennis.** In press. Characterization of DC1, a broad host range Bcep22-like podovirus. Appl. Environ. Microbiol. doi: 10.1128/AEM.07097-11.

179. **Lynch, K. H., P. Stothard, and J. J. Dennis.** Submitted. Comparative analysis of two convergently evolved *Burkholderia cenocepacia*-specific bacteriophages. BMC Genomics manuscript 1715325926619906.

180. **Ma, Q., X. Sun, S. Gong, and J. Zhang.** 2010. Screening and identification of a highly lipolytic bacterial strain from barbecue sites in Hainan and characterization of its lipase. Ann. Microbiol. **60:**429-437.

181. **Magnusson, L. U., A. Farewell, and T. Nyström.** 2005. ppGpp: A global regulator in *Escherichia coli.* Trends Microbiol. **13:**236-242.

182. **Mahenthiralingam, E., J. Bischof, S. K. Byrne, C. Radomski, J. E. Davies, Y. Av-Gay, and P. Vandamme.** 2000. DNA-based diagnostic approaches for identification of *Burkholderia cepacia* complex, *Burkholderia vietnamiensis*, *Burkholderia multivorans*, *Burkholderia stabilis,* and *Burkholderia cepacia* genomovars I and III. J. Clin. Microbiol. **38:**3165-3173.

183. **Mahenthiralingam, E., T. Coenye, J. W. Chung, D. P. Speert, J. R. W. Govan, P. Taylor, and P. Vandamme.** 2000. Diagnostically and experimentally useful panel of strains from the *Burkholderia cepacia* complex. J. Clin. Microbiol. **38:**910-913.

184. **Mahenthiralingam, E., L. Song, A. Sass, J. White, C. Wilmot, A. Marchbank, O. Boaisha, J. Paine, D. Knight, and G. L. Challis.** 2011. Enacyloxins are products of an unusual hybrid modular polyketide synthase encoded by a cryptic *Burkholderia ambifaria* genomic island. Chem. Biol. **18:**665-677.

185. **Mahenthiralingam, E., T. A. Urban, and J. B. Goldberg.** 2005. The multifarious, multireplicon *Burkholderia cepacia* complex. Nat. Rev. Microbiol. **3:**144-156.

186. **Mahillon, J., and M. Chandler.** 1998. Insertion sequences. Microbiol. Mol. Biol. Rev. **62:**725-774.

187. **Makarova, K. S., D. H. Haft, R. Barrangou, S. J. J. Brouns, E. Charpentier, P. Horvath, S. Moineau, F. J. M. Mojica, Y. I. Wolf, A. F. Yakunin, J. van der Oost, and E. V. Koonin.** 2011. Evolution and classification of the CRISPR-Cas systems. Nat. Rev. Microbiol. **9:**467-477.

188. **Malott, R. J., A. Baldwin, E. Mahenthiralingam, and P. A. Sokol.** 2005. Characterization of the *cciIR* quorum-sensing system in *Burkholderia cenocepacia*. Infect. Immun. **73:**4982-4992.

189. **Malott, R. J., B. R. Steen-Kinnaird, T. D. Lee, and D. P. Speert.** 2012. Identification of hopanoid biosynthesis genes involved in polymyxin resistance in *Burkholderia multivorans*. Antimicrob. Agents Chemother. **56:**464-471.

190. **Mann, N. H., M. R. J. Clokie, A. Millard, A. Cook, W. H. Wilson, P. J. Wheatley, A. Letarov, and H. M. Krisch.** 2005. The genome of S-PM2, a "photosynthetic" T4-type bacteriophage that infects marine *Synechococcus* strains. J. Bacteriol. **187:**3188-3200.

191. **Marchler-Bauer, A., and S. H. Bryant.** 2004. CD-Search: Protein domain annotations on the fly. Nucleic Acids Res. **32:**W327-W331.

192. **Marinelli, L. J., M. Piuri, Z. Swigoňová, A. Balachandran, L. M. Oldfield, J. C. van Kessel, and G. F. Hatfull.** 2008. BRED: A simple and powerful tool for constructing mutant and recombinant bacteriophage genomes. PLoS ONE **3:**e3957.

193. **Markov, D., G. E. Christie, B. Sauer, R. Calendar, T. Park, R. Young, and K. Severinov.** 2004. P2 growth restriction on an *rpoC* mutant is suppressed by alleles of the Rz1 homolog *lysC*. J. Bacteriol. **186:**4628-4637.

194. **Matinkhoo, S., K. H. Lynch, J. J. Dennis, W. H. Finlay, and R. Vehring.** 2011. Spray-dried respirable powders containing bacteriophages for the treatment of pulmonary infections. J. Pharm. Sci. **100:**5197-5205.

195. **Matsuda, T., T. A. Freeman, D. W. Hilbert, M. Duff, M. Fuortes, P. P. Stapleton, and J. M. Daly.** 2005. Lysis-deficient bacteriophage therapy decreases endotoxin and inflammatory mediator release and improves survival in a murine peritonitis model. Surgery **137:**639-646.

196. **Matsumoto, H., Y. Itoh, S. Ohta, and Y. Terawaki.** 1986. A generalized transducing phage of *Pseudomonas cepacia*. J. Gen. Microbiol. **132:**2583-2586.

