DISSERTATION

# LINKING MOSQUITO MIDGUT AND VIRUS POPULATION BIOLOGY AT THE MOLECULAR AND CELLULAR LEVEL

Submitted by

Emily Anne Fitzmeyer

Department of Microbiology, Immunology, and Pathology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2024

Doctoral Committee:

Advisor: Gregory D. Ebel

Rebekah Kading
Mark Stenglein
Brooke Anderson

ABSTRACT

LINKING MOSQUITO MIDGUT AND VIRUS POPULATION BIOLOGY AT THE CELLULAR AND MOLECULAR LEVEL

Vector competence (VC) refers to the efficiency of pathogen transmission by vectors. Each step in infection of a mosquito vector constitutes a barrier to transmission that may impose bottlenecks on virus populations. West Nile virus (WNV) is maintained by multiple mosquito species with varying VC. However, the extent that bottlenecks and VC are linked is poorly understood. Similarly, quantitative analyses of mosquito-imposed bottlenecks on virus populations are limited.

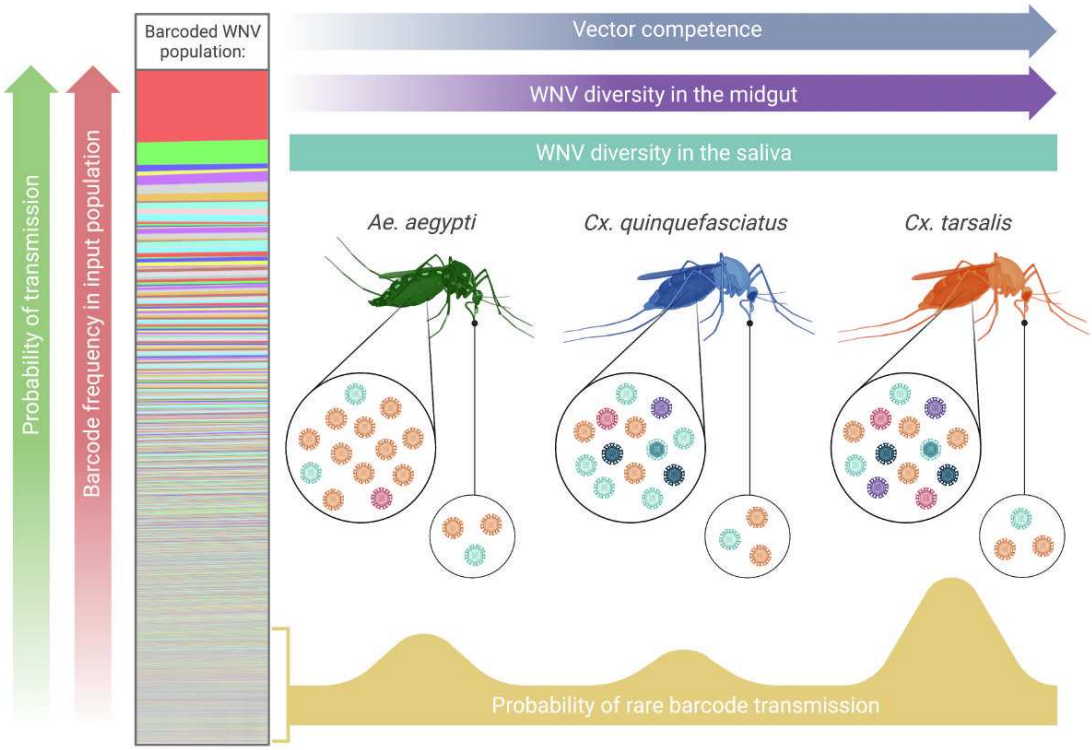



**Figure 1. Graphical Abstract.** Overview of chapter 2 findings.

We used molecularly barcoded WNV to quantify tissue-associated population bottlenecks in three variably competent WNV vectors. Our results confirm strong population bottlenecks during mosquito infection that are capable of dramatically reshaping virus population structure in a non-selective manner. In addition, we found that mosquitoes with differing VC uniquely shape WNV population structure: highly competent vectors are more likely to contribute to the maintenance of rare viral genotypes. These findings have important implications for arbovirus emergence and evolution.

The mosquito midgut functions as a key interface between virus and vector. However, studies of midgut physiology and associated virus infection dynamics are scarce, and in *Culex tarsalis* - the primary vector of West Nile virus (WNV) in the contiguous United States - nonexistent. We performed single-cell RNA sequencing on dissociated, WNV-infected *Cx. tarsalis* midguts. We identified populations of distinct midgut cell-types consistent with existing descriptions of insect midgut physiology and found that all midgut cell populations were permissive to WNV infection. However, we observed high levels of viral RNA suggesting enhanced replication in enteroendocrine cells and cells enriched for mitochondrial genes. In addition, we found no significant upregulation of mosquito immune genes associated with WNV infection at the whole-midgut level, rather, a significant positive correlation between immune gene expression and WNV viral RNA load at the individual cell level. These findings illuminate the midgut infection dynamics of WNV, providing insight into cell-type specific enhancement of, and immune response to, WNV infection in a primary vector.

ACKNOWLEDGEMENTS

I would first like to thank my father. He successfully imparted to me two things; a curiosity about how things work, and a deep-seated belief that if something wasn't working, I could fix it. These two tenets of my personality have been invaluable to me throughout my life and career. I miss him immensely.

As much as I owe to my father, I owe more to the women in my life who, both professionally and personally, shaped me and guided me on this path. Professionally – Dr. Elizabeth Greguske took a chance on me, introduced me to research, and in doing so changed the course of my life. Dr. Sonja Best and Dr. Rebecca Broekel trained me, fostered my passion for the work, and allowed me the level of independence required to grow as a young scientist. Personally – I am who I am because I stood behind a wall of women. My mother, sister, and aunts loved me, encouraged me, and uplifted me.

I want to thank all Ebel lab members, past and present, for being phenomenal colleagues and inspiring scientists. They say that during your doctoral work your environment matters more than the subject of your research; that when everything is falling apart it's the people around you that keep you sane and motivated. This was certainly true for me, and I am extremely grateful to have worked with the people I did.

I want to thank my committee members; Dr. Rebekah Kading, Dr. Brooke Anderson, and Dr. Mark Stenglein for their guidance and support. I would also, of course, like to thank my advisor, Dr. Gregory Ebel. Greg is the kind of mentor who remembers what it feels like to suffer in graduate school and seeks to break the cycle by treating his students with compassion. I am grateful to him for his patience and humanity.

I want to give a special shoutout to Dr. Emily Gallichotte. Emily's passion for everything she does makes her an incredible mentor, scientist, and friend. Her work ethic continually astounds me, and her guidance and assistance were invaluable to the work herein. I consider myself privileged to have worked with her.

Much of the work described in this dissertation is published or in preparation for publication and, as such, I would like to acknowledge co-authors James Weger-Lucarelli, Marylee Kapuscinski, Zaid Abdo, Kyra Pyron, Michael Young, Taru Dutt, Silvain Pinaud, Barb Graham, Jessica Hill, Sarah Helene Merkling, Hunter Ogg, Corey Campbell, Marcela Henao-Tamayo – and all others who facilitated this work – Nicole Sexton, Reyes Murrieta, Matthew Aliota, Kasen Riemersma, and Vineet Menachery.

I would like to thank the homies. The Montana homies: Arielle, Kris, Ashley, Josh, and Michael. Y'all are so special to me. My pursuit of this career is what brought you all into my life, and the time spent sitting around campfires, running around in the mountains, and celebrating life with y'all is the greatest gift this pursuit has given me. The FoCo homies: Darby, Laura, Bekah, James, Ariel, Dominique, JP, and Addie. The hometown crew is the source of my sanity. It's sharing the weekly rants and rambles with y'all that made grad school just a little more tolerable.

I want to thank my therapist, Emma. I truly doubt that I would have been capable of achieving this goal without the emotional work and healing that she fostered.

I would like to thank Chloe, Beebs, The Beeb, Beeby, Clo Clo, Cloobie, Cloobie Doobie Doo, Muffin, The Clo-inator, The Cloo, Sometimes All I Think About is Cloo, Beebit, Cleebit, Beebaroni, Blueberry Loaf, The Bubba, Bebe, EhBepBepBepBep, My Sweet Pea, My Sweet Girl, Little Honey, My Little Old Lady, Poop Face, Poop Face Poop Butt, Bib. I started graduate school anticipating that she wouldn't live to see the end of this 5-year endeavor and,

subsequently, that at some point during this process I would get completely derailed by the loss of my beloved cat. Fortunately, Beebs celebrated her 21st birthday this February and is with us still. Thanks for hanging out, Bib.

I would like to thank Portugal. The Man. for putting on some soul-altering shows at Red Rocks.

I would like to thank Hot Wok for making the best Chinese food in town. Gone too soon, but never forgotten; sweet and sour chicken, hot and sour soup, and a large side of white rice brought directly to my door every time I had a hangover.

Finally, not to get all Matthew McConaughey on you, but I would like to thank every version of myself. Every time my morale and motivation dropped due to stress and mental exhaustion, I was able to find more energy, more motivation, more passion, more joy, and more resilience within myself. That would not have been possible without every layer of who I am and who I've been. I carry them with me, and I did this for them. I did this because I love them.

An unfortunate addendum – Beebs lived until March 31, 2024, twelve days after the successful defense of my thesis. She had been my companion and friend for 20 years and it almost seemed that she waited to see me through before saying goodbye. Caring for her was the greatest privilege of my life. She has my whole heart.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

*The confusion is saying*
*"I don't know"*
*But the minute you are quiet*
*You find out that in truth you do know*
*For in you, you know*
*Plane after plane will open to you*
*I want to know who I really am*

*As if in each of us*
*There once was a fire*
*And for some of us*
*There seem as if there are only ashes now*
*But when we dig in the ashes*
*We find one ember*

*And very gently we fan that ember*
*Blow on it, it gets brighter*
*And from that ember we rebuild the fire*
*Only thing that's important is that ember*

*That's what you and I are here to celebrate*
*That though we've lived our life totally involved in the world*
*We know*
*We know that we're of the spirit*

*The ember gets stronger*
*Flame starts to flicker a bit*
*And pretty soon you realize that all we're going to do for eternity*
*Is sit around the fire*

- *Sit Around the Fire: Jon Hopkins / Trevor H Oswalt / Richard Alpert*

# CHAPTER 1: INTRODUCTION

## 1.1 The expansion of West Nile virus throughout North America, and the resulting health burden in the United States

Virus evolution and vector physiology play direct roles in the emergence and spread of new virus variants. The emergence and re-emergence of arthropod-borne viruses like West Nile virus (WNV) and others, exemplifies this. WNV was introduced to North America in 1999 and rapidly spread throughout North and Central America (1,2). In 2002 a WNV variant (WN02), which possessed three coding changes distinguishing it from the historical NY99 strain, demonstrated enhanced replication efficiency resulting in higher virus titers in mosquitoes earlier in infection (1–4). WN02 caused higher proportions of infection and transmission in *Culex tarsalis* and *Culex pipiens* mosquitoes (the primary vectors of WNV in the western and northeastern/central United States respectively) as well as earlier transmission in both species (3,5). It is hypothesized that these factors contributed to the displacement of NY99 by WN02, which has become the dominant strain in much of N. America (3). Since WNV became endemic to the United States it has caused over 56,000 reported human cases – roughly half of which were diagnosed as West Nile neuroinvasive disease (WNND) – and ~2,700 deaths (6,7). It is projected that 80% of human WNV cases are asymptomatic, and that fewer than 1% of cases have neurological manifestations (8). The high number of reported WNND cases therefore indicates that greater than 3 million persons in the United States may have been infected with WNV. Serological studies support these estimates (9).

## 1.2 Competent Vectors of WNV

Vector competence (VC) refers to the efficiency of pathogen infection and transmission of/by a vector (10–12). WNV is primarily transmitted by mosquitoes of the genus *Culex* – several species of which have been identified as highly competent WNV vectors throughout the

distribution of the virus (13). In North America *Cx. tarsalis*, *Cx. quinquefasciatus,* and *Cx. pipiens* have been identified as competent WNV vectors (1,14,15). However, vector competence is not the only determinant of a vector's ability to transmit a virus. The population density of a given vector species, the frequency of that species' contact with infected hosts, their longevity and the extrinsic incubation period (EIP) of the pathogen together determine vectorial capacity, i.e., the transmission potential of a virus:vector pairing (16,17). A vector can have relatively low competence for a virus and still play a role in transmission due to high vectorial capacity (18). Further, WNV has a historical propensity to rapidly adapt to new vectors. For these reasons, the study of WNV evolution should not be limited to highly competent vectors, and include species that have transmission potential for the virus due to their vectorial capacity (15).

### 1.3 WNV evolution and the enzootic transmission cycle

The rapid spread and fixation of the WN02 strain in nature highlights the biological importance of virus evolution with respect to virus:vector relationships. The current evolutionary rate of WNV is approximately $3.6e^{-4}$ – $8.2e^{-3}$ substitutions/site/year – a rate which is largely attributed to the RNA dependent RNA polymerase (RdRp) which lacks proofreading mechanisms and introduces replication-associated changes to the virus genome (19). This error prone mechanism of replication facilitates the maintenance of WNV in nature as genetically and phenotypically complex mutant swarms – or quasispecies (20–22). The adaptability conferred by these highly diverse swarms is pivotal to viral success in a replication cycle that alters between vertebrate and invertebrate hosts (**Figure 1.1**) (12,20,23). However, the observed evolutionary rate of WNV falls short of the theoretical evolutionary rate that could be achieved through unchecked RdRp error introduction. This indicates that other forces lend genetic stability to WNV and play a role in WNV evolution (19,24).

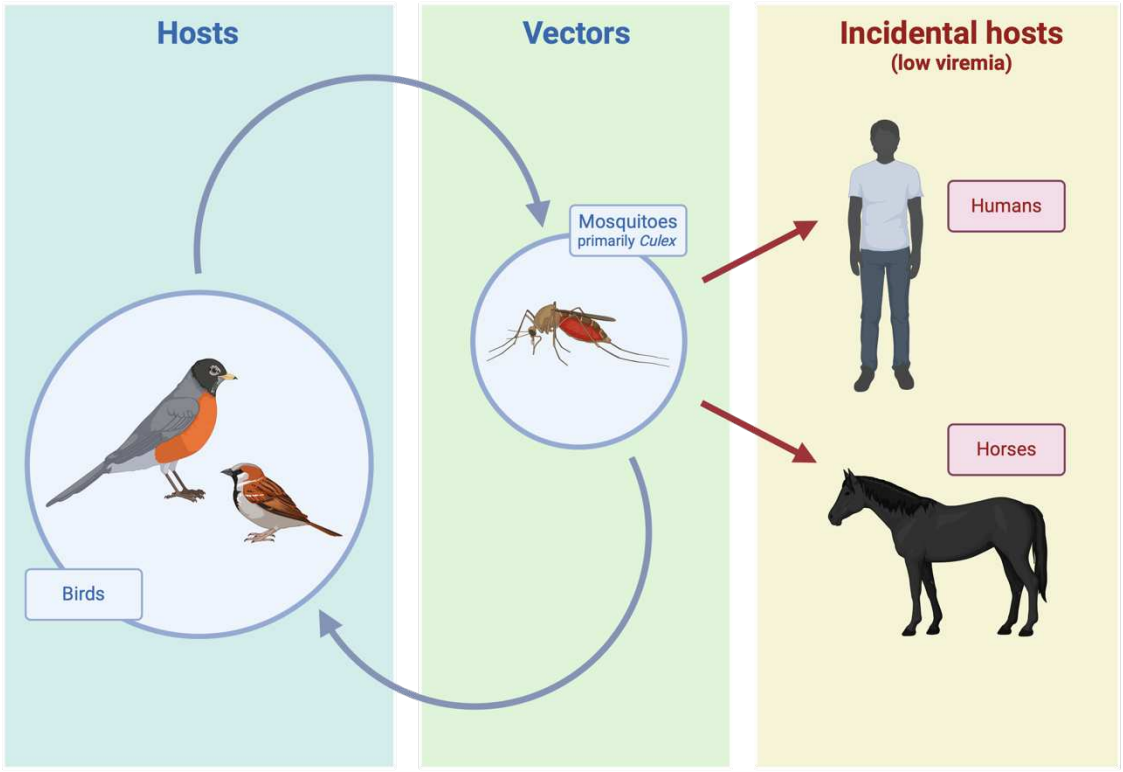



**Figure 1.1. The enzootic transmission cycle of West Nile virus.** WNV is maintained in nature in a transmission cycle between mosquitoes (primarily of the genus *Culex*) and birds. Humans and horses are incidental 'dead end' hosts; WNV infection in these hosts does not cause sufficient viremia to infect mosquitoes. (Created with biorender.com)

WNV persists in a replication cycle that alternates between vertebrate and invertebrate hosts (**Figure 1.1**) (25). During the transmission cycle of WNV in nature, mosquitoes feed on viremic birds – including American Crows, American robins, and house sparrows – become infected, and subsequently transmit WNV to naïve birds and mammals (**Figure 1.1**) (16,26–28). While the adaptability conferred by highly diverse WNV populations is thought to be pivotal to viral success in this replication cycle, previous work has demonstrated that avian infection imposes strong purifying selection on WNV populations (4,20,23,24,29–31). This purifying selection contributes to the overall genetic stability of WNV (24,29–31).

