GENOMIC ANALYSIS OF *Burkholderia* AND *Rhodococcus equi* BACTERIOPHAGES

A Senior Scholars Thesis

by

ROBERT CHARLES ORCHARD II

Submitted to the Office of Undergraduate Research
Texas A&M University
in partial fulfillment of the requirements for the designation as

UNDERGRADUATE RESEARCH SCHOLAR

April 2008

Major: Microbiology

GENOMIC ANALYSIS OF *Burkholderia* AND *Rhodococcus equi* BACTERIOPHAGES

A Senior Scholars Thesis

by

ROBERT CHARLES ORCHARD II

Submitted to the Office of Undergraduate Research
Texas A&M University
in partial fulfillment of the requirements for the designation as

UNDERGRADUATE RESEARCH SCHOLAR

Approved by:

Research Advisor: Elizabeth Summer
Associate Dean for Undergraduate Research: Robert C. Webb

April 2008

Major: Microbiology

ABSTRACT

Genomic analysis of *Burkholderia* and *Rhodococcus equi* bacteriophages (April 2008)

Robert Charles Orchard II
Department of Biology
Texas A&M University

Research Advisor: Dr. Elizabeth Summer
Department of Biochemistry and Biophysics

Recently there has been an increase in bacterial infections that are resistant to traditional antibiotics. With this upward trend of resistance, many scientists are turning to alternative treatments like phage therapeutics. In this report there is discussion of isolating and characterizing novel phages of *Rhodococcus equi* and *Burkholderia cenocepacia*. *Burkholderia cenocepacia* is part of the *Burkholderia cepacia* complex (BCC), which causes pulmonary infections in cystic fibrosis patients (10). *Rhodococcus equi* is an intracellular pathogen which invades the macrophages of immunocompromised individuals such as young foals. While phylogenetically *R. equi*, a Gram-positive bacterium, and *Burkholderia*, Gram-negative bacteria, are unrelated, they both occupy the same ecological niche as soil saprophytes. Therefore, it is possible to isolate novel phages from the soil for both of these bacteria. Using a soil enrichment procedure, it has been possible to isolate and amplify 12 novel *R. equi* bacteriophages. These bacteriophages all appear to have the same morphologies as siphophages and similar genome sizes. Five of the *R. equi* phage genomes were pooled together with

other bacteriophages and pyrosequenced using 454 LifeSciences Technology. These phages are mosaic and sequencing traditional plasmid libraries were sufficient to deconvolute the 454 data. Also reported is the complete genomic sequencing of BcepNY3, a *Burkholderia cenocepacia* bacteriophage via a Whole Genome Shotgun Approach. BcepNY3's 47,382 bp genome was found to encode 70 proteins and 1 tRNA. It was determined that BcepNY3 is part of the previously described Bcep781 family of phages (26). Surprisingly, it is also related to the *Xanthomonas oryzae* phage OP_2, and proposed in this report is a possible evolutionary connection between OP_2 and the Bcep781 family of phages.

ACKNOWLEDGMENTS

I would like to first thank my research advisor Dr. Elizabeth Summer. Dr. Summer has provided me with a wealth of knowledge and support during my time at Texas A&M University. Her motivation and devotion to training undergraduates to become successful researchers are typically overlooked. I am fortunate enough to be one of the many students who are indebted to her dedication, and more importantly, her patience.

I would also like to thank Dr. Noah Cohen and Dr. Carlos Gonzalez for their work in providing the *Rhodococcus equi* and *Burkholderia* strains used respectively. I am grateful for Kevin Attenhofer and Amy Coffey for their work on the BcepNY3 genomic project. I was lucky enough to have worked with them in the Phage Genomics course and have used their insight and intelligence in constructing this report. Dr. Jason Gill and Dr. Christos Savva graciously provided the electron microscopy images.
I would like to thank my friends Van, Charley, and Thomas for their support of my goals, especially when the times were tough. I especially want to thank Megan for her assistance in reviewing this report and helping me to relax this past year.

During my time at Texas A&M University, I have been blessed to have met and interacted with Dr. Ryland Young, Dr. Michael Manson, Dr. John Gold, and Dr. Rita Moyes, all of whom have molded me into the scientist that I am. I want to thank all of those previously listed along with Dr. Summer and Dr. Gill for their support of my

career goals. Also, I would like to particularly thank Dr. Young for allowing me the opportunity to research in his lab.

Lastly, I would like to thank my parents and brothers for their continuous support and dedication.

NOMENCLATURE

BCC	*Burkholderia cepacia* complex
bp	Base pairs
CF	Cystic fibrosis
gp	Gene product
hyp.	Hypothetical
kb	Kilobases
LB	Luria-Bertani
OD	Optical density
pfu	Plaque forming units
λ-dil	Lambda dilution buffer

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION: PHAGE THERAPY- A SOLUTION TO ANTIBIOTIC RESISTANCE

An overview of Burkholderia

Burkholderia is a Gram-negative bacteria and a member of the *Betaproteobacteria* class. *Burkholderia* species occupy a wide variety of niches. Ecologically, most *Burkholderia* species are either plant pathogens or soil saprophytes, but *Burkholderia mallei* and *Burkholderia pseudomallei* are prominent animal and human pathogens (9). In humans, *B. mallei* and *B. psuedomallei* are the causative agents of glanders and melioidosis respectively and are potential bioterrorism agents (12). Because of their significance to global security, there is an emphasis on investigating *B. mallei* and *B. psuedomallei*. While these *Burkholderia* species are important, perhaps an even greater focus has been on the species in the *Burkholderia cepacia* complex (BCC) because of its diversity and its tendency to act as an opportunistic pulmonary pathogen.

The BCC is a complicated grouping of species that were once classified as *Burkholderia cepacia* genomovars. The current listing of BCC members include: *Burkholderia*

This thesis follows the style of Journal of Bacteriology.

cepacia, *Burkholderia multivorans*, *Burkholderia cenocepacia*, *Burkholderia stabilis*, *Burkholderia vietnamiensis*, *Burkholderia dolosa*, *Burkholderia ambifaria*, *Burkholderia anthina*, and *Burkholderia pyrrocinia* (10). BCC members have diverse ecological niches as they can be found as soil saprophytes, mutualistic plant symbiots, plant pathogens, and animal pathogens (9). Most are also capable of colonizing immunocompromised human lungs. The most prevalent cases of lungs inhabited by the BCC are in cystic fibrosis (CF) patients. It is estimated that about 10% of all CF patients have a BCC infection and even up to 40% have BCC infections in some regional areas (13). While BCC colonization is not as predominant as *Psuedomonas aeruginosa* colonization in CF patients, BCC is resistant to most antibiotics used to treat *P. aeruginosa* infections (13). BCC's resistance to multiple antibiotics has two implications. First, BCC infections are more difficult to treat and therefore more hazardous to CF patients. Second, to treat these dangerous infections, traditional antibiotics typically are not an appropriate means to fully clear the infection.

An overview of *Rhodococcus equi*

Rhodococcus equi is a Gram-positive bacterium and a member of the group Mycolata. Mycolata are defined as having a genome rich in guanine-cytosine nucleotides and cell envelopes that contain mycolic acid linked to arabinogalactans (20). Notable Mycolata pathogens include *Nocardia*, *Corynebacterium, Mycobacterium*, and *Rhodococcus.* The best characterized of these pathogens is *Mycobacterium tuberculosis*, the etiological agent of tuberculosis. *M. tuberculosis* is phagocytized by alveolar macrophages and is

capable of surviving in the macrophage and will eventually cause cell lysis. *R. equi* manifests a similar molecular pathology in young foals and recently in immunocompromised individuals, which results in pyogranulomatous pneumonia (22). Due to *R. equi*'s ability to remain in macrophages, it has been quite difficult to treat with antibiotics. *R. equi* is already resistant to β-lactams, excluding carbapenems (11). Currently, foals are treated by a combination of rifampin and erythromycin due to their ability to penetrate into macrophages (24). There is a great need for innovative treatments because of the increased resistance of *R. equi* to carbapenems such as imipenem, and a reduction of the efficiency of rifampin in foals derived from acquired resistance (23, 27).

Phage therapeutics- a plausible solution

With the increasing number of resistant bacteria, and the declining production of novel classes of antibiotics, many scientists are investigating alternative methods to treat bacterial infections. One popular idea is using bacteriophages, viruses that infect bacteria, to treat bacterial infections. This idea of phage therapy is not nascent, as it has been used in the Eastern European countries and the Soviet Union, while Western countries have focused on antibiotics. While phage therapy in these countries has been slightly successful, historically there have been significant numbers of clinical failures, which are attributed to the dearth of understanding of phage biology (15). With decades of phage biology research completed, a greater understanding of the molecular aspects of

bacteriophages has been achieved. This foundation of phage biology can be used to help develop therapeutic agents derived from bacteriophages.

A basic understanding of the phage life cycle

Bacteriophages are viruses that infect a specific bacterium or a specific bacterial strain. Because bacteriophages are viruses, they are totally dependent upon their host for the energy and machinery to generate new progeny. The phage life cycle is depicted in FIG. 1. It is important to remember that not all phage have the ability to lysogenize, but all phage can proceed through the lytic pathway. This information is vital when considering phage for phage therapy. Lysogenic phages are inappropriate to use for therapeutics for two major reasons. First, it is obvious that if a population of phage lysogenizes, the bacterial hosts will not be immediately killed. Therefore, the efficiency of using lysogenic phage to clear a bacterial infection will be low.

While the inefficiency of lysogenic phages is a hindrance for using them in phage therapeutics, the major deterrent in using these phages is their ability to carry virulence factors such as morons. Morons are genetic elements that are found in between a group of genes on the opposite strand. The term moron comes from the fact that there is "more DNA" and the "on" is added as part of the tradition of naming genetic elements. When a phage lysogenizes, these morons typically increase the fitness of the bacterium. Some elements prevent infection from other phage, while others convert innocuous bacteria to dangerous pathogens. The latter includes the phage encoded toxins of *Clostridum*

botulinim (botulism), *Corynebacterium diptheriae* (diphtheria), *Escherichia coli* (Shiga toxin), *Streptococcus pyogenes* (scarlet fever) and *Streptococcus aureus* (food poisoning). The ability of lysogenic phage to exacerbate bacterial infections is a major concern for phage therapy, because any therapeutic phage must be constitutively lytic and not temperate.

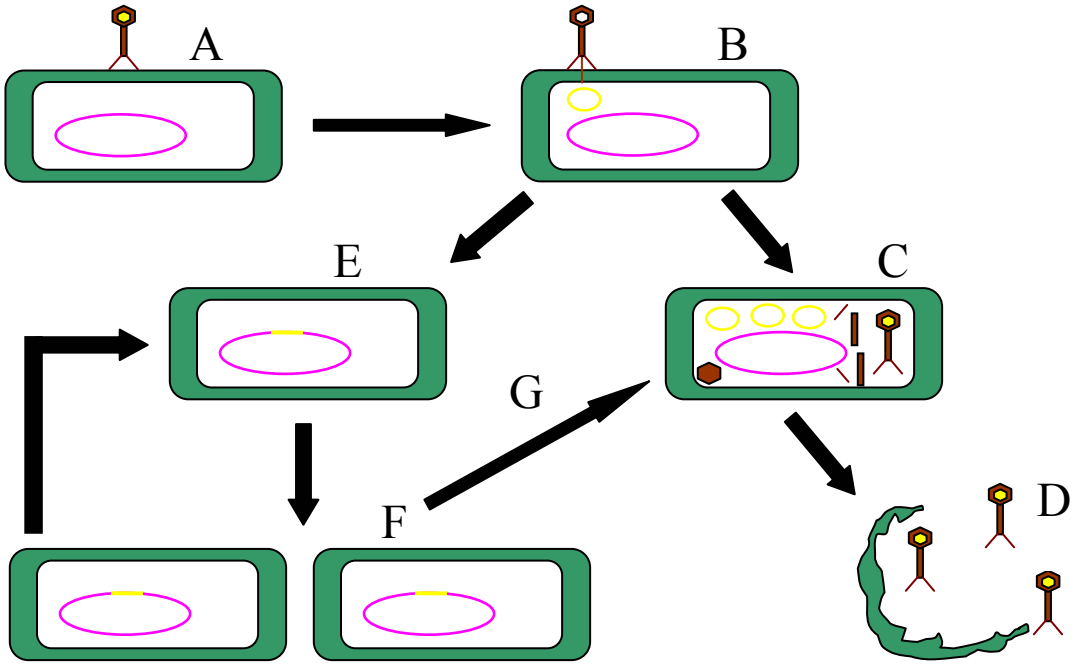

FIG. 1. A depiction of the phage life cycle,
A) Bacteriophage recognizes and adsorbs to its bacterial host.
B) The bacteriophage injects its DNA into the cell. At this point if the phage is temperate, it has a choice to lysogenize (leftward arrow) or proceed directly through the lytic cycle (rightward arrow).
C) Initial steps of the lytic pathway involve replication of the phage genome followed by the production and assembly of the structural proteins.
D) Phage mediated lysis of the cell leads to the loss of a membrane potential and death of the cell. Release of progeny virions allows for the infection of other cells.
E) After injection of the virion genome, temperate phages may choose to proceed through the lysogenic pathway and integrate into the host genome.
F) After lysogeny the bacterial cell will continue to grow and replicate and in doing so will replicate the integrated prophage as well. This process continues until some environmental factor triggers the phage excision (G).
G) An environmental signal causes the phage to be excised and proceed through the lytic pathway

D*igging for therapeutic phage*

Bacteriophages are the most numerous biological entities on the planet. It is estimated that there are more than 10^{30} bacteriophages in the biosphere (7). With the high concentration of phage in the biosphere, it can be assumed that novel phage from any ecosystem can be isolated. While the bacteria of interest, *Burkholderia* and *R. equi,* are phylogenetically unrelated, they both share the same ecological niche as soil saprophytes. Therefore, a high concentration of bacteriophages for both *Burkholderia* and *R. equi* should be able to be isolated from the soil.

