From unique killers to a jumbo genome – isolation and characterization of phages that infect the plant pathogen, *Agrobacterium tumefaciens*

A Dissertation

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Hedieh Attai

The undersigned, appointed by the dean of the Graduate School, have examined the thesis entitled

From unique killers to a jumbo genome – isolation and characterization of phages that infect the plant pathogen, *Agrobacterium tumefaciens*

presented by Hedieh Attai,

a candidate for the degree of Doctor of Philosophy,

and hereby certify that, in their opinion, it is worthy of acceptance.

_____

Dr. Pamela Brown, Major Advisor

_____

Dr. David Braun

_____

Dr. Mirela Milescu

_____

Dr. Scott Peck

ACKNOWLEDGEMENTS

I am so blessed to have such a large number of people to thank, who have been behind me and supported me during this challenging period.

Huge thank you to my advisor, Dr. Pamela Brown, for believing I could do things that I never dreamed I would be capable of. Our discussions have taught me to be a better scientist and have forever shaped the way I think. You have set such a strong example for me and been an inspiration as a woman in science. Thank you for being such a kind and understanding mentor and for advocating for me since day one. I could not have done any of this without you.

To Dr. George Smith, thank you for sharing your admiration of phage, for being such a patient teacher and mentor, and for challenging me to have a deeper understanding of my science.

The Division of Biological Sciences has been the ideal learning environment; every professor, staff member, classmate, and colleague has been friendly and supportive and I wish I could thank them all individually.

Thank you to my committee members Dr. David Braun, Dr. Mirela Milescu, and Dr. Scott Peck for being so supportive. Special thanks to Debbie Allen and Dr. Mark Hannink for providing me with advice when I needed it. To everyone at the MU EM Core Facility for sharing my excitement when we imaged new phages. Thank you to my collaborators, Rob Lavigne, Maarten Boon, and Jean-Paul Noben at KU Leuven who contributed mass spectrometry analysis of the structural proteins of Atu_ph07 and

Christopher Yost at University of Regina who helped us with the transposon mutagenesis sequencing of PPH-expressing *Agrobacterium*.

Thank you to the National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health (NIH), which provided me a T32 training grant under award number T32GM008396, and the U.S. Department of Education Graduate Assistance in Areas of National Need (GAANN) Fellowship.

I would like to give a special thanks to Wanda Figueroa-Cuilan, Michelle Williams, and Matt Howell—having you three to share this experience with made it worth so much more. We learned everything about *Agrobacterium* together, had countless laughs in the deep confines of fourth floor Tucker, and somewhere in the meanwhile, we grew into real scientists. Words cannot describe how grateful I am for Wanda; our friendship has been so special to me and has held me together through the ups and downs of grad school. Michelle has challenged me in so many ways, intellectually and personally, and I always looked forward to walking in to lab and sharing a coffee and conversation with her. Matt provided tons of technical advice and always helped keep me calm when I felt like the world was falling apart if an experiment did not work. My team of undergrads (or "Team Phage") has been awesome to work with and has contributed significant data to this project: Jeanette Rimbey, Helen Blaine, Kenya Phillips, Courtney Buchanan, and Karlie Schaphorst. Thank you to the people who made it a pleasure to come in to lab everyday: Sue Ann Flores, Jeremy Daniel, Chris Richards, Gustavo Santiago-Collazo, and Jacob Bouchier. The overall positivity and

encouragement I received from the Brown lab was a unique experience that I will always treasure.

I would also like to thank Dr. Shawna Reed, my graduate student mentor when I was an undergrad trainee, for giving me such a strong foundation in research. Dr. Matthew Welch and Dr. David Bermudes both took a chance on me as a budding scientist and allowed me to join their labs, and for that, I am forever grateful.

Outside of the lab, I have met some truly wonderful people in Columbia, MO, who have been essential in helping me survive grad school and always manage to put a smile on my face. Those people are: David Porciani and Alessandra Cecchini-Porciani (who convinced me that I have what it takes to be a postdoc), Andrea Saltos, Amir Mehdi Mofrad, Morgan Halane, Mark and Sam Schroeder, Marco Navarro, Emilia Assante, Melissa Zaidi, Lisa Shepard, Amanda Howland, James Hopfenblatt McAtee, Carlos Martinez-Villar, Selim Sukhtaiev, Armen Kazarian, Elisavet Levogianni, Andrick Payen, Eleni Bickell, and Angelia Chen. I would also like to thank my dear friends from home, who continuously sent me kind words of encouragement from hundreds of miles away: Sahra Mirbabaee, Betty Zong, Parinaz Fozouni, Alejandra Gonzalez, and Jinal Hicks. Special thanks to Andrew Oliver, who has given me so much support over the years, and who has expanded my knowledge about phage and bioinformatics; I will never forget how you listened and provided your feedback as I practiced my first big conference talk over and over until it was perfect.

To my dear Elizabeth Cappa, my best friend of 20 years, who lost her battle to cystic fibrosis just a few months ago, you are the bravest, strongest warrior I know.

Thanks for making me laugh harder than anyone, for being the loyal friend I could always count on, the best listener with the wisest advice, and for teaching me what it means to be resilient. I dedicate my research to you and hope to always make you proud.

None of this work would have been possible without the unconditional support of my family. I owe the deepest gratitude to my loving mom, Helen Attai, who has been my biggest advocate, whose daily words of encouragement have given me strength, and whose advice I cherish and have made me who I am today. My dad, Behnam Attai, has taught me to be curious about the world around us; he encouraged me to love science at a young age and given me the support to pursue my dreams. My sister, Halleh Attai, whose work ethic I strive to emulate, has been my biggest role model in life, and has always believed in me. My precious grandparents, including Mamani, a strong feminist from whom I take my extroverted nature from, and Bababozorg, who shared his love of science and philosophy with me. Thank you to my Kansas City family, especially Daiee Jamshid and Maryam Joon, for giving me a home away from home. To my aunts, uncles, and dear cousins, I love every single one of you and I'm so thankful to have such a strong support system.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

From unique killers to a jumbo genome – isolation and characterization of phages that infect the plant pathogen, *Agrobacterium tumefaciens*

Hedieh Attai

Dr. Pamela Brown, Dissertation Advisor

Abstract

Bacteriophages and their lytic peptides can protect plants from phytopathogens such as *Agrobacterium tumefaciens*. To better understand mechanisms of phage-mediated host killing, we isolated and characterized five lytic bacteriophages with activity against *A. tumefaciens* C58. These phages come in different shapes and sizes—from T7-like phages with podoviral morphology and isometric heads to T4-like phages with myoviral morphology and a contractile tail—and exhibit varying host ranges and killing efficiencies. The smallest *Agrobacterium* phages are phiKMV-like phages in the T7 superfamily that are efficient at killing their hosts. Their lethality can be attributed to their expression of a unique endolysin, called Phage Peptidoglycan Hydrolase (PPH). The atypical domain structure of PPH, along with the absence of obvious accessory proteins, suggest PPH may function independently to mediate host cell lysis. Contrary to the narrow host range of the phage, expression of PPH from an inducible promoter inhibits cell growth and blocks cell division in a broad range of bacteria including *Agrobacterium*, *Sinorhizobium*, and *Escherichia* strains. Another member of the *Podoviridae* family, Atu_ph08, carries remnants of a lysogen and shares 60.2% identity with *Agrobacterium* genomospecies 3. The T4-like

phages in our collection are not potent killers. Atu_ph04 is unique T4-like phage that is similar to a group of rhizophages. The largest *Agrobacterium* phage, Atu_ph07, has a head diameter of 146 nm, an extended tail length of 136 nm, and a genome of 490 kbp. Our results indicate a high degree of morphological and genomic diversity and also suggest novel mechanisms of host cell killing remain to be uncovered.

Chapter 1

Hidden Gems of the Microbial World: Bacteriophages and their Enzymes

ABSTRACT

Bacteriophages are virtually everywhere—phages have been identified in nearly every environment on this planet and are numerically dominant in many environments. Phages and their bacterial hosts have been coevolving for billions of years, enabling some phages to become efficient killers of their hosts. With the spread of antibiotic-resistant bacteria, it is crucial that we better understand the diverse mechanisms employed by phages to slow growth and lyse bacterial cells. Some phages express proteins called endolysins to lyse their hosts. Better understanding how these proteins function should reveal their antimicrobial potential and enable the development of strategies to prevent or treat bacterial infections. Furthermore, given the diversity of phages that remain unexplored it is likely that additional mechanisms of host cell killing remain to be discovered. Finally, in some cases, phage cocktails have been shown to control the spread of diseases in plants and humans. To further explore the diversity of phages and in hope of isolating phages suitable for biocontrol, here we isolate and characterize phages that infect the bacterial plant pathogen, *Agrobacterium tumefaciens*.

INTRODUCTION

For over 3 billion years, microbes and viruses have inhabited the earth and shaped the environment we recognize today. Bacteria are found nearly everywhere. Our environment has so-called “good” bacteria, those that colonize our microbiomes or provide nutrients to plants, and “bad” bacteria, which cause disease. But the underlying entities that influence bacterial diversity are less understood, often even referred to as “dark matter” [1]. Viruses are the most abundant biological entities on this planet, yet

much remains to be discovered about them and the roles that they play in our environments [2]. Viruses that infect bacteria are called bacteriophages, or phages. Current estimates suggest that there are 10^8 phages in every milliliter of water [3] and that 31 billion phages traverse cells in the epithelial layer in the human body each day [4]. Bacteriophages, the term literally meaning "bacterial eaters," were first discovered in 1915 and scientists quickly realized their potential as antimicrobials. For example, in the 1940s Eli Lilly sold seven phage products for human use, including "Staphylo-jel," which was used to prevent or treat *Staphylococcus* infections [5]. However, the discovery and commercialization of antibiotics overshadowed the need for phage research and commercial therapeutic phage production ceased in the Western World. Still, some scientists recognized the value of studying these phages and scientific research has validated the important role of phages in driving microbial diversity and genetic recombination [6], as well as their potentially therapeutic properties.

Phages Influence Microbial Diversity

Phages are able to drive microbial diversity because of their unique lifestyles. These non-living organisms are entirely dependent on their bacterial hosts for survival. The majority of double-stranded DNA tailed phages are characterized as having either lytic and temperate lifestyles (Figure 1-1), though there are additional phage lifestyles. The lytic cycle is initiated by phage attachment to a bacterial host cell. The phage will recognize a specific receptor (or receptors) on, or protruding from, the bacterium [7]. The phage will then inject its genetic material through its tail and into its host by poking a needle through the cell membrane(s) [8]. The phage genome translocates into the host,

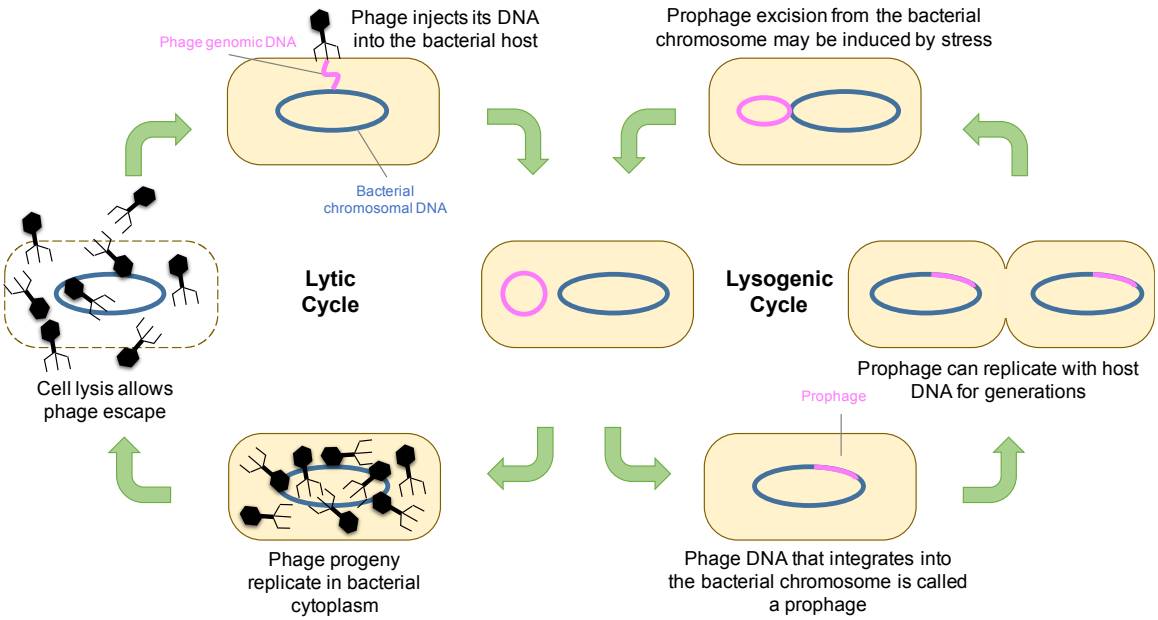


Figure 1-1. Phages can undergo the lytic or lysogenic cycle. In the lytic cycle, phage DNA will be injected into the host, where it will replicate and be packaged into new phage heads. Phage lysis proteins will burst the cell wall to allow progeny to infect the next host. In the lysogenic cycle, phage DNA can integrate into the host chromosome, where it becomes a prophage. This prophage can replicate along with the bacterial DNA for generations until a stressor, such as UV or chemical damage to DNA induces prophage excision and entry into the lytic cycle.

replicates in the host cytoplasm, and subsequent transcription and translation results in the production of proteins involved in the assembly of phage progeny. New phage particles are assembled and packaged with genetic material. To allow the new phages to escape the cell, the phages may express enzymes that destroy the host cell wall. The bacterial cell bursts and a new round of phage infection is initiated. The process of killing microbes leads to a release in organic matter that is consumed by other organisms. It is estimated that viral infections release up to 10^9 tons of carbon each day, contributing to 25% of the carbon used in photosynthesis [9].

Temperate Phages are Major Contributors to Horizontal Gene Transfer

Not all phages rapidly kill their hosts—the genomes of temperate phages are capable of entering the lysogenic cycle (Figure 1-1), remaining dormant by either incorporating into the host genome as a prophage or remaining as a separate plasmid with a phage-derived origin of replication in the host cell. The phage genome can remain latent in the host indefinitely, until phage replication is activated by a stress response. Once activated by a stressor, such as chemical treatment or UV damage to DNA, the phage enters the lytic cycle. Temperate phages can undergo both lytic and lysogenic cycles. In fact, *Bacillus* phages can communicate with one another during the switch from lysogeny to lysis [10]. The presence of prophages in the environment is vast—it has been estimated that half of marine bacterial genomes contain prophages [11].

The process in which a phage transfers DNA from one bacterium to another is generally mediated by temperate phages and is called transduction. Generalized transduction occurs when the phage packages random DNA from the bacterial host rather

than phage DNA. During the next round of infection, the phage injects the bacterial DNA (rather than phage DNA) into the host cell, enabling genetic transfer. This process is well characterized in *Enterobacteria* phage P22, in which the phage DNA sequence used for the initiation of packaging is homologous to sequences in the host genome, enabling the phage to package host DNA fragments [12]. In specialized transduction, which is modelled in *Enterobacteria* phage lambda, the prophage is excised from a specific site in the bacterial genome, and adjacent bacterial DNA is excised as well [12]. Transduction is one of the major contributors to horizontal gene transfer and subsequent microbial evolution. Each year, 10^{24} genes are transferred from virus to host in the oceans [13].

DIVERSITY OF PHAGES

Phage diversity is so wide that there is no single, universal gene that is conserved throughout all phages [14]. Instead, the phylogeny of phages is determined using signature genes, including the large terminase, portal vertex protein, and the major capsid protein [15]. Phages are most often categorized by their morphology and relation to well-characterized phages, such as the coliphages T4, T7, or lambda. These phages have been key model organisms in the phage field and have led to critical findings in molecular biology. T4 is known for its icosahedral head and visible tail, whereas T7 shares a similarly shaped head with a much smaller, less noticeable tail. Together, these make up major superfamilies of phages that look similar but may infect very unrelated hosts.

The morphology of phages has been a key consideration when classifying them into families. Electron microscopy of isolated phages allows researchers to observe their morphologies. The order *Caudovirales* is comprised of tailed dsDNA phages, phages

with a long contractile tail, such as T4, are classified as *Myoviridae*, phages with short noncontractile tails, like T7, are *Podoviridae*, and phages with long noncontractile tails, such as lambda, are *Siphoviridae*. This form of classification system was defined by Hans Ackermann [16] and is still used by the International Committee on Taxonomy of Viruses (ICTV) [17].

PHAGE GENOMES

The size of phage genomes can range from about 3 kbp to the recently identified 540 kbp Megaphages [18,19]. Phages with genomes larger than 200 kbp are considered jumbo phages, and are often more closely related to other jumbo phages than phages that share a host [20].

Phage genomes are often modular with regions encoding for functions such as virion structure, DNA replication, DNA modification, and control of transcription. Phage genomes of the same family encode a set of conserved, or core, genes that are involved in essential processes [18]. Highly conserved genes encode proteins involved in virion structure and phage morphogenesis. Phage morphogenesis in the *Caudovirales* initiates with the formation of the prohead. DNA is packaged through the portal vertex and cleaved by the terminase proteins. Next, the head completion proteins, as well as the tail subunits and tail fibers are assembled [21]. The baseplate lysozyme is necessary for DNA entry into the host cell. Other core genes are involved in DNA replication, transcription, translation and posttranslational control, nucleotide metabolism, and morphogenesis. DNA replication, repair, and recombination genes include DNA polymerase and helicase. RNA polymerase sigma factor initiates transcription. Ribonucleotide reductase enzymes

reduce ribonucleotides into deoxyribonucleotides, a step that is essential to DNA synthesis [22].

Phages of the same superfamily often share core genes, while having variability in hyperplastic regions, whose genes are mostly uncharacterized. Gene shuffling contributes to this variability. Genes in the hyperplastic regions are hypothesized to allow the phage to adapt to their hosts [23].

PHAGES AS BIOCONTROL AGENTS

As the rate of antibiotic resistant bacteria increases, there is an increasing need for a new approach to finding antimicrobials [24]. Since phages have natural antagonistic properties against bacteria, they are being investigated for their role in treating and preventing bacterial diseases in animals as well as plants [25–27]. In humans, phage therapy has been shown to be successful in cases as a last resort, when all antibiotics fail to treat a patient. In one particularly striking case, a man named Tom Patterson emerged from his coma after receiving a phage therapy cocktail to treat his multi-drug resistant *Acinetobacter baumannii* [28].

In plants, phages can be applied to prevent infection. From an agricultural standpoint, plant diseases lead to a 10% decrease in global food production [29] and the global food supply is predicted to be insufficient for the expected population increase. In several cases, phage biocontrol has been shown to lead to an improved crop yield [30]. For instance, phage treatment led to a 59% reduction of *Xanthomonas axonopodis*-causing citrus canker [31]. Phage mixtures were able to prevent tomato wilt caused by *Ralstonia solanacearum*, as seen in multiple studies [32,33]. Coinoculation of phage with

pathogenic *Pectobacterium* and *Dickeya* reduced soft rot on potato slices by 80% [34]. Finally, phage treatments also reduced the rotting and blackleg in potatoes caused by *Dickeya solani* [35].

There are a number of important considerations as phage biocontrol strategies move from the lab into the field. The timing of application is a key consideration [36]. Phages cannot undo the damage caused by a phytopathogen and therefore should be viewed as a means to prevent, rather than cure, an infection. Also, to avoid UV exposure which can cause DNA damage to the phage, it is preferred that phage application occur at night. Finally, it is necessary to prescreen the host range of phages to ensure that beneficial microbes in the soil and rhizosphere will not be adversely impacted.

PHAGE ENDOLYSINS

Due to centuries of coevolution between bacteria and phages, some phages have become armed with a reservoir of genes effective at adapting to their environments and killing their hosts. Understanding the mechanism in which phages lyse their host cells is an important step to discovering a new source of antimicrobials. In order to allow the release of new phage progeny, many phages express proteins that hydrolyze the bacterial cell wall.

The Canonical Endolysin-Holin-Spanin Mechanism

This process is well-characterized in lambda phage, and the endolysin-holin-spanin mechanism is considered to be the canonical mechanism by which phage-mediated cell lysis occurs. Lambda phage infects the Gram negative bacteria, *Escherichia*

coli. Gram negative bacteria are enclosed by an inner membrane, periplasm that contains the peptidoglycan (PG) cell wall, and an outer membrane. Cell wall PG is composed of alternating sugars N-acetylmuramic acid (MurNAc) and N-acetylglucosamine (GlcNAc), as well as peptide chains that are crosslinked together to form a lattice. Phages express three classes of proteins for the lysis process: endolysins, holins, and spanins (Figure 1-2) [37]. First, holins oligomerize to form a hole in the inner membrane [38]. The endolysin is now able to travel through this newly formed pore, into the periplasmic space, where it cleaves PG. Endolysins vary in their specific enzymatic activity, but the lambda phage is a murein transglycosylase, which cleaves the glycosidic bond between the MurNAc and GlcNAc [39,40]. The spanin proteins comprise an i-spanin, or inner membrane spanin, and an o-spanin, or outer-membrane spanin. These spanins contract, forcing the inner and outer membranes to fuse, creating a pore in the envelope, which completes the lysis of the host cell [41].

Non-canonical Modes of Phage-Mediated Lysis

The diversity of phages can be appreciated by considering numerous other mechanisms of phage-mediated cell death that have been discovered. Signal-arrest-release (SAR) endolysins encode an N-terminal transmembrane domain that anchors to the inner membrane until signaled by an enzyme called the pinholin [42]. Pinholins form small pores that trigger a shift in the proton motive force, causing the release of the active SAR endolysin into the periplasm, where it hydrolyzes the bacterial cell wall [43].

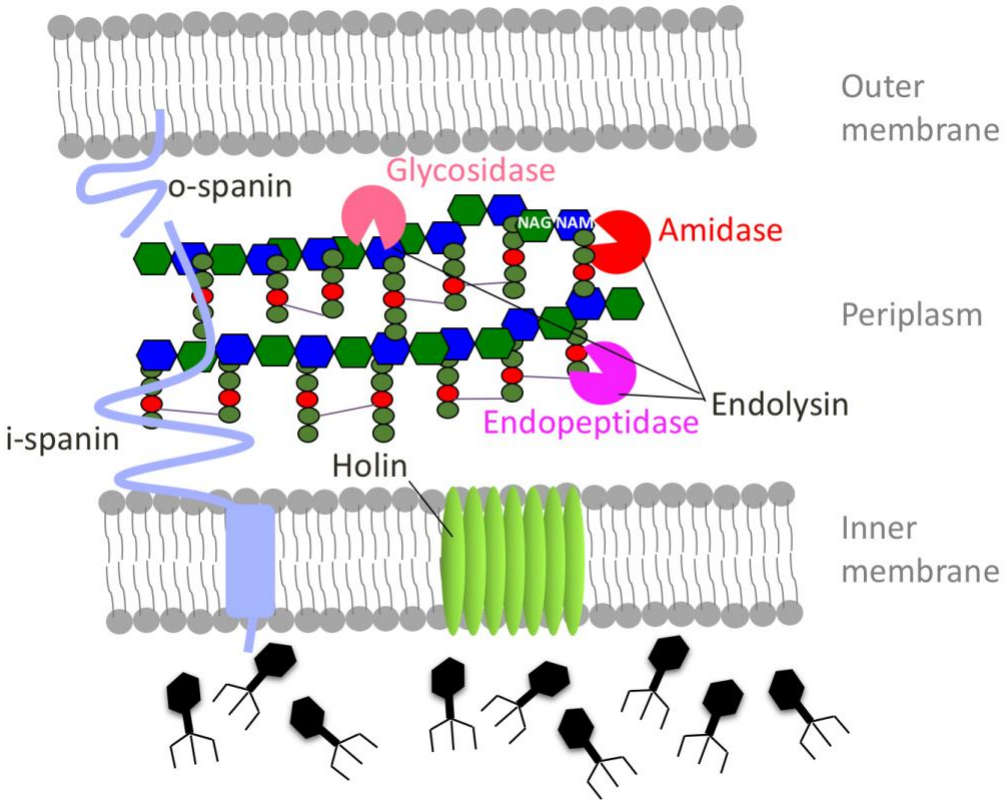


Figure 1-2. The canonical endolysin-holin-spanin mechanism of phage-mediated cell lysis. Phage lambda expresses holins that form a pore in the inner membrane, allowing an endolysin to hydrolyze the cell wall PG. Three common mechanisms of PG lysis include amidase, endopeptidase, or glycosidase activity. Once the cell wall is destroyed, the spanin proteins are activated to fuse the inner and outer membranes and complete the lysis process.

Small, single-stranded nucleic acid phages, such as coliphages ϕX174 and Qβ, use a different approach to kill their hosts, involving the expression of single lysis proteins. Single lysis proteins are sufficient to induce cell death because they target proteins involved in cell wall biogenesis, a process essential to the cell's survival [44]. The phage ϕX174 genome encodes lysis protein E which blocks the essential protein MraY and the phage Qβ genome encodes A_2, which targets MurA. Both MraY and MurA are involved in PG precursor biosynthesis.

Applications of Lysis Proteins

While phages and bacteria coevolve, bacteria can gain resistance against certain phages. They can mutate the surface-exposed receptors that are recognized by phage tail fibers, such as flagella, pili, or lipopolysaccharides. In contrast, they cannot drastically modify the structure of PG because the cell wall is necessary for the maintenance of cell shape and structural integrity. Therefore, it can be argued that phage endolysins may be outstanding candidates for therapy against multidrug resistant bacterial infections [45]. Other advantages to using phage-derived peptides is their rapid action and synergy when combined with antibiotics or other antimicrobial strategies [46].

The use of lysin enzymes as antibiotics, or "enzybiotics," was first mentioned by Nelson et al. [47]. Though the application of endolysins are arguably better suited for Gram positive bacteria that lack an outer membrane and thus have an exposed cell wall, recent studies have shown the promise for treating Gram negative bacteria with a combination of endolysins and outer membrane permeabilizers. Polycationic agents, such as polymyxin E, and chelators, such as EDTA, have been shown to effectively

permeabilize the outer membrane by disrupting lipopolysaccharides [48,49]. Another approach to increase permeabilization of the outer membrane is to engineer fusion proteins. The modular structure of endolysins allow for the construction of such protein fusions [50]. One particularly striking example is the engineering of outer-membrane penetrating endolysins, dubbed Artilysins [51]. Artilysins contain seven peptides that destabilize lipopolysaccharides fused to highly active endolysins. *In vitro* and *in vivo* studies found that Artilysins are effective in killing *Pseudomonas aeruginosa* and *Acinetobacter baumannii*. Artilysins led to a 4 to 5 log reduction in bacteria *in vitro* and, in combination with EDTA, Artilysins could rescue 40% of infected worms.

A CASE STUDY: *AGROBACTERIUM* AS AN AGENT OF DISEASE

Phages that infect human pathogens and marine bacteria have been widely identified, yet there is an underrepresentation of phages that infect the order *Rhizobiales* and family *Rhizobiaceae*, which includes many soil-dwelling bacteria [52]. These bacteria are quite important for plant growth, their impacts ranging from symbiosis in which the bacteria fixes nitrogen for the host plant to pathogenesis. There are currently 18 *Rhizobium* phages and 15 *Sinorhizobium* phages in the database, reflecting the importance of these beneficial microbes to soil and plant health. Only one *Agrobacterium* phage, named 7-7-1, had been isolated prior to our research, despite the importance of *Agrobacterium* in agriculture [53]. *Myoviridae* phage 7-7-1 infects the non-pathogenic *Agrobacterium* sp. H13-3 by attaching to its flagella [54]. To gain a better representation of *Agrobacterium* phages, and to potentially find phages that prevent disease, we sought to isolate phages that infect the phytopathogenic *A. tumefaciens* strain C58.

While *Agrobacterium tumefaciens* is widely known to genetically modify plants, it is also known to plant pathologists as the third most scientifically and economically important plant pathogen [55]. *A. tumefaciens* causes crown gall disease in plants, which manifests as galls, or tumors, that are comprised of overproliferating plant cells [56]. Uncontrolled plant cell division leads to a block in nutrient transport throughout the plant and a misallocation of resources, leading to a decrease in crop yield [57]. Pathogenic *A. tumefaciens* contains a Ti plasmid, which encodes its virulence genes, and is required for the formation of crown galls [58]. The Ti plasmid has a Vir region that encodes a type IV secretion system and a fragment called T-DNA, which integrates into the plant host cell's genome [59]. The genes encoded on the T-DNA include enzymes that overproduce plant hormones, including auxin and cytokinin, and proteins involved in synthesizing opines, which serve as nutrients, providing carbon and energy for the bacteria [60]. Scientists have taken advantage of *Agrobacterium*'s ability to integrate DNA into plant cells and now use this system to deliver genes to plants [61].

Currently, the commercially available biocontrol to prevent crown gall disease is a strain of *A. radiobacter* called K84. This strain produces a toxin called agrocin 84, which inhibits the pathogenic *A. tumefaciens* and *A. rhizogenes* [62]. However, not all strains of *A. tumefaciens* are sensitive to agrocin 84 [63]. Another concern is that conjugation between *A. radiobacter* and *A. tumefaciens* can enable *A. radiobacter* strains to acquire the Ti plasmid and become pathogenic [64]. Thus, there is a need for alternative biocontrol options to prevent *A. tumefaciens* infections, and here we explore the possibility of treatments using phage or their proteins as viable biocontrol options.

CONCLUSIONS

It has become evident that bacteriophages play an important role in our environment through the shuffling of genes that has led to microbial diversity. In a society that is becoming increasingly threatened by antibiotic resistance, it is vital that alternative approaches are explored. We have only uncovered the tip of the iceberg of phage diversity, and the vast reservoir of genes which have evolved to kill bacteria are only beginning to be understood. The diverse mechanisms of host cell killing employed by phages have the potential for providing a source of novel antimicrobials. In our study, we isolated *Agrobacterium* phages to be used as biocontrol agents to prevent crown gall disease in plants. Not only did we isolate and characterize new phages that infect the phytopathogen, but we discovered new genes that have potential as antimicrobials.

REFERENCES

1. Youle, M.; Haynes, M.; Rohwer, F. Scratching the surface of biology's dark matter. *Viruses Essent. Agents Life* **2012**, *9789400748*, 61–81, doi:10.1007/978-94-007-4899-6.
2. Paez-Espino, D.; Eloe-Fadrosh, E. A.; Pavlopoulos, G. A.; Thomas, A. D.; Huntemann, M.; Mikhailova, N.; Rubin, E.; Ivanova, N. N.; Kyrpides, N. C. Uncovering Earth's virome. *Nature* **2016**, *536*, 425–430, doi:10.1038/nature19094.
3. Bergh, Ø.; Børsheim, K. Y.; Bratbak, G.; Heldal, M. High abundance of viruses found in aquatic environments. *Nature* **1989**, *340*, 467–468, doi:10.1038/340467a0.
4. Nguyen, S.; Baker, K.; Padman, B. S.; Patwa, R.; Dunstan, R. A.; Weston, T. A.; Schlosser, K.; Bailey, B.; Lithgow, T.; Lazarou, M.; Luque, A.; Rohwer, F.; Blumberg, R. S.; Barr, J. J. Bacteriophage transcytosis provides a mechanism to cross epithelial cell layers. *MBio* **2017**, *8*, 1–14.
5. Kaźmierczak, Z.; Górski, A.; Dabrowska, K. Facing antibiotic resistance: *Staphylococcus aureus* phages as a medical tool. *Viruses* **2014**, *6*, 2551–2570, doi:10.3390/v6072551.
6. Koskella, B.; Brockhurst, M. A. Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol. Rev.* **2014**, *38*, 916–931, doi:10.1111/1574-6976.12072.
7. Bertozzi Silva, J.; Storms, Z.; Sauvageau, D. Host receptors for bacteriophage adsorption. *FEMS Microbiol. Lett.* **2016**, *363*, 1–11, doi:10.1093/femsle/fnw002.
8. Leiman, P. G.; Shneider, M. M. Contractile tail machines of bacteriophages. *Adv*

Exp Med Biol **2012**, *726*, 93–114, doi:10.1007/978-1-4614-0980-9_5.

9. Breitbart, M. Marine viruses: truth or dare. *Ann. Rev. Mar. Sci.* **2012**, *4*, 425–448, doi:10.1146/annurev-marine-120709-142805.

10. Erez, Z.; Steinberger-Levy, I.; Shamir, M.; Doron, S.; Stokar-Avihail, A.; Peleg, Y.; Melamed, S.; Leavitt, A.; Savidor, A.; Albeck, S.; Amitai, G.; Sorek, R. Communication between viruses guides lysis-lysogeny decisions. *Nature* **2017**, *541*, 488–493, doi:10.1038/nature21049.

11. Paul, J. H. Prophages in marine bacteria: Dangerous molecular time bombs or the key to survival in the seas? *ISME J.* **2008**, *2*, 579–589, doi:10.1038/ismej.2008.35.

12. Clokie, M. R. J. *Bacteriophages methods and protocols, volume 1: isolation, characterization, and interactions*; 2009; Vol. 531; ISBN 9781588296825 (hbk.)r1588296822 (hbk.).

13. Rohwer, F.; Thurber, R. V. Viruses manipulate the marine environment. *Nature* **2009**, *459*, 207–212, doi:10.1038/nature08060.

14. Rohwer, F.; Edwards, R. The phage proteomic tree: a genome-based taxonomy for phage. *J. Bacteriol.* **2002**, *184*, 4529–4535, doi:10.1128/jb.184.16.4529-4535.2002.

15. Adriaenssens, E. M.; Cowan, D. A. Using signature genes as tools to assess environmental viral ecology and diversity. *Appl. Environ. Microbiol.* **2014**, *80*, 4470–4480, doi:10.1128/AEM.00878-14.

16. Ackermann, H. W. Phage classification and characterization. *Methods Mol. Biol.* **2009**, *501*, 127–140, doi:10.1007/978-1-60327-164-6.

17. Aiewsakun, P.; Adriaenssens, E. M.; Lavigne, R.; Kropinski, A. M.; Simmonds, P.

Evaluation of the genomic diversity of viruses infecting bacteria, archaea and eukaryotes using a common bioinformatic platform: Steps towards a unified taxonomy. *J. Gen. Virol.* **2018**, *99*, 1331–1343, doi:10.1099/jgv.0.001110.

18. Hatfull, G. F.; Hendrix, R. W. Bacteriophages and their genomes. *Curr Opin Virol* **2011**, *1*, 298–303, doi:10.1016/j.coviro.2011.06.009.

19. Devoto, A. E.; Santini, J. M.; Olm, M. R.; Anantharaman, K.; Munk, P.; Tung, J.; Archie, E. A.; Turnbaugh, P. J.; Seed, K. D.; Blekhman, R.; Aarestrup, F. M.; Thomas, B. C.; Banfield, J. F. Megaphage infect *Prevotella* and variants are widespread in gut microbiomes. *Nat. Microbiol.* **2018**, 356790, doi:10.1101/356790.

20. Yuan, Y.; Gao, M. Jumbo bacteriophages: an overview. *Front. Microbiol.* **2017**, *8*, doi:10.3389/fmicb.2017.00403.

21. Fokine, A.; Rossmann, M. G. Molecular architecture of tailed double-stranded DNA phages. *Bacteriophage* **2014**, *4*, e28281, doi:10.4161/bact.28281.

22. Sakowski, E. G.; Munsell, E. V; Hyatt, M.; Kress, W.; Williamson, S. J.; Nasko, D. J.; Polson, S. W.; Wommack, K. E. Ribonucleotide reductases reveal novel viral diversity and predict biological and ecological features of unknown marine viruses. *Proc Natl Acad Sci U S A* **2014**, *111*, 15786–15791, doi:10.1073/pnas.1401322111.

23. Comeau, A. M.; Bertrand, C.; Letarov, A.; Tetart, F.; Krisch, H. M. Modular architecture of the T4 phage superfamily: a conserved core genome and a plastic periphery. *Virology* **2007**, *362*, 384–396, doi:10.1016/j.virol.2006.12.031.

24. Laxminarayan, R.; Duse, A.; Wattal, C.; Zaidi, A. K. M.; Wertheim, H. F. L.;

Sumpradit, N.; Vlieghe, E.; Hara, G. L.; Gould, I. M.; Goossens, H.; Greko, C.; So, A. D.; Bigdeli, M.; Tomson, G.; Woodhouse, W.; Ombaka, E.; Peralta, A. Q.; Qamar, F. N.; Mir, F.; Kariuki, S.; Bhutta, Z. A.; Coates, A.; Bergstrom, R.; Wright, G. D.; Brown, E. D.; Cars, O. Antibiotic resistance-the need for global solutions. *Lancet Infect. Dis.* **2013**, *13*, 1057–1098, doi:10.1016/S1473-3099(13)70318-9.

25. Domingo-Calap, P.; Delgado-Martínez, J. Bacteriophages: protagonists of a post-antibiotic era. *Antibiotics* **2018**, *7*, 66, doi:10.3390/antibiotics7030066.

26. Lin, D. M.; Koskella, B.; Lin, H. C. Phage therapy: An alternative to antibiotics in the age of multi-drug resistance. *World J Gastrointest Pharmacol Ther* **2017**, *8*, 162–173.

27. Potera, C. Phage renaissance. *Environ. Heal. Perspect.* **2013**, *121*, A48–A49.

28. Schooley, R. T.; Biswas, B.; Gill, J. J.; Hernandez-Morales, A.; Lancaster, J.; Lessor, L.; Barr, J. J.; Reed, S. L.; Rohwer, F.; Benler, S.; Segall, A. M.; Taplitz, R.; Smith, D. M.; Kerr, K.; Kumaraswamy, M.; Nizet, V.; Lin, L.; Mccauley, M. D.; Strathdee, S. A.; Benson, C. A.; Pope, R. K.; Leroux, B. M.; Picel, A. C.; Mateczun, A. J.; Cilwa, K. E.; Regeimbal, J. M.; Estrella, L. A.; Wolfe, D. M.; Henry, M. S.; Quinones, J.; Salka, S.; Bishop-Lilly, K. A.; Young, R.; Hamilton, T. Development and use of personalized bacteriophage-based therapeutic cocktails to treat a patient with a disseminated resistant *Acinetobacter baumannii* infection. *Antimicrob. Agents Chemother.* **2017**, *61*, e00954-17.

29. Strange, R. N.; Scott, P. R. Plant disease: a threat to global food security. *Annu Rev Phytopathol* **2005**, *43*, 83–116, doi:10.1146/annurev.phyto.43.113004.133839.

30. Buttimer, C.; McAuliffe, O.; Ross, R. P.; Hill, C.; O'Mahony, J.; Coffey, A. Bacteriophages and bacterial plant diseases. *Front. Microbiol.* **2017**, *8*, doi:10.3389/fmicb.2017.00034.

31. Balogh, B.; Canteros, B. I.; Stall, R. E.; Jones, J. B. Control of citrus canker and citrus bacterial spot with bacteriophages. *Plant Dis.* **2008**, *92*, 1048–1052.

32. Iriarte, F. B.; Obradovic, A.; Wernsing, M. H.; Jackson, L. E.; Balogh, B.; Hong, J. A.; Momol, M. T.; Jones, J. B.; Vallad, G. E. Soil-based systemic delivery and phyllosphere in vivo propagation of bacteriophages: Two possible strategies for improving bacteriophage persistence for plant disease control. *Bacteriophage* **2012**, *2*, 215–224, doi:10.4161/bact.23530.

33. Bae, J. Y.; Wu, J.; Lee, H. J.; Jo, E. J.; Murugaiyan, S.; Chung, E.; Lee, S.-W. Biocontrol potential of a lytic bacteriophage PE204 against bacterial wilt of tomato. *J. Microbiol. Biotechnol.* **2012**, *22*, 1613–1620, doi:10.4014/jmb.1208.08072.

34. Czajkowski, R.; Ozymko, Z.; De Jager, V.; Siwinska, J.; Smolarska, A.; Ossowicki, A.; Narajczyk, M.; Lojkowska, E. Genomic, proteomic and morphological characterization of two novel broad host lytic bacteriophages ΦPD10.3 and ΦPD23.1 infecting pectinolytic *Pectobacterium* spp. and *Dickeya* spp. *PLoS One* **2015**, *10*, 1–23, doi:10.1371/journal.pone.0119812.

35. Adriaenssens, E. M.; van Vaerenbergh, J.; Vandenheuvel, D.; Dunon, V.; Ceyssens, P. J.; de Proft, M.; Kropinski, A. M.; Noben, J. P.; Maes, M.; Lavigne, R. T4-related bacteriophage LIMEstone isolates for the control of soft rot on potato caused by "Dickeya solani." *PLoS One* **2012**, *7*, e33227,

doi:10.1371/journal.pone.0033227.

36. Jones, J. B.; Vallad, G. E.; Iriarte, F. B.; Obradovic, A.; Wernsing, M. H.; Jackson, L. E.; Balogh, B.; Hong, J. C.; Momol, M. T. Considerations for using bacteriophages for plant disease control. *Bacteriophage* **2012**, *2*, 208–214, doi:10.4161/bact.23857.

37. Young, R. Phage lysis: three steps, three choices, one outcome. *J. Microbiol.* **2014**, *52*, 243–258, doi:10.1007/s12275-014-4087-z.

38. Dewey, J. S.; Savva, C. G.; White, R. L.; Vitha, S.; Holzenburg, A.; Young, R. Micron-scale holes terminate the phage infection cycle. *Proc Natl Acad Sci U S A* **2010**, *107*, 2219–2223, doi:10.1073/pnas.0914030107.

39. Bie'nkowska-Szewczyk, K.; Lipi'nska, B.; Taylor, A. The R gene product of bacteriophage λ is the murein transglycosylase. *MGG Mol. Gen. Genet.* **1981**, *184*, 111–114, doi:10.1007/BF00271205.

40. Schmelcher, M.; Donovan, D. M.; Loessner, M. J. Bacteriophage endolysins as novel antimicrobials. *Future Microbiol.* **2012**, *7*, 1147–1171, doi:10.2217/fmb.12.97.Bacteriophage.

41. Berry, J.; Rajaure, M.; Pang, T.; Young, R. The spanin complex is essential for lambda lysis. *J Bacteriol* **2012**, *194*, 5667–5674, doi:10.1128/JB.01245-12.

42. Xu, M.; Struck, D. K.; Deaton, J.; Wang, I. N.; Young, R. A signal-arrest-release sequence mediates export and control of the phage P1 endolysin. *Proc Natl Acad Sci U S A* **2004**, *101*, 6415–6420, doi:10.1073/pnas.0400957101.

43. Cahill, J.; Young, R. *Phage lysis: multiple genes for multiple barriers*; 1st ed.; Elsevier Inc., 2018; Vol. 103;.

44. Bernhardt, T. G.; Wang, I. N.; Struck, D. K.; Young, R. Breaking free: "Protein antibiotics" and phage lysis. *Res. Microbiol.* **2002**, *153*, 493–501, doi:10.1016/S0923-2508(02)01330-X.

45. Abdelkader, K.; Gerstmans, H.; Saafan, A.; Dishisha, T.; Briers, Y. The preclinical and clinical progress of bacteriophages and their lytic enzymes: the parts are easier than the whole. *Viruses* **2019**, *11*, 96, doi:10.3390/v11020096.

46. Briers, Y.; Lavigne, R. Breaking barriers: expansion of the use of endolysins as novel antibacterials against Gram-negative bacteria. *Future Microbiol.* **2015**, *10*, 377–390.

47. Nelson, D.; Loomis, L.; Fischetti, V. A. Prevention and elimination of upper respiratory colonization of mice by group A streptococci by using a bacteriophage lytic enzyme. *Proc Natl Acad Sci U S A* **2001**, *98*, 4107–4112, doi:10.1073/pnas.061038398.

48. Thummeepak, R.; Kitti, T.; Kunthalert, D.; Sitthisak, S. Enhanced antibacterial activity of *Acinetobacter baumannii* bacteriophage ØABP-01 endolysin (LysABP-01) in combination with colistin. *Front. Microbiol.* **2016**, *7*, 1–8, doi:10.3389/fmicb.2016.01402.

49. Briers, Y.; Walmagh, M.; Lavigne, R. Use of bacteriophage endolysin EL188 and outer membrane permeabilizers against *Pseudomonas aeruginosa*. *J. Appl. Microbiol.* **2011**, *110*, 778–785, doi:10.1111/j.1365-2672.2010.04931.x.

50. Gerstmans, H.; Criel, B.; Briers, Y. Synthetic biology of modular endolysins. *Biotechnol Adv* **2017**, doi:10.1016/j.biotechadv.2017.12.009.

51. Briers, Y.; Walmagh, M.; Puyenbroeck, V. Van; Cornelissen, A.; Cenens, W.;

Aertsen, A.; Oliveira, H. Engineered Endolysin-Based "Artilysins" To Combat Multidrug-Resistant Gram-Negative Pathogens. *MBio* **2014**, *5*, e01379-14, doi:10.1128/mBio.01379-14.Editor.

52. Johnson, M. C.; Tatum, K. B.; Lynn, J. S.; Brewer, T. E.; Lu, S.; Washburn, B. K.; Stroupe, M. E.; Jones, K. M. *Sinorhizobium meliloti* Phage PhiM9 Defines a New Group of T4 Superfamily Phages with Unusual Genomic Features but a Common T=16 Capsid. *J Virol* **2015**, *89*, 10945–10958, doi:10.1128/JVI.01353-15.

53. Kropinski, A. M.; Van Den Bossche, A.; Lavigne, R.; Noben, J. P.; Babinger, P.; Schmitt, R. Genome and proteome analysis of 7-7-1, a flagellotropic phage infecting *Agrobacterium* sp H13-3. *Virol. J.* **2012**, *9*, doi:10.1186/1743-422X-9-102.

54. Wibberg, D.; Blom, J.; Jaenicke, S.; Kollin, F.; Rupp, O.; Scharf, B.; Schneiker-Bekel, S.; Sczcepanowski, R.; Goesmann, A.; Setubal, J. C.; Schmitt, R.; Pühler, A.; Schlüter, A. Complete genome sequencing of *Agrobacterium* sp. H13-3, the former *Rhizobium lupini* H13-3, reveals a tripartite genome consisting of a circular and a linear chromosome and an accessory plasmid but lacking a tumor-inducing Ti-plasmid. *J. Biotechnol.* **2011**, *155*, 50–62, doi:10.1016/j.jbiotec.2011.01.010.

55. Mansfield, J.; Genin, S.; Magori, S.; Citovsky, V.; Sriariyanum, M.; Ronald, P.; Dow, M.; Verdier, V.; Beer, S. V; Machado, M. A.; Toth, I.; Salmond, G.; Foster, G. D. Top 10 plant pathogenic bacteria in molecular plant pathology. *Mol Plant Pathol* **2012**, *13*, 614–629, doi:10.1111/j.1364-3703.2012.00804.x.

56. Pulawska, J. Crown Gall of Stone Fruits and Nuts, Economic Significance and Diversity of its Causal Agents: Tumorigenic *Agrobacterium* spp. *J. Plant Pathol.*

2010, *92*, S1.87-S1.98.

57. Escobar, M. A.; Dandekar, A. M. Agrobacterium tumefaciens as an agent of disease. *Trends Plant Sci.* **2003**, *8*, 380–386, doi:10.1016/S1360-1385(03)00162-6.

58. Watson, B.; Currier, T. C.; Gordon, M. P.; Chilton, M.-D.; Nester, E. W. Plasmid required for virulance of *Agrobacterium tumefaciens*. *J. Bacteriol* **1975**, *123*, 255–264.

59. Bourras, S.; Rouxel, T.; Meyer, M. *Agrobacterium tumefaciens* Gene Transfer: How a Plant Pathogen Hacks the Nuclei of Plant and Nonplant Organisms. *Phytopathology* **2015**, *105*, 1288–1301, doi:10.1094/PHYTO-12-14-0380-RVW.

60. Guyon, P.; Chilton, M. D.; Petit, A.; Tempe, J. Agropine in "null-type" crown gall tumors: Evidence for generality of the opine concept. *Proc Natl Acad Sci U S A* **1980**, doi:10.1073/pnas.77.5.2693.

61. Gelvin, S. *Agrobacterium*-mediated plant transformation: the biology behind the "gene-jockeying" tool. *Microbiol. Mol. Biol. Rev.* **2003**, *67*, 16–37, doi:10.1128/MMBR.67.1.16-37.2003.

62. Kim, J. G.; Park, B. K.; Kim, S. U.; Choi, D.; Nahm, B. H.; Moon, J. S.; Reader, J. S.; Farrand, S. K.; Hwang, I. Bases of biocontrol: sequence predicts synthesis and mode of action of agrocin 84, the Trojan horse antibiotic that controls crown gall. *Proc Natl Acad Sci U S A* **2006**, *103*, 8846–8851, doi:10.1073/pnas.0602965103.

63. Kerr, A.; Panagopoulos, C. G. Biotypes of Agrobacterium radiobacter var. tumefaciens and their biological control. *J. Phytopathol.* **1977**, doi:10.1111/j.1439-0434.1977.tb03233.x.

64. Süle, S.; Kado, C. I. Agrocin resistance in virulent derivatives of *Agrobacterium*

tumefaciens harboring the pTi plasmid. *Physiol. Plant Pathol.* **1980**, *17*, 347–356.

Chapter 2

Expression of a peptidoglycan hydrolase from lytic bacteriophages Atu_ph02 and Atu_ph03 triggers lysis of *Agrobacterium tumefaciens*

Author Contributions

HA and JR conducted experiments for Figure 2-1. HA contributed data to Figures 2-2 through 2-5. GPS contributed to the development of methods for the initial isolation of phages. HA, GPS, and PB contributed to writing and editing of the manuscript.

Published as

Attai, H., Rimbey, J., Smith, G. P., and Brown, P. J. B. (2017). Expression of a peptidoglycan hydrolase from lytic bacteriophages Atu_ph02 and Atu_ph03 triggers lysis of *Agrobacterium tumefaciens*. *Appl. Environ. Microbiol.* 83, e01498-17. doi:10.1128/AEM.01498-17.

ABSTRACT

In order to provide food security, innovative approaches to preventing plant disease are currently being explored. Here, we demonstrate that lytic bacteriophages and phage lysis proteins are effective at triggering lysis of the phytopathogen *Agrobacterium tumefaciens*. Phages Atu_ph02 and Atu_ph03 were isolated from wastewater and induce lysis of C58-derived strains of *A. tumefaciens*. Co-inoculation of *A. tumefaciens* with phage onto potato discs limits tumor formation. The genomes of Atu_ph02 and Atu_ph03 are nearly identical and ~42% identical to T7 supercluster phages. *In silico* attempts to find a canonical lysis cassette were unsuccessful; however, we found a putative phage peptidoglycan hydrolase (PPH), which contains a C-terminal transmembrane domain. Remarkably, endogenous expression of *pph* in the absence of additional phage genes causes a block in cell division and subsequent lysis of *A. tumefaciens* cells. When the presumed active site of the N-acetylmuramidase domain carries an inactivating mutation, PPH expression causes extensive cell branching due to a block in cell division but does not trigger rapid cell lysis. In contrast, mutation of positively charged residues at the extreme C-terminus of PPH causes more rapid cell lysis. Together, these results suggest that PPH causes a block in cell division and triggers cell lysis through two distinct activities. Finally, the potent killing activity of this single lysis protein can be modulated suggesting that it could be engineered to be an effective enzybiotic.

IMPORTANCE

Characterization of bacteriophages such as Atu_ph02 and Atu_ph03 that infect plant pathogens such as *Agrobacterium tumefaciens* may be the basis of new biocontrol

strategies. First, cocktails of diverse bacteriophages could be used as a preventative measure to limit plant diseases caused by bacteria; a bacterial pathogen is unlikely to develop resistance simultaneously to multiple bacteriophage species. The specificity of bacteriophage treatment for their hosts is an asset in complex communities such as orchards where it would be detrimental to harm the symbiotic bacteria in the environment. Second, bacteriophages are potential sources of enzymes that efficiently lyse bacterial cells. These phage proteins may have a broad specificity, but since proteins do not replicate as phages do, their effect is highly localized, providing an alternative to traditional antibiotic treatments. Thus, studies of lytic bacteriophages that infect *A. tumefaciens* may provide insights for designing preventative strategies against bacterial pathogens.

