2017

The Role of Phage tRNAs in the Evolution of Codon Usage Biases in Giant Pseudomonas Phage phiKZ and EL

Stephanie Steidel
Loyola University Chicago

LOYOLA UNIVERSITY CHICAGO

THE ROLE OF PHAGE TRNAS IN THE EVOLUTION OF CODON USAGE BIASES IN GIANT *PSEUDOMONAS PHAGE PHIKZ* AND *EL*

A THESIS SUBMITTED TO

THE FACULTY OF THE GRADUATE SCHOOL

IN CANDIDACY FOR THE DEGREE OF

MASTER OF SCIENCE

PROGRAM IN BIOLOGY

BY

STEPHANIE C. STEIDEL

CHICAGO, IL

MAY 2017

ACKNOWLEDGEMENTS

Enormous thanks to my mother who has always been my biggest ally. Thanks to my committee, whose assistance and unending patience has been pivotal to the completion of this work. Thanks especially to Dr. Catherine Putonti who is an absolute power house of getting-things-done. Thanks to the Biology Department faculty and staff who have been quintessential resources throughout my graduate career. Thanks especially to Dr. Meghan Helfgott for all her support. Thanks to my fellow graduate students for all their humor and encouragement. Thanks to all the members of the Putonti Lab who have been endlessly kind and helpful.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

As the most abundant and diverse biological agents in the biosphere phage have significant roles in microbial ecology, acting both as lethal bacterial parasites and vehicles of horizontal gene transfer. Phage/host coevolution drives optimization of phage codon usage for use of host translational machinery, thus lowered correspondence between phage and host codon usage reduces viral fitness. Some phage may partially bypass host translational selection on their codon usage by encoding their own tRNAs, although the effects of these tRNAs on phage codon usage and translation has not been well examined. This work explores the influence of phage encoded tRNAs on viral codon usage via 1) codon usage analysis of *Pseudomonas aeruginosa* phage phiKZ and EL; 2) attempted engineering of phiKZ populations with deoptimized proline tRNAs; and 3) engineering and experimental evolution of mutant phiKZ strain, B1, with duplicated aspartic acid, methionine, and proline tRNAs.

CHAPTER ONE

INTRODUCTION

Bacteriophage

Viruses are the most abundant biological agents on Earth, in fact, an organism has yet to be found that does not have some form of virus utilizing it as a host. Viruses are model parasites which use the resources of a host organism to meet their own survival needs at the expense of host fitness. Viral parasitism is so extreme that viruses themselves are not considered living, rather most depend entirely on host cells to meet all their metabolic needs, including reproduction. It follows that viral genomes are highly compact, coding only for factors essential to commandeering hosts and converting them into virus factories. Although nonliving, viral species evolve and adapt to changes in their environment to promote their own continuation just as any living population does.

Bacteriophage (phage), viruses that only infect bacteria, are the most abundant, complex, and possibly oldest of the viruses. These most prolific of viruses have served as essential tools in biological research. Studies in phage led to the identification of DNA as the hereditary molecule, breaking the genetic code, characterizing RNA, and a number of other discoveries upon which molecular biology was founded (Clark & March 2006). The biosphere is thought to harbor as many as 10^{31} phage virions at any given time and these particles overturn almost entirely in a matter of days (Hatfull 2015). In all environments, phage are found to outnumber their bacterial hosts (Bailly-Bechet 2007). The tremendous numbers and broad environmental range of phage

spawns enormous diversity. No two identical strains of phage have ever been isolated twice in nature (Pedulla *et al*. 2003).

Phage demonstrate a wide array of morphologies, genome structures and life cycle types. The most common and most recognized phage structure consists of an icosahedral head, or capsid, with a tail and tail fibers. However, this common virion structure is not the rule. Many phage lack tails or tail fibers, some have lipoprotein envelopes, and some lack capsids entirely, having instead a fibrous virion structure (Parasion *et al*. 2014). Phage are in part classified based on their morphology which includes differences in capsid size, tail length, and tail type (contractile or non-contractile), as well as their host range and life cycle (Ackerman & Nguyen 1983, Hendrix *et al*. 1999).

The genomes of phage also come in a variety of forms. Phage genomes may be single or double stranded, circular or linear, singular or segmented, and may be composed of RNA or DNA. Due to differences in their molecular stability, each genome composition is subject to a different mutational rate. RNA and single stranded genomes mutate at a greater rate than double stranded or DNA genomes (Duffy *et al.* 2008, Sanjuán *et al*. 2010). The majority of known phage have double stranded (ds) DNA genomes, the most stable genome structure and the easiest phage type to isolate in the lab (Serwer *et al*. 2007, Hatfull 2008). Order *Caudovirales*, dsDNA tailed phage, make up what has long been considered to be the largest monophyletic group of biological agents (Hertveldt *et al*. 2005, Hendrix *et al*. 1999).

Phage further demonstrate a range of lifestyles including lytic, lysogenic, chronic infective, and pseudolysogenic (Mills *et al*. 2013), of which lytic and lysogenic life cycles are most common. An illustrated overview of the lytic and lysogenic life cycles is shown in Figure 1. All phage begin their life cycle with the adsorption process wherein the phage comes in contact

with an appropriate bacterial host: one for which the phage's capsid, tail or tail fibers are specifically receptive, allowing the phage to attach itself to the bacterial cell wall. Phage are unable to actively seek out hosts and so depend on host populations exceeding a threshold density to maintain infection rates (De Paepe *et al*. 2014). Adsorbed phage proceed to puncture the cell's outer covering to inject its genome into the host cytoplasm. The cell may be penetrated enzymatically, mechanically, or through some combination of the two (Briers *et al*. 2007, Paradis-Bleau *et al*. 2007, Aksyuk *et al*. 2011). After injection, lytic and lysogenic phage diverge in their infection strategy.

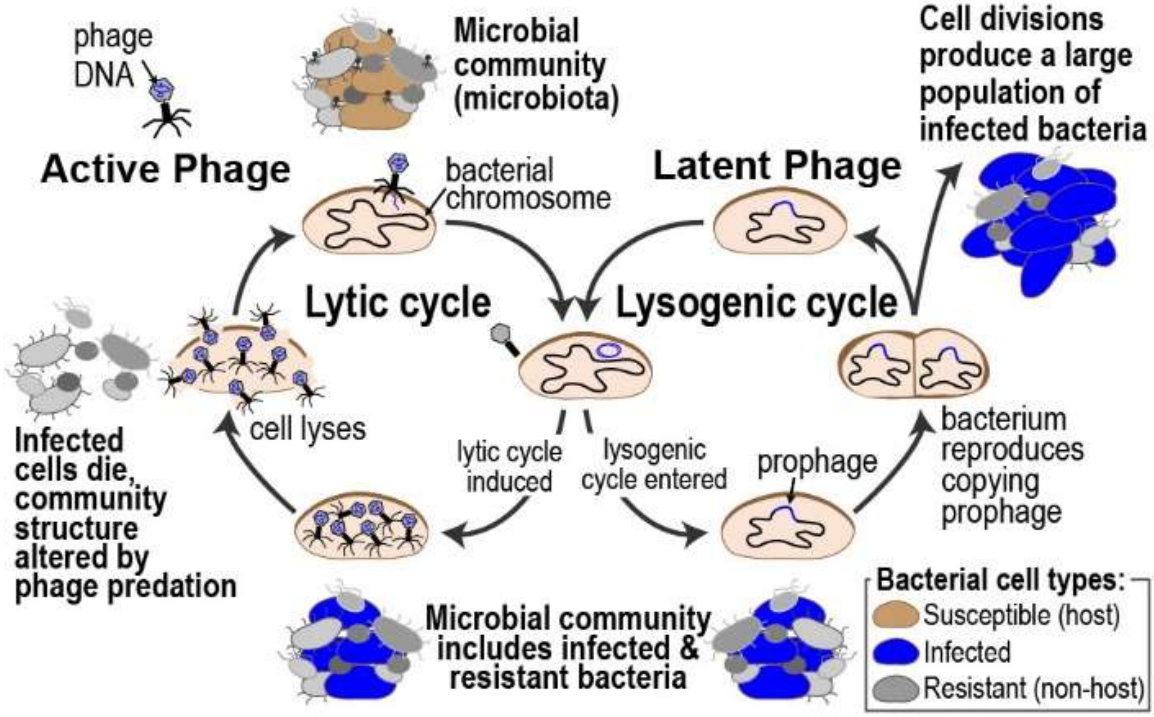

Figure 1. Diagram illustrating the lytic and lysogenic life cycles of phage.

Most phage are considered virulent, meaning they are strictly lytic. Following genome injection, virulent phage quickly hijack the host cell's machinery to produce the components of phage progeny. Phage begin their replication by directly altering their host's metabolism, selectively activating and repressing host proteins to maximize production of phage components

(Van den Bossche *et al*. 2016). Once the structural proteins of phage progeny have assembled, a copy of the phage genome is then packaged in the newly formed capsid. Within phage order *Caudovirales*, Myoviruses (family *Myoviridae*) form capsids by specific cleavage of structural proteins performed by a phage encoded prohead protease which causes the capsid to shift from a rounded shape to its mature icosahedral form (Thomas *et al*. 2012, Lecoutere *et al*. 2009). After generating numerous progeny, virulent phage complete their life cycle by lysing, or bursting, the host cell thus releasing progeny into the extracellular environment where the next phage generation may find suitable new hosts in which to restart the process.

Temperate phage, the second most common phage type, use a lysogenic life cycle wherein the phage alternates between lytic and lysogenic states. After injection, temperate phage either immediately begin the lytic cycle or initiate lysogenic infection. When entering a lysogenic state, phage typically splice their genomes into the genomes of their hosts. Once the phage genome is incorporated into the host genome, the phage is referred to as a prophage and has very limited gene expression (De Paepe *et al*. 2014). The prophage is propagated as the host cell undergoes divisions, being replicated and vertically passed to host progeny. Prophage may be triggered (induced) to enter the lytic cycle either spontaneously or by some environmental stressor on the host such as nutrient deprivation or exposure to antibiotics (Wagner & Waldor 2002). Once induced, the prophage excises itself from the host genome and enters the lytic cycle.

Phage Ecology

As vehicles of horizontal gene transfer and natural bactericidal agents, phage are major drivers of diversity for their surrounding microbiota. Phage and host have a predator/prey relationship: as one population grows, the other declines. In natural systems, phage are thought to adhere to a "Kill the Winner" growth model wherein phage replicate in the most abundant and fastest

growing host available until that host is depleted (De Paepe *et al*. 2014). This Kill the Winner system may be a major factor in the maintenance of bacterial diversity, as it never allows a single bacterial player to dominate a given ecosystem (Mills *et al.* 2013).

These interactions are not limited only to environmental systems. In recent years as more is understood about the human microbiome, it has become increasingly apparent that phage play a distinct role in maintaining the balance of human microbiota. For example, the sudden induction of prophage can cause sudden crashes of probiont populations in the human gut microbiome, opening up the floor for pathogenic bacteria and disease in a manner similar to the dysbiosis of gut microflora sometimes observed following antibiotic use (Mills *et al*. 2013). Furthermore, the first direct link between human health and regular phage activity was recently identified in a study which found disturbances of healthy core phage communities in the gastrointestinal tract of individuals with GI disease (Manrique *et al*. 2016).

Phage Genome Structure: Lifestyles of the Genetically Promiscuous

Despite their abundance and ecological importance, little is known about phage diversity and evolution. The majority of known phage have not had their genomes sequenced and of those that have been sequenced, few have had their genomes fully annotated (Pope *et al*. 2014a). To date, there are 2,269 phage genomes in the NCBI RefSeq collection (https://www.ncbi.nlm.nih.gov/pubmed/26553804). The majority of phage sequences that are currently available are for dsDNA phage, primarily *Mycobacterium* phage (for which the largest sequence collection is available) and *Escherichia coli* phage (Pope *et al*. 2014a, Krylov *et al*. 2003, Murphy *et al*. 2013). Our current understanding of phage species largely comes from laboratory studies of *E. coli* phage T4 and lambda, which have each proven to be indispensable molecular biology tools (Krylov *et al*. 2003). Expanding knowledge of phage genome structure

beyond that of a few model phage groups will be essential to fully utilizing phage as tools for molecular biology and genetic engineering.

Phage reproduce rapidly, many virulent phage cycling through hosts in a matter of minutes or hours, and with low fidelity, producing large numbers of variants with each generation as a result of replication and packaging errors. The majority of these variants likely do not survive (Pedulla *et al.* 2003). Additionally, phage undergo a tremendous amount of horizontal gene transfer with other phage, their hosts, and any environmental microbes they might otherwise encounter. Phage acquire and exchange significant portions of genetic material through both homologous and non-homologous recombination with hosts and any coinfecting phage or prophage (Murphy *et al.* 2013, Pedulla *et al*. 2003). Further, phage can act as general transducers, carrying fragmented environmental, host, or other phage's genetic material in their capsids along with their genomes and transferring it to both hosts and coinfecting phage (Wagner & Waldor 2002). Many phage possess genes sharing homology with a hugely diverse range of microbes, indicating that phage are likely migratory, moving through a range of hosts over time (Pope *et al*. 2014b, Hatfull 2015), and that phage likely gain new genetic material via horizontal gene transfer from a shared pool of genetic elements within the local microbial community which experiences regular turn over (Krylov *et al*. 2003).

Phage genomes are described as mosaic, consisting of a series of sequence cassettes that have been swapped, mutated, and reswapped many times over. Each genomic mosaic is structurally distinct, consisting of large fixed blocks containing both fixed essential and variable nonessential genes. Blocks are often only recognizable by gene order and function, sharing little or no nucleic acid or amino acid sequence similarity between phage groups (Hendrix *et al*. 1999, Hertveldt *et al*. 2005). Although illegitimate recombination occurs throughout phage genomes on

a grand scale, work in lambda phage indicates that only phage that happen to undergo recombination at select genome regions remain functional (Pedulla *et al*. 2003). While overall genome structure in terms of functional gene size and order indicates that most phage likely descend from a common ancestor, the degree of horizontal gene transfer going on among phage makes identifying clear phylogenetic relationships between phage based on sequence homology nearly impossible (Hendrix *et al.* 1999, Krylov *et al*. 2003). Further complicating phage phylogeny, no genes are conserved amongst all phages or even amongst all viruses, unlike all cellular life.

Phage-Host Coevolution

Phage evolution is largely shaped by interactions with host species. Just as the phage is shaped by its host, the host is in turn shaped by the phage. Phage and host are locked in a vicious red queen coevolutionary dynamic: for each new defense the host develops against phage infection, the phage counters with its own resistance strategies. Due to the swift and lethal nature of phage infections, bacteria have developed a number of defenses against phage both to prevent and halt infection. Bacteria may mask or change their surface receptors to prevent phage adsorption and subsequent infection. Once phage have successfully injected their genomes, bacteria have a number of systems in place which target foreign genetic material for destruction including DNA restriction and modification systems, and systems inhibiting phage replication (Mills *et al*. 2013). Host bacteria may degrade their own tRNAs to prevent phage translation upon infection (Pope *et al.* 2014a). Many bacteria are found with CRISPR/Cas immune systems which retain horizontally transferable cassettes of DNA from previous foreign DNA exposure and mobilize Cas enzymes to destroy similar DNA when it is later encountered (Fineran & Charpentier 2012). Although phage resistance is of high importance for bacteria, greater levels of phage resistance

come at a fitness cost (Mills *et al.* 2013) so bacterial populations must strike a balance between phage resistance and overall fitness.

For every defense developed by their hosts, phage evolve new offensive strategies. Phage have a competitive advantage over their hosts from the start having all the natural characteristics of parasitic organisms: minimal survival needs, rapid reproduction, and the generation of prolific numbers of progeny (Murphy *et al*. 2013). Additionally, even the most genomically stable phage (those with dsDNA genomes) demonstrate significantly higher mutation rates than most bacteria (Mills *et al.* 2013). Thus, for every defense bacteria develop against phage invasion, by sheer chance of numbers and diversity, there is likely to be some virion among the phage population that can bypass that defense. Beyond out competing hosts in terms of their rapid evolution, many phage are found to encode genes that directly counter host immune systems such as phage encoded DNA methyltransferases which methylate specific sites in the genome, protecting them from cleavage by host restriction endonucleases. Ironically, phage likely inherited these enzymes via horizontal gene transfer from host bacteria which use methyltransferases to protect their genomes against their own restriction endonucleases (Murphy *et al*. 2013). Phage are also found to dodge host restriction endonuclease systems by simply evolving to exclude restriction sites for those endonucleases the host produces. The absence of specific restriction sites in phage has proven to be a useful clue for tracking phage's past use of alternate hosts as phage are often found to lack restriction sites for enzymes their current hosts do not produce (Krylov *et al*. 2003). Phage are also found to encode inhibitors for host RNA degradation systems that might otherwise stunt phage replication (Van den Bossche *et al.* 2016).

The coevolutionary bond between phage and host goes well beyond competition as phage's propensity for swapping genetic material is not limited to other phage, phage both give

and take genetic elements from their bacterial hosts. In fact, phage have been found to be responsible for the transduction of a number of virulence factors to otherwise harmless bacteria (Pedulla *et al*. 2003). Some prophage have been found to contribute capsid and tail fiber sequences to their hosts which enable the bacteria to newly adhere to eukaryotic tissues. Phage may also transfer phage encoded hyaluronidase genes (used to penetrate bacterial hosts) to host genomes which may allow bacteria to penetrate the extracellular matrix components of animal tissues. Phage can also facilitate the transfer of exotoxin genes and regulatory factors between bacteria, altering bacterial pathogenicity (Wagner & Waldor 2002). Phage have been found to pass cholera, Shiga, diphtheria, and Panton-Valentine toxin genes to their hosts, initiating pathogenicity in a process known as "lysogenic conversion" (De Paepe *et al*. 2014). Additionally, while there are no known phage-encoded antibiotic resistance genes, their role as general transducers permits them to ferry any such genes they might encounter between hosts (Wagner & Waldor 2002).

Codon Usage Bias

Microbes often exhibit biased codon usage wherein some synonymous codons are used preferentially over others, this is the basis of the Genome Hypothesis which states that synonymous codon usage preferences are similar within a given species' genome but vary between separate species (Ikemura 1985, West & Iglewski 1988). Codon usage biases are generally found to be similar among closely related organisms and these biases become more distinct with taxonomical distance (Ikemura 1985). Codon usage biases may evolve to optimize RNA folding energies, to maximize translational efficiency, or may in some cases simply be a byproduct of genome composition (Wua *et al*. 2004). Especially in genomes with significant GC skew (being either highly GC or AT enriched), compositional bias appears to be the predominant

determinant of genomic codon usage (Gupta & Ghosh 2001). Synonymous codon usage has significant effects on mRNA folding, splicing, and half-life as well as playing a role in sequence recognition by DNA binding proteins (Martínez *et al*. 2016).

Examination of codon usage can be highly informative with regards to gene expression level and tRNA abundance. Codon usage bias is most prevalent in highly expressed genes and conserved genes, while overall codon usage is observed to be more uniform in lowly and moderately expressed genes (Gutiérrez *et al*. 1996, Pan *et al*. 1998, Lafay *et al.* 2000, Bailly-Bechet *et al*. 2007). Comparing the level of codon usage bias present in a group of genes can thus be indicative of differences in expression level among genes as those genes with stronger biases are likely expressed at higher rates than genes demonstrating weaker biases. Variation in the degree of codon usage bias among genes with different expression levels arises as a result of the strong selective pressure on highly expressed genes to utilize codons that optimize their translational rate and efficiency, thereby reducing the metabolic cost of these genes' high expression. Similarly, strong biases occur in highly conserved genes which are under strong selection to optimize fidelity of translation, as these genes are often essential to organism survival. Translational selection is less rigorous in areas of the genome with low expression where the metabolic costs of translation are lower, allowing the effects of genetic drift and mutation to predominantly shape synonymous codon usage (Bulmer 1991).

Strong codon biases have, in most cases, been found to match cells' tRNA pools: preferred codons are those recognized by the cell's most abundant tRNA species, and rare codons correspond to less available tRNAs (Ikemura 1985, Pan *et al*. 1998, Martínez *et al*. 2016). Codons correspondent to tRNAs available in large concentrations can be translated quickly and with greater fidelity; the more frequently these types of codons are used, the faster

the average rate of elongation will be. In contrast, rare codons for which few charged tRNAs are available can cause the translational machinery to either slow or halt, significantly reducing the overall translational rate, total protein output, and can affect the final protein structure (Wua *et al*. 2004). The effects of rare codon usage can be positive in some instances, such as when slowed translation enables proper folding of regions of the polypeptide as the protein is being produced without interference from the complete polymer (Albers & Czech 2016).

Codon usage biases are thought to form primarily by selection for translational efficiency. Translation is highly metabolically expensive, consequently efficiency of translation directly corresponds with cell growth rates. Organisms growing under competitive growth conditions face tremendous selective pressure to use resources efficiently so it follows that codon usage bias is most prominent in these organisms (Lafay *et al*. 2000). The optimization of translation is a balancing act: selection favors highly biased genes using the most abundant tRNAs enabling fast and high fidelity translation whereas genetic drift pushes genes toward a state of random codon usage and lower accuracy translation (Tuller 2012). Translation may also be optimized by favoring the use of codons with intermediate codon/anticodon interaction energy in high expression genes as strong interaction energy can also stall the translation rate but weak energies reduce translational accuracy (Grosjean & Fiers 1982).

Phage's total (or near total for phage with self-encoded tRNAs) dependence on host translational machinery has a significant effect on how phage codon usage biases evolve (Cardinale *et al*. 2013). All viruses evolve under strong selective pressure to adhere to host codon usage both to evade host immunity and to ensure translational efficiency. Phage are no exception as most sequenced phage have been found to match or nearly match both their hosts' codon usage preferences and GC content (Carbone 2008, Lucks *et al*. 2008, Cardinale & Duffy

2011). Total matching of host GC content is rare, however, as phage tend to have more AT rich genomes than bacteria, a common compositional trait of parasitic DNA (Carbone 2008, Lucks *et al*. 2008). Viruses with highly mismatched codon usage compared to their host suffer fitness loss (Grosjean & Fiers 1982, Bull *et al*. 2012). Fitness loss as a result of deoptimization of virus/host codon usage correspondence can be substantial, so much so that viral codon deoptimization has been used as a method to attenuate viral strains used in vaccines (e.g., Mueller *et al*. 2006, Bull 2015). Alteration of synonymous codon usage may reduce viral fitness by several mechanisms including reducing translational efficiency, increasing the translational error rate, altering mRNA secondary structure, and modifying viral regulatory signals. The effects of codon deoptimization are individually small, but fitness loss accumulates linearly with the number of altered codons. A virus engineered with hundreds of deoptimized codons will suffer substantial fitness loss and can only regain wild type fitness through a long series of individual reversions to preferred codons. Viruses attenuated in this way are thus especially suited for vaccines as fitness recovery is expected to be too slow and gradual to allow virulence to redevelop during the vaccine exposure period (Bull 2015).

Phage, like all viruses, are under strong selection pressure to utilize codon biases similar to those of the host as these codons are optimized for the host translational machinery upon which the phage is often largely dependent (Carbone 2008). Curiously, some phage are found to have significantly different codon usage and GC content than their hosts, suggesting the level of influence of host metabolism on shaping phage genome composition varies widely among different phage types (Sahu *et al*. 2004). Disparities between phage and host codon usage may arise as a result of replication biases of phage encoded DNA polymerases as phage without their own polymerase are found to more closely match host biases than other phage groups (Kunisawa

et al. 1998) or may indicate that the phage has only recently migrated into and begun to coevolve with its current host (Sokolova *et al*. 2014). Alternatively, phage codon usage may differ from the host's in order to reduce translational competition with the host for tRNA use (Albers & Czech 2016), either by utilizing phage encoded tRNAs corresponding to preferred phage codons or evolving to favor codons infrequently used by the host. While most viruses are constrained to correspondence with host codon usage to maintain their fitness, uniquely those phage species which encode their own tRNAs may partially escape translational selection pressure by providing their own translational machinery.

