Abstract

JAQUA, STEPHEN ALEXANDER. Characterization of the Bacteriophage HJ1031. (Under the direction of Professor Paul Hamilton).

This project was begun with the intent to isolate a prophage from a pathogenic Gram positive bacterial host. After isolating the host, it was our intent to analyze the bacteriophage genome by purifying a phage particle and cloning the target endolysin genes. In the process of searching for an appropriate bacteriophage, we isolated and characterized a bacteriophage that packaged segments of the host chromosome not related to bacteriophage function. This rendered our previous strategy inappropriate for analyzing this bacteriophage.

To resolve this, we turned to analyzing the structural proteins of our bacteriophage, which we named HJ1031. We created Transmission Electron Microscopy images to describe the morphology of the bacteriophage. We then created pure samples of intact bacteriophage particles through CsCl centrifugation, and separated the structural proteins by SDS-PAGE analysis. The polyacrylamide gel containing these phage proteins was turned over to the NCSU Mass Spectrometry Facility for MALDI-TOF MS/MS mass spectrometry, along with a proteome compiled for them from the UniProt Knowledge Base. After mass spectrometry analysis, likely identities were generated for three major structural proteins, and proportions for bacteriophage structural proteins were derived. These protein identities can serve as a basis for future research in identifying the genes for HJ1031.
Characterization of the Bacteriophage HJ1031

by

Stephen Alexander Jaqua

A thesis submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirement for the Degree of Master of Science

Microbiology

Raleigh, North Carolina

2014

APPROVED BY:

______________________________  ________________________________
Dr. Eric Miller                Dr. Michael Hyman

______________________________
Dr. Paul Hamilton, Committee Chair
DEDICATION

This humble work is dedicated to my Principle Investigator, mentor, and fellow chili gourmet, Paul Hamilton. Not only would this work not have been possible without you, I don’t know if I would have made it through graduate school under anyone else. Thank you for your patience and guidance.

This is also dedicated with gratitude to everyone in my department, and whom I have met while at NCSU. If we’ve spoken even once, I’ve learned something from you. Of particular note are Akimbolada Oyagunwa, Justin Bradshaw, Jason Whitham, Kimberly Rochester, John Hannah, and my roommate David Pilgreen.

Finally, this is dedicated to my large (and ever growing) family. I would never have made it to the start of this project without your love and support. I’ll be spending more time with you in the future.
BIOGRAPHY

Stephen A. Jaqua was born in 1986. At 10 years old, he moved with his family to western North Carolina. He graduated from Western Carolina University in 2009 with a B.S. in Biology and a B.A. in Philosophy. Mr. Jaqua enrolled in 2011 at North Carolina State University, and worked under Dr. Paul Hamilton in his lab while completing his Master’s Degree. He is lucky to have the friends and family that he does.
# TABLE OF CONTENTS

LIST OF TABLES .................................................................................................................. v
LIST OF FIGURES ................................................................................................................ vi

Characterization of the Bacteriophage HJ1031 ................................................................. 1

  Literature Review .............................................................................................................. 1
  References Cited for Literature Review ......................................................................... 11
  Materials and Methods ..................................................................................................... 20
  Isolation of Host Strain .................................................................................................... 20
  Mitomycin C Induction of Host Strain ............................................................................. 20
  16S rRNA Identification of Host 11-1 ........................................................................... 21
  Isolation of HJ1031 Particles .......................................................................................... 22
  Characterization of HJ1031 Proteins ............................................................................. 23
  Isolation of HJ1031 Derived DNA ................................................................................. 24
  Exonuclease and Endonuclease Digestion Analysis of HJ1031 Derived DNA .......... 25
  Endonuclease Digestion Analysis of Host Genomic DNA .............................................. 26
  Sonication Shearing of HJ1031 DNA .......................................................................... 27
  Mass Spectrometry Analysis of HJ1031 Proteins Using Matrix Assisted Laser
    Desorption/Ionization Time of Flight ............................................................................. 29
  Analysis of Mass Spectrometry and Identification of HJ1031
    Structural Protein Genes ............................................................................................ 31
  Electron Microscopy of HJ1031 ..................................................................................... 32
  Results .............................................................................................................................. 34
  Induction of Phage and Analysis of HJ1031 DNA ......................................................... 34
  Analysis of HJ1031 Proteins .......................................................................................... 48
  Discussion ......................................................................................................................... 55
  References Cited for Materials and Methods, and Discussion .................................... 65
LIST OF TABLES

Materials and Methods

Table 1. Sequencing of PCR Amplified DNA from Alkaline Lysis Miniprep of Litmus 38i-HJ1031 Derived DNA .................................................................35
Table 2. BLAST Analysis of Cloned Sequences .....................................................39
Table 3. 16S rRNA Identification of 11-1 .............................................................40
Table 4. Quantitation of HJ1031 Particle Size from TEM .................................42
LIST OF FIGURES

Materials and Methods

Figure 1. $\text{OD}_{600}$ Measurements from Mitomycin C Exposure ........................................30
Figure 2A. Enzymatic Qualification of HJ1031 Packaged Nucleic Acid .........................32
Figure 2B. Enzymatic Qualification for HJ1031 Packaged DNA

  Physical Structure ...........................................................................................................33

Figure 3. Endonuclease Digestion of Genomic $B. \text{pumilus}$ DNA

  and HJ1031 Derived DNA ................................................................................................35

Figure 5. Transmission Electron Microscopy of HJ1031 .................................................41
Figure 6. SDS-PAGE Analysis of HJ1031 Proteins Purified by

  CsCl Gradient Centrifugation and Dialysis ..................................................................43

Figure 7. MALDI-TOF Mass Spectrometry of

  HJ1031 Protein Fragments ............................................................................................45
Characterization of the Bacteriophage HJ1031

Literature Review

1.1 Introduction

The purpose of this literature review is to present background information on lysogenic bacteriophages, their role in bacterial genomic structure, and their presence in the *Bacillus* genus.

1.2 Overview of Lysogenic Bacteriophages

Bacteriophages are viruses that infect bacteria, in contrast with viruses that infect archaea or eukaryotic cells (1). The first major division for classifying bacteriophages, simply called “phages” for short, is whether they are filamentous or non-filamentous. Filamentous phages such as M13 have long filament-like bodies and are secreted from the host cell in a non-destructive manner. The other broad division is non-filamentous phage, which have a defined head and tail. They are not secreted by the host cell, but destroy the host on their egress (2). These were the phages whose effects were first observed by Frederick Twort and Felix d’Herelle (3). An “ultra-microscopic” agent was observed to be cultured between generations and create lytic plaques on bacterial lawns, “eating” the bacteria and earning them the title bacteriophages.

Beyond the major divisions of filamentous and non-filamentous, the propagation cycle of a bacteriophage can follow one of two broadly classified cycles. In a virulent or lytic cycle, immediately upon infecting a host cell the phage will begin reproducing infectious agents as quickly as possible. This leads to the lysis of the cell and the further infection of host cells by newly produced viruses. The alternative is a temperate or lysogenic cycle. In the lysogenic cycle the phage will infect a host and enter a period of
latency. During latency it will integrate into the host chromosome or maintain its genomic information as a stable plasmid, while not producing viral particles (4).

Lysogenic bacteriophages are among the most common biological agents on earth. They are found in the majority of species of bacteria and have increasingly become the target of biomedical research (5, 6) over the last three decades. Lysogenic bacteriophages are phages that integrate into the host chromosome and remain in a dormant state. They remain in this lysogenic state due to encoded repressor proteins, which prevent active transcription and synthesis of the phage until appropriate conditions arise (4, 5, 7, 8, 9, 10, 11). These conditions include stressors such as heat shock, UV radiation, or oxidative damage to host chromosomal DNA. The stressor may degrade the phage repressor protein, or the host response to chromosomal damage may degrade the repressor protein. Active transcription of phage genetic material begins after the degradation of the repressor protein as the bacteriophage enters the lytic cycle. A successful bacteriophage will assemble its own structural proteins, package its own infectious genetic material, and produce late stage proteins like holins and endolysins to help in its egress. Phages are closely associated with virulence factors in several major pathogenic bacteria, such as the *Salmonella enterica* Typhimurium serovar, *Streptococcus pyogenes*, and *Escherichia coli* O157:H7 (4, 12, 13, 14, 15, 16). The association arises from the proximity of known phage integrase sequences with virulence factor genes.

### 1.3 Phage Integration and Maintenance of Lysogeny

The repression of bacteriophages in a lysogenic state is an extensively covered subject of research. The Lambda phage is a model for lysogenic phages, which uses *E. coli* as its host strain. After initial infection of a host cell and successful expression of early transcripts, Lambda phage and lambdoid phages express late-early transcripts.
These transcripts are necessary for the successful integration of a phage into the host genome and the maintenance of a lysogenic state. Two important proteins are the Int and cl protein, which are used for integration at an attachment site and for repression of critical genes, respectively (5, 7, 17). Integration requires an attachment site on both the infecting phage (attP) and on the bacterial host genome (attB), which allows the recombination event that results in phage integration (18). This attachment site is a region of high homology between the host chromosome and the integrating DNA. Integration requires the host protein HIF (host integration factor) and Int, the Lambda integrase. These proteins stabilize the conformation of host DNA and cause a break at the attB site. The incoming Lambda DNA is then annealed to the complementary sites on the attB site, and ligated into place. The bacterial attachment site attB is separated as the phage integrates, resulting in a lysogenic phage genome flanked by what are now termed attL and attR sites on the left and right.

Super-infection and multiple lysogeny events in the same cell are determined by many factors. The first factor is if an infecting phage alters the cell expression of the surface receptor. Some phages have demonstrated an ability to repress surface receptor proteins, preventing super-infection (19). Another critical determinant is the presence of lysogenic phages already expressing repressor proteins. This mechanism was little understood at first (20, 21). Cells with integrated prophages showed decreased competence and lowered levels of DNA integration. This effect was also demonstrated to be additive; the more lysogenic phages present, the more that competence and integration were diminished. This is likely due to the constitutive expression by prophages of repressor proteins like cl binding to homologous sequences found on incoming DNA. The more integrated prophages present, the more repressor is likely to be present. If RecA proteins are not currently activated by single-stranded DNA, repressor binding to incoming DNA is less likely to allow integration. This is further supported by experiments in which cell lines have been cured of lysogenic prophage
and cryptic phage expression has been nullified, leading to no alteration of integration frequency or competence (22). If free cl expressed by the lysogenic phage does not bind to the incoming lambdoid super-infecting DNA, in rare cases the secondary phage DNA recombines with the already-present lysogen (23).

If the phage or phages already present in the cell do not prevent super-infection, additional phage may lysogenize or integrate multiple times in the same genome (24, 25). If super-infection occurs with two phages simultaneously it may result in tandem integration at the same site, while sequential super-infection can result in singular integration at multiple loci in the host genome (25). After integration into the host genome, cl binds as a repressor (10). This repressor prevents active transcription of phage genes and keeps the phage in a lysogenic state.

The mechanism for this repression is widely conserved in function. RecA protein acts as a major antirepressor of the SOS system genes by acting as an inducer in the autoproteolytic cleavage of LexA (26, 27, 28, 29, 30). This leads to an increase in SOS protein expression including RecA. RecA also induces the autolytic cleavage of cl, the repressor of Lambda phage transcription. When activated RecA is present at high levels, it induces the autoproteolysis of cl (10). cl acts as a repressor when present as a dimer, but monomers are incapable of repression. RecA induction of cl autoproteolysis ends cl repression of genes. Variability is demonstrated across different phages, particularly in the C-terminus region of the cl protein. Among certain classes of Lambda-like phages, the variable C-terminus region has been shown to destabilize the cl protein by making it a target for proteolysis. The absence of a C-terminus region or nonsense mutation in this region impairs the destabilizing effect, creating a gain of function for the bacteriophage which is now capable of repression and lysogeny (7). This would ultimately favor a mutant for long-term maintenance of the phage DNA in the host cell. Since lysogenic phage prevent super-infection by members of the same virus species
(31), the integration of a mutant C-terminus cl into the genome may increase host fitness.

A further example of C-terminus variability impacting the efficacy of cl repression is found in the lactococcal phage TP901-1. In this phage, the C-terminus of the cl demonstrates greater flexibility in its tertiary structure, allowing triplets of dimerized cl to form hexameric repressor compounds (9). The C-terminus region is again the target for inducing or inhibiting repression and lysogeny, as the expression of the modifier of repression (MOR) protein targets this region. After binding the C-terminus, MOR causes dissociation between the dimers, opening them to interactions with RecA.

Proteins expressed at low concentration by a phage in lysogeny do not always act exactly as the cl repressor does. The λ434 hex protein directly interferes with RecA cleavage of phage 434 and P22 repressor, but not the autolytic cleavage of LexA or cl repressor (11). This indicates a species-specific method of preventing super-infection and a RecA-independent path of regulatory control, but not a global prevention of RecA induced lytic activity.