197. **Matsuzaki, S., M. Yasuda, H. Nishikawa, M. Kuroda, T. Ujihara, T. Shuin, Y. Shen, Z. Jin, S. Fujimoto, M. D. Nasimuzzaman, H. Wakiguchi, S. Sugihara, T. Sugiura, S. Koda, A. Muraoka, and S. Imai.** 2003. Experimental protection of mice against lethal *Staphylococcus aureus* infection by novel bacteriophage ϕMR11. J. Infect. Dis. **187:**613-624.

198. **McIntyre, K., M. Muller, S. Ota, A. Stephenson, and E. Tullis.** 2009. Epidemic of *Burkholderia cenocepacia* ET12 in Toronto adult cystic fibrosis clinic: lessons learned. International *Burkholderia cepacia* Working Group 13th Annual Meeting, abstract V-c.

199. **McMenamin, J. D., T. M. Zaccone, T. Coenye, P. Vandamme, and J. J. LiPuma.** 2000. Misidentification of *Burkholderia cepacia* in US cystic fibrosis treatment centers: An analysis of 1,051 recent sputum isolates. Chest **117:**1661-1665.

200. **McNeely, D., J. E. Moore, J. S. Elborn, B. C. Millar, J. Rendall, and J. S. Dooley.** 2009. Isolation of *Burkholderia cenocepacia* and *Burkholderia vietnamiensis* from human sewage. Int. J. Environ. Health Res. **19:**157-162.

201. **Medhekar, B., and J. F. Miller.** 2007. Diversity-generating retroelements. Curr. Opin. Microbiol. **10:**388-395.

202. **Merabishvili, M., J-P. Pirnay, G. Verbeken, N. Chanishvili, M. Tediashvili, N. Lashkhi, T. Glonti, V. Krylov, J. Mast, L. Van Parys, R. Lavigne, G. Volckaert, W. Mattheus, G. Verween, P. De Corte, T. Rose, S. Jennes, M. Zizi, D. De Vos, and M. Vaneechoutte.** 2009. Quality-controlled small-scale production of a well-defined bacteriophage cocktail for use in human clinical trials. PLoS ONE **4:**e4944.

203. **Merril, C. R., D. Scholl, and S. L. Adhya.** 2003. The prospect for bacteriophage therapy in Western medicine. Nat. Rev. Drug Discov. **2:**489-497.

204. **Miller, E. S., E. Kutter, G. Mosig, F. Arisaka, T. Kunisawa, and W. Ruger.** 2003. Bacteriophage T4 genome. Microbiol. Mol. Biol. Rev. **67:**86-156.

205. **Miller, R. V., and M. Day.** 2008. Contribution of lysogeny, pseudolysogeny, and starvation to phage ecology, p. 114-143. *In* S. T.

Abedon (ed.), Bacteriophage Ecology. Cambridge University Press, Cambridge.

206. **Minakhin, L., E. Semenova, J. Liu, A. Vailov, E. Severinova, T. Gabisonia, R. Inman, A. Mushegian, and K. Severinov.** 2005. Genome sequence and gene expression of *Bacillus anthracis* bacteriophage Fah. J. Mol. Biol. **354:**1-15.

207. **Mohanty, G., and S. Mukherji.** 2008. Biodegradation rate of diesel range *n*-alkanes by bacterial cultures *Exiguobacterium aurantiacum* and *Burkholderia cepacia*. Int. Biodeterior. Biodegrad. **61:**240-250.

208. **Moon, S., Y. Byun, H-J. Kim, S. Jeong, and K. Han.** 2004. Predicting genes expressed via -1 and +1 frameshifts. Nucleic Acids Res. **32:**4884-4892.

209. **Moore, J. E., B. C. Millar, J. Xu, M. Crowe, A. O. B. Redmond, and J. S. Elborn.** 2002. Misidentification of a genomovar of *Burkholderia cepacia* by *recA* restriction fragment length polymorphism. J. Clin. Pathol. **55:**309-311.

210. **Moradpour, Z., and A. Ghasemian.** 2011. Modified phages: Novel antimicrobial agents to combat infectious diseases. Biotechnol. Adv. **29:**732-738.

211. **Mücke, M., G. Grelle, J. Behlke, R. Kraft, D. H. Krüger, and M. Reuter.** 2002. EcoRII: A restriction enzyme evolving recombination functions? EMBO J. **21:**5262-5268.

212. **Mühlenhoff, M., K. Stummeyer, M. Grove, M. Sauerborn, and R. Gerardy-Schahn.** 2003. Proteolytic processing and oligomerization of bacteriophage-derived endosialidases. J. Biol. Chem. **278:**12634-12644.

213. **Myers, E. W., and W. Miller.** 1988. Optimal alignments in linear space. Comput. Appl. Biosci. **4:**11-17.

214. **Nakayama, K., S. Kanaya, M. Ohnishi, Y. Terawaki, and T. Hayashi.** 1999. The complete nucleotide sequence of ϕCTX, a cytotoxin-converting phage of *Pseudomonas aeruginosa*: Implications for phage evolution and horizontal gene transfer via bacteriophages. Mol. Microbiol. **31:**399-419.

215. **Nilsson, A. S., and E. Haggård-Ljungquist.** 2007. Evolution of P2-like phages and their impact on bacterial evolution. Res. Microbiol. **158:**311-317.

