

Bacteriophages in the honey bee gut and amphibian skin microbiomes: investigating the interactions between phages and their bacterial hosts

Emma Kathryn Rose Bueren

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Biological Sciences

Lisa Belden, Chair
David Haak
Bryan Hsu
Richard Fell
Dana Hawley

April 29th, 2024
Blacksburg, Virginia

Keywords: bacteriophage, prophage, viral communities, viromes, auxiliary metabolic genes, lysogenic conversion, horizontal gene transfer, *Apis mellifera*, gut microbiome, skin microbiome, amphibian, *Janthinobacterium lividum*, chitinase

Bacteriophages in the honey bee gut and amphibian skin microbiomes: investigating the interactions between phages and their bacterial hosts

Emma K. Bueren

Public Abstract

The microbial communities of animals, called "microbiomes", play important roles in the health of animals. The bacteria in these microbiomes can help strengthen the immune system, provide resistance to dangerous pathogens, and break down nutrients. However, bacteria are not alone in the microbiome; viruses are also present. Surprisingly, the vast majority of the world's viruses, even those living inside animals, infect bacteria. These viruses, called "bacteriophages" or "phages", can impact the bacterial communities in a microbiome. Phages can be grouped in to two broad categories based on lifecycle. Lytic phages kill the bacterial host directly after infection. Temperate phages, on the other hand, can either immediately kill the host like lytic phages or alternatively, become a part of the bacterial genome and live as prophages. Phages with both lifecycles can sometimes carry genes that, although not essential to the phage, may change the traits of the bacteria during infection. For example, some phages carry toxin genes, which bacteria use to cause disease in animals. Other phages might carry genes that provide antibiotic resistance or alter the metabolism of the infected bacteria. If a phage gene benefits the infected bacteria, the bacteria may begin interacting with its environment in a new way or may even become more abundant. Alternatively, phages that directly kill infected bacteria may have a negative effect on bacterial population sizes. To begin unraveling how phages influence bacterial species in microbiomes, I investigated two different animal systems: the Western honey bee (*Apis mellifera*) gut microbiome and the amphibian skin microbiome. I first identified prophages of several common bacterial species that reside in the honey bee gut (Chapter 2). Prophages were more common in certain bacterial species than others, and some possessed genes associated with the breakdown of sugars or pollen, suggesting they help honey bees process their food. Using similar techniques, I then identified prophages in bacteria isolated from the skin microbiomes of several amphibian species common in the eastern United States (American bullfrogs, Eastern newts, Spring peepers, and American toads) (Chapter 3). Most notably, the bacteria *Janthinobacterium lividum* may benefit from prophages that carry genes for potentially antifungal chitinase enzymes that destroy the fungal cell wall. Finally, I returned to the honey bee gut microbiome system by investigating how honey bee bacteriophage communities change over large geographic distances (Chapter 4). This study, which examined honey bees from 12 apiaries sampled from the east to west coast of the United States, looks primarily at lytic phage and temperate phage that are not integrated as prophage, but are instead seeking a bacterial host to infect. I found that nearby apiaries tended to have more similar communities of bacteriophages, compared to apiaries far away. Additionally, most bacteriophage communities carry genes associated with the breakdown of sugars like sucrose. Overall, these three studies show that phages, and especially prophages,

contribute to the genetic landscape of the microbiome by broadly providing bacterial hosts with access to a diverse set of genes.

Bacteriophages in the honey bee gut and amphibian skin microbiomes: investigating the interactions between phages and their bacterial hosts

Emma K. Bueren

Academic Abstract

The bacteria in host-associated microbial communities influence host health through various mechanisms, such as immune stimulation or the release of metabolites. However, viruses that target bacteria, called bacteriophages (phages), may also shape the animal microbiome. Most phage lifecycles can be classified as either lytic or temperate. Lytic phages infect and directly kill bacterial hosts and can directly regulate bacterial population size. Temperate phages, in contrast, have the potential to undergo either a lytic cycle or integrate into the bacterial genome as a prophage. As a prophage, the phage may alter bacterial host phenotypes by carrying novel genes associated with auxiliary metabolic functions, virulence-enhancing toxins, or resistance to other phage infections. Lytic phages may also carry certain auxiliary metabolic genes, which are instead used to takeover bacterial host functions to better accommodate the lytic lifecycle. In either case, the ability to alter bacterial phenotypes may have important ramifications on host-associated communities. This dissertation focused on the genetic contributions that phages, and particularly prophages, provide to the bacterial members of two separate host-associated communities: the honey bee (*Apis mellifera*) gut microbiome and the amphibian skin microbiome. My second chapter surveyed publicly available whole genome sequences of common honey bee gut bacterial species for prophages. It revealed that prophage distribution varied by bacterial host, and that the most common auxiliary metabolic genes were associated with carbohydrate metabolism. In chapter three, this bioinformatic pipeline was applied to the amphibian skin microbiome. Prophages were identified in whole genome bacterial sequences of bacteria isolated from the skin of American bullfrogs (*Lithobates catesbeianus*), eastern newts (*Notophthalmus viridescens*), Spring peepers (*Pseudacris crucifer*) and American toads (*Anaxyrus americanus*). Prophages were additionally identified in publicly available genomes of non-amphibian isolates of *Janthinobacterium lividum*, a bacteria found both on amphibian skin and broadly in the environment. In addition to a diverse set of predicted prophages across amphibian bacterial isolates, several *Janthinobacterium lividum* prophages from both amphibian and environmental isolates appear to encode a chitinase-like gene undergoing strong purifying selection within the bacterial host. While identifying the specific function of this gene would require *in vitro* isolation and testing, its high homology to chitinase and endolysins suggest it may be involved in the breakdown of either fungal or bacterial cellular wall components. Finally, my fourth chapter revisits the honey bee gut system by investigating the role of geographic distance in bacteriophage community similarity. A total of 12 apiaries across a transect of the United States, from Virginia to Washington, were sampled and honey bee viromes were sequenced, focusing on the lytic and actively lysing temperate community of phages. Although each apiary possessed many unique

bacteriophages, apiaries that were closer together did have more similar communities. Each bacteriophage community also carried auxiliary carbohydrate genes, especially those associated with sucrose degradation, and antimicrobial resistance genes. Combined, the results of these three studies suggest that bacteriophages, and particularly prophages, may be contributing to the genetic diversity of the bacterial community through nuanced relationships with their bacterial hosts.

Dedication

To my parents: you are always so very proud of me.

Acknowledgments

I would like to first thank my advisor, Dr. Lisa Belden, who in addition to letting viruses infiltrate her lab, was also an invaluable resource for guidance these past five years. I have grown immensely as a person and as a scientist under her mentorship, becoming more confident in my skills and abilities to ask more ambitious questions. I would also like to express my deepest affection and gratitude for the other members of the Belden lab during my time at Tech (Dr. Angie Estrada, Dr. Korin Jones, Dr. Emma Bradford, Dr. Casey Gregory, Noah Wax, Mychala Snead, Arturo Roman, Tiffany Duong, Olivia Sacci, and Maddie Brown). Dr. Korin Jones has been and continues to be a shining role model and a wonderful friend in times when I've needed both.

Additionally, I would like to thank my committee members (Dr. David Haak, Dr. Richard Fell, Dr. Bryan Hsu, Dr. Dana Hawley) for their unwavering support and numerous contributions. Dr. David Haak showed me how to open my computer's terminal and then gave me the tools to run wild (even as my scripts were running mostly error codes). Dr. Richard Fell, who teaches me something new about bees every time I see him, opened his backyard and bee hives to me. Dr. Bryan Hsu showed me the ropes of phage isolation and virome preparation and graciously let me borrow all sorts of equipment and reagents. Dr. Dana Hawley supported and encouraged me with her insight and birds-eye view of my research in the greater scheme of things.

I would also like to thank Dr. Frank Aylward and Dr. Alaina Weinheimer, who were both instrumental in the development of the bioinformatic pipeline I use in chapters two and four. Dr. Frank Aylward also graciously allowed me access to his lab and his centrifuge. Dr. Alaina Weinheimer held my hand through much of the early growing pains as I became a "real" bioinformatician. She also provided me with several invaluable python scripts I still use regularly. I would also like to thank Hollyn Franklin and the other members of the Hsu lab for helping me troubleshoot and set up phage protocols.

A heartfelt thank you to the Biological Sciences faculty as a group for creating such a supportive and cheerful environment. Whenever I was feeling down, I always seemed to magically run in to a faculty member who knew just what to say. I also want to thank the graduate and postdoc community of the department for their warmth, enthusiasm, and sometimes much needed commiseration. I am especially grateful to my friends on the Derring 4th floor whose offices were always open to me. Kait Malewicz, my office mate, went above and beyond the call of friendship and unlocked the office door more than once when I locked myself out. I also want to thank the Microbiology Teaching Lab Staff (Katrina Lasley, Jordan Knapp, Georgia Murn, Hannah Bradley, and the numerous other TAs) for always having so much fun with me. I also must thank the undergraduates I taught over the years for constantly reawakening my wonder and reminding me how much I love microbiology.

I would also like to thank the Interfaces of Global Change and the Global Change Center. Both the curriculum and the amazing community made me a bigger thinking and more worldly scientist. I

am also incredibly grateful for the yearlong fellowship I received from the IGC during my fourth year, which provided some much-needed time to refocus on my research. The IGC also championed the efforts of Flip the Fair, an outreach event organized by several graduate students (Abbey Lewis, Amanda Hensley, Grace O'Malley, Heather Wander, Gates Palissery, Carla López Lloreda, Claudia Perez, Sophie Drew and me). It was an incredibly rewarding experience, and I am in awe of my other organizers, who taught me so much about how to put on such a gigantic event.

I am also deeply grateful to my hometown friends, who always make me feel so loved. Thank you all for staying touch after I ran across the country. Jaclyn Hanson, Christian Reyes, Makayla Kinsella, Colin Piwtorak, Ali Cho, Ricco Torres, Salina Johnson, Amanda Dahl, and many, many others, I don't think I could have done it without your support, jokes, and your overwhelming confidence in my abilities. I am also so grateful to my newer Blacksburg friends, and particularly my roommates, who provided much needed social enrichment when I was holed up in my room. I'd also like to include a shoutout to the many healthcare workers over the past five years who helped me get a handle on my migraines, as well as Becky Barger and Dr. Deb Jenkins for listening and keeping me sane.

None of this would have happened without the continuous love and support of my parents, Paul and Sara Bueren. I am so grateful you bought me a microscope for Christmas and then let me play around with pond water. Thank you for instilling that love of science in me and for truly embracing the dorky weird kid energy. I am also grateful to my gigantic extended family of aunts, uncles, and cousins, who have always been so enthusiastic and interested in my work. Also included in my family is, of course, Jalapeño, my 9lb calico cat who is a stalwart defender of bedtime and a small, swift-moving lovable entity of chaos.

Finally, I'd like to thank Evi Emmenegger, who hired me fresh out of college, showed me how to be brave, and made me realize how fun microbiology is when you mix it with ecology.

Table of Contents

Public Abstract	ii
Academic Abstract	iv
Dedication	vi
Acknowledgments.....	vii
Table of Contents.....	ix
Chapter One: Introduction.....	1
1. Host-symbiont interactions.....	1
2. Symbiont-symbiont interactions.....	3
3. Studying bacteriophages in animal host systems.....	6
4. Summarized doctoral research	10
References.....	10
Chapter Two: Characterization of prophages in bacterial genomes from the honey bee (<i>Apis mellifera</i>) gut microbiome	17
Abstract	17
Introduction.....	17
Materials and Methods	19
Results	22
Discussion	25
Conclusion.....	29
Figure Legends	30
Supplementary Information	36
References.....	38
Chapter Three: Putative prophages of bacterial isolates collected from amphibian skin	43
Abstract	43
Introduction.....	43
Methods	46
Results	51
Discussion	59
Conclusion.....	61
Supplemental Information	61
References.....	64
Chapter Four: Bacteriophage in the gut microbiome of honey bees (<i>Apis mellifera</i>) across the United States	69
Abstract	69
Introduction.....	69
Methods	71
Results	74
Discussion	81
Supplementary Materials	83
References.....	87
Chapter Five: Synthesis	93
References.....	95

Chapter One: Introduction

1. Host-symbiont interactions

1.1 Individual host-symbiont interactions within the mutualism-parasitism continuum

Plants and animals are hosts to a diverse community of microbial symbionts that live on or inside their respective hosts. With the realization that each animal or plant is an entire microbial ecosystem, the concept of the “holobiont” has been established, which combines a “macrobe” host and their “microbe” symbionts into a single entity, similar to the concept of superorganisms (Youle et al., 2013). One of the earliest recognized holobiont systems was that of coral and their photosynthetic algae, as the coral cannot survive without the algae, despite the fact that they are separate species (Rosenberg et al., 2007).

Not all symbiotic relationships are positive; symbiont interactions fall on a continuum between mutualism and parasitism. Mutualism occurs when both the host and the symbiont benefit, which is exhibited by the relationship between legumes and nitrogen-fixing bacteria, called rhizobia. The legumes host the rhizobia in specially made root nodes and provide them with carbon in exchange for access to bioavailable nitrogen fixed by the rhizobia (Regus et al., 2015). Alternatively, in parasitism, only the host or symbiont benefits at the expense of the other. Within the context of disease, pathogens are parasites, as they exploit host resources to reproduce. However, all symbiotic relationships are context-dependent and may change over time in response to varying conditions. If host health or environmental conditions create a cost too high for one of the symbionts, or if an alternate “cost-free” source appears, a mutualistic relationship breaks down (Sachs & Simms, 2006). For example, that same mutualist relationship between legumes and rhizobia can shift to parasitism if bioavailable nitrogen is already abundant (meaning the presence of the rhizobia imposes an unnecessary cost and is a drain of resources on the legume), or if the rhizobia exploit the plant through excessive nodulation (Nishimura et al., 2002; Sachs, Quides & Wendlandt, 2018). Alternatively, some parasites may impose a cost, but provide protection against other, more costly parasites. Humans co-infected with *Plasmodium falciparum* (a causative agent of malaria) and *Schistosoma haematobium* (a chronic urinary helminth) experience a reduction in fevers caused by malaria (Dumbo et al., 2014). In this case, it is possible that the long-term costs of the chronic schistosomiasis may be outweighed by the benefit of reduced malarial disease, which can be fatal. Other symbiotic relationships may vary throughout the life stage of the individual host or symbiont, displaying aspects of both mutualism and parasitism at different times. The endophytic fungus *Epichloë bromicola* considerably improves the growth of the grass *Bombus erectus* by increasing drought tolerance and the production of bioprotective compounds that deter predators in exchange for host nutrients and protection; however, it also renders the grass effectively infertile by preventing flowering (Groppe et al., 1999; Eaton, Cox & Scott, 2011). When *B. erectus* begins to redirect energetic resources for flowering, the fungus grows unrestrained and parasitizes the flowers, shifting the relationship from that of mutualism to parasitism (Eaton, Cox & Scott, 2011).

In some cases, the host may control a mutualist to prevent overexploitation and the transition to parasitism. Females of the wasp *Allodynerus delphinalis* transfer a mutualist mite *Ensliniella parasitica* to new nests in a specialized organ, the acarinarium, which functions essentially as a mite-carrying pocket (Okabe & Makino, 2010). Although this mite feeds on the hemolymph of juvenile wasps, it also protects them from the larger threat of the parasitoid *Melittobia acasta* (Okabe & Makino, 2010). Okabe and Makino (2010) demonstrated that high mite infection levels are detrimental to the survival of larval male wasps, who pupate faster and have less fat reserves than female larval wasps. However, this threshold is never passed in the wild, possibly due to the limited size of the acarinarium, which prevents too many mites from colonizing a single female wasp (and nest builder). Furthermore, Okabe and Makino (2010) argue that the mite has no reason to challenge this acarinarium control, as mite overpopulation provides little benefit compared to the risk of juvenile death, which would prevent further mite transmission. While the long-term evolution of a mutualistic relationship to a parasitic one is relatively rare due to enforced constraints from both the host and mutualist, it can occur. The reversal of a leafflower-leafflower moth mutualism facilitated by a braconid wasp that preys on the moth larvae suggests that parasites outside of a mutualism may be a possible trigger for the transition from mutualism to parasitism (Kawakita, Mochizuki & Kato, 2015).

At the center of this context-dependent mutualism-parasitism continuum, commensalism occurs. A commensal relationship provides a one-sided benefit to either the host or symbiont, but without detriment to the other. For example, animals and plants host a variety of microorganisms that at first glance do not appear to provide any obvious benefits (or costs); however, it may be that some of these so-called commensals are not as neutral as first thought. Human gut commensals appear to shape the immune system (Lee & Mazmanian, 2010; Cho & Blaser, 2012), and *Drosophila melanogaster* gut commensals appear to influence mate choice (Sharon et al., 2010).

1.2 Host-microbiome interactions

A single host participates in multiple host-symbiont relationships, and the community of all microbes (mutualistic, commensal, and parasitic) is referred to as a "microbiome". Recently, more research has been dedicated to understanding the interactions between a host and their microbiome, with a particular focus on bacterial communities, which have a major role in host health. Studies with germ-free (GF) mice have demonstrated that the microbiome is a critical component in the development of a well-functioning immune system, as immune function is altered in GF mice, and is corrected with the return of a normal gut microbiota (Chen & Kasper, 2014; Cabrera-Perez, Badovinac & Griffith, 2017). Although the microbiome of mammalian neonates is still establishing itself, early microbial colonization is key for some specific aspects of immune functioning. Mice deprived of a microbiome in the first few days have higher abundances of certain immune cells, which accumulate and can lead to increased mortality from autoimmune diseases even after a normal microbiome is returned (Olszak et al., 2012).

In addition to its role in regulating the immune system, the physical presence of the microbiome can also prevent the establishment of more pathogenic bacteria. For example, antibiotic treatments indiscriminately remove many commensal bacteria, as well as targeted pathogens. The groups of bacteria first able to colonize this environment often grow disproportionately, as there is little competition, and this can cause dysbiosis. A classic case of this type of dysbiosis is *Clostridium difficile*-associated diarrhea (CDAD), which is caused by an overgrowth of the pathogenic *C. difficile* and often occurs after antibiotic treatment. As antibiotics are the primary treatment for CDAD, successful treatment can be difficult. Recurring episodes of CDAD are common, which in turn leads to the development of antibiotic-resistant *C. difficile* (Khosravi & Mazmanian, 2013). However, reestablishment of a healthy microbiome through fecal transplants, in which the microbiome of a healthy individual without CDAD is introduced to a CDAD patient, can have an initial cure rate of 91%, with 98% of individuals cured after a second round of antibiotics (either with or without a second fecal transplant) (Brandt et al., 2012).

The bacterial community within the human gut is highly diverse and varies in each human host. It is debated whether or not humans and many other hosts have a “core” community of gut symbionts, which are present in all healthy guts, or if they are more variable (Caporaso et al., 2011; Faith et al., 2013). There is some evidence that certain “core” types of microbes are associated with healthy microbiomes (Huse et al., 2012; Henderson et al., 2015). However, this core may not necessarily be determined by specific taxa, but instead by the functional abilities of the microbes— i.e. their abilities to metabolize specific carbohydrates or produce certain metabolites (Shafquat et al., 2014). Furthermore, an unusual or irregular microbiome, in which these core bacteria or functional groups are reduced (known as “dysbiosis”), can be correlated with disease (Carding et al., 2015). Recent efforts have begun to characterize the fungal (mycobiome) and viral (virome) communities in humans as well (Scarpellini et al., 2015; Huseyin et al., 2017).

2. Symbiont-symbiont interactions

2.1 Interactions within the microbiome

It is important to consider that microbial symbionts also interact with each other. Much effort has been devoted to developing network models that attempt to untangle the competitive and facilitative interactions between different bacterial symbionts in a host-associated community. These network models focus on how each positive or negative symbiont-symbiont interaction shapes the collective microbiome. In other words, these models look for correlative relationships between symbionts indicating co-occurrence (when both symbionts occur together at high relative abundance) or co-exclusion (where high relative abundance of one symbiont is associated with the low relative abundance of the other) to predict whether relationships are cooperative or competitive (Faust & Raes, 2012; Berry & Widder, 2014). Many of these interactions are resource-driven. Two bacteria may compete for the same resource, leading to reduced growth in either both or one of the symbionts. Alternatively, the waste-product of one bacterium may facilitate the growth of another (Berry & Widder, 2014). These positive and negative interactions may directly or indirectly impact

other symbionts in the community, which in turn, may also impact host health (as described in section 1.2).

2.2 Microbiome-bacteriophage interactions

Like eukaryotic organisms, bacteria can also be infected by viruses. These viruses, called bacteriophages (or phages), are found ubiquitously in bacteria both in the environment and within host microbiomes (Weinbauer, 2004; Rosario & Breitbart, 2011; Pride et al., 2012). In environmental samples, they are the most abundant biological agent present, outnumbering bacteria, fungi and eukaryotic viruses (Weinbauer, 2004). Within hosts, they often outnumber other biological agents, and frequently make up the largest numerical proportion of the host virome (which includes eukaryotic viruses) (Reyes et al., 2010; Pride et al., 2012). A phage typically has one of two distinct life cycles, and is either lytic or lysogenic (also referred to as virulent and temperate life cycles, respectively). Lytic phages infect a bacterium and then replicate until the bacterial cell bursts, releasing phage progeny. Lysogenic phages, however, either follow the lytic pathway immediately or infect through lysogeny, in which they integrate into the host's genome, called the prophage stage. The prophage replicates with the host genome when it divides, remaining in the host and all host progeny until the lytic life cycle is induced. What causes a prophage to return to a lytic life cycle is unclear, although certain stimuli, such as exposure to ultraviolet radiation and temperature changes, can sometimes trigger lysis (Weinbauer, 2004).

In the past, research on the ecology and impact of bacteriophages in microbial communities has primarily focused on environmental systems, such as the ocean (Weinbauer, 2004). Notably, lytic phages play a significant role in the biogeochemical cycles of carbon, nitrogen and other nutrients (referred to as the “viral shunt”) in the ocean by lysis (Zimmerman et al., 2020). In these environments, phages may also modulate bacteria populations during infection, both by direct top-down control (predation), and via the takeover of bacterial machinery, resulting in significant changes to typical bacterial metabolism prior to infection (Weinbauer, 2004; Breitbart et al., 2018; Zimmerman et al., 2020). In environmental communities, where predatory phages often directly modulate bacterial populations, microbial populations can be modelled using traditional predator-prey Lotka-Volterra dynamics which depict an cyclical and inverse relationship between populations of predator (phage) and prey (bacterial host) (Weinbauer, 2004; Abedon, 2022). Environmental bacteriophages have been proposed to follow a “Kill the Winner” strategy, in which phage target the most abundant bacteria in order to release the most phage progeny, which massively reduces the bacterial population while a new bacteria fills the niche and is then targeted by their own respective phage, continuing the cycle indefinitely (Rodriguez-Brito et al., 2010; De Paepe et al., 2014)

In host-associated environments, the ability of phages to alter bacterial populations has been primarily studied within the field of phage therapy, in which bacteriophages are used as targeted therapies against specific pathogenic bacteria (Matsuzaki et al., 2005). The therapeutic use and impact of phages on dysbiosis and microbiome community structure is still a burgeoning field.

However, early research shows that phages may modulate bacterial symbionts in the gut microbiome (Shin, Whon & Bae, 2015; Hsu et al., 2019). Not only do bacterial species decline after exposure to specific targeted phages, but this effect may have cascading effects on the overall bacterial community, as the targeted bacteria may be repressing or promoting the growth of other species (Hsu et al., 2019).

Some research suggests the Kill-the-Winner strategy may be less applicable within the host-associated systems in certain scenarios due to the higher number of lysogenic phages, which do not always kill a bacterial host, but instead integrate as a prophage (Reyes et al., 2010). In these cases, reduced phage lysis may result in other strategies, such as “Piggyback-the-Winner”, in which a prophage enhances the survival of its host via various genetic benefits, allowing both host and prophage to proliferate in greater numbers (Silveira & Rohwer, 2016). In some scenarios however, phage may still manipulate populations of closely related bacterial communities through occasional spontaneous induction of prophages (which kills the bacterial host and releases progeny phage). This theory, known as “Kill-Your-Relatives,” suggests that although one member of the bacterial community is sacrificed to release infective phages, the result is more beneficial than negative. Closely related, competing taxa that do not carry the prophage may experience a population crash due to infection and lysis via the induced prophage (Gamage et al., 2003; De Paepe et al., 2014).

Given the significant impact of phage in certain environments like the ocean, as well as their ability to influence bacterial populations, one might ask whether microbiome-bacteriophage interactions have any connection to dysbiosis-related disease or have other clinical significance to the host. Although this question is still ambiguous, exploratory studies have found that human patients with dysbiosis-related diseases, such as Crohn’s disease and ulcerative colitis, also have unusual phage communities compared to controls, with reduced phage diversity and differing dominant phage families, but higher phage abundance overall (Wagner et al., 2013; Zuo et al., 2019). It is unknown if these phage community differences have a role in driving dysbiosis, or if they are symptomatic of dysbiosis. As many of gut phages exist as prophages, it is possible that greater environmental stress, such as inflammation, is what leads to an induction of the lytic phage cycle, or alternatively, that the induction of the lytic cycle releases cellular components, which then trigger host inflammation (Stecher, Maier & Hardt, 2013; Zuo et al., 2019).

2.3 Genetic contributions and potential benefits that phage provide to bacterial hosts

In addition to directly impacting the bacterial community structure via predation, phages are also capable of altering the microbiome genetically. This most often occurs via lysogenic conversion, a process in which the prophage provides a gene that alters bacterial phenotype, or via generalized transduction, in which a defective phage “infects” a new cell with novel genes from the original host, rather than the phage genome (De Paepe et al., 2014). Many pathogens owe some of their virulence to lysogenic conversion. For example, both *Vibrio cholera* (the causative agent of cholera) and Shiga-toxin producing *Escherichia coli* induce disease via prophage-encoded toxins, while phage integration of *Pseudomonas aeruginosa* in patients with cystic fibrosis can result in a

more virulent motile phenotype (Hyman, 2008; Davies et al., 2016). Surveys of gene composition in viral communities in certain environments, like pig digestive tracts and other agricultural-associated environments, may also harbor viruses capable of transferring antibiotic resistance genes (ARGs) to bacterial hosts, although the ability to accurately predict ARGs remains somewhat controversial (Enault et al., 2017; Wu et al., 2022; Zhang et al., 2023).

Even when lacking notable virulence factors, the mere presence of prophages may provide a significant benefit to a bacterial host in certain environments, even in spite of the risk of potential lysis (Paul, 2008). The presence of a prophage can often prevent further infection of similar phage, known as superinfection exclusion (Bondy-Denomy et al., 2016; Taylor et al., 2019). Additionally, as the bacterial host replicates, the phage may become defective due to the pressures of natural selection removing its ability to excise and lyse (Bobay, Touchon & Rocha, 2014). In many cases, bacteria instead have pieces of degraded or “domesticated” prophage genomes to retain phage immunity or other benefits without risk (Bobay, Touchon & Rocha, 2014). In some cases however, risk is inextricably tied to the provided benefit, as is the case with many prophage-encoded toxins; Shiga-toxin producing *E. coli* only releases said toxin while undergoing phage lysis (Gamage et al., 2003; Hyman, 2008).

In addition to superinfection exclusion, phages may also encode a variety of metabolic genes that have no clear link to phage infection, but may re-program bacterial metabolic processes to better suit the needs of phage replication during active infection (Breitbart et al., 2018). These genes, known as auxiliary metabolic genes, which are initially derived from bacterial hosts, may be additionally useful to the bacterial host while the phage is integrated as a prophage, particularly if the gene is novel or if the prophage has been degraded (Forcone et al., 2021; Huang et al., 2021). Research demonstrating that viral-encoded AMG has true benefits to the bacterial host is in its infancy, but an early study has shown that prophage-encoded genes associated with bioremediation are upregulated in the guts of earthworms living in contaminated environments, suggesting this notion is not without merit (Xia et al., 2023).

3. Studying bacteriophages in animal host systems

3.1 Using honey bee and amphibian microbiomes to study bacteriophage interactions

In general, the host-associated bacterial communities of animals play a significant role in the health and success of their hosts, providing beneficial functions, such as immune priming, disease resistance, or the production of metabolites (Harris et al., 2009; Lee & Mazmanian, 2010; Emery, Schmidt & Engel, 2017; Haase et al., 2018; Fallet et al., 2022). However, the vast majority of bacteriophage host-associated research has relied on either human or murine systems (Pride et al., 2012; Wagner et al., 2013; Hsu et al., 2019; Zuo et al., 2019; Boling et al., 2020; Baaziz et al., 2022; Inglis, Roach & Edwards, 2024). These studies have yielded insight on the ways bacteriophages might modulate microbiomes in their respective hosts but leaves our understanding of how phages may influence microbiomes incomplete. To clarify our understanding of the broader rules of phages in host-associated communities, the exploration of many types of host-associated

communities, in terms of both host and anatomical location, will be necessary. Two examples of such models are the gut microbiomes of the European honey bee (*Apis mellifera*) and broadly, the skin microbiome of amphibians. The bacterial components of each have been well studied, and both microbiomes can have a significant impact on the health of the animals (Kueneman et al., 2014; Engel et al., 2016; Emery, Schmidt & Engel, 2017; Bates et al., 2022). However, the mechanisms that shape bacterial community structures and function in these systems are less well understood. Investigating phages in these systems may yield deeper insight into how these host-associated communities influence the health of their specific hosts.

3.2 Honey bees as a model system for the gut microbiome

Honey bees are a critical agricultural species and pollinate 115 major crops (vanEngelsdorp & Meixner, 2010). In 2009 alone, honey bee pollination of crops in the United States contributed \$17 billion to the United States economy (Calderone, 2012). Despite their economic importance, managed honey bees populations have experienced several notable declines in the past few decades (vanEngelsdorp & Meixner, 2010). From 1961 to 2008, the number of honey-producing *A. mellifera* colonies declined by 61% in the US crops (vanEngelsdorp & Meixner, 2010). Although there has been an increase in the number of managed colonies since 2006, this growth is severely hindered by high levels of colony failure; replacing dead colonies is expensive and time-consuming (Kulhanek et al., 2017). The cause of this decline is multifaceted, and includes factors such as exposure to pesticides, changing climate, and reduction in habitat, in addition to a suite of pathogens that often co-occur in hives, and the *Varroa* mite (*Varroa destructor*), an external parasite and viral vector, which were first reported to parasitize the European honey bee in Russia in 1949, but became a significant issue in the USA in the late 1980s (vanEngelsdorp & Meixner, 2010; Mondet et al., 2020). To improve the health of honey bees, researchers began investigating ways in which the gut microbiome could be leveraged to either mitigate disease or enhance the health of colonies.

The majority of honey bee gut microbiome research has focused on adult worker bees, as they make up the majority of the hive. In the warm seasons, the aptly named caste of worker bees spend the first few weeks of their adult lives participating in domestic hive tasks (feeding larvae as nurse bees, attending to the queen, and maintaining and protecting the hive). After ~18 -21 days, they enter the second worker phase as foragers that collect food (nectar, honey) and water for the hive until death approximately ~21 days later (Abrol, 2013). In contrast, queens and drones have essentially a single task for their life-time: to reproduce. Newly adult queens leave the hive to go on mating flights, where they copulate with multiple males (drones) from other hives. Afterwards, they return to the hive and lay eggs for the remainder of their life. These caste roles are decided at or a few days after egg-laying, with unfertilized eggs hatching into drones and fertilized eggs hatching into worker bees or queens, depending on the specifics of developmental diet (Abrol, 2013). Regardless of their caste, all eggs hatch as larvae (white grubs) and are fed by young adult nurse bees until they become pupae. At the pupal stage, they are enclosed within a hive cell and undergo metamorphosis before emerging as adult bees (see Figure 1). Directly after emergence,

the gut microbiome of the new adult bee is considered to be nearly sterile, as the process of pupation involves an initial ejection of gut contents before metamorphosis begins (Engel et al., 2016). Afterwards, the guts of newly adult bees are quickly colonized by a highly conserved community of approximately nine bacterial phylotypes, which dominate the guts of their sister worker bees and the hive environment (Engel et al., 2016).

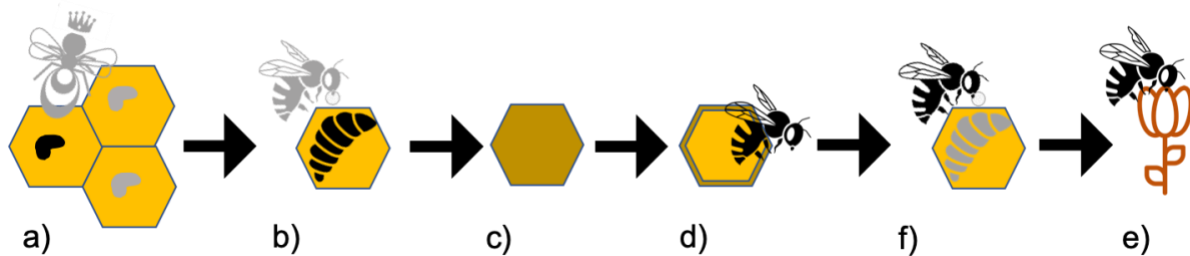


Figure 1: The life cycle of a single honey bee worker (drawn in black). First, a queen lays a fertilized egg in an individual brood cell **(a)**. The egg develops into a larvae which is fed by nurse worker bees **(b)**. Once the larvae is large enough, the cell is covered with a brood cap and the larvae becomes a pupae and undergoes metamorphosis **(c)**. Once the pupae develops into a young adult worker bee, it chews its way out of the cap and joins the hive **(d)**. As a young adult, the bee assists with in-hive tasks (such as feeding larval bees) **(f)**. After a few weeks, the bee begins leaving the hive to forage for food and water **(e)**.

Five out of these eight phylotypes are considered to be the “core” group of bacteria that are found in nearly every adult honey bee: *Snodgrassella alvi*, *Gillamella apicola*, *Bifidobacterium asteroides* and broadly, species from the group *Lactobacillus* Firmicutes-4 and Firmicutes-5, some of which have been recently taxonomically reclassified into new genera such as *Apilactobacillus* and *Bombilactobacillus* (Engel et al., 2016; Zheng et al., 2020). The remaining three (*Frischella perrara*, *Bartonella apis*, and *Bombella apis* (which now includes *Parasaccharibacter apium*)) represent a smaller part of the overall bacterial population and have lower prevalence across bees (Engel et al., 2016; Smith et al., 2021). These bacterial phylotypes are specific to different regions of the digestive tract, which consists of several distinct regions from the crop to the midgut to the ileum and rectum.

The hindgut of the bee (the ileum and the rectum) carries the highest bacterial load, particularly of the core bacterial species. The midgut often hosts the rarer commensals, which are specific to the midgut or associated with the hive environment (Anderson & Ricigliano, 2017). The midgut and hindgut are linked by the pylorus, which is colonized by *F. perrara*. While *F. perrara* appears to be the causative agent of scabbing in the pylorus, it elicits a strong immune response and may serve as a critical immune primer (Emery, Schmidt & Engel, 2017). Bees colonized by *F. perrara* exhibit a more diverse collection of antimicrobial peptides (AMPs) and pattern recognition

receptors (PRRs) which may provide greater protection against later microbial challenges (Emery, Schmidt & Engel, 2017). Although the microbiome of the honey bee is well characterized, less is known about the strain diversity within each phylotype or how specific strains of these phylotypes interact with each other (Ellegaard & Engel, 2019).

Furthermore, outside of the specific characterization of phages that target *Paenibacillus larvae*, a bacterial pathogen that causes the disease American foulbrood (Merrill et al., 2014; Beims et al., 2015), only a few assessments of honey bee gut phages have been conducted (Bonilla-Rosso et al., 2020; Deboutte et al., 2020; Busby et al., 2022). These initial surveys establish an excellent foundation for further phage studies, as they suggest honey bee gut phages are highly diverse, unique to individual hives, and carry potential metabolic genes.

3.4 Amphibians as a model system for the skin microbiome

Broadly, amphibian populations have experienced a notable decline in the past few decades, much of which can be contributed to global emergence of the fungal pathogen *Batrachochytrium dendrobatidis* (Bd) (Lips, 2016; Fisher & Garner, 2020). Bd targets the skin of amphibians, disrupting regular skin functions, such as osmoregulation and respiration (Van Rooij et al., 2015). The host range of Bd is incredibly vast, but often impacts species with varying degrees of severity, due to innate species immunity, local climate factors, and host microbiota (Alan Pounds et al., 2006; Harris et al., 2009; Van Rooij et al., 2015; Bates et al., 2022). The inhibition of Bd via bacterial skin microbiota was initially attributed to specific bacterial species, such as *Janthinobacterium lividum* or *Serratia* spp., which produce secondary antifungal metabolites like violacein and prodigiosin (Woodhams et al., 2018). However, other species of bacteria may also produce metabolites or other molecules that reduce infection and mortality from Bd (Becker et al., 2015). Additionally, in many animal host-associated systems, certain communities of skin microbiota may indirectly inhibit pathogens via the stimulation of the host immune system or through competition over resources and space (Reid et al., 2011; Becker et al., 2015).

Amphibian species and life stage (ex: tadpole, metamorph, adult) are strong predictors for bacterial communities found on amphibian skin (Kueneman et al., 2014). In addition to host species, environmental factors appear to be powerful drivers of microbial community composition (Kueneman et al., 2014; Hughey et al., 2017). However, nothing is known about the impact bacteriophages may have on the amphibian skin microbiome. Limited early studies of human skin viromes suggest that the majority of bacteriophages on skin are highly transient in nature, but phages able to successfully colonize may carry genes beneficial to the bacterial host to secure stability (Graham et al., 2023). If bacteriophages similarly use host-benefiting genes to colonize the amphibian skin microbiome, then it is plausible phages may be enriched for antifungal or other auxiliary metabolic genes.

4. Summarized doctoral research

The host-associated microbial communities of animals can have profound effects on the health of the host. However, our understanding of these impacts is hindered by our lack of understanding the viral mediators which both prey upon and partner with bacterial hosts. To continue unraveling the complex interactions and genetic exchanges between bacteria and phages in host-associated systems, I investigated the honey bee gut microbiome and the amphibian skin microbiome for evidence of interactions between bacteria and phages that could shape microbial communities. For my first chapter, I characterized the prophages of core honey bee gut bacteria, determining which species seemed to frequently partner with prophages (either intact or degraded), as well as identify the presence of auxiliary metabolic genes associated with carbohydrate breakdown. In my second chapter, I applied the techniques developed in my first chapter to a set of bacterial isolate genomes from the amphibian skin microbiome and searched for the presence of prophage-encoded antifungal genes, in addition to auxiliary metabolic genes. Additionally, this study compared prophages of amphibian-isolated to prophages of environmentally-isolated *Janthinobacterium lividum*. Finally, my third chapter investigated the viral communities of 12 honey bee apiaries distributed across a transect in the United States from Washington to Virginia. I assessed how distance and other environmental factors shape the taxonomic composition of the viromes and identified how these viruses may be impacting the bacterial community, either through the acquisition of beneficial genes or possibly via predation.