Lambda and lambdoid phages are not the only phages capable of lysogeny. Lysogenic phage for the species Bacillus thuringiensis and Bacillus subtilis are the subjects of much research. Phages studied in B. subtilis share common characteristics including tail-sheath motif homology and repressor mechanisms, indicating a common distant heritage (30, 32, 33). Instead of relying on the cl protein, these phages rely on LexA and LexA-like proteins for repression (34). LexA binds to conserved sequences, like the upstream operator regions for din boxes (27), or a unique repeating sequence for repressor binding (35). The SOS response protein RecA will interact with LexA bound to DNA, and initiate autoproteolysis (27). This results in the lysogenic phage no longer being repressed and entering a lytic phase. LexA repression may be imperfect, given that RecA may not be the only factor capable of initiating autoproteolysis. The
variability between LexA-like proteins also means different levels of activity and reactivity (32), and the homology of the operator to a consensus sequence may effect LexA binding. This may account for the low levels of lysogenic bacteria seen in Bacillus species even in the absence of an induced SOS response (32, 36). Even if LexA were to act only as a repressor and bind perfectly to operators, the LexA homologues with a helix-turn-helix (HTH) motif have displayed different kinds of activity, modulating phage gene expression by promoting or repressing transcription (27). The conservation or diversity of integral protein motifs for phages has become a tool for the study of horizontal gene transfer as well as phage and host phylogeny (37).

1.4 Phage in Bacterial Genomic Structure

The association with virulence factors like bacterial toxins, antibiotic resistance, and effector proteins, is an indicator of the critical role that phages may play in bacterial populations and evolution. Specialized transduction involves the creation of a functional virion particle that is packaged with replicative phage DNA and additional primary host DNA, which is successfully recombined into a second host. General transduction is the packaging of only initial host DNA and successful recombination into a secondary host (38). The degree of success and specificity in packaging only phage DNA into infectious particles may vary from species to species. As a general rule, the closer a host gene is to the site of phage integration, the more likely it is to be packaged in specialized transduction.

This is particularly true of 'additional' DNA. It is considered 'additional' because it is not strictly related the ordinary phage DNA that would be packaged. This DNA consists of pieces of non-phage DNA situated between genes in the prophage genome. Flanked by prophage genes, they are very likely to undergo specialized transduction (4, 12, 35, 38, 39, 40, 41). This tangential DNA can be thought of as tolerated mutations rather than beneficial or detrimental. They typically present no immediate benefit to
the phage that carries them, and may not provide any immediate benefit to the host infected by the phage. They are tolerated so long as they are not detrimental. If they provided no benefit to either the bacteriophage or the host species then eventually the gene would be extinguished. They are potential recombination sites inside the phage, acting as a carrier mechanism for recombinant DNA. This is in contrast to a mutation in phage function or structure called mosaicism. Mosaicism is the recombining of entire genes or proteins from more than one phage in order to form new gene and protein sequences that could not be produced in the absence of either phage. Mosaicism may increase the virulence of the phage or the potential host range for the bacteriophage by altering receptor specificity (41).

Horizontal gene transfer and recombination is the driving force behind mosaic phage genome structures. Phages demonstrate synteny in gene organization, with head genes 5' to genes for tail structure in Siphoviridae (42). Head genes include the major capsid subunit, portal proteins, and terminases. Tail proteins include the major tail subunit, minor tail proteins, and the tapemeasure protein. The tapemeasure protein is typically the largest gene in the genome, since as an individual unit its length must equal the cumulative length of the phage tail. This synteny is evidence of distant common heritage among phages (42, 43), and allows for recombination and integration events among incoming phage. This leads to phages with new structures capable of different host selection and altering the target range for the exchange of non-essential DNA (4, 12, 35, 38, 39, 40, 41).

Cryptic phages are phages that produce non-functional or incomplete virion particles. Their integration is associated with major pathogenicity islands (14) responsible for a range of virulence factors (4, 12, 16, 40). This would indicate the common exchange of genes through recombination with intervening DNA sequences, as the cryptic phage has degraded so much over continuous integration in the host chromosome that DNA not related to phage function interrupts phage genes with
greater frequency. It would also account for maintaining cryptic phages in the absence of other benefits. These intervening sequences recombine with lytic or functional phages that newly infect the host, exchange genetic material, and egress to new populations to infect. In this way, a cryptic phage acts as a transfer station. Intervening DNA that is not related to phage function can be distinguished from ancestral phage DNA by G+C content (4, 41) when it interrupts a region of well conserved phage genes. However, as the long-term integration of the cryptic phage in the host genome may lead to base substitutions bringing the G + C content in line with the host genomic G + C content, this method of distinguishing ‘true’ ancestral phage DNA from non-phage DNA becomes less reliable (5).

Understanding the role that phages play in recombination and shaping bacterial evolution has lead research into the role of gene transfer agents (GTA) (43, 44, 45). A defective phage is distinguished from a cryptic phage in that it maintains structurally functional virion particles capable of egress from the host cell, but is not capable of forming replicative infectious virion particles. These virions may be capable of adsorping to a secondary host cell, and may be capable of penetrating their carried DNA into the secondary host. The inserted DNA will not be capable of producing a second generation of infectious virion particles, but can result in the transfer of genes.

Currently, four well-defined GTAs are described in the literature, with closely related hosts displaying their own similar GTAs. The first GTA characterized was found in Rhodopseudomonas capsulata, termed RcGTA (11, 43, 45), and its closely related cousins Roseovarius nubinhibens and Reugeria mobilis have produced GTAs upon induction. The GTA VSH-1 is found in Brachyspira hyodysenteriae, a spirochaete, and was shown in laboratory experiments to be capable of general transduction while not producing infectious siphovirus particles (44, 45, 46). The delta-proteobacteria Desulfovibrio desulfcans produces a non-replicative podovirus phage particle, Dd1, and demonstrates a gene cluster with specific conformation for podoviridae (43). The
commonality of GTAs between the Rhodobacterales in the alpha-proteobacteria class suggests that the ancestor of these GTAs was associated with the distant common progenitor of these hosts. This by no means indicates that all GTAs are old enough to have a single common ancestor. The first indication is that they are governed by different regulatory mechanisms. RcGTA is under regulation by the genes cckA and ctrA, and is not induced by DNA damage from mitomycin C (45), whereas VSH-1 is induced by mitomycin C damage. While it is possible that the RcGTA and Dd1 are related based on the distant relation of their alpha-proteobacteria and delta-proteobacteria hosts, the fact that RcGTA is a siphovirus-like particle and Dd1 is a podovirus-like particle makes it highly unlikely that they have a common ancestor. GTAs have also been demonstrated in the archaeon Methanococcus voltae, producing a tailed phage. This is peculiar, since tailed phage are not common among archae (43). All four of these GTAs have the ability to transduce genes from an initial host to a target organism.

Two other significant defective phages are the GTA-like elements BLP and PBSX. These two are termed GTA-like elements because there is currently no evidence that they can play a direct role in transduction, but do demonstrate random host DNA packaging of fragments smaller than would be necessary to produce a replicative particle. BLP (bacteriophage-like particle) was discovered by the observation of a repeating 14 kb extrachromosomal DNA strand resistant to initial DNase I degradation (47, 48). The other significant GTA-like element is PBSX (Phage Bacillus subtilis, X), a defective phage found in Bacillus subtilis (49), which will be discussed further in Section 1.5. Plaque formation has not been demonstrated for either of these defective phages.

In summary, prophages may be maintained by the host over successive generations to increase or sustain gene pool diversity for the host. The prophage may be maintained in order to recombine with lytic phages and modify the structure of the lytic phage virion particle, changing the host range. The prophage may also be maintained for the presence of intervening DNA sequences not related to phage
function, leading to the recombination of non-phage genetic material that benefits the host population. Defective phages, including GTAs, may play a direct role in transduction but not produce replicative phage particles.

1.5 Defective Phage in Bacillus

Several important bacteriophages have been discovered in the Bacillus genus. However, not every phage is capable of successful egress, adsorption, infection, and integration into a new host. These are defective phages. In 1968, Okamoto, et. al described a defective phage isolated from Bacillus subtilis, called PBSX. (49). This phage regularly packaged random portions of host genomic DNA into PBSX particles after induction by mitomycin C. Initial experiments showed that PBSX regularly packaged double stranded linear bands of 13 kb DNA fragments from the host genome in a head-packing manner with non-staggered cuts (50), and produced non-infectious particles that could not penetrate the DNA successfully into a new host genome (51).

This DNA was demonstrated to be the original host genomic DNA (52), and did not require new synthesis of DNA when PBSX entered the lytic phase. It packaged DNA at random and did not favor any particular region of the chromosome (52, 53, 54). The presence of this defective phage demonstrates an interesting stage in the evolutionary development of a phage, between a lytic parasite capable of successfully producing infectious replicative particles and cryptic phage so degraded that they may have actually lost some recombination specificity (55, 56). The defective PBSX-like phages are widespread and conserved among different Bacillus species and close relatives (49, 52, 57, 58, 59, 60) and so appear to serve some evolutionary purpose. This work describes the isolation and characterization of a PBSX-like bacteriophage from a Bacillus pumilus strain isolated from the environment.
References


60. Tsutsumi Y, Hirokawa H, Shishido K. 1990. A new defective phage containing a randomly selected 8 kilobase-pairs fragment of host chromosomal DNA inducible in a strain of *Bacillus natto*. *FEMS Microbiology Letters.* **72** 41-46
Materials and Methods

2.1 Isolation of Host Strain

Samples were collected from the teat of a cow suffering from mastitis at the North Carolina State University Dairy on October 31st, 2011, by wiping sterile cotton swabs on the animal. Cotton swabs were placed in sealed 50 mL conical tubes containing 5 mL lysogeny broth (1% NaCl, 1% tryptone, 0.5% yeast extract, pH 7.0), and stored at room temperature. Samples were returned to the lab, and streaked out on lysogeny agar (lysogeny broth + 2% agar) plates. Samples were incubated overnight at 37°C. Multiple microbiological organisms were present in each plate. Isolated colonies were obtained by further streaking on new LB agar plates incubated at 37°C overnight. Potential hosts were then screened by Gram staining. After heat fixing on a slide, potential hosts were saturated with crystal violet for 1 min, rinsed with water, saturated with Gram’s Iodine for 1 min, rinsed with water, de-colorized with acetone: alcohol (50:50) for 10 sec, rinsed with water, saturated with safranin for 1 min, rinsed with water, and dried. Gram positive organisms were then tested for mitomycin C induction of prophages.

2.2 Mitomycin C Induction of Host Lysis

To screen for prophage activation, candidate phage hosts were exposed to mitomycin C during early log phase to induce phage transcription and production. 3 mL lysogeny broth was inoculated with isolated colonies of host bacteria, incubated at 37°C 250 RPM. Optical density was measured in triplicate every hour in a BIO-RAD SmartSpec 3000 spectrophotometer. 15 μL of 0.25 mg/mL mitomycin C was added to the culture to a final concentration of 0.8 μg mitomycin C/mL 2 hr after inoculation, when early log phase (OD₆₀₀=0.2-0.5) had been achieved. Measurements were taken until 8 hr after mitomycin C induction.
A potential phage-containing strain demonstrated repeatable and significant decrease in growth as measured by optical density when comparing cultures grown under optimal conditions and those grown under identical conditions with mitomycin C exposure. This isolate was designated 11-1.

2.3 16S rRNA Identification of Host 11-1

In order to identify 11-1, the DNA sequence of the 16S ribosomal RNA gene was determined. To isolate genomic DNA, a 5 mL culture of 11-1 was grown to stationary phase. The culture was centrifuged at 2,500X g 4°C for 25 min in order to pellet the cells. Supernatant was discarded, and the pellet was suspended in 560 μL of TE Buffer. 5μL 10 mg/mL RNase A and 5 μL 10 mg/mL Lysozyme were added, and the solution was incubated at 37°C for 1 hr, followed by heat inactivation at 65°C for 25 min. Solution was then cooled to room temperature. 30 μL 10% SDS and 5μL 10 mg/mL RNase A were added to the solution and incubated at 37°C for 1 hr, followed by heat inactivation at 65°C for 25 min.

The DNA was extracted with an equal volume of phenol: chloroform: Isoamyl alcohol (25:24:1). The mixture was centrifuged at 14,000X g for 2 min at room temperature. The aqueous layer was removed and this process was repeated two more times, until the layer of denatured protein no longer emerged at the interface. The aqueous layer was precipitated by mixing 1/10 volume of 3 M sodium acetate (NaOAc) and 2 volumes of 95% ethanol, mixing thoroughly, and incubating at -20°C for 2 hr. This mixture was centrifuged at 14,000X g 2 min at room temperature. A small white pellet emerged, and supernatant was removed by aspiration. The tube was turned on its side to allow addition ethanol to evaporate. The pellet was then rinsed with 1 mL 70% ethanol, centrifuged for another 2 min at 14,000X g at room temperature, and the supernatant was removed by aspiration. The tube was again laid on its side and the ethanol allowed to evaporate. When all ethanol had
evaporated, the pellet was dissolved in 50 μL TE Buffer and quantified using the Thermo Scientific Nanodrop 1000 Spectrophotometer.