Phage Encoded tRNAs

Large, dsDNA viruses, most commonly phage, are the only viruses known to encode their own tRNAs but little is known about the exact function of these genes. The most widely accepted hypothesis regarding the function of phage tRNAs is that they are used to supplement the host tRNA pool with self-encoded tRNAs corresponding to codons preferred in the phage genome which occur infrequently in the host (Albers & Czech 2016). Phage vary dramatically in their tRNA gene number, some having only one tRNA while some marine cyanophage have been discovered with as many as 41 tRNA genes (Bailly-Bechet *et al*. 2007, Albers & Czech 2016). The majority of phage with self-encoded tRNAs have only a small complement of tRNA genes, so relief from host imposed translational selection is likely only present for a few codons. However, all phage tRNAs have highly atypical structure compared to those of other organisms (Pope *et al*. 2014a) and with tRNA modification comes room for wobble base pairing (Agris *et al*. 2007, Das & Lyngdoh 2012). Wobble base pairing may allow phage with few tRNAs to translate many of their own preferred codons without having to compete for use of host tRNAs.

Aside from providing phage with some level of independence from host translational machinery, it has also been proposed that phage tRNAs may be used as insertion points for homing endonucleases (Brok-Volchanskaya *et al*. 2008). Phage tRNAs may have a role in facilitating adaptation to new hosts, providing phage with some self-selected translational machinery to work with as they evolve in a new cellular environment (Sahu *et al.* 1987, Pope *et al.* 2014b). Much of the available evidence points to phage using their tRNAs to improve their translational efficiency in a given host as many phage genomes are found to be enriched with codons for self-encoded tRNAs (Kunisawa 2000, Bailly-Bechet *et al.* 2007). Whatever their actual function, phage tRNAs appear to have functional importance as tRNA deletion has been found to reduce phage fitness (Wilson 1973, Sahu *et al.* 2004, Bailly-Bechet *et al*. 2007)

Giant *Pseudomonas* Phage phiKZ and EL

Pseudomonas aeruginosa is a Gram negative bacterium with a 6.3 megabase pair (Mbp), GC enriched genome (Grocock & Sharp 2002). It is host to a large number of known phage, the best studied of which belong to the *Phikzvirus* genus of the *Myoviridae* family. Although more studied than most *P. aeruginosa* phage, much about the genomes and coevolutionary relationship of the *Phikzviruses* with their host remains either unknown or unexamined.

Pseudomonas phage phiKZ is the largest *Phikzvirus* with a 280,334 bp dsDNA, circularly permuted genome with 306 identified open reading frames (ORFs). phiKZ is a virulent phage with an unusually large icosahedral capsid, 120-130 nm in diameter, and has a long, 185 nm contractile tail with short tail fibers (Mesyanzhinov *et al*. 2002, Krylov *et al.* 2007). It has long been known to be a general transducing phage because its large isocahedral capsid allows space for non-phage (often host) DNA (Jusupova *et al.* 1982). phiKZ shares minimal amino acid sequence similarity and almost no nucleotide sequence homology with even its closest relatives

in the *Phikzvirus* genus, which in turn are all highly diverged from all other Myoviruses at both the amino acid and nucleotide levels. phiKZ along with the other members of the *Phikzvirus* genus are thus an evolutionarily unique and distinct branch of the *Myoviridae* family (Mesyanzhinov *et al*. 2002). The phylogenetic relationships between the *Phikzviruses* are shown in Figure 2.

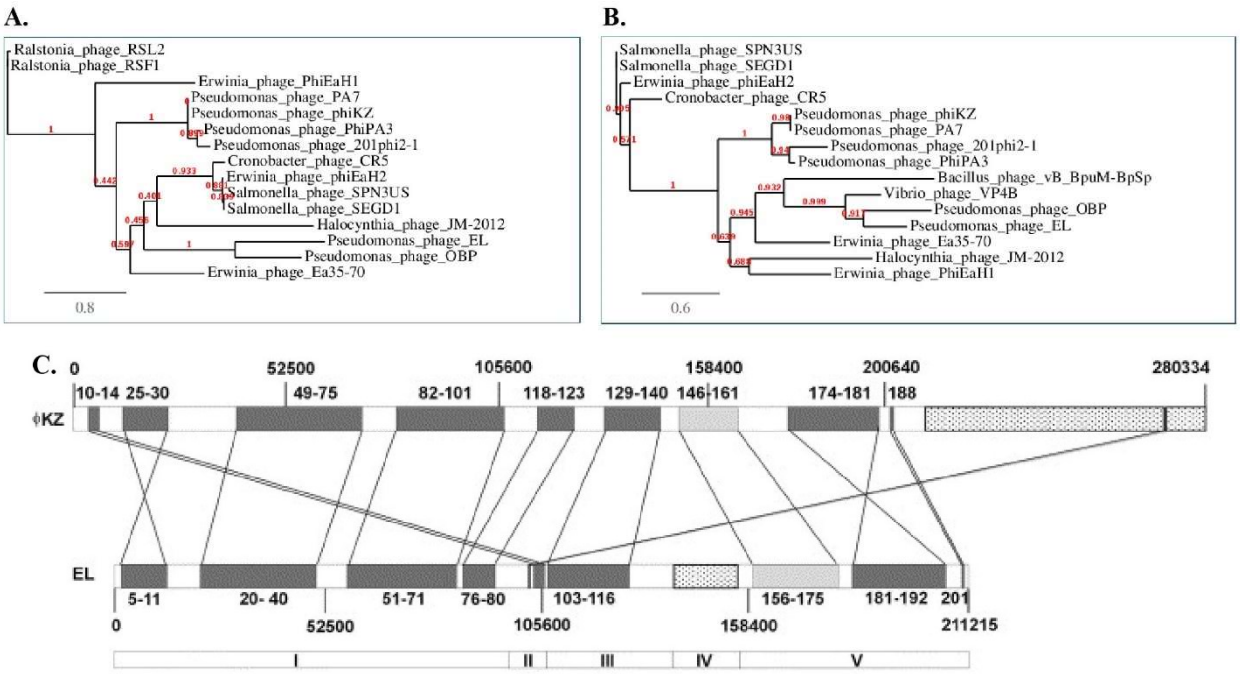

Figure 2. Genome comparisons of members of the *Phikzvirus* genus. In the two phylogenetic trees based upon **(A)** the major capsid protein and **(B)** the TerL protein, the branch lengths are proportional to the number of substitutions per site; branch support values are indicated on the branches in red (image reproduced from Adriaenssens *et al*. 2016). **(C)** Schematic comparison of the EL and phiKZ genomes (image reproduced from Hertveldt *et al*. 2005). Dark grey boxes indicate regions of proteins sharing amino acid similarity between the two genomes. Light grey boxes indicate regions sharing amino acid similarity, but with reshuffled gene order within the regions. Regions with dotted boxes are unique to the specific phage.

phiKZ is unusual among phage for a number of reasons. First, phiKZ has a highly dissimilar GC content compared to its host (Hertveldt *et al*. 2005). phiKZ's genome is highly AT enriched, having a GC content of only 36.8 % while its host has a GC content of 65-67.2%, depending on the strain (West & Iglewski 1988, Krylov *et al*. 2003). phiKZ is the only phage to

have been found to encode a multi-subunit RNA polymerase (Yakunina *et al*. 2015). Additionally, phiKZ has a rod-like protein structure in its capsid known as the inner body which sits at a 20° angle to the tail shaft. phiKZ virions with visible inner body structures are shown in Figure 3. The inner body is thought to be used as a spool for the phage's large genome and may also participate in storage of proteins used to initiate infection as the inner body is injected into the host along with the genome (Thomas *et al*. 2012, Wu *et al*. 2012). The inner body structure appears to be conserved among the *Phikzviruses* (Thomas *et al.* 2012, Sokolova *et al.* 2014).

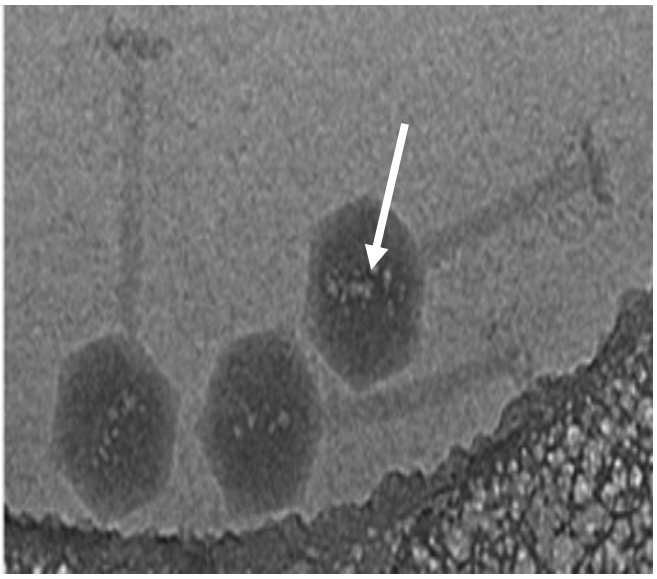

Figure 3. Scanning electron micrograph of phiKZ virions with inner body structures visible in their capsids (image reproduced from Sokolova *et al*. 2014).

Although phiKZ is one of the most studied *Pseudomonas* phage species, its genome remains poorly annotated, due in large part to its limited homology with other microbes (Hertveldt *et al*. 2005, Thomas *et al*. 2012). Of phiKZ's 306 genes, only about 30 have been functionally annotated, most of which are structural proteins that are better conserved amongst Myoviruses than other proteins (Lecoutere *et al*. 2009, Thomas *et al*. 2012, Thomas & Black 2013, Mesyanzhinov *et al*. 2002, Krylov *et al*. 2003, Kurochkina *et al.* 2009, Aksyuk *et al*. 2011,

Sycheva *et al*. 2012, Fokine *et al*. 2008, Briers *et al*. 2007, Briers *et al*. 2008, Paradis-Bleau *et al.* 2007, Yakunina *et al.* 2015).

phiKZ's endolysin, gene product (gp) 144, is among the best studied of phiKZ's proteins (Briers *et al*. 2007, Briers *et al*. 2008, Fokine *et al*. 2008, Paradis-Bleau *et al*. 2007). phiKZ's gp144 is its primary endolysin used to rupture *P. aeruginosa* cells during the lytic stage of phage infection. phiKZ also has a secondary endolysin, gp181, which is used to enzymatically penetrate its host following adsorption allowing the phage's tail tube to puncture the host and inject the phage genome (Briers *et al.* 2007, Briers *et al*. 2008, Paradis-Bleau *et al*. 2007). Phage endolysins are of special interest as purified phage endolysins can be used to clear bacterial infections in a species specific manner (Nelson *et al*. 2001). Among the other non-structural proteins of phiKZ that have been functionally annotated are its non-virion RNA polymerase proteins (gp55, gp68, gp71, gp72, gp73, gp74, and gp123) (Yakunina *et al*. 2015), the virion RNA polymerase beta subunit (gp180) (Thomas *et al*. 2012), the prohead protease (gp175) (Thomas *et al*. 2012, Thomas & Black 2013), and "Dip", a protein used to competitively inhibit host initiated RNA degradation (gp37) (Van den Bossche *et al*. 2016). The phiKZ genome also encodes for six different tRNAs which correspond to amino acids methionine, asparagine, aspartic acid, threonine, proline and leucine utilizing anticodons CAU, GUU, GUC, UGU, UGG, and UAA respectively (Mesyanzhinov *et al*. 2002) (Figure 4).

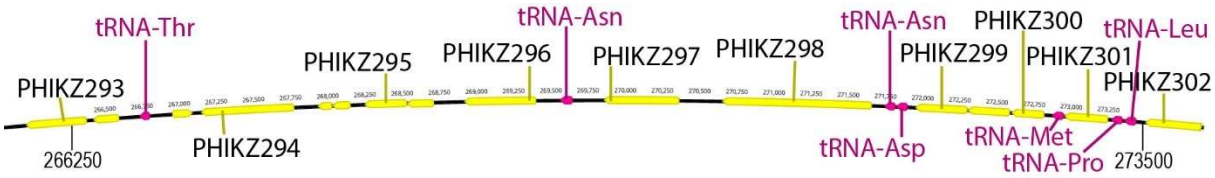

Figure 4. Location of tRNAs within the phiKZ genome sequence.

While little is known about phiKZ compared to more studied phage like T4 and lambda, even less is understood about *Pseudomonas phage EL.* EL has a 211,215 bp circularly permuted dsDNA genome with 201 ORFs (Lecoutere *et al*. 2009). EL has a capsid diameter of 140 nm and a 200 nm long tail (Hertveldt *et al.* 2005). Like phiKZ, EL is virulent and has a substantially more AT rich genome than *P. aeruginosa* with a GC content of 49.3%. EL and phiKZ are highly similar in terms of both their morphology and proteome. Further, amino acid sequence similarity and shared genomic structural motifs indicate that EL shares a recent common ancestor with phiKZ (Hertveldt *et al*. 2005, Lecoutere *et al*. 2009). Although EL has long been considered a member of the *Phikzvirus* genus, a proposal to create a new genus, *ELvirus*, with *Pseudomonas phage EL* as its type virus is currently under review with the International Committee on the Taxonomy of Viruses (Adriaenssens *et al*. 2016). As the phylogenetic trees (Figure 2A and B) for members of the *Phikzvirus* genus show, phiKZ and EL are significantly diverged from each other despite sharing similar 'blocks' of coding regions (Figure 2C).

The EL DNA polymerase, gp21, is the first DNA polymerase to be identified among the *Phikzviruses* as their polymerases are extremely diverged from all other known DNA polymerases (Kazlauskas & Venclovas 2011). Perhaps the most interesting known feature of EL's genome is gp146, an ortholog of GroEL chaperonins, the first chaperonin protein found to be encoded in a phage genome (Hertveldt *et al*. 2005, Kurochkina *et al*. 2012). The structural genes of EL have been identified experimentally by SDS-PAGE but few have been annotated (Lecoutere *et al*. 2009). The penetrative and lytic endolysins of EL, gp183 and gp181, have been identified based on amino acid sequence similarities with those of phiKZ (Briers *et al*. 2007). EL encodes a single tRNA for amino acid threonine using anticodon CGU (Hertveldt *et al.* 2005). Beyond these, almost no EL genes have been functionally annotated, both because EL has

seldom been examined and, as is also the case with phiKZ, EL's genes and proteins demonstrate little to no homology with any other known organisms.

Evaluation of the Effects of Phage tRNAs on Codon Usage

This work aims to investigate how codon usage evolves in *P. aeruginosa* large phage with respect to phage-encoded tRNAs by examining the codon usage of two of the most extensively studied *P. aeruginosa* phage, phiKZ and EL. The genomes of these phage have been fully sequenced, assembled, and sufficiently annotated to allow for comparisons to be made between genes of different types and expected levels of expression. Codon usage analysis will be used to determine whether these phage species demonstrate any apparent preference for codons corresponding to their own tRNAs and to compare the degree of translational selection for codon usage correspondence with host bacteria experienced by phage with different numbers of self-encoded tRNAs. In order to examine the functions of phage tRNAs in more detail, two mutant phiKZ strains were designed with either altered tRNA anticodons or copy numbers. Engineering of a phiKZ mutant strain with a deoptimized proline tRNA intended to observe the effects of altered self-tRNA isoreceptor type on phage fitness was started but was not successfully completed. A mutant strain of phiKZ with a tandemly duplicated region containing its aspartic acid, methionine, and proline tRNAs was engineered to increase the relative abundance of these tRNAs. Evaluation of the effects of tRNA duplication on phage codon usage following experimental evolution will help clarify the relationship between phage tRNA availability and phage codon usage.

CHAPTER TWO

METHODS

Part I: Analysis of Phage Codon Usage

Sequence Selection

The coding sequences for the RefSeq records of *Pseudomonas phage phiKZ* (Accession: NC_004629.1) and *Pseudomonas phage EL* (Accession: NC_007623.1) were manually pulled from NCBI's GenBank. Coding sequences were retained for analysis only if they exceeded 150 bp (50 codons) in length in order to avoid analyzing genes that might demonstrate codon usage biases resulting solely from small total codon counts.

In order to ensure that sequences for phiKZ were essential genes to phage fitness and therefore reflective of the phage's native codon usage, sequences for phiKZ were retained if they were either experimentally confirmed to code for functional gene products or if their protein products shared more than 90% amino acid sequence identity with two closely related phiKZ strains, *Pseudomonas phage KTN4* (Accession: KU521356.1) and *Pseudomonas phage PA7* (Accession: JX233784.1), for which complete genome sequences are available. Maintenance of these genes to such a high threshold of similarity across multiple strains was considered evidential that these genes are likely essential to phage fitness. Similar selection criteria could not be applied to EL as whole genome sequences for near EL relatives were not available.

Those sequences for EL sharing more than 90% amino acid sequence identity with *P. aeruginosa* strains were rejected as such high sequence similarity indicates that these genes were

likely horizontally acquired from the host and thus do not reflect the codon usage of the phage. No phiKZ sequences were identified with 90% or more amino acid sequence similarity with *P. aeruginosa* proteins. Sequence similarity to related phage strains and *P. aeruginosa* was evaluated from each gene product's UniProt UniRef90 database entry which lists clusters of UniProtKB sequences which share at least 90% sequence identity and 80% sequence overlap with the cluster's longest sequence. Identical sequences in UniProtKB are combined in UniRef databases so clusters included sequences from multiple, unspecified *P. aeruginosa* strains.

Sequences for both phiKZ and EL were rejected if the sequence possessed properties of mobile elements including HNH endonucleases and recognizable intein sequences in order to avoid inclusion of foreign sequences in the codon usage analysis. Mobile elements were identified previously during comparative sequence analysis of the phiKZ and EL genomes (Hertveldt *et al*. 2005).

The number of genes meeting the aforementioned selection criteria are listed in Table 1. In total 258 coding regions passed all filters for phiKZ and 166 passed all filters for EL. A complete listing of these coding regions, including their gene identification number and predicted product, is located in Appendix A (for phiKZ) and Appendix B (for EL).

Criteria for Gene Selection	**phiKZ**	**EL**
Total number of genes	306	201
Number of genes >150 bp in length	305	201
Number of experimentally verified genes	74	64
Number of genes with mobile element motifs	4	7
Number of non-experimentally verified genes sharing >90% amino acid identity with KTN4 and PA7	184	N/A
Number of genes sharing >90% amino acid identity with *P. aeruginosa*	0	28
Total number of genes retained for codon usage analysis	**258**	**166**

Table 1. Number of genes meeting the criteria for selection for codon analysis.

Codon Usage and Statistical Analyses

Codon usage and statistical analyses were performed using CodonW 1.4 (Peden 1999, Peden 2005) available from http://codonw.sourceforge.net/, except where otherwise noted. CodonW is a bioinformatics software tool that can be used to calculate codon counts, relative synonymous codon usage, codon adaptation index, effective number of codons, correspondence analysis, and a number of other analyses on codon usage for a given set of input sequences.

Relative Synonymous Codon Usage

Relative synonymous codon usage (RSCU) values are calculated as the ratio of observed codon usage to expected codon usage wherein all codons are used equally for each amino acid. RSCU values normalize codon usage across sequences which differ in amino acid composition and length. Codons with RSCU values greater than 1 are used more often than expected and codons with RSCU values less than 1 are used less frequently than expected (Sharp & Li 1987). RSCU values were calculated for both the phiKZ and EL sequence sets.

RSCU values of codon usage for 40 known high expression genes of *P. aeruginosa* PAO1 (Sharp *et al*. 2005) were previously calculated utilizing internal laboratory software developed in C++ based on the reference sequence of *P. aeruginosa* PAO1 downloaded from NCBI (GenBank: NC_002516). These RSCU values and other codon usage metrics are available through the Codon Bias Database at www.cbdb.info (Hilterbrand *et al*. 2012).

Correspondence Analysis

Correspondence analysis (COA), a form of multivariate statistical analysis, was performed for both codon counts and RSCU values of the phiKZ and EL coding sequence sets. COA plots genes along 59 continuous axes based on trends in their use of the 59 sense codons for which synonymous codons exist; thus, stop codons and codons for methionine and

tryptophan are excluded. This analysis is based on a contingency table with measures of codon usage (e.g. codon count or RSCU values) as columns and genes as rows (Peden 1999).

The 59 dimensional plot is built by measuring the variance between individual row and column values and the variance between individual values and the average row and column profiles (determined from the marginal frequency distribution over the sum of all rows or all columns) using a chi-square metric. The total variance or inertia of each row or column is calculated as the weighted sum of the chi-square distances calculated between each row value and the average row profile or between each column value and the average column profile. COA then decomposes the 59 dimensional plot produced down to four axes of variation which have maximum inertia but minimum variance between the axes and the original row and column values. These are the axes with the strongest correlation coefficients which maximize the correspondence between genes and their codon usage. Each of the four resulting axes of variation accounts for subsequently less variation than the last. Axis 1 accounts for the most variation in codon usage among the genes, axis 2 accounts for the second most variation (often substantially less than axis 1), and so on (Sourial *et al*. 2010).

COA thus effectively reduces high dimensional data down to a more interpretable, low dimensional data set explaining the variance among the original data.

Frequencies of Nucleotides at Third Codon Positions

The relative frequencies of each nucleotide at the third codon position were calculated for all analyzed genes. Calculations were made for GC_3, A_3, T_3, G_3, and C_3. GC_3 is a measure of the relative frequency of guanine and cytosine at the third codon position calculated as (G+C) / (G + C + A + T). A_3, T_3, G_3, and C_3 measure the relative frequency of adenine, thiamine, guanosine

and cytosine at the third codon position respectively. These frequencies are calculated as A / (G + C + A + T), T / (G + C + A + T), G / (G + C + A + T), and C / (G + C + A + T).

Effective Number of Codons

The effective number of codons, or N_c, values were calculated for each gene. N_c is a measurement of the non-uniformity of synonymous codon usage in a sequence (Wright 1990). N_c values range from 20-61, where 20 represents an extreme bias wherein only one codon is used for each amino acid in the gene and 61 represents a case where synonymous codons appear to be used randomly (i.e., no bias present).

Expected N_c values under random codon usage were used as a basis of comparison for the N_c values observed. Expected N_c values were calculated using the following adjusted version of Wright's formula: $N_c = 2.5 - s + (29.5 / (s^2 + (1 – s)^2)$, where $s = GC_3$. The resulting expected N_c values are more accurate and thus more closely demonstrate the relationship between observed N_c values and their expected values than those calculated by Wright's original formula (Liu 2013).

Correlation of Axis 1 Position with Third Codon Position Content

Spearman's rank correlation of genes' axis 1 positions for correspondence analysis of RSCU values with genes' axis 1 positions from correspondence analysis of codon counts, as well as genes' axis 1 positions (determined by correspondence analysis of codon counts) with GC_3, A_3, T_3, G_3, and C_3 values (Ghosh *et al*. 2000) was performed in R 3.1 (R Core Team 2013).

Part II: Engineering Mutant phiKZ Populations

Strategy I: Generating phiKZ Populations with Deoptimized Proline tRNAs

PhiKZ Growth, Harvest, and DNA Extraction

P. aeruginosa str. ATCC 15692 was obtained from ATCC (Manassas, VA). Liquid cultures were grown for a minimum of 18 h at 37°C, shaking at 120 rpm in LB broth (10 g bacto-tryptone, 5 g yeast extract, 10 g NaCl per 1 L dH_2O). Liquid cultures exceeding 3 ml in volume were grown with sterile glass beads to disrupt biofilm formation. Bacterial lawns were prepared by gently swirling together a 3:1 ratio of warm, liquid LB soft agar (SA) (LB broth + 0.7% agarose) to turbid *P. aeruginosa* liquid culture until well mixed and pouring approximately 1.5 ml of the mixture over a 1.7% agar LB plate. The mixture was quickly spread evenly using a cool, flame sterilized bacterial spreader. The plate was then allowed to solidify. *Pseudomonas phage phiKZ* (HER153) was obtained from the Félix d'Hérelle Reference Center for Bacterial Viruses (Quebec City, Canada). Liquid phiKZ was taken from stock culture (stored in 50/50 v/v glycerol at -80°C) and was spotted onto the dried lawn in 10 µl increments and allowed to dry. Five to seven spots were added per plate. Plates were incubated at 37°C (18-24 h).