Characterizing phage using genomics

The rapid advances of biotechnology has increased the availability of sequencing technology, which has led to the sequencing of more than 150 bacteriophage genomes (5). The growth of phage genomics has allowed researchers to gain better insight into the biological machinery of phage. Using a genomic approach to characterize an entire phage proteome will not only give a greater understanding of phage biology, but also determine if a specific phage carries the genes required for lysogeny. As this applies to phage therapeutics, genomics allows for a detailed analysis and screening of potential candidates.

Here in this report, novel bacteriophages of *Burkholderia* and *R. equi* have been isolated from the soil and their genomes analyzed. The objective is to first determine if these

phage can be used as therapeutic agents, then to use the information obtained to further understand the molecular biology of bacteriophages.

CHAPTER II

METHODS

Rhodococcus equi phage hunt

Generating a soil rinse-ate

Phages were isolated using an enrichment procedure with soil from farms and paddocks in Kentucky where horses have developed *R. equi* infections. The first step in this procedure is making a soil rinse-ate. 600 g of rich soil was mixed with 50 mL of Luria-Bertani (LB) media. This mixture was incubated over night at 30° C with shaking. At this point the flask contains a plentiful mixture of assorted microorganisms (bacteria, bacteriophages, amoebas, amoeboid viruses, etc.). Chloroform was added so that the solution was 0.3% by volume. Chloroform kills the majority of the living cells in the solution. To isolate only virus particles, a series of centrifugations and filtrations were preformed. Large soil debris was pelleted by a low centrifugation of 2,988 *g* for 5 minutes. After centrifugation, the supernatant was filtered through Mira cloth to remove any remaining large particles. The filtered supernatant was then centrifuged again at a low speed at 11,952 *g* for 10 minutes. The new supernatant was then filtered through a 0.45 μm filter followed by a 0.22 μm filter. This new solution is termed rinse-ate.

Phage enrichment and isolation

To amplify *R. equi* viruses from the rinse-ate, 25 mL of the rinse-ate, 25-mL of LB broth, and 5 mL of *R. equi* at an OD_{550} of 0.05 were mixed together and incubated at 30° C overnight with shaking. Different enrichments were performed where soil *R. equi* strains and clinical *R. equi* isolates were used. After the overnight incubation, chloroform was added so that the mixture was at 0.3% chloroform by volume. The solution was then centrifuged at a low speed for 10 minutes and then filtered through a 0.45 μm filter followed by a 0.22 μm filter. 5 μL of the enrichment was spotted onto a bacterial lawn of *R. equi* on LB plates to test for bacteriophages.

Phage isolation

To create pure phage stocks, the enrichments were diluted ten-fold serially from 10^{-1} to 10^{-6} and 100 μL of each dilution was mixed with 100 μL of the respected *R. equi* host within an OD_{550} range of 0.5 to 1.1. The mixture incubated at room temperature for 10 minutes to allow time for phage adsorption. After the 10 minutes, the phage-bacteria solution was added to 3 mL of either tryptone nutrient agar or LB T-top agar (tryptone nutrient broth/LB and 0.5% agar), vortexed briefly and plated onto LB plates. After top agar solidification, the plates were incubated overnight at 30°C. The following day individual plaques were picked from the plate with the most well dispersed plaques and resuspended in 1 mL of λ-dil (10 mM Tris pH 7.6, 5 mM $MgSO_4$, 0.08% gelatin, pH 7.4). Taking this pick-ate solution the same procedure of diluting and plating was

preformed. The end result is a 2nd pick-ate that is considered a pure solution of phage, which means it only contains one species of phage.

Phage amplification

2nd pick-ates were titered and plaque forming units (pfu) per mL were determined. To amplify the phage stock, 8 plates were made of dilutions resulting in 104 and 105 plaques per plate. The plating procedure was the same as listed previously. 5 mL of λ-dil was added to the plates and the t-top was scraped off and placed into an oakridge tube. Chloroform was added to a final concentration of 0.3%. This solution was placed on a shaking table to stir the contents for at least 45 minutes. Bacteria and agar were pelleted by centrifugating at 17,211 *g* for 10 minutes. The supernatant was then filtered through a 0.45 m filter and then a 0.22 m filter to isolate the virions. This phage lysate was then titered. The ideal titer is 10^9 pfu/mL so that ample DNA can be isolated. If the newly amplified phage stock was not at a high enough concentration, the amplification procedure was repeated with the phage lysate used in lieu of the pick-ate.

Host range studies of the enrichments and purified phages

5 μL of solutions from various sources were spotted onto lawns of different strains of *R. equi* on LB plates and incubated at 30 °C overnight. Sources included purified phage stocks, enrichments, λ-dil, and the rinse-ate.

Examining phage genomes

DNA was isolated from the phage lysates by first adding 40 µL of a nuclease mix (0.25 mg/mL RNAse A, 0.25 mg/mL DNAse I, 150mM NaCl, and 50% glycerol) to 10 mL of phage lysate and incubating for 30 minutes at 28°C. The virions were precipitated by adding 4 mL of a phage precipitate solution (PEG 8000). This mixture was placed on ice for 30 minutes. The solution was then pelleted by using a low speed centrifugation at 11,952 *g* for 10 minutes. The pellet was drained and the DNA was purified by using Promega's Wizard DNA Cleanup Kit (Promega cat. # A7280).

CHEF gel of isolated DNA

10 µL of phage DNA was placed in a 1% agarose gel and the gel was loaded unto a contour-clamped homogeneous electric field (CHEF) and electrophoresed under the conditions listed in Table 1. A CHEF gel is a specific way of using pulse-field electrophoresis to separate large DNA molecules (8).

TABLE 1. CHEF gel conditions

Condition	Setting
Angle	120°
Pulse time	50-90 s
Ramping factor	0 (linear)
Run time	22 hrs.
Voltage gradient	6V/cm

Phage restriction digest

To determine if the phage isolated were unique, DNA restriction digests were performed with three endonucleases: EcoRI, EcoRV, and HindIII. To digest the DNA standard procedures were used for each respective endonuclease with the reaction continuing overnight at 37°C. Fragments were viewed by electrophoresis on a 1% agarose gel.

Pooling DNA for 454 LifeSciences sequencing

DNA from 16 bacteriophages was pooled together and sent to 454 LifeSciences (www.454.com) to sequence. The genome sizes ranged from 34 kb to 250 kb and the total length of genomic sequence was $1.267 * 10^6$ base pairs. To ensure that each genome was present in equal quantities, the samples were normalized to 210 ng per genome equivalent. For this case a genome equivalent was defined as 50 kb. There were 24 genome equivalents in the DNA pool. Dr. Elizabeth Summer kindly generated the DNA pool for 454 LifeSciences.

Sequencing phage genomic libraries

Phage genomic libraries were graciously provided by Dr. Elizabeth Summer. Libraries were comprised of Lucigen's *E. cloni* cells with Lucigen's pSMART-LC KAN with genomic phage DNA inserted. Individual colonies from the plasmid library were picked and placed in a 96-well block containing LB broth supplemented with kanomycin (30 mg/L). The block was incubated at 32°C overnight. The following day each block was

placed in a low speed centrifuge at 120 g for 10 minutes. After removing the supernatant, the plasmids were isolated using the Qiagen Miniprep Kit. After isolating the plasmids, the inserts were sequenced using standard ABI BigDye Termination methods with the primers SR1 (GGTCAGGTATGATTTAAATGGTCAGT) and SL1 (CAGTCCAGTTACGCTGGAGTC). Purification of the reaction began by adding 30 μL of 75% isopropanol to each well. The plate was then centrifuged at 480 *g* for 2 hours and the supernatant was removed. Plasmids were precipitated by adding 100 μL of 70% ethanol to each well. The plate was centrifuged again at 480 *g* for 30 minutes and pellets were resuspended in 10 μL of formamide. The reactions were then resolved on an ABI3100 capillary sequencer (Laboratory for Plant Genome Technology, Texas A&M).

Assembly of phage genomes

Sequence assembly was performed using the computer program Sequencher (Gene Codes Corporation) was used. Assembled 454 contigs were assigned to their respective phage by comparing them to the plasmid library sequences. Primer walking was used to close gaps and to resequence ambiguities. Primer walking uses designed internal primers to amplify the region of interest using traditional PCR methods with the phage genome as the template. The listing of all primers used in this project is located in Table 2. After purifying the PCR product, the product was then sequenced using the ABI BigDye Termination method, previously described. If the sequencing reaction did not yield a fragment long enough, new primers were created based on this new sequencing data. The PCR and Sequencing procedures were then repeated.

Predicting genes

Protein coding regions were predicted initially using GeneMark.hmm (http://opal.biology.gatech.edu/GeneMark/) (2). Genes were refined through usage of the program Artemis (http://www.sanger.ac.uk/Software/Artemis/) (25). Changes were made based upon homology and presence and distance of Shine-Dalgarno sequences. To determine homology and putative functions of the predicted genes, protein sequences were compared with those of the NCBI database by BLASTP (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi?PAGE=Proteins&PROGRAM=blastp). To determine structural features like transmembrane domains and signal sequences the programs TMHMM (http://www.cbs.dtu.dk/services/TMHMM/) and LipoP (http://www.cbs.dtu.dk/services/LipoP/) were used, respectively. t-RNA were predicted using tRNAscan-SE (http://lowelab.ucsc.edu/tRNAscan-SE/). Genomic maps were constructed using DNA Master (http://cobamide2.bio.pitt.edu/computer.htm).

TABLE 2. Primers used for BcepNY3 sequencing project

Name	Sequence
NY3.1	GCTAAGATTGCGCGACCTAC
NY3.2	ACTCCTCGATGCTGTCGAAT
NY3.3	CTGCGCGAACTTCTTGAAC
NY3.4	CTTCCATGTTGGCTCCTACG
NY3.5	TGGGATAAGCGATGAAGGAT
NY3.6	GCAAGGAGAACATCCTCGTA
NY3.7	CAATCCAGCATCGCAGTCT
NY3.8	GAGAAGAGGTTGCCGAACTG
NY3.9	GCGCTTCTCGCAAATCTTC
NY3.10	CTTTGCGACTTCGGCATAGT
NY3.11	CCGACGTACTCGACTGGATG
NY3.12	TATTCCAGGACAGCCGATTC
NY3.A.R	GAACCCGGTCACAAGGTCTA
NY3A.L	GCTGCAGCATGAACAGCA
NY3.B.R	CATATTTCGCGAACGTGTCC
NY3.B.L	GAGTACTTGACGCCCCGATA
NY3.Y	GTGGTCGCTCTGCGAGTATT
NY3.Z	CGCAGTTGCATGTACGTCTG
NY3.1GL	GCAACCGGTTACGATTGTTT
NY3.1GR	GTGCGCAACTTGACGTGAT
NY3.2GL	GTGACCTTGTTCGCATCGT
NY3.2GR	CCGGTACGTTTCTTTCCTTG
NY3.3GL	TCGAGCTTCTTCCCGTAGAC
NY3.3GR	AACGGTGTGTCGTTCAACCT

CHAPTER III

RESULTS

Rhodococcus equi phages

Phage hunt

The first round of phage enrichments used 4 soil strains of *R. equi* that were isolated from Kentucky soils (Hil-n-Dale 1C, Milstream Barn, Hil-n-Dale 5B, and Venture 16 acre). The bacterial strains were isolated from the same location as the soil used for the enrichments. All four of these enrichments had *R. equi* phages and one phage from each was isolated. After determining that phage could be isolated from the soil, another round of enrichments was conducted with clinical equine isolates of *R. equi*. The strains used were 04-172, 04-181, 04-195, 04-200, 05-300, 05-305, 05-306, 06-383. All strains are VapA and ChoE positive, which are two *R. equi* virulence factors. Table 3 describes the strains of *R. equi* used in the enrichment. All 8 of the enrichments with clinical isolates yielded at least one phage isolated. Out of the 12 isolated phages, 11 were able to be grown to a high titer.

Prophages are dormant viruses whose DNA is inside the bacteria. Prophages can become lytic in response to external stimuli. To ensure that the phages isolated were in fact from the soil and not induced prophages, 2 mL of host strains were centrifuged at 11,952 *g* and 5 μL spotted onto bacterial lawns. In all cases the supernatants did not

have a clearing effect, demonstrating that the phages isolated were indeed from the soil. The enrichment procedure was effective in generating a concentration of phages that could be measured. The enrichment process was determined to be essential because 5 μL spots of the rinse-ate did not have any clearing effect on any bacterial lawns. Therefore, the enrichment process is an appropriate and necessary step in isolating novel *R. equi* bacteriophages from the soil.

TABLE 3. Summary of *R. equi* strains

Strain	Source	VapA and ChoE
Hil-n-Dale 1C	Soil	Unknown
Milstream Barn	Soil	Unknown
Hil-n-Dale 5B	Soil	Unknown
Venture 16 acre	Soil	Unknown
04-172	Equine	Positive
04-181	Equine	Positive
04-195	Equine	Positive
04-200	Equine	Positive
05-300	Equine	Positive
05-305	Equine	Positive
05-306	Equine	Positive
06-383	Equine	Positive

Phage characterization

Phages were characterized first based upon morphology. Dr. Jason Gill and Dr. Christos Savvaa have generously provided electron microscopy images of phages Reqi4, ReqiDocB7, ReqiPepy6, ReqiPine5, ReqiPoco6, ReqiRob04-200, ReqiRob06-325, and ReqiZip11 (FIG. 2.). All the bacteriophages have long flexible tails, which are characteristics of siphophages. Reqi4 has a very long and flimsy tail and there are several instances where the phage head was not attached to the phage tail (FIG. 2). This flimsy tail could explain why it has been unfeasible to grow it up to a high titer, but it could also be an artifact of staining.