Purified 11-1 genomic DNA served as the template material for a PCR reaction to amplify the 16S RNA gene. The PCR mixture was 2 μL 11-1 DNA, 1 μl 100 nanomolar fD1 primer (5’CCGAATTTCGTCGACAGAGTTTGA3’)(1), 1 μL nanomolarD1 primer (5’CCCGGGATCCAAGCTTAAGGAGGTG3’), 21 μL dH2O, and 25 μL Econ Taq Plus 2X Master Mix (Lucigen) were mixed on ice. This mixture was then run on a standard PCR protocol. PCR conditions were an initial denaturing step at 95°C for 5 min; then at 95°C for 1 min, 60°C for 1 min, and 72°C for 4 min, cycled 30 times; a final extension step at 72°C for 5 min; and storage at 4°C.

PCR mixture results were purified using the QIAquick PCR purification Kit (QIAGen, catalogue number 28106). Samples were suspended in 50 μL TE buffer and quantified using the Thermo Scientific NANODROP 1000 Spectrophotometer. PCR product was then sent to Eton Bioscience for sequencing using primers rD1 and fD1.

2.4 Isolation of HJ1031 Particles

To produce phage on an observable scale, 50 μL of an overnight culture of 11-1 was inoculated in 50 mL lysogeny broth and incubated at 37°C 250 RPM until early log phase was reached (OD=0.2-0.5). Mitomycin C was added to a final concentration of 0.5 μg mitomycin C/mL and the culture was incubated at 37°C 250 RPM for 2 hr. Post-induction, cultures were poured into BD Falcon 50 ml polypropylene conical tubes and centrifuged in a Sorvall Heraeus 70056445 Swinging Bucket Rotor at 3,500 RPM (2,500X g) 4°C for 50 min. The supernatant was poured into new 50 mL conical tubes and mixed 1:1 with 30% Polyethylene glycol (PEG) 1.5 M NaCl. The mixture was incubated overnight on ice. The mixture was centrifuged at 3,500 RPM (2,500X g) 4°C 50 min. A yellow pellet formed at the
bottom point of the cone, with a light film forming over the rest of the cone. The supernatant was gently poured into a waste container. The pellet was centrifuged again at 3,500 RPM 4°C 5 min. Any remaining supernatant was aspirated out. The pellet was suspended in 10⁻¹ original volume of mycobacteria phage buffer (MPB, 0.01 M Tris, 0.01 M MgSO₄, 0.068 M NaCl). Initial samples were mixed 1:1 with glycerol and stored at -80°C for long-term storage. Samples to be used within 7 days of induction were not mixed with glycerol, and were stored at 4°C.

2.5 Characterization of HJ1031 Protein

After initial isolation, HJ1031 particles were purified by CsCl step-gradient centrifugation. CsCl gradients were prepared by dissolving CsCl in MPB. CsCl solutions of densities ranging from 1.2-1.7 g/mL in 0.1 g/mL intervals were created. Layers of 1 mL CsCl solution were placed in Thermo Scientific UltraCrimp tubes with the 1.7 g/mL layer at the base and the 1.2 g/mL layer at the top. A further 2 mL of 1.2 g/mL were placed above this. A solution of HJ1031 particles was then added to the tube. The remainder of the tube was filled with light mineral oil until it reach capacity, and was then crimped shut. Centrifugation was performed in a Sorvall T 1270 rotor using a Beckman L-8 55 centrifuge. Samples were centrifuged at 36,200 RPM (120,000X g) 4°C for 3 hr. An opalescent blue band emerged between 1.3 g/mL and 1.4 g/mL. The band was removed using an 18-gauge needle and syringe. The sample was dialyzed overnight at 4°C in 1000X volume phosphate buffered saline (PBS, 137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄) using a Thermo Scientific Slide-A-Lyzer Dialysis Cassette (Thermo Scientific, #66380). A 20 μL sample was mixed with an equal volume of Laemmli sample buffer (62.5 mM Tris-HCl pH 6.8, 25% glycerol, 2% SDS, 0.01% Bromophenol blue) and denatured at 95°C for 3 min. 20 μL of sample was then run on a 12% acrylamide Mini-PROTEAN TGX Precast Gel (Biorad) at 120 V for 60 min. The gel was then stained using Coomassie
Brilliant Blue (29.9% methanol, 20% glacial acetic acid, 0.1% Coomassie Brilliant Blue R 250, 50% dH2O) on a rocker at 25 RPM for 2 hr, followed by destaining using a 30% methanol and 10% acetic acid solution on a rocker at 25 RPM for 30 min.

2.6 Isolation of HJ1031 Derived DNA

After isolation of HJ1031 particles as described in section 2.4, DNase I (final concentration 0.5μg/mL) and RNase A (final concentration 0.5 μg/mL) were added, then incubated at 37°C for 1 hr. Enzymes were heat inactivated at 65°C for 25 min. Samples were centrifuged in a Sorvall Legend Microcentrifuge 21 at 7,200 RPM (5,000X g) 10 min. A pellet formed comprised of cells, denatured proteins, and cellular components, and the supernatant containing phage HJ1031 was transferred to a new 1.5 mL microcentrifuge tube. This supernatant was centrifuged further at 7,200 RPM (5,000X g) 10 min. A smaller pellet formed, made of anything not pelleted by the first round of centrifugation. Samples were then transferred to a new 1.5 mL microcentrifuge tube.

10 μL 10% SDS and 10 μL 0.5 M EDTA were added to 500 μL samples of HJ1031 particles. Samples were incubated at 65°C for 60 min. Samples were then mixed 1:1 with phenol: chloroform: Isoamyl alcohol (25:24:1), pipetted to mix thoroughly, then centrifuged in a Sorvall Legend Microcentrifuge 21 at 21,000X g (14,500 RPM) 25°C for 2 min. A layer of denatured protein formed between the organic and aqueous layers. The aqueous layer was removed and phenol chloroform extraction was repeated three times or until the denatured protein layer did not emerge at the interface.

Aqueous layers was finally moved to an eppendorf tube by itself for ethanol precipitation. Samples were mixed with 1/10 volumes of 3 M sodium acetate (NaOAc), then mixed with 2 volumes of 95% ethanol (500 μL aqueous layer, 50 μL 3 M NaOAc, 1000 μL 95% ethanol), stored at -20°C overnight. Samples were
centrifuged in a Sorvall Legend Microcentrifuge 21 at 21,000X g 25°C for 2 min. Pellet formation occurred and supernatant was aspirated out, and eppendorf microcentrifuge tubes were laid on their side to air dry until all ethanol had completely evaporated. An additional 1 mL of 70% ethanol was added to the tube and incubated at room temperature for 5 min, followed by centrifugation at 21,000X g 25°C for 2 min. Ethanol was aspirated out and tubes were laid open on their sides until all ethanol had evaporated. Pellets were then suspended in 50 μL Tris-EDTA buffer (TE buffer, 100 mM Tris-Cl, 10 mM EDTA, pH 7.0).

Samples were quantified for purity and concentration using a Thermo Scientific NANODROP 1000 Spectrophotometer.

2.7 Exonuclease and Endonuclease Digestion Analysis of HJ1031 Derived DNA

After purification and quantification of HJ1031 derived DNA as described in section 2.6, HJ1031 derived nucleic acids were incubated in separate digestions reactions with RNase A and DNase I at a final concentration of 0.1 mg enzyme/mL. Digestions were incubated at 37°C for 1 hr, followed by heat inactivation at 65°C for 25 min. Samples were separated by gel electrophoresis at 95 V for 1 hr on 1% or 2% agarose gels in Tris-acetate EDTA buffer (TAE buffer, 40 mM Tris-acetate, 1 mM EDTA), followed by staining using ethidium bromide (EtBr, 0.1 mg/ml). These were visualized using a Gel Doc XR+ System.

Samples of Litmus38i plasmid, HJ1031 packaged DNA and HindIII-digested λ DNA were quantified to produce bands of similar intensity on a 1% TAE agarose gel. This required 500 ng of HindIII-digested λ DNA, 400 ng of Litmus38i DNA, and 500 ng of HJ1031 packaged DNA. Samples were each then placed in incubation solutions and adjusted to a volume of 10 μL. Samples were incubated at 37°C for 1 hr with no enzyme, with RecJf, with Lambda Exonuclease, or with Exonuclease III. Samples incubate with Lambda Exonuclease or Exonuclease III were incubated with the
buffer provided by NEB. Samples incubated without an exonuclease or with RecJf were incubated with NEB Buffer 2. After 1 hr, the samples were heat inactivated at 65°C for 20 min, then separated by gel electrophoresis on a 1% TAE agarose gel.

DNA derived from HJ1031 as described in section 2.5 was used for substrate in endonuclease digestions. Digestions were performed using 15 μL HJ1031 derived DNA (0.5-1 μg/mL), 2 μL NEB 10X Buffer, 2 μL dH₂O, and 1 μL of endonuclease. Digests were incubated for 1 hr. Digests with BamHI, BfuCI, BglII, BspHI, EcoRI, and Sau3AI were incubated at 37°C. Digests with BstUI were incubated at 60°C. Digests were then loaded onto 1% or 2% agarose gels and were separated by gel electrophoresis at 95 V for 1 hr in TAE buffer. Gels were stained using EtBr, and visualized on a Gel Doc XR+ System.

2.8 Endonuclease Digestion Analysis of Host Genomic DNA

50 μL of an overnight culture of 11-1 was inoculated in 5 mL LB broth and incubated at 37°C 250 RPM until optical density reached 0.3. At this point, 500 μL of cell culture was aliquoted into an eppendorf tube. 10 μL 10mg/mL Lysozyme was added to the aliquot and incubated at 37 °C for 1 hr. Lysozyme was then inactivated at 65°C for 25 min. 10 μL 10μg/mL RNase A was added to solution and incubated at 37°C for 1 hr, followed by heat inactivation at 65°C for 25 min. 10 μL 0.5 M EDTA was added along with 20 μL 10% SDS. The solution was then incubated at 65°C for 1 hr. After 1 hr, the aliquot was cooled to room temperature. 250 μL of phenol: chloroform: Isoamyl alcohol (25:24:1) was added and gently mixed. The solution was centrifuged in the Sorvall RT Legend Microcentrifuge 21 at 21,000X g (14,500 RPM) 2 min at room temperature. The aqueous layer was moved to a new tube, and phenol: chloroform: Isoamyl alcohol extraction was performed two more times or until denatured protein bands disappeared.
After phenol: chloroform: Isoamyl alcohol extraction, ethanol precipitation of the aqueous layer was performed as described in section 2.5. The DNA pellet was suspended in 50 μL TE Buffer. Quantification and purity were assessed using the Thermo Scientific NANODROP 1000 Spectrophotometer as described in section 2.5.

Samples of 11-1 host genomic DNA were incubated with endonucleases as described in section 2.6, using identical conditions. Digestion results were separated by gel electrophoresis on 1% or 2% agarose gels at 95 V for 1 hr in TAE buffer. Gels were stained using EtBr, followed by visualization using a Gel Doc XR+ System.

2.9 Sonication Shearing of HJ1031 Derived DNA

Samples of HJ1031 derived DNA were purified as described in section 2.6, to a concentration of 0.5-1.0 μg/ml. Sonication was performed with a Heat Systems Ultrasonics Inc. W-370 Sonicator. 20 μL samples were sonicated in an ice water bath for 10 sec bursts, ranging from a total of 10 sec to 120 sec. Samples were separated by gel electrophoresis on a 1% agarose gel at 95 V for 1 hr in TAE buffer. The gel was then stained using EtBr followed by visualization using a Gel Doc XR+ System.

After confirming DNA shearing by sonication, HJ1031 derived DNA samples were sonicated for two 10 sec bursts. Our sheared HJ1031 derived DNA fragments were blunted using the NEB Quick Blunting Kit, using 8 μL of sonicated HJ1031 DNA and adjusted to a reaction volume of 10 μL for the total solution. The reaction was incubated at room temperature for 30 min, followed by heat inactivation at 65°C for 25 min.

Litmus 38i plasmid DNA was digested with StuI. The digestion reaction was 4 μg Litmus 38i, 1 μL 10X NEB Buffer 4, 10 units StuI (1 μL at 10,000 units/mL) in a total reaction volume of 10 μL. Digestion was incubated at 37°C for 1 hr, followed by heat inactivation at 65°C for 20 min.
At this point, the 10 µL volume of StuI digest Litmus 38i was added to a mixture with the 10 µL volume sonication-sheared, blunted HJ1031 derived DNA fragments, and they were ligated using T4 Ligase. The reaction mixture was 10 µL HJ1031 solution, 10 µL Litmus 38i solution, 2.5 µL 10X T4 Ligase Buffer, 400 units T4 DNA Ligase (1 µL at 400,000 units/mL), and 1.5 µL dH₂O, to a total volume of 25 µL. The ligase reaction was incubated overnight at 16°C. The ligation was heat inactivated at 65°C for 25 min. The ligation reaction was then digested using StuI at 37°C for 1.5 hr to eliminate recircularized vector DNA from the ligation, followed by heat inactivation at 65°C for 25 min. 2 µg of Litmus 38i was digested with StuI at 37°C for 1.5 hr, followed by heat inactivation at 65°C for 25 min. Samples of each digestion reaction were separated by gel electrophoresis on a 1% agarose gel at 95 V for 1 hr in TAE buffer. This was followed by staining using EtBr and visualization using a Gel Doc XR+ System.