References

- Abedon ST. 2022. *Bacteriophages as Drivers of Evolution: An Evolutionary Ecological Perspective*. Cham: Springer International Publishing. DOI: 10.1007/978-3-030-94309-7.
- Abrol DP. 2013. Biology of the honeybee. In: *Beekeeping : A Compressive Guide to Bees and Beekeeping*. Jodhpur, India: Scientific Publisher [India];,.
- Alan Pounds J, Bustamante MR, Coloma LA, Consuegra JA, Fogden MPL, Foster PN, La Marca E, Masters KL, Merino-Viteri A, Puschendorf R, Ron SR, Sánchez-Azofeifa GA, Still CJ, Young BE. 2006. Widespread amphibian extinctions from epidemic disease driven by global warming. *Nature* 439:161–167. DOI: 10.1038/nature04246.
- Anderson KE, Ricigliano VA. 2017. Honey bee gut dysbiosis: a novel context of disease ecology. *Current Opinion in Insect Science* 22:125–132. DOI: 10.1016/j.cois.2017.05.020.
- Baaziz H, Baker ZR, Franklin HC, Hsu BB. 2022. Rehabilitation of a misbehaving microbiome: phages for the remodeling of bacterial composition and function. *iScience* 25:104146. DOI: 10.1016/j.isci.2022.104146.
- Bates KA, Sommer U, Hopkins KP, Shelton JMG, Wierzbicki C, Sergeant C, Tapley B, Michaels CJ, Schmeller DS, Loyau A, Bosch J, Viant MR, Harrison XA, Garner TWJ, Fisher MC. 2022. Microbiome function predicts amphibian chytridiomycosis disease dynamics. *Microbiome* 10:44. DOI: 10.1186/s40168-021-01215-6.
- Becker MH, Walke JB, Cikanek S, Savage AE, Mattheus N, Santiago CN, Minbiole KPC, Harris RN, Belden LK, Gratwicke B. 2015. Composition of symbiotic bacteria predicts survival in

- Panamanian golden frogs infected with a lethal fungus. *Proceedings of the Royal Society B: Biological Sciences* 282:20142881. DOI: 10.1098/rspb.2014.2881.
- Beims H, Wittmann J, Bunk B, Spröer C, Rohde C, Günther G, Rohde M, von der Ohe W, Steinert M. 2015. Paenibacillus larvae-Directed Bacteriophage HB10c2 and Its Application in American Foulbrood-Affected Honey Bee Larvae. *Applied and Environmental Microbiology* 81:5411–5419. DOI: 10.1128/AEM.00804-15.
- Berry D, Widder S. 2014. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Frontiers in Microbiology* 5. DOI: 10.3389/fmicb.2014.00219.
- Bobay L-M, Touchon M, Rocha EPC. 2014. Pervasive domestication of defective prophages by bacteria. *Proceedings of the National Academy of Sciences* 111:12127–12132. DOI: 10.1073/pnas.1405336111.
- Boling L, Cuevas DA, Grasis JA, Kang HS, Knowles B, Levi K, Maughan H, McNair K, Rojas MI, Sanchez SE, Smurthwaite C, Rohwer F. 2020. Dietary prophage inducers and antimicrobials: toward landscaping the human gut microbiome. *Gut Microbes* 11:721–734. DOI: 10.1080/19490976.2019.1701353.
- Bondy-Denomy J, Qian J, Westra ER, Buckling A, Guttman DS, Davidson AR, Maxwell KL. 2016. Prophages mediate defense against phage infection through diverse mechanisms. *The ISME Journal* 10:2854–2866. DOI: 10.1038/ismej.2016.79.
- Bonilla-Rosso G, Steiner T, Wichmann F, Bexkens E, Engel P. 2020. Honey bees harbor a diverse gut virome engaging in nested strain-level interactions with the microbiota. *Proceedings of the National Academy of Sciences*. DOI: 10.1073/pnas.2000228117.
- Brandt LJ, Aroniadis OC, Mellow M, Kanatzar A, Kelly C, Park T, Stollman N, Rohlke F, Surawicz C. 2012. Long-Term Follow-Up of Colonoscopic Fecal Microbiota Transplant for Recurrent Clostridium difficile Infection: *American Journal of Gastroenterology* 107:1079–1087. DOI: 10.1038/ajg.2012.60.
- Breitbart M, Bonnain C, Malki K, Sawaya NA. 2018. Phage puppet masters of the marine microbial realm. *Nature Microbiology* 3:754–766. DOI: 10.1038/s41564-018-0166-y.
- Busby TJ, Miller CR, Moran NA, Van Leuven JT. 2022. Global Composition of the Bacteriophage Community in Honey Bees. *mSystems* 7:e01195-21. DOI: 10.1128/msystems.01195-21.
- Cabrera-Perez J, Badovinac VP, Griffith TS. 2017. Enteric immunity, the gut microbiome, and sepsis: Rethinking the germ theory of disease. *Experimental Biology and Medicine* 242:127–139. DOI: 10.1177/1535370216669610.
- Calderone NW. 2012. Insect Pollinated Crops, Insect Pollinators and US Agriculture: Trend Analysis of Aggregate Data for the Period 1992–2009. *PLOS ONE* 7:e37235. DOI: 10.1371/journal.pone.0037235.
- Caporaso JG, Lauber CL, Costello EK, Berg-Lyons D, Gonzalez A, Stombaugh J, Knights D, Gajer P, Ravel J, Fierer N, Gordon JI, Knight R. 2011. Moving pictures of the human microbiome. *Genome Biology* 12:R50. DOI: 10.1186/gb-2011-12-5-r50.
- Carding S, Verbeke K, Vipond DT, Corfe BM, Owen LJ. 2015. Dysbiosis of the gut microbiota in disease. *Microbial Ecology in Health & Disease* 26. DOI: 10.3402/mehd.v26.26191.
- Chen VL, Kasper DL. 2014. Interactions between the intestinal microbiota and innate lymphoid cells. *Gut Microbes* 5:129–140. DOI: 10.4161/gmic.27289.
- Cho I, Blaser MJ. 2012. The human microbiome: at the interface of health and disease. *Nature Reviews Genetics* 13:260–270. DOI: 10.1038/nrg3182.
- Davies EV, James CE, Williams D, O'Brien S, Fothergill JL, Haldenby S, Paterson S, Winstanley C, Brockhurst MA. 2016. Temperate phages both mediate and drive adaptive evolution in pathogen

- biofilms. *Proceedings of the National Academy of Sciences of the United States of America* 113:8266–8271. DOI: 10.1073/pnas.1520056113.
- De Paepe M, Leclerc M, Tinsley CR, Petit M-A. 2014. Bacteriophages: an underestimated role in human and animal health? *Frontiers in Cellular and Infection Microbiology* 4. DOI: 10.3389/fcimb.2014.00039.
- Deboutte W, Beller L, Yinda CK, Maes P, Graaf DC de, Matthijssens J. 2020. Honey-bee-associated prokaryotic viral communities reveal wide viral diversity and a profound metabolic coding potential. *Proceedings of the National Academy of Sciences*. DOI: 10.1073/pnas.1921859117.
- Doumbo S, Tran TM, Sangala J, Li S, Doumtabe D, Kone Y, Traoré A, Bathily A, Sogoba N, Coulibaly ME, Huang C-Y, Ongoiba A, Kayentao K, Diallo M, Dramane Z, Nutman TB, Crompton PD, Doumbo O, Traore B. 2014. Co-infection of Long-Term Carriers of *Plasmodium falciparum* with *Schistosoma haematobium* Enhances Protection from Febrile Malaria: A Prospective Cohort Study in Mali. *PLoS Neglected Tropical Diseases* 8. DOI: 10.1371/journal.pntd.0003154.
- Eaton CJ, Cox MP, Scott B. 2011. What triggers grass endophytes to switch from mutualism to pathogenism? *Plant Science* 180:190–195. DOI: 10.1016/j.plantsci.2010.10.002.
- Ellegaard KM, Engel P. 2019. Genomic diversity landscape of the honey bee gut microbiota. *Nature Communications* 10:1–13. DOI: 10.1038/s41467-019-08303-0.
- Emery O, Schmidt K, Engel P. 2017. Immune system stimulation by the gut symbiont *Frischella perrara* in the honey bee (*Apis mellifera*). *Molecular Ecology* 26:2576–2590. DOI: 10.1111/mec.14058.
- Enault F, Briet A, Bouteille L, Roux S, Sullivan MB, Petit M-A. 2017. Phages rarely encode antibiotic resistance genes: a cautionary tale for virome analyses. *The ISME Journal* 11:237–247. DOI: 10.1038/ismej.2016.90.
- Engel P, Kwong WK, McFrederick Q, Anderson KE, Barribeau SM, Chandler JA, Cornman RS, Dainat J, de Miranda JR, Doublet V, Emery O, Evans JD, Farinelli L, Flenniken ML, Granberg F, Grasis JA, Gauthier L, Hayer J, Koch H, Kocher S, Martinson VG, Moran N, Munoz-Torres M, Newton I, Paxton RJ, Powell E, Sadd BM, Schmid-Hempel P, Schmid-Hempel R, Song SJ, Schwarz RS, vanEngelsdorp D, Dainat B. 2016. The Bee Microbiome: Impact on Bee Health and Model for Evolution and Ecology of Host-Microbe Interactions. *mBio* 7:e02164-15, /mbio/7/2/e02164-15.atom. DOI: 10.1128/mBio.02164-15.
- Faith JJ, Guruge JL, Charbonneau M, Subramanian S, Seedorf H, Goodman AL, Clemente JC, Knight R, Heath AC, Leibel RL, Rosenbaum M, Gordon JI. 2013. The Long-Term Stability of the Human Gut Microbiota. *Science* 341. DOI: 10.1126/science.1237439.
- Fallet M, Montagnani C, Petton B, Dantan L, de Lorgeril J, Comarmond S, Chaparro C, Toulza E, Boitard S, Escoubas J-M, Vergnes A, Le Grand J, Bulla I, Gueguen Y, Vidal-Dupiol J, Grunau C, Mitta G, Cosseau C. 2022. Early life microbial exposures shape the *Crassostrea gigas* immune system for lifelong and intergenerational disease protection. *Microbiome* 10:85. DOI: 10.1186/s40168-022-01280-5.
- Faust K, Raes J. 2012. Microbial interactions: from networks to models. *Nature Reviews Microbiology* 10:538–550. DOI: 10.1038/nrmicro2832.
- Fisher MC, Garner TWJ. 2020. Chytrid fungi and global amphibian declines. *Nature Reviews Microbiology* 18:332–343. DOI: 10.1038/s41579-020-0335-x.
- Forcone K, Coutinho FH, Cavalcanti GS, Silveira CB. 2021. Prophage Genomics and Ecology in the Family Rhodobacteraceae. *Microorganisms* 9:1115. DOI: 10.3390/microorganisms9061115.
- Gamage SD, Strasser JE, Chalk CL, Weiss AA. 2003. Nonpathogenic *Escherichia coli* Can Contribute to the Production of Shiga Toxin. *Infection and Immunity* 71:3107–3115. DOI: 10.1128/IAI.71.6.3107-3115.2003.

- Graham EH, Tom WA, Neujahr AC, Adamowicz MS, Clarke JL, Herr JR, Fernando SC. 2023. The persistence and stabilization of auxiliary genes in the human skin virome. *VIROLOGY JOURNAL* 20:49. DOI: 10.1186/s12985-023-02012-3.
- Groppe K, Steinger T, Sanders I, Schmid B, Wiemken A, Boller T. 1999. Interaction between the endophytic fungus *Epichloë bromicola* and the grass *Bromus erectus*: effects of endophyte infection, fungal concentration and environment on grass growth and flowering. *Molecular Ecology* 8:1827–1835. DOI: 10.1046/j.1365-294x.1999.00772.x.
- Haase S, Haghighi A, Wilck N, Müller DN, Linker RA. 2018. Impacts of microbiome metabolites on immune regulation and autoimmunity. *Immunology* 154:230–238. DOI: 10.1111/imm.12933.
- Harris RN, Brucker RM, Walke JB, Becker MH, Schwantes CR, Flaherty DC, Lam BA, Woodhams DC, Briggs CJ, Vredenburg VT, Minbiole KPC. 2009. Skin microbes on frogs prevent morbidity and mortality caused by a lethal skin fungus. *The ISME Journal* 3:818–824. DOI: 10.1038/ismej.2009.27.
- Henderson G, Cox F, Ganesh S, Jonker A, Young W, Janssen PH. 2015. Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. *Scientific Reports* 5:1–15. DOI: 10.1038/srep14567.
- Hsu BB, Gibson TE, Yeliseyev V, Liu Q, Lyon L, Bry L, Silver PA, Gerber GK. 2019. Dynamic Modulation of the Gut Microbiota and Metabolome by Bacteriophages in a Mouse Model. *Cell Host & Microbe* 25:803–814.e5. DOI: 10.1016/j.chom.2019.05.001.
- Huang D, Yu P, Ye M, Schwarz C, Jiang X, Alvarez PJJ. 2021. Enhanced mutualistic symbiosis between soil phages and bacteria with elevated chromium-induced environmental stress. *Microbiome* 9:150. DOI: 10.1186/s40168-021-01074-1.
- Hughey MC, Pena JA, Reyes R, Medina D, Belden LK, Burrowes PA. 2017. Skin bacterial microbiome of a generalist Puerto Rican frog varies along elevation and land use gradients. *PeerJ* 5:e3688. DOI: 10.7717/peerj.3688.
- Huse SM, Ye Y, Zhou Y, Fodor AA. 2012. A Core Human Microbiome as Viewed through 16S rRNA Sequence Clusters. *PLOS ONE* 7:e34242. DOI: 10.1371/journal.pone.0034242.
- Huseyin CE, O'Toole PW, Cotter PD, Scanlan PD. 2017. Forgotten fungi—the gut mycobiome in human health and disease. *FEMS Microbiology Reviews* 41:479–511. DOI: 10.1093/femsre/fuw047.
- Hyman P. 2008. Phage ecology of bacterial pathogenesis. In: Abedon ST ed. *Bacteriophage Ecology: Population Growth, Evolution, and Impact of Bacterial Viruses*. Advances in Molecular and Cellular Microbiology. Cambridge: Cambridge University Press, 353–386. DOI: 10.1017/CBO9780511541483.017.
- Inglis LK, Roach MJ, Edwards RA. 2024. Prophages: an integral but understudied component of the human microbiome. *Microbial Genomics* 10:001166. DOI: 10.1099/mgen.0.001166.
- Kawakita A, Mochizuki K, Kato M. 2015. Reversal of mutualism in a leafhopper–leafhopper moth association: the possible driving role of a third-party partner. *Biological Journal of the Linnean Society* 116:507–518. DOI: 10.1111/bij.12633.
- Khosravi A, Mazmanian SK. 2013. Disruption of the gut microbiome as a risk factor for microbial infections. *Current opinion in microbiology* 16:221–227. DOI: 10.1016/j.mib.2013.03.009.
- Kueneman JG, Parfrey LW, Woodhams DC, Archer HM, Knight R, McKenzie VJ. 2014. The amphibian skin-associated microbiome across species, space and life history stages. *Molecular Ecology* 23:1238–1250. DOI: 10.1111/mec.12510.
- Kulhanek K, Steinhauer N, Rennich K, Caron DM, Sagili RR, Pettis JS, Ellis JD, Wilson ME, Wilkes JT, Tarpy DR, Rose R, Lee K, Rangel J, vanEngelsdorp D. 2017. A national survey of managed honey bee 2015–2016 annual colony losses in the USA. *Journal of Apicultural Research* 56:328–340. DOI: 10.1080/00218839.2017.1344496.

- Lee YK, Mazmanian SK. 2010. Has the Microbiota Played a Critical Role in the Evolution of the Adaptive Immune System? *Science* 330:1768–1773. DOI: 10.1126/science.1195568.
- Lips KR. 2016. Overview of chytrid emergence and impacts on amphibians. *Philosophical Transactions of the Royal Society B: Biological Sciences* 371:20150465. DOI: 10.1098/rstb.2015.0465.
- Matsuzaki S, Rashel M, Uchiyama J, Sakurai S, Ujihara T, Kuroda M, Ikeuchi M, Tani T, Fujieda M, Wakiguchi H, Imai S. 2005. Bacteriophage therapy: a revitalized therapy against bacterial infectious diseases. *Journal of Infection and Chemotherapy* 11:211–219. DOI: 10.1007/s10156-005-0408-9.
- Merrill BD, Grose JH, Breakwell DP, Burnett SH. 2014. Characterization of *Paenibacillus* larvae bacteriophages and their genomic relationships to firmicute bacteriophages. *BMC Genomics* 15:745–762. DOI: 10.1186/1471-2164-15-745.
- Mondet F, Beaurepaire A, McAfee A, Locke B, Alaux C, Blanchard S, Danka B, Le Conte Y. 2020. Honey bee survival mechanisms against the parasite *Varroa destructor*: a systematic review of phenotypic and genomic research efforts. *International Journal for Parasitology* 50:433–447. DOI: 10.1016/j.ijpara.2020.03.005.
- Nishimura R, Hayashi M, Wu G-J, Kouchi H, Imaizumi-Anraku H, Murakami Y, Kawasaki S, Akao S, Ohmori M, Nagasawa M, Harada K, Kawaguchi M. 2002. HAR1 mediates systemic regulation of symbiotic organ development. *Nature* 420:426–429. DOI: 10.1038/nature01231.
- Okabe K, Makino S. 2010. Conditional Mutualism Between *Allodynerus delphinalis* (Hymenoptera: Vespidae) and *Ensliniella parasitica* (Astigmata: Winterschmidtidae) May Determine Maximum Parasitic Mite Infestation. *Environmental Entomology* 39:424–429. DOI: 10.1603/EN09208.
- Olszak T, An D, Zeissig S, Vera MP, Richter J, Franke A, Glickman JN, Siebert R, Baron RM, Kasper DL, Blumberg RS. 2012. Microbial Exposure During Early Life Has Persistent Effects on Natural Killer T Cell Function. *Science* 336:489–493. DOI: 10.1126/science.1219328.
- Paul JH. 2008. Prophages in marine bacteria: dangerous molecular time bombs or the key to survival in the seas? *The ISME Journal* 2:579–589. DOI: 10.1038/ismej.2008.35.
- Pride DT, Salzman J, Haynes M, Rohwer F, Davis-Long C, White RA, Loomer P, Armitage GC, Relman DA. 2012. Evidence of a robust resident bacteriophage population revealed through analysis of the human salivary virome. *The ISME Journal* 6:915–926. DOI: 10.1038/ismej.2011.169.
- Regus JU, Gano KA, Hollowell AC, Sofish V, Sachs JL. 2015. Lotus hosts delimit the mutualism–parasitism continuum of *Bradyrhizobium*. *Journal of Evolutionary Biology* 28:447–456. DOI: 10.1111/jeb.12579.
- Reid G, Younes JA, Van der Mei HC, Gloor GB, Knight R, Busscher HJ. 2011. Microbiota restoration: natural and supplemented recovery of human microbial communities. *Nature Reviews. Microbiology* 9:27–38. DOI: 10.1038/nrmicro2473.
- Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, Rohwer F, Gordon JI. 2010. Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* 466:334–338. DOI: 10.1038/nature09199.
- Rodriguez-Brito B, Li L, Wegley L, Furlan M, Angly F, Breitbart M, Buchanan J, Desnues C, Dinsdale E, Edwards R, Felts B, Haynes M, Liu H, Lipson D, Mahaffy J, Belen Martin-Cuadrado A, Mira A, Nulton J, Pasic L, Rayhawk S, Rodriguez-Mueller J, Rodriguez-Valera F, Salamon P, Srinagesh S, Thingstad TF, Tran T, Thurber RV, Willner D, Youle M, Rohwer F. 2010. Viral and microbial community dynamics in four aquatic environments. *Isme Journal* 4:739–751. DOI: 10.1038/ismej.2010.1.
- Rosario K, Breitbart M. 2011. Exploring the viral world through metagenomics. *Current Opinion in Virology* 1:289–297. DOI: 10.1016/j.coviro.2011.06.004.

- Rosenberg E, Koren O, Reshef L, Efrony R, Zilber-Rosenberg I. 2007. The role of microorganisms in coral health, disease and evolution. *Nature Reviews Microbiology* 5:355–362. DOI: 10.1038/nrmicro1635.
- Sachs JL, Quides KW, Wendlandt CE. 2018. Legumes versus rhizobia: a model for ongoing conflict in symbiosis. *New Phytologist* 219:1199–1206. DOI: 10.1111/nph.15222.
- Sachs JL, Simms EL. 2006. Pathways to mutualism breakdown. *Trends in Ecology & Evolution* 21:585–592. DOI: 10.1016/j.tree.2006.06.018.
- Scarpellini E, Ianiro G, Attili F, Bassanelli C, De Santis A, Gasbarrini A. 2015. The human gut microbiota and virome: Potential therapeutic implications. *Digestive and Liver Disease* 47:1007–1012. DOI: 10.1016/j.dld.2015.07.008.
- Shafquat A, Joice R, Simmons S, Huttenhower C. 2014. Functional and phylogenetic assembly of microbial communities in the human microbiome. *Trends in microbiology* 22:261–266. DOI: 10.1016/j.tim.2014.01.011.
- Sharon G, Segal D, Ringo JM, Hefetz A, Zilber-Rosenberg I, Rosenberg E. 2010. Commensal bacteria play a role in mating preference of *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences* 107:20051–20056. DOI: 10.1073/pnas.1009906107.
- Shin N-R, Whon TW, Bae J-W. 2015. Proteobacteria: microbial signature of dysbiosis in gut microbiota. *Trends in Biotechnology* 33:496–503. DOI: 10.1016/j.tibtech.2015.06.011.
- Silveira CB, Rohwer FL. 2016. Piggyback-the-Winner in host-associated microbial communities. *npj Biofilms and Microbiomes* 2:1–5. DOI: 10.1038/npjbiofilms.2016.10.
- Smith EA, Anderson KE, Corby-Harris V, McFrederick QS, Parish AJ, Rice DW, Newton ILG. 2021. Reclassification of seven honey bee symbiont strains as *Bombella apis*. *International Journal of Systematic and Evolutionary Microbiology* 71:004950. DOI: 10.1099/ijsem.0.004950.
- Stecher B, Maier L, Hardt W-D. 2013. “Blooming” in the gut: how dysbiosis might contribute to pathogen evolution. *Nature Reviews Microbiology* 11:277–284. DOI: 10.1038/nrmicro2989.
- Taylor VL, Fitzpatrick AD, Islam Z, Maxwell KL. 2019. Chapter One - The Diverse Impacts of Phage Morons on Bacterial Fitness and Virulence. In: Kielian M, Mettenleiter TC, Roossinck MJ eds. *Advances in Virus Research*. Academic Press, 1–31. DOI: 10.1016/bs.aivir.2018.08.001.
- Van Rooij P, Martel A, Haesebrouck F, Pasmans F. 2015. Amphibian chytridiomycosis: a review with focus on fungus-host interactions. *Veterinary Research* 46:137. DOI: 10.1186/s13567-015-0266-0.
- vanEngelsdorp D, Meixner MD. 2010. A historical review of managed honey bee populations in Europe and the United States and the factors that may affect them. *Journal of Invertebrate Pathology* 103:S80–S95. DOI: 10.1016/j.jip.2009.06.011.
- Wagner J, Maksimovic J, Farries G, Sim WH, Bishop RF, Cameron DJ, Catto-Smith AG, Kirkwood CD. 2013. Bacteriophages in Gut Samples From Pediatric Crohn’s Disease Patients Metagenomic Analysis Using 454 Pyrosequencing. *Inflammatory Bowel Diseases* 19:1598–1608. DOI: 10.1097/MIB.0b013e318292477c.
- Weinbauer MG. 2004. Ecology of prokaryotic viruses. *FEMS Microbiology Reviews* 28:127–181. DOI: 10.1016/j.femsre.2003.08.001.
- Woodhams DC, LaBumbard BC, Barnhart KL, Becker MH, Bletz MC, Escobar LA, Flechas SV, Forman ME, Iannetta AA, Joyce MD, Rabemananjara F, Gratwicke B, Vences M, Minbiole KPC. 2018. Prodigiosin, Violacein, and Volatile Organic Compounds Produced by Widespread Cutaneous Bacteria of Amphibians Can Inhibit Two Batrachochytrium Fungal Pathogens. *Microbial Ecology* 75:1049–1062. DOI: 10.1007/s00248-017-1095-7.

- Wu R, Cao Z, Jiang Y, Chen W, Sun Y, Li Q, Mi J, Deng L, Liao X, Feng Y, Lan T, Ma J. 2022. Early life dynamics of ARG and MGE associated with intestinal virome in neonatal piglets. *Veterinary Microbiology* 274:109575. DOI: 10.1016/j.vetmic.2022.109575.
- Xia R, Sun M, Balcázar JL, Yu P, Hu F, Alvarez PJJ. 2023. Benzo[a]pyrene stress impacts adaptive strategies and ecological functions of earthworm intestinal viromes. *The ISME Journal* 17:1004–1014. DOI: 10.1038/s41396-023-01408-x.
- Youle M, Knowlton N, Rohwer F, Gordon J, Relman DA. 2013. Superorganisms and Holobionts: Looking for a term for the functional entity formed by a macrobe and its associated symbiotic microbes and viruses? The term is “holobiont.” *Microbe Magazine* 8:152–153. DOI: 10.1128/microbe.8.152.1.
- Zhang Y, Kitazumi A, Liao Y-T, de los Reyes BG, Wu VCH. 2023. Metagenomic investigation reveals bacteriophage-mediated horizontal transfer of antibiotic resistance genes in microbial communities of an organic agricultural ecosystem. *Microbiology Spectrum* 11:e00226-23. DOI: 10.1128/spectrum.00226-23.
- Zheng J, Wittouck S, Salvetti E, Franz CMAP, Harris HMB, Mattarelli P, O’Toole PW, Pot B, Vandamme P, Walter J, Watanabe K, Wuyts S, Felis GE, Gänzle MG, Lebeer S. 2020. A taxonomic note on the genus *Lactobacillus*: Description of 23 novel genera, emended description of the genus *Lactobacillus* Beijerinck 1901, and union of *Lactobacillaceae* and *Leuconostocaceae*. *International Journal of Systematic and Evolutionary Microbiology* 70:2782–2858. DOI: 10.1099/ijsem.0.004107.
- Zimmerman AE, Howard-Varona C, Needham DM, John SG, Worden AZ, Sullivan MB, Waldbauer JR, Coleman ML. 2020. Metabolic and biogeochemical consequences of viral infection in aquatic ecosystems. *Nature Reviews Microbiology* 18:21–34. DOI: 10.1038/s41579-019-0270-x.
- Zuo T, Lu X-J, Zhang Y, Cheung CP, Lam S, Zhang F, Tang W, Ching JYL, Zhao R, Chan PKS, Sung JJY, Yu J, Chan FKL, Cao Q, Sheng J-Q, Ng SC. 2019. Gut mucosal virome alterations in ulcerative colitis. *Gut* 68:1169–1179. DOI: 10.1136/gutjnl-2018-318131.

Chapter Two: Characterization of prophages in bacterial genomes from the honey bee (*Apis mellifera*) gut microbiome

Emma K. Bueren, Alaina R. Weinheimer, Frank O. Aylward, Bryan B. Hsu, David C. Haak, and Lisa K. Belden

Bueren EK, Weinheimer AR, Aylward FO, et al (2023) Characterization of prophages in bacterial genomes from the honey bee (*Apis mellifera*) gut microbiome. PeerJ 11:e15383.
<https://doi.org/10.7717/peerj.15383>

Abstract

The gut of the European honey bee (*Apis mellifera*) possesses a relatively simple bacterial community, but little is known about its community of prophages (temperate bacteriophages integrated into the bacterial genome). Although prophages may eventually begin replicating and kill their bacterial hosts, they can also sometimes be beneficial for their hosts by conferring protection from other phage infections or encoding genes in metabolic pathways and for toxins. In this study, we explored prophages in 17 species of core bacteria in the honey bee gut and two honey bee pathogens. Out of the 181 genomes examined, 431 putative prophage regions were predicted. Among core gut bacteria, the number of prophages per genome ranged from 0 – 7 and prophage composition (the compositional percentage of each bacterial genome attributable to prophages) ranged from 0 – 7%. *Snodgrassella alvi* and *Gilliamella apicola* had the highest median prophages per genome (3.0 ± 1.46 ; 3.0 ± 1.59), as well as the highest prophage composition ($2.58\% \pm 1.4$; $3.0\% \pm 1.59$). The pathogen *Paenibacillus larvae* had a higher median number of prophages (8.0 ± 5.33) and prophage composition ($6.40\% \pm 3.08$) than the pathogen *Melissococcus plutonius* or any of the core bacteria. Prophage populations were highly specific to their bacterial host species, suggesting most prophages were acquired recently relative to the divergence of these bacterial groups. Furthermore, functional annotation of the predicted genes encoded within the prophage regions indicates that some prophages in the honey bee gut encode additional benefits to their bacterial hosts, such as genes in carbohydrate metabolism. Collectively, this survey suggests that prophages within the honey bee gut may contribute to the maintenance and stability of the honey bee gut microbiome and potentially modulate specific members of the bacterial community, particularly *S. alvi* and *G. apicola*.

Introduction

Bacteriophages, viruses that infect bacteria, are ubiquitous and the most numerous biological entities on Earth (Hendrix, 2002). In animals, bacteriophages are known to shape microbial communities, such as the gut or skin microbiome (Hannigan et al., 2015). Bacteriophages are

broadly classified in two forms, based on reproductive strategy. Lytic bacteriophages infect and immediately kill their bacterial hosts via cell lysis. In contrast, temperate bacteriophages can undergo either a lytic cycle or a lysogenic cycle, in which the phage integrates into the host genome as a prophage and replicates along with the hosts genome until its triggered to revert to a lytic lifecycle. Many phages found in animal microbial communities are lysogenic (Minot et al., 2011; Kim & Bae, 2018). Some prophages encode genes in auxiliary metabolic pathways or encode non-metabolic accessory genes ("morons"), like virulence factors, that are beneficial to their bacterial hosts and can enhance bacterial fitness (Fortier & Sekulovic, 2013; Forcone et al., 2021; Huang et al., 2021). For example, the human pathogens *Vibrio cholera* and *E. coli* O157:H7 can cause disease via toxins encoded by prophages (Fortier & Sekulovic, 2013).

While prophages in pathogenic bacteria have been extensively studied, the roles of prophages in other animal-associated bacteria are less clear. It is generally difficult to characterize phage-bacteria interactions in the mammalian gut due to the complexity of these microbial communities, although inroads have been made in murine models (Hsu et al., 2019). In contrast, the bacterial community in the gut of honey bees (*Apis mellifera*) is significantly more constrained, consisting primarily of a core group of nine bacterial phylotypes: *Bartonella apis*, *Bombella apis* (previously known as *Parasaccharibacter apium*), *Frischella perrara*, *Snodgrassella alvi*, *Bifidobacterium* spp., *Bombilactobacillus* spp. (also referred to as *Lactobacillus* Firmicutes-4), *Lactobacillus* Firmicutes-5, *Gilliamella apicola*, and *Commensalibacter* spp. (also known as Alpha2.1) (Kwong & Moran, 2016). Some of these core microbiota, like *Gilliamella apicola*, *Lactobacillus* Firmicutes-5, *Bombilactobacillus* spp., and *Bifidobacterium* spp., ferment a wide variety of carbohydrates, while others, like *F. perrara*, stimulate bee immune function (Kwong & Moran, 2016; Emery, Schmidt & Engel, 2017).

Studies using both culture and sequence-based approaches reveal the numerous roles of phages in honey bee-associated bacteria. For instance, some strains of the pathogen *Paenibacillus larvae*, which causes American Foulbrood, acquire additional virulence through the prophage encoded toxin *P1x1* (Ebeling, Fünfhaus & Genersch, 2021). Additionally, many phage targeting *P. larvae* have been isolated and characterized for the purpose of experimental phage therapy (Beims et al., 2015; Abraham et al., 2016; Stamereilers et al., 2018; Ribeiro et al., 2019). To capture the broader diversity of phages present in honey bee guts and include those that cannot be easily isolated, recent studies have used metagenomic surveys of viruses, in which all double-stranded DNA viruses in a sample are sequenced to characterize the "virome". These virome studies have shown that the phage community (consisting of both temperate and lytic phages) is far more diverse than the bacterial community and appears to encode auxiliary metabolic genes (Bonilla-Rosso et al., 2020; Deboutte et al., 2020; Busby et al., 2022). The majority of phages in these viromes were novel, and thus were not taxonomically classified. Those that were classified were primarily from the recently dissolved *Podoviridae*, *Myoviridae*, and *Siphoviridae* families (Bonilla-Rosso et al., 2020; Deboutte et al., 2020), leaving their current taxonomic standing unknown. However, viromes may not fully capture the phages present in the bee gut that exist primarily as prophages

because the preparation of viromes involves filtering out bacterial cells to enrich for free viral particles (Huang et al., 2021).

As prophages can both protect the bacterial host from infection by other phages and improve host fitness by expanded metabolic versatility, they have the potential to impact the physiology, defense, and evolution of core honey bee gut bacteria (Deboutte et al., 2020). To understand the roles of prophages in the regulation and function of honey bee gut bacteria, we examined publicly available honey bee gut bacterial genomes for prophage regions and compared the occurrence of prophages among bacterial species. We then assessed prophage diversity with pairwise average nucleotide identity calculations and gene-sharing networks with reference phages. To explore potential benefits the prophages confer their hosts, we functionally annotated their genes and searched for those in relevant metabolic pathways.

Materials and Methods

Bacterial Genomes. Publicly available genomes (complete and fragmented) of bacteria originally isolated from the gut of the honey bee, *A. mellifera*, were downloaded from GenBank between March and July, 2020 (n = 181) (Table 1, Table S1). The nine core bacterial phylotypes are represented in this dataset (n = 151 total strains across 17 species), as were two bacterial pathogens (n=30). Each genome was from a uniquely named strain, and in the case of duplicate strain names in NCBI, the most complete and recent genome was used.

Prophage Identification. To detect putative prophages, we used a combination of VirSorter2 (v2.2.2; (Guo et al., 2021, p. 2)), CheckV (v.0.7.0 (Nayfach et al., 2021)) and VIBRANT (v1.2.1; (Kieft, Zhou & Anantharaman, 2020)). First, all genomes were analyzed with VirSorter2 (settings --include-groups "dsDNAphage,ssDNA,NCLDV,laviviridae"). Resulting viral regions were retained if they scored at least 0.5 for double stranded DNA phage (dsDNA phage, n = 539). Host genome regions flanking these viral sequences were then trimmed with the CheckV 'contamination' command. To exclude highly degraded phages, trimmed sequences were retained only if they were at least 5 kb in length (n = 462). Of these resulting sequences, a region was considered a putative prophage if it scored at least 0.9 with VirSorter2 (n = 357). Additionally, those with VirSorter scores of 0.5-0.9 were further analyzed with VIBRANT and were retained as putative prophage if VIBRANT also classified these sequences as virus (n = 74). This resulted in a final set of 431 putative prophages.

Prophage Abundance and Prophage Composition in Honey Bee Bacterial Isolates. Two metrics for prophage presence were assessed: the absolute abundance of prophages (total number of distinct regions) in a single bacterial genome and the portion of the bacterial genome that was composed of prophage (referred to as prophage composition). Differences in prophage abundance and composition among bacterial species were tested using Kruskal-Wallis tests. All graphs were visualized using ggplot2 and PNWColors in R (Wickham, 2016; Lawlor, 2020).

Larger host genomes often contain more prophage regions (Touchon, Bernheim & Rocha, 2016). We examined this pattern with host genome size vs. prophage abundance and host genome size vs. prophage composition using Spearman's correlation with the R library package *cocor* (Diedenhofen & Musch, 2015) and the Meng, Rosenthal and Rubin's z test (Meng, Rosenthal & Rubin, 1992; Myers & Sirois).

Functional Annotation of Prophage Region Genes and Detection of Intact Phages. The 431 putative prophages were first annotated using *prokka* (v. 1.14.6) against the Prokaryotic Virus Remote Homologous Groups database (PHROGs) database (downloaded from <http://millardlab.org/2021/11/21/phage-annotation-with-phrogs/> on July 20, 2022) (Seemann, 2014; Terzian et al., 2021). If an amino acid sequence had no hit to the PHROGs database (e value < 10⁻⁶), or was classified by PHROGs as “other” or “unknown function”, additional functions were then detected by aligning the amino acid sequences to hidden Markov profiles from the EggNOG 5.0 databases for bacteria, eukarya, archaea, and viruses (Huerta-Cepas et al., 2019) using *hmmsearch* (-E 0.001) with HMMER (v. 3.1b2) (hmmerr.org, (Eddy, 2011)). Hits with bitscores above 30 were retained (Roux et al., 2016).

Functional composition of prophage genes were compared based on the COG category assignment within the EggNOG annotation files for bacteria, eukarya, archaea, and viruses databases. EggNOG hits were manually re-assigned to the category of “Phage-associated” if the descriptive EggNOG name included the words: phage, baseplate, capsid, integrase, tail, tape, lysozyme, portal, holin, N-Acetylmuramoyl-L-alanine amidase, transposase, virus, or viral. Sequences that did not match to anything in the EggNOG database, or that were classified as belonging to multiple COG categories, were re-classified as COG Category S, Function Unknown. Functional composition of prophages using the combined results of PHROGs and EggNOG were then compared among bacterial host species by summing the proteins of all prophages detected in genomes of that species and visualized with the packages *ggplot2* and *microshades* (Wickham, 2016; Dahl et al., 2022).

To estimate intactness of prophage regions, trimmed sequences were predicted as intact using the following (i) those identified as intact by PHASTER (Arndt et al., 2016), (ii) those that encoded at least 3 or more different phage hallmark genes (see: cornerstone genes in (Zhou et al., 2011)) and additionally encoded an integrase and/or transposase.

Identification of Unique Putative Prophages. To identify prophage regions likely belonging to the same population (boundary defined in (Roux et al., 2019)), genomes of the 431 putative prophages were dereplicated via pairwise average nucleotide identity (ANI) analysis using *dRep* (v3.0.0 (Olm et al., 2017)) (settings: --ignoreGenomeQuality -l 5000 -pa 0.95 -sa 0.95 -nc 0.85) with *LASTn* (v1080; (Kielbasa et al., 2011)). Prophage pairs were considered to be from the same population if the ANI was above 95% over 85% of the shorter sequence (Roux et al., 2019). The representative prophage region for each population was selected with *dRep*. In one case, a population cluster contained two prophages that were identified from different bacterial host species. In that case, both prophage sequences were retained as representatives. As a result, a total

of 237 putative prophage regions were unique to themselves and did not share high similarity with other sequences. The remaining 194 sequences had high similarity with at least one other prophage sequence, resulting in 66 distinct populations of phage. After dereplication, a total of 303 representative phage were used for downstream functional annotation and taxonomic analysis.

Viral Clustering and Classification. Phages lack a universally conserved, high-resolution phylogenetic marker gene to classify them. As such, prophages are often grouped into clusters via gene-sharing network-based approaches or whole genome sequence alignments to reference phage (Bolduc et al., 2017). Dereplicated prophages in this study (n = 303), along with 24,289 phage reference genomes downloaded from the INfrastructure for a PHAge REference Database (INPHARED) on January 25, 2023 (Cook et al., 2021) were first classified with gene-sharing network approaches using vConTACT2 (v0.9.22; --rel-mode Diamond --db 'None' --pcs-mode MCL --vcs-mode ClusterON). vContact2 (v0.9.22) groups phages into subclusters, indicating phage are likely to be within the same viral genera, and broader viral clusters that indicate subfamily relatedness (relatedness somewhere between family and genus level) (Bin Jang et al., 2019). The resulting networks were visualized using edge-weighted spring-embedded models in Cytoscape (Shannon et al., 2003; Bolduc et al., 2017; Bin Jang et al., 2019). Although vConTACT2 can provide taxonomic classification if target phage cluster with reference phage, for our dataset, few of the resulting clusters included references phages that enabled classification at any taxonomic level. Therefore, while vConTACT2 gene-sharing networks were used to assess the relationships among honey bee prophages, we ended up relying more on whole genome alignments to reference phage to assign broader taxonomy.

To assign taxonomy based on similarities to reference viral groups, we compared the average amino acid identity (AAI) of proteins between our dereplicated putative prophages and phage from a dereplicated *Caudoviricetes* database (amino acids downloaded from the INPHARED on November 5, 2022, taxonomic metadata downloaded from INPHARED on March 16, 2023) (Cook et al., 2021). The 16,253 reference genomes were first dereplicated with dRep (--ignoreGenomeQuality -p 32 -l 20000 -pa 0.90 --SkipSecondary inphared_derepout_90mash) resulting in 2,600 representative reference genomes. Amino acid sequences of proteins encoded by the reference were predicted using prodigal (default on each genome). A BLAST database was made of the reference proteins (makeblastdb). The honey bee gut prophages protein predictions (produced via prodigal, described above) were aligned to this reference database with BLASTp (E value 0.001). Taxonomic class was assigned at the most precise rank possible using each prophage's AAI similarity to its top hitting reference phage. If a putative prophage and a reference phage shared greater than 70% AAI across 85% of proteins, it was classified as belonging to the same genus as the reference (Turner, Kropinski & Adriaenssens, 2021). If a phage did not meet this threshold but still shared at least 30% AAI over 50% of proteins, it was classified as the same family as the reference phage (Turner, Kropinski & Adriaenssens, 2021). If a prophage could not be classified using the above thresholds, but its top reference hit was to a *Caudoviricetes*

class phage, it was classified broadly as *Caudoviricetes*. Any putative prophage with a top reference hit to a completely unclassified phage was labeled as Unclassified.

Results

Distribution of Prophages Across Core Bacterial Hosts. Bacterial host species significantly varied in both the number and composition of prophages across all bacterial isolates, both core and pathogens (Kruskal-Wallis, $p < 0.0001$; Fig. 1 & Table S2). A total of 269 high-confidence prophage regions were predicted across the 151 core bacterial isolates before dereplication. Bacterial genomes ranged in size from 1.3 Mb – 3.5 Mb (Table S2). Among the core bacterial species, the number of prophages per isolate varied from 0 to 7 and prophage composition of the bacterial genome varied from 0 – 7.01% (Table S2).

The core bacteria with the highest median number of prophages and prophage composition was *S. alvi* (3.0 ± 1.46 prophages; $2.58\% \pm 1.40$ prophage composition) and *G. apicola* (3.0 ± 1.58 prophages; $2.04\% \pm 1.48$ prophage composition) (Fig. 1A, Fig. 1B, Table S2). Notably, while *Bombella apis*, *Bartonella apis*, and *F. perrara* had the second highest median number of prophages per genome, *L. melliventris* had the second highest prophage composition (Fig. 1A; Fig. 1B; Table S2). All other core species had a median number of 1 prophage or less and a prophage composition of less than 1% (Fig. 1A; Fig. 1B, Table S2). These included *Bifidobacterium asteroides*, *Lactobacillus mellifer*, *Lactobacillus apis*, and *Lactobacillus kimbladii*, which had a median of 0 prophages per genome (Fig. 1A; Fig. 1B; Table S2).