Five different virions' tail length and head length and width were measured and averaged for each phage (Table 4). There appears to be two different groups of phages based upon tail length. The first group includes Reqi4 and ReqiDocB7 which have tails averaging 443 nm and 473 nm respectively. The other examined phages belong to a group that has significant shorter tails averaging approximately 280 nm. The head measurements do show some variation, but due to their small size and their sensitivity to osmotic changes, it is difficult to obtain accurate measurements. Therefore, it is not surprising that the data provided does not show any distinct groupings.

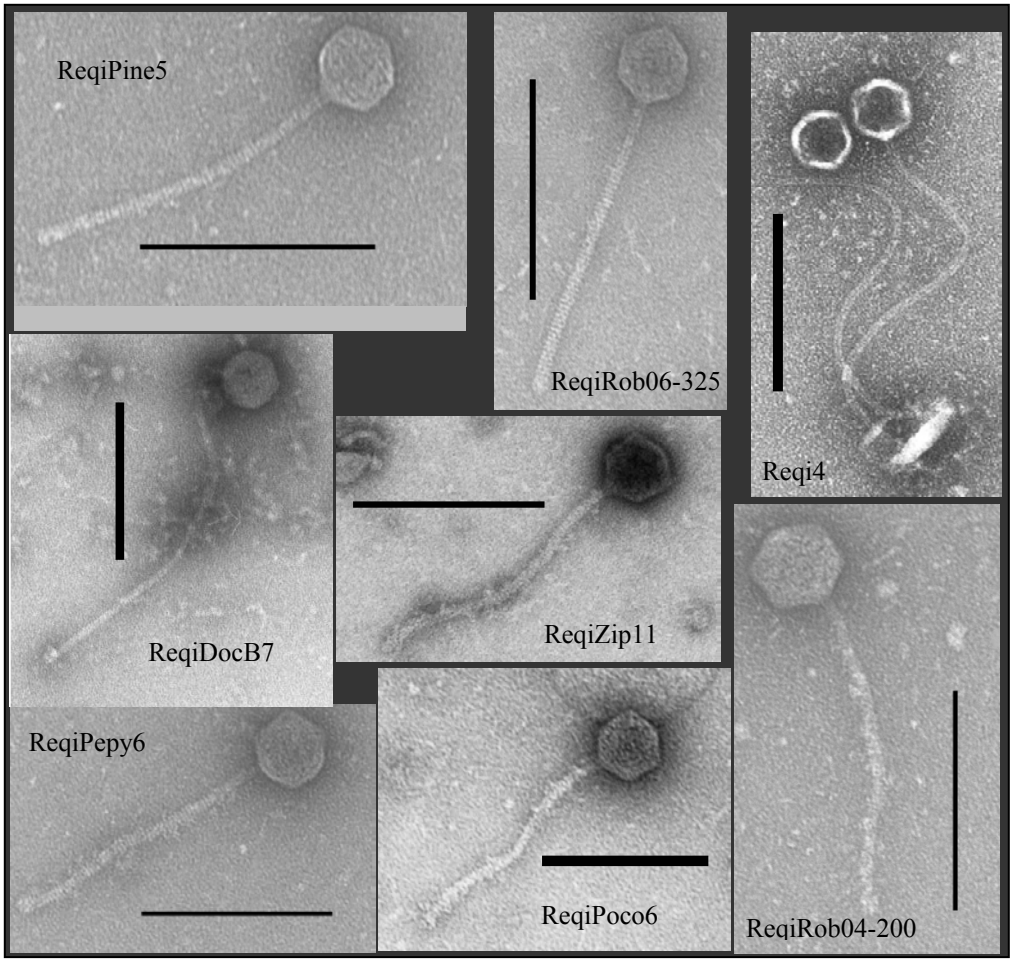

FIG. 2. Electron microscopy images of *R. equi* bacteriophages.
These are negatively stained electron microscopy of *R. equi* bacteriophages. All scale bars represent 200 nm.

TABLE 4. Measurements of *R. equi* bacteriophages

Phage	Tail length	Head length	Head width
ReqiDocB7	473 nm	72 nm	70 nm
ReqiPoco6	281 nm	75 nm	71 nm
ReqiZip11	274 nm	67 nm	73 nm
Reqi4	443 nm	60 nm	55 nm
ReqiRob-06-325	272 nm	64 nm	61 nm
ReqiRob-04-200	298 nm	80 nm	73 nm
ReqiPepy6	285 nm	80 nm	74 nm
ReqiPine5	272 nm	66 nm	65 nm

Genomic differences

Genomic differences were analyzed to determine the number of distinct phage types and the diversity of the phage isolated. Genome size of phage ReqiDocB7, ReqiPoco6, and ReqiZip11 was determined by a CHEF gel (Fig. 3a). All 3 of these phages have a similar genome size of approximately 75 kb. Diversity was then assessed by analyzing restriction digestion patterns of these phage genomes (Fig. 3b). The EcoRV digest patterns reveal that ReqiDocB7 is clearly distinct from ReqiPoco6 and ReqiZip11. This data supports the previously mentioned morphological differences (significant differences in tail lengths) between ReqiDocB7 and both ReqiPoco6 and ReqiZip11. There appears to be small differences between ReqiPoco6 and ReqiZip11 in the HindIII digest, but the ReqiPoco6 bands are too faint to make a confident conclusion.

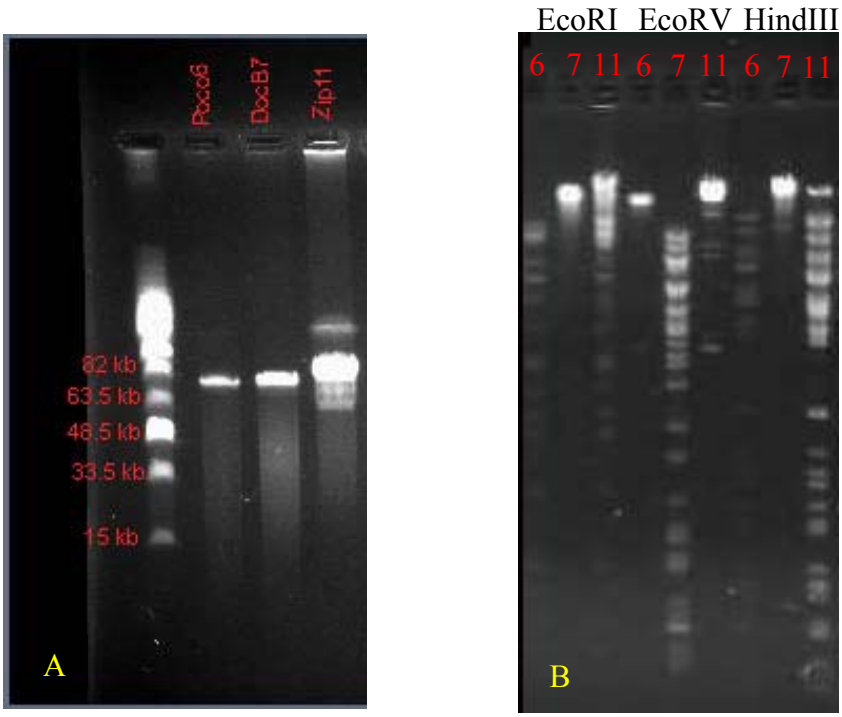

FIG. 3. CHEF gel and restriction digest of *R. equi* bacteriophages
(A) A CHEF gel of ReqiPoco6, ReqiDocB7 and ReqiZip11 DNA, stained with ethidium bromide (B). A DNA restriction digest of ReqiPoco6 (6), ReqiDocB7 (7) and ReqiZip11 (11) with enzymes EcoRI, EcoRV, and HindIII and stained with ethidium bromide.

Sequencing results of *R. equi* bacteriophages

5 of the 16 bacteriophage genomes submitted for 454 LifeSciences pyrosequencing were *R. equi* phage (ReqiDocB7, ReqiPepy6, ReqiPine5, ReqiPoco6, and ReqiZip11). Genomic libraries for ReqiDocB7, ReqiPepy6, ReqiPoco6, and ReqiZip11 were sequenced. Assembling all plasmid reads with all 454 contigs yielded 454 contigs that

had plasmid reads from multiple phages. This demonstrates that these *R. equi* bacteriophages have mosaic characteristics. Out of these phages, ReqiPine5 was clearly distinct, due to a dearth of other phage plasmid reads aligning with its 454 contigs. ReqiZip11 and ReqiPoco6 had the greatest number of plasmid reads coinciding in the same 454 contigs. This mosaicism confirms the restriction digest result that these phages are similar. Deconvoluting the mosaicism was made possible based upon the individual plasmid reads. Individual sequencing projects were set up in Sequencher and 454 contigs that did not have plasmid reads spanning the entire contig were removed from the phage's project. The plasmid reads and 454 contigs were then reassembled and the process continued. This method led to identification of ReqiDocB7, ReqiPepy6, ReqiPine5, ReqiPoco6, and ReqiZip11's 454 contigs. ReqiPine5's genome assembled into one contig and was identified by default. The 454 data demonstrates that the phages analyzed are similar, but clearly are distinct phage species. Table 5 summarizes the characterization of the isolated *R. equi* phages.

TABLE 5. A basic overview of the isolated *R. equi* bacteriophages

Phage	*R. equi* host strain	Morphology	DNA isolated	Sequencing project
ReqiDocB7	Hil-n-Dale 1C	Siphophage	Yes	Started
ReqiPoco6	Milstream Barn	Siphophage	Yes	Started
ReqiZip11	Hil-n-Dale 5B	Siphophage	Yes	Started
Reqi4	Venture 16 acre	Siphophage	No	Not Started
ReqiRob04-172	04-172	Unknown	No	Not Started
ReqiRob04-181	04-181	Unknown	No	Not Started
ReqiRob04-195	04-195	Unknown	No	Not Started
ReqiRob04-200	04-200	Siphophage	No	Not Started
ReqiRob05-300	04-300	Unknown	No	Not Started
ReqiRob06-325	06-325	Siphophage	No	Not Started
ReqiPine5	05-305	Siphophage	Yes	Not Started
ReqiPepy6	05-306	Siphophage	Yes	Started

Host range studies

Host ranges were determined for phages Reqi4, ReqiDocB7, ReqiPoco6, and ReqiZip11. At the time of the experiment, the other 8 phages had not been isolated, so instead the host range of their respective enrichments was determined. The estimated titer or all sources used is listed in Table 6. The phages and enrichments were tested against all 8 previously mentioned clinical hosts. Also tested was λ-dil to ensure that the dilution solution was not contaminated. As mention previously, the supernatants and rinse-ate was tested to determine the validity and efficacy of an enrichment procedure for *R. equi* bacteriophages. An example of a plate used in the host range study is in FIG. 4. The strong clearing by the sources indicates that the phages are competent for replication in the host. Table 7 summarizes the host range results.

TABLE 6. Titer of *R. equi* phages and enrichments

Source[a]	Titer
Reqi4	10^5 pfu/mL
ReqiDocB7	$>10^7$
ReqiPoco6	$>10^7$
ReqiZip11	$>10^7$
E-172	10^5 pfu/mL
E-181	10^6 pfu/mL
E-195	10^5 pfu/mL
E-200	10^7 pfu/mL
E-300	10^5 pfu/mL
E-305	10^7 pfu/mL
E-306	10^5 pfu/mL
E-325	10^7 pfu/mL

a E represents an enrichment

FIG. 4. Host range plates

These plates are characteristic of the results of the host range experiment. The enrichments (#E), supernatants (S), rinse-ate (R), λ-dil (λ), Reqi4 (Φ4), ReqiDocB7 (Φ7), ReqiPoco6 (Φ6), and ReqiZip11 (Φ11) are all respectively labeled.

TABLE 7. Host range results

R. equi strains								
Spot Source	04-172	04-181	04-195	04-200	05-300	05-305	05-306	06-325
Reqi4	-	-	+/-	-	+/-	-	-	-
ReqiPoco6	+	+	+	+	+	+	+	+
ReqiDocB7	+	+	+	+	+	+	+	+
ReqiZip11	+	+	+	+	+	+/-	+	+
E-04-172	+	+	+	+	+	+	+	+/-
E-04-181	+	+	+	+	+	+	+	+
E-04-195	+	+	+	+	+	+	+	+
E-04-200	+	+	+	+	+	+	+	+
E-05-300	+	+	+	+	+	+	+	+
E-05-305	+/-	+/-	+/-	+/-	+/-	+/-	+/-	+
E-05-306	+	+	+	+	+	+	+	+
E-06-325	+/-	+	+/-	+/-	+	+/-	+/-	+

strong clearing (+), weak clearing (+/-), no clearing (-)

The enrichments demonstrated a very broad host range, but this could be the result of these solutions being crude (containing multiple species of phages). This data can be used in the future when the host range of the respective purified phages is determined. If it differs from this data, then one can take the respective host and enrichment and select for that phage. Surprisingly, ReqiDocB7, ReqiPoco6, and ReqiZip11, all of whom were enriched with soil strains of *R. equi*, were competent for growing on all clinical strains. This demonstrates their potential usage in therapeutics due to their broad spectra of hosts. Reqi4, which has not been able to be grown up to a high titer, showed very weak clearing, and may be due to its flimsy tail.

Genomic analysis of BcepNY3

BcepNY3 was previously isolated by Dr. Jason Gill. BcepNY3 was sequenced solely using the traditional genomic plasmid libraries. The genome was found to be circularly permuted, and has a linear length of 47,382 bp with 63.6% GC content. The genome was found to encode 70 proteins and 1 tRNA encoding for tyrosine and recognizing UAC. Only 23 of the 70 predicted protein encoding genes have a putative function, but only 1 of the remaining 47 did not have a homologue in the database. Table 8 describes the predicted proteome of BcepNY3 which has been submitted to Genbank (accession number: NC_009604.1). FIG. 5 is a genomic map of BcepNY3 with proteins colored by function and significant proteins labeled. Overall, BcepNY3 is very similar to Bcep1 another *Burkholderia cenocepacia* bacteriophage, which is part of the previously described Bcep781 family (26). Due to BcepNY3's high homology and mosaicism between Bcep1 and the other Bcep781 phages, we can conclude that BcepNY3 is part of the Bcep781 family.