216. **Nishiyama, E., Y. Ohtsubo, Y. Nagata, and M. Tsuda.** 2010. Identification of *Burkholderia multivorans* ATCC 17616 genes induced in soil environment by *in vivo* expression technology. Environ. Microbiol. **12:**2539-2558.

217. **Nørskov-Lauritsen, N., H. K. Johansen, M. G. Fenger, X. C. Nielsen, T. Pressler, H. V. Olesen, and N. Høiby.** 2010. Unusual distribution of *Burkholderia cepacia* complex species in Danish cystic fibrosis clinics may stem from restricted transmission between patients. J. Clin. Microbiol. **48:**2981-2983.

218. **Nzula, S., P. Vandamme, and J. R. W. Govan.** 2000. Sensitivity of the *Burkholderia cepacia* complex and *Pseudomonas aeruginosa* to transducing bacteriophages. FEMS Immunol. Med. Microbiol. **28:**307-312.

219. **Nzula, S., P. Vandamme, and J. R. W. Govan.** 2002. Influence of taxonomic status on the *in vitro* antimicrobial susceptibility of the *Burkholderia cepacia* complex. J. Antimicrob. Chemother. **50:**265-269.

220. **Odegrip, R., A. S. Nilsson, and E. Haggård-Ljungquist.** 2006. Identification of a gene encoding a functional reverse transcriptase within a highly variable locus in the P2-like coliphages. J. Bacteriol. **188:**1643-1647.

221. **O'Grady, E. P., D. F. Viteri, R. J. Malott, and P. A. Sokol.** 2009. Reciprocal regulation by the CepIR and CciIR quorum sensing systems in *Burkholderia cenocepacia.* BMC Genomics **10:**441.

222. **Ohtsubo, Y., H. Genka, H. Komatsu, Y. Nagata, and M. Tsuda.** 2005. High-temperature-induced transposition of insertion elements in *Burkholderia multivorans* ATCC 17616. Appl. Environ. Microbiol. **71:**1822-1828.

223. **Oppenheim, A. B., O. Kobiler, J. Stavans, D. L. Court, and S. Adhya.** 2005. Switches in bacteriophage lambda development. Annu. Rev. Genet. **39:**409-429.

224. **Ortega, X., A. Silipo, M. S. Saldfas, C. C. Bates, A. Molinaro, and M. A. Valvano.** 2009. Biosynthesis and structure of the *Burkholderia cenocepacia* K56-2 lipopolysaccharide core oligosaccharide: Truncation of the core

oligosaccharide leads to increased binding and sensitvity to polymyxin B. J. Biol. Chem. **284:**21738-21751.

225. **Palmer, K. L., and M. S. Gilmore.** 2010. Multidrug-resistant enterococci lack CRISPR-*cas*. mBio **1:**e00227-10.

226. **Panlilio, A. L., C. M. Beck-Sague, J. D. Siegel, R. L. Anderson, S. Y. Yetts, N. C. Clark, P. N. Duer, K. A. Thomassen, R. W. Vess, B. C. Hill, O. C. Tablan, and W. R. Jarvis.** 1992. Infections and pseudoinfections due to povidone-iodine solution contaminated with *Pseudomonas cepacia*. Clin. Infect. Dis. **14:**1078-1083.

227. **Parr Jr., T. R., R. A. Moore, L. V. Moore, and R. E. W. Hancock.** 1987. Role of porins in intrinsic antibiotic resistance of *Pseudomonas cepacia*. Antimicrob. Agents Chemother. **31:**121-123.

228. **Paul, V., S. Sundarrajan, S. Rajagopalan, S. Hariharan, N. Kempashanaiah, S. Padmanabhan, B. Sriram, and J. Ramachandran.** 2011. Lysis-deficient phages as novel therapeutic agents for controlling bacterial infection. BMC Microbiol. **11:**195.

229. **Payne, G. W., P. Vandamme, S. H. Morgan, J. J. LiPuma, T. Coenye, A. J. Weightman, T. H. Jones, and E. Mahenthiralingam.** 2005. Development of a *recA* gene-based identification approach for the entire *Burkholderia* genus. Appl. Environ. Microbiol. **71:**3917-3927.

230. **Payne, R. J. H., and V. A. A. Jansen.** 2001. Understanding bacteriophage therapy as a density-dependent kinetic process. J. Theor. Biol. **208:**37-48.

231. **Peeters, E., H. J. Nelis, and T. Coenye.** 2009. *In vitro* activity of ceftazidime, ciprofloxacin, meropenem, minocycline, tobramycin and trimethoprim/sulfamethoxazole against planktonic and sessile *Burkholderia cepacia* complex bacteria. J. Antimicrob. Chemother. **64:**801-809.

232. **Peeters, E., A. Sass, E. Mahenthiralingam, H. Nelis, and T. Coenye.** 2010. Transcriptional response of *Burkholderia cenocepacia* J2315 sessile cells to treatments with high doses of hydrogen peroxide and sodium hypochlorite. BMC Genomics **11:**90.

233. **Pilacinski, W., E. Mosharrafa, and R. Edmundson.** 1977. Insertion sequence IS*2* associated with *int*-constitutive mutants of bacteriophage lambda. Gene **2:**61-74.

234. **Plasterk, R. H. A., T. A. M. Ilmer, and P. van de Putte.** 1983. Site-specific recombination by Gin of bacteriophage Mu: Inversions and deletions. Virology **127:**24-36.