INTRODUCTION

Crop damage caused by bacterial phytopathogens poses a threat to food security worldwide (1). *Agrobacterium tumefaciens* is one of the top three most scientifically and economically important bacterial plant pathogens (2) and is responsible for significant economic losses in stone fruit and nut production (3, 4). *A. tumefaciens* causes crown gall disease by transforming plant cells with constitutively-expressed genes for production of phytohormones and opines. Opines serve as a custom food source for the bacteria. Increased hormone production causes plant cells to locally over-proliferate (5, 6), leading to tumor formation and reduced transport of water and nutrients throughout the plant. Thus, infected plants often do not achieve maximal crop yields.

Current commercially available biocontrol involves the application of *Agrobacterium radiobacter* strain K84, which releases a bacteriocin called agrocin to

outcompete *A. tumefaciens*. However, only a limited number of *A. tumefaciens* strains are sensitive to agrocin (7, 8) and sensitive strains of *A. tumefaciens* can become resistant (9). Therefore, alternative methods of biocontrol are emerging, including selection and breeding of resistant crops (10-12), chemical treatments (13), and isolation of additional antagonist organisms (14). In this work, we consider the possibility of lytic bacteriophages and phage-encoded lysis proteins as options for *A. tumefaciens* biocontrol.

Bacteriophages and phage-derived proteins have recently been employed against several plant pathogens (15, 16). Lytic phages contain a large reservoir of genes specifically involved in killing their host cells and are attractive as biocontrol agents. Lytic bacteriophages have shown promise in protecting tomato plants from wilting caused by *Ralstonia solanacearum* (17, 18), oranges from citrus canker caused by *Xanthomonas axonopodis* (19), leeks from bacterial blight caused by *Pseudomonas syringae* pv. *porri* (20), and kiwi from canker caused by *Pseudomonas syringae* pv. *actinidiae* (21). While these studies indicate the potential of bacteriophages to serve as biocontrol agents, further optimization of phage replication and lysis will be necessary to effectively scale for use in crop fields and orchards. As of yet, bacteriophages have not been employed as biocontrol against *A. tumefaciens*.

As research on bacteriophages as biocontrol agents has recently grown in popularity, so has the study of phage proteins responsible for their antimicrobial activity (22). The term "enzybiotics," first coined in 2001, refers to the direct application of phage endolysins, peptidoglycan hydrolase enzymes that target the bacterial cell wall, to susceptible hosts (23). Benefits of using endolysins include their direct mode of action, low incidence of resistance, and potential for protein optimization (24). The use of

exogenously applied peptidoglycan hydrolases to directly kill bacteria is more established for Gram positive bacteria since they lack an outer membrane barrier; however, in at least some cases, exogenous application of endolysins has been shown to effectively lyse Gram negative bacteria (25). Additional strategies for targeting Gram negative pathogens include co-application of the endolysin with an outer membrane permeabilizer such as ethylenediaminetetraacetic acid (EDTA) (26), or engineering an endolysin to gain the ability to lyse bacteria from the outside by absorption through the outer membrane (27).

Lytic bacteriophages that infect *A. tumefaciens* have been isolated from soil and sewage samples (28-30) suggesting that there is untapped potential for using bacteriophage or endolysins as biocontrol agents against the pathogen; however, only *Agrobacterium* sp H13-3 phage 7-7-1 has been subject to genomic characterization (31). In this work, we isolate and describe 2 lytic bacteriophages that specifically infect a subset of *A. tumefaciens* strains. These bacteriophages are closely related and contain a novel endolysin with potential antimicrobial activity.

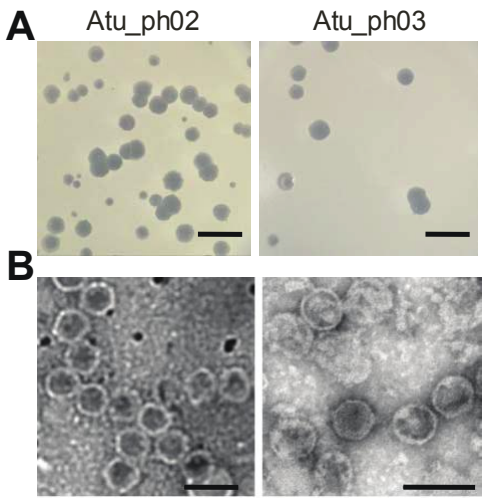
RESULTS AND DISCUSSION

Isolation and characterization of bacteriophages that infect *A. tumefaciens* C58.

Wildtype strains of *A. tumefaciens* are well known for their ability to cause crown gall disease and this ability is dependent on the presence of the tumor inducing plasmid, pTi (3, 32, 33). In this work, *A. tumefaciens* strain C58 was selected as the host strain for isolation of bacteriophage since it was isolated from a cherry tree tumor (34), the complete genome sequence is available (35, 36), and it has been widely studied as a pathogen (37).

Using a modified phage enrichment protocol (38), we isolated two bacteriophages, called Atu_ph02 and Atu_ph03, from samples obtained from the Columbia, MO, regional wastewater treatment plant. Presence of bacteriophage in supernatants from cleared *A. tumefaciens* cultures was confirmed by spot and classic plaque assays. Virions were concentrated and partially purified by precipitation with polyethylene glycol (PEG) and differential centrifugation. Phages Atu_ph02 and Atu_ph03 form large and clear plaques on a lawn of *A. tumefaciens* (Figure 2-1A). Transmission electron microscopy (TEM) of the virions reveals icosahedral heads with diameters ~58 nm, and short tails (Figure 2-1B). Subterminal tail fibers are not visible through TEM. This morphology suggests that these bacteriophages are podoviruses (39).

The host ranges of the phages were assessed by spotting phage stock dilutions on a range of bacteria (Figure 2-1C). Each bacteriophage exhibits a narrow host range, only infecting a subset of *A. tumefaciens* strains. Both phages infect C58-derived strains (C58, EHA105, EHA101, GV3101, NTL4) with the exception of AGL-1, which carries a mutation in *recA* (40). This suggests that RecA, an enzyme responsible for homologous recombination and DNA repair, may be required for efficient bacteriophage infection. Atu_ph02 and Atu_ph03 cannot infect *A. tumefaciens* strains that are not derived from C58 (LBA4404 and Chry5) or other tested species including *Agrobacterium vitis*, *Sinorhizobium meliloti, Caulobacter crescentus,* and *Escherichia coli*. A narrow-host range is considered to be an important asset when assessing the potential of bacteriophages as biocontrol agents against phytopathogens, as it minimizes harm to other beneficial microbes in the rhizosphere. Ideally, a cocktail of lytic bacteriophages



C

	Strain	Atu_ph02	Atu_ph03
Gram-negative, Alphaproteobacteria, *A. tumefaciens*, C58-derived	C58	+	+
	EHA105	+	+
	EHA101	+	+
	GV3101	+	+
	NTL4	+	+
	AGL1	-	+*
	Chry5	-	-
	LBA4404	-	-
	Agrobacterium vitis S4	-	-
	Sinorhizobium meliloti RM41	-	-
	Caulobacter crescentus CB15	-	-
	Escherichia coli DH5α	-	-

Figure 2-1. Characterization of plaque and bacteriophage morphologies. (A) Plaques formed on a lawn of *Agrobacterium tumefaciens* C58 are shown for each bacteriophage. Scale bars = 10 mm. (B) Transmission electron micrographs reveal the morphology of each bacteriophage. Scale bars = 100 nm. (C) Specificities of bacteriophages were determined by spotting dilutions of phage on a lawn of the host bacterium. + indicates that plaques were observed and – indicates that plaques were not observed. *Plaques were observed only at titers ~1000 times higher than required for plaque formation on other host strains. Strain AGL1 contains an insertion mutation in *recA* to stabilize recombination plasmids.

requiring different host factors would be deployed to reduce the incidence of host resistance (21).

Phage treatment causes cell lysis and results in reduced pathogenicity of *A. tumefaciens.*

To further assess the potential of these bacteriophages as biocontrol agents, we measured the effect of phage infection on *A. tumefaciens* grown in liquid medium, growth on agarose pads, and on *A. tumefaciens*-induced tumor formation on potato discs. Bacterial growth curves indicate the rate at which phages can inhibit the growth of their bacterial hosts. *A. tumefaciens* cells infected with Atu_ph02 or Atu_ph03 at a multiplicity of infection (MOI) of 0.001 grew for ~3 h post infection prior to the onset of cell lysis (Figure 2-2A). Timelapse microscopy shows uninfected WT cells form microcolonies within 7 h (Figure 2-2B top) whereas cells infected with Atu_ph02 (Figure 2-2B center) or Atu_ph03 (Figure 2-2B bottom) at an MOI of 0.01 initially grow and divide but begin lysing 5 h after infection. Since phages are released after the first cell lyses, the remaining cells are subsequently infected and all cells in the field are lysed within the next 2 h (Figure 2-2B center and bottom). In other representative fields, cells which are not initially infected form relatively large microcolonies; however, cells on the periphery of these microcolonies later lyse. Since this lysis event is never observed in the absence of phage, we infer that these cells are susceptible to the phage particles which have likely diffused through the agarose.

The qualitative potato tumor assay uses potato discs to mimic wound sites and evaluate virulence of *A. tumefaciens* cells on plant hosts (41). Potato discs inoculated

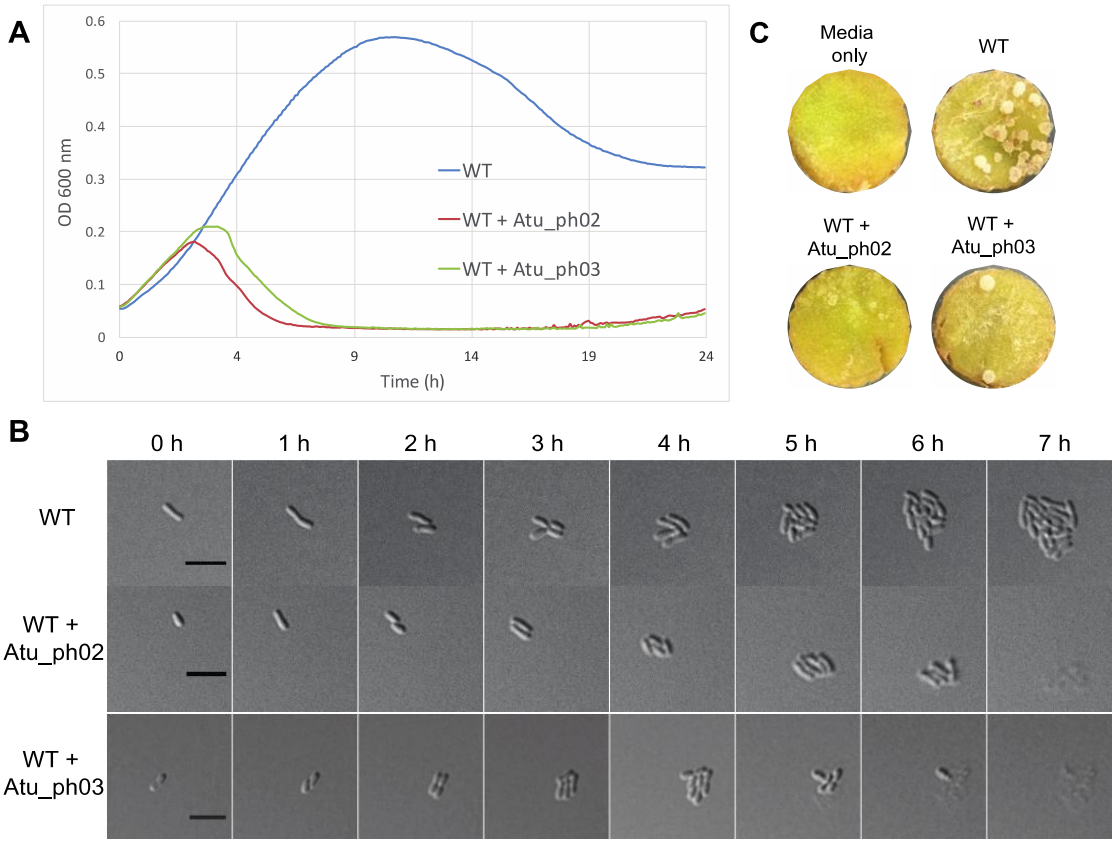


Figure 2-2. Bacteriophages Atu_ph02 and Atu_ph03 lyse *A. tumefaciens* cells. (A) Growth curve of *A. tumefaciens* strain C58 (WT) infected with Atu_ph02 and Atu_ph03 at MOI 0.001. Results from a representative growth curve are shown. Each line is the average of four replicate wells. (B) Time-lapse microscopy showing the growth of uninfected WT cells (upper panel), WT cells infected with Atu_ph02 at an MOI of 0.01 (center panel), and WT cells infected with Atu_ph03 at an MOI of 0.01 (lower panel). Scale bars = 5 µm. (C) Representative potato discs treated with media (top left), WT *A. tumefaciens* cells (top right), a mixture of WT *A. tumefaciens* cells and Atu_ph02 at MOI 1.0 (bottom left), and a mixture of WT *A. tumefaciens* cells and Atu_ph03 at MOI 1.0 (bottom right) after 14 days of incubation in a humid chamber. White spots on the potato discs are *A. tumefaciens* induced tumors.

with *A. tumefaciens* formed tumors after 14 days of infection (Figure 2-2C, top right). Co-inoculation with Atu_ph02 or Atu_ph03 at an MOI of 1.0 reduces the number of tumors formed on the potato disc (Figure 2-2C, bottom). Since bacteria readily evolve resistance to individual bacteriophages, phages Atu_ph02 or Atu_ph03 alone are unlikely to be effective biocontrol agents. However, these phages may be valuable as a component of a bacteriophage cocktail. Together, the growth curve, microscopy, and potato tumor assay in Figure 2-2 show that Atu_ph02 and Atu_ph03 are lytic phages capable of rapidly killing *A. tumefaciens* and potentially protecting plants from infection.

Phages Atu_ph02 and Atu_ph03 belong to the T7 supercluster. Although Atu_ph02 and Atu_ph03 are similar in morphology (Figure 2-2B), lysis rate (Figure 2-2A), and genome size (Supplemental Figure 2-S1A), the genomes are not identical based on restriction fragment pattern analysis (Supplemental Figure 2-S1B). Therefore, we sequenced both genomes in order to gain insights into the mechanism of phage-mediated host cell lysis.

Phages Atu_ph02 and Atu_ph03 contain nearly identical small genomes of ~45 kbp encoding 55 and 58 open reading frames (ORFs), respectively (Figure 2-3, Table 2-1, Supplemental Table 2-S1). Dot plot analysis of the Atu_ph02 and Atu_ph03 sequences reveals that these genomes are almost entirely syntenic, with only a few regions indicative of small deletions or insertions (Supplemental Figure 2-S2A). Comparison of the 52 shared protein sequences reveals a high degree of similarity: 23 are 100% identical, another 23 are 90-99.88% identical, and the remaining 6 are 49%-88% identical (Supplemental Figure 2-S2B and Supplemental Table 2-S1).

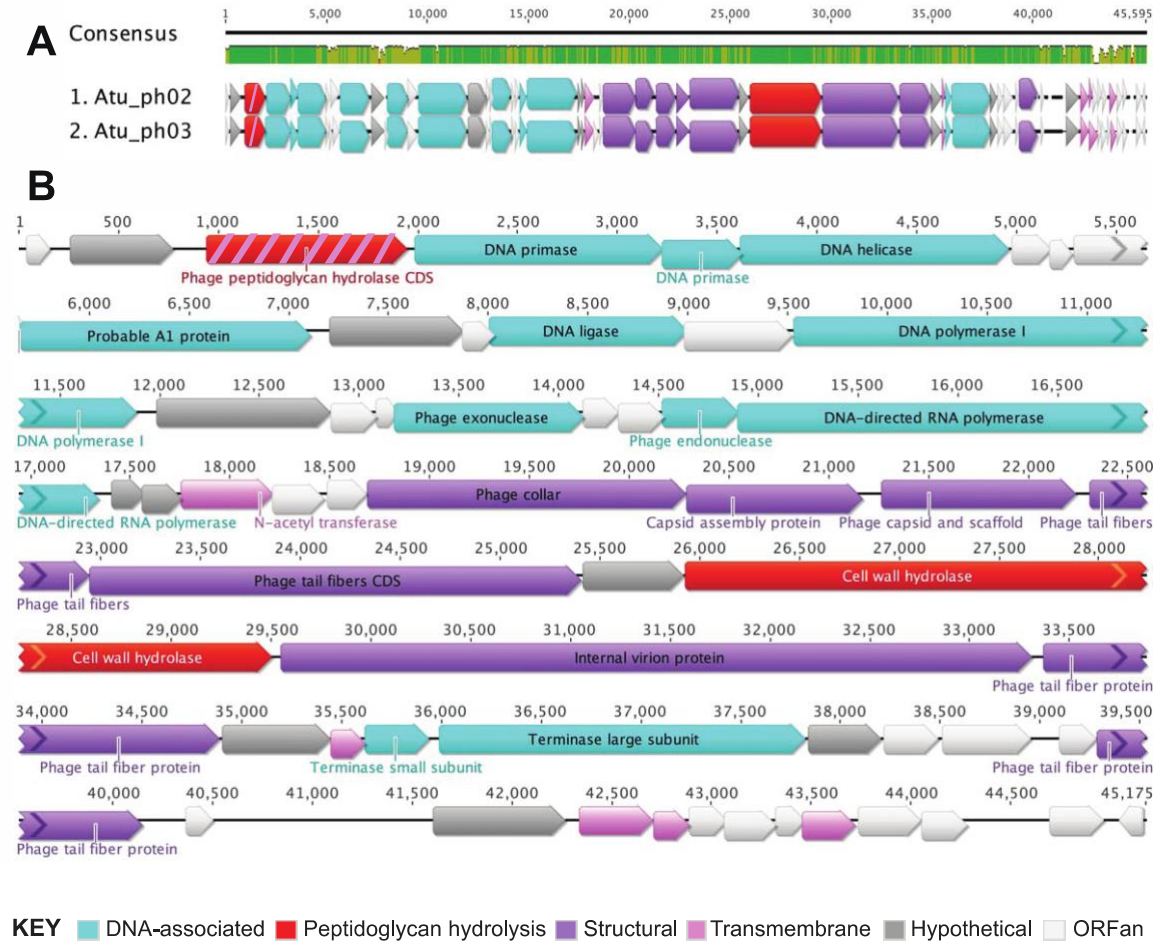


Figure 2-3. Genome organization of Atu_ph02 and Atu_ph03. (A) Genome alignments of Atu_ph02 and Atu_ph03 created using the MUSCLE plugin in Geneious. Consensus identity in green = 100% identical, gold = 30-100%, red = <30%, no color = 0% (B) Gene annotations for the Atu_ph03 phage genome. Color-coding indicates functional classification of the open reading frames. Protein PPH, which is classified as both a peptidoglycan hydrolysis protein (red) and a transmembrane protein (pink) is colored red with pink stripes

The genomes of Atu_ph02 and Atu_ph03 are organized in functional blocks (Figure 2-3 and Supplemental Table 2-S2) including genes encoding DNA-associated proteins (Figure 2-3B, light blue arrows) and genes predicted to function in phage morphogenesis (Figure 2-3B, purple arrows). Remarkably, 60% of the ORFs encode hypothetical proteins of unknown function (Figure 2-3B, white and grey arrows), ~70% of which (23 in Atu_ph02 and 26 in Atu_ph03) are ORFans (Figure 2-3B, white arrows), meaning they share no significant homology with existing proteins in the non-redundant database (42).

Whole genome comparisons using the Atu_ph03 nucleic acid sequence reveal that this phage genome is ~42% identical to both the T7-like *Rhizobium etli* phage RHEph02 and Phage MedPE-SW-cel-C56. Whole genome alignments identified syntenic regions among these phage genomes (Figure 2-4A). Similar to *Rhizobium etli* phages RHEph02, RHEph03, RHEph08 and RHEph09 (38), phages Atu_ph02 and Atu_ph03 can be classified as members of the T7 supercluster based on the similarity of the genome organization and the presence of core T7 genes. There are 4 conserved core T7 genes, predicted to encode the T7-like RNA polymerase, large terminase, and structural proteins (Supplemental Table 2-S2). These 4 core genes are also present in RHEph02, RHEph08, and Phage MedPE-SWcel-C56 (Supplemental Table 2-S2). A phylogenetic analysis using the gene encoding the large terminase reveals that *Agrobacterium* phages Atu_ph02 and Atu_ph03, *Rhizobium etli* phages, and Phage Med-SWcel-C56 form a distinct clade and share a common ancestor with other T7-like bacteriophages that target non-enteric hosts, including *Pseudomonas* phage ϕKMV (Figure 2-4B). Furthermore, this phylogeny supports the classification of these *Agrobacterium* phages within the ϕKMV-like cluster

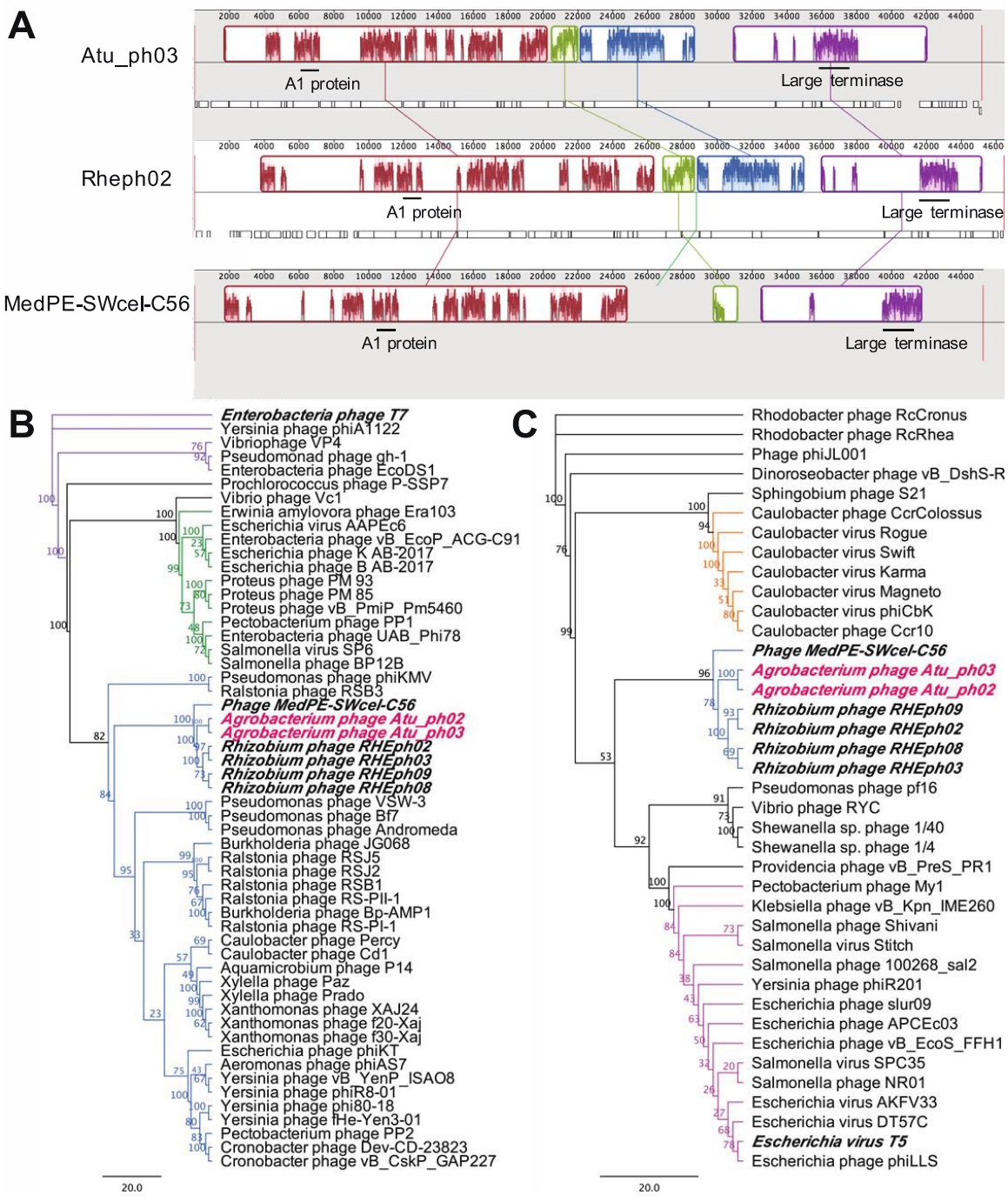


Figure 2-4. Genome wide syntenic mapping and key protein phylogenies. (A) Whole genome alignments of Atu_ph03 with *Rhizobium* phage RHEph02 and phage MedPE-SWceI-C56. Position of genes encoding the A1 protein and large terminase are shown

with black bars. (B) Phylogenetic analysis of the large terminase subunit from Atu_ph02 and Atu_ph03 and other T7-like phages. Purple nodes indicate T7-like phages, green indicate Sp-6-like phages, and blue indicate ϕKMV-like phages. (C) Phylogenetic analysis of the probable A1 protein from Atu_ph02 and Atu_ph03, closely related *Rhizobium* phages, phage MedPE-SWcel-C56, and phages belonging to T5 family. Blue nodes indicate ϕKMV-like phages, orange indicate phi-Cbk-like phages, and pink indicate T5-like phages. Scale bars represent the number of amino acid substitutions per site. Numerical value on each node represents the bootstrap value of 100 replicates.

of phages, which includes characterized phages that infect Alphaproteobacteria such as *Caulobacter* phages ϕCd1 (43) and Percy (44), and *Ralstonia* phage ϕRSB1(45). Furthermore, whole genome comparisons using Atu_ph03 and *Pseudomonas* phage ϕKMV reveal that the genomes are 42% identical. A total of 13 of the predicted proteins in the Atu_ph03 genome have homologous proteins in ϕKMV which share at least 24% identity, including key proteins that function in DNA metabolism and virion structure and assembly (Supplemental Table 2-S2).

Phage Atu_ph02 and Atu_ph03 contain a putative A1 protein. A surprising observation within the genomes of Atu_ph02 and Atu_ph03 is the presence of a homolog of the A1 protein from phage T5 (Figure 2-3B, Supplemental Table 2-S2). The probable A1 protein is conserved within the clade containing the *Rhizobium etli* phages and Phage Med-SWcel-C56 suggesting that the ancestor of these phages acquired the gene horizontally since it is not prevalent among ϕKMV-like bacteriophages (Figure 2-4C, Supplemental Table 2-S2). The putative A1 protein (Gp10) from Atu_ph03 is 486 amino acids and is 38% identical to the phage T5 A1 protein (Supplemental Table 2-S2). In phage T5, A1 mutants are defective in multiple processes, including degrading host DNA, downregulating pre-early gene expression (46), and completion of T5 DNA transfer into the host cell (47). The probable A1 proteins have no readily identifiable functional domains, transmembrane domains, or signal peptides, thus we cannot speculate on function of the A1 protein in phages Atu_ph02 and Atu_ph03. Homologs to the A1 protein are found in other non-T5 phages including *Caulobacter* phage ϕCBK (48, 49) (Figure 2-4C). In phage ϕCBK and related phages, the A1 protein is located with the

DNA replication module, suggesting that this protein may function to alter ϕCBK gene expression through an interaction with host RNA polymerase (49). The putative A1 proteins in *Agrobacterium* phages Atu_ph02 and Atu_ph03 join a small family of A1-related proteins found in distinct clades of non-T5-like phages (Figure 2-4C) and bacterial genomes, though the function remains uncharacterized in all cases (48).

Atu_ph02 and Atu_ph03 phage lysis proteins. Many bacteriophages that infect Gram-negative hosts contain a lysis cassette consisting of an endolysin and accessory proteins (50). Most of these endolysins have a globular structure containing a single enzymatic active domain (EAD) and cannot reach the periplasm independently (50-52). In the canonical holin-endolysin system, endolysins accumulate in the cytoplasm until a sufficient quantity of holins are inserted in the inner membrane to form homo-oligomeric pores that allow endolysins to enter the periplasm (50). An alternative strategy is used by signal-arrest-release (SAR) endolysins that contain an N-terminal type II signal anchor, which embeds the inactive enzyme in the inner membrane until pinholins cause membrane depolarization and release of the endolysin to the periplasm (50). Both the holin-endolysin and pinholin-SAR endolysin systems rely on spanins to fuse the inner and outer membranes to complete cell lysis. In addition to these strategies, *in silico* and experimental analysis suggest that some endolysins contain N-terminal signal sequences which may enable delivery to the periplasm via the Sec machinery (51, 53, 54).

Among the phage lysis proteins, endolysins are most readily identified in phage genomes due to the presence of peptidoglycan hydrolase domains. Thus, the genomes of Atu_ph02 and Atu_ph03 were searched for putative endolysin proteins. All of the

predicted ORFs were translated and searched for the presence of domains encoding peptidoglycan hydrolases. Two candidate peptidoglycan hydrolases were identified (Figure 2-3, red arrows; Table 2-S2). Gp35 contains a putative cell wall hydrolase (pfam07486) and is located in close proximity to the putative tail proteins and internal virion proteins suggesting that this hydrolase may function as a virion associated lysin. Gp3 contains a putative N-acetylmuramidase domain (DUF3380) suggesting that it may hydrolyze peptidoglycan.

To determine if either Gp35 or Gp3 are part of a canonical lysis cassette, we searched for potential accessory proteins, such as holins or pinholins, by screening each putative protein sequence for the presence of transmembrane domains. Phage genome Atu_ph03 encodes a total of 6 predicted transmembrane proteins (Figure 2-3, pink arrows); however, none of these proteins are in close proximity to the putative peptidoglycan hydrolases. This observation suggests that Atu_ph03 may not contain a canonical lysis cassette. Remarkably, one of the transmembrane-containing proteins also contains the putative phage peptidoglycan hydrolase domain DUF3380 (Figure 2-3, red and pink striped arrows). The sequence of PPH is 100% identical in phages Atu_ph02 and Atu_ph03 (Figure 2-3A, Supplemental Table 2-S1) and is not present in the genomes of the closely related *Rhizobium etli* phages or Phage MedPE-SWcel-C56 (Supplemental Table 2-S2). To determine if PPH contributes to cell lysis, *pph* (*gp*3) was subject to bioinformatic and genetic characterization.

Phage peptidoglycan hydrolase is sufficient to induce *A. tumefaciens* cell lysis. PPH contains a domain (DUF3380) that is found in bacterial and viral proteins that bind

peptidoglycan (Figure 2-5A and Supplemental Figure 2-S3). Recent characterization of an endolysin (Gp110) from a *Salmonella* phage 10 containing DUF3380 revealed that this domain functions as a N-acetylmuramidase and cleaves the β1-4 glycosidic bond between N-acetylmuramic acid and N-acetylglucosamine in peptidoglycan (55). Predicted PPH transmembrane topology according to a hidden Markov model (56) orients the transmembrane domain with the EAD in the periplasm and the short C-terminal tail in the cytosol (Figure 2-5A). Remarkably, if PPH functions as an endolysin, this suggests that PPH may utilize an atypical mechanism of entering the periplasm and causing host cell lysis.

In light of these observations, we sought to determine if expression of PPH is sufficient for cell lysis. A plasmid containing *pph* under the control of a *lac* promoter (pSRKKm-Plac-PPH) was introduced into WT *A. tumefaciens* cells allowing observations of cell viability, cell growth, cell morphology, and cell lysis under conditions where *pph* is induced by the presence of isopropyl β-D-1-thiogalactopyranoside (IPTG) (Figure 2-5). Growth curve analysis shows that induction of *pph* leads to growth inhibition within 4 hours (Figure 2-5C), in contrast to normal growth exhibited by uninduced cells (Supplemental Figure 2-S4A). To measure viability, we grew *A. tumefaciens* containing pSRKKm-Plac-PPH in the absence of IPTG to mid-exponential phase. Next, we spotted dilutions of *A. tumefaciens* containing pSRKKm-Plac-PPH onto plates with and without IPTG. In the presence of inducer, there is a 5-log loss in viability of *A. tumefaciens* cells (Figure 2-5D) compared to the same strain in the absence of inducer (Supplemental Figure 2-S4B). These results suggest that accumulation of PPH is sufficient to inhibit *A. tumefaciens* growth. Next, timelapse microscopy was used to determine if *pph* induction

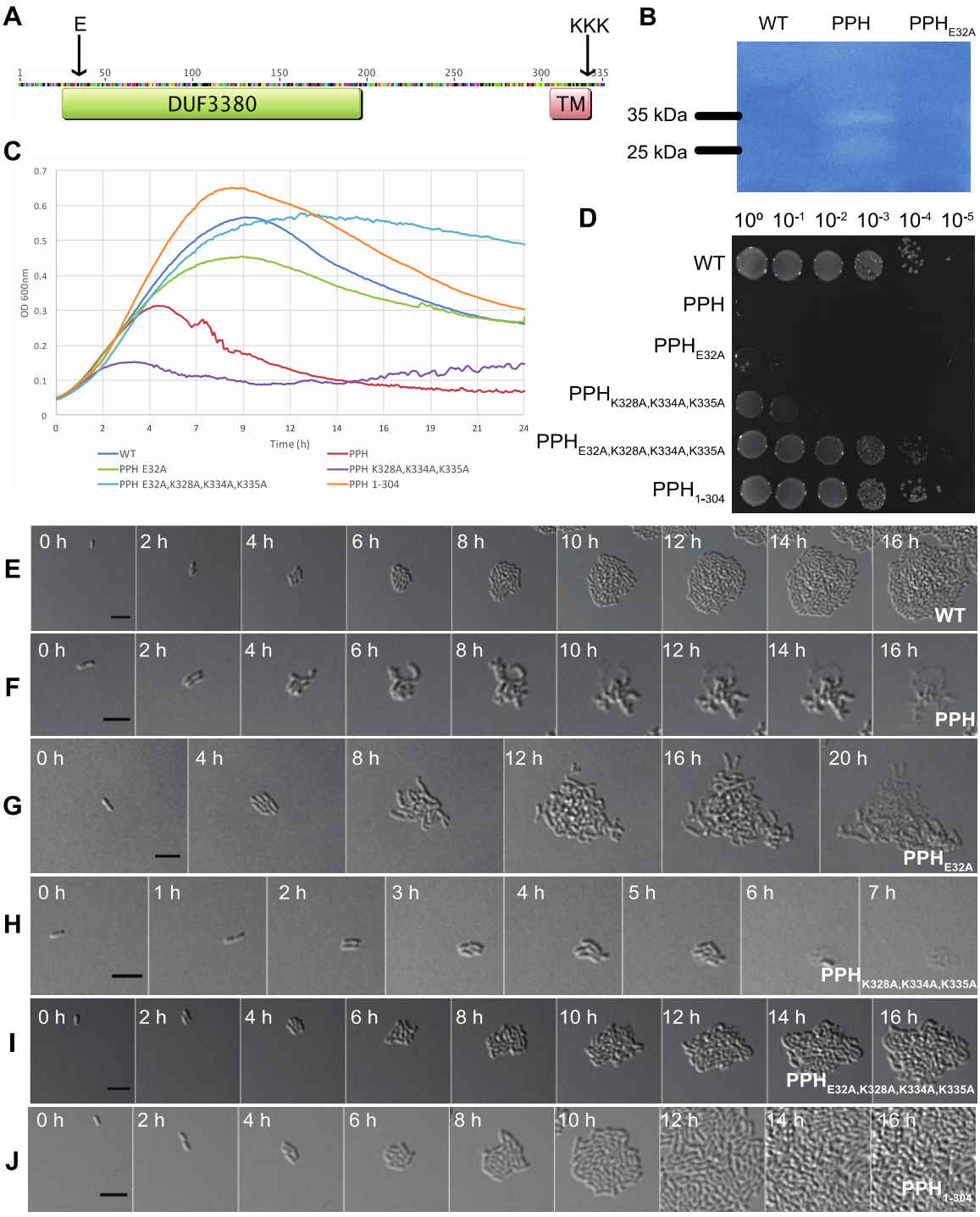


Figure 2-5. Characterization of phage peptidoglycan hydrolase (PPH) and its effect on *A. tumefaciens*. (A) Predicted PPH protein topology. Arrows indicate sites subject to point mutagenesis. (B) Zymogram of 30 µg whole cell lysates from *A. tumefaciens* lacking

PPH, expressing PPH, and expressing PPH_{E32A}. (C) Growth curve of *A. tumefaciens* growth when expressing plasmid pSRKKm-Plac with variants of *pph* under induced conditions. (D) Cell viability of *A. tumefaciens* containing plasmids to express unmutated and variant PPH grown under induced conditions. Time-lapse microscopy of an *A. tumefaciens* cell containing (E) empty pSRKKm-Plac vector, (F) pSRKKm-Plac-PPH, (G) pSRKKm-Plac-PPH_{E32A}, (H) pSRKKm-Plac-$PPH_{K328A,K334A,K335A}$, (I) pSRKKm-Plac-$PPH_{E32A,K328A,K334A,K335A}$, and (J) pSRKKm-Plac-$PPH_{1-304}$. Cells were induced for PPH expression for 1 hour prior to initiation of time-lapse microscopy. Scale bars = 5 μm.

triggers lysis of *A. tumefaciens* cells. *A. tumefaciens* cells with an empty pSRKKm-Plac plasmid grew and divided to form microcolonies (Figure 2-5E). In contrast to cells infected with Atu_ph02 or Atu_ph03, which lyse rapidly with little change in cellular morphology (Figure 2-2C), PPH induction causes cells to elongate and branch prior to cell lysis (Figure 2-5F). The branching phenotype observed when *pph* is induced in *A. tumefaciens* cells is reminiscent of cells exhibiting a block in cell division (57-63). This observation suggests that PPH may have a dual function: blocking cell division and triggering cell lysis.

We hypothesized that the peptidoglycan hydrolase activity is necessary for PPH induction to trigger cell lysis. Alignments of DUF3380 sequences from bacterial and viral proteins reveal the presence of a conserved glutamate (E), which is presumed to be a catalytic residue, followed by a conserved serine (S) (55) (Supplemental Figure 2-S3). To determine if the predicted catalytic glutamate functions in cell lysis during PPH induction, we mutated PPH to PPH_{E32A} and characterized cell growth (Figure 2-5C), cell viability (Figure 2-5D), and cell lysis of the PPH_{E32A} strain (Figure 2-5G). Cultures of *A. tumefaciens* producing PPH_{E32A} become turbid and increase in optical density based on growth curve analysis (Figure 2-5C), but are not viable when spotted on media containing inducer (Figure 2-5D). Timelapse microscopy of cell growth when PPH_{E32A} is induced explains these seemingly contradictory observations (Figure 2-5G). When PPH_{E32A} is expressed, *A. tumefaciens* cell division is blocked and large, extensively branched cells form. This unusual morphology causes both an increase in light scattering in growth curve analysis and a marked decrease in cell viability. Cells lyse ~4 h later post induction when expressing PPH_{E32A} than when they express PPH (compare Figures 2-5F-G).

Together, these observations suggest that the predicted N-acetylmuramidase domain contributes to PPH-mediated cell lysis. To further test this hypothesis, we assessed the ability of PPH to clear peptidoglycan embedded in an SDS-PAGE gel. A zymogram loaded with equal amounts of protein from whole cell lysates of WT cells, WT cells expressing PPH, and WT cells expressing PPH_{E32A} reveals bands of clearing at the expected size of 35 kDa when PPH is expressed (Figure 2-5B, Supplemental Figure 2-S5). A smaller band of clearing (~25 kDa) is also observed when PPH is expressed and may indicate a degradation product of the PPH. In contrast, no clearing is observed when the PPH_{E32A} variant is expressed (Figure 2-5B, Supplemental Figure 2-S5). The lack of clearing in PPH_{E32A} shows the important role this residue plays in peptidoglycan cleavage. These results suggest that PPH has peptidolglycan hydrolyzing activity, although it is possible that PPH stimulates other peptidoglycan hydrolases in *A. tumefaciens*. Future work using purified proteins will be necessary to confirm these results.

The observation that expression of PPH_{E32A} causes a dramatic cell morphology including very large, branched cells (Figure 2-5G) suggests that PPH causes a block in cell division that is independent of the peptidoglycan hydrolase activity. Since divisome assembly is initiated in the cytoplasm (64) and the predicted topology of PPH suggests that only the C-terminal tail would extend into the cytoplasm, we examined this sequence for any remarkable features and observed that this region is lysine-rich. To determine if the positively charged C-terminus functions in regulation of cell division and timing of cell lysis, we constructed a plasmid encoding a PPH variant in which the lysines have been mutated to alanines ($PPH_{K328A,K334A,K335A}$) and assessed cell growth (Figure 2-5C),

cell viability (Figure 2-5D), and cell lysis (Figure 2-5H). The growth curve reveals that expression of $PPH_{K328A,K334A,K335A}$ causes a growth defect that is significantly more pronounced than that of wild-type PPH (Figure 2-5C) and spotting of cells expressing $PPH_{K328A,K334A,K335A}$ under inducing conditions results in a decrease in cell viability (Figure 2-5D). Remarkably, timelapse microscopy reveals that cells producing $PPH_{K328A,K334A,K335A}$ lyse rapidly (Figure 2-5H). This phenotype is strikingly similar to cells infected with phages Atu_ph02 and Atu_ph03 (Figure 2-2C). Unlike cells expressing PPH, cells that express $PPH_{K328A,K334A,K335A}$ lyse rapidly without a block in cell division. This observation suggests that the positively-charged lysines function in the regulation of PPH-mediated cell killing by contributing to the block in cell division but are not required for lysis.

To confirm the roles of the catalytic E32 and the positively charged lysine residues, the loss of function mutations were combined to create $PPH_{E32A,K328A,K334A,K335A}$. Induction of $PPH_{E32A,K328A,K334A,K335A}$ restores normal growth (Figure 2-5C) and viability (Figure 2-5D). Timelapse microscopy reveals that cells producing $PPH_{E32A,K328A,K334A,K335A}$ continue to elongate and divide, producing microcolonies (Figure 2-5I). Some cells expressing $PPH_{E32A,K328A,K334A,K335A}$ appear to be hypercurved, swollen, or bulging, suggesting that this variant of PPH causes relatively minor defects in the cell wall or cell growth. Together, these data suggest that we have identified important residues responsible for the PPH-mediated block in cell division and cell lysis.

Finally, we sought to determine the contribution of the C-terminal transmembrane domain in PPH-mediated cell lysis. We truncated PPH to remove the C-terminal TM domain and cytosolic tail and expressed PPH_{1-304} in *A. tumefaciens*. Expression of PPH_{1-}

$_{304}$ does not impair growth (Figure 2-5C), viability (Figure 2-5D), or the ability to form microcolonies (Figure 2-5J). Consistent with the predicted topology of PPH, these observations suggest that the cells do not lyse since the peptidoglycan hydrolase does not reach the periplasm. We hypothesize that these cells do not exhibit a block in cell division since the positively-charged residues at the extreme C-terminus of PPH are also absent.

While the possibility of a dual function of PPH in peptidoglycan hydrolysis and blocking cell division is intriguing, additional work is needed to determine if a block in cell division occurs during infection with phage Atu_ph02 or Atu_ph03 and if PPH contributes to a delay in cell division. Since we are artificially expressing PPH in *A. tumefaciens*, we cannot yet determine if this phenotype is an artifact of protein expression or representative of PPH induction during phage infection; however, the ability to abolish the cell division defect and induce rapid lysis by mutating the positively charged C-terminus may suggest a biological role for this region of the peptide. A dual function for PPH in peptidoglycan hydrolysis and blocking cell division is consistent with descriptions of single lysis proteins in phages with smaller genomes. For example, Coliphages ϕX174, MS2, and Qβ encode lysis proteins E, L, and A_2, respectively (65-68). Protein E causes host cell lysis by inhibiting the activity of a host protein involved in peptidoglycan biosynthesis (66). The C-terminus of PPH may have a similar function leading to inhibition of a cell division protein. Indeed, *A. tumefaciens* cells expressing PPH exhibit a branching phenotype similar to that in FtsZ-depleted cells (63) (Figure 2-5F). The inhibition of cell division causes cells to increase in volume, which may benefit the phage by maximizing burst size (69). While it remains unclear if or how the phages

Atu_ph02 and Atu_ph03 regulate the timing of lysis during infection, it appears that the C-terminus of PPH may contribute to the regulation of cell lysis.

Overall, we find that PPH is a potent inhibitor of *A. tumefaciens* cell growth and viability, revealing the possibility that this protein may be engineered to be an even more potent antimicrobial. Future studies will be aimed at characterizing the enzymatic activity of PPH, identifying host factors required for PPH-mediated cell division blocks, and addressing the specificity of PPH as an antimicrobial.

Conclusions. Our laboratory has isolated and characterized two closely related bacteriophages that specifically infect *A. tumefaciens* strains derived from C58. These phages, Atu_ph02 and Atu_ph03, are lytic and lead to host cell lysis. While the potato tumor assay shows that these phages offer partial protection from tumor formation, the use of these phages for biocontrol may be limited unless additional lytic phages are used in combination therapies. In order to investigate the mechanism of host cell lysis, the genomes Atu_ph02 and Atu_ph03 were sequenced, revealing the presence of a putative atypical endolysin, termed phage peptidoglycan hydrolase (PPH). PPH is sufficient for lysis of *A. tumefaciens* cells and appears to have a dual function in disrupting the divisome assembly or function and triggering cell lysis. Mutational analyses suggest that a putative N-acetylmuramidase domain contributes to cell lysis while a positively charged C-terminus causes a block in cell division. The transmembrane domain is hypothesized to aid in delivery of the peptidoglycan hydrolase domain to the periplasm and is necessary for rapid PPH-induced cell lysis. Understanding the mechanism by which PPH blocks cell division, the specific host factor it targets, and its enzymatic

activity on the cell wall will elucidate the mode of action of PPH and determine if PPH shows promise as an enzybiotic. More detailed characterization will be necessary to confirm if PPH disrupts cell division during phage infection and to determine if PPH functions as an endolysin late during the phage infection cycle. This work demonstrates that bacteriophages have evolved additional mechanisms to kill their host cells and illustrates the value of exploring bacteriophage genomes as a source of candidate enzybiotics.

MATERIALS AND METHODS

Bacterial strains and culture conditions. The bacterial strains used in the study are listed in Table 2- 2. *Agrobacterium tumefaciens* strains and *Sinorhizobium meliloti* were cultured in Luria-Bertani (LB) broth, with the exception of *A. tumefaciens* LBA4404, which was cultured using yeast mannitol (YM) media (70). *Agrobacterium vitis* was grown in potato dextrose media (Difco) and *Caulobacter crescentus* was cultured on peptone yeast extract (PYE) media (71). All of these strains were grown in liquid culture at 28°C with shaking. *Escherichia coli* was cultured in LB broth at 37°C. When necessary, solid media were prepared with 1.5% agar. Kanamycin was used at a working concentration of 300 μg/ml for *A. tumefaciens* and 50 μg/ml for *E. coli*. Isopropyl-β-D-thiogalactosidase (IPTG) was used as an inducer at a concentration of 1 mM.

Clonal isolation of bacteriophage strains. Bacteriophages capable of infecting *A. tumefaciens* strain C58 were isolated from wastewater samples using an enrichment

protocol adapted from Santamaría *et al.* (38) and described in detail in the supplemental methods.

Plaque assays. Classic whole plate plaque assays were performed using the soft agar overlay method (72). To complete whole plate plaque assays, 100 µl cells (OD_{600} = ~0.2) were incubated with 100 µl of diluted phage for 15 min at 28°C. Phage solutions were serial diluted in phage dilution buffer, a mixture of Dulbecco's phosphate-buffered saline (DPBS, Mediatech, Inc., Manassas, VA, USA) with 2 percent gelatin (1:20). The mixture of cells and phage was then added to 3 ml of soft agar prior to overlay. The soft agar containing bacteriophage and cells was poured onto a room-temperature LB plate containing 1% agar and swirled gently to spread the soft agar evenly across the plate. For host range testing, plaque assays were completed by spotting phage on lawns of bacterial cells. In the spot assays, 100 µl cells (OD_{600} = ~0.2) were added to 0.3% soft agar and overlayed on solid medium. After solidification, 10 µl of phage serial dilutions in DPBS-gelatin were spotted on the soft agar. Plates were incubated for 1–2 days and observed for plaque formation. Spot assays were used for host range testing with appropriate adjustments to the base medium and soft agar. Media used for each strain is listed in Table 2-2.

Partial purification of virions and preparation of virion DNA. Virions from 1-liter cleared lysates were enriched and concentrated to 1.5 ml by 2 successive precipitations with 10 percent polyethylene glycol (73) and differential centrifugation (17,000 × *g* for 10 min.; 288,000 × *g* for 2 h), as detailed in supplemental methods. Since some non-

virion lysate components co-purify with virions during this procedure, we consider these virions to be only partially purified. Virion DNA was prepared from partially purified virions by 2 phenol extractions, chloroform extraction, and ethanol precipitation, as detailed in supplemental methods.

Transmission electron microscopy. Virion morphology was observed by applying a small volume of concentrated partially purified virions onto a carbon-coated Formvar grid, negatively stained with 2% uranyl acetate. Specimens were observed on a JEOL JEM-1400 transmission electron microscope at 120 kV. Capsid diameters of ten virions from each phage strain were measured using ImageJ (74).

Growth curves. Growth curves were performed in LB medium by infecting C58 cells at an optical density (OD_{600}) of 0.05 with bacteriophage at an MOI of 0.001 in liquid culture. The turbidity of these cultures, represented by their OD_{600}, was recorded every 5 min during a 24 h interval while the cells grew at 28°C. Cultures were shaken for 1 min prior to each reading. The OD_{600} was measured using a BioTek Synergy H1 Hybrid Reader. For the growth curve with induction of PPH or PPH variants, cells were grown in LB medium without inducer for 16 h, then diluted to an OD_{600} of 0.05. Where indicated, 1 mM IPTG was added to the cultures just prior to taking the initial reading.

Time-lapse microscopy. *A. tumefaciens* strain C58 cells were grown to an OD_{600} of 0.2 and infected with Atu_ph02 and Atu_ph03 at an MOI of 0.01. Infected cells were incubated at room temperature for 15 minutes to allow phage attachment before 1-µl

portions were added to a 1% agarose pad containing LB as described previously (57, 75). Cells were imaged using a 60× oil immersion objective (1.4 NA) by differential interference microscopy every 5–10 minutes for 24 hours using a Nikon Eclipse TiE equipped with a QImaging Rolera em-c^2 1 K electron-multiplying charge-coupled-device (EMCCD) camera and Nikon Elements imaging software. Cells containing pSRKKm, pSRKKm-PPH, pSRKKm-PPH_{E32A}, pSRKKm-$PPH_{K328A,K334A,K335A}$, pSRKKm-$PPH_{E32A,K328A,K334A,K335A}$, and pSRKKm-$PPH_{1-304}$ were grown to an OD_{600} of 0.2, then induced with 1 mM IPTG. Cells were placed in the 28°C shaker for 1 h prior to imaging.

Potato tumor assay. To test for phage protection from *A. tumefaciens*-mediated plant transformation, potato tumor assays adapted from Morton and Fuqua (41) were used. Briefly, red-skinned, organically grown potatoes were rinsed, peeled, and sterilized prior to cutting discs. Sterilization consisted of soaking the potatoes in 1.05% sodium hypochlorite for 20 min and exposing each side to UV for 20 min using a Cole-Parmer SK-97505-30 lamp emitting at 254 nm with an irradiance of 900 μW/cm^2 at the work surface. Potatoes were then cut into cylinders and sliced into discs with diameter 2 cm and thickness 0.5 cm. Discs were overlaid with 100 μl cells (OD_{600} = 0.2) resuspended in DPBS-gelatin. When indicated, cells were premixed with Atu_ph02 or Atu_ph03 at an MOI of 1.0. Potatoes were incubated at room temperature for 10–20 days in a humid chamber and tumor formation was observed.