Of these 269 putative prophage regions, 90 were estimated to be intact. Bacterial hosts with the highest median number of estimated intact prophages were *G. apicola* (1.0 ± 1.1), *S. alvi* (1.0 ± 0.60) and *L. melliventris* (1 ± 0.58) (Fig. 1C; Table S2). All three isolates of *L. melliventris*, 71% of *G. apicola* isolates ($n = 38$) and 56% of *S. alvi* isolates ($n = 34$) possessed at least one estimated intact prophage (Table S3). The second highest median of intact prophage was found in *L. kullabergensis* (0.5 ± 0.71), with one of the two *L. kullabergensis* isolates possessing an intact prophage (Fig. 1C; Table S3). All other core species had a median of 0 intact prophages (Fig. 1C, Table S2). However, intact prophages did still occur in these isolates at lower frequencies (Table S3). Intact prophage composition varied slightly compared to prophage abundance, with *L. melliventris* having the highest composition ($1.81\% \pm 1.25$), followed by *S. alvi* ($1.16\% \pm 1.09$), *L. kullabergensis* ($0.97\% \pm 1.37$), and *G. apicola* ($0.89\% \pm 1.31$) (Fig. 1D; Table S2).

Distribution of Prophages Across Pathogenic Bacterial Hosts. A total of 162 prophage regions were predicted across the 30 bacterial isolates from the two pathogens, *P. larvae* and *M. plutonius*, before dereplication. The bacterial isolate genomes ranged in size from 2.0 Mb – 4.8 Mb (Table S2). The number of predicted prophage regions ranged from 1 to 22, and prophage composition ranged from 0.96% - 13.85% of the bacterial genome (Table S2). *P. larvae* had the highest median number of prophages per genome (8.0 ± 5.33) and prophage composition ($6.40\% \pm 3.08$) across all bacterial species, including the core bacteria (Fig. 1A; Fig. 1B; Table S2). The median number

of prophages found per genome of *M. plutonius* was 2.0 ± 0.64 , and its median prophage composition was $2.77\% \pm 0.83$ (Fig. 1A; Fig. 1B; Table S2).

Of the 13 *P. larvae* isolates analyzed, all had at least one intact prophage, with a median of 4 ± 1.98 intact prophage per isolate (TableS-Freq; Fig. 1C; Table S2). *M. plutonius* had a median of 1 ± 0.35 intact prophage, with 16 of 17 isolates possessing an intact prophage (Fig. 1C; Table S2; TableS-Freq). The total intact prophage composition of *P. larvae* ($4.57\% \pm 3.10$) was higher than *M. plutonius* ($1.71\% \pm 0.97$) (Fig. 1D; Table S2).

Prophage Abundance Is More Correlated with Bacterial Genome Size than Prophage Composition. Bacterial genome size was positively correlated with both the number of prophages (Spearman's $\rho = 0.55$, $p \leq 1.6e-15$, Fig. 2A) and prophage composition (Spearman's $\rho = 0.34$, $p = 2e-6$, Fig. 2B). Meng, Rosenthal and Rubin's correlation comparison determined that bacterial genome size had a stronger correlation with the number of prophages than with prophage composition ($z = 5.9819$, $p\text{-value} < 0.001$, 95% CI: 0.18-0.35) (Meng, Rosenthal & Rubin, 1992).

Prophage-encoded genes varied by bacterial host. Nearly half (48.4%) of all predicted protein-encoding genes of the dereplicated prophages were classified broadly as phage-associated. A total of 2.3% of all genes were classified by PHROGs specifically as “Moron, auxiliary metabolic genes, and host takeover” (Table S5). *S. alvi* had the highest frequency of genes in this category (66 genes, 3.3% frequency), due to its large collection of phage toxins and antitoxins (Fig. 3). These toxins shared homology to Doc-like toxins (10 genes), BstA abortive infection systems (5 genes), HicB-like toxin-antitoxin systems (17 genes), and a RelE-like toxin (1 gene), among other undescribed toxins (33 genes) (Table S7). *P. larvae* had the second highest number of genes (124 genes, 2.9% frequency), in the PHROGs category: moron, auxiliary metabolic genes, and host takeover (Fig. 3). *P. larvae* prophages possessed a variety of phage toxins, antitoxins, and additionally, genes associated with bacteriocin production or immunity (Table S7). Some *P. larvae* prophages also possessed a gene associated with quaternary ammonium compound-resistance proteins (2 genes), or to ABC transporters (7 genes), indicating the presence of antimicrobial resistance genes (Table S7). One other antimicrobial-associated gene in a *P. larvae* prophage, originally classified as “unknown” by PHROGS, was annotated as a bacitracin resistance protein, BacA, by the eggNOG database (Table S7). Additionally, *P. larvae* prophages possessed genes for flavodoxin, a NifU-like Fe-S cluster assembly protein, and a phosphoadenosine phosphosulfate (PAPs) reductase (Table S7). The PAPs reductase also occurred frequently in prophages isolates from *G. apicola* (13 genes), and once in a *B. asteroides* prophage (Table S7).

The majority of remaining predicted protein-encoding genes (49.0%) were classified broadly as Function Unknown (COG Category S or no hits to the COG database) after failing to match to the PHROGs database (Table S5). Coding regions that were able to be classified broadly by COG were distributed into the categories of Information Storage and Processing (1.6%), Metabolism (0.5%), or Cellular Processing and Signaling (0.5%) (Table S4). Within these broader categories, COG category K (Transcription — Information Storage and Processing) occurred most frequently

at 0.8% (Table S5). Amino Acid Transport and Metabolism (COG category E) and Carbohydrate Transport and Metabolism (COG category G) were the most commonly occurring metabolic categories, each corresponding to 0.1% of genes (Table S5, Table S7). Genes associated with amino acid metabolism were found in prophages of *G. apicola* (8 genes, 0.4% frequency) and *P. larvae* (5 genes, 0.1% frequency) (Fig. 3, Table S6). Five different prophages of *P. larvae* appear to possess a gene homologous for a thermolysin metalloproteinase, while two share a gene associated with glycine amidinotransferase activity (Table S7). The amino acid genes associated with *G. apicola* prophages are more varied but appear to be associated with arginosuccinate and glutamate (Table S7). Prophage genes associated with carbohydrate metabolism were found in several bacterial hosts: *M. plutonius* (3 genes, 0.8% frequency), *G. apis* (4 genes, 0.7%), *L. kunkeei* (1 gene, 0.41% frequency), *Bartonella apis* (1 gene, 0.3% frequency), *G. apicola* (4 genes, 0.2% frequency), and *P. larvae* (2 genes, <0.1% frequency) (Fig. 3, Table S6). Interestingly, a single *M. plutonius* prophage possessed all 3 carbohydrate-associated proteins identified for that bacterial host: a mannitol dehydrogenase, a MFS/sugar transport protein, and a protein associated with the dehydration of D-mannotate (Table S7). Similarly, a single *G. apicola* prophage possessed all four of the carbohydrate-associated proteins identified for that bacterial host species: three proteins associated with the glycerate kinase family and one associated with alpha-ribazole phosphatase activity (Table S7). In contrast, three separate prophages of *G. apis* possessed a gene for alpha-galactosidase, with one also carrying a gene for an exopolysaccharide biosynthesis protein (Table S7). Additionally, two separate prophage species of *P. larvae* both possessed a gene associated with the D-gluconate metabolic process (Table S7). The remaining COG metabolic categories appear to be more rare, although they did still occur (Fig. 3, Table S6). For example, the sole *B. indicum* prophage that was identified possessed genes associated with nucleotide metabolism (a purine-cytosine permease) and secondary metabolite metabolism (isochorismatase) (Table S7).

Closely Related Prophages Typically Share Host Species. Using vConTACT2, a total of 208 honey bee prophages from the dereplicated prophages (n = 303) were grouped into 56 subclusters, with the remaining prophages detected as outliers (n=82) or unable to be incorporated into the network (n = 6) (Fig. 4). The bacterial host strongly predicted clustering, with prophages of the same host species typically forming exclusive subfamily clusters (Fig. 4A). Only 10 of the 62 predicted subclusters included a prophage from more than one host, via either multiple honey bee bacterial species or the inclusion of non-honey bee reference phage (Fig. 4A). In several of these mixed clusters, the bacterial host genus was shared among isolates even if the species was not (Table S8).

In the resulting gene sharing network, honey bee prophages within the same clusters, and therefore often from the same bacterial host genome, were most closely positioned to each other (Figure 4B). However, prophages from different host species sometimes appeared to share genetic similarities. For example, prophages from *G. apicola* and *G. apis* were closely positioned, as were some of the prophages from *S. alvi* and *Bombella apis*, and *Bartonella apis* (Fig. 4B). Interestingly, most of the prophages from *Latobacillus* spp., appear to share genes with each other, but prophages

of *L. kunkei* appear very diverse potentially due to few genes shared among these prophages (Fig. 4B). Furthermore, *L. mellis* prophages appear to share some similarity to prophages isolated from *M. plutonius* and *P. larvae* (Fig. 4B).

Most Prophages from Honey Bee Symbionts are Unclassified Caudoviricetes. Of the 303 distinct prophage populations, 59.4% (180 prophage regions) were only able to be classified as *Caudoviricetes*, while an additional 31.7% prophages (96 prophage) were unable to be taxonomically classified using top reference hits (Fig. 5). Only the remaining 27 prophage regions (0.9%) were able to be classified at a family level or lower (Fig. 5). A total of 12 (16%) *G. apicola* prophages (n = 75) belonged to the family *Peduoviridae* (Class: *Caudoviricetes*), while one of three *F. perrara* prophages was predicted to belong to the *Mesyanzhinoviridae* family (Class: *Caudoviricetes*) (Fig. 5). The most common taxonomic ranks assigned to prophage from *P. larvae* isolates (n = 83) outside of *Caudoviricetes* (69 prophage, 66.3%) or unclassified (21 prophage, 20.2%) were the genera *Fernvirus* (7 prophage, 6.7%), *Vegasvirus* (3 prophage, 2.9%) and *Halyconevirus* (2 prophage, 1.9%) (Class: *Caudoviricetes*) (Fig 5). A single *P. larvae* prophage each (1%) was classified as either the genera *Lilyvirus* or *Dragolirvirus* (Class: *Caudoviricetes*).

Discussion

By analyzing publicly available genomes, 303 distinct putative prophage regions were predicted from honey bee-associated bacteria. Although these predicted prophage regions have not been empirically confirmed to be intact and active, this is the first targeted survey of dsDNA prophage distribution among core honey bee bacterial symbionts. While some of these predicted prophage regions may be genomic relics of prophages, the minimum threshold of 5 kb ensures that all identified prophages, active or relic, could possess genes that may affect their bacterial hosts. Additionally, it should be noted that this study targeted specifically dsDNA phage. The prevalence of prophages with single stranded DNA genomes, like those found in the families *Microviridae* or *Inoviridae*, have yet to be determined (Székely & Breitbart, 2016).

The wide presence of distinct prophage regions found in this study supports the broader trend of high viral diversity in the honey bee gut. However, some of the most common hosts of prophage identified in this study differ from the predicted hosts of viral particles revealed in metagenomic studies of the gut virome (Bonilla-Rosso et al., 2020; Deboutte et al., 2020; Busby et al., 2022). This is possibly because viral metagenomes filter out bacterial cells to enrich for free viral particles. The differences in the host ranges identified with metagenomic approaches and our study of individual isolate genomes may be partially due to the phenomenon of superinfection exclusion, in which bacteria that contain lysogens are protected from additional infection of phages (Bondy-Denomy et al., 2016). For instance, we found that *S. alvi* contains a median of three prophage regions, and about half of all *S. alvi* isolates possessed at least one intact prophage. However, *S. alvi* phages have been relatively rare in metagenomic virome analyses of the honey bee gut (Bonilla-Rosso et al., 2020; Busby et al., 2022). It is possible that bacterial hosts like *S. alvi* may be underrepresented in metagenomic studies due to the challenge of predicting hosts of viral

contigs using CRISPR spacers (Dion et al., 2021). However, *S. alvi* prophages may also limit the population of lytic phages targeting *S. alvi*, or intact prophages in *S. alvi* may excise infrequently; these factors could make detection of *S. alvi* prophage in metagenomic studies less likely. Furthermore, in one virome study, 25% of the phage identified in the honey bee gut virome were predicted to infect *Bifidobacterium* spp. (Bonilla-Rosso et al., 2020), but few *B. asteroides* and other *Bifidobacterium* spp. genomes analyzed in our study had estimated intact prophage; only three out of 12 *B. asteroides* isolates possessed an intact prophage, while the single *B. indicum* isolate and the two *B. coryneforme* isolates did not possess any. This may suggest that some honey bee symbionts, like *Bifidobacterium* spp., are targeted predominantly by lytic phages, or alternatively, highly active temperate phages that frequently excise from their host.

In contrast, *G. apicola* has a high number of prophage regions per genome, which is consistent with other virome reports that assessed the abundance of phage (both lytic and temperate) that existed outside of the bacterial cell at the time of sampling (Bonilla-Rosso et al., 2020; Busby et al., 2022). In this study, *G. apicola* genomes had a higher frequency of intact prophage than many other core bacteria. It is possible that unlike *S. alvi*, these prophages may frequently release progeny phage into the host gut, contributing to the overall composition of the virome. Alternatively, the prophage of *G. apicola* may fail to provide significant superinfection exclusion and protection from lytic phage. A similar scenario may be true for *L. melliventris*, which had a predicted intact prophage in every isolate and was also predominant in one of the same virome studies (Busby et al., 2022). Investigating the lifestyle of phage identified in metagenomic studies may further elucidate if certain bacterial hosts are primarily targeted by a specific lifestyle of phage. In at least one study, the majority of *Bifidobacterium* spp. were targeted by phage predicted to be lytic, while *Gilliamella* spp. phage were primarily predicted to be temperate (Bonilla-Rosso et al., 2020). Exploration of additional bacterial metagenomic studies would also further clarify the relationship between certain phylotypes of bacteria and prophage, as our study is limited by having few sequenced isolates of several of the bacterial species. For example, conclusions about *L. melliventris* should be interpreted cautiously, as only three genomes were publicly available at the time of our analysis. Furthermore, additional genomes from species less represented in this study, such as *F. perrara*, many of the other *Latcobacillus* spp., and *Bifidobacterium* spp., would clarify which bacteria harbor prophages more frequently and to what degree these prophages remain excisable. It should also be noted that the bacterial species with more sequenced isolates, such as *Gilliamella* sp., and *S. alvi*, typically had a higher median abundance of prophages, possibly indicating a bias in this analysis.

The prevalence of prophages across bacterial species is also associated with bacterial traits, such as pathogenicity (Knowles et al., 2016; Silveira & Rohwer, 2016; Touchon, Bernheim & Rocha, 2016). The two bacterial pathogens in our study, *P. larvae* (American Foulbrood) and *M. plutonius* (European Foulbrood), noticeably differ from each other in terms of both prophage number and prophage composition. *P. larvae* has a higher number of prophages and prophage composition compared to both *M. plutonius* and all core bacterial species. In contrast, the number of prophages

and prophage composition of *M. plutonius* is closer to the core species *S. alvi* and *L. melliventris*. While the prophage-encoded toxin *P1x1* found in some strains of *P. larvae* is known to contribute to virulence, it is not the sole cause of virulence; genotypes lacking *P1x1* can still cause severe disease (Ebeling, Fünfhaus & Genersch, 2021). However, the high abundance and diversity of putative prophage regions in nearly all *P. larvae* isolates may imply that there are more prophage-encoded virulence factors to be discovered. For example, the presence of *P. larvae* prophage regions containing possible antimicrobial resistance genes, as well as carbohydrate, iron and sulfur metabolism genes, may indicate additional ways phage provide fitness to the pathogen (discussed in Ribeiro et al., 2022). The presence of three separate carbohydrate-associated genes in a single *M. plutonius* prophage indicates that prophage-encoded auxiliary metabolic pathways provide additional benefit to *M. plutonius*, as well. This is particularly interesting because *M. plutonius* strains are not typically able to metabolize the sugar mannitol, but can metabolize D-mannose (Arai et al., 2012). The phage-encoded mannitol dehydrogenase may break down mannitol into D-mannose, while a second phage-encoded protein may then dehydrate the D-mannose to release energy. The third phage-encoded protein, a sugar transporter, could have a role in shuttling these sugars in or out of the bacterial cell. Combined, these three prophage-encoded genes may serve to enhance the pathogenicity of *M. plutonius* 82 by providing an additional sugar to exploit. However, the lower prophage abundance and composition in *M. plutonius* may indicate that prophages play a less significant role in the virulence of *M. plutonius* compared to *P. larvae*.

Functional annotation of the predicted prophages also suggests that some commensal bacterial, such as *Gilliamella* spp., may receive auxiliary metabolic benefits from their prophages. *Gilliamella* spp. contribute to the breakdown of pollen by primarily targeting pectin (Engel, Martinson & Moran, 2012). It is possible that the glycerate kinases and the alpha-ribazole phosphatase found in *G. apicola* prophage, or the alpha-galactosidase found in three separate *G. apis* prophages assist with the breakdown of pectin or the production of other metabolites. Some isoforms of alpha-galactosidase have been shown to be associated with pectin hydrolysis in fruits (Soh, Ali & Lazan, 2006). The predicted glycerate kinase may be used to support bacterial glycolysis pathways, while the alpha-ribazole phosphatase enzyme may contribute to cobalamin biosynthesis, an important secondary metabolite (Doughty, Hayashi & Guenther, 1966; O'Toole, Trzebiatowski & Escalante-Semerena, 1994). Additionally, an exopolysaccharide biosynthesis coding region found in a *G. apis* prophage may contribute to biofilm formation. Biofilms produced by *Gilliamella* spp. may inhibit pathogen invasion in the midgut (Engel, Martinson & Moran, 2012). As a result, it is possible that prophage presence in *Gilliamella* spp. may directly benefit its bacterial host, and consequently the honey bee. Additionally, several *G. apicola* prophages appear to carry a PAPS reductase gene, which may enhance the host's ability to metabolize sulfur (Mara et al., 2020).

Given that phage typically have a high host-specificity, along with the ability to recombine with host DNA and other co-infecting phages, it is not surprising that the prophage regions found in the same host species are most closely related to each other (Casjens, 2003; Brüssow, Canchaya &

Hardt, 2004). At the same time, phages from entirely different bacterial host species may occasionally transfer genes horizontally to each other (Stecher, Maier & Hardt, 2013), and this may be more likely to occur between phages with hosts that are physically near each other. Prophages identified from bacterial hosts in the midgut and ileum (*Bombella apis*, *Bartonella apis*, *Gilliamellia* spp., *F. perrara*, and *S. alvi*) were positioned more closely in the network, and therefore more likely to share genes with each other, than prophages of bacteria in the rectum (*Bifidobacterium* spp., *Lactobacillus* spp.), possibly because the phages are either more closely related to one another or due to increased horizontal gene transfer between nearby phage communities (Anderson & Ricigliano, 2017). The idea of spatially-separated phage communities in the gut is supported by a recent study investigating the gut virome of the marine invertebrate *Ciona intestinalis*, which found phage communities were localized to specific regions (Leigh et al., 2018). Due to the inherent difficulty in classifying phages, and the limited number of International Committee on Taxonomy of Viruses (ICTV) classified reference phage genomes relative to the vast diversity of bacteriophages in nature, it is unsurprising that a large portion of predicted prophage regions in our study remain unclassified (Bin Jang et al., 2019). Less than ten percent of the predicted prophages were classifiable at a family level or lower. A few prophage predicted from *P. larvae* shared taxonomic similarity to reference *P. larvae* phage isolated from other studies, but the majority were only classifiable as belonging to *Caudoviricetes* order, indicating that a greater diversity of *P. larvae* phage may exist (Stamereilers et al., 2018).

Although honey bee gut bacterial communities contain relatively few bacterial phylotypes, prophages may contribute to the high bacterial strain diversity within the honey bee gut microbiome (Ellegaard et al., 2015; Ellegaard & Engel, 2019), as 237 out of the 303 unique prophage regions were detected in only one bacterial isolate. This may indicate that in some bacterial species, susceptibility to specific prophage integration may vary depending on host strain genotype, possibly driving both viral and bacterial evolution or diversification. By potentially impacting their host's evolution, active honey bee gut prophages may affect general honey bee health and dysbiosis, particularly within the context of antibiotic and pesticide exposure. In other host-associated and environmental systems, antibiotics and other chemicals can induce prophage excision (Danovaro et al., 2003; Allen et al., 2011). Frequent exposure to either antibiotics or pesticides, something that honey bees must contend with (Kakumanu et al., 2016; Raymann, Shaffer & Moran, 2017), may result in a gut state that is constantly perturbed, leading to prophage induction and further gut perturbation. On the other hand, prophages in the honey bee gut may stabilize bacterial communities against antimicrobials and pesticides, as prophages can also enhance the fitness of their bacterial partners via beneficial accessory genes that provide resistance to antibiotics or other environmental pollutants (Wang et al., 2010; Huang et al., 2021).

The diverse prophage community identified in our study supports the well-established idea that prophages are common and present in a wide range of bacterial hosts (Touchon, Bernheim & Rocha, 2016). However, some of the most common bacterial hosts of the prophage community identified in this study do not entirely reflect that of the wider virome studies of honey bee guts,

indicating that prophages may remain cryptic despite more targeted viral sequencing approaches. As a result, the role prophages play in bacterial evolution and community dynamics will require further untangling both within the honey bee system and beyond.

Conclusion

As the first step in identifying how the prophage community interacts with and modulates bacterial communities of the honey bee gut, a survey of prophage in publicly available honey bee gut-associated bacterial genomes was conducted. This study found that prophage abundance and composition vary across the core honey bee gut bacterial species, with *S. alvi* and *G. apicola* possessing the highest prophage number and percent phage composition, and *L. melliventris* having the highest frequency of likely intact prophage. Interestingly, there was some discrepancy between the most commonly predicted bacterial hosts for prophages compared to phage isolated from metagenomic studies. This may indicate that certain nuances of prophage-bacteria interactions are not able to be captured by viromes, which primarily focus on free viral particles.

Additionally, the prophages of some commensal bacterial hosts, such as *Gilliamella* spp., may provide auxiliary metabolic genes for carbohydrate metabolism, which could possibly benefit the honey bee host. These results set the foundation for targeted exploration of the prophage infecting core bacterial species through culture-based methods.

Figure Legends

	Phylotype	Species	Total Isolates (n)
Core, n=150	Acetobacter, n=6	Bombella apis	6
	Bartonella apis, n=6	Bartonella apis	6
	Bifidobacterium spp., n=15	Bifidobacterium asteroides	12
		Bifidobacterium coryneforme	2
		Bifidobacterium indicum	1
	Firmicutes-4, n=6	Lactobacillus mellifer	1
		Lactobacillus mellis	5
	Firmicutes-5, n=11	Lactobacillus apis	3
		Lactobacillus helsingborgensis	2
		Lactobacillus kimbladii	1
		Lactobacillus kullabergensis	2
		Lactobacillus melliventris	3
	Lactobacillus kunkeei, n=10	Lactobacillus kunkeei	10
	Frischella perrara, n=2	Frischella perrara	2
	Gilliamella spp., n=61	Gilliamella apicola	38
		Gilliamella apis	23
Pathogen, n=30	Snodgrassella alvi, n=34	Snodgrassella alvi	34
	Pathogen, n=30	Melissococcus plutonius	17
		Paenibacillus larvae	13

Table 1: Honey bee-associated bacterial isolates

The number of isolates (n) in each phylotype and species of the 181 publicly available genomes of unique bacterial isolates from the NCBI GenBank that were downloaded March-July 2020.

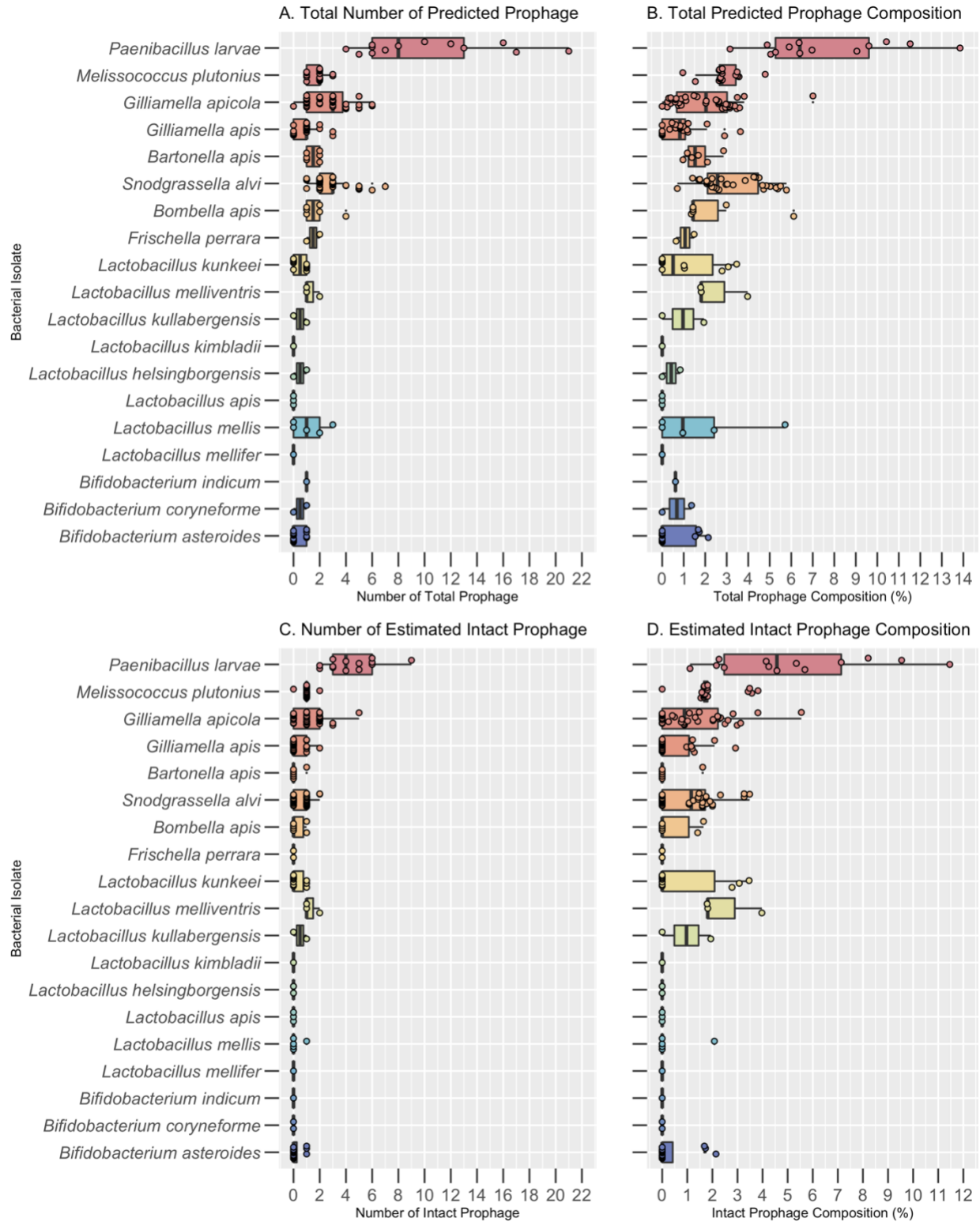


Figure 1: Prophage frequency and composition across bacterial hosts

(A) The predicted number of total prophage regions, (B) total prophage composition (the percent of the bacterial genome composed of prophages), (C) number of intact prophage regions, and (D) intact prophage composition per bacterial isolate. Dots represent individual bacterial isolates, while the center line indicates the median value for each bacterial species. The boxes span the interquartile range and whiskers represent the 25th and 75th quartile ranges. Values below or above the 25th and 75th quartiles are indicated by the larger colored circles behind individual dots. The minimum, maximum, and median values are listed in Table S2.

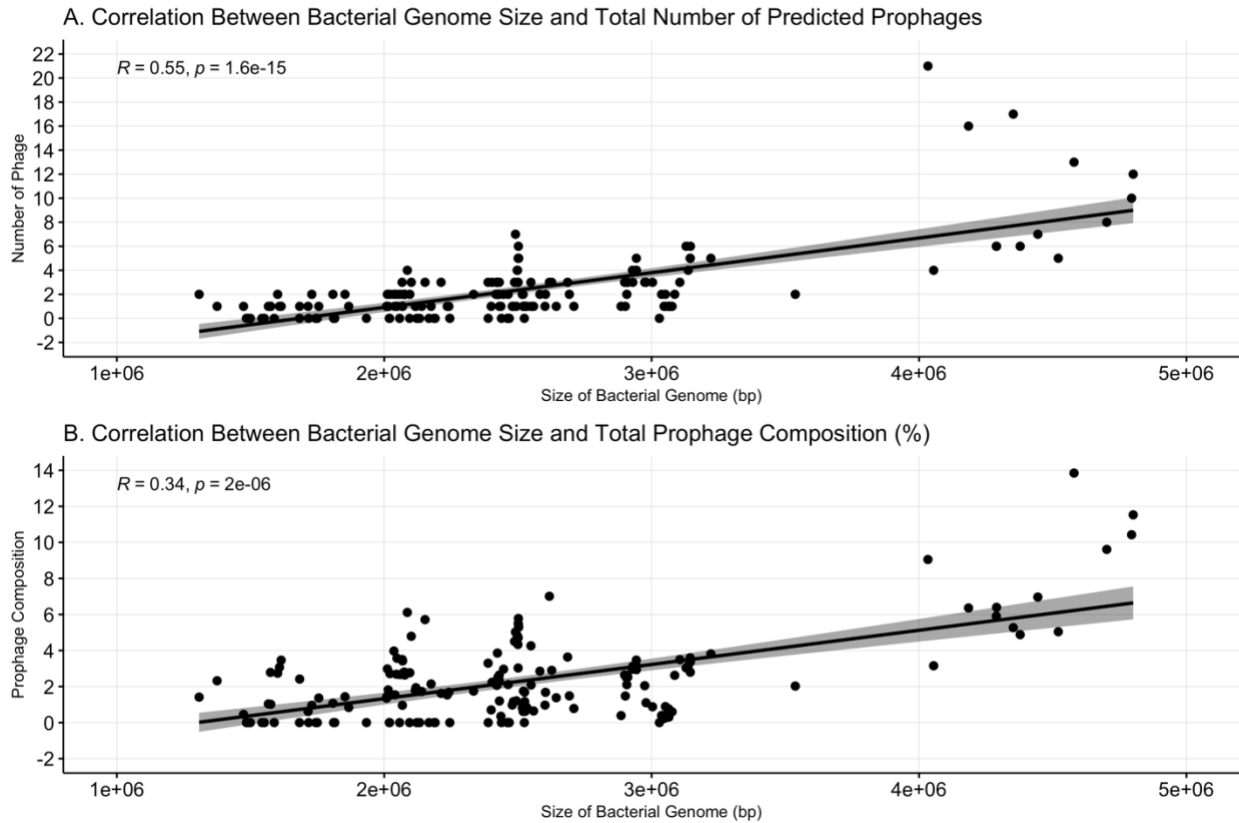


Figure 2: Bacterial genome size correlates to prophage number and composition

(A) The number of prophages identified and the (B) percent prophage composition is positively associated with bacterial genome size. Dots represent individual bacterial genomes.

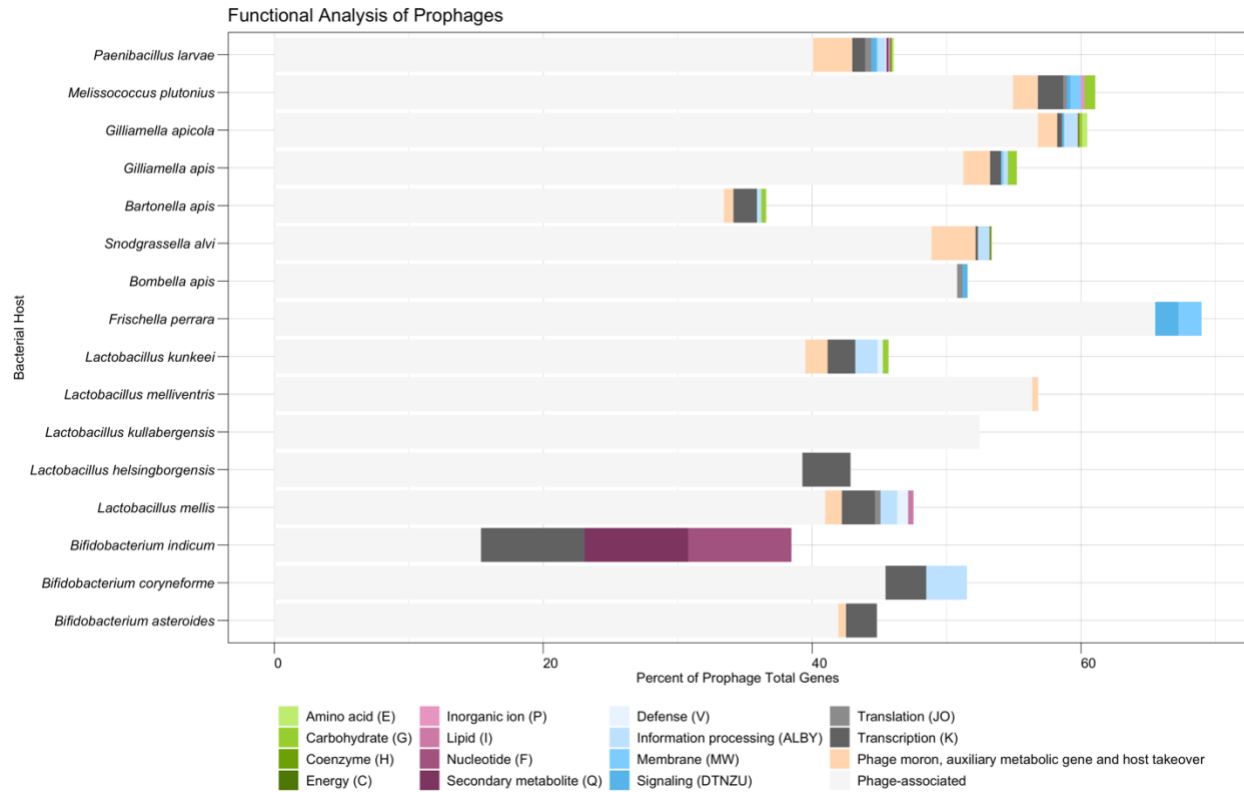


Figure 3: Functional analysis of prophages found in honey bee-associated bacteria

The distribution of phage-associated and COG categories assigned to prophage genes, organized by bacterial host. Similar COG categories are grouped together for visual clarity. X-axis indicates percent. COG Category S (Unknown or No Hits) is not shown.

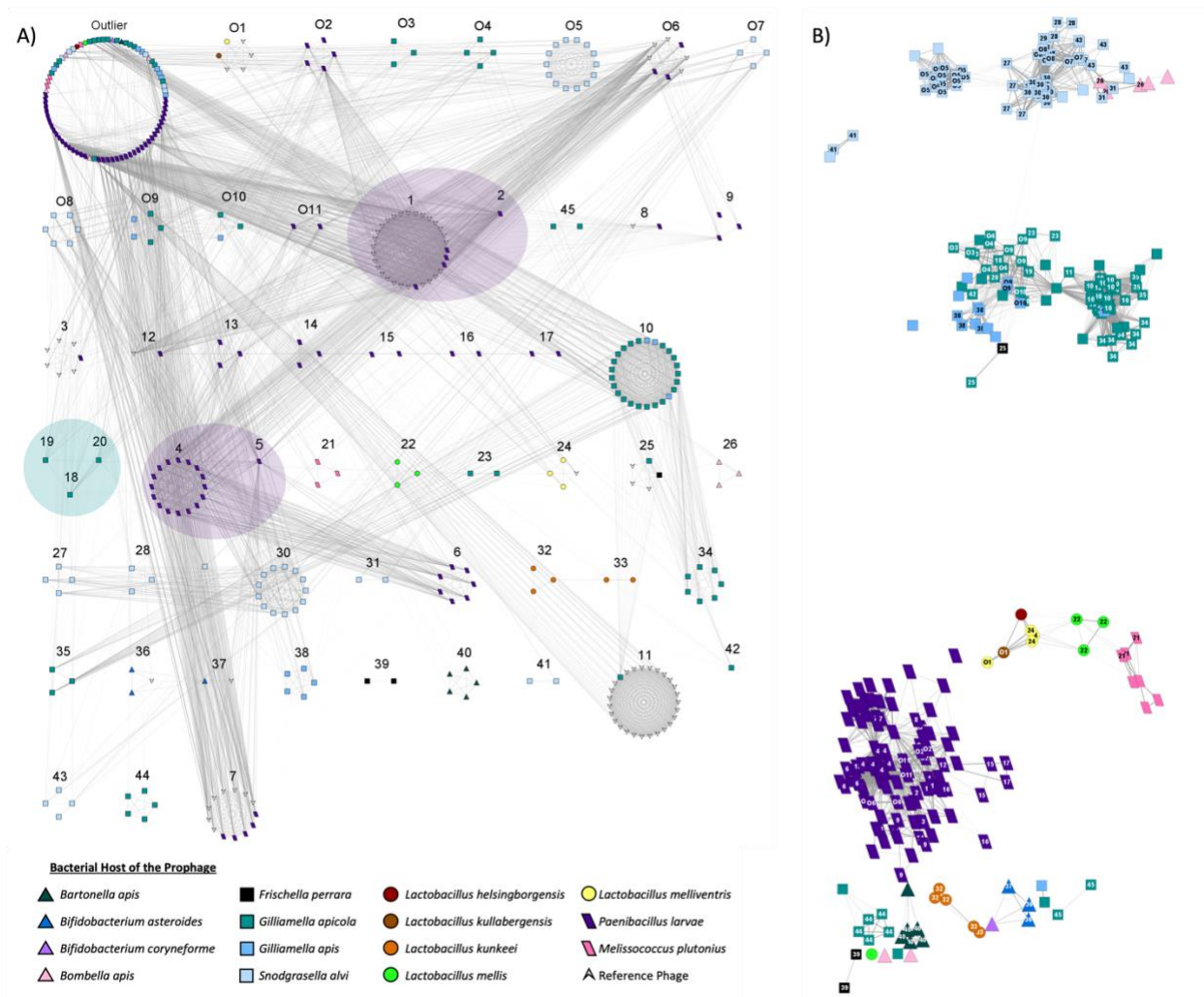


Figure 4: Viral clustering and network analysis based on shared genes

Nodes represent individual prophages. Bacterial hosts are indicated by the node's color and shape. (A) Prophages are separated based on sub-family viral clusters. Numbers above clusters indicate a specific subcluster. Distance between subclusters has no significance, but viral subclusters that come from the same greater viral cluster (indicated by transparent oval overlay) are more related to one another than phage originating from other clusters. Subclusters O6, 1, 7, 8, and 12 contain reference *P. larvae* phage from external studies. Prophage within the outlier cluster are not related to one another. (B) The layout of the network is based on shared genes, with strength indicated by opacity of the network edges. The numbers on nodes indicate subcluster. Blank nodes indicate outliers, while subclusters starting with "O" indicate unique overlapping subclusters.