TABLE 8. Coding regions of BcepNY3

Gene	F/R	Function	Homology	Amino Acid
BcepNY3 gp01	R	Hyp. conserved	Similar to Bcep1 gp2	MPLIEGKSDKSRSENIRTEVEAGKSPKQAEAI GYAVQRRAQHGADFARDCDMNLRHVMDV AKDYKR
BcepNY3 gp02	R	Hyp. conserved	Similar to Bcep1 gp3	MAGTLTTANSTMYCTTEALFPTAQRIQGYA ADDAFDPDAVENGEYSMGIDGTLSAGFVFN EVPLTITLQADSPSLAQFEQIWMYEFQNRTKL QQDLTITNPAVGKRYEYKRGFMRSFKAAAG KKILQPAVIVFVFNQLQFTPIA
BcepNY3 gp03	R	Hyp. conserved	Similar to Bcep1 gp4	MPTTTIPIDQIVQMLPGVIGAGGAPGRLTGLV LTQDTSIQPGQLADFFQKTDVENWFGGLSNE AVIADAYFPGIVNGGQLPYDLKFARYVAAD APASVYGIPLTGVTLTQLQGYSGTLTVTTAA QHVSSNISLAAATSFANAATLIEAAFTSPDFV VSYDALRNRFVVNTNATGTAAAISAVTGTN NLADELGLSAAAGASLQAAGVAADTPASAM NRAVGLSRNWATFTTAWTAVIADRLALASW NSGQAYKYMYVAPDLEPASIVTNNSASFGA QVFAAPYQGTLPLYGDQATAGAVMGYAASI NFQLRNGRTVLAFRQFNAGVPATAHDLGTA NALRSNNYTYIGAYANAANNYTIAYDGKLS GKFLWVDTYLDQIYLNAELQRAEFEAMLAY NSLPYNEDGYTALYRAGVDVIDAAVTSGIIR AGVTLTNSQLQQIDAAAGVAGAGQLVQMR GWYFLIGDPANPGQARQNRTTPTCTLWYSD GGSIQELTIGSNAVI
BcepNY3 gp04	R	Hyp. conserved	Similar to Bcep1 gp5	MTNPVTLRPSEDEVFDTLWGWVTSLFDPAL ASQIAKADQNATSTLYGTYALIRPGVREALN QTIRTYDATAGTVSNELHTGYWYQVDCYGP QAPDWANTIAAMWRTMWSADALRGTALIP LYADQPQQLNIVNGENQFEQRYMVKLHAQV NQVATAPQQFFTEVPATTATPVDIVPLD
BcepNY3 gp05	R	Hyp. conserved	Similar to Bcep1 gp6	MNLHDIVRGAITQVNPDEAGTMFVSTGRTN VRGILTPTFSSIDAQLQIQAQKHTPLQHERGA LYTNSFLTVYAYGKFDDLSRPLGKGGDFAAF RGGWWYITQFLEWWPDWCAFEVTQQLNAA NIQTLLGYLQNGANLPVGLPPLPPGATTTP
BcepNY3 gp06	R	Hyp. conserved	Similar to Bcep1 gp7	MSVTRRGLTLPKDRYRSMSVKAGVLAGATY PDESGKKLADGTILTKDPRAGLPVAMIAMAL NYGTSKLPARPFMEKTITDRSAEWIKGLTVM MTMGYDAEVAMGQIGQAMKDDIKTTISEWP ADNSADWAGKKGFNHGLIWTSHLLNSVEQE IVK
BcepNY3 gp07	R	Hyp. conserved	Similar to Bcep1 gp8	MSTPPYRITFDPAGFIAEYPEFATVATPRLQA MFNQAQTALLDNTGGSPVTDDNVLRELFNM LVAHLLTLFGATPTSANSRPPGRLSSAAEGTV SSSFEFKLPEGSAIAPWYNQTQYGAMFWMA TARYRSARYMVSGGSGIGTARAYGQPTIQIP GGV
BcepNY3 gp08	R	Hyp. conserved	Similar to Bcep1 gp9	MAAVESKKKTGNGMVRVACKLPHGLNVRM PDGRTIELNGLHSRQAVAGHGMTYIPAKDW DAIQVVYAEAKWLRNEHVFAFADADDAAA MAEEREQVNAGFNPIDPKNPGIHGGVTIQRE GAKDPNAE

TABLE 8. Continued

Gene	F/R	Function	Homology	Amino Acid
BcepNY3 gp09	R	DNA Methylase	DNA methylase N-4/N-6, similar to Bcep1 gp10	MSEIANRCELMFGDCLLAMHELPAQSVDLV LCDLPYGTTRNRWDTPLDLSRLWVAYRHVC KPGAPVLLFAQTPFDKVLGASNLPELRYEWI WEKTNATGFLNAKRAPLKAHENILVFCDRA PTYRPIKTSGHVRKTSTRLGYSSNYGAQAVS SYDSTERYPRSVLRFASDKQRSKLHPTQKPV ALLEYLIRTHAAPGAVVLDNCMGCASTALA AMQAGCAFIGIENDVEHFETAQRRVRDYRS
BcepNY3 gp10	R	Hyp. conserved	Similar to Bcep1 gp11	MKAIEELAALAPAIGAMRHPSARDMQQKLN DVIAELRREYPETGVRAAWAAVVDTLRAVR PDWDETAGKPSMDNAVAAIRSMAAQCSAQE LQSLQVQLTPPAGVTDYDLETIAAAIGVLYA NDPVDDPIKLRLVDLFKSLRAHAPRT
BcepNY3 gp11	R	Major capsid protein	major structural protein, similar to Bcep1 gp13	MRDAQRIQNLARAGVILPRSVKNVSTPLAEY AMDAADLSPHLSSTGSSGIPNYLTTYVDPSVI DILVAPMKAAELVGESKKGDWTTLVAAFIT AEPTTTVATYGDYSSDGDSGTNINYPQRQSY FFQTWTRWGERELEMAGAGRVDLASELNYS SALGLAKFLNGSYLFGVAGLENYGLINDPSL SAPITATTPWSGSPAVEAVVNEVVTLFQVLQ TQSQGIITQEAVLHMGLPPTAMSDLSKTNQY GLSAAAKLKEIFPKLEFVTIPEYDTASGRLVQ LWAPRVEGKDTATCGFTEKMRAHSIERYSSY FRQKKSAGTWGAVIFRPFAVAQMIGV
BcepNY3 gp12	R	Minor capsid protein	minor structural protein, similar to Bcep1 gp14	MPFQKQVYITPAQGIAGDFASSNPMIYKLSSN GKMIADSSGVTVGTFAVLNADGTVTSKPGA APSSTSRIGFVHREMNAQIVTYLAEFGNTIQP GMPVALFGTGDFFANADVVAGSPSRGTKIL WDVVAGQINVGGTVSATLLDTGYILISESAT VNSLIQISNTGA
BcepNY3 gp13	R	Prohead protease	Prohead protease, COG3566, similar to Bcep1 gp 15	MPEVCFAFDKQTARSFDADGRMRVRDCILST AEVNPYRGREVVGYADLGLDPNRVYDLYRD PTELGHPDTLKSFEGLPLMIKHVAQTADNPR KEYVGGSVHNVRFDGKHLRGDLLVWDGHA IDLIESDELSDLSCGYRYVPVMRSGDADGQA YDGRMTAIRGNHVALVDDGRASGAHVADA AFREPRAPNPTLNGDNAMPFPENEQPGAGAP PAAGAQPADAAPAGAEGGGNELATIGAALK QLVEQNAQAHAAILQKLEQLGGGAAPAPGA QDGEPDGVRSPQGSEDDEMPGAEDNEEEGP TPRADDEHESAEDNELNPAGGENPTVKPNPH EGYAARGEAPPFGAMDAKSVRTAIDTAVAN ERKRAAAVEQAKRDVRYVLGGDIALDSASQ IYREALTQIGVDVSQVAKGSERAAWQAASA ASVAAAHGRTIQPAHAMDSAGANEAASRID ANLAKIKVRG
BcepNY3 gp13'	R	Scaffold protein	Scaffold Protein, similar to Bcep1 gp15'	MPFPENEQPGAGAPPAAGAQPADAAPAGAE GGGNELATIGAALKQLVEQNAQAHAAILQK LEQLGGGAAPAPGAQDGEPDGVRSPQGSED DEMPGAEDNEEEGPTPRADDEHESAEDNEL NPAGGENPTVKPNPHEGYAARGEAPPFGAM DAKSVRTAIDTAVANERKRAAAVEQAKRDV RYVLGGDIALDSASQIYREALTQIGVDVSQV AKGSERAAWQAASAASVAAAHGRTIQPAHA MDSAGANEAASRIDANLAKIKVRG
BcepNY3 gp14	F	Hyp. conserved	Similar to Bcep1 gp16	MLGSAQVARGYRKYEKDESTAKAWDDAAR KQLFRLNGMAGNW

TABLE 8. Continued

Gene	F/R	Function	Homology	Amino Acid
BcepNY3 gp15	R	Head portal	Mu gp30 homologue, similar to Bcep1 gp17	MARAPLRMPGRDDKVLAPVSVDRQTEAMY RRKLERAVALMAASYLRSIERKYGRALEAN VDTGRLPDIAQDASAQAPGAASSDMFDEMK RLRDYWQRYFDTFAREVTTGAFEDLYVDNQ RMWQSRLRNAGFDIKLDMTPSQRLVMEAKV QENVALIRSISQEYHTAVEGEVLRHFIAGRDL KGLQDKLVERGKVTTNRAAFIARDQCNKAT AQFNSARQRELGLHWATWQHSSAGKEPRPN HVRAGREKWIFNTQVGIDFGDKFGSVLPGEA INCRCSSRTIIPGMGRTPGGREFDPGALGEITG FPGAYREAA
BcepNY3 gp16	R	NUDIX Hyrdolase	COG3567, NUDIX hydrolase domain containing protein, similar to Bcep1 gp18	MSRRNAKKRTQLAHTGRRPEVAKAAALAA AATIATATAAQPVPADMGRRGALNALDAAP VAEPSPSLRLARQFEVDVSNYTPRERRAASY ALDFNGTSMDALSFVTSSGFPGFPTLVLLAQ LPEYRAMHEVLADECIRTWGEAIGGTKEKA DTSGLAAGGNAASTSDGDQLKQINDEIERLRI RDAVRTTVIHDQAFGRAHPYFKIKGDDQIMD TPLVPRPYTVPKGSFQGLRVVEPYWVTPNNY NSINPVADDFYKPSTWWMIGTEVHATRLHTI VSRPVGDMLKPTYSFAGISMTQLAMPYIDN WLRTRQSVSDIVKQFSVSGILMDLAQALMPG ANVDLSMRAELINRYRDNRNILFLDKATEEF FQFNTPLSGLDALQAQAQEQMSAVSHIPLIKL LGITPTGLNASSEGEIRVWYDYVRAYQRNAL QQLMNDVIVMIQLSLFGAVDPSIKWQWNAL RELDDLEVAESRYKQAQSDVLYVQEQVIRPD QVAARLNTEPDGPYAGKLDANDDPGVPADD DIDGVLTYVQRLAEGGDTGAPGGARAGATA PPTVANVNANVKPREAGAQDAAMRAAGAV YVVDGKVLLMKRPAGDWGLPAGKVEGNET PEEAARRETREETGYDHDGELVPLGKFDGFF HAFVAHLEPFDVELNDEHTAFDWFNPDELPH PLHRDTAAIVDAACKALDRLERA
BcepNY3 gp17	R	TerL	TerL, similar to Bcep1 gp19	MGANAALVEPPVVAASSPAALGTPSAVVDP VAFLINLARTNFAAFVSLVHRPRYRHSAFSA RVCAEIDKFIDDLLEGKRPVLMLTAPPQHGK SSLISRCLAPYLYGRLTGLLPAVRIANATYAL PLARRNATDAKSIMKEPVYRAVFPHVSLIGF KGGKDTSNEFDVPAGGEFRGVGVGGPLTGF SIDVGIIDDATKNAEEALSAVVQDGLENWYD SVLLTRLQQLSGVILIGTPWSANDLLARVRR KMEGQPNFTLLSFPALNDPDQIGYNPDLPLG ALVPHLHSADKLREMRRNISEFWWSAMYQQ VPLSEFGAIFPREHLQYYHAADLPKQFVRVI MSCDATFKDGQASDFVFVGVWGKTADERV WLIDWRREKLAFMATAQAIADLKRKHAAVS RVYIEEAANGAALIDMLKKHFPMLEGVPPLG SKEARAHAVAWVWSNNCVMLPHPDERPGIG PVVNEITSFPDTVTGHDDSVDGMTIALHQLC LRTPIAAMITRDILNKA
BcepNY3 gp18	R	Hyp. conserved	Similar to Bcep1 gp20	MADLQPIKASLVADLVDLINENLADLLKVRT ATCPDCGGSGTVGGERKWNDRGGTDVFDD GTLTTCVTCGGVGAIERFEIDHDLLKSRRFGR YVEGFDVKHGIIVPKMRSKDKAFAMLVKLL GFDKAVIEVANGASFVDTVSDEQRAVVVEQ LKELAAMGLLDGR