Cell viability assays. Serial dilutions of *A. tumefaciens* cells containing plasmids for expression of *pph* and *pph* variants were spotted on plates in the presence and absence of

IPTG to test cell viability during *pph* induction. Colonies were inoculated overnight in LB with kanamycin, and diluted to OD_{600} 0.05. Cells were then serially diluted and 4 µl of each dilution was spotted onto LB plates containing Kan and either IPTG (for induction) or 1% glucose (for maximal repression). Plates were grown for 2 days at 28°C and imaged.

Genome sequencing and assembly. Libraries for genome sequencing were constructed from virion DNA following the manufacturer's protocol and reagents supplied in Illumina's TruSeq DNA PCR-Free sample preparation kit (#FC-121-3001). Briefly, 1 microgram of DNA was sheared using standard Covaris methods to generate average fragmented sizes of 350 bp. The resulting 3' and 5' overhangs were converted to blunt ends by an end repair reaction using 3' to 5' exonuclease and polymerase activities, followed by size selection (350 bp) and purification with magnetic sample purification beads. A single adenosine nucleotide was added to the 3' ends of the blunt fragment followed by the ligation of Illumina indexed paired-end adapters. The adaptor-ligated library was purified twice with magnetic sample purification beads. The purified library was quantified using KAPA library quantification kit (KK4824) and library fragment size confirmed by Fragment Analyzer (Advanced Analytical Technologies, Inc.). Libraries were diluted, pooled and sequenced using a paired-end, 75 base read length according to Illumina's standard sequencing protocol for the MiSeq. Library preparation and sequencing was conducted by the University of Missouri DNA Core Facility.

Genome annotation. Protein-coding regions were annotated by RAST server (76). Proteins of interest were analyzed by TMHMM (56) and SignalP 4.1 (77). Whole genome alignments were created using the Mauve (78) plugin in Geneious version 8.1. Phylogenetic trees were constructed using a ClustalW (79) protein alignment and creating a PhyML (80) tree, version 3.0, as a Geneious plugin using the Geneious Tree Builder with the following settings: Le Gascuel substitution model with 100 bootstrap models. Nucleotide alignments were used to determine percent identities between genomes, using the MUSCLE (81) alignment in Geneious.

Nucleotide sequence accession numbers. Genome sequences of phages Atu_ph02 and Atu_ph03 have been deposited in the GenBank database with nucleotide accession numbers MF403005 and MF403006, respectively.

Construction of plasmids for characterization of PPH. See Table 2-3 for a list of all primers used in plasmid construction and sequencing. All variants of *pph* were cloned into vector pSRKKm-Plac-sf*gfp*, which allows for inducible expression of target genes under the control of the *lac* promoter using IPTG as the inducer (63). To construct pSRKKm-PPH, PCR using Phusion High-Fidelity DNA Polymerase (Thermo Scientific) was performed on the Atu_ph02 genomic DNA using primers PPH *Nde*I F and PPH *Bam*HI R. PCR products were gel purified using the GeneJET Gel Extraction Kit (Thermo Scientific). Amplified gene products and the pSRKKm-Plac-sf*gfp* plasmid (63) were digested with *Nde*I and *Bam*HI overnight at 37°C and subsequently gel purified. Digested vector and insert were ligated using T4 DNA ligase (Invitrogen). The ligation

reaction was incubated at 4°C overnight. The ligation was transformed into DH5α chemically competent *E. coli* cells (Invitrogen) and selected for on LB agar plates containing kanamycin. Plasmid DNA was extracted using the GeneJET Plasmid Miniprep kit (Thermo Scientific) prior to electroporating competent *A. tumefaciens* cells as described previously (82). The pSRKKm-PPH vector was sequenced using the pSRK Forward Sequencing primer at the MU DNA Core Facility. To perform site-directed mutagenesis on PPH, the Q5 Site-Directed Mutagenesis Kit was used according to the protocol (New England Biolabs). To construct PPH_{E32A}, primers "PPH E32A F" and "PPH E32A R" were used. For $PPH_{K328A\ K334A\ K335A}$, "PPH K328A F" and "PPH K328A R" were used first, followed by "PPH K334A F" and "PPH K334A R," and lastly "PPH K335A F" and "PPH K335A R." To construct $PPH_{1\text{-}304}$, "PPH NdeI F" and "$PPH_{1\text{-}304}$ BamHI R" primers were used. Generated constructs were sequenced at the MU DNA Core Facility using the "pSRKKm Forward Sequencing," "PPH NdeI F," and "PPH linker For NdeI" primers.

SDS-PAGE and Zymography. Whole cell lysates were prepared using 100 ml cultures of exponential-phase cells (OD 0.3-0.6) grown with inducer for 3 h for WT cells, WT cells expressing PPH, or WT cells expressing PPH_{E32A}. Cells were centrifuged at 4,300 × g for 15 min at 4°C and the cell pellets stored at -20°C overnight. The next day, pellets were resuspended in 8 ml B-PER Bacterial Protein Extraction Reagent (Thermo Scientific) with the addition of 1/6 of a crushed protease inhibitor tablet (Thermo Scientific). Cells were lysed by sonication (4 pulses comprised of a 10 s burst, followed by 20 s burst). Cell debris was pelleted at 17,000 x g for 15 min at 4°C. Soluble proteins

in the supernatants were quantified using the Pierce BCA Protein Assay Kit (Thermo Scientific). 30 μg total protein from each whole cell lysate sample was boiled for 5 min and loaded onto two gels (one with embedded peptidoglycan), which ran at 30 V for 30 min, followed by 100 V for 90 min. For zymography, SDS-PAGE gels were embedded with peptidoglycan. 1 L culture of *A. tumefaciens* strain C58 was autoclaved and peptidoglycan was harvested by centrifugation at 7,000 x g for 20 min at 4°C. Pellets containing peptidoglycan were resuspended in 10 ml D-PBS, 1X with calcium and magnesium (Corning Cellgro). 500 μl peptidoglycan was added to a 12% SDS polyacrylamide gel. After running the gel, the running gel was incubated in 25 mM Tris (pH 8), 1% Triton X-100 at 28°C overnight to renature the proteins and enable peptidoglycan hydrolysis. The gel was stained in 1:50 0.1% methylene blue in 0.01% KOH for 3 h and destained in 0.01% KOH. For SDS-PAGE, the running gel was incubated in Coomassie Blue dye (0.25% w/v Brilliant Blue R-250 (FisherBiotech), 10% acetic acid, 5% methanol) for 30 s, then destained (7.5% acetic acid, 50% methanol) overnight with shaking.

ACKNOWLEDGEMENTS

We thank Zhanyuan Zhang and the MU Plant Transformation Core Facility for providing *Agrobacterium* strains, and Kenya Phillips and Courtney Buchanan for assisting in the characterization of Atu_ph03. We appreciate the efforts of Dr. Tommi White, Dr. Martin Schauflinger, and DeAna Grant of the MU Electron Microscopy Core for help with the transmission electron microscopy. We thank Nathan Bivans and the MU DNA Core for assistance sequencing the bacteriophages and William Spollen and the MU Research

Informatics core for assistance with genome assembly and GenBank submission. Finally, we thank members of the Brown lab, especially Helen Blaine, and Chiqian Zhang for feedback during the preparation of this manuscript.

This research is supported by start-up funds, a Research Council Grant (URC 14-051) and a Research Board Grant (3786-2) from the University of Missouri to PJBB. HA has been supported by National Institute of General Medical Sciences of the National Institutes of Health under award number T32GM008396. JR was supported by a Monsanto Undergraduate Research Fellowship.

REFERENCES

1. **Strange RN, Scott PR.** 2005. Plant disease: a threat to global food security. Annu Rev Phytopathol **43:**83-116.
2. **Mansfield J, Genin S, Magori S, Citovsky V, Sriariyanum M, Ronald P, Dow M, Verdier V, Beer SV, Machado MA, Toth I, Salmond G, Foster GD.** 2012. Top 10 plant pathogenic bacteria in molecular plant pathology. Mol Plant Pathol **13:**614-629.
3. **Escobar MA, Dandekar AM.** 2003. *Agrobacterium tumefaciens* as an agent of disease. Trends Plant Sci **8:**380-386.
4. **Pulawska J.** 2010. Crown gall of stone fruits and nuts, economic significance and diversity of its causal agents: tumorigenic *Agrobacterium* Spp. J Plant Pathol **92:**S87-S98.

5. **Bourras S, Rouxel T, Meyer M.** 2015. *Agrobacterium tumefaciens* gene transfer: how a plant pathogen hacks the nuclei of plant and nonplant organisms. Phytopathology **105:**1288-1301.

6. **Kado CI.** 1976. The tumor-inducing substance of *Agrobacterium tumefaciens*. Annu Rev Phytopathol **14:**265-308.

7. **Kerr A, Panagopoulos CG.** 1977. Biotypes of *Agrobacterium radiobacter* var. *tumefaciens* and their biological control. Phytopathol z **90:**172-179.

8. **Kerr A, Roberts WP.** 1976. *Agrobacterium*: correlations between and transfer of pathogenicity, octopine and nopaline metabolism and bacteriocin 84 sensitivity. Physiol plant path **9:**205-211.

9. **Sule S, Kado CI.** 1980. Agrocin resistance in virulent derivative of *Agobacterium tumefaciens* harboring the pTi plasmid. Physiol plant path **17:**347-356.

10. **Beneddra T, Picard C, Petit A, Nesme X.** 1996. Correlation between susceptibility to crown gall and sensitivity to cytokinin in aspen cultivars. Phytopathology **86:**225-231.

11. **Bliss FA, Almehdi AA, Dandekar AM, Schuerman PL, Bellaloui N.** 1999. Crown gall resistance in accessions of 20 *Prunus* species. HortScience **34:**326-330.

12. **Reynders-Aloisi S, Pelloli G.** 1998. Tolerance to crown gall differs among genotypes of rose rootstocks. HortScience **33:**296-297.

13. **Anand A, Uppalapati SR, Ryu CM, Allen SN, Kang L, Tang Y, Mysore KS.** 2008. Salicylic acid and systemic acquired resistance play a role in attenuating

crown gall disease caused by *Agrobacterium tumefaciens*. Plant Physiol **146:**703-715.

14. **Dandurishvili N, Toklikishvili N, Ovadis M, Eliashvili P, Giorgobiani N, Keshelava R, Tediashvili M, Vainstein A, Khmel I, Szegedi E, Chernin L.** 2011. Broad-range antagonistic rhizobacteria *Pseudomonas fluorescens* and *Serratia plymuthica* suppress *Agrobacterium* crown gall tumours on tomato plants. J Appl Microbiol **110:**341-352.

15. **Jones JB, Vallad GE, Iriarte FB, Obradovic A, Wernsing MH, Jackson LE, Balogh B, Hong JC, Momol MT.** 2012. Considerations for using bacteriophages for plant disease control. Bacteriophage **2:**208-214.

16. **Buttimer C, McAuliffe O, Ross RP, Hill C, O'Mahony J, Coffey A.** 2017. Bacteriophages and bacterial plant diseases. Front Microbiol **8:**34.

17. **Iriarte FB, Obradovic A, Wernsing MH, Jackson LE, Balogh B, Hong JA, Momol MT, Jones JB, Vallad GE.** 2012. Soil-based systemic delivery and phyllosphere in vivo propagation of bacteriophages: Two possible strategies for improving bacteriophage persistence for plant disease control. Bacteriophage **2:**215-224.

18. **Fujiwara A, Fujisawa M, Hamasaki R, Kawasaki T, Fujie M, Yamada T.** 2011. Biocontrol of *Ralstonia solanacearum* by treatment with lytic bacteriophages. Appl Environ Microbiol **77:**4155-4162.

19. **Balogh B, Canteros BI, Stall KE, Jones JB.** 2008. Control of citrus canker and citrus bacterial spot with bacteriophages. Plant Disease **92:**1048-1052.

20. **Rombouts S, Volckaert A, Venneman S, Devlercq B, Vandenheuvel D, Allonsius CN, Van Malderghem C, Jang HB, Briers Y, Noben JP, Klumpp J, Van Vaerenbergh J, Maes M, Lavigne R.** 2016. Characterization of noval bacteriophages for biocontrol of bacterial blight in leek caused by *Pseudomonas syringae* pv. *porri*. Frontiers in Microbiology **7**.

21. **Frampton RA, Taylor C, Holguin Moreno AV, Visnovsky SB, Petty NK, Pitman AR, Fineran PC.** 2014. Identification of bacteriophages for biocontrol of the kiwifruit canker phytopathogen *Pseudomonas syringae* pv. *actinidiae*. Appl Environ Microbiol **80:**2216-2228.

22. **Roach DR, Donovan DM.** 2015. Antimicrobial bacteriophage-derived proteins and therapeutic applications. Bacteriophage **5:**e1062590.

23. **Nelson D, Loomis L, Fischetti VA.** 2001. Prevention and elimination of upper respiratory colonization of mice by group A streptococci by using a bacteriophage lytic enzyme. Proc Natl Acad Sci U S A **98:**4107-4112.

24. **Keary R, McAuliffe O, Ross RP, Hill C, O'Mahony J, Coffey A.** 2013. Bacteriophages and their endolysins for control of pathogenic bacteria., p. 1028-1040. *In* Méndez-Vilas A (ed.), Microbial Pathogens and Strategies for Combating Them: Science, Technology and Education. Formatex Research Center., Badajoz, Spain.

25. **Lim JA, Shin H, Heu S, Ryu S.** 2014. Exogenous lytic activity of SPN9CC endolysin against gram-negative bacteria. J Microbiol Biotechnol **24:**803-811.

26. **Guo M, Feng C, Ren J, Zhuang X, Zhang Y, Zhu Y, Dong K, He P, Guo X, Qin J.** 2017. A Novel Antimicrobial Endolysin, LysPA26, against *Pseudomonas aeruginosa*. Front Microbiol **8:**293.

27. **Wang S, Gu J, Lv M, Guo Z, Yan G, Yu L, Du C, Feng X, Han W, Sun C, Lei L.** 2017. The antibacterial activity of *E. coli* bacteriophage lysin lysep3 is enhanced by fusing the *Bacillus amyloliquefaciens* bacteriophage endolysin binding domain D8 to the C-terminal region. J Microbiol.

28. **Boyd RJ, Hildebrandt AC, Allen ON.** 1970. Specificity patterns of *Agrobacterium tumefaciens* phages. Arch Mikrobiol **73:**324-330.

29. **Roslycky EB, Allen ON, McCoy E.** 1965. Growth characteristics of phages of *Agrobacterium radiobacter*. Can J Microbiol **11:**95-101.

30. **Lotz W, Mayer F.** 1972. Electron microscopical characterization of newly isolated *Rhizobium lupini* bacteriophages. Can J Microbiol **18:**1271-1274.

31. **Kropinski AM, Van den Bossche A, Lavigne R, Noben JP, Babinger P, Schmitt R.** 2012. Genome and proteome analysis of 7-7-1, a flagellotropic phage infecting *Agrobacterium* sp H13-3. Virol J **9:**102.

32. **Watson B, Currier TC, Gordon MP, Chilton MD, Nester EW.** 1975. Plasmid required for virulence of *Agrobacterium tumefaciens*. J Bacteriol **123:**255-264.

33. **Christie PJ, Gordon JE.** 2014. The *Agrobacterium* Ti Plasmids. Microbiol Spectr **2**.

34. **Hamilton RH, Fall MZ.** 1971. The loss of tumor-initiating ability in *Agrobacterium* tumefaciens by incubation at high temperature. Experientia **27:**229-230.

35. **Goodner B, Hinkle G, Gattung S, Miller N, Blanchard M, Qurollo B, Goldman BS, Cao Y, Askenazi M, Halling C, Mullin L, Houmiel K, Gordon J, Vaudin M, Iartchouk O, Epp A, Liu F, Wollam C, Allinger M, Doughty D, Scott C, Lappas C, Markelz B, Flanagan C, Crowell C, Gurson J, Lomo C, Sear C, Strub G, Cielo C, Slater S.** 2001. Genome sequence of the plant pathogen and biotechnology agent *Agrobacterium tumefaciens* C58. Science **294:**2323-2328.

36. **Wood DW, Setubal JC, Kaul R, Monks DE, Kitajima JP, Okura VK, Zhou Y, Chen L, Wood GE, Almeida NF, Jr., Woo L, Chen Y, Paulsen IT, Eisen JA, Karp PD, Bovee D, Sr., Chapman P, Clendenning J, Deatherage G, Gillet W, Grant C, Kutyavin T, Levy R, Li MJ, McClelland E, Palmieri A, Raymond C, Rouse G, Saenphimmachak C, Wu Z, Romero P, Gordon D, Zhang S, Yoo H, Tao Y, Biddle P, Jung M, Krespan W, Perry M, Gordon-Kamm B, Liao L, Kim S, Hendrick C, Zhao ZY, Dolan M, Chumley F, Tingey SV, Tomb JF, Gordon MP, Olson MV, Nester EW.** 2001. The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. Science **294:**2317-2323.

37. **Gohlke J, Deeken R.** 2014. Plant responses to *Agrobacterium tumefaciens* and crown gall development. Front Plant Sci **5:**155.

38. **Santamaria RI, Bustos P, Sepulveda-Robles O, Lozano L, Rodriguez C, Fernandez JL, Juarez S, Kameyama L, Guarneros G, Davila G, Gonzalez V.** 2014. Narrow-host-range bacteriophages that infect *Rhizobium etli* associate with distinct genomic types. Appl Environ Microbiol **80:**446-454.

39. **Ackermann HW.** 2009. Phage classification and characterization. Methods Mol Biol **501:**127-140.

40. **Lazo GR, Stein PA, Ludwig RA.** 1991. A DNA transformation-competent *Arabidopsis* genomic library in *Agrobacterium*. Biotechnology (N Y) **9:**963-967.

41. **Morton ER, Fuqua C.** 2012. Phenotypic analyses of *Agrobacterium*. Curr Protoc Microbiol **Chapter 3:**Unit 3D 3.

42. **Yin Y, Fischer D.** 2008. Identification and investigation of ORFans in the viral world. BMC Genomics **9:**24.

43. **West D, Lagenaur C, Agabian N.** 1976. Isolation and characterization of *Caulobacter crecentus* bacteriophage phi Cd1. J Virol **17:**568-575.

44. **Lerma RA, Tidwell TJ, Cahill JL, Rasche ES, Kuty Everett GF.** 2015. Complete genome sequence of *Caulobacter crescentus* podophage Percy. Genome Announc **3**.

45. **Kawasaki T, Shimizu M, Satsuma H, Fujiwara A, Fujie M, Usami S, Yamada T.** 2009. Genomic characterization of *Ralstonia solanacearum* phage phiRSB1, a T7-like wide-host-range phage. J Bacteriol **191:**422-427.

46. **McCorquodale DJ, Lanni YT.** 1970. Patterns of protein synthesis in *Escherichia coli* infected by amber mutants in the first-step-transfer DNA of T5. J Mol Biol **48:**133-143.

47. **McCorquodale DJ, Warner HR.** 1988. Bacteriophage T5 and Related Phages, p. 439-475. *In* Calendar R (ed.), The Bacteriophages. Plenum Press, New York, NY.

48. **Davison J.** 2015. Pre-early functions of bacteriophage T5 and its relatives. Bacteriophage **5:**e1086500.

49. **Gill JJ, Berry JD, Russell WK, Lessor L, Escobar-Garcia DA, Hernandez D, Kane A, Keene J, Maddox M, Martin R, Mohan S, Thorn AM, Russell DH, Young R.** 2012. The *Caulobacter crescentus* phage phiCbK: genomics of a canonical phage. BMC Genomics **13:**542.

50. **Young R.** 2014. Phage lysis: three steps, three choices, one outcome. J Microbiol **52:**243-258.

51. **Oliveira H, Melo LD, Santos SB, Nobrega FL, Ferreira EC, Cerca N, Azeredo J, Kluskens LD.** 2013. Molecular aspects and comparative genomics of bacteriophage endolysins. J Virol **87:**4558-4570.

52. **Catalao MJ, Gil F, Moniz-Pereira J, Sao-Jose C, Pimentel M.** 2013. Diversity in bacterial lysis systems: bacteriophages show the way. FEMS Microbiol Rev **37:**554-571.

53. **Kakikawa M, Yokoi KJ, Kimoto H, Nakano M, Kawasaki K, Taketo A, Kodaira K.** 2002. Molecular analysis of the lysis protein Lys encoded by *Lactobacillus plantarum* phage phig1e. Gene **299:**227-234.

54. **Sao-Jose C, Parreira R, Vieira G, Santos MA.** 2000. The N-terminal region of the *Oenococcus oeni* bacteriophage fOg44 lysin behaves as a bona fide signal peptide in *Escherichia coli* and as a *cis*-inhibitory element, preventing lytic activity on oenococcal cells. J Bacteriol **182:**5823-5831.

55. **Rodriguez-Rubio L, Gerstmans H, Thorpe S, Mesnage S, Lavigne R, Briers Y.** 2016. DUF3380 domain from a *Salmonella* phage endolysin shows potent N-acetylmuramidase activity. Appl Environ Microbiol **82:**4975-4981.

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

57. **Brown PJ, de Pedro MA, Kysela DT, Van der Henst C, Kim J, De Bolle X, Fuqua C, Brun YV.** 2012. Polar growth in the Alphaproteobacterial order Rhizobiales. Proc Natl Acad Sci U S A **109:**1697-1701.

58. **Fujiwara T, Fukui S.** 1972. Isolation of morphological mutants of *Agrobacterium tumefaciens*. J Bacteriol **110:**743-746.

59. **Fujiwara T, Fukui S.** 1974. Unidirectional growth and branch formation of a morphological mutant, *Agrobacterium tumefaciens*. J Bacteriol **120:**583-589.

60. **Kahng LS, Shapiro L.** 2001. The CcrM DNA methyltransferase of *Agrobacterium tumefaciens* is essential, and its activity is cell cycle regulated. J Bacteriol **183:**3065-3075.

61. **Latch JN, Margolin W.** 1997. Generation of buds, swellings, and branches instead of filaments after blocking the cell cycle of *Rhizobium meliloti*. J Bacteriol **179:**2373-2381.

62. **Su S, Stephens BB, Alexandre G, Farrand SK.** 2006. Lon protease of the alpha-proteobacterium *Agrobacterium tumefaciens* is required for normal growth, cellular morphology and full virulence. Microbiology **152:**1197-1207.

63. **Figueroa-Cuilan W, Daniel JJ, Howell M, Sulaiman A, Brown PJ.** 2016. Mini-Tn7 Insertion in an Artificial *att*Tn7 Site Enables Depletion of the Essential Master Regulator CtrA in the Phytopathogen *Agrobacterium tumefaciens*. Appl Environ Microbiol **82:**5015-5025.

64. **Lutkenhaus J, Pichoff S, Du S.** 2012. Bacterial cytokinesis: From Z ring to divisome. Cytoskeleton (Hoboken) **69:**778-790.

65. **Bernhardt TG, Roof WD, Young R.** 2000. Genetic evidence that the bacteriophage phi X174 lysis protein inhibits cell wall synthesis. Proc Natl Acad Sci U S A **97:**4297-4302.

66. **Bernhardt TG, Struck DK, Young R.** 2001. The lysis protein E of phi X174 is a specific inhibitor of the MraY-catalyzed step in peptidoglycan synthesis. J Biol Chem **276:**6093-6097.

67. **Chamakura KR, Tran JS, Young R.** 2017. MS2 Lysis of *Escherichia coli* Depends on Host Chaperone DnaJ. J Bacteriol **199**.

68. **Reed CA, Langlais C, Wang IN, Young R.** 2013. A(2) expression and assembly regulates lysis in Qbeta infections. Microbiology **159:**507-514.

69. **Tran TA, Struck DK, Young R.** 2005. Periplasmic domains define holin-antiholin interactions in t4 lysis inhibition. J Bacteriol **187:**6631-6640.

70. **Vincent J.** 1970. A manual for the practical study of root-nodule bacteria. Blackwell Scientific, Oxford.

71. **Poindexter JS.** 1964. Biological properties and classification of the *Caulobacter* group. Bacteriol Rev **28:**231-295.

72. **Adams MH.** 1959. Bacteriophages. Interscience Publishers, Inc., New York, NY.

73. **Yamamoto KR, Alberts BM, Benzinger R, Lawhorne L, Treiber G.** 1970. Rapid bacteriophage sedimentation in the presence of polyethylene glycol and its application to large-scale virus purification. Virology **40:**734-744.

74. **Schneider CA, Rasband WS, Eliceiri KW.** 2012. NIH Image to ImageJ: 25 years of image analysis. Nat Methods **9:**671-675.

75. **Howell M, Daniel, J.J., Brown, P.J.B.** 2017. Live cell fluorescence microscopy to observe essential processes during microbial cell growth. J Vis Exp **in press**.

76. **Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O.** 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics **9:**75.

77. **Petersen TN, Brunak S, von Heijne G, Nielsen H.** 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods **8:**785-786.

78. **Darling AC, Mau B, Blattner FR, Perna NT.** 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res **14:**1394-1403.

79. **Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG.** 2007. Clustal W and Clustal X version 2.0. Bioinformatics **23:**2947-2948.

80. **Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O.** 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol **59:**307-321.

81. **Edgar RC.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32:**1792-1797.

82. **Morton ER, Fuqua C.** 2012. Genetic manipulation of *Agrobacterium*. Curr Protoc Microbiol **Chapter 3:**Unit 3D 2.

83. **Simon R, Priefer U, Puhler A.** 1983. A broad host range mobilization system for *in vivo* genetic engineering: transposon mutagenesis in Gram negative bacteria. Nat Biotechnol **1:**784-791.

84. **Luo ZQ, Clemente TE, Farrand SK.** 2001. Construction of a derivative of *Agrobacterium tumefaciens* C58 that does not mutate to tetracycline resistance. Mol Plant Microbe Interact **14:**98-103.

85. **Bush AL, Pueppke SG.** 1991. Characterization of an Unusual New *Agrobacterium tumefaciens* Strain from Chrysanthemum morifolium Ram. Appl Environ Microbiol **57:**2468-2472.

86. **Slater SC, Goldman BS, Goodner B, Setubal JC, Farrand SK, Nester EW, Burr TJ, Banta L, Dickerman AW, Paulsen I, Otten L, Suen G, Welch R, Almeida NF, Arnold F, Burton OT, Du Z, Ewing A, Godsy E, Heisel S, Houmiel KL, Jhaveri J, Lu J, Miller NM, Norton S, Chen Q, Phoolcharoen W, Ohlin V, Ondrusek D, Pride N, Stricklin SL, Sun J, Wheeler C, Wilson L, Zhu H, Wood DW.** 2009. Genome sequences of three *Agrobacterium* biovars

help elucidate the evolution of multichromosome genomes in bacteria. J Bacteriol **191:**2501-2511.

87. **Weidner S, Baumgarth B, Gottfert M, Jaenicke S, Puhler A, Schneiker-Bekel S, Serrania J, Szczepanowski R, Becker A.** 2013. Genome Sequence of *Sinorhizobium meliloti* Rm41. Genome Announc **1**.

88. **Nierman WC, Feldblyum TV, Laub MT, Paulsen IT, Nelson KE, Eisen JA, Heidelberg JF, Alley MR, Ohta N, Maddock JR, Potocka I, Nelson WC, Newton A, Stephens C, Phadke ND, Ely B, DeBoy RT, Dodson RJ, Durkin AS, Gwinn ML, Haft DH, Kolonay JF, Smit J, Craven MB, Khouri H, Shetty J, Berry K, Utterback T, Tran K, Wolf A, Vamathevan J, Ermolaeva M, White O, Salzberg SL, Venter JC, Shapiro L, Fraser CM.** 2001. Complete genome sequence of *Caulobacter crescentus*. Proc Natl Acad Sci U S A **98:**4136-4141.

TABLES

Table 2-1. Summary of key genomic features

Bacteriophage	Genome length (bp)	GC content (%)	Number of ORFs	Coding density (%)	Number of hypothetical proteins	Number of ORFans[a]
Atu_ph02	45,423	54.8	55	92.9	32	23
Atu_ph03	45,175	54.7	58	93.8	36	26

[a] ORFans are predicted proteins that do not have significant hits in the nr database. Hypothetical proteins share homology with proteins in the nr database (42)

Table 2-2. Bacterial strains and plasmids used in this study.

Strain or plasmid	**Sequence or relevant characteristics**	**Growth Medium**	**Reference/Source**
Plasmids			
PSRKKm-Plac-sf*gfp*	Km^r; broad host range vector containing $lacI^q$ and *lac* promoter		(63)
pSRKKm-Plac-PPH	Phage Peptidoglycan Hydrolase (PPH) inserted into pSRKKm-Plac-sf*gfp*		This study
pSRKKm-Plac-PPH_{E32A}	PPH predicted catalytic residue mutated		This study
pSRKKm-Plac-$PPH_{K328A,K334A,K335A}$	PPH regulatory residues mutated		This study
pSRKKm-Plac-$PPH_{E32A,K328A,K334A,K335A}$	PPH catalytic and regulatory residues mutated		This study
pSRKKm-Plac-PPH_{1-304}	PPH truncation to remove TM domain		This study
***E. coli* strains**			
DH5α	Cloning strain; Gammaproteobacterium	LB	Life Technologies
S17-1	Sm^r; RP4-2, Tc::Mu,Km-Tn7, for plasmid mobilization	LB	(83)
***A. tumefaciens* strains**			
C58	Nopaline type strain; pTiC58; pAtC58	LB	(32)
C58 pSRKKm-Plac-sf*gfp*	C58 transformed with empty pSRKKm plasmid	LB	(63)
C58 pSRKKm-Plac-PPH	C58 transformed with pSRKKM-Plac-PPH	LB	This study
C58 pSRKKm-Plac-PPH_{E32A}	C58 transformed with pSRKKm-Plac-PPH_{E32A}	LB	This study
C58 pSRKKm-Plac-$PPH_{K328A,K334A,K335A}$	C58 transformed with pSRKKm-Plac-$PPH_{K328A,K334A,K335A}$	LB	This study
C58 pSRKKm-Plac-$PPH_{E32A,K328A,K334A,K335A}$	C58 transformed with pSRKKm-Plac-$PPH_{E32A,K328A,K334A,K335A}$	LB	This study
C58 pSRKKm-Plac-PPH_{1-304}	C58 transformed with pSRKKm-Plac-PPH_{1-304}	LB	This study
EHA105	C58 derived; succinamopine strain; T-DNA deletion derivative of pTiBo542	LB	MU Plant Transformation Core Facility
EHA101	C58 derived; nopaline strain; T-DNA deletion derivative of pTiBo542	LB	MU Plant Transformation Core Facility

GV3101	C58 derived; nopaline strain	LB	MU Plant Transformation Core Facility
NTL4	C58 derived; nopaline-agrocinopine strain; Δ*tetRA*	LB	(84)
AGL-1	C58 derived; succinamopine strain; T-DNA deletion derivative of pTiBo542	LB	MU Plant Transformation Core Facility
LBA4404	Ach5 derived; octopine strain; T-DNA deletion derivative of pTiAch5	YM	MU Plant Transformation Core Facility
Chry5	Succinamopine strain; pTiChry5	LB	(85)
Other bacterial strains			
Agrobacterium vitis S4	Vitopine strain; pTiS4; pSymA; pSymB	Potato dextrose	(86)
Sinorhizobium meliloti RM41	Rhizopine strain; pSymA; pSymB; pRme41a	LB	(87)
Caulobacter crescentus CB15	Alphaproteobacterium	PYE	(88)

Table 2-3. Synthesized DNA primers used in this study.

Synthesized DNA Primers	Sequence
PPH NdeI F	5’ -GTA CCA TAT GTG CAA CCA AAG -3’
PPH BamHI R	5’ -TCA GGA TCC TTA TTT CTT CCA -3’
$PPH_{1\text{-}304}$ BamHI R	5’ -TCA GGA TCC TTG AGG AAC -3’
PPH K328A F	5’ -GCG GCA TAC ATC CAC -3’
PPH K328A R	5’ -GTA CCC TGC GTA GGC -3’
PPH K334A F	5’ -GCG AAA TAA GGA TCC -3’
PPH K334A R	5’ -CCA GTG GAT GTA TGC -3’
PPH K335A F LacGFP	5’ -GCA TAA GGA TCC GCT -3’
PPH K335A R	5’ -CGC CCA GTG GAT GTA -3’
PPH E32A F	5’ -GCG AGT GCA GGC AAA -3’
PPH E32A R	5’ -CTT GTC CAC GAT GGC -3’
pSRK Forward Sequencing	5’ -AAT GTG AGT TAG CTC ACT CAT TAG GCA -3’
PPH 31 For NdeI	5’ -ATA CAT ATG GGG GCT GGT GCC -3’
PPH linker For NdeI	5’ -ATA CAT ATG AGC AAG GCT GGT AAT -3’

CHAPTER 2: SUPPLEMENTAL MATERIAL

SUPPLEMENTAL METHODS

Clonal isolation of bacteriophage strains. Water samples were filtered by passing through a 0.45 μm membrane (Millipore Ultrafree – CL, Low-binding Durapore PVDP membrane) and 890 μl of filtrate was mixed with 100 μl 10X LB and 10 μl *A. tumefaciens* C58 at a starting OD_{600} of ~0.2. Cultures were incubated at 28°C in LB broth for 16 h while shaking. Cultures that appeared clear were screened for plaque formation. If the culture was turbid, supernatants were collected by centrifugation at 3,000 x g for 10 min and 100 μl filtrate were mixed with 100 μl bacteria (OD_{600} ~0.2) for another round of amplification. If cultures remained turbid after 5 rounds of amplification, the filtrate was considered to be negative for lytic activity. Filtrates that caused clearing of the bacterial culture within 5 rounds of amplification were examined for evidence of phage activity. Filtrates were screened for phage activity using a spot assay for detection of plaques. Whole plate plaque assays were performed using filtrates which produced plaques in the spot assay. Individual plaques were suspended in Dulbecco's phosphate-buffered saline (DPBS, Mediatech, Inc., Manassas, VA, USA) with gelatin added (1:20). Three rounds of purification comprised of selecting individual plaques after whole plate plaque assays were completed for each bacteriophage to ensure homogenous bacteriophage populations.

Concentration and partial purification of virions. Concentrated phage stocks were produced by polyethylene glycol (PEG) precipitation. For PEG precipitation, filtered lysates were scaled up to 1 L and centrifuged at 11,000 x g for 20 min at 4°C to remove

bacterial cells. To the supernatants, 400 µl 10 mg/ml RNase A (Sigma) and 1 ml 3.45 mg/ml DNase I (Sigma) were added for removal of bacterial genomic DNA and RNAs. After 1 h of stirring at room temperature, NaCl (final concentration of 500 mM) and 10% w/w PEG 8000 (Fisher) were added and the solution was stirred for 2 h until dissolved. Bacteriophages were precipitated for 16 h at 4°C. Precipitated bacteriophages were collected by centrifugation at 11,000 x g for 30 min at 4°C and resuspended in 30 ml DPBS. The bacteriophage solution was incubated with shaking overnight at 4°C. Insoluble material was removed and the supernatant was recovered. NaCl (final concentration 0.5M) and PEG 8000 (10% w/w) were added to the supernatant and the solution rotated for 2 h at 4°C. Bacteriophages were precipitated, collected by centrifuging for 20 min at 17,000 x g at 4°C, and resuspended 8 ml DPBS. The bacteriophage solution was rotated overnight at 4°C. This viscous solution was centrifuged at 17,000 x g for 2 min at 4°C and the supernatant was collected. The bacteriophage solution was mixed with 300 µl 10 mM phenol red and 22 ml DPBS and centrifuged at 17,000 x g for 30 min and the supernatant was collected. The supernatant was overlayed with a 2 ml sucrose (5% w/w) cushion and ultracentrifuged at 141,000 x g for 2 h at 4°C. The supernatant was removed and the pellet was washed in DPBS and dissolved in 1 ml DPBS. All phage stocks were stored at 4°C.

Preparation of virion DNA. Two 500-µl portions of partially purified virions in 1.5-ml microtubes were extracted twice with 500 µl neutralized phenol (water-saturated phenol shaken twice with 1/10 vol 1 M Tris.HCl pH 8, discarding the upper phase each time) and once with 500 µl chloroform (a 24:1 mixture of chloroform and isoamyl alcohol), each

time discarding the organic (lower) phases. To the final aqueous phases were added 40 µl 3 M sodium acetate, pH adjusted to 6 with acetic acid, and 1 ml ethanol; precipitates were pelleted by a 10-min centrifugation in a microfuge; supernatants were aspirated; pellets were gently washed by adding 1 ml freezer-cold 70% v/v ethanol and aspirating the liquid; pellets were air-dried, dissolved in 100 µl TE (10 mM Tris.HCl pH 7.5, 1 mM Na_2EDTA), and centrifuged 10 min in a microfuge to clear insoluble material; supernatants were pooled and stored at -20°C.

DNA restriction analysis. Phage genomic DNA was digested with restriction endonucleases from New England Biolabs using the standard protocol. All reactions contained 2.5 µg DNA and were incubated at 37°C for 2 h. Restriction patterns were analyzed on a 0.7% agarose gel, which ran at 100 V for 1 h and was stained with SYBR Safe DNA Gel Stain (Thermo Scientific).

SUPPLEMENTAL TABLES

Table 2-S1. Comparison of gene products encoded in Atu_ph02 and Atu_ph03

Atu_ph03		Atu_ph02		
gene product	length in AA[a]	gene product	length in AA[a]	percent identity[b]
gp1	41	gp1	41	72.5
gp2	173	gp2	173	100
gp3	336	gp3	336	100
gp4	415	gp4	409	99.26
gp5	134	gp5	134	100
gp6	449	gp6	449	100
gp7	68	gp7	68	83.58
gp8	39	gp8	163	77.78
gp9	123	gp8	163	92.62
gp10	486	gp9	486	99.79
gp11	224	gp10	210	69.78
gp12	55			
gp13	331	gp11	337	93.75
gp14	180	gp12	181	88.27
gp15	786	gp13	787	98.09
gp16	291	gp14	291	100
gp17	77	gp15	77	100
gp18	38	gp16	38	100
gp19	317	gp17	317	98.73
gp20	61	gp18	61	100
gp21	77	gp19	77	96.05
gp22	128	gp20	128	100
gp23	816	gp21	816	99.88
gp24	57	gp22	57	100
gp25	66	gp23	66	100
gp26	155	gp24	155	100
gp27	88			
gp28	68	gp25	68	100
gp29	533	gp26	533	99.62
gp30	296	gp27	296	99.66

Atu_ph03		Atu_ph02		
gene product	length in AA[a]	gene product	length in AA[a]	percent identity[b]
gp31	327	gp28	327	100
gp32	212	gp29	212	99.53
gp33	823	gp30	823	99.64
gp34	169	gp31	169	99.4
gp35	1192	gp32	1192	99.83
gp36	1255	gp33	1255	99.68
gp37	507	gp34	507	98.62
gp38	181	gp35	181	98.33
gp39	59	gp36	59	100
gp40	107	gp37	107	100
gp41	612	gp38	623	100
gp42	122	gp39	122	100
gp43	91	gp40	91	100
gp44	149	gp41	149	99.32
gp45	64	gp42	64	100
gp46	289	gp43	289	98.96
gp47	47	gp44	47	94.59
gp48	222	gp45	222	98.19
gp49	124	gp46	124	100
gp50	61	gp47	60	93.22
		gp48	104	
gp51	65	gp49	65	85.94
gp52	88	gp50	88	49.33
gp53	50			
gp54	89	gp51	105	59.00
gp55	107	gp52	107	93.40
gp56	78	gp53	78	98.70
gp57	91	gp54	95	100
gp58	40	gp55	40	100

[a]Length of each gene product is given in amino acids (AA)
[b]Percent identites were determined by blastx analysis using each predicted ORF in Atu_ph03 as query against the protein database for Atu_ph02

Table 2-S2. Similarity of putative Atu_ph03 proteins to proteins in select bacteriophages

		Similarity of putative proteins in select bacteriophages E-value (percent identity)[a]						
Atu_ph03		Rhe_phe2	Rhe_phe8	MedPE-SWcel-C56	T7	phiKMV	T5	Putative Function[b]
gene product	length in AA							
gp1	41	-	-	-	-	-	-	ORFan
gp2	173	gp014 $8e^{-10}$ (35%)	gp013 $2e^{-09}$ (34%)	-	-	-	-	Hypothetical protein
gp3	336	-	-	-	-	-	-	Hypothetical peptidoglycan binding protein (PPH)
gp4	415	-	-	-	-	-	-	DNA primase
gp5	134	gp020 $8e^{-24}$ (42%)	gp020 $9e^{-24}$ (42%)	gp17 $5e^{-08}$ (28%)	-	-	-	DNA primase
gp6	449	gp021 $2e^{-141}$ (49%)	gp021 $2e^{-141}$ (49%)	gp19 $1e^{-124}$ (43%)	-	gp15 $7e^{-59}$ (33%)	-	DNA helicase
gp7	68	-	-	-	-	-	-	ORFan
gp8	39	-	-	-	-	-	-	ORFan
gp9	123	-	-	-	-	-	-	ORFan
gp10	486	gp022 $3e^{-159}$ (49%)	gp022 $3e^{-159}$ (48%)	gp20 $6e^{-119}$ (43%)	-	-	gp004 $7e^{-103}$ (38%)	A1 protein
gp11	224	-	-	-	-	-	-	Hypothetical protein
gp12	55	-	-	-	-	-	-	ORFan
gp13	331	gp026 $2e^{46}$ (35%)	gp026 $2e^{-46}$ (34%)	gp23 $3e^{-21}$ (28%)	-	-	-	ATP-dependent DNA ligase
gp14	180	-	-	-	-	-	-	ORFan
gp15	786	gp028 0.0 (52%)	gp028 0.0 (52%)	gp26 0.0 (49%)	-	gp 19 $5e^{-114}$ (32%)	-	DNA-directed DNA polymerase
gp16	291	gp029 $9e^{-91}$ (52%)	gp029 $8e^{-92}$ (53%)	gp28 $3e^{-39}$ (35%)	-	gp21 $3e^{-27}$ (33%)	-	Hypothetical protein
gp17	77	-	-	-	-	-	-	ORFan
gp18	38	-	-	-	-	-	-	ORFan
gp19	317	gp031 $2e^{-125}$ (56%)	gp031 $2e^{-125}$ (56%)	gp29 $3e^{-76}$ (41%)	-	gp22 $6e^{-28}$ (32%)	-	5'-3' exonuclease
gp20	61	-	-	-	-	-	-	ORFan
gp21	77	-	-	-	-	-	-	ORFan
gp22	128	gp034 $2e^{-20}$ (42%)	gp034 $1e^{-26}$ (42%)	gp32 $1e^{-25}$ (37%)	-	gp23 $2e^{-16}$ (38%)	-	Recombination endonuclease VII
gp23	816	gp036 0.0 (49%)	gp037 0.0 (48%)	gp33 0.0 (42%)	gp1 $3e^{-123}$ (49%)	gp26 $1e^{-96}$ (29%)	-	T7-like RNA polymerase
gp24	57	gp037 $4e^{-15}$	gp038 $3e^{-13}$	-	-	-	-	Hypothetical protein

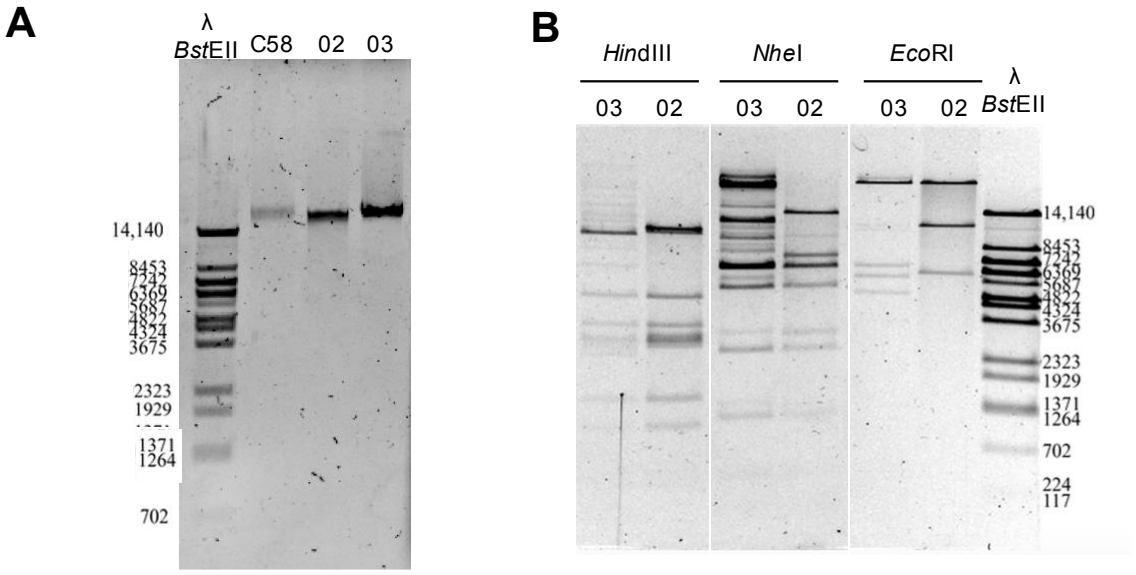
		(60%)	(60%)					
gp25	66	-	-	-	-	-	-	Hypothetical protein
gp26	155	-	-	-	-	-	-	N-acetyltransferase
gp27	88	-	-	-	-	-	-	ORFan
gp28	68	-	-	-	-	-	-	ORFan
gp29	533	gp043 $1e^{-173}$ (50%)	gp042 $9e^{-176}$ (48%)	gp37 $9e^{-104}$ (36%)	gp8 $8e^{-55}$ (29%)	gp30 $3e^{-37}$ (29%)	-	Tail-head connector protein
gp30	296	gp043 $1e^{-41}$ (38%)	gp044 $1e^{-40}$ (38%)	gp38 $8e^{-19}$ (31%)	-	gp31 0.001 (30%)	-	Capsid assembly protein
gp31	327	gp044 $7e^{-154}$ (66%)	gp045 $2e^{-153}$ (66%)	gp39 $8e^{-123}$ (53%)	-	gp32 $3e^{-33}$ (27%)	-	Major capsid protein
gp32	212	gp045 $5e^{-60}$ (43%)	gp046 $5e^{-60}$ (43%)	-	-	-	-	Tail tubular protein A
gp33	823	gp046 0.0 (44%)	gp047 0.0 (44%)	gp41 $9e^{-92}$ (29%)	gp12 $5e^{-62}$ (27%)	gp34 $1e^{-46}$ (28%)	-	Tail tubular protein B
gp34	169	gp047 $4e^{-18}$ (40%)	gp048 $4e^{-18}$ (40%)	-	-	-	-	Internal virion protein
gp35	1192	gp048 $1e^{-68}$ (43%)	gp049 $1e^{-68}$ (43%)	-	-	-	-	Cell wall hydrolase; M15 peptidase
gp36	1255	gp049 0.0 (33%)	gp050 0.0 (33%)	gp44 $5e^{-78}$ (27%)	-	gp37 $1e^{-19}$ (24%)	-	Internal virion protein
gp37	507	gp050 $3e^{-29}$ (46%)	gp051 $2e^{-28}$ (46%)	gp45 $1e^{-19}$ (41%)	-	-	-	Tail fiber protein
gp38	181	-	-	-	-	-	-	Hypothetical protein
gp39	59	gp052 $3e^{-13}$ (41%)	gp053 $3e^{-12}$ (36%)	-	-	-	-	Hypothetical protein
gp40	107	gp053 $2e^{-20}$ (41%)	-	gp47 $1e^{-09}$ (33%)	-	gp42 $5e^{-05}$ (33%)	-	Terminase, small subunit
gp41	612	gp054 0.0 (65%)	gp055 0.0 (64%)	gp48 0/0 (58%)	gp19 $4e^{-87}$ (35%)	gp43 $1e^{-132}$ (40%)	-	Terminase, large subunit
gp42	122	gp055 $1e^{-14}$ (37%)	gp056 $1e^{-14}$ (37%)	-	-	-	-	Hypothetical protein
gp43	91	-	-	-	-	-	-	ORFan
gp44	149	-	-	-	-	-	-	ORFan
gp45	64	-	-	-	-	-	-	ORFan
gp46	289	-	-	-	-	-	-	Hypothetical protein
gp47	47	-	-	-	-	-	-	ORFan
gp48	222	gp006 $1e^{-25}$ (34%)	gp004 $6e^{-26}$ (35%)	-	-	-	-	Hypothetical protein
gp49	124	-	-	-	-	-	-	ORFan
gp50	61	-	-	-	-	-	-	ORFan

gp51	65	-	-	-	-	-	-	ORFan
gp52	88	-	-	-	-	-	-	ORFan
gp53	50	-	-	-	-	-	-	ORFan
gp54	89	-	-	-	-	-	-	ORFan
gp55	107	-	-	-	-	-	-	ORFan
gp56	78	-	-	-	-	-	-	ORFan
gp57	91	-	-	-	-	-	-	ORFan
gp58	40	-	-	-	-	-	-	ORFan

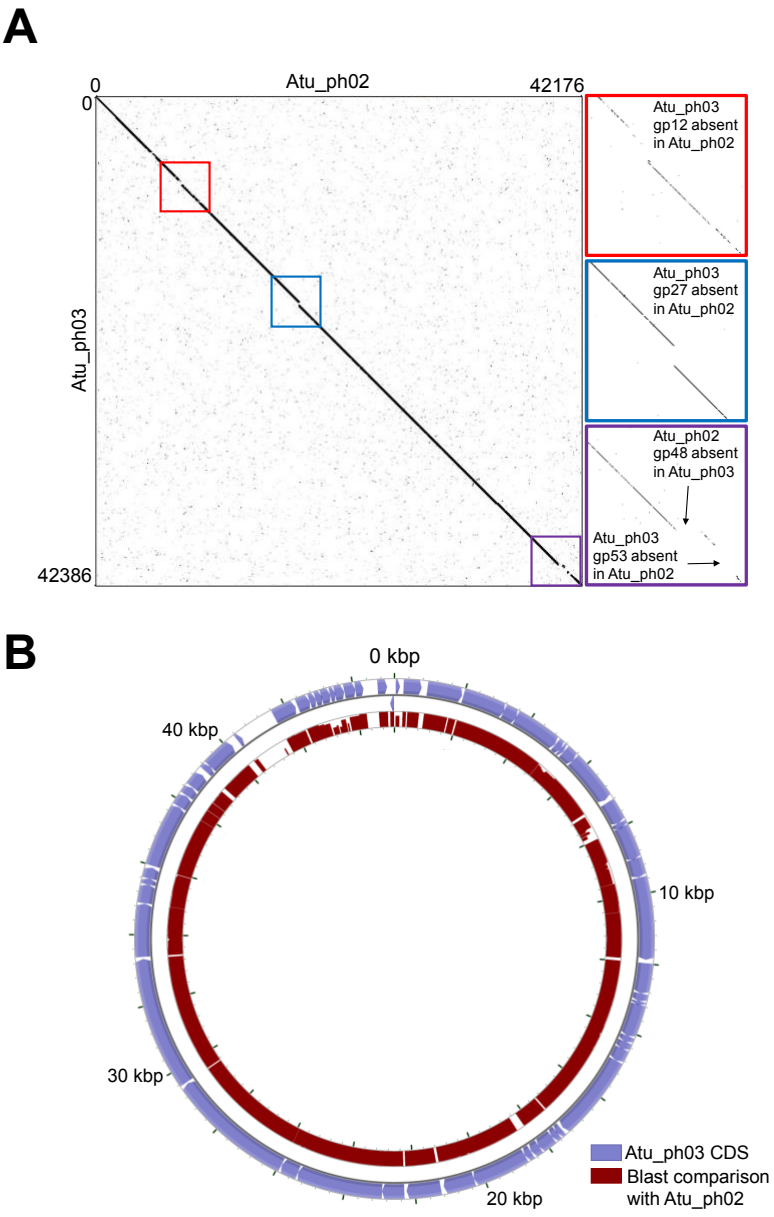
[a]E-values and percent identity were determined by blastp analysis using each predicted ORF in Atu_ph03 as query against the protein databases for Rhizobium phage RHEph02 (taxid:1220602), Rhizobium phage RHEph08 (taxid:1220715), Phage MedPE-SWcel-C56 (taxid: 1871314), Bacteriophage T7 (taxid: 10760), Bacteriophage phiKMV (taxid: 204270) and Bacteriophage T5 (taxid: 10726). – indicates that a significant hit was not detected in the pairwise comparison.