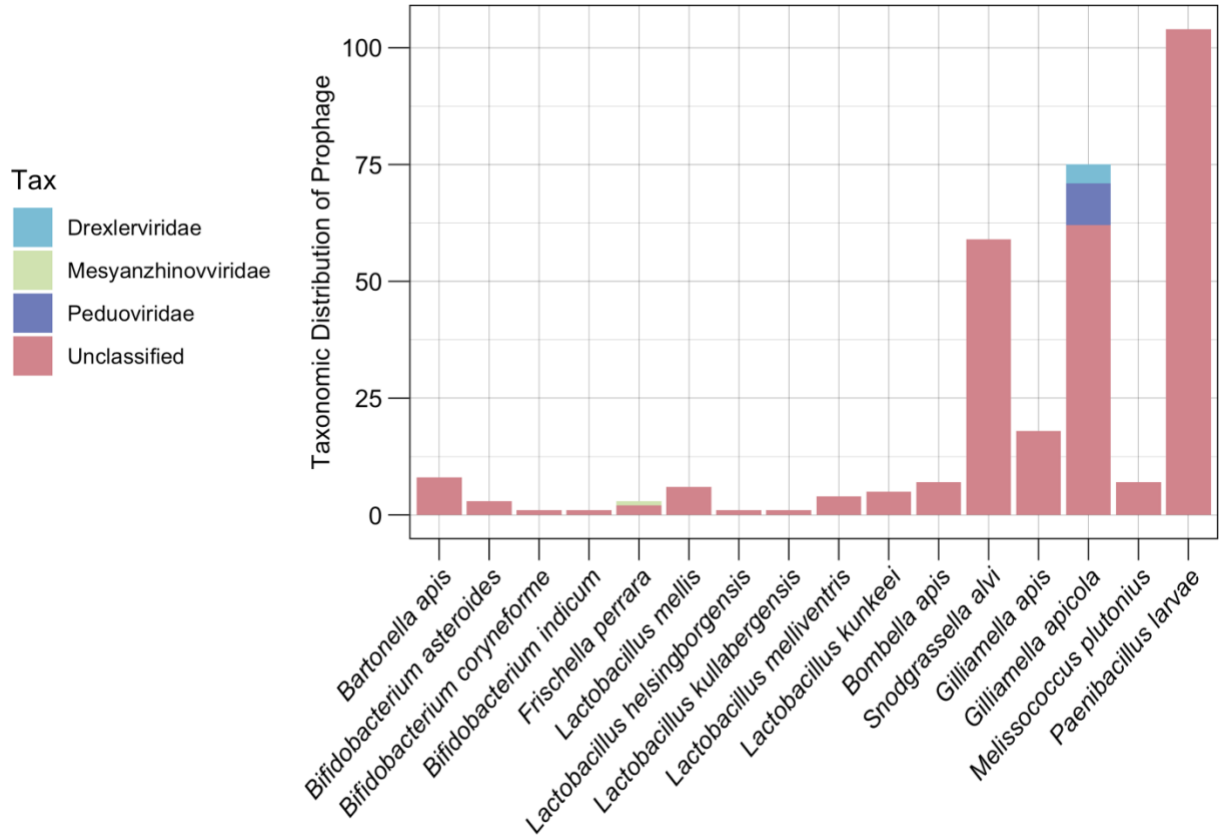
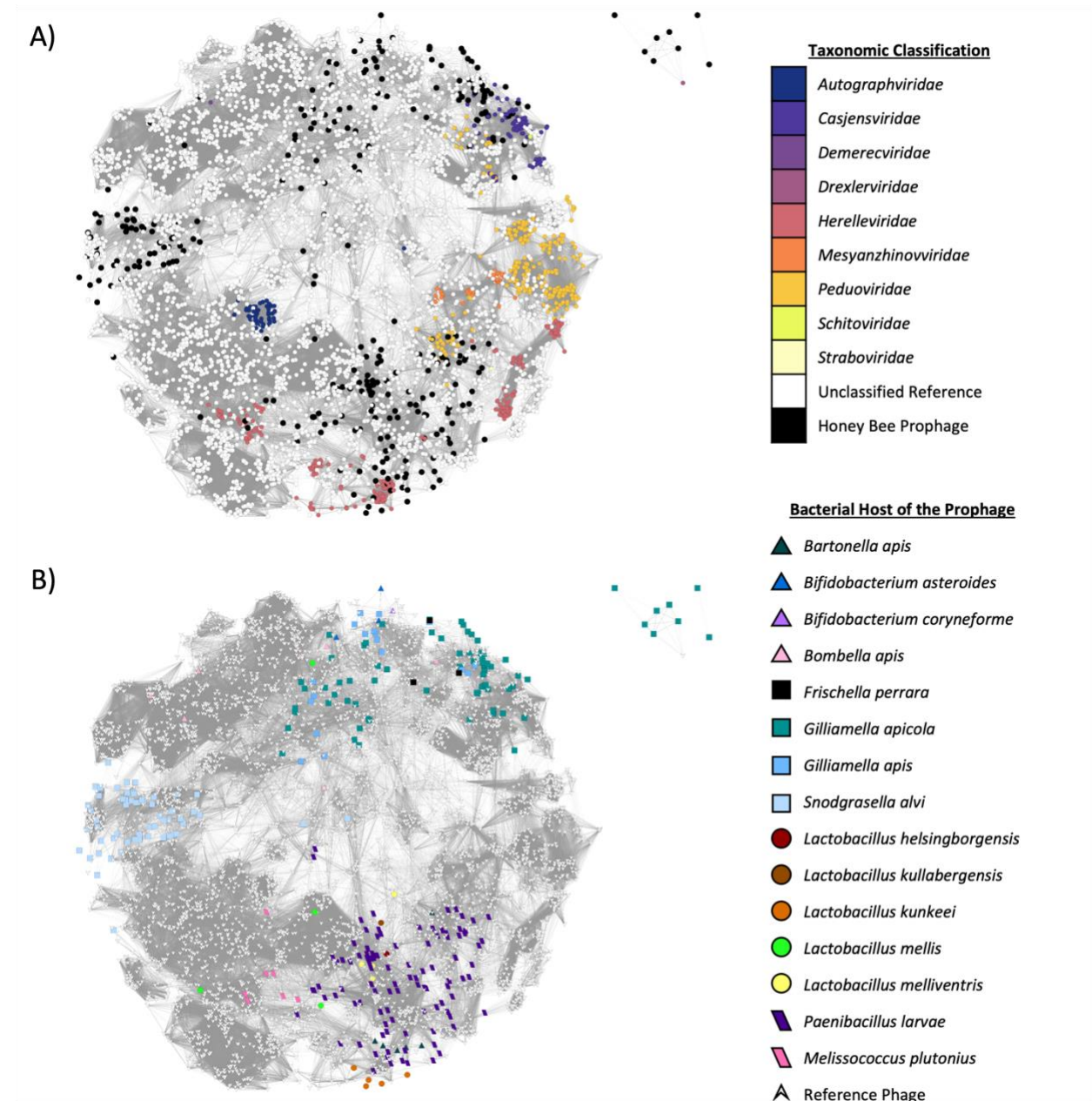


Figure 5: Taxonomic classification of prophages

The lowest taxonomic rank (genus, family, or class) assigned to each prophage based on amino acid similarity to its closest reference phage. Genus and family classifications were assigned if prophages shared AAI across all proteins with a reference above respective thresholds. Prophage which did not meet these thresholds but still matched to a *Caudoviricetes* reference were broadly classified as *Caudoviricetes* class. Prophage which matched to unclassified reference phage were grouped as Unclassified.

Supplementary Information



Supplementary Figure 1: Prophage networks in the context of INPHARED

A vConTACT2 network created with both honey bee isolated prophages and reference phage from the INPHARED database (January 2023). Visually, only reference phage that are third neighbors or less to a honey bee prophage are represented. Nodes indicate individual phages and the opacity of edges represent the strength of shared gene interactions. (A) Honey bee isolated prophages were unable to be taxonomically classified by vConTACT2, but often clustered near other taxonomically classified reference phage. Larger black circles are prophages identified from honey bee symbionts, while smaller colored circles depict the taxonomic family of reference phages. (B)

Honey bee isolated prophages (host indicated by shape and color) tended to cluster near other phage from the same bacterial host, but were not spatially separated from reference phages.

Data Availability

Sequences and additional supplementary data (described below) are available at both:

Virginia Tech's Data Repository: <https://doi.org/10.7294/22203001>

PeerJ: <https://doi.org/10.7717/peerj.15383>

Code and raw data are available at Github: https://github.com/ebueren/HBProphage_PeerJ

Supplementary Table 1 (.csv): NCBI Genbank accession numbers

A list of all NCBI bacterial isolates used and their accession numbers.

Supplementary Table 2 (.xlsx file): Summary statistics of prophages across bacterial hosts

The minimum, median $\pm$ SD, and maximum number of total and estimated intact prophage regions per genome, as well as total and estimated intact prophage composition per genome. Additionally, the minimum, median and maximum lengths of the bacterial genomes are included.

Supplementary Table 3 (.csv): Percentage of bacterial isolates with intact prophage

The percent of individual isolates within a bacterial species that possessed at least one estimated intact prophage.

Supplementary Table 4 (.csv file): Frequency of broad COG categories across all prophages.

Supplementary Table 5 (.csv file): Frequency of specific COG categories across all prophages.

Supplementary Table 6 (.csv file): Frequency of specific COG categories across prophages, grouped by bacterial species.

Supplementary Table 7 (.csv file): All predicted protein annotations, with a final column "AMG_Moron", indicating whether a protein is predicted to be a potential AMG, a possible moron, or neither.

Supplementary Table 8: The vConTACT2 cluster assigned to each phage.

Statements & Declarations

Funding

This research was supported by the National Science Foundation (MCB-1817736).

Competing Interests

The authors declare that there is no conflict of interest or competing interests to disclose.

References

- Abraham J, Bousquet A-C, Bruff E, Carson N, Clark A, Connell A, Davis Z, Dums J, Everington C, Groth A, Hawes N, McArthur N, McKenney C, Oufkir A, Pearce B, Rampal S, Rozier H, Schaff J, Slehrta T, Carson S, Miller ES. 2016. Paenibacillus larvae Phage Tripp Genome Has 378-Base-Pair Terminal Repeats. *Genome Announcements* 4:e01498-15, /ga/4/1/e01498-15.atom. DOI: 10.1128/genomeA.01498-15.
- Allen HK, Looft T, Bayles DO, Humphrey S, Levine UY, Alt D, Stanton TB. 2011. Antibiotics in Feed Induce Prophages in Swine Fecal Microbiomes. *Mbio* 2:e00260-11. DOI: 10.1128/mBio.00260-11.
- Anderson KE, Ricigliano VA. 2017. Honey bee gut dysbiosis: a novel context of disease ecology. *Current Opinion in Insect Science* 22:125–132. DOI: 10.1016/j.cois.2017.05.020.
- Arai R, Tominaga K, Wu M, Okura M, Ito K, Okamura N, Onishi H, Osaki M, Sugimura Y, Yoshiyama M, Takamatsu D. 2012. Diversity of Melissococcus plutonius from Honeybee Larvae in Japan and Experimental Reproduction of European Foulbrood with Cultured Atypical Isolates. *PLoS ONE* 7:e33708. DOI: 10.1371/journal.pone.0033708.
- Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Research* 44:W16–W21. DOI: 10.1093/nar/gkw387.
- Beims H, Wittmann J, Bunk B, Spröer C, Rohde C, Günther G, Rohde M, von der Ohe W, Steinert M. 2015. Paenibacillus larvae-Directed Bacteriophage HB10c2 and Its Application in American Foulbrood-Affected Honey Bee Larvae. *Applied and Environmental Microbiology* 81:5411–5419. DOI: 10.1128/AEM.00804-15.
- Bin Jang H, Bolduc B, Zablocki O, Kuhn JH, Roux S, Adriaenssens EM, Brister JR, Kropinski AM, Krupovic M, Lavigne R, Turner D, Sullivan MB. 2019. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nature Biotechnology* 37:632–639. DOI: 10.1038/s41587-019-0100-8.
- Bolduc B, Jang HB, Doulier G, You Z-Q, Roux S, Sullivan MB. 2017. vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect *Archaea* and *Bacteria*. *PeerJ* 5:e3243. DOI: 10.7717/peerj.3243.
- Bondy-Denomy J, Qian J, Westra ER, Buckling A, Guttman DS, Davidson AR, Maxwell KL. 2016. Prophages mediate defense against phage infection through diverse mechanisms. *The ISME Journal* 10:2854–2866. DOI: 10.1038/ismej.2016.79.
- Bonilla-Rosso G, Steiner T, Wichmann F, Bexkens E, Engel P. 2020. Honey bees harbor a diverse gut virome engaging in nested strain-level interactions with the microbiota. *Proceedings of the National Academy of Sciences*. DOI: 10.1073/pnas.2000228117.

- Brüssow H, Canchaya C, Hardt W-D. 2004. Phages and the Evolution of Bacterial Pathogens: from Genomic Rearrangements to Lysogenic Conversion. *Microbiology and Molecular Biology Reviews* 68:560–602. DOI: 10.1128/MMBR.68.3.560-602.2004.
- Busby TJ, Miller CR, Moran NA, Van Leuven JT. 2022. Global Composition of the Bacteriophage Community in Honey Bees. *mSystems* 7:e01195-21. DOI: 10.1128/msystems.01195-21.
- Casjens S. 2003. Prophages and bacterial genomics: what have we learned so far?: Prophage genomics. *Molecular Microbiology* 49:277–300. DOI: 10.1046/j.1365-2958.2003.03580.x.
- Cook R, Brown N, Redgwell T, Rihtman B, Barnes M, Clokie M, Stekel DJ, Hobman J, Jones MA, Millard A. 2021. INfrastructure for a PHAge REference Database: Identification of Large-Scale Biases in the Current Collection of Cultured Phage Genomes. *PHAGE* 2:214–223. DOI: 10.1089/phage.2021.0007.
- Dahl EM, Neer E, Bowie KR, Leung ET, Karstens L. 2022. microshades: An R Package for Improving Color Accessibility and Organization of Microbiome Data. *Microbiology Resource Announcements* 11:e00795-22. DOI: 10.1128/mra.00795-22.
- Danovaro R, Armeni M, Corinaldesi C, Mei ML. 2003. Viruses and marine pollution. *Marine Pollution Bulletin* 46:301–304. DOI: 10.1016/S0025-326X(02)00461-7.
- Deboutte W, Beller L, Yinda CK, Maes P, Graaf DC de, Matthijnsens J. 2020. Honey-bee-associated prokaryotic viral communities reveal wide viral diversity and a profound metabolic coding potential. *Proceedings of the National Academy of Sciences*. DOI: 10.1073/pnas.1921859117.
- Diedenhofen B, Musch J. 2015. cocor: A Comprehensive Solution for the Statistical Comparison of Correlations. *PLOS ONE* 10:e0121945. DOI: 10.1371/journal.pone.0121945.
- Dion MB, Plante P-L, Zufferey E, Shah SA, Corbeil J, Moineau S. 2021. Streamlining CRISPR spacer-based bacterial host predictions to decipher the viral dark matter. *Nucleic Acids Research* 49:3127–3138. DOI: 10.1093/nar/gkab133.
- Doughty CC, Hayashi JA, Guenther HL. 1966. Purification and Properties of d-Glycerate 3-Kinase from *Escherichia coli*. *Journal of Biological Chemistry* 241:568–572. DOI: 10.1016/S0021-9258(18)96874-2.
- Ebeling J, Fünfhaus A, Genersch E. 2021. The Buzz about ADP-Ribosylation Toxins from *Paenibacillus* larvae, the Causative Agent of American Foulbrood in Honey Bees. *Toxins* 13:151. DOI: 10.3390/toxins13020151.
- Eddy SR. 2011. Accelerated Profile HMM Searches. *PLOS Computational Biology* 7:e1002195. DOI: 10.1371/journal.pcbi.1002195.
- Ellegaard KM, Engel P. 2019. Genomic diversity landscape of the honey bee gut microbiota. *Nature Communications* 10:1–13. DOI: 10.1038/s41467-019-08303-0.
- Ellegaard KM, Tamarit D, Javelind E, Olofsson TC, Andersson SG, Vásquez A. 2015. Extensive intra-phylogroup diversity in lactobacilli and bifidobacteria from the honeybee gut. *BMC Genomics* 16. DOI: 10.1186/s12864-015-1476-6.
- Emery O, Schmidt K, Engel P. 2017. Immune system stimulation by the gut symbiont *Frischella perrara* in the honey bee (*Apis mellifera*). *Molecular Ecology* 26:2576–2590. DOI: 10.1111/mec.14058.
- Engel P, Martinson VG, Moran NA. 2012. Functional diversity within the simple gut microbiota of the honey bee. *Proceedings of the National Academy of Sciences* 109:11002–11007. DOI: 10.1073/pnas.1202970109.
- Forcone K, Coutinho FH, Cavalcanti GS, Silveira CB. 2021. Prophage Genomics and Ecology in the Family Rhodobacteraceae. *Microorganisms* 9:1115. DOI: 10.3390/microorganisms9061115.
- Fortier L-C, Sekulovic O. 2013. Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence* 4:354–365. DOI: 10.4161/viru.24498.

- Guo J, Bolduc B, Zayed AA, Varsani A, Dominguez-Huerta G, Delmont TO, Pratama AA, Gazitúa MC, Vik D, Sullivan MB, Roux S. 2021. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* 9:37. DOI: 10.1186/s40168-020-00990-y.
- Hannigan GD, Meisel JS, Tyldsley AS, Zheng Q, Hodkinson BP, SanMiguel AJ, Minot S, Bushman FD, Grice EA. 2015. The Human Skin Double-Stranded DNA Virome: Topographical and Temporal Diversity, Genetic Enrichment, and Dynamic Associations with the Host Microbiome. *mBio* 6:e01578-15. DOI: 10.1128/mBio.01578-15.
- Hendrix RW. 2002. Bacteriophages: Evolution of the Majority. *Theoretical Population Biology* 61:471–480. DOI: 10.1006/tpbi.2002.1590.
- Hsu BB, Gibson TE, Yeliseyev V, Liu Q, Lyon L, Bry L, Silver PA, Gerber GK. 2019. Dynamic Modulation of the Gut Microbiota and Metabolome by Bacteriophages in a Mouse Model. *Cell Host & Microbe* 25:803-814.e5. DOI: 10.1016/j.chom.2019.05.001.
- Huang D, Yu P, Ye M, Schwarz C, Jiang X, Alvarez PJJ. 2021. Enhanced mutualistic symbiosis between soil phages and bacteria with elevated chromium-induced environmental stress. *Microbiome* 9:150. DOI: 10.1186/s40168-021-01074-1.
- Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, von Mering C, Bork P. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research* 47:D309–D314. DOI: 10.1093/nar/gky1085.
- Kakumanu ML, Reeves AM, Anderson TD, Rodrigues RR, Williams MA. 2016. Honey Bee Gut Microbiome Is Altered by In-Hive Pesticide Exposures. *Frontiers in Microbiology* 7. DOI: 10.3389/fmicb.2016.01255.
- Kieft K, Zhou Z, Anantharaman K. 2020. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* 8:90. DOI: 10.1186/s40168-020-00867-0.
- Kielbasa SM, Wan R, Sato K, Horton P, Frith MC. 2011. Adaptive seeds tame genomic sequence comparison. *Genome Research* 21:487–493. DOI: 10.1101/gr.113985.110.
- Kim M-S, Bae J-W. 2018. Lysogeny is prevalent and widely distributed in the murine gut microbiota. *The ISME Journal* 12:1127–1141. DOI: 10.1038/s41396-018-0061-9.
- Knowles B, Silveira CB, Bailey BA, Barott K, Cantu VA, Cobián-Güemes AG, Coutinho FH, Dinsdale EA, Felts B, Furby KA, George EE, Green KT, Gregoracci GB, Haas AF, Haggerty JM, Hester ER, Hisakawa N, Kelly LW, Lim YW, Little M, Luque A, McDole-Somera T, McNair K, de Oliveira LS, Quistad SD, Robinett NL, Sala E, Salamon P, Sanchez SE, Sandin S, Silva GGZ, Smith J, Sullivan C, Thompson C, Vermeij MJA, Youle M, Young C, Zgliczynski B, Brainard R, Edwards RA, Nulton J, Thompson F, Rohwer F. 2016. Lytic to temperate switching of viral communities. *Nature* 531:466–470. DOI: 10.1038/nature17193.
- Kwong WK, Moran NA. 2016. Gut microbial communities of social bees. *Nature Reviews Microbiology* 14:374–384. DOI: 10.1038/nrmicro.2016.43.
- Lawlor J. 2020. jakelawlor/PNWColors: Initial Release. DOI: 10.5281/zenodo.3971033.
- Leigh BA, Djurhuus A, Breitbart M, Dishaw LJ. 2018. The gut virome of the protochordate model organism, *Ciona intestinalis* subtype A. *Virus Research* 244:137–146. DOI: 10.1016/j.virusres.2017.11.015.
- Mara P, Vik D, Pachiadaki MG, Suter EA, Poulos B, Taylor GT, Sullivan MB, Edgcomb VP. 2020. Viral elements and their potential influence on microbial processes along the permanently stratified Cariaco Basin redoxcline. *The ISME Journal* 14:3079–3092. DOI: 10.1038/s41396-020-00739-3.
- Meng X, Rosenthal R, Rubin DB. 1992. Comparing correlated correlation coefficients. *Psychological Bulletin* 111:172–175. DOI: 10.1037/0033-2909.111.1.172.

- Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, Wu GD, Lewis JD, Bushman FD. 2011. The human gut virome: Inter-individual variation and dynamic response to diet. *Genome Research* 21:1616–1625. DOI: 10.1101/gr.122705.111.
- Myers L, Sirois MJ. SPEARMAN RANK CORRELATION COEFFICIENT. :2.
- Nayfach S, Camargo AP, Schulz F, Eloë-Fadrosh E, Roux S, Kyrpides NC. 2021. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nature Biotechnology* 39:578–585. DOI: 10.1038/s41587-020-00774-7.
- Olm MR, Brown CT, Brooks B, Banfield JF. 2017. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *The ISME Journal* 11:2864–2868. DOI: 10.1038/ismej.2017.126.
- O'Toole GA, Trzebiatowski JR, Escalante-Semerena JC. 1994. The cobC gene of *Salmonella typhimurium* codes for a novel phosphatase involved in the assembly of the nucleotide loop of cobalamin. *Journal of Biological Chemistry* 269:26503–26511. DOI: 10.1016/S0021-9258(18)47223-7.
- Raymann K, Shaffer Z, Moran NA. 2017. Antibiotic exposure perturbs the gut microbiota and elevates mortality in honeybees. *PLOS Biology* 15:e2001861. DOI: 10.1371/journal.pbio.2001861.
- Ribeiro HG, Melo LDR, Oliveira H, Boon M, Lavigne R, Noben J-P, Azeredo J, Oliveira A. 2019. Characterization of a new podovirus infecting *Paenibacillus* larvae. *Scientific Reports* 9:20355. DOI: 10.1038/s41598-019-56699-y.
- Ribeiro HG, Nilsson A, Melo LDR, Oliveira A. 2022. Analysis of intact prophages in genomes of *Paenibacillus* larvae: An important pathogen for bees. *Frontiers in Microbiology* 13:903861. DOI: 10.3389/fmicb.2022.903861.
- Roux S, Adriaenssens EM, Dutilh BE, Koonin EV, Kropinski AM, Krupovic M, Kuhn JH, Lavigne R, Brister JR, Varsani A, Amid C, Aziz RK, Bordenstein SR, Bork P, Breitbart M, Cochrane GR, Daly RA, Desnues C, Duhaime MB, Emerson JB, Enault F, Fuhrman JA, Hingamp P, Hugenholtz P, Hurwitz BL, Ivanova NN, Labonté JM, Lee K-B, Malmstrom RR, Martinez-Garcia M, Mizrahi IK, Ogata H, Páez-Espino D, Petit M-A, Putonti C, Rattei T, Reyes A, Rodriguez-Valera F, Rosario K, Schriml L, Schulz F, Steward GF, Sullivan MB, Sunagawa S, Suttle CA, Temperton B, Tringe SG, Thurber RV, Webster NS, Whiteson KL, Wilhelm SW, Wommack KE, Woyke T, Wrighton KC, Yilmaz P, Yoshida T, Young MJ, Yutin N, Allen LZ, Kyrpides NC, Eloë-Fadrosh EA. 2019. Minimum Information about an Uncultivated Virus Genome (MIUViG). *Nature Biotechnology* 37:29–37. DOI: 10.1038/nbt.4306.
- Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, Poulos BT, Solonenko N, Lara E, Poulain J, Pesant S, Kandels-Lewis S, Dimier C, Picheral M, Searson S, Cruaud C, Alberti A, Duarte CM, Gasol JM, Vaqué D, Bork P, Acinas SG, Wincker P, Sullivan MB. 2016. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* 537:689–693. DOI: 10.1038/nature19366.
- Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. DOI: 10.1093/bioinformatics/btu153.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* 13:2498–2504. DOI: 10.1101/gr.1239303.
- Silveira CB, Rohwer FL. 2016. Piggyback-the-Winner in host-associated microbial communities. *npj Biofilms and Microbiomes* 2:1–5. DOI: 10.1038/npjbiofilms.2016.10.
- Soh C-P, Ali ZM, Lazan H. 2006. Characterisation of an α -galactosidase with potential relevance to ripening related texture changes. *Phytochemistry* 67:242–254. DOI: 10.1016/j.phytochem.2005.09.032.

- Stamereilers C, Fajardo CP, Walker JK, Mendez KN, Castro-Nallar E, Grose JH, Hope S, Tsourkas PK. 2018. Genomic Analysis of 48 *Paenibacillus* larvae Bacteriophages. *Viruses* 10:377. DOI: 10.3390/v10070377.
- Stecher B, Maier L, Hardt W-D. 2013. “Blooming” in the gut: how dysbiosis might contribute to pathogen evolution. *Nature Reviews Microbiology* 11:277–284. DOI: 10.1038/nrmicro2989.
- Székely AJ, Breitbart M. 2016. Single-stranded DNA phages: from early molecular biology tools to recent revolutions in environmental microbiology. *FEMS Microbiology Letters* 363:fnw027. DOI: 10.1093/femsle/fnw027.
- Terzian P, Olo Ndela E, Galiez C, Lossouarn J, Pérez Bucio RE, Mom R, Toussaint A, Petit M-A, Enault F. 2021. PHROG: families of prokaryotic virus proteins clustered using remote homology. *NAR Genomics and Bioinformatics* 3:lqab067. DOI: 10.1093/nargab/lqab067.
- Touchon M, Bernheim A, Rocha EP. 2016. Genetic and life-history traits associated with the distribution of prophages in bacteria. *The ISME Journal* 10:2744–2754. DOI: 10.1038/ismej.2016.47.
- Turner D, Kropinski AM, Adriaenssens EM. 2021. A Roadmap for Genome-Based Phage Taxonomy. *Viruses* 13:506. DOI: 10.3390/v13030506.
- Wang X, Kim Y, Ma Q, Hong SH, Pokusaeva K, Sturino JM, Wood TK. 2010. Cryptic prophages help bacteria cope with adverse environments. *Nature Communications* 1:147. DOI: 10.1038/ncomms1146.
- Wickham H. 2016. *ggplot2*. Cham: Springer International Publishing. DOI: 10.1007/978-3-319-24277-4.
- Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: A Fast Phage Search Tool. *Nucleic Acids Research* 39:W347–W352. DOI: 10.1093/nar/gkr485.

Chapter Three: Putative prophages of bacterial isolates collected from amphibian skin

Emma K. Bueren, Noah Wax, Jenifer B. Walke, David C. Haak, and Lisa K. Belden

Abstract

The bacterial skin microbiome of amphibians has been extensively characterized in response to the decline of global amphibian populations. However, the bacterial viruses (bacteriophages) of the amphibian skin microbiome have yet to be characterized, despite the fact that they may influence bacterial population structure and serve as vehicles of horizontal gene transfer. Upon infection, some phages integrate directly into the bacterial genome, and while they remain capable of killing the bacterial host, they may provide beneficial genes to the bacteria, such as resistance to infection from additional phages, competitive toxins, or auxiliary metabolic genes. To assess the prophages of the amphibian skin microbiome, 46 whole bacterial genome sequences across 14 bacterial genera, each isolated from one of four amphibians (*Lithobates catesbeianus*, *Notophthalmus viridescens*, *Pseudacris crucifer* and *Anaxyrus americanus*) were bioinformatically surveyed. A total of 72 unique prophages were identified, and the number of prophages appeared to vary among genera, with *Hafnia* spp. isolates possessing the highest median number of phage per isolate. Prophages from several bacterial hosts appear to encode genes associated with sulfur metabolism (phosphoadenosine phosphosulfate reductases). Additionally, prophages primarily found in *Janthinobacterium lividum* appear to encode genes with a high homology to chitinase, a proposed antifungal enzyme that can degrade fungal cell walls. To determine if these predicted chitinase-like genes were unique to amphibian-isolated *J. lividum*, an additional 21 publicly available *J. lividum* genomes isolated from the environment were assessed for prophages as well. Results suggest that *J. lividum* isolation source has little impact on prophage abundance or genetic relatedness, and that these predicted chitinase-like genes are common across *J. lividum* isolates. Notably, these predicted chitinase-like endolysins appear to be highly conserved across individual prophages, and are selectively retained by the bacteria host, which suggests the gene provides some benefit to the bacterial host. The results of this survey of the virial component of the amphibian skin microbiome suggest the presence of a diverse community of amphibian prophages, which may encode potential auxiliary metabolic genes that benefit their bacterial hosts, and thus ultimately impact the structure of these skin bacterial communities.

Introduction

Bacteriophages (phages) are viruses that infect bacterial hosts. They are highly abundant, and may play important roles in many ecosystems (Hendrix, 2002; Weinbauer, 2004). For example, within aquatic and soil systems, phages influence bacterial population dynamics and help maintain biogeochemical cycles (Thingstad, 2000; Weinbauer, 2004). Within the ocean, phages with lytic lifestyles that infect, replicate within, and lyse their hosts immediately engage in a “kill-the-

winner” strategy, resulting in a cyclical rise and fall of dominance among bacterial hosts that maintains bacterial diversity in the community (Winter et al., 2010). By lysing their hosts, lytic phages also cycle carbon and other nutrients back into the ocean, a process referred to as the viral shunt (Roux et al., 2016). Biogeochemical cycles may be further altered by the presence of phage auxiliary metabolic genes (AMGs), which are genes with non-specific viral function that lytic viruses encode to potentially alter bacterial metabolic processes to better suit viral replication (Breitbart et al., 2018).

Not all phages follow an exclusively lytic life cycle; temperate phages, for example, are capable of undergoing both immediate infection and lysis of the host, or alternatively, integrating their viral genomes into the bacterial host genome. Once integrated, the virus is known as a "prophage", and it does not replicate until bacterial excision and lytic infection is triggered (Canchaya et al., 2003). While the presence of a prophage is an inherent risk (lysis), they may also enhance the competitiveness of their bacterial hosts via the presence of novel beneficial genes that can alter bacterial phenotypes (lysogenic conversion). Examples of lysogenic conversion include phage-encoded virulence factors and resistance against other phages (superinfection exclusion) (Brüssow, Canchaya & Hardt, 2004; Fortier & Sekulovic, 2013; Wendling, 2023). Under some conditions, non-pathogenic bacterial species possessing prophages, even without any obvious beneficial genes, can outcompete identical strains without prophages (Oh et al., 2019; Wendling, Refardt & Hall, 2021). Temperate phages and prophages also provide AMGs to bacterial hosts, which may enhance bacterial fitness in competition-heavy or harsh environments, such as metal-contaminated fields (Huang et al., 2021). These host-benefiting genes may not require a prophage that is capable of excision and lysis; therefore, over evolutionary time, the bacterial host may degrade any harmful prophage genes, like those associated with lysis or excision, while retaining genes that provide a benefit (Khan, Burmeister & Wahl, 2020). As a result, many prophages become “degraded” or “cryptic”, meaning they are no longer capable of lysis or infection; instead, a degraded prophage is now a permanent genetic element of the bacterial genome (Bobay, Touchon & Rocha, 2014). Phage-encoded genes not useful to the bacterial host are further degraded over evolutionary time, while beneficial genes are maintained or repurposed for other tasks by the bacterial host (Bobay, Touchon & Rocha, 2014). However, other prophages, like the toxin-encoding prophages in *Vibrio cholera* and *Escherichia coli* O157:H7, are only useful to the bacterial host if the prophage remains a fully functional temperate phage and is capable of excision and lysis (Fortier & Sekulovic, 2013).

Until recently, our understanding of phages in host-associated communities has primarily focused on one-to-one interactions between phages and pathogens. For example, the field of phage therapy has investigated how individual phages can reduce the population of specific pathogens, while research in lysogenic conversion has focused on pathogen virulence (Matsuzaki et al., 2005; Fortier & Sekulovic, 2013). Nearly half of all bacterial genomes that have been sequenced have at least one prophage, and prophages are heavily present in every part of the human microbiome (Touchon, Bernheim & Rocha, 2016; Inglis, Roach & Edwards, 2024). While research on the downstream benefits that phage-encoded genes may have on animal hosts is still in its infancy,

one metatranscriptomic study found that in polluted soils, there was upregulation of phage-encoded metabolic genes in the bacterial community within earthworm guts, suggesting that phages may play a role in earthworm soil bioremediation (Xia et al., 2023).

We know less about the role of phages in other host-associated microbiomes, despite our understanding that the microbiome influences host health (Cho & Blaser, 2012; Thomas et al., 2017). For example, the composition of the amphibian skin microbiome can impact the outcome of exposure to the fungal pathogen *Batrachochytrium dendrobatidis* (Bd) (Harris et al., 2009; Joneson et al., 2011; Becker et al., 2015; Bates et al., 2022). Bd, which causes the disease chytridiomycosis, has contributed to the decline of over 500 amphibian species globally (Lips, 2016; Fisher & Garner, 2020). Although Bd is highly transmissible, some amphibian skin commensals are capable of inhibiting Bd growth via production of secondary metabolites, such as violacein and prodigiosin (Woodhams et al., 2018). *Janthinobacterium lividum*, which produces violacein, is frequently put forth as a potential probiotic to be used against chytrid (Woodhams et al., 2018). *J. lividum* is found globally in soil and water, in addition to host-associated environments, and can be diverse in its metabolic capabilities (Ramdass & Rampersad, 2021). Evidence suggests that while violacein is the primary mechanism of chytrid inhibition, *J. lividum* may possess additional antifungal genes (Wax et al., 2023).

Many genes may act as possible mediators between bacteria, fungi, and animal hosts. For instance, the presence of metal-binding genes, such as iron siderophores and zinc metalloproteases, may be associated with Bd inhibition due to increased resource competition between bacterial isolates and Bd (Bates et al., 2022; Wax et al., 2023). Other amphibian-associated bacterial isolates encode a known antifungal enzyme, chitinase, which can break down the chitin that is present in the fungal cell wall (Bates et al., 2022; Kumar et al., 2022; Wax et al., 2023). Serine proteases, which are common across most domains of life, appear to be both a virulence factor for many fungal pathogens, including Bd, while at the same time, may possess antifungal properties (Joneson et al., 2011; Ellison et al., 2014). While violacein and prodigiosin are not phage-encoded, more common genes like serine proteases, metalloproteases, and chitinases are sometimes found in viral genomes, including phages (Yamada et al., 2010; Tang et al., 2013; Busico-Salcedo & Owens, 2013; Argov et al., 2019; Khilyas et al., 2021; Abulikemu et al., 2021).

Beneficial prophage-encoded metabolic genes or antifungal genes in the amphibian skin microbiome may enhance bacterial host fitness, shaping the amphibian skin microbial community, and potentially impacting amphibian health. In this study, the genomes of 46 bacterial isolates collected from the skin of four species of amphibian were assessed for prophages using bioinformatic methods. Then, an additional 17 publicly available *Janthinobacterium lividum* genomes from environmental sources were assessed to determine if isolation source resulted in prophage specialization, such as altered number of prophages or the presence of potentially antifungal genes.

Methods

Bacterial Isolation and Chytrid Inhibition Assays

Four species of amphibians (American bullfrogs-*Lithobates catesbeianus*, Eastern newts-*Notophthalmus viridescens*, Spring peepers-*Pseudacris crucifer*, and American toads-*Anaxyrus americanus*) were swabbed in southwest Virginia to isolate skin bacteria, as previously described (Walke et al., 2015). Briefly, each amphibian was rinsed with sterile water and then swabbed twice with separate sterile rayon swabs. The first swab was used for culture-independent 16S rRNA gene amplicon sequencing to assess bacterial community structure, while the second swab was plated onto R2A media for bacterial isolation. A total of 666 isolates were collected and taxonomically identified by Sanger sequencing of the full-length 16S rRNA gene (Walke et al., 2015). These isolates were then challenged against Bd in a 96-well plate assay, and Bd growth inhibition was quantified from 0 (no inhibition) to 100 (complete inhibition), with negative values indicating facilitation of Bd growth (Walke et al., 2017).

Isolate whole genome sequencing

Whole genome sequencing was completed on a set of 62 of these 666 isolates, representing 15 genera, and isolates that had both high and low Bd inhibition values (detailed DNA extraction and sequencing methods in Wax et al. 2023). Wax et al. (2023) used data from 40 of those isolate whole genomes produced via Illumina short-read sequencing (PRJNA922551) to assess potential genomic variation that could explain variation in the ability of bacterial isolates to inhibit Bd growth. Seven additional isolates in this set were also sequenced with Illumina short-reads, but as the resulting whole genome sequence identity did not match the initial 16S-based designation, they were not used in the study. The 15 remaining bacterial isolates of the 62 were sequenced using the PacBio RS II long-read platform. One long-read assembly was removed due to low quality, while six other assemblies were removed due to discrepancy between assigned whole genome taxonomy and initial 16S-based designation. An additional two isolates had genomes nearly identical to short read genomes, and so are not included to prevent redundancy. The remaining 6 long-read whole genomes included in this dataset were not used in Wax et al. (2023) and are reported here for the first time.

Sequencing was performed at the Duke Center for Genomic and Computational Biology. The short-read sequencing used the Illumina Hi-Seq 4000 platform, 2 x150bp on a single lane; for the previously published short-read genomes, raw data and assemblies are available from Wax et al. (2023) and under PRJNA922551. The long-read sequences were produced using the PacBio RS II platform to approximately 100X coverage. Long-read genome assemblies were generated using canu v2.0 and circularized using OrganellePBA (Soorni et al., 2017; Koren et al., 2017). Raw PacBio reads and assemblies are available upon request (Supplementary Data 1).

Prophage Survey

The final set of 46 whole genome sequences were assessed for prophages using the same viral pipeline as in Bueren et al., (2023). Briefly, isolate genomes were first analyzed with VirSorter2 (Guo et al., 2021) and trimmed with CheckV (Nayfach et al., 2021). Only sequences classified by VirSorter2 to be most likely double-stranded DNA viruses, and that remained larger than 5 kb after CheckV trimming, were kept for further analysis. The final set of putative prophages (n=85) consisted of all high scoring (>0.9) VirSorter2 sequences (n=77) and any medium scoring sequences (0.5-0.9) that were additionally classified as "viral" by a secondary program, VIBRANT (Kieft, Zhou & Anantharaman, 2020) (n=8).

Prophage Abundance and Composition Measurements

For each bacterial isolate, the total number of putative prophages (prophage abundance) and the total portion of the bacterial genome composed of putative prophage regions (prophage composition) were calculated. Prophage composition was assessed because bacterial genome size is associated with prophage abundance (Touchon, Bernheim & Rocha, 2016; Bueren et al., 2023). Spearman correlations between genome size and total abundance or total composition were calculated. Isolates were then grouped together at the genus level to identify the median prophage abundance and prophage composition based on bacterial host genera.

Functional Annotation

Initially, all putative prophage sequences were annotated using prokka (v.1.14.6) and the phage-specific Prokaryotic Virus Remote Homologous Groups database (PHROGs) (downloaded from <http://millardlab.org/2021/11/21/phage-annotation-with-phrogs/> on July 20, 2022) (Seemann, 2014; Terzian et al., 2021). All hits to the PHROGs database were broadly classified as "Phage-associated", except for three PHROG categories: "other", "unknown function", and "moron, auxiliary metabolic gene, and host takeover". Additionally, all protein sequences from prokka were then searched using EggNOG-mapper (<http://eggno-mapper.embl.de/>). Any hits from EggNOG-mapper that included multiple COG categories were manually reclassified as "unknown". All hits were matched to the original COG 2020 database, which includes the COG category "[X]: Mobilome: Prophage and Mobile Elements". EggNOG hits that belonged to COG category X or contained any of the words: phage, baseplate, capsid, integrase, tail, tape, lysozyme, portal, holin, N-Acetylmuramoyl-L-alanine amidase, transposase, virus, or viral, were manually reclassified to the category of "Phage-associated." Any proteins that had no match to the EggNOG database, or were classified as multiple COG categories (ex: EH), were reclassified as belonging to the COG category S (Function Unknown). The PHROGs and EggNOG annotations were then combined, with priority given to the PHROGs database. In cases where PHROGs was unable to match to a protein or classified a protein broadly as "other" or "unknown function," the EggNOG annotation was used. Finally, the PHROG and EggNOG database entries of "moron, takeover, potential AMG" proteins were manually inspected and reclassified to either "Predicted AMG" or "Host Takeover or Defense" to differentiate between the two. The frequency of each protein category

was calculated by determining the number of proteins associated with each category relative to all predicted proteins.

Additionally, intactness of each putative prophage was assessed. A putative prophage was estimated to be intact if it was considered “complete” by PHASTER (Arndt et al., 2016) or it possessed at least 3 hallmark genes and either an integrase or a transposase (Bueren et al., 2023). A putative prophage that was not intact was considered degraded. These degraded prophages can still contribute important genes to the host and may in some cases indicate a longer evolutionary history with the host, and thus were retained in the analyses. A total of 41 intact and 44 degraded prophages were identified across the amphibian isolates.

Identification of Potential Antifungal Compounds

To identify potential antifungal proteins, both PHROGs and EggNOG hits were searched for broad terms associated with fungal cell components or potential antifungals identified in previous studies (fungal, fungi, chitin, chitinase, zinc, metal, m36, serine, aspartyl, ABC transporters, superoxide dismutase, violacein, prodigiosin, indole-3-carboxaldehyde, indole, tryptophan, vocs, peptidyl-nucleosides) (Joneson et al., 2011; Lima, Colombo & de Almeida Junior, 2019; Bates et al., 2022). After manual inspection, all remaining genes were classified into the antifungal categories that were present in the dataset: chitinases, serine proteases, metalloproteases, or ABC transporters. These genes were flagged as potentially antifungal, but were not reclassified in terms of predicted function via PHROGs or EggNOG. Additionally, all nucleotide sequences were also analyzed with antiSMASH (v.7.0.0) to identify potential metabolic biosynthesis pathways, but no hits were found.

The amino acid sequences of amphibian prophage-encoded chitinases, possible antifungal compounds previously identified in bacterial genomes from amphibians in Wax et al., (2023), and which had the highest frequency of all potential antifungal compounds in this dataset, were selected for additional analysis. Conserved domains of all putative predicted chitinases were assessed using InterProScan to further differentiate function between chitinase and phage endolysins, a protein which phages use during infection and replication to lyse bacterial hosts. Both proteins target similar structures in chitinase and peptidoglycan respectively, and as a result, bioinformatic predictions about chitinolytic activity been proven wrong by *in vitro* testing (Park et al., 2014). All initially predicted chitinase-like proteins from amphibian isolates, including those lacking conserved chitinase domains, were aligned using MACSE v.2 (Ranwez et al., 2018). An additional 14 reference proteins of similar homology and with *in vitro* verified functions as either chitinases or bacterial endolysins were included in the alignment (Supp Data, 1). This alignment was used to produce a maximum likelihood phylogenetic tree in RAXmL, which was bootstrapped 100 times (Stamatakis, 2014).