TABLE 8. Continued

Gene	F/R	Function	Homology	Amino Acid
BcepNY3 gp19	F	Hyp. conserved	Similar to Salmonella phage MB78 13.6 kDa late protein, protein similar to Bcep1 gp21	MLRLRDLQHQGQVMSNPGVVEHSTLSFSGR ERWRNCPASVTLSKGMPDNSSPAAAEGTCA HTVGEFYVRQHFDLPGAAPRGTEAPLQAVPE GLDLEGKTVEEWNDDLRRHGKAYRDFIISLI PPGVEAFVSLEQRVAAKTIDGRLFGTADCLI WCPGARVLIVVDYKYGFMVVDVGTAEKPN AQLAAYAVAALDSCTLQANGVMLAVFQPR RNIGEPGHKVYLSAEHVAAEQQRMREEVVR VDKATASPDLFIVAGDHCRYCKAKPACPRM QDALQIAFDVNAGRRSILDMPEDDLIALYSA RSGVKSLWEDVEQRIELLAQRGHDALTIKTS PGRRMWRNAKAAALTLLALDRTDLLQPVAL SEAIAHIPEALHDDLIGKSRDSQSIVVKTPAAP GAVADTFAKYAKSVDTTQDKA
BcepNY3 gp20	F	Hyp. conserved	Similar to Bcep1 gp22	MSTIDKLAGYEAILTHHSIITPQINKLKPTKPA EFYALIALPAAAQADLWAILCERATSAFGHA NNFEHGIKTNATSKKPIAGVPGDALVVRAAS QYAPEIYDADGTLLNPQNPAHLQTIKAKFFA GTRVRTILTPFHWTFQGRNGVSFNLAGIMLV PSEAQRLAIGGVDTASAFKKFAQPGTGGVPA TAGAPTDAAAAFAAGGNPDAAGGTLPANPN PFAQQTGSAAGAGGNPFL
BcepNY3 gp21	F	Rus	Holliday junction resolvase Rus, similar to Bcep1 gp23	MPTTRYSLTLPFPPSLNRAYRAVAGRVVLSK AARQYGVAVRNALPAGRVERIAGRLRVVVT VHPPARLVGRAWDVANREKLLSDALTKAGF WRDDSQIDSFRVDRGEFLETRPAGCAVVDVE VLAPVRFFP
BcepNY3 gp22	F	Hyp. conserved	Similar to Bcep1 gp24	MGKDKFIENANGILAAFRAALEQYGLDVTN MHTSGDLVLLASKYTAVYGKKLEPGVEPKF NFVDIALEQFNSQLTILLMRYAYAALPVTFTR VPAPPGVSA
BcepNY3 gp23	F	Rz	Rz, similar to Bcep1 gp25	MIQFYAIGAGALAALVAVWAIVRKLTGAGR AAGTAQGESRAIDDANKVTASAASAADAGR QQLATEIQDNAQASDDYAARVLGAGSVQDG ATAVNDAIGRANTRDRAAR
BcepNY3 gp24	F	Rz1	Rz1, similar to Bcep1 gp26	MRKLLTIMLLACSALAACKTAPLPSTTRSAA PTHGIVPRVECDALLTDAGNIPAYPVPAEGA DAQAYARAQQLWAIRAIRIIDDERAGRRAAV ECFARLRAAGLIH
BcepNY3 gp25	F	Hyp. conserved	Similar to Bcep1 gp27	MNFNLSSFANAMAAAVPTIAAVGADTGVIH SFVAGVIASTEQAYAAAGAGAHKKAAVLAA AETFVTSIGHEWSTVAPHVESFIEVAVGAYN LAATLVPGLPAVPTGTATGFVNAVENEVKA VAAVAAPLVTAFENTFGGAKPAPVVTQPVS VPAAAAQPLPAAAAPLAGGL
BcepNY3 gp26	F	Endolysin	Endolysin, similar to Bcep1 gp28	MAAPLIVGASGRAVVFLQSRLGLAQSGQFD AGVATALRQWQEAHGMTPDGVYGSQTNAV MTARALSDIADAAARLRVDVPAFQAIIQVET TGSGFLPDGRPRILLERHKVWAATSPAQRVL LGAQDCNPTPGGYATGPDADARGAGEWVR FERVAAVTGDEVAAQCCSWGLGQVMGANY ATCGFTNAVGLMFASALNERAQLDVMVRFA LPQAGLLGALRAHQWAAVARIWNGPNFAIN RYDTKLAEAYTALTSQ
BcepNY3 gp27	F	Hyp. conserved	Similar to Bcep1 gp29	MNQLEIDWSAAPAWASWAAQDADGGVFWF SQEPTLSRVTSRWHCRDDDTTCSAALVGVV RVGPAAKWRESLTRRPPAFDKEQPAANFDPE GMLRAQIGHLLEDMTAEQLDALYAHVRQQL DAAKFDELNRQLDKPFGEVPREWQLKAFEA FLDGNPIEYRDVARAASWFAAPSPNWAPSLR YRVKPKQS

TABLE 8. Continued

Gene	F/R	Function	Homology	Amino Acid
BcepNY3 gp28	R	Holin	Holin, similar to Bcep1 gp31	MNDRFDTMFTSLGQVAIAFLGGLIGALMRRE ASTWQTAILGACGAGFVGFLVAKLCHASGL SDDWTFVMVGVSGWLGAERTISYLERLFAA RLGIEQAAAAEPPPAAAGDKEKQS
BcepNY3 gp29	R	Hyp. conserved	Similar to Bcep1 gp32	MNNYAIVKDGTVVNIVVWDGRAAWAPPSG TNAVPIPADESVTLGDTYNGVKFQSAGAPA
BcepNY3 gp30	R	Tail fiber	possible tail fiber, gpH similar to Bcep1 gp33	MTAQTNQPAKFLVPFAQNDSSRVELPVTTAD ATRASQSLGFPPSTMQPPEAGGVPPQGEDFN GALNQVARIAWWLMLGGQFPFDSAFATATQ IGGYPKGAALQSADALGSWISTADNNSANPD TSTDPAGGGYVPGYQYGTTALTGLTGGTVT LTNAQAAKATVTLAGALTSALTLVVPTWLK HWTVTNNTTGAFAATIKTAAGTGIAIPQNGA PTPVVGDGANIVQVGENIAQATRLTQGVRLD QLYGTAKGFLRATSSGSTTVPPNVTTIYVSGS AGGAGGGGCATVGTPNIMSGAGGGGAGQFT EWQALTVVPGETLTYTIGGAGNGGAPNVPG GAGGNTTITGSVSGLLLSLTGGVGGSPGVGG AYSVYTAGGNGGAGAPAGGYGQDTGPQGV GANGGAGGSNPFGGGGTAVRGSQIGNTMLA GLAATGYGCGGGGCGGVYAAGNTTIPTSTG NSGGAGTPGVLLFAW
BcepNY3 gp31	R	Hyp. conserved	Similar to Bcep1 gp34	MTDYLGRTVQKQYSNSPVLLALLASFDQWV DPTKFSADFLANVWDISTAQGFGLDIWGRIL GRSRLFQVAQTPGNNFGFFSTGGTPWKPWG QAPFYGGQAGGTVAFALQDAYYRKLLLVKA AANIARCDCPSINALMRSMFGDRGKCYVGY DIAHPMDIAYHYEFFPTAIEKAIIESGLFPQPA GTNPHYVYKTLTYAPFGFRTMNGGTNPNVV VGFNQNPFYSA
BcepNY3 gp32	R	Hyp. conserved	gpW, similar to Bcep1 gp35	MTTTNVPQPTFTPTGLQLAGEQAILAGVQAD QVAAFATAGKTLSTELTTPQGQLASSEAFIV AAWQALFAQLIANVDPLTSSGAYQDALGRIY FMTRNPAVAATIPGVVVTGTPGVTAAAGTL QARSPDGSLWSNQTDVTYDATTGNATVTYV AAVAGVGPVATPNTLKIYQQVNGWLGIANP NGSVAGVDVESRSEFETRRQESVSIGGIGQA ANVRAAVLAVPGVTDCYVYNNGSDSAINYG ATNYPIPAHSVAISVTGGADADVALAINSKL DCGCGFSGQGTTTVTVMDSINYPPPYPQYPV RFVRPPTVEVYFNVQVAQLPNVPATYIQDIQ KAIVAAFVSGFSTDDGKITLSRARIGMQLIGA AYKPVVAVLDPNLIPVNIFIGTHANPTGESIT MGIDQQPTIANLNISVSQIAV
BcepNY3 gp33	R	Hyp. conserved	Similar to Bcep1 gp36	MNLPEPNPKEHWPFQRRVLAAGVDAARNSY IKAQCGDARRMALAHQALLQIRIDRLTRP
BcepNY3 gp34	R	Hyp. conserved	Similar to Bcep22 gp42	MKTYHRMHTELVLLAGGYLEVACPNPEAPP LRRHWQIRRTVDYRSIPWC
BcepNY3 gp35	R	Hyp. conserved	Similar to Bcep1 gp37	MTPTPLSALQPSAITTDGSDVINVRGELFRIR VAQILSSTGFRIDGAVRATYCLTFDEAVAWM S
BcepNY3 gp36	R	Baseplate assembly protein	baseplate assembly protein gpV, similar to Bcep1 gp38	MADRAPTPYNSPFEAQFHEGRAQEWMIVKLI REIHTATPVEVKAVRIIDDRVGFVDVLPLLEE TDTNDAVIEQSLIYNVPFLRVQGGQSAVVLD PAEGDIGLAVFAERDATALATTLQAGPSATK RAYSSADGFYFGGFLNGAPTQWVKFLAGAA GIDIHTPGDLTLSAAGAITLTSGGATTINAASF VVNAPTTFNDTIAGTKTGAGSVQFAAPVGAP DFIDGNGVRHATHIHDDPQGSQVGTPHN

TABLE 8. Continued

Gene	F/R	Function	Homology	Amino Acid
BcepNY3 gp37	R	Hyp. conserved	Similar to Bcep1 gp39	MTPIRYPLDFTPWQPGRIIGFAGSWRPFSQSE YEASAQADAAITTAMYRAKHGLPPHLGGAH G
BcepNY3 gp38	R	Hyp. conserved	Similar to Bcep1 gp40	MTFNPFQERRVRVTVTVTRPDAQGNQEPDIY TFVEHRMRIAVSLGGAQYGNARVEIFGVPLD TMNQIARIWQDVLTPITSDTISIDVWNGQDY VPFYQGTIAYSYIDPESMPFVPLVIEANASYA LMAEVMSPYSNAGPVKLSDVLKAVCTPAGF AVDYSASTTDYMLTNVRLTGSAADQIRAAIA QFPNLTYDVSLQRVQVRDSQVSMFADAVPV NAGNGMQKAPRYSTSGITFTSLFNAQIRPGA ALAIDTSIAYINRTQWIAAVVQHTLEPNYPNG VWSTAVAAQGYGKRDGTAVAPGTSTATAP
BcepNY3 gp39	R	Hyp. conserved	Similar to Bcep781 gp39	MAIRQHVRCRECTRRRALPKVLTLYLRVPRC DCGARNWRPDKHMNQRDNGATRCDCAGY WFPHRRGCLYCHYRRDGTMRMFGDYDFAD RNYDPELGWT
BcepNY3 gp40	R	Hyp. conserved	Similar to Bcep1 gp41	MDYRVKRQIGAIVGGAVLALVIVLAARDAH ADPAAVQLIAAFDRVNNACLGAPLRADGKY APACVERDRAARALERAGYERSRHDVWFRS ADMRTLYALTRVATADARRPEDAPRGLRAL VRLAGLSDVELLAIWQQHASELRDGDPAAW AVASELMGQISAAHPNDPRFMVD
BcepNY3 gp41	R	Hyp. conserved	Similar to Bcep1 gp42	MEEQHIQDWGNGWNDAMRGFVQRPDQSLD YYVGYADAMADAYRPPTIH
BcepNY3 gp42	R	Hyp. conserved	Similar to Bcep1 gp43	MKGSQRRDGQKCYAVCFVGDEPLEVRVKQP RGWYTLWLQGEPFKNRIAMDAIKAARAA
BcepNY3 gp43	R	Hyp. conserved	Similar to Bcep1 gp44	MDDVLDFLKFIGALALALVIILTAIGGGVGGI VYVSNRSECAQYHTLSGRATYFSWSTDCLV RNDDGKWVMLAAFKGNTADVTVRNK
BcepNY3 gp44	R	Tape measure protein	gpT, tape measure, similar to Bcep1 gp45	MADAANANIVDELVVTLELDARQYENADKT IDKLVTKTEKKAVENDKKAKKRHEAQKKRF DETKRAAASLGGGLMKLAGIAGAVLGVGAG AAGLAGMVLALTNTETALRRAAVATGMTN RQMSALSSTARRLGSDAKAGGDAFAGLARE QQLAPLTGQAPNLQALASFGVNIQQDVPHM LQQLQRTYRASSPQQQGFMESRLTAAGVSPD VIVAMKASVDALDAFNQSMRETSEENRQAL DAFYDAVSTVSNNLRNMANVVMTVAAPYV QRFGEYLSDATAELADFENDVAAAGGGVNG FMKVLEQRTPRLAALLHDLASGLGFLGEAV DVAVFGWQTIAKALEWGWDKLAGSKIGQFF GLDKAGQAVVSTVKQTWREAVATAREFGP APVASTFGIAVDDPNRVKLSAGAAARLAGG VAGATAASSPSKLDDSRLYVMNALIARGASI DQAAAITANGVRESGLVPSAYNTNGGGQGA QGIFQWRGSRIDAFRRRYGIDPRFGTVDQQID FLMSDPDERRRMGLALGAGGSAAALGTRFS QIFEAHGKAAEDAARGPIAQRFADDYRARFA ANDQSTGGAPISINGPVTVVANDPRQLVNGI QRQSGVQNYASGQR
BcepNY3 gp45	R	Hyp. conserved	E, similar to Bcep1 gp46	MAGLPTVAAFDVLALIASELPSLNAPTPIYAI VASDTFFPLTIPSSWGEFDVKYEYALSDYPTE QGGFGVFNKVRRPSSIDVTLVKTGSDLARFA WLAAIQQMQANNPLQLYTLISPQGIYLDFTIG RMSHDTRADKGSNMLYLTLTFVEIPQIASSV DGLSNTVDAKSGPIEQIGRLFTRAATAAETAL INIGNFLTS