Interestingly, mosquito infection is the evolutionary antithesis to avian infection; mosquitoes serve as sites of virus diversification for WNV (4,24,30,31). Indeed, WNV populations derived from mosquito samples contain approximately double the genetic diversity of WNV populations derived from avian samples, and contain significantly more unique variants and single-nucleotide polymorphisms (SNPs) (4,30). WNV populations that have been experimentally passed in alternating mosquito and avian hosts, mimicking the natural transmission cycle, retain the high genetic diversity seen in mosquito passed populations, and have $d_N/d_S$ ratios indicative of purifying selection (31). This suggests that alternating pressures applied to WNV during this transmission cycle drive WNV evolution (24,30–32).

### 1.4 The impact of bottlenecks on virus populations during systemic mosquito infection

Virus evolution is a broad phrase that encompasses a variety of evolutionary mechanisms and, while the importance of selective pressure is known, the impact of stochastic mechanisms of evolution cannot be ignored. Every stage of the enzootic transmission cycle of arthropod-borne viruses like WNV involves stochastic reductions in virus population diversity, termed bottlenecks (**Figure 1.2**). However, these bottlenecks are brought into sharp relief during mosquito infection. During mosquito infection a virus population must infect and escape from the midgut, infect the salivary glands, and escape/shed into the saliva for transmission to occur (16,20,25,32,33). Previous studies have demonstrated that stringent population bottlenecks are associated with each of these infection and escape barriers (10,14,20,25,34–37).

One of the earliest experimental demonstrations of the population bottlenecks imposed on virus populations during mosquito infection was performed in WNV-infected *Cx. pipiens* (14). This work revealed that WNV population diversity was broadly reduced during infection of and transmission by the vector, but that all variants within the initial population had a non-zero chance of surviving intrahost bottlenecks (14). These foundational findings are further supported by the work described in this dissertation (10). Shortly thereafter a study confirmed that

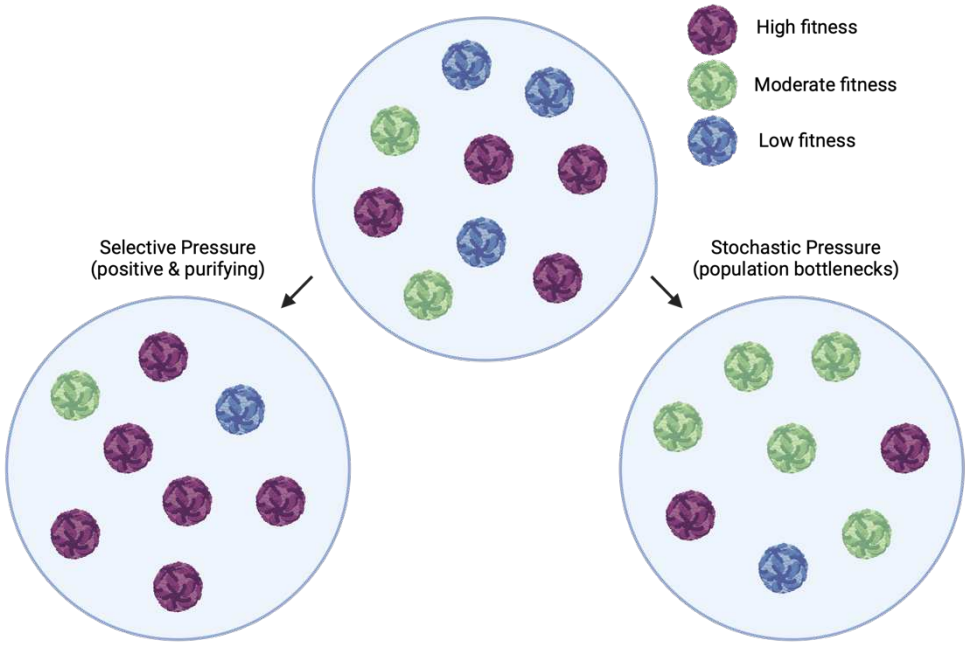



**Figure 1.2. Graphical representation of how selective and stochastic pressures shape virus populations.** (Created with biorender.com)

Venezuelan equine encephalitis virus (VEEV) is similarly impacted by intrahost bottlenecks in its primary vector *Cx. taeniopus* (35). This study also demonstrated that initial virus titer in the infectious bloodmeal is correlated with efficient midgut infection and escape and thus impacts the stringency of those bottlenecks (35). Further work with VEEV in *Cx. taeniopus* revealed that an expansion of virus population diversity occurs following each intrahost bottleneck, and that certain mutations in individual variants appear to confer an increased probability of surviving bottlenecks (36). Bottlenecks followed by intra-tissue population expansion were also examined in variably competent vectors of WNV (15). Investigators compared WNV diversification in *Cx. tarsalis*, *Cx. pipiens*, *Cx. quinquefasciatus*, and *Ae. aegypti* mosquitoes and found that, while population bottlenecks occurred in all species, virus diversification was increased in *Cx. tarsalis* and *Cx. quinquefasciatus* (15). In conjunction with the population expansion that proceeds a bottleneck event, the reduction of diversity imposed by a bottleneck plays a crucial role in

directing the evolution of a virus population. Indeed, it has been hypothesized that, in individual hosts, stochastic forces play a more dominant role in virus evolution than selective pressures (32,38).

This growing body of literature demonstrates that bottlenecks alter the frequency of virus variants in populations, influence the evolution of virus populations, and strongly influence virus evolution at the individual-host level (14,32,38). Ultimately the biologically relevant variant within a WNV population is one that successfully overcomes the infection and escape barriers associated with the midgut and salivary glands and enters the saliva where it has a chance to establish infection in new hosts. In this way bottlenecks shape transmissible virus populations and play a role in the emergence of new or rare virus variants. Thus, quantitatively measuring population bottlenecks during virus infection and transmission, to evaluate the impact on virus population diversity and composition, is of great importance (32). This is particularly important for arboviruses such as WNV, which are transmitted by vectors that have been shown to impose significant bottlenecks on arbovirus populations during infection and transmission (14,15). Thus, one of the primary goals of this dissertation was to quantify the impact of infection and escape barrier associated bottlenecks on WNV populations throughout mosquito infection. To do this we used a barcoded West Nile virus, which allowed us to quantify population richness and complexity in a variety of infected mosquito tissues and secretions at unprecedented resolution.

### 1.5 The barcoded virus method

Due to error-prone replication, RNA viruses such as Zika virus (ZIKV), WNV, influenza A virus (IAV), and simian and human immunodeficiency viruses (SIV and HIV respectively) exist in nature as genetically and phenotypically complex mutant swarms (14,21,39). The ability of RNA viruses to be maintained in nature as a mutant swarm is thought to promote adaptive plasticity and facilitate the evolution and emergence of these viruses (21).

First published in PLOS Pathogens, April 2023

Investigating swarm dynamics during infection, transmission, and treatment is therefore of great significance. Studying intrahost virus population dynamics typically requires identifying intrahost single nucleotide variants (iSNVs) using various approaches to whole-genome sequencing (15). While these approaches are well suited to exploring virus diversification and measuring how natural selection shapes the virus genome, they are not well suited to quantitatively assess reductions in virus diversity. Barcoded viruses are a rapidly expanding technology that allows researchers to quantitatively characterize aspects of virus population dynamics with greater sensitivity and resolution than can be achieved with computational haplotype reconstruction (40–42).

A barcoded virus is engineered to contain a sequence motif (i.e., a barcode) that can be used to distinguish one otherwise identical virus genome from another (**Figure 1.3**).

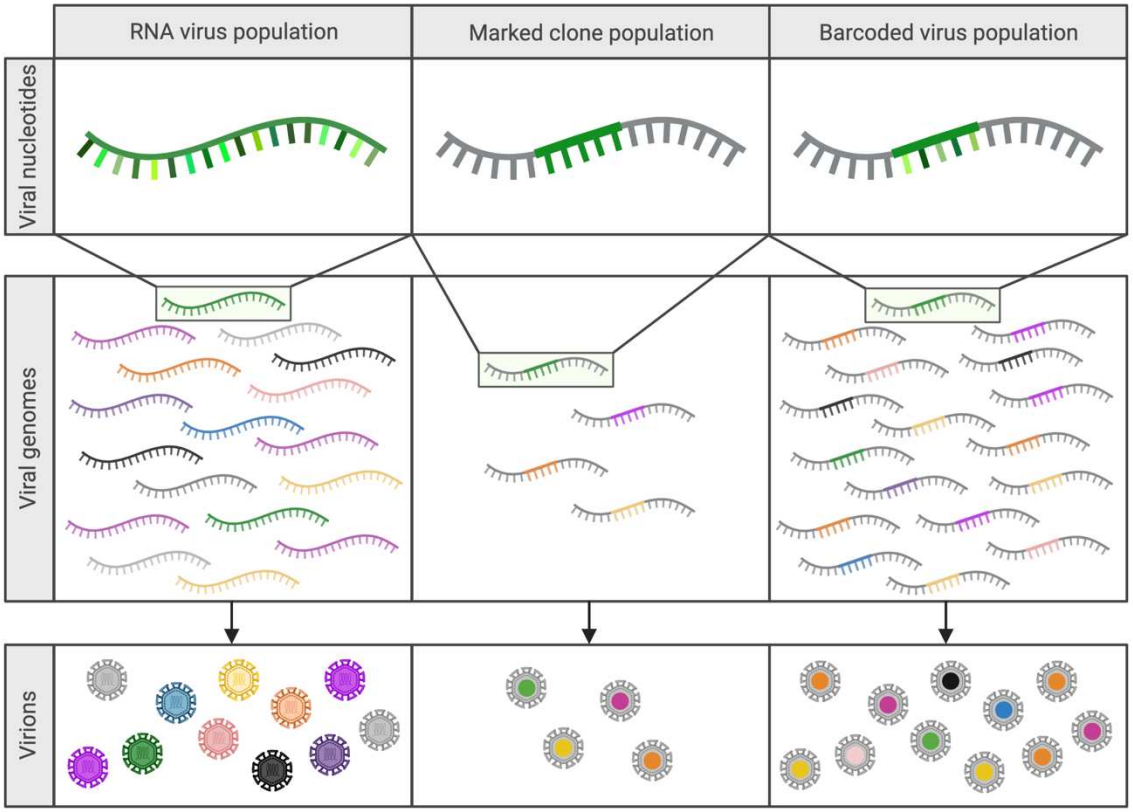


**Figure 1.3. Barcoded viruses are highly effective mimics of RNA virus mutant swarms**. Marked clones and barcoded viruses allow investigators to study changes to population diversity in a variety of systems without the variable fitness observed between viruses in natural RNA virus swarms. (Created with Biorender.com)

While technical approaches to generating barcoded virus populations vary by virus and study, barcoded viruses generally contain a series of coding-neutral alterations to the genome (37,42,43). Barcodes commonly range from 9 to >40 nucleotides in length and are either placed in a non-coding region of the genome (e.g., 3' or 5' untranslated regions) or into coding sequences that contain a series of adjacent codons with fully synonymous third positions (i.e., Leu, Val, Ser, Pro, Thr and Ala). The barcodes are introduced to the viral genome via site directed mutagenesis (SDM) or custom synthetic gene fragments (14,42,44). Early iterations of barcoded viruses were marked clones; the marked (i.e., barcoded) regions of which were introduced individually by methods such as SDM (14,35,45). Due to the constraints associated with developing each unique virus individually, marked clone populations often contain a small fraction - less than 20 unique clones - of the diversity that can now be achieved in barcoded virus populations (14,35) (**Figure 1.3**). Current applications of barcoded virus technology use degenerate nucleotides that allow all possible mutations to occur at the site of insertion and, upon amplification and rescue of the recombinant virus, can theoretically generate several million unique barcodes without the burden of generating clones individually (37,42). Due to the nature of barcode generation, this technology is currently limited to viruses for which efficient reverse genetics, cloning, and rescue systems already exist.

Uniquely barcoded viruses within a population function as analogues for naturally occurring variant genomes within a mutant swarm, making them powerful tools for mimicking RNA virus populations (21,43). Barcoded virus populations are high-diversity representations of virus populations and thus allow for a more precise characterization of population dynamics than can be achieved by characterizing naturally occurring diversity (40,42,44). Since barcodes

are introduced to the virus genome without altering the coding sequence, they theoretically have a minimal impact on fitness. This distinguishes them from true quasispecies virus populations, which contain naturally occurring mutations, that may confer significant phenotypic variability (21). Defining natural variants requires whole-genome sequencing and is typically accomplished using short-read sequencing, and assigning reads to individual genomes (binning) to identify variants. This method of variant identification, when employed to quantify population diversity, is extremely sensitive to errors introduced during sample preparation and sequencing (46).

While the fitness-neutral nature of the barcode population model precludes it from capturing the full range of evolutionary processes (such as positive selection), stochastic forces, such as bottlenecks, have a significant impact on virus evolution within individual hosts (38). Barcoded viruses are extremely well suited to quantifying these stochastic forces shaping virus populations, as the neutrality of the barcode allows investigators to examine population dynamics without the variable of fitness (14,35,43,44). Barcoded viruses also provide solutions for examining infection dynamics in systems where intra-host virus diversity is typically constrained or too low to uniquely identify viral lineages, or where independent tracking of multiple infection events, which would be difficult to track by conventional methods, is required (42,47). Additionally, the function of barcodes as unique identifiers allows for determination of the clonal origin of viruses throughout infection and allows for the identification and quantification of variant analogues within a population with extremely high resolution and sensitivity (40–42,48). Finally, using barcode-specific probes, the population dynamics of barcoded viruses can be tracked during transmission events and analyzed by qRT-PCR without the need for deep-sequencing (35).

The predominant use of barcoded viruses is as mimics of highly diverse virus populations in studies of population dynamics during infection, transmission, and treatment. Early work successfully demonstrated the value of synthetic swarm viruses as analogues for

virus populations by using marked clone populations to characterize the impact of stochastic forces, such as bottlenecks (i.e., random and rapid reduction of diversity in a virus population), on poliovirus populations during neuroinvasion in mice, and on WNV and Venezuelan equine encephalitis virus (VEEV) populations within relevant mosquito vectors (14,35,45). This work shed light on the adaptive potential of these pathogens, highlighted the important role infection plays in shaping virus populations, and established synthetic swarms as powerful molecular tools (14,35,45). Since this foundational work with poliovirus, WNV, and VEEV, barcoded viruses have been utilized to answer questions about population dynamics across numerous virus families.

### 1.6 Barcoded viruses in action

*bcZIKV*. Barcoded Zika virus (bcZIKV) was used to characterize ZIKV infection dynamics in pregnant and nonpregnant macaques (47). Low complexity barcode populations persisted in pregnant animals after typical resolution of infection in non-pregnant animals, indicating that an anatomical reservoir had been established in the pregnant macaques (47). This work provided proof-of-concept for the use of bcZIKV *in vivo* to examine virus populations throughout infection and highlighted the potential of barcoded viruses to probe the impacts of anatomical reservoirs, and bottlenecks on virus populations. This work also identified a cumulative reduction in bcZIKV population diversity associated with intra-host bottlenecks during infection in *Ae. aegypti* mosquitoes (**Figure 1.4A**) (37). bcZIKV has also been instrumental in determining the impact of transmission modes between vertebrates and mosquitoes on ZIKV evolution and was used to demonstrate how direct vertebrate transmission chains could promote enhanced ZIKV virulence (49). Further, bcZIKV was used to identify diversity in individual plaque forming units, demonstrating that ZIKV could potentially be transmitted as multigenome aggregates (50).