Dereplication

All prophage sequences were dereplicated using dRep v3.0.0 to determine if any prophages were present in more than one bacterial isolate (Olm et al., 2017). Putative prophage regions (intact + degraded) sharing more than 95% average nucleotide identity across more than 85% of the shorter sequence were considered to be from the same prophage population (Roux et al., 2019). The longest prophage sequence was used as the reference prophage sequence for all dereplicated datasets.

Taxonomic Classification

Prophage taxonomic classification was assigned by comparing average amino acid identity between dereplicated putative prophage and a database of dereplicated *Caudoviricetes* phages (sequences downloaded from INPHARED on November 5, 2022, with taxonomic information updated on March 16, 2023) (Bueren et al, 2023). A BLASTp search (E value 0.001) of the putative prophage was run against the *Caudoviricetes* database. Taxonomic class was assigned using the putative prophage's top hit, with taxonomic rank determined by AAI. High AAI similarity (>70% across 85% shared proteins) was classified as the same genus as the reference, while lower AAI similarity (>30% over 50% of shared proteins) was classified as sharing the same family. If neither condition was met, but the top reference protein had a classified Order, then that was assigned to the prophage. Otherwise, putative prophages were labeled "Unclassified".

*Comparison between amphibian- and environmentally-isolated *Janthinobacterium* prophages*

In addition to 5 sequences of *J. lividum* isolated from amphibians reported here and in Wax et al. (2023), 17 publicly-available *J. lividum* isolate genomes from NCBI (June 2023) were assessed for prophages, and analyzed in the same manner as described above, in order to determine if the presence of the potentially antifungal chitinase was unique to amphibian-isolated *J. lividum*. This additional survey resulted in 40 (24 estimated intact and 16 degraded) additional prophages. Based on NCBI metadata, these 17 additional isolates and their respective prophages were classified broadly as either environmental (from soil, water, or food), or host-associated (from plants or sponges). All our *J. lividum* isolates collected from amphibian skin were classified as host-associated (amphibian). The abundance and genome composition of prophage were compared across each specific category ("Environmental", "Amphibian", "Non-Amphibian Host"), as were gene functional predictions. To assess prophage similarity between isolation sources, VContact2 was used to produce similarity networks comparing all *J. lividum* prophages to each other and against the viral RefSeq211 database. To determine if isolation source impacted relatedness between bacterial genomes, all *J. lividum* isolates were additionally clustered into groups of 98% ANI similarity using dRep and visualized with a cladogram.

Calculation of dN/dS rates of chitinase in Janthinobacterium isolates

To identify whether chitinase genes encoded by *J. lividum* prophages were undergoing positive or purifying selection within the bacterial host, nucleotide sequences of intact and degraded GH18 or GH19 chitinases from *J. lividum* prophages were aligned and cleaned using MACSE v.2 (Ranwez et al., 2018). All four alignments were then tested for dN/dS ratios with null and site-based models (M0, M3, M7, M8) using codeML in the EasyCodeML wrapper (Jeffares et al., 2015).

Results

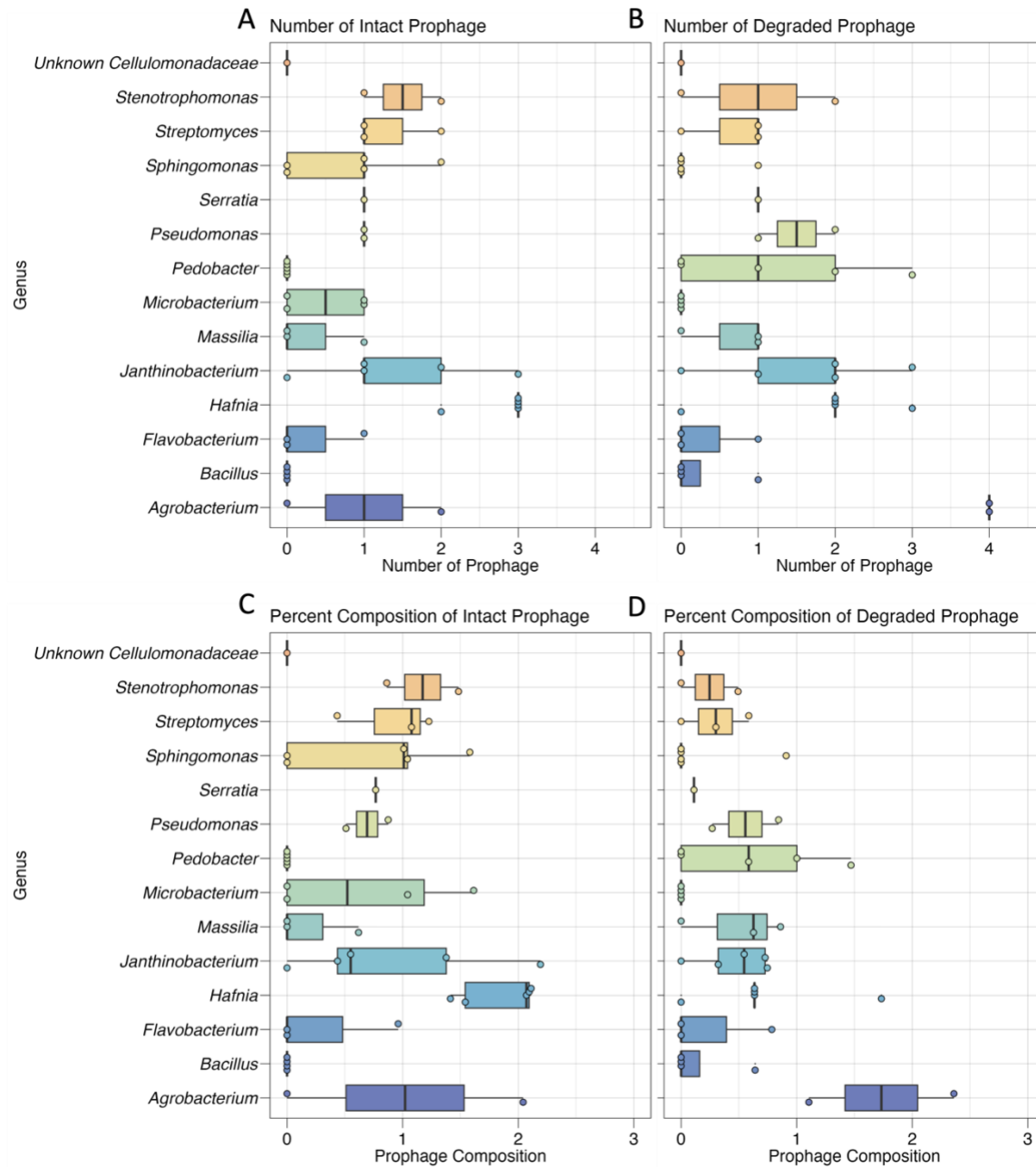


Figure 1: The abundance (absolute number) for all estimated intact (A) and degraded (B) prophages, as well as the prophage composition (total percent of prophage in the bacterial genome) across all estimated intact (C) and degraded prophages (D). Individual dots represent a single isolate and black bars represent the median.

Prophage Vary Across Bacterial Species Associated with Amphibian Skin

Bacterial genome size does not correlate with combined (intact and degraded) prophage abundance ($R=0.23$, $p=0.13$) or prophage composition ($R<0.01$, $p=1$) (Fig. S1). Abundance and composition of prophages appears to vary across bacterial host genera (Fig. 1). *Hafnia* spp. had the highest median of intact prophages (3 ± 0.4 , $n = 5$ isolates), followed by *Stenotrophomonas* spp., which had 1.5 ± 0.70 median intact phage ($n = 2$ isolates). No intact prophages were found in *Pedobacter* spp. ($n= 5$ isolates) or in the single isolate in the Cellulomonadaceae family. Intact prophage abundance closely reflected intact prophage composition (Fig. 1A, 1C). Median intact prophage composition was the highest in *Hafnia* spp. ($2.1\% \pm 0.3$), followed by *Stenotrophomonas* spp. isolates ($1.2\% \pm 0.4$) and *Streptomyces* isolates ($1.1\% \pm 0.4$) (Fig. 1A).

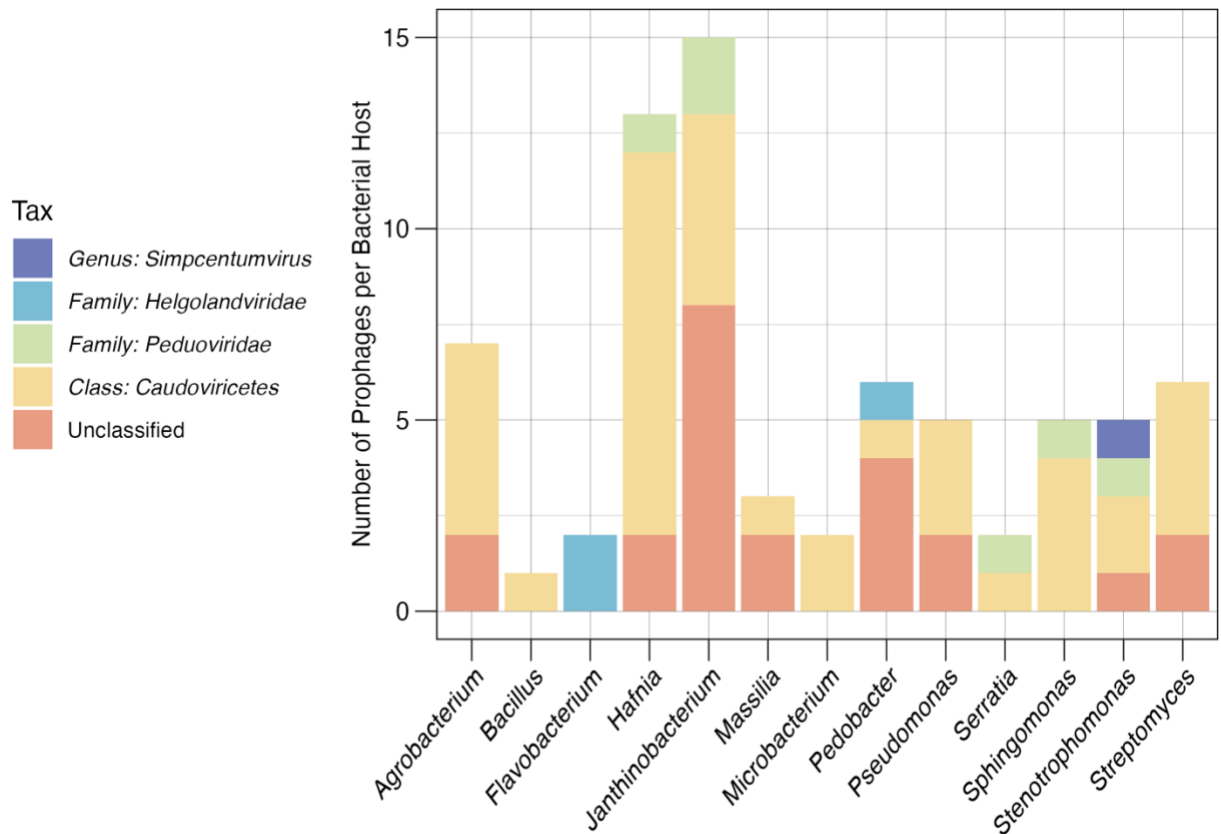


Figure 2. The number of dereplicated prophages identified per bacterial host. Stacked colored bars indicate the lowest possible taxonomic rank of each identified prophage.

Most Amphibian-Isolated Prophage are Unique to Bacterial Host

From the original 85 identified prophages (both degraded and intact), 72 were determined to be unique representative species, with the remaining 13 prophages exist within species clusters (Supplementary Data 1). Prophages from various *Hafnia* isolates were often closely related, forming species clusters of greater than 95% ANI similarity, while most prophages in other bacterial hosts were unique to each isolate. Slightly more than half (54.2%, n = 39) of these dereplicated prophages could be classified as unspecified phages belonging to the *Caudoviricetes* class, while another 32% (n=23) were completely unclassified (Fig. 2). Only 10 (14%) prophages were able to be classified at the family level (*Peduoviridae*, *Helgolandviridae*) or lower (*Simpcentumvirus*) (Fig. 2).

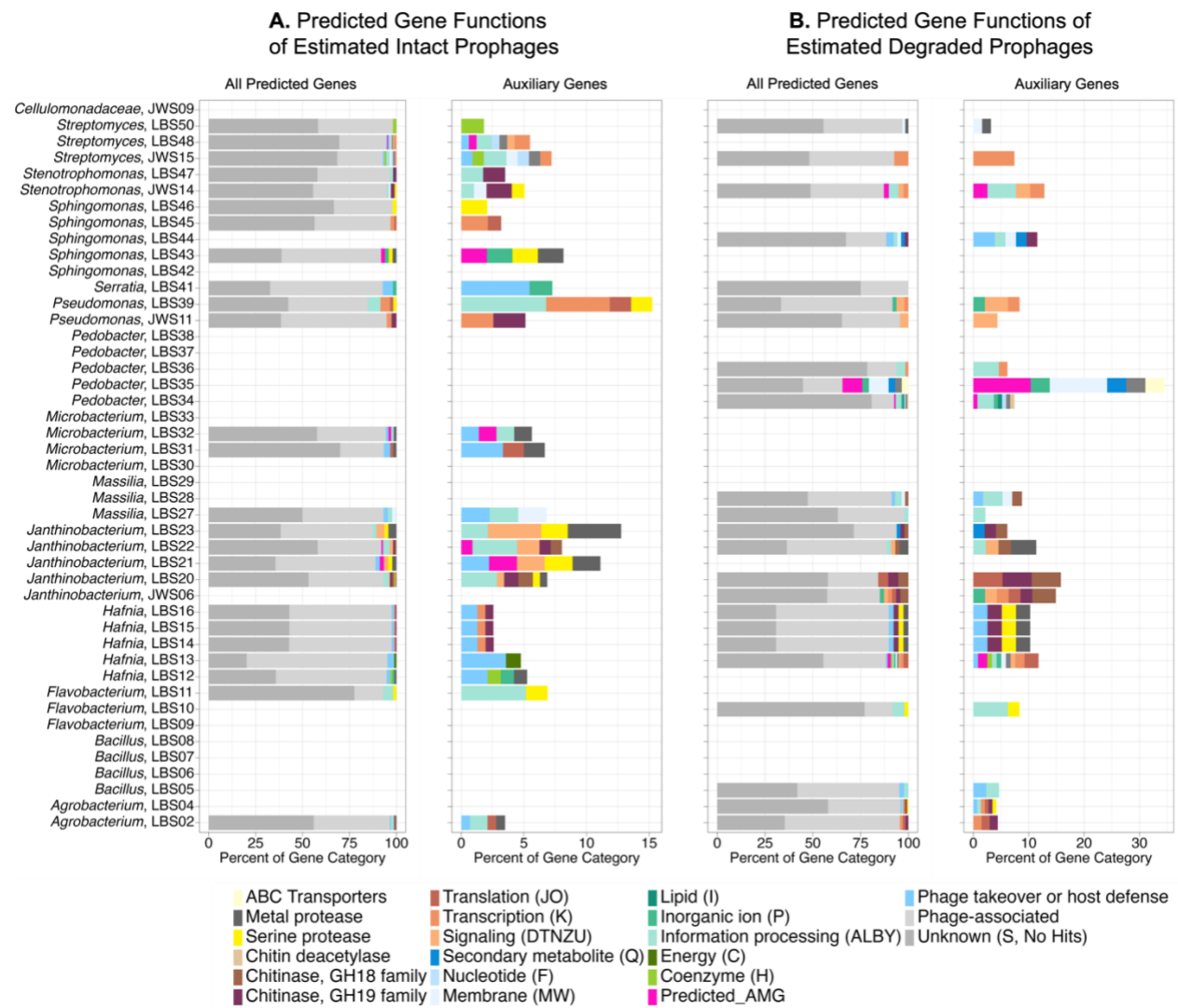


Figure 3. The functional composition of intact (A) and degraded (B) prophages by isolate, with complete predicted gene functions (left) and an enhanced view of the distribution of genes classified as potential auxiliary or antifungal genes (right). Antifungal classifications visually supersede initial functional categories assigned by EggNOG or PHROGs in this figure. The y-axis is grouped by bacterial host. White space indicates no estimated intact or degraded prophages were predicted for that isolate.

Potential auxiliary metabolic functions of amphibian-associated prophages

Approximately half (52.8%, $n = 1860$) of all predicted (intact and degraded) prophage-encoded proteins ($n = 3521$) were classified as unknown, and an additional 41.5% ($n = 1460$) were recognizable phage-associated proteins used in basic phage functions, such as tail proteins and integrases (Fig. 3, Supplementary Data 1). The combined COG categories of information processing (ALBY) (1.33%, $n=47$) and Transcription (K) (0.6%, $n=22$) were both common protein classifications. PHROGS identified 0.9% ($n=31$) of all proteins as belonging to either phage takeover or bacterial host defense; these were most commonly membrane-associated proteins ($n=10$) or toxin-antitoxin systems ($n=6$).

Broadly, metabolism, including likely AMGs based on the PHROGs database, consisted of 1.2% of all protein predictions ($n=44$). For example, phosphoadenosine phosphosulfate (PAPS) reductases appeared in two intact prophages from different *Janthinobacterium* isolates, and in different degraded prophages found in *Hafnia*, *Stenotrophomonas*, and *Pedobacter* isolates. A degraded *Pedobacter* prophage from a separate isolate also possessed 4 other AMGs (4.2% of *Pedobacter* prophage proteins): a 2OG-Fe(II) oxygenase, an ABC transporter, a cyanobacterial phosphoribosylglycinamide formyltransferase, and a Nol-like carbamoyltransferase. An intact *Sphingomonas* prophage also carried a 2OG-Fe(II) oxygenase. The same degraded *Hafnia* prophage with a PAPS reductase also carried a porphyrin-associated protein. Different estimated intact prophages in a *Streptomyces* isolate and a *Microbacterium* isolate were predicted to possess a phosphoheptose isomerase and flavodoxin protein, respectively.

Carbohydrate metabolism and transport (G) was the most common COG metabolic protein classification (0.3%, $n=10$). Nine of these were estimated to be GH18 chitinase-like proteins, and belonged to both intact and degraded prophages in *Janthinobacterium* and one degraded *Masilla* prophage. The tenth carbohydrate-associated protein was a GH19 chitinase encoded by a degraded *Sphingobacterium* prophage. Proteins predicted to contribute to inorganic ion metabolism and transport were present (0.23%, $n=8$), with bacterial hosts including *Hafnia*, *Sphingobacterium*, *Serratia*, *Pseudomonas*, *Pedobacter*, and *Janthinobacterium*. These proteins appear to be primarily involved in metal transport and binding, such as COG1629, an outer membrane receptor for iron transport found in two separate degraded prophages hosted by different *Pedobacter* isolates. An intact prophage hosted by a *Hafnia* isolate also appears to possess a protein associated

with the uptake of either iron or zinc. Two degraded prophages, from a *Janthinobacterium* and a *Pseudomonas* isolate, also appear to possess a potassium-binding protein (COG1652).

A total of 4 (0.1%) proteins were also predicted to be involved in coenzyme metabolism and transport. Two intact prophages from different *Streptomyces* isolates were predicted to encode either a phosphopantothenoylcysteine synthetase/decarboxylase or a 2-polyprenyl-3-methyl-5-hydroxy-6-methoxy-1,4-benzoquinol methylase, which are associated with CoA or Ubiquione biosynthesis respectively. An intact *Hafnia* prophage hosted by a different isolate encodes a methenyl tetrahydrofolate cyclohydrolase, while a degraded *Hafnia* prophage was predicted to encode a protein for molybdopterin or thiamine biosynthesis. A different intact *Hafnia* prophage also possesses the only protein COG associated with energy metabolism (<0.1%, n=1), as it shares homology to cytochrome b561. Predicted amino acid metabolism (<0.1%, n=1), lipid metabolism (<0.1%, n=1), and nucleotide metabolism (<0.1%, n=3) were also identified.

Prophage-encoded genes that may confer fungal growth inhibition

In addition to functional annotation using COG and PHROG categories, all proteins were searched for terms associated with Bd growth inhibition: serine proteases, metallo-proteases (including zinc), ABC transporters, and chitinase. The most common antifungal group in our analysis was chitinase. As discussed above, *J. lividum* prophages and one *Masilla* prophage encoded a total of 9 GH18 chitinase-like proteins. In addition to the single GH19 chitinase-like protein encoded by a *Sphingomonas* prophage, an additional 17 proteins were annotated by EggNog-mapper as GH19 chitinase-like proteins (n=18). A prophage in one *Pedobacter* isolate may also encode a protein described as a peptidoglycan, xylan or chitin deacetylase (COG0726). Predicted chitinases tended to share the greatest homology with other chitinases from prophage infecting the same bacterial host (Fig. 4). Additional domain verification with InterProScan determined that conserved chitinase domains were predicted in 10 of the 19 chitinase-like prophages (Fig. 4). In contrast, none of the GH18 chitinase-like predicted proteins possessed a conserved chitinase domain.

Metalloproteases and zinc-associated proteins (n=15) were the second most common, followed by serine proteases (n=13). The majority of metalloproteases and zinc-associated proteins appear to be associated with transcription (various zinc-finger proteins, n=11), but three intact prophages hosted by a single *Agrobacterium* isolate and two different *Microbacterium* isolates share homology with M78 family Zn-dependent peptidase. The majority of serine proteases appear to be involved in phage packaging and structure, such as the phage capsid scaffolding protein (GPO) serine peptidase (PHROG_73, n=8) and the periplasmic serine protease, CLpP class (PHROG_53, n=3). Only two ABC transporters were identified, carried by degraded prophages found in *Pedobacter* and *Hafnia* isolates.

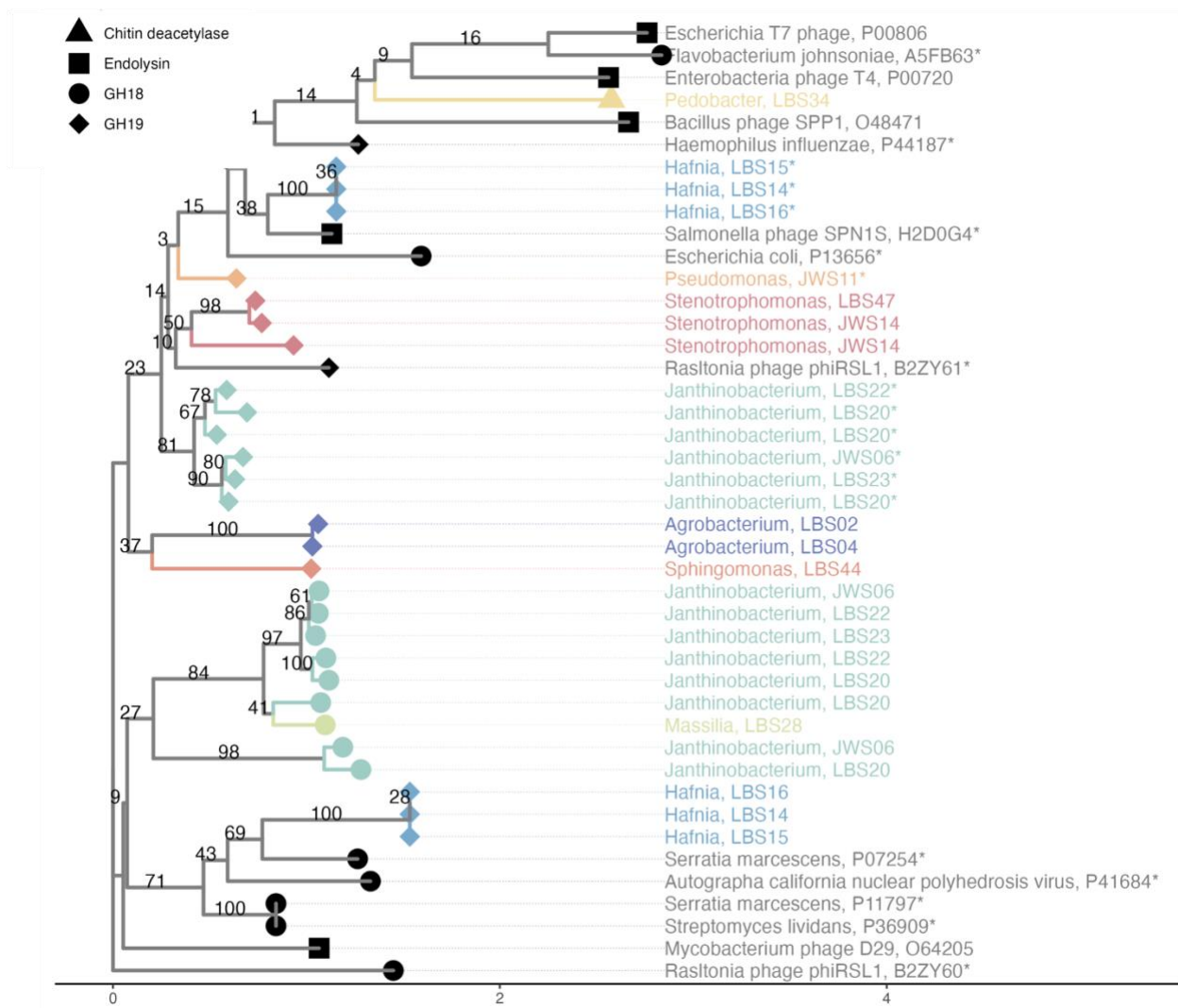


Figure 4: A phylogenetic maximum likelihood tree of all chitinase-like proteins found in prophages predicted from our amphibian-associated bacterial isolates. The shape of the tips indicate predicted class of chitinase. Colored tips indicate chitinase-like proteins identified from prophages of amphibian-isolated bacterial hosts, while black tips indicate reference proteins with known functions (indicated by tip shape) derived from various bacterial and viral isolates. Tip labels indicate the source of each chitinase-like protein; predicted chitinase-like prophages of amphibian bacterial isolates are labeled with bacterial host and isolate identifier. Reference proteins are labeled with a description of the sequence and a UniProt ID. Asterisk at the end of tips indicates the protein was predicted to possess a InterPro conserved chitinase domain.

Isolation source of J. lividum does not affect prophage densities and functional compositions, but key genes may correlate with estimated intactness

In the full set of 22 *J. lividum* genomes, there was no significant statistical relationship between prophage abundance or composition and the source of isolation (amphibian host, environmental, non-amphibian host) (Fig. 5). Additionally, sharing the same isolation source had no effect on genetic similarity between prophages or bacterial genomes (Fig. S1, Fig. S2).

Genes predicted to encode a chitinase-like protein were common among prophages isolated from *Janthinobacterium*, regardless of isolation source (Fig. 5). Of all intact *Janthinobacterium* prophages, 1.7% of genes were Chitinase-like proteins, while in degraded *Janthinobacterium* prophages, chitinase-like proteins were 5.4% of all predicted proteins. Of the GH19 chitinase-like proteins (n=16), 15 contained predicted conserved chitinase domains, while no GH18 chitinase-like proteins possessed a conserved chitinase domain (Supplementary Data 1).

In contrast, metal proteases were more common among estimated intact prophages (1.3%), compared to degraded prophages (0.6%). Serine proteases were only found in intact *Janthinobacterium* prophages (n=13, 0.7%). PHROGs also predicted six PAPS reductases and one queueine tRNA-ribosyltransferase in intact *Janthinobacterium* prophages (combined as Predicted_AMG in Fig. 5, 0.4%). Four different degraded prophages are predicted to encode a cytoplasmic potassium binding protein (0.8%) compared to just one intact prophage (0.1%).

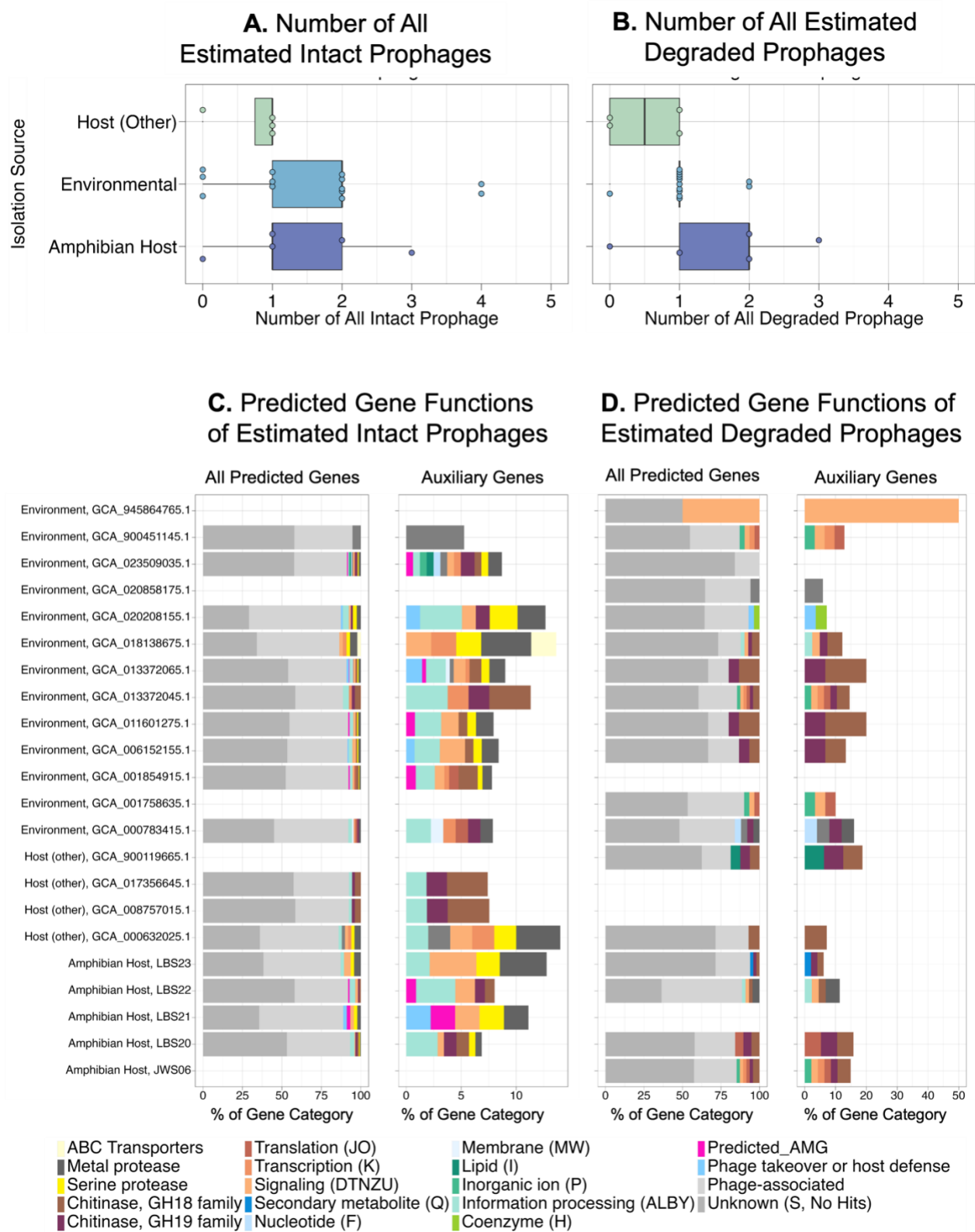


Figure 5: The abundance (absolute number) for all estimated intact (A) and degraded (B) *Janthinobacterium lividum* prophages. Individual dots represent a single isolate and black bars

represent the median. The functional composition of prophages from each *J. lividum* isolate, separated by intact (C) and degraded (D) prophages. All predicted gene functions are depicted in the left panel, while a focused view of potential auxiliary or antifungal genes of interest are on the right. Antifungal classifications visually supersede initial functional categories assigned by EggNOG or PHROGs in this figure.

Prophage-encoded chitinases in J. lividum appear to undergo purifying selection

The null model (M0) was the best fitting model to describe the rate of nonsynonymous to synonymous mutation for each type of chitinase gene, regardless of whether they were derived from estimated intact or degraded prophages (Table S1). Selection of the null model indicates that selection forces were even across all codon sites. Both GH18 chitinase (intact $\omega = 0.11124$, degraded $\omega = 0.12787$) and GH19 chitinase (intact $\omega = 0.06531$, degraded $\omega = 0.07915$) appear to undergo strong purifying selection, favoring synonymous mutations that leave amino acids intact (Table S1).

Discussion

This initial survey provides evidence that prophages are present in the amphibian skin microbiome and may influence the bacterial community by providing an array of auxiliary genes that could indirectly impact the amphibian host. The number of estimated intact prophages within a bacterial host appears to vary across bacterial genera, at least for the subset of isolates we examined. This suggests that certain bacterial species within the amphibian microbiome, like *Hafnia*, may experience more frequent phage integration events, as the bacterial host has not yet had evolutionary time to degrade its prophage into a nonfunctional genomic relic (Khan, Burmeister & Wahl, 2020). Alternatively, it may be that many *Hafnia* prophages provide a fitness advantage requiring phage intactness, as is the case with the CTX phage in *C. vibrio* (Pant, Das & Bhadra, 2020). In contrast, other bacterial taxa, like *Bacillus*, appear to have no intact prophage at all, suggesting that any phage-bacteria relationships are likely with phage that rarely integrate, or are exclusively lytic, which may impact the bacterial population via top-down control (Hsu et al., 2019).

Overall, identified prophages were unique to their bacterial hosts. No prophages appeared in two different bacterial isolates, except for nearly all prophages of *Hafnia* spp., and one prophage found in *J. lividum*. The 15 total phage identified for *Hafnia* spp. belonged to only 5 different phage species clusters (as defined by Roux et al., 2019). The presence of several reoccurring prophage in *Hafnia* suggests that phage integration likely occurred before strain divergence (Bobay, Touchon & Rocha, 2014). Additionally, the majority of these double-stranded DNA phage appear to belong to the taxonomic class *Caudoviricetes*, but were typically unable to be classified further.

A recent study on the human skin virome has suggested that auxiliary genes may be critical to the long term establishment of viruses in transient environments (Graham et al., 2023). In this study, we found evidence that some prophages associated with the amphibian skin microbiome appear to

possess auxiliary metabolic genes (AMGs), which may be useful to their bacterial hosts. For instance, phosphoadenosine phosphosulfate (PAPS) reductases were one of the most frequently predicted AMGs, appearing in prophages of multiple bacterial species, and may contribute to bacterial sulfur and amino acid metabolisms (Mara et al., 2020). Prophages from *Pedobacter* isolates appeared to encode a variety of AMGs, but were highly degraded and lacked nearly all key phage proteins outside of tail proteins, which are known to be enriched in degraded prophage (Khan, Burmeister & Wahl, 2020). Interestingly, a single predicted *Pedobacter* prophage encodes two cyanophage-associated AMGs, a Nol-like carbamoyltransferase (COG2192, KEG K00612) previously linked to fucose metabolism, and a phosphoribosylglycinamide formyltransferase associated with purine metabolism (Mergaert et al., 1996; Sullivan et al., 2005, 2010). These genes are typically found in phages that infect bacteria in the phylum *Cyanobacteria*, so their presence in *Pedobacter* (phylum: *Bacteroidota*) is interesting.

We also found evidence that putative chitinase-like genes are under purifying selection even in estimated degraded *J. lividum* prophage genomes. Chitin, which makes up the cell walls of fungi, as well as the exoskeletons of insects, is an abundant nitrogen and carbohydrate source in nature (Kumar et al., 2022). Thus, predicted phage-encoded chitinases may play a role in bacterial metabolism, like the photosynthetic genes that undergo purifying selection in cyanobacteria prophages (Zeidner et al., 2005). Alternatively, these phage-encoded chitinases may act as antifungals (Kumar et al., 2022). However, chitinases also share a high homology with phage endolysins, which break down peptidoglycan in bacterial cell walls during phage replication (Park et al., 2014). Core phage genes, such as endolysin and tail proteins, undergo purifying selection within bacterial hosts as well, as phage proteins can be weaponized into competitive toxins, such as bacteriocins, to use against other strains of bacteria (Bobay, Touchon & Rocha, 2014; Takala et al., 2023). Additional downstream phylogenetic and conserved domain homology suggests that the *J. lividum* prophage genes, which share homology to GH19 chitinases, may have chitinolytic function, while the predicted GH18 chitinases are more likely misidentified endolysins that target bacterial peptidoglycan. However, even with the additional *in silico* methods employed here, other studies have shown that *in vitro* testing of function is required, as endolysins often masquerade as GH19 chitinases in functional databases (Park et al., 2014; Orlando, 2020). Therefore, the function of these predicted chitinase-like proteins remains ambiguous without further testing. However, the strong purifying selection in bacterial hosts suggests these genes are beneficial to the bacterial hosts, as function appears to be conserved even in degraded prophages.

The presence of GH18 and GH19 prophage-encoded chitinase-like proteins across *J. lividum* isolates regardless of isolation source (amphibian, environmental, other hosts) suggests that whatever benefits are conferred to the bacterial host are not exclusive to the amphibian skin microbiome, and may be broadly advantageous in many environments. In general, the bacterial members of the amphibian skin community colonize hosts from the environment (Kueneman et al., 2014). As a result, *J. lividum* and its prophages may not evolve to specifically adapt to the amphibian host. Whether this is true of other amphibian skin microbiome bacteria and prophages

remains to be investigated. While a deeper examination of the interactions between other amphibian skin associated bacteria and their prophages will need to be conducted, the relationship between *J. lividum* and its prophages, as well as the presence of a diverse range of auxiliary metabolic genes in the prophages of several other bacterial species, suggests that prophages in the amphibian skin microbiome may have important impacts on bacterial communities, and potentially amphibian host health.

Conclusion

This is the first study to investigate the presence of phage in the amphibian skin microbiome by exploring the genetic contributions of prophages (phages embedded in bacterial genomes). Notably, *J. lividum* and other bacteria appear to possess prophage-encoded chitinase-like genes, which may be used by the bacterial host as competitive toxins, auxiliary metabolic genes, or antifungal agents. In *J. lividum*, these prophage-encoded chitinases can be found in bacterial isolates from both amphibian hosts and environmental sources. Additionally, these chitinase-like proteins undergo strong purifying selection, suggesting a benefit to the bacterial host. Overall, this study identified and characterized phages associated with the amphibian skin microbiome, as well as examined how the amphibian skin bacterial community may be granted additional functional capabilities via viral-encoded genes.

Supplemental Information

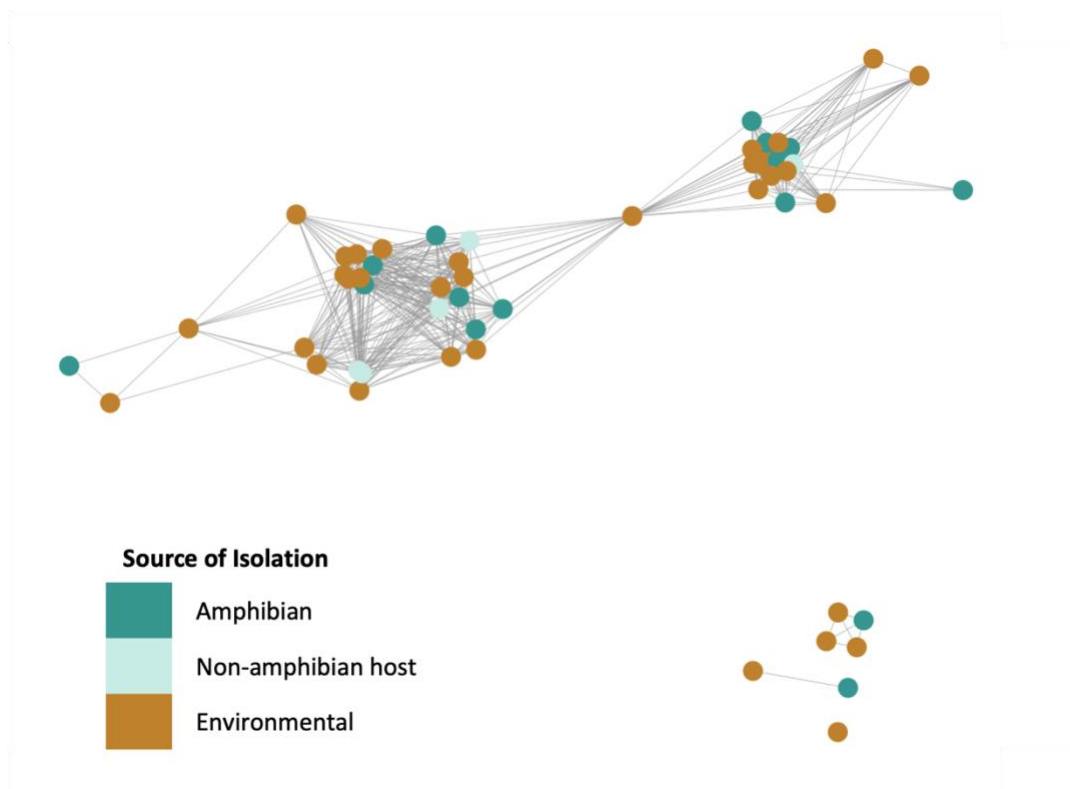


Figure S1. A network of relatedness based on shared proteins among *J. lividum* prophages from different isolation sources (indicated by color).