[b] ORFans are predicted proteins that do not have significant hits in the nr database (1). Hypothetical proteins share homology with proteins in the nr database.

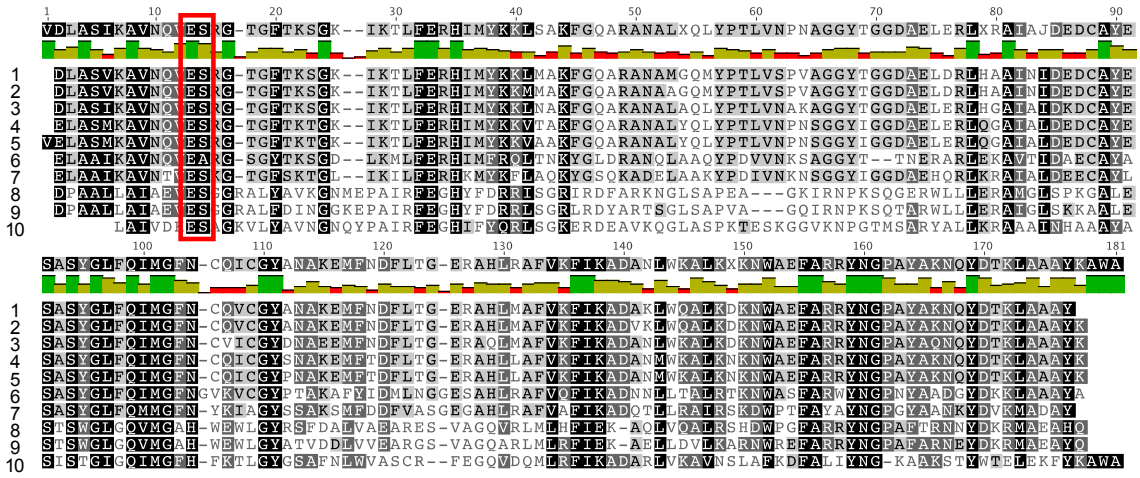
SUPPLEMENTAL FIGURES



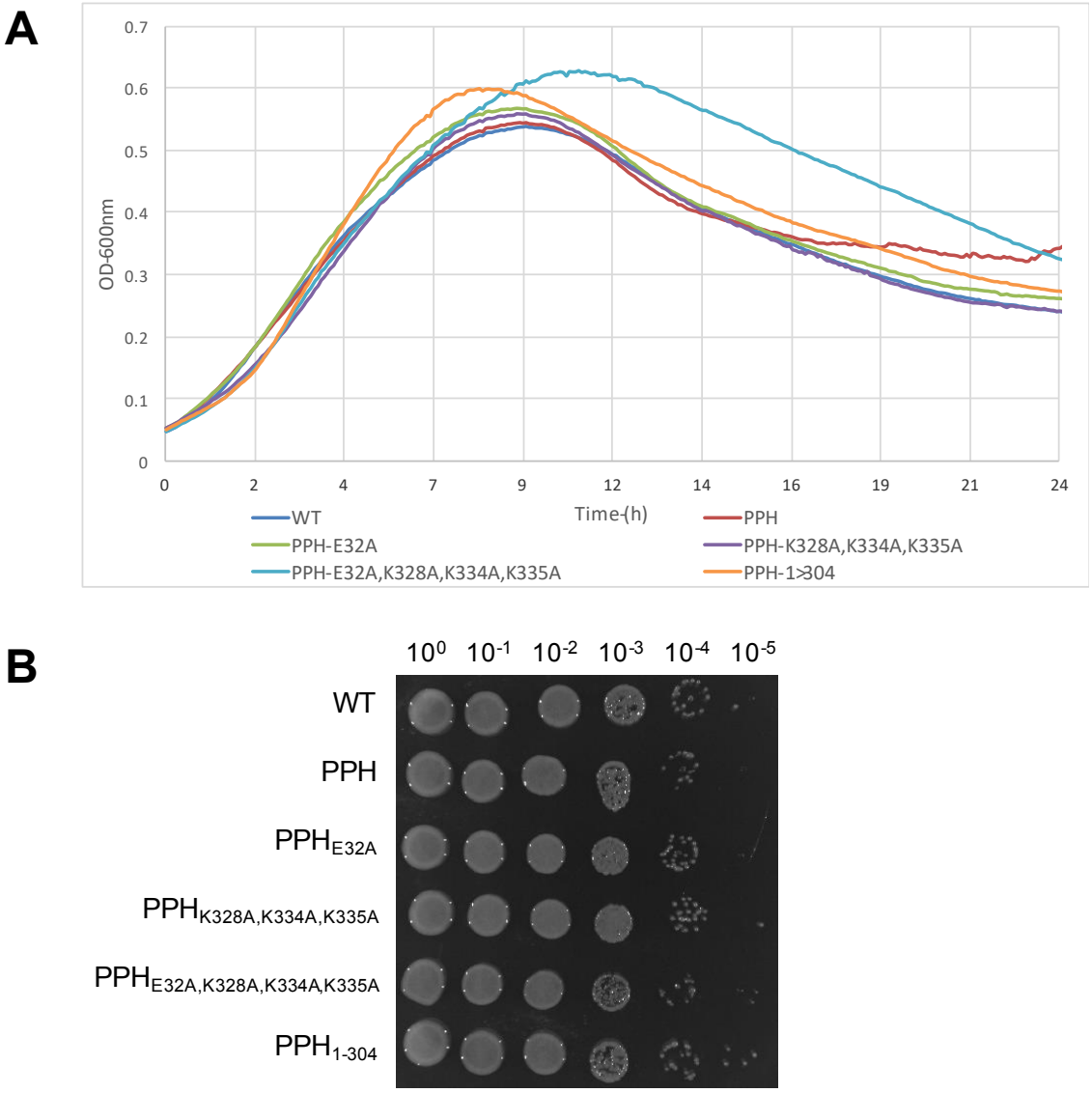
Supplemental Figure 2-S1. Initial characterization of phage genomic DNA shows Atu_ph02 and Atu_ph03 are distinct. (A) Agarose gel containing undigested genomic DNA extracted from *A. tumefaciens* strain C58, phage Atu_ph02 (02), and phage Atu_ph03 (03). (B) Restriction fragment pattern analysis of Atu_ph02 (02) and Atu_ph03 (03) genomic DNA digested with *EcoR*I, *Nhe*I, and *Hin*dIII.



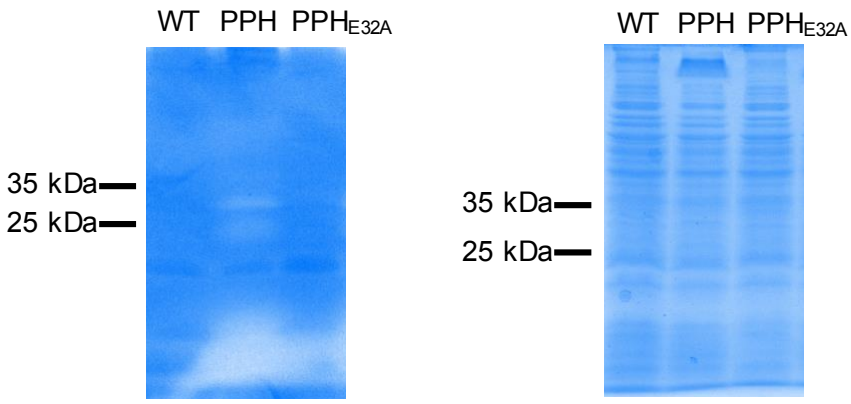
Supplemental Figure 2-S2. Phage Atu_ph03 and Atu_ph02 are very similar. (A) Dot plot analysis comparing the nucleotide sequences of phages Atu_ph03 and Atu_ph02 genomes. Insets highlight areas of difference. (B) Blast analysis of the protein sequences encoded in Atu_ph02 and Atu_ph03. The CDS of Atu_ph03 (blue arrows) and the Blast comparison with Atu_ph02 (maroon) are shown.



Supplemental Figure 2-S3. Clustal alignment of DUF3380 domains from various phage proteins with similarity to the DUF3380 from *Salmonella* phage 10 endolysin and bacterial PG-binding proteins with similarity to the DUF3380 found in PPH. Conserved ES residues are present in all sequences and shown in a red box. Consensus identity for the sequence is mapped along the top of the alignment. Green = 100% identical, gold = 30-100%, red = <30%, no color = 0%. (1) *Dickeya* virus Limestone putative endolysin YP_007237392.1 (aa 90-260), (2) *Shigella* phage Ag3 hypothetical protein YP_003358573.1 (aa 90-261), (3) *Klebsiella* phage 0507-KN2-1 phage-encoded PG binding protein YP_008531963.1 (aa 90-261), (4) *Salmonella* phage 10 endolysin gp110 ANK36008.1 (aa 90-261), (5) *Salmonella* phage Vil phage encoded PG-binding protein YP_004327457.1 (aa 90-261), (6) *Serratia* phage phiMAM1 PG-binding protein YP_007349105.1 (aa 90-261), (7) *Erwinia* phage phiEa2809 putative PG binding protein YP_009147574.1 (90-261), (8) *Brucella abortus* hypothetical protein WP_006091019.1 (20-190), (9) *Ochrobacterium anthropis* PG-binding protein WP_061347616.1 (aa 20-190), (10) *Agrobacterium* phage Atu_ph03 PPH (aa 26-196).



Supplemental Figure 2-S4. Growth of *A. tumefaciens* with plasmids to express variants of *pph* under uninduced conditions. (A) Growth curve of *A. tumefaciens* growth when expressing plasmid pSRKKm with variants of *pph* under uninduced conditions. (B) Cell viability of *A. tumefaciens* containing plasmids to express variants of *pph* grown under uninduced conditions.



Supplemental Figure 2-S5. Clearing of peptidoglycan is observed when PPH is expressed in *A. tumefaciens*. Zymogram (left) and SDS polyacrylamide gel (right) loaded with 30 μg whole cell lysates of *A. tumefaciens* lacking PPH, expressing PPH, and expressing PPH_{E32A}.

SUPPLEMENTAL REFERENCES

1. **Yin Y, Fischer D.** 2008. Identification and investigation of ORFans in the viral world. BMC Genomics **9:**24.

Chapter 3

Larger than life: Isolation and genomic characterization of a jumbo phage that infects the bacterial plant pathogen, *Agrobacterium tumefaciens*

Author Contributions

HA, MB, KP, and J-PN conducted experiments. HA and PB designed experiments and analyzed data for Figures 3-1 through 3-6. KP did preliminary experiments involving the initial isolation of Atu_ph07. MB and J-PN contributed data to Figure 3-7. HA, MB, RL, and PB contributed to writing and editing of the manuscript.

Published as

Attai, H., Boon, M., Phillips, K., Noben, J.-P., Lavigne, R., Brown, P. J. B. (2018). Larger than life: Isolation and genomic characterization of a jumbo phage that infects the bacterial plant pathogen, *Agrobacterium tumefaciens*. *Frontiers in Microbiology*, 9. doi: 10.3389/FMICB.2018.01861

ABSTRACT

Agrobacterium tumefaciens is a plant pathogen that causes crown gall disease, leading to the damage of agriculturally-important crops. As part of an effort to discover new phages that can potentially be used as biocontrol agents to prevent crown gall disease, we isolated and characterized phage Atu_ph07 from Sawyer Creek in Springfield, MO, using the virulent *Agrobacterium tumefaciens* strain C58 as a host. After surveying its host range, we found that Atu_ph07 exclusively infects *Agrobacterium tumefaciens*. Time-lapse microscopy of *A. tumefaciens* cells subjected to infection at a multiplicity of infection (MOI) of 10 with Atu_ph07 reveals that lysis occurs within 3 h. Transmission electron microscopy (TEM) of virions shows that Atu_ph07 has a typical *Myoviridae* morphology with an icosahedral head, long tail, and tail fibers. The sequenced genome of Atu_ph07 is 490 kbp, defining it as a jumbo phage. The Atu_ph07 genome contains 714 open reading frames (ORFs), including 390 ORFs with no discernable homologs in other lineages (ORFans), 214 predicted conserved hypothetical proteins with no assigned function, and 110 predicted proteins with a functional annotation based on similarity to conserved proteins. The proteins with predicted functional annotations share sequence similarity with proteins from bacteriophages and bacteria. The functionally annotated genes are predicted to encode DNA replication proteins, structural proteins, lysis proteins, proteins involved in nucleotide metabolism, and tRNAs. Characterization of the gene products reveals that Atu_ph07 encodes homologs of sixteen T4 core proteins and is closely related to Rak2-like phages. Using ESI-MS/MS, the majority of predicted structural proteins could be experimentally confirmed and 112 additional virion-associated proteins were identified. The genomic characterization of Atu_ph07 suggests

that this phage is lytic and the dynamics of Atu_ph07 interaction with its host indicate that this phage may be suitable for use as a biocontrol agent.

Keywords: *Agrobacterium*, jumbo bacteriophage, mass spectrometry, genomics, biocontrol, Atu_ph07

Introduction

Bacteriophages, or phages, are the most abundant biological entities on the planet (Clokie et al., 2011). Phages are viruses that specifically infect bacteria, often causing lysis. Phage-mediated host cell lysis of bacteria impacts environments both directly through release of dissolved organic carbon and micronutrients and indirectly by modulation of the microbial communities (Srinivasiah et al., 2008). Phages also contribute to horizontal gene transfer and host cell evolution. Although bacteria-phage coevolution significantly drives gene diversity and microbial evolution (Koskella and Brockhurst, 2014), there is a huge number of viral genes that encode proteins of unknown function (Hatfull, 2015). Thus, research on phage biology, phage-host interactions, and phage-derived enzymes has recently reemerged, particularly in the context of medical and biotechnology applications (Santos et al., 2018).

In one such application, phage cocktails can be used as a form of biocontrol against plant pathogens, as seen in successful experiments with *Xanthomonas* species, *Ralstonia solanacearum*, *Pseudomonas syringae*, and *Dickeya solani* (Adriaenssens et al., 2012; Buttimer et al., 2017b; Rombouts et al., 2016). *Agrobacterium tumefaciens* is a Gram-

negative bacterium that causes crown gall disease in flowering plants (Dandekar, 2003) and phage cocktails may be a viable option to improve biocontrol of this phytopathogen; however, there are only three well-characterized *Agrobacterium* phages: Atu_ph02, Atu_ph03, and 7-7-1 (Attai et al., 2017; Kropinski et al., 2012). When searching for additional lytic phages with potential to serve as biocontrol agents against *A. tumefaciens*, we isolated a unique jumbo phage, Atu_ph07, with a dsDNA genome size of 490,380 bp.

Jumbo phages have genomes exceeding 200 kbp (Hendrix, 2009) and are less frequently isolated since they are often eliminated during common size-exclusion isolation methods due to their large size (Yuan and Gao, 2017). Most jumbo phages are members of the *Myoviridae* family and contain visible tails. The largest known phage genome belongs to *Bacillus* phage G at 497 kbp (Donelli et al., 1975), followed by *Salicola* phage SCTP-2 at 440 kbp, *Xanthomonas* phage XacN1 at 384 kbp (Yoshikawa et al., 2018), *Pectobacterium* phage CBB at 378 kbp (Buttimer et al., 2017a), *Cronobacter* phage vB_CsaM_GAP32 at 358 kbp (Abbasifar et al., 2014), and *Serratia* phage BF at 357 kbp (Casey et al., 2017). Recently, some of these T4-like jumbo phages (CBB, vB_CsaM_GAP32, BF, vB_KleM-Rak2, K64-1, 121Q, vB_Eco_slurp01, PBECO4) were classified into a new phylogenetic clade called 'Rak2-like viruses' (Buttimer et al., 2017a; Yoshikawa et al., 2018), named after Enterobacteria phage Rak2 (Simoliunas et al., 2013).

In this work, we use phenotypic, genomic, and proteomic approaches to characterize phage Atu_ph07 (formal name according to Kropinski et al., 2009:

vB_AtuM_Atu_ph07). Based upon comparative genome analysis and phylogenetic analysis, Atu_ph07 clusters just outside the Rak2-like phages. Though most Atu_ph07 ORFs encode as-yet uncharacterized hypothetical proteins, this work identifies functions for some key proteins, including experimentally validated structural proteins, and compares those proteins to homologs in related phages.

Materials and Methods

Bacterial strains and culture conditions. Strains used in this study are shown in Table 3-1. *Agrobacterium tumefaciens* strains were cultured in Lysogeny Broth (LB), with the exception of *A. tumefaciens* strain LBA4404, which was grown in yeast mannitol (YM) medium. *Agrobacterium vitis* was cultured using potato dextrose media (Difco), *Rhizobium rhizogenes* was grown in mannitol glutamate yeast (MGY) medium and *Caulobacter crescentus* was grown in peptone-yeast extract (PYE) medium (Poindexter, 1964). *Sinorhizobium meliloti* was grown in LB (Weidner et al., 2013). These strains were grown at 28°C. *Escherichia coli* was grown in LB at 37°C. Liquid cultures were grown with shaking and solid medium was prepared with 1.5% agar.

Clonal isolation of bacteriophage Atu_ph07. Atu_ph07 was isolated from Sawyer Creek in Springfield, MO using *Agrobacterium tumefaciens* strain C58 as the host. Atu_ph07 was isolated using an enrichment protocol (Santamaría et al., 2014) adapted as described previously (Attai et al., 2017).

Partial purification of virions. Virions were concentrated and partially purified from 2 L lysate by polyethylene glycol (PEG) precipitation (Yamamoto et al., 1970) and differential centrifugation. All centrifugations and incubations were performed at 4°C. The starting lysate was distributed evenly into six 500-ml centrifuge bottles and centrifuged at 5,000 rpm for 20 min to pellet bacterial cells. The supernatants were poured into fresh bottles, which were centrifuged as before to pellet residual bacterial cells. The doubly-cleared supernatants were treated with DNase I (Sigma D5025) at a final concentration of 3.5 μg/ml with stirring for 1 h at room temperature to digest bacterial DNA. Solid NaCl was added to a final concentration of 0.5 M with stirring; when the salt was fully dissolved, solid PEG 8000 (Fisher BP233-1) was added gradually to a final concentration of 10% w/w with constant stirring; stirring was continued for another 2 h. The suspension was distributed evenly into six 500-ml centrifuge bottles, which were refrigerated overnight before being centrifuged at 8,000 rpm for 10 min to pellet the virions; supernatants were decanted and discarded; the bottles were centrifuged again and residual supernatants were removed by aspiration. To each bottle, 20 ml 1 X Dulbecco's phosphate-buffered saline with magnesium and calcium (DPBS; Fisher) was added and the pellets dissolved by gentle shaking at 4°C overnight.

The six dissolved pellets were distributed evenly into four 50-ml disposable plastic conical centrifuge tubes, which were centrifuged for 10 min at 5,000 rpm to pellet insoluble material. The cleared supernatants were pooled and 1/9 vol of 5 M NaCl was added. Solid PEG 8000 was added gradually with stirring to a final concentration of 10% w/w, and stirring was continued for another 1 h at room temperature. The suspension was

distributed evenly into 4 50-ml centrifuge tubes, which were centrifuged 20 min at 10,000 rpm to pellet virions. The supernatants were decanted and discarded, and the tubes were centrifuged again briefly. Residual supernatants were aspirated and discarded. To each tube, 15 ml DPBS was added, and the tubes were rotated overnight at 4°C to dissolve the pellets. The tubes were centrifuged for 10 min at 5,000 rpm to pellet insoluble material. The supernatants combined and 0.6 ml of sterile 10-mM phenol red (pH ~7) was added to color the solution cherry-red. This solution was layered onto 2-ml cushions of 5% w/w sucrose in DPBS in 6 14×89 mm Ultraclear centrifuge tubes (Beckman 331372) and centrifuged at 40,000 rpm for 20 min in a Beckman SW41Ti rotor. The clear supernatants were aspirated gently. 1 ml DPBS was added to each tube and the pellets were dissolved by periodic vortexing and standing overnight at 4°C. The dissolved pellets were transferred to 6 1.5-ml microtubes, which were vortexed to complete dissolution and centrifuged at 6,000 rpm for 5 min to pellet insoluble material. The supernatants were transferred to fresh 1.5-ml microtubes, which were centrifuged at 13,000 rpm for 30 min to pellet virions. The supernatants were aspirated, the tubes were centrifuged, and the residual supernatants were aspirated. The pellets were resuspended in 1 ml DPBS by periodic vortexing and standing overnight at 4°C. Finally, the tubes were centrifuged at 6,000 rpm for 5 min to pellet insoluble material and the cleared supernatants were transferred to fresh 1.5-ml microtubes and stored at 4°C. The solutions were notably turbid. Virion concentration was estimated at 10^{12} physical particles/ml by scanning a 1/10 dilution spectrophotometrically, assuming that intact virions have about the same molar absorption coefficient at 260 nm as do naked 490,380-bp DNA molecules. The infective titer was 7×10^{10} plaque forming units (pfu)/ml.

Plaque assays. Whole-plate plaque assays were performed with the soft agar overlay method (Attai et al., 2017). Briefly, 100 µl cells, grown at an optical density at 600 nm (OD_{600}) of ~0.2 and diluted to OD_{600} of 0.05, were mixed with 100 µl phage for 15 min at room temperature prior to dilution to allow attachment. This mixture of cells and phage were serially diluted in LB and added to 3 ml melted 0.15% or 0.3% LB-soft agar. The solution was then overlaid onto a 1% LB-agar plate and swirled for even distribution. For host range testing, serial dilutions of phage were spotted onto a bacterial lawn. A mixture of 100 µl cells (OD_{600} of ~0.2) and 0.3% LB-soft agar was overlaid onto a 1% LB-agar plate. Once the cells solidified, 5 µl of phage dilutions were spotted onto the soft agar. Plates were incubated for 1 to 2 days to allow plaque formation.

Preparation of virion DNA. Virion DNA was prepared essentially as described (Attai et al., 2017). A 500-µl portion of partially purified virions was pipetted into a 1.5-ml microtube and extracted twice with neutralized phenol (liquefied phenol equilibrated twice with 1/10 vol 1 M Tris-HCl pH 8, discarding the small upper phase each time) and once with chloroform:isoamyl alcohol (24:1 v/v) as follows: 500 µl neutralized phenol or chloroform:isoamyl alcohol was added to the microtube, the microtube was vigorously vortexed, and the phases were separated by centrifugation at 13,000 rpm for 2 min; most of the lower (organic) phase was removed and discarded, the microtube was centrifuged as before, and the upper (aqueous) phase containing the DNA was transferred to a fresh microtube, taking care to avoid residual bottom layer and interphase material. Next, 40 µl 3 M sodium acetate (pH adjusted to 6 with acetic acid) and 1 ml 100% ethanol were

mixed with the final extract to precipitate the DNA. The precipitate was pelleted by centrifugation at 13,000 rpm for 10 min and washed gently with 1 ml freezer-cold 70% ethanol. The pellet was air-dried, dissolved in 100 µl 1 mM Tris-HCl pH 7.5, 100 µM Na_2EDTA, and stored at -20°C.

Growth curves. Growth curves were performed by growing bacteria at a starting OD_{600} of 0.05, in LB. Cells were mixed with purified Atu_ph07 in liquid media at the MOIs indicated. Cell growth was measured by the culture turbidity, represented by the absorbance at OD_{600}. Measurements were taken every 10 min for 36 h. Cells were grown at 28°C and shaken for 1 min prior to each reading. The OD_{600} was measured using a BioTek Synergy H1 Hybrid reader. Results were taken in quadruplicate and averaged. For host range testing, Atu_ph07 was added to cells at an MOI of 10.

Time-lapse microscopy. *A. tumefaciens* strain C58 cells were grown to an OD_{600} of 0.2 and infected with Atu_ph07 at an MOI of 10. Infected cells were incubated at room temperature for 15 min to allow phage attachment and 1 µl infected cells were spotted on a 1% agarose pad containing LB, as previously described (Howell et al., 2017). Cells were imaged using a 60× oil immersion objective (1.4 numerical aperture) by differential interference microscopy every 10 min for 24 h using a Nikon Eclipse TiE equipped with a QImaging Rolera EM-C^2 1 K electron-multiplying charge-coupled-device (EMCCD) camera and Nikon Elements imaging software.

Transmission electron microscopy. Virion morphology was observed by applying a small volume of concentrated purified virions onto a freshly, glow-discharged carbon-coated TEM grid and negatively stained with 2% Nano-W (Nanoprobes, LLC, Brookhaven NY). Specimens were observed on a JEOL JEM-1400 transmission electron microscope at 120 kV. Capsid diameters of 100 virions were measured using ImageJ (v.2.0.0) (Schneider et al., 2012). Head lengths were measured from the top of the phage head vertex to the top of the neck (n=114). Head widths were measured from the right vertex of the head to the left vertex, approximately equidistant between the top of the head vertex and top of the neck and perpendicular to the tail (n=118). Tail lengths were measured from the bottom of the head vertex to the baseplate (n=102). Contracted tails were also measured (n=12).

Genome sequencing and assembly. Libraries for genome sequencing were constructed from virion DNA following the manufacturer's protocol and reagents supplied in Illumina's TruSeq DNA PCR-free sample preparation kit (FC-121-3001). Briefly, 2.4 μg of DNA was sheared using standard Covaris methods to generate average fragmented sizes of 350 bp. The resulting 3' and 5' overhangs were converted to blunt ends by an end repair reaction using 3'-to-5' exonuclease and polymerase activities, followed by size selection (350 bp) and purification with magnetic sample purification beads. A single adenosine nucleotide was added to the 3' ends of the blunt fragments followed by the ligation of Illumina indexed paired-end adapters. The adaptor-ligated library was purified twice with magnetic sample purification beads. The purified library was quantified using a KAPA library quantification kit (KK4824), and library fragment sizes were confirmed

by Fragment Analyzer (Advanced Analytical Technologies, Inc.). Libraries were diluted, pooled, and sequenced using a paired-end 75-base read length according to Illumina's standard sequencing protocol for the MiSeq. Library preparation and sequencing were conducted by the University of Missouri DNA Core facility.

Genome annotation. Protein-coding regions were annotated by RAST server (Aziz et al., 2008) and PSI-BLAST (Altschul et al., 1997) with an e-value cut-off of 1e-03. Proteins of interest were analyzed by TMHMM (Krogh et al., 2001) and SignalP 4.1 (Petersen et al., 2011). The presence of tRNAs were detected by tRNAscan-SE (version 2.0) (Lowe and Chan, 2016). Codon usage was analyzed by Geneious (v.11.0.5) (Kearse et al., 2012). Pairwise (%) nucleotide identity was determined using the Mauve plugin in Geneious (Darling et al., 2004).

16S rRNA gene amplification. 16S rRNA gene sequences were amplified by colony PCR using One*Taq* DNA Polymerase (New England Biolabs) and universal primers, 27F and 1492R (Hogg and Lehane, 1999; Turner et al., 1999). Amplified DNA was purified using the GeneJET PCR Purification Kit (Thermo Scientific) and sequenced by the MU DNA Core facility.

Phylogenetic and gene product analysis. Homologs of the major capsid protein, large terminase, and portal vertex protein were identified by BLASTp using an E-value cutoff of 1e-03. Protein alignment was performed by Geneious using ClustalW (v.2.1) and the BLOSUM matrix (Kearse et al., 2012; Larkin et al., 2007). Maximum-likelihood trees

based on phylogeny (PhyML) were built using a Geneious plugin with 100 bootstrap models (Guindon et al., 2010). For the 16S rRNA tree, a ClustalW nucleotide alignment and a neighbor-joining tree were created in Geneious using the Jukes-Cantor genetic distance model. These trees were imported and annotated in iTOL (v3) (Letunic and Bork, 2016).

SDS-PAGE and electron spray ionization mass spectrometry (ESI-MS/MS). Starting from a PEG purified phage stock of $>10^{10}$ pfu/ml, a protein pellet was obtained by chloroform:methanol extraction (1:1:0.75 [vol/vol/vol]). The pellet was resuspended in loading buffer (40% Glycerol [vol/vol], 4% SDS [wt/vol], 200 mM Tris-HCl pH 6.8, 8 mM EDTA, 0.4% Bromophenol blue [wt/vol]) and heated at 95°C for 5 minutes before loading on a 12% SDS-PAGE gel. After separation by gel electrophoresis, virion proteins were visualized by staining in Gelcode™ Blue Safe Protein Stain (Thermo Scientific). Fragments covering the full lane of the gel were subsequently isolated and subjected to trypsin digestion as described by Shevchenko et al., 1996. The samples were then analyzed by nano-liquid chromatography-electrospray ionization tandem mass spectrometry (nanoLC-ESI-MS/MS) and peptides were identified, based on a database containing all predicted phage proteins, using the search engines SEQUEST [v 1.4.0.288] (ThermoFinnigan) and Mascot [v 2.5] (Matrix Science).

Lysogen induction and detection assays. To test if Atu_ph07 produces lysogens, we attempted to induce C58 cells that survived Atu_ph07 infection with mitomycin C and ultra-violet (UV) irradiation. Ten survivor strains were isolated by streak-purifying 3

times on LB-agar and were confirmed to survive Atu_ph07 infection by conducting spot assays. Survivor cells were grown to OD_{600} of 0.4-0.5 and mixed with mitomycin C (Fisher) at a final concentration of 0.5 μg/ml for 2 h at 28°C with shaking or grown to an OD_{600} of 0.6-0.7 irradiated with UV for a time ranging from 3 s to 120 s. In each case, cells were centrifuged at 7,500 rpm for 10 min. The supernatant was filtered through a 0.45 μm column and centrifuged at 4,000 rpm for 10 min. The flow-through was spotted (5 μl) on a lawn of C58 (OD_{600} = 0.2, 0.3% LB-agar) and incubated overnight at 28°C to be observed for plaque formation in comparison to the Atu_ph07 control.

To find prophages in the genomes of survivor strains, we attempted to PCR amplify genes from Atu_ph07 that are not present in C58 using two sets of primers, which amplified nicotinate phosphoribosyltransferase (CDS 242) and adenine-specific methyltransferase (CDS 399). Primers to amplify nicotinate phosphoribosyltransferase (1,299 bp) were 5' ATG ATC GAT ATC GCA ACA 3' (forward) and 5' TTA GAC AAT TAG AGG TGC 3' (reverse) and adenine-specific methyltransferase (759 bp) primers were 5'ATG CAA ATT GGT AAT GGG 3' (forward) and 5' TTA AAA TTC AAA TAG CCC 3' (reverse). PCR was performed using One*Taq* DNA Polymerase (New England Biolabs). DNA from Atu_ph07 or *A. tumefaciens* were used as positive and negative controls, respectively.

Accession number. The genome sequence of Atu_ph07 has been deposited in the GenBank database with the nucleotide accession number MF403008.

RESULTS AND DISCUSSION

Isolation and characterization of Atu_ph07. *Agrobacterium tumefaciens* strain C58 (Watson et al., 1975) was used as a host strain to isolate phages from environmental samples. We isolated phage Atu_ph07 from a water sample from Sawyer Creek in Springfield, MO using a modified phage enrichment protocol (Santamaría et al., 2014). Following filtration of water samples, we noticed the clearing of bacterial cultures after two rounds of incubation with C58. The presence of phage was apparent after performing plaque assays. Virions were concentrated and partially purified using polyethylene glycol (PEG) precipitation and differential centrifugation. Atu_ph07 appeared to make small, turbid plaques on 0.3% soft agar (Figure 3-1A, left), and larger, clearer plaques on 0.15% soft agar (Figure 3-1A, right). Lowering the agar concentration allows propagation of jumbo phages by promoting phage diffusion through the medium. Growth curves of *A. tumefaciens* infected with Atu_ph07 at different multiplicities of infection (MOIs) show that Atu_ph07 inhibits growth of its host after two hours (Figure 3-1B; Supplementary Figure 3-S1A). Time-lapse microscopy of *A. tumefaciens* cells infected with Atu_ph07 at an MOI of 10 shows that Atu_ph07 causes cell lysis within 3 hours (Figure 3-1C). Transmission electron microscopy (TEM) of virions revealed icosahedral heads (length 146 ± 0.6 nm and width 152 ± 0.8 nm) and long tails (136 ± 0.5 nm), as shown in Figure 3-1D. Phage tails appear to be contractile, as shorter tails with an average length of 77 ± 2.6 nm were observed and some heads also appear to be empty. Lastly, the TEMs indicate the presence of tail fibers and tail-associated 'hairy' whiskers (Figure 3-1D, inset). Similar features have also been observed in Enterobacteria phage vB_PcaM_CBB (Buttimer et al., 2017a). While the 'hairy' whiskers are tail-associated in Enterobacteria

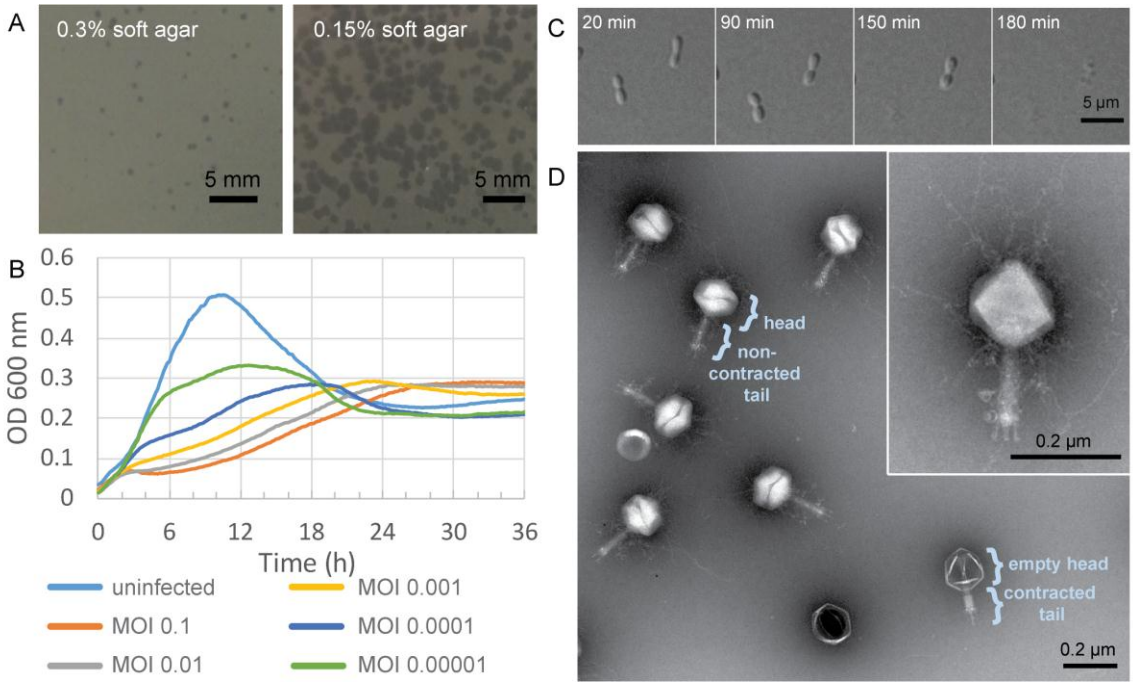


Figure 3-1. Characterization of Atu_ph07. (A) Atu_ph07 forms small plaques on a lawn of *A. tumefaciens* on 0.3% soft agar (left) and larger plaques on 0.15% soft agar (right). (B) Growth curve of *A. tumefaciens* infected with Atu_ph07 at different MOIs. (C) Time-lapse microscopy of *A. tumefaciens* cells infected with Atu_ph07 at an MOI of 10. (D) TEM image of Atu_ph07 shows the phage is in the family *Myoviridae*.

phage vB_PcaM_CBB, these long, thin appendages appear to be primarily capsid-associated in Atu_ph07. Together, the morphology confirms that Atu_ph07 belongs to the family *Myoviridae* (Ackermann, 2009).

Host range of Atu_ph07. Since plaque formation by Atu_ph07 is inconsistent (Figure 3-1A), growth curves in the presence or absence of Atu_ph07 at an MOI of 10 were used to assess the host range of the phage (Supplementary Figure 3-S1). Susceptible test strains have a decreased growth rate and growth yield in the presence of phage when compared to growth in the absence of phage (Figure 3-2, Supplementary Figure 3-S1A), and plaques are formed when an undiluted phage stock with an infective titer of 7×10^{10} pfu/ml is spotted on the test strain. In contrast, resistant strains have comparable growth curves in the presence or absence of phage (Figure 3-2, Supplementary Figure 3-S1B), and no plaques develop when the phage stock is spotted on lawns of the test strain.

To determine if similarity of the bacterial strains plays a role in phage infectivity, we acquired or sequenced the 16S rRNA gene of the host strains and constructed a phylogenetic tree (Figure 3-2). The strains of *A. tumefaciens* form a monophyletic clade consistent with the grouping of these closely related strains into Agrobacteria biovar 1 based on biochemical tests and pathogenicity assays (Keane et al., 1970; Panagopoulos and Psallidas, 1973). While it remains debated if biovar 1 comprises a single species (Sawada et al., 1993; Young et al., 2006) or a complex of related species (Costechareyre et al., 2010; Mougel et al., 2002; Portier et al., 2006), the bacterial strains in this group are heterogeneous, comprising at least nine genomospecies (G1 – G9) (Mougel et al., 2002). *A. tumefaciens* C58, which was the host used to isolate Atu_ph07, belongs to the

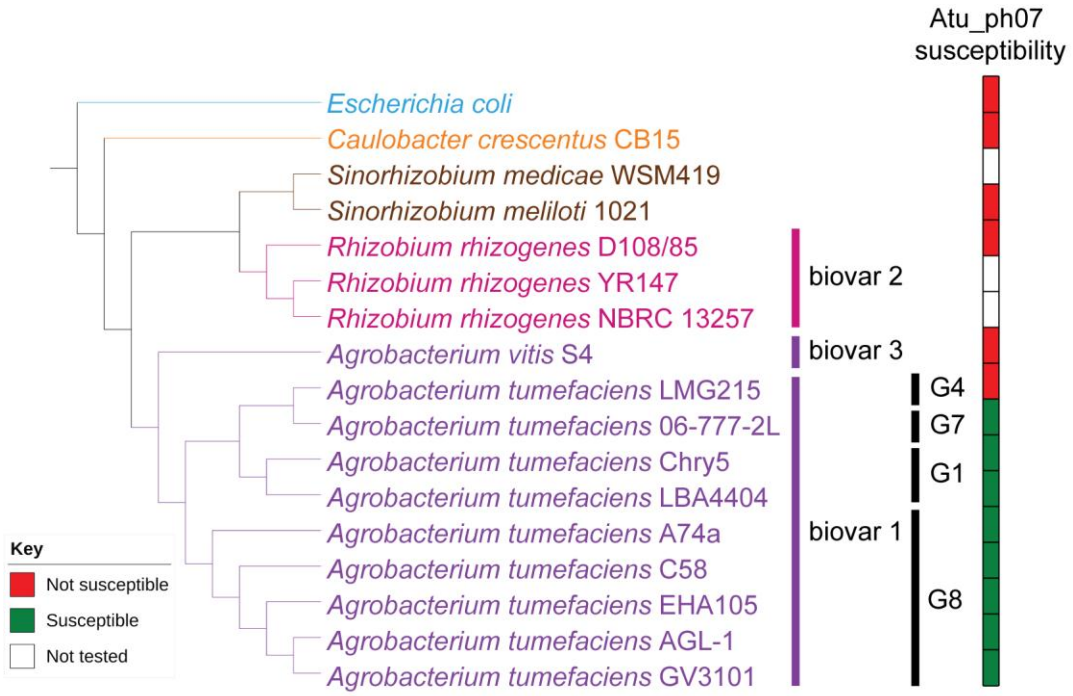


Figure 3-2. Host range of Atu_ph07. Phylogenetic tree of Alphaproteobacteria tested for Atu_ph07 susceptibility in this study was constructed using 16S rRNA sequences. *Agrobacterium* strains are represented in purple, *Rhizobium* strains in pink, *Sinorhizobium* strains in brown, *Caulobacter* in orange, and the outgroup, *E. coli*, is in blue. Agrobacteria biovars and genomospecies are indicated. Phage susceptibility (green) or resistance (red) for each strain is indicated.

G8 genomospecies (Mougel et al., 2002) and each G8 strain tested was susceptible to Atu_ph07 (Figure 3-2). While other Agrobacteria biovar 1 strains belonging to G1 (LBA4404 and Chry5) are susceptible to Atu_ph07, this is not a universal phenotype as strain LMG215 is resistant to Atu_ph07 (Figure 3-2). Representative isolates from biovars 2 and 3, as well as other Alphaproteobacterial strains, are not susceptible to Atu_ph07 infection (Figure 3-2). Thus, the host range of Atu_ph07 appears to be restricted within a subset of Agrobacteria biovar 1.

Genome analysis and phylogeny. The Atu_ph07 genome is 490,380 bp in length, leading to the classification of Atu_ph07 as a jumbo phage. Like other agriculturally-relevant jumbo phages, Atu_ph07 has a low G+C content (37.1%) (Almpanis et al., 2018). The genome was annotated using a combination of Rapid Annotation using Subsystem Technology (RAST) (Aziz et al., 2008) and manual annotation based on PSI-BLAST analysis (Altschul et al., 1997). Atu_ph07 contains 714 open reading frames (ORFs), including 390 ORFans (no discernable homologs in other lineages), 214 conserved hypothetical proteins with no assigned function, and 110 predicted proteins with assigned functions based on similarity to conserved proteins (Table 3-2, Supplementary Table 3-S1, Figure 3-3).

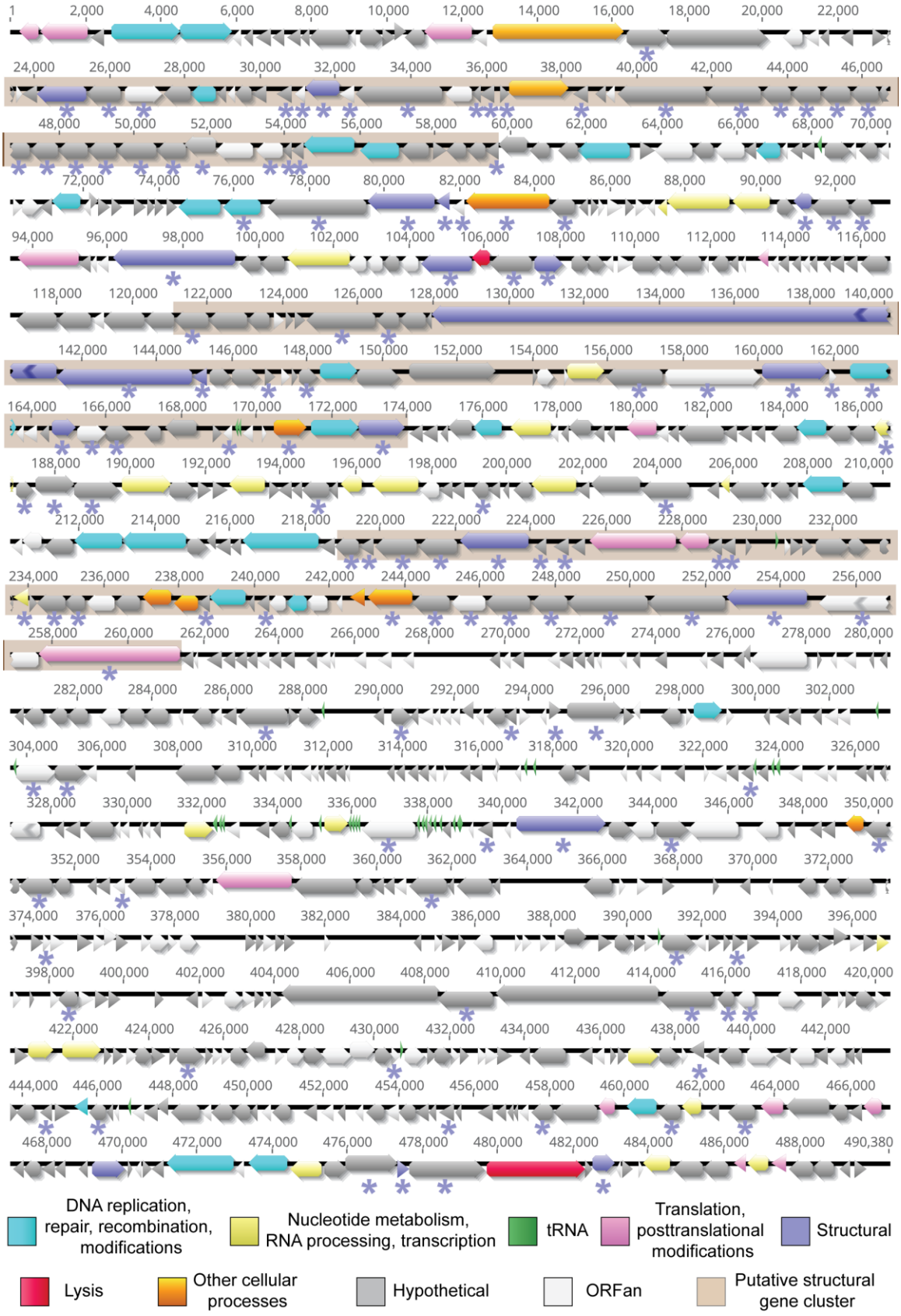
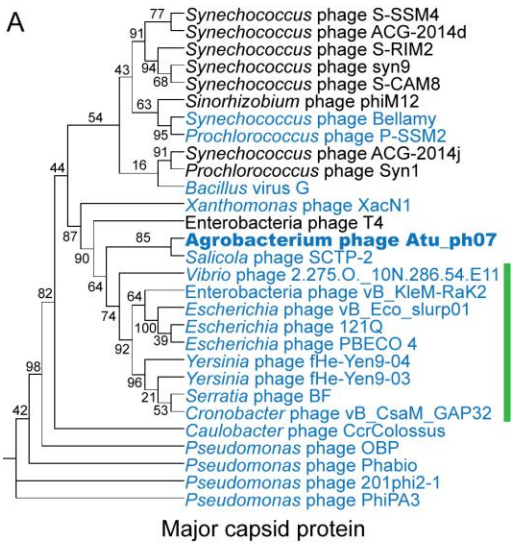
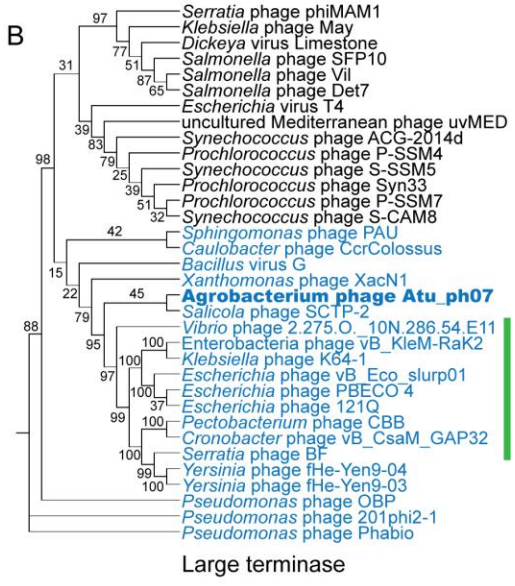

1 2,000 4,000 6,000 8,000 10,000 12,000 14,000 16,000 18,000 20,000 22,000
24,000 26,000 28,000 30,000 32,000 34,000 36,000 38,000 40,000 42,000 44,000 46,000
48,000 50,000 52,000 54,000 56,000 58,000 60,000 62,000 64,000 66,000 68,000 70,000
72,000 74,000 76,000 78,000 80,000 82,000 84,000 86,000 88,000 90,000 92,000
94,000 96,000 98,000 100,000 102,000 104,000 106,000 108,000 110,000 112,000 114,000 116,000
118,000 120,000 122,000 124,000 126,000 128,000 130,000 132,000 134,000 136,000 138,000 140,000
142,000 144,000 146,000 148,000 150,000 152,000 154,000 156,000 158,000 160,000 162,000
164,000 166,000 168,000 170,000 172,000 174,000 176,000 178,000 180,000 182,000 184,000 186,000
188,000 190,000 192,000 194,000 196,000 198,000 200,000 202,000 204,000 206,000 208,000 210,000
212,000 214,000 216,000 218,000 220,000 222,000 224,000 226,000 228,000 230,000 232,000
234,000 236,000 238,000 240,000 242,000 244,000 246,000 248,000 250,000 252,000 254,000 256,000
258,000 260,000 262,000 264,000 266,000 268,000 270,000 272,000 274,000 276,000 278,000 280,000
282,000 284,000 286,000 288,000 290,000 292,000 294,000 296,000 298,000 300,000 302,000
304,000 306,000 308,000 310,000 312,000 314,000 316,000 318,000 320,000 322,000 324,000 326,000
328,000 330,000 332,000 334,000 336,000 338,000 340,000 342,000 344,000 346,000 348,000 350,000
352,000 354,000 356,000 358,000 360,000 362,000 364,000 366,000 368,000 370,000 372,000
374,000 376,000 378,000 380,000 382,000 384,000 386,000 388,000 390,000 392,000 394,000 396,000
398,000 400,000 402,000 404,000 406,000 408,000 410,000 412,000 414,000 416,000 418,000 420,000
422,000 424,000 426,000 428,000 430,000 432,000 434,000 436,000 438,000 440,000 442,000
444,000 446,000 448,000 450,000 452,000 454,000 456,000 458,000 460,000 462,000 464,000 466,000
468,000 470,000 472,000 474,000 476,000 478,000 480,000 482,000 484,000 486,000 488,000 490,380
DNA replication, repair, recombination, modifications
Nucleotide metabolism, RNA processing, transcription
tRNA
Translation, posttranslational modifications
Structural
Lysis
Other cellular processes
Hypothetical
ORFan
Putative structural gene cluster

Figure 3-3. The annotated genome of Atu_ph07. ORFs are represented by functional categories in corresponding colors. Regions shaded in beige represent putative structural protein clusters as identified using ESI-MS/MS analysis. Proteins detected by ESI-MS/MS analysis are indicated with an asterisk (*) below the corresponding ORF.

Due to the high degree of divergence, comparative genome analysis of jumbo phages is challenging. Based on nucleotide identity, Atu_ph07 is most similar to *Synechococcus* phage S-SSM7 (Sullivan et al., 2010). However, the genomes are only 13.1% identical and do not share collinear blocks. Since whole genome alignments did not reveal phages similar to Atu_ph07, we next constructed phylogenetic trees using the sequence of proteins conserved in many jumbo phages. There is no universal gene present in all phages, therefore signature gene products including the major capsid protein, large terminase subunit, and portal vertex protein were selected for the phylogenetic analysis (Adriaenssens and Cowan, 2014). These phylogenies place Atu_ph07 among the jumbo phages in the T4-superfamily (Figure 3-4). Although the genome of Atu_ph07 only shares 33 homologous ORFs with the genome of bacteriophage T4 (Supplementary Table 3-S2), core proteins involved in phage morphogenesis and DNA replication, recombination, and repair were identified (Supplementary Table 3-S3) (Miller et al., 2003). The phylogenetic trees are consistent with the recent characterization of *Xanthomonas* phage XacN1 (Yoshikawa et al., 2018), which suggested that XacN1, *Salicola* phage SCTP-2 and Atu_ph07 are distantly related to the Rak2-like jumbo phages.