The Litmus 38i – HJ1031 derived DNA ligation mixture was transformed using a high-efficiency protocol (NEB) and NEB 5-α’ F’Iα cells. 0.1 mL of cells was spread over plates containing ampicillin 50 µg/mL, X-GAL 40µg/mL, IPTG 0.1mM. Plates were incubated overnight at 37°C.

Colonies that showed positive phenotype for Litmus 38i insertion (white colonies) were inoculated in 2 mL LB media 50 µg/mL ampicillin and incubated overnight at 37°C 250 RPM. These colonies were used in colony PCR using Litmus38iFWD (5’GATTAGCCTAGGTGCTTAAGCGATCGA3’) and Litmus38iREV (5’GATTAGCGATTAAGTTGGTACGCA 3’) primers. Colonies incubated overnight were also used for alkaline lysis minipreparation (2) and suspended in 50 µL TE Buffer. DNA purity and concentration were quantified using a Thermo Scientific NANODROP 1000 Spectrophotometer as described in section 1.5.

After quantification, plasmids isolated from alkaline lysis minipreparation were incubated with StuI at 37°C for 1 hr to confirm cloning of HJ1031 by the absence of a
viable *Stul* digestion site, followed by heat inactivation at 65°C for 25 min. A sample of pure Litmus 38i was also digested with *Stul* at 37°C for 1 hr, followed by heat inactivation at 65°C for 25 min. Samples were loaded onto 1% or 2% agarose gels and separated by gel electrophoresis at 95 V for 1 hr in TAE buffer. Gels were then stained using EtBr followed by visualization using a Gel Doc XR+ System. Samples from minipreparation were incubated in a double digest reaction to further confirm the insertion of HJ1031 DNA, using *EcoRI* and *Hpal*, which flank the *Stul* site on plasmid Litmus 38i. The digestion was incubated at 37°C for 1 hr followed by heat inactivation at 65°C for 25 min. Results were then loaded onto 1% or 2% agarose gels and run at 95 V for 1 hr in TAE buffer. Gels were then stained using EtBr followed by visualization using a Gel Doc XR+ System.

After endonuclease digestion mapping, plasmids isolated from alkaline lysis minipreparation were placed in 10 μL aliquots and sequenced by Eton Bioscience using the primers Litmus38iFWD and Litmus38iREV. Sequences were in an Applied Bioinformatics (ABI) file format. Nucleotide sequences were analyzed using the National Center for Biotechnology Information Basic Local Alignment Search Tool (BLAST), using the Nucleotide collection (nr/nt) searching for highly similar sequences (megablast).

### 2.10 Mass Spectrometry Analysis of HJ1031 Proteins Using Matrix Assisted Laser Desorption/Ionization Time of Flight

After purifying HJ1031 particles using CsCl centrifugation, samples were analyzed by SDS-PAGE as described in section 2.5. Three significant bands emerged after staining with Coomassie Brilliant Blue. The sample was then given to the NC State Mass Spectrometry Facility for mass spectrometry analysis.

The bands were excised using a sterile razor blade (collecting 1-2 mm of gel) and collected in polypropylene tubes. 100 μL of 50:50 100 mM NH₄HCO₃ : ACN
(acetonitrile) solution was added and incubated at room temperature for 30 min. The destaining solution was aspirated out and the bands incubated for a further 20 min at room temperature to dehydrate, until the gel bands turned white. 50 μL of 10 mM DTT (DL-Dithiothreitol) was mixed with each gel band and incubated at 56°C for 30 min. Samples were cooled to room temperature and DTT was aspirated out, to be replaced by 500 μL ACN and incubated at room temperature for 20 min. ACN was removed and replaced by 50 μL of 55 mM IAA (iodoacetamide) solution, then incubated at room temperature in the dark for 20 min. IAA was removed from the sample and the sample was dried using a Speed-Vac for 10 min.

Samples were then digested with 100 μL of 20 ng/μL trypsin solution incubated at 37°C overnight. 100 μL of 5% formic acid (v/v) :ACN (1:2) was added to each digest and incubated at 37°C for 15 min. Digested peptides were transferred to new tubes, with 50 μL ACN added and incubated at room temperature for 10 min. The peptide sample was then extracted and moved to the corresponding sample vial. Samples were rehydrated with 50 μL of 100 mM NH₄CO₃ at room temperature for 10 min, then peptide solution was extracted to the corresponding sample vial. 50 μL of ACN were added and incubated at room temperature for 10 min to dehydrate the sample, then moved to the corresponding sample vial. Peptide solutions were then dried using a Speed-Vac for 10 min and stored at -20°C until MS analysis.

Liquid chromatography separation and mass spectrometry detection were performed using an Eksigent (Dublin, CA) nano-LC-1D+ system and Nanoflex cHiPLC with an autosampler (Dublin, CA) coupled to a hybrid LTQ Orbitrap XL mass spectrometer from Thermo Scientific, Inc. (San Jose, CA). The nanoLC was operated in reversed-phase mode. The trap column employed was a NanoFlex cHiPLC column with a 200 μm i.d. and 2 cm of ChromeXP C18-CL stationary phase (3μm and 120 Å particle size). The analytical column employed was a NanoFlex cHiPLC column with 75 μm i.d. and 15 cm of the same stationary phase. mLCo solvents used were mobile
phase A \([\text{H}_2\text{O}/\text{CAN/HCOOH} \ (98/2/0.1\% \text{ by volume})]\) and mobile phase B \([\text{CAN/H}_2\text{O/HCOOH} \ (98/2/0.1\% \text{ by volume})]\). Blank runs (injection of mobile phase A) were performed after every sample run, so as to minimize carryover. Sample and blank injections were performed at 6 \(\mu\text{L}\) on the column. Each of the samples had 1 technical nanoLC/MS run. Analytical separations were performed on the nano-flow pump at 300 nL/min, initially maintaining 2\% B, with gradient elution of tryptic peptides.

The MS method consisted of nine events: a precursor scan followed by eight data dependent tandem MS scans of the 1\textsuperscript{st}-8\textsuperscript{th} most abundant peaks in the ion trap. A high resolving power precursor scan of the eluted peptides was obtained using the Orbitrap (60,000 resolution) with a pre-determined number of the most abundant ions selected for the MS/MS in the ion trap through dynamic exclusion. This method aimed at coverage of low and high abundance peptides/proteins. The instrument was externally tuned and calibrated according to the manufacturer’s protocol. Internal calibration was also achieved via lock mass.

Data files were processed by MASCOT Distiller (Matrix Science, INC). Both MS and MS/MS data were interrogated. MASCOT: Batch searching of nanoLC/MS/MS data was performed using a modified \textit{Bacillus pumilus} database. A Pearl script was used to create a reverse sequence database of the modified \textit{B. pumilus} database. The original proteomic database was the Uniprot proteomic files for \textit{Bacillus pumilus} SAFR-032 (Taxon identifier 315750) and \textit{Bacillus pumilus} ATCC 7061 (Taxon identifier 536229). Corresponding target and reverse sequences were combined into one FASTA file for MASCOT searching of nanoLC/MS/MS data to account for the false discovery rate (FDR). An FDR of <1\% is considered adequate and was used in determining reliable protein identifications. The following is a list of variable and fixed amino acid modifications allowed in all of the database searches. Variable modifications included M oxidation and N, Q deamidation. Fixed modifications
 included C carbamindomethylation. MASCOT search parameters were: Maximum missed cleavages 2, Peptide charge 1+, 2+, and 3+, Peptide tolerance ± 5 ppm, and MS/MS tolerance ± 0.6 Da.

2.11 Analysis of Mass Spectrometry and Identification of HJ1031 Structural Protein Genes

A report was compiled from the results of the mass spectrometry described in section 2.10. Samples were analyzed by the abundance of spectral lines produced, as indicative of the probability of protein identity when matched against the protein reference database. The protein reference database was the proteomic report compiled by Uniprot for the organisms Bacillus pumilus SAFR-032 (Taxon identifier 315750) and Bacillus pumilus ATCC 7061 (Taxon identifier 536229).

2.12 Electron Microscopy of HJ1031

HJ1031 was induced using mitomycin C and isolated as described in section 1.2. After induction, cultures were centrifuged at 2,500X g 25°C 50 min. At this point a sample of supernatant was collected through a Fisherbrand PVDF 0.45μM filter. A second sample of 1 mL was collected at this point after further centrifugation at 14,000X g 25°C 10 min. A small pellet formed. The supernatant was aspirated off, and the pellet was suspended in fresh LB media. A third sample of 1 mL lysate supernatant was collected at this point and centrifuged at 14,000X g 25°C 10 min, then was collected through a Fisherbrand PVDF 0.45 μM filter. The remaining lysate supernatant was then precipitated using PEG as described in section 2.3.

After the PEG precipitation, the supernatant was aspirated out completely, and the pellet was suspended in 10⁻¹ original volume LB media. A sample was collected through a Fisherbrand PVDF 0.45 μM filter. Another PEG precipitated sample was centrifuged at 14,000X g 25°C 10 min. A small pellet formed. The
supernatant was removed, the pellet was suspended in the prior volume of LB media, then collected through a Fisherbrand PVDF 0.45 µM filter.

All samples collected this way were stained on a carbon type B on 300 mesh support grid (PELCO, #01813). 10µL were placed on the grid, and left standing for 90 sec. Excess fluid was removed by the application of a paper wedge to the edge of the grid. The grid was washed using 10 µL dH₂O on top of the grid, left standing for 60 sec. Excess dH₂O was removed using paper. The grids were washed a second time with dH₂O, with excess dH₂O being removed by filter paper. The sample were then stained using 2% Uranyl Acetate (UA). 10 µL UA were applied to the sample and then immediately wicked away using filter paper. The samples air-dried at room temperature for 2 min. Samples were stored in a vacuum desiccator for 2 days. Samples were visualized on a JEOL 100S Transmission Electron Microscope.

Particle sizes were measured using the scale bar from the TEM image. Ten whole and intact phages were measured, in order to ensure that shearing forces were not responsible for a distortion of the phage or variance in particle size.
Results

3.1 Induction of Phage and Analysis of HJ1031 DNA

Initial culturing of isolates taken from the NCSU Dairy showed a diverse microbial community. Organisms were given a numerical identifier and isolated through two rounds of T-streaking. Isolates were then Gram stained. 11-1 proved to be a Gram-positive bacillary organism. As a Gram-positive isolate, 11-1 was then screened for hosting a phage by induction with mitomycin C.

Figure 1. OD$_{600}$ Measurement from Mitomycin C Exposure. Figure 1 demonstrates the OD$_{600}$ values averaged for five cultures of individually measured cultures of 11-1. Cultures not exposed to mitomycin C are indicated by 11-1 Avg. Cultures exposed to mitomycin C are listed are indicated by MC Avg.
At sufficiently high concentrations, mitomycin C causes single-strand breaks in DNA, activating RecA and leading to the autoproteolysis of repressors like LexA that govern the SOS response and are responsible for phage lysogeny. Figure 1 shows two distinct trends for cultures of 11-1. Those grown under optimal conditions undergo continuous growth up to six hours after inoculation under the conditions described in Section 2.2. By contrast those cultures exposed to mitomycin C at two hours after inoculation (during early log phase) show a maximum OD\textsubscript{600} value from 1-2 hr after mitomycin C induction, followed by a decrease in optical density. The damage done by mitomycin C to the DNA of the host is not a guarantee of prophage induction, but demonstrates that mitomycin C exposure at this level was enough to alter growth curves. The next step in confirming that the change in growth was due to the induction of a prophage would be the isolation of a phage particle that contains nucleic acid.

A culture of 11-1 was grown and treated with mitomycin C. 2 hr after induction with mitomycin C, cells and cellular debris were removed by centrifugation. Having cleared the lysate of large cellular debris and intact cells, lysate was incubated with the crowding agent PEG which will precipitate phage particles. The resulting pellet was suspended in a phage buffer to maintain the integrity of any phage present, and incubated with DNaseI and RNase A to remove any nucleic acids left in solution from lysed cells but not packaged into phage particle. After degrading any free nucleic acids, further centrifugation removed cellular debris larger than phages. The resulting suspension should contain only phage particles at this point.

Phage particles were treated with a combination of SDS and EDTA and incubated at 65°C. This will denature the capsid proteins and release the phage-packaged nucleic acid. To qualify the nucleic acid as RNA or DNA, this nucleic acid was treated with DNase I or RNase A and run on a 1% agarose gel. Due to the high
specificity of these enzymes, degradation of the nucleic acid is evidence of its composition.