First published in PLOS Pathogens, April 2023

*bcIAV*. Barcoded influenza A virus (bcIAV) has been employed to study infection routes and their associated bottlenecks, evolution of the NS1 gene, virus replication in different regions of the respiratory tract, and the impact of compartmentalized replication on virus population

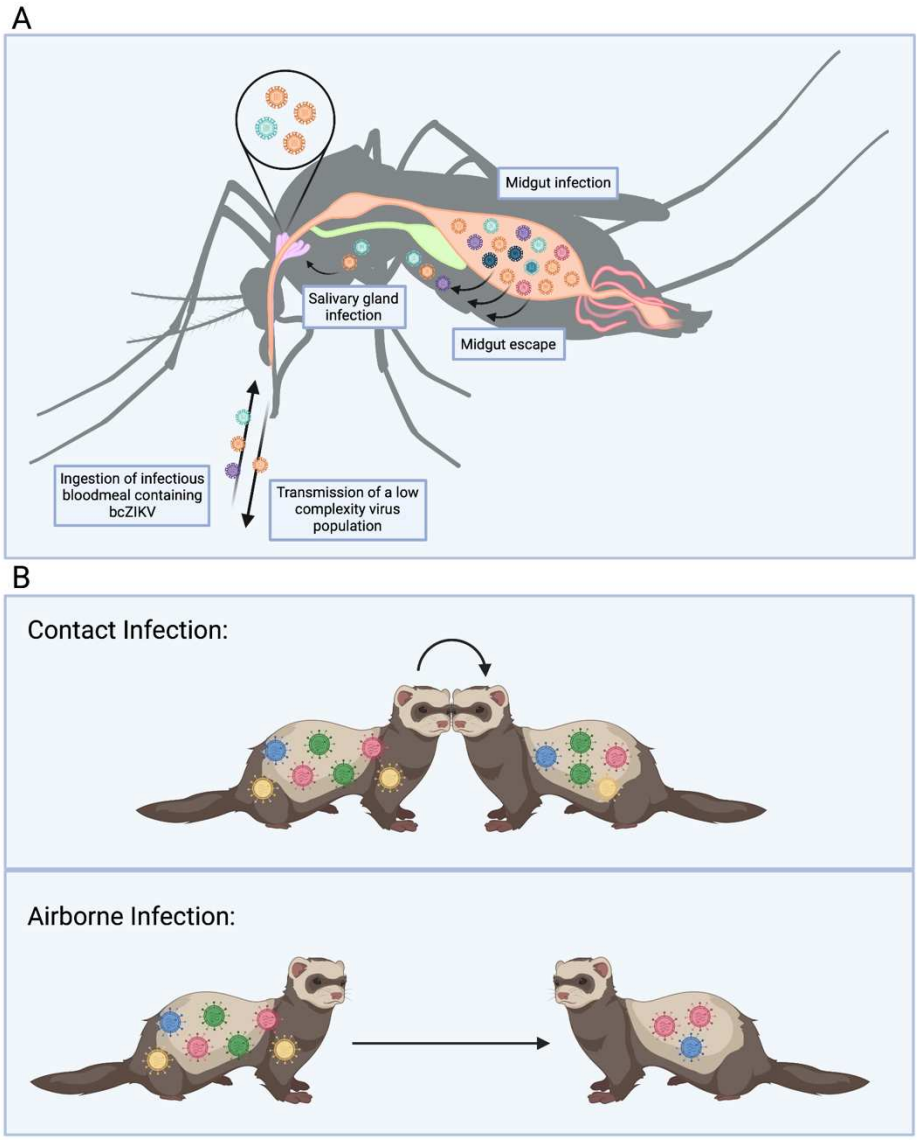



**Figure 1.4. Barcoded viruses mimic virus population dynamics during transmission and infection events.** (A) Intrahost bottlenecks in *Ae. aegypti* rapidly reduce the complexity of a bcZIKV population during mosquito infection. (B) A more complex bcIAV population is transmitted between ferrets via contact infection than airborne infection. (Created with Biorender.com)

dynamics (39,43,44,51). Studies of bcIAV in ferrets allowed for a high-resolution characterization of physiological bottlenecks and demonstrated that aerosol transmission represents a severe and highly restrictive bottleneck (**Figure 1.4B**) (43). Additionally, bcIAV infection of ferret lungs revealed that a series of bottlenecks within the lungs results in genetically distinct virus population "islands" that are heavily impacted by founder effects (43,44). Further work with bcIAV in the respiratory tract established that transmissible droplets are generated in upper respiratory tissues, providing an anatomical target for viral load reduction to prevent IAV transmission (51). bcIAV has also been instrumental in demonstrating the adaptive plasticity of the IAV NS1 gene, and in quantifying reassortment of IAV upon multiple infection (39,52).

*bcCVB*. Coxsackievirus B3 (CVB3) is an enterovirus that can penetrate the gastrointestinal tract and cause systemic infection (53). A highly rich barcoded CVB3 (bcCVB3) population was used to study CVB3 population dynamics in mice and allowed researchers to quantify the impact of the gastrointestinal (GI) barrier on a CVB3 population, demonstrating a significant reduction in barcode diversity upon infection of and replication in extraintestinal tissues (53).

*bcSIV & bcHIV.* Barcoded SIV and HIV (bcSIV and bcHIV respectively) have been used in numerous studies to explore retrovirus population dynamics during transmission, infection, immune escape, treatment, and reactivation and rebound following latency (40–42,48,54–57). Using bcSIV investigators have shown that intrarectal (IR) inoculation of macaques results in 70 to 560-fold less complex SIV populations when compared to intravenous inoculation, demonstrating that infection barriers associated with IR challenge impose a population bottleneck on SIV populations (57). Further studies with bcSIV quantified the impact of the host immune response on an SIV population and the level of replication preceding the generation of escape mutations in unique lineages within an SIV population (41). Studies of reactivation and

rebound following latency or the interruption of combination antiretroviral therapy (cART) with bcSIV established an estimated rate of reactivation in viral reservoirs, an approximate viral load per reactivated latent cell, and determined that the viral lineages that were dominant in the population pre-treatment tend to reactivate first during treatment interruption (42,54,55). Finally, studies of bcHIV populations in mice revealed the efficacy of latency-reversing agents (LRAs) in reducing the diversity of rebound populations and delaying rebound upon interruption of cART (48).

*bcAAV*. Adeno-associated viral (AAV) vectors are a clinically relevant mode of therapeutic gene-transfer that have had success in reprogramming certain cell types in animal models (58). Given the high demand for this therapeutic, there is a growing need to develop and screen AAVs for increased or tissue-specific transduction efficiency (58–61). Barcoded AAVs (bcAAV) are frequently employed as high-throughput screening tools for recombinant AAV vector pools that allow investigators to quantify multiple AAV genome and transcript abundances in parallel in tissues of interest (59,61).

The barcoded virus approach to studying virus evolution is reliable, allows for investigation of population dynamics with unprecedented depth and ease, and has been successfully adapted to study these phenomena in a wide range of virus families and hosts. Barcoded viruses have proven to be valuable tools in assessing viral replication dynamics, progeny production, and polyinfection (i.e., infection of a single cell with multiple unique genomes) at the single-cell level (62,63). Finally, barcoded virus technology has the potential to be highly valuable for computational modelling of infection by providing quantitative estimates of host- and environment-dependent patterns of genetic restriction.

The use of barcoded WNV facilitated the quantification of bottlenecks described in this dissertation. This work revealed that the mosquito midgut plays a significant role in stochastically restricting WNV population size and variant success during infection in a species

dependent manner. However, the factors that contribute to this tissue-level bottleneck are unknown.

### 1.7 The importance of the midgut

For hematophagous disease vectors, like mosquitoes, the midgut functions as a critical interface between pathogen and vector (33,64). Several studies have demonstrated that the mosquito midgut plays an important role in systemic mosquito infection, the mosquito immune response, intrahost virus population dynamics, and vector competence (16,65). The dynamics of cell populations within mosquito tissues is a key area of study in the spheres of both mosquito physiology and vector-biology. Work examining insect midgut tissue at single-cell resolution has shed light on the complexity of this organ by identifying distinct cell populations that comprise the midgut and characterizing their functions (66–68). It is hypothesized that the permissibility of midgut cells contributes to the midgut infection barrier and thus impacts the ability of pathogens to establish infection in a mosquito (65). Additionally, it is postulated that the midgut is the most important organ with respect to determining vector competence (16). For these reasons, in conjunction with our finding that the midgut plays a central role in WNV infection dynamics in mosquitoes, one of the goals of this dissertation was to examine midgut physiology in a primary vector of WNV at single-cell resolution.

### 1.8 Single-cell and single-nucleus RNA sequencing of insect tissues

*Midgut.* Recent single-cell RNA sequencing (scRNA-seq) studies in the *Drosophila* midgut have expanded the model system to the study of intestinal stem cell renewal, cellular response to microbiome and nutrient fluctuations, and cell type localization within the gut (66,67). These foundational studies characterized and identified canonical gene markers for midgut cell populations and elucidated the complexity of midgut tissue (66,67). Through this work investigators also linked cell expression patterns with findings derived from morphological studies of midgut cells, i.e., enrichment of ribosomal protein genes in intestinal stem

cell/enteroblast (ISC/EB) populations; cells that have been demonstrated to have a high ribosome abundance via electron micrograph visualization (66,69). Further, this work in *Drosophila* paved the way for studies of midgut cell dynamics in disease vectors. Indeed, one of the earliest studies of single-cell population dynamics in mosquitoes – a study characterizing *Aedes aegypti* midgut cell populations – draws heavily upon the *Drosophila* midgut cell atlas (68). This work in *Ae. aegypti* employed single-nucleus RNA sequencing (snRNAseq) to identify midgut cell populations and examine the response of distinct cell types to bloodmeal ingestion and digestion (68). Through this approach investigators determined that the proportion of ISC, enterocyte (nutrient adsorptive cells), and EC-like populations were significantly increased by bloodmeal consumption (68). Additionally, they noted significant upregulation of genes associated with metabolic processes in the midguts of blood fed mosquitoes (68). snRNA-seq has also been instrumental in characterizing mechanisms of *Bombyx mori* nucleopolyhedrovirus (BmNPV) immune escape in the midgut of silkworm larvae (70). snRNA-seq of the silkworm larvae midgut during BmNPV infection revealed that BmNPV inhibited expression of antiviral factors, and stimulated enrichment of genes involved in replication – demonstrating the importance of the midgut as a site of virus replication and immune activation (70).

*Hemocyte.* Studies of infection dynamics and innate immunity in insects have profoundly advanced our understanding of innate immunology (71). Hemocytes – invertebrate blood and immune cells – are central to the insect innate immune response (72,73). Single-cell studies of *Drosophila* hemocytes revealed the diversity of this class of cells by identifying unique immune gene profiles/microbial responses, cell cycle gene profiles, and metabolic states (74,75). Further, investigators identified a novel immune response mechanism against parasitic wasp eggs in a subset of *Drosophila* hemocytes that expressed components of fibroblast growth factor (75). Additionally, scRNA-seq facilitated the classification of hemocytes in *Anopheles gambiae* and *Ae. aegypti* mosquitoes by allowing investigators to examine gene expression

patterns of hemocyte subtypes and identify novel subtype markers (72,73,76). One novel subtype, termed ‘megacyte’, was found to play a central role in mosquito immune priming by facilitating hemocyte differentiation (76). These findings underscore that single-cell sequencing is a powerful tool for teasing out cellular functions not evident at the population level (75).

*Brain.* Single-cell transcriptomics has also been applied to studies of the *Drosophila* brain to investigate neuron subtypes, cellular diversity in specific brain regions, and the gene expression patterns of an aging brain (77–79). Examining the *Drosophila* olfactory system at the cellular level allowed investigators to characterize the transcriptomes of several classes of neurons and identify transcription factors and cell-surface molecules that distinguish neuron identity (77). Further work in the *Drosophila* brain assigned gene expression profiles to the major brain regions and cell-types and demonstrated that neurons can be typed based on genes associated with neurotransmitters and neuromodulators (79). Finally, the cellular resolution afforded by scRNA-seq enabled investigators to discover that, in *Drosophila*, aging does not impact cell fate despite cell-type specific indicators of aging in transcriptional profiles and aging associated declines in gene expression (78).

### 1.9 The importance of characterizing infection at cellular resolution

Bulk RNA sequencing, qRT-PCR, and equivalent assays provide invaluable population-level information about the host response to virus infection. However, they don’t provide sufficient resolution to identify permissive and refractory cell types, cell-specific replication kinetics, and gene enrichment associated with infection in individual cells (80). Cell-to-cell variability in infection has been described as early as 1952 and continues to be a key area of study (81). Previous work examining replication and evolutionary dynamics of a diverse vesicular stomatitis virus (VSV) population in baby hamster kidney (BHK) cells demonstrated that individual infected cells produced varying rates of spontaneous VSV mutation, and varying levels of genetic diversity in VSV progeny (82). Additionally, scRNA-seq of dengue virus and

Zika virus infected human hepatoma (Huh7) cells facilitated the identification of novel antiviral and proviral factors by allowing investigators to examine host gene expression correlated with viral RNA (vRNA) load in individual cells (80,83). Therefore, one of the goals of this dissertation was to characterize WNV infection of *Cx. tarsalis* midguts at single-cell resolution.

### 1.10 Goals of dissertation

The overarching goal of this dissertation was to characterize WNV infection dynamics in the mosquito at the whole-tissue and single-cell level, with a specific focus on the midgut. The aims of this dissertation are as follows:

**AIM1:** Test the hypothesis that bottleneck stringency is linked to vector competence by quantifying the impact of infection and escape barrier associated population bottlenecks in variably competent vectors of WNV.

**AIM2:**

**Subaim 2.1**: Test the hypothesis that distinct cell populations in the *Cx. tarsalis* midgut can be characterized by scRNA-seq; identifying distinct cell types and generating a midgut cell atlas (i.e., map of cell type and function) for *Cx. tarsalis*, and contributing to the annotation of the genome via ortholog identification.

**Subaim 2.2 –** Test the hypothesis that cell populations in the *Cx. tarsalis* midgut play distinct roles in the establishment of WNV infection by characterizing WNV infection of the *Cx. tarsalis* midgut at the single-cell level; comparing vRNA levels in specific cell types and identifying host gene expression correlated with WNV vRNA load in individual cells.

# CHAPTER 2: LOSS OF WEST NILE VIRUS GENETIC DIVERSITY DURING MOSQUITO INFECTION DUE TO SPECIES-DEPENDENT POPULATION BOTTLENECKS

## 2.1 Introduction

West Nile virus (WNV, *Flaviridae: flavivirus*), an arthropod-borne virus (arbovirus), is the leading cause of arboviral disease in the contiguous United States (84). Since the introduction of WNV to the Western Hemisphere in 1999, it has spread rapidly and undergone dramatic population-level genetic changes as it adapted to local transmission cycles (3,4,13,85,86). Due to error-prone replication, RNA viruses such as WNV exist as genetically complex mutant swarms (20–22,36). This confers significant adaptive plasticity to these viruses as they encounter different selective environments within vertebrate and invertebrate hosts, enter and spread within new environments, and adapt to new environmental conditions (4,20,23,25,86,87). Previous work demonstrated that mosquito infection dramatically shapes WNV populations via selective pressures and single nucleotide polymorphism (SNP) accumulation in the viral genome, indicating that mosquito infection has a significant impact on the evolution of WNV (4,15,22,30,86). However, stochastic processes such as genetic drift, that are not generally selective, also have a major impact on RNA virus evolution within individual hosts (20,32,38) .

During infection of the mosquito vector, WNV sequentially infects the mosquito midgut and salivary glands, and ultimately is released into the saliva (14,20,25,34,35,37). Infection of and escape from the midgut and salivary glands constitute the main barriers to virus transmission within an individual mosquito (15,20,88). These barriers also impose stochastic reductions on the richness and complexity of the virus population (i.e., bottlenecks), which can negate changes that occur selectively by randomly selecting variants for onward transmission

First published in iScience, August 2023

(14,15,20,35). Several studies have clearly demonstrated that tissue-level bottlenecks play an important role in shaping virus population structure in mosquitoes (14,35,37). Moreover, it has been established that mechanisms of evolution – such as bottlenecks – that alter the frequency of virus variants within the mosquito vector, are determinants of emergence (29).

WNV is transmitted in nature by mosquito species of varying vector competence (VC) – the capacity to become infected with and transmit the virus. VC clearly influences the rate of virus evolution during mosquito infection (11,12,15,16). While it is known that bottlenecks play an important role in virus evolution, it remains unknown whether vector competence impacts these stochastic mechanisms of intrahost evolution (15,35,37,89). Additionally, quantitative characterization of bottlenecks and their impact on virus populations within the mosquito are scarce (20,32). Therefore, we determined the impact of vector competence on the severity of tissue-associated bottlenecks in three variably competent vectors of WNV.