A
Synechococcus phage S-SSM4
Synechococcus phage ACG-2014d
Synechococcus phage S-RIM2
Synechococcus phage syn9
Synechococcus phage S-CAM8
Sinorhizobium phage phiM12
Synechococcus phage Bellamy
Prochlorococcus phage P-SSM2
Synechococcus phage ACG-2014j
Prochlorococcus phage Syn1
Bacillus virus G
Xanthomonas phage XacN1
Enterobacteria phage T4
Agrobacterium phage Atu_ph07
Salicola phage SCTP-2
Vibrio phage 2.275.O._10N.286.54.E11
Enterobacteria phage vB_KleM-RaK2
Escherichia phage vB_Eco_slurp01
Escherichia phage 121Q
Escherichia phage PBECO 4
Yersinia phage fHe-Yen9-04
Yersinia phage fHe-Yen9-03
Serratia phage BF
Cronobacter phage vB_CsaM_GAP32
Caulobacter phage CcrColossus
Pseudomonas phage OBP
Pseudomonas phage Phabio
Pseudomonas phage 201phi2-1
Pseudomonas phage PhiPA3
Major capsid protein


B
Serratia phage phiMAM1
Klebsiella phage May
Dickeya virus Limestone
Salmonella phage SFP10
Salmonella phage Vil
Salmonella phage Det7
Escherichia virus T4
uncultured Mediterranean phage uvMED
Synechococcus phage ACG-2014d
Prochlorococcus phage P-SSM4
Synechococcus phage S-SSM5
Prochlorococcus phage Syn33
Prochlorococcus phage P-SSM7
Synechococcus phage S-CAM8
Sphingomonas phage PAU
Caulobacter phage CcrColossus
Bacillus virus G
Xanthomonas phage XacN1
Agrobacterium phage Atu_ph07
Salicola phage SCTP-2
Vibrio phage 2.275.O._10N.286.54.E11
Enterobacteria phage vB_KleM-RaK2
Klebsiella phage K64-1
Escherichia phage vB_Eco_slurp01
Escherichia phage PBECO 4
Escherichia phage 121Q
Pectobacterium phage CBB
Cronobacter phage vB_CsaM_GAP32
Serratia phage BF
Yersinia phage fHe-Yen9-04
Yersinia phage fHe-Yen9-03
Pseudomonas phage OBP
Pseudomonas phage 201phi2-1
Pseudomonas phage Phabio
Large terminase

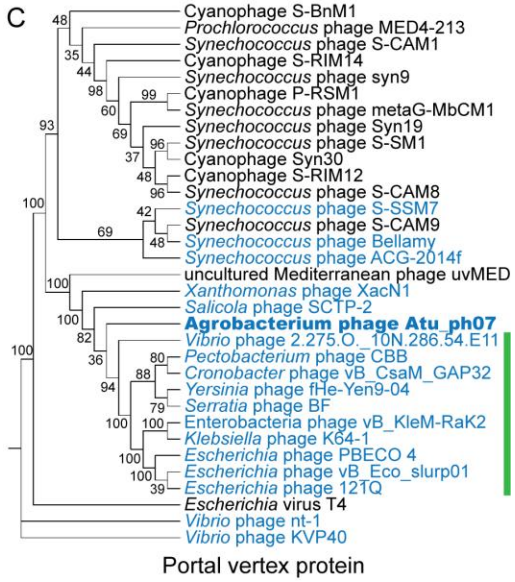

C
Cyanophage S-BnM1
Prochlorococcus phage MED4-213
Synechococcus phage S-CAM1
Cyanophage S-RIM14
Synechococcus phage syn9
Cyanophage P-RSM1
Synechococcus phage metaG-MbCM1
Synechococcus phage Syn19
Synechococcus phage S-SM1
Cyanophage Syn30
Cyanophage S-RIM12
Synechococcus phage S-CAM8
Synechococcus phage S-SSM7
Synechococcus phage S-CAM9
Synechococcus phage Bellamy
Synechococcus phage ACG-2014f
uncultured Mediterranean phage uvMED
Xanthomonas phage XacN1
Salicola phage SCTP-2
Agrobacterium phage Atu_ph07
Vibrio phage 2.275.O._10N.286.54.E11
Pectobacterium phage CBB
Cronobacter phage vB_CsaM_GAP32
Yersinia phage fHe-Yen9-04
Serratia phage BF
Enterobacteria phage vB_KleM-RaK2
Klebsiella phage K64-1
Escherichia phage PBECO 4
Escherichia phage vB_Eco_slurp01
Escherichia phage 121Q
Escherichia virus T4
Vibrio phage nt-1
Vibrio phage KVP40
Portal vertex protein

Figure 3-4. Phylogenetic comparison of Atu_ph07 and related phages. Phylogenetic trees of phages based on alignments of the (A) major capsid protein, (B) large terminase, and (C) portal vertex protein. Jumbo phages are labeled in blue. Bootstrap values of 100 replicates are represented by circles. Rak2-like phages are indicated by a green line.

The ORFs in the Atu_ph07 genome were compared to those in XacN1, SCTP-2 and Rak2-like phage genomes (Figure 3-5, Supplementary Table 3-S2). The genome of phage SCTP-2 has the highest level of gene conservation with 141 genes (19.7% of Atu_ph07 ORFs) in common with the Atu_ph07 genome. Overall, the gene composition of Atu_ph07 is not well conserved with the Rak2-like phages (Figure 3-5, Supplementary Table 3-S2), consistent with proposal that Atu_ph07, together with SCTP-2, may belong to a new clade that comprises a sister group to the Rak2-like phages (Yoshikawa et al., 2018).

Functional annotation. Most of the 110 Atu_ph07 ORFs that can be assigned to a functional annotation are predicted to function in phage morphogenesis and replication. The functional annotation is enriched in proteins involved in DNA replication, modification, recombination or repair (Figure 3-3, light blue arrows), nucleotide metabolism (Figure 3-3, yellow arrows), translation and posttranslational proteins (Figure 3-3, pink arrows) and structural proteins (Figure 3-3, purple arrows) (Supplementary Table 3-S1).

DNA Replication, Repair, and Recombination. Atu_ph07 encodes 26 enzymes involved in DNA replication, repair, and recombination (Figure 3-3, light blue arrows, Supplementary Table 3-S1). The majority of these enzymes are all highly conserved in XacNI, SCTP-2, and the Rak2-like phages (Figure 3-5, Supplementary Table 3-S2). These highly conserved enzymes include homologs of six of the T4 core proteins involved in DNA replication, repair, and recombination (Petrov et al., 2010). The

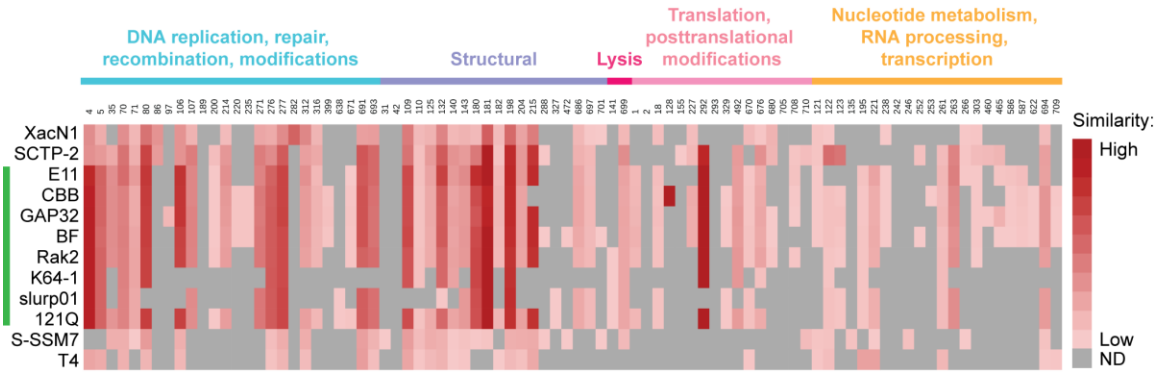


Figure 3-5. Similarity of annotated gene products in Atu_ph07 and related phages. Heat map displaying Atu_ph07 gene products compared with homologs in 12 related phages, including members of the Rak2-like phages (indicated with a green line). Intensity of the red color indicates the degree of similarity among homologs. Grey boxes indicate that a homolog with an E-value smaller than 1e-03 was not detected (ND). Gene products are organized by functional category and Atu_ph07 gp numbers are indicated.

enzymes with homology to the T4 core proteins are predicted to function as part of the DNA helicase-primase complex (gp70), DNA polymerases (gp276, gp277), sliding clamp loader (gp312), and recombination-related endonucleases (gp691, gp693) (Supplementary Table 3-S3). One of the endonucleases (gp693) is directly upstream of a protein (gp694) with similarity to the RNA polymerase sigma factor for late transcription. This gene product shares similarity with XacNI, SCTP-2, T4, and most of the Rak2-like phages (Supplementary Table 3-S2). The Atu_ph07 genome also encodes a predicted DNA polymerase III alpha subunit (gp316) and epsilon subunit (gp671) suggesting that the polymerase may contribute to both DNA replication and 3'-5' exonuclease activity. The DNA polymerase III subunits are conserved in most of the Rak2-like phages (Figure 3-5, Supplementary Table 3-S2). Atu_ph07 encodes three type II topoisomerase proteins involved in chromosome partitioning (gp4, gp5, gp200) (Kato et al., 1992). Gp4 is highly conserved with DNA gyrase subunit B encoded by the Rak2-like phages and gp5 encodes topoisomerase IV subunit A, also well-conserved in the Rak2-like phages (Figure 3-5). Together, the presence of these highly conserved genes suggests that Atu_ph07 encodes the proteins necessary to complete phage DNA replication and DNA-related functions including recombination and repair.

Nucleotide metabolism. To supplement the nucleotide pool required for phage DNA and RNA synthesis, T4-like phage genomes encode enzymes for nucleotide metabolism (Petrov et al., 2010). The Atu_ph07 genome contains several enzymes predicted to contribute to nucleotide metabolism (Figure 3-3, Supplementary Table 3-S1). These include both alpha and beta subunits (gp122, gp123) of ribonucleotide reductase (RNR)

of class 1a (*nrdA* and *nrdB*), which are involved in oxygen-dependent nucleotide metabolism of ribonucleotides into deoxyribonucleotides, a step that is needed for DNA replication (Dwivedi et al., 2013). Phages generally acquire RNR proteins from their hosts to provide them an evolutionary advantage. The alpha subunit of Atu_ph07 is similar to the RNR alpha subunit encoded by *A. tumefaciens* strain B6. RNR proteins catalyze nucleotide metabolism with the help of glutaredoxin and thioredoxin (Sengupta, 2014). Two putative glutaredoxin proteins are encoded by Atu_ph07 (gp121, gp266), one of which (gp121) is directly adjacent to the alpha subunit of RNR. Thioredoxin is encoded by gp221.

T4-like phages require ATP and NADH/NAD^+ for important processes like DNA synthesis, transcription, and translation. To metabolize NAD^+, phages use nicotinamide-adenine dinucleotide pyrophosphatase (NUDIX) hydrolases (Bessman et al., 1996; Lee et al., 2017). This family of enzymes is involved in housekeeping functions of the cell, including the hydrolysis of unwanted nucleotides or removal of excess metabolites. Atu_ph07 encodes three putative members of the NUDIX hydrolase superfamily (gp257, gp303, gp557). Other putative proteins involved in nucleotide metabolism include NadR (gp587) and PnuC (gp586). NadR transcriptionally regulates NAD biosynthesis and PnuC is a membrane transporter that allows nicotinamide mononucleotide (NMN) uptake (Foster et al., 1990; Kurnasov et al., 2002).

tRNA genes and tRNA processing genes. The genome of Atu_ph07 encodes 33 tRNA genes, including 32 canonical tRNAs corresponding to all amino acids except asparagine and threonine (Figure 3-6). The remaining tRNA is a suppressor with an anticodon of

UCA indicating read-through of opal (UGA) stop codons. The UGA stop codon is abundant in both the phage (N=267) and *A. tumefaciens* (N= 2,923) genomes suggesting that there are several potential genes targeted by the suppressor tRNA. With the exception of the suppressor tRNA, all of the tRNAs encoded in the Atu_ph07 genome are also found in the *A. tumefaciens* genome suggesting that the tRNAs do not improve decoding capacity; however, some of the phage tRNAs correspond to codons that are more frequently used in the phage genome (Figure 3-6). This observation is consistent with the notion that phage-encoded tRNAs allow translation to be optimized for the codon usage of the phage genome (Bailly-Bechet et al., 2007). In addition to the tRNA genes, the Atu_ph07 genome encodes four tRNA processing proteins. These tRNA processing proteins include tRNA nucleotidyltransferase (gp18) and $tRNA^{His}$-5'-guanylyltransferase (gp227), which are involved in tRNA maturation. Putative peptidyl-tRNA hydrolases (gp680, gp256) function to decrease the pool of peptidyl-tRNAs formed throughout the initiation, elongation, and termination stages of translation.

Clp-like proteins. The Atu_ph07 phage genome encodes seven putative members of the Clp family of proteins that function to degrade proteins, including ClpX (gp2), a prophage Clp protease-like protein (gp128), an ATP-dependent Clp protease ATP-binding subunit (gp155), ClpA (gp292), ClpB (gp492), ATP-dependent Clp protease proteolytic subunit (gp676), and ClpS (gp708). The Clp proteases may contribute to virion assembly or have alternative functions. For example, during phage lambda DNA replication, the ClpX/ClpP protease removes the O protein from the origin of replication

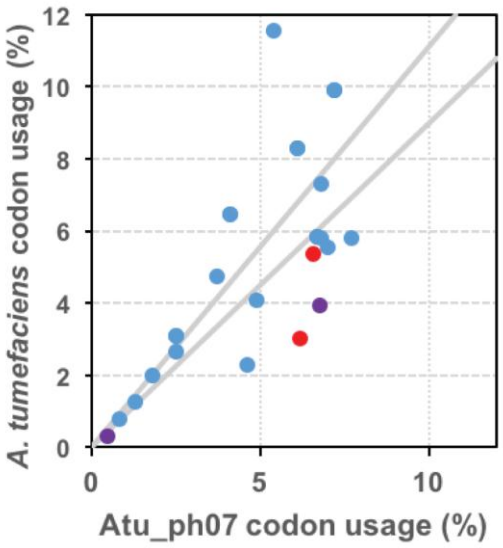


Figure 3-6. tRNAs are encoded in the Atu_ph07 genome. Table (left) shows amino acids, codons, and tRNAs encoded in the Atu_ph07 genome. Graphical representation of codon bias (right) of phage Atu_ph07 and its host *A. tumefaciens* strain C58. Data points represent the usage of each codon in the Atu_ph07 and *A. tumefaciens* genomes. Red points represent codons only found in *A. tumefaciens*, purple points represent codons only found in Atu_ph07, and blue points represent codons found in both genomes. Grey lines outline the region in which codon usage in both genomes is similar.

(Zylicz et al., 1998) and the activity of ClpX/ClpP has been associated with slowing down DNA replication of the phage under poor growth conditions (Węgrzyn et al., 2000). In the Atu_ph07 genome, *clpX* (gp2) is located in close proximity to genes predicted to encode topoisomerase proteins (gp4-5) suggesting that the ClpX protein may function in the regulation of phage DNA replication. In *E. coli*, ClpS is an adaptor protein that modifies the substrate specificity of the ClpA/ClpP protease and contributes to degradation or refolding of protein aggregates (Dougan et al., 2002). ClpA (gp292) is highly conserved in most of the Rak2-like phages (Figure 3-5). The presence of the ClpB (gp492) and DnaJ (gp293) chaperones, which also function in the removal of protein aggregates (Mogk et al., 1999), further suggests that Atu_ph07 may help its host to survive the stress of phage infection long enough for the phage replication cycle to be completed. Together, these observations suggest that Clp proteins likely contribute to diverse aspects of phage biology potentially including virion assembly, DNA replication, and proteolytic clearance of protein aggregates.

Structural Proteins. Based on homology, the genome of Atu_ph07 was predicted to encode 20 proteins involved in phage morphogenesis and structure (Figure 3-3, Supplementary Table 3-S1). Candidate structural genes encode proteins for head morphogenesis and structure (gp125, gp143, gp198, gp204, gp215), baseplate (gp181, gp182, gp686, gp697), tail sheath (gp132, gp140), and tail fibers (gp31, gp288, gp327, gp472). All five of the candidate head proteins, one of the baseplate wedge subunits (gp182), and both of the tail sheath proteins share significant homology with T4 core structural proteins (Supplementary Table 3-S2, 3-S3). The Atu_ph07 genome does not

encode proteins with similarity to T4 core tail fiber proteins (Supplementary Table 3-S3); however short tail fibers are evident when Atu_ph07 is observed under transmission electron microscopy (Figure 3-1D). One of the tail fiber proteins (gp472) is most closely related (51% identity) to a tail fiber protein encoded in the genome of *Agrobacterium* phage Atu_ph02 (Attai et al., 2017), suggesting that these phages may share an entry route into *Agrobacterium* cells.

Based upon the complex morphology of Atu_ph07, we hypothesized that the genome annotation likely underestimates the quantity of proteins involved in phage morphogenesis and structure. To experimentally identify additional structural proteins, electrospray ionization mass spectrometry (ESI-MS/MS) was used (Figure 3-7, Supplementary Table 3-S4). Overall, the proteomic analysis supported the annotation of structural proteins, as all of the head, neck, tail fiber, and most of the tail (5/7) proteins could be identified in the virion proteome. As expected, the most abundant protein observed in the particle proteome is the major capsid protein (gp215, fragment 6 in Figure 3-7). A total of 131 proteins were found among the phage virion proteins, with sequence coverages higher than 5%, or more than one identified unique peptide (Supplementary Table 3-S4). About 78% (102/131) of these proteins do not have an assigned function. The majority of the virion proteins are encoded in three large clusters as shaded in beige in Figure 3-3 (CDS 31-76, CDS 170-215 and CDS 284-329), while the remaining proteins reside in smaller clusters or as separate genes spread across the genome. Notably, CDS 31-76 contains 29 identified proteins of which only two were predicted, indicative of the vast amount of potentially unique structural proteins in

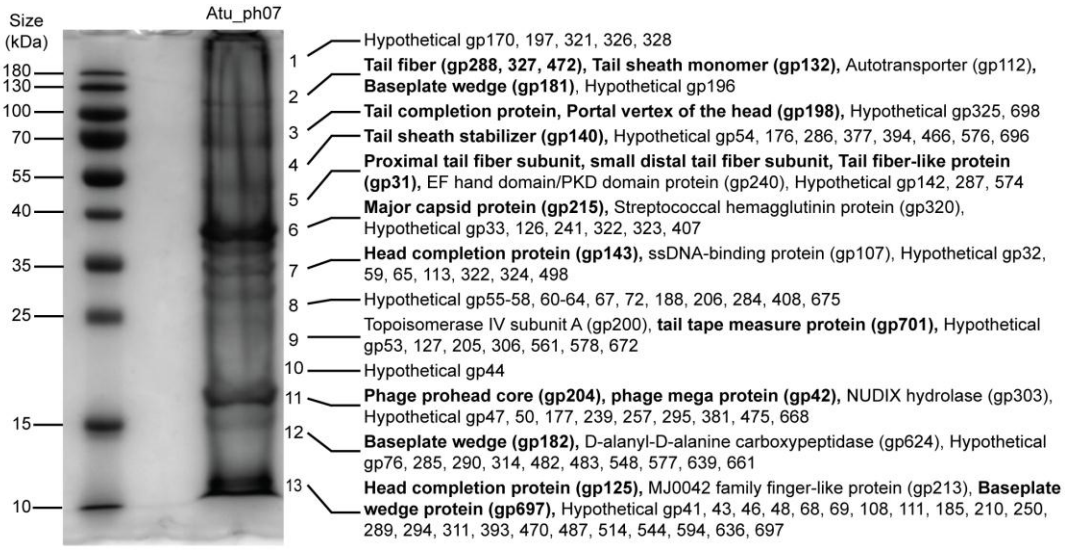


Figure 3-7. SDS-PAGE of Atu_ph07 structural proteins as identified by ESI-MS/MS. Phage proteins were separated by size and excised from the gel for proteomic analysis. Numbers at the right of the gel indicate the position of bands which were excised from the gel. Proteins identified in each band are listed. Bold font indicates validation of annotated structural proteins.

Atu_ph07. The lack of similarity with known virion proteins should not be surprising, as only a small number of jumbo phages with whisker-like structures have been described to date.

Atu_ph07 induces cell lysis. The genome content is insufficient to confidently predict the phage life cycle, however the phage can induce lysis (Figure 3-1B-C). To assess the possibility of a lysogenic phase, we isolated ten variant clones of *A. tumefaciens* strain C58 that survive exposure to Atu_ph07 at high MOI. Lysogens were not induced from any of the ten unsusceptible variant clones by exposing them to UV irradiation or mitomycin C (see Materials and Methods for experimental details). Furthermore, we were unable to PCR-amplify two distinctive Atu_ph07 genes, nicotinate phosphoribosyltransferase (CDS 242) and adenine-specific methyltransferase (CDS 399), from the unsusceptible variants. Consistent with these observations, no integrase- or Cro-like genes, which are required for lysogeny in several temperate phage species, could be identified in the Atu_ph07 genome.

Since we observe that Atu_ph07 can induce cell lysis (Figure 3-1C), we searched for genes encoding candidate lysis proteins in the genome. Atu_ph07 contains two predicted lysozymes (gp141 and gp699). Gp141 is in close proximity to the predicted phage head completion protein and tail sheath monomer, indicating that it may be involved in phage DNA entry. Gp699 has homologues in the Rak2-like phages and is adjacent to three predicted structural proteins—the baseplate hub subunit (gp686), baseplate wedge protein (gp697), and tail tape measure protein (gp701).

To widen our search for candidate lysis proteins, we searched for transmembrane (TM) proteins which can be indicative of the presence of the canonical endolysin-holin-spanin system of host cell lysis (Young, 2014). A TMHMM analysis (Krogh et al., 2001) of the predicted proteins identified 40 predicted TM proteins, only three of which have putative functions: ribonucleotide reductase of class 1a beta subunit (gp123), ribosyl nicotinamide transporter PnuC (gp586), and peptidyl tRNA hydrolase (gp680). Eight of these encoded TM proteins (gp369-gp376) appear consecutively on the genome at ~280 kbp, however each of these are ORFs with no detectable similarity in the database, or ORFans (Yin and Fischer, 2008), making us unable to predict their function as a unit at this time. Thus, at present, the mechanism of Atu_ph07-mediated host cell lysis remains unknown.

CONCLUSION

Several jumbo phages have been recently characterized, many encoding a large number of hypothetical proteins. Recently, a group of T4-like phages have been categorized into a new monophyletic group called 'Rak2-like'. While phage Atu_ph07 clusters just outside this group, many genes share homology with core genes in the Rak2-like phage genomes. Atu_ph07 infects a subset of *Agrobacterium tumefaciens* strains (Figure 3-2) and its ability to infect this plant pathogen makes it a candidate for biocontrol.

The phage biology of Atu_ph07 is likely to be remarkable in its own right. While the genome encodes genes for DNA replication, transcription, translation, nucleotide metabolism, as well as over 130 experimentally confirmed structural proteins, many more molecular mechanisms remain to be unraveled. Understanding the modes through which

a non-living entity can acquire, store, replicate, and express such a vast number of genes to promote its life cycle is a fascinating aspect of phage biology unique to jumbo phages. A logical assumption considering the coding density of this phage is that many of the putative hypothetical proteins have functional significance and provide jumbo phages with an evolutionary advantage in specific ecological niches. Continued exploration of jumbo phages will help elucidate the mechanisms in which diverse bacteriophages have evolved to thrive as the most abundant biological entities in the world.

Conflict of Interest Statement

This research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

HA, MB, KP, and J-PN conducted experiments. All authors designed experiments and analyzed data. HA, MB, RL, and PB contributed to writing and editing of the manuscript.

Funding

This research is supported by startup funds, a research council grant (URC 14-051), and a research board grant (3786-2) from the University of Missouri to PJBB. This work was supported in part by the "Excellence in Electron Microscopy" Award provided by the University of Missouri Electron Microscopy Core and the Office of Research. HA has been supported by the National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health (NIH) under award number T32GM008396 and the U.S.

Department of Education Graduate Assistance in Areas of National Need (GAANN) Fellowship. KP was supported by the IMSD EXPRESS Program via grant number R25GM056901 from the NIGMS, a component of the NIH. RL and MB are supported by a GOA grant from KU Leuven and JPN was supported by Hercules Foundation project R-3986.

Acknowledgements

We thank Jeff Chang at Oregon State University for providing *Agrobacterium* and *Rhizobium* strains, as well as their 16S rRNA sequences. We thank Zhanyuan Zhang at the MU Plant Transformation Core facility for providing *Agrobacterium* strains. We thank Tommi White, DeAna Grant, and Martin Schauflinger of the MU Electron Microscopy Core for help with the transmission electron microscopy. We thank Nathan Bivens and the MU DNA Core for assistance with sequencing the bacteriophages and William Spollen and the MU Research Informatics Core for assistance with genome assembly and GenBank submission. We thank Michelle Williams and Blackman Water Labs for donating water samples from which we isolated Atu_ph07. George Smith provided valuable technical assistance during the initial purification and characterization of Atu_ph07. Finally, we thank George Smith, Linda Chapman, and members of the Brown lab for their feedback during the preparation of this manuscript.

References

Abbasifar, R., Griffiths, M. W., Sabour, P. M., Ackermann, H. W., Vandersteegen, K., Lavigne, R., et al. (2014). Supersize me: *Cronobacter sakazakii* phage GAP32.

Virology 460–461, 138–146. doi:10.1016/j.virol.2014.05.003.

Ackermann, H. W. (2009). Phage classification and characterization. *Methods Mol. Biol.* 501, 127–140. doi:10.1007/978-1-60327-164-6.

Adriaenssens, E. M., and Cowan, D. A. (2014). Using signature genes as tools to assess environmental viral ecology and diversity. *Appl. Environ. Microbiol.* 80, 4470–4480. doi:10.1128/AEM.00878-14.

Adriaenssens, E. M., van Vaerenbergh, J., Vandenheuvel, D., Dunon, V., Ceyssens, P. J., de Proft, M., et al. (2012). T4-related bacteriophage LIMEstone isolates for the control of soft rot on potato caused by "*Dickeya solani*." *PLoS One* 7. doi:10.1371/journal.pone.0033227.

Almpanis, A., Swain, M., Gatherer, D., and Mcewan, N. (2018). Correlation between bacterial G + C content , genome size and the G + C content of associated plasmids and bacteriophages. *Microb. Genomics* 4, 0–7. doi:10.1099/mgen.0.000168.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST:a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi:10.1093/nar/25.17.3389.

Attai, H., Rimbey, J., Smith, G. P., and Brown, P. J. B. (2017). Expression of a peptidoglycan hydrolase from lytic bacteriophages Atu_ph02 and Atu_ph03 triggers lysis of *Agrobacterium tumefaciens*. *Appl. Environ. Microbiol.* 83, e01498-17. doi:10.1128/AEM.01498-17.

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9, 75. doi:10.1186/1471-2164-9-75.

Bailly-Bechet, M., Vergassola, M., and Rocha, E. (2007). Causes for the intriguing presence of tRNAs in phages. *Genome Res*. 17, 1486–1495. doi:10.1101/gr.6649807.

Bessman, M. J., Frick, D. N., and O'Handley, S. F. (1996). The MutT proteins or "Nudix" hydrolases, a family of versatile, widely distributed "housecleaning" enzymes. *J*. *Biol*. *Chem*. 271, 25059–25062. doi:10.1074/jbc.271.41.25059.

Bush, A. L., and Pueppke, S. G. (1991). Characterization of an unusual new *Agrobacterium tumefaciens* strain from *Chrysanthemum morifolium* ram. *Appl*. *Environ*. *Microbiol*. 57, 2468–2472.

Buttimer, C., Hendrix, H., Oliveira, H., Casey, A., Neve, H., McAuliffe, O., et al. (2017a). Things are getting hairy: Enterobacteria bacteriophage vB_PcaM_CBB. *Front*. *Microbiol*. 8, 1–16. doi:10.3389/fmicb.2017.00044.

Buttimer, C., McAuliffe, O., Ross, R. P., Hill, C., O'Mahony, J., and Coffey, A. (2017b). Bacteriophages and bacterial plant diseases. *Front*. *Microbiol*. 8. doi:10.3389/fmicb.2017.00034.

Casey, E., Fitzgerald, B., Mahony, J., Lugli, G. A., Ventura, M., and Sinderen, D. Van (2017). Genome sequence of *Serratia marcescens* phage BF. *Genome Announc*. 5, e00211-17. doi:10.1128/genomeA.00211-17.

Clokie, M. R. J., Millard, A. D., Letarov, A. V., and Heaphy, S. (2011). Phages in nature. *Bacteriophage* 1, 31–45. doi:10.4161/bact.1.1.14942.

Costechareyre, D., Rhouma, A., Lavire, C., Portier, P., Chapulliot, D., Bertolla, F., et al. (2010). Rapid and efficient identification of *Agrobacterium* species by *recA* allele analysis. *Microb*. *Ecol*. 60, 862–872. doi:10.1007/s00248-010-9685-7.

Dandekar, M. A. E. and A. M. (2003). *Agrobacterium tumefaciens* as an agent of disease. *Trends Plant Sci*. doi:10.1016/S1360-1385(03)00162-6.

Darling, A. C. E., Mau, B., Blattner, F. R., and Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res*. 14, 1394–1403. doi:10.1101/gr.2289704.

Donelli, G., Dore, E., Frontali, C., and Grandolfo, M. E. (1975). Structure and physico-chemical properties of bacteriophage G. III. A homogeneous DNA of molecular weight 5 × 108. *J. Mol. Biol*. 94. doi:10.1016/0022-2836(75)90321-6.

Dougan, D. A., Reid, B. G., Horwich, A. L., and Bukau, B. (2002). ClpS, a substrate modulator of the ClpAP machine. *Mol. Cell* 9, 673–683. doi:10.1016/S1097-2765(02)00485-9.

Dwivedi, B., Xue, B., Lundin, D., Edwards, R. A., and Breitbart, M. (2013). A bioinformatic analysis of ribonucleotide reductase genes in phage genomes and metagenomes. 1–17.

Foster, J. W., Park, Y. K., Penfound, T., Fenger, T., and Spector, M. P. (1990). Regulation of NAD metabolism in *Salmonella typhimurium*: Molecular sequence analysis of the bifunctional *nadR* regulator and the *nadA-pnuC* operon. *J. Bacteriol*. 172, 4187–4196. doi:10.1128/jb.172.8.4187-4196.1990.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol*. 59, 307–321. doi:10.1093/sysbio/syq010.

Hatfull, G. F. (2015). Dark matter of the biosphere: the amazing world of bacteriophage

diversity. *J. Virol.* 89, 8107–8110. doi:10.1128/JVI.01340-15.

Hendrix, R. W. (2009). Jumbo bacteriophages. *Curr. Top. Microbiol. Immunol.* 328, 229–240. doi:10.1007/978-3-540-68618-7-7.

Hogg, J. C., and Lehane, M. J. (1999). Identification of bacterial species associated with the sheep scab mite (*Psoroptes ovis*) by using amplified genes coding for 16S rRNA. *Appl. Environ. Microbiol.* 65, 4227–4229.

Howell, M., Daniel, J. J., and Brown, P. J. B. (2017). Live cell fluorescence microscopy to observe essential processes during microbial cell growth. *J. Vis. Exp.*, e56497. doi:doi:10.3791/56497.

Kato, J., Suzuki, H., and Ikeda, H. (1992). Purification and characterization of DNA topoisomerase IV in *Escherichia coli*. *J. Biol. Chem.* 267, 25676–25684.

Keane, B. P. J., Kerr, A., and Newt, P. B. (1970). Crown gall of stone fruit identification and nomenclature of *Agrobacterium* isolates. *Aust. J. biol. Sci.* 23, 585–596. doi:10.1071/BI9700585.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi:10.1093/bioinformatics/bts199.

Koskella, B., and Brockhurst, M. A. (2014). Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol. Rev.* 38, 916–931. doi:10.1111/1574-6976.12072.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to

complete genomes. *J Mol Biol* 305, 567–580. doi:10.1006/jmbi.2000.4315.

Kropinski, A. M., Prangishvili, D., and Lavigne, R. (2009). Position paper: The creation of a rational scheme for the nomenclature of viruses of Bacteria and Archaea. *Environ. Microbiol.* 11, 2775–2777. doi:10.1111/j.1462-2920.2009.01970.x.

Kropinski, A. M., Van Den Bossche, A., Lavigne, R., Noben, J. P., Babinger, P., and Schmitt, R. (2012). Genome and proteome analysis of 7-7-1, a flagellotropic phage infecting *Agrobacterium* sp H13-3. *Virol. J.* 9. doi:10.1186/1743-422X-9-102.

Kurnasov, O. V., Polanuyer, B. M., Ananta, S., Sloutsky, R., Tam, A., Gerdes, S. Y., et al. (2002). Ribosylnicotinamide kinase domain of NadR protein: Identification and implications in NAD biosynthesis. *J. Bacteriol.* 184, 6906–6917. doi:10.1128/JB.184.24.6906-6917.2002.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., Mcgettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi:10.1093/bioinformatics/btm404.

Lee, J. Y., Li, Z., and Miller, E. S. (2017). *Vibrio* phage KVP40 Encodes a Functional NAD^+ Salvage Pathway. *J. Bacteriol.* 199, JB.00855-16. doi:10.1128/JB.00855-16.

Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44, W242–W245. doi:10.1093/nar/gkw290.

Lowe, T. M., and Chan, P. P. (2016). tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res* 44, W54-7. doi:10.1093/nar/gkw413.

Luo, Z. Q., Clemente, T. E., and Farrand, S. K. (2001). Construction of a derivative of

Agrobacterium tumefaciens C58 that does not mutate to tetracycline resistance. *Mol. Plant. Microbe. Interact.* 14, 98–103. doi:10.1094/MPMI.2001.14.1.98.

Miller, E. S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T., and Ruger, W. (2003). Bacteriophage T4 genome. *Microbiol. Mol. Biol. Rev.* 67, 86–156. doi:10.1128/mmbr.67.1.86-156.2003.

Mogk, A., Tomoyasu, T., Goloubinoff, P., Rüdiger, S., Röder, D., Langen, H., et al. (1999). Identification of thermolabile *Escherichia coli* proteins: prevention and reversion of aggregation by DnaK and ClpB. *EMBO J.* 18, 6934–6949. doi:10.1093/emboj/18.24.6934.

Mougel, C., Thioulouse, J., Perrière, G., and Nesme, X. (2002). A mathematical method for determining genome divergence and species delineation using AFLP. *Int. J. Syst. Evol. Microbiol.* 52, 573–586. doi:10.1099/00207713-52-2-573.

Nierman, W. C., Feldblyum, T. V, Laub, M. T., Paulsen, I. T., Nelson, K. E., Eisen, J. A., et al. (2001). Complete genome sequence of *Caulobacter crescentus*. *Proc Natl Acad Sci U S A* 98, 4136–4141. doi:10.1073/pnas.061029298\r061029298 [pii].

Panagopoulos, C. G., and Psallidas, P. G. (1973). Characteristics of Greek isolates of *Agrobacterium tumefaciens* (E. F. Smith & Townsend) Conn. *J. Appl. Bacteriol.* 36, 233–240. doi:10.1111/j.1365-2672.1973.tb04096.x.

Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8, 785–786. doi:10.1038/nmeth.1701.

Petrov, V. M., Ratnayaka, S., Nolan, J. M., Miller, E. S., and Karam, J. D. (2010). Genomes of the T4-related bacteriophages as windows on microbial genome

evolution. *Virol*. *J*. 7, 1–19. doi:10.1186/1743-422X-7-292.

Poindexter, J. S. (1964). Biological properties and classification of the *Caulobacter* group. *Bacteriol*. *Rev*. 28, 231–295.

Portier, P., Fischer-Le Saux, M., Mougel, C., Lerondelle, C., Chapulliot, D., Thioulouse, J., et al. (2006). Identification of genomic species in *Agrobacterium* biovar 1 by AFLP genomic markers. *Appl*. *Environ*. *Microbiol*. 72, 7123–7131. doi:10.1128/AEM.00018-06.

Rombouts, S., Volckaert, A., Venneman, S., Declercq, B., Vandenheuvel, D., Allonsius, C. N., et al. (2016). Characterization of novel bacteriophages for biocontrol of bacterial blight in leek caused by *Pseudomonas syringae* pv. *porri*. *Front*. *Microbiol*. 7, 1–15. doi:10.3389/fmicb.2016.00279.

Santamaría, R. I., Bustos, P., Sepúlveda-Robles, O., Lozano, L., Rodríguez, C., Fernández, J. L., et al. (2014). Narrow-host-range bacteriophages that infect *Rhizobium etli* associate with distinct genomic types. *Appl*. *Environ*. *Microbiol*. 80, 446–454. doi:10.1128/AEM.02256-13.

Santos, S. B., Costa, A. R., Carvalho, C., Nóbrega, F. L., and Azeredo, J. (2018). Exploiting bacteriophage proteomes: the hidden biotechnological potential. *Trends Biotechnol*., 1–19. doi:10.1016/j.tibtech.2018.04.006.

Sawada, H., Ieki, H., Oyaizu, H., and Matsumoto, S. (1993). Proposal for rejection of *Agrobacterium tumefaciens* and revised descriptions for the genus *Agrobacterium* and for *Agrobacterium radiobacter* and *Agrobacterium rhizogenes*. *Int*. *J*. *Syst*. *Bacteriol*. 43, 694–702. doi:10.1099/00207713-43-4-694.

Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ : 25

years of image analysis HISTORICAL commentary NIH Image to ImageJ : 25 years of image analysis. *Nat*. *Methods* 9, 671–675. doi:10.1038/nmeth.2089.

Sengupta, R. (2014). Thioredoxin and glutaredoxin-mediated redox regulation of ribonucleotide reductase. *World J*. *Biol*. *Chem*. 5, 68. doi:10.4331/wjbc.v5.i1.68.

Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996). Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Anal*. *Chem*. 68, 850–858. doi:10.1021/ac950914h.

Simoliunas, E., Kaliniene, L., Truncaite, L., Zajanckauskaite, A., Staniulis, J., Kaupinis, A., et al. (2013). *Klebsiella* phage vB_KleM-RaK2 - a giant singleton virus of the family Myoviridae. *PLoS One* 8, e60717. doi:10.1371/journal.pone.0060717.

Slater, S. C., Goldman, B. S., Goodner, B., Setubal, J. C., Farrand, S. K., Nester, E. W., et al. (2009). Genome sequences of three *Agrobacterium* biovars help elucidate the evolution of multichromosome genomes in bacteria. *J*. *Bacteriol*. 191, 2501–2511. doi:10.1128/JB.01779-08.

Srinivasiah, S., Bhavsar, J., Thapar, K., Liles, M., Schoenfeld, T., and Wommack, K. E. (2008). Phages across the biosphere: contrasts of viruses in soil and aquatic environments. *Res*. *Microbiol*. 159, 349–357. doi:10.1016/j.resmic.2008.04.010.

Sullivan, M. B., Huang, K. H., Ignacio-Espinoza, J. C., Berlin, A. M., Kelly, L., Weigele, P. R., et al. (2010). Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Env*. *Microbiol* 12, 3035–3056. doi:10.1111/j.1462-2920.2010.02280.x.

Turner, S., Pryer, K. M., Miao, V. P. W., and Palmer, J. D. (1999). Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA

sequence analysis. *J. Eukaryot. Microbiol*. 46, 327–338. doi:10.1111/j.1550-7408.1999.tb04612.x.

Watson, B., Currier, T. C., Gordon, M. P., Chilton, M.-D., and Nester, E. W. (1975). Plasmid required for virulance of *Agrobacterium tumefaciens*. *J. Bacteriol* 123, 255–264.

Węgrzyn, A., Czyż, A., Gabig, M., and Węgrzyn, G. (2000). ClpP/ClpX-mediated degradation of the bacteriophage λ O protein and regulation of λ phage and λ plasmid replication. *Arch. Microbiol*. 174, 89–96. doi:10.1007/s002030000177.

Weidner, S., Baumgarth, B., Göttfert, M., Jaenicke, S., Pühler, A., Schneiker-Bekel, S., et al. (2013). Genome sequence of *Sinorhizobium meliloti* Rm41. *Genome Announc*. 1, e00013-12. doi:10.1128/genomeA.00013-12.

Yamamoto, K. R., Alberts, B. M., Benzinger, R., Lawhorne, L., and Treiber, G. (1970). Rapid bacteriophage sedimentation in the presence of polyethylene glycol and its application to large-scale virus purification. *Virology* 40, 734–744.

Yin, Y., and Fischer, D. (2008). Identification and investigation of ORFans in the viral world. *BMC Genomics* 9, 24. doi:10.1186/1471-2164-9-24.

Yoshikawa, G., Askora, A., Blanc-Mathieu, R., Kawasaki, T., Li, Y., Nakano, M., et al. (2018). *Xanthomonas citri* jumbo phage XacN1 exhibits a wide host range and high complement of tRNA genes. *Sci. Rep*. 8, 1–10. doi:10.1038/s41598-018-22239-3.

Young, J. M., Pennycook, S. R., and Watson, D. R. W. (2006). Proposal that *Agrobacterium radiobacter* has priority over *Agrobacterium tumefaciens*. Request for an Opinion. *Int. J. Syst. Evol. Microbiol*. 56, 491–493. doi:10.1099/ijs.0.64030-0.

Young, R. (2014). Phage lysis: three steps, three choices, one outcome. *J. Microbiol*. 52,

243–258. doi:10.1007/s12275-014-4087-z.

Yuan, Y., and Gao, M. (2017). Jumbo bacteriophages: an overview. *Front*. *Microbiol*. 8. doi:10.3389/fmicb.2017.00403.

Zylicz, M., Liberek, K., Wawrzynow, A., and Georgopoulos, C. (1998). Formation of the preprimosome protects lambda O from RNA transcription-dependent proteolysis by ClpP/ClpX. *Proc*. *Natl*. *Acad*. *Sci*. *U*. *S*. *A*. 95, 15259–63. doi:10.1073/pnas.95.26.15259.

Tables

Table 3-1. Bacterial strains used in this study

Strain or plasmid	**Relevant characteristics**	**Growth medium**	**Reference or source**
A. tumefaciens strains			
C58	Nopaline type strain; pTiC58; pAtC58	LB	(Watson et al., 1975)
EHA105	C58 derived, succinamopine strain, T-DNA deletion derivative of pTiBo542	LB	MU plant transformation core facility
GV3101	C58 derived, nopaline strain	LB	MU plant transformation core facility
NTL4	C58 derived, nopaline-agrocinopine strain, Δ*tetRA*	LB	(Luo et al., 2001)
AGL-1	C58 derived, succinamopine strain, T-DNA deletion derivative of pTiBo542 ΔrecA	LB	MU plant transformation core facility
LBA4404	Ach5 derived, octopine strain, T-DNA deletion derivative of pTiAch5	YM	MU plant transformation core facility

Chry5	Succinamopine strain, pTiChry5	LB	(Bush and Pueppke, 1991)
LMG215	*Agrobacterium* biovar 1, genomospecies 4, isolated from hops in 1928	LB	Chang lab at Oregon State University
LMG232	*Agrobacterium* biovar 1, genomospecies 1, isolated from beet in 1963	LB	Chang lab at Oregon State University
A74a	*Agrobacterium* biovar 1, genomospecies 8, isolated from Pennsylvania lavender in 2003	LB	Chang lab at Oregon State University
06-777-2L	*Agrobacterium* biovar 1, genomospecies 7, isolated from Marguerite Daisy in 2006	LB	Chang lab at Oregon State University
Other bacterial strains			
A. vitis S4	Vitopine strain, pTiS4, pSymA, pSymB	Potato dextrose	(Slater et al., 2009)
Rhizobium rhizogenes D108/85	*Agrobacterium* biovar 2, isolated from Michigan blueberry 1985	MGY	Chang lab at Oregon State University

Caulobacter crescentus CB15	Alphaproteobacterium	PYE	(Nierman et al., 2001)
Sinorhizobium meliloti 1021	Rhizopine strain, pSymA, pSymB, pRme41a	LB	(Weidner et al., 2013)
Escherichia coli DH5α	Gammaproteobacterium	LB	Life Technologies

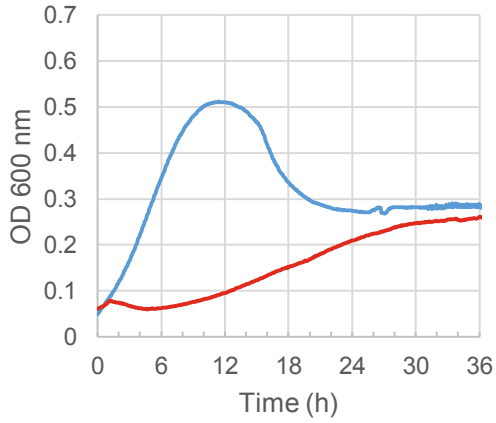
Table 3-2. Summary of key genomic features of Atu_ph07

Genome length (bp)	**G+C content (%)**	**Number of ORFs**	**Coding density (%)**	**Number of hypothetical proteins**	**Number of ORFs with predicted functions**	**Number of ORFans**	**Number of tRNAs**
490,380	37.1	714	83.6	214	110	390	33

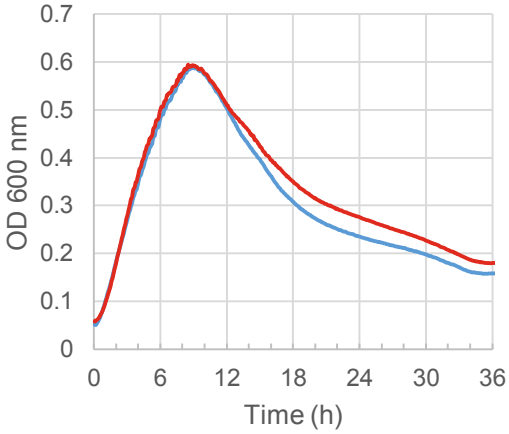
SUPPLEMENTARY FIGURES

Supplementary Figure 3-S1. Representative growth curves of *A. tumefaciens* strains C58 (A) and LMG215 (B) in the absence (blue) or presence (red) of phage Atu_ph07 at MOI 10. Experiment was performed in duplicate and averaged.

A



B



SUPPLEMENTARY TABLES

Supplementary Table 3-S1. Atu_ph07 genes categorized by predicted function.