A substantial difference is demonstrated by all four samples present in Figure 1A. Lane 1 shows the relative size of host strain 11-1 genomic DNA, compared to the distinct banding of untreated nucleic acid packaged by HJ1031 in Lane 2. Treatment of HJ1031 packaged nucleic acid with DNase I results in the complete degradation of the nucleic acid, while the distinct band remains after treatment with RNase A. This is evidence that the nucleic acid packaged into the head of HJ1031 is DNA, and given the resolution of the band appears to be a uniform size.
Figure 2A. Enzymatic Qualification of HJ1031 Packaged Nucleic Acid. Lane 1, 11-1 DNA; Lane 2, HJ1031 packaged nucleic acid; Lane 3, HJ1031 packaged nucleic acid treated with DNase I; Lane 4, HJ1031 packaged nucleic acid treated with RNase A; Lane 5, 1 kb ladder (NEB N3232L)
Figure 2B. Enzymatic Qualification for HJ1031 Packaged DNA Physical Structure. Lane 1, λ HindIII-digested DNA ladder; Lane 2, λ ladder incubated with RecJf; Lane 3, λ ladder incubated with Lambda Exonuclease; Lane 4, λ ladder incubated with Exonuclease III; Lane 5, Litmus38i; Lane 6, Litmus38i incubated with RecJf; Lane 7, Litmus38i incubated with Lambda Exonuclease; Lane 8, Litmus38i incubated with Exonuclease III; Lane 9, HJ1031 packaged DNA; Lane 10, HJ1031 packaged DNA incubated with RecJf; Lane 11, Hj10 packaged DNA incubated with Lambda Exonuclease; Lane 12, HJ1031 packaged DNA incubated with Exonuclease III

Having established that the nucleic acid packaged in HJ1031 particles is DNA, Figure 2B shows the enzymatic qualification of the physical structure of this DNA. The different exonucleases were used because of their substrate specificity. RecJf acts on single-stranded DNA, and will also remove between 17-30 base from double-stranded DNA with a 5’ overhang. Lambda Exonuclease prefers phosphorylated 5’
double-stranded DNA as its substrate, although single-stranded DNA will be degraded at a significantly lower rate as it acts in a 5’→3’ manner. It does not initiate digestion at nicks or gaps, so it will only work on double stranded breaks or linear DNA. Exonuclease III is not active on single-stranded DNA and prefers blunted ends or recessed 3’ ends of double-stranded DNA as its substrate, and completely degrades it until it cannot be visualized. It will also remove nucleotides at nicks or gaps.

The first four lanes are HindIII-digested λ DNA. RecJf digestion, as expected, shows no major observable change at this scale. HindIII-digested λ DNA is mainly comprised of dsDNA. The removal of the small ssDNA ends of the HindIII fragments is not visible in this gel system. However, digestion by Lambda Exonuclease or Exonuclease III degrades the DNA significantly, to the point where it cannot be visualized. Digestion by HindIII create a 5’ overhang of duplex DNA, leaving the preferred substrate of duplex DNA for Lambda Exonuclease and the preferred substrate of recessed 3’ termini for Exonuclease III. The digestion by Lambda Exonuclease also suggests that degradation by Exonuclease III is not simply due to nicks or gaps in the DNA, but is due only to its structure as a linear molecule.

This behavior is contrasted by the incubation of Litmus38i, a circular DNA plasmid, with these same exonucleases. Lane 5 shows untreated Litmus38i. RecJf (Lane 6) also shows no major change in DNA concentration, indicating a lack of activity. Lane 7 also shows no change in DNA concentration, indicating that Lambda Exonuclease is not active against circular DNA that does not show double stranded breaks that would create the substrate of 5’ phosphorylated DNA. However, Lane 8 shows degradation of the upper DNA bands. The upper bands are likely to be Litmus38i with nicks or single stranded breaks that result in some uncoiling of the DNA strand. This would be open to digestion by Exonuclease III which has exonucleic activity against nicks or single stranded breaks. Taken with the
untreated DNA found in lanes 1 and the exonucleic activity demonstrated in lanes 2, 3, 5, and 6, this creates a clear contrast of the exonuclease reactions against linear and circular DNA.

This serves as a template to compare HJ1031 packaged DNA incubated with these same exonucleases. Lane 9 shows an untreated sample of the DNA packaged by HJ1031. Lane 10 shows no significant reaction with RecJf, while Lambda Exonuclease and Exonuclease III show complete degradation of the DNA. This indicates that the DNA packaged by HJ1031 is double-stranded linear DNA.

Having determined the structures of the DNA packaged by HJ1031, further research using enzymatic digestion was performed to determine the composition of this DNA. The banding pattern from an endonuclease-digested HJ1031 genome would allow cloning of the phage into a plasmid vector. This in turn would allow sequencing and identification of target genes like the lysin gene.

Figure 3. Endonuclease Digestion of Genomic *B. pumilus* DNA and of HJ1031 Derived DNA. Lane 1, digestion with *BamH*I; Lane 2, digestion with *BfuCl*; Lane 3, digestion with *BglII*; Lane 4, digestion with *BspHI*; Lane 5, digestion with *BstUI*; Lane 6, digestion with *EcoRI*; Lane 7, digestion with *Sau3AI*; Lane 8, 1 kb DNA ladder.
HJ1031 was anticipated to synthesize and package its own genomic material. This highly conserved sequence would produce discrete bands after digestion with endonucleases. Digestion with a wide variety of endonucleases did not produce discrete bands, but instead produced smears on an agarose gel of varying intensity. It was unclear at first if this was a failure on the part of the endonucleases due to reaction conditions (including DNA processing like methylation) or if there was another cause for the lack of a distinct banding pattern.

Figure 3 shows two gels. The gel on the left is genomic DNA from 11-1 that has been digested with endonucleases. The gel on the right is DNA derived from HJ1031 particles that has been digested by those same endonucleases. The two sources of DNA show similar banding patterns, with the difference being the upper limit in size. The 11-1 DNA has an upper limit in the unresolvable range (>10 kb), while the upper limit for DNA derived from HJ1031 is at the 8 kb mark. Two of the endonucleases used, BfuCI and Sau3AI, are isoschizomers sharing a common digestion site 4 base pairs long. The other endonucleases used here have palindromic sites 6 base pairs long. The HJ1031 derived DNA mirrors the 11-1 DNA in frequency of palindromic sites, indicating that there is no region selectivity when packaging DNA.

These digestions also tell us about post-excision processing and modification. BfuCI, BstUI, EcoRI, and Sau3AI are all CpG methylation-sensitive digestion endonucleases, while BspHI is dam methylation sensitive. The similar banding patterns between digested 11-1 genomic DNA and HJ1031 packaged DNA indicates not only that the fragments are random and equal portions of the 11-1 genome, but that CpG methylation and dam methylase activity are not a part of post-excision processing. Taken together with the information from Figure 2, this indicates that no significant post-excision DNA processing is performed by HJ1031. The HJ1031
DNA appears to be made of random samplings of genomic 11-1 DNA. To test this hypothesis, HJ1031 packaged DNA would have to be cloned and sequenced.

To achieve the fragmentation of HJ1031 DNA for cloning, we chose sonication. It results in the random fragmentation of the DNA. Some of these fragments may have overhanging ends. These random sonicated fragments were blunted using the NEB Quick Blunting Kit. The fragments were ligated with vector Litmus 38i that had been digested with StuI, an endonuclease that performs a blunt restriction. This left the Litmus 38i open to ligation with blunted random fragments of HJ1031 DNA. Blue/white phenotype screening indicated any vectors that had seen interruption of the multiple cloning site (MCS), through the successful cloning of some small fragment of HJ1031 packaged DNA. These clones were then prepared by alkaline lysis miniprep and amplified using PCR with primers designed to anneal to Litmus 38i. PCR products were then sequenced.

### Table 1. Sequencing of PCR Amplified DNA from Alkaline Lysis Miniprep of Litmus38i-HJ1031 Recombinant DNA

<table>
<thead>
<tr>
<th>Sample</th>
<th>Oligonucleotide</th>
<th>Unique Sequence</th>
</tr>
</thead>
</table>
| 8      | Litmus38i FWD   | GGTGTAAAGTGCCGATATTGAGGAATCCTTTACACAGCTAGGACTCGCAATGCTTGCAGCGATTGCTATTGTATACCTCGTTCTCGTATTACATTTGGCGGAGGTCTTGCACCATTGGCGATTCTGTTCTCACTTCCATTTACGATTATTGGTGCACTCGTTGGATTGTGTAGCGAAAGAAACCATCAGCTTTAATGCCATCTCATAGTGAGTCGTATTAGTGAGCGAGCCAGCGCCAGTGGGTAATACGACTCACTATAGGATGGCATTTAAGCTGGTGTTTCTTTCGCTCAAACAATCCCAAATCGACCCAAATATATGTAATGACGAGAACGAGGATACAATAGCAATTGCTGCAAGCAGTCTAGCTGTGAAGATATATACGAGACCCGCGCCAAATATGTAATGACGAGAACGAGGATACCAATATAGCAATTGCTGCAAGCAGTCTAGCTGTGAAGATATATACGAGACCCGCGCCAAATATGTAATGACGAGAACGAGGATACCAATATAGCAATTGCTGCAAGCAGTCTAGCTGTGAAGATATATACGAGACCCGCGCCAAATATGTAATGACGAGAACGAGGATACCAATATAGCAATTGCTGCAAGCAGTCTAGCTGTGAAGATATATACGAGACCCGCGCCAAATATGTAATGACGAGAACGAGGATACCAATATAGCAATTGCTGCAAGCAGTCTAGCTGTGAAGATATATACGAGACCCGCGCCAAATATGTAATGACGAGAACGAGGATACCAATATAGCAATTGCTGCAAGCAGTCTAGCTGTGAAGATATATACGAGACCCGCGCCAAATATGTAATGACGAGAACGAGGATACCAATATAGCAATTGCTGCAAGCAGTCTAGCTGTGAAGATATATACGAGACCCGCGCCAAATATGTAATGACGAGAACGAGGATACCAATATAGCAATTGCTGCAAGCAGTCTAGCTGTGAAGATATATACGAGACCCGCGCCAAATATGTAATGACGAGAACGAGGATACCAATATAGCAATTGCTGCAAGCAGTCTAGCTGTGAAGATATATACGAGACCCGCGCCAAATATGTAATGACGAGAACGAGGATACCAATATAGCAATTGCTGCAAGCAGTCTAGCTGTGAAGATATATACGAGACCCGCGCCAAATATGTAATGACGAGAACGAGGATACCAATATAGCAATTGCTGCAAGCAGTCTAGCTGTGAAGATATATACGAGACCCGCGCCAAATATGTAATGACGAGAACGAGGATACCAATATAGCAATTGCTGCAAGCAGTCTAGCTGTGAAGATATATACGAGACCCGCGCCAAATATGTAATGACGAGAACGAGGATACCAATATAGCAATTGCTGCAAGCAGTCTAGCTGTGAAGATATATACGAGACCCGCGCCAAATATGTAATGACGAGAACGAGGATACCAATATAGCAATTGCTGCAAGCAGTCTAGCTGTGAAGATATATACGAGACCCGCGCCAAATATGTAATGACGAGAACGAGGATACCAATATAGCAATTGCTGCAAGCAGTCTAGCTGTGAAGATATATACGAGACCCGCGCCAAATATGTAATGACGAGAACGAGGATACCAATATAGCAATTGCTGCAAGCAGTCTAGCTGTGAAGATATATACGAGACCCGCGCCAAATATGTAATGACGAGAACGAGGATACCAATATAGCAATTGCTGCAAGCAGTCTAGCTGTGAAGATATATACGAGACCCGCGCCAAATATGTAATGACGAGAACGAGGATACCAATATAGCAATTGCTGCAAGCAGTCTAGCTGTGAAGATATATACGAGACCCGCGCCAAATATGTAATGACGAGAACGAGGATACCAATATAGCAATTGCTGCAAGCAGTCTAGCTGTGAAGATATATACGAGACCCGCGCCAAATATGTAATGACGAGAACGAGGATACCAATATAGCAATTGCTGCAAGCAGTCTAGCTGTGAAGATATATACGAGACCCGCGCCAAATATGTAATGACGAGAACGAGGATACCAATATAGCAATTGCTGCAAGCAGTCTAGCTGTGAAGATATATACGAGACCCGCGCCAAATATGTAATGACGAGAACGAGGATACCAATATAGCAATTGCTGCAAGCAGTCTAGCTGTGAAGATATATACGAGACCCGCGCCAAATATGTAATGACGAGAACGAGGATACCAATATAGCAATTGCTGCAAGCAGTCTAGCTGTGAAGATATATACGAGACCCGCGCCAAATATGTAATGACGAGAACGAGGATACCAATATAGCAATTGCTGCAAGCAGTCTAGCTGTGAAGATATATACGAGACCCGCGCCAAATATGTAATGACGAGAACGAGGATACCAATATAGCAATTGCTGCAAGCAGTCTAGCTGTGAAGATATATACGAGACCCGCGCCAAATATGTAATGACGAGAACGAGGATACCAATATAGCAATTGCTGCAAGCAGTCTAGCTGTGAAGATATATACGAGACCCGCGCCAAATATGTAATGACGAGAACGAGGATACCAATATAGCAATTGCTGCAAGCAGTCTAGCTGTGAAGATATATACGAGACCCGCGCCAAATATGTAATGACGAGAACGAGGATACCAATATAGCAATTGCTGCAAGCAGTCTAGCTGTGAAGATATATACGAGACCCGCGCCAAATATGTAATGACGAGAACGAGGATACCAATATAGCAATTGCTGCAAGCAGTCTAGCTGTGAAGATATATACGAGACCCGCGCCAAATATGTAATGACGAGAACGAGGATACCAATATAGCAATTGCTGCAAGCAGTCTAGCTGTGAAGATATATACGAGACCCGCGCCAAATATGTAATGACGAGAACGAGGATACCAATATAGCAATTGCTGCAAGCAGTCTAGCTGTGAAGATATATACGAGACCCGCGCCAAATATGTAATGACGAGAACGAGGATACCAATATAGCAATTGCTGCAAGCAGTCTAGCTGTGAAGATATATACGAGACCCGCGCCAAATATGTAATGACGAGAACGAGGATACCAATATAGCAATTGCTGCAAGCAGTCTAGCTGTGAAGATATATACGAGACCCGCGCCAAATATGTAATGACGAGAACGAGGATACCAATATAGCAATTGCTGCAAGCAGTCTAGCTGTGAAGATATATACGAGACCCGCGCCAAATATGTAATGACGAGAACGAGGATACCAATATAGCAATTGCTGCAAGCAGTCTAGCTGTGAAGATATATACGAGACCCGCGCCAAATATGTAATGACGAGAACGAGGATACCAATATAGCAATTGCTGCAAGCAGTCTAGC...
Table 1. Continued.