We utilized a highly diverse, molecularly barcoded WNV (bcWNV) to mimic an RNA virus mutant swarm, allowing us to determine the extent to which bottlenecks and vector competence are related (**Figure 2.1**) (62,90). Using bcWNV, we quantitatively measured population bottlenecks in three WNV vectors with varying vector competence; *Cx. tarsalis*, *Cx. quinquefasciatus,* and *Aedes aegypti*. *Culex tarsalis* is the most competent WNV vector of the three species, followed by *Cx. quinquefasciatus* and *Ae. aegypti* (**Figure 2.1)** (11,15). We examined alterations in virus population structure by measuring changes to barcode diversity in midguts, salivary glands and expectorated saliva from each of these species (**Figure 2.1)**. We quantified how bottlenecks associated with infection, dissemination, and transmission influence the overall diversity of a WNV swarm, which gave insight into the impact of vector competence on the stringency of bottlenecks that occur during mosquito infection. In addition to providing a comprehensive and quantitative analysis of WNV population dynamics in three relevant

mosquito species, we determined the probability of infection and transmission for barcodes of varying frequencies in the input population – demonstrating the effect initial proportion in the

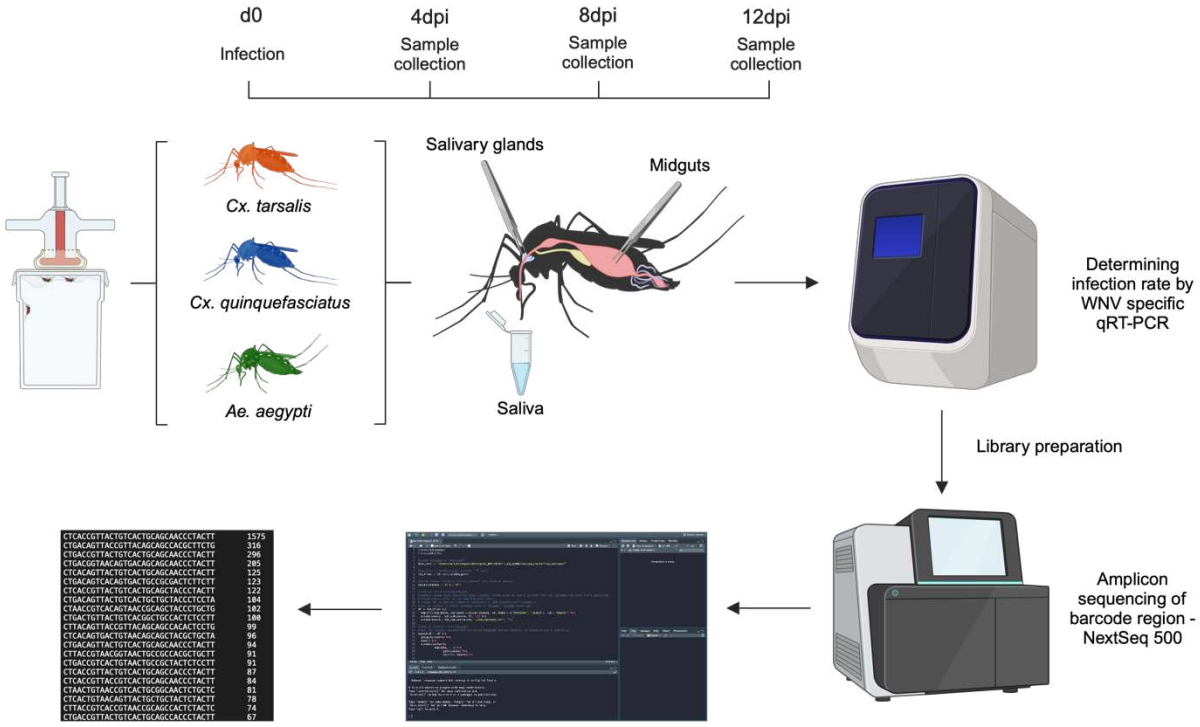



**Figure 2.1. Quantifying the impact of intrahost bottlenecks using bcWNV.** Graphical overview of midgut, salivary gland, and saliva collection, sample validation, amplicon sequencing, and analyses.

population has on successful virus transmission, and highlighting the stochastic nature of bottlenecks and their influence on virus transmission potential.

## 2.2 Results

*Molecular characterization of barcode diversity.* We utilized a virus barcoding method that has been employed previously with Zika virus (37,50). Barcoded-WNV contains degenerate nucleotides at the third coding position of 11 consecutive codons in the NS4b region of the genome (62). These changes are synonymous and do not alter the NS4b amino acid sequence

(**Figure 2.2A**). We deep-sequenced the barcode region of our bcWNV stock and confirmed the barcodes contained varied proportions of A, T, C and G only at the third nucleotide positions,

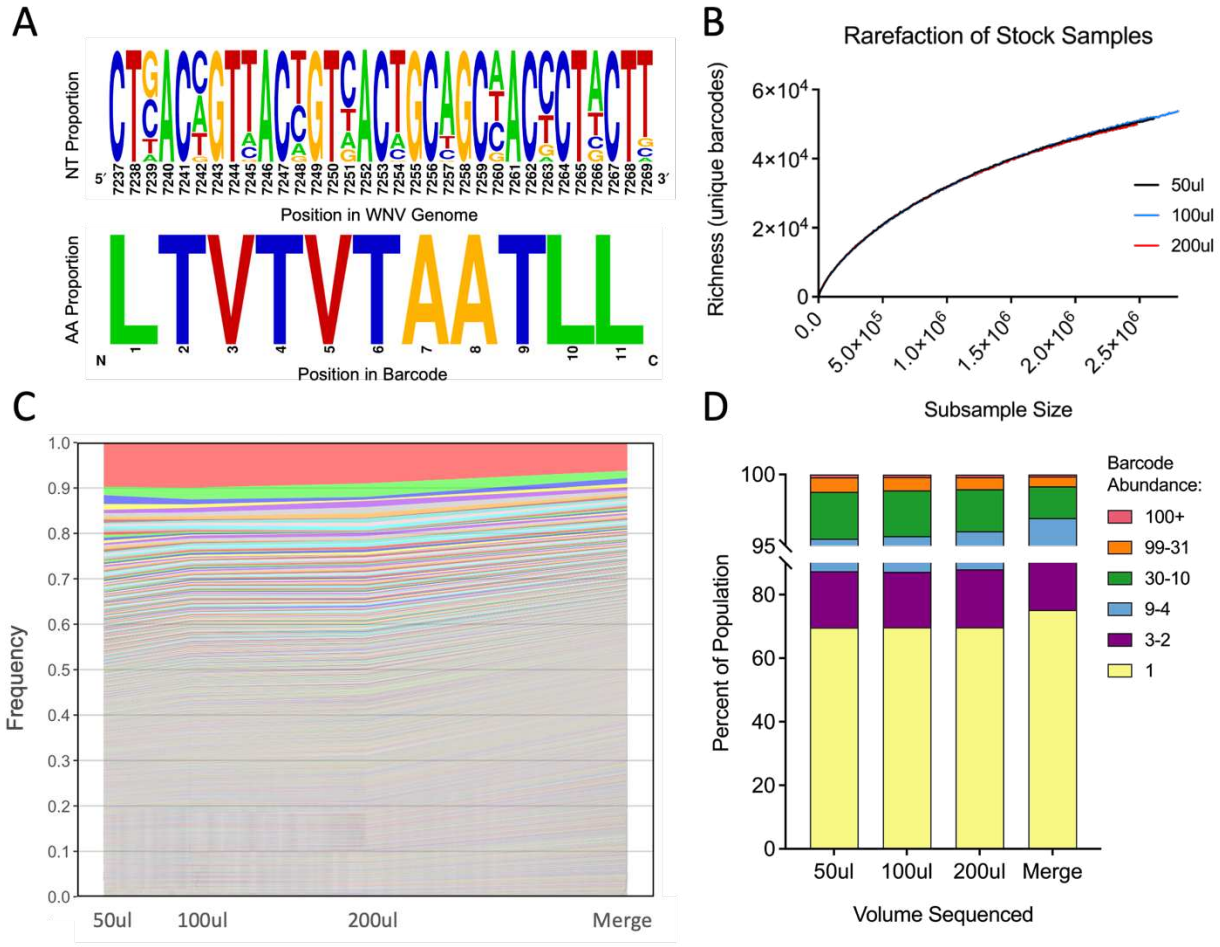



**Figure 2.2. Confirming the diversity of the molecular barcode. (A)** The barcode, located in the NS4B region of the WNV genome, contains degenerate nucleotides inserted at every third codon position for 11 consecutive codons. Degenerate nucleotides produce synonymous changes and do not alter the encoded amino acid. (created with weblogo.berkeley.edu) **(B)** Rarefaction analysis of sequenced bcWNV stock samples estimates barcode richness as a function of sequencing depth. RNA was extracted from 50, 100, and 200ul of virus, deep sequenced and (**C)** barcode frequencies were determined from abundance values and ranked from most to least frequent across three stock samples and a merged population containing unique barcodes from all three stock populations **(D)** and number and frequency of detected barcodes was determined.

with no alterations to the barcode region amino acid sequence (**Figure 2.2A**). We then performed a bootstrap rarefaction to examine barcode richness as a function of read number

(presented as subsample size) and found that sample richness increased as subsample size increased, with no notable flattening of the curve for any sample volume (50ul, 100ul, 200ul) (**Figure 2.2B**). This demonstrated that our sampling volume of 50ul and target coverage of 1.2 million paired-end reads per sample captured a subset of the most abundant barcodes, but not the full richness of the barcode population in each sample. We then evaluated barcode abundance and diversity of the bcWNV stock (**Figure 2.2C, D**). When visualizing individual barcodes, dominant barcodes (those comprising ~10% of total barcodes) were consistently identified across all sample volumes, whereas low frequency barcodes (those detected only once) were variably captured (**Figure 2.2C**). The percent of barcodes occurring at certain abundances was similar across sample volumes, with barcodes occurring only once comprising ~70% of the population, whereas those occurring 100+ times comprised only ~0.2% (**Figure 2.2D**). To maximize the richness of our input population for downstream analyses, we merged the three stock samples by barcode (**Figure 2.2C, D**) and, for barcodes that overlapped between samples, retained the highest detected abundances. This merged input population was used as our input reference for all proceeding analyses.

*Barcoded WNV replication in cells and mosquitoes.* To determine whether introduction of the barcode affects virus replication, we infected mammalian (Vero) and mosquito (CT, *Cx. tarsalis* derived) cells with bcWNV and the parental recombinant wild type WNV (rWT-WNV) at two multiplicities of infection (MOI) and measured extracellular viral RNA (vRNA) (**Figure 2.3A**). vRNA levels were not significantly different between the barcoded and WT viruses at both MOIs, in both cell types at all days post-infection (**Figure 2.3A**). We next exposed *Cx. tarsalis*, *Cx. quinquefasciatus*, and *Ae. aegypti* to an infectious bloodmeal containing bcWNV, collected midguts, salivary glands, and saliva on multiple days post-infection and measured vRNA. All species became infected with bcWNV. However, midgut infection, dissemination (as measured by infection of salivary glands), and transmission (shedding into the saliva) rates were variable

across species (**Figure 2.3B, Table S2.1)**. At 8 days post-infection, *Cx. tarsalis* consistently had the highest infection rates of salivary glands, and saliva, followed by *Ae. aegypti* and *Cx.*

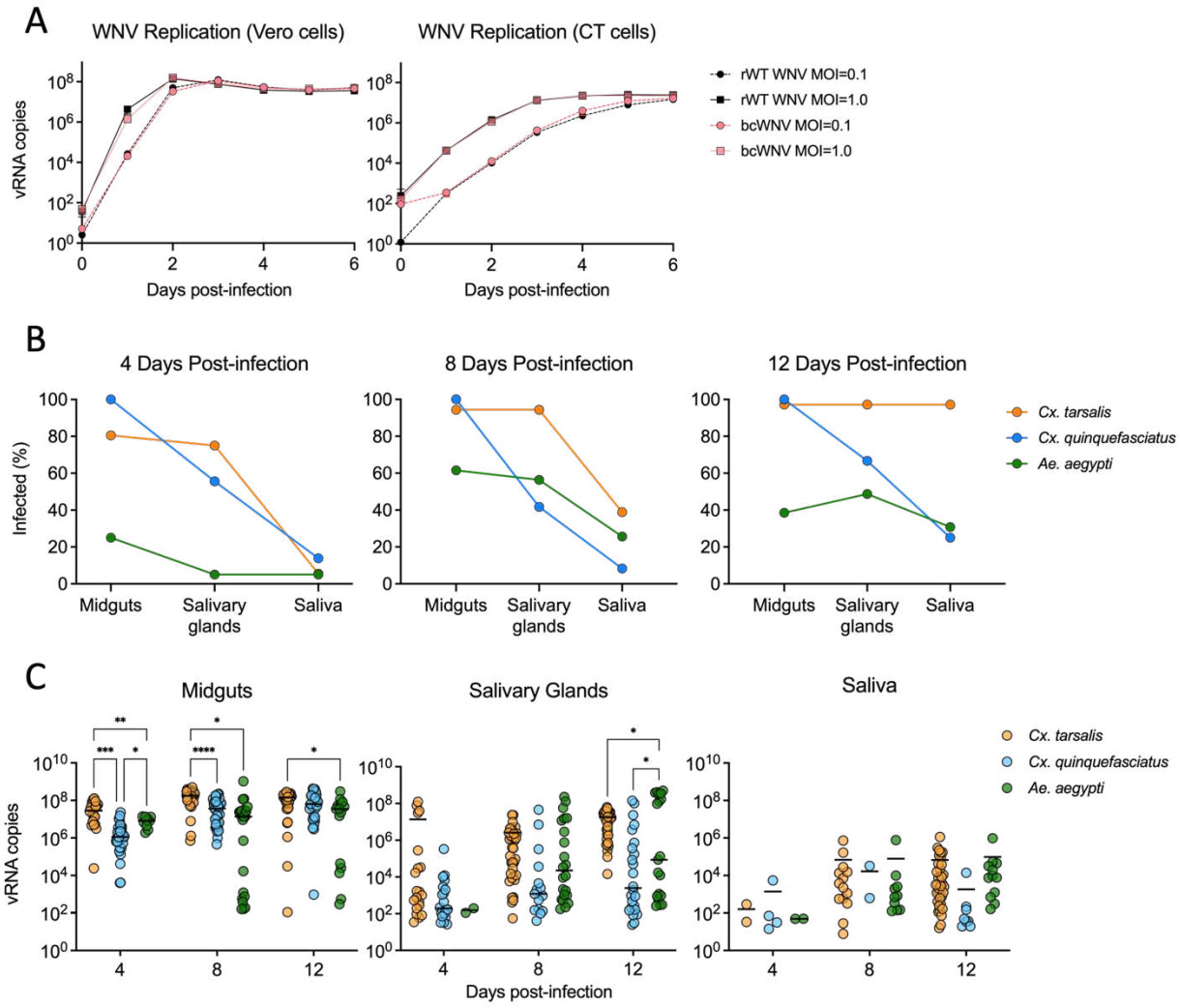



**Figure 2.3. Barcoded WNV efficiently infects, and replicates in cells and mosquitoes. (A)** bcWNV (red) replication was evaluated relative to rWT-WNV (black) at MOIs of 0.01 (dashed line) and 1.0 (solid line) in mammalian (Vero) and insect cells (CT, *Culex tarsalis*-derived embryonic cell line) (n=3, error bars denote mean ± SD). vRNA copies = WNV envelope copies per reaction. *Cx. tarsalis* (orange) and *quinquefasciatus* (blue), and *Ae. aegypti* (green) mosquitoes were infected with bcWNV. At the indicated days, saliva, salivary glands and midguts were collected from each species. **(B)** Infection rates of midguts, salivary glands, and saliva were determined for all species at all timepoints. **(C)** Levels of viral RNA were measured via qRT-PCR (n=35-40), solid lines denote median. vRNA copies = WNV envelope copies per sample. Within each time point, statistical significance between mosquito species was determined by two-way ANOVA with Tukey's multiple comparisons test, * = $p<0.05$, ** = $p<0.005$, *** = $p<0.0005$, **** = $p<0.0001$. Only significant relationships are shown.

*quinquefasciatus* (**Figure 2.3B**). By 12 days post-infection *Cx. quinquefasciatus* salivary gland infection rates overtook *Ae. aegypti*, however, *Ae. aegypti* maintained slightly higher transmission rates (**Figure 2.3B**). We examined vRNA levels in all samples and found that, on average, *Cx. tarsalis* attained the highest peak vRNA levels in the midgut and salivary glands (**Figure 2.3C**). While vRNA peak days and levels varied by species, lower competence vectors (*Cx. quinquefasciatus* and *Ae. aegypti*), despite having lower transmission rates (**Figure 2.3B**), did not have significantly lower levels of vRNA in the saliva compared to the higher competence species (*Cx. tarsalis)* (**Figure 2.3C**).