PhiKZ plaques were harvested by gently removing the SA in the plaque area with a cool, flame sterilized scoopula. All plaques from a single plate were placed in a sterile 1.5 ml microcentrifuge tube filled with 0.8% saline solution. Plaque samples were then flatbed vortexed for 3 min then centrifuged for 1 min at 10,000× g. The supernatant was removed to a fresh 1.5ml microcentrifuge tube and mixed with 20 µl 10% chloroform in order to destroy any remaining host cells. PhiKZ samples were stored at 4°C.

Viral DNA was extracted from phiKZ samples using a MO Bio Laboratories UltraClean Microbial DNA Isolation Kit (Carlsbad, CA). The manufacturer's protocol was followed exactly

with the following alterations to increase DNA yield: optional heating step 3 in which samples were placed in a 70°C heating block for 10 min was used and only 30 µl (rather than the recommended 50 µl) of MD5 elution buffer (TE buffer) was used in the final protocol step. DNA samples were stored at -20°C.

PCR Amplification of Genomic Region Targeted for Mutagenesis

Due to the large size of phiKZ's genome (>280 Kbp) a subsection of phiKZ's genome was amplified by PCR for use in site directed mutagenesis. All PCR primers were designed using Primer3 (Rozen & Skaletsky 1998) based on the phikZ reference sequence retrieved from NCBI (Accession: NC_004629.1).

The following primer pair was designed: 5'-GGGGAATTCGCCACTTACCCGATGGT-3' (phiKZFUNKfor) and 5'-GGGTCTAGACCGTCGCGAAATATCCA-3' (phiKZFUNKrev). These primers amplify a 3,405 bp region (hereafter referred to as phiKZFUNK) of phiKZ's genome containing its proline, leucine, threonine and arginine tRNAs. phiKZFUNK spans base pairs 271,140 through 274,544 of phiKZ's genome. The forward primer includes an added restriction site for EcoRI (5'-GAATTC-3') and the reverse primer includes an added restriction site for XbaI (5'-TCTAGA-3'), which are underlined in each primer sequence above. These restriction sites allow phiKZFUNK to be inserted into cloning vectors. All PCR primers were synthesized by Eurofins (Huntsville, AL).

The 50 µl PCR mixture for the production of phiKZFUNK consisted of 5 µl 10x Invitrogen PCR Reaction Buffer (-MgCl2), 1 µl 10mM dNTPs, 1 µl 100µM phiKZFUNKfor primer, 1 µl 100 µM phiKZFUNKrev primer, 1 µl Invitrogen KB extender, 1.5 µl 50 mM $MgCl_2$, 1 µl template phiKZ DNA (minimum concentration 150 µg/ml), 0.2 µl Invitrogen Platinum Taq Polymerase, and nuclease-free water to 50 µl. All DNA concentrations were determined using a

Thermo Fisher Scientific (Waltham, MA) Qubit-Fluorometer and dsDNA BR Assay Kit following the manufacturer's protocol exactly.

PhiKZFUNK was generated using the following thermalcycling conditions: 94°C 5 min, 40 cycles of: 94°C 1 min, 55°C 1 min, 62.6°C 5 min, and 62.6°C 10 min. Two negative controls using nuclease-free water and whole *P. aeruginosa* cells in place of phiKZ DNA were run for the phiKZFUNK primers using the same thermalcycling conditions.

Amplification of phiKZFUNK was confirmed by running PCR products in a 1.2% agarose gel containing ethidium bromide for 45 min at 90 V with a DNA ladder, yielding a single bright band of the expected ~3,400 bp size. PCR products were purified using an EZNA Cycle Pure Spin kit (Omega Bio-tek, Norcross, GA). The kit protocol was followed exactly, with only 30 µl of elution buffer used in the final protocol step to improve DNA yield. Samples were stored at -20°C.

Cloning phiKZFUNK

The purified phiKZFUNK PCR product and pUC19 vectors were digested using Thermo Fisher Scientific (Waltham, MA) Fast Digest XbaI (5'-TCTAGA-3') and EcoRI (5'-GAATTC-3') restriction enzymes, following the manufacturer's plasmid DNA digest protocol exactly. Restriction enzymes were heat inactivated for 20 min at 65°C.

XbaI/EcoRI double digested phiKZFUNK was then ligated into XbaI/EcoRI double digested pUC19 vectors using Stratagene (San Diego, CA) T4 DNA Ligase by mixing 0.5 µl pUC19 (500 ng/µl), 1 µl phiKZFUNK (minimum concentration 100 µg/ml), 1 µl 10x Ligase Buffer, 1 µl 10mM ATP, 0.5 µl 4U/µl T4 DNA Ligase, and 6 µl of nuclease-free water. The ligation mixture was incubated for 1 h at 25°C. The ligase was heat inactivated for 10 min at 65°C.

Plasmids with successful insertions were separated from remaining ligation reactants by gel electrophoresis using a 1.2% agarose gel, run at 90 V for 45 min. Gel bands corresponding to the size of the plasmid with phiKZFUNK inserts (~6,400 bp) were cut from the gel. Gel band DNA was purified using a Thermo Fisher Scientific (Waltham, MA) GeneJET Gel Extraction Kit following the manufacturer's protocol exactly. DNA yield was improved by using only 30 µl of elution buffer in the final protocol step.

Purified pUC19 plasmids containing phiKZFUNK were transformed into JM109 chemically competent *E. coli* (Promega, Madison, WI) by the following procedure: (1) 5 µl plasmid DNA was added to 50 µl gently thawed JM109 *E. coli* cells; (2) cells were allowed to rest on ice for 30 min; (3) cells were heat shocked in a 42°C water bath for 30 s and immediately placed back on ice for 2 min; and (4) cells were added to 250 µl super optimal broth with catabolite repression (SOC) media (0.5% yeast extract, 2% bacto-tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM glucose) prewarmed to 37°C and incubated 1 h at 37°C, shaking at 120 rpm.

60 µl of the inoculated SOC was then mixed with 3 ml warm, liquid LB SA media premixed with 30 µl Ampicillin, 20 µl X-Gal and 10 µl IPTG to allow blue/white and antibiotic resistance screening of transformants. Inoculated LB SA was then spread evenly over an LB plate and allowed to dry. Transformants were allowed to grow for 18 h at 37°C. Resulting white colonies were individually collected using micropipette 200 µl capacity tips then ejected into 3 ml LB broth. Cultures were grown for 24-48 h at 37°C, shaking at 120 rpm.

Cloned vectors were extracted from transformant cultures using a Thermo Fisher Scientific (Waltham, MA) GeneJET Plasmid Miniprep Kit. The manufacturer's protocol was

followed exactly, using only 40 µl elution buffer in the final step to improve DNA yield. Samples were stored at -20°C.

Site Directed Mutagenesis

Miniprep samples with a minimum concentration of 100 µl/ml were used as templates for site directed mutagenesis PCR to alter the proline tRNA anticodon. Mutagenic PCR was performed using a Thermo Fisher Scientific (Waltham, MA) Phusion Site Directed Mutagenesis Kit. Mutagenesis primers were designed in Primer3 (Rozen & Skaletsky 1998), with the third position of the anticodon manually altered from base A to base C. This alteration generated the desired point mutation in the proline tRNA. Mutagenesis primers were designed to change the proline anticodon from 5'-UGG-3' to 5'-GGG-3'. The forward primer designed was 5'-[Phosphate]GCTCCC**C**AAGCAGGCGCGCTACCA-3' (SuperProF) (the altered base is indicated) and the reverse primer designed was 5'-[Phosphate]AGACTCGAACTCCCGACATCCT-3' (SuperProR). Mutagenesis primers must be phosphorylated at their 5' ends. Primers were synthesized by Eurofins (Huntsville, AL). Mutant phiKZFUNK/pUC19 vectors generated by mutagenic PCR are hereafter referred to as SuperPro amplicons.

Mutagenesis PCRs were performed using 50 µl PCR solutions consisting of 10 µl 5x Phusion HF Buffer, 1µl 10 mM dNTPs, 1 µl 100 µM SuperProF, 1 µl 100 µM SuperProR, 1 µl template phiKZFUNK/pUC19 DNA (minimum concentration 100 µg/ml), 0.5 µl 2 U/µl Phusion II Hot Start DNA Polymerase, and 35.5 µl nuclease-free water.

SuperPro amplicons were generated using the following thermalcycling conditions: 98°C 1 min, 25 cycles of: 98°C 30 s, 60.4°C 1 min, 66.7°C 8 min, and 66.7°C for 10 min. A negative

control using only nuclease-free water in place of a DNA template using the same thermalcycling conditions was also run.

Mutagenic PCR generated both the desired 6,091 bp SuperPro amplicon and a truncated amplicon (~500 bp) when visualized in a 1.2% agarose gel. The SuperPro amplicon was isolated using a Thermo Fisher Scientific (Waltham, MA) GeneJET Gel Extraction Kit. The kit protocol was followed exactly using only 30 µl of elution buffer in the final protocol step to increase DNA yield.

SuperPro amplicons were circularized using Stratagene (San Diego, CA) T4 DNA Ligase. A 10 µl ligation solution consisting of 1 µl purified SuperPro amplicon (minimum concentration 100 µg/ml), 2 µl 10x Fast Ligation Buffer (w/ATP), and 6.5 µl of nuclease-free water was mixed. Then 0.5 µl 5U/µl T4 DNA Ligase was added and the solution was vortexed a second time. The ligation solution was incubated at 25°C for 1 h then allowed to chill on ice for 10 min before being cloned using the previously described procedure used to clone phiKZFUNK/pUC19 vectors. Twenty white colonies generated were each grown in 3 ml LB broth for 48 h at 37°C, shaking at 125 rpm. Plasmid DNA was extracted from 1 ml of each *E. coli* culture by the previously described miniprep procedure used to extract cloned phiKZFUNK/pUC19.

SuperPro minipreps were prepared for sequencing by mixing 1 µl of miniprep sample with 0.5 µl of a 100 µM sequencing primer and 3.5 µl nuclease-free water. Each miniprep sample was sequenced twice, once with the forward sequencing primer, ProleuF with sequence 5'-GAATAGCGTAAAGAGCTGTTCC-3' and a second time with the reverse sequencing primer ProleuR with sequence 5'-TGCGAAGTGACCGATGTC-3'. Sequencing primers were designed using Primer3 (Koressar & Remm 2007, Untergasser *et al*. 2012) and were synthesized by

Eurofins (Huntsville, AL). These primers cover the region of phiKZ's genome spanning base pairs 273,189 through 273,561 and fall within 500 bp of phiKZ's proline and leucine tRNAs, effectively targeting these tRNAs for Sanger sequencing with read lengths of ~1,000 bp. Sanger sequencing was performed by GeneWiz sequencing services (South Plainfield, NJ). Sequencing results confirmed that two of the twenty miniprep samples sequenced contained the desired point mutation.

The remaining *E. coli* culture with plasmids confirmed by sequencing to have the desired mutation was grown in 500 ml LB broth for 24 h at 37°C, shaking at 125 rpm in an Erlenmeyer flask. A Thermo Fisher Scientific (Waltham, MA) GeneJET Plasmid Midiprep Kit was then used to extract SuperPro vectors following the manufacturer's protocol exactly. Samples were stored at -20°C.

Incorporating the Mutant Proline tRNA into the phiKZ Genome

Several strategies for restriction digest of whole wild type phiKZ genomes and SuperPro vectors were developed which would allow the removal of the wild type phiKZ proline tRNA in one restriction fragment. This would then permit a matching restriction fragment containing the SuperPro mutant proline tRNA to be ligated into the phiKZ genome in its place. Restriction enzyme digest sites within the phiKZ genome were identified using the NEBcutter tool (Vincze *et al*. 2003).

Initial restriction digest designs aimed to functionally shred the phiKZ genome, breaking it into fifteen to twenty pieces using groups of five or more restriction enzyme types in order to generate fragment sizes near or below 20 Kbp, allowing them to be resolved by standard agarose gel electrophoresis which can resolve fragments up to approximately 20 Kbp in size. Challenges

presented by full digest with multiple restriction enzymes coupled with the necessity to accurately ligate together tens of fragments motivated pursuit of an alternative restriction design.

Restriction digest designs targeted to the region of the phiKZ genome containing its tRNAs were considered the best methods by which the wild type and mutant proline tRNAs might by exchanged using a restriction digest and ligation approach. The two targeted digest designs begin with the digestion of the phiKZ genome with its only two-cutter restriction enzyme, SgsI, splitting the genome into a 54,258 bp fragmenting containing the phage tRNAs and a 226,076 bp fragment. Mutant SuperPro vectors would also be digested with SgsI, opening an SgsI site at the left end of the phiKZFUNK insert. The two phiKZ genome fragments would then be separated from each other and purified using pulsed field gel electrophoresis (PFGE). The ~54 Kbp fragment would then be further digested by one of the following two methods to allow insertion of the SuperPro mutant genome section.

One targeted method to achieve SuperPro insertion involved digestion of the ~54 Kbp genome fragment and SgsI digested SuperPro vectors with SphI, opening a restriction site at the right end of the phiKZFUNK insert, allowing separation of the mutant inserts from their pUC19 vectors. Following SphI digest, the ~54 Kbp fragment would be broken into six restriction fragments ranging in size from 2,488 bp to 17,244 bp. Digestion at the SphI site furthest left in the ~54 Kbp fragment produces a 2,488 bp restriction fragment which would be removed and replaced with purified SgsI/SphI digested SuperPro mutant segments. All other fragments would be kept and used to ligate the phiKZ genome back together.

The second targeted method for SuperPro insertion involved serial digestion of the ~54 Kbp genome section with restriction enzymes KasI, BsiWI and MfeI. The genome fragment would first be digested with KasI and the resulting 28,303 bp and 25,955 bp fragments would be

separated by PFGE. The purified 28,303 bp fragment would then be digested with BsiWI and the resulting 4,245 bp and 24,058 bp fragments would then be separated by gel electrophoresis. Finally, the purified 4,245 bp fragment would be digested with MfeI, producing 2,120 bp and 2,715 bp fragments to be separated by agarose gel electrophoresis. The 2,120 bp fragment would be discarded and replaced with SgsI/MfeI double digested SuperPro mutant fragments. Remaining fragments would be used to ligate the phiKZ genome back together.

While the aforementioned methods were attempted, this strategy was abandoned due to complications in working with the phage that could not be overcome. Further discussion of this strategy is included in Chapter Four. An effective strategy to replace the wild type proline tRNA of phiKZ with the SuperPro mutant tRNA was not identified that could be used to finish engineering this mutant phage strain.

Strategy II: Duplicating a Subset of phiKZ tRNAs

PCR Amplification of the Genomic Region to be Duplicated

In order duplicate a region of the phiKZ genome containing three of its six tRNAs (those for aspartic acid, methionine and proline), a 1,604 bp section of this region (genome location 271,799-273,403 bp), hereafter referred to as Looper, was identified. Primers used to amplify Looper were designed using Primer3 (Koressar & Remm 2007, Untergasser *et al*. 2012). Forward primer 5'-CGACTGAGCTATCACGGAAT-3' and reverse primer 5'-GGGGGCGCGCCAATCTTCCGGCTACGGTCTT-3' were designed. The reverse primer includes an added restriction site for SgsI (5'-GGCGCGCC-3'), underlined in the primer sequence. An additional SgsI site occurs naturally at the left end of the Looper sequence. SgsI sites flanking Looper amplicons allow Looper duplicates to be inserted into the phiKZ genome

via the SgsI site in the Looper amplicon sequence. Primers were synthesized by Eurofins (Huntsville, AL).

Looper PCR reactions consisted of 5 µl 10x Invitrogen PCR buffer, 0.5 µl 100 mM dNTPs, 1 µl 100 µM forward primer, 1 µl 100 µM reverse primer, 1.5 µl 50 mM MgCl2, 1 µl template phiKZ DNA (minimum concentration 100 µg/ml), 0.2 µl 2 U/µl Invitrogen Platinum Taq Polymerase, and 35.5 µl nuclease-free water.

Looper amplicons were generated using the following thermalcycling conditions: 98°C 5 min, 40 cycles of: 98°C 1 min, 62.7°C 1 min, 72°C 4 min, and 72°C for 10 min. Two negative controls using nuclease-free water and *P. aeruginosa* DNA in place of the phiKZ DNA template were run using the same thermalcycling conditions.

Looper amplicons were confirmed by running PCR products through a 1.2% agarose gel at 90 V for 45 min. Single bands of the expected ~1,600 bp size confirmed Looper amplification. Confirmed Looper PCR products were purified using the EZNA Cycle Pure Spin kit (Omega Bio-tek, Norcross, GA), following the manufacturer's protocol exactly, using only 30 µl of elution buffer in the final protocol step to increase DNA yield.

SgsI Digest and Ligation of phiKZ Genomic DNA and Looper

phiKZ DNA (minimum concentration 100 µg/ml) and Looper (minimum concentration 100 µg/ml) were each digested with Thermo Fisher Scientific (Waltham, MA) Fast Digest SgsI following the manufacturer's plasmid DNA digest protocol. The 20 µl digest solution consisted of 2 µl 10x Fast Digest Buffer, 2 µl of either phiKZ or Looper DNA (each with a minimum concentration of 100 µg/ml), 1 µl 10 U/µl SgsI, and 15 µl nuclease-free water. The digest solution was mixed and then incubated at 37°C for 10 min. SgsI enzymes were heat deactivated at 65°C for 20 min.

SgsI digested phiKZ genomes and Looper amplicons were ligated together to generate putative mutant phage genomes in solutions consisting of 2 µl Fermentas (Waltham, MA) T4 Ligase Buffer, 1 µl 4U/µl Fermentas T4 DNA ligase, 1 µl SgsI digested Looper (489 µg/ml), 5 µl SgsI digested genomic phiKZ DNA (10.1 µg/ml), and nuclease-free water to a 20 µl volume. Ligation solutions were mixed and incubated at 25°C for 1 h. The ligase was then heat inactivated at 60°C for 10 min. Solutions were chilled on ice for 5 min then stored at -20°C.

Transformation of Looper/phiKZ Genome Ligation Products into Spheroplasts

Chemically competent *P. aeruginosa* cells (spheroplasts) were made following the protocols of Benzinger *et al*. (1971) utilizing *P. aeruginosa* grown in LB broth for 18 h at 37°C, shaking at 125 rpm. Spheroplasts were stored at -20°C prior to use.

Putative mutant phage genomes were transformed into these *P. aeruginosa* spheroplasts by the following procedure: (1) 5 µl Looper/phiKZ genome ligation product was added to 50 µl gently thawed *P. aeruginosa* spheroplasts on ice; (2) spheroplasts were rested on ice for 30 min; (3) cells were heat shocked in a 42°C water bath for 30 s and immediately placed back on ice for 2 minutes; and (4) transformed spheroplasts were placed in 500 µl of prewarmed 37°C SOC media and incubated for 3 h at 37°C, shaking at 120 rpm. One set of transformants was refrigerated at 4°C after the initial 3 h growth period and a second set grew an additional 15 h.

Putative mutant phage was extracted by lysing spheroplasts in a 1:10 solution of transformed spheroplast culture with 1% saline. The resulting lysate was then spotted onto *P. aeruginosa* bacterial lawns in 5 µl increments and allowed to dry. Lawns were grown for 18 h at 37°C. Resulting plaques were harvested by the methods described previously for harvest of wild type phiKZ. Phage lysate was then diluted by 10x dilution series from 10^1 to 10^5 and spotted onto lawns to resolve individual plaques. Forty individual plaques were isolated from these

lawns and placed in separate microcentrifuge tubes filled with 1.5 ml 0.8% saline. Tubes were flatbed vortexed for 10 min followed by centrifugation at 10000× g for 30 s. The resulting lysate was transferred to a new microcentrifuge tube then 5 µl of 10% chloroform was added. Tubes were vortexed 5 s to lyse any remaining host cells. 100 µl lysate was added to 1 ml turbid *P. aeruginosa* culture (24 h growth) in a microcentrifuge tube and incubated at 37°C, shaking at 125 rpm. After 18 h, 10 µl of 10% chloroform was added to each phage culture followed by a 5 s vortex. 250 µl of the resulting lysate was used for DNA extraction following the exact method used to extract wild type phiKZ DNA (including additional heating), as previously described.

Confirmation of the Looper Duplication by PCR

Mutant phage populations containing the desired Looper duplication were confirmed by PCR using a series of primers targeted to the specific architecture of a single Looper insertion in the phiKZ genome. Five total primers pairs were designed, two positive control primer pairs confirming the presence of phiKZ DNA and three pairs which confirmed the presence of the desired mutation. All primers were designed using Primer3 (Koressar & Remm 2007, Untergasser *et al*. 2012) and were synthesized by Eurofins (Huntsville, AL).

One control primer pair targeted phiKZ's methionine tRNA and the second targeted a phiKZ capsid protein, ORF 27. The tRNA control primers were pkzTRNA-L with sequence 5'-ACATCCTCGCCGATAAGTTC-3' and pkzTRNA-R with sequence 5'-TACGGAGTGTAGCGCAGTTG-3', with an expected amplicon size of 954 bp. The ORF 27 control primers were pkzControl-L with sequence 5'-CACGCGTGTAATCAAGACC-3'and pkzControl-R with sequence 5'-CCTACTCGTTGGCCAAGTC-3', with an expected 964 bp amplicon size.

Primers confirming the Looper duplication were designed utilizing the sequence of the region of the phiKZ genome containing the Looper sequence duplicate manually inserted at the SgsI site sequence. Confirmation primers targeted the joint sequence between the normal genome and the repeated Looper sequence. Three confirmation primer sets were designed and each functioned such that the forward and reverse primers would only generate amplicons if the desired duplication was present. If the Looper duplication were not present, the forward primer would hybridize upstream of the reverse primer and no amplification would occur.

The three sets of confirmation primers were as follows: Set 1 consisted of forward primer pkzDup1-L* with sequence 5'-CAACTGCGCTACACTCCG-3' and reverse primer pkzDup1-R with sequence 5'-GAACTTATCGGCGAGGATG-3' (expected amplicon size 680 bp), Set 2 consisted of forward primer pkzDup1-L* with sequence 5'-CAACTGCGCTACACTCCG-3' and reverse primer pkzDup1.2-R with sequence 5'-CACCTGACCGATGAAGTAACC-3' (expected amplicon size 761 bp), and Set 3 consisted of forward primer pkzDup2-L with sequence 5'-GAACTTATCGGCGAGGATG-3' and reverse primer pkzDup2-R with sequence 5'-GCCATGTGGATAGGTTACAG-3' (expected amplicon size 947 bp).

Control and confirmation primers were first tested using wild type phiKZ DNA. 50 µl PCR solutions for each primer pair consisted of 25 µl Amresco (Solon, OH) 2x Ready PCR Mix, 1 µl 100 µM forward primer, 1 µl 100 µM reverse primer, 1 µl phiKZ DNA (100 µg/ml minimum), and 22 µl nuclease-free water.

PCR was carried out using the following thermalcycling conditions: 95°C 2 min, 30 cycles of: 95°C 30 s, 55°C 30 s, and 68°C 1 min, with a final extension at 84°C for 7 min. Negative control PCRs utilizing nuclease-free water or *P. aeruginosa* DNA in place of phiKZ DNA were run using the same thermalcycling conditions.

PCR products were visualized by running products through a 1.2% agarose gel with ethidium bromide at 90 V for 30 min. Amplicons of the expected lengths appeared as bright bands for both control primer sets and no bands appeared for reactions using the three confirmation primer sets. No amplification was observed for any of the negative controls for any of the five primer pairs.

The same PCR reaction mixes and thermalcycling conditions described to test the control and confirmation primer pairs on ancestral phiKZ DNA were prepared and used to test for the presence of the Looper duplication in the DNA extracted from 40 isolated mutant phage plaques. Twenty-five of the 40 isolates were tested before the Looper mutant was successfully confirmed via PCR. Negative control PCRs using nuclease-free water or *P. aeruginosa* DNA in place of phage DNA were also run utilizing the same thermalcycling conditions.

Putative mutant phage DNA sample B1 produced amplicons of the expected size for both control primers and confirmation primers sets 1 and 3 (primer set 2 produced no amplicon), as visualized through a 1.2% agarose gel. PCR was run on B1 DNA again twice (prepared and run separately) using the same PCR solution and thermalcycling conditions along with negative controls (Figure 5). Primer set 2 consistently failed to produce an amplicon and thus may not be a compatible primer pair. Phage B1 belonged to the phage group isolated from spheroplasts grown for only 3 h in SOC media following heat shock transformation, rather than those grown for 18 h (methods described previously).