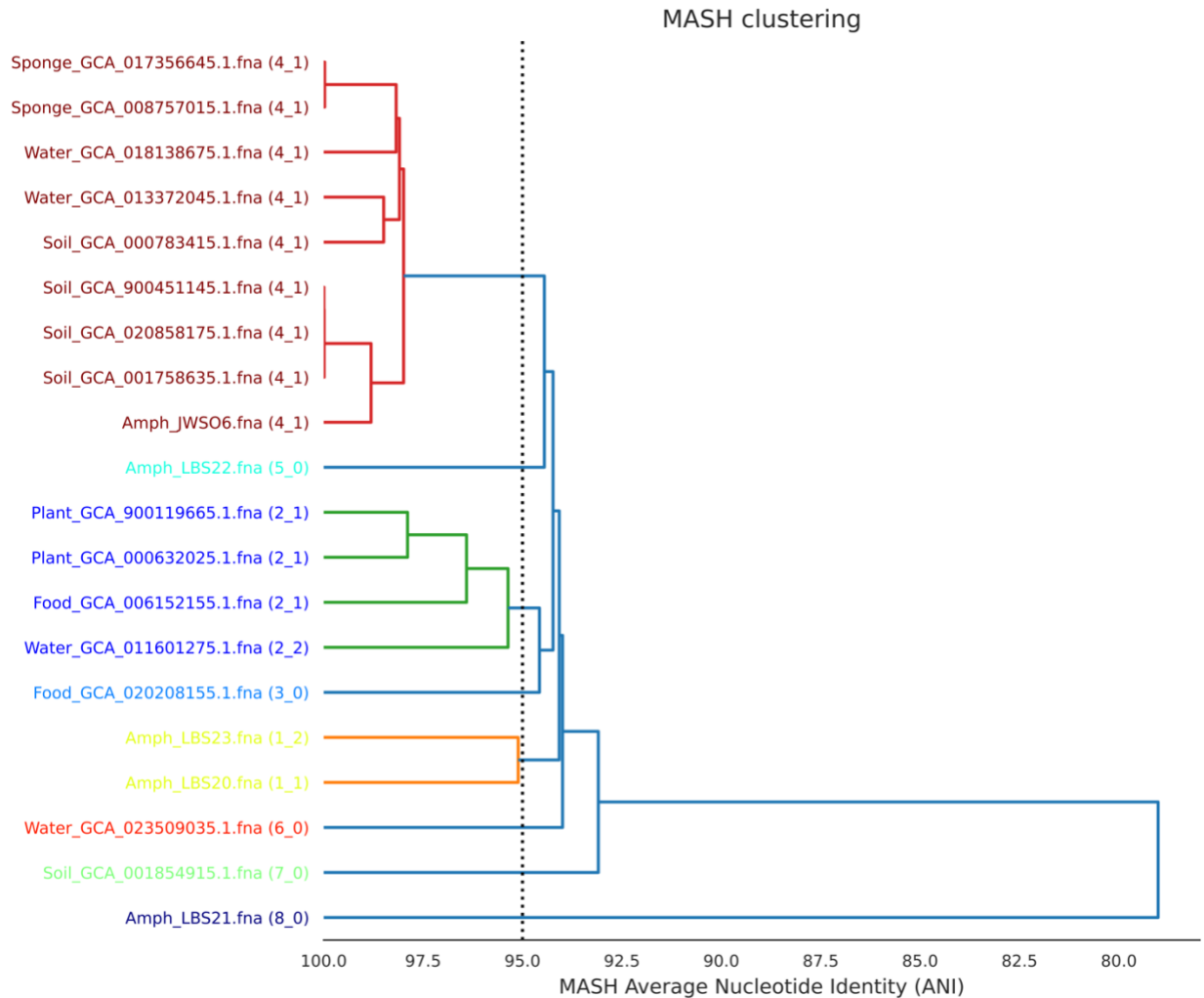


Figure S2. A dendrogram of *J. lividum* genomes based on average nucleotide similarity. Colors indicate different clusters of genomes which shared 95% or more MASH ANI. The tips are labeled with a description of the isolate source (Environment: “Water”, “Soil”, “Food”; Amphibian Host: “Amph”, or Non-Amphibian Host: “Sponge”, “Plant), the NCBI accession number or isolate ID, and the cluster of each *J. lividum*.

Site model (SM). - GH18 Chitinases, Estimated Degraded *J. lividum* prophages

Model	np	Ln L	Estimates of parameters				Model compared	LRT P-value	Positive sites
M3	37	-9420.386188	p:	0.07686	0.68956	0.23358	M0 vs. M3	0.000000000	[] Not Allowed
M0	33	-9526.642484	ω:	0.00140	0.07929	0.39862			
M2a	36	-9466.116578	ω0:	0.12787			M1a vs. M2a	1.000000000	[] Not Allowed
M1a	34	-9466.116578	p:	0.88967	0.10138	0.00895			
			ω:	0.10322	1.00000	1.00000	M7 vs. M8	0.003002677	318 W 0.691,396 T 0.607,413 Q 0.701 Not Allowed
M8	36	-9425.009076	p0=0.96849	p=1.03316	q=6.52517				
M7	34	-9430.817327	ω=	1.00000			M8a vs. M8	0.994358151	Not Allowed
			(p1= 0.03151)						
M8a	35	-9425.009051	p=	0.87827	q=	4.58498			
			p0=0.96849	p=1.03358	q=6.52732				
			(p1= 0.03151)	ω= 1.00000					

Site model (SM). - GH18 Chitinase, Estimated Intact *J. lividum* prophages

Model	np	Ln L	Estimates of parameters				Model compared	LRT P-value	Positive sites
M3	49	-7192.252157	p:	0.51962	0.44125	0.03913	M0 vs. M3	0.000000000	[] Not Allowed
M0	45	-7283.855900	ω:	0.02507	0.19572	0.80115			
M2a	48	-7235.494385	ω0:	0.11124			M1a vs. M2a	0.999999000	[] Not Allowed
M1a	46	-7235.494386	p:	0.92482	0.00797	0.06721			
			ω:	0.09294	1.00000	1.00000	M7 vs. M8	0.000375407	199 G 0.767,203 A 0.696,253 V 0.659 Not Allowed
M8	48	-7192.860948	p:	0.92482	0.07518				
M7	46	-7200.748447	ω:	0.09294	1.00000		M8a vs. M8	1.000000000	Not Allowed
			p0=0.97313	p=0.75915	q=5.50657				
M8a	47	-7192.860948	ω=	1.00000					
			(p1= 0.02687)						
			p=	0.67051	q=	4.00522			
			p0=0.97313	p=0.75918	q=5.50687				
			(p1= 0.02687)	ω= 1.00000					

Site model (SM). - GH 19 Chitinases, Estimated Degraded *J. lividum* prophages

Model	np	Ln L	Estimates of parameters				Model compared	LRT P-value	Positive sites
M3	25	-1793.31747	p:	0.77514	0.19751	0.02735	M0 vs. M3	0	[] Not Allowed
M0	21	-1843.83025	ω:	0.00857	0.30017	1.79121			
M2a	24	-1798.87772	ω0:	0.07915			M1a vs. M2a	1	[] Not Allowed
M1a	22	-1798.87772	p:	0.88421	0.04081	0.07498			
			ω:	0.02564	1	1	M7 vs. M8	0.000516401	26 D 0.900,112 T 0.958*,113 A 0.718 Not Allowed
M8	24	-1793.71939	p:	0.88421	0.11579				
M7	22	-1801.28802	ω:	0.02564	1		M8a vs. M8	0.996431764	Not Allowed
			p0=0.92393	p=0.04070	q=0.32453				
M8a	23	-1793.7194	ω=	1.00000					
			(p1= 0.07607)						
			p=	0.12949	q=	0.78497			
			p0=0.92393	p=0.04223	q=0.34210				
			(p1= 0.07607)	ω= 1.00000					

Site model (SM). -- GH19 Chitinases, Estimated Intact *J. lividum* prophages

Model	np	Ln L	Estimates of parameters				Model compared	LRT P-value	Positive sites
M3	25	-1793.31747	p:	0.77514	0.19751	0.02735	M0 vs. M3	0	[] Not Allowed
M0	21	-1843.83025	ω:	0.00857	0.30017	1.79121			
M2a	24	-1798.87772	ω0:	0.07915			M1a vs. M2a	1	[] Not Allowed
M1a	22	-1798.87772	p:	0.88421	0.04081	0.07498			
			ω:	0.02564	1	1	M7 vs. M8	0.000516401	26 D 0.900,112 T 0.958*,113 A 0.718 Not Allowed
M8	24	-1793.71939	p:	0.88421	0.11579				
M7	22	-1801.28802	ω:	0.02564	1		M8a vs. M8	0.996431764	Not Allowed
			p0=0.92393	p=0.04070	q=0.32453				
M8a	23	-1793.7194	ω=	1.00000					
			(p1= 0.07607)						
			p=	0.12949	q=	0.78497			
			p0=0.92393	p=0.04223	q=0.34210				
			(p1= 0.07607)	ω= 1.00000					

Supplemental Table 1: Likely ratio hood testing of dN/dS models used to test each

All other supplementary data (Supplementary Data 1) is available upon request via zip file.

References

- Abulikemu S, Yesilyurt A, Gencer D, Usta M, Nalcacioglu R. 2021. Comparison of the potential activities of viral and bacterial chitinases. *Egyptian Journal of Biological Pest Control* 31:91. DOI: 10.1186/s41938-021-00435-0.
- Argov T, Sapir SR, Pasechnek A, Azulay G, Stadnyuk O, Rabinovich L, Sigal N, Borovok I, Herskovits AA. 2019. Coordination of cohabiting phage elements supports bacteria–phage cooperation. *Nature Communications* 10:5288. DOI: 10.1038/s41467-019-13296-x.
- Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Research* 44:W16–W21. DOI: 10.1093/nar/gkw387.
- Bates KA, Sommer U, Hopkins KP, Shelton JMG, Wierzbicki C, Sergeant C, Tapley B, Michaels CJ, Schmeller DS, Loyau A, Bosch J, Viant MR, Harrison XA, Garner TWJ, Fisher MC. 2022. Microbiome function predicts amphibian chytridiomycosis disease dynamics. *Microbiome* 10:44. DOI: 10.1186/s40168-021-01215-6.
- Becker MH, Walke JB, Cikanek S, Savage AE, Mattheus N, Santiago CN, Minbiole KPC, Harris RN, Belden LK, Gratwicke B. 2015. Composition of symbiotic bacteria predicts survival in Panamanian golden frogs infected with a lethal fungus. *Proceedings of the Royal Society B: Biological Sciences* 282:20142881. DOI: 10.1098/rspb.2014.2881.
- Bobay L-M, Touchon M, Rocha EPC. 2014. Pervasive domestication of defective prophages by bacteria. *Proceedings of the National Academy of Sciences* 111:12127–12132. DOI: 10.1073/pnas.1405336111.
- Breitbart M, Bonnain C, Malki K, Sawaya NA. 2018. Phage puppet masters of the marine microbial realm. *Nature Microbiology* 3:754–766. DOI: 10.1038/s41564-018-0166-y.
- Brüssow H, Canchaya C, Hardt W-D. 2004. Phages and the Evolution of Bacterial Pathogens: from Genomic Rearrangements to Lysogenic Conversion. *Microbiology and Molecular Biology Reviews* 68:560–602. DOI: 10.1128/MMBR.68.3.560-602.2004.
- Bueren EK, Weinheimer AR, Aylward FO, Hsu BB, Haak DC, Belden LK. 2023. Characterization of prophages in bacterial genomes from the honey bee (*Apis mellifera*) gut microbiome. *PeerJ* 11:e15383. DOI: 10.7717/peerj.15383.
- Busico-Salcedo N, Owens L. 2013. Virulence Changes to Harveyi Clade Bacteria Infected with Bacteriophage from *Vibrio owensii*. *Indian Journal of Virology* 24:180–187. DOI: 10.1007/s13337-013-0136-1.
- Canchaya C, Proux C, Fournous G, Bruttin A, Brüssow H. 2003. Prophage Genomics. *Microbiology and Molecular Biology Reviews* 67:238–276. DOI: 10.1128/MMBR.67.2.238-276.2003.
- Cho I, Blaser MJ. 2012. The human microbiome: at the interface of health and disease. *Nature Reviews Genetics* 13:260–270. DOI: 10.1038/nrg3182.
- Ellison AR, Savage AE, DiRenzo GV, Langhammer P, Lips KR, Zamudio KR. 2014. Fighting a Losing Battle: Vigorous Immune Response Countered by Pathogen Suppression of Host Defenses in the Chytridiomycosis-Susceptible Frog *Atelopus zeteki*. *G3 Genes|Genomes|Genetics* 4:1275–1289. DOI: 10.1534/g3.114.010744.

- Fisher MC, Garner TWJ. 2020. Chytrid fungi and global amphibian declines. *Nature Reviews Microbiology* 18:332–343. DOI: 10.1038/s41579-020-0335-x.
- Fortier L-C, Sekulovic O. 2013. Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence* 4:354–365. DOI: 10.4161/viru.24498.
- Graham EH, Tom WA, Neujahr AC, Adamowicz MS, Clarke JL, Herr JR, Fernando SC. 2023. The persistence and stabilization of auxiliary genes in the human skin virome. *VIROLOGY JOURNAL* 20:49. DOI: 10.1186/s12985-023-02012-3.
- Guo J, Bolduc B, Zayed AA, Varsani A, Dominguez-Huerta G, Delmont TO, Pratama AA, Gazitúa MC, Vik D, Sullivan MB, Roux S. 2021. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* 9:37. DOI: 10.1186/s40168-020-00990-y.
- Harris RN, Brucker RM, Walke JB, Becker MH, Schwantes CR, Flaherty DC, Lam BA, Woodhams DC, Briggs CJ, Vredenburg VT, Minbiole KPC. 2009. Skin microbes on frogs prevent morbidity and mortality caused by a lethal skin fungus. *The ISME Journal* 3:818–824. DOI: 10.1038/ismej.2009.27.
- Hendrix RW. 2002. Bacteriophages: Evolution of the Majority. *Theoretical Population Biology* 61:471–480. DOI: 10.1006/tpbi.2002.1590.
- Hsu BB, Gibson TE, Yeliseyev V, Liu Q, Lyon L, Bry L, Silver PA, Gerber GK. 2019. Dynamic Modulation of the Gut Microbiota and Metabolome by Bacteriophages in a Mouse Model. *Cell Host & Microbe* 25:803–814.e5. DOI: 10.1016/j.chom.2019.05.001.
- Huang D, Yu P, Ye M, Schwarz C, Jiang X, Alvarez PJJ. 2021. Enhanced mutualistic symbiosis between soil phages and bacteria with elevated chromium-induced environmental stress. *Microbiome* 9:150. DOI: 10.1186/s40168-021-01074-1.
- Inglis LK, Roach MJ, Edwards RA. 2024. Prophages: an integral but understudied component of the human microbiome. *Microbial Genomics* 10:001166. DOI: 10.1099/mgen.0.001166.
- Jeffares DC, Tomiczek B, Sojo V, dos Reis M. 2015. A Beginners Guide to Estimating the Non-synonymous to Synonymous Rate Ratio of all Protein-Coding Genes in a Genome. In: Peacock C ed. *Parasite Genomics Protocols*. Methods in Molecular Biology. New York, NY: Springer, 65–90. DOI: 10.1007/978-1-4939-1438-8_4.
- Joneson S, Stajich JE, Shiu S-H, Rosenblum EB. 2011. Genomic Transition to Pathogenicity in Chytrid Fungi. *PLOS Pathogens* 7:e1002338. DOI: 10.1371/journal.ppat.1002338.
- Khan A, Burmeister AR, Wahl LM. 2020. Evolution along the parasitism-mutualism continuum determines the genetic repertoire of prophages. *PLOS Computational Biology* 16:e1008482. DOI: 10.1371/journal.pcbi.1008482.
- Khilyas IV, Sorokina AV, Markelova MI, Belenikin M, Shafigullina L, Tukhbatova RI, Shagimardanova EI, Blom J, Sharipova MR, Cohen MF. 2021. Genomic and phenotypic analysis of siderophore-producing *Rhodococcus qingshengii* strain S10 isolated from an arid weathered serpentine rock environment. *Archives of Microbiology* 203:855–860. DOI: 10.1007/s00203-020-02057-w.
- Kieft K, Zhou Z, Anantharaman K. 2020. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* 8:90. DOI: 10.1186/s40168-020-00867-0.
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research* 27:722–736. DOI: 10.1101/gr.215087.116.
- Kueneman JG, Parfrey LW, Woodhams DC, Archer HM, Knight R, McKenzie VJ. 2014. The amphibian skin-associated microbiome across species, space and life history stages. *Molecular Ecology* 23:1238–1250. DOI: 10.1111/mec.12510.

- Kumar M, Chakdar H, Pandiyan K, Thapa S, Shahid M, Singh A, Srivastava AK, Saxena AK. 2022. Bacterial chitinases: genetics, engineering and applications. *World Journal of Microbiology and Biotechnology* 38:252. DOI: 10.1007/s11274-022-03444-9.
- Lima SL, Colombo AL, de Almeida Junior JN. 2019. Fungal Cell Wall: Emerging Antifungals and Drug Resistance. *Frontiers in Microbiology* 10.
- Lips KR. 2016. Overview of chytrid emergence and impacts on amphibians. *Philosophical Transactions of the Royal Society B: Biological Sciences* 371:20150465. DOI: 10.1098/rstb.2015.0465.
- Mara P, Vik D, Pachiadaki MG, Suter EA, Poulos B, Taylor GT, Sullivan MB, Edgcomb VP. 2020. Viral elements and their potential influence on microbial processes along the permanently stratified Cariaco Basin redoxcline. *The ISME Journal* 14:3079–3092. DOI: 10.1038/s41396-020-00739-3.
- Matsuzaki S, Rashed M, Uchiyama J, Sakurai S, Ujihara T, Kuroda M, Ikeuchi M, Tani T, Fujieda M, Wakiguchi H, Imai S. 2005. Bacteriophage therapy: a revitalized therapy against bacterial infectious diseases. *Journal of Infection and Chemotherapy* 11:211–219. DOI: 10.1007/s10156-005-0408-9.
- Mergaert P, D’Haeze W, Fernández-López M, Geelen D, Goethals K, Promé J-C, Van Montagu M, Holsters M. 1996. Fucosylation and arabinosylation of Nod factors in *Azorhizobium caulinodans*: involvement of nolKnodZ as well as noeC and/or downstream genes. *Molecular Microbiology* 21:409–419. DOI: 10.1046/j.1365-2958.1996.6451366.x.
- Nayfach S, Camargo AP, Schulz F, Eloë-Fadrosch E, Roux S, Kyrpides NC. 2021. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nature Biotechnology* 39:578–585. DOI: 10.1038/s41587-020-00774-7.
- Oh J-H, Lin XB, Zhang S, Tollenaar SL, Özçam M, Dunphy C, Walter J, van Pijkeren J-P. 2019. Prophages in *Lactobacillus reuteri* Are Associated with Fitness Trade-Offs but Can Increase Competitiveness in the Gut Ecosystem. *Applied and Environmental Microbiology* 86:e01922-19. DOI: 10.1128/AEM.01922-19.
- Olm MR, Brown CT, Brooks B, Banfield JF. 2017. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *The ISME Journal* 11:2864–2868. DOI: 10.1038/ismej.2017.126.
- Orlando M. 2020. Biochemical and biophysical analysis of two Antarctic lysozyme endolysins and in silico exploration of glycoside hydrolase 19 sequence space. Università degli Studi di Milano-Bicocca,.
- Pant A, Das B, Bhadra RK. 2020. CTX phage of *Vibrio cholerae*: Genomics and applications. *Vaccine* 38:A7–A12. DOI: 10.1016/j.vaccine.2019.06.034.
- Park Y, Lim J-A, Kong M, Ryu S, Rhee S. 2014. Structure of bacteriophage SPN1S endolysin reveals an unusual two-module fold for the peptidoglycan lytic and binding activity. *Molecular Microbiology* 92:316–325. DOI: 10.1111/mmi.12555.
- Ramdass AC, Rampersad SN. 2021. Molecular signatures of *Janthinobacterium lividum* from Trinidad support high potential for crude oil metabolism. *BMC Microbiology* 21:287. DOI: 10.1186/s12866-021-02346-4.
- Ranwez V, Douzery EJP, Cambon C, Chantret N, Delsuc F. 2018. MACSE v2: Toolkit for the Alignment of Coding Sequences Accounting for Frameshifts and Stop Codons. *Molecular Biology and Evolution* 35:2582–2584. DOI: 10.1093/molbev/msy159.
- Roux S, Adriaenssens EM, Dutilh BE, Koonin EV, Kropinski AM, Krupovic M, Kuhn JH, Lavigne R, Brister JR, Varsani A, Amid C, Aziz RK, Bordenstein SR, Bork P, Breitbart M, Cochrane GR, Daly RA, Desnues C, Duhaime MB, Emerson JB, Enault F, Fuhrman JA, Hingamp P, Hugenholtz P, Hurwitz BL, Ivanova NN, Labonté JM, Lee K-B, Malmstrom RR, Martinez-Garcia M, Mizrachi IK, Ogata H, Páez-Espino D, Petit M-A, Putonti C, Rattei T, Reyes A,

- Rodriguez-Valera F, Rosario K, Schriml L, Schulz F, Steward GF, Sullivan MB, Sunagawa S, Suttle CA, Temperton B, Tringe SG, Thurber RV, Webster NS, Whiteson KL, Wilhelm SW, Wommack KE, Woyke T, Wrighton KC, Yilmaz P, Yoshida T, Young MJ, Yutin N, Allen LZ, Kyrpides NC, Elie-Fadrosh EA. 2019. Minimum Information about an Uncultivated Virus Genome (MIUViG). *Nature Biotechnology* 37:29–37. DOI: 10.1038/nbt.4306.
- Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, Poulos BT, Solonenko N, Lara E, Poulain J, Pesant S, Kandels-Lewis S, Dimier C, Picheral M, Searson S, Cruaud C, Alberti A, Duarte CM, Gasol JM, Vaqué D, Bork P, Acinas SG, Wincker P, Sullivan MB. 2016. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* 537:689–693. DOI: 10.1038/nature19366.
- Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. DOI: 10.1093/bioinformatics/btu153.
- Soorni A, Haak D, Zaitlin D, Bombarely A. 2017. Organelle_PBA, a pipeline for assembling chloroplast and mitochondrial genomes from PacBio DNA sequencing data. *BMC Genomics* 18:49. DOI: 10.1186/s12864-016-3412-9.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313. DOI: 10.1093/bioinformatics/btu033.
- Sullivan MB, Coleman ML, Weigele P, Rohwer F, Chisholm SW. 2005. Three Prochlorococcus Cyanophage Genomes: Signature Features and Ecological Interpretations. *PLOS Biology* 3:e144. DOI: 10.1371/journal.pbio.0030144.
- Sullivan MB, Huang KH, Ignacio-Espinoza JC, Berlin AM, Kelly L, Weigele PR, DeFrancesco AS, Kern SE, Thompson LR, Young S, Yandava C, Fu R, Krastins B, Chase M, Sarracino D, Osburne MS, Henn MR, Chisholm SW. 2010. Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environmental Microbiology* 12:3035–3056. DOI: 10.1111/j.1462-2920.2010.02280.x.
- Takala TM, Mokhtari S, Ahonen SL, Wan X, Saris PEJ. 2023. Wild-type Lactococcus lactis producing bacteriocin-like prophage lysins. *Frontiers in Microbiology* 14.
- Tang F, Bossers A, Harders F, Lu C, Smith H. 2013. Comparative genomic analysis of twelve Streptococcus suis (pro)phages. *Genomics* 101:336–344. DOI: 10.1016/j.ygeno.2013.04.005.
- Terzian P, Olo Ndela E, Galiez C, Lossouarn J, Pérez Bucio RE, Mom R, Toussaint A, Petit M-A, Enault F. 2021. PHROG: families of prokaryotic virus proteins clustered using remote homology. *NAR Genomics and Bioinformatics* 3:lqab067. DOI: 10.1093/nargab/lqab067.
- Thingstad TF. 2000. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnology and Oceanography* 45:1320–1328. DOI: 10.4319/lo.2000.45.6.1320.
- Thomas S, Izard J, Walsh E, Batich K, Chongsathidkiet P, Clarke G, Sela DA, Muller AJ, Mullin JM, Albert K, Gilligan JP, DiGuilio K, Dilbarova R, Alexander W, Prendergast GC. 2017. The Host Microbiome Regulates and Maintains Human Health: A Primer and Perspective for Non-Microbiologists. *Cancer Research* 77:1783–1812. DOI: 10.1158/0008-5472.CAN-16-2929.
- Touchon M, Bernheim A, Rocha EP. 2016. Genetic and life-history traits associated with the distribution of prophages in bacteria. *The ISME Journal* 10:2744–2754. DOI: 10.1038/ismej.2016.47.
- Walke JB, Becker MH, Hughey MC, Swartwout MC, Jensen RV, Belden LK. 2015. Most of the Dominant Members of Amphibian Skin Bacterial Communities Can Be Readily Cultured. *Applied and Environmental Microbiology* 81:6589–6600. DOI: 10.1128/AEM.01486-15.
- Walke JB, Becker MH, Hughey MC, Swartwout MC, Jensen RV, Belden LK. 2017. Dominance-function relationships in the amphibian skin microbiome. *Environmental Microbiology* 19:3387–3397. DOI: 10.1111/1462-2920.13850.

- Wax N, Walke JB, Haak DC, Belden LK. 2023. Comparative genomics of bacteria from amphibian skin associated with inhibition of an amphibian fungal pathogen, *Batrachochytrium dendrobatidis*. *PeerJ* 11:e15714. DOI: 10.7717/peerj.15714.
- Weinbauer MG. 2004. Ecology of prokaryotic viruses. *FEMS Microbiology Reviews* 28:127–181. DOI: 10.1016/j.femsre.2003.08.001.
- Wendling CC. 2023. Prophage mediated control of higher order interactions - Insights from multi-level approaches. *Current Opinion in Systems Biology* 35:100469. DOI: 10.1016/j.coisb.2023.100469.
- Wendling CC, Refardt D, Hall AR. 2021. Fitness benefits to bacteria of carrying prophages and prophage-encoded antibiotic-resistance genes peak in different environments. *Evolution* 75:515–528. DOI: 10.1111/evo.14153.
- Winter C, Bouvier T, Weinbauer MG, Thingstad TF. 2010. Trade-Offs between Competition and Defense Specialists among Unicellular Planktonic Organisms: the “Killing the Winner” Hypothesis Revisited. *Microbiology and Molecular Biology Reviews* : *MMBR* 74:42–57. DOI: 10.1128/MMBR.00034-09.
- Woodhams DC, LaBumard BC, Barnhart KL, Becker MH, Bletz MC, Escobar LA, Flechas SV, Forman ME, Iannetta AA, Joyce MD, Rabemananjara F, Gratwicke B, Vences M, Minbiole KPC. 2018. Prodigiosin, Violacein, and Volatile Organic Compounds Produced by Widespread Cutaneous Bacteria of Amphibians Can Inhibit Two *Batrachochytrium* Fungal Pathogens. *Microbial Ecology* 75:1049–1062. DOI: 10.1007/s00248-017-1095-7.
- Xia R, Sun M, Balcázar JL, Yu P, Hu F, Alvarez PJJ. 2023. Benzo[a]pyrene stress impacts adaptive strategies and ecological functions of earthworm intestinal viromes. *The ISME Journal* 17:1004–1014. DOI: 10.1038/s41396-023-01408-x.
- Yamada T, Satoh S, Ishikawa H, Fujiwara A, Kawasaki T, Fujie M, Ogata H. 2010. A jumbo phage infecting the phytopathogen *Ralstonia solanacearum* defines a new lineage of the Myoviridae family. *Virology* 398:135–147. DOI: 10.1016/j.virol.2009.11.043.
- Zeidner G, Bielawski JP, Shmoish M, Scanlan DJ, Sabehi G, Béjà O. 2005. Potential photosynthesis gene recombination between *Prochlorococcus* and *Synechococcus* via viral intermediates. *Environmental Microbiology* 7:1505–1513. DOI: 10.1111/j.1462-2920.2005.00833.x.

Chapter Four: Bacteriophage in the gut microbiome of honey bees (*Apis mellifera*) across the United States

Emma Bueren, Emma Bradford, Richard Fell, Jenifer Walke, David C. Haak, Lisa Belden

Abstract

The microbial communities associated with animal hosts, known as microbiomes, play a significant role in the health of their host, such as by stimulating the immune system or producing important metabolites. While the bacterial communities of some animal microbiomes can vary widely between individuals, the core bacterial community of the honey bee (*Apis mellifera*) is remarkably consistent in terms of taxonomic composition and functional capabilities across individual bees. However, less is known about how bacteriophage (viruses that infect bacteria) communities of honey bees, which are known to be diverse and change across bees. This study used double-stranded DNA viromes to examine the taxonomic and genetic structure of bacteriophage communities in honey bee guts collected across the entire United States, from Virginia to Washington. Notably, this study found that the viral composition at each site was highly specialized and largely taxonomically unclassifiable. However, neighboring sites often shared a greater subset of overlapping viral species compared to more distant sites. In addition, phage at most sites were predicted to predominantly infect bacteria in the genera *Bifidobacterium* and *Lactobacillus*, two core honey bee gut bacterial commensals. Viromes at several sites also appear to harbor a wide range of auxiliary metabolism genes and other potentially host-benefitting genes, such as genes for antibiotic resistance. While some auxiliary genes were present only at individual sites, others, particularly those associated with sucrose metabolism, were more widely found across the US. This finding suggests that honey bee bacteriophages may consistently play a role in sucrose metabolism, while also carrying a diverse array of secondary host-altering genes that may be specialized to individualized apiaries. Overall, these results suggest that bacteriophage communities of honey bees at an apiary-level may experience distance decay in community composition, but that certain auxiliary functions, like sucrose metabolism, may be selected for consistently, even across large geographic distances.

Introduction

Viruses that infect bacteria, also known as bacteriophages or phages, are the most common biological entity in nearly all environments (Hendrix, 2002), and can play important roles in microbial community dynamics. For instance, in the ocean, predatory phages (lytic phages), which immediately kill their bacterial hosts upon infection, exert top-down control on bacterial populations (Rodriguez-Brito et al., 2010). In addition to the direct manipulation of bacterial populations, this lysis of bacteria also results in the return of carbon and other valuable resources to the environment in a process called the viral shunt (Breitbart et al., 2018). Bacteriophages also

contribute to the genetic diversity of bacterial communities; some temperate phage integrate into the bacterial genome rather than killing the host, which can result in the acquisition of novel genes for the host (Canchaya et al., 2003). This process, called lysogenic conversion, can result in phenotypic changes, such as the development of motility or a virulent toxin, in the bacterial host (Fortier & Sekulovic, 2013; Tsao et al., 2018). When integrated into the bacterial host, temperate phages are called prophages, and they retain the ability to lyse and kill their bacterial host unless the bacterial host is able to degrade key viral genes (Buckling & Rainey, 2002). The relationship between bacterial hosts and prophages oscillates between mutualism and parasitism depending on the pair, as potential competitive advantages, such as immunity to other phage or increased metabolic capabilities, may outweigh the cost of carrying a potentially deadly passenger (Paul, 2008; Khan, Burmeister & Wahl, 2020; Huang et al., 2021; Wendling, 2023).

Although phages are also ubiquitous in microbiomes, far less is understood about their impact on bacterial communities in these host-associated systems. As in the environment, lytic phage infection in host-associated bacterial communities may directly decrease a host population, which in turn may alter the populations of other bacterial community members (Hsu et al., 2019; Federici, Nobs & Elinav, 2021). Bacterial lysis may also result in a phage feedback loop in host-associated systems, as increased lysis in a system like the gut can trigger host immune responses that induce prophage excision (Stecher, Maier & Hardt, 2013; Federici, Nobs & Elinav, 2021). As agents of horizontal gene transfer, host-associated phage may hasten the progression of disease by providing virulence-enhancing motility genes to pathogens or alternatively, valuable bioremediation genes to commensal gut bacteria (Tsao et al., 2018; Xia et al., 2023).

The honey bee gut microbiome has become an important system for investigating host-microbe interactions. Notably, the bacterial community is comprised of relatively few highly-conserved taxa, with approximately 10 key bacterial phylotypes globally (Moran, 2015; Engel et al., 2016). Additionally, the honey bee gut microbiome has significant impacts on honey bee health, as bacteria participate in the stimulation of the immune system, breakdown of nutrients, defense against pathogens, and even learning and social behaviors via the gut-brain axis (Engel, Martinson & Moran, 2012; Moran, 2015; Kwong & Moran, 2016; Emery, Schmidt & Engel, 2017; Zhang et al., 2022). Honey bees are also critically important agricultural pollinators, contributing millions of dollars yearly to the US economy, despite continued reports of struggling managed honey bee hives (vanEngelsdorp & Meixner, 2010; Calderone, 2012). Initial research on honey bee gut phage communities suggests that viral communities are diverse and may carry diverse metabolic genes, particularly those associated with the breakdown of carbohydrate (Bonilla-Rosso et al., 2020; Deboutte et al., 2020; Busby et al., 2022; Bueren et al., 2023).

In this study, we sampled 12 apiaries across a large transect of land in the United States (Virginia to Washington state). We then characterize the viral communities of double-stranded DNA phage present outside bacterial cells, such as lytic and temperate phages seeking new bacterial hosts to infect. We show that the phage community of apiaries are highly taxonomically diverse. However, neighboring apiaries are more similar to each other than geographically distant apiaries. Despite

the diversity in phages between sites, the distribution of bacterial hosts and functional capabilities were similar across all hosts, suggesting localized speciation of phage at each apiary. Apiaries also had a variety of phage-encoded metabolic genes, particularly those associated with carbohydrate metabolism and the breakdown of sugars like sucrose, as well as genes associated with antimicrobial resistance.

Methods

Hive Sampling

As part of a larger study investigating geographic variation in honey bee gut metagenomes across the United States, a total of 13 apiaries were sampled in June, 2022 (Fig. 1). The beekeeper of each apiary also filled out a brief survey detailing any supplementary treatments or feeding in the past year, as well as geographic information (latitude, longitude, elevation) (Supplementary Data 1). For each of the 13 sites, we aimed to collect two 50ml falcon tubes of bees from each of 7 overwintered hives (at site 3, only 5 hives were sampled). In each hive, bees were collected from the frame next to the brood nest. Sampled bees were kept on ice until they could be returned to the Belden lab at Virginia Tech (Blacksburg, VA) for storage at -70°C , or were frozen at the local site and then shipped overnight to the Belden lab and stored at -70°C after arrival.

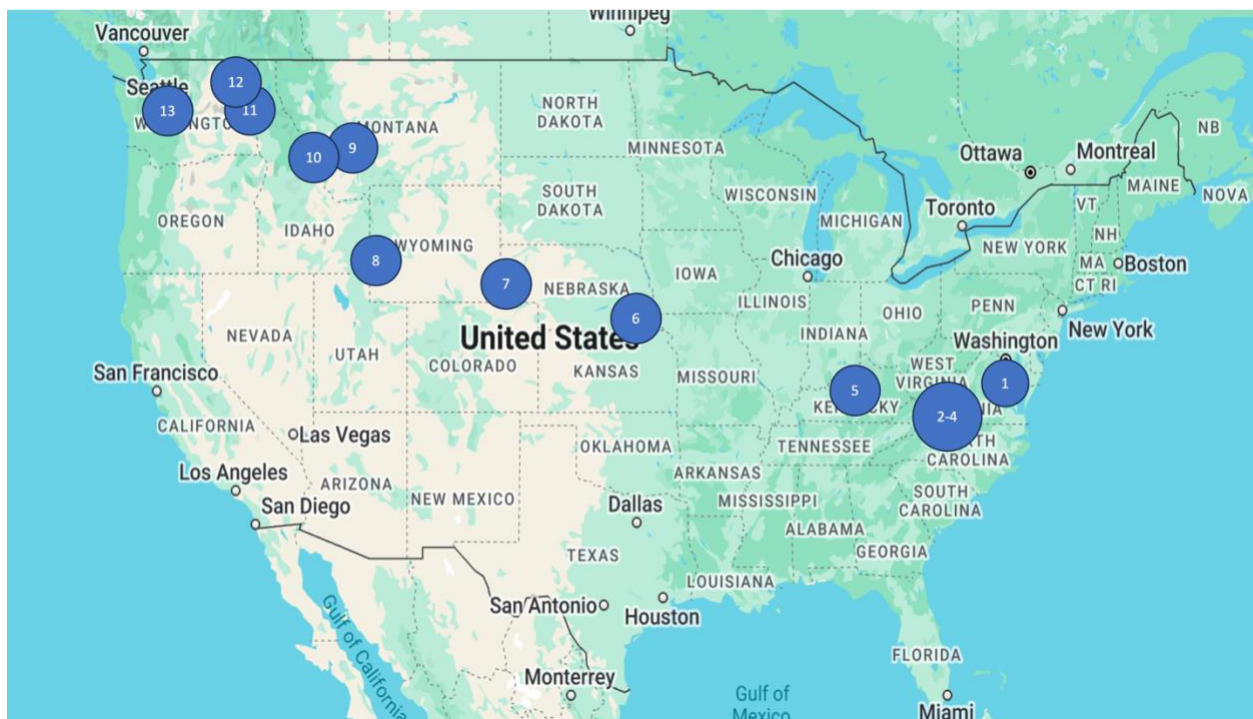


Figure 1. A map of apiary sampling locations, indicated by blue circles. Numbers on blue circles indicate site number.

Dissection

For this study, 8 bee guts were sampled from each hive and pooled together by site, resulting in 13 apiary-level viromes consisting of 56 bees. For the site 3 only 5 hives were sampled, so a total of 11 bees per hive were sampled/hive, resulting in 55 bees in the total virome for that site.

For dissections, bees were briefly thawed in sterile petri dishes and forceps were bleach and flame sterilized between each bee. Gloves were changed or bleached between hives. Individual bees were surface sterilized in 5% bleach and briefly rinsed three times in sterilized MilliQ water (Engel et al., 2013). Whole bee guts ($n = 8/\text{hive}$; 5 or 7 hives/apiary) were dissected and pooled into a 1.5ml microcentrifuge tube of 500 μl of phage buffer and kept on ice. Pooled hive samples were then homogenized using a handheld tissue grinder with sterilized pestles (Fisherbrand Pellet Pestle, 12-1413-61). After homogenization, samples from the same apiary were further pooled into a 15ml centrifuge tube, resulting in a single sample per apiary.

Virome Extraction

Viromes were then prepared for viral DNA extraction using a protocol modified from Bikel et al. (2022). Apiary samples were centrifuged at 4550g for 30 minutes. The supernatant was then vacuum filtered using a 20 μm nylon filter to remove remaining biological debris, before passing through both 0.45 μm and 0.22 μm filters. Samples were then concentrated in Amplicon Ultra-4 100K centrifugal filters by centrifuging at 4550g until only 200 μl remained (MilliporeSigma™ UFC800324). Samples were then rinsed six times by adding 3ml of phage buffer and spun until approximately 200 μl remained. The last rinse was centrifuged until only 150 μl remained, and that was transferred to a 1.5ml microcentrifuge tube. The walls of the filter were rinsed twice with 50 μl of phage buffer and transferred to the final sample. The sample was then stored overnight at 4°C.

The next day, 1 μl of DNase (Thermo Scientific™ EN0521) was added to each sample. Samples were incubated at 37.8°C for two hours before moving to 65°C for 10 minutes to deactivate the DNase (Bikel et al., 2022). After cooling to room temperature, each sample was extracted using the Qiagen Blood and Tissue Extraction Kit and eluted into 100 μl of molecular grade water (QIAGEN, Valencia, CA). DNA concentration of each sample was determined using a qubit fluorometer with a high sensitivity kit at three time points: before DNase was added (median: 37.6 ng/ μl $\pm$ 8.67 SD), after inactivation, but before DNA extraction (median: 4.18 $\pm$ 1.86), and after extraction (median: 0.81 ng/ μl $\pm$ 0.47 SD). Extracted DNA quality was assessed using a high sensitivity Agilent 2100 TapeStation in the Fralin Life Sciences Genomic Core Lab.

Sequencing and Assembly

Samples were briefly concentrated using a speedvac before being sent to the Duke Center for Genomic and Computational Biology for library construction via Kappa Hyperprep and sequencing on the Illumina Mi-Seq platform, using 2 x 250bp reads. This sequencing generated

8.7Gbp of sequences across the 13 samples, for an average of 670Mbp per sample (range: 562-1,035Mbp), with an average quality score of 34.

Reads were first cleaned using TrimGalore and visualized using FastQC (Andrew, 2010; Krueger, 2012). Cleaned reads were then aligned to the reference honey bee genome (GCF_003254395.2) using BWA (v 0.7.17) and all mapped hits were removed (Li & Durbin, 2010; Wallberg et al., 2019). Unmapped paired reads were *de novo* assembled using MEGAHIT with default settings, resulting in a total 198,456 of contigs across sites (Supplementary Data 1) (v1.2.9) (Li et al., 2015).