*Mosquito bottlenecks reduce richness and diversity of bcWNV populations.* We next examined differences in barcode richness (number of unique barcodes) and complexity (Shannon's index) within 10 individual mosquitoes per species at 4, 8, and 12 days post-infection using next-generation sequencing (**Figure 2.4A-C, Figure S2.1**). At 8 days post-infection, bcWNV population richness and complexity decreased significantly between the midgut and salivary glands ($p<0.005$), and midgut and saliva ($p<0.001$) in *Cx. tarsalis* (highest VC) (**Figure 2.4A**). A decrease in richness and complexity was seen between the midgut and salivary glands, ($p=0.002$) and midgut and saliva ($p<0.05$) in *Cx. quinquefasciatus* (**Figure 2.4B**). In *Ae. aegypti* the only significant decreases in richness and complexity were between the midgut and saliva ($p=0.02$, $p=0.01$) (**Figure 2.4C**). No significant reductions in richness nor complexity were observed between the salivary glands and saliva for any species (**Figure 2.4A-C**). We observed similar trends at day 12 post-infection, and no significant reductions in population diversity in any species at day 4 post-infection (**Figure S2.1**).

We then examined what percent of population richness is retained after each bottleneck at all timepoints combined (**Figure 2.4D**). *Culex* species retained significantly more population richness (unique barcodes) between the input and midgut populations than *Ae. aegypti*

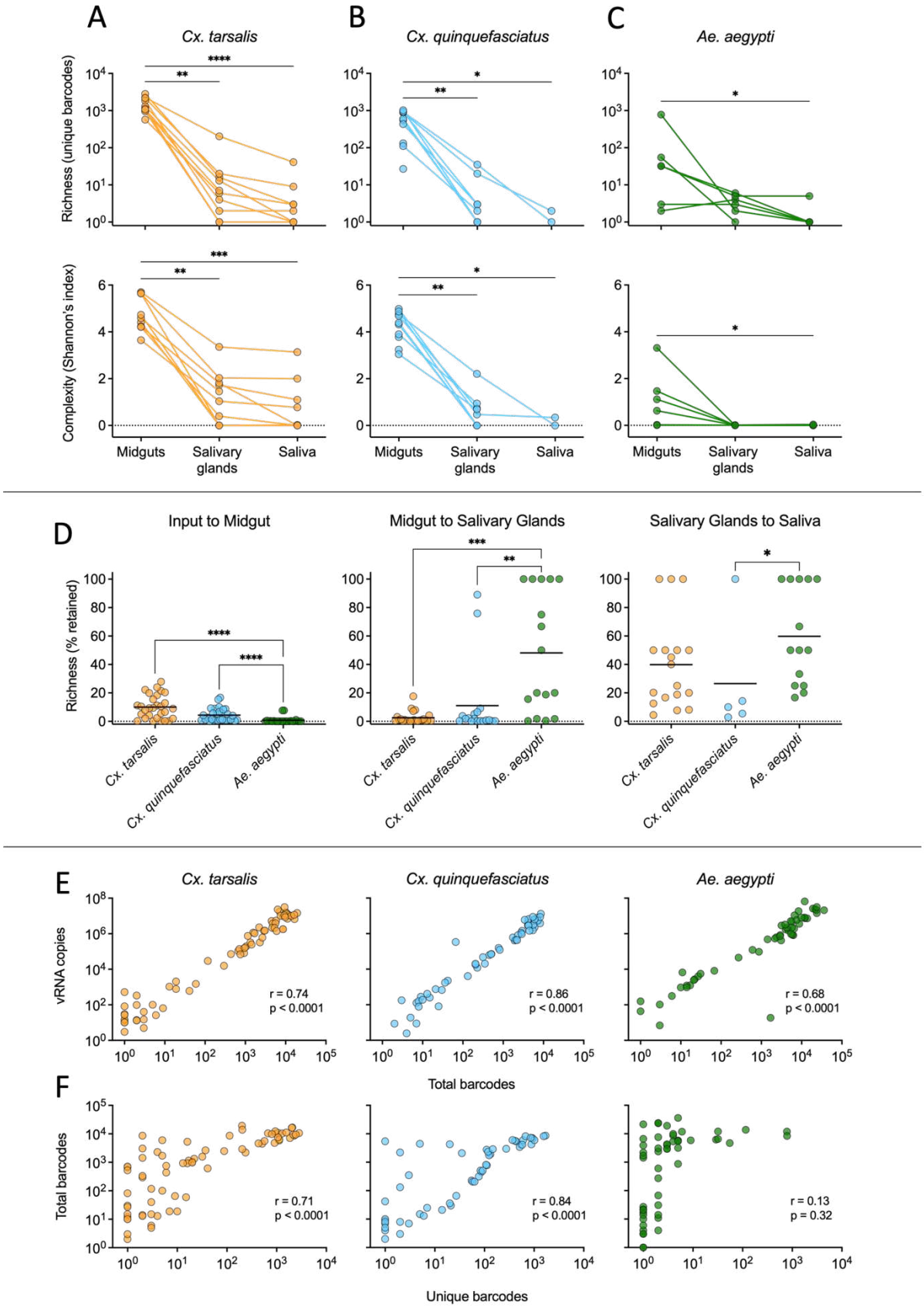

A
Cx. tarsalis
B
Cx. quinquefasciatus
C
Ae. aegypti
Richness (unique barcodes)
Complexity (Shannon's index)
Midguts
Salivary glands
Saliva
D
Input to Midgut
Midgut to Salivary Glands
Salivary Glands to Saliva
Richness (% retained)
Cx. tarsalis
Cx. quinquefasciatus
Ae. aegypti
E
F
vRNA copies
Total barcodes
Unique barcodes
r = 0.74
p < 0.0001
r = 0.86
p < 0.0001
r = 0.68
p < 0.0001
r = 0.71
p < 0.0001
r = 0.84
p < 0.0001
r = 0.13
p = 0.32

**Figure 2.4. Quantifying the impact of intrahost bottlenecks on population richness and complexity.** 8-day post-infection bcWNV population richness and complexity in midgut, salivary gland, and saliva samples from *Cx. tarsalis* **(A)**, *Cx. quinquefasciatus* **(B)**, and *Ae. aegypti* **(C)** mosquitoes. Lines connecting points indicate that samples were collected from the same mosquito. Diversity indices were generated by applying the Shannon function from the QSutils package in R to barcode abundance vectors. Dashed line represents 0. (**D**) Percent of preceding unique barcode population retained after each bottleneck in each species (all timepoints included). Solid lines denote median. **(A-D)** Statistical significance determined using Kruskal-Wallis with Dunn's multiple comparisons test, * = p<0.05, ** = p<0.005, *** = p<0.0005, **** = p<0.0001. Only significant comparisons shown. **(E)** Pearson's correlation between vRNA and total barcodes, and **(F)** unique and total barcodes in all samples from each species. Pearson's r values and significance values are displayed on plots.

(p<0.0001), with *Cx. tarsalis* and *Cx. quinquefasciatus* retaining 10.0% and 4.4% respectively, compared to only 0.8% in *Ae. aegypti* (**Figure 2.4D**). *Ae. aegypti* retained significantly more population richness between the midgut and salivary gland populations than both *Culex* species (p<0.005), and between the salivary gland and saliva populations than *Cx. quinquefasciatus* (p=0.03) (**Figure 2.4D**). *Cx. tarsalis* sustained the greatest loss of richness between the midgut and salivary glands (97.5%), while *Cx. quinquefasciatus* and *Ae. aegypti* sustained the greatest loss of richness between the input and midgut populations (95.6% and 99.2% respectively). To further examine barcode dynamics and to ensure differences in richness/complexity were not an artifact of total barcodes detected, we compared vRNA, total barcodes and unique barcodes for all samples from each species (**Figure 2.4E, F**). We saw a significant, positive linear correlation between vRNA and total barcodes in all samples from all species, demonstrating that vRNA and total barcodes are analogous (**Figure 2.4E**). However, while we observed a significant, positive linear correlation between unique and total barcodes in all *Cx. tarsalis* and *Cx. quinquefasciatus* samples, (r>0.7, p<0.0001) there was no significant linear correlation between unique and total barcodes in *Ae. aegypti* samples (r=0.13, p=0.32) (**Figure 2.4F**).

*Visual representation of barcode population dynamics.* Visual inspection of barcode population dynamics in individual mosquitoes that expectorated detectable barcodes reinforced our quantitative findings in **Figure 2.4**. We saw decreases in barcode population complexity at days

4, 8, and 12 post-infection in individual mosquitoes for each species (**Figure S2.2**), with two representative plots for each species shown (**Figure 2.5A**).

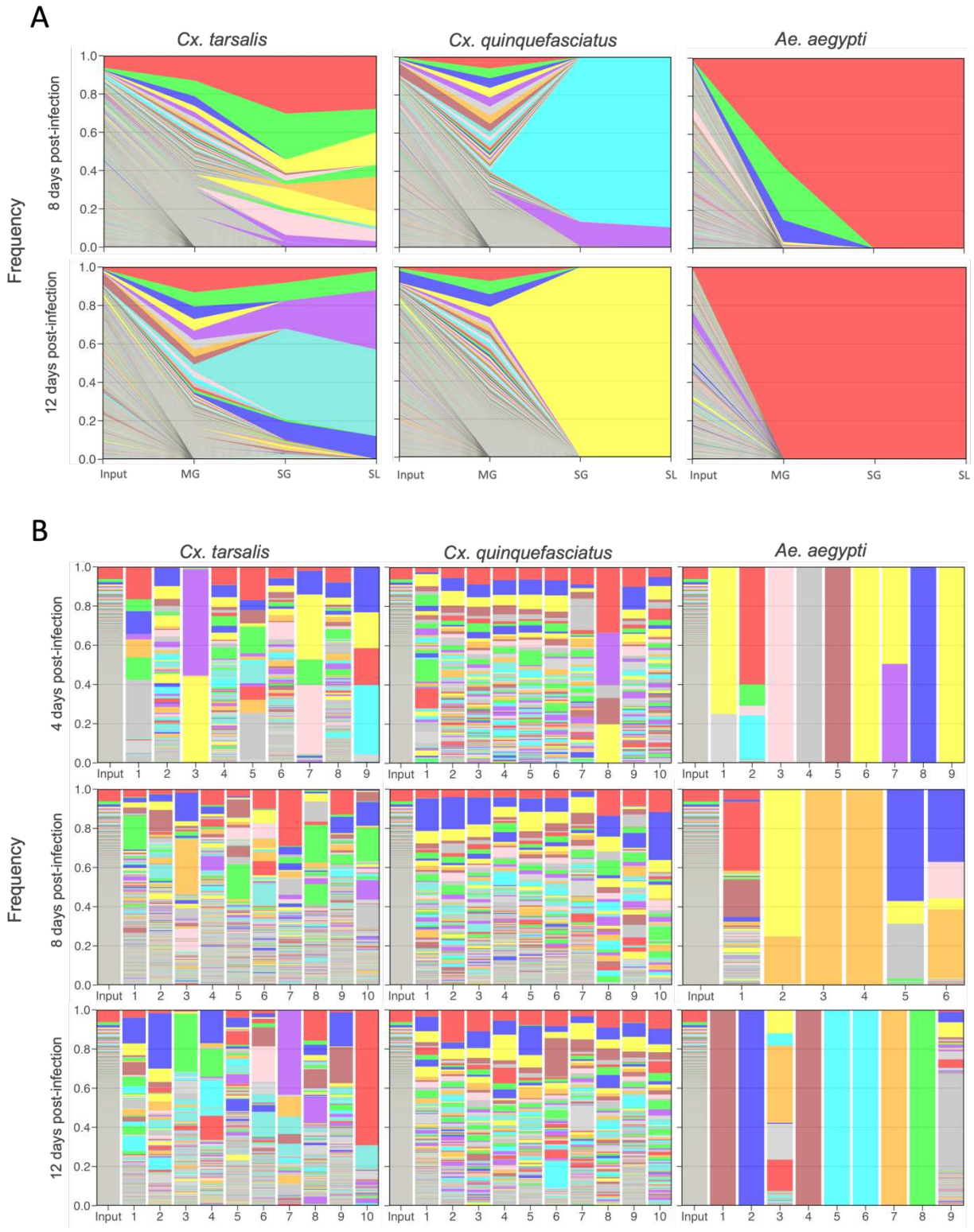


**Figure 2.5**. **Visualizing barcode dynamics across bottlenecks.** (**A**) bcWNV barcode dynamics during mosquito infection in individual *Cx. tarsalis*, *Cx. quinquefasciatus*, and *Ae. aegypti* mosquitoes at days 8 and 12 post-infection. Barcodes were ranked from most to least frequent by each midgut sample. Color represents barcode rank and is not consistent across graphs. MG = midgut, SG = salivary glands, SL = saliva. **(B)** bcWNV barcode maintenance in all midgut samples containing detectable barcodes, at each timepoint post-bloodfeed in *Cx. tarsalis*, *Cx. quinquefasciatus*, and *Ae. aegypti* mosquitoes. Barcodes were ranked from most to least frequent by the input population. Color represents both barcode rank and barcode sequence and is consistent within and across graphs.

Midgut infection results in mosquito species-independent population bottlenecks - barcode populations in the midgut are always less complex than the input population (**Figure 2.5A**). Following this initial bottleneck, the barcode population is rapidly reduced to a small subset of midgut barcodes as the population infects the salivary glands, and sheds into the saliva (**Figure 2.5A**). Importantly, high barcode abundance in the input or midgut does not guarantee progression to the salivary glands and saliva (**Figure 2.5A**).

Certain barcodes in our bcWNV stock were present at higher frequency than most, permitting examination of the relationship between starting population frequency and the likelihood of detection in various mosquito tissues and secretions. We visually compared all midguts to the input barcode population by ranking midgut barcodes by input abundance. In both *Culex* species, barcodes that are present at high frequencies in the input (i.e., the red barcode found at the top of the input column) are consistently maintained in midgut infection, and often become a highly abundant barcode in the midgut population (**Figure 2.5B**). The low complexity of barcode populations in *Ae. aegypti* midguts (consistent with **Figure 2.5A**) precludes inference from this visual comparison, as midgut populations in this species are frequently comprised of fewer than 10 unique barcodes (**Figure 2.5B**).

*Differences in diversity of populations establishing infection in the midgut across species.* We next compared richness and complexity of bcWNV populations across species to determine whether vector competence impacts bottleneck stringency. Midgut richness in *Cx. tarsalis* was

significantly higher than *Cx. quinquefasciatus* at 8 days post-infection (p<0.05), and significantly higher than *Ae. aegypti* at all timepoints (p<0.002) (**Figure 2.6A**).

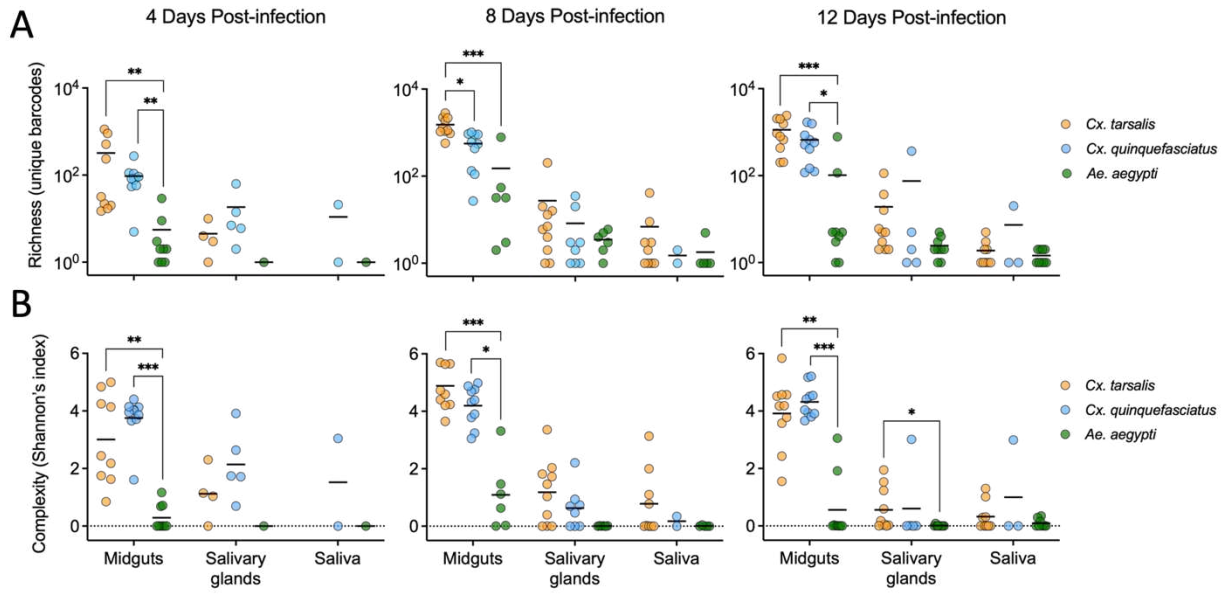



**Figure 2.6. Bottleneck severity, post-midgut infection, does not vary between vectors of differing competence.** Comparing bcWNV population richness (**A**) and complexity (**B**) in midgut, salivary gland, and saliva samples across *Cx. tarsalis* (orange), *Cx. quinquefasciatus* (blue), and *Ae. aegypti* (green) mosquitoes at days 4, 8, and 12 post-infection. Dashed line represents 0 and was added for easier visualization of low complexity values. Within each mosquito sample type, statistical significance between mosquito species was tested using Kruskal-Wallis with Dunn's multiple comparisons test, * = p<0.05, ** = p<0.005, *** = p<0.0005, **** = p<0.0001. Solid lines denote median. Only significant comparisons shown.