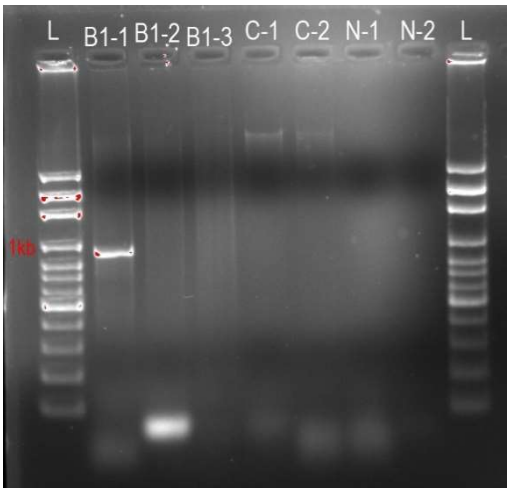

Figure 5. Agarose gel confirmation of B1 mutant for the duplication insertion within the tRNA cassette of phiKZ.

Propagation of Engineered phiKZ

In order to track the evolution of codon usage in phage B1, multiple lineages were serially transferred through naïve *P. aeruginosa* cultures. Prior to beginning propagation, ancestral phiKZ and B1 phage populations were titered. Each phage was serially diluted in a series of 12 PCR tubes containing 100 µl each of 0.8% saline. 10 µl of stock phage lysate was added to the first tube and mixed by pipetting, then 10 µl of the first dilution was added to the second tube and mixed by pipetting, and this process was repeated for the following ten tubes. 10 µl of each dilution was spotted onto *P. aeruginosa* lawns then allowed to dry. These lawns vary slightly from those described previously: 1 ml turbid *P. aeruginosa* liquid culture was added to 3 ml warm liquid LB SA and was spread evenly by oscillating the LB plate, creating a higher density lawn. Plates were incubated for 18 h at 37°C. Higher degree dilutions allowed individual plaques to be counted to quantify the number of Plaque Forming Units (PFU) in each phage stock solution.

P. aeruginosa grown for 24 h was also titered using the previously described dilution series. Each dilution was spread entirely over separate LB plates using a cooled, flame sterilized

bacterial spreader. Plates were incubated for 18 h at 37°C. Higher level dilutions allowed individual colonies to be counted in order to quantify the number of Colony Forming Units (CFU) in *P. aeruginosa* stock after 24 h growth.

Six B1 lineages were serially transferred five times by the following procedure: 1) each B1 lineage was started by inoculating 3 ml *P. aeruginosa* liquid LB culture (~10^9 CFU/ml) with 100 µl of B1 phage (~10^6 PFU/ml) for an initial multiplicity of infection (MOI) less than 0.1; 2) phage cultures were grown at 37°C, shaking at 130 rpm for 2.25 h; 3) phage cultures were then vortexed with 20 µl of 10% chloroform in order to lyse the host cells; and 4) 1 ml of the resulting lysate was then inoculated into 3 ml of naïve ~10^9 CFU/ml *P. aeruginosa* liquid culture and grown for 2.25 h at 37°C, shaking at 130 rpm. Steps 2-4 were repeated three more times. Alongside the B1 lineages, two control lines of ancestral phiKZ phage were serially transferred by the same procedure used for B1, starting with a phage concentration of ~10^5 PFU/ml. Un-inoculated *P. aeruginosa* was grown alongside each of the experimental lines as a negative control. Remaining lysate from each transfer of both B1 and wild type phiKZ was serially diluted via a 10x dilution series for 10^1 to 10^8 and spotted onto higher-density *P. aeruginosa* lawns in 10 µl increments. Dilutions were used to titer the resulting lysate and ensure that phage was still present after each transfer. All remaining lysate not used to inoculate bacteria or in dilution series was stored at 4°C.

Sequencing the Evolved Engineered phiKZ Lines

From the dilution series for each line after 5 transfers, six individual plaques were picked up using a micropipette tip and placed into a microcentrifuge tube filled with 700 µl 0.8% saline. This included six individual plaques from each of the six evolved B1 lines and each of the two evolved phiKZ ancestral control lines.

Tubes were vortexed for 10 min to release plaques from the SA into solution. 700 µl turbid *P. aeruginosa* (24 h growth) was then added to each tube. All tubes were incubated for 6-7 h at 37°C, shaking at 120 rpm. 250 µl of each phage culture was then place in a microcentrifuge tube, 10 µl of 10% chloroform was added and tubes were vortexed for 5 s. The remaining phage culture was placed back in the incubator to continue growth for a full 24 h. Phage DNA was extracted from the 250 µl samples of phage lysate by the same procedure used to isolate wild type phiKZ as described previously. Successful DNA extraction was confirmed by running DNA through a 1.2% agarose gel at 90 V for 40 min.

In order to examine the effects of the engineered duplication on phage codon usage, a set of 22 genes was identified for direct sequencing. Genes of interest were selected based on their status as either structural genes or as genes with RSCU values below 1 for codons recognized by phiKZ's aspartic acid and proline tRNAs. These genes were expected to be among the most likely candidate genes to reflect any changes in codon usage resulting from changed self-tRNA abundance produced by the Looper duplication. Twenty primers were designed to PCR amplify these genes of interest. All primers were designed using Primer3 (Koressar & Remm 2007, Untergasser *et al.* 2012) and were designed to generate similar amplicon lengths (~1,000 bp) and have compatible melting temperatures, allowing PCR with all or most primers to be run using the same thermalcycling conditions. Primer sequences and their expected amplicon sizes are listed in Table 2. Primers were synthesized by Eurofins (Huntsville, AL).

Gene	Primer Name	Genome Loc.	Ampl. Size	Primer Sequence (5'-3')
026	Orf026-L	14932-14949	1059	TCACGACGAATCCAGGTG
	Orf026-R	15974-15991		TGCAGAATGGTGAACGTG
028	Orf028-L	19298-19315	1018	TTGCATCGTCAGAACGTG
	Orf028-R	20298-20316		TGGAGCTTTGAACTCGATG
032	Orf032-L	25303-25323	1074	CCAGTCCTAGTTGCTGCGATA

Gene	Primer Name	Genome Loc.	Ampl. Size	Primer Sequence (5'-3')
	Orf032-R	26358-26377		GGAATGCTGATCACTCTGCT
035	Orf035_L	28309-28330	948	CATTATCCAGATTGGTGTATCG
	Orf035-R	29238-29257		GTTGCTTCACTTGCACCTTC
052	Orf052-L	43836-43854	1010	CAACTTGTGCCTGTTGAGC
	Orf052-R	44826-44846		AGGTGGTTGATCCTAACATCC
083	Orf083-L	80465-80483	1033	ATCCATCGTGCTAATCGTG
	Orf083-R	81481-81498		GGTGCCACATCAGAATGG
084	Orf084-L	82012-82031	1013	AGCGTGACTTAGGTCCATTG
	Orf084-R	83005-83025		TGTCATCATCAATTCCTGGTT
086	Orf086-L	83743-83760	1056	TGGTTGAAGGCGAAGAAG
	Orf086-R	84778-84799		GCTGCTAGATATTGTGCTTGAG
106-107*	Orf106-7-L	108233-108251	1084	TGATAGCAGTGGCGAAGAG
	Orf106-7-R	109300-109317		GGTACCAGCCAGCATGAG
110-111	Orf110-11-L	111285-111303	581	CTGGCAGACTGGCTATGTG
	Orf110-11-R	111847-111866		ATACCCCTTGCACCGATCAA
120	Orf120-1-L	118380-118400	1043	ACAGCGTTGTTAGTCAAGGTC
	Orf120-1-R	119408-119423		CGCCATCAGCTGAACC
	Orf120-2-L	119257-119274	1317	CGTTCTACTCGGCGGTAA
	Orf120-2-R	120557-120574		GGCATTACCGTCTTCCAA
121-122*	Orf121-22-L	120842-120862	1118	TGAAGAATTCCATCCTGTTCA
	Orf121-22-R	121943-121960		CAACCATTGCAGCAGCTT
145	Orf145-1-L	147369-147386	1110	GCAGCTGGTCAAGTCGAT
	Orf145-1-R	148461-148479		CTAACGGATAACCGCCAGT
	Orf145-2-L	148380-148398	1182	GTTGTAGAACACGCCGATG
	Orf145-2-R	149546-149562		TGCCATCCTTGATGGTG
177	Orf177-L	180657-180674	924	TGAGGTGCAACATCACCA
	Orf177-R	181564-181581		TTGTTGGCTCAGCACGTA
183	Orf183-L	197845-197867	928	TGGTGTAGGTCTAGCATTAACTG
	Orf183-R	198753-198773		TGCATAGTTACCCATCCACCA
198	Orf198-L	206204-206225	945	CAATTGCTCATTGCATTATTGA
	Orf198-R	207130-207149		TTCACAACGTCGATCTTCAC
209	Orf209-L	216756-216773	1093	AACAAGCGATTGCAGCTC
	Orf209-R	217832-217849		GAATAGCGCTGGTGGATG
217-218*	Orf217-18-L	221152-221171	1019	GGTCTTGAGCTATTCGGTTG
	Orf217-18-R	222154-222171		CCACCACCAACATTAGGC

Table 2. Primer pairs used to amplify phiKZ genes of interest for sequencing. Left primers are marked L and right primers are marked R. *Primer sets that did not amplify ancestral phiKZ DNA.

These 20 sequencing primers were tested using ancestral phiKZ DNA. Each PCR solution consisted of 25 µl Amresco (Solon, OH) 2x PCR Ready Mix, 0.5 µl 100 µM forward primer (indicated by "L" in the primer name), 0.5 µl 100 µM reverse primer (indicated by "R" in the primer name), 1 µl phage DNA, and 22 µl nuclease-free water.

All PCRs were run using the following thermalcycling conditions: 95°C for 10 min, 40 cycles of: 95° for 30 s, 54°C for 30 s, and 68°C for 1 min 20 s, and a final extension at 68°C for 5 min. Negative controls using nuclease-free water in place of DNA were run using the same thermalcycling conditions. Amplification was confirmed by running 5 µl of each PCR product through a 1.2% agarose gel. Three of the primer pairs designed (indicated in Table 2 by asterisks) did not produce amplicons and were thus removed from the set.

Primer sets 120-1 and 145-1 were used to identify three DNA samples from each lineage that contained phage DNA with sufficiently high concentration to run PCR for all other primer sets. PCR utilizing the remaining 15 primer pairs (17 pairs total per isolate covering 16 genes of interest) were run following the same reaction mix and thermalcycling conditions used for the control phiKZ DNA. Negative controls using nuclease-free water and positive controls using ancestral phiKZ DNA in place of experimental DNA were also run using the same thermalcycling conditions.

Amplification of the desired PCR products was confirmed by running all products through a 1.2% agarose gel. Confirmed PCR products were purified using an EZNA (Omega Bio-tek, Norcross, GA) Cycle Pure Spin Kit following the manufacturer's protocol exactly, using only 30 µl of elution buffer in the final protocol step to boost DNA yield.

Sanger sequencing was performed on all purified PCR samples obtained from one of the replicate B1 lineages. Sequencing was conducted by GeneWiz (South Plainfield, NJ).

Sequence Analysis

All sequences were aligned against the ancestral phiKZ genome sequence using Geneious version 10.0.09 (available at http://www.geneious.com/) to identify mutations.

CHAPTER THREE

CODON USAGE ANALYSIS: RESULTS & DISCUSSION

Codon Usage Analysis of phiKZ

The overall Relative Synonymous Codon Usage (RSCU) values were calculated for the 258 phiKZ genes selected for analysis (listed in Appendix A). These values are listed in Table 3. RSCU values are calculated as the ratio of observed codon usage to expected codon usage where RSCU values greater than 1 are indicative of codons that are used more often than expected (Sharp & Li 1987). Twenty-six of the codons within the phiKZ genes have RSCU values greater than one, all of which have A or U at their third position. This bias toward codons with A or U at the third position reflects the overall AT enrichment of the phiKZ genome (%GC = 36.8).

The codon usage of *P. aeruginosa* has previously been found to be shaped by weak translational selection. Although it is not as strongly translationally biased as other bacterial species, e.g. *E. coli*, there is evidence of selection for translational efficiency via the use of more abundant tRNAs within the coding sequences of highly expressed genes (HEGs) (Sharp *et al*. 2005). Thus correspondence between phiKZ and *P. aeruginosa*'s codon usage is expected if the phage's codon usage is under selection for translational efficiency. phiKZ's codon usage correspondence with its host appears to be weak, however. None of phiKZ's overrepresented codons are also overrepresented in the overall host genome based on RSCU values previously calculated for the overall *P. aeruginosa* genome (Grocock & Sharp 2002) and only four of phiKZ's overrepresented codons are overrepresented among *P. aeruginosa*'s HEGs. Notably, the

four codons overrepresented in both phiKZ and *P. aeruginosa*'s HEGs are the only A or U ending codons overrepresented in *P. aeruginosa*'s HEGs (indicated by '**'), suggesting that this correspondence may be an artifact of phiKZ's.

AA	Codon	N	RSCU	AA	Codon	N	RSCU
Phe	UUU	1706	1.17	Ser	UCU	1166	1.58
	UUC	1218	0.83		UCC	287	0.39
Leu	UUA	2099	2.08		UCA	1202	1.63
	UUG	412	0.41		UCG	205	0.28
	CUU	1148	1.14	Pro	CCU	1054	1.42
	CUC	300	0.3		CCC	191	0.26
	CUA	1685	1.67		CCA	1397	1.88
	CUG	400	0.4		CCG	337	0.45
Ile	AUU	3099	1.77	Thr	ACU	2281	1.98
	AUC	1432	0.82		ACC	715	0.62
	AUA	714	0.41		ACA	1331	1.15
Met	AUG	1939	1		ACG	292	0.25
Val	GUU	2189	1.75	Ala	GCU**	2068	1.82
	GUC	406	0.32		GCC	479	0.42
	GUA	1901	1.52		GCA	1665	1.46
	GUG	515	0.41		GCG	335	0.29
Tyr	UAU	2371	1.52	Cys	UGU	487	1.43
	UAC	751	0.48		UGC	194	0.57
TER	UAA	174	2.02	TER	UGA	56	0.65
	UAG	28	0.33	Trp	UGG	942	1
His	CAU	1085	1.52	Arg	CGU**	1517	2.64
	CAC	341	0.48		CGC	365	0.63
Gln	CAA	1737	1.37		CGA	490	0.85
	CAG	791	0.63		CGG	245	0.43
Asn	AAU	3268	1.47		AGA	476	0.83
	AAC	1187	0.53		AGG	358	0.62
Lys	AAA	3186	1.39	Ser	AGU	1209	1.64
	AAG	1397	0.61		AGC	359	0.49
Asp	GAU	4062	1.64	Gly	GGU**	2687	2.58
	GAC	895	0.36		GGC	544	0.52
Glu	GAA**	3484	1.52		GGA	636	0.61
	GAG	1114	0.48		GGG	305	0.29

Table 3. Pooled RSCU values and codon counts for the 258 phiKZ genes analyzed. AA indicates the encoded amino acid. N is the number of occurrences of the codon across all examined genes.

Codon Usage Variation Among phiKZ Genes

To analyze the codon usage variation among individual phiKZ genes, N_c and GC_3 values were calculated for each of the genes analyzed. N_c values ranged from 32.22, with a mean value of 43.85 (standard deviation 5.06). Low N_c values, $N_c < 40$, indicate strong codon usage bias while higher N_c values indicate weak or no codon usage bias (D'Andrea *et al*. 2011). Only 47 (18.2%) of the phiKZ genes analyzed demonstrated N_c values below 40 and none demonstrated extreme codon usage bias ($N_c < 30$). Thus, phiKZ does not demonstrate strong codon usage bias overall. The wide range and high standard deviation of phiKZ's N_c values demonstrates substantial variation in the level of codon usage bias among individual genes. GC_3 values of phiKZ ranged from 0.12-0.492 with a mean value of 0.237 (standard deviation 0.063). The phiKZ genes analyzed demonstrated a maximum GC content at their third codon positions below 50%, with 181 of the 258 genes (70%) having GC_3 values below 0.25. These GC_3 values further demonstrate that phiKZ has a strong preference for A and U ending codons.

The N_c and GC_3 values of the phiKZ genes were plotted against each other to further examine the variation in codon usage among the sequence set analyzed (Wright 1990). phiKZ's N_c versus GC_3 plot, shown in Figure 6, is superimposed over a bell curve representing the expected N_c values if codon usage were solely determined by GC_3 values, in which case codon usage is random and determined by genomic composition. Most phiKZ genes fall to the GC poor end of the plot as is expected for this GC poor gene set. Genes whose plot points deviate from (either falling above or below) the expected N_c versus GC_3 curve demonstrate codon usage biases shaped by factors beyond compositional constraint, most likely translational selection. Figure 6 shows that the majority of phiKZ genes fall on or near the expected curve, indicating that the majority of these genes' codon usage is primarily the result of compositional constraint.

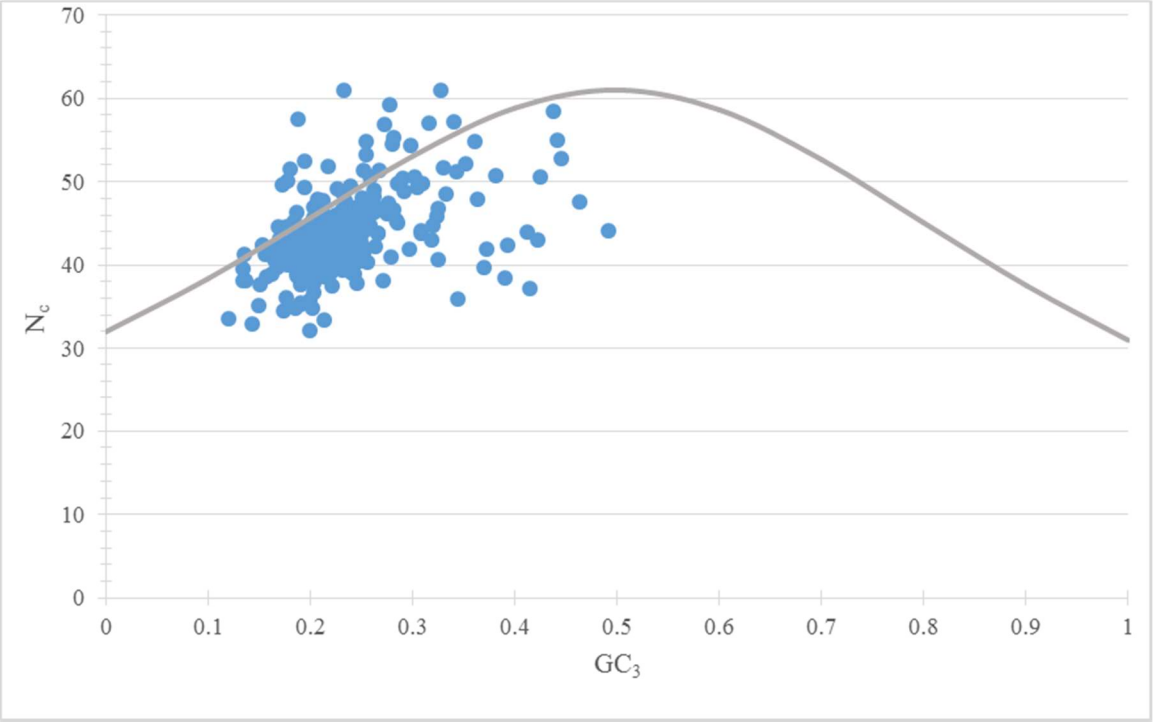

Figure 6. Plot of N_c versus GC_3 values for phiKZ genes. The continuous curve represents the expected N_c distribution given no selection (random codon usage).

A subset of genes fall above or far below the expected N_c versus GC_3 curve, indicating that these genes may be subject to stronger translational selection than those genes that fall on or near the expected curve. Genes with N_c values greater than expected and those outlying genes with N_c >10 below the expected value for their GC_3 content are considered likely to demonstrate codon usage determined more by translational selection than compositional constraint (Wright 1990, Peden 1999). Those genes with codon usage that appears to be translationally selected based on their relationship to the expected N_c distribution are listed in Table 4 along with their predicted functions. Notably, while it is expected that structural phage genes experience greater translational selection than non-structural genes due to both their expected high expression levels and conservation (Carbone 2008), only 6 of the 44 genes identified in Table 4 are known structural genes so the deviation from compositional constraint observed among these genes may result from factors outside both composition and translational selection.

Observed v. Expected N_c value	Gene	Predicted function
N_c value >10 below expected for GC_3	10	Predicted protein
	14	Predicted protein
	15	Predicted protein
	54	Predicted protein
	55	Major non-virion RNA polymerase protein
	61	Predicted protein
	71	Minor non-virion RNA polymerase protein
	110	Predicted protein
	111	Predicted protein
	115	Predicted protein
	217	Predicted protein
	218	Predicted protein
	263	Predicted protein
N_c value greater than expected for GC_3	31	Predicted protein
	53	Predicted protein
	62	Predicted protein
	79	**Capsid protein**
	81	Predicted protein
	96	**Capsid protein**
	102	Predicted protein
	132	**Tail protein**
	137	Predicted protein
	140	Predicted protein
	141	Predicted protein
	171	Predicted protein
	181	**Tail protein**
	183	Predicted protein
	194	Predicted protein
	198	Predicted protein
	207	Predicted protein
	209	Predicted protein
	210	Predicted protein
	224	**Tail protein**
	234	Predicted protein
	243	Predicted protein
	253	Predicted protein
	265	Predicted protein
	267	Predicted protein
	269	Predicted protein
	275	Predicted protein
	276	Predicted protein

	289	Predicted protein
	298	**Tail protein**
	302	Predicted protein

Table 4. phiKZ genes with N_c values indicating codon usage selection outside of compositional constraint. Structural genes and their predicted functions are shown in bold.

Correspondence Analysis of phiKZ Codon Usage

Correspondence analysis (COA) is a multivariate statistical technique frequently used to identify how the degree of codon usage bias varies among sets of genes. COA simplifies high dimensional data, like codon usage data, by reducing the data down to a more interpretable, low dimensional set demonstrating the relationships between the original data points (Suzuki *et al*. 2008). When analyzing codon usage, COA plots codon usage data along 59 axes corresponding to all sense codons with synonyms (termination codons and codons without synonyms, those for methionine and tryptophan, are excluded from analysis), to identify how codon usage varies among individual genes. The relationships among genes based on their codon usage identified from the initial 59 dimensional plot are then consolidated down to four major axes of variation. The first axis accounts for the most variation among the genes' codon usage, while each subsequent axis accounts for less and less variation. Gene sets whose first axis of variation accounts for twice as much or more variation than the second axis demonstrate substantial differences in the codon usage biases observed among individual genes and are expected to demonstrate strong translational selection for their codon usage (Peden 1999).

COA enables identification of the groups of genes within a set with the most and least biased codon usage. In species for which gene expression levels are known the relationship between genes' level of codon usage bias and expression level can be gauged. If codon usage bias is found to positively correlate with gene expression levels, then a preferred set of codons

for the genes analyzed can be calculated based on the differences in codon usage between the high and low bias gene sets identified by COA.

COA was performed for both codon count (CCOA) and RSCU values (RCOA) of phiKZ genes in order to compare results between the two and to correct for any distortion yielded by either analysis alone (Peden 1999, Perrière & Thiolouse 2002). RCOA is often preferred over CCOA for codon usage analyses because the variation among genes observed for codon counts may also reflect differences in genes' amino acid compositions and lengths. CCOA is most confounded by genes whose products are highly hydrophobic (GRAVY > 0.3), which are found to cluster together along the axes of variation apart from other genes. RCOA bypasses this issue with CCOA as RSCU values normalize the codon usage of genes for both amino acid composition and length. However, RCOA can also result in skewed codon usage comparisons. In RCOA, codons that do not appear in a gene are automatically corrected to have a non-zero RSCU value by assigning the absent codon a relative frequency of one over the number of synonymous codons available for its amino acid, causing a distortion in the axes positions of genes with absent codons (Perrière & Thioulouse 2002). This distortion primarily becomes a problem when the set of genes analyzed has an overall deficiency for one or more codon types, usually commonly rare amino acids like cysteine, skewing the positions of all codon deficient genes in the set. Performing both CCOA and RCOA then comparing their results is especially important when analyzing phage genomes as many phage genes, based solely on their small size, do not contain all codons, potentially causing data distortion for either CCOA or RCOA depending on the genome's overall codon distribution.