Viral OTU Prediction and Taxonomic Identification

To isolate putative viral sequences, contigs were analyzed with GeNOMAD (v1.5.0) using the end-to-end command (Camargo et al., 2023). Viral regions (vOTUs) predicted to have lower than 2% false positive rate were retained (n=23,898). All vOTUs larger than 1,000bp (n =8,676) were pooled and dereplicated at 95% ANI similarity using dRep (v1.2.9; command:--ignoreGenomeQuality -pa 0.95 -sa 0.95 -nc 0.85 -l 1000 --multiround_primary_clustering (n = 6,643) (Olm et al., 2017). Based on GeNOMAD taxonomic predictions, we removed vOTUs associated exclusively with eukaryotic taxonomic groups (n = 225), and therefore likely to be viruses other than phages. The bacterial hosts of the remaining prokaryotic vOTUs (n = 6,418) were predicted using IPHOP (v.1.3.3) with default settings (Lu et al., 2021; Coutinho et al., 2021; Roux et al., 2023).

Prokaryotic vOTUs were then combined with the complete January 2024 INPHARED database to create a custom reference database (prokaryotic virus database) of predicted vOTUs and known phage (Cook et al., 2021). This custom database was then mapped against the reads of each apiary virome using CoverM (v0.6.1) (settings: "--min-read-percent-identity 90 --min-covered-fraction 0") (Aroney et al., 2024). Reads that mapped to the custom prokaryotic virus database were retained for functional analysis. Raw read-mapping counts were then normalized to determine relative abundance of phage taxa across samples. Any reads that mapped directly to a sequence in the INPHARED database were assigned a predicted host based on available INPHARED metadata at the genus level.

Virome Functional Annotation

Protein-encoding genes of the assembled metagenomes prior to vOTU identification were predicted and annotated with with MetaCerberus (v1.1.3) (--fraggenescan --hmm "KOFam_all, COG, VOG, PHROG, CAZy") and then dereplicated at 95% ANI using CD-HIT EST (v. 4.8.1; -c .95 -n -l 250) (Fu et al., 2012; Figueroa III et al., 2024). These genes, then mapped against the cleaned phage reads using CoverM (Aroney et al., 2024). Raw counts were normalized to determine relative abundance across sites. Gene classifications from KEGG, FOAM, PHROGs, COG, and CAZY were merged into equivalent sections, primarily using the standard COG hierarchy (Supplementary Data 1). Genes that had multiple functions (ex: amino acid metabolism and carbohydrate metabolism) were reclassified to the lowest level possible (ex: Unspecified

metabolism). Any genes that had at least one metabolic category, in addition to a non-metabolic category, were also reclassified as “Unspecified metabolism.”

Statistical Analysis

Statistical tests were conducted in R (v4.3.2) with Rstudio. The alpha diversity of each apiary was calculated to determine a typical range of observed vOTUs (richness) and Simpson’s diversity (species evenness). After calculation with the R packages phyloseq and microbiome, the metrics were averaged and standard deviation was calculated (McMurdie & Holmes, 2013; Lahti, 2019). Linear models and Pearson’s correlation tests were used to determine if elevation or longitude had an impact on either alpha diversity metric.

To determine if geographic distance between apiaries was associated with viral community composition, the beta diversity between apiaries was calculated with both Bray-Curtis and Jaccard’s distances and ordinated using NMDS plots. The latitude and longitude of each site was then used to create a Haversine distance matrix with the package geosphere (<https://github.com/rspatial/geosphere>). Then, using vegan, a Mantel test was performed for each variable using both Bray-Curtis and Jaccard distances with a spearman correlation and 9999 permutations (Dixon, 2003). To test the relationship between longitude and community variance, the first and second axis of PCoAs (Principal Coordinate Analyses) using Bray-Curtis and Jaccard distances were extracted and Spearman correlations were run.

To better demonstrate pairwise Bray-Curtis and Jaccard distances between sites, values were also visualized with a heatmap. Heatmaps were also created to better visualize relative abundance of vOTUs and predicted proteins across sites. Beta diversity metrics (Bray-Curtis and Jaccard dissimilarity) and ordinations (NMDS, PCoA) were conducted with phyloseq. All alpha and beta diversity analyses were conducted with normalized relative abundance data.

Results

Phage communities at the apiary level are diverse, but possess a smaller subset of core viruses

A total of 11289 unique vOTUs were identified across all apiaries (Fig. 2, Fig. 3). Across apiaries, vOTU richness ranged between 2001 – 5476, with a mean of 4010 vOTUs $\pm$ 1000 SD. The Inverse Simpson’s diversity index ranged between 0.57 - 0.99, with a mean of 0.91 $\pm$ 0.13 SD (Fig. 2). Elevation had no statistical significance on either alpha diversity metric (Fig. S1). However, decreasing longitude (west to east) correlated with higher numbers of observed number of viral regions ($R = 0.6, 0.03$) (Fig. 2B). Inverse Simpson’s diversity had no statistical relationship to longitude (Fig. 2D).

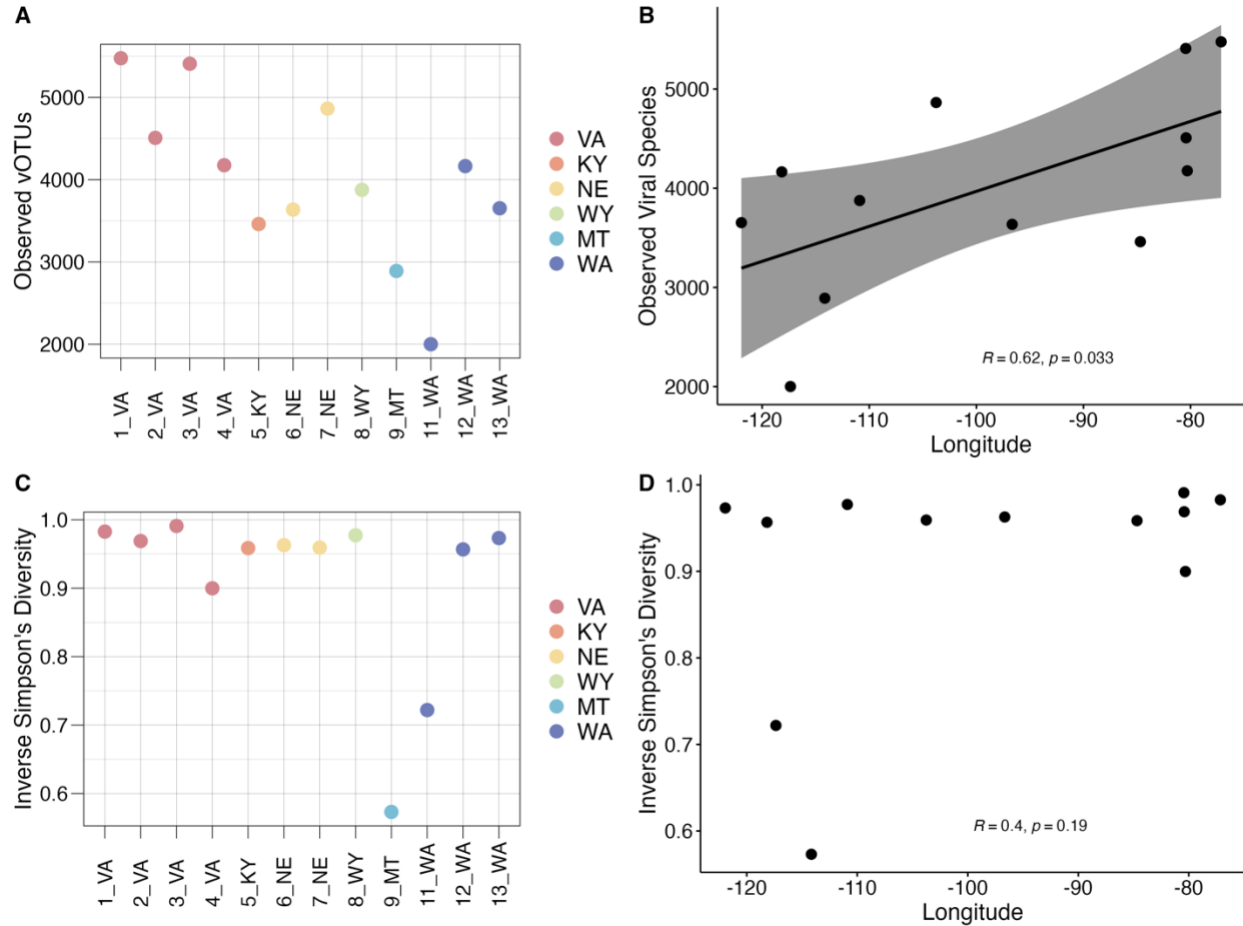


Figure 2. (A) The richness of each apiary, represented by observed vOTUs (B) correlates with longitude. Eastern sites are associated with a higher number of observed viral species. (C) Inverse Simpson's diversity, which represents community evenness, (D) does not correlate with longitude.

The majority of vOTUs were only classifiable as Phylum: *Uroviricota* ($n = 5826$) Class: *Caudoviricetes* (Phylum: *Uroviricota*) ($n = 3086$). All lower ranks classified belongs to the class *Caudoviricetes*. The most common families of vOTUs were *Autographiviridae* ($n=577$), *Straboviridae* ($n = 405$), *Drexleviridae* ($n=152$) and *Herelleviridae* ($n=138$). An additional 895 vOTUs were classified to other taxonomic families, while 210 remained completely unclassifiable (S Data).

No vOTU was shared across all 12 apiaries, but 1,029 vOTUs were shared across 11 out of the 12 (Fig. 3). The combined relative abundance of these 1,029 vOTUs was 6% - 79% of the entire bacteriophage community at each apiary.

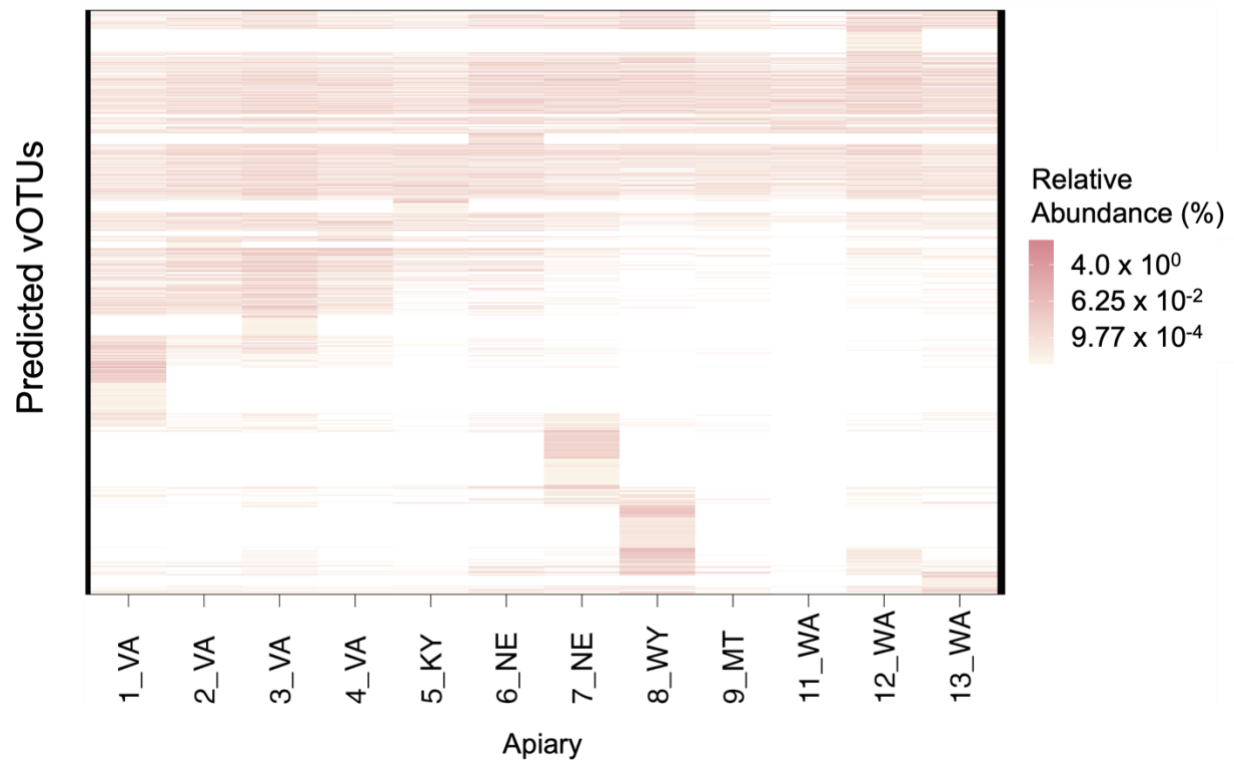


Figure 3. Color saturation of the lines in each site's column indicates relative abundance (percentage) of individual vOTUs, with no color indicating absence.

Geographically close apiaries are more likely to share similar phage communities

There is increased community dissimilarity with distance between sites (Bray-Curtis, Mantel $r = 0.14$, $p < 0.025$; Jaccard, Mantel $r = 0.21$, $p < 0.022$) (Fig. 4A; Fig. S3). Each site has a relatively unique profile of vOTUs (Fig. 3), but nearby sites, such as the three closest Virginia sites (Sites 2-4) tended to share lower pairwise Bray-Curtis and Jaccard values (Fig. 4B, Fig. S3B). Longitude did not statistically correlate with the most powerful driver of community variance (Bray-Curtis PCoA Axis 1: 14.6%, $p > 0.05$), but did correlate with the second most powerful driver of variance (Bray-Curtis PCoA Axis 2: 13.3%, $R=0.68$, < 0.02) (Fig. 4C, Fig. S2). Testing with Jaccard's index yielded comparable results (Fig. S3, Fig. S4).

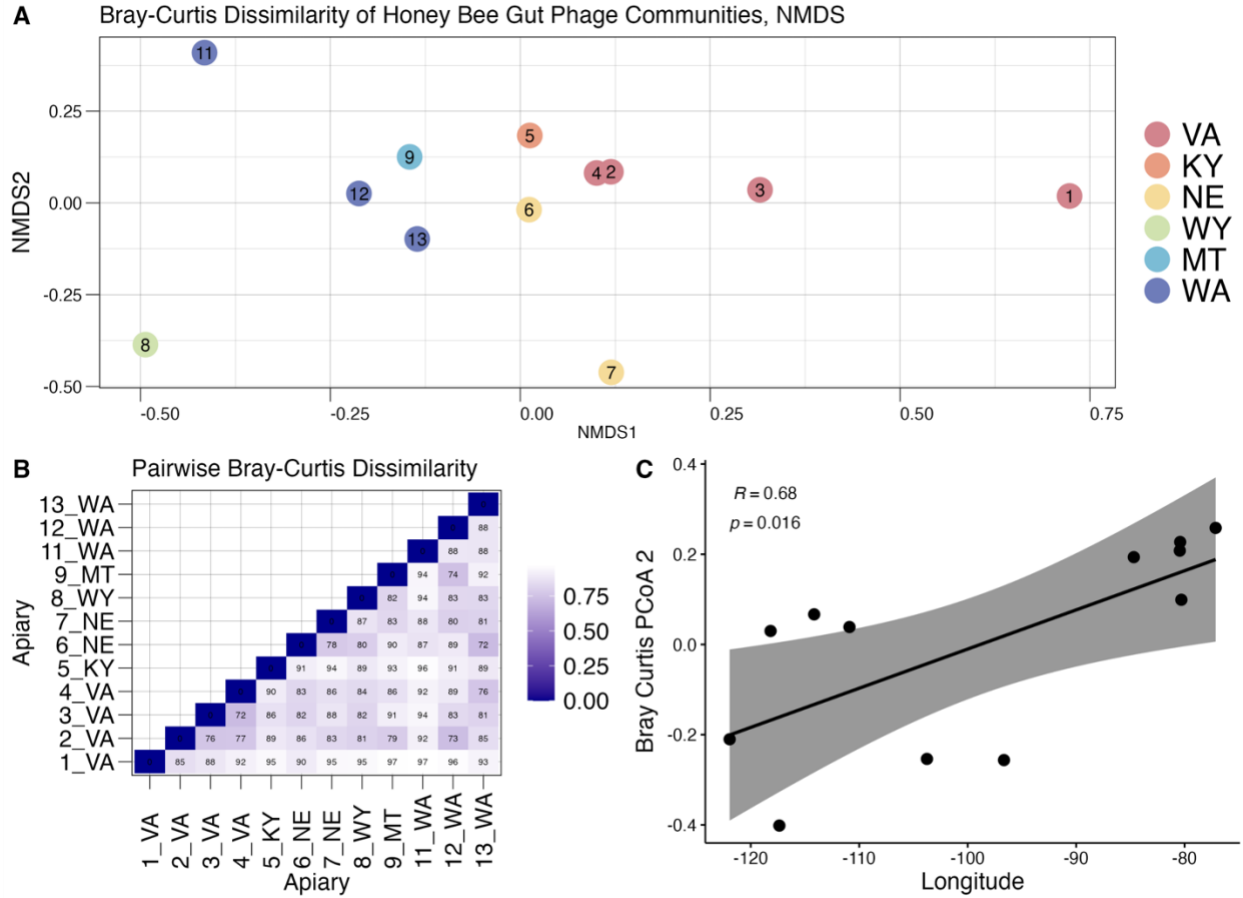


Figure 4. (A) An NMDS plot visualizing Bray-Curtis dissimilarity between phage communities of apiaries. The color and number on each point indicates state and site number of each apiary sampled. (B) A heatmap of the Bray-Curtis dissimilarity values visualized in 3B. Boxes with darker saturation and numbers closer to 0 indicate a higher similarity between sites. Numbers in each box represent the decimal Bray-Curtis value. (C) Sample site longitude correlates with the second strongest driver of community variance (PCoA Axis 2).

Bifidobacterium and *Lactobacillus* spp. are the most common predicted phage host

The viral communities of most apiaries were predominantly predicted to infect *Bifidobacterium* spp. (3.6-37.0%) and *Lactobacillus* spp. (4.5-34.1%) (Fig. 5). However, at Site 11_WA, the more than half of the phage population was predicted to infect *Spiroplasma* spp. (58.2%). At all other sites, the relative abundance of vOTUs predicted to infect *Spiroplasma* spp. were low (6_NE: 2.4%, 13_WA: 1.2%, Others < 0.2%). At lower relative abundances at many sites, vOTUs were also predicted to target *Gilliamella* spp. (1.0%-4.3%, 10/12 sites), *Acinetobacter* (1.1%-3.0%, 5/12 sites), and *Melissococcus* spp., (2.4 - 6.7%, 4/12 sites).

Among the 1,029 vOTUs present at 11 out of 12 sites, 441 vOTUs (42.9%) were predicted to infect *Lactobacillus* spp.. Other common predicted hosts were *Gilliamella* spp. (175 vOTUs, 17.0%) and *Bifidobacterium* spp. (117 vOTUs, 11.4%).

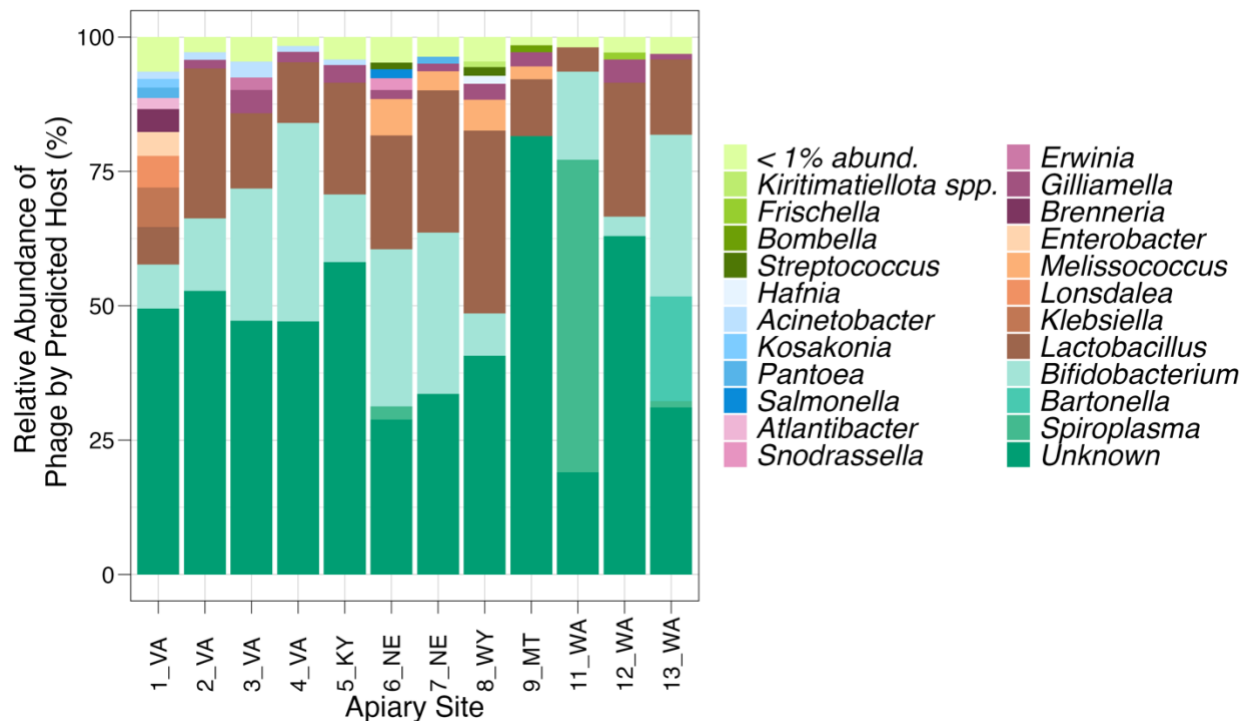


Figure 5. The relative abundance of vOTUs at each site pooled by predicted bacterial host.

Phage-encoded metabolic genes are primarily carbohydrate focused

Genes of unspecified metabolism ($n = 54$) and genes associated with carbohydrate metabolism were the most common categories of predicted auxiliary metabolic genes ($n = 136$ total) (Fig. S2). While the category of putative carbohydrate metabolism genes was found in all apiaries, individual viromes typically had unique AMGs shared with only a few other samples (Fig. 6A, Table S2). Several viromes had genes associated specifically with sucrose, such as a sucrose porin, levansucrase *sacB* and *sacC*. Genes associated with the degradation of plant cell components like pectin and starch were also identified (polygalacturonase and pullulanase, respectively). Other carbohydrate associated genes appear to target fructose (phosphofructokinase: L-glutamine-D-fructose-6-phosphate aminotransferases, Mannose/fructose-specific phosphotransferase component, or glucose (enolase and glucose-1-phosphate thymidyltransferase).

Phage encode both antimicrobial and antimicrobial resistance genes

Phage also appear to encode genes associated with the synthesis and resistance to antibiotics (n = 9 total) (Fig. 6B, Table S2). Genes associated with antibiotic production include beta-lactamase inhibitor proteins, which were found in every virome, and DTDP-glucose 4,6-dehydratase and Glucose-1-phosphate thymidyltransferase, which are both associated with streptomycin production. The majority of putative antibiotic resistance genes were associated with drug efflux pumps. However, proteins associated with resistance to specific antibiotics, such as trimethoprim (Dihydrofolate reductase), cephalosporin (Cephalosporin hydroxylase), and beta-lactams (PASTA domain) were also present. Additionally, several viromes appear to possess phage-encoded tellurite resistance.

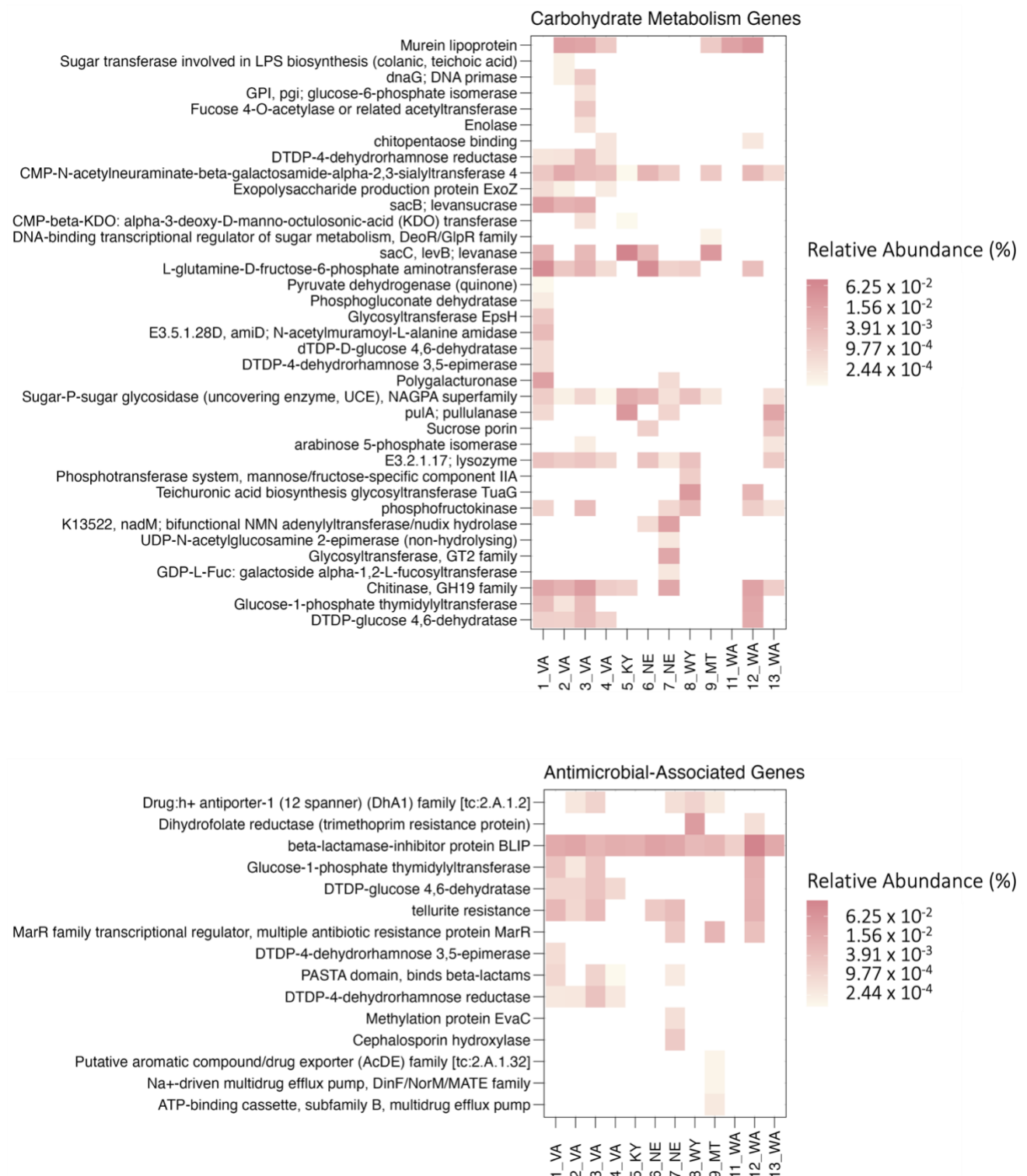


Figure 6. The distribution of (top) carbohydrate-associated metabolism and (bottom) antimicrobial resistance or biosynthesis genes in each community. Increasing color saturation indicates the relative abundance of the gene within the entire functional community of each apiary.

Discussion

We compared apiary-level dsDNA lytic bacteriophages from the honey bee gut community across a large (~3800 km) geographic transect. While phage communities are distinct in specific apiaries, our results suggest that there is distance decay in these communities, with apiaries that are closer together being more likely to share phage species. Although host-associated viromes are known for high diversity, the impact geographic distance has on community similarity is beginning to be recognized, and sometimes correlates more strongly with community differences than host phenotypes, such as disease state (Van Espen et al., 2021; Li et al., 2022; Tobin, Hill & Shkoporov, 2023). Recently, a similar effect of distance was found in the RNA phage community of native bees along the eastern United States seaboard (Robinson, Dolezal & Newton, 2024).

We found a core phage community appears to be present at most sites across our transect. This is surprising given the high level of unique bacteriophage taxa found in our study and in other studies (Bonilla-Rosso et al., 2020; Deboutte et al., 2020; Busby et al., 2022). While pathogenic viruses are capable of moving hive to hive through shared flower resources, the majority of the honey bee gut bacterial community is received from hive mates shortly after adults emerge from their brood cells and before they start leaving the hive (Graystock, Goulson & Hughes, 2015; Kwong & Moran, 2016). It seems unlikely, therefore, that bees are acquiring phage from the environment. The core phage community is more likely a result of the highly conserved bacterial taxa in the honey bee gut (Kwong & Moran, 2016), which may provide some stability and consistency to the phage community. However, there is also suspected to be remarkable strain diversity within the honey bee bacterial taxa, which may contribute to increased phage diversity and speciation in individual apiaries (Ellegaard et al., 2015; Ellegaard & Engel, 2019).

Despite sharing some viral taxonomic species, the richness of neighboring apiaries can vary widely. Interestingly, the observed number of predicted viral species decreased as sites moved across the United States, east to west. This may indicate that external factors, such as diet or climate variation, may contribute to the overall richness of the viral community. Notably, the evenness of the viral community was not affected, suggesting that community evenness may be more consistent across apiaries, regardless of location. The alpha diversity of honey bee viromes may correlate with bee health status or dysbiosis, something first noted due to increased abundance and dominance of free phage (infective lytic and temperate phage) in sickly queen bees (Caesar et al., 2024). Unusually high or low levels of alpha diversity of phages in the human guts has also been noted to correlated with diseases such as irritable bowel (Zuo et al., 2019; Clooney et al., 2019). Interestingly, the site with the lowest vOTU richness (Site 11_WA) also has a high relative abundance of phage targeting the bacterial host *Spiroplasma* spp., which is a genera containing two common honey bee pathogens (Zheng & Chen, 2014). However, the high relative abundance of *Spiroplasma* targeting phage may not necessarily reflect a high relative abundance of *Spiroplasma*. The interactions between phage and bacterial community structure are still being disentangled, although in some cases one may be a reflection or driver of the other (Shkoporov,

Turkington & Hill, 2022). Part of this issue may be the context dependent nature of phage-bacteria interactions, with some outcomes increasing bacterial fitness and some decreasing fitness.

Our results, as well as other bee virome studies previously, show that based on phage host predictions, bee gut phage predominantly infect *Bifidobacterium spp.*, and *Lactobacillus spp.*, two of the more typically abundant bacterial species in the gut (Bonilla-Rosso et al., 2020). We additionally found low levels of phage targeting other common honey bee commensal, such as *Gilliamella spp.*, and *Acinetobacter spp.*. Most *Bifidobacterium spp.* associated with the guts of honey bees have been estimated to carry few prophages (Bueren et al., 2023). Given both *Bifidobacterium spp.*, is typically abundant in the honey bee gut, the phages which target *Bifidobacterium spp.*, may be utilizing a “Kill-The-Winner” strategy which targets dominant bacterial species to maximize phage replication. Interestingly, despite their overall high abundance in viromes, *Bifidobacterium spp.* phages were a notably small proportion of the core community of viral species found at nearly all sites. *Bifidobacterium spp.*, phages may be highly diverse due to rapid evolution to escape the defenses of coevolving bacterial hosts (Buckling & Rainey, 2002). In contrast, a high number of core phages (17%) were predicted to infect *Gilliamella spp.* in spite of low levels of predicted *Gilliamella* phages at each site. This may suggest that *Gilliamella spp.*, which are predicted to have a high number of intact prophages, may periodically release genetically conserved prophages via spontaneous induction or in response to environmental triggers (Nanda, Thormann & Frunzke, 2015; Bueren et al., 2023). Low levels of prophage induction may be moderately beneficial to a bacterial population, as it can lower populations levels of bacterial competitors lacking the prophage (Gamage et al., 2003; De Paepe et al., 2014; Shkoporov, Turkington & Hill, 2022).

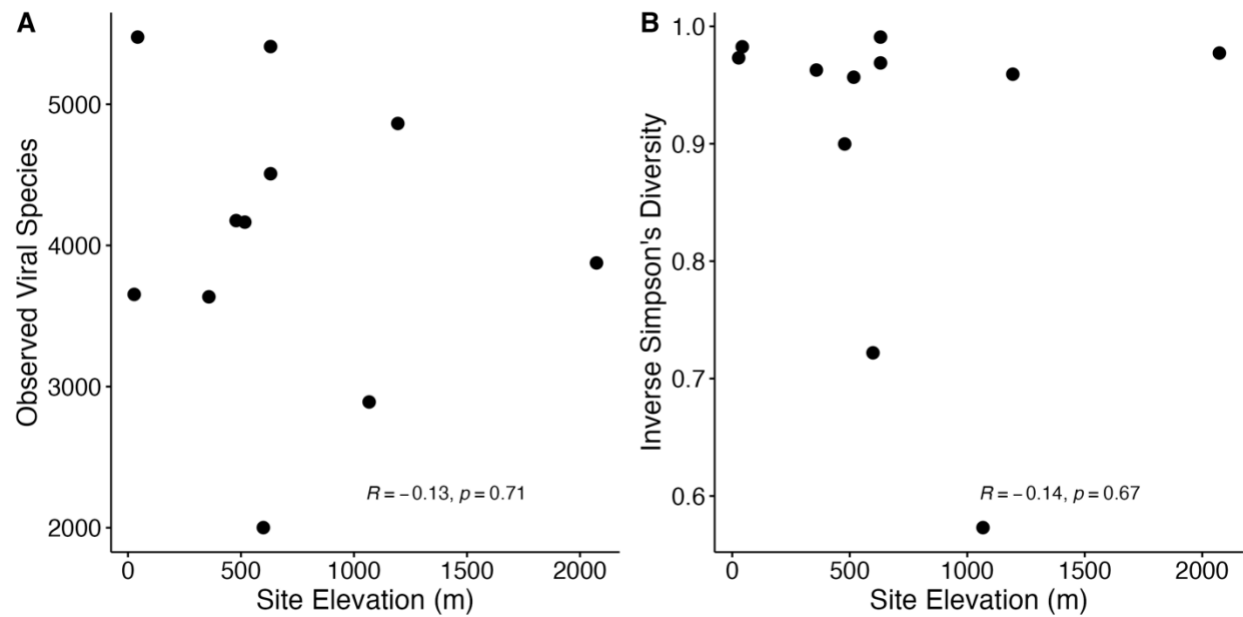
Broadly, all viromes across the US appear to encode genes associated with carbohydrate metabolism, but the specific genes present varied among viromes. For example, honey bee bacterial commensals may benefit from prophage-encoded genes to degrade pectin, a component of pollen (Bueren et al., 2023). Two apiary viromes (1_VA, 6_NE) encoded polygalacturonases, a pectinase associated with the stimulation and eventual bursting of pollen (Christensen, Munkres & Vannette, 2021). Other viromes encoded genes associated with sucrose metabolism, a common component of the honey bee diet (Taylor et al., 2019). Most common were the enzymes levansucrase *sacB*, which breaks apart the glucose and fructose bond of sucrose to release glucose and form levan, and levanase *sacC*, which releases fructose (Martínez-Fleites et al., 2005; Domżał-Kędzia et al., 2023). Levanase enzymes may also be beneficial to the bacterial host beyond metabolism, as high levels of levan can have negative impacts on Gram-negative bacteria (Pelicic, Reyrat & Gicquel, 1996). Several other carbohydrate-metabolism associated genes present in at least one apiary virome may also be involved in the production of antimicrobials. For example, Glucose-1-phosphate thymidyltransferase is associated with the synthesis of both polyketide sugars and streptomycin (Sivaraman et al., 2002).

The beta-lactamase inhibitor protein (BLIP) gene was the most prevalent gene associated with antimicrobial synthesis across viromes. This small protein enhances the efficacy of beta-lactam

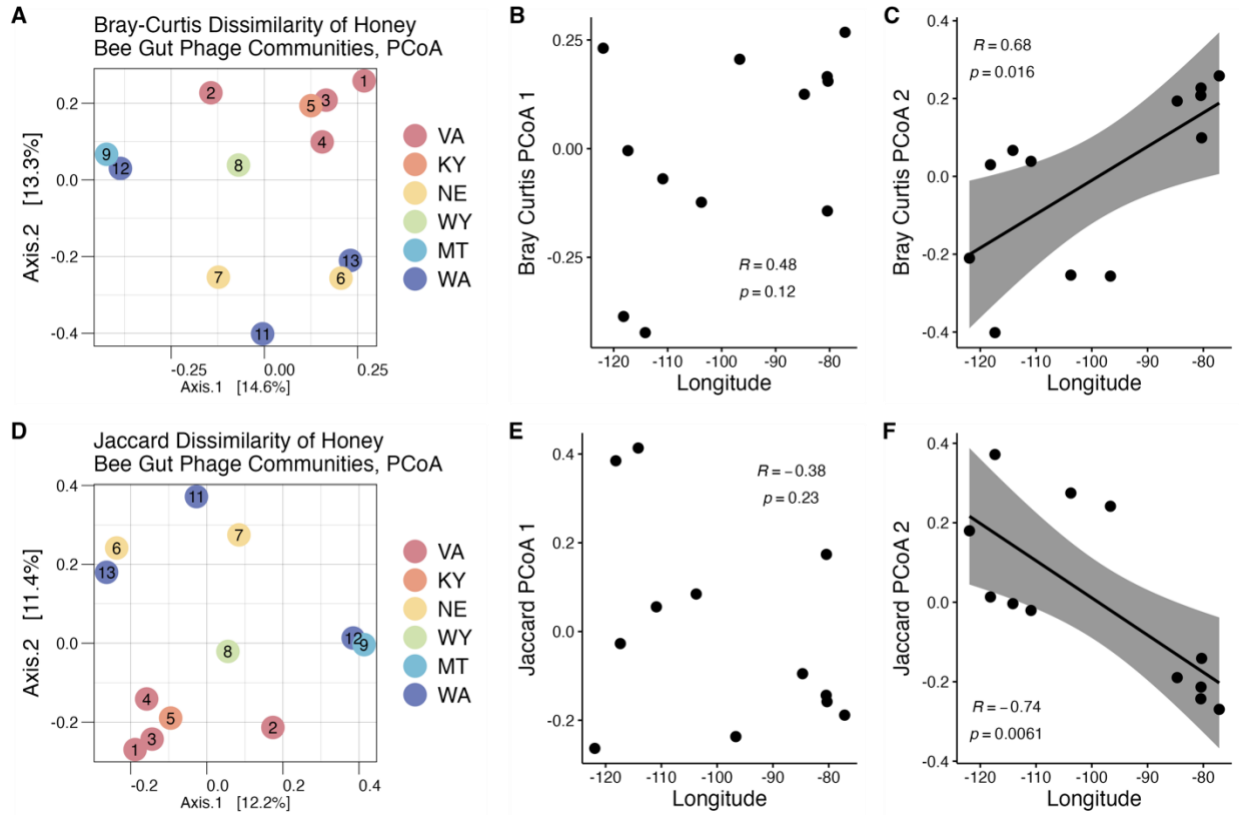
antibiotics by blocking common inhibition receptors (Drawz & Bonomo, 2010). Notably, two viromes additionally encoded beta-lactam antibiotic resistance via PASTA domains, suggesting warring bacterial communities (Yeats, Finn & Bateman, 2002). The most common encoded putative antibiotic resistance genes are multidrug efflux pumps, which likely are multipurpose (Nikaido & Pagès, 2012). Interestingly, putative resistance to cephalosporin and trimethoprim were also identified each in a single virome, although neither antibiotic is used in beekeeping. Resistance to these antibiotics may have developed as a response to antagonistic antibiotic-producing bacteria, rather than via beekeeper treatment or environmental exposures (Ortiz-Alvarado et al., 2020). Several viromes also encoded resistance to the antimicrobial tellurite, which is often associated with enhanced bacterial virulence and may be another response to bacterial competition (Moreno Switt et al., 2013; Turkovicova et al., 2016).

While the specific relationship between a bacteria and phage may range from antagonist to mutualistic, the broad presence of metabolic and antimicrobial genes suggest that the phage community is likely a reservoir of genetic diversity for bacterial hosts (Shkoporov, Turkington & Hill, 2022). Exclusively lytic phages, which may use such genes to temporarily reprogram bacterial hosts or pick up antibiotic resistance genes during replication due to incorrect packaging, are common vehicles of horizontal gene transfer; similarly, temperate phages harboring such genes may integrate as prophages upon infection, granting novel genes which alter the phenotype of bacterial host (Shkoporov, Turkington & Hill, 2022; Wendling, 2023). As a result, phages contribute to rapid strain diversification and evolution of their bacterial hosts (Stecher, Maier & Hardt, 2013; Shkoporov, Turkington & Hill, 2022). Ultimately, this study reveals that while honey bee gut bacteriophage taxa are highly diverse and change across large geographic distances, a core community of phage, as well as common classes of potentially auxiliary genes encoded by the virome, remain consistent.