235. **Platt, R., D. L. Reynolds, and G. J. Phillips.** 2003. Development of a novel method of lytic phage delivery by use of a bacteriophage P22 site-specific recombination system. FEMS Microbiol. Lett. **223:**259-265.

236. **Poteete, A. R.** 2001. What makes the bacteriophage λ Red system useful for genetic engineering: Molecular mechanism and biological function. FEMS Microbiol. Lett. **201:**9-14.

237. **Pouillot, F., H. Blois, and F. Iris.** 2010. Genetically engineered virulent phage banks in the detection and control of emergent pathogenic bacteria. Biosecur. Bioterror. **8:**155-169.

238. **Pujol, M., X. Corbella, J. Carratala, and F. Gudiol.** 1992. Community-acquired bacteremic *Pseudomonas cepacia* pneumonia in an immunocompetent host. Clin. Infect. Dis. **15:**887-888.

239. **Ramasubbu, N., L. M. Thomas, C. Ragunath, and J. B. Kaplan.** 2005. Structural analysis of dispersin B, a biofilm-releasing glycoside hydrolase from the periodontopathogen *Actinobacillus actinomycetemcomitans*. J. Mol. Biol. **349:**475-486.

240. **Ramette, A., J. J. LiPuma, and J. M. Tiedje.** 2005. Species abundance and diversity of *Burkholderia cepacia* complex in the environment. Appl. Environ. Microbiol. **71:**1193-1201.

241. **Ravin, V. K., and M. G. Shulga.** 1970. Evidence for extrachromosomal location of prophage N15. Virology **40:**800-807.

242. **Reik, R., T. Spilker, and J. J. LiPuma.** 2005. Distribution of *Burkholderia cepacia* complex species among isolates recovered from persons with or without cystic fibrosis. J. Clin. Microbiol. **43:**2926-2928.

243. **Ren, J., M. Kotaka, M. Lockyer, H. K. Lamb, A. R. Hawkins, and D. K. Stammers.** 2005. GTP cyclohydrolase II structure and mechanism. J. Biol. Chem. **280:**36912-36919.

244. **Rhoads, D. D., R. D. Wolcott, M. A. Kuskowski, B. M. Wolcott, L. S. Ward, and A. Sulakvelidze.** 2009. Bacteriophage therapy of venous leg ulcers in humans: results of a phase I safety trial. J. Wound Care **18:**237-243.

245. **Rice, P., L. Longden, and A. Bleasby.** 2000. EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet. **16:**276-277.

246. **Rice, S. A., C. H. Tan, P. J. Mikkelsen, V. Kung, J. Woo, M. Tay, A. Hauser, D. McDougald, J. S. Webb, and S. Kjelleberg.** 2009. The biofilm life cycle and virulence of *Pseudomonas aeruginosa* are dependent on a filamentous prophage. ISME J. **3:**271-282.

247. **Rich, D. P., M. P. Anderson, R. J. Gregory, S. H. Cheng, S. Paul, D. M. Jefferson, J. D. McCann, K. W. Klinger, A. E. Smith, and M. J. Welsh.** 1990. Expression of cystic fibrosis transmembrane conductance regulator corrects defective chloride channel regulation in cystic fibrosis airway epithelial cells. Nature **347:**358-363.

248. **Riedel, K., M. Hentzer, O. Geisenberger, B. Huber, A. Steidle, H. Wu, N. Høiby, M. Givskov, S. Molin, and L. Eberl.** 2001. *N*-acylhomoserine-lactone-mediated communication between *Pseudomonas aeruginosa* and *Burkholderia cepacia* in mixed biofilms. Microbiology **147:**3249-3262.

249. **Robinson, M., and P. T. B. Bye.** 2002. Mucociliary clearance in cystic fibrosis. Pediatr. Pulmonol. **33:**293-306.

250. **Rolain, J., P. François, D. Hernandez, F. Bittar, H. Richet, G. Fournous, Y. Mattenberger, E. Bosdure, N. Stremler, J. Dubus, J. Sarles, M. Reynaud-Gaubert, S. Boniface, J. Schrenzel, and D. Raoult.** 2009. Genomic analysis of an emerging multiresistant *Staphylococcus aureus* strain rapidly spreading in cystic fibrosis patients revealed the presence of an antibiotic inducible bacteriophage. Biol. Direct **4:**1.

251. **Ronning, C. M., L. Losada, L. Brinkac, J. Inman, R. L. Ulrich, M. Schell, W. C. Nierman, and D. DeShazer.** 2010. Genetic and phenotypic diversity in *Burkholderia*: Contributions by prophage and phage-like elements. BMC Microbiol. **10:**202.

252. **Routier, S. J. M.** 2010. DC1, a podoviridae with a putative cepacian depolymerase enzyme, MSc thesis.

253. **Ryley, H. C., and I. J. M. Doull.** 2003. *Burkholderia cepacia* complex infection in patients with cystic fibrosis: Laboratory investigations, epidemiology and clinical management. Rev. Med. Microbiol. **14:**15-24.

254. **Saiman, L., and J. Siegel.** 2004. Infection control in cystic fibrosis. Clin. Microbiol. Rev. **17:**57-71.

255. **Saini, L. S., S. B. Galsworthy, M. A. John, and M. A. Valvano.** 1999. Intracellular survival of *Burkholderia cepacia* complex isolates in the presence of macrophage cell activation. Microbiology **145:**3465-3475.

256. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:**406-425.