| 16 | Litmus38i FWD | GCGTCGACCCTCTTAGTCAAGGCGGATTTCTTTTATTGCCATTTGACACATTAGCTTTTTTTTCTTTTCTTTTCTTTGATCCCTTTACCAGGTATACGCAACATAGGCTTACTGCTTTGACGCCTTGATTCGATGCCTGCGAATTTAATTTTTTCTGTACAGGCTCTAGGCCCATCTTGCTGCTAATCGCATTGCCGACATCAAATCAAAGCCGGACATCTGGTTTCCATCCATGTAACTGAATGGGTGGAATTCACCTGATGCTGCAAAAATAATTTCCCATCCTCCAAAATCGGCGACCCTTTCGCTGTAGAAATTACGTTTATCTCCGTTCCCATTTGTTTCATTTGAACCGCACGAAAGACAAGTAATACTGCAACCGAGAAAAGGGCTGGGACCCAAAACCGTCTTTTTTAAATAATAAAGCCTATATGAGTGATCTATTACG |
| 38 | Litmus38i FWD | GCGTCGACCCTCTTAGTCAAGGCGGATTTCTTTTATTGCCATTTGACACATTAGCTTTTTTTTCTTTTCTTTTCTTTGATCCCTTTACCAGGTATACGCAACATAGGCTTACTGCTTTGACGCCTTGATTCGATGCCTGCGAATTTAATTTTTTCTGTACAGGCTCTAGGCCCATCTTGCTGCTAATCGCATTGCCGACATCAAATCAAAGCCGGACATCTGGTTTCCATCCATGTAACTGAATGGGTGGAATTCACCTGATGCTGCAAAAATAATTTCCCATCCTCCAAAATCGGCGACCCTTTCGCTGTAGAAATTACGTTTATCTCCGTTCCCATTTGTTTCATTTGAACCGCACGAAAGACAAGTAATACTGCAACCGAGAAAAGGGCTGGGACCCAAAACCGTCTTTTTTAAATAATAAAGCCTATATGAGTGATCTATTACG |
| 40 | Litmus38i FWD | GCGTCGACCCTCTTAGTCAAGGCGGATTTCTTTTATTGCCATTTGACACATTAGCTTTTTTTTCTTTTCTTTTCTTTGATCCCTTTACCAGGTATACGCAACATAGGCTTACTGCTTTGACGCCTTGATTCGATGCCTGCGAATTTAATTTTTTCTGTACAGGCTCTAGGCCCATCTTGCTGCTAATCGCATTGCCGACATCAAATCAAAGCCGGACATCTGGTTTCCATCCATGTAACTGAATGGGTGGAATTCACCTGATGCTGCAAAAATAATTTCCCATCCTCCAAAATCGGCGACCCTTTCGCTGTAGAAATTACGTTTATCTCCGTTCCCATTTGTTTCATTTGAACCGCACGAAAGACAAGTAATACTGCAACCGAGAAAAGGGCTGGGACCCAAAACCGTCTTTTTTAAATAATAAAGCCTATATGAGTGATCTATTACG |
| 41 | Litmus38i FWD | GCGTCGACCCTCTTAGTCAAGGCGGATTTCTTTTATTGCCATTTGACACATTAGCTTTTTTTTCTTTTCTTTTCTTTGATCCCTTTACCAGGTATACGCAACATAGGCTTACTGCTTTGACGCCTTGATTCGATGCCTGCGAATTTAATTTTTTCTGTACAGGCTCTAGGCCCATCTTGCTGCTAATCGCATTGCCGACATCAAATCAAAGCCGGACATCTGGTTTCCATCCATGTAACTGAATGGGTGGAATTCACCTGATGCTGCAAAAATAATTTCCCATCCTCCAAAATCGGCGACCCTTTCGCTGTAGAAATTACGTTTATCTCCGTTCCCATTTGTTTCATTTGAACCGCACGAAAGACAAGTAATACTGCAACCGAGAAAAGGGCTGGGACCCAAAACCGTCTTTTTTAAATAATAAAGCCTATATGAGTGATCTATTACG |
| 42 | Litmus38i FWD | GCGTCGACCCTCTTAGTCAAGGCGGATTTCTTTTATTGCCATTTGACACATTAGCTTTTTTTTCTTTTCTTTTCTTTGATCCCTTTACCAGGTATACGCAACATAGGCTTACTGCTTTGACGCCTTGATTCGATGCCTGCGAATTTAATTTTTTCTGTACAGGCTCTAGGCCCATCTTGCTGCTAATCGCATTGCCGACATCAAATCAAAGCCGGACATCTGGTTTCCATCCATGTAACTGAATGGGTGGAATTCACCTGATGCTGCAAAAATAATTTCCCATCCTCCAAAATCGGCGACCCTTTCGCTGTAGAAATTACGTTTATCTCCGTTCCCATTTGTTTCATTTGAACCGCACGAAAGACAAGTAATACTGCAACCGAGAAAAGGGCTGGGACCCAAAACCGTCTTTTTTAAATAATAAAGCCTATATGAGTGATCTATTACG |
| 43 | Litmus38i FWD | GCGTCGACCCTCTTAGTCAAGGCGGATTTCTTTTATTGCCATTTGACACATTAGCTTTTTTTTCTTTTCTTTTCTTTGATCCCTTTACCAGGTATACGCAACATAGGCTTACTGCTTTGACGCCTTGATTCGATGCCTGCGAATTTAATTTTTTCTGTACAGGCTCTAGGCCCATCTTGCTGCTAATCGCATTGCCGACATCAAATCAAAGCCGGACATCTGGTTTCCATCCATGTAACTGAATGGGTGGAATTCACCTGATGCTGCAAAAATAATTTCCCATCCTCCAAAATCGGCGACCCTTTCGCTGTAGAAATTACGTTTATCTCCGTTCCCATTTGTTTCATTTGAACCGCACGAAAGACAAGTAATACTGCAACCGAGAAAAGGGCTGGGACCCAAAACCGTCTTTTTTAAATAATAAAGCCTATATGAGTGATCTATTACG |
| 45 | Litmus38i FWD | GCGTCGACCCTCTTAGTCAAGGCGGATTTCTTTTATTGCCATTTGACACATTAGCTTTTTTTTCTTTTCTTTTCTTTGATCCCTTTACCAGGTATACGCAACATAGGCTTACTGCTTTGACGCCTTGATTCGATGCCTGCGAATTTAATTTTTTCTGTACAGGCTCTAGGCCCATCTTGCTGCTAATCGCATTGCCGACATCAAATCAAAGCCGGACATCTGGTTTCCATCCATGTAACTGAATGGGTGGAATTCACCTGATGCTGCAAAAATAATTTCCCATCCTCCAAAATCGGCGACCCTTTCGCTGTAGAAATTACGTTTATCTCCGTTCCCATTTGTTTCATTTGAACCGCACGAAAGACAAGTAATACTGCAACCGAGAAAAGGGCTGGGACCCAAAACCGTCTTTTTTAAATAATAAAGCCTATATGAGTGATCTATTACG |
| 46 | Litmus38i FWD | GCGTCGACCCTCTTAGTCAAGGCGGATTTCTTTTATTGCCATTTGACACATTAGCTTTTTTTTCTTTTCTTTTCTTTGATCCCTTTACCAGGTATACGCAACATAGGCTTACTGCTTTGACGCCTTGATTCGATGCCTGCGAATTTAATTTTTTCTGTACAGGCTCTAGGCCCATCTTGCTGCTAATCGCATTGCCGACATCAAATCAAAGCCGGACATCTGGTTTCCATCCATGTAACTGAATGGGTGGAATTCACCTGATGCTGCAAAAATAATTTCCCATCCTCCAAAATCGGCGACCCTTTCGCTGTAGAAATTACGTTTATCTCCGTTCCCATTTGTTTCATTTGAACCGCACGAAAGACAAGTAATACTGCAACCGAGAAAAGGGCTGGGACCCAAAACCGTCTTTTTTAAATAATAAAGCCTATATGAGTGATCTATTACG |
| 49 | Litmus38i FWD | GCGTCGACCCTCTTAGTCAAGGCGGATTTCTTTTATTGCCATTTGACACATTAGCTTTTTTTTCTTTTCTTTTCTTTGATCCCTTTACCAGGTATACGCAACATAGGCTTACTGCTTTGACGCCTTGATTCGATGCCTGCGAATTTAATTTTTTCTGTACAGGCTCTAGGCCCATCTTGCTGCTAATCGCATTGCCGACATCAAATCAAAGCCGGACATCTGGTTTCCATCCATGTAACTGAATGGGTGGAATTCACCTGATGCTGCAAAAATAATTTCCCATCCTCCAAAATCGGCGACCCTTTCGCTGTAGAAATTACGTTTATCTCCGTTCCCATTTGTTTCATTTGAACCGCACGAAAGACAAGTAATACTGCAACCGAGAAAAGGGCTGGGACCCAAAACCGTCTTTTTTAAATAATAAAGCCTATATGAGTGATCTATTACG |

The PCR products from clones were sequenced and shown in Table 1. The StuI half sites flanking the unique sequences found in the MCS are underlined, with an
additional 20 nucleotides of vector sequence. Successfully cloned sequence length ranged from 37-383 base pairs in length. Clones shorter than 37 base pairs in length may have been too short to produce any meaningful data. Clones had an upper maximal limit of 8 kb in length (the length of an entire DNA fragment packaged into HJ1031) and a practical upper limit that was even smaller since it becomes difficult to clone fragments into plasmid vectors past a certain size.

Having this information, we proceeded to access the NCBI databanks for homology to known sequences. This may reveal whether this was DNA containing phage-related genes or if our earlier hypothesis of packaging host chromosome was correct.

<table>
<thead>
<tr>
<th>Sample</th>
<th>BLASTN Organism</th>
<th>Gene</th>
<th>% Identity</th>
<th>e-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>8 forward</td>
<td><em>Bacillus pumilus</em> SAFR-032</td>
<td>swrC</td>
<td>93</td>
<td>3e-74</td>
</tr>
<tr>
<td>8 reverse</td>
<td><em>Bacillus pumilus</em> SAFR-032</td>
<td>swrC</td>
<td>91</td>
<td>3e-59</td>
</tr>
<tr>
<td>16</td>
<td><em>Bacillus pumilus</em> SAFR-032</td>
<td>BPUM_1825</td>
<td>86</td>
<td>1e-110</td>
</tr>
<tr>
<td>38</td>
<td><em>Bacillus pumilus</em> SAFR-032</td>
<td>BPUM_3296</td>
<td>93</td>
<td>9e-126</td>
</tr>
<tr>
<td>42</td>
<td><em>Bacillus pumilus</em> SAFR-032</td>
<td>spo II D</td>
<td>90</td>
<td>1e-35</td>
</tr>
<tr>
<td>46</td>
<td><em>Bacillus intermedius</em>, strain 3-19</td>
<td>ywfA</td>
<td>96</td>
<td>2e-26</td>
</tr>
<tr>
<td>49</td>
<td><em>Bacillus pumilus</em>, SAFR-032</td>
<td>ywf0</td>
<td>94</td>
<td>4e-23</td>
</tr>
<tr>
<td>49</td>
<td><em>Bacillus pumilus</em>, SAFR-032</td>
<td>pabC</td>
<td>91</td>
<td>2e-20</td>
</tr>
</tbody>
</table>
All searches using the DNA sequence of the cloned fragments returned *Bacillus pumilus* as a candidate for homology, except for sample 43. No sample was 100% identical to previously sequenced *B. pumilus* genes, with clones showing homology ranging from 86%-96%. Search results for samples 40, 41, 43, and 45 are not shown in Table 2. Owing to their short size, they yielded large numbers of results that cannot be displayed here.