A plot of phiKZ gene positions on the first and second major axes of variation for CCOA is shown in Figure 7 and the same plot for RCOA is shown in Figure 8.

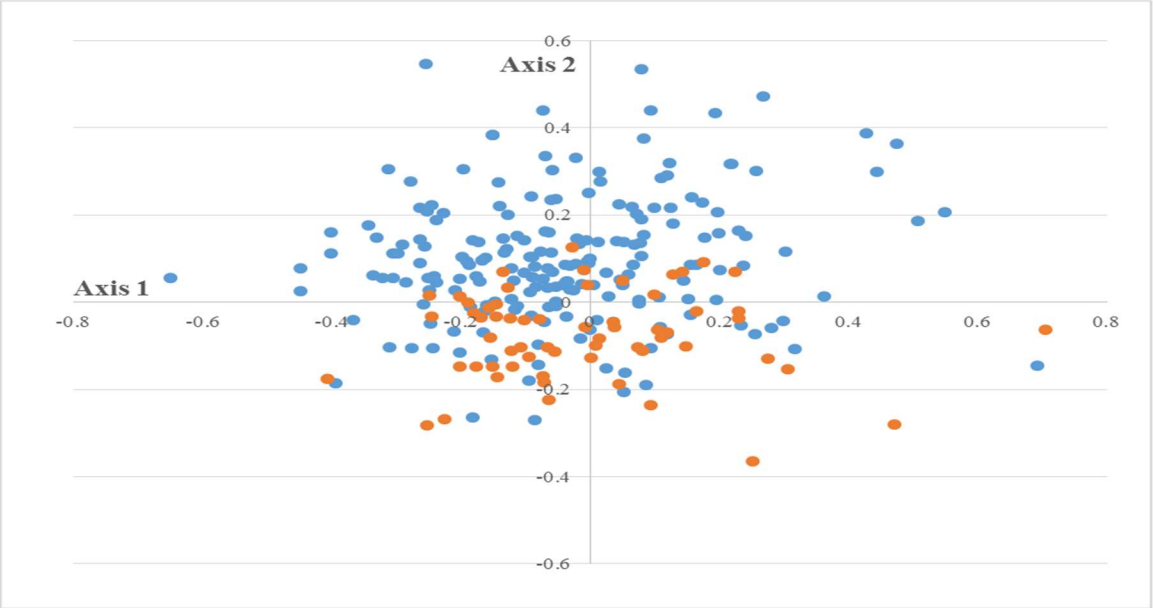

Figure 7. Plot of phiKZ genes along the first two major axes of variation for CCOA. Structural genes are shown in orange and non-structural genes are shown in blue.

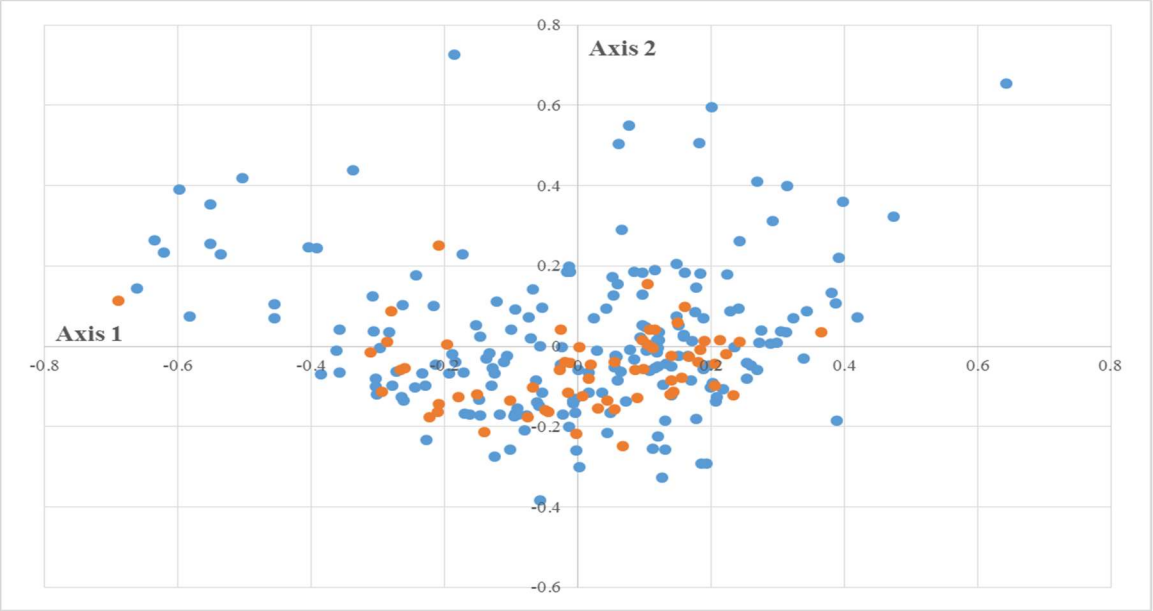

Figure 8. Plot of phiKZ genes along the first two major axes of variation for RCOA. Structural genes are shown in orange and non-structural genes are shown in blue.

CCOA and RCOA axis 1 values are found to be strongly negatively correlated with each other ($R = -0.899$, $p<0.01$), indicating that gene positions on the axes of variation for both COA types are highly similar and analysis of codon usage bias of genes based on either of these

analyses will provide similar results. As is expected for CCOA and RCOA of the same sequence set, gene positions on a given axis are mirrored between the two analyses. Cysteine codons are rare in this gene set (N < 700) with 38 genes (14.7%) having no cysteine codons at all, which is expected to cause a marked distortion in the axis positions calculated by RCOA for genes deficient in cysteine. Only 3 genes in this set have high GRAVY scores (GRAVY > 0.3), so amino acid composition effects on CCOA are expected to be minimal. Thus, for purposes of this analysis, CCOA will be used to assess the variation in codon usage bias among genes.

CCOA axis 1 accounts for 12.5% of the total variation observed for this data set and axis 2 accounts for 7% of the observed total variation. Axis 1 accounts for less than twice the amount of codon usage variation observed compared to axis 2, indicating that any translational selection experienced by this gene set is moderate or weak. Axis 1 accounts for less total variation both overall and in comparison to axis 2 than is typically observed for genomes under strong translational selection (Peden 1999). The genes of phiKZ are spread broadly across CCOA axis 1 and demonstrate minimal clustering to either the positive or negative ends of the axis indicating that the individual codon usage patterns of these genes vary substantially. Structural genes, shown in orange, demonstrate a similar pattern of spread about both axis 1 and axis 2 as non-structural genes, shown in blue, indicating that degree of codon usage bias varies as much among structural genes as it does among non-structural genes.

The orientation of genes on axis 1 is arbitrary. Genes demonstrating strong codon usage bias will fall to one end of axis 1, while genes with the least biased codon usage will fall to the other end. The average N_c values of the 20% (N=51) of genes farthest left and the 20% of genes farthest right on axis 1 were compared to determine which end of axis 1 demonstrates greater codon usage bias. Lower N_c values indicate higher levels of codon usage bias. The genes lying to

the right of axis 1 with an average N_c of 42.64 are observed to have greater levels of codon usage bias than those to the far left of axis 1 with an average N_c value of 47.51.

The 20% of genes located on the far right of axis 1 include those phiKZ structural genes which are expected to be present in the greatest copy number in each virion including the major and minor capsid proteins (the first and second most abundant proteins of each virion), the tail sheath protein, and the more abundant inner body proteins. The orientation of these high copy number, putatively highly expressed genes on the end of axis 1 associated with greater codon usage bias may indicate that degree of codon usage bias corresponds with gene expression level in phiKZ. However, as the gene expression profile of phiKZ is not known, no definitive conclusions can be drawn about the relationship between codon usage bias and gene expression level.

Gene positions on axis 1 demonstrated no significant correlation with T_3, G_3, or C_3 values. Axis 1 gene position was found to positively correlate with GC_3 ($R=0.563$, $p<0.01$) and negatively correlate with A_3 ($R=-0.703$, $p<0.01$). These results identify a moderate trend toward greater numbers of G and C ending codons among genes to the right of axis 1 and a strong trend toward enrichment with A ending codons to the left of axis 1. A greater incidence of G and C ending codons among genes to the right of axis 1 where codon usage bias is strongest indicates that the codon usage of these genes corresponds to a greater degree with host codon usage and thus may be subject to stronger translational selection than other phiKZ genes. Absence of axis 1 correlation with T_3, G_3, C_3 and only a moderate correlation with GC_3 all point toward substantial compositional similarity among phiKZ's genes. Overall, phiKZ's codon usage appears to primarily result from strong compositional constraint to maintain a highly AT enriched genome. While translational selection appears to weakly influence codon usage among phiKZ's high bias

gene set and may account for the deviation from compositional constraint observed for those genes listed in Table 4, the effects of translational selection are either weak or confounded by compositional effects throughout the remainder of the phiKZ genome.

It is interesting to note that these findings are similar to those obtained from a previous codon usage analysis performed on phiKZ (Sau *et al*. 2005), although that study examined all 306 putative phiKZ ORFs. While in this case selective analysis of a subset of genes does not appear to have significantly altered the results obtained, great care must be taken when selecting genes for codon usage analysis, especially in phage, if the results are to accurately reflect true codon usage biases. As with any genome, phage contain mobile element sequences that do not reflect the native genome composition which must be excluded from codon usage analysis (Lafay *et al*. 2000). Further, phage genomes are spattered with recent, horizontally acquired genes which do not reflect the genome's composition as they have only been under the same selective and mutational pressures as the native genes for a short period of evolutionary time.

Codon Usage Analysis of EL

RSCU values calculated for the 166 EL genes analyzed are listed in Table 5; a listing of these 166 EL genes can be found in Appendix B. Thirty codons are overrepresented (RSCU>1) in the EL genome, eleven of which are A or U ending. Despite having a GC content of just under 50%, EL demonstrates an apparent preference for G and C ending codons. Thirteen of EL's twenty-nine overrepresented sense codons are also overrepresented in the overall *P. aeruginosa* genome. Seven of these thirteen genes are also overrepresented among the HEGs of *P. aeruginosa*. In total, seventeen of EL's overrepresented sense codons are also overrepresented among the HEGs of *P. aeruginosa*. Based on RSCU values alone, EL's codon usage appears to correspond to a significant extent with that of its host.

AA	Codon	N	RSCU	AA	Codon	N	RSCU
Phe	UUU	835	0.72	Ser	UCU	407	0.76
	UUC***	1500	1.28		UCC***	633	1.19
Leu	UUA	377	0.45		UCA	225	0.42
	UUG	1243	1.48		UCG***	556	1.04
	CUU	432	0.52	Pro	CCU	779	1.26
	CUC*	1012	1.21		CCC*	700	1.13
	CUA	216	0.26		CCA	250	0.4
	CUG***	1753	2.09		CCG***	743	1.2
Ile	AUU	1492	1.45	Thr	ACU	495	0.57
	AUC***	1441	1.4		ACC***	1791	2.07
	AUA	152	0.15		ACA	306	0.35
Met	AUG	1248	1		ACG	863	1
Val	GUU*	992	1.01	Ala	GCU	849	0.98
	GUC**	1022	1.04		GCC***	1051	1.22
	GUA	636	0.65		GCA	463	0.54
	GUG***	1292	1.31		GCG*	1086	1.26
Tyr	UAU	797	0.77	Cys	UGU	256	1.26
	UAC***	1273	1.23		UGC	149	0.74
TER	UAA	92	1.66	TER	UGA	53	0.96
	UAG	21	0.38	Trp	UGG	685	1
His	CAU	499	0.93	Arg	CGU**	925	2.03
	CAC***	576	1.07		CGC	337	0.74
Gln	CAA	1030	1.06		CGA	290	0.64
	CAG	914	0.94		CGG*	868	1.9
Asn	AAU	1154	0.8		AGA	138	0.3
	AAC***	1738	1.2		AGG	181	0.4
Lys	AAA	1508	0.96	Ser	AGU*	933	1.75
	AAG***	1618	1.04		AGC	449	0.84
Asp	GAU	1770	1.1	Gly	GGU**	1498	1.58
	GAC	1435	0.9		GGC	490	0.52
Glu	GAA**	2304	1.25		GGA	587	0.62
	GAG	1395	0.75		GGG	1210	1.28

Table 5. RSCU values for the 166 EL genes analyzed. AA indicates amino acid and N indicates the number of codons.

* Codons which have RSCU>1 for overall *P. aeruginosa* genes.

** Codons which have RSCU>1 for highly expressed *P. aeruginosa* genes.

*** Codons which have RSCU>1 for both overall and highly expressed *P. aeruginosa* genes.

Codon Usage Variation Among EL Genes

To explore variation in codon usage among individual EL genes, N_c and GC_3 values were calculated for each gene. N_c values ranged broadly from 30.66-61, with a mean value of 51.81 and standard deviation of 6.157. EL genes demonstrate a slightly wider range of N_c values than phiKZ but overall EL N_c values are similarly high. Only 9 of EL's 166 analyzed genes have N_c values below 40, indicating that the majority of EL genes demonstrate only minor to moderate levels of codon usage bias. There is a high standard deviation among the N_c values of EL genes, so while no EL genes demonstrate strong codon usage bias ($N_c < 30$), the level of codon usage bias varies substantially among individual genes.

GC_3 values of EL genes ranged between 0.442-0.618 with a mean of 0.565 and a standard deviation of 0.058. While EL's genome has a GC content of 49.3%, on average EL appears to favor G or C ending codons in more than 50% of cases. 142 of the 166 EL genes analyzed (85.5%) have GC_3 values exceeding 0.5. These findings follow from observations that high EL RSCU values were predominantly associated with G and C ending codons, further indicating that compositional constraint does not appear to be the main driving force behind EL's codon usage.

A plot of N_c versus GC_3 values for EL's genes, shown in Figure 9, was created to further examine the relationship between gene composition and EL's codon usage. The N_c versus GC_3 plot for EL is superimposed over a curve representing the expected N_c versus GC_3 plot under random codon usage allowing a comparison to be made between the observed N_c values and the expected N_c values for a given GC_3 value.

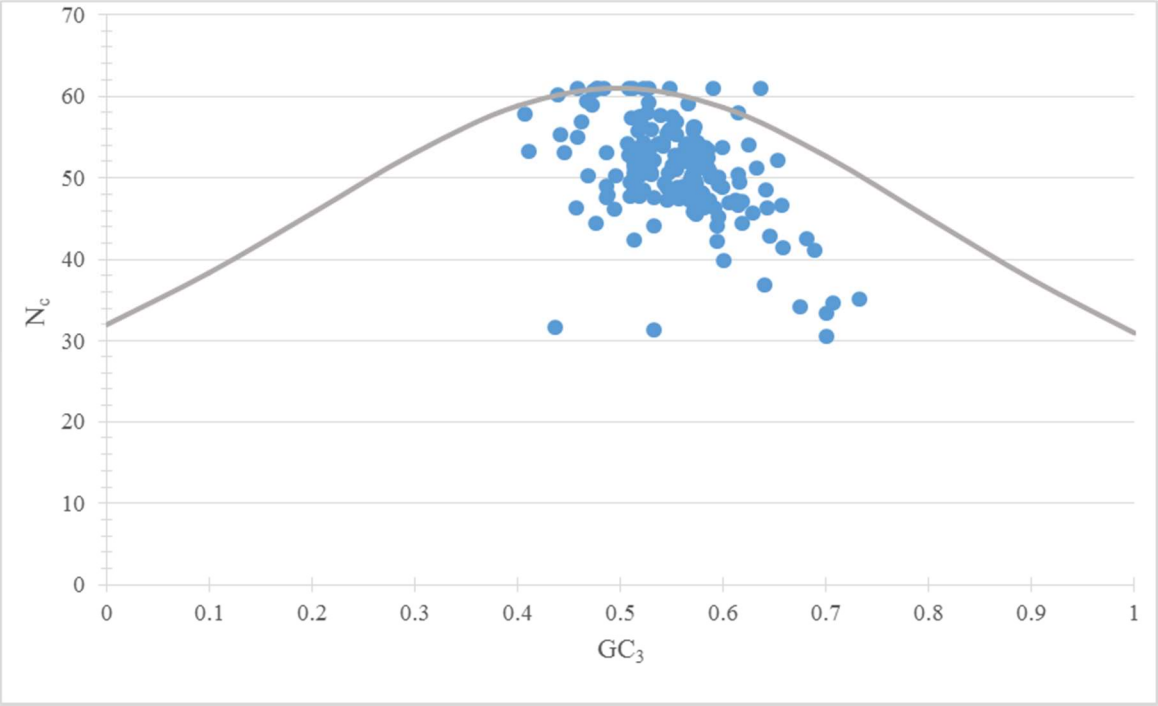

Figure 9. Plot of N_c versus GC_3 values for EL genes. The continuous curve represents the expected N_c distribution given no selection (random codon usage).

The great majority of EL genes fall far below (N_c >10 below the expected value for a given GC_3) the expected N_c versus GC_3 curve. Results of the N_c versus GC_3 plot strongly indicate that composition is not the primary determinant of EL's codon usage. These results along with observations from EL's RSCU values that a substantial proportion of EL's codon usage corresponds to the codon usage of its host's high expression genes indicate that EL's codon usage is most likely determined to a significant extent by translational selection.

Correspondence Analysis of EL Codon Usage

COA for both codon count and RSCU values was performed for the EL genes analyzed in order to further explore trends in codon usage bias among EL genes. CCOA and RCOA results for EL were compared to ensure that the trends in codon usage identified by COA accurately reflect the relationships among genes' individual codon usage and are not reflective of biases caused by gene length, amino acid composition, or rare codons. A plot of EL genes' positions on the first

and second major axes of variation identified by CCOA are shown in Figure 10 and the same plot for RCOA is shown in Figure 11.

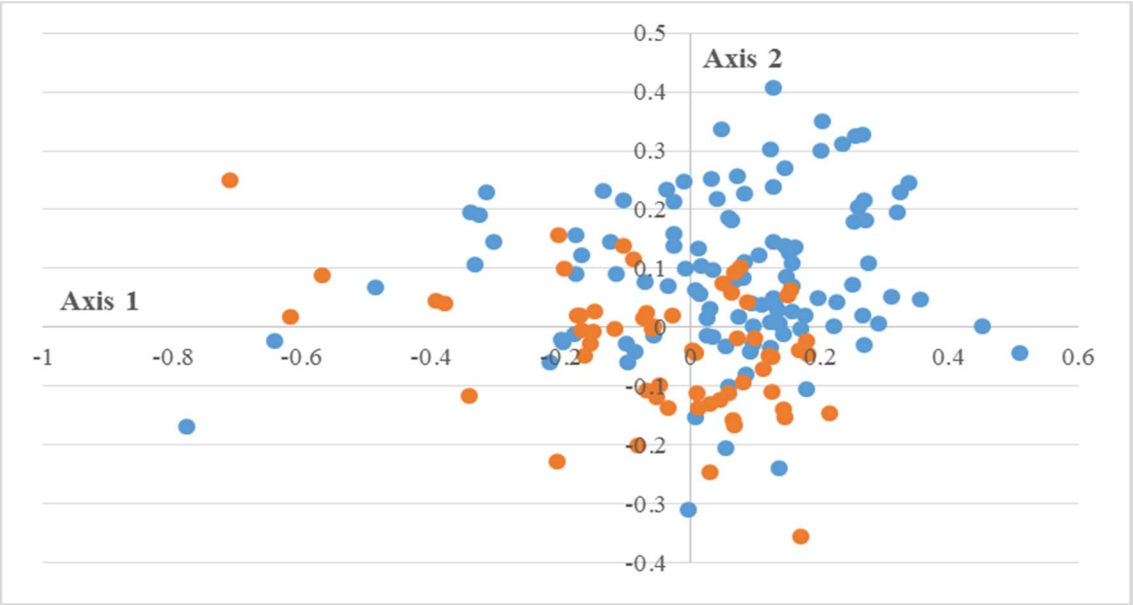

Figure 10. Plot of EL genes along the first and second major axes of variation for CCOA. Structural genes are shown in orange and non-structural genes are shown in blue.

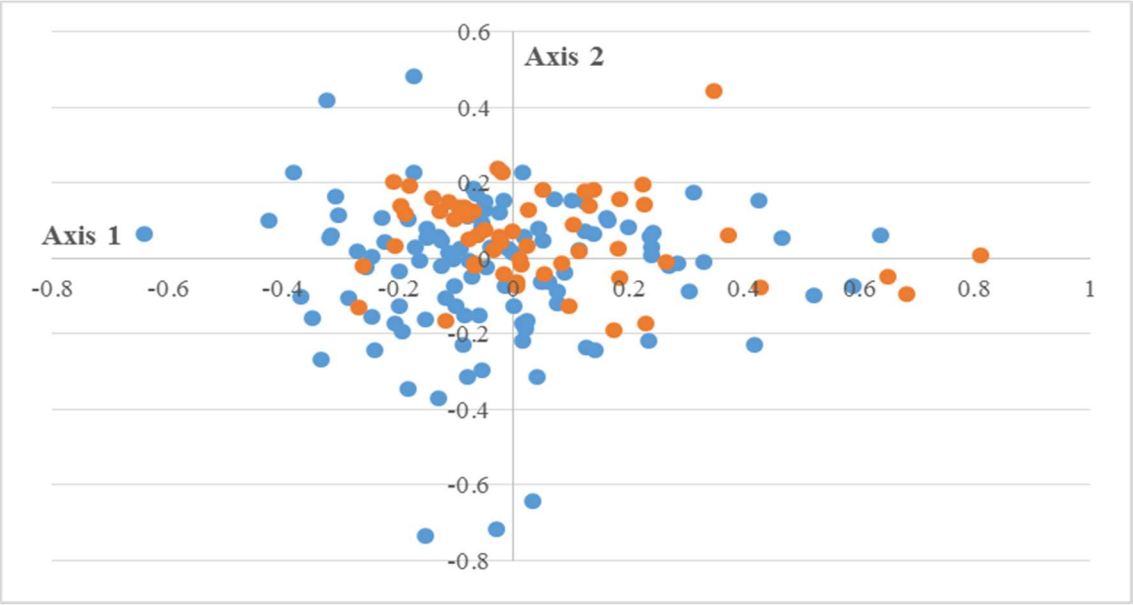

Figure 11. Plot of EL genes along the first and second major axes of variation for RCOA. Structural genes are shown in orange and non-structural genes are shown in blue.

Axis 1 values for CCOA and RCOA were found to be strongly negatively correlated with each other ($R= -0.89$, $p<0.01$), thus analysis of codon usage bias based on either of these analyses will produce similar results. As was the case for phiKZ, cysteine codons are rare in the EL genome with fewer than 450 total cysteine codons among the 166 genes analyzed. Only one EL gene was found to have a high GRAVY score (GRAVY > 0.3). EL's deficiency in cysteine codons is expected to cause a marked distortion in the axis 1 and 2 positions of cysteine deficient genes for RCOA. As the overwhelming majority of EL genes are not strongly hydrophobic, amino acid effects on gene position for CCOA are not expected to distort the relationships among genes. Therefore, all further exploration of the variation in codon usage bias among EL genes are based on CCOA.

CCOA axis 1 accounts for 14.1% of the total variation observed for the codon usage among the EL gene set while axis 2 accounts for 5.9% of the total variation observed. Axis 1 accounts for more than double the codon usage variation observed compared to axis 2, indicating that translational selection is a significant factor influencing the variation in codon usage bias observed (Peden 1999). While the EL genes are spread broadly across CCOA axis 1, indicating a high degree of variation in the degree of codon usage bias observed among individual genes, most EL genes are found to cluster on the positive end of axis 1. Structural genes are found to have a more limited range of spread across axis 1 than non-structural genes. Non-structural genes are found in a range of axis 1 positions from -0.6 – 0.5 while structural genes are found in a smaller range across axis 1 covering positions -0.6 – 0.2. Thus, structural genes appear to predominate toward the left end of axis 1.