257. **Sajjan, S. U., L. A. Carmody, C. F. Gonzalez, and J. J. LiPuma.** 2008. A type IV secretion system contributes to intracellular survival and replication of *Burkholderia cenocepacia*. Infect. Immun. **76:**5447-5455.

258. **Sajjan, S. U., and J. F. Forstner.** 1992. Identification of the mucin-binding adhesin of *Pseudomonas cepacia* isolated from patients with cystic fibrosis. Infect. Immun. **60:**1434-1440.

259. **Sajjan, U. S., F. A. Sylvester, and J. F. Forstner.** 2000. Cable-piliated *Burkholderia cepacia* binds to cytokeratin 13 of epithelial cells. Infect. Immun. **68:**1787-1795.

260. **Sakaguchi, Y., T. Hayashi, K. Kurokawa, K. Nakayama, K. Oshima, Y. Fujinaga, M. Ohnishi, E. Ohtsubo, M. Hattori, and K. Oguma.** 2005. The genome sequence of *Clostridium botulinum* type C neurotoxin-converting phage and the molecular mechanisms of unstable lysogeny. Proc. Natl. Acad. Sci. USA **102:**17472-17477.

261. **Sambrook, J., and D. W. Russell.** 2001. Extraction of bacteriophage λ DNA from large-scale cultures using proteinase K and SDS. Molecular Cloning: A Laboratory Manual, Third Edition. **1:**2.56-2.58.

262. **Schattner, P., A. N. Brooks, and T. M. Lowe.** 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res. **33:**W686-W689.

263. **Schell, M. A., L. Lipscomb, and D. DeShazer.** 2008. Comparative genomics and an insect model rapidly identify novel virulence genes of *Burkholderia mallei*. J. Bacteriol. **190:**2306-2313.

264. **Schroeder, S. G., and C. T. Samudzi.** 1997. Structural studies of EcoRII methylase: Exploring similarities among methylases. Protein Eng. **10:**1385-1393.

265. **Seed, K. D., and J. J. Dennis.** 2005. Isolation and characterization of bacteriophages of the *Burkholderia cepacia* complex. FEMS Microbiol. Lett. **251:**273-280.

266. **Seed, K. D., and J. J. Dennis.** 2008. Development of *Galleria mellonella* as an alternative infection model for the *Burkholderia cepacia* complex. Infect. Immun. **76:**1267-1275.

267. **Seed, K. D., and J. J. Dennis.** 2009. Experimental bacteriophage therapy increases survival of *Galleria mellonella* larvae infected with clinically relevant strains of the *Burkholderia cepacia* complex. Antimicrob. Agents Chemother. **53:**2205-2208.

268. **Segal, B. H., T. L. Leto, J. I. Gallin, H. L. Malech, and S. M. Holland.** 2000. Genetic, biochemical, and clinical features of chronic granulomatous disease. Medicine **79:**170-200.

269. **Shah, S. A., and R. A. Garrett.** 2011. CRISPR/Cas and Cmr modules, mobility and evolution of adaptive immune systems. Res. Microbiol. **162:**27-38.

270. **Sheehan, M. M., E. Stanley, G. F. Fitzgerald, and D. van Sinderen.** 1999. Identification and characterization of a lysis module present in a large proportion of bacteriophages infecting *Streptococcus thermophilus*. Appl. Environ. Microbiol. **65:**569-577.

271. **Shelly, D. B., T. Spilker, E. J. Gracely, T. Coenye, P. Vandamme, and J. J. LiPuma.** 2000. Utility of commercial systems for identification of *Burkholderia cepacia* complex from cystic fibrosis sputum culture. J. Clin. Microbiol. **38:**3112-3115.

272. **Silipo, A., A. Molinaro, T. Ieranò, A. De Soyza, L. Sturiale, D. Garozzo, C. Aldridge, P. A. Corris, C. M. A. Khan, R. Lanzetta, and M. Parrilli.**

2007. The complete structure and pro-inflammatory activity of the lipooligosaccharide of the highly epidemic and virulent gram-negative bacterium *Burkholderia cenocepacia* ET-12 (strain J2315). Chem. Eur. J. **13:**3501-3511.

273. **Silva, A. J., and J. A. Benitez.** 2006. A *Vibrio cholerae* relaxed (*relA*) mutant expresses major virulence factors, exhibits biofilm formation and motility, and colonizes the suckling mouse intestine. J. Bacteriol. **188:**794-800.

274. **Smith, H. W., and M. B. Huggins.** 1982. Successful treatment of experimental *Escherichia coli* infections in mice using phage: its general superiority over antibiotics. J. Gen. Microbiol. **128:**307-318.

275. **Smith, M. C. M., and H. M. Thorpe.** 2002. Diversity in the serine recombinases. Mol. Microbiol. **44:**299-307.

276. **Snyder, L., and W. Champness.** 2003. Molecular Genetics of Bacteria. ASM Press, Washington, DC.

277. **Söding, J., A. Biegert, and A. N. Lupas.** 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. **33:**W244-W248.

278. **Soothill, J., C. Hawkins, and D. Harper.** 2011. Bacteriophage-containing therapeutic agents. Patent US 2011/0020290 A1.

279. **Speert, D. P.** 2002. Advances in *Burkholderia cepacia* complex. Paediatr. Respir. Rev. **3:**230-235.