CDS #	RAST assigned function	Updated assigned function	Length (bp)	Category
1	Translation initiation factor 3 CDS		504	Translation
2	ATP-dependent Clp protease ATP-binding subunit ClpX CDS		1,260	Posttranslational modification
4	Phage DNA topoisomerase large subunit (EC 5.99.1.3) #T4-like gp60 #T4 GC1464 CDS	DNA gyrase subunit B	1,815	DNA replication
5	Topoisomerase IV subunit A (EC 5.99.1.-) CDS		1,362	DNA replication
18	tRNA nucleotidyltransferase (EC 2.7.7.21) (EC 2.7.7.25) CDS		1,266	Translation
20	internalin, putative CDS		3,507	Other
31	Hypothetical protein CDS	Phage tail fiber-like protein	1,290	Structural
35	Hypothetical protein CDS	Exonuclease type II	681	DNA replication
42	conserved phage mega protein		921	Structural
49	multi-sensor signal transduction histidine kinase CDS		1,611	Other
70	DNA primase/helicase, phage-associated CDS		1,323	DNA replication
71	Hypothetical protein CDS	DNA primase subunit	1,023	DNA replication
80	Hypothetical protein CDS	RNA-DNA + DNA-DNA helicase	1,383	DNA replication
86	Hypothetical protein CDS	2'-5' RNA ligase	645	DNA replication
tRNA	tRNA-Gln-TTG		125	tRNA
97	putative type II DNA modification enzyme (methyltransferase) CDS		768	DNA modification
106	Hypothetical protein CDS	RecA-like protein	1,107	DNA recombination and repair
107	Hypothetical protein CDS	ssDNA-binding protein	1002	DNA replication
109	Gp17 terminase DNA packaging enzyme large subunit CDS		1,809	Structural
110	Phage terminase, large subunit CDS		372	Structural
112	autotransporter CDS		2,235	Other
121	Hypothetical protein CDS	Glutaredoxin	237	Nucleotide metabolism
122	Ribonucleotide reductase of class Ia (aerobic), alpha subunit (EC 1.17.4.1) CDS		1,641	Nucleotide metabolism
123	Ribonucleotide reductase of class Ia (aerobic), beta subunit (EC 1.17.4.1) CDS		999	Nucleotide metabolism
125	Phage head completion protein CDS		483	Structural

128	Prophage Clp protease-like protein CDS		1,656	Posttranslational modification
132	Phage tail sheath monomer CDS		3,261	Structural
135	NAD synthetase (EC 6.3.1.5) / Glutamine amidotransferase chain of NAD synthetase CDS		1,671	Nucleotide metabolism
140	Hypothetical protein CDS	Tail sheath stabilizer and completion protein	1,374	Structural
141	Phage lysin (EC 3.2.1.17) # Phage lysozyme or muramidase (EC 3.2.1.17) CDS		450	Lysis
143	Hypothetical protein CDS	Head completion adaptor; neck	771	Structural
155	Hypothetical protein CDS	ATP-dependent Clp protease ATP-binding subunit	285	Posttranslational modification
180	T4-like phage structural protein CDS		13,350	Structural
181	Hypothetical protein CDS	Baseplate wedge	3,597	Structural
182	Hypothetical protein CDS	Baseplate wedge	390	Structural
189	ulcer associated adenine specific DNA methyltransferase CDS		978	DNA modification
191	Phage protein CDS		2,289	Hypothetical
195	T4-like phage protein, T4 GC1542 CDS	Nucleotidyltransferase	993	Nucleotide metabolism
198	Phage portal vertex of the head #T4-like phage Gp20 CDS		1,755	Structural
200	Topoisomerase IV subunit A (EC 5.99.1.-) CDS		1,170	DNA replication
204	Phage prohead core protein CDS - Gp21		633	Structural
208	Phage protein CDS		819	Hypothetical
tRNA	tRNA-Leu-CAA		78	tRNA
tRNA	tRNA-Met-CAT		76	tRNA
213	MJ0042 family finger-like protein CDS		876	Other
214	DNA double-strand break repair Rad50 ATPase CDS		1,245	DNA recombination and repair
215	Phage major capsid protein of Caudovirales (T4-like gp23) CDS		1,230	Structural
219	Conserved protein CDS		651	Hypothetical
220	Adenine-specific methyltransferase (EC 2.1.1.72) CDS		783	DNA modification
221	Thioredoxin, phage-associated CDS		1,047	Nucleotide metabolism
227	tRNAHis-5'-guanylyltransferase CDS		789	Translation
235	Adenine-specific methyltransferase (EC 2.1.1.72) CDS		810	DNA modification
238	ADP-ribose pyrophosphatase (EC		498	Nucleotide

	3.6.1.13) CDS			metabolism
240	EF hand domain/PKD domain protein CDS		1,017	Hypothetical
242	Nicotinate phosphoribosyltransferase (EC 2.4.2.11) CDS		1,299	Nucleotide metabolism
246	Thymidylate synthase thyX (EC 2.1.1.-) CDS		951	Nucleotide metabolism
252	HNH endonuclease family protein CDS		558	Nucleotide metabolism
253	protein serine-threonine phosphatase CDS		1,257	Transcription
261	RNA ligase, phage-associated #T4-like RnlA #T4 GC1653 CDS		1,194	Transcription
263	Sll7028 protein CDS	Zn dependent hydrolase	1,290	Nucleotide metabolism
266	Glutaredoxin CDS		258	Nucleotide metabolism
271	DNA polymerase I (EC 2.7.7.7) CDS		1,062	DNA replication
276	DNA polymerase (EC 2.7.7.7), phage-associated #T4-like phage gp43 #T4 GC0178 CDS		1,254	DNA replication
277	DNA polymerase (EC 2.7.7.7), phage-associated CDS		1,698	DNA replication
279	Phage protein CDS		249	Hypothetical
282	DNA ligase (EC 6.5.1.2) CDS		1,995	DNA replication
288	Phage tail fiber protein CDS		1,833	Structural
292	ATP-dependent Clp protease ATP-binding subunit ClpA CDS		2,298	Posttranslational modification
293	Chaperone protein DnaJ CDS		807	Posttranslational modification
tRNA	tRNA Ser-GCT		93	tRNA
303	Probable NUDIX hydrolase CDS		393	Nucleotide metabolism
309	Phosphate starvation-inducible protein PhoH, predicted ATPase CDS		765	Other
310	Hypothetical protein CDS	Phosphoglycolate phosphatase	621	Other
312	Replication factor C small subunit CDS		948	DNA replication
316	Hypothetical protein CDS	DNA polymerase III alpha subunit	534	DNA replication
319	Cyanophage-encoded pyrophosphatase, MazG #T4 GC0184 CDS		408	Other
320	Streptococcal hemagglutinin protein CDS		1,182	Other
327	Phage tail fiber protein CDS		2,130	Structural
329	Extracellular serine proteinase precursor (EC 3.4.21.-) CDS		3,756	Posttranslational modification
tRNA	tRNA-Pseudo-ACC		74	tRNA

297	Conserved protein CDS		297	Hypothetical
393	Phage protein CDS		312	Hypothetical
394	no significant homology		1,455	Hypothetical
399	Adenine-specific methyltransferase (EC 2.1.1.72) CDS		759	DNA modification
tRNA	tRNA-Gly-GCC		75	tRNA
tRNA	tRNA-Met-CAT		78	tRNA
tRNA	tRNA-Asp-GTC		78	tRNA
tRNA	tRNA-Arg-TCT		78	tRNA
tRNA	tRNA-His-GTG		75	tRNA
tRNA	tRNA-Ser-TGA		92	tRNA
tRNA	tRNA-Tyr-GTA		119	tRNA
tRNA	tRNA-Leu-TAG		90	tRNA
tRNA	tRNA-Gly-TCC		72	tRNA
tRNA	tRNA-Gln-TTG		75	tRNA
tRNA	tRNA-Pro-TGG		77	tRNA
tRNA	tRNA-Ser-TGA		109	tRNA
460	Hypothetical protein CDS	HNH endonuclease	765	Nucleotide metabolism
465	Phage-associated homing endonuclease CDS		609	Nucleotide metabolism
tRNA	tRNA-Phe-GAA		76	tRNA
tRNA	tRNA-Gln-CTG		74	tRNA
tRNA	tRNA-Pro-TGG		78	tRNA
tRNA	tRNA-Val-TAC		75	tRNA
tRNA	tRNA-Glu-TTC		77	tRNA
tRNA	tRNA-Met-CAT		72	tRNA
tRNA	tRNA-Trp-CCA		76	tRNA
tRNA	tRNA-Tyr-GTA		82	tRNA
tRNA	tRNA-Ile-GAT		80	tRNA
tRNA	tRNA-Pro-TGG		78	tRNA

tRNA	tRNA-Phe-GAA		76	tRNA
tRNA	tRNA-Arg-ACG		78	tRNA
tRNA	tRNA-Ala-TGC		72	tRNA
472	Phage tail fiber protein CDS		2,361	Structural
481	Guanosine-3',5'-bis(Diphosphate) 3'-pyrophosphohydrolase (EC 3.1.7.2) CDS		456	Nucleotide metabolism
492	ClpB protein CDS		1,986	Posttranslational modification
tRNA	tRNA-Lys-CCT		73	tRNA
554	Conserved protein CDS		384	Hypothetical
557	Hypothetical protein CDS		306	Hypothetical
573	Autotransporter adhesin CDS	Hypothetical protein	4,152	Hypothetical
575	High-affnity carbon uptake protein Hat/HatR CDS		4,317	Hypothetical
586	Ribosyl nicotinamide transporter, PnuC-like CDS		678	Nucleotide metabolism
587	Nicotinamide-nucleotide adenylyltransferase, NadR family (EC 2.7.7.1) / Ribosylnicotinamide kinase (EC 2.7.1.22) CDS		1,026	Nucleotide metabolism
tRNA	tRNA-Sup-TCA		72	tRNA
622	Hypothetical protein CDS	HNH endonuclease	789	Nucleotide metabolism
624	Putative phage-related protein precursor CDS	D-alanyl-D-alanine carboxypeptidase	399	Other
638	FIG006762: Phosphoglycerate mutase family CDS	RNAse HI	369	DNA replication
tRNA	tRNA-Cys-GCA		76	tRNA
670	ADP-ribose 1""-phosphate phosphatase related protein CDS		438	Posttranslational modification
671	Polymerase epsilon subunit CDS		792	DNA replication
673	Uncharacterized protein COG3236 CDS		537	Hypothetical
676	ATP-dependent Clp protease proteolytic subunit (EC 3.4.21.92) CDS		615	Posttranslational modification
677	Phage protein CDS		1,125	Hypothetical
680	Peptidyl-tRNA hydrolase, archaeal type (EC 3.1.1.29) CDS		468	Translation
686	Hypothetical protein CDS	Baseplate hub subunit	843	Structural
691	Phage recombination-related endonuclease Gp46 CDS		1,776	DNA recombination and repair
693	Phage recombination related endonuclease CDS		1,023	DNA recombination and repair

694	Hypothetical protein CDS	RNA polymerase sigma factor	834	Transcription
697	Hypothetical protein CDS	Baseplate wedge	315	Structural
699	Hypothetical protein CDS	Lysozyme	2,598	Lysis
701	PE-PGRS virulence associated protein CDS	Tail tape measure protein	618	Structural
705	Threonyl-tRNA synthetase (EC 6.1.1.3) CDS		708	Translation
708	ATP-dependent Clp protease adaptor protein ClpS CDS		303	Posttranslational modification
709	Thymidine kinase (EC 2.7.1.21) CDS		549	Nucleotide metabolism
710	16 kDa heat shock protein A CDS	Molecular chaperone IbpA, HSP20 family	402	Posttranslational modification

Supplementary Table 3-S2. Atu_ph07 gene products compared with 12 related phages.

gp	XacNI	SCTP-2	BF	CBB	GAP32	121Q	K64-1	E11	Rak2	slurp01	S-SSM7	T4
DNA replication, repair, recombination												
4	1.00E-65	1.00E-62	1.00E-179	4.00E-176	5.00E-179	6.00E-177	2.00E-177	2.00E-167	4.00E-177	2.00E-178		3.00E-41
5	1.00E-34	4.00E-53	3.00E-101	4.00E-107	4.00E-107	5.00E-107	2.00E-107	3.00E-109	4.00E-107	9.00E-107		5.00E-36
35		3.00E-37	1.00E-65	8.00E-66	2.00E-65	5.00E-58		2.00E-63	3.00E-63		6.00E-31	
70	6.00E-64	4.00E-90	2.00E-79	1.00E-76	2.00E-77	4.00E-75	8.00E-77	7.00E-96	8.00E-77	2.00E-75	4.00E-31	4.00E-29
71	2.00E-24	7.00E-31	9.00E-40	1.00E-36	4.00E-35	8.00E-38	1.00E-42	7.00E-52	3.00E-42	5.00E-38	1.00E-04	
80	2.00E-91	9.00E-138	9.00E-146	1.00E-147	3.00E-146	2.00E-145	5.00E-141	2.00E-151	7.00E-141		6.00E-51	1.00E-26
86	1.00E-59	2.00E-52										
97					2.00E-24							
106	2.00E-60	6.00E-59	2.00E-141	8.00E-154	5.00E-153	6.00E-143		2.00E-155	3.00E-140		2.00E-31	7.00E-24
107	2.00E-28	4.00E-81	7.00E-74	2.00E-70	9.00E-74	1.00E-66		8.00E-71	6.00E-73	8.00E-67		
189												
200		5.00E-15	5.00E-06	8.00E-06	4.00E-05	5.00E-05		4.00E-06	5.00E-07			
214	2.00E-18	7.00E-55	4.00E-48	4.00E-53	6.00E-51	3.00E-49		2.00E-62	3.00E-50			
220			1.00E-09	5.00E-09	6.00E-09							
235			4.00E-09	4.00E-09	3.00E-09							
271	1.00E-39	6.00E-80	5.00E-86	2.00E-85	1.00E-83	2.00E-87		2.00E-75	1.00E-87	1.00E-87		
276	5.00E-36	6.00E-67	2.00E-116	3.00E-115	6.00E-116	7.00E-116	8.00E-118	3.00E-109	5.00E-117	7.00E-116	4.00E-15	2.00E-11
277	5.00E-68	2.00E-116	2.00E-133	5.00E-137	9.00E-139	1.00E-135	8.00E-133	2.00E-137	6.00E-133	2.00E-135	6.00E-33	2.00E-26
282	1.00E-108											
312	4.00E-70	5.00E-48	4.00E-49	4.00E-51	3.00E-52	2.00E-47		3.00E-61	4.00E-46		5.00E-33	3.00E-18
316	6.00E-25	1.00E-37	1.00E-26	1.00E-27	2.00E-27	3.00E-24		5.00E-22	5.00E-26			
399			1.00E-09	1.00E-11	3.00E-13							
638						2.00E-05				4.00E-06		
671			9.00E-08	5.00E-08	8.00E-09	4.00E-09		0.001	3.00E-07			
691	7.00E-48	1.00E-120	6.00E-124	8.00E-124	1.00E-125	3.00E-113		3.00E-119	1.00E-109	6.00E-113	5.00E-48	7.00E-50
693	2.00E-49	1.00E-100	1.00E-91	2.00E-92	4.00E-91	2.00E-93		2.00E-101	2.00E-96	1.00E-93	6.00E-27	8.00E-15
Structural												

31											2.00E-04	
42												
109	2.00E-83	1.00E-91	4.00E-134	1.00E-130	3.00E-131	3.00E-130	2.00E-122	7.00E-133	4.00E-123		9.00E-49	2.00E-42
110	7.00E-19	1.00E-14	3.00E-12	4.00E-11	3.00E-12	8.00E-12	6.00E-14	4.00E-14	2.00E-14		1.00E-18	4.00E-08
125	5.00E-27	4.00E-51	2.00E-36	1.00E-37	5.00E-37	3.00E-35		3.00E-44	2.00E-36		7.00E-33	6.00E-20
132	3.00E-72	1.00E-116	6.00E-116	6.00E-117	3.00E-114	2.00E-103	5.00E-103	8.00E-142	5.00E-103	9.00E-05	3.00E-30	1.00E-28
140	1.00E-22	3.00E-59	1.00E-44	3.00E-43	2.00E-47	8.00E-42	3.00E-44	5.00E-54	5.00E-45		1.00E-05	1.00E-04
143	1.00E-27	8.00E-51	4.00E-55	3.00E-53	1.00E-54	3.00E-53		2.00E-68	5.00E-53	3.00E-53	3.00E-12	2.00E-07
180	1.00E-19	2.00E-85	2.00E-132	6.00E-130	1.00E-121	1.00E-127	5.00E-136	4.00E-162	4.00E-136	6.00E-128	5.00E-05	
181	3.00E-112	0	0	1.00E-169	2.00E-173	0	0	0	0	0	2.00E-09	
182	7.00E-17	1.00E-11	7.00E-23	3.00E-25	4.00E-25	3.00E-25		1.00E-21	2.00E-21			2.00E-04
198	3.00E-100	2.00E-164	7.00E-161	3.00E-166	3.00E-165	6.00E-159	4.00E-158	1.00E-167	1.00E-156	5.00E-162	7.00E-31	5.00E-20
204	1.00E-44	2.00E-54	1.00E-43	9.00E-44	2.00E-44	5.00E-46		1.00E-54	3.00E-42		9.00E-34	5.00E-14
215	6.00E-80	1.00E-154	2.00E-164		4.00E-163	1.00E-158		3.00E-163	5.00E-157		3.00E-35	1.00E-38
288		3.00E-06	0.001								3.00E-05	
327						2.00E-04				2.00E-04		
472			9.00E-06								1.00E-04	
686	2.00E-23	1.00E-30	5.00E-38	5.00E-41	3.00E-41	7.00E-38		4.00E-40	2.00E-32	3.00E-38		
697	3.00E-05	2.00E-22	3.00E-20	3.00E-21	8.00E-21	5.00E-20		5.00E-19	1.00E-22	2.00E-20		
701	2.00E-04		5.00E-04								3.00E-04	
Lysis												
141						5.00E-04	0.001		0.001	4.00E-04		
699	3.00E-21	5.00E-65	1.00E-39	1.00E-39	9.00E-41	4.00E-40	5.00E-40	5.00E-44	5.00E-40	4.00E-40	7.00E-19	
Translation												
1			5.00E-30	4.00E-24	1.00E-25	1.00E-23		2.00E-21	1.00E-24			
2												
18	2.00E-15		3.00E-13	1.00E-14	7.00E-15	4.00E-12		2.00E-15	2.00E-13	6.00E-12		
128												
155		3.00E-04										
227	5.00E-15	6.00E-21	3.00E-33	9.00E-36	1.00E-33							
292		5.00E-170	0	0	0	0	0	0	0			
293												
329			5.00E-05	3.00E-04	1.00E-05	3.00E-05				1.00E-05		

492		6.00E-52	6.00E-51	1.00E-50	2.00E-53	1.00E-51	7.00E-56	8.00E-50	3.00E-55			
670	3.00E-19		3.00E-12	2.00E-09	2.00E-09	4.00E-13		3.00E-08	1.00E-12	2.00E-13		1.00E-08
676	5.00E-08	5.00E-61	1.00E-38	1.00E-37	2.00E-39	3.00E-36		6.00E-33	3.00E-43			
680	7.00E-27	3.00E-26	5.00E-04		3.00E-04				3.00E-07			
705												
708		9.00E-10		2.00E-06	4.00E-04			6.00E-11	3.00E-07			
710		2.00E-16									2.00E-18	
Nucleotide Metabolism												
121	8.00E-10	4.00E-09	1.00E-05	1.00E-05	9.00E-06			5.00E-06			3.00E-10	8.00E-08
122	5.00E-06	3.00E-96	1.00E-13	4.00E-14	2.00E-14	4.00E-14	7.00E-10		7.00E-10	5.00E-14	1.00E-12	1.00E-15
123		1.00E-89	8.00E-09	1.00E-10	4.00E-13	3.00E-09			1.00E-05	3.00E-09		
135											3.00E-11	
195			2.00E-09	1.00E-08	3.00E-07	2.00E-10	4.00E-10	7.00E-12	3.00E-10	2.00E-10		4.00E-46
221			1.00E-35	5.00E-34	3.00E-33	4.00E-32		3.00E-43	5.00E-33	1.00E-32	0.001	4.00E-46
238	0.001		7.00E-04	4.00E-04	3.00E-04			3.00E-04				
242												
246											1.00E-13	
252		3.00E-14	0.001									
253				0.001								
261		7.00E-18	9.00E-16	7.00E-19	2.00E-23	8.00E-17	5.00E-17	1.00E-24	1.00E-17	5.00E-17		5.00E-24
263		1.00E-74	2.00E-61	3.00E-69	4.00E-67	2.00E-65		5.00E-79	1.00E-65	2.00E-66		
266	5.00E-06	1.00E-08						9.00E-05			4.00E-04	
303	3.00E-07	6.00E-17	1.00E-06	4.00E-06	4.00E-06			1.00E-05	2.00E-04			
460		2.00E-19			7.00E-06							
465		2.00E-25	0.001		2.00E-29			5.00E-09				
586			0.001	4.00E-05	1.00E-06			1.00E-11				
587			5.00E-08	4.00E-11	3.00E-10			4.00E-16				
622			3.00E-04	2.00E-04	3.00E-04							
694	2.00E-17	2.00E-47	2.00E-58	2.00E-58	2.00E-57	1.00E-52		3.00E-46	1.00E-49	1.00E-52	5.00E-12	3.00E-06
709			9.00E-04	3.00E-05	1.00E-04							6.00E-08
Total	**84**	**141**	**121**	**117**	**120**	**101**	**31**	**102**	**100**	**63**	**40**	**33**

Supplementary Table 3-S3. T4 core proteins found in Atu_ph07. *Atu_ph07 matches with E-values above 1E-10 are considered "yes" matches and those between 1E-10 and 1E-03 are "unresolved." Matches with E-values lower than 1E-03 were not considered significant.

T4 protein	T4 protein function	Match in Atu_ph07*	Identity (%)	E-value	Query cover (%)	Atu_ph07 protein name	Atu_ph07 gp #
Phage morphogenesis							
gp4	head completion protein	yes	37	6.00E-21	92	head completion protein	125
gp5	baseplate lysozyme hub component	unresolved	26	0.001	18	hypothetical protein	699
gp13	head completion protein	yes	20	3.00E-08	93	hypothetical protein	143
gp15	tail completion protein	yes	20	8.00E-06	72	hypothetical protein	140
gp17	subunit of the terminase for DNA packaging	yes; two	28	7.00E-43	68	terminase DNA packaging enzyme large	109
"			31	2.00E-08	47	terminase large subunit	110
gp18	tail tube subunit	yes; two	32	8.00E-29	39	tail sheath monomer	132
"			37	4.00E-05	9	structural protein	180
gp20	head portal vertex protein	yes	25	4.00E-21	51	portal vertex of the head	198
gp21	prohead core protein and protease	yes	33	8.00E-15	66	prohead core protein	204
gp22	prohead core protein	unresolved	19	5.00E-03	94	DNA double-strand break repair Rad50 ATPase	214
gp23	precursor of major head protein	yes	31	8.00E-14	99	major capsid protein	215
gp25	base plate wedge subunit	yes	28	2.00E-05	66	hypothetical protein	182
gp34	proximal tail fiber protein subunit	unresolved	26	4.50E-02	7	EF hand domain/PKD domain protein	240
gp36	small distal tail fiber protein subunit	unresolved	23	2.80E-01	98	hypothetical protein	574
DNA replication, repair, and recombination							
gp43	DNA polymerase	yes; two	26	4.00E-27	28	DNA polymerase	277
"			25	4.00E-12	32	DNA polymerase	276

gp44	sliding clamp loader complex tetramer	yes	27	3.00E-19	90	replication factor C small subunit	312
gp41	helicase-primer complex hexamer	yes	25	4.00E-30	79	DNA primase/helicase	70
gp46	subunit of a recombination nuclease complex required for initiation of DNA replication	yes; two	25	6.00E-51	98	recombination-related endonuclease	691
"			25	1.00E-03	29	hypothetical protein	108
gp47	subunit of a recombination nuclease complex required for initiation of DNA replication	yes	29	8.00E-16	62	recombination-related endonuclease	693
UvsW	recombination DNA-RNA helicase, DNA-dependent ATPase	yes	23	2.00E-27	72	hypothetical protein	80
Auxillary metabolism							
nrdA	subunit of an aerobic ribonucleotide reductase complex	yes	23	2.00E-16	60	ribonucleotide reductase of class Ia (aerobic)	122
nrdB	subunit of an aerobic ribonucleotide reductase complex	unresolved	21	1.00E-03	75	ribonucleotide reductase of class Ia (aerobic)	123
Gene expression							
gp55	sigma factor for late transcription	unresolved	31	2.00E-07	50	hypothetical protein	694

Supplementary Table 3-S4. Amino acids, anticodons, and tRNAs encoded in the Atu_ph07 genome.

Amino acid	Anticodon	Number of tRNAs
Ala	TGC	1
Arg	ACG	1
	TCT	1
Asp	GTC	1
Cys	GCA	1
Gln	TTG	2
	CTG	1
Glu	TTC	1
Gly	GCC	1
	TCC	1
His	GTG	1
Ile	GAT	1
Leu	TAG	1
	CAA	1
Lys	CCT	1
Met	CAT	3
Phe	GAA	2
Pro	TGG	3
Ser	TGA	2
	GCT	1
Trp	CCA	1
Tyr	GTA	2
Val	TAC	1
Suppressor	TCA	1

Supplementary Table 3-S5. Bacteriophage Atu_ph07 structural proteins identified by ESI-MS/MS

Gp #	RAST assigned function	Updated assigned function	Band Nº (most abundant)	Protein MW (kDa)	Nº of unique peptides	Sequence coverage, %
20	*Putative internalin*		*1*	*127,12*	*1*	*1.11*
21	Hypothetical protein		5	39,09	3	10.50
31	**Hypothetical protein**	**Phage tail fiber-like protein**	**1,2,3,4,5,6 (5)**	**45,47**	**9**	**27.70**
32	Hypothetical protein		1,2,3,4,5,6,7 (7)	34,14	7	27.32
33	Hypothetical protein		1,2,3,4,5,6,7,8,9,10,11,12,13 (6)	37,24	21	64.88
40	*Hypothetical protein*		*13*	*15,00*	*1*	*8.33*
41	Hypothetical protein		8,9,13 (13)	10,62	3	35.10
42	**Conserved phage mega protein**		**7,8,9,10,11,12,13 (11)**	**34,60**	**13**	**42.78**
43	Hypothetical protein		8,9,12,13 (13)	14,72	2	24.20
44	Hypothetical protein		1,2,3,4,5,6,7,8,9,10,11,12,13 (10)	86,18	21	27.33
46	Hypothetical protein		13	9,14	3	54.70
47	Hypothetical protein		6,7,8,9,10,11,12 (11)	13,97	2	22.10
48	Hypothetical protein		13	9,11	3	37.00
50	Hypothetical protein		1,2,3,4,5,6,7,8,9,10,11,12,13 (11)	14,36	7	57.00
53	Hypothetical protein		2,3,4,5,6,7,8,9,11,12,13 (9)	87,08	31	45.52
54	Hypothetical protein		1,2,3,4 (4)	56,49	15	30.83
55	Hypothetical protein		1,2,5,6,7,8 (8)	29,65	3	12.11
56	Hypothetical protein		1,3,4,5,6,7,8 (8)	30,93	11	46.55
57	Hypothetical protein		3,4,5,6,7,8 (8)	30,16	9	38.19
58	Hypothetical protein		3,4,5,6,7,8 (8)	29,20	7	36.52
59	Hypothetical protein		3,4,5,6,7 (7)	31,19	7	34.10
60	Hypothetical protein		4,5,6,7,8,9 (8)	29,55	10	42.30
61	Hypothetical protein		4,5,7,8 (8)	29,96	9	45.30
62	Hypothetical protein		1,3,4,5,6,7,8,9 (8)	29,61	12	44.40
63	Hypothetical protein		1,3,4,5,6,7,8,9,10,11,12,13 (8)	41,40	13	38.92
64	Hypothetical protein		4,5,6,7,8 (8)	29,96	4	17.20
65	Hypothetical protein		3,4,5,7 (7)	32,16	10	36.80
67	Hypothetical protein		1,2,3,4,5,6,7,8,9,10,11,12,13 (8)	26,32	16	66.98
68	Hypothetical protein		9,11,12,13 (13)	9,95	5	63.82
69	Hypothetical protein		13	11,73	1	11.80
72	*Hypothetical protein*		*8*	*27,59*	*1*	*4.60*
76	Hypothetical protein		11,12 (12)	14,88	4	33.90
106	*Hypothetical protein*	RecA-like protein	*3,5 (5)*	*40,53*	*1*	*3.26*
107	Hypothetical protein	ssDNA-binding protein	5,6,7,8,11,13 (7)	38,06	3	13.21
108	Hypothetical protein		1,2,3,4,5,6,7,8,9,10,11,12,13 (13)	97,96	25	31.24
109	**Gp17 terminase DNA packaging enzyme large subunit**		**11,12 (11)**	**68,13**	**4**	**10.46**
110	***Phage terminase large subunit***		***12,13 (13)***	***14,29***	***1***	***7.32***
111	Hypothetical protein		9,13 (13)	12,80	2	23.90
112	Autotransporter		1,2,3,4,5 (2)	76,62	15	29.56
113	Hypothetical protein		1,2,3,4,5,6,7,8,9,11,12,13 (7)	25,06	11	61.70
122	*Ribonucleotide reductase of class Ia (aerobic) alpha subunit*		*3*	*62,00*	*1*	*1.83*

	(EC 1.17.4.1)					
125	***Phage head completion protein***		***13***	***18,50***	***1***	***8.12***
126	Hypothetical protein		4,6 (6)	36,51	4	15.20
127	Hypothetical protein		1,2,3,4,5,6,7,8,9, 12,13 (9)	21,37	7	52.83
132	**Phage tail sheath monomer**		**1,2,3,4,5,6,7,8,9,10,11,12,13 (2)**	**117,53**	**31**	**41.14**
133	*Hypothetical protein*		*9*	*20,79*	*1*	*4.86*
140	**Hypothetical protein**	**Tail sheath stabilizer and completion protein**	**1,2,3,4,5,8 (4)**	**51,67**	**9**	**26.91**
142	Hypothetical protein		3,4,5,6 (5)	41,07	3	10.20
143	**Hypothetical protein**	**Head completion adaptor; neck**	**3,4,5,6,7 (7)**	**28,45**	**4**	**18.40**
170	Hypothetical protein		1,2,3,4,5,6,7,8,9 (1)	31,89	8	35.16
176	Hypothetical protein		1,2,3,4,5,6,7 (4)	66,07	14	36.20
177	Hypothetical protein		1,2,3,4,6,7,8,9,10,11 (11)	21,66	2	11.60
181	**Hypothetical protein**	**Baseplate wedge**	**1,2,3 (2)**	**136,79**	**11**	**11.85**
182	**Hypothetical protein**	**Baseplate wedge**	**12,13 (12)**	**15,07**	**3**	**22.50**
183	*Hypothetical protein*		*9*	*21,97*	*1*	*4.74*
185	Hypothetical protein		13	10,13	2	12.80
188	Hypothetical protein		7,8 (8)	18,34	4	26.20
196	Hypothetical protein		1,2,3,4,13 (2)	57,94	14	30.20
197	Hypothetical protein		1,2,3,4,5 (1)	87,30	15	23.12
198	**Phage portal vertex of the head T4-like phage Gp20**		**1,2,3,4,5 (3)**	**66,31**	**20**	**43.00**
199	Hypothetical protein		12,13	9,01	1	27.30
200	Topoisomerase IV subunit A (EC 5.99.1-)		3,4,5,6,7,8,9,10,11,12,13 (9)	44,81	4	16.96
204	**Phage prohead core protein**		**1,2,3,4,5,6,7,8,9,10,11,12,13 (11)**	**22,97**	**7**	**36.20**
205	Hypothetical protein		4,5,6,7,8,9,10,11,12,13 (9)	29,01	10	43.60
206	Hypothetical protein		4,5,6,7,8,9 (8)	25,47	6	28.00
210	Hypothetical protein		13	8,75	1	20.30
213	MJ0042 family fingerlike protein		2,3,4,5,7,8,9,10,11,12,13 (13)	32,88	7	20.60
215	**Phage major capsid protein of Caudovirales (T4-like gp23)**		**1,2,3,4,5,6,7,8,9,10,11,12,13 (6)**	**43,51**	**10**	**40.56**
238	*ADP-ribose pyrophosphatase (EC_3_6_1_13)*		*9*	*18,68*	*1*	*6.06*
239	Hypothetical protein		1,4,5,6,7,8,9,10,11,12 (11)	15,80	3	40.60
240	EF hand domain PKD domain protein		5,6 (5)	36,31	2	13.90
241	Hypothetical protein		1,2,3,4,5,6 (6)	43,70	6	18.24
250	Hypothetical protein		7,8,13 (13)	26,25	2	7.66
257	Hypothetical protein		1,2,3,5,6,7,8,9,10,11,12 (11)	16,73	4	43.26
264	*Hypothetical protein*		*5,7 (5)*	*46,80*	*2*	*4,56*
275	*Hypothetical protein*		*2,5,7,8 (8)*	*27,85*	*1*	*3.63*
284	Hypothetical protein		1,3,4,5,6,7,8,9 (8)	25,12	7	50.20
285	Hypothetical protein		12	14,30	3	23.60
286	Hypothetical protein		2,3,4,5 (4)	38,73	8	33.29
287	Hypothetical protein		1,2,3,4,5,6,7,8 (5)	38,49	4	18.70
288	**Phage tail fiber protein**		**1,2,3,4,5,6,7,8,9,10,11,12,13 (2)**	**61,03**	**14**	**35.90**
289	Hypothetical protein		5,7,8,9,11,12,13 (13)	12,51	2	26.40
290	*Hypothetical protein*		*12*	*15,56*	*1*	*15.20*
294	Hypothetical protein		4,5,7,8,9,10,11,12,13 (13)	12,38	4	42.20
295	Hypothetical protein		11	14,92	3	18.90

303	Probable NUDIX hydrolase		3,11 (11)	14,75	1	10.00
305	Hypothetical protein		12	27,33	2	8.75
306	Hypothetical protein		9,10,12 (9)	20,39	2	16.40
311	Hypothetical protein		13	11,83	3	42.30
314	Hypothetical protein		12	12,81	1	16.20
320	Streptococcal hemagglutinin protein		1,3,4,5,6 (6)	40,73	4	18.80
321	Hypothetical protein		1,2,3,4 (1)	36,02	2	8.81
322	Hypothetical protein		2,4,6,7 (6,7)	33,63	2	7.3
323	Hypothetical protein		1,2,3,4,5,6,7,8,9,11,13 (6)	42,40	17	62.20
324	Hypothetical protein		1,2,3,4,5,6,7 (7)	34,76	6	26.50
325	Hypothetical protein		1,2,3,4 (3)	77,23	13	26.03
326	Hypothetical protein		1,2,3,4,5 (1)	75,68	9	16.70
327	**Phage tail fiber protein**		**1,2,3,4,5,6,7,8,9 (2)**	**72,15**	**14**	**37.40**
328	Hypothetical protein		1,2,7,9 (1)	93,95	8	12.74
329	*Extracellular serine proteinase precursor (EC 3.4.21-)*		*1,2,13 (1)*	*139,38*	*3*	*2.08*
377	Hypothetical protein		1,2,3,4,5 (4)	49,73	6	16.70
381	Hypothetical protein		11	20,00	2	15.90
390	*Hypothetical protein*		*13*	*11,75*	*1*	*9.00*
393	Phage protein		12,13 (13)	11,30	2	29.10
394	No significant homology		1,2,3,4,8 (4)	55,10	18	41.01
407	Hypothetical protein		1,2,3,4,5,6,7,8 (6)	39,67	17	44.40
408	Hypothetical protein		1,2,3,4,5,6,7,8,9 (8)	30,60	6	24.78
443	*Hypothetical protein*		*8,12 (12)*	*13,04*	*1*	*7.48*
466	Hypothetical protein		1,2,3,4,5,6,7,8,9,11,12,13 (4)	52,32	21	62.80
470	Hypothetical protein		13	14,60	2	18.50
472	**Phage tail fiber protein**		**1,2,3 (2)**	**80,58**	**14**	**28.10**
475	Hypothetical protein		1,2,3,4,5,6,7,8,9,10,11,12,13 (11)	31,66	14	39.54
482	Hypothetical protein		7,8,9,10,11,12,13 (12)	31,68	9	36.70
483	Hypothetical protein		5,6,7,8,9,10,11,12 (12)	33,72	16	55.80
487	Hypothetical protein		1,4,5,6,7,8,9,10,11,12,13 (13)	12,99	8	50.90
498	Hypothetical protein		1,2,3,4,5,6,7,8,9,10,11,12,13 (7)	35,16	23	75.50
514	Hypothetical protein		12,13 (13)	7,10	2	31.70
544	Hypothetical protein		4,5,6,7,8,9,10,12,13 (13)	31,80	9	32.06
548	Hypothetical protein		12	13,65	3	40.00
561	Hypothetical protein		1,2,3,4,5,6,7,8,9,10,11,13 (9)	17,79	7	52.16
574	Hypothetical protein		1,5 (5)	52,57	6	18.10
576	Hypothetical protein		1,4,5 (4)	52,20	4	7.23
577	Hypothetical protein		1,2,3,4,5,6,7,8,9,10,11,12,13 (12)	14,42	5	52.10
578	Hypothetical protein		2,3,4,5,6,7,8,9,10,11 (9)	19,16	4	43.90
594	Hypothetical protein		3,7,8,9,12,13 (13)	23,87	15	57.51
606	*Hypothetical protein*		*13*	*8,62*	*1*	*13.90*
616	*Hypothetical protein*		*1*	*32,06*	*1*	*4.67*
624	Putative phage related protein precursor	D-alanyl-D-alanine carboxypeptidase	12,13 (12)	14,79	1	9.85
627	*Hypothetical protein*		*8*	*27,46*	*1*	*3.69*
636	Hypothetical protein		13	11,55	2	21.60
639	Hypothetical protein		12	16,38	3	29.60
661	Hypothetical protein		12	14,72	1	10.40
668	Hypothetical protein		1,2,3,4,5,6,7,8,9,10,11,12 (11)	20,34	8	50.59
672	Hypothetical protein		9,10 (9)	22,56	3	23.15
675	Hypothetical protein		1,2,3,4,5,6,7,8,9,10,11,12,13 (8)	24,87	3	20.60
696	Hypothetical protein		2,3,4,5,7,8 (4)	50,11	5	13.90

697	**Hypothetical protein**	**Baseplate wedge**	**9,10,11,12,13 (13)**	**12,32**	**3**	**38.50**
698	Hypothetical protein		1,2,3 (3)	72,75	7	11.50
701	**PE-PGRS virulence associated protein**	**Tail tape measure protein**	**5,6,7,8,9 (9)**	**21,77**	**8**	**68.80**
707	*Hypothetical protein*		*9*	*24,28*	*1*	*3.98*

[a]*Italics:* Proteins with only a single peptide or low coverage

Bold: Proteins annotated as structural proteins within the head, neck, tail or tail fibers

Chapter 4

Isolation and Characterization T4- and T7-like Phages that Infect the Bacterial Plant Pathogen, *Agrobacterium tumefaciens*

Author Contributions

HA conducted experiments. HA and PB designed experiments and analyzed data for all figures. HA and PB contributed to writing and editing of the manuscript.

To be submitted as

Attai, H., Brown, P. J. B. (2019) Isolation and Characterization T4- and T7-like Phages that Infect the Bacterial Plant Pathogen, *Agrobacterium tumefaciens*.

Abstract

In the rhizosphere, bacteria-phage interactions are likely to have important impacts on the ecology of microbial communities and microbe-plant interactions. To better understand the dynamics of Agrobacteria-phage interactions we have isolated diverse bacteriophages which infect the bacterial plant pathogen, *Agrobacterium tumefaciens*. Here, we complete genomic characterization of *Agrobacterium tumefaciens* phages Atu_ph04 and Atu_ph08. Atu_ph04, a T4-like phage belonging to the *Myoviridae* family, was isolated from waste water and has a 143,349 bp genome which encodes 223 predicted open reading frames (ORFs). Based on phylogenetic analysis and whole genome alignments, Atu_ph04 is a member of a newly described T4 superfamily which contains other *Rhizobiales*-infecting phages. Atu_ph08, a member of the *Podoviridae* T7-like family, was isolated from waste water, has a 59,034 bp genome and encodes 75 ORFs. Based on phylogenetic analysis and whole genome alignments, Atu_ph08 may form a new T7 superfamily which includes *Sinorhizobium* phage PCB5 and *Ochrobactrum* phage POI1126. Atu_ph08 is predicted to have lysogenic activity, as we found evidence of an integrase and several transcriptional repressors with similarity to proteins in transducing phage P22. Together, this data suggests that *Agrobacterium* phages are diverse in morphology, genomic content, and lifestyle.

Keywords: *Agrobacterium tumefaciens*; bacteriophage; phage; biocontrol

Introduction

Agrobacterium tumefaciens is a plant pathogen that causes damage to crops worldwide [1]. This Gram negative bacterium transforms plant cells, which results in over-proliferation of host cells, causing crown gall disease in the form of tumors that block the plant from receiving proper nutrients. The interactions between *Agrobacterium* and plants have been studied extensively, leading to innovations in plant biotechnology [2,3]. In contrast, little is known about the natural predators of *Agrobacterium*. Studies of bacteriophages that prey upon bacterial plant pathogens such as *Agrobacterium* should reveal effective biocontrol strategies for host cell killing that can be exploited to limit phytopathogenesis [4,5]. With the rise of antibiotic resistant bacteria, there has been an increased interest in phage research; however, the diversity of phages that infect soil bacteria is under-sampled relative to phages of human pathogens and marine environments [6,7]. Understanding the diversity of phages in soil is important because of their impact on host populations, community interactions, and biogeochemical cycles [8].

Here, we sought to further explore the diversity of phages that infect *Agrobacterium tumefaciens*. Currently, there are 4 characterized lytic phages that infect *Agrobacterium*: 7-7-1 [9], Atu_ph02 and Atu_ph03 [10], and Atu_ph07, a jumbo phage [11]. Phage 7-7-1 and Atu_ph07 are T4-like *Myoviridae* and Atu_ph02 and Atu_ph03 are T7-like *Podoviridae*. Here, we report characteristics of 2 additional phages, Atu_ph04 and Atu_ph08, and compare them to related phages, including the extensively-characterized *Escherichia* phage T4 [12,13] and P1.

Materials and Methods

Bacterial Strains and Culture Conditions

Strains used in this study are shown in Table 4-1. *Agrobacterium tumefaciens* strains were cultured in Lysogeny Broth (LB), with the exception of *A. tumefaciens* strain LBA4404, which was grown in yeast mannitol (YM) medium. *Agrobacterium vitis* was cultured using potato dextrose media (Difco), *Rhizobium rhizogenes* was grown in mannitol glutamate yeast (MGY) medium and *Caulobacter crescentus* was grown in peptone-yeast extract (PYE) medium [14]. These strains were grown at 28°C. *Escherichia coli* was grown in LB at 37°C. Liquid cultures were grown with shaking and solid medium was prepared with 1.5% agar.

Phage Isolation and Purification

Phage Atu_ph04 was isolated from an effluent sample from a waste water treatment plant in Columbia, MO, while Atu_ph08 was isolated from a waste water sample from Reno, Nevada. *A. tumefaciens* strain C58 was used as a host strain, using the multiple-enrichment isolation method as described previously [10,15].

Plaque Assays

Whole-plate plaque assays were performed with the soft agar overlay method [10]. Briefly, 100 µl cells, grown at an optical density of 600 nm (OD_{600}) of ~0.2 and diluted to OD_{600} of 0.05, were mixed with 100 µl phage for 15 min at room temperature prior to dilution to allow attachment. This mixture of cells and phage were serially diluted in LB and added to 3 ml of melted 0.3% LB-soft agar. The solution was then overlaid onto a 1% LB-agar plate and swirled for even distribution. For host range testing, serial dilutions of phage were spotted onto a bacterial lawn. A mixture of 100 µl

cells (OD_{600} of ~0.2) and 0.3% LB-soft agar was overlaid onto a 1% LB-agar plate. Once the cells solidified, 5 µl of phage dilutions were spotted onto the soft agar. Plates were incubated for 1–2 days to allow plaque formation.

Preparation of Virion DNA, Genome Sequencing, and Genome Assembly

DNA was isolated from purified virions using phenol-chloroform extraction as described previously [10]. Libraries for genome sequencing were constructed from virion DNA following the manufacturer's protocol and reagents supplied in Illumina's TruSeq DNA PCR-free sample preparation kit (FC-121-3001) [10]. The purified library was quantified using a KAPA library quantification kit (KK4824), and library fragment sizes were confirmed by Fragment Analyzer (Advanced Analytical Technologies, Inc.). Libraries were diluted, pooled, and sequenced using a paired-end 75-base read length according to Illumina's standard sequencing protocol for the MiSeq. Library preparation and sequencing were conducted by the University of Missouri DNA core facility.

DNA Restriction Analysis

Phage genomic DNA was digested with restriction endonucleases from New England Biolabs using the standard protocol. All reactions contained 500 ng DNA, which was incubated for 2 h at 37°C. Digested DNA was analyzed on a 0.7% agarose gel. Gel electrophoresis was performed at 100 V for 1 h and stained with SYBR Safe DNA Gel Stain (Thermo Scientific).

Growth Curves

Growth curves were performed by growing bacteria at a starting OD_{600} of 0.05 in LB. Cells were mixed with purified phage in liquid media at the MOIs indicated. Cell growth was measured by the culture turbidity, represented by the absorbance at OD_{600}. Measurements were taken every 10 min for 36 h. Cells were grown at 28°C and shaken for 1 min prior to each reading. The OD_{600} was measured using a BioTek Synergy H1 Hybrid reader. Results were taken in quadruplicate and averaged.

Transmission Electron Microscopy

Virion morphology was observed by applying a small volume of concentrated purified virions onto a freshly, glow-discharged carbon-coated TEM grid and negatively stained with 2% Nano-W (Nanoprobes, LLC, Brookhaven NY) or 2% uranyl acetate. Specimens were observed on a JEOL JEM-1400 transmission electron microscope at 120 kV. Capsid diameters of Atu_ph04 (n=103 virions) and Atu_ph08 (n=61 virions), as well as tails of Atu_ph04 (n=15 virions) and Atu_ph08 (n=15 virions) were measured using ImageJ (v.2.0.0) [16].

Genome Annotation

The sequences were annotated by the RAST server [17] and ORFs with no homology in the database, or ORFans, were defined as having an e-value greater than 1e-03 by PSI-BLAST v 2.8.1 [18]. All gene products were analyzed by TMHMM [19]. The presence of tRNAs were detected by tRNAscan-SE (version 2.0) [20]. G + C content was

analyzed by Geneious (v.11.0.5) [21]. Pairwise (%) nucleotide identity was determined using the Mauve plugin in Geneious [22].

Phylogenetic Analysis

Homologs of the large terminase subunit in Atu_ph08 and portal vertex protein in Atu_ph04 were identified by BLASTp using an E-value cutoff of 1e-03. Protein alignment was performed by Geneious using ClustalW (v.2.1) and the BLOSUM matrix [21,23]. Maximum-likelihood trees based on phylogeny (PhyML) were built using a Geneious plugin with 100 bootstrap models [24].

GenBank Accession Number

The genome sequences of *Agrobacterium* phage Atu_ph04 and Atu_ph08 are available in GenBank under accession numbers MF403007 and MF403009, respectively.

Results and Discussion

Phage Atu_ph08 has Higher Lytic Activity than Atu_ph04

Waste water includes agricultural runoff and provides an enriched mixture of bacterial populations making this a prime environment for isolation of bacteriophages. We isolated phages that infect *A. tumefaciens* from waste water using a phage enrichment protocol as described previously [10]. Infection of *A. tumefaciens* C58 with Atu_ph04 or Atu_ph08 results in the formation of small, clear plaques (Figure 4-1A) or larger, clear plaques (Figure 4-1B), respectively. Negative-staining transmission electron microscopy (TEM) of Atu_ph04 reveals an icosahedral head and tail (Figure 4-1C) classifying

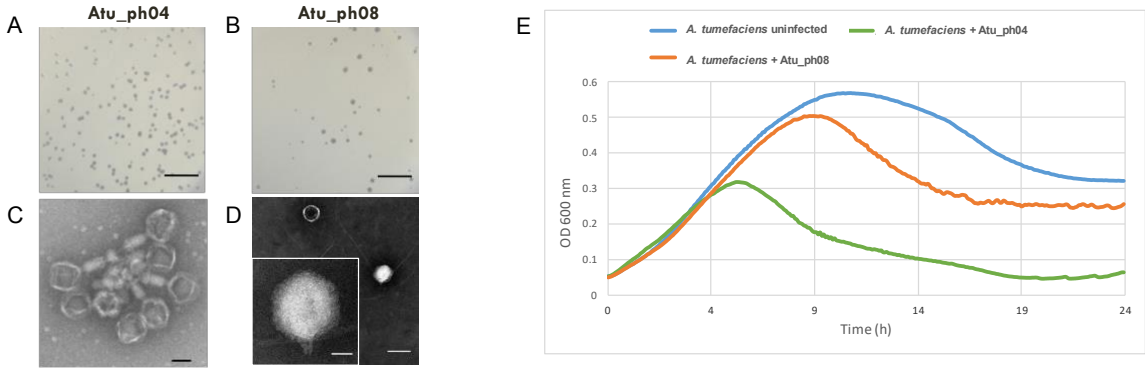


Figure 4-1. Characterization of Atu_ph04 and Atu_ph08. Plaque assays of Atu_ph04 (**A**) and Atu_ph08 (**B**). Scale bars represent 10 mm. Transmission electron microscopy of (C) Atu_ph04 shows it is in the *Myoviridae* family. Scale bars represent 100 nm. (D) Atu_ph08 is in the family *Podoviridae*. Scale bar (right) represents 100 nm and scale bar in inset represents 25 nm. (E) Growth curve of *A. tumefaciens* C58 cells growing in the presence and absence of phage at an MOI of 0.001.

Atu_ph04 in the family *Myoviridae* [25]. The average capsid head diameter of Atu_ph04 is 84.7 nm and its tail length is 79.8 nm. TEM of Atu_ph08 reveals the presence of an icosahedral head with an average diameter of 65.0 nm and a short, stubby tail with a length of 21.9 nm (Figure 4-1D) indicating that this phage belongs to the *Podoviridae*.

Growth curves of *A. tumefaciens* strain C58 infected with Atu_ph04 and Atu_ph08 at an MOI of 0.001 reveals that Atu_ph04 begins to exhibit lethal activity at 4 h post-infection, whereas the modest lytic activity of Atu_ph08 is observable after 8 h post-infection (Figure 4-1E). While both phages exhibit lytic activity, Atu_ph04 would be preferred for biocontrol purposes because it significantly reduces cell turbidity.

Host Ranges of Atu_ph04 and Atu_ph08 are Limited to A. tumefaciens Strains

Host range was determined by performing plaque assays of phage dilutions and is summarized in Table 4-2. Atu_ph04 causes lysis of most C58-derived *A. tumefaciens* strains, including C58, EHA101, EHA105, GV3101, but does not infect AGL-1. Furthermore, Atu_ph04 is able to lyse NTL4 and LBA4404 but unable to infect *A. tumefaciens* Chry5 or other bacterial species. Atu_ph08 lyses C58-derived *A. tumefaciens*, however it is only moderately infective in AGL-1. Atu_ph08 does not infect Chry5 or other bacterial species. This host range is comparable to the range of other *A. tumefaciens*-infecting phages described. The narrow range suggests that Atu_ph04 and Atu_ph08 will not disrupt other, beneficial bacterial strains in the rhizosphere, an important consideration when selecting phages for biocontrol.

Genomic Characteristics of Atu_ph04

The genome of Atu_ph04 is 143,349 bp in length, with a G + C content of 49.4% (Figure 4-2, Supplementary Table 4-S1, Table 4-3). Interestingly, attempts to digest the Atu_ph04 genomic DNA with 9 different restriction enzymes failed, despite the presence of the restriction sites in the genome sequence, suggesting that the DNA may be modified (Supplementary Figure 4-S1). The genome of Atu_ph04 encodes 223 open reading frames (ORFs), of which, 73 have predicted functions, 83 are ORFans, meaning they have no obvious homologs, and 67 conserved hypothetical proteins. Atu_ph04 only encodes one predicted tRNA, but its anticodon is undetermined, as predicted by tRNAscan-SE v 2.0 [20].

Of the 73 gene products with predicted functions encoded by Atu_ph04, many include structural proteins such as the portal vertex of the head (gp72), the major capsid protein (gp76), and a T4-like phage large terminase (gp53). The Atu_ph04 major capsid protein shares 76% identity with *Sinorhizobium* phage phiM9 major head subunit, gp23, as characterized by Johnson, et al [26]. Atu_ph04 also encodes DNA synthesis proteins, including DNA topoisomerase (gp110, gp113), nucleotide metabolism proteins, such as ribonucleotide reductase of class 1a alpha (gp24) and beta subunits (gp25), and proteins involved in translation, like RNA polymerase sigma factor (gp89, 119).

Phylogenetic Analysis Shows Atu_ph04 is Closely Related to T4-Like Sinorhizobium Phage phiM9 and Rhizobium Phage vB_RleM_P10VF

Phage Atu_ph04 shares pairwise identity with *Rhizobium* phage vB_RleM_P10VF (21.6%) and *Sinorhizobium* phage phiM9 (19.7%) and whole genome alignments constructed using Mauve [22] reveal that the three genomes contain blocks of

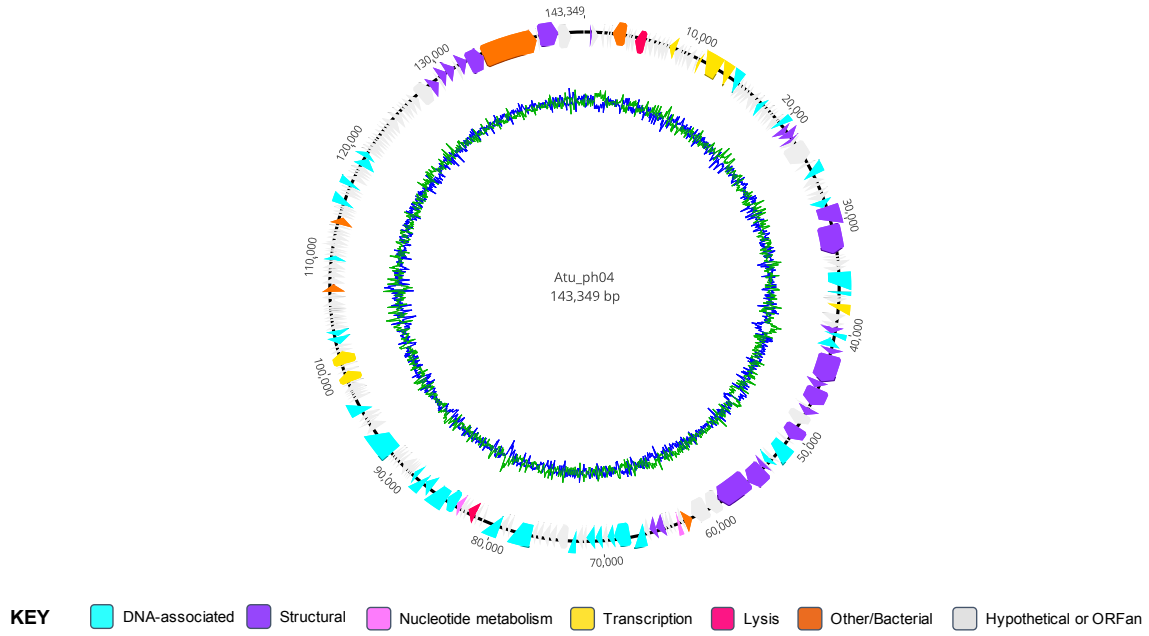


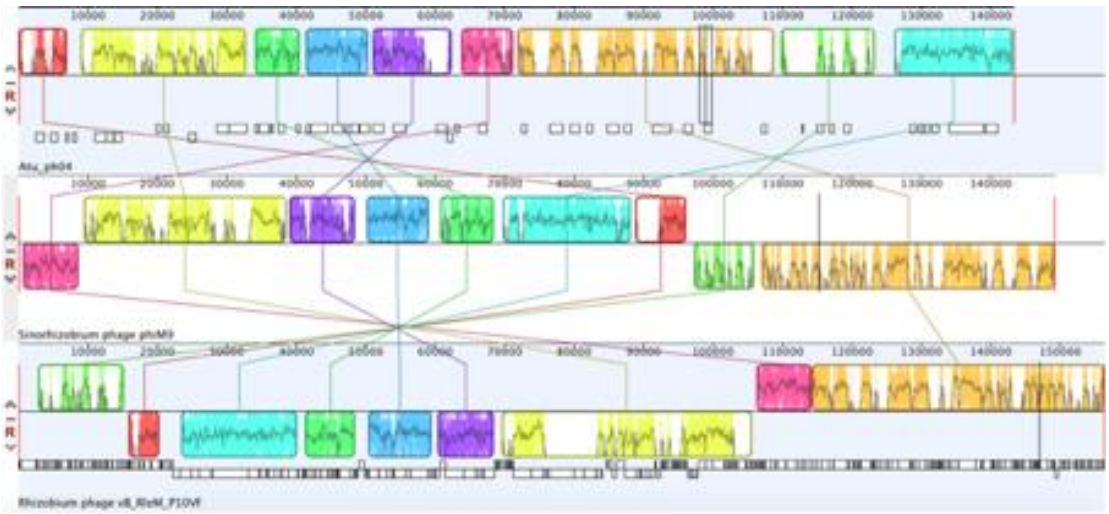
Figure 4-2. Genome annotation of Atu_ph04, color-coded by functional annotation. G + C content represented by inner circle: AT=green and GC=blue.

genomic synteny (Figure 4-3A) suggesting that Atu_ph04 joins this recently-described group of T4 superfamily phages [26]. This analysis is consistent with the phylogenetic tree built using an alignment of the portal vertex protein (Figure 4-3B). This group of rhizophages is clustered into a larger group of cyanophages and *Synechococcus* phages. Comparative analysis of the gene products of Atu_ph04 with those of several representative T4-like phages confirms a relatively high degree of gene conservation among *Rhizobium* phage vB_RleM_P10VF and *Sinorhizobium* phage phiM9 (Supplementary Table 4-S2).