The average N_c values of the 20% (N=33) of EL genes farthest left on axis 1 and the 20% of genes farthest right on axis 1 were compared to determine which end of axis 1 demonstrates a

higher average degree of codon usage bias. Those 20% of genes farthest left on axis 1 with an average N_c value of 44.54 were found to demonstrate greater codon usage bias than those genes farthest right on axis 1 with an average N_c of 54.66. As previously noted, EL's structural genes have a left leaning distribution on axis 1 compared to non-structural genes, indicating that in EL's case structural genes appear to demonstrate stronger codon usage bias on average than other genes. Stronger codon usage bias is expected among structural genes as they are predicted to be among the more highly expressed and conserved of phage genes. Notably, EL's major and minor capsid proteins, EL78 and EL63 respectively, are among the 20% of genes on the far left of axis 1. As was the case for phiKZ, no information is available on the expression levels of EL genes, however finding the putatively most highly expressed and conserved genes of phage EL among the high bias gene set may point toward an association between degree of codon usage bias and gene expression level. The expression profile of phage EL will need to be defined in order to draw any definitive conclusions as to the relationship between gene expression and level of codon usage bias.

The positions of EL genes along axis 1 were found to show no significant correlation with G_3 values. Gene positions on axis 1 were found to positively correlate with A_3 ($R=0.578$, $p<0.01$) and T_3 ($R=0.334$, $p<0.01$) and negatively correlate with C_3 ($R=-0.523$, $p<0.01$) and GC_3 ($R=-0.59$, $p<0.01$). These results indicate that as genes become more highly biased in their codon usage (falling more leftward on axis 1), they become more enriched in G and especially C ending codons while genes demonstrating the least codon usage bias (those farthest right on axis 1) are more enriched with A and T ending codons than other genes. High bias genes are expected to be under stronger translational selection than other genes. The trend toward greater use of G and C ending codons among EL's high bias genes further substantiates prior observations that EL's

codon usage appears to be markedly shaped by the translational influence of its host, which demonstrates a strong preference for G and C ending codons.

Comparison of phiKZ and EL Codon Usage

phiKZ and EL, despite sharing the same host and having apparently descended from a single common ancestor, have dramatically different codon usage profiles, which is not unusual as significant variability in codon usage even among more closely related phage is regularly observed (Sahu *et al*. 2005). More interestingly, based on these analyses, these phage species face very different types of selective pressure on their genome structure. phiKZ's codon usage primarily appears to be the result of compositional bias while EL's codon usage appears to be substantially influenced by host translational selection. The apparent low selective pressure to adopt host codon usage or approximate genome composition indicate that phiKZ is at least partially metabolically independent from its host in a way that EL is not. The level of disparity between phage and host codon usage has been found to positively correlate with phage tRNA copy number among *Aeromonas* phage (Prabhakaran *et al*. 2014). Such may be the case with phiKZ and EL, where phiKZ's larger set of tRNAs facilitates its discordance with host codon usage and GC content. Further, the region of phiKZ's genome containing its tRNAs accounts for the majority of the size difference between phiKZ and EL's genomes and is entirely unique to phiKZ sharing no structural or sequence similarities with any regions of the EL genome (Hertveldt *et al*. 2005). Thus the tRNA containing region of the phiKZ genome is the most probable genomic attribute to account for major differences in host based selection pressure faced by phage phiKZ and EL which otherwise have functionally highly similar genomes.

The Effects of tRNA Availability on phiKZ and EL Codon Usage

At the forefront of apparent differences between phiKZ and EL that might lead to the observed variation in their level of host dependence is their self-encoded tRNAs. EL encodes a single tRNA for threonine, while phiKZ encodes six tRNAs for amino acids methionine, asparagine, aspartic acid, threonine, proline, and leucine. While, both phage have a threonine tRNA, their anticodons correspond to different codons: codon ACA for phiKZ and ACG for phage EL. Differences in tRNA compliments between phage of similar composition is thought to result from differences in host range or growth rate (Delesalle *et al*. 2016). phiKZ and EL have nearly identical host range and growth rates so some other factor is likely to have influenced their variable tRNA acquisition and maintenance.

In order to explore the influences of both self and host translational selection on the codon usage of phage phiKZ and EL, the RSCU values calculated for the sense codons of each phage were compared to the tRNA abundances of *P. aeruginosa* (Chan & Lowe 2009) and the available phage tRNAs (each present in a single copy number). Results of this comparison are shown in Table 6. Notably, phiKZ is computationally predicted to have a second copy of its asparagine tRNA, shown in Figure 4, however to date this putative tRNA is not included in the annotation of phiKZ's genome and is therefore not considered in this analysis. For the purpose of this comparison, tRNA abundance is considered to be directly proportional to tRNA copy number, as is usually found to be the case in bacteria (Kanaya *et al.* 2001).

AA	Codon	RSCU		*P. aeruginosa* tRNA copy number	AA	Codon	RSCU		*P. aeruginosa* tRNA copy number
		phiKZ	EL				phiKZ	EL	
Phe	UUU	1.17	0.72		Ser	UCU	1.58	0.76	
	UUC	0.83	1.28	1		UCC	0.39	1.19	1
Leu	UUA*	2.08	0.45			UCA	1.63	0.42	1
	UUG	0.41	1.48	1		UCG	0.28	1.04	1
	CUU	1.14	0.52		Pro	CCU	1.42	1.26	
	CUC	0.3	1.21	1		CCC	0.26	1.13	1
	CUA	1.67	0.26	1		CCA*	1.88	0.4	1
	CUG	0.4	2.09	2		CCG	0.45	1.2	1
Ile	AUU	1.77	1.45		Thr	ACU	1.98	0.57	
	AUC	0.82	1.4	4		ACC	0.62	2.07	1
	AUA	0.41	0.15			ACA*	1.15	0.35	1
Met	AUG*	1	1	4		ACG†	0.25	1	1
Val	GUU	1.75	1.01		Ala	GCU	1.82	0.98	
	GUC	0.32	1.04	1		GCC	0.42	1.22	2
	GUA	1.52	0.65	2		GCA	1.46	0.54	4
	GUG	0.41	1.31			GCG	0.29	1.26	
Tyr	UAU	1.52	0.77		Cys	UGU	1.43	1.26	
	UAC	0.48	1.23	1		UGC	0.57	0.74	1
His	CAU	1.52	0.93		Trp	UGG	1	1	1
	CAC	0.48	1.07	2	Arg	CGU	2.64	2.03	3
Gln	CAA	1.37	1.06	1		CGC	0.63	0.74	
	CAG	0.63	0.94			CGA	0.85	0.64	
Asn	AAU	1.47	0.8			CGG	0.43	1.9	1
	AAC*	0.53	1.2	2		AGA	0.83	0.3	1
Lys	AAA	1.39	0.96	2		AGG	0.62	0.4	1
	AAG	0.61	1.04		Ser	AGU	1.64	1.75	
Asp	GAU	1.64	1.1			AGC	0.49	0.84	1
	GAC*	0.36	0.9	4	Gly	GGU	2.58	1.58	
Glu	GAA	1.52	1.25	3		GGC	0.52	0.52	3
	GAG	0.48	0.75			GGA	0.61	0.62	1
						GGG	0.29	1.28	1

Table 6. Comparison of phiKZ and EL RSCU values with tRNA copy numbers in *P. aeruginosa*. Codons indicated by an asterisk (UUA, AUG, AAC, GAC, CCA, and ACA) are recognized by phiKZ tRNAs while the codon denoted by † (ACG) is recognized by the tRNA of EL.

Twenty-six of phiKZ's codons are used more often than expected (RSCU > 1), sixteen of which are codons for which *P. aeruginosa* does not encode an isoreceptive tRNA. Additionally, seven of the phiKZ codons corresponding to the twelve most abundant *P. aeruginosa* tRNAs, those with two or more copies (excluding those for methionine), are underrepresented in phiKZ (RSCU < 1). Two of these seven codons, aspartic acid codon GAC and asparagine codon AAC, are also recognized by phiKZ's self-encoded tRNAs. These results further indicate that phiKZ's codon usage does not demonstrate either strong translational selection or codon usage correspondence with its host. Only phiKZ's leucine tRNA isoreceptor is not also produced by *P. aeruginosa*, indicating that in this case phage tRNAs do not appear to function primarily to supplement for the absence of host isoreceptive tRNAs, although this may be the underlying function of phiKZ's leucine tRNA.

Only nine of the twenty-nine overrepresented EL sense codons are recognized by tRNAs not encoded by *P. aeruginosa*, indicating that EL's codon usage corresponds to host tRNA availability to a greater extent than phiKZ's. Further, seven of the twelve codons recognized by *P. aeruginosa*'s most abundant tRNAs are overrepresented in the EL genome.

RSCU values of phiKZ and *P. aeruginosa* are compared to phiKZ's available tRNAs in Table 7. These data demonstrate that threonine, proline and leucine codons which correspond to phiKZ's tRNAs are all either underrepresented, and even rare in most cases (RSCU < 0.1), in the host genome, both overall and among HEGs. The opposite applies to phiKZ's asparagine and aspartic acid tRNAs which each correspond to codons which are overrepresented in *P. aeruginosa* but are underrepresented in phiKZ. phiKZ's proline, leucine, and threonine codons thus appear to boost translation of codons preferred in the phage but rarely used by the host.

phiKZ's asparagine and aspartic acid tRNAs have no clear relationship with phiKZ's codon usage.

Amino acid	Anticodon (5'-3')	Codon (5'-3')	Locus (in bp)	phiKZ RSCU	*P. aeruginosa* RSCU (Overall)	*P. aeruginosa* RSCU (HEG)
Threonine	UGU	ACA	266,724 - 266,797	1.15	0.08	0.06
Asparagine	GUU	AAC	269,724 - 269,644	0.53	1.72	1.83
Aspartic acid	GUC	GAC	271,828 - 271,899	0.36	1.61	1.44
Methionine	CAU	AUG	272,875 - 272,948	1	1	1
Proline	UGG	CCA	273,271 - 273,346	1.88	0.17	0.07
Leucine	UAA	UUA	273,359 - 273, 440	2.08	0.01	0.01

Table 7. phiKZ's tRNAs compared with phiKZ and *P. aeruginosa* RSCU values for codons recognized by phiKZ tRNAs. RSCU values for all *P. aeruginosa* genes and for HEGs are included.

Threonine codon ACG which corresponds to EL's self-encoded tRNA is used as often as expected in the EL genome. A comparison of EL and host RSCU values for codon ACG are shown in Table 8. These data indicate that codon ACG is underrepresented in the overall *P. aeruginosa* genome and is rare in its high expression genes while ACG occurs as often as expected in phage EL. These data indicate that EL's tRNA may play a role in supplementing the tRNA pool of the host for a codon utilized more by the phage than the host.

Amino acid	Anticodon (5'-3')	Codon (5'-3')	EL RSCU	*P. aeruginosa* RSCU (Overall)	*P. aeruginosa* RSCU (HEG)
Threonine	CGU	ACG	1	0.61	0.12

Table 8. EL's tRNA compared with EL and *P. aeruginosa* RSCU values for the codon recognized by EL's tRNA. RSCU values for all *P. aeruginosa* genes and for HEGs are included.

These data demonstrate a mixed relationship between phage codon usage and their self-encoded tRNAs. The proline, leucine and threonine tRNAs of phiKZ and the threonine tRNA of EL all appear to give the phage machinery to translate codons which are used significantly more by the phage than the host. On the other hand, phiKZ's asparagine and aspartic acid tRNAs correspond to codons which are underrepresented in phiKZ and overrepresented in the host, indicating that they do no function to optimize translation of their cognate codons. The majority of phage tRNAs examined in this work appear to function to optimize translation of codons preferentially used by the phage. Similar variable self-encoded tRNA usage has also been observed for *Mycobacterium* phage Bxz1. Most Bxz1 tRNAs recognize codons preferentially used by the phage while its other tRNAs correspond to codons used infrequently in both the phage and its host (Sahu *et al*. 2004).

The codon usage within the phiKZ genome is an outlier to the common trend in phage-host codon usage correspondence (Carbone 2008, Lucks *et al.* 2008), although similar deviations have been observed for some eukaryotic viruses (e.g., Gu *et al*. 2004). While phiKZ's discordance with its host's codon usage is perhaps an anomaly, insight can be derived from another phage with similarly mismatched phage-host codon usage – the Vibriophage KVP40 (Sau *et al*. 2007). Like phiKZ, KVP40 exhibits codon usage distinct from its *Vibrio* hosts, *V. cholera* and *V. parahaemolyticus* (Matsuzaki *et al.* 1992). Notably, however, KVP40 has a much broader known host range than phiKZ. In addition to its ability to infect multiple *Vibrio* spp., itself characteristic of a broad host-range phage, it can also infect *Photobacterium leiognathi* species (Matsuzaki *et al.* 1992). Thus, perhaps its codon usage evolved to correspond with this other host, or is an aggregate of the hosts it can infect. KVP40 encodes 25 functional tRNAs and five pseudo-tRNAs. KVP40 has clear codon usage biases both for self tRNAs as well as *Vibrio*

preferred tRNAs (Miller *et al.* 2003). This suggests that phage-encoded tRNAs may supplement host translational machinery for phages which can infect across bacterial taxa. The presence of the pseudo-tRNAs may be artifacts of past biosynthetic needs during infection of a former host taxon. Thus it is reasonable to speculate that phiKZ may have a broader host range than is currently recognized. While *P. aeruginosa* is the only host species identified for phiKZ to date, phiKZ is unable to infect all *P. aeruginosa* strains. Moreover, phiKZ and its close relative *Pseudomonas phage KTN4* (>99% genome homology) differ in their ability to infect different *P. aeruginosa* strains (Danis-Wlodarczyk *et al.* 2016). Based upon the results of the codon analysis performed here, the host-range of phiKZ is called into question. Of particular interest would be testing phiKZ's ability to infect putative bacterial host taxa selected from their codon usage correspondence with phiKZ.

CHAPTER FOUR

MUTANT PHAGE ENGINEERING AND EXPERIMENTAL EVOLUTION: RESULTS & DISCUSSION

Strategy I: Engineering of phiKZ Mutants with Deoptimized Proline tRNAs

As the codon usage analyses found, the coding regions of phiKZ clearly favor the codon CCA relative to CCC (Table 3). The phiKZ proline tRNA was thus targeted for engineering, altering the anticodon from 5'-UGG-3' to 5'-GGG-3'. Through this targeted engineering strategy, the number of cognate tRNAs for the anticodon UGG is now zero as *P. aeruginosa* spp. do not include a tRNA for this anticodon (Table 6) (Chan & Lowe 2009). The *P. aeruginosa* host also includes a tRNA for the anticodon GGG. Consequently, if the phage is under translational selection (utilizing available tRNAs), this engineered phiKZ strain will be under strong selection to favor the codon GGG.

The region containing the proline tRNA was amplified and site directed mutagenesis was successful in changing the anticodon sequence. This was confirmed by sequencing (the protocol for engineering this tRNA is described in detail within Chapter Two). A number of restriction digest designs were created in an effort to remove the section of phiKZ's genome containing the proline tRNA and replace it with the mutant version, however, the region of phiKZ's genome containing its tRNAs is bereft of restriction sites and the genome as a whole, due to its large size, has only three restriction sites that occur fewer than fifteen times, making insertion of the mutant proline tRNA by digest and ligation a design challenge.

Restriction digest with SgsI, the only two-cutter in the phiKZ genome might be the most promising option for replacing the wild type tRNA by the restriction and ligation method, as one SgsI site occurs within the SuperPro mutant subsection of the phiKZ genome (the region amplified and engineered) and the other occurs well outside the tRNA region. Digest with SgsI splits the phiKZ genome into two large pieces, one ~50 Kbp in size, containing the tRNAs, and the remainder of the genome in a ~230 Kbp section. Two digest approaches were designed for paring down the resulting ~50 Kbp fragment in order to replace the section containing phiKZ's proline tRNA with the corresponding SuperPro mutant section. Details regarding the specific protocols for each approach are detailed in Chapter Two. The first approach developed entailed successfully recovering seven distinct fragments, the majority of which would have matching SphI digested ends. Thus, the process of reassembling the genome via ligation, although feasible, is thwarted by the fact that the fragments must ligate together in the correct order and orientation in order for the phage to be viable. Therefore, the second approach was developed in which three serial digestions would be performed. Accurately ligating the phage genome back together is simplified by this approach as different restriction enzymes are used to generate each fragment, giving each fragment a unique set of ends.

Given the greater likelihood of success with this second approach, it was explored. This process, while circumventing the issues of the first approach, was highly contingent on being able to retain DNA concentrations after a series of gel purifications as each digest would need to be separated by gel and purified in preparation for the next digest/ligation. Thus, high concentration stocks of phiKZ genomic DNA were produced for the initial digest. This is itself a challenge. With each infected host cell, phiKZ typically produces relatively few progeny, an aspect of viral fitness common amongst the *Phikzviruses* (Krylov & Zhazykov 1978, Danis-

Wlodarczyk *et al*. 2016) and other large *Pseudomonas*-infecting phages (Lu *et al*. 2013). This is the typical burst size for infection at MOI < 1-5. While a higher MOI increases the likelihood of infection and thus the likelihood of more progeny, MOI > 5 has a drastic negative effect on phage viability, the exact cause of which is a matter of conjecture. At these high MOI, phiKZ transitions into a pseudolysogenic state, hypothesized to be the result of a change in the conditions in DNA polymerase activity (Krylov *et al*. 2013), severely reducing viral production. Through successive digests and purifications, which themselves lend to DNA loss, sufficient concentrations of the required fragments were unable to be produced.

A more promising and less invasive method for insertion of a point mutation such as this into a phage genome of this size is via homologous recombination. Phage are naturally prone to both homologous and non-homologous recombination events; engineered phage can even be used to induce recombination events in bacteria to generate mutations (Thomason *et al*. 2007). It follows that phage might be used to mutate itself by recombination. Homologous recombination between wild type phiKZ genomes and cloning vectors containing mutant tRNAs would be expected to occur if the two were transformed into the same host and allowed to replicate simultaneously. While recombination rates are generally infrequent (having a 5-30% chance of success) for small mutations flanked by short, homologous ends, recombination in this instance would involve swapping only a single point mutation located in a ~3 kb cloned region of the viral genome (Thomason *et al*. 2007). This approach would require that selection for antibiotic resistance be included such that only host cells including a plasmid harboring the engineered phiKZ region are viable. Next, the phage would have to be capable of infecting these host cells as changes in the efficacy of *Pseudomonas* phage has been observed as a result of the presence of plasmids (Krylov *et al*. 2013).

The greatest challenge of this technique would be identifying successfully mutated phage. Populations grown in the presence of mutant vectors would need to be serially diluted in order to isolate single PFUs. Plaques resulting from a single PFU would then be grown (ensuring the presence of a single genotype), have their DNA extracted, and the mutation would be confirmed by sequencing. Alternatively, recent attempts to incorporate fluorescent markers within phage genomes could be explored (e.g., Vinay *et al*. 2015). While this would allow for expedient recognition of engineered phage, the introduction of additional DNA could affect phage fitness via unintended consequences of engineering (e.g. altered genome packaging, genome-capsid interactions, and protein-protein interactions).

Strategy II: Duplicating a Subset of phiKZ tRNAs

Experimental Evolution of Engineered Phage

A population of phiKZ phage, B1, was engineered with a tandem duplication of a subsection of the ancestral phage genome termed Looper. The Looper duplication is functionally a 1,604 bp insertion in the phiKZ genome, giving phage populations with this mutation additional copies of phiKZ's aspartic acid, methionine, and proline tRNAs as well as phiKZ ORFs 299, 300 and 301. This duplication is expected to be at least somewhat deleterious to the phage as their increased genome size will increase replication time and has the potential to affect the phage's DNA packaging. Additionally, the Looper sequence includes three ORFs with unknown functions with confirmed expression (NCBI record Accession: NC_004629.1). The duplication of these ORFs may deleteriously affect phage fitness. The original map of the tRNA-rich region of the phiKZ genome (Figure 4) is now shown with the engineered duplicate in Figure 12. It is worth noting that the putative second asparagine tRNA is not included within the duplicated region; it occurs before the SgsI site in the amplicon.

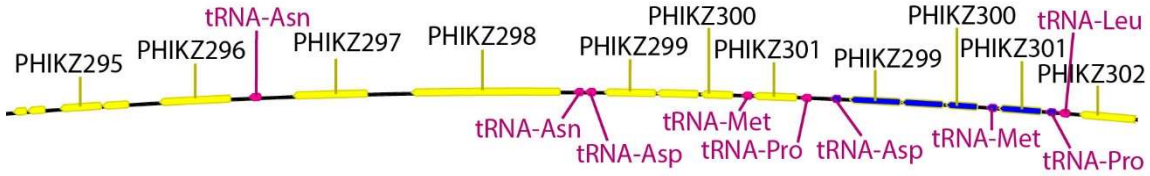

Figure 12. Engineered Looper duplication. The duplicated coding regions are indicated in blue.

In order to examine the effects of the Looper duplication on mutant phage populations, six replicate lineages of the mutant B1 phage were serially transferred through naïve *P. aeruginosa* populations in parallel with two control lineages of ancestral phiKZ. Five transfers were conducted; phage were collected after 2.25 h of co-culture. This duration was specifically selected as the ancestral phiKZ has a measured 'generation' (time from adsorption to burst) of 2.5 h (unpublished results). Transferring lysate at 2.25 h thus provides a selective advantage to phage capable of translating efficiently and quickly lysing their host. Consequently, through the experimental design itself, the phage were under selection for translational efficiency which in turn should increase the likelihood of fixing mutations which confer a codon for which there are more readily available tRNAs.

Comparative phage fitness was assessed by titering each population for all phage lineages after each transfer. Dilution series for each successive transfer of both experimental and control phage lineages were spotted onto *P. aeruginosa* lawns (see details in Chapter Two). The titer of each phage population was calculated in PFU/ml from these dilution series as a means to track phage fitness over time and to compare the fitness of B1 lineages with each other and the control lines. Changes in phage titer over the course of the selection experiment are shown in Figure 13.

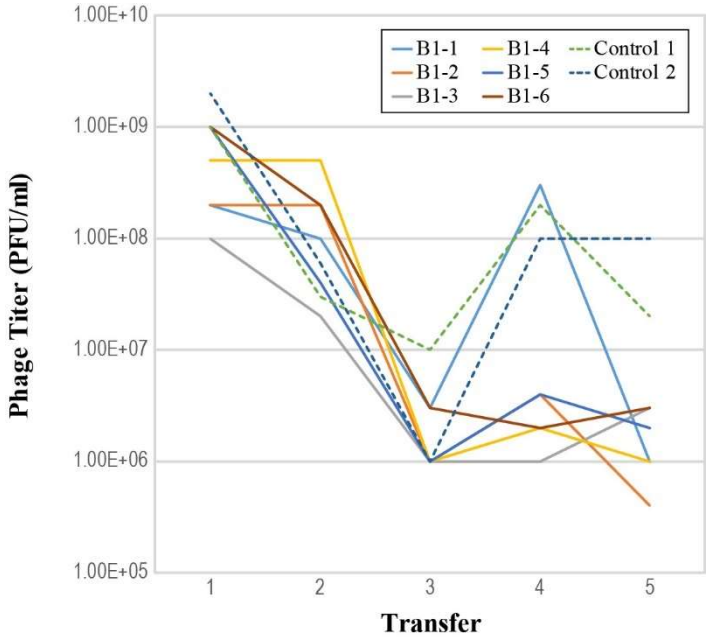

Figure 13. Changes in phage titer over five transfers.

All phage lines, both experimental and control, were found to experience a similar drop in the size of the phage population in culture following the third transfer. This drop is likely the result of the experimental parameters imposing selection for a reduced generation time. The fact that the control lines and replicate engineered lines drop at a similar rate suggests that this decrease in the phage population cannot be attributed to the engineered mutation. It is likely that the majority of phage had not yet completed lysis at the time of the second transfer, reducing third transfer population sizes. After the third transfer, however, the control lines recovered as did one of the six engineered lines, B1-1. By the next transfer, however, the engineered line population again decreased. Greater reduction in titer among all of the B1 lineages compared to the control lineages over the course of all five transfers is likely the result of reduced translational efficiency among these phage as a result of their increased genome size. Tracking the titers of B1 and ancestral phiKZ lines over more transfers would serve to further elucidate the observed trends in phage fitness over time.