280. **Speert, D. P., D. Henry, P. Vandamme, M. Corey, and E. Mahenthiralingam.** 2002. Epidemiology of *Burkholderia cepacia* complex in patients with cystic fibrosis, Canada. Emerg. Infect. Dis. **8:**181-187.

281. **Spilker, T., A. Baldwin, A. Bumford, C. G. Dowson, E. Mahenthiralingam, and J. J. LiPuma.** 2009. Expanded multilocus sequence typing for *Burkholderia* species. J. Clin. Microbiol. **47:**2607-2610.

282. **Sullivan, M. B., M. L. Coleman, P. Weigele, F. Rohwer, and S. W. Chisholm.** 2005. Three *Prochlorococcus* cyanophage genomes: Signature features and ecological interpretations. PLoS Biol. **3:**e144.

283. **Sullivan, M. B., K. H. Huang, J. C. Ignacio-Espinoza, A. M. Berlin, L. Kelly, P. R. Weigele, A. S. DeFrancesco, S. E. Kern, L. R. Thompson, S. Young, C. Yandava, R. Fu, B. Krastins, M. Chase, D. Sarracino, M. S. Osburne, M. R. Henn, and S. W. Chisholm.** 2010. Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. Environ. Microbiol. **12:**3035-3056.

284. **Summer, E. J.** 2009. Preparation of a phage DNA fragment library for whole genome shotgun sequencing. Methods Mol. Biol. **502:**27-46.

285. **Summer, E. J., J. Berry, T. A. T. Tran, L. Niu, D. K. Struck, and R. Young.** 2007. Rz/Rz1 lysis gene equivalents in phages of Gram-negative hosts. J. Mol. Biol. **373:**1098-1112.

286. **Summer, E. J., J. J. Gill, C. Upton, C. F. Gonzalez, and R. Young.** 2007. Role of phages in the pathogenesis of *Burkholderia*, or 'Where are the toxin genes in *Burkholderia* phages?'. Curr. Opin. Microbiol. **10:**410-417.

287. **Summer, E. J., C. F. Gonzalez, M. Bomer, T. Carlile, A. Embry, A. M. Kucherka, J. Lee, L. Mebane, W. C. Morrison, L. Mark, M. D. King, J. J. LiPuma, A. K. Vidaver, and R. Young.** 2006. Divergence and mosaicism among virulent soil phages of the *Burkholderia cepacia* complex. J. Bacteriol. **188:**255-268.

288. **Summer, E. J., C. F. Gonzalez, T. Carlisle, L. M. Mebane, A. M. Cass, C. G. Savva, J. LiPuma, and R. Young.** 2004. *Burkholderia cenocepacia* phage BcepMu and a family of Mu-like phages encoding potential pathogenesis factors. J. Mol. Biol. **340:**49-65.

289. **Sunenshine, R., M. Schultz, M. G. Lawrence, S. Shin, B. Jensen, S. Zubairi, A. M. Labriola, A. Shams, J. Noble-Wang, M. J. Arduino, F. Gordin, and A. Srinivasan.** 2009. An outbreak of postoperative gram-negative bacterial endophthalmitis associated with contaminated trypan blue ophthalmic solution. Clin. Infect. Dis. **48:**1580-1583.

290. **Susskind, M. M., and D. Botstein.** 1978. Molecular genetics of bacteriophage P22. Microbiol. Rev. **42:**385-413.

291. **Taplin, D., D. C. J. Bassett, and P. M. Mertz.** 1971. Foot lesions associated with *Pseudomonas cepacia*. Lancet **2:**568-571.

292. **Taylor, C. M., M. Beresford, H. A. S. Epton, D. C. Sigee, G. Shama, P. W. Andrew, and I. S. Roberts.** 2002. *Listeria monocytogenes relA* and *hpt* mutants are impaired in surface-attached growth and virulence. J. Bacteriol. **184:**621-628.

293. **Thiel, K.** 2004. Old dogma, new tricks - 21st Century phage therapy. Nat. Biotechnol. **22:**31-36.

294. **Tomich, M., A. Griffith, C. A. Herfst, J. L. Burns, and C. D. Mohr.** 2003. Attenuated virulence of a *Burkholderia cepacia* type III secretion mutant in a murine model of infection. Infect. Immun. **71:**1405-1415.

295. **Tomich, M., C. A. Herfst, J. W. Golden, and C. D. Mohr.** 2002. Role of flagella in host cell invasion by *Burkholderia cepacia*. Infect. Immun. **70:**1799-1806.

296. **van der Oost, J., M. M. Jore, E. R. Westra, M. Lundgren, and S. J. J. Brouns.** 2009. CRISPR-based adaptive and heritable immunity in prokaryotes. Trends Biochem. Sci. **34:**401-407.

297. **van Kessel, J. C., and G. F. Hatfull.** 2007. Recombineering in *Mycobacterium tuberculosis*. Nat. Methods **4:**147-152.

298. **van Kessel, J. C., L. J. Marinelli, and G. F. Hatfull.** 2008. Recombineering mycobacteria and their phages. Nat. Rev. Microbiol. **6:**851-857.

299. **Vandamme, P., and P. Dawyndt.** 2011. Classification and identification of the *Burkholderia cepacia* complex: Past, present and future. Syst. Appl. Microbiol. **34:**87-95.