Atu_ph04 is a T4-like Phage but Lacks Several T4 Core Proteins

Though Atu_ph04 is placed in the T4 superfamily, Atu_ph04 only shares 4.5% pairwise identity with *Enterobacteria* phage T4. To determine the relationship between Atu_ph04 and T4, we performed a comparative analysis matching T4 core proteins with the Atu_ph04 genome (Supplementary Table 4-S3). The genome of Atu_ph04 encodes putative homologs of 14 of the 22 T4 core proteins (with an E-value > 1E-03); however, it is missing key T4 core proteins, including some structural proteins. Though the Atu_ph04 genome encodes a T4-like gp21, the prohead core protein, it does not encode gp22, another prohead core protein that is essential in phage T4 [12]. Similar to phages phiM9 and vB_RleM_P10VF, Atu_ph04 also has a split T4 gp5 baseplate hub protein (gp54, gp213). The Atu_ph04 genome also lacks obvious homologs of T4-like tail fibers (T4 gp34, 36). The absence of T4-like tail fibers in the Atu_ph04 genome (Supplementary Table 4-S3) may be compensated by the presence of gp222, a predicted tail fiber protein and that is conserved in phiM9 and vB_RleM_P10VF (Supplementary

A



B

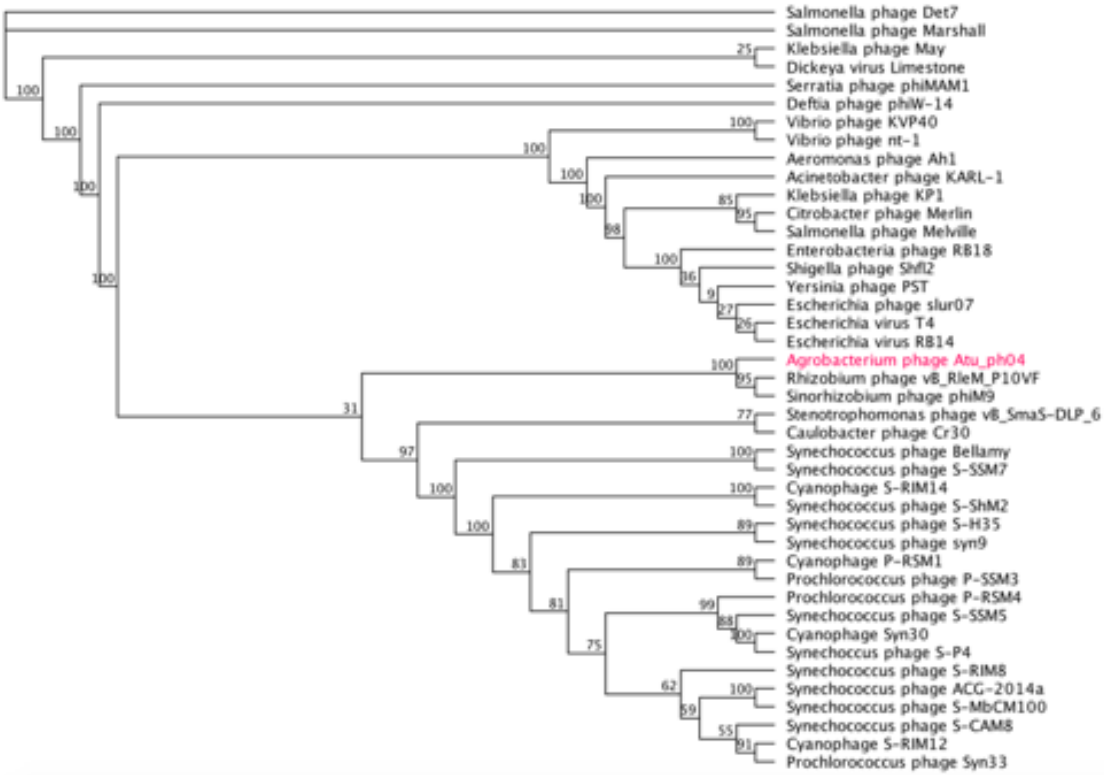


Figure 4-3. Phylogenetic analysis of Atu_ph04 with its relatives. (A) Mauve genome alignment of Atu_ph04, *Sinorhizobium* phage phiM9, and *Rhizobium* phage RleM_P10VF. (B) Phylogenetic tree of portal vertex protein.

Table 4-S2). This difference in tail fiber proteins likely allows this group of rhizophages to infect a different host than T4 does.

Another feature of Atu_ph04, phiM9, and vB_RleM_P10VF genomes is the lack of genes encoding T4 protein gp33, which is involved in late transcription. Instead, it is hypothesized that phiM9 and vB_RleM_P10VF encode an RpoE stress response sigma factor, which compensates for the missing protein [26]. In the Atu_ph04 genome, not only is T4 protein gp33 missing, but the core sigma factor for late transcription protein gp55 is also not encoded. The Atu_ph04 genome encodes a DNA-directed RNA polymerase RpoE sigma factor (gp89) that shares 20.3% pairwise identity with the sigma factor in phiM9. It also encodes gp119, a putative sigma factor for late transcription, which shares 49% identity with the one encoded by phiM9. Additionally, the Atu_ph04 genome encodes T4 core protein NrdA (gp24), the alpha subunit of ribonucleotide reductase class 1a, which is involved in nucleotide metabolism. Yet, instead of *nrdB*, which encodes the beta subunit in T4, it encodes a presumably diverged ribonucleotide reductase class 1a, beta subunit homolog (gp25), adjacent to its alpha partner. Together, these data suggest that the rhizophages have diverged from the T4-phages with respect to regulation of transcription throughout the phage replication cycle and nucleotide metabolism.

Major Gene Categories of Atu_ph04

The Atu_ph04 genome encodes 24 predicted structural gene products, including 2 putative tail fiber proteins (gp1, 222), 4 tail completion and sheath proteins (gp66, 70, 71, 218), 11 baseplate subunits (gp41, 42, 43, 54, 82, 83, 84, 93, 94, 213, 219), 4 capsid head

proteins (gp69, 72, 74, 76), 1 terminase (gp53), and 2 neck proteins (gp215, 216). Protein VrlC (gp220) is predicted to be responsible for the structure of double-layered, or double ring-like, baseplates [27,28], which are a feature of some T4-like phages but not T4 itself.

Atu_ph04 has an abundance of genes involved in DNA replication, repair, and recombination. It encodes 34 DNA-associated proteins involved in DNA replication, repair, and recombination. The DNA replication proteins include 2 DNA primases (gp26, gp195), single-stranded DNA-binding proteins (gp47, 67), ribonuclease H (gp63) [29], DNA helicase (gp78), two topoisomerase subunits (gp110, 113), and 3 sliding clamp loader subunits (gp122, 123, 124). The DNA polymerase is predicted to be gp133. There is a cluster of DNA-associated proteins: DNA primase/helicase (gp97), a putative holliday junction resolvase (gp98), 5'-deoxynucleotidase (gp100), a deoxynucleotide monophosphate kinase (gp101), and deoxycytidylate 5-hydroxymethyltransferase (gp104).

The presence of 3 putative homing endonucleases (gp52, 58, 68) in close proximity to the large terminase (gp53) is consistent with the hypothesis that these endonucleases are involved in DNA packaging [30]. Gp60 shares similarity with T4 protein DenV, which is responsible for the removal of pyrimidine dimers caused by UV damage, a process necessary for DNA repair [31].

Several proteins involved in nucleotide metabolism are often encoded by phages. The Atu_ph04 genome encodes 6 proteins involved in this process. These include the MutT/Nudix family protein (gp17), a putative glutaredoxin (gp23), ribonucleotide reductase alpha (gp24) and beta (gp25) subunits, thymidylate synthase (gp145), and GT1 glycosyltransferase (gp148).

Atu_ph04 also encodes several genes that enhance the survival of their bacterial hosts. One such example is the phosphate starvation-inducible protein PhoH (gp87), which is suggested to enhance the phosphate metabolism in the host under stress [32]. Another bacterial gene product (gp6) encodes UDP-galactopyranose mutase, which is involved in the synthesis of the essential bacterial cell wall component, galactofuranose [33]. Finally, Atu_ph04 encodes 2 putative lysis proteins: gp10, which is an N-acetylmuramoyl-L-alanine amidase, and gp116, which is a predicted hydrolase of the conserved HD superfamily consistent with our classification of Atu_ph04 as a lytic phage.

Atu_ph08 Genomic Summary

The genome of Atu_ph08 is 59,034 bp in length, with a G + C content of 59.7% (Figure 4-4, Table 4-3, Supplementary Table 4-S4). The Atu_ph08 genome encodes 75 ORFs, only 3 of which are ORFans (gp45, gp63, gp75). Of the 75 ORFs, 43 encode conserved hypothetical proteins and 32 have predicted functions. Atu_ph08 does not contain any obvious tRNA-encoding genes.

Gene Organization of Atu_ph08

The Atu_ph08 genome encodes 8 predicted structural proteins (Figure 4-4, purple arrows), including two potential major capsid proteins (gp31, 36). The tail fiber proteins gp23, 28), the portal protein (gp15) and the large terminase (gp13). Remarkably, the Atu_ph08 genome does not encode any gene products involved in DNA replication, such as DNA polymerase, with the exception of the DarB-like gp21, suggesting that it may use

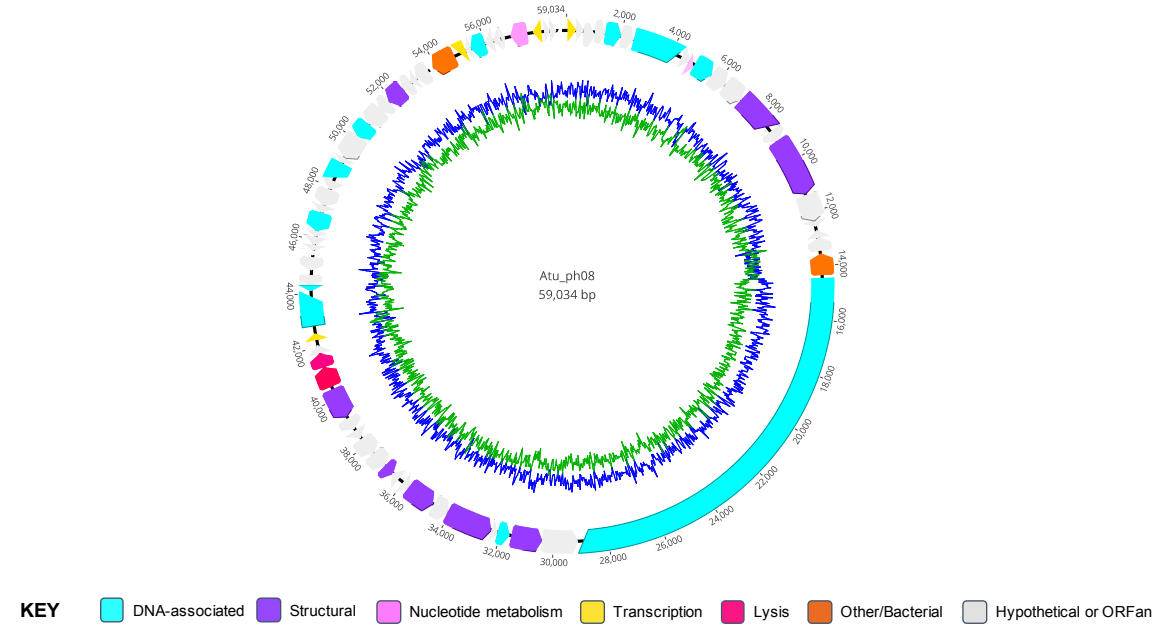


Figure 4-4. Genome annotation of Atu_ph08, color-coded by functional annotation. G + C content represented by inner circle: AT=green and GC=blue.

host machinery to replicate its DNA. The genome does encode several gene products predicted to be involved in DNA modification. These include gp7, which is a cytosine-specific DNA methylase and a NERD domain-containing protein (gp10), predicted to be involved in DNA processing [34]. Other DNA modification proteins include: N-acetyltransferase (gp24), 3'-5' exoribonuclease (gp49), methyltransferase (gp53), a metal-dependent phosphohydrolase (gp56), and a class I SAM-dependent methyltransferase (gp67).

Atu_ph08 also encodes transcription regulators, including the GcrA cell cycle regulator (gp5), which activates transcription at methylated promoter sequences by interacting with RNA polymerase, previously characterized in *Caulobacter crescentus* [35]. The putative GcrA regulator in the Atu_ph08 genome is 89.74% identical to a hypothetical protein (WP_080842116.1) in *Agrobacterium* genomospecies 3. The GcrA protein is conserved within the Alphaproteobacteria [36], as well as phiCbk-like *C. crescentus* phages [37], suggesting that phage may have acquired this protein from their hosts potentially enabling the phage to upregulate host DNA replication machinery.

There are two predicted genes involved in posttranslational modifications. Gp71 is predicted to be a Clp protease, and gp9 contains a PRK12775 domain, which is predicted to be involved in amino acid transport and metabolism.

Atu_ph08 has Some Features of a Temperate Phage and Shares High Homology with A. tumefaciens genomospecies 3

The genome of Atu_ph08 shares most of its genes with *A. tumefaciens* and *Rhizobium* species, leading us to hypothesize that Atu_ph08 and the Alphaproteobacteria

have exchanged genes through horizontal gene transfer. Furthermore, the G + C content of the genomes of *A. tumefaciens* and phage Atu_ph08 are similar (~59%), in contrast with the G + C content of the other *Agrobacterium* phages, which are all lower. An initial analysis of the *Agrobacterium* genomospecies 3 strain CFBP 6623 genome (Accession number: NZ_LT009723) reveals the existence of 3 intact prophage regions and 1 incomplete prophage at the 1.5 million bp [38]. Mauve genome alignment of Atu_ph08 with this region in *Agrobacterium* genomospecies 3 strain CFBP 6623 (1,555,808-1,601,554 bp) revealed a 60.2% pairwise identity between the genomes (Figure 4-5).

Interestingly, while attempts to UV induce lysogens from *A. tumefaciens* C58 cells infected with Atu_ph08 have been unsuccessful thus far, the Atu_ph08 genome encodes an integrase (gp41) and an XRE transcriptional regulator (gp1). The XRE transcriptional regulator belongs to a family of transcriptional regulators that contains Cro and cI repressors [39], suggesting that Atu_ph08 may exhibit lysogenic activity or be derived from an ancestor with lysogenic activity. The Atu_ph08 integrase shares 34% identity to the integrase encoded by *Salmonella* phage vB_SemP_Emek, which is a P22-like phage. P22 is a transducing phage that encodes the C2 repressor, so we sought to determine if the Atu_ph08 genome encodes a transcriptional repressor. Remarkably, gp65, annotated as a transcriptional regulator, shares 28% identity with the C2 repressor in vB_SemP_Emek. Directly upstream of the gene encoding the integrase is the gene encoding an Arc family phage regulatory protein (gp42), which acts as a transcriptional repressor in phage P22 [40]. Directly downstream of these genes is another peculiar gene encoding an AlpA family phage regulatory protein (gp40). AlpA has been characterized in *E. coli* to suppress sensitivity to UV light [41]. The presence of these genes strongly

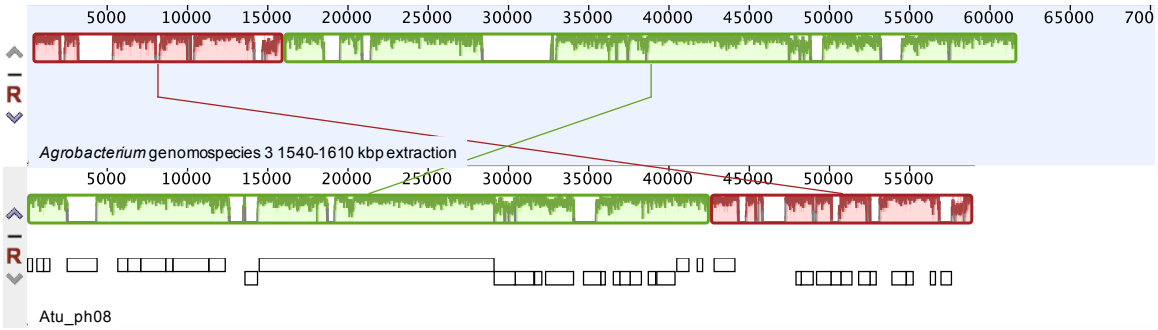


Figure 4-5. Mauve genome alignment of 1540-1610 kbp region of *Agrobacterium* genomospecies 3 and Atu_ph08.

suggests that Atu_ph08 may be lysogenic and it should be explored as a candidate transducing phage for *A. tumefaciens*.

The Atu_ph08 Genome is Highly Syntenic with the Genome of the T7-Like Sinorhizobium Phage, PBC5

Phylogenetic analysis of Atu_ph08 reveals that it is closely related to *Sinorhizobium* phage PBC5 and *Ochrobactrum* phage POI1126. The Atu_ph08 genome shares 38.2% pairwise identity with *Sinorhizobium* phage PBC5 and 24.0% identity with *Ochrobactrum* phage POI1126. The large terminase tree (Figure 4-6A) shows that Atu_ph08 forms a distinct group with PBC5 and POI1126, and is placed within a larger group with T7-like *Burkholderia* phage Bcepmigl and *Erwinia* phage PEp14. These phages are distant relatives of the T7-superfamily of *Podoviridae* phages. The close relation to PBC5 and POI1126 are observed in the Mauve genome alignment (Figure 4-6B). These alignments show evidence that genomic rearrangements have taken place among phages in this family. The mosaicism of phage genomes is a common result of horizontal gene transfer [42].

The Atu_ph08 Genome Encodes a DarB-like Protein, Commonly Found Among PBC5-Like Phages

The Atu_ph08 genome encodes a 4,877 aa gene product (gp21), previously discussed in the context of this phage family in Gill et al. [43], which has 4 major domains that suggest it may have helicase and methylase activity (Supplementary Figure 4-S2A). This unusually large gene product is described as a DarB homolog. DarB, or

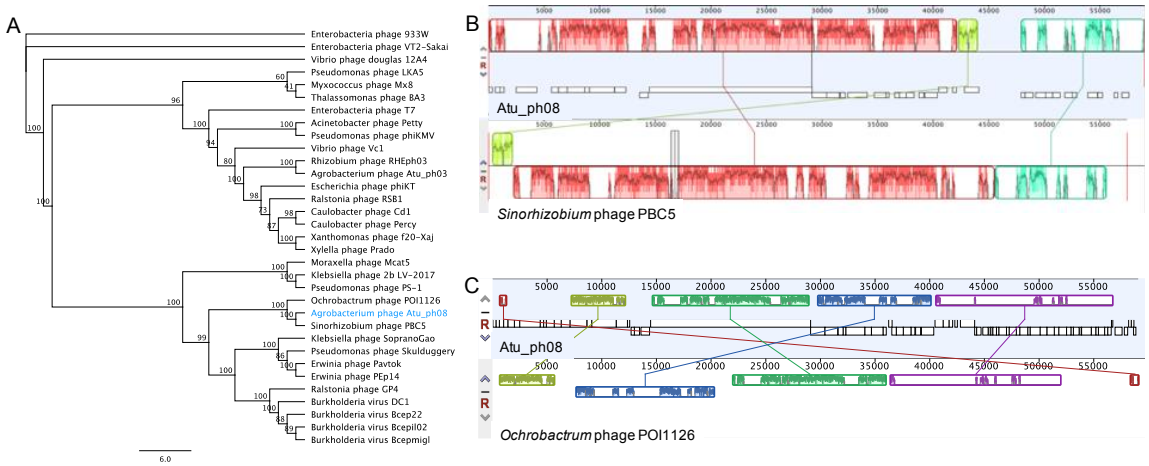


Figure 4-6. Relatives of Atu_ph08. (A) Phylogenetic tree of large terminase protein. Mauve genome alignment of Atu_ph08 with (B) *Sinorhizobium* phage PBC5 and (C) *Ochrobactrum* phage POI1126.

defense against restriction, is an *Escherichia* phage P1 protein that protects the phage from host restriction enzymes, *Eco*B and *Eco*K [44]. In phage P1, DarB is prepackaged inside the capsid, allowing DNA methylation to occur immediately upon infection, protecting the DNA from host killing by restriction [45,46].

The DarB-like protein in Atu_ph08 is 21.6% identical to the DarB-like protein of *Burkholderia* phage Bcep22 and is predicted to have both methyltransferase and helicase domains. Similar to Bcep22, Atu_ph08 does not have a DarA homolog encoded in the genome, which was thought to be required for DarB incorporation into the capsid. The DarB protein in Bcep22 also has a lytic transglycosylase domain on its N-terminus. The Atu_ph08 DarB protein appears to have an N-terminal cell wall hydrolase domain followed by a peptidase domain.

This DarB-like protein appears to be conserved in several T7-like phages (Supplementary Figure 4-S2B). A bioinformatic search found that *Agrobacterium* phages Atu_ph02 and Atu_ph03 also have a DarB-like protein. Since Atu_ph02 and Atu_ph03 share a host with Atu_ph08, acquisition of similar proteins to protect phage DNA from *A. tumefaciens* restriction and modification systems that destroy foreign DNA is plausible. Remarkably, DarB homologs are often found on mobile genetic elements, including the Ti plasmid of *A. tumefaciens,* suggesting that DarB likely confers a benefit to invading foreign DNAs [43].

The Atu_ph08 Genome Encodes a Putative Holin-Endolysin Cassette

The genome of Atu_ph08 encodes three possible gene products involved in cell lysis, which are consecutively located (gp37-9). The first, gp37, encodes a lysozyme-like

domain. Directly adjacent, gp38 shares homology with a putative 3TM holin, named after a family of holins for gene-transfer release with 3 transmembrane domains, encoded by Alphaproteobacterium *Mesorhizobium australicum*. All three genes are predicted to encode transmembrane domains—gp37 contains 1, gp38 contains 2, and gp39 contains 3. As holins are typically located in the inner membrane where they form a pore, it is likely that gp38 exhibits holin activity.

Conclusions

In this study, we characterize two additional *Agrobacterium* phages, which is important given the under-sampling of phages from soil and rhizosphere environments. Despite sharing a common host, no conserved proteins are identified among all the *Agrobacterium* phage genomes, suggesting that the phages may not share mechanisms of host entry or lysis. Atu_ph04 forms a group with *Rhizobium* phage RleM_P10VF and *Sinorhizobium* phage phiM9, which are in the T4 superfamily, and Atu_ph08 is closely related to *Sinorhizobium* phage PBC5 and *Ochrobactrum* phage POI1126, which are T7-like. Through our comparative analysis, we found that Atu_ph08 may be a temperate phage, as it encodes several genes that are commonly expressed in phages that undergo the lysogenic cycle. Together, this data, along with previously published data on *Agrobacterium* phages, illustrates the diversity of phages that share a common host and provides examples of the breadth of genes these phages express, which can further our understanding of microbial diversity. Further studies are required to understand the impact these phages play in the environment where they reside.

Author Contributions: HA conducted experiments. HA and PB designed experiments, analyzed data and contributed to writing and editing of the manuscript.

Acknowledgements: George Smith provided valuable technical assistance during the initial purification and characterization of Atu_ph04 and Atu_ph08. We thank Kenya Phillips for assisting in the isolation and characterization of Atu_ph08 and Courtney Buchanan for assisting in characterization of Atu_ph04. We thank Tommi White, Martin Schauflinger, and DeAna Grant of the MU Electron Microscopy Core for help with the transmission electron microscopy. We thank Nathan Bivens and the MU DNA Core for assistance with sequencing the bacteriophages and William Spollen and the MU Research Informatics Core for assistance with genome assembly and GenBank submission. We thank Zhanyuan Zhang at the MU Plant Transformation Core facility for providing *Agrobacterium* strains. Finally, we thank members of the Brown lab, especially Michelle Williams, for feedback during the preparation of the manuscript.

Funding: This research is supported by startup funds, a research council grant (URC 14-051), and a research board grant (3786-2) from the University of Missouri to PJBB. HA has been supported by the National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health (NIH) under award number T32GM008396 and the U.S. Department of Education Graduate Assistance in Areas of National Need (GAANN) Fellowship.

Conflicts of Interest: This research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Pulawska, J. Crown gall of stone fruits and nuts, economic significance and diversity of its causal agents: tumorigenic *Agrobacterium* spp. *J. Plant Pathol.* **2010**, *92*, S1.87-S1.98.
2. Sardesai, N.; Subramanyam, S. *Agrobacterium*: a genome-editing tool-delivery system. In *Agrobacterium Biology: From Basic Science to Biotechnology*; Gelvin, S. B., Ed.; Springer International Publishing: Cham, 2018; pp. 463–488 ISBN 978-3-030-03257-9.
3. Anand, A.; Jones, T. J. Advancing *Agrobacterium*-based crop transformation and genome modification technology for agricultural biotechnology. In *Current Topics in Microbiology and Immunology*; 2018 ISBN 0070-217X (Print)r0070-217X (Linking).
4. Buttimer, C.; McAuliffe, O.; Ross, R. P.; Hill, C.; O'Mahony, J.; Coffey, A. Bacteriophages and bacterial plant diseases. *Front. Microbiol.* **2017**, *8*, doi:10.3389/fmicb.2017.00034.
5. Dy, R. L.; Rigano, L. A.; Fineran, P. C. Phage-based biocontrol strategies and their application in agriculture and aquaculture. *Biochem. Soc. Trans.* **2018**, BST20180178, doi:10.1042/BST20180178.
6. Pratama, A. A.; van Elsas, J. D. The 'neglected' soil virome – potential role and impact. *Trends Microbiol.* **2018**, *26*, 649–662, doi:10.1016/j.tim.2017.12.004.

7. Domingo-Calap, P.; Delgado-Martínez, J. Bacteriophages: protagonists of a post-antibiotic era. *Antibiotics* **2018**, *7*, 66, doi:10.3390/antibiotics7030066.

8. Williamson, K. E.; Fuhrmann, J. J.; Wommack, K. E.; Radosevich, M. Viruses in soil ecosystems: an unknown quantity within an unexplored territory. *Annu. Rev. Virol.* **2017**, *4*, 201–219, doi:10.1146/annurev-virology-101416-041639.

9. Kropinski, A. M.; Van Den Bossche, A.; Lavigne, R.; Noben, J. P.; Babinger, P.; Schmitt, R. Genome and proteome analysis of 7-7-1, a flagellotropic phage infecting *Agrobacterium* sp H13-3. *Virol. J.* **2012**, *9*, doi:10.1186/1743-422X-9-102.

10. Attai, H.; Rimbey, J.; Smith, G. P.; Brown, P. J. B. Expression of a peptidoglycan hydrolase from lytic bacteriophages Atu_ph02 and Atu_ph03 triggers lysis of *Agrobacterium tumefaciens*. *Appl. Environ. Microbiol.* **2017**, *83*, e01498-17, doi:10.1128/AEM.01498-17.

11. Attai, H.; Boon, M.; Phillips, K.; Noben, J.-P.; Lavigne, R.; Brown, P. J. B. Larger than life: Isolation and genomic characterization of a jumbo phage that infects the bacterial plant pathogen, *Agrobacterium tumefaciens*. *Front. Microbiol.* **2018**, *9*, 1861, doi:10.3389/FMICB.2018.01861.

12. Miller, E. S.; Kutter, E.; Mosig, G.; Arisaka, F.; Kunisawa, T.; Ruger, W. Bacteriophage T4 genome. *Microbiol. Mol. Biol. Rev.* **2003**, *67*, 86–156, doi:10.1128/mmbr.67.1.86-156.2003.

13. Kutter, E.; Bryan, D.; Ray, G.; Brewster, E.; Blasdel, B.; Guttman, B. From host to phage metabolism: Hot tales of phage T4's takeover of *E. coli*. *Viruses* **2018**, *10*, 1–17, doi:10.3390/v10070387.

14. Poindexter, J. S. Biological properties and classification of the *Caulobacter* group. *Bacteriol. Rev.* **1964**, *28*, 231–295.

15. Santamaría, R. I.; Bustos, P.; Sepúlveda-Robles, O.; Lozano, L.; Rodríguez, C.; Fernández, J. L.; Juárez, S.; Kameyama, L.; Guarneros, G.; Dávila, G.; González, V. Narrow-host-range bacteriophages that infect *Rhizobium etli* associate with distinct genomic types. *Appl. Environ. Microbiol.* **2014**, *80*, 446–454, doi:10.1128/AEM.02256-13.

16. Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W. NIH Image to ImageJ : 25 years of image analysis HISTORICAL commentary NIH Image to ImageJ : 25 years of image analysis. *Nat. Methods* **2012**, *9*, 671–675, doi:10.1038/nmeth.2089.

17. Aziz, R. K.; Bartels, D.; Best, A. A.; DeJongh, M.; Disz, T.; Edwards, R. A.; Formsma, K.; Gerdes, S.; Glass, E. M.; Kubal, M.; Meyer, F.; Olsen, G. J.; Olson, R.; Osterman, A. L.; Overbeek, R. A.; McNeil, L. K.; Paarmann, D.; Paczian, T.; Parrello, B.; Pusch, G. D.; Reich, C.; Stevens, R.; Vassieva, O.; Vonstein, V.; Wilke, A.; Zagnitko, O. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **2008**, *9*, 75, doi:10.1186/1471-2164-9-75.

18. Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI-BLAST:a new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402, doi:10.1093/nar/25.17.3389.

19. Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **2001**, *305*, 567–580, doi:10.1006/jmbi.2000.4315.

20. Lowe, T. M.; Chan, P. P. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res* **2016**, *44*, W54-7, doi:10.1093/nar/gkw413.

21. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; Thierer, T.; Ashton, B.; Meintjes, P.; Drummond, A. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649, doi:10.1093/bioinformatics/bts199.

22. Darling, A. C. E.; Mau, B.; Blattner, F. R.; Perna, N. T. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **2004**, *14*, 1394–1403, doi:10.1101/gr.2289704.

23. Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; Mcgettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J. D.; Gibson, T. J.; Higgins, D. G. Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948, doi:10.1093/bioinformatics/btm404.

24. Guindon, S.; Dufayard, J.-F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **2010**, *59*, 307–321, doi:10.1093/sysbio/syq010.

25. Ackermann, H. W. Phage classification and characterization. *Methods Mol. Biol.* **2009**, *501*, 127–140, doi:10.1007/978-1-60327-164-6.

26. Johnson, M. C.; Tatum, K. B.; Lynn, J. S.; Brewer, T. E.; Lu, S.; Washburn, B. K.; Stroupe, M. E.; Jones, K. M. *Sinorhizobium meliloti* phage phiM9 defines a new

group of T4 superfamily phages with unusual genomic features but a common T=16 capsid. *J Virol* **2015**, *89*, 10945–10958, doi:10.1128/JVI.01353-15.

27. Habann, M.; Leiman, P. G.; Vandersteegen, K.; Van den Bossche, A.; Lavigne, R.; Shneider, M. M.; Bielmann, R.; Eugster, M. R.; Loessner, M. J.; Klumpp, J. *Listeria* phage A511, a model for the contractile tail machineries of SPO1-related bacteriophages. *Mol. Microbiol.* **2014**, *92*, 84–99, doi:10.1111/mmi.12539.

28. Nováček, J.; Šiborová, M.; Benešík, M.; Pantůček, R.; Doškař, J.; Plevka, P. Structure and genome release of Twort-like *Myoviridae* phage with a double-layered baseplate. *Proc. Natl. Acad. Sci.* **2016**, *113*, 9351–9356, doi:10.1073/pnas.1605883113.

29. Mueser, T. C.; Nossal, N. G.; Hyde, C. C. Structure of bacteriophage T4 RNase H, a 5’ to 3’ RNA-DNA and DNA-DNA exonuclease with sequence similarity to the RAD2 family of eukaryotic proteins. *Cell* **1996**, *85*, 1101–1112, doi:10.1016/S0092-8674(00)81310-0.

30. Kala, S.; Cumby, N.; Sadowski, P. D.; Hyder, B. Z.; Kanelis, V.; Davidson, A. R.; Maxwell, K. L. HNH proteins are a widespread component of phage DNA packaging machines. *Proc. Natl. Acad. Sci.* **2014**, *111*, 6022–6027, doi:10.1073/pnas.1320952111.

31. McMillan, S.; Edenberg, H. J.; Radany, E. H.; Friedberg, R. C.; Friedberg, E. C. *denV* gene of bacteriophage T4 codes for both pyrimidine dimer-DNA glycosylase and apyrimidinic endonuclease activities. *J. Virol.* **1981**, *40*, 211–223.

32. Sullivan, M. B.; Coleman, M. L.; Weigele, P.; Rohwer, F.; Chisholm, S. W. Three *Prochlorococcus* cyanophage genomes: signature features and ecological

interpretations. *PLoS Biol* **2005**, *3*, e144, doi:10.1371/journal.pbio.0030144.

33. Tanner, J. J.; Boechi, L.; Andrew McCammon, J.; Sobrado, P. Structure, mechanism, and dynamics of UDP-galactopyranose mutase. *Arch. Biochem. Biophys.* **2014**, *544*, 128–141, doi:10.1016/j.abb.2013.09.017.

34. Grynberg, Marcin and Godzik, A. NERD: a DNA processing-related domain present in the anthrax virulence plasmid, pXO1. *Trends Biochem. Sci.* **2004**, *29*, 103–106, doi:10.1016/j.tibs.2004.01.001.

35. Haakonsen, D. L.; Yuan, A. H.; Laub, M. T. The bacterial cell cycle regulator GcrA is a σ70 cofactor that drives gene expression from a subset of methylated promoters. *Genes Dev.* **2015**, *29*, 2272–2286, doi:10.1101/gad.270660.115.In.

36. Fioravanti, A.; Fumeaux, C.; Mohapatra, S. S.; Bompard, C.; Brilli, M.; Frandi, A.; Castric, V.; Villeret, V.; Viollier, P. H.; Biondi, E. G. DNA Binding of the cell cycle transcriptional regulator GcrA depends on N6-adenosine methylation in *Caulobacter crescentus* and other Alphaproteobacteria. *PLoS Genet.* **2013**, *9*, doi:10.1371/journal.pgen.1003541.

37. Gill, J. J.; Berry, J. D.; Russell, W. K.; Lessor, L.; Escobar-Garcia, D. A.; Hernandez, D.; Kane, A.; Keene, J.; Maddox, M.; Martin, R.; Mohan, S.; Thorn, A. M.; Russell, D. H.; Young, R. The *Caulobacter crescentus* phage phiCbK: Genomics of a canonical phage. *BMC Genomics* **2012**, *13*, 1–20, doi:10.1186/1471-2164-13-542.

38. Arndt, D.; Grant, J. R.; Marcu, A.; Sajed, T.; Pon, A.; Liang, Y.; Wishart, D. S. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* **2016**, *44*, W16–W21, doi:10.1093/nar/gkw387.

39. Barragán, M. J. L.; Bláquez, B.; Zamarro, M. T.; Mancheño, J. M.; García, J. L.; Díaz, E.; Carmona, M. BzdR, a repressor that controls the anaerobic catabolism of benzoate in *Azoarcus* sp. CIB, is the first member of a new subfamily of transcriptional regulators. *J. Biol. Chem.* **2005**, *280*, 10683–10694, doi:10.1074/jbc.M412259200.

40. Knight, K. L.; Bowie, J. U.; Vershon, A. K.; Kelley, R. D.; Sauer, R. T.; Vershong, A. K.; Sauer, R. T.; Vershon, A. K.; Kelley, R. D.; Sauer, R. T.; Vershong, A. K.; Sauer, R. T. The Arc and Mnt repressors. *J. Biol. Chem.* **1989**, *264*, 3639–3642.

41. Trempy, J. E.; Kirby, J. E.; Gottesman, S. Alp suppression of Lon: dependence on the *slpA* gene. *J. Bacteriol.* **1994**, *176*, 2061–2067, doi:10.1128/jb.176.7.2061-2067.1994.

42. Hatfull, G. F.; Hendrix, R. W. Bacteriophages and their genomes. *Curr Opin Virol* **2011**, *1*, 298–303, doi:10.1016/j.coviro.2011.06.009.

43. Gill, J. J.; Summer, E. J.; Russell, W. K.; Cologna, S. M.; Carlile, T. M.; Fuller, A. C.; Kitsopoulos, K.; Mebane, L. M.; Parkinson, B. N.; Sullivan, D.; Carmody, L. A.; Gonzalez, C. F.; LiPuma, J. J.; Young, R. Genomes and characterization of phages Bcep22 and BcepIL02, founders of a novel phage type in *Burkholderia cenocepacia*. *J Bacteriol* **2011**, *193*, 5300–5313, doi:10.1128/JB.05287-11.

44. Iida, S.; Streiff, M. B.; Bickle, T. A.; Arber, W. Two DNA antirestriction systems of bacteriophage P1, darA, and darB: characterization of darA-phages. *Virology* **1987**, doi:10.1016/0042-6822(87)90324-2.

45. Łobocka, M. B.; Rose, D. J.; Guy Plunkett, I.; Rusin, M.; Samojedny, A.; Lehnherr, H. ̈rg; Yarmolinsky, M. B.; Blattner, F. R. Genome of bacteriophage P1.

J. Bacteriol. **2004**, *186*, 7032–7068, doi:10.1128/JB.186.21.7032.

46. Piya, D.; Vara, L.; Russell, W. K.; Young, R.; Gill, J. J. The multicomponent antirestriction system of phage P1 is linked to capsid morphogenesis. *Mol. Microbiol.* **2017**, *105*, 399–412, doi:10.1111/mmi.13705.

47. Watson, B.; Currier, T. C.; Gordon, M. P.; Chilton, M.-D.; Nester, E. W. Plasmid required for virulance of *Agrobacterium tumefaciens*. *J. Bacteriol* **1975**, *123*, 255–264.

48. Luo, Z. Q.; Clemente, T. E.; Farrand, S. K. Construction of a derivative of *Agrobacterium tumefaciens* C58 that does not mutate to tetracycline resistance. *Mol. Plant. Microbe. Interact.* **2001**, *14*, 98–103, doi:10.1094/MPMI.2001.14.1.98.

49. Bush, A. L.; Pueppke, S. G. Characterization of an unusual new *Agrobacterium tumefaciens* strain from *Chrysanthemum morifolium* ram. *Appl. Environ. Microbiol.* **1991**, *57*, 2468–2472.

50. Slater, S. C.; Goldman, B. S.; Goodner, B.; Setubal, J. C.; Farrand, S. K.; Nester, E. W.; Burr, T. J.; Banta, L.; Dickerman, A. W.; Paulsen, I.; Otten, L.; Suen, G.; Welch, R.; Almeida, N. F.; Arnold, F.; Burton, O. T.; Du, Z.; Ewing, A.; Godsy, E.; Heisel, S.; Houmiel, K. L.; Jhaveri, J.; Lu, J.; Miller, N. M.; Norton, S.; Chen, Q.; Phoolcharoen, W.; Ohlin, V.; Ondrusek, D.; Pride, N.; Stricklin, S. L.; Sun, J.; Wheeler, C.; Wilson, L.; Zhu, H.; Wood, D. W. Genome sequences of three *Agrobacterium* biovars help elucidate the evolution of multichromosome genomes in bacteria. *J. Bacteriol.* **2009**, *191*, 2501–2511, doi:10.1128/JB.01779-08.

51. Nierman, W. C.; Feldblyum, T. V; Laub, M. T.; Paulsen, I. T.; Nelson, K. E.; Eisen, J. A.; Heidelberg, J. F.; Alley, M. R.; Ohta, N.; Maddock, J. R.; Potocka, I.;

Nelson, W. C.; Newton, A.; Stephens, C.; Phadke, N. D.; Ely, B.; DeBoy, R. T.; Dodson, R. J.; Durkin, A. S.; Gwinn, M. L.; Haft, D. H.; Kolonay, J. F.; Smit, J.; Craven, M. B.; Khouri, H.; Shetty, J.; Berry, K.; Utterback, T.; Tran, K.; Wolf, A.; Vamathevan, J.; Ermolaeva, M.; White, O.; Salzberg, S. L.; Venter, J. C.; Shapiro, L.; Fraser, C. M. Complete genome sequence of *Caulobacter crescentus*. *Proc Natl Acad Sci U S A* **2001**, *98*, 4136–4141, doi:10.1073/pnas.061029298\r061029298 [pii].

Tables

Table 4-1. Bacterial strains used in this study.

Strain or plasmid	**Relevant characteristics**	**Growth medium**	**Reference or source**
A. tumefaciens strains			
C58	Nopaline type strain; pTiC58; pAtC58	LB	[47]
EHA105	C58 derived, succinamopine strain, T-DNA deletion derivative of pTiBo542	LB	MU plant transformation core facility
EHA101	C58 derived, nopaline strain, T-DNA deletion derivative of pTiBo542	LB	MU plant transformation core facility
GV3101	C58 derived, nopaline strain	LB	MU plant transformation core facility
NTL4	C58 derived, nopaline-agrocinopine strain, Δ*tetRA*	LB	[48]
AGL-1	C58 derived, succinamopine strain, T-DNA deletion derivative of pTiBo542 ΔrecA	LB	MU plant transformation core facility
LBA4404	Ach5 derived, octopine strain, T-DNA deletion derivative of pTiAch5	YM	MU plant transformation core facility
Chry5	Succinamopine strain, pTiChry5	LB	[49]

Other bacterial strains			
A. vitis S4	Vitopine strain, pTiS4, pSymA, pSymB	Potato dextrose	[50]
Caulobacter crescentus CB15	Alphaproteobacterium	PYE	[51]
Escherichia coli DH5α	Gammaproteobacterium	LB	Life Technologies

Table 4-2. Host range testing of Atu_ph04 and Atu_ph08. (S) indicates strain is susceptible to phage infection, (I) indicates strain has an intermediate phenotype and is only somewhat susceptible at a reduced MOI, and (R) indicates that the strain is resistant to phage infection.

Strain	**Susceptibility to Phage**	
	Atu_ph04	**Atu_ph08**
A. tumefaciens C58	S	S
A. tumefaciens EHA105	S	S
A. tumefaciens EHA101	S	S
A. tumefaciens GV3101	S	S
A. tumefaciens NTL4	S	S
A. tumefaciens AGL-1	R	I
A. tumefaciens LBA4404	I	I
A. tumefaciens Chry5	R	R

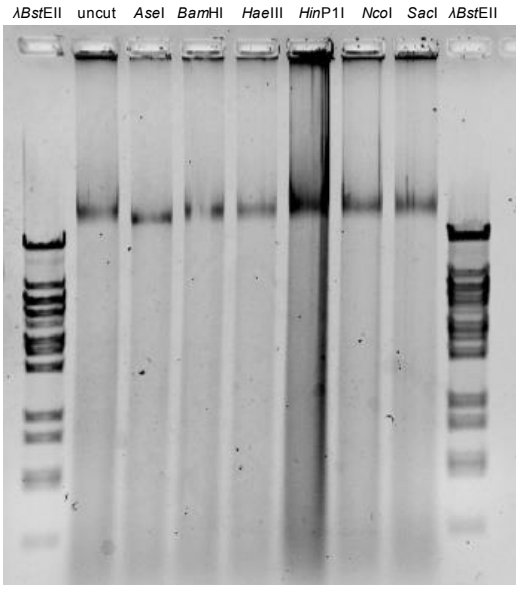
A. vitis S4	R	R
C. crescentus CB15	R	R
E. coli DH5α	R	R

Table 4-3. Summary of key genomic features of Atu_ph04 and Atu_ph08.

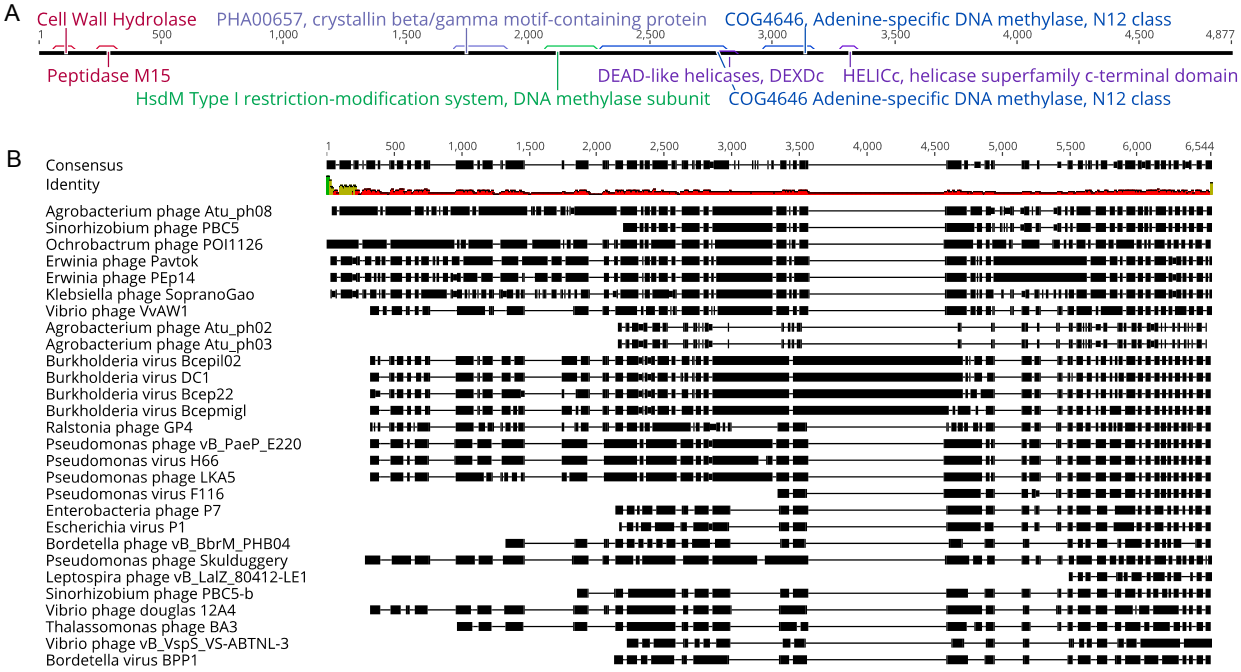
Phage	**Genome length (bp)**	**G+C content (%)**	**Number of ORFs**	**Number of hypothetical proteins**	**Number of ORFs with predicted functions**	**Number of ORFans**	**Number of tRNAs**
Atu_ph04	143,349	49.4	223	67	73	83	1
Atu_ph08	59,034	59.7	75	43	32	3	0

SUPPLEMENTARY FIGURES

Supplementary Figure 4-S1: Restriction fragment analysis of digested Atu_ph04 genomic DNA loaded onto a 0.7% agarose gel.



Supplementary Figure 4-S2: Analysis of DarB-like protein in Atu_ph08. (A) Domain structure of DarB-like protein in Atu_ph08. (B) ClustalW alignment of DarB-like proteins in other T7-like phages. Dark black blocks indicate regions of synteny.



SUPPLEMENTARY TABLES

Supplementary Table 4-S1: Atu_ph04 genes organized by predicted function.