Sixteen genes, listed in Table 2, which were expected to be among the most responsive genes to the engineered tRNA duplications, were PCR amplified and sequenced from DNA obtained from clonal plaques grown from isolates picked after plating the lysate of one B1 lineage collected after the fifth transfer. Genes of interest were either structural genes or genes with RSCU<1 for codons correspondent to phiKZ's aspartic acid and/or proline tRNAs. Structural genes are expected be the most influenced by changes in translational pressure as prior work found phage-host codon usage correspondence within structural genes to be greatest (Carbone 2008). Synonymous mutations which increase the number of codons in structural genes that are recognized by those tRNAs with abundances increased by the Looper mutation are expected to have a fitness benefit for the phage by increasing both the rate and accuracy of these genes' translation.

Increased abundances of phiKZ's aspartic acid and proline tRNAs are expected to promote optimization of phage genes to these more available tRNAs and to relieve any selection pressure preventing optimization of those phiKZ genes with RSCU<1 for codons corresponding to these tRNAs. The codon usage of ancestral phiKZ populations indicates that phiKZ preferentially uses codons corresponding to its proline tRNA but not its aspartic acid tRNA (Table 3). Increased abundance of phiKZ's proline tRNA in B1 mutants is expected to bolster this preference. If the phage preferentially uses its tRNAs over those of the host, then increased aspartic acid tRNA abundance in B1 is expected to promote the optimization of phage codon usage to its aspartic acid tRNA via accumulation of synonymous mutations increasing the number of codons correspondent to this tRNA.

Examination of the resulting sequence data found that no mutations occurred in any of the samples obtained from the sequenced B1 replicate. While a rapid response to genetic

engineering of codons was previously observed within 5 generations (Kula *et al*., unpublished results), this was in a significantly smaller (5kbp) and single stranded DNA virus. phiKZ's larger, dsDNA genome, given its lower mutation rate (Sanjuán *et al.* 2010), did not respond to selective pressure rapidly. Consequently, sequencing of the remaining engineered and control lineages was not performed as all lineages are expected to bear similar results.

Predicted Responses to Genomic Engineering

Newly acquired mutations in the evolving B1 populations which compensate for any deleterious effects of the Looper duplication will have a fitness advantage over competing B1 variants. Small populations of phage such as those used in the experimental evolution of B1 are extremely competitive; phage variants that occur in sufficient numbers with beneficial mutations are expected to quickly outcompete both other variants and the ancestral strain (Abedon 2008). Any synonymous mutations which increase B1's use of codons corresponding to the duplicated aspartic acid and proline tRNAs are expected to be especially beneficial as use of more abundant tRNAs is associated with increased translational speed and accuracy, which may compensate for any fitness loss experienced by increased replication time.

Selective pressure to maximize fitness following a large insertion event and the evolutionary bottlenecking of each phage population following each transfer into naïve host is expected to accelerate the evolutionary rate of B1 phage populations (Bull *et al*. 2003). Further, while there is no information available on the background mutation rates of giant dsDNA phage such as phiKZ, smaller dsDNA phage T7 (~40 Kbp genome) populations have been found to quickly accumulate mutations both under normal growth conditions (Yin 1998) and under conditions of stress (Heineman & Brown 2012). Experimental evolution of phage has primarily been performed in the small ssDNA phage phiX174 (e.g., Poon & Chao 2005, Domingo-Calap *et*

al. 2009), which has also been found to rapidly accumulate mutations within short periods of time (Bull *et al.* 2000). In contrast to phiKZ, phiX174 does not encode for its own tRNAs. Thus, in an investigation of the role of selection for translational efficiency in phiX174, engineering was performed to deoptimize the codon usage within its genes while tRNA availabilities were maintained. This is a direct reversal of what was explored here, where the codon usage was maintained but the tRNAs available were altered. Evolution of these engineered phiX174 lines saw a rapid recovery of fitness and phage-host codon compatibility (Kula *et al*., unpublished results). It is for this reason that the immediate response in phiKZ was studied here in detail.

Despite the expectation that both the experimental parameters used and selection pressure to regain fitness following a probable deleterious mutation would promote rapid fixation of beneficial synonymous mutations in the engineered phiKZ populations, no mutations were identified among sixteen genes sequenced from one of the replicate B1 lineages. These results indicate that even when grown under conditions that favor mutation, phiKZ evolves much more slowly than phage with smaller genomes. Thus, if the effects of the engineered tRNA duplication on phiKZ's codon usage are to be observed, B1 populations will likely need to be grown over the course of hundreds of generations. Although giant viruses like phiKZ have not previously been evolved experimentally, leaving their mutation rates as yet undefined, mutation rates among viruses are usually found to decline with increased genome size (Duffy *et al*. 2008); the largest of the known giant viruses (mimiviruses with ~1.2 Mega base pair dsDNA genomes) may have the lowest mutation rates among all viruses (Holmes 2011).

Future Directions

The B1 mutants engineered in this work provide the groundwork for an in-depth examination of the effects of self-encoded tRNA abundance on the evolution of phage codon

usage. The full effects of increased tRNA abundances on the codon usage of B1 may be examined by the same methods utilized here, expanding the number of replicates and transfers performed. Additionally, in depth fitness assays measuring the burst size of phage B1 and its population doubling time will provide further insights into the fitness effects of the tRNA engineering.

Importance of Examining phiKZ Evolution

Phage have become increasingly popular candidates for use either as alternatives or as additives to antibiotics in medicine, food production industries, and agriculture with the growing rise of multi antibiotic resistant bacteria. Medical use of phage therapy has become an especially popular area of phage research. The use of phage to treat infections has a number of potential benefits: phage are self-limiting, can be used locally rather than systemically, and they are host specific (Clark & March 2006). Perhaps the greatest potential benefit of phage therapy is that unlike conventional antibiotics phage can penetrate and break down biofilms (Nelson *et al*. 2001, Glonti *et al*. 2010, Mills *et al*. 2013). Biofilm infections are particularly difficult to treat act as the biofilm structure acts as a bacterial refuge that often prevents complete infection clearance, leading to chronic infections and increased antibiotic resistance among those bacteria that remain. Biofilms are notoriously difficult to treat (Donlan & Costerton 2002, Mills *et al*. 2013). Phage have been found to be able to effectively penetrate or break down biofilms and can infect both planktonic and vegetative cells (Glonti *et al*. 2010, Mills *et al*. 2013). Further, phage cocktails can be used to effectively resolve multispecies biofilm infections (Knoll & Mylonakis 2014).

However, there a number of issues with phage therapy. First, bacteria are just as likely to develop resistance to therapeutic phage mixtures as they are to antibiotics (Krylov *et al*. 2013).

Phage themselves are highly antigenic so they may induce inflammatory responses in some patients (Mills *et al.* 2013, Knoll & Mylonakis 2014, De Paepe *et al*. 2014), or may only be effective in patients for a single use without stimulating immune system clearance of the phage before it can work its way through the host population (Clark & March 2006). There is also concern that phage lysis of some bacteria may release high concentrations of bacterial endotoxins into the patient's system, leading to toxicity and inflammation, especially in immunocompromised patients (Knoll & Mylonakis 2014, Clark & March 2006). Phage themselves have been found to be immunosupressive in some cases for mammal cells, suppressing T cell and NFKB cell activation, as well as both humoral and cell mediated immune responses (Mills *et al.* 2013).

P. aeruginosa is at the forefront of proposed targets for phage therapy treatment (Glonti *et al*. 2010, Murphy *et al*. 2013). *P. aeruginosa* causes a number of pulmonary infections in immunocompromised individuals susceptible to opportunistic infections and is the predominant cause of lung infections in cystic fibrosis patients (Mesyanzhinov *et al.* 2002). *P. aeruginosa* forms mucous-like alginate biofilms that compound problems caused by the viscous, difficult to expel mucus that is phenotypic of cystic fibrosis. The mucus and biofilm act synergistically to protect the bacteria from patients' immune systems and any administered antimicrobials (Glonti *et al*. 2010). Further complicating matters, *P. aeruginosa* has become increasingly resistant to multiple antibiotics including penicillins, cephalosporins, tetraclines, and vancomycin (Mesyanzhinov *et al.* 2002, Krylov *et al.* 2013). phiKZ is widely considered the most promising candidate for use in phage therapy against *Pseudomonas* infections due to its ability to infect up to 50% of known *P. aeruginosa* strains (Krylov *et al*. 2003) and because it can be grown and purified at high titer with minimal cost (Mesyanzhinov *et al.* 2002). Recent work with a phiKZ-

like virus, OMKO1, presents an interesting phage-antibiotic combination approach of combatting *P. aeruginosa* infections in which phage selection produced an evolutionary trade-off: resistance to phage increased sensitivity to antibiotics (Chan *et al*. 2016).

The greatest issue facing phage therapy is our lack of knowledge regarding phage ecology and evolution. As previously noted, the majority of known phages remain only partially characterized and utilize only a small set of model hosts. Use of phage therapy against a broad range of pathogenic bacteria will not be feasible until more is understood about phage genome structure and how to identify and functionally annotate phage genes. While phage genomes remain cryptic and phage ecology poorly examined, large scale implementation of phage as antibacterial agents in medicine, agriculture and industry is premature at best as the effects of introducing dense phage populations into the human body or the environment are not yet reliably predictable. If phage are ever to become viable antibacterial agents, studies in phage must become broader in scope, expanding well beyond those phage occupying laboratory model hosts.

CHAPTER FIVE

CONCLUSIONS

Phage, being almost entirely dependent on their hosts' translational machinery are usually found to have similar codon usage biases as those of their hosts, although this is not always the case. Codon usage analysis of the giant *P. aeruginosa* phages *Pseudomonas phage phiKZ* and *Pseudomonas phage EL* performed here demonstrates that despite sharing a host and having similar proteomes, the codon usage of these phage is shaped by distinct forms of selection. The analyses performed here demonstrate that phiKZ's codon usage is primarily determined by its strongly AT enriched composition whereas EL's codon usage appears to be largely shaped by translational selection to use codons preferentially utilized by its host. phiKZ's greater independence from host selection pressure compared to phage EL may in part be explained by phiKZ's larger compliment of tRNAs. Indeed, phiKZ's threonine, proline and leucine tRNAs were found to correspond to codons which are preferentially used in phiKZ but which are conspicuously rare in the host. phiKZ's tRNAs alone do not account for the extreme disparity observed between its codon usage and that of *P. aeruginosa*, perhaps indicating that phiKZ's codon usage is shaped by a broader host-range than previously thought.

The phiKZ strain engineered with duplicated aspartic acid, methionine, and proline tRNAs provides a mechanism to study the effects on tRNA abundance and selection for translational efficiency. Furthermore, the selection experiment was designed to increase selection for translational efficiency by propagating only those virions which were efficient in their

infection of the host. Substantial reductions in phage titer produced by replicate B1 lineages over the course of five transfers indicate that the experimental regime was effective in increasing selection for translational efficiency. This is supported by the observation that these reductions occurred both within the control and engineered phiKZ lines. Following five transfers, no mutations were found to have occurred in phiKZ genes with codon usage expected to be especially influenced by translational selection, providing experimental evidence that giant dsDNA viruses like phiKZ have lower mutation rates than their smaller relatives. Due to phiKZ's slow evolutionary rate, evaluation of the effects of the engineered tRNA duplication will necessitate long term evolution of B1 lineages. Experimental evolution such as this has the potential to examine the unique constraints of phage-encoded tRNAs on viral fecundity. Use of *Pseudomonas phage phiKZ* is particularly attractive for such work given its potential use in phage therapy.

APPENDIX A

258 PHIKZ GENES SELECTED FOR CODON USAGE ANALYSIS

(STRUCTURAL GENES INDICATED IN BOLD)

Gene	Predicted Function	Gene	Predicted Function
1	Predicted protein	153	**Tail protein**
2	Predicted protein	157	**Tail protein**
3	Predicted protein	158	Predicted protein
4	Predicted protein	159	Predicted protein
5	Predicted protein	160	**Tail protein**
6	Predicted protein	161	Predicted protein
7	Predicted protein	162	**Inner Body protein**
8	Predicted protein	163	**Inner Body protein**
9	Predicted protein	164	**Tail protein**
10	Predicted protein	165	Predicted protein
14	Predicted protein	166	Predicted protein
15	Predicted protein	167	Predicted protein
16	Predicted protein	168	Predicted protein
17	Predicted protein	169	Predicted protein
18	Predicted protein	170	Predicted protein
19	Predicted protein	171	Predicted protein
22	Predicted protein	172	Predicted protein
25	Predicted protein	173	Predicted protein
26	**Tail protein**	174	**Tail protein**
27	**Tail protein**	175	Prohead Protease
28	**Capsid protein**	176	**Capsid protein**
29	**Tail protein**	177	**Capsid protein**
30	**Tail sheath protein**	178	**Capsid protein**
31	Predicted protein	180	Virion RNA Polymerase Beta Subunit
32	**Capsid protein**	181	**Tail protein**
33	Predicted protein	182	Predicted protein
34	**Tail tube protein**	183	Predicted protein
35	**Capsid protein**	184	**Penetrative endolysin (tail needle)**
37	RNA Degradosome Inhibitor	186	Predicted protein
39	Tubulin FtsZ / GTPase	187	Predicted protein
48	Predicted protein	188	Predicted protein
49	Predicted protein	189	Predicted protein
50	Predicted protein	190	Predicted protein
51	Predicted protein	191	Predicted protein
52	**Capsid protein**	194	Predicted protein
53	Predicted protein	195	Predicted protein
54	Predicted protein	197	Predicted protein
55	Major non-virion RNA polymerase protein	198	Predicted protein
58	Predicted protein	199	Predicted protein

Gene	Predicted Function	Gene	Predicted Function
59	Predicted protein	200	Predicted protein
60	Predicted protein	201	**Tail protein**
61	Predicted protein	202	**Tail protein**
62	Predicted protein	203	**Capsid protein**
63	Predicted protein	204	Predicted protein
64	Predicted protein	205	Predicted protein
65	Predicted protein	206	Predicted protein
66	Predicted protein	207	Predicted protein
67	Predicted protein	208	Predicted protein
68	Minor non-virion RNA polymerase protein	209	Predicted protein
69	Predicted protein	210	Predicted protein
70	Predicted protein	211	Predicted protein
71	Minor non-virion RNA polymerase protein	212	Predicted protein
72	Minor non-virion RNA polymerase protein	213	Predicted protein
73	Minor non-virion RNA polymerase protein	214	Predicted protein
74	Predicted protein	215	Predicted protein
75	Predicted protein	216	Predicted protein
77	Predicted protein	217	Predicted protein
78	Predicted protein	218	Predicted protein
79	**Capsid protein**	219	Predicted protein
80	**Capsid protein**	220	Predicted protein
81	Predicted protein	221	Predicted protein
82	Predicted protein	222	Predicted protein
83	**Capsid protein**	223	Predicted protein
84	**Capsid protein**	224	**Tail protein**
85	**Tail protein**	226	Predicted protein
86	**Capsid protein**	227	Predicted protein
87	**Tail protein**	228	Predicted protein
88	Predicted protein	229	Predicted protein
89	**Inner Body protein**	230	Predicted protein
90	**Inner Body protein**	231	Predicted protein
91	**Tail protein**	232	Predicted protein
92	**Inner Body protein**	233	Predicted protein
93	**Inner Body protein**	234	Predicted protein
94	**Inner Body protein**	235	Predicted protein
95	**Inner Body protein**	242	Predicted protein

Gene	Predicted Function	Gene	Predicted Function
96	**Capsid protein**	243	Predicted protein
97	**Inner Body protein**	244	**Tail protein**
98	Predicted protein	245	Predicted protein
99	**Capsid protein**	246	Predicted protein
100	Predicted protein	247	Predicted protein
101	**Tail protein**	248	Predicted protein
102	Predicted protein	249	Predicted protein
103	Predicted protein	250	Predicted protein
104	Predicted protein	251	Predicted protein
106	Predicted protein	252	Predicted protein
107	Predicted protein	253	Predicted protein
108	Predicted protein	254	Predicted protein
109	Predicted protein	255	Predicted protein
110	Predicted protein	256	Predicted protein
111	Predicted protein	257	Predicted protein
112	Predicted protein	258	Predicted protein
113	Predicted protein	260	Predicted protein
115	Predicted protein	262	Predicted protein
117	Predicted protein	263	Predicted protein
118	Predicted protein	264	Predicted protein
119	**Tail protein**	265	Predicted protein
120	**Major capsid protein (most abundant protein)**	266	Predicted protein
121	Predicted protein	267	Predicted protein
122	Predicted protein	268	Predicted protein
123	Major non-virion RNA polymerase protein	269	Predicted protein
124	Predicted protein	270	Predicted protein
125	Predicted protein	272	Predicted protein
126	Predicted protein	274	Predicted protein
127	**Tail protein**	275	Predicted protein
128	**Capsid protein**	276	Predicted protein
129	**Portal protein**	277	Predicted protein
130	**Tail protein**	278	Predicted protein
131	**Tail protein**	279	Predicted protein
132	**Tail protein**	280	Predicted protein
133	**Tail protein**	281	Predicted protein
134	**Tail protein**	282	Predicted protein
135	**Tail protein**	283	Predicted protein
136	Predicted protein	284	Predicted protein

Gene	Predicted Function	Gene	Predicted Function
137	Predicted protein	285	Predicted protein
138	Predicted protein	286	Predicted protein
139	**Tail protein**	287	Predicted protein
140	Predicted protein	288	Predicted protein
141	Predicted protein	289	Predicted protein
142	Predicted protein	290	Predicted protein
143	**Tail protein**	291	Predicted protein
144	**Lytic endolysin**	293	Predicted protein
145	**Minor capsid protein (second most abundant protein)**	298	**Tail protein**
146	**Capsid protein**	299	Predicted protein
147	Predicted protein	300	Predicted protein
148	Predicted protein	301	Predicted protein
149	**Capsid protein**	302	Predicted protein
150	Predicted protein	304	Predicted protein
151	Predicted protein	305	Predicted protein
152	Predicted protein	306	Predicted protein

APPENDIX B

166 EL GENES SELECTED FOR CODON USAGE ANALYSIS

(STRUCTURAL GENES IN BOLD)

Gene	Predicted Function	Gene	Predicted Function
2	Predicted protein	106	**Structural protein**
3	Predicted protein	107	Predicted protein
4	Predicted protein	108	Predicted protein
5	**Structural protein**	110	Predicted protein
6	**Structural protein**	111	Predicted protein
7	**Structural protein**	112	**Structural protein**
8	**Structural protein**	113	**Tail fiber protein**
9	**Structural protein**	114	**Tail fiber protein**
10	Predicted protein	115	**Tail fiber protein**
12	Predicted protein	116	**Structural protein**
13	**Structural protein**	117	Predicted protein
14	**Structural protein**	119	Predicted protein
15	Predicted protein	121	Predicted protein
17	Predicted protein	122	Predicted protein
18	Predicted protein	123	Predicted protein
19	Predicted protein	125	Predicted protein
21	DNA polymerase	126	Predicted protein
22	**Structural protein**	127	Predicted protein
23	Predicted protein	128	Predicted protein
26	Predicted protein	129	Predicted protein
28	**Structural protein**	130	Predicted protein
29	Predicted protein	131	Predicted protein
30	Predicted protein	132	Predicted protein
31	Predicted protein	133	Predicted protein
32	Predicted protein	134	Predicted protein
34	Predicted protein	135	Predicted protein
35	Predicted protein	136	Predicted protein
36	Predicted protein	138	Predicted protein
38	Predicted protein	139	Predicted protein
41	Predicted protein	140	Predicted protein
42	Predicted protein	141	Predicted protein
43	Predicted protein	142	Predicted protein
44	**Structural protein**	143	Predicted protein
45	**Structural protein**	144	**Structural protein**
46	Predicted protein	145	GroEL-like chaperonin
47	Predicted protein	146	Predicted protein
48	Predicted protein	147	**Structural protein**
51	Predicted protein	148	Predicted protein
52	**Structural protein**	149	Predicted protein

Gene	Predicted Function	Gene	Predicted Function
54	**Structural protein**	150	Predicted protein
55	**Structural protein**	151	Predicted protein
56	**Structural protein**	152	Predicted protein
57	**Structural protein**	153	Predicted protein
58	**Structural protein**	154	**Structural protein**
59	**Structural protein**	155	**Structural protein**
60	**Structural protein**	156	**Structural protein**
61	**Structural protein**	157	**Structural protein**
62	**Structural protein**	158	**Structural protein**
63	**Minor capsid protein**	159	Predicted protein
64	**Structural protein**	162	Predicted protein
65	Predicted protein	163	Predicted protein
66	**Structural protein**	164	Predicted protein
67	Predicted protein	165	**Structural protein**
68	**Structural protein**	167	Predicted protein
69	**Structural protein**	168	**Structural protein**
70	Predicted protein	169	**Structural protein**
71	**Structural protein**	170	Predicted protein
72	Predicted protein	171	Predicted protein
73	Predicted protein	172	**Structural protein**
75	Predicted protein	173	Predicted protein
76	Predicted protein	175	**Structural protein**
77	**Structural protein**	176	Predicted protein
78	**Major capsid protein**	177	Predicted protein
79	Predicted protein	178	Predicted protein
80	Predicted protein	179	Predicted protein
81	Predicted protein	180	Predicted protein
82	Predicted protein	181	**Structural protein**
83	Predicted protein	182	**Structural protein**
85	Predicted protein	183	**Structural protein**
86	Predicted protein	184	**Structural protein**
87	Predicted protein	186	**Structural protein**
90	Predicted protein	188	**Structural protein**
91	Predicted protein	189	**Structural protein**
92	Predicted protein	190	**Structural protein**
93	Predicted protein	191	**Structural protein**
95	Predicted protein	193	Predicted protein
96	Predicted protein	194	Predicted protein
98	Predicted protein	196	Predicted protein

Gene	Predicted Function	Gene	Predicted Function
99	Predicted protein	197	Predicted protein
100	Predicted protein	198	Predicted protein
103	**Structural protein**	199	Predicted protein
104	**Structural protein**	200	Predicted protein
105	Predicted protein	201	Predicted protein

LITERATURE CITED

1. Abedon ST. 2008. Bacteriophage Ecology: Population Growth, Evolution, and Impact of Bacterial Viruses. [Cambridge, UK]: Cambridge University Press: 154-70.

2. Agris PF, FAP Vendeix and WD Graham. "tRNA's Wobble Decoding of the Genome: 40 Years of Modification." *J Mol Biol* 366 (2007): 1–13.

3. Albers S and A Czech. "Exploiting tRNAs to Boost Virulence." *Life* 6 (2016): 4.

4. Ackermann HW and TM Nguyen. "Sewage coliphages studied by electron microscopy." *Appl Environ Microbiol* 45 (1983): 1049–1059.

5. Adriaenssens EM, Kropinski AM, Kuhn JH, Lavigne R, Krylov V. "A reassessment of the genus *Phikzvirus* with the addition of one (1) new species and the removal of one (1) species and the criterion of one (1) new genus for this species." International Committee on Taxonomy of Viruses (ICTV): Pending Proposals. (2016).

6. Aksyuk AA, LP Kurochkina, A Fokine, F Forouhar, VV Mesyanzhinov, L Tong, and MG Rossmann. "Structural conservation of the *Myoviridae* phage tail sheath protein fold." *Structure* 19 (2011): 1885-94.

7. Bailly-Bechet M, M Vergassola, and E Rocha. "Causes for the Intriguing Presence of tRNAs in Phages." *Genome Res* 17 (2007): 1486-1495.

8. Benzinger R, I Kleber, and R Huskey. "Transfection of *Escherichia coli* Spheroplasts I: General Facilitation of Double-Stranded Deoxyribonucleic Acid Infectivity by Protamine Sulfate." *J Virol* 7 (1971): 646-50.

9. Briers Y, G Volckaert, A Cornelissen, S Lagaert, CW Michiels, K Hertveldt, and R Lavigne. "Muralytic activity and modular structure of the endolysins of *Pseudomonas aeruginosa* bacteriophages phiKZ and EL." *Mol Microbiol* 65 (2007): 1334-44.