*Cx. quinquefasciatus* midgut richness was significantly higher than *Ae. aegypti* midgut richness at days 4 and 8 post-infection (p<0.05) (**Figure 2.6A**). Population complexity in *Culex* midguts was significantly higher than *Ae. aegypti* midguts at all time points (p<0.05) (**Figure 2.6B**). The only significant difference in salivary gland population complexity was observed between *Cx. tarsalis* and *Ae. aegypti* at 12 days post-infection (p<0.05) (**Figure 2.6B**). No significant differences in saliva population diversity were observed between species. To ensure that variability in sequencing coverage did not impact barcode abundance measurements, a subset of sample files were down-sampled to 750,000 reads. We replicated these analyses and

findings on these down-sampled files (data not shown); confirming that coverage variability did not impact our results.

*Barcode dominance in the input does not guarantee transmission.* Having determined that input frequency of a barcode influences barcode maintenance in the midgut of *Culex* species, we examined input frequency's impact on barcode transmission (i.e., detection in the saliva). We plotted the input frequency of every barcode detected in all saliva samples, with the frequencies of the most (upper dashed line, 0.062) and least (lower dashed line, 0.000039) frequent barcodes in the input (**Figure 2.7A**). Transmissible barcodes had a large range of input frequencies, including one at the lowest input frequency, and we saw no significant differences in the starting frequency of transmissible barcodes between species (**Figure 2.7A**). Next, for all transmissible barcodes, we assessed the relationship between proportion in the input population and proportion of successful transmission events (number of saliva samples that contained a barcode/total positive saliva samples) (14). There was a significant positive relationship between input proportion and proportion of successful transmission in both *Cx. tarsalis* and *Cx. quinquefasciatus* ($r=0.62$, $p<0.0001$ and $r=0.52$, $p=0.0002$ respectively) (**Figure 2.7B**), however, no significant relationship was observed in *Ae. aegypti* ($r=0.37$, $p>0.05$) (**Figure 2.7B**).

*Barcode infection and transmission probability is predominantly influenced by frequency in the input.* We next wanted to quantify the likelihood of barcode infection and transmission based on frequency in the input. We calculated the probability a barcode would survive from the input to midgut, midgut to salivary glands, and salivary glands to saliva, based on that barcode's abundance in the preceding population and the size of the proceeding population.

We found that barcodes of higher abundance in the input population consistently had greater probabilities of infecting the midgut and salivary glands, and eventually entering the saliva (transmission) (**Figure 2.7C-E, Table 1**). Barcodes of all frequencies had a significantly

higher probability of infecting the midgut in *Cx. tarsalis* (p<0.0001), followed by *Ae. aegypti,* with the exception of the most abundant barcode quantile (p<0.0001) (**Figure 2.7C, Table 2.1**). *Cx. quinquefaciatus* had the lowest midgut infection probabilities for all barcodes present at

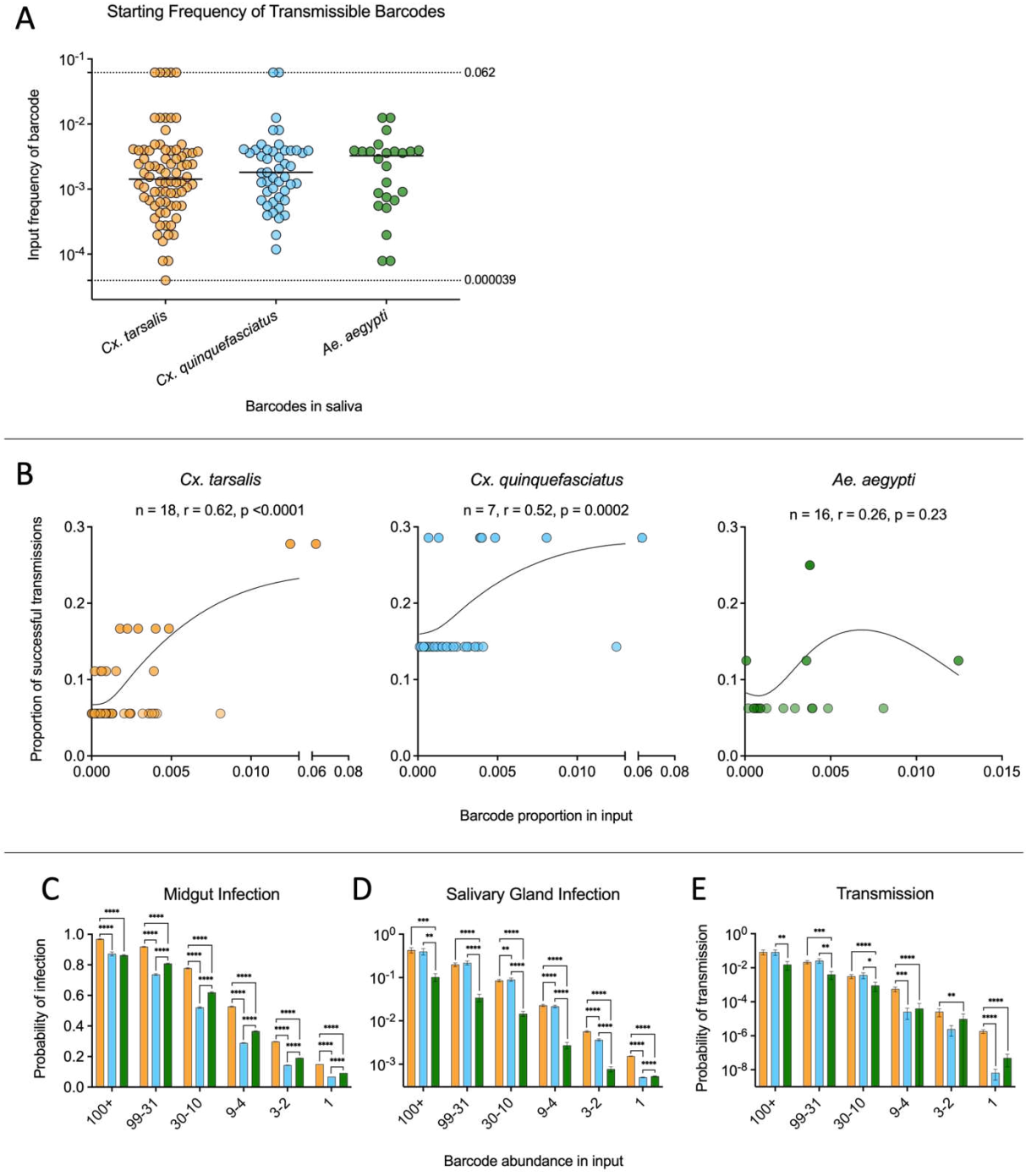


**Figure 2.7. The impact of barcode input frequency on infection and transmission probability.** All points correspond to unique barcodes not mosquitoes. **(A)** The input frequency of all barcodes detected in all saliva samples from each species were plotted and compared by Kruskal-Wallis with Dunn's multiple comparisons test. Dashed lines represent the highest and lowest barcode frequencies in the input. Solid lines denote median. **(B)** Relationship between barcode proportion in the input and proportion of successful transmission events (number of saliva samples that contained the barcode/total positive saliva samples) for all species was determined using Spearman's correlation – Spearman's r values and significance values are displayed on plots, with splines added for visualization. **(C)** Midgut infection probabilities for *Cx. tarsalis* (orange), *Cx. quinquefasciatus* (blue), and *Ae. aegypti* (green) by abundance in the input population. **(D)** Salivary gland infection probabilities for *Cx. tarsalis* (orange), *Cx. quinquefasciatus* (blue), and *Ae. aegypti* (green) by abundance in the input population. **(E)** Transmission probabilities for *Cx. tarsalis* (orange), *Cx. quinquefasciatus* (blue), and *Ae. aegypti* (green) by abundance in the input population. All error bars denote ± SEM. Significance between mosquito species for infection/transmission probabilities was determined using Kruskal-Wallis with Dunn's multiple comparisons test. * = p<0.05, ** = p<0.005, *** = p<0.0005, **** = p<0.0001. Only significant comparisons shown.

**Table 2.1. Barcode infection and transmission probabilities by input frequency.** Average midgut and salivary gland infection probabilities, and transmission probabilities for *Cx. tarsalis* (orange), *Cx. quinquefasciatus* (blue), and *Ae. aegypti* (green) by barcode frequency in the input population. Conditional formatting on probability values denotes highest (green) and lowest (red) values.

| | Barcode abundance in input: | 100+ | 99-31 | 30-10 | 9-4 | 3-2 | 1 |
|---|---|---|---|---|---|---|---|
| | Mean frequency: | 0.00558 | 0.00105 | 0.00032 | 0.00011 | 0.00005 | 0.00002 |
| | Frequency range: | 0.033 - 0.0021 | 0.0021 - 0.00065 | 0.00063 - 0.00021 | 0.00018 - 8.34E-05 | 6.26E-05 - 4.17E-05 | 2.09E-05 |
| | n total: | 3747 | 3527 | 3468 | 3379 | 3639 | 30188 |
| | n unique: | 14 | 70 | 224 | 613 | 1590 | 30188 |
| Midgut Infection Probability | *Cx. tarsalis* | 0.96850 | 0.91871 | 0.77861 | 0.52734 | 0.29778 | 0.14982 |
| | *Cx. quinquefasciatus* | 0.87194 | 0.73747 | 0.52123 | 0.28915 | 0.14412 | 0.06806 |
| | *Ae. aegypti* | 0.86273 | 0.80854 | 0.61992 | 0.36584 | 0.19000 | 0.09171 |
| Salivary Gland Infection Probability | *Cx. tarsalis* | 0.42726 | 0.19785 | 0.08515 | 0.02281 | 0.00570 | 0.00154 |
| | *Cx. quinquefasciatus* | 0.39668 | 0.21768 | 0.08945 | 0.02160 | 0.00366 | 0.00050 |
| | *Ae. aegypti* | 0.10180 | 0.03418 | 0.01462 | 0.00273 | 0.00077 | 0.00053 |
| Transmission Probability | *Cx. tarsalis* | 0.08356 | 0.02184 | 0.00313 | 0.00056 | 2.5969E-05 | 1.8317E-06 |
| | *Cx. quinquefasciatus* | 0.08235 | 0.02606 | 0.00367 | 2.5663E-05 | 2.4516E-06 | 6.6257E-09 |
| | *Ae. aegypti* | 0.01531 | 0.00400 | 0.00091 | 3.9735E-05 | 9.6169E-06 | 4.8920E-08 |

frequencies ≤0.0021 (≤99 barcodes in input). Barcodes present at frequencies >0.00063 (>30) in the input did not have significantly different probabilities of infecting the salivary glands of one *Culex* species over the other. However, barcodes present at frequencies ≤0.00018 (≤9) were significantly more likely to establish infection in *Cx. tarsalis* salivary glands (p<0.0001) **(Figure 2.7D)**. The probability of infecting *Ae. aegypti* salivary glands was significantly lower than both *Culex* species for all barcode frequencies, apart from barcodes with an abundance of 1 (p<0.003) **(Figure 2.7D)**. Transmission probabilities were comparable between *Culex* species

for barcodes at frequencies ≥0.00021 (≥99), but barcodes at frequencies of 0.00018-0.000083 and 0.00002 (9-4, and 1) had a greater probability of transmission in *Cx. tarsalis* (p≤0.0004) **(Figure 2.7E, Table 2.1)**. *Cx. tarsalis* consistently had significantly higher transmission probabilities than *Ae. aegypti* for barcodes present at ≤0.0021 (≤99) (p<0.009) **(Figure 2.7E, Table 2.1)**. Barcodes at frequencies of ≤0.00018 (≤9) had the lowest transmission probability in *Cx. quinquefaciatus* (**Figure 2.7E, Table 2.1**). The highest observed probability of transmission for barcodes of the lowest frequency was $1.8e^{-6}$ in *Cx. tarsalis* – significantly higher than *Cx. quinquefaciatus* and *Ae. aegypti* (p<0.0001). (**Figure 2.7E, Table 2.1**). This is consistent with our finding that *Cx. tarsalis* was the only species to expectorate a barcode of the lowest frequency (**Figure 2.7A**).

### 2.3 Discussion

In nature, both selective and stochastic processes control virus population structure (15,20,32,38,91). Arboviruses, including WNV, experience extreme population bottlenecks as they cycle between mosquito and avian hosts (30,32). The degree of WNV virus diversification, the strength of purifying and positive selection, and the 'tightness' of population bottlenecks is host dependent (15,20,24). However, precise quantification of the degree to which different vectors impose bottlenecks on virus populations remains elusive (32,35). Additionally, the degree to which population bottleneck size is dependent on mosquito vector competence has not been quantified using high resolution molecular tools. This is a significant shortcoming in the field considering the rapid globalization of arboviruses in recent decades, and the frequency with which they encounter new vectors of varying VC in different environments (24,87,92–94). Therefore, we sought to determine how tissue-associated bottlenecks, and their impact on virus population dynamics, varied between three WNV vectors of varying competence. The selected species represent high competence (*Cx. tarsalis*) moderate competence (*Cx. quinquefasciatus*) and low competence (*Ae. aegypti*) vectors of WNV (11,15,26,27). *Cx. tarsalis* and *Cx.*

*quinquefasciatus* are known enzootic vectors of WNV that, in addition to being competent vectors, have high vectorial capacity for WNV due to their propensity to feed on birds, with mammal feeding occurring in a context-dependent manner (16,26–28). *Ae. aegypti* have low vectorial capacity for WNV due to their preference for feeding on humans and other mammals (95,96). However, despite being somewhat refractory to infection, *Ae. aegypti* are capable of transmitting WNV – making them an ideal model for bottlenecks in a low competence vector (11). This, in conjunction with WNV's historical propensity to rapidly adapt to new ecological niches, makes *Ae. aegypti*:WNV a vector:virus pairing of great interest (3,13).

We used a previously described barcoded West Nile virus (bcWNV) which mimics a genetically diverse mutant swarm, allowing us to quantify the impact of mosquito bottlenecks on viral populations and measure stochastic genetic restriction of the swarm (37,62,90). Sequencing our bcWNV stock confirmed a high degree of barcode genetic diversity and preservation of the native amino acid sequence. Importantly, we used unique molecular identifiers (UMIs) in our cDNA primers to uniquely tag RNA molecules and ensure that barcode abundance calculations were free of sequencing errors and amplification bias (**Figure S2.3**) (97,98). Characterization of bcWNV revealed that bcWNV replicates comparably to unbarcoded WNV *in vitro* and was infectious to, and transmissible by, the mosquitoes in our study. Notably, results from vector competence studies conducted herein with bcWNV are extremely similar to those in a prior publication using unbarcoded WNV (15). These data indicate that bcWNV is highly genetically diverse and phenotypically comparable to wild-type WNV and therefore a suitable tool for our studies. Because the introduction of the barcode is coding-neutral and maintains similar fitness between each uniquely barcoded virus, the individual fitness of each barcode in our 'input' WNV population (i.e. that has been successfully produced and propagated) is likely approximately equal (14,35,62,90). However, RNA secondary structure and codon usage can impact the fitness of RNA viruses and result in alterations to the fitness of any

given barcode (99–101). In nature, fitness would likely play a role in every step of mosquito infection and transmission (20,25). However, intrahost bottlenecks within mosquitoes make survival and maintenance of any variant challenging for even the most fit viruses in a population (25,32). Removing fitness as a variable by using a swarm of comparably fit barcoded viruses, allowed us to specifically quantify the impact of stochastic pressures on a virus population within mosquitoes.

We hypothesized based on prior literature that the diversity of barcode populations would decrease during stages of systemic mosquito infection (i.e., from input to midgut, midgut to salivary glands, etc.) (14,15,20,34–37). We also reasoned that bottleneck size would decrease with decreasing vector competence (e.g., high diversity in highly competent *Cx. tarsalis* and lower diversity in less competent *Ae. aegypti*) (11,16). We observed overall decreases in barcode diversity between the midgut and saliva populations in all species – confirming the occurrence of bottlenecks, consistent with previous studies of bottlenecks in mosquitoes, in all species of interest (14,35,37). However, *Ae. aegypti* was the only species in which we did not observe significant decreases in barcode diversity between the midgut and salivary glands due to the extremely low barcode diversity found in *Ae. aegypti* midguts. We found that the complexity of barcode populations infecting the mosquito midgut were significantly higher in *Culex* species than *Ae. aegypti* on all sampled days post infection. We also saw, on average, 13-fold and 5-fold higher barcode richness in *Cx. tarsalis* and *Cx. quinquefasciatus* midguts compared to *Ae. aegypti*. In contrast, we observed minimal differences in barcode diversity between salivary gland and saliva populations across different mosquito species with different VC. It seems likely that this is due to the severe initial reduction of genetic diversity imposed on the bcWNV population during midgut escape (35,88). This indicates that bottlenecks occurring downstream of midgut infection negate differences in diversity observed in the midguts of these mosquitoes, as all species end up expectorating

comparably diverse virus populations. Moreover, these data support the observation that maintenance of WNV genetic diversity upon mosquito infection is associated with the degree of vector competence – with a larger proportion of standing viral genetic variation from the input population present in the midguts of our more competent vectors. This founding population is then subject to additional, non-species dependent, bottlenecks during systemic infection of the mosquito.