CDS #	RAST annotated function	Updated assigned function	Length (bp)	Category
1	hypothetical protein CDS	putative T7-like tail fiber protein	186	Structural
2	hypothetical protein CDS		600	Hypothetical
3	hypothetical protein CDS		321	Hypothetical
4	hypothetical protein CDS		246	Hypothetical
5	hypothetical protein CDS		210	Hypothetical
6	UDP-galactopyranose mutase (EC 5.4.99.9) CDS		1173	Other/Bacterial
7	hypothetical protein CDS		315	Hypothetical
8	hypothetical protein CDS		159	Hypothetical
9	hypothetical protein CDS		354	Hypothetical
10	N-acetylmuramoyl-L-alanine amidase (EC 3.5.1.28) CDS		942	Lysis
11	hypothetical protein CDS		216	Hypothetical
12	hypothetical protein CDS		516	Hypothetical
13	hypothetical protein CDS		366	Hypothetical
14	Phage protein CDS		351	Hypothetical
15	hypothetical protein CDS		399	Hypothetical
16	hypothetical protein CDS		213	Hypothetical
17	Phosphohydrolase (MutT/nudix family protein) CDS		588	Nucleotide metabolism
18	hypothetical protein CDS		600	Hypothetical

19	hypothetical protein CDS		249	Hypothetical
20	hypothetical protein CDS		309	Hypothetical
21	hypothetical protein CDS		546	Hypothetical
22	hypothetical protein CDS		198	Hypothetical
23	hypothetical protein CDS	putative glutaredoxin	282	Nucleotide metabolism
24	Ribonucleotide reductase of class Ia (aerobic), alpha subunit (EC 1.17.4.1) CDS		1662	Nucleotide metabolism
25	Ribonucleotide reductase of class Ia (aerobic), beta subunit (EC 1.17.4.1) CDS		1053	Nucleotide metabolism
26	Phage-associated DNA primase (EC 2.7.7.-) #gp61 CDS		1041	DNA replication
27	hypothetical protein CDS		339	Hypothetical
28	hypothetical protein CDS		402	Hypothetical
29	hypothetical protein CDS		354	Hypothetical
30	hypothetical protein CDS		297	Hypothetical
31	hypothetical protein CDS		510	Hypothetical
32	hypothetical protein CDS		480	Hypothetical
33	hypothetical protein CDS	putative uracil DNA glycosylase	468	DNA repair
34	hypothetical protein CDS		273	Hypothetical
35	hypothetical protein CDS		405	Hypothetical
36	hypothetical protein CDS		255	Hypothetical
37	hypothetical protein CDS		219	Hypothetical

38	hypothetical protein CDS		291	Hypothetical
39	hypothetical protein CDS		207	Hypothetical
40	Phage DNA end protector during packaging CDS		627	DNA associated
41	hypothetical protein CDS	putative baseplate tail tube initiator	537	Structural
42	Phage baseplate hub subunit # T4-like gp26 CDS		702	Structural
43	hypothetical protein CDS	putative baseplate hub assembly catalyst	177	Structural
44	hypothetical protein CDS		381	Hypothetical
45	hypothetical protein CDS		1812	Hypothetical
46	hypothetical protein CDS		273	Hypothetical
47	Single stranded DNA-binding protein, phage-associated CDS		1011	DNA replication
48	hypothetical protein CDS		555	Hypothetical
49	hypothetical protein CDS		312	Hypothetical
50	hypothetical protein CDS		288	Hypothetical
51	hypothetical protein CDS		540	Hypothetical
52	hypothetical protein CDS	putative homing endonuclease	654	DNA replication
53	Phage terminase, large subunit #T4-like phage Gp17 CDS		1734	Structural
54	T4-like phage baseplate hub + tail lysozyme CDS		2562	Structural
55	hypothetical protein CDS		126	Hypothetical
56	hypothetical protein CDS		285	Hypothetical
57	hypothetical protein CDS		507	Hypothetical
58	T4-like phage protein, T4 GC1630 CDS	putative homing endonuclease	459	DNA replication

59	probable ATP-dependent helicase CDS		1614	DNA replication
60	Phage endonuclease CDS	denV endonuclease V, N-glycosylase UV repair enzyme	318	DNA repair
61	hypothetical protein CDS		264	Hypothetical
62	hypothetical protein CDS		414	Hypothetical
63	Phage ribonuclease H (EC 3.1.26.4) #T4-like phage Rnh #T4 GC0870 CDS		909	DNA replication
64	hypothetical protein CDS		720	Hypothetical
65	hypothetical protein CDS		438	Hypothetical
66	hypothetical protein CDS	tail completion and sheath stabilizer protein	522	Structural
67	Single stranded DNA-binding protein, phage-associated #T4-like phage Gp32 CDS		486	DNA replication
68	hypothetical protein CDS	putative homing endonuclease	765	DNA replication
69	Phage head completion protein CDS		459	Structural
70	Phage tail sheath CDS		2517	Structural
71	hypothetical protein CDS	putative tail tube monomer	627	Structural
72	Phage portal vertex of the head #T4-like phage Gp20 CDS		1575	Structural
73	hypothetical protein CDS		153	Hypothetical
74	Phage prohead core protein CDS		666	Structural
75	MJ0042 family finger-like protein CDS		1131	Hypothetical
76	Phage major capsid protein of Caudovirales CDS		1332	Structural
77	hypothetical protein CDS		183	Hypothetical

78	DNA helicase, phage-associated CDS		1503	DNA replication
79	hypothetical protein CDS		324	Hypothetical
80	hypothetical protein CDS	putative terminase DNA packaging enzyme small subunit	504	DNA associated
81	hypothetical protein CDS		186	Hypothetical
82	hypothetical protein CDS	putative base plate wedge subunit	387	Structural
83	gp6 baseplate wedge subunit CDS		1782	Structural
84	hypothetical protein CDS	baseplate wedge subunit-like protein	3114	Structural
85	hypothetical protein CDS		1227	Hypothetical
86	Phage protein CDS		1545	Hypothetical
87	Phosphate starvation-inducible protein PhoH, predicted ATPase CDS		783	Other/Bacterial
88	hypothetical protein CDS		348	Hypothetical
89	RNA polymerase ECF-type sigma factor CDS		483	Transcription
90	hypothetical protein CDS		294	Hypothetical
91	hypothetical protein CDS		345	Hypothetical
92	hypothetical protein CDS		150	Hypothetical
93	hypothetical protein CDS	putative baseplate tail tube cap	780	Structural
94	hypothetical protein CDS	putative base plate wedge component	579	Structural
95	hypothetical protein CDS		414	Hypothetical
96	Phage recombination protein CDS		1164	DNA associated
97	hypothetical protein CDS	DNA primase/helicase	1434	DNA replication

98	hypothetical protein CDS	putative holliday junction resolvase	624	DNA associated
99	hypothetical protein CDS		435	Hypothetical
100	hypothetical protein CDS	5'-deoxynucleotidase	621	DNA associated
101	hypothetical protein CDS	putative deoxynucleotide monophosphate kinase	786	DNA associated
102	hypothetical protein CDS		363	Hypothetical
103	hypothetical protein CDS		465	Hypothetical
104	Deoxycytidylate 5-hydroxymethyltransferase (EC 2.1.2.8) CDS		744	DNA modification
105	hypothetical protein CDS		1017	Hypothetical
106	hypothetical protein CDS		642	Hypothetical
107	hypothetical protein CDS		594	Hypothetical
108	hypothetical protein CDS		423	Hypothetical
109	hypothetical protein CDS		378	Hypothetical
110	Phage DNA topoisomerase large subunit (EC 5.99.1.3) #T4-like gp60 #T4 GC1464 CDS		1938	DNA replication
111	hypothetical protein CDS		306	Hypothetical
112	hypothetical protein CDS		552	Hypothetical
113	Topoisomerase IV subunit A (EC 5.99.1.-) CDS		1365	DNA replication
114	hypothetical protein CDS		120	Hypothetical
115	hypothetical protein CDS		336	Hypothetical
116	COG1896: Predicted hydrolases of HD superfamily CDS		723	Lysis
117	hypothetical protein CDS		336	Hypothetical

118	hypothetical protein CDS		360	Hypothetical
119	hypothetical protein CDS	sigma factor for late transcription	513	Transcription
120	hypothetical protein CDS	recombination endonuclease subunit	951	DNA associated
121	Phage recombination-related endonuclease Gp46 CDS		1647	DNA associated
122	hypothetical protein CDS	sliding clamp DNA polymerase accessory protein	729	DNA replication
123	Replication factor C small subunit CDS	Sliding clamp loader subunit	948	DNA replication
124	hypothetical protein CDS	putative clamp loader subunit	405	DNA replication
125	hypothetical protein CDS		177	Hypothetical
126	hypothetical protein CDS		159	Hypothetical
127	hypothetical protein CDS		267	Hypothetical
128	hypothetical protein CDS		690	Hypothetical
129	hypothetical protein CDS		258	Hypothetical
130	hypothetical protein CDS		381	Hypothetical
131	hypothetical protein CDS		210	Hypothetical
132	hypothetical protein CDS		234	Hypothetical
133	DNA polymerase (EC 2.7.7.7), phage-associated #T4-like phage gp43 #T4 GC0178 CDS		2640	DNA replication
134	hypothetical protein CDS		768	Hypothetical
135	hypothetical protein CDS		120	Hypothetical
136	hypothetical protein CDS		303	Hypothetical
137	hypothetical protein CDS		546	Hypothetical

138	DNA ligase, phage-associated CDS		1257	DNA replication
139	hypothetical protein CDS		336	Hypothetical
140	hypothetical protein CDS		213	Hypothetical
141	hypothetical protein CDS		189	Hypothetical
142	hypothetical protein CDS		201	Hypothetical
143	hypothetical protein CDS		594	Hypothetical
144	hypothetical protein CDS		165	Hypothetical
145	Phage protein CDS	predicted alternative thymidylate synthase	984	Nucleotide metabolism
146	hypothetical protein CDS		282	Hypothetical
147	hypothetical protein CDS		165	Hypothetical
148	hypothetical protein CDS	putative GT1 glycosyltransferase protein	1164	Nucleotide metabolism
149	hypothetical protein CDS		405	Hypothetical
150	FIG00451076: hypothetical protein CDS		255	Hypothetical
151	hypothetical protein CDS	putative homing endonuclease	714	DNA associated
152	hypothetical protein CDS	putative ParB-like nuclease domain containing protein	450	DNA associated
153	hypothetical protein CDS		474	Hypothetical
154	hypothetical protein CDS		216	Hypothetical
155	hypothetical protein CDS		303	Hypothetical
156	hypothetical protein CDS		303	Hypothetical

157	hypothetical protein CDS		267	Hypothetical
158	hypothetical protein CDS		183	Hypothetical
159	hypothetical protein CDS		246	Hypothetical
160	hypothetical protein CDS		501	Hypothetical
161	hypothetical protein CDS		156	Hypothetical
162	hypothetical protein CDS		357	Hypothetical
163	Glycine-rich cell wall structural protein 1.8 precursor CDS		657	Other/Bacterial
164	hypothetical protein CDS		480	Hypothetical
165	hypothetical protein CDS		489	Hypothetical
166	hypothetical protein CDS		279	Hypothetical
167	hypothetical protein CDS		555	Hypothetical
168	hypothetical protein CDS		135	Hypothetical
169	hypothetical protein CDS	DNA repair exonuclease	489	DNA repair
170	hypothetical protein CDS		267	Hypothetical
171	hypothetical protein CDS		207	Hypothetical
172	hypothetical protein CDS		195	Hypothetical
173	hypothetical protein CDS		198	Hypothetical
174	hypothetical protein CDS		201	Hypothetical
175	hypothetical protein CDS		417	Hypothetical
176	hypothetical protein CDS		363	Hypothetical
177	hypothetical protein CDS		300	Hypothetical
178	hypothetical protein CDS		198	Hypothetical
179	hypothetical protein CDS		225	Hypothetical

180	hypothetical protein CDS	von Willebrand factor type A domain containing protein	702	Other/Bacterial
181	hypothetical protein CDS		225	Hypothetical
182	hypothetical protein CDS		303	Hypothetical
183	hypothetical protein CDS		207	Hypothetical
184	hypothetical protein CDS		276	Hypothetical
185	Polymerase epsilon subunit CDS		858	DNA associated
186	hypothetical protein CDS		285	Hypothetical
187	hypothetical protein CDS		291	Hypothetical
188	hypothetical protein CDS		264	Hypothetical
189	dCMP deaminase (EC 3.5.4.12); Late competence protein ComEB CDS		600	DNA associated
190	hypothetical protein CDS		510	Hypothetical
191	hypothetical protein CDS		210	Hypothetical
192	hypothetical protein CDS		234	Hypothetical
193	hypothetical protein CDS		462	Hypothetical
194	Phage protein CDS	putative exonuclease	825	DNA associated
195	hypothetical protein CDS	DNA primase	531	DNA replication
196	hypothetical protein CDS		282	Hypothetical
197	hypothetical protein CDS		363	Hypothetical
198	hypothetical protein CDS		372	Hypothetical
199	hypothetical protein CDS		231	Hypothetical

200	hypothetical protein CDS		234	Hypothetical
201	hypothetical protein CDS		261	Hypothetical
202	hypothetical protein CDS		423	Hypothetical
203	hypothetical protein CDS		300	Hypothetical
204	hypothetical protein CDS		330	Hypothetical
205	hypothetical protein CDS		744	Hypothetical
206	hypothetical protein CDS		609	Hypothetical
207	hypothetical protein CDS		231	Hypothetical
208	hypothetical protein CDS		201	Hypothetical
209	hypothetical protein CDS		435	Hypothetical
210	hypothetical protein CDS		510	Hypothetical
211	hypothetical protein CDS		579	Hypothetical
212	hypothetical protein CDS		1347	Hypothetical
213	T4-like phage baseplate hub + tail lysozyme CDS		876	Structural
214	hypothetical protein CDS		297	Hypothetical
215	Phage neck protein #Gp13 CDS		756	Structural
216	Gp14 neck protein CDS		843	Structural
217	hypothetical protein CDS		360	Hypothetical
218	Phage tail assembly CDS		882	Structural
219	hypothetical protein CDS	baseplate wedge	1461	Structural
220	Phage virulence-associated VriC protein CDS	VrlC protein	5088	Structural
221	hypothetical protein CDS		207	Hypothetical
222	Phage tail fibers CDS		1791	Structural

223	FIG00920814: hypothetical protein CDS		1149	Hypothetical

Supplementary Table 4-S2: Comparative analysis of Atu_ph04 gene products with related phages.

gp	Functional annotation	vB_RleM_P10VF	phiM9	Cr30	syn9	syn30	syn33	T4	Atu_ph07	phiM12	CcrColossus	phiN3	KVP40	Melville
6	UDP-galactopyranose mutase	3.00E-125	1.00E-128											
9	hypothetical	9.00E-22												
10	N-acetylmuramoyl-L-alanine amidase	4.00E-90	1.00E-87											
11	hypothetical	6.00E-12	5.00E-14											
12	hypothetical	1.00E-31	2.00E-38											
13	hypothetical		4.00E-27											
14	hypothetical	1.00E-26	1.00E-21											
15	hypothetical	9.00E-17								4.00E-11				
17	Phosphohydrolase (MutT/nudix family protein)			7.00E-16						5.00E-23	9.00E-24	6.00E-23		
20	hypothetical										2.00E-11			
23	hypothetical	2.00E-23	1.00E-20											
24	Ribonucleotide reductase of class Ia (aerobic), alpha subunit	0	0		3.00E-17	2.00E-16	2.00E-16	9.00E-17	0		5.00E-70		2.00E-17	1.00E-16
25	Ribonucleotide reductase of class Ia (aerobic), beta subunit	4.00E-156	1.00E-162						3.00E-104					
26	Phage-associated DNA primase	3.00E-87	2.00E-108		4.00E-42	1.00E-41	8.00E-34	2.00E-37		1.00E-37		8.00E-37	7.00E-22	9.00E-34
27	hypothetical	1.00E-32	1.00E-43											
28	hypothetical	6.00E-29	3.00E-36											
29	hypothetical	9.00E-10	1.00E-23											
30	hypothetical	3.00E-24	4.00E-21											
33	hypothetical	1.00E-17	1.00E-19											
36	hypothetical	1.00E-06												
40	Phage DNA end protector during packaging	3.00E-102	3.00E-100	2.00E-17				6.00E-29		2.00E-25				3.00E-28
41	putative baseplate tail tube initiator	7.00E-56	3.00E-67		3.00E-06		6.00E-08						2.00E-08	
42	Phage baseplate hub subunit	1.00E-80	3.00E-84										3.00E-05	9.00E-07

43	hypothetical	3.00E-18	5.00E-13	3.00E-04										
44	hypothetical	8.00E-17	2.00E-19											
45	hypothetical	2.00E-122	1.00E-141											
46	hypothetical		7.00E-07											
47	Single stranded DNA-binding protein	6.00E-125	4.00E-105	1.00E-25		3.00E-26	7.00E-28			3.00E-24		2.00E-24	5.00E-16	
48	hypothetical	7.00E-12	2.00E-16											
49	hypothetical	5.00E-20	7.00E-14											
50	hypothetical	5.00E-16	1.00E-14											
51	hypothetical	4.00E-66	9.00E-70											
52	hypothetical	5.00E-22	2.00E-36	6.00E-10				2.00E-13		2.00E-07		3.00E-07		
53	Phage terminase, large subunit	0	0	2.00E-114	9.00E-114	7.00E-117	1.00E-112	2.00E-105	3.00E-33	5.00E-120		8.00E-120	2.00E-99	2.00E-108
54	T4-like phage baseplate hub + tail lysozyme	0	0	8.00E-16				6.00E-14		6.00E-18		2.00E-17	7.00E-20	7.00E-16
58	putative homing endonuclease	4.00E-79	2.00E-66							3.00E-72	6.00E-60	2.00E-71	2.00E-57	7.00E-60
59	probable ATP-dependent helicase	0	0											
60	Phage endonuclease	1.00E-31	4.00E-37					2.00E-18					2.00E-07	
62	hypothetical	8.00E-36	2.00E-32						7.00E-20		1.00E-24			
63	Phage ribonuclease H	2.00E-135	8.00E-140	1.00E-28						3.00E-32				
64	hypothetical	6.00E-53	3.00E-56											
65	hypothetical	9.00E-16	2.00E-27											
66	tail completion and sheath stabilizer protein	3.00E-55	9.00E-55					2.00E-08						7.00E-12
67	Single stranded DNA-binding protein	4.00E-52	5.00E-48			1.00E-14	3.00E-14	3.00E-08		9.00E-15				1.00E-11
68	hypothetical		2.00E-23											
69	Phage head completion protein	1.00E-76	2.00E-77	9.00E-08	5.00E-36	6.00E-35	6.00E-37	5.00E-34	1.00E-25	4.00E-36			2.00E-40	1.00E-41
70	Phage tail sheath	0	0	6.00E-47	2.00E-46	1.00E-43	4.00E-42	3.00E-31		4.00E-61		3.00E-61	3.00E-34	2.00E-33
71	putative tail tube monomer	5.00E-87	2.00E-98		3.00E-14	2.00E-16	7.00E-18							
72	Phage portal vertex of the head	0	0	9.00E-98	3.00E-109	7.00E-113	3.00E-76	5.00E-97	7.00E-19	3.00E-87		4.00E-87	4.00E-86	3.00E-93

74	Phage prohead core protein	7.00E-93	7.00E-94	3.00E-26	8.00E-34	5.00E-34	3.00E-33	1.00E-14		3.00E-38		3.00E-38	2.00E-30	2.00E-16
75	MJ0042 family finger-like protein	2.00E-52	4.00E-55											
76	Phage major capsid protein	0	0	2.00E-71	1.00E-83	4.00E-81	9.00E-85	5.00E-48	5.00E-31	1.00E-92		1.00E-92	2.00E-55	3.00E-48
78	DNA helicase	0	0	2.00E-78	1.00E-68	3.00E-66	7.00E-65	2.00E-69	3.00E-18	2.00E-67		4.00E-66	2.00E-72	1.00E-71
80	putative terminase DNA packaging enzyme small subunit	6.00E-48	2.00E-43											
82	putative base plate wedge subunit	5.00E-58	2.00E-57											
83	baseplate wedge subunit	0	0	2.00E-54	5.00E-46	2.00E-46	1.00E-43			1.00E-55		2.00E-56		
84	baseplate wedge subunit-like protein	0	0							5.00E-06		5.00E-06		
85	hypothetical		0.001											
87	Phosphate starvation-inducible protein PhoH, predicted ATPase	2.00E-99	2.00E-111	1.00E-23	1.00E-21	2.00E-21	1.00E-22		2.00E-54		8.00E-29	1.00E-17	3.00E-29	
92	hypothetical	1.00E-05												
93	putative baseplate tail tube cap	5.00E-123	5.00E-135											
94	putative base plate wedge component	2.00E-60	4.00E-66											
95	hypothetical	2.00E-28	2.00E-27											
96	Phage recombination protein	0	0	7.00E-55				2.00E-71	5.00E-17	8.00E-65		5.00E-65	5.00E-72	5.00E-71
97	DNA primase/helicase	0	0	6.00E-79	4.00E-83	5.00E-84	1.00E-84	2.00E-56	1.00E-27	9.00E-76		1.00E-74	1.00E-68	2.00E-56
98	putative holliday junction resolvase	3.00E-63	8.00E-51	1.00E-13										
100	hypothetical	1.00E-70	1.00E-59											
101	putative deoxynucleotide monophosphate kinase	6.00E-58	4.00E-66											
103	hypothetical	4.00E-71	3.00E-65											
104	Deoxycytidylate 5-hydroxymethyltransferase	9.00E-79	2.00E-85											
105	hypothetical	5.00E-89	9.00E-91											
106	hypothetical	3.00E-10	7.00E-20											

107	hypothetical			3.00E-09			8.00E-16							
109	hypothetical	1.00E-28												
110	Phage DNA topoisomerase large subunit	0	0					5.00E-26	3.00E-146				2.00E-49	5.00E-53
113	Topoisomerase IV subunit A	0	0					8.00E-33	1.00E-98				2.00E-37	3.00E-31
119	sigma factor for late transcription	1.00E-47	2.00E-51											
120	recombination endonuclease subunit	1.00E-67	4.00E-70											
121	Phage recombination-related endonuclease	6.00E-164	2.00E-173	9.00E-56	7.00E-59	1.00E-59	5.00E-62	6.00E-65		4.00E-59		8.00E-59	5.00E-28	1.00E-59
122	sliding clamp DNA polymerase accessory protein	6.00E-58	8.00E-70											
123	Replication factor C small subunit	5.00E-128	4.00E-117					7.00E-05	2.00E-31				7.00E-51	6.00E-04
124	putative clamp loader subunit	4.00E-50	4.00E-48	3.00E-09	5.00E-05	3.00E-05								
129	hypothetical	2.00E-36	2.00E-30											
131	hypothetical		2.00E-05											
133	DNA polymerase (EC 2.7.7.7), phage-associated	0	0	3.00E-89	2.00E-109	6.00E-103	1.00E-105	2.00E-81	8.00E-23	8.00E-104		9.00E-104	6.00E-74	2.00E-80
134	hypothetical	2.00E-90	5.00E-102											
138	DNA ligase, phage-associated	8.00E-145	1.00E-153										6.00E-26	
139	hypothetical	2.00E-21	3.00E-14											
143	hypothetical	1.00E-43	7.00E-38										6.00E-13	
145	predicted alternative thymidylate synthase										3.00E-39			
148	putative GT1 glycosyltransferase protein	2.00E-127	1.00E-118											
150	FIG00451076: hypothetical protein	7.00E-22	8.00E-20	1.00E-22										
151	putative homing endonuclease		4.00E-41											
152	putative ParB-like nuclease domain containing protein	1.00E-22	5.00E-19											
153	hypothetical	3.00E-61	2.00E-59											
157	hypothetical	2.00E-08												

159	hypothetical	4.00E-12	3.00E-13						3.00E-23					
160	hypothetical	8.00E-07												
164	hypothetical	7.00E-13	2.00E-23											
165	hypothetical	3.00E-30	4.00E-37											
166	hypothetical		0.001						2.00E-04					
167	hypothetical	1.00E-09												
168	hypothetical	4.00E-04												
169	DNA repair exonuclease	1.00E-25	2.00E-33						3.00E-16	2.00E-31		1.00E-31		
180	von Willebrand factor type A domain containing protein	1.00E-81												
182	hypothetical								6.00E-16	7.00E-27			4.00E-04	
185	Polymerase epsilon subunit	7.00E-112	3.00E-112						7.00E-28					
188	hypothetical	5.00E-14												
189	dCMP deaminase (EC 3.5.4.12); Late competence protein ComEB	1.00E-49	1.00E-49					9.00E-10					3.00E-14	4.00E-11
190	hypothetical	2.00E-09	1.00E-10											
192	hypothetical	0.001	7.00E-04											
194	putative exonuclease	7.00E-112	1.00E-100	1.00E-26	6.00E-22	5.00E-21	1.00E-21		2.00E-14			3.00E-22		
201	hypothetical	2.00E-07	7.00E-07											
202	hypothetical	3.00E-27												
206	hypothetical	6.00E-14							2.00E-12					
207	hypothetical	2.00E-06												
209	hypothetical	2.00E-11	3.00E-20											
212	hypothetical	0	0				2.00E-06							
213	T4-like phage baseplate hub + tail lysozyme	3.00E-147	3.00E-141	2.00E-18	8.00E-13	3.00E-13	4.00E-14	3.00E-15		4.00E-12		4.00E-12		9.00E-16
215	Phage neck protein	4.00E-106	3.00E-106	3.00E-25	4.00E-24	7.00E-20		7.00E-09		5.00E-24		2.00E-24	5.00E-04	4.00E-06
216	Gp14 neck protein	3.00E-122	7.00E-109	1.00E-14	1.00E-08					2.00E-15		3.00E-15		
217	hypothetical	8.00E-37	6.00E-31											
218	Phage tail assembly	1.00E-82	3.00E-106	7.00E-24	9.00E-24	6.00E-19	2.00E-26			2.00E-18			5.00E-14	

219	baseplate wedge	4.00E-147	2.00E-170	7.00E-09										
220	Phage virulence-associated VriC protein	0	0	3.00E-28	6.00E-41	4.00E-43	2.00E-42			9.00E-69		4.00E-67		
221	hypothetical		7.00E-17	9.00E-05										
222	Phage tail fibers	8.00E-175	2.00E-167											2.00E-06
223	FIG00920814: hypothetical protein	5.00E-66	7.00E-76											
	TOTAL	**115**	**109**	**32**	**23**	**23**	**24**	**25**	**22**	**30**	**7**	**25**	**29**	**26**

Supplementary Table 4-S3: T4 core proteins found in Atu_ph04. *Atu_ph04 matches with E-values above 1E-10 are considered "yes" matches and those between 1E-10 and 1E-03 are "unresolved." Matches with E-values lower than 1E-03 were not considered significant.

T4 protein	T4 protein function	Match in Atu_ph04*	Identity (%)	E-value	Query cover (%)	Atu_ph04 protein name	Atu_ph04 gp #
Phage morphogenesis							
gp4	head completion protein	yes	39	1e-35	96	Head completion protein	69
gp5	baseplate lysozyme hub component	yes; two	45	2e-16	17	T4-like phage baseplate hub + tail lysozyme	213
"			30	1e-15	32	T4-like phage baseplate hub + tail lysozyme	54
gp13	head completion protein	unresolved	21	6e-08	99	Neck protein	215
gp15	tail completion protein	yes	25	1e-11	74	Phage tail assembly	218
gp17	subunit of the terminase for DNA packaging	yes	37	7e-107	81	Terminase large subunit	53
gp18	tail tube subunit	yes	28	9e-32	91	Tail sheath	70
gp20	head portal vertex protein	yes	36	2e-98	85	Portal vertex of the head	72
gp21	prohead core protein and protease	yes	33	6e-16	66	Prohead core protein	74
gp22	prohead core protein	no					
gp23	precursor of major head protein	yes	34	4e-17	96	Major capsid protein	76
gp25	base plate wedge subunit	no					
gp34	proximal tail fiber protein subunit	no					
gp36	small distal tail fiber protein subunit	no					
DNA replication, repair, and recombination							
gp43	DNA polymerase	yes	28	7e-83	97	DNA polymerase	133
gp44	sliding clamp loader complex tetramer	yes	32	6e-47	99	Replication factor C small subunit	123
gp41	helicase-primer complex hexamer	no					
gp46	subunit of a recombination nuclease complex	yes	29	2e-66	99	Phage recombinatio	121

	required for initiation of DNA replication					n-related endonuclease	
gp47	subunit of a recombination nuclease complex required for initiation of DNA replication	no					
UvsW	recombination DNA-RNA helicase, DNA-dependent ATPase	yes	33	8e-71	82	DNA helicase	78
Auxillary metabolism							
nrdA	subunit of an aerobic ribonucleotide reductase complex	yes	23	4e-18	58	Ribonucleotide reductase of 1a (aerobic), alpha subunit	24
nrdB	subunit of an aerobic ribonucleotide reductase complex	no					
Gene expression							
gp55	sigma factor for late transcription	no					

Supplementary Table 4-S4: Atu_ph08 genes organized by predicted function.

CDS #	RAST annotated function	Updated assigned function	Length (bp)	Category
1	Phage protein CDS	XRE transcriptional regulator	300	Transcription
2	hypothetical protein CDS		297	Hypothetical
3	Phage protein CDS		417	Hypothetical
4	FIG00451076: hypothetical protein CDS	DUF2312 domain-containing protein	396	Hypothetical
5	hypothetical protein CDS	GcrA cell cycle regulator	588	Transcription
6	hypothetical protein CDS		495	Hypothetical
7	C-5 cytosine-specific DNA methylase CDS		1866	DNA modification
8	hypothetical protein CDS		351	Hypothetical
9	hypothetical protein CDS	putative PRK12775-containing protein	246	Posttranslational modification
10	hypothetical protein CDS	NERD domain-containing protein	741	DNA processing
11	Phage protein CDS		627	Hypothetical
12	Phage protein CDS		804	Hypothetical
13	probable terminase large subunit CDS		1557	Structural
14	Phage protein CDS		429	Hypothetical
15	Phage portal protein CDS		2232	Structural
16	Phage protein CDS		1011	Hypothetical
17	hypothetical protein CDS		195	Hypothetical
18	hypothetical protein CDS		294	Hypothetical
19	hypothetical protein CDS		477	Hypothetical
20	Methyl-accepting chemotaxis protein I (serine chemoreceptor protein) CDS		786	Other/Bacterial
21	helicase, Snf2 family CDS	Adenine-specific DNA methylase, N12 class	14634	DNA modification

22	Phage protein CDS		1311	Hypothetical
23	Phage protein CDS	tail fiber domain-containing protein	1161	Structural
24	Phage protein CDS	N-acetyltransferase	450	DNA modification
25	hypothetical protein CDS		246	Hypothetical
26	Phage protein CDS	putative virion structural protein	1740	Structural
27	hypothetical protein CDS	DUF4376 domain-containing protein	603	Hypothetical
28	Phage tail fibers CDS		1074	Structural
29	Phage protein CDS		279	Hypothetical
30	hypothetical protein CDS		312	Hypothetical
31	Phage protein CDS	major capsid protein	417	Structural
32	Phage protein CDS	DUF4238 domain-containing protein	627	Hypothetical
33	Phage protein CDS		702	Hypothetical
34	hypothetical protein CDS		363	Hypothetical
35	Phage protein CDS		462	Hypothetical
36	Phage protein CDS	N4-gp56 family major capsid protein	1131	Structural
37	protein of unknown function DUF847 CDS	secretion activator protein; lysozyme-like protein	753	Lysis
38	hypothetical protein CDS	Holin of 3TMs, for gene-transfer release	525	Lysis
39	Phage protein CDS		333	Hypothetical
40	hypothetical protein CDS	AlpA family phage regulatory protein	231	Transcription
41	Integrase CDS		1299	DNA recombination
42	hypothetical protein CDS	Arc family DNA-binding protein	192	DNA-associated
43	hypothetical protein CDS		402	Hypothetical
44	hypothetical protein CDS		498	Hypothetical
45	hypothetical protein CDS		189	Hypothetical
46	hypothetical protein CDS	DUF551 domain-containing protein	186	Hypothetical

47	hypothetical protein CDS		324	Hypothetical
48	hypothetical protein CDS		219	Hypothetical
49	hypothetical protein CDS	3'-5' exoribonuclease	705	DNA-associated
50	hypothetical protein CDS		264	Hypothetical
51	hypothetical protein CDS		645	Hypothetical
52	Phage protein CDS		330	Hypothetical
53	Phage DNA modification methyltransferase CDS		753	DNA modification
54	hypothetical protein CDS		219	Hypothetical
55	Bacteriophage protein gp37 CDS	DUF5131 family protein	921	Hypothetical
56	COG1896: Predicted hydrolases of HD superfamily CDS	metal-dependent phosphohydrolase	606	DNA modification
57	Phage protein CDS		687	Hypothetical
58	hypothetical protein CDS		426	Hypothetical
59	Phage-related protein CDS	morphogenetic protein	708	Structural
60	Phage protein CDS		399	Hypothetical
61	hypothetical protein CDS		231	Hypothetical
62	hypothetical protein CDS		525	Hypothetical
63	hypothetical protein CDS		159	Hypothetical
64	Phage protein CDS	C4-dicarboxylate ABC transporter substrate-binding protein	882	Other/Bacterial
65	Transcriptional regulator CDS	C2-like repressor protein	417	Transcription
66	hypothetical protein CDS		219	Hypothetical
67	hypothetical protein CDS	class I SAM-dependent methyltransferase	582	DNA modification
68	hypothetical protein CDS		288	Hypothetical
69	Phage protein CDS		309	Hypothetical

70	hypothetical protein CDS		123	Hypothetical
71	Predicted periplasmic protein CDS	Clp protease	651	Posttranslational modification
72	hypothetical protein CDS	XRE family transcriptional regulator	306	Transcription
73	hypothetical protein CDS		240	Hypothetical
74	hypothetical protein CDS		243	Hypothetical
75	hypothetical protein CDS		207	Hypothetical

Supplementary Table 4-S5: Comparative analysis of Atu_ph08 gene products with related phages.

gp	Functional annotation	PBC5	POI1126	SopranoGao	Pavtok	DC1	Bcep22	PS-1	phiKMV	T7	Atu_ph03
1	XRE transcriptional regulator	7.00E-44									
3	Hypothetical	4.00E-26	1.00E-10								
4	Hypothetical	1.00E-38	5.00E-20								
10	NERD domain-containing protein		1.00E-32								
11	Hypothetical	8.00E-66									
12	Hypothetical	1.00E-77									
13	probable terminase large subunit	0	0	1.00E-143	5.00E-145	2.00E-127	1.00E-127	2.00E-100			
14	Hypothetical	6.00E-72	2.00E-40	3.00E-17	1.00E-17		9.00E-15				
15	Phage portal protein	0	0	0	0	0	0				
16	Hypothetical	7.00E-79	6.00E-07	9.00E-11	2.00E-18	4.00E-18	2.00E-19				
17	Hypothetical					6.00E-05					
21	Adenine-specific DNA methylase, N12 class; helicase, Snf2 family	0	0	0	0	0	0				5.00E-48
22	Hypothetical	1.00E-106	2.00E-45	5.00E-08	2.00E-20						
23	tail fiber domain-containing protein	0	4.00E-160	1.00E-110	2.00E-114	1.00E-75	3.00E-73				
24	N-acetyltransferase	3.00E-84	2.00E-43	9.00E-36	1.00E-35	3.00E-41	1.00E-39				
25	Hypothetical		2.00E-13			5.00E-08	2.00E-08				
26	putative virion structural protein	0	0	2.00E-140	1.00E-148	0.00E+00	1.00E-179				
28	Phage tail fibers	2.00E-31	3.00E-28	9.00E-23	4.00E-17	1.00E-26	2.00E-26				
29	Hypothetical	7.00E-38	1.00E-31			2.00E-18	3.00E-18				
31	major capsid protein	1.00E-46	5.00E-49			2.00E-36	6.00E-38				
32	Hypothetical	4.00E-76	6.00E-28	7.00E-30	1.00E-33	8.00E-45	2.00E-37				
33	Hypothetical	6.00E-105	2.00E-54	4.00E-45	2.00E-38	8.00E-29	9.00E-23				
34	Hypothetical		3.00E-12								
35	Hypothetical		2.00E-31	4.00E-12		2.00E-24	5.00E-34				

36	N4-gp56 family major capsid protein	0	0	2.00E-155	0	3.00E-179	0				
37	secretion activator protein; lysozyme-like protein		7.00E-61								
38	Holin of 3TMs, for gene-transfer release		8.00E-18								
39	Hypothetical	7.00E-26									
41	Integrase	3.00E-179									
42	Arc family DNA-binding protein		1.00E-22								
52	Hypothetical	5.00E-31									
55	Hypothetical	2.00E-89	7.00E-34								
59	morphogenetic protein	8.00E-29	1.00E-27								
60	Hypothetical	2.00E-32									
64	C4-dicarboxylate ABC transporter substrate-binding protein	2.00E-165									
65	Transcriptional regulator	7.00E-39									
67	class I SAM-dependent methyltransferase	7.00E-91									
69	Hypothetical	1.00E-21									
72	XRE family transcriptional regulator		8.00E-14								
	TOTAL	**30**	**27**	**14**	**13**	**16**	**16**	**1**	**0**	**0**	**1**

Supplementary Table 4-S6: Atu_ph08 gene products present in other *Agrobacterium* phages.

Atu_ph 08 gene product	Atu_ph08 protein function	Phage Match	Identity (%)	E-value	Query cover (%)	Protein name	Phage gp#
gp4	Hypothetical	7-7-1	81	9e-40	59	Hypothetical protein	15
gp21	Helicase	Atu_ph03	36	4e-48	6	Cell wall hydrolyse	35
“		Atu_ph02	36	4e-48	6	Cell wall hydrolyse	32
“		Atu_ph07	31	3e-18	3	Mega protein	42
gp37	protein of unknown function DUF847	7-7-1	28	6e-28	91	Putative membrane protein	10
gp56	Predicted hydrolases of HD superfamily	Atu_ph04	29	3e-20	92	putative HD superfamily hydrolase	116
gp65	Transcriptional regulator	7-7-1	47	2e-17	47	Putative transcriptiona l regulator	68
gp67	Hypothetical protein	Atu_ph07	33	1e-20	80	Hypothetical protein	599

Chapter 5

Phages in the Phuture

Bacteriophages are the most abundant biological entities in the world, and yet knowledge about them is vastly underrepresented in the field of biology. Despite being major drivers of horizontal gene transfer and microbial diversity, phages that infect soil bacteria are especially under-studied. The plant pathogen *Agrobacterium tumefaciens* is one of the most important agricultural bacteria [1], causing crown gall disease in plants and leading to decreased crop yield [2]. Prior this work, there was only one characterized *Agrobacterium* phage genome sequence [3]. In an effort to find phages to use as biocontrol, we have isolated and characterized 5 novel phages that infect *Agrobacterium tumefaciens*. All of the phages in this study were isolated from environmental water or waste water samples and have been classified as dsDNA phages in the order *Caudovirales*.

Phages Atu_ph02 and Atu_ph03 Encode a Unique Endolysin

Agrobacterium phages Atu_ph02 and Atu_ph03 are in the family *Podoviridae*, sharing a similar shape comprised of an icosahedral head and short tail [4]. Their genomes share high nucleotide identity and are ~45 kbp in size. Our phylogenetic analysis determined that they are phi-KMV-like phages that are closely related to a group of *Rhizobium* phages. Because of the phages' high lytic activity, we were interested in identifying their endolysin. A gene product annotated as phage peptidoglycan hydrolase (PPH) appeared to have an unusual domain structure, with an N-acetylmuramidase domain on its N-terminal, and a transmembrane domain on its C-terminal, and therefore we decided to investigate it further.

After expressing PPH under an inducible promoter in *A. tumefaciens*, it became evident that PPH is sufficient to cause cell lysis. Interestingly, we found that PPH causes cell branching prior to lysis. Since branching is a phenotype that occurs when the cell division machinery is

disrupted, we hypothesized that PPH is interacting with the divisome. Site-specific mutagenesis of the conserved glutamate and three lysine residues showed that the glutamate is necessary for PPH-mediated cell death and that the positively charged C-terminus is involved in negative regulation of the protein. Mutating the positive residues increased the lytic properties of the protein, showing that we can enhance the antimicrobial properties of PPH.

Agrobacterium phage Atu_ph07 is a Jumbo Phage

Sequencing of phage Atu_ph07 revealed that its genome is 490 kbp, making it one of the largest phages identified [5]. Atu_ph07 is classified in the *Myoviridae* family of phages, with an icosahedral head that is 146 nm in diameter and a 136 nm tail. Its closest relative is *Salicola* phage SCTP-2. Our comparative genome analysis shows that Atu_ph07 is closely related to a group of Rak2-like phages. Since the genome of Atu_ph07 encodes over 700 ORFs, the majority of which have unknown functions, we performed mass spectrometry to help identify the structural genes.

Characterization of Agrobacterium phages Atu_ph04 and Atu_ph08

Two phages, Atu_ph04 and Atu_ph08, were isolated and characterized from waste water samples. Atu_ph04, a member of the *Myoviridae* family, has a 143 kbp sized genome. Its closest relatives are *Rhizobium* phage vB_RleM_P10VF and *Sinorhizobium* phage phiM9. Phage Atu_ph08 has a 59 kbp genome, a *Podoviridae*, and closely related to *Sinorhizobium* phage PBC5 and *Ochrobactrum* phage POI1126. The Atu_ph08 genome encodes genes commonly found in transducing phages, including an integrase and transcriptional repressors. Further studies will be required to determine if Atu_ph08 is capable of transduction.

LIMITATIONS OF GENOMIC-BASED PHAGE CHARACTERIZATION

The high degree of phage diversity and the vast numbers of genes that encode proteins of unknown functions found in phage genomes suggest that a genomic-based approach to phage characterization is inherently limited in scope. Functional characterization of phages is needed to determine how they can be exploited for biological and biotechnology purposes.

In contrast to most other biological entities, there is no universal gene that is shared among phages to enable rapid phylogenetic classification of phages. Thus, studies of phage phylogeny are often performed by comparing signature genes to build trees and using comparative whole genome analyses. Using comparative genome analysis generally involves the comparison of core genes in a phage group. The genetic mosaicism of phages is evident during this analysis, as genes are often rearranged within the genome and non-core genes are thought to be non-essential.

Currently, the standard method of annotating a phage genome involves analyzing the sequence through the Rapid Annotations using Subsystems Technology (RAST) server, which compares each open reading frame to others in the database [6]. Annotations are created by grouping proteins into families based on their function. While this method has been shown to give a decent initial prediction, it is incapable of predicting the function of hypothetical proteins or those with no homology [7]. Manual annotations using BLAST and comparative genomic analysis can improve the genome annotations; however, biological characterization of hypothetical genes is required to expand the functional annotations in the database. It is often difficult to discover the true origin of phage genes or to predict the host range based on the phage genome sequence alone. For this reason, a combination of genomic analysis and phenotypic

analysis, including electron microscopy, host range testing, and mass spectrometry can help gain a better understanding of the phage biology.

Increased knowledge about the functions of phage proteins is needed to understand the role of phages in the environment, as drivers of microbial diversity, and as potential biotechnology tools. In particular, phage proteins with antimicrobial properties are of increasing importance as the global health crisis related to the rise in antibiotic resistance among bacteria continues to emerge. Here we outline some future directions to expand upon our genomic characterization of Agrophages. We expect that these studies should further our understanding of how Agrophages may contribute to host cell killing or serve as agents of genetic transfer.

FUTURE DIRECTIONS

Potential of PPH as an Antimicrobial

To further explore the potential of PPH as an antimicrobial, we would purify the protein and determine its efficacy in causing lysis of diverse bacteria following exogenous protein treatments. If purification of an active version of this membrane protein is difficult, we would propose to truncate the protein and purify the soluble protein portion containing the catalytic N-acetylmuramidase domain. Fusion of the active N-acetylmuramidase domain with a domain that allows protein entry into the outer membrane of Gram negative bacteria may be necessary for PPH to induce lysis following external application. Such studies should reveal if PPH may be a viable antimicrobial, as has been proposed for other phage endolysins [8,9].

Transposon Mutagenesis to Understand the Cell Division Block Caused by PPH

One major outstanding question that arose from this work is how PPH, the putative endolysin expressed by Atu_ph02 and Atu_ph03, causes cells to branch prior to lysis, and why mutating the positively-charged C-terminal of PPH eliminates PPH's ability to cause cell branching. Since the phenotype observed when PPH is expressed in *A. tumefaciens* is reminiscent of *A. tumefaciens* growth with a depleted divisome [10], we hypothesized that PPH is interacting with cell division machinery to cause the abnormal phenotype.

To determine if PPH is interacting with any host cell proteins, we initiated a transposon mutagenesis screen. We used a *mariner* transposon system to mutagenize PPH-expressing *A. tumefaciens* at random sites in the genome [11,12]. To identify survivor strains that are no longer killed by PPH expression, we plan to perform high-throughput sequencing of survivor colonies. We expect that this screen will elucidate host factors that contribute to PPH-mediated lysis. In addition, we have transposon mutagenized PPH_{E32A} and $PPH_{K328A,K334A,K335A}$ strains of *A. tumefaciens*. Since PPH_{E32A} lacks the predicted catalytic site for cell wall hydrolysis, we expect that mutants surviving following expression of PPH may reveal the host proteins which interact with PPH and cause the changes in cell morphology. Finally, since $PPH_{K328A,K334A,K335A}$ causes rapid host lysis, we expect to identify mutants that are blind to PPH expression. Overall, we expect that this data will not only reveal host targets of PPH, but also illuminate how the host responds, perhaps via stress responses, when exposed to this phage lysis protein. Finally, the host targets of PPH, if not already known, may be candidate targets for novel antimicrobial drugs or peptides.

Exploring the Transducing Ability of Atu_ph08

The lifestyle of Atu_ph08 merits further study. Our identification of an integrase and transcriptional repressors in the Atu_ph08 genome suggest that this phage may be temperate and capable of undergoing both lytic and lysogenic cycles. The high degree of similarity of the Atu_ph08 genome with the genome of an *Agrobacterium* genomospecies 3 isolate suggests that Atu_ph08 is certainly derived from a temperate ancestor. Since there is not a known transducing phage that infects *A. tumefaciens* and transducing phages are useful for genetic manipulation of bacteria, further studies should be completed to determine if Atu_ph08 has transducing abilities.

CONCLUDING REMARKS

Despite the success of phage application as a biocontrol agent against phytopathogens [13], this is the first study describing phages that infect, *A. tumefaciens*. The number of unknown genes revealed by the characterization of *Agrobacterium* phages uncovers the vast gap in knowledge in regard to phages that reside in the rhizosphere. Though we initially sought to isolate phages as conventional biocontrol agents, we found that phage genomes contain an untapped reservoir of genes encoding proteins of unknown function, likely including novel lysis proteins.

It is clear that the so-called canonical endolysin-holin-spanin mechanism is far from the only method that phages use to kill their hosts [14]. Non-canonical endolysins and single lysis proteins have only recently been well characterized [15], and the known diversity of phages suggests that novel mechanisms of lysis are awaiting discovery. With numerous new phage genomes available annually due to programs such as SEA-PHAGES [16], the mining of new phage lysis proteins could potentially contribute to the engineering of novel phage-based therapeutic strategies. In the last decade, the use of endolysins and peptides with shuffled

endolysin domains has shown promise as a feasible therapeutic approach [17,18]. Indeed, phage therapy in humans is gaining a renewed interest in the West [19] and recent case studies highlight that phages can save lives [20,21]. Thus, as the numbers of multi-drug resistant bacterial pathogens continue to rise it is vital to invest in fundamental research about diverse phages from many environments. While explorations of phages that infect soil dwelling bacteria may not improve conventional phage therapy approaches, which typically use naturally-occurring phages to lyse bacteria at the site infection, biotechnological advances have further expanded the repertoire of phage therapeutics to include the use of bioengineered phages and purified phage lytic proteins. Modern advances in phage research are likely to yield crucial insights necessary for further development and innovation of antibacterial therapies.

REFERENCES

1. Mansfield, J.; Genin, S.; Magori, S.; Citovsky, V.; Sriariyanum, M.; Ronald, P.; Dow, M.; Verdier, V.; Beer, S. V; Machado, M. A.; Toth, I.; Salmond, G.; Foster, G. D. Top 10 plant pathogenic bacteria in molecular plant pathology. *Mol Plant Pathol* **2012**, *13*, 614–629, doi:10.1111/j.1364-3703.2012.00804.x.
2. Pulawska, J. Crown gall of stone fruits and nuts, economic significance and diversity of its causal agents: tumorigenic *Agrobacterium* spp. *J. Plant Pathol.* **2010**, *92*, S1.87-S1.98.
3. Kropinski, A. M.; Van Den Bossche, A.; Lavigne, R.; Noben, J. P.; Babinger, P.; Schmitt, R. Genome and proteome analysis of 7-7-1, a flagellotropic phage infecting *Agrobacterium* sp H13-3. *Virol. J.* **2012**, *9*, doi:10.1186/1743-422X-9-102.
4. Attai, H.; Rimbey, J.; Smith, G. P.; Brown, P. J. B. Expression of a peptidoglycan hydrolase from lytic bacteriophages Atu_ph02 and Atu_ph03 triggers lysis of *Agrobacterium tumefaciens*. *Appl. Environ. Microbiol.* **2017**, *83*, e01498-17, doi:10.1128/AEM.01498-17.
5. Attai, H.; Boon, M.; Phillips, K.; Noben, J.-P.; Lavigne, R.; Brown, P. J. B. Larger than life: Isolation and genomic characterization of a jumbo phage that infects the bacterial plant pathogen, *Agrobacterium tumefaciens*. *Front. Microbiol.* **2018**, *9*, 1861, doi:10.3389/FMICB.2018.01861.
6. Aziz, R. K.; Bartels, D.; Best, A. A.; DeJongh, M.; Disz, T.; Edwards, R. A.; Formsma, K.; Gerdes, S.; Glass, E. M.; Kubal, M.; Meyer, F.; Olsen, G. J.; Olson, R.; Osterman, A. L.; Overbeek, R. A.; McNeil, L. K.; Paarmann, D.; Paczian, T.; Parrello, B.; Pusch, G. D.; Reich, C.; Stevens, R.; Vassieva, O.; Vonstein, V.; Wilke, A.; Zagnitko, O. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **2008**, *9*, 75,

doi:10.1186/1471-2164-9-75.

7. McNair, K.; Aziz, R. K.; Pusch, G. D.; Overbeek, R.; Dutilh, B. E.; Edwards, R. Phage genome annotation using the RAST pipeline. *Methods Mol. Biol.* **2018**, *1681*, 231–238, doi:10.1007/978-1-4939-7343-9_17.

8. Briers, Y.; Walmagh, M.; Puyenbroeck, V. Van; Cornelissen, A.; Cenens, W.; Aertsen, A.; Oliveira, H. Engineered endolysin-based "Artilysins" to combat multidrug-resistant Gram-negative pathogens. *MBio* **2014**, *5*, e01379-14, doi:10.1128/mBio.01379-14.Editor.

9. Zampara, A.; Sørensen, M. C. H.; Grimon, D.; Antenucci, F.; Briers, Y.; Brøndsted, L. Innolysins: A novel approach to engineer endolysins to kill Gram-negative bacteria. *bioRxiv* **2018**, *408948*, doi:http://dx.doi.org/10.1101/408948.

10. Howell, M. L.; Aliashkevich, A.; Sundararajan, K.; Daniel, J. J.; Lariviere, P. J.; Goley, E.; Cava, F.; Brown, P. J. B. *Agrobacterium tumefaciens* divisome proteins regulate the transition from polar growth to cell division. *Mol. Microbiol.* **2019**, doi:10.1111/mmi.14212.

11. Perry, B. J.; Yost, C. K. Construction of a mariner-based transposon vector for use in insertion sequence mutagenesis in selected members of the *Rhizobiaceae*. *BMC Microbiol.* **2014**, *14*, 1–11, doi:10.1186/s12866-014-0298-z.

12. Perry, B. J.; Akter, M. S.; Yost, C. K. The use of transposon insertion sequencing to interrogate the core functional genome of the legume symbiont *Rhizobium leguminosarum*. *Front. Microbiol.* **2016**, *7*, 1–21, doi:10.3389/fmicb.2016.01873.

13. Buttimer, C.; McAuliffe, O.; Ross, R. P.; Hill, C.; O'Mahony, J.; Coffey, A. Bacteriophages and bacterial plant diseases. *Front. Microbiol.* **2017**, *8*, doi:10.3389/fmicb.2017.00034.

14. Young, R. Phage lysis: three steps, three choices, one outcome. *J. Microbiol.* **2014**, *52*, 243–258, doi:10.1007/s12275-014-4087-z.

15. Bernhardt, T. G.; Wang, I. N.; Struck, D. K.; Young, R. Breaking free: "Protein antibiotics" and phage lysis. *Res. Microbiol.* **2002**, *153*, 493–501, doi:10.1016/S0923-2508(02)01330-X.

16. Jordan, T. C.; Burnett, S. H.; Carson, S.; Caruso, S. M.; Clase, K.; DeJong, R. J.; Dennehy, J. J.; Denver, D. R.; Dunbar, D.; Elgin, S. C. R.; Findley, A. M.; Gissendanner, C. R.; Golebiewska, U. P.; Guild, N.; Hartzog, G. A.; Grillo, W. H.; Hollowell, G. P.; Hughes, L. E.; Johnson, A.; King, R. A.; Lewis, L. O.; Li, W.; Rosenzweig, F.; Rubin, M. R.; Saha, M. S.; Sandoz, J.; Shaffer, C. D.; Taylor, B.; Temple, L.; Vazquez, E.; Ware, V. C.; Barker, L. P.; Bradley, K. W.; Jacobs-Sera, D.; Pope, W. H.; Russell, D. A.; Cresawn, S. G.; Lopatto, D.; Bailey, C. P.; Hatfull, G. F. A broadly implementable research course for first-year undergraduate students. *MBio* **2014**, *5*, 1–8, doi:10.1128/mBio.01051-13.Editor.

17. Briers, Y.; Lavigne, R. Breaking barriers: expansion of the use of endolysins as novel antibacterials against Gram-negative bacteria. *Future Microbiol.* **2015**, *10*, 377–390.

18. Abdelkader, K.; Gerstmans, H.; Saafan, A.; Dishisha, T.; Briers, Y. The preclinical and clinical progress of bacteriophages and their lytic enzymes: the parts are easier than the whole. *Viruses* **2019**, *11*, 96, doi:10.3390/v11020096.

19. Kortright, K. E.; Chan, B. K.; Koff, J. L.; Turner, P. E. Phage therapy: a renewed approach to combat antibiotic-resistant bacteria. *Cell Host Microbe* **2019**, *25*, 219–232, doi:10.1016/j.chom.2019.01.014.

20. Schooley, R. T.; Biswas, B.; Gill, J. J.; Hernandez-Morales, A.; Lancaster, J.; Lessor, L.;

Barr, J. J.; Reed, S. L.; Rohwer, F.; Benler, S.; Segall, A. M.; Taplitz, R.; Smith, D. M.; Kerr, K.; Kumaraswamy, M.; Nizet, V.; Lin, L.; Mccauley, M. D.; Strathdee, S. A.; Benson, C. A.; Pope, R. K.; Leroux, B. M.; Picel, A. C.; Mateczun, A. J.; Cilwa, K. E.; Regeimbal, J. M.; Estrella, L. A.; Wolfe, D. M.; Henry, M. S.; Quinones, J.; Salka, S.; Bishop-Lilly, K. A.; Young, R.; Hamilton, T. Development and use of personalized bacteriophage-based therapeutic cocktails to treat a patient with a disseminated resistant *Acinetobacter baumannii* infection. *Antimicrob. Agents Chemother.* **2017**, *61*, e00954-17.

21. Chan, B. K.; Turner, P. E.; Kim, S.; Mojibian, H. R.; Elefteriades, J. A.; Narayan, D. Phage treatment of an aortic graft infected with *Pseudomonas aeruginosa*. *Evol. Med. Public Heal.* **2018**, *2018*, 60–66, doi:10.1093/emph/eoy005.

Vita

Hedieh Attai was born and raised in Los Angeles, California. She graduated Granada Hills Charter High School in 2008. She next went to College of the Canyons, where she earned her Associates Degree in Biological and Physical Sciences. She transferred to University of California, Berkeley, where she worked in the laboratory of Dr. Matthew Welch. Hedieh graduated in 2012, earning her Bachelor Degree in Molecular and Cell Biology, with an emphasis in Immunology. For the next two years, Hedieh worked as a laboratory technician in the laboratory of Dr. David Bermudes at California State University, Northridge.

Hedieh started graduate school at the University of Missouri in Fall 2014, when she joined the laboratory of Dr. Pamela Brown. During this time, she taught 2 semesters of Introduction to Cell Biology discussion courses, as well as 2 semesters of Microbiology Lab. Hedieh mentored 5 undergraduate students. She received an NIH T32 Training Grant in 2015, and a U.S. Department of Education GAANN Fellowship in 2017. Hedieh gave 5 oral presentations, 3 of which took place at international conferences. In total, Hedieh gave 12 poster presentations. During graduate school, Hedieh participated in the Biology Graduate Student Association (BGSA) as a Social Chair, then Secretary.

Hedieh has accepted a postdoctoral position at the Center for Innovative Phage Applications and Therapeutics (IPATH) University of California, San Diego in the laboratory of Dr. David Pride.