300. **Vandamme, P., D. Henry, T. Coenye, S. Nzula, M. Vancanneyt, J. J. LiPuma, D. P. Speert, J. R. W. Govan, and E. Mahenthiralingam.** 2002. *Burkholderia anthina* sp. nov. and *Burkholderia pyrrocinia*, two additional

Burkholderia cepacia complex bacteria, may confound results of new molecular diagnostic tools. FEMS Immunol. Med. Microbiol. **33:**143-149.

301. **Vandamme, P., B. Holmes, T. Coenye, J. Goris, E. Mahenthiralingam, J. J. LiPuma, and J. R. W. Govan.** 2003. *Burkholderia cenocepacia* sp. nov. - A new twist to an old story. Res. Microbiol. **154:**91-96.

302. **Vandamme, P., B. Holmes, M. Vancanneyt, T. Coenye, B. Hoste, R. Coopman, H. Revets, S. Lauwers, M. Gillis, K. Kersters, and J. R. W. Govan.** 1997. Occurrence of multiple genomovars of *Burkholderia cepacia* in cystic fibrosis patients and proposal of *Burkholderia multivorans* sp. nov. Int. J. Syst. Evol. Microbiol. **47:**1188-1200.

303. **Vandamme, P., E. Mahenthiralingam, B. Holmes, T. Coenye, B. Hoste, P. De Vos, D. Henry, and D. P. Speert.** 2000. Identification and population structure of *Burkholderia stabilis* sp. nov. (formerly *Burkholderia cepacia* genomovar IV). J. Clin. Microbiol. **38:**1042-1047.

304. **Vanlaere, E., A. Baldwin, D. Gevers, D. Henry, E. De Brandt, J. J. LiPuma, E. Mahenthiralingam, D. P. Speert, C. Dowson, and P. Vandamme.** 2009. Taxon K, a complex within the *Burkholderia cepacia* complex, comprises at least two novel species, *Burkholderia contaminans* sp. nov. and *Burkholderia lata* sp. nov. Int. J. Syst. Evol. Microbiol. **59:**102-111.

305. **Vanlaere, E., J. J. LiPuma, A. Baldwin, D. Henry, E. D. Brandt, E. Mahenthiralingam, D. Speert, C. Dowson, and P. Vandamme.** 2008. *Burkholderia latens* sp. nov., *Burkholderia diffusa* sp. nov., *Burkholderia*

arboris sp. nov., *Burkholderia seminalis* sp. nov., and *Burkholderia metallica* sp. nov., novel species within the *Burkholderia cepacia* complex. Int. J. Syst. Evol. Microbiol. **58:**1580-1590.

306. **Verheust, C., G. Jensen, and J. Mahillon.** 2003. pGIL01, a linear tectiviral plasmid prophage originating from *Bacillus thuringiensis* serovar *israelensis*. Microbiology **149:**2083-2092.

307. **Vermis, K., M. Brachkova, P. Vandamme, and H. Nelis.** 2003. Isolation of *Burkholderia cepacia* complex genomovars from waters. Syst. Appl. Microbiol. **26:**595-600.

308. **Vermis, K., T. Coenye, J. J. LiPuma, E. Mahenthiralingam, H. J. Nelis, and P. Vandamme.** 2004. Proposal to accommodate *Burkholderia cepacia* genomovar VI as *Burkholderia dolosa* sp. nov. Int. J. Syst. Evol. Microbiol. **54:**689-691.

309. **Vermis, K., T. Coenye, E. Mahenthiralingam, H. J. Nelis, and P. Vandamme.** 2002. Evaluation of species-specific *recA*-based PCR tests for genomovar level identification within the *Burkholderia cepacia* complex. J. Med. Microbiol. **51:**937-940.

310. **Viklund, H., and A. Elofsson.** 2008. OCTOPUS: Improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. Bioinformatics **24:**1662-1668.

311. **Vincze, T., J. Posfai, and R. J. Roberts.** 2003. NEBcutter: A program to cleave DNA with restriction enzymes. Nucleic Acids Res. **31:**3688-3691.

312. **Waldor, M. K., and J. J. Mekalanos.** 1996. Lysogenic conversion by a filamentous phage encoding cholera toxin. Science **272:**1910-1913.

313. **Wang, C., M. Villion, C. Semper, C. Coros, S. Moineau, and S. Zimmerly.** 2011. A reverse transcriptase-related protein mediates phage resistance and polymerizes untemplated DNA *in vitro*. Nucleic Acids Res. **39:**7620-7629.

314. **Wang, I-N., D. L. Smith, and R. Young.** 2000. Holins: The protein clocks of bacteriophage infections. Annu. Rev. Microbiol. **54:**799-825.

315. **Weigele, P. R., W. H. Pope, M. L. Pedulla, J. M. Houtz, A. L. Smith, J. F. Conway, J. King, G. F. Hatfull, J. G. Lawrence, and R. W. Hendrix.** 2007. Genomic and structural analysis of Syn9, a cyanophage infecting marine *Prochlorococcus* and *Synechococcus*. Environ. Microbiol. **9:**1675-1695.

316. **Westwater, C., L. M. Kasman, D. A. Schofield, P. A. Werner, J. W. Dolan, M. G. Schmidt, and J. S. Norris.** 2003. Use of genetically engineered phage to deliver antimicrobial agents to bacteria: An alternative therapy for treatment of bacterial infections. Antimicrob. Agents Chemother. **47:**1301-1307.