Reduction of input population richness upon midgut infection varied by species with *Ae. aegypti* sustaining the greatest reduction, followed by *Cx. quinquefasciatus* and *Cx. tarsalis*, further demonstrating the stringency of the midgut infection barrier in *Ae. aegypti*. Previous work with barcoded Zika virus (bcZIKV) – a virus for which *Ae. aegypti* is a highly competent vector – saw no significant decrease in bcZIKV population diversity between the input population and midgut (37). This further supports that *Ae. aegypti*'s low competence as a WNV vector impacts the diversity of the barcode population establishing infection in the midgut. Interestingly, we noted the most dramatic loss of unique barcodes between the midgut and salivary glands of *Cx. tarsalis*, our most highly competent vector. It has been previously shown that when the midgut is readily infected, as in *Cx. tarsalis*, a severe bottleneck occurs during midgut escape, which supports this finding (20,35,88). The most severe reduction in barcode richness between the salivary glands and saliva was seen in *Cx. quinquefasciatus,* which aligns with our findings that *Cx. quinquefasciatus* produced the fewest transmission positive saliva samples out of the three species (**Figure 2.2B**) and that low frequency barcodes have the lowest probability of transmission in *Cx. quinquefasciatus* (**Figure 2.6E**). Further, this finding aligns with a study of WNV vector competence that noted high midgut infection rates and low dissemination and transmission rates in *Cx. quinquefasciatus* (26).

The cause of the significant differences in midgut population diversity between *Culex* and *Ae. aegypti* remain to be determined. *Ae. aegypti* was the only species in which no

significant linear relationship existed between total and unique barcodes – indicating an overall dearth of barcode population diversity in that species. Interestingly, despite the low barcode diversity in *Ae. aegypti*, all *Ae. aegypti* samples contained more total barcodes than *Cx. quinquefasciatus*, and *Ae. aegypti* salivary glands contained more total barcodes than *Cx. tarsalis*. The lack of linear relationship between total and unique barcodes, in conjunction with the high number of total barcodes identified in *Ae. aegypti* tissues, demonstrates that low diversity in *Ae. aegypti* midguts is not due to poor virus replication. This is further supported by our finding that *Ae. aegypti* and *Cx. quinquefaciatus* achieved comparable peak vRNA levels in the midgut.

Successful midgut infection hinges upon a virus entering and replicating in midgut cells (16,88,102). Thus, the strong bottleneck/genetic restriction in *Aedes* midguts compared to *Culex* could be due to a smaller subset of midgut cells that are susceptible to infection. Further, genetic bases for infection and escape barriers have been identified and shown to have significant influence on the success of vector:virus pairings (16,34,88). It is possible that the *Ae. aegypti*:WNV pairing is impacted by unknown cellular, and genetic factors (16,33). Single-cell analyses of mosquito midguts at early timepoints post-infection could identify differences in cell population involvement in, and genetic response to midgut infection. This would elucidate the establishment of population diversity between different species, and provide key insight into the physiology of infection within arbovirus vectors. Additionally, the mosquito microbiome is known to have a profound impact on both vector competence and vectorial capacity (16,103). Recent work has demonstrated that pre-activation of mosquito innate immune system by the microbiome can promote antiviral defense (104,105). This innate immune activity can, in turn, impede virus transmission (104). While our understanding of mosquito microbiome interactions with pathogens remains limited for most mosquito:virus pairings, the microbiomes of *Culex* species and *Ae. aegypti* are known to differ (103,106,107).

An important remaining question concerns the likelihood of transmission of minority population variants given the bottlenecks that occur in mosquitoes. In other words: at what population frequencies does transmission of a WNV variant become either very likely or very unlikely? Previous work demonstrated that mosquitoes expectorate a unique virus population each time they feed, and that conditions that impact the frequency of virus variants within the mosquito vector are causal factors of emergence (29). This knowledge, combined with the stringent reductions in barcode population diversity that we observed upon escape into the saliva, suggests that bottlenecks play a role in the transmission of rare or unique variants regardless of their fitness.

We found a significant relationship between barcode frequency and proportion of successful transmission events in *Culex* mosquitoes, the most significant WNV vectors included in this study. However, despite the relationship between starting frequency and successful transmission, higher barcode frequency (~3%) in the input does not guarantee transmission, as we only identified our most frequent input barcode in 7 saliva samples. This is consistent with studies of bcZIKV in *Ae. aegypti*, that demonstrated that the highest frequency barcode in the input population was not consistently the highest frequency in the saliva (37). We found that infection and transmission probability decreased with barcode input frequency in all species – this aligns with previous work demonstrating that in a stochastic system, the main variable that influences viral dominance and/or maintenance throughout infection is frequency in the input population (14,38). However, we observed that, while rare, barcodes of the lowest frequency can escape into the saliva, and noted that rare barcodes (i.e., low frequency barcodes – present in the input population at frequencies from $2.1e^{-5}$ – $6.3e^{-5}$) had a non-zero chance of midgut and salivary gland infection and transmission in all species. This finding is supported by work characterizing WNV population dynamics in *Cx. pipiens*, which demonstrated that all variants in a starting population had a chance of surviving intrahost bottlenecks (14). Further, this finding

highlights the stochasticity of mosquito transmission, as previous studies have demonstrated that stochastic transmission is marked by a non-zero chance of transmission for all viruses in a population (37,43). Interestingly, low frequency barcodes had a significantly higher chance of establishing infection in, and being transmitted by, *Cx. tarsalis* when compared to our lower competence vectors. This indicates that the most competent vector is the most likely to become infected with and promote the transmission of rare variants in a virus population. Given that the virus population that is being expectorated by a mosquito can ultimately establish infection in a new host, these findings suggest that mosquito infection, particularly in highly competent vectors, serves as an opportunity for rare virus variants to be transmitted and establish infection in a new host, which has important implications for WNV evolution and emergence (29,36). Moreover, bottlenecks and transmission of novel, low-frequency virus variants that go on to establish infection in new hosts, alter virus population structure in a non-selective manner.

The severity and variability of the midgut infection and escape barriers across species makes it difficult to compare salivary gland associated bottlenecks across species. The salivary glands are critical sites of virus infection and replication, as they are the compartment directly preceding virus transmission (34,88). Previous work has demonstrated the impact replication in the salivary glands has on WNV population diversification and the subsequent diversity of the transmissible population (15,29,32). Future work in this area could better examine salivary gland bottlenecks by infecting mosquitoes via intra-thoracic injection and bypassing the midgut infection and escape barriers; an approach used previously in a study characterizing the impact of intrahost bottlenecks on Venezuelan Equine Encephalitis virus in its primary vector (35). This approach was not undertaken as part of this work because our focus was on biological transmission mechanisms, and as such we opted to mimic natural infection as closely as possible. Additionally, limitations associated with laboratory colony-derived mosquitoes may

impact vector competence (108). Future studies with outbred or field-collected mosquitoes could be warranted to confirm the phenotype.

The ongoing globalization of arboviruses has guaranteed that they will continue to encounter new vectors of varying VC (94). Additionally, it is likely that as viruses adapt to their local ecology and vectors, the interactions between virus and mosquito species (e.g., VC, bottleneck stringency, etc.) may be altered as well. The work presented here demonstrates that vector competence is linked to the stringency of intrahost population bottlenecks and that this in turn influences the maintenance of WNV population diversity upon midgut infection. Ultimately the virus that is relevant in nature is the one that gets transmitted; while vector competence does not appear to impact the diversity of transmissible virus populations, infection of a highly competent mosquito vector can promote the transmission of new or rare virus variants.

**2.4 Methods**

*Cells.* Vero (ATCC CCL-81) cells were maintained in Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal bovine serum (FBS), 1% penicillin-streptomycin (10,000 U/ml), and 50 µg/mL gentamicin at 37°C with 5% $CO_2$ (18). CT, *Cx. tarsalis*-derived embryonic cell line cells were maintained in Schneider's insect medium with 10% FBS, 1% penicillin-streptomycin (10,000 U/ml), and 50 µg/mL gentamicin at 28°C (37,109).

*Mosquito infection.* Mosquito studies were conducted using three species of laboratory colony-derived mosquitoes; *Culex quinquefasciatus* (>50 passages)*, Culex tarsalis* (>50 passages) and *Ae. aegypti* (>30 passages). West Nile virus infections in mosquitoes were performed under A-BSL3 conditions. Larvae were raised on a diet of powdered fish food. Mosquitoes were maintained at 26–27°C (*Culex*) or 28-29°C (*Aedes*) with a 16:8 (*Culex*) or 12:12 (*Aedes*) light:dark cycle (12). All species were maintained at 70-80% relative humidity, with water and sucrose provided *ad libitum*. *Cx. tarsalis*, *Cx. quinquefasciatus*, and *Ae. aegypti* mosquitoes

were exposed to an infectious bloodmeal containing a 1:1 dilution of defibrinated calf's blood and bcWNV stock at a final concentration of 3-6e$^7$ pfu/ml. A final concentration of 1mM ATP was added to *Ae. aegypti* bloodmeals to encourage feeding to repletion. All bloodmeals were provided in a hog's gut membrane feeder, warmed by circulating 37°C water. Following 40-60 minutes of feeding, mosquitoes were anesthetized at -20°C and engorged females were separated into cartons corresponding to each timepoint and maintained on sucrose.

*Generation of rWT-WNV and bcWNV P2 stocks.* The parental rWT-WNV (epidemic lineage I strain, 3356) and bcWNV P2 stocks were generated by infecting confluent flasks of Vero cells at an MOI of ~0.5 with rWT-WNV or bcWNV P1 stock, as previously described (110,111). Supernatant was harvested from cells 3 days post-infection, aliquoted into single use stocks and stored at -80°C. Titers for each stock were determined by standard Vero cell plaque assay (50). RNA was extracted from the bcWNV P2 stock at volumes of 50, 100, and 200ul using a Direct-zol miniprep plus kit (Zymo) and sequenced as described below.

*Growth curves.* Vero and CT (*Culex tarsalis* derived) cells were infected with bcWNV and rWT-WNV at MOIs of 0.1 and 1.0, for one hour, washed three times with PBS and fresh growth media added. Supernatant was sampled daily and RNA was extracted using the Omega MagBind Viral DNA/RNA 96 kit on the KingFisher Flex Magnetic Particle Processor (ThermoFisher). qRT-PCR was performed using One-Step SuperScript qRT-PCR kit (Invitrogen) and WNV envelope primer and probes (**Table S2.2**). RNA copies/ml was extrapolated using an RNA standard. The RNA standard was generated by amplifying ~1kb of sequence surrounding the envelope region of the WNV genome using a forward primer containing a T7 promoter sequence. The resulting amplicon was transcribed with T7 polymerase, and RNA was extracted and quantified.

*Collection of mosquito tissues.* At indicated time points, mosquitoes were cold-anesthetized and legs and wings were removed prior to salivating on capillaries containing immersion oil for 30 minutes. Capillaries were broken off into tubes containing 100ul of mosquito diluent and centrifuged at 15,000xg for 5 minutes at 4°C. Mosquitoes were then transferred to a dish containing 1X phosphate buffered saline (PBS) for salivary gland and midgut dissection. Salivary glands and midguts were rinsed in clean PBS before being transferred to individual tubes containing 200ul of mosquito diluent and a stainless-steel bead for homogenization. Tissue samples were homogenized using RetschMixerMill MM400 at a frequency of 29 cycles/second for 1 minute. RNA was extracted using the Omega MagBind Viral DNA/RNA 96 kit on the KingFisher Flex Magnetic Particle Processor (ThermoFisher). Presence of viral RNA was assessed by qRT-PCR as described above. Midgut, salivary gland, and saliva samples from 10 mosquitoes per species per timepoint were selected for next-generation sequencing with preference given to those with detectable vRNA in their saliva.

*Library preparation and sequencing.* cDNA was generated using Superscript IV RT enzyme and kit (ThermoFisher), and a previously described custom cDNA primer containing a unique molecular identifier (97,98) (**Table S2.2, Figure S2.3**). cDNA was purified using AMPure XP beads (Beckman) at a 1X concentration. All bead purifications were performed using the KingFisher Flex Magnetic Particle Processor. The barcoded region of the genome was amplified via PCR using custom primers and KAPA HiFi HotStart mastermix (Roche) (**Table S2.2**). Amplified samples were purified using AMPure XP beads at a 0.6X concentration. Samples were uniquely indexed with Illumina i5 and i7 indices via PCR using custom primers and KAPA HiFi HotStart mastermix (**Table S2.2**). Second-round amplification products were purified using AMPure XP beads at a 0.6x concentration. Samples underwent a final library amplification using a KAPA Library Amplification Kit (Roche) and were purified using AMPure XP beads at a 0.7X concentration prior to quantification and pooling. Sample concentrations

were determined using Qubit dsDNA High Sensitivity kit (Thermo Fisher) and pooled by mass at 2ng per sample. The final pool was concentrated using AMPure XP beads at a concentration of 1.4X. Size distribution (~350 bp for this library) and concentration of the final pool were verified by TapeStation D1000 HS Screen Tape assay (Agilent) and KAPA library quantification (Roche) respectively. A pilot run was performed to determine target coverage and stock diversity. The pilot library was denatured and a 15% PhiX control was spiked in. The library was loaded at a concentration of 1.5pM and sequenced using a NextSeq 550 system, 300 cycle (2x150 paired-end) mid output kit (Illumina). A 40% PhiX control was spiked into all libraries to offset the low diversity of the library and prevent low sequence quality. All libraries were loaded at a concentration of 1.5pM and sequenced using a NextSeq 550 system, 300 cycle high output kit (Illumina). Sequencing data files can be accessed through Mendeley Data; "cxt_bcWNV_amplicon_seq", Mendeley Data, V1, doi: 10.17632/88xx6596wz.1, "cxq_bcWNV_amplicon_seq", Mendeley Data, V1, doi: 10.17632/s7kthpysb7.1, "aea_bcWNV_amplicon_seq", Mendeley Data, V1, doi: 10.17632/556zktntym.1.

*Analyses.* The preprocessing and analyses of next generation sequencing data was performed as previously described (62,97). Briefly, sequence data was demultiplexed and trimmed of adaptor and index sequences. Paired-end reads were merged and aligned to the WNV genome, reads were grouped by UMI and consensus sequences for unique RNA molecules were generated (**Figure S2.3**). The barcode region was identified by upstream and downstream flanking regions, extracted from the consensus sequences, and abundance determined using custom Perl and Ruby scripts. Barcode population richness and complexity (Shannon's index) for each sample were determined using custom R scripts (R version 4.2.2.); richness was determined by summing the number of unique barcodes detected in a sample, while complexity was determined by applying the Shannon function from the QSutils (quasispecies diversity) package to barcode abundance vectors (112).

When calculating richness retention (**Figure 2.3D**) there were salivary gland and saliva samples in *Ae. aegypti* that appeared to retain greater than 100% of the unique barcodes identified in the preceding population – a finding consistent with previous studies employing a barcoded virus system (37). This system does not allow for the generation of novel barcodes after amplification and rescue of the initial barcoded virus stock, and any barcodes introduced by mutation are removed from the data-set prior to analysis. While some barcodes were found to be truly "parentless" (i.e., present in one compartment but not present in the upstream compartment) we found that many of these parentless barcodes were present in the raw reads from the sample file associated with the upstream compartment. These missing parent barcodes were present at low enough levels that they were removed from the data-set by previously described Primer ID cutoff rates (97,98). For every case where richness retention exceeded 100% due to low-level parent barcodes being filtered out of the data-set, we adjusted the percent retention value to 100%.