TABLE 8. Continued

Gene	F/R	Function	Homology	Amino Acid
BcepNY3 gp46	R	Hyp. conserved	Similar to Bcep1 gp47	MTMNETISAVIDASAPRDVGKRFRVNEISPLV LSGYMLRLVAALKGAEYETLVSELREARAD TENPDAALNVVLRTLSGSDPAAVHALVTDLL THVDIAADPRHPEGFRPLLPTDIRELKTLGDV LIALVVLNLGE
BcepNY3 gp47	R	Hyp. conserved	Similar to Bcep1 gp48	MITLTELPALVADRYARAALKAIDADPDSGI AGLAFQHYAEVAKLPDAHAYLLPFVRGTVA GSEFSLEFHLKGWRNIDRLEQAALMLHVAFL TDREPLDIPVAMQGAALMAGGSDLRVTFCSP AIAAVLNSRHADYVQLETVLSTEDVYNLTEC INVEAAREYAARKRDK
BcepNY3 gp48	R	Hyp. conserved	Similar to Bcep1 gp49	MKDEYVDVDGKRYCRTTFYRDSGHHVVMF TAFTAKGVWRVLPPRCAKIRARIDSALDNPQ TSD
BcepNY3 gp49	R	Hyp. conserved	Similar to Bcep1 gp50	MIKTPQQLWIARRDAANTAGLAATIRRTLAG PQLSPFYAIGQQAFRDGVPFARNESDPWRAG WLAAAGATRALG
BcepNY3 gp50	R	Tail spike	Tail Spike, similar to Bcep1 gp51	MNMKKILSRVFACAIVLAAFSQFALGQFAPG QILTAAQLNAAFANVLPIAGGTLTGPLTATTL TATTSVSAPAVNATASMSTPALVVTTSLTGA AKAALTGTATSALTTDQAQITGQLNNPTATA DYIYGAALGAIGQPVFDGAGVRGVATQTAG STNWTTNGVAGYVLNNQAGATATKSAVGL LGINICAVDNCQSWGTSTIVSDVLGFTGVSA GVGRQLYGNESDLNVSSPNTNGIAYMAAGT ALTPSAYLVGFQVSNLYGGTGAKWNYAFSS PDGNSTNAMYIGLTATSGNNLSSQPIVFNYT DAGGAKHAPQLLATPNGLQTTAGALIPGSAN STALGLPALPWGTIYGTTAVISDTSASNQATL TVAAPNDTQGASIKFQGNGVTTPTKTIRVQN GAFQWVNDAFNNIIATLSDSGSLSVNGNITTP TGTGALPVYTATGASLNGPHAVQGSNALSA GAATVTFSGAAAFSSASSYVCTANDTTAAN AVRVNQTSGTSVTFVGTGADSIQYRCVGN
BcepNY3 gp51	R	Hyp. conserved	Similar to Bcep1 gp52	MSLIVIPIKATANQIASTVLDGLNAQIALMTT DFGLFADVTYNGTRVATGRLCLDRTDINAA KYLGLPQPLFFADLQGTSDPAWTGFGTRYLL CYGTPPAAVTAPAPTLAFATEGRLDIDFVLD KSILG
BcepNY3 gp52	R	Hyp. conserved	Similar to Bcep1 gp53	MDSLALDPATWDITTDAYGNLATVGDATPG DHSGGAYRMAQDVACRCMSWLGEVYYDTT QGVRYEQVLGLAPNLVLVQALFVNEALKVA PVAQAVANLTYTAGAQRRVTGQLVVSDGSV TTGVVQL

TABLE 8. Continued

Gene	F/R	Function	Homology	Amino Acid
BcepNY3 gp53	F	DNA Primase	DNA primase domain protein, VirE, similar to Bcep1 gp54	MQIQFAQVTATNCELTKTFQIGHNGQLDSSA IAHMTEGFARIRAIEDVGQLRGVLEALTPHD AITCGIPQRGDTPLTTRAGAEFRRDAVARTN EAFTYPYGAALFPIDVDVEGDAFQSVAAVLD ALESASPWLRDVHRVARPSSSSYVGARGLRG VHVYCGVTNGADIPALAKRMQIEQWAAGH GHVKISRSGALLVRQLADALVYQPSRLMFES SPVLHDVTRDIPPDQAFVERAPDPLAGRPGA WRLNGLLDVGKLPRLREIDERRFVTQAKQA KDAKRRDAKRIAIEYQTQNAIASGLEPEAGE RFGLLAIRALGDAKLPASWEVHVKDIGRVTV ANILDALPASLGFQCADPFDTWRPDLDAKHF GKAEIVMLNGLPGIWSHKLQQFFAFDADPAA DLGTPLAMAAEKLCGLIEYPESSKRAAPFVN VMHALKCLFDEIDARATVHAATGEMRLEGV PPEAELIDALSRVGCAGVTPATVKTAIETLAA SNFVDPWRDAMIALPQWDGTQRLDTFFVDL CDALPSDALTATTQLLFAGIVKRQLQPGAPL PVVPVLIGPGGTGKSYFVEQLAAALKFPQPP ALAFTDTIRMTMEAATSGIAELAEMSGMGR RETEEIKLWTTDTSDTYRAPYERRPSAHPRRF ALIGTANKHETNHDATGNRRFMPVFVNRPID PNWHVEALQLFAEAKARFVEPDGEYARLVR RASALVKEYNDADMRDGIGMPITDLDDILPP VLRALHRQHGPRIPSAELRAMLDRTPSGRQA HARAIAGWLLTRGWQPIRSAAARFYDAPQA FIDNLIDEKINSALNSTSSPFNTP
BcepNY3 gp54	F	Hyp. conserved	Similar to Bcep1 gp55	MTWAEFKAKMEARGVTDDTPIGYIDIGCGA DVDVDFRDDGSVCVSDA
BcepNY3 gp55	F	Hyp. conserved	Similar to Bcep1 gp56	MNEQRVHEHPSGRLLPDANFTVAEAARFLV GLPAPLTTSDLGKLGKDCSMVCRIKGLQWTI RPTPLALWPSERAYPAAVIQEVFKTNPMTAP YVPQPKTQEQTA
BcepNY3 gp56	F	Hyp. conserved	Similar to Bcep1 gp57	MNAIDALGEAIEQALEECPTNEVLAFLTGAF VGLITELARRHGADASQDIKIDGGKNRDITIH AAKAA
BcepNY3 gp57	F	Hyp. conserved	Similar to Bcep1 gp58	MMARFVYIHGSLHSVVHSDAEAWAIIRRFPL GTLYEVRDAAGQPIDEFEPF
BcepNY3 gp58	F	Hyp. conserved	Similar to Bcep1 gp59	MRGLSMYARSFENPGQYSTNSPESANGAHA ARPTRTRLKTGKGRNATPTKRDRSKKVKYG AAAGYGSVT
BcepNY3 gp59	F	Helicase	DNA or RNA helicases of superfamily II, Uvs helicase, similar to Bcep1 gp60	MTRKTPRWYQAEAADELFGALATAKNTNPI AAIVTGGGKSLLNAMLIERIAQTWPQARVMS LAPSMELIKQNVEEARAFWSPAMVARLGIYC AGLRMKDRMHQYIFATPQSVARQAKRFGPF DFVLVDEAHLFNIDMKTARTIVDTFRAANPH VRFVGMTATDFIMKGLKAVPLTQCGLFDAK VYDLTSGRNFNRLVREGYLSPVVSPSLRFPQI DTGGVKTKGGDFDEAELARRAMDVTRECV RVALEHAPDRKHFMWFAVNIEHAHMIEQAL LDAGESAVVIHGELEKSERVSGVDEYLKKQH RHIVSVAMLTTGFNAPFVDCLVALRPTRSLV LWRQIVGRGLRPYAGKENILVLDAGGNFGR HGAINEEIGAGDSRAGLWVCTSEEVRSPFPV KLPDGTTAPGRERSGIRFPINSPEQPEFDLRVI LGLMEPDSEGCGYLNDAEHMTCRQCGRPRQ GFLSVRVKRAANERGVLAESDSYEIHDEDSV VLRDEACREVRQLPVHEMQIAPEGNSVLKFE FGTDFGPYRLRLDFDRTTADNKWYAFARKF YEKATGRKVPTEAYRVLLQRELIPKPIDITLT KYEDGQVFLTELRFLRNEQLESFKYDPNY

TABLE 8. Continued

Gene	F/R	Function	Homology	Amino Acid
BcepNY3 gp60	F	Hyp. conserved	Similar to Bcep1 gp61	MIRTTDESRADALKNAAWRLKRDLERGNVG AETSHVYMADLRVLLEAVEQHEAAPAETAM LDWAVGRWDAEVKNRPLTNVHRRSLDDTW RQVVRHCGGDDVSLIGPRHDDLVAASAPAPS ASLEGTGNGADERMAFDSWARDIRMDVSVS LEGHYTDAATAYALAAWRARAAASQPAAE PTIPAELHHDTAKLVRRFARALANKLLSAQR KYGYSDNWMRDDWADECRAELVRHIQKGD PRDVAAYCAFLWHHNESTAAAAAAGQEAV ATLHDDGHYTWNPKVPRPDGYDRAGWRMS VYTAPPAQVATRQGLTFEDWFKAELVADRM HGTDEEVARIAWNAARAPLDHDGA
BcepNY3 gp61	F	Hyp. conserved	Similar to Bcep1 gp62	MTAPDRVNGVPRIRLERECASKNRYPDEIMA RASGLHHQDRNKLDALWVYSCKHCAGWHL TRRDNGPRWRV
BcepNY3 gp62	F	XRE transcription regulator	XRE xenobiotic response element family of transcriptional regulators, similar to Bcep1 gp63	MSFAIQLIKARNARAWSAPELSRRAGVRHSL IYDFEADKRLPNLKTLLRLADALGVTLDWLC GRTPE
BcepNY3 gp63	F	Hyp. conserved	Similar to Bcep1 gp64	MKRFIVFAYDEYYPGGGWNDFVTTVDEFDA ADQEARRQLASEAGTKDYAHIVDTVTGRVW EYA
BcepNY3 gp64	F	Hyp. Novel	Hypothetical novel	MLQVPKLRVCVMPHRNRIAAWFPDQPLDGK LFVGNDTRLTVQYMYGRRLRKLTQRLR
BcepNY3 gp65	F	DNA Polymerase	COG0749, DNA Pol I, similar to Bcep1 gp66	MTERLLFLDFETSSHTDLTEHGLGRYLADQS TRAYCFTFRLPGMSTADLWEVGQRVPEVITR HVRAGHLFVAHNAPFDFWIWNTVLRRQRGY HDLPELQIGQVRCSAARARYNGLPGSLAGAC EALGLPVQKDTEGAKWMKEIAANPDWTPA DHPEHFARTYKYALIDTDAMVGVWENTVPL PAREQAYFEMDMRINARGIGVDVEAAQAME DLKAFAEAQLDYEMAYLTDGGVLAVSEVEK IKTYAATLGEDMDDAGRETLKKIAARDNLP DSLRQLIELRLDASRAPKKSAAILRAHVGGR LQHSTIYHGALSGRSTARGAGGAQTLNTARP RPGKKTADCEAILDACLRHDRAYLSSPEVGPI LAALADAQRQLFRATQPGHVLVGADLSGIE ARFSPWIAGDLELLEAFEKGVDPYKLAAAAI FQVTYEAVTKDQRQIGKVAMLALTYGGGAG AFVSMAANYGVHLPPEQVDEIVLNWRAARP AFERWWSLCEYSVLMALDQPNREITMPIGRD FCSHVTFVHDGRALRMQLPSGRAISYHNARL HLEPGANVPIAIYDKPEGYIETLDRKILSNNM TQGLARDLFWSVLLDVDRVEQIVHHVYDEA LLEVREEVAELRCDQLVERMCRGEAWCPGL PLGAEGWFGLRWRKD
BcepNY3 gp66	F	Hyp. conserved	Similar to Bcep1 gp67	MTDLTATLAERGARYGKFEDHAVIAQGLKD QMWATEGWSRLAPDQRQALEVIQDKVARIL NGDPDYTDNWHDIAGYSRLVEDRLNERAPL CAAADLARPRTVGEQRAPERGWIEWKGGEC PVPADTLVEYVLRDYYKGCKRAGGLSWGDT GSDSNIVAYRLLPKQG
BcepNY3 gp67	F	Hyp. conserved	Similar to Bcep1 gp68	MPQPTNTELAARYKKLPVDQFTITELTDALG VEVSAALLGTSKRAVYTVRNTNVLGIERHKL LIDAVRSKETECRERLLFMLQRRERREASRA AKAARADDKQ

TABLE 8. Continued

Gene	F/R	Function	Homology	Amino Acid
BcepNY3 gp68	F	Hyp. conserved	Similar to Bcep1 gp69	MKFEFDSIEELQGFLEFARNVSAIFRAAPTEP YPDRASWPDDIALVPDAHLTPEERAAAQHE AQVAATGALLSSGLSIAVQSAAPDSASVAQF GTSVKPPRKRRTKAEIEADKAEAANTPPDET KAPSDAQATAATSPLAGANPFDASTSNPPAA TAALNEAPAGTTGTLSALGSDVRAEAQSAV DGLGADVDPQTYMQLRVAEMGGAFDAREH MKKCVEFIGALGKARYDEAFDLAGTSRSVA TYTPADCAKHIAALEYLQLSSKGA
BcepNY3 gp69	F	Hyp. Conserved	Similar to Bcep1 gp70	MYIGALLILLAAFAFASAGLYYVAAILVLFG KLLGALVVGIVGGCRLFAKGAWLALLGLAW IVSRLAGIAWELAHRAAQEGDAATASAHRA GMVAVEWLKRTYVKREKRERGL

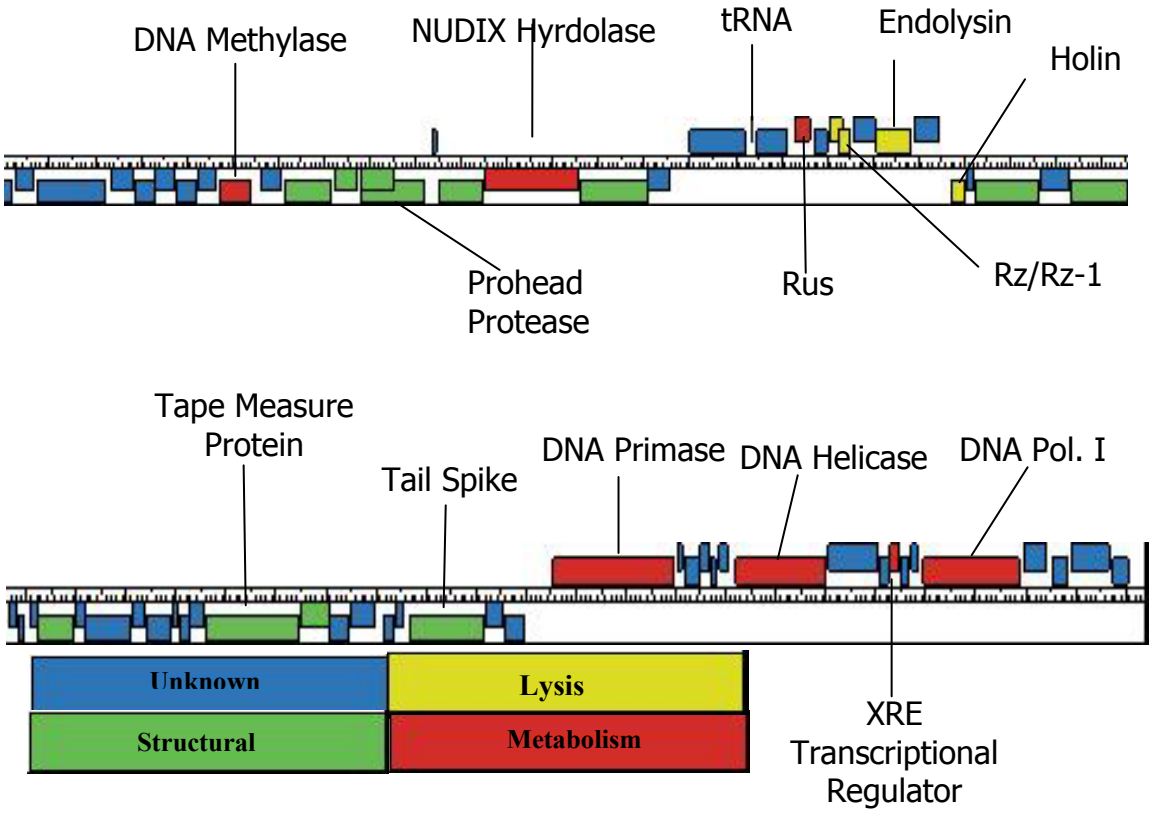

FIG. 5. Genomic map of BcepNY3

Genomic map of BcepNY3 divided in half. Boxes represent ORFs and are color coded based upon function: unknown, blue; lysis, yellow; structural, green; metabolism, red.

Metabolism

BcepNY3 has 5 genes involved in DNA metabolism: Holliday junction resolvase (NY3 gp21, DNA methylase (NY3 gp9), DNA primase (NY3 gp53), helicase (NY3 gp59), and a DNA polymerase I (NY3 gp65). Prior to recombining, a cruciform structure forms with the two strands of DNA and endonucleases, called Holliday junction resolvases, cleave part of the structure, allowing the strands to recombine. BcepNY3's Holliday junction resolvase is highly similar to Bcep1 gp23, both of which are homologues of RusA in coliphage 82 (26). Methylation of DNA can be used as a regulatory function or to protect bacteriophages from endonucleases. The functional role of BcepNY3 gp9, a predicted DNA methylase, is unknown like its relatives in the Bcep781 family (26). For DNA replication to initiate, it requires a primer provided by DNA primase. Bcep NY3 gp53 demonstrated homology to Bcep1 gp54, a previously predicted DNA primase (26). As observed with the Bcep781-like phages, BcepNY3 has a gene (NY3 gp59) that shows significant homology to a T4 helicase protein, UvsW, which is involved in DNA-DNA and likely even DNA-RNA unwinding at branch points during phage replication and transcription (4, 26). NY3 gp65 is predicted to be a homologue of DNA polymerase I based upon its homology with Bcep1 gp66. Even though it has a conserved DNA polymerase I domain, the functionality of this gene along with the Bcep781 family and other similar phage homologues is unknown (26). Although all members of the Bcep781-like phages were annotated for a DNA polymerase III beta-clamp unit, BcepNY3 gp14 only received a BLAST return for Bcep1 gp16, most likely because Bcep1 gp16 differs significantly from gp15 of Bcep781 and Bcep43 (26). In

prokaryotes, the DNA polymerase III beta-clamp clamps on to the lagging strand, keeping the replication complex together and increasing the processivity of the replication reaction. As observed in Bcep1 gp16, the BcepNY3 gp15 has a significantly short sequence in comparison to the typical length in bacterial genomes, which leaves the functional capability, and thus the annotation, of BcepNY3 gp15 unclear.

Along with DNA metabolism genes, BcepNY3 has two other genes involved in metabolism: an XRE transcriptional regulator (BcepNY3 gp62) and a NUDIX hydrolase (BcepNY3 gp16). XRE stands for xenobiotic response element. XRE transcriptional regulators respond to foreign chemicals or substances and alter transcription of specific genes. It is unknown what BcepNY3 gp62 actually senses or regulates. NUDIX hydrolases hydrolyze organic pyrophosphate bonds on a wide variety of substrates. Recently, it has been shown that there is a linear relationship with the number of NUDIX hydrolases in a bacterium and the linear size of the bacterial genome (17). It was found that *Burkholderia cepacia* has more NUDIX hydrolases for its genomic size than expected (19). It can be suggested that the Bcep781's NUDIX hydrolases were acquired through the large available pool of hydrolases in the BCC. Also, because NUDIX hydrolases have broad applications, the actual metabolic function of BcepNY3 gp16 is unclear.

Morphogenesis

In the Bcep781 family, there are a predicted 21 genes that are related to the tail or baseplate themselves or the assembly of these structures (26). BcepNY3 has homologues to all 21 of these proteins in the region BcepNY3 gp30 to BcepNY3 gp52 with an extra gene that is homologous to only Bcep781.

The tails of bacteriophages are vital in recognizing and adhering to the bacterium. BcepNY3 gp30 is predicted to be a tail fiber, due to a 99% similarity with Bcep1 gp33. The tail spike is required for adsorption to the bacterium. BcepNY3 gp50 is 100% identical to Bcep1 gp51, a predicted tail spike. It is noteworthy that BcepNY3's tail genes are highly identical to Bcep1 and only slightly (less than 50%) identical to the other members of the Bcep781 family (Bcep781 and Bcep43). Because tail fibers and spikes require a degree of host specificity, it is reasonable that BcepNY3's tail related genes would be more identical to Bcep1 due to the fact that they both are pathogens of *Burkholderia cenocepacia*, whereas Bcep781 and Bcep43 infect *Burkholderia cepacia*. To connect the tail fibers to the base of the tail, phages use a baseplate protein as a connecting device. In BcepNY3 the predicted baseplate assembly protein is BcepNY3 gp37, based upon its homology with Bcep1 gp38.

To regulate the length of the tail, bacteriophages employ a protein called the tape measure protein. The tail is constructed around the tape measure protein until it is as long as the protein itself. In BcepNY3 the tape measure protein is BcepNY3 gp45. Tail

assembly is dependent upon a frameshift protein in lambda. An unknown mechanism causes the frameshift to occur on a slippery sequence (18). These frameshift mutations are highly conserved among tailed phages (28). Summer *et al*. found the sequence GGCAAC in Bcep781 gp45, which is a slippery sequence that has a -1 frameshift (26). The GGCAAC sequence found in *Yersinia* lambda, a phage of *Yersinia pestis*, originally reads G-A-K, but then shifts back one base pair and reads G-K-V (28). However, the slippery sequence is absent in the Bcep1 homologue, Bcep1 gp46 (26). BcepNY3 gp45 is 99% identical to Bcep1 gp46. After performing a search of the BcepNY3 gp45 gene, the *Yersinia* lambda slippery sequence was absent with the same single base pair change GGCGAAC as in the Bcep1 gp46. Therefore, if BcepNY3 gp45 and Bcep1 gp46 are the frameshift proteins they must have a unique mechanism compared to the rest of the Bcep781 family and should be similar to each other. A gene in this tail and baseplate cluster is BcepNY3 gp48. BcepNY3 gp48 is 100% identical to Bcep1 and has no homologue in Bcep781 or Bcep43. Because of its placement and it being solely present *Burkholderia cenocepacia* phages and not in *Burkholderia cepacia* phages, it may be a host specific protein.

The terminase protein is used in regulating the packaging of DNA into the capsid. The large subunit (TerL) has mechanical activities which include cutting and an ATPase site for energy usage. The small subunit (TerS) is the *cos* recognition domain of the protein. BcepNY3 gp18 is 99% homologous to Bcep1 gp19, which Summer *et al.* predicted as TerL. Just as the TerS protein could not be accurately predicted in the other Bcep781

family members, the BcepNY3 genome and the NCBI database currently do not have enough information to confidently annotate TerS in BcepNY3 (26).

The process to develop a mature capsid is a complicated process that requires several proteins. These proteins have been predicted in BcepNY3. BcepNY3 gp11 and BcepNY3 gp12 are nearly 100% identical to the major and minor structural head proteins of Bcep1, respectively (26). The minor capsid protein here is the decoration protein, which is added to the exterior capsid to stabilize it. Bcep1's major capsid protein was determined by using PredictProtein software to justify that its secondary structure is similar to lambda E, the capsid protein of lambda (26). Prohead proteases are used to convert a precursor (prohead) to a mature phage head. BcepNY3 gp13 is 100% identical to Bcep1 gp15, a prohead protease (26). Bcep1 gp15 was one of nearly 200 head maturation prohead proteases determined by Cheng *et al.* (6). The morphogenesis genes are highly identical to the Bcep781-like phage genes and also have the same linear order. Therefore, the Summer *et al*. paper demonstrates that Bcep781-like phage have the following gene order: terL, minor head protein(s), prohead protease, decorator protein, and major capsid protein, which is the same order found in BcepNY3.

Lysis

In *Caudovirales* host cell lysis requires, at a minimum, two proteins: an endolysin and a holin. Endolysins are small enzymes that degrade cell wall peptidoglycan at a specified time during the infection cycle. BcepNY3 gp26 encodes a primary sequence homologue

of Bcep781 gp27, which has previously been demonstrated to be an authentic endolysin (29). An expect value of 3E-133 and 97% identity with Bcep781 gp27 indicate that the lytic function of this sequence is likely to be conserved in BcepNY3 gp26.

Holins are hydrophobic integral membrane proteins that create an opening in the bacterial membrane, allowing endolysins to attack the peptidoglycan as described above. Holins are grouped into two classes based on their primary structure: Class I holins, which generally have more than 95 residues and form three transmembrane helices; and class II holins, which are smaller and form only two transmembrane helices (14). BcepNY3 gp28 shares significant structural similarities with Bcep1 gp31, a known holin. Additionally, sequencing of BcepNY3 and Bcep1 has shown a high degree of similarity in the locations of these genes with respect to the theoretical endolysin coding regions. Further TMHMM analysis of BcepNY3 gp28 suggest that it is a class I holin, with three transmembrane domains and a predicted N-out, C-in topology.

Rz and Rz1 facilitate in lysis by rupturing the outer membrane. The Rz protein, BcepNY3 gp26, was predicted by GeneMark and annotated by its homology to the Bcep781-like phages. Additionally, TMHMM predicted a transmembrane domain for BcepNY3 gp23, confirming the Rz annotation, as Rz's N-terminus is imbedded in the inner membrane. Rz1, BcepNY3 gp24 was not predicted by GeneMark but was manually predicted by the presence of a strong Shine-Dalgarno in the middle of the Rz gene, a conserved cysteine at the N-terminus, and by the SPII lipoprotein predicted by

LipoP. This cysteine is modified and results in the formation an Rz-1 lipoprotein embedded in the outer membrane.

BcepNY3's relationship to Bcep1

In the entire BcepNY3 genome, only 3 of the 70 predicted proteins do not have a Bcep1 homologue. Out of these 67 related genes, 30 are 100% identical and 32 are greater than 90% identical. Not only are these genes identical, but the genome organization is identical as well (FIG. 6).

BcepNY3's relationship to OP_2

BcepNY3 has 28 putative homologues with the *Xanthomonas oryzae* phage OP_2. All of the 28 homologues showed less than 60% identity; however, the gene order appears to be strongly conserved (FIG. 6). This strong homology is interesting for phage that have different geographical locations and different host specificity (*X. oryzae* is a *Gammaproteobacterium* and *B. cenocepacia* is a *Betaproteobacterium*).

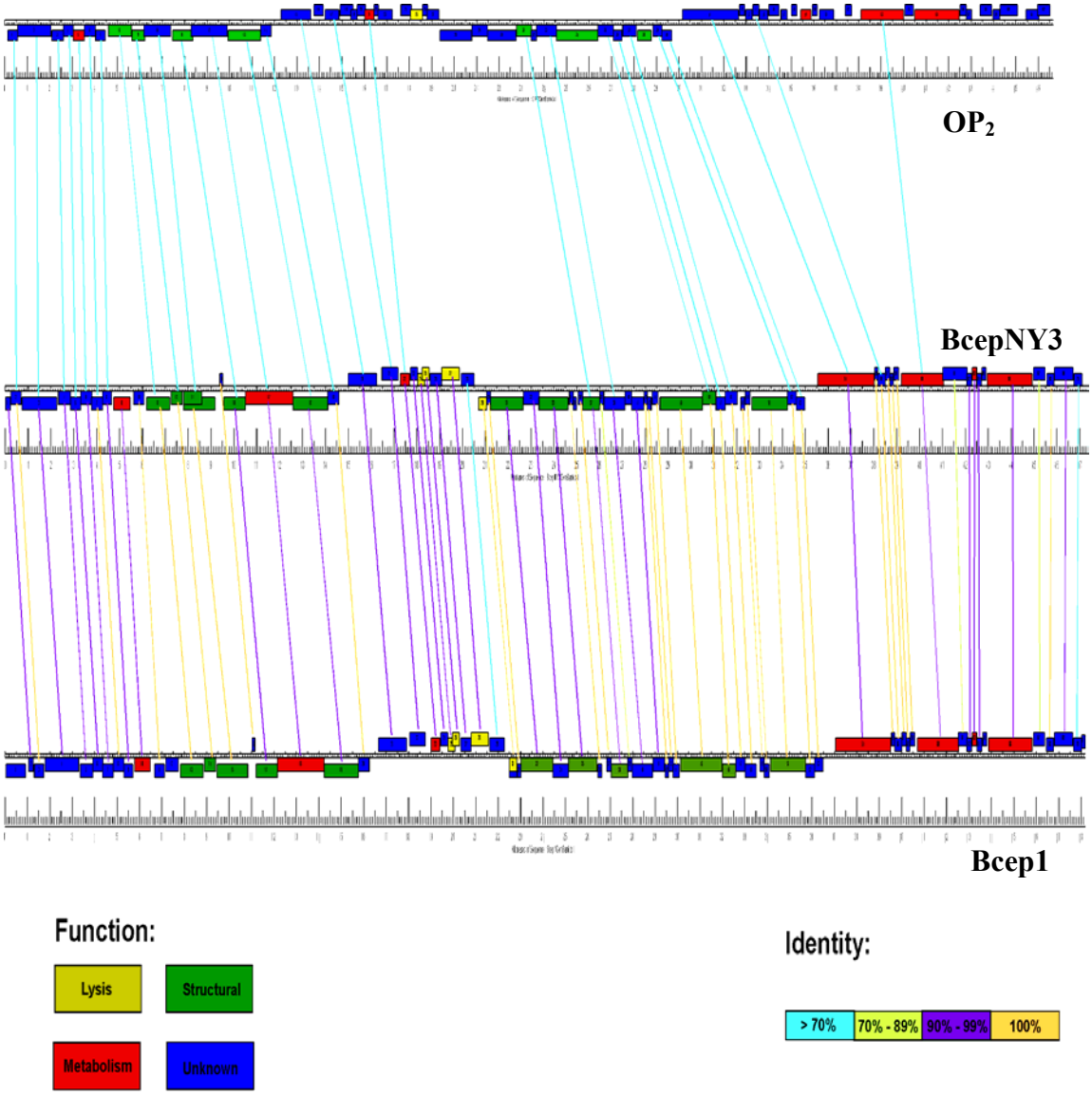

FIG 6. Comparison map of BcepNY3 to OP_2 and Bcep1

Three genomic maps are represented here. From top to bottom they are as follows: OP2 BcepNY3, and Bcep1. The map consists of a bar in the center with boxes drawn as a predicted ORF. ORFs above the bar are in the forward direction while ORFs below the bar are found on the reverse strand. The function of the genes are color coded based upon function as follows: yellow, lysis; green, structural; red, metabolism; blue, unknown. Lines are drawn from BcepNY3 to homologues in Bcep1 and OP2. The color of the line represents the percent identity of the homologue to BcepNY3 as follows: light blue, less than or equal to 70% identity; lime green, 70% to 89% identity; purple, 90%-99% identity; orange, 100% identical

CHAPTER IV

SUMMARY AND CONCLUSIONS

The *Rhodococcus equi* phage hunt yielded twelve novel phages isolated, making these the first reported *R. equi* phage isolated. Surprisingly, these phages all appear to have similar morphologies, despite the fact that viruses are quite numerous and diverse, especially in local environments (3). The dearth of diversity found in these bacteriophages raises several new questions. First, is the enrichment procedure leading to a bias in phage morphologies isolated? While this seems like a logical possibility, there have been other reports of using the same enrichment procedure with other bacteria that yielded other morphologies of phages (26). Another concern with the enrichments used is that soil from different locations was pooled together. Therefore, one location might have extremely virulent bacteriophages that are dominating the pool bacteriophages. Work has already begun on this topic by using enrichments with soil from one location and characterizing these bacteriophages. One other possibility is that these siphophages are the most fit virus for *Rhodococcus equi*. Investigating this notion requires a wider study of *R. equi* bacteriophages and their interactions with the bacterium itself that is currently unavailable because these 12 *R. equi* bacteriophages are the first reported. One important aspect to not overlook is that while these bacteriophages are all siphophages, their genomes have the potential to be quite diverse. Therefore, it is important to obtain genomic sequencing and annotation for these bacteriophages to help answer these questions.

Perhaps the most significant findings of the *R. equi* phage hunt are the broad host ranges of the soil enriched bacteriophages on clinical hosts. The genomic projects of these bacteriophages has already been initiated and appear to have great potential for therapeutic usage. Work is currently underway to use these bacteriophages and perhaps some of the clinically enriched ones to generate a cocktail of bacteriophages to be used as a prophylactic spray.

Horses are only susceptible to *R. equi* infections during their first 4 to 5 months and afterwards become resistant to infection due to the development of a stronger immune system. Due to the intracellular mechanism of pathogenesis and the short time frame of susceptibility, it would be more appropriate to use these phages in a prophylactic therapeutic spray, that could be opportunistically used, rather than direct treatment of foal infections. There is a lot of work needed in this area to determine the proper concentration of each specific phage in the cocktail and the frequency of sprays required to sustain lower levels of *R. equi* infections.

With the genomic project of BcepNY3, a *Burkholderia cenocepacia* bacteriophage, complete, it can be concluded that BcepNY3 is part of the previously described Bcep781 family of phages (26). One surprising discovery is that BcepNY3, and therefore the other Bcep781 phages, is related to the *Xanthomonas oryzae* phage OP_2. *X. oryzae* is the causative agent of leaf blight in rice (1). At first glance it would not be expected that

OP_2 and BcepNY3 have a strong homology because their hosts are a *Gammaproteobacterium* and a *Betaproteobacterium* respectively. Also, BcepNY3 was isolated in soils from New York and OP_2, while the exact location of isolation is unknown, it can be assumed that it was most likely isolated from soils in Asia due to *X. oryzae*'s prevelance there (17). However, upon further analysis there appears to be a potential link for horizontal gene transfer between the two phages. *Burkholderia cepacia*, which is thought to be found ubiquitously in soils around the world, is believed to enhance its virulence by exchanging genes with *Burkholderia psuedomallei* (16). *B. pseudomallei* is the causative agent of meleiodosis, a disease that plagues Southeast Asia. Therefore, there is a geographical link between the host bactieria. Also, another *Burkholderia* species, *Burkholderia platnarii*, is a pathogen of rice like *X. oryzae*, which leads to another possible avenue for gene transfer to occur (21). Therefore, there are both geographic and pathogenic relationships between the two host of BcepNY3 and OP_2, demonstrating a potential evolutionary relationship between these two bacteriophages and perhaps others of the same clade.

REFERENCES

1. **Adhikari, T. B., C. Cruz, Q. Zhang, R. J. Nelson, D. Z. Skinner, T. W. Mew, and J. E. Leach.** 1995. Genetic diversity of *Xanthomonas oryzae* pv. oryzae in Asia. Applied and Environmental Microbiology **61:**966-971.

2. **Besemer, J., and M. Borodovsky.** 1992. Heuristic approach to deriving models for gene finding. Nucleic Acids Res. **27:**3911-3920.

3. **Breitbarta, M., and F. Rohwer.** 2005. Here a virus, there a virus, everywhere the same virus? Trends in Microbiology **13:**278-284.

4. **Carles-Kinch, K., J. W. George, and K. N. Kreuzer.** 1997. Bacteriophage T4 UvsW protein is a helicase involved in recombination, repair and the regulation of DNA replication origins. Journal of EMBO **16:**4142-4151.

5. **Casjens, S. R.** 2005. Comparative genomics and evolution of the tailed-bacteriophages. Current Opinion in Microbiology **8:**451-458.

6. **Cheng, H., N. Shen, J. Pei, and N. V. Grishin.** 2004. Double stranded DNA bacteriophage prohead protease is homologous to herpesvirus protease. Protein Science **13:**2260-2269.

7. **Chibani-Chennoufi, S., A. Bruttin, M-L. Dillmann, and H. Bru¨ssow.** 2004. Phage-host interaction: an ecological perspective. Journal of Bacteriology **186:**3677-3686.

8. **Chu, G., and K. Gunderson.** 1991. Separation of large DNA by a variable-angle contour-clamped homogeneous electric field apparatus. Analytical Biochemistry **194:**439-446.

9. **Coenye, T., and P. Vandamme** 2003. Diversity and significance of *Burkholderia* species occupying diverse ecological niches Environmental Microbiology **5:**719-729.

10. **Coenye, T., P. Vandamme, J. R. W. Govan, and J. J. Lipuma.** 2001. Taxonomy and identification of the *Burkholderia cepacia* complex. Journal of Clinical Microbiology **39:**3427-3436.

11. **Decré, D., A. Buré, B. Pangon, A. Phillippon, and E. Bergogne-Bérézin.** 1991. In-vitro susceptibility of *Rhodococcus equi* to 27 antibiotics. The Journal of Antimicrobial Chemotherapy **28:**311-313.

12. **Gilad, J., I. Harary, T. Dushnitsky, D. Schwartz, and Y. Amsalem.** 2007. *Burkholderia mallei* and *Burkholderia pseudomallei* as bioterrorism agents: national aspects of emergency preparedness. The Israel Medical Association Journal **9:**499-503.

13. **Govan, J. R. W., and V. Deretic.** 1996. Microbial pathogenesis in cystic fibrosis: mucoid *Pseudomonas aeruginosa* and *Burkholderia cepacia*. Microbiological Reviews **60:**539-574.

14. **Grundling, A., U. Blasi, and R. Young.** 2000. Biochemical and genetic evidence for three transmembrane domains in the class I holin, lambda S. Journal of Biological Chemistry **275:**769-776.

15. **Hanlon, G. W.** 2007. Bacteriophages: an appraisal of their role in the treatment of bacterial infections. International Journal of Antimicrobial Agents **30:**118-128.

16. **Holmes A, J. Govan, and R. Goldstein.** 1998. Agricultural use of *Burkholderia* (*Pseudomonas*) *cepacia*: a threat to human health? Emerging Infectious Diseases **4:**221-227.

17. **Inoue, Y., T. Matsuura, T. Ohara, and K. Azegami.** 2006. Sequence analysis of the genome of OP_2, a lytic bacteriophage of *Xanthomonas oryzae* pv. oryzae. The Journal of General Plant Pathology **72:**104-110.

18. **Levin, M. E., R. W. Hendrix, and S. R. Casjens.** 1993. Programmed translational frameshift is required for the synthesis of a bacteriophage λ tail assembly protein. Journal of Molecular Biology **234:**124-139.

19. **McLennan, A. G.** 2006. The Nudix hydrolase superfamily. Cellular and Molecular Life Sciences **63:**123-143.

20. **Meijer W. M., and J. F. Prescott** 2004. *Rhodococcus equi*. Veterniary Research **35:**323-396.

21. **Mitchell, R. E., and K. L. Teh.** 2005. Antibacterial iminopyrrolidines from *Burkholderia plantarii*, a bacterial pathogen of rice. Organic and Biomolecular Chemistry **3:**3540-3543.

22. **Mosser, D. M., and M. K. Hondalus.** 1996. *Rhodococcus equi*: an emerging opportunistic pathogen. Trends in Microbiology **4:**29-33.

23. **Nordmann, P., M. H. Nicolas, and L. Gutmann.** 1993. Penicillin-binding proteins of *Rhodococcus equi*: potential role in resistance to imipenem. The Journal of Antimicrobial Chemotherapy **37:**1406-1409.

24. **Prescott, J. F., and C. R. Sweeney.** 1985. Treatment of *Corynebacterium equi* pneumonia of foals: a review. Journal of the American Veterinary Medical Association **187:**725-728.

25. **Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M. A. Rajandream, and B. Barrell.** 2000. Artemis: sequence visualization and annotation. Bioinformatics **16:**944-945.

26. **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. Journal of Bacteriology **188:**255-268.

27. **Takai S., K. Takeda, Y. Nakano, T. Karasawa, Y. Sasaki, J. Furugoori, S. Tsubaki, T. Higuchi, T. Anzai, R. Wada, and M. Kamada.** 1997. Emergence of rifampin-resistant *Rhodococcus equi* in an infected foal. Journal of Clinical Microbiology **35:**1904-1908.

28. **Xu, J., R. W. Hendrix, and R. L.Duda.** 2004. Conserved translational frameshift in dsDNA bacteriophage tail assembly genes. Molecular Cell **16:**11-21.

29. **Young, R.** 1992. Bacteriophage lysis: mechanism and regulation. Microbiology Reviews **56:**430-481.

CONTACT INFORMATION

Name:	Robert Charles Orchard II
Professional Address:	c/o Dr. Elizabeth Summer Department of Biochemistry and Biophysics Texas A&M University College Station, TX 77843
Email Address:	RcOrchard@gmail.com m
Education:	B.S. Microbiology, Texas A&M University, May 2008 Summa Cum Laude Undergraduate Research Scholar