317. **Wichman, H. A., M. R. Badgett, L. A. Scott, C. M. Boulianne, and J. J. Bull.** 1999. Different trajectories of parallel evolution during viral adaptation. Science **285:**422-424.

318. **Wichman, H. A., L. A. Scott, C. D. Yarber, and J. J. Bull.** 2000. Experimental evolution recapitulates natural evolution. Phil. Trans. R. Soc. Lond. **355:**1677-1684.

319. **Wilbur, W. J., and D. J. Lipman.** 1983. Rapid similarity searches of nucleic acid and protein data banks. Proc. Natl. Acad. Sci. USA **80:**726-730.

320. **Willner, D., M. R. Haynes, M. Furlan, N. Hanson, B. Kirby, Y. W. Lim, P. B. Rainey, R. Schmieder, M. Youle, D. Conrad, and F. Rohwer.** In press. Case studies of the spatial heterogeneity of DNA viruses in the cystic fibrosis lung. Am. J. Respir. Cell Mol. Biol. doi: 10.1165/rcmb.2011-0253OC.

321. **Winstanley, C., M. G. I. Langille, J. L. Fothergill, I. Kukavica-Ibrulj, C. Paradis-Bleau, F. Sanschagrin, N. R. Thomson, G. L. Winsor, M. A. Quail, N. Lennard, A. Bignell, L. Clarke, K. Seeger, D. Saunders, D. Harris, J. Parkhill, R. E. W. Hancock, F. S. L. Brinkman, and R. C. Levesque.** 2009. Newly introduced genomic prophage islands are critical determinants of in vivo competitiveness in the Liverpool Epidemic Strain of *Pseudomonas aeruginosa*. Genome Res. **19:**12-23.

322. **Woods, D. E., J. A. Jeddeloh, D. L. Fritz, and D. DeShazer.** 2002. *Burkholderia thailandensis* E125 harbors a temperate bacteriophage specific for *Burkholderia mallei*. J. Bacteriol. **184:**4003-4017.

323. **Wright, A., C. H. Hawkins, E. E. Änggård, and D. R. Harper.** 2009. A controlled clinical trial of a therapeutic bacteriophage preparation in chronic

otitis due to antibiotic-resistant *Pseudomonas aeruginosa*; A preliminary report of efficacy. Clin. Otolaryngol. **34:**349-357.

324. **Xu, J., R. W. Hendrix, and R. L. Duda.** 2004. Conserved translational frameshift in dsDNA bacteriophage tail assembly genes. Mol. Cell **16:**11-21.

325. **Xu, K. D., M. J. Franklin, C-H. Park, G. A. McFeters, and P. S. Stewart.** 2001. Gene expression and protein levels of the stationary phase sigma factor, RpoS, in continuously-fed *Pseudomonas aeruginosa* biofilms. FEMS Microbiol. Lett. **199:**67-71.

326. **Yabuuchi, E., Y. Kosako, H. Oyaizu, I. Yano, H. Hotta, Y. Hashimoto, T. Ezaki, and M. Arakawa.** 1992. Proposal of *Burkholderia* gen. nov. and transfer of seven species of the genus *Pseudomonas* homology group II to the new genus, with the type species *Burkholderia cepacia* (Palleroni and Holmes 1981) comb. nov. Microbiol. Immunol. **36:**1251-1275.

327. **Yang, C-J., T-C. Chen, L-F. Liao, L. Ma, C-S. Wang, P-L. Lu, Y-H. Chen, J-J. Hwan, L-K. Siu, and M-S. Huang.** 2008. Nosocomial outbreak of two strains of *Burkholderia cepacia* caused by contaminated heparin. J. Hosp. Infect. **69:**398-400.

328. **Young, R., I-N. Wang, and W. D. Roof.** 2000. Phages will out: Strategies of host cell lysis. Trends Microbiol. **8:**120-128.

329. **Yun, T., and D. Vapnek.** 1977. Electron microscopic analysis of bacteriophages P1, P1Cm, and P7. Determination of genome sizes,

sequence homology, and location of antibiotic resistance determinants. Virology **77:**376-385.

330. **Zafar, N., R. Mazumder, and D. Seto.** 2002. CoreGenes: A computational tool for identifying and cataloging "core" genes in a set of small genomes. BMC Bioinformatics **3:**12.

331. **Zhang, Y., S. E. Cottet, and S. E. Ealick.** 2004. Structure of *Escherichia coli* AMP nucleosidase reveals similarity to nucleoside phosphorylases. Structure **12:**1383-1394.

332. **Zhou, J., Y. Chen, S. Tabibi, L. Alba, E. Garber, and L. Saiman.** 2007. Antimicrobial susceptibility and synergy studies of *Burkholderia cepacia* complex isolated from patients with cystic fibrosis. Antimicrob. Agents Chemother. **51:**1085-1088.

333. **Ziermann, R., B. Bartlett, R. Calendar, and G. E. Christie.** 1994. Functions involved in bacteriophage P2-induced host cell lysis and identification of a new tail gene. J. Bacteriol. **176:**4974-4984.

334. **Zughaier, S. M., H. C. Ryley, and S. K. Jackson.** 1999. A melanin pigment purified from an epidemic strain of *Burkholderia cepacia* attenuates monocyte respiratory burst activity by scavenging superoxide anion. Infect. Immun. **67:**908-913.

335. **Zuker, M.** 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. **31:**3406-3415.