We merged all sample files by barcode and determined that the entire data-set contained >30,000 unique barcodes. As mentioned previously, this system does not allow for the generation of novel barcodes within the mosquito – leading us to determine that ~20,000 unique barcodes not initially identified in our bcWNV stock population were present in the stock, but not detected. We merged the additional unique barcodes into our merged input population to create an estimated input population. Since we could infer presence but not abundance, we preserved the abundances of the merged input population and assigned the minimum abundance of 1 to all additional unique barcodes. This estimated input population was used for all probability calculations. All probability calculations are derived from barcode abundance, population size (number of total barcodes in the current population), and subsample size (number of total barcodes in the downstream population) and thus are based upon but not

entirely representative of the data. Infection and transmission probabilities were calculated via custom R scripts, using the following equation:

$$\left(1-\left(1-\left(\frac{x_{i1}}{n_1}\right)\right)^{n_2}\right)\times\left(1-\left(1-\left(\frac{x_{i2}}{n_2}\right)\right)^{n_3}\right)\times\ldots\times\left(1-\left(1-\left(\frac{x_{ib-1}}{n_{b-1}}\right)\right)^{n_b}\right)$$

$x_{in}$ = *abundance of unique barcode in n population*
$n_n$ = *total barcodes in n population*
$n_{n+1}$ = *total barcodes in downstream population*

These calculations were performed for all barcodes in all mosquitoes individually and then averaged to produce mean probabilities for each barcode in each species. We divided all barcodes present in the estimated input population at abundances >1 into quintiles containing ~3500 total barcodes (± 250 barcodes to prevent dividing abundance groups) (**Table 1**). We then averaged the mean probabilities for each species within each of the abundance quantiles (**Table 1**).

All scripts used for data processing, analysis, and visualization are available on GitHub (https://github.com/fitz-meyer/bcWNV_analysis.git). Statistical analyses were performed in GraphPad Prism version 9.4.1. Normality of datasets was measured by Shapiro-Wilk test in R. Differences in sample richness and complexity by sample type and species were measured by Kruskal-Wallis with Dunn's multiple comparisons test. Differences in infection and transmission probability were measured by Kruskal-Wallis with Dunn's multiple comparisons test. Relationships between unique and total identified barcodes, and starting frequency and proportion of successful transmission were measured by Pearson's and Spearman's correlation respectively. Additional statistical details can be found in the text and figure legends.

# CHAPTER 3: A SINGLE-CELL ATLAS OF THE *CULEX TARSALIS* MIDGUT DURING WEST NILE VIRUS INFECTION

## 3.1 Introduction

Arthropod-borne viruses represent a severe and ever-growing public health threat (113). Mosquito-borne viruses alone are estimated to cause over 400 million infections globally each year (114). In order for transmission of a mosquito-borne virus to occur, a mosquito must first become infected with a virus via ingestion of an infectious bloodmeal after feeding on a viremic host (20,115). Said virus must establish infection in the mosquito midgut before it can escape the midgut and disseminate into the body cavity, and eventually enter the salivary glands and saliva – where transmission can occur (20,115). The mosquito midgut is a complex organ comprised of a variety of cell-types with distinct functions (64,116). It is a site of digestion, nutrient absorption, endocrine signaling, and innate immune activity (64,116). The midgut is also the site of infection and escape barriers that strongly influence virus population dynamics (20,64). Previous studies have demonstrated that successful infection of the midgut epithelium, and replication and immune evasion therein, is essential for establishing disseminated infection in a mosquito (115,117). In these ways, for hematophagous disease vectors like mosquitoes, the midgut serves as a critical interface between vector and pathogen.

*Cx. tarsalis* is the primary vector of West Nile virus (WNV) in much of North America (7,118,119). WNV is the most epidemiologically important arbovirus in North America, causing ~2,700 deaths from 1999 to 2022 (1,7,118). Despite the importance of *Cx. tarsalis* as a vector of WNV and other important human viruses, studies examining the cellular composition of its midgut, and WNV infection dynamics therein, are nonexistent. The recent publication of the full *Cx. tarsalis* genome, in conjunction with the growing body of work demonstrating the successful application of single-cell RNA sequencing methodologies in insect models has made it possible to address this significant knowledge gap (66–68,71–73,76,120,121). Therefore, we performed

single-cell RNA sequencing (scRNA-seq) on dissociated midgut cells from both mock and WNV-infected *Cx. tarsalis* mosquitoes to gain a better understanding of how the *Cx. tarsalis* midgut functions as the interface between vector and WNV.

We utilized a scRNA-seq approach previously demonstrated to be flavivirus RNA inclusive, which allowed us to detect WNV viral RNA (vRNA) in addition to *Cx. tarsalis* host transcripts (122). Through this approach we identified distinct midgut populations corresponding with midgut cell-types previously described in *Drosophila* and *Aedes aegypti* midguts – enterocyte (nutrient absorption cells), enteroendocrine (secretory cells), cardia (peritrophic matrix secreting cells), intestinal stem cell/enteroblast (undifferentiated progenitor cells), proliferating intestinal stem cell/enteroblast, visceral muscle cells, and hemocyte (immune cells) – and characterized the infection and replication dynamics of WNV within each population (64,66,68,76,116). We found that WNV infects all midgut cell-types, with evidence suggesting enhanced replication in enteroendocrine cells and cells enriched for mitochondrial genes. Additionally, we characterized the *Cx. tarsalis* immune response to WNV infection at both the whole-midgut and single-cell level. This study has bolstered our understanding of how WNV establishes infection in a highly competent vector, and elucidated the midgut physiology of *Cx. tarsalis*.

### 3.2 Results

*Single-cell sequencing of female Cx. tarsalis midguts on the 10X Genomics platform identified distinct cell clusters.* Using the 10X Genomics platform we performed scRNA-seq on dissociated mock and WNV-infected *Cx. tarsalis* midgut pools at 4 and 12 days post-infection (dpi) (**Figure 3.1**). We recovered an average of 2,416 cells per pool with an average coverage of 255,000 reads per cell, which were mapped to the *Cx. tarsalis* genome (**Supplemental Tables S3.2, S3.3**). Following quality control (QC) filtering, we retained data for 12,886 cells at 4dpi (7,386 WNV-infected, 5,500 mock), and 9,301 cells at 12dpi (4,609 WNV-infected, 4,692

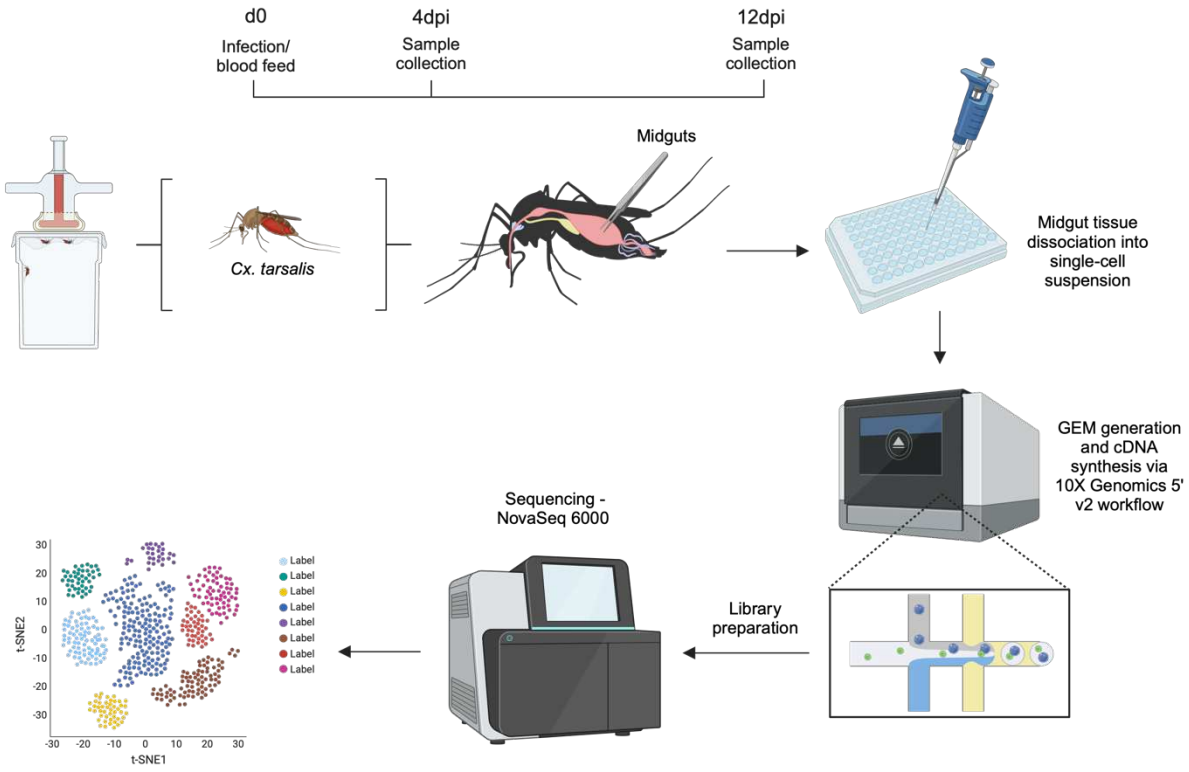
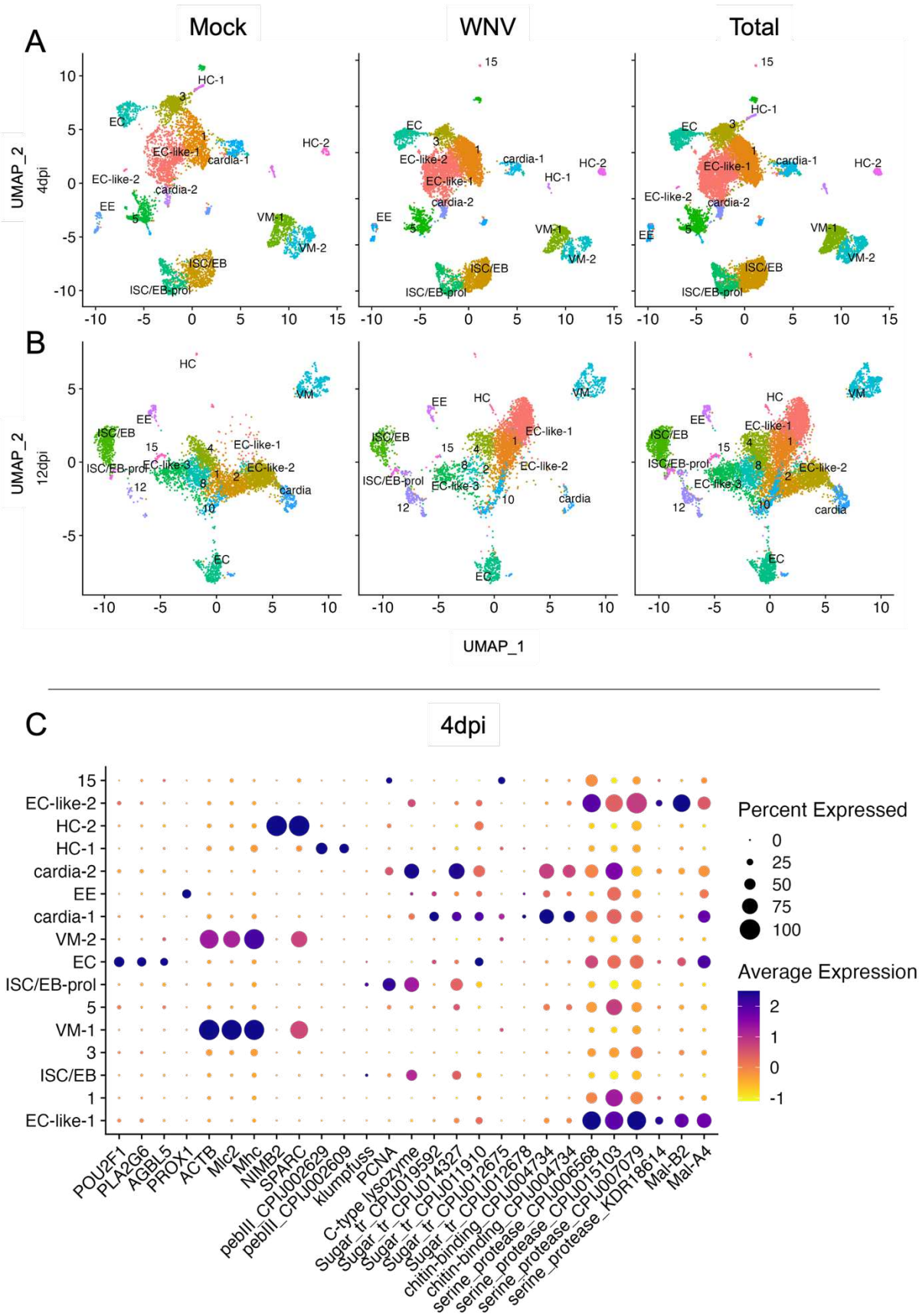



**Figure 3.1. Single-cell sequencing of *Cx. tarsalis* midguts.** Graphical overview of midgut collection, dissociation, and single-cell RNA sequencing.

mock) for downstream analyses (**Supplemental Table S3.4**). Cells retained after QC contained an average of 597 (611 WNV-infected, 580 mock) and 407 (448 WNV-infected, 367 mock) unique genes per cell at days 4 and 12dpi respectively.

*Identification of distinct midgut cell populations via canonical markers.* All cell-type identifications, with indicated exceptions, were based on cluster markers that were conserved between mock and WNV-infected conditions (**Supplemental Table S3.5**). Guided clustering in Seurat generated 15 and 16 distinct clusters of cells at 4 and 12dpi respectively (**Figure 3.2A-B**). We noted less separation between clusters at 12dpi, a pattern which was further sustained by fewer identifying cluster markers resulting in fewer total identified clusters at 12dpi compared

to 4dpi (**Figure 3.2A-B**). Between both time points, the cell types of 22 out of 31 total cell clusters were identified using canonical gene markers and gene enrichment patterns previously identified in *Drosophila* and *Ae. aegypti* midguts (**Figure 2A-D**) (66–68,73,76). Clusters 7 in both the 4 and 12dpi datasets were identified as enterocytes (EC) due to significant expression of POU2F1 (nubbin), PLA2G6 (phospholipase A2), and AGBL5 (zinc carboxypeptidase) (**Figure 3.2C-D**). Cluster 10 (4dpi) was identified as enteroendocrine cells (EE) by the expression of PROX1 (homeo-prospero domain) (**Figure 2C**). At 12dpi, PROX1 did not appear as a conserved cluster marker in our dataset, however cluster 13 was identified as EE by high PROX1 expression visualized via feature expression map (**Figure S3.2**). Due to high expression of Mlc2 (light chain), Mhc (myosin), and ACTB (actin), clusters 4 and 8 (4dpi) and 9 (12dpi) were identified as visceral muscle cells (VM); VM-1, VM-2, and VM respectively (**Figure 3.2C-D**). These structural/cytoskeletal genes, and other orthologous structural/cytoskeletal genes, are present in a variety of cell types. As such, we further confirmed the cell type of clusters 4, 8, and 9 by visualizing high expression of Mlc2, Mhc, and ACTB in all VM clusters via feature expression map (**Figure S3.3**). We identified cluster 13 (4dpi) and 16 (12dpi) as hemocytes (HC-2 and HC respectively) based on expression of NIMB2 and SPARC (**Figure 3.2C-D**) (73,76). SPARC is highly expressed in VM cells as well as hemocytes and thus SPARC was only considered a hemocyte marker in conjunction with low expression of VM cell markers (**Figure 3.2C-D).** We identified cluster 12 (4dpi) as HC-1 due to high expression of two pebIII genes shared with the 12dpi HC cluster (**Figure 3.2C-D**). Clusters 0 and 14 (4dpi), 3, 0, and 6 (12dpi) were identified as EC-like based on enrichment for several serine protease and alpha amylase genes (**Figure 3.2C-D**) (66,68). Clusters 9 (4dpi) and 11 (12dpi) were enriched for sugar transport and chitin binding genes, as well as several EC-like genes, and as such were identified as cardia-1 and cardia respectively (**Figure 3.2C-D**). Cluster


A
Mock
WNV
Total
UMAP_2
4dpi
HC-1
HC-2
EC
EC-like-1
EC-like-2
cardia-1
cardia-2
EE
5
VM-1
VM-2
ISC/EB
ISC/EB-prol
15
3
1
B
12dpi
HC
VM
EE
ISC/EB
15
4
EC-like-1
ISC/EB-prol
EC-like-3
EC-like-2
8
2
12
cardia
10
EC
UMAP_1
C
4dpi
15
EC-like-2
HC-2
HC-1
cardia-2
EE
cardia-1
VM-2
EC
ISC/EB-prol
5
VM-1
3
ISC/EB
1
EC-like-1
POU2F1
PLA2G6
AGBL5
PROX1
ACTB
Mlc2
Mhc
NIMB2
SPARC
pebIII_CPIJ002629
pebIII_CPIJ002609
klumpfuss
PCNA
C-type lysozyme
Sugar_tr_CPIJ019592
Sugar_tr_CPIJ014327
Sugar_tr_CPIJ011910
Sugar_tr_CPIJ012675
Sugar_tr_CPIJ012678
chitin-binding_