10. Briers Y, A Cornelissen, A Aertsen, K Hertveldt, CW Michiels, G Volckaert, and R Lavigne. "Analysis of outer membrane permeability of *Pseudomonas aeruginosa* and bactericidal activity of endolysins KZ 144 and EL 188 under high hydrostatic pressure." *FEMS Microbiol Lett* 280 (2008): 113-9.

11. Brok-Volchanskaya VS, FA Kadyrov, DE Sivogrivov, PM Kolosov, AS Sokolov, MG Shlyapnikov, VM Kryukov, and IE Granovsky. “Phage T4 SegB protein is a homing endonuclease required for the preferred inheritance of T4 tRNA gene region occurring in co-infection with a related phage.” *Nucleic Acids Res* 36 (2008): 2094-2105.

12. Bulmer M. “The selection-mutation-drift theory of synonymous codon usage.” *Genetics* 129 (1991): 897-907.

13. Bull JJ, MR Badgett, and HA Wichman. “Big-Benefit Mutations in a Bacteriophage Inhibited with Heat.” *Mol Biol Evol* 17 (2000): 942-950

14. Bull JJ, MR Badgett, D Rokyta, and IJ Molineux. “Experimental evolution yields hundreds of mutations in a functional viral genome.” *J Mol Evol* 57 (2003): 241-8.

15. Bull JJ, IJ Molineux, and CO Wilke. “Slow Fitness Recovery in a Codon-Modified Viral Genome.” *Mol Biol Evol* 29 (2012): 2997-3004.

16. Bull JJ. “Evolutionary reversion of live viral vaccines: Can genetic engineering subdue it?” *Virus Evol* 1 (2015): 1-10.

17. Carbone A. “Codon Bias is a Major Factor Explaining Phage Evolution in Translationally Biased Hosts.” *J Mol Evol* 66 (2008): 210-223.

18. Cardinale DJ and S Duffy. “Single-stranded genomic architecture constrains optimal codon usage.” *Bacteriophage* 1 (2011): 219-224.

19. Cardinale DJ, K DeRosa, and S Duffy. “Base Composition and Translational Selection are Insufficient to Explain Codon Usage Bias in Plant Viruses.” *Viruses* 5 (2013): 162-181.

20. Chan BK, M Sistrom, JE Wertz, KE Kortright, D Narayan, and PE Turner. “Phage selection restores antibiotic sensitivity in MDR *Pseudomonas aeruginosa*.” *Scientific Reports* 6 (2016): 26717.

21. Chan PP and TM Lowe. “GtRNAdb: A database of transfer RNA genes detected in genomic sequence.” *Nucl Acids Res* 37 (2009): D93-D97.

22. Clark JR and JB March. “Bacteriophages and biotechnology: vaccines, gene therapy and antibacterials.” *Trends Biotechnol* 24 (2006): 212-8.

23. Costerton JW, PS Stewart, and EP Greenberg. “Bacterial biofilms: a common cause of persistent infections.” *Science* 284 (1999): 1318-22.

24. D’Andrea L, RM Pintó, A Bosch, H Mustoc, and J Cristina. “A detailed comparative analysis on the overall codon usage patterns in Hepatitis A virus.” *Virus Research* 157 (2011): 19-24.

25. Danis-Wlodarczyk K, D Vandenheuvel, HB Jang, Y Briers, T Olszak, M Arabski, S Wasik, M Drabik, G Higgins, J Tyrrell, BJ Harvey, JP Noben, R Lavigne, and Z Drulis-Kawa. "A proposed integrated approach for the preclinical evaluation of phage therapy in *Pseudomonas* infections." *Sci Rep* 6 (2016): 28115.

26. Das G and RH Lyngdoh. "Role of wobble base pair geometry for codon degeneracy: Purine-type bases at the anticodon wobble position." *J Mol Model* 18.8 (2012): 3805–3820.

27. Delesalle VA, NT Tanke, AC Vill, and GP Krukonis. "Testing hypotheses for the presence of tRNA genes in mycobactriophage genomes." *Bacteriophage* 6 (2016): e1219441.

28. De Paepe M, M Leclerc, CR Tinsley, and M-A Petit. "Bacteriophages: an underestimated role in human and animal health?" *Front Cell Infect Microbiol* 4 (2014): 39.

29. Domingo-Calap P, JM Cuevas, and R Sanjuán. "The fitness effects of random mutations in single-stranded DNA and RNA bacteriophages." *PLoS Genetics* 5 (2009): e1000742.

30. Donlan RM and JW Costerton. "Biofilms: Survival Mechanisms of Clinically Relevant Microorganisms." *Clin Microbiol Rev* 15 (2002): 167-193.

31. Duffy S, LA Shackelton, and EC Holmes. "Rates of evolutionary change in viruses: Patterns and determinants." *Nat Rev Genet* 9 (2008): 267-276.

32. Fineran PC and E Charpentier. "Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information." *Virology* 434.2 (2012): 202-9.

33. Fokine A, AJ Battisti, VD Bowman, AV Efimov, LP Kurochkina, PR Chipman, VV Mesyanzhinov, and MG Rossmann. "Cryo-EM Study of the *Pseudomonas* Bacteriophage φKZ." *Structure* 15 (2007): 1099-1104.

34. Ghosh TC, SK Gupta, and S Majumdar. "Studies on codon usage in *Entamoeba histolytica*." *Int J Parasitology* 30 (2000):715-722.

35. Glonti T, N Chanishvili, and RW Taylor. "Bacteriophage-derived enzyme that depolymerizes the alginic acid capsule associated with cystic fibrosis isolates of *Pseudomonas aeruginosa." J Appl Microbiol* 108 (2010): 695-702.

36. Grocock RJ and PM Sharp. "Synonymous codon usage in *Pseudomonas aeruginosa* PA01." *Gene* 289 (2002): 131-9.

37. Grosjean H and W Fiers. "Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes." *Gene* 18 (1982): 199-209.

38. Gu W, T Zhou, J Ma, X Sun, and Z Lu. "Analysis of synonymous codon usage in SARS Coronavirus and other viruses in the Nidovirales." *Virus Res* 101 (2004): 155-161.

39. Gupta SK and TC Ghosh. "Gene expressivity is the main factor in dictating the codon usage variation among the genes in *Pseudomonas aeruginosa.*" *Gene* 273 (2001): 63-70.

40. Gutiérrez G, L Márquez and A Marín. "Preference for Guanosine at First Codon Position in Highly Expressed *Escherichia Coli* Genes: A Relationship with Translational Efficiency." *Nucl Acids Res* 24 (1996):2525-2527.

41. Hatfull GF. "Bacteriophage Genomics." *Curr Opin Microbiol* 11 (2008): 447-453.

42. Hatfull GF. "Dark Matter of the Biosphere: the Amazing World of Bacteriophage Diversity." *J Virol* 89 (2015): 8107-8110.

43. Heineman RH and SP Brown. "Experimental Evolution of a Bacteriophage Virus Reveals the Trajectory of Adaptation Across a Fecundity/Longevity Trade-Off." *PLoS ONE* 7 (2012): e46322.

44. Hendrix RW, MCM Smith, RN Burns, ME Ford, and GF Hatfull. "Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage." *Proc Natl Acad Sci USA* 96 (1999): 2192-2197.

45. Hertveldt K, R Lavigne, E Pleteneva, N Sernova, L Kurochkina, R Korchevskii, J Robben, V Mesyanzhinov, VN Krylov, and G Volckaert. "Genome comparison of *Pseudomonas aeruginosa* large phages." *J Mol Biol* 354 (2005): 536-545.

46. Hilterbrand A, J Saelens, and C Putonti. "CBDB: The codon bias database." *BMC Bioinformatics* 13 (2012): 62.

47. Holmes EC. "What does virus evolution tell us about virus origins?" *J Virol* 85 (2011): 5247-51.

48. Ikemura T. "Codon Usage and tRNA Content in Unicellular and Multicellular Organisms." *Mol Biol Evol* 2 (1985): 13-34.

49. Jusupova AB, TG Plotnikova, and VN Krylov. "Detection of transduction by virulent bacteriophage φKZ of *Pseudomonas aeruginosa* chromosomal markers in the presence of plasmid RMS148." *Genetika* 11 (1982): 1799-1802.

50. Kanaya S, M Kinouchi, T Abe, Y Kudo, Y Yamada, T Nishi, H Mori, and T Ikemura. "Analysis of codon usage diversity of bacterial genes with a self-organizing map (SOM): Characterization of horizontally transferred genes with emphasis on the *E. coli* O157 genome." *Gene* 276 (2001): 89-99.

51. Kazlauskas D and C Venclovas. “Computational analysis of DNA replicases in double-stranded DNA viruses: Relationship with the genome size.” *Nucleic Acids Res* 39 (2011): 8291-8305.

52. Kearse M, R Moir, A Wilson, S Stones-Havas, M Cheung, S Sturrock, S Buxton, A Cooper, S Markowitz, C Duran, T Thierer, B Ashton, P Mentjies, and A Drummond. “Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data.” *Bioinformatics* 28 (2012): 1647-1649.

53. Knoll BM and E Mylonakis. “Antibacterial Bioagents Based on Principles of Bacteriophage Biology: An Overview.” *Clin Infect Dis* 58 (2014): 528-534.

54. Koressaar T and M Remm. “Enhancements and modifications of primer design program Primer3.” *Bioinformatics* 23 (2007): 1289-1291.

55. Krylov V, E Pleteneva, M Bourkaltseva, O Shaburova, G Volckaert, N Sykilinda, L Kurochkina, and V Mesyanzhinov. “Myoviridae bacteriophages of *Pseudomonas aeruginosa*: a long and complex evolutionary pathway.” *Res Microbiol* 154 (2003): 269-275.

56. Krylov VN, DM Dela Cruz, and K Hertveldt, H-W Ackermann. ‘*φKZ-like viruses*’, a proposed new genus of Myovirus bacteriophages.” *Arch Virol* 152 (2007): 1955-1959.

57. Krylov VN and IZh Zhazykov. “*Pseudomonas* bacteriophage phiKZ—possible model for studying the genetic control of morphogenesis.” *Genetika* 14 (1978): 678-685.

58. Krylov VN, O Shaurova, S Kryolov, and E Pleteneva. “A genetic approach for the development of new therapeutic phages to fight *Pseudomonas aeruginosa* in wound infections.” *Viruses* 5 (2013): 15-53.

59. Kunisawa T, S Kanaya, and E Kutter. “Comparison of Synonymous Codon Distribution Patterns of Bacteriophage and Host Genomes.” *DNA Res* 5 (1998): 319-326.

60. Kunisawa T. “Functional Role of Mycobacteriophage Transfer RNAs.” *J Theor Bio* 205 (2000): 167-170.

61. Kurochkina LP, AA Aksyuk, MY Sachkova, NN Sykilinda, and VV Mesyanzhinov. “Characterization of tail sheath protein of giant bacteriophage φKZ *Pseudomonas aeruginosa*.” *Virology* 395 (2009): 312-317.

62. Lafay B, JC Atherton, and PM Sharp. “Absence of translationally selected synonymous codon usage bias in *Helicobacter pylori.*” *Microbiology* 146 (2000): 851-60.

63. Lecoutere E, PJ Ceyssens, KA Miroshnikov, VV Mesyanzhinov, VN Krylov, JP Noben, J Robben, K Hertveldt, G Volckaert, and R Lavigne. “Identification and comparative

analysis of the structural proteomes of phiKZ and EL, two giant *Pseudomonas aeruginosa* bacteriophages." *Proteomics* 9 (2009): 3215-3219.

64. Liu X. "A more accurate relationship between effective number of codons' and GC_ss under assumptions of no selection." *Comput Biol Chem* 42 (2013): 35-39.

65. Lu S, S Le, Y Tan, J Zhu, M Li, X Rao, L Zou, S Li, J Wang, X Jin, G Huang, L Zhang, X Zhao, and F Hu. "Genomic and proteomic analyses of the terminally redundant genome of the *Pseudomonas aeruginosa* phage PAP1: Establishment of the genus PaP1-like phages." *PLoS ONE* 8 (2013): e62933.

66. Lucks JB, DR Nelson, GR Kudla, and JB Plotkin. "Genome Landscapes and Bacteriophage Codon Usage." *PLoS Comput Biol* 4 (2008): e1000001.

67. Manrique P, B Bolduc, ST Walk, J van der Oost, WM de Vos, and MJ Young. "Healthy human gut phageome." *Proc Natl Acad Sci USA* 113 (2016): 10400-10405.

68. Martínez MA, A Jordan-Paiz, S Franco, and M Nevot. "Synonymous Virus Genome Recoding as a Tool to Impact Viral Fitness." *Trends Microbiol* 24 (2016): 134-147.

69. Matsuzaki S, S Tanaka, T Koga, and T Kawata. "A broad-host-range vibriophage, KVP40, isolated from sea water." *Microbiol Immunol* 36 (1992): 93-97.

70. Mesyanzhinov VV, J Robben, B Grymonprez, VA Kostyuchenko, MV Bourkaltseva, NN Sykilinda, VN Krylov, and G Volckaert. "The genome of bacteriophage phiKZ of *Pseudomonas aeruginosa*." *J Mol Biol* 317 (2002): 1-19.

71. Miller ES, JF Heidelberg, JA Eisen, WC Nelson, AS Durkin, A Ciecko, TV Feldblyum, O White, IT Paulsen, WC Nierman, J Lee, B Szczypinski, and CM Fraser. "Complete genome sequence of the broad-host-range vibriophage KVP40: comparative genomics of a T4-related bacteriophage." *J Bacteriol* 185 (2003): 5220-5233.

72. Mills S, F Shanahan, C Stanton, C Hill, A Coffey, and RP Ross. "Movers and shakers: Influence of bacteriophages in shaping the mammalian gut microbiota." *Gut Microbes* 4 (2013): 4-16.

73. Mueller S, D Papamichail, JR Coleman, S Skiena, and E Wimmer. "Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity." *J Virol* 80 (2006): 9687-9696.

74. Murphy J, J Mahony, S Ainsworth, A Nauta, and D van Sinderen. "Bacteriophage Orphan DNA Methyltransferases: Insights from Their Bacterial Origin, Function, and Occurrence." *Appl Environ Microbiol* 79 (2013): 7547-7555.

75. Nelson D, L Loomis, and VA Fischetti. “Prevention and elimination of upper respiratory colonization of mice by group A *Streptococci* by using a bacteriophage lytic enzyme.” *Proc Natl Acad Sci USA* 98 (2001): 4107-4112.

76. Pan A, C Dutta, and J Das. “Codon usage in highly expressed genes of *Haemophillus influenzae* and *Mycobacterium tuberculosis*: translational selection versus mutational bias.” *Gene* 215 (1998): 405-413.

77. Paradis-Bleau C, I Cloutier, L Lemieux, F Sanschagrin, J Laroche, M Auger, A Garnier, and RC Levesque. “Peptidoglycan lytic activity of the Pseudomonas aeruginosa phage phiKZ gp144 lytic transglycosylase.” *FEMS Microbiol Lett* 266 (2007): 201-209.

78. Parasion S, M Kwiatek, R Gryko, L Mizak and A Malm. “Bacteriophages as an Alternative Strategy for Fighting Biofilm Development.” *Pol J Microbiol* 63 (2014): 137-145.

79. Peden J. 1999. Analysis of Codon Usage (PhD thesis). [Nottingham (England)]: University of Nottingham.

80. Peden J. CodonW 1.4. (2005) URL: https://sourceforge.net/projects/codonw/

81. Pedulla ML, ME Ford, JM Houtz, T Karthikeyan, C Wadsworth, JA Lewis, D Jacobs-Sera, J Falbo, J Gross, NR Pannunzio, W Brucker, V Kumar, J Kandasamy, L Keenan, S Bardarov, J Kriakov, JG Lawrence, WR Jacobs Jr, RW Hendrix, and GF Hatfull. “Origins of highly mosaic mycobacteriophage genomes.” *Cell* 113 (2003): 171-182.

82. Perrière G and J Thioulouse. “Use and misuse of correspondence analysis in codon usage studies.” *Nucleic Acids Res* 30 (2002): 4548-4555.

83. Poon A and L Chao. “The rate of compensatory mutation in the DNA bacteriophage phiX174.” *Genetics* 170 (2005): 989-999.

84. Pope WH, KR Anders, M Baird, CA Bowman, MM Boyle, GW Broussard, T Chow, KL Clase, S Cooper, KA Cornely, RJ DeJong, VA Delesalle, L Deng, D Dunbar, NP Edgington, CM Ferreira, KW Hafer, GA Hartzog, JR Hatherill, LE Hughes, K Ipapo, GP Krukonis, CG Meier, DL Monti, MR Olm, ST Page, CL Peebles, CA Rinehart, MR Rubin, DA Russell, ER Sanders, M Schoer, CD Shaffer, J Wherley, E Vazquez, H Yuan, D Zhang, SG Cresawn, D Jacobs-Sera, RW Hendrix, and GF Hatfull. “Cluster M Mycobacteriophages Bongo, PegLeg, and Rey with Unusually Large Repertoires of tRNA Isotypes.” *J Virol* 88 (2014a): 2461-2480.

85. Pope WH, D Jacobs-Sera, DA Russell, DHF Rubin, A Kajee, ZNP Msibi, MH Larsen, WR Jacobs, Jr, JG Lawrence, RW Hendrix, and GF Hatfull. “Genomics and Proteomics of Mycobacteriophage Patience, an Accidental Tourist in the *Mycobacterium* Neighborhood.” *mBio* 5 (2014b): e02145-14.

86. Prabhakaran R, S Chithambaram, and X Xia. "*Aeromonas* phages encode tRNAs for their overused codons." *Int J of Computational Biol Drug Des* 7 (2014): 168-82.

87. R Core Team. "R: A language and environment for statistical computing." R Foundation for Statistical Computing (2013) URL: http://www.R-project.org/

88. Sahu K, SK Gupta, TC Ghosh, and SJ Sau. "Synonymous codon usage analysis of the mycobacteriophage Bxz1 and its plating bacteria *M. smegmatis*: identification of highly and lowly expressed genes of Bxz1 and the possible function of its tRNA species." *Biochem Mol Biol* 37 (2004): 487-492.

89. Sahu K, SK Gupta, S Sau, and TC Ghosh. "Comparative analysis of the base composition and codon usages in fourteen mycobacteriophage genomes." *J Biomol Struct Dyn* 23 (2005): 63-71.

90. Sanjuán R, MR Nebot, N Chirico, LM Mansky, and R Belshaw. "Viral Mutation Rates." *J Virol* 84 (2010): 9733-9748.

91. Sau K, S Sau, SC Mandal, and TC Ghosh. "Factors influencing the synonymous codon and amino acid usage bias in AT-rich *Pseudomonas aeruginosa* phage PhiKZ." *Acta Biochim Biophys Sin* 37 (2005): 625-33.

92. Sau K, SK Gupta, S Sau, SC Mandal, and TC Ghosh. "Studies of synonymous codon and amino acid usage biases in the broad-host range bacteriophage KVP40." *J Microbiol* 45 (2007): 58-63.

93. Serwer P, SJ Hayes, JA Thomas, and SC Hardies. "Propagating the missing bacteriophages: a large bacteriophage in a new class." *Virol J* 4 (2007): 21.

94. Sharp PM and W-H Li. "The codon adaptation index - a measure of directional synonymous codon usage bias, and its potential applications." *Nucleic Acids Res* 15.3 (1987): 1281-95.

95. Sharp PM, E Bailes, RJ Grocock, JF Peden, and RE Sockett. "Variation in the strength of selected codon usage bias among bacteria." *Nucleic Acids Res* 33 (2005): 1141-1153.

96. Sokolova OS, OV Shaburova, EV Pechnikova, AK Shaytan, SV Krylov, NA Kiselev, and VN Krylov. "Genome packaging in EL and Lin68, two giant phiKZ-like bacteriophages of *P. aeruginosa*." *Virology* 468-470 (2014): 472-478.

97. Sourial N, C Wolfson, B Zhu, J Quail, J Fletcher, S Karunananthan, K Bandeen-Roche, F Béland, and H Bergman. "Correspondence analysis is a useful tool to uncover the relationships among categorical variables." *J Clin Epidemiol* 63.6 (2010): 638-46.

98. Suzuki H, CJ Brown, LJ Forney, and EM Top. “Comparison of Correspondence Analysis Methods for Synonymous Codon Usage in Bacteria.” *DNA Res* 15 (2008): 357-365.

99. Sycheva LV, MM Shneider, NN Sykilinda, MA Ivanova, KA Miroshnikov, and PG Leiman. “Crystal structure and location of gp131 in the bacteriophage phiKZ virion.” *Virology* 434 (2012): 257-264.

100. Tuller T. “The Effect of Dysregulation of tRNA Genes and Translation Efficiency utations in Cancer and Neurodegenertion.” *Front Genet* 3 (2012): 201.

101. Thomas JA, ST Weintraub, W Wu, DC Winkler, N Cheng, AC Steven, and LW Black. “Extensive proteolysis of head and inner body proteins by a morphogenetic protease in the giant *Pseudomonas* aeruginosa phage ϕKZ.” *Mol Microbiol* 84.2 (2012): 324-39.

102. Thomas JA and LW Black. “Mutational Analysis of the *Pseudomonas aeruginosa* Myovirus ϕKZ Morphogenetic Protease gp175.” *J Virol* 87 (2013): 8713-8725.

103. Thomason L, DL Court, M Bubunenko, N Costantino, H Wilson, S Datta, and A Oppenheim. “Recombineering: Genetic engineering in bacteria using homologous recombination.” *Curr Protoc Mol Biol* 106 (2007): 1.16.1-39.

104. Untergasser A, I Cutcutache, T Koressaar, J Ye, BC Faircloth, M Remm, and SG Rozen. “Primer3 – New capabilities and interfaces.” *Nucleic Acids Res* 40 (2012): e115.

105. Van den Bossche A, SW Hardwick, P-J Cevssens, H Hendrix, M Voet, T Dendooven, KJ Bandyra, M DeMaeyer, A Aertsen, J-P Noben, BF Luisi, and R Lavigne. “Structural elucidation of a novel mechanism for the bacteriophage-based inhibition of the RNA degradosome.” *eLife* 5 (2016): e16413.

106. Vinay M, N Franche, G Grégori, J-R Fantino, F Pouillot, and M Ansaldi. “Phage-based fluorescent biosensor prototypes to specifically detect enteric bacteria such as *E. coli* and *Salmonella enterica* Typhimurium.” *PLoS ONE* 10 (2015): e0131466.

107. Vincze T, J Posfai, and RJ Roberts. “NEBcutter: a program to cleave DNA with restriction enzymes.” *Nucleic Acids Res* 31 (2003): 3688-3691.

108. Wagner PL and MK Waldor. “Bacteriophage Control of Bacterial Virulence.” *Infect Immun* 70 (2002): 3985-3993.

109. West SE and BH Iglewski. “Codon usage in *Pseudomonas aeruginosa*.” *Nucleic Acids Res* 16 (1988): 9323-9335.

110. Wilson JH. “Function of the bacteriophage T4 transfer RNA’s.” *J Mol Biol* 74 (1973): 753-757.

111. Wright F. “The ‘Effective number of codons’ used in a gene.” *Gene* 87 (1990): 23-29.

112. Wu W, JA Thomas, N Cheng, LW Black, and AC Steven. “Bubblegrams Reveal the Inner Body of Bacteriophage ϕKZ.” *Science* 335 (2012): 182.

113. Wua X, H Jörnvalla, KD Berndt, and U Oppermann. “Codon optimization reveals critical factors for high level expression of two rare codon genes in *Escherichia coli*: RNA stability and secondary structure but not tRNA abundance.” *Biochem Biophys Res Commun* 313 (2004): 89-96.

114. Yakunina M, T Artamonova, S Borukhov, KS Makarova, K Severinov, and L Minakhin. “A non-canonical multisubunit RNA polymerase encoded by a giant bacteriophage.” *Nucleic Acids Res* 43 (2015): 10411-10420.

115. Yin J. “Evolution of Bacteriophage T7 in a Growing Plaque.” J of Bacteriology 175.5 (1993): 1272-77.

VITA

Stephanie Steidel grew up in northwest Missouri. She attended Loyola University of Chicago, obtaining a Bachelor of Science degree in Biology with a Molecular Biology focus.