None of these genes are related to phages. Based on the banding patterns in Figure 3 and the gene homology identified in Table 2, HJ1031 is not packaging DNA that contains phage-related genes.
Sequencing the 16S rRNA gene gave unambiguous information about the identity of 11-1. The PCR product is from a 2734 bp region of the 16S ribosomal RNA gene. Table 3 shows only four hits from each BLASTN query of the amplified gene, sequenced by primer fD1 and rD1. For the sequence using primer fD1, the first 100 matches have an identity score from 95%-96%, and an E value ranging from 1e-
155 to 5e-159. For the sequence using primer rD1, the first 100 matches have an identity score of 99% and an E value of 0. Many of these strains have not even been given formal names. The small variations in scores were for members of different but closely related species, like *Bacillus aerophilus* and *Bacillus stratosphericus*. 11-1 is certainly a member of the *Bacillus* genus, and appears to be a *B. pumilus* strain.

Coupled with the data in Table 2 that shows HJ1031 packaging host DNA, it is likely that variations found in cloned fragments in Table 2 are the results of ordinary mutation in regions that are not as well conserved as 16S ribosomal RNA gene. Since analyzing the genes that encode HJ1031 by sequencing the DNA it packaged proved extremely impractical, it became important to analyze HJ1031 by its protein structure. Identifying structural proteins in the viral particle will provide insight into the identity and function of HJ1031, and help to classify it by taxonomy. To move forward, we proceeded first by Transmission Electron Microscopy (TEM) to understand these proteins as part of a whole structure.
3.2. Analysis of HJ1031 Proteins

This image from TEM showed the greatest resolution of our phage particle. It was produced using the methods described in 2.13. The sheath protein is approximately 189 nm in length, connected by a neck protein with an icosahedral head 33 nm in diameter. At the other end of the sheath, tail fiber proteins are clearly visible and well defined.
Table 4. Quantitation of HJ1031 Particle Size from TEM

<table>
<thead>
<tr>
<th>Phage Sample</th>
<th>Tail Length</th>
<th>Tail Diameter</th>
<th>Neck Length</th>
<th>Head Diameter</th>
<th>Vertex to Vertex</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>187 nm</td>
<td>17 nm</td>
<td>11 nm</td>
<td>33 nm</td>
<td>35 nm</td>
</tr>
<tr>
<td>2</td>
<td>190 nm</td>
<td>16 nm</td>
<td>10 nm</td>
<td>34 nm</td>
<td>35 nm</td>
</tr>
<tr>
<td>3</td>
<td>191 nm</td>
<td>16 nm</td>
<td>10 nm</td>
<td>34 nm</td>
<td>36 nm</td>
</tr>
<tr>
<td>4</td>
<td>184 nm</td>
<td>17 nm</td>
<td>9 nm</td>
<td>34 nm</td>
<td>35 nm</td>
</tr>
<tr>
<td>5</td>
<td>200 nm</td>
<td>15 nm</td>
<td>10 nm</td>
<td>31 nm</td>
<td>33 nm</td>
</tr>
<tr>
<td>6</td>
<td>188 nm</td>
<td>18 nm</td>
<td>11 nm</td>
<td>33 nm</td>
<td>39 nm</td>
</tr>
<tr>
<td>7</td>
<td>182 nm</td>
<td>19 nm</td>
<td>8 nm</td>
<td>33 nm</td>
<td>35 nm</td>
</tr>
<tr>
<td>8</td>
<td>197 nm</td>
<td>20 nm</td>
<td>9 nm</td>
<td>35 nm</td>
<td>36 nm</td>
</tr>
<tr>
<td>9</td>
<td>195 nm</td>
<td>17 nm</td>
<td></td>
<td>33 nm</td>
<td>36 nm</td>
</tr>
<tr>
<td>10</td>
<td>175 nm</td>
<td>15 nm</td>
<td>11 nm</td>
<td>33 nm</td>
<td>39 nm</td>
</tr>
<tr>
<td>Average</td>
<td>188.9 nm</td>
<td>17 nm</td>
<td>9.8 nm</td>
<td>33.3 nm</td>
<td>35.9 nm</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>6.91 nm</td>
<td>1.55 nm</td>
<td>2.99 nm</td>
<td>1.00 nm</td>
<td>5.6 nm</td>
</tr>
</tbody>
</table>

Measurements were taken from 10 different samples of intact, whole phage particles taken from three TEM images. The images were chosen based on the clarity and resolution. Specific phage samples were chosen for being whole and intact, indicating that shearing forces had not distorted them. A neck length could not be determined for sample 9 in Table 4. Measurements were taken by cropping an image of each phage particle with the legend bar, and comparing the unit length as defined by the imaging software.

Sample averages and standard deviations are listed at the bottom of Table 4. The proportionately large standard deviation listed for measuring the phage head from vertex to vertex may be due to both user error and the difficulty of achieving precise resolution on the vertices. This visual confirmation of the major structure of HJ1031 allowed us to anticipate information about banding patterns and intensity that HJ1031 would provide on an SDS-PAGE gel.

After TEM images of HJ1031 were created, ultrapure samples of HJ1031
particles had to be created in a CsCl gradient for SDS-PAGE and mass spectrometry analysis. In a CsCl gradient centrifugation measurement, pure protein samples typically have a density of 1.2, while pure DNA will have a density of 1.7. Due to the large heads of functional viruses, the proportion of DNA to protein is very high and will result in a band of purified virus particles with a density of 1.5-1.6. Due to its small head size, the defective phage HJ1031 has a higher proportion of protein to DNA than a functional virus and had a band form between 1.3-1.4 in CsCl centrifugation.

![Figure 6. SDS-PAGE Analysis of HJ1031 Proteins Purified by CsCl Gradient Centrifugation and Dialysis.](image)

Lane 1, blank; Lane 2, blank; Lane 3, blank; Lane 4, HJ1031 proteins; Lane 5, blank; Lane 6, molecular weight marker (Promega, V8491).
A 20 μL sample was loaded onto the gel in Figure 6 to prevent overload, allowing us to visualize distinct bands from HJ1031 samples purified as described in Material and Methods, section 2.4. Three major bands are present, at 50 kD, 32 kD, and between 20 kD and 15 kD. A faint band is present below the largest band, but was not determined to be distinct or intense enough to merit mass spectrometry analysis. This gel also is the result of loading a 20 μL sample of purified HJ1031. For further research, overloading the gel with as much sample as a lane could hold may also reveal fainter bands for structural proteins that are present but not as abundant.
Figure 7. MALDI-TOF Mass Spectrometry of HJ1031 Protein Fragments.
The three protein bands seen in Figure 6 were analyzed by mass spectrometry, and a report was generated to identify the proteins based on a reference database provided. The database was the proteomes of *B. pumilus* strains ATCC 7061 and SAFR-032, which had been selected because of their similarity to those strains identifies by using a BLAST nucleotide search for 16S rRNA gene homology. The report described the charge and spectral count from each ionized particle. The spectral count is a highly reliable quantification of relative proteins present after tryptic digests, analyzed by tandem MS/MS analysis. This analysis allows the user to quantify the relative amount of protein of a given mass with high precision, and that mass is correlated with a known proteome database. By quantifying the matches for a proteome by using spectral count, a high degree of certainty can be attained for identifying the protein bands. Each band was matched for its highest spectral count, and a difference of at least 25% in spectral count from the other bands for the same protein identity.

From these results it was surmised that band H1 was most likely protein B4AHJ5|B4AHJ5_BACPU, from *B. pumilus* ATCC 7061. This is a phage tail-sheath protein, a repeating polymer. The high incidence of repetition would account for the relative band intensity present in Figure 6. H2 was likeliest to be protein A8FGQ6|A8FGQ6_BACP2 from *B. pumilus* SAFR-032, a capsid protein. The fact that these two proteins from the same phage show a resemblance to hypothetical proteins found in different host strains is indicative of the mosaicism that occurs in bacteriophages, and may be evidence that this is a phage that has not been characterized before. The host 11-1 may also be a previously unsequenced strain of *B. pumilus*, which would account for difference in homology between the cloned fragments discussed in Table 2 and the comparable genes in the NCBI database.
Band H3 had identical spectral counts for proteins A8FGP9|A8FGP9_BACP2 from B. pumilus SAFR-032 and B4AHJ4|B4AHJ4_BACPU from B. pumilus ATCC 7061 with a count of 56 each. H3 also had a count of 53 for protein B4AH42|B4AH42_BACPU from B. pumilus ATCC 7061, leaving no statistical difference in the likelihood of any of these candidate proteins being an identical match for our protein. All three of these proteins are portal proteins. However the first two results, while identified through different proteomes, code for identical amino acid sequences, and can be considered redundant. Even with the difference between the protein identity with 56 spectral lines and the protein identity with 53 spectral lines, a TBLASTN search yields identical results.

It is also relevant that the nature of this mass spectrometry identifies critical proportions of phage components. As demonstrated in Figure 5, HJ1031 has an icosahedral head. Icosahedrons have 20 facets. At a spectral line count for H2 (the hypothetical capsid protein) of 99, this is nearly perfectly divisible by 5, yielding the ideal repeat of this protein 20 times. The relative protein abundance of H3 (our hypothetical portal protein) is 56. We feel that it is reasonable to round up to 60, given protein fragmentation. Dividing 60 by 5 gives 12 repeating units, the ordinary arrangement for bacteriophage portal proteins, which are present at each of the 12 vertices on the icosahedral head.

Continuing with these proportions, the H1 protein (hypothetical tail-sheath) has a spectral count of 290. Given room for error and peptide fragmentation, it is reasonable to assume that the spectral line count could easily range from 280 to 310. Divided by 5, this means that our hypothetical sheath protein is repeated from 56 to 62 times. This would mean that an individual HJ1031 particle’s structural proteins would have ~60 tail-sheath proteins subunits, 20 capsid proteins subunits, and 12 portal proteins subunits. This also opens up possibilities for future virus characterization. If the repeating structure of a single protein is known, the spectral
count abundance could be used to create well-defined estimates of virus structural proportions.
Discussion

4.1 HJ1031 and PBSX

The bacteriophage presented in this thesis, HJ1031, is a previously uncharacterized bacteriophage. Mitomycin C induction activates it from a lysogenic state in host 11-1 to a lytic state. It packages double stranded linear DNA fragments of unmodified host chromosomal DNA, approximately 8 kb in length. There is no evidence as of yet that there is any bias in what part of the host chromosome is packaged. From the top of its head to the bottom of its tail it is 232 nm in length, and has a relatively small head for a bacteriophage.

Mass spectrometry analysis, as indicated in Section 3.2 (specifically Figure 7 and the explanation of results), was used to give the most probable identity of the structural proteins in HJ1031. After determining the most likely protein identities, primers were reverse engineered from a tBLASTn search using these proteins. These primers were used in the PCR amplification of 11-1 genomic DNA. Successful amplification and sequencing from the primers designed for protein band H2 (the capsid subunit protein) yielded a useful product. After sequencing, this 867 bp gene yielded 99% homology to Bacillus pumilus strain CCTCC AB94180 defective phage PBP180, and 91% homology to B. pumilus SAFR-032. In B. pumilus SAFR-032 this gene is xkdG, a hypothetical phage capsid protein. When using pBLAST to search for homology to the protein for SDS-PAGE band H2, xkdG had been the 15th hit. The amplified gene sequence showed greater homology than the likely identity from a pBLAST search. We may need to modify or refine our reference protein database for future work on HJ1031.

As discussed in section 1.4, phage structural gene arrangement demonstrates synteny. Written from 5’ to 3’ on the negative strand, the relevant genes in B. pumilus SAFR-032 are xkdG, an intervening sequence of 1.7 kb made of
five open reading frames, xkdK, and xkdM. The genes xkdK and xkdM are contiguous to each other, and are hypothesized to be the phage tail-sheath protein and portal protein, respectively. They also demonstrate high homology to two of the open reading frames mentioned in B. pumilus SAFR-032. pBLAST searches for likely proteins identified by mass spectrometry for bands H1 and H3 also gave homology to xkdK and xkdM, respectively. However, CLUSTAL-Omega alignments were not able to produce primers capable of successfully amplifying the other structural protein genes at this time. This may be either due to mosaicism in HJ1031, or because the sequence encoding HJ1031 has diverged from the sequences of other PBSX-like phage present in B. pumilus strains. Successful amplification of the genes responsible for protein bands H1 and H3 gives an anchor point for further research into the other genes for HJ1031, since other phage genes are likely going to be close together on a chromosome.

Initial attempts were made to use HJ1031 for transduction (data not shown). High-density cultures of 11-1 were plated on media containing Rifampicin. Comparison to plating on non-selective media also gave a comparative mutation rate of 1.8 e-9. Rifampicin-resistant isolates were then grown in liquid media in the presence of rifampicin, and underwent mitomycin C induction and phage isolation as previously described. These purified phage particles from rifampicin resistance 11-1 were then incubated with 11-1 that had not been selected for rifampicin resistance. This new culture was then plated on both ordinary LB media and on LB with rifampicin. If HJ1031 had acted as a transducing agent, it would have encapsulated the rifampicin resistance mutation and successfully integrated this into the new host chromosome, showing an increased incidence of rifampicin resistance. Our target culture did no see an increase in rifampicin resistance. At this point, we do not have data indicating that HJ1031 can act as a transducer. This is in line with the work of Anderson and Bott (3).
HJ1031’s discovery was incidental. HJ1031 and close relative are easy to overlook in the search for bacteriophages characterized by standard techniques. The phage cannot currently propagate a second generation, and does not form plaques on a bacterial lawn. Another unique challenge in characterizing HJ1031 is that the DNA packaged in virion particles is not consistent from virion to virion, leaving standard molecular cloning techniques and enzymatic characterization impotent. These characteristics are what lead to characterization by MALD-TOF MS, TEM, and whole-sequence analysis. This does not mean that HJ1031 is not useful, and raises questions about the persistence of HJ1031 in the genome. Three potential reasons for HJ1031 presence in *B. pumilus* arise. To understand these possibilities, a brief discussion of a close relative to HJ1031 is necessary.