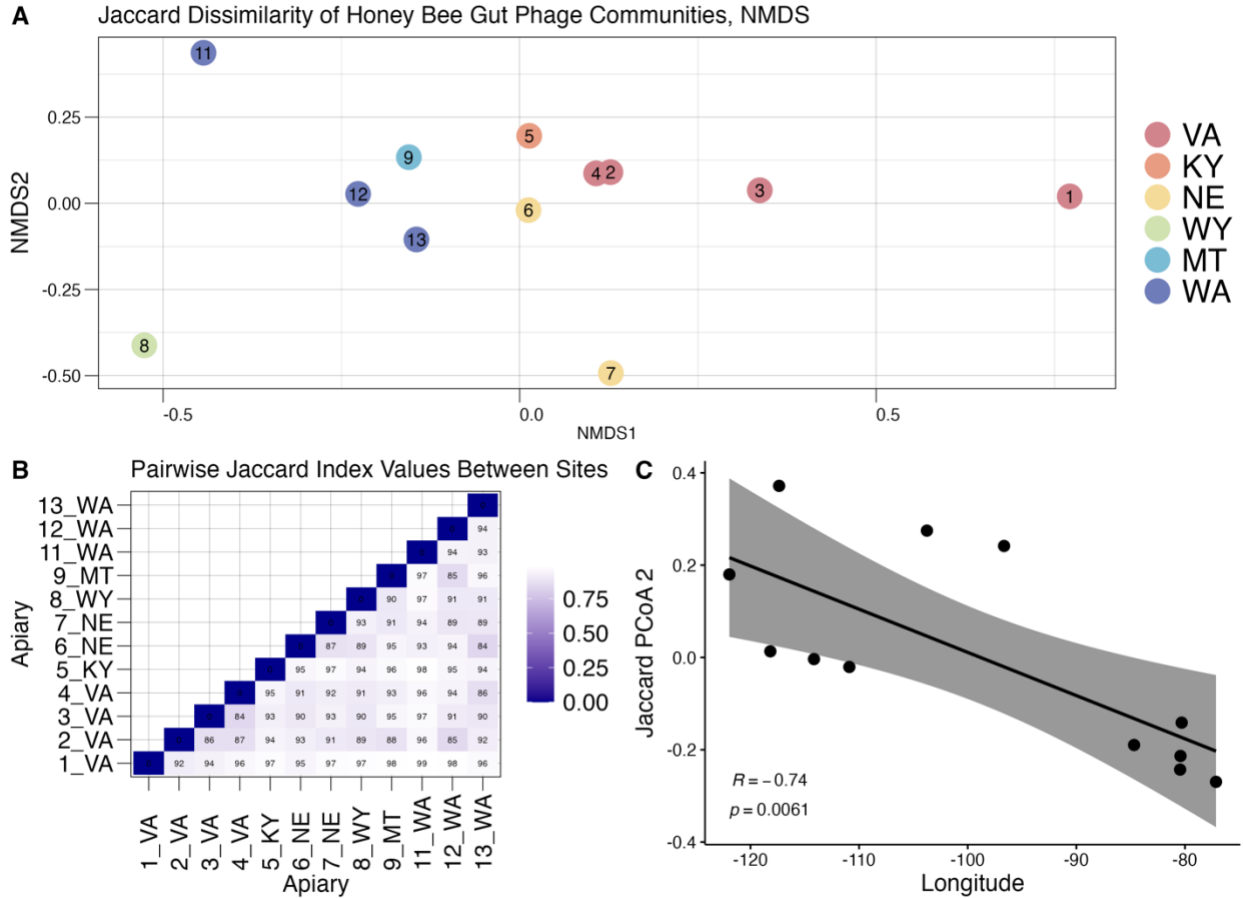
Supplementary Materials



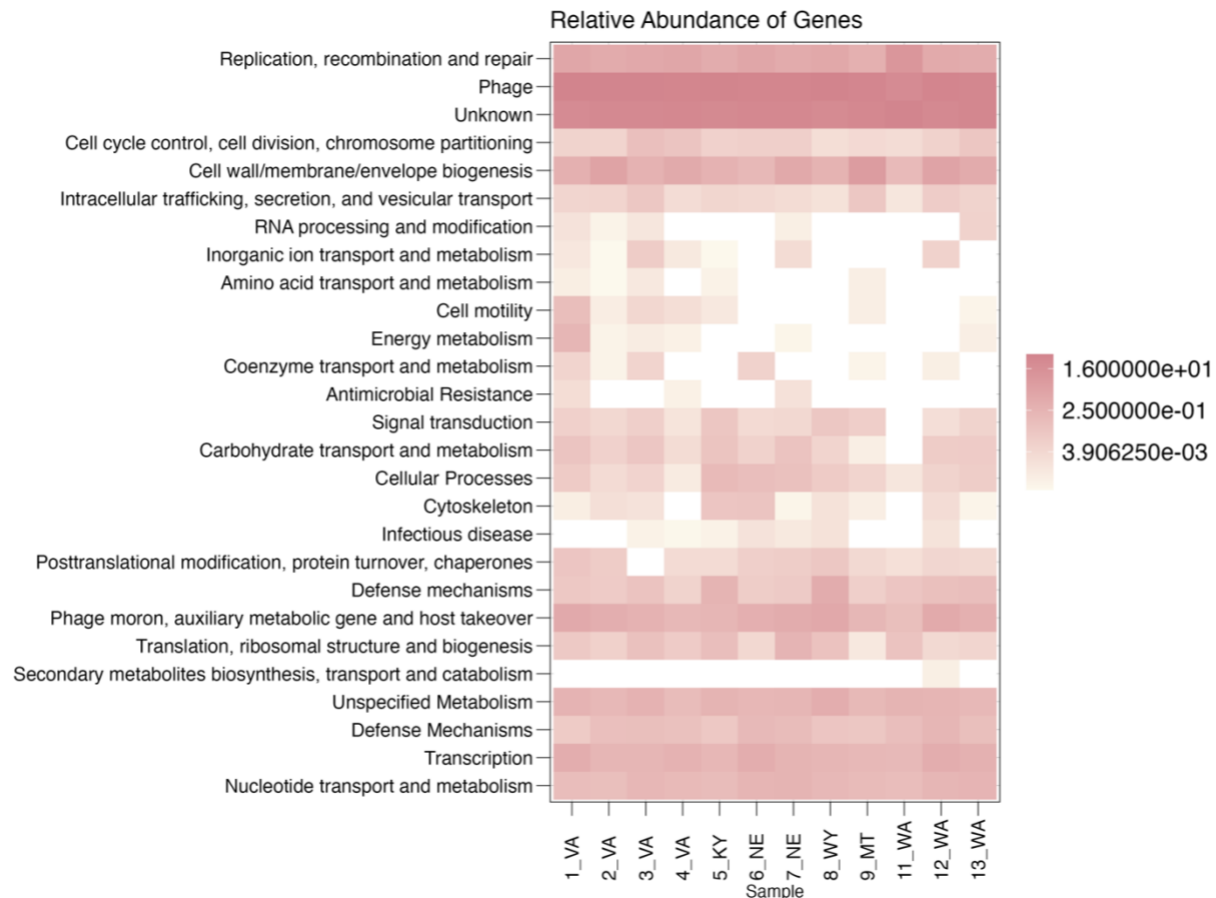
Supplementary Figure 1. The observed viral species (A) and inverse Simpson's diversity (B) did not correlate with the changes in elevation (m) of 11 of the sites. This analysis excludes Site 5, which did not have its elevation recorded at the time of sampling.



Supplementary Figure 2. Principal coordinate analysis (PCoA) of Bray-Curtis (A) and Jaccard (D) distances. X and Y axes represent the two strongest explanations for community variations, with total contributions represented by percentages. Longitude does not appear to correlate to PCoA axis 1 for either Bray-Curtis (B) or Jaccard dissimilarity (E) but the second axis of both Bray-Curtis (C) and Jaccard (F) PCoAs do correlate with longitude.



Supplementary Figure 3. An NMDS plot visualizing Jaccard dissimilarity between phage communities of apiaries. The color and number on each point indicates state and site number of each apiary sampled. (B) A heatmap of the Jaccard dissimilarity values visualized in 3B. Boxes with darker saturation and numbers closer to 0 indicate a higher similarity between sites. Numbers in each box represent the decimal Jaccard value. (C) Sample site longitude correlates with the second strongest driver of community variance (PCoA Axis 2).



Supplementary Figure 4. The relative abundance of genetic functional categories in each phage community. Increasing color saturation indicates the increasing relative abundance.

Additional supplementary information (Supplementary Data 1) related to sequencing, assembly, viral prediction protein annotation, and genetic sequences are available upon request via zip file or depository link.

References

- Andrew S. 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- Aroney STN, Newell RJP, Nissen J, Camargo AP, Tyson GW, Woodcroft BJ. 2024. CoverM: Read coverage calculator for metagenomics. DOI: 10.5281/zenodo.10531254.
- Bikel S, Gallardo-Becerra L, Cornejo-Granados F, Ochoa-Leyva A. 2022. Protocol for the isolation, sequencing, and analysis of the gut phageome from human fecal samples. *STAR Protocols* 3:101170. DOI: 10.1016/j.xpro.2022.101170.

- Bonilla-Rosso G, Steiner T, Wichmann F, Bexkens E, Engel P. 2020. Honey bees harbor a diverse gut virome engaging in nested strain-level interactions with the microbiota. *Proceedings of the National Academy of Sciences*. DOI: 10.1073/pnas.2000228117.
- Breitbart M, Bonnain C, Malki K, Sawaya NA. 2018. Phage puppet masters of the marine microbial realm. *Nature Microbiology* 3:754–766. DOI: 10.1038/s41564-018-0166-y.
- Buckling A, Rainey PB. 2002. Antagonistic coevolution between a bacterium and a bacteriophage. *Proceedings of the Royal Society B-Biological Sciences* 269:931–936. DOI: 10.1098/rspb.2001.1945.
- Bueren EK, Weinheimer AR, Aylward FO, Hsu BB, Haak DC, Belden LK. 2023. Characterization of prophages in bacterial genomes from the honey bee (*Apis mellifera*) gut microbiome. *PeerJ* 11:e15383. DOI: 10.7717/peerj.15383.
- Busby TJ, Miller CR, Moran NA, Van Leuven JT. 2022. Global Composition of the Bacteriophage Community in Honey Bees. *mSystems* 7:e01195-21. DOI: 10.1128/msystems.01195-21.
- Caesar L, Rice DW, McAfee A, Underwood R, Ganote C, Tarry DR, Foster LJ, Newton ILG. 2024. Metagenomic analysis of the honey bee queen microbiome reveals low bacterial diversity and Caudoviricetes phages. *mSystems* 9:e01182-23. DOI: 10.1128/msystems.01182-23.
- Calderone NW. 2012. Insect Pollinated Crops, Insect Pollinators and US Agriculture: Trend Analysis of Aggregate Data for the Period 1992–2009. *PLOS ONE* 7:e37235. DOI: 10.1371/journal.pone.0037235.
- Camargo AP, Roux S, Schulz F, Babinski M, Xu Y, Hu B, Chain PSG, Nayfach S, Kyrpides NC. 2023. You can move, but you can't hide: identification of mobile genetic elements with geNomad. :2023.03.05.531206. DOI: 10.1101/2023.03.05.531206.
- Canchaya C, Proux C, Fournous G, Bruttin A, Brüssow H. 2003. Prophage Genomics. *Microbiology and Molecular Biology Reviews* 67:238–276. DOI: 10.1128/MMBR.67.2.238-276.2003.
- Christensen SM, Munkres I, Vannette RL. 2021. Nectar bacteria stimulate pollen germination and bursting to enhance microbial fitness. *Current Biology* 31:4373-4380.e6. DOI: 10.1016/j.cub.2021.07.016.
- Clooney AG, Sutton TDS, Shkoporov AN, Holohan RK, Daly KM, O'Regan O, Ryan FJ, Draper LA, Plevy SE, Ross RP, Hill C. 2019. Whole-Virome Analysis Sheds Light on Viral Dark Matter in Inflammatory Bowel Disease. *Cell Host & Microbe* 26:764-778.e5. DOI: 10.1016/j.chom.2019.10.009.
- Cook R, Brown N, Redgwell T, Rihtman B, Barnes M, Clokie M, Stekel DJ, Hobman J, Jones MA, Millard A. 2021. INfrastructure for a PHAge REference Database: Identification of Large-Scale Biases in the Current Collection of Cultured Phage Genomes. *PHAGE* 2:214–223. DOI: 10.1089/phage.2021.0007.
- Coutinho FH, Zaragoza-Solas A, López-Pérez M, Barylski J, Zielezinski A, Dutilh BE, Edwards R, Rodriguez-Valera F. 2021. RaFAH: Host prediction for viruses of Bacteria and Archaea based on protein content. *Patterns (New York, N.Y.)* 2:100274. DOI: 10.1016/j.patter.2021.100274.
- De Paepe M, Leclerc M, Tinsley CR, Petit M-A. 2014. Bacteriophages: an underestimated role in human and animal health? *Frontiers in Cellular and Infection Microbiology* 4. DOI: 10.3389/fcimb.2014.00039.
- Deboutte W, Beller L, Yinda CK, Maes P, de Graaf DC, Matthijssens J. 2020. Honey-bee-associated prokaryotic viral communities reveal wide viral diversity and a profound metabolic coding potential. *Proceedings of the National Academy of Sciences* 117:10511–10519. DOI: 10.1073/pnas.1921859117.
- Dixon P. 2003. VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science* 14:927–930. DOI: 10.1111/j.1654-1103.2003.tb02228.x.

- Domżał-Kędzia M, Ostrowska M, Lewińska A, Łukaszewicz M. 2023. Recent Developments and Applications of Microbial Levan, A Versatile Polysaccharide-Based Biopolymer. *Molecules* 28:5407. DOI: 10.3390/molecules28145407.
- Drawz SM, Bonomo RA. 2010. Three Decades of β -Lactamase Inhibitors. *Clinical Microbiology Reviews* 23:160–201. DOI: 10.1128/cmr.00037-09.
- Ellegaard KM, Engel P. 2019. Genomic diversity landscape of the honey bee gut microbiota. *Nature Communications* 10:1–13. DOI: 10.1038/s41467-019-08303-0.
- Ellegaard KM, Tamarit D, Javelind E, Olofsson TC, Andersson SG, Vásquez A. 2015. Extensive intra-phylogroup diversity in lactobacilli and bifidobacteria from the honeybee gut. *BMC Genomics* 16. DOI: 10.1186/s12864-015-1476-6.
- Emery O, Schmidt K, Engel P. 2017. Immune system stimulation by the gut symbiont *Frischella perrara* in the honey bee (*Apis mellifera*). *Molecular Ecology* 26:2576–2590. DOI: 10.1111/mec.14058.
- Engel P, James RR, Koga R, Kwong WK, McFrederick QS, Moran NA. 2013. Standard methods for research on *Apis mellifera* gut symbionts. *Journal of Apicultural Research* 52:1–24. DOI: 10.3896/IBRA.1.52.4.07.
- Engel P, Kwong WK, McFrederick Q, Anderson KE, Barribeau SM, Chandler JA, Cornman RS, Dainat J, de Miranda JR, Doublet V, Emery O, Evans JD, Farinelli L, Flenniken ML, Granberg F, Grasis JA, Gauthier L, Hayer J, Koch H, Kocher S, Martinson VG, Moran N, Munoz-Torres M, Newton I, Paxton RJ, Powell E, Sadd BM, Schmid-Hempel P, Schmid-Hempel R, Song SJ, Schwarz RS, vanEngelsdorp D, Dainat B. 2016. The Bee Microbiome: Impact on Bee Health and Model for Evolution and Ecology of Host-Microbe Interactions. *mBio* 7:e02164-15, /mbio/7/2/e02164-15.atom. DOI: 10.1128/mBio.02164-15.
- Engel P, Martinson VG, Moran NA. 2012. Functional diversity within the simple gut microbiota of the honey bee. *Proceedings of the National Academy of Sciences* 109:11002–11007. DOI: 10.1073/pnas.1202970109.
- Federici S, Nobs SP, Elinav E. 2021. Phages and their potential to modulate the microbiome and immunity. *Cellular & Molecular Immunology* 18:889–904. DOI: 10.1038/s41423-020-00532-4.
- Figuerola III JL, Dhungel E, Bellanger M, Brouwer CR, White III RA. 2024. MetaCerberus: distributed highly parallelized HMM-based processing for robust functional annotation across the tree of life. *Bioinformatics* 40:btac119. DOI: 10.1093/bioinformatics/btac119.
- Fortier L-C, Sekulovic O. 2013. Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence* 4:354–365. DOI: 10.4161/viru.24498.
- Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28:3150–3152. DOI: 10.1093/bioinformatics/bts565.
- Gamage SD, Strasser JE, Chalk CL, Weiss AA. 2003. Nonpathogenic *Escherichia coli* Can Contribute to the Production of Shiga Toxin. *Infection and Immunity* 71:3107–3115. DOI: 10.1128/IAI.71.6.3107-3115.2003.
- Graystock P, Goulson D, Hughes WOH. 2015. Parasites in bloom: flowers aid dispersal and transmission of pollinator parasites within and between bee species. *Proceedings of the Royal Society B: Biological Sciences* 282:20151371. DOI: 10.1098/rspb.2015.1371.
- Hendrix RW. 2002. Bacteriophages: Evolution of the Majority. *Theoretical Population Biology* 61:471–480. DOI: 10.1006/tpbi.2002.1590.
- Hsu BB, Gibson TE, Yeliseyev V, Liu Q, Lyon L, Bry L, Silver PA, Gerber GK. 2019. Dynamic Modulation of the Gut Microbiota and Metabolome by Bacteriophages in a Mouse Model. *Cell Host & Microbe* 25:803–814.e5. DOI: 10.1016/j.chom.2019.05.001.

- Huang D, Yu P, Ye M, Schwarz C, Jiang X, Alvarez PJJ. 2021. Enhanced mutualistic symbiosis between soil phages and bacteria with elevated chromium-induced environmental stress. *Microbiome* 9:150. DOI: 10.1186/s40168-021-01074-1.
- Khan A, Burmeister AR, Wahl LM. 2020. Evolution along the parasitism-mutualism continuum determines the genetic repertoire of prophages. *PLOS Computational Biology* 16:e1008482. DOI: 10.1371/journal.pcbi.1008482.
- Krueger F. 2012. Trim Galore, available at https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.
- Kwong WK, Moran NA. 2016. Gut microbial communities of social bees. *Nature Reviews Microbiology* 14:374–384. DOI: 10.1038/nrmicro.2016.43.
- Lahti L. 2019. microbiome R package. Available at [http://bioconductor.org/packages/microbiome\(development version\)/](http://bioconductor.org/packages/microbiome(development version)/) (accessed April 17, 2024).
- Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* 26:589–595. DOI: 10.1093/bioinformatics/btp698.
- Li S, Guo R, Zhang Y, Li P, Chen F, Wang X, Li J, Jie Z, Lv Q, Jin H, Wang G, Yan Q. 2022. A catalog of 48,425 nonredundant viruses from oral metagenomes expands the horizon of the human oral virome. *iScience* 25. DOI: 10.1016/j.isci.2022.104418.
- Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31:1674–1676. DOI: 10.1093/bioinformatics/btv033.
- Lu C, Zhang Z, Cai Z, Zhu Z, Qiu Y, Wu A, Jiang T, Zheng H, Peng Y. 2021. Prokaryotic virus host predictor: a Gaussian model for host prediction of prokaryotic viruses in metagenomics. *BMC Biology* 19:5. DOI: 10.1186/s12915-020-00938-6.
- Martínez-Fleites C, Ortíz-Lombardía M, Pons T, Tarbouriech N, Taylor EJ, Arrieta JG, Hernández L, Davies GJ. 2005. Crystal structure of levansucrase from the Gram-negative bacterium *Gluconacetobacter diazotrophicus*. *Biochemical Journal* 390:19–27. DOI: 10.1042/BJ20050324.
- McMurdie PJ, Holmes S. 2013. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLOS ONE* 8:e61217. DOI: 10.1371/journal.pone.0061217.
- Moran NA. 2015. Genomics of the honey bee microbiome. *Current Opinion in Insect Science* 10:22–28. DOI: 10.1016/j.cois.2015.04.003.
- Moreno Switt AI, Orsi RH, den Bakker HC, Vongkamjan K, Altier C, Wiedmann M. 2013. Genomic characterization provides new insight into Salmonella phage diversity. *BMC Genomics* 14:481. DOI: 10.1186/1471-2164-14-481.
- Nanda AM, Thormann K, Frunzke J. 2015. Impact of Spontaneous Prophage Induction on the Fitness of Bacterial Populations and Host-Microbe Interactions. *Journal of Bacteriology* 197:410–419. DOI: 10.1128/JB.02230-14.
- Nikaido H, Pagès J-M. 2012. Broad-specificity efflux pumps and their role in multidrug resistance of Gram-negative bacteria. *FEMS Microbiology Reviews* 36:340–363. DOI: 10.1111/j.1574-6976.2011.00290.x.
- Olm MR, Brown CT, Brooks B, Banfield JF. 2017. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *The ISME Journal* 11:2864–2868. DOI: 10.1038/ismej.2017.126.
- Ortiz-Alvarado Y, Clark DR, Vega-Melendez CJ, Flores-Cruz Z, Domingez-Bello MG, Giray T. 2020. Antibiotics in hives and their effects on honey bee physiology and behavioral development. *Biology Open* 9:bio053884. DOI: 10.1242/bio.053884.

- Paul JH. 2008. Prophages in marine bacteria: dangerous molecular time bombs or the key to survival in the seas? *The ISME Journal* 2:579–589. DOI: 10.1038/ismej.2008.35.
- Pelicic V, Reytrat JM, Gicquel B. 1996. Expression of the *Bacillus subtilis* sacB gene confers sucrose sensitivity on mycobacteria. *Journal of Bacteriology* 178:1197–1199.
- Robinson CRP, Dolezal AG, Newton ILG. 2024. Host species and geography impact bee-associated RNA virus communities with evidence for isolation by distance in viral populations. *ISME Communications* 4:ycad003. DOI: 10.1093/ismeco/ycad003.
- Rodriguez-Brito B, Li L, Wegley L, Furlan M, Angly F, Breitbart M, Buchanan J, Desnues C, Dinsdale E, Edwards R, Felts B, Haynes M, Liu H, Lipson D, Mahaffy J, Belen Martin-Cuadrado A, Mira A, Nulton J, Pasic L, Rayhawk S, Rodriguez-Mueller J, Rodriguez-Valera F, Salamon P, Srinagesh S, Thingstad TF, Tran T, Thurber RV, Willner D, Youle M, Rohwer F. 2010. Viral and microbial community dynamics in four aquatic environments. *Isme Journal* 4:739–751. DOI: 10.1038/ismej.2010.1.
- Roux S, Camargo AP, Coutinho FH, Dabdoub SM, Dutilh BE, Nayfach S, Tritt A. 2023. iPhoP: An integrated machine learning framework to maximize host prediction for metagenome-derived viruses of archaea and bacteria. *PLOS Biology* 21:e3002083. DOI: 10.1371/journal.pbio.3002083.
- Shkoporov AN, Turkington CJ, Hill C. 2022. Mutualistic interplay between bacteriophages and bacteria in the human gut. *Nature Reviews Microbiology* 20:737–749. DOI: 10.1038/s41579-022-00755-4.
- Sivaraman J, Sauvé V, Matte A, Cygler M. 2002. Crystal Structure of Escherichia coli Glucose-1-Phosphate Thymidyltransferase (RffH) Complexed with dTTP and Mg²⁺*. *Journal of Biological Chemistry* 277:44214–44219. DOI: 10.1074/jbc.M206932200.
- Stecher B, Maier L, Hardt W-D. 2013. “Blooming” in the gut: how dysbiosis might contribute to pathogen evolution. *Nature Reviews Microbiology* 11:277–284. DOI: 10.1038/nrmicro2989.
- Taylor MA, Robertson AW, Biggs PJ, Richards KK, Jones DF, Parkar SG. 2019. The effect of carbohydrate sources: Sucrose, invert sugar and components of mānuka honey, on core bacteria in the digestive tract of adult honey bees (*Apis mellifera*). *PLOS ONE* 14:e0225845. DOI: 10.1371/journal.pone.0225845.
- Tobin CA, Hill C, Shkoporov AN. 2023. Factors Affecting Variation of the Human Gut Phageome. *Annual Review of Microbiology* 77:363–379. DOI: 10.1146/annurev-micro-032421-105754.
- Tsao Y-F, Taylor VL, Kala S, Bondy-Denomy J, Khan AN, Bona D, Cattoir V, Lory S, Davidson AR, Maxwell KL. 2018. Phage Morons Play an Important Role in *Pseudomonas aeruginosa* Phenotypes. *Journal of Bacteriology* 200. DOI: 10.1128/JB.00189-18.
- Turkovicova L, Smidak R, Jung G, Turna J, Lubec G, Aradska J. 2016. Proteomic analysis of the TerC interactome: Novel links to tellurite resistance and pathogenicity. *Journal of Proteomics* 136:167–173. DOI: 10.1016/j.jprot.2016.01.003.
- Van Espen L, Bak EG, Beller L, Close L, Deboutte W, Juel HB, Nielsen T, Sinar D, De Coninck L, Frithioff-Bøjsøe C, Fonvig CE, Jacobsen S, Kjærgaard M, Thiele M, Fullam A, Kuhn M, Holm J-C, Bork P, Krag A, Hansen T, Arumugam M, Matthijnssens J. 2021. A Previously Undescribed Highly Prevalent Phage Identified in a Danish Enteric Virome Catalog. *mSystems* 6:10.1128/msystems.00382-21. DOI: 10.1128/msystems.00382-21.
- vanEngelsdorp D, Meixner MD. 2010. A historical review of managed honey bee populations in Europe and the United States and the factors that may affect them. *Journal of Invertebrate Pathology* 103:S80–S95. DOI: 10.1016/j.jip.2009.06.011.
- Wallberg A, Bunikis I, Pettersson OV, Mosbech M-B, Childers AK, Evans JD, Mikheyev AS, Robertson HM, Robinson GE, Webster MT. 2019. A hybrid de novo genome assembly of the honeybee, *Apis mellifera*, with chromosome-length scaffolds. *BMC genomics* 20:275. DOI: 10.1186/s12864-019-5642-0.

- Wendling CC. 2023. Prophage mediated control of higher order interactions - Insights from multi-level approaches. *Current Opinion in Systems Biology* 35:100469. DOI: 10.1016/j.coisb.2023.100469.
- Xia R, Sun M, Balcázar JL, Yu P, Hu F, Alvarez PJJ. 2023. Benzo[a]pyrene stress impacts adaptive strategies and ecological functions of earthworm intestinal viromes. *The ISME Journal* 17:1004–1014. DOI: 10.1038/s41396-023-01408-x.
- Yeats C, Finn RD, Bateman A. 2002. The PASTA domain: a β -lactam-binding domain. *Trends in Biochemical Sciences* 27:438–440. DOI: 10.1016/S0968-0004(02)02164-3.
- Zhang Z, Mu X, Cao Q, Shi Y, Hu X, Zheng H. 2022. Honeybee gut *Lactobacillus* modulates host learning and memory behaviors via regulating tryptophan metabolism. *Nature Communications* 13:2037. DOI: 10.1038/s41467-022-29760-0.
- Zheng H-Q, Chen YP. 2014. Detection of *Spiroplasma melliferum* in honey bee colonies in the US. *Journal of Invertebrate Pathology* 119:47–49. DOI: 10.1016/j.jip.2014.03.006.
- Zuo T, Lu X-J, Zhang Y, Cheung CP, Lam S, Zhang F, Tang W, Ching JYL, Zhao R, Chan PKS, Sung JJY, Yu J, Chan FKL, Cao Q, Sheng J-Q, Ng SC. 2019. Gut mucosal virome alterations in ulcerative colitis. *Gut* 68:1169–1179. DOI: 10.1136/gutjnl-2018-318131.

Chapter Five: Synthesis

My dissertation identifies the potential importance of viral genetic contributions from bacteriophage in two host-mediated bacterial communities: the honey bee gut and the amphibian skin microbiome. While it is well known that phages contribute beneficial genes to pathogens (Brüssow, Canchaya & Hardt, 2004; Fortier & Sekulovic, 2013), this mechanism has been understudied in the commensal bacteria of microbiomes (Baugher, Durmaz & Klaenhammer, 2014; Oh et al., 2019). My research has shown that, similar to environmental phages, host-associated phages also carry auxiliary metabolic genes and other potentially beneficial genes that may contribute to bacterial host fitness.

Chapter two of my dissertation investigates the presence and contributions of prophages in bacteria from the honey bee gut microbiome. This chapter confirms that within the honey bee gut system, prophages are present within the bacterial community and are more common in some bacterial hosts than others, something which has been reported broadly in a variety of other environments (Inglis, Roach & Edwards, 2024). I found that that some bacterial taxa, such as the genus *Gilliamella*, may be more strongly impacted by the genetic contributions of prophages. The types of fitness impacts of prophages may additionally vary based on bacterial host, some bacterial hosts, like *Gilliamella* spp., carried multiple prophage-encoded carbohydrate-associated metabolism genes, while other bacterial hosts, like *Snodgrassella alvi*, primarily had prophage-encoded genes predicted to protect from additional phage infection. This was not true of every bacterial host, however, as many taxa possessed prophages which lacked obvious beneficial genes, or instead had what seemed to be a random assortment of putative beneficial genes, rather than those sharing similar functions.

Chapter three, which focuses on the prophages of bacteria isolated from the amphibian skin microbiome, furthers some of the broader conclusions drawn from chapter two. In particular, it identifies the presence of a chitinase-like prophage-encoded gene in many different prophages of the bacterial host *Janthinobacterium lividum*. Based on mutation rates, this gene appears to undergo purifying selection by *J. lividum*. This suggests the gene is potentially useful to the bacteria. The chitinase-like gene's homology to known proteins suggests it may assist in either the breakdown of fungal cell walls, releasing carbon in addition to potentially being antifungal (Lima, Colombo & de Almeida Junior, 2019; Abulikemu et al., 2021; Kumar et al., 2022), or may alternatively, be a phage endolysin repurposed as a bacterial toxin (Park et al., 2014; Takala et al., 2023). Other bacterial taxa associated with the amphibian skin microbiome do not have any highly prevalent prophage-encoded genes. This suggests that *J. lividum*'s relationship with its prophages could be relatively unique in the system. Interestingly, while only a small amount of work has been done in skin viromes of any host-associated systems, one study of the human virome does suggest that phages may be highly transient and only those with valuable genes are fixed in the community (Graham et al., 2023). While the chitinase-like gene is also found in environmental *J. lividum* prophages, it may be those environments are sufficiently transient as well.

The final chapter of my dissertation returns to the honey bee gut system, but transitions away from prophages and instead explores the diversity of lytic and unintegrated temperate phages. This study is the first to show in the honey bee gut system that increased geographic distance between sites is associated with decreased similarity of phage communities. It also shows that while broadly, phages carry auxiliary metabolic genes associated with carbohydrate metabolism, the specific genes may be unique to sites and nearby neighbors. The diversity and prevalence of carbohydrate metabolism genes suggests that whether it is through shorter periods of lytic infection or potentially long-term lysogenic conversation, phages capable of altering host metabolic pathways will often target carbohydrate pathways. Given the diet of bees is heavily focused on sugar, it is logical that phages may carry carbohydrate auxiliary metabolic genes (Dolezal & Toth, 2018). Interestingly, I also found, that compared to the genes identified in the honey bee prophages, there was a higher prevalence of antimicrobial genes in the virome. Increased antimicrobial associated genes in the honey bee gut virome may suggest that while prophages of the honey bee gut system (chapter two) predominantly encode beneficial genes associated with carbohydrate metabolism and superinfection exclusion, phage lifestyles that more frequently lyse hosts (ex: lytic phages) may still harbor some genes which benefit the bacterial host. Additionally, phages in the viral community outside of bacterial cells may be recently induced prophages or temperate phages capable of shifting symbiotic strategies. Therefore, some phages in the virome may harbor beneficial genes useful in alternate lifecycles like prophage integration. Other viruses, like lytic phages, may possess antimicrobial resistance genes (ARGs) coincidentally, which are acquired during error-prone replication and repackaging in bacterial cells (Colomer-Lluch, Jofre & Muniesa, 2011; Touchon, Moura de Sousa & Rocha, 2017). Lytic phages may also carry auxiliary metabolic genes which are used to bacterial metabolic functions during phage takeover (Breitbart et al., 2018). The presence of these genes in phage which frequently lyse their bacterial hosts may still have beneficial effects on the bacterial hosts, as failed or incomplete lytic infection can result in the acquisition of new genes without bacterial death (Stecher, Maier & Hardt, 2013).

All together, these studies highlight the importance of understanding all portions of the viral component of microbiomes. Chapters two and three present evidence that prophages may be influencing bacterial species within the host-associated communities of the honey bee gut and amphibian skin microbiome through beneficial genes. Previous research has primarily focused on the role of phage-encoded virulence genes, such as toxins that cause disease in animal hosts (Waldor & Mekalanos, 1996; Brüssow, Canchaya & Hardt, 2004; Fortier & Sekulovic, 2013; Ebeling, Fünfhaus & Genersch, 2021). These studies provide further evidence to a burgeoning literature (Bondy-Denomy & Davidson, 2014; Oh et al., 2019; Wendling, Refardt & Hall, 2021) suggesting that lysogenic conversion may enhance the fitness of commensal bacteria in host-associated systems and could have downstream impacts on the health of the animal host. These auxiliary metabolic genes and other potentially beneficial genes may contribute to bacterial strain diversity via lysogenic conversion (De Paepe et al., 2014; Baugher, Durmaz & Klaenhammer, 2014; Zhang et al., 2020).

Surveys of already sequenced bacterial isolates (chapters two and three) may be a useful and cost-effective way to gain initial insight into the viral interactions of a host-associated community, prior to more costly metagenomic sequencing (chapter four). However, the complementary, but not identical, results of chapters two and four suggests that neither method (prophage identification nor virome sequencing) provides a complete picture of viral interactions within a host-associated community. For example, both chapters two and four identify the presence of carbohydrate-associated auxiliary-metabolic genes, but chapter four additionally identifies the presence of antimicrobial genes. Chapter two also shows that prophages, either intact or degraded, are predominantly infecting certain bacterial taxa, such as *Gilliamella* spp. and *S. alvi*, while chapter four shows that free virions are predicted to primarily target *Lactobacillus* spp. and *Bifidobacterium* spp. This suggests that bacterial hosts may be predominantly infected by phages with similar life cycle proclivities, or alternatively, that current phage host identification methods are unable to successfully identify phages of *Gilliamella* spp., and *S. alvi*. Additionally, metagenomic studies of viromes, such as the one conducted in chapter four, provide more direct answers to some questions, such as how distance or other experimental factors may affect viral communities, relative to prophage surveys. Functional analysis of viromes may also indicate the ways in which viruses undergoing active lytic life cycles may reshape bacterial metabolic functions during phage takeover or alternatively, the types of genes which are being passed around during phage-mediated horizontal gene transfer (Breitbart et al., 2018). Therefore, both methods (prophage identification and virome sequencing) will be required to tease out the interactions between bacteria and phages in host-associated systems. The validation of bioinformatic predictions using *in vivo* and *in vitro* methods (Park et al., 2014; Oh et al., 2019; Orlando, 2020; Wendling, Refardt & Hall, 2021; Xia et al., 2023) will be critical next steps in advancing the field of phage ecology, both broadly and within host-associated systems.

Ultimately, this dissertation provides evidence that bacteriophages in host-associated microbiomes, such as the honey bee gut and the amphibian skin microbiome, contribute beneficial genes to the bacterial community, which in turn may shape the broader host-associated community. Additionally, it demonstrates that the distribution of prophages in a host-associated system are not necessarily well-reflected by viral metagenomic studies, with each study providing valuable unique information. Going forward, efforts to unite our understanding of both integrated bacteriophage and actively infectious and replicating phage in host-associated communities will be key to further unravel the nuanced interactions between bacteriophage and their hosts within the animal microbiome.

References

- Abulikemu S, Yesilyurt A, Gencer D, Usta M, Nalcacioglu R. 2021. Comparison of the potential activities of viral and bacterial chitinases. *Egyptian Journal of Biological Pest Control* 31:91. DOI: 10.1186/s41938-021-00435-0.
- Baugher JL, Durmaz E, Klaenhammer TR. 2014. Spontaneously Induced Prophages in *Lactobacillus gasseri* Contribute to Horizontal Gene Transfer. *Applied and Environmental Microbiology* 80:3508–3517. DOI: 10.1128/AEM.04092-13.
- Bondy-Denomy J, Davidson AR. 2014. When a virus is not a parasite: the beneficial effects of prophages on bacterial fitness. *Journal of Microbiology* 52:235–242. DOI: 10.1007/s12275-014-4083-3.
- Breitbart M, Bonnain C, Malki K, Sawaya NA. 2018. Phage puppet masters of the marine microbial realm. *Nature Microbiology* 3:754–766. DOI: 10.1038/s41564-018-0166-y.
- Brüssow H, Canchaya C, Hardt W-D. 2004. Phages and the Evolution of Bacterial Pathogens: from Genomic Rearrangements to Lysogenic Conversion. *Microbiology and Molecular Biology Reviews* 68:560–602. DOI: 10.1128/MMBR.68.3.560-602.2004.
- Colomer-Lluch M, Jofre J, Muniesa M. 2011. Antibiotic Resistance Genes in the Bacteriophage DNA Fraction of Environmental Samples. *PLOS ONE* 6:e17549. DOI: 10.1371/journal.pone.0017549.
- De Paepe M, Leclerc M, Tinsley CR, Petit M-A. 2014. Bacteriophages: an underestimated role in human and animal health? *Frontiers in Cellular and Infection Microbiology* 4. DOI: 10.3389/fcimb.2014.00039.
- Dolezal AG, Toth AL. 2018. Feedbacks between nutrition and disease in honey bee health. *Current Opinion in Insect Science* 26:114–119. DOI: 10.1016/j.cois.2018.02.006.
- Ebeling J, Fünfhaus A, Genersch E. 2021. The Buzz about ADP-Ribosylation Toxins from *Paenibacillus* larvae, the Causative Agent of American Foulbrood in Honey Bees. *Toxins* 13:151. DOI: 10.3390/toxins13020151.
- Fortier L-C, Sekulovic O. 2013. Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence* 4:354–365. DOI: 10.4161/viru.24498.
- Graham EH, Tom WA, Neujahr AC, Adamowicz MS, Clarke JL, Herr JR, Fernando SC. 2023. The persistence and stabilization of auxiliary genes in the human skin virome. *VIROLOGY JOURNAL* 20:49. DOI: 10.1186/s12985-023-02012-3.
- Inglis LK, Roach MJ, Edwards RA. 2024. Prophages: an integral but understudied component of the human microbiome. *Microbial Genomics* 10:001166. DOI: 10.1099/mgen.0.001166.
- Kumar M, Chakdar H, Pandiyan K, Thapa S, Shahid M, Singh A, Srivastava AK, Saxena AK. 2022. Bacterial chitinases: genetics, engineering and applications. *World Journal of Microbiology and Biotechnology* 38:252. DOI: 10.1007/s11274-022-03444-9.
- Lima SL, Colombo AL, de Almeida Junior JN. 2019. Fungal Cell Wall: Emerging Antifungals and Drug Resistance. *Frontiers in Microbiology* 10.
- Oh J-H, Lin XB, Zhang S, Tollenaar SL, Özçam M, Dunphy C, Walter J, van Pijkeren J-P. 2019. Prophages in *Lactobacillus reuteri* Are Associated with Fitness Trade-Offs but Can Increase Competitiveness in the Gut Ecosystem. *Applied and Environmental Microbiology* 86:e01922-19. DOI: 10.1128/AEM.01922-19.
- Orlando M. 2020. Biochemical and biophysical analysis of two Antarctic lysozyme endolysins and in silico exploration of glycoside hydrolase 19 sequence space. *Università degli Studi di Milano-Bicocca*.
- Park Y, Lim J-A, Kong M, Ryu S, Rhee S. 2014. Structure of bacteriophage SPN1S endolysin reveals an unusual two-module fold for the peptidoglycan lytic and binding activity. *Molecular Microbiology* 92:316–325. DOI: 10.1111/mmi.12555.
- Stecher B, Maier L, Hardt W-D. 2013. “Blooming” in the gut: how dysbiosis might contribute to pathogen evolution. *Nature Reviews Microbiology* 11:277–284. DOI: 10.1038/nrmicro2989.

- Takala TM, Mokhtari S, Ahonen SL, Wan X, Saris PEJ. 2023. Wild-type *Lactococcus lactis* producing bacteriocin-like prophage lysins. *Frontiers in Microbiology* 14.
- Touchon M, Moura de Sousa JA, Rocha EP. 2017. Embracing the enemy: the diversification of microbial gene repertoires by phage-mediated horizontal gene transfer. *Current Opinion in Microbiology* 38:66–73. DOI: 10.1016/j.mib.2017.04.010.
- Waldor MK, Mekalanos JJ. 1996. Lysogenic Conversion by a Filamentous Phage Encoding Cholera Toxin. *Science* 272:1910–1914. DOI: 10.1126/science.272.5270.1910.
- Wendling CC, Refardt D, Hall AR. 2021. Fitness benefits to bacteria of carrying prophages and prophage-encoded antibiotic-resistance genes peak in different environments. *Evolution* 75:515–528. DOI: 10.1111/evo.14153.
- Xia R, Sun M, Balcázar JL, Yu P, Hu F, Alvarez PJJ. 2023. Benzo[a]pyrene stress impacts adaptive strategies and ecological functions of earthworm intestinal viromes. *The ISME Journal* 17:1004–1014. DOI: 10.1038/s41396-023-01408-x.
- Zhang Y, Liao Y-T, Salvador A, Sun X, Wu VCH. 2020. Prediction, Diversity, and Genomic Analysis of Temperate Phages Induced From Shiga Toxin-Producing *Escherichia coli* Strains. *Frontiers in Microbiology* 10. DOI: 10.3389/fmicb.2019.03093.