PBSX is a defective phage discovered in the 1960s (4) by researchers working on phage presence in *B. subtilis* (termed *B. natto* at the time). Discovery showed the protein encapsulated presence of a DNA fragments, 13 kb in length, produced by host *B. subtilis* undergoing mitomycin C induction (5). No chromosome regional selectivity was observed (5) and generalized transduction could not be demonstrated reliably (3). Its major repressor, Xre (PBSX repressor), is highly homologous to LexA (6).

### 4.2 HJ1031 May Directly Benefit the Host

HJ1031 may directly benefit 11-1 with its lysogeny. A major contribution prophages make to host fitness is the prevention of super-infection by related phages (7, 8). By repressing surface protein receptors, prophages can prevent other phages from adsorbing to the surface of the host, making infection impossible. They may also repress transmembrane or periplasmic proteins necessary to act as a channel for DNA injected by the new phage, so that even those new phage which do adsorp to the host surface are not capable of infecting it.
A second contribution to host fitness is the presence or induction of factors for the survival of the host under adverse conditions, either through direct transduction of survival elements or altered repressor activity. As previously discussed, bacteriophages often transduce genes related to host pathogenicity such as biofilm formation (9, 8, 10) and sporulation (11). It is possible that the lytic phase of ordinarily defective PBSX-like phages is correlated with stress conditions on *Bacillus* hosts that also produce late log-phase or oxidative-damage response conditions (12).

This is further supported by the LexA repressor protein, the *Bacillus* homolog to the Lambda cl repressor protein. *B. subtilis* and *B. pumilus* share 87% homology in LexA structure (13). LexA represses the SOS response and responds to RecA-induced autoproteolysis. LexA also represses *Bacillus* prophages in a lysogenic state. This conservation of form and function opens different possibilities for how *Bacillus* phage use lysogeny to promote host and self propagation. Through the use of antirepressors a phage can determine a commitment point for entering the lytic phage (14). These antirepressors have been demonstrated to allosterically hinder LexA and repressor cognates. This effect is two fold: it extends the range of genes that become open for transcription, while also promoting the induction of multiple phages in a polylysogenic host. Multiple phages becoming lytic will lead to greater odds of transduction or mosaicism events. Since PBSX is under repression by Xre, antirepressor activity is likely to be governed in a similar manner, and so polylysogenic *B. subtilis* will see large shifts in transcription activity based on the presence or absence of PBSX. This same principle likely applies to PBSX-like phages, such as HJ1031.

A more direct example of phage repressor genes directly benefiting or altering host transcription events is demonstrated by the *skin* element (sigmaK
intervening element) (15). SigmaK is an RNA polymerase holoenzyme associated with sporulation events (16). The skin element resembles PBSX Xre and LexA, but does not respond to RecA activity with autoproteolysis. It acts to repress yqaF-N. If overexpressed, these genes will lead to abnormal chromosome segregation in cell division, and cell filamentation. The skin element remains as a necessary regulator for the benefit of the host, even though surrounding regions show no functional prophage genes. Since the skin repressor is impervious to RecA activity, it is likely that in the distant past a prophage infecting B. subtilis became unable to enter the lytic phase, and has gradually decayed. Only those elements still useful to the host have been retained. Given the high similarity between skin and Xre, and the close relationship between B. subtilis and B. pumilus, PBSX and HJ1031 are likely to play a role in gene repression. Minor differences in amino acid sequence and three dimensional structure may have an impact on sequence recognition (17). This implicates HJ1031 in a variety of gene regulation processes (18, 19) beyond its own lysogeny. This is one possible reason for the persistence of HJ1031 as a defective phage.

4.3 HJ1031 May Play a Role in Population Dynamics

The second major possibility for the continued existence of HJ1031 and other PBSX-like phages is a direct role in transduction or bacterial populations. This discussion of prophage activation has focused on RecA activity leading to autoproteolysis of repressor proteins like LexA, Xre, and skin. However, further research is emerging concerning alternative pathways for prophage activation. The first of these alternative pathways would be RecA-independent stress induction of prophages. A change in electrochemical gradient across a cell membrane or the altering of surface protein conformation by cation chelation can trigger a series of event altering cellular transcription (20). Imamovic and Muniesa demonstrated that
by chelating the cellular environment of recA- mutants, they were capable of achieving RecA-independent phage activation comparable to mitomycin C induced levels. The proteins RcsA and DsrA have been implicated in phage activation in place of RecA under these conditions; both proteins are involved in cell adhesion and osmotic regulation, leading to the production of RpoS. RpoS is the stress sigma factor, aiding in the transcription of stress-associated genes but not necessarily single-stranded breaks in DNA.

Another RecA-independent path may involve presence of double-stranded breaks. Double stranded breaks do not induce significant RecA activity, but does lead to a twenty-fold increase in RecN activity (21). RecN in turn leads to the activation of PBSX genes under ordinary PBSX regulation, with similar activation of SOS genes. This may likely be a response to the failure of a host cell to replicate or move past a stalled replication fork. This leaves PBSX and PBSX-like phages, including HJ1031, activated by DNA alkylating agents, UV radiation, osmotic shock, or replication failure.

Having established alternative pathways to RecA activation, population dynamics becomes an increasingly likely factor in HJ1031 persistence. The Bacillus genus demonstrates quorum-sensing behavior (22), under a complex web of feedback loops. However, some significant trends have been established. Up-regulation of comK leads to the transcription of a host of competence factors as well as recA. RecA is necessary in this process so that incoming DNA can be recombined into the new host genome. Increased RecA presence or activity will naturally lead to a certain increase in SOS activity and lysogenic phage activation.

This is supported by the works of Ghosh, et al (23). Examining a complex microbial soil community, it was observed that firmicute prophage induction exceeds their representation in the soil community. This hints that auto-inducer (AI-2) activity is a significant pathway for prophage induction. AI-2 concentration will
be directly related to population density, achieved in late-log and stationary phase of a *Bacillus* community. This would occur at the same time as competence is achieved by a small portion of the *Bacillus* and sporulation also occurs in the population (22, 24, 25). This correlation between stationary phase behavior and phage activation may be linked.

The first possible contribution of HJ1031 to stationary phase *B. pumilus* is the addition of extracellular DNA (eDNA) to biofilm formation. eDNA serves as a viscous substance that helps a population of cells adhere to each other and to initiate quorum sensing behavior. This could account for the packaging of random DNA fragments into the headspace of the PBSX-like phage virion. It would be released into the extracellular environment and, after degradation of the capsid protein, be a potential source of material for eDNA.

Another possible reason for maintaining whole but non-infectious virion particles lies in the competition of sporulating organisms for limited resources. PBSX and PBSX-like phages have demonstrated an inability not only to penetrate a new host with packaged genetic material (4, 26), but are unable to even adsorb to their own host strain (27, 4). However, the defective phage can adsorb to other strains and act as bacteriocins (28, 5, 4, 26, 11). With the release of defective phages, a specific strain of *Bacillus subtilis* may ensure the destruction of similar but non-identical strains of *Bacillus*, which would be the strongest competitors for limited resources. There is further evidence that phage activation will occur during stationary phase as evidence uncovers that timing plays a role in phage infection and replication. The multiplicity of infection for a phage and host is meaningless without information about the cell density and the time during which phage and host interact (29). By waiting until appropriate cell density is achieved so that competence and sporulation events activate the prophage bacteriocin, the phage does not become active until an optimum period of cell density and immobility in
competing host strains. The destruction of competing strains will not only eliminate strong competition for resources and promote the host fitness, but contribute to the strength of the extracellular matrix through the contents of the lysed cells.

4.4 HJ1031 May Participate in Horizontal Gene Transfer

The third significant possibility for the persistence of HJ1031 and PBSX-like phages in Bacillus is their role in horizontal gene transfer (HGT). As discussed earlier, phages play an immense role in accelerating evolution. At present, HJ1031 and other PBSX-like phage have not demonstrated canonical transduction. This does not mean that HJ1031 cannot play a role in HGT.

The most likely path for HJ1031 involvement in HGT is as an integration site for infection and recombination by other phages capable of transduction (3, 30, 31, 32, 33). As previously discussed, instances of polylysogenic hosts are identified at an increasing rate. After prevention of superinfection has been overcome, the most successful integration sites tend to be where prophages have already integrated (34, 35, 36). This presents HJ1031 as a dock for successful temperate phage. The intervening sequences of DNA that are not essential for a phage’s life cycle would then allow for the exchange of genetic material without necessarily interrupting an already functional and integral gene (7, 24, 37).

While this is a possibility, we feel that this answer is insufficient. It does not explain the persistence of structural genes found in HJ1031. A host looking only to use a phage as an exchange site for DNA would over time lose structural elements of the phage that result in the lysis of the host. It is then likely that the integration and subsequent activation of a successful temperate phage may include the assembly of mosaic structural virions, assembled from pieces of the successful phage and HJ1031. This would give some phages an adaption to a new host range, conferring an advantage to the phage and host through enhanced transduction opportunities.
Finally, HJ1031 virion particles may play an indirect role in HGT through altered transformation. As discussed in section 4.3, eDNA plays a critical role in the stability of biofilms and extracellular matrixes during quorum sensing behavior. This same behavior triggers competence. It is possible that HJ1031 persists because of its ability to lyse a host cell, contributing to the biofilm and presenting further eDNA for those cells which do become competent (24, 25). The virions which do package DNA but do not immediately adsorp to a new host cell may lie in the biofilm until their capsid gradually degrades exposing the 8 kb fragment to the extracellular matrix, contributing further to this process. Both the incidence of transformation of biofilm eDNA and the integrity of HJ1031 particles under these conditions should be researched further before more suppositions are made.

4.5 Concluding Thoughts

Further investigation into the genome of HJ1031 is warranted, and is becoming increasingly feasible. Our strategies in the work of this thesis were performed based on an inability to sequence the genome of HJ1031 based on what the virion particle was packaging. As whole genome sequencing continues to become more available and inexpensive, it may become a primary technique in identifying these phages. This is an option that our lab is currently pursuing, and the results may be included in future works. In addition to describing the entire genome of host 11-1 and the genome of HJ1031 that lies inside this, there is an addition benefit. We are able to characterize the DNA packaged by HJ1031 in a more quantitative way then our current strategy of random-fragmentation cloning. This technique should reveal both the proportions of 11-1 DNA fragments packaged into HJ1031 and the location of these fragments on the chromosome.

More research should also be guided towards HJ1031’s potential role as a GTA. While previous work indicates that PBSX and PBSX-like phages are not capable
of infecting new hosts with their DNA, (3) this conclusion may be upended. The first step in further investigation in our lab would be to determine adsorption rates, and whether or not a target culture must be from a different strain. This could be important as prophages are known to prevent super-infection, and one of the mechanisms for this is by down-regulating expression of the target receptor, preventing adsorption (7). The marker for transduction, rifampicin resistance, is easy to induce. However, it may alter production levels of HJ1031, further complicating transduction processes.

Characterization of defective phages like HJ1031 and PBSX was limited in the past by available technology. While individual proteins could be isolated, genuine statistical analysis of the composition of packaged DNA was nearly impossible. This is compounded by the fact that, as a defective phage, HJ1031 does not have perfect homology with other related phages (given evidence by high homology in the capsid protein, but relatively low homology in both the tail-sheath and portal proteins identified through MALDI-TOF MS). As online genomic analysis tools have become available and genome sequencing has become geometrically cheaper and faster, these difficulties can be overcome.

Stream-lining the identification and characterization of these phages will become important. In the future, protocols may be used to precipitate and identify any extracellular, packaged DNA present after mitomycin C induction of a potential host culture. Upon identifying repeatable, packaged nucleic acid elements it may become simpler to proceed to direct whole genome sequencing. The potential applications are significant.

Whole genome sequencing of the host would identify phage structural genomes. Compiling that information in online resources like GenBank reveals critical points of both form and functions in these proteins. These could be used to expand the range of current phage research, from protein panning in phage-display
to the increase of narrow spectrum antibiotics that could be used in both research and clinical settings. Structural genes are also likely to be contiguous or near any genetic exchange sites for the previously discussed DNA that is not essential for phage function. Sampling and sequencing of this DNA would prove relevant in a clinical setting in helping to predict the likely HGT of antibiotic resistance genes through bacterial communities and their human host populations.
References


34. **Fogg PCM, Gossage SM, Smith DL, Saunders JR, McCarthy AJ, Allison HE.**
Doi10.1099/mic.0.2007/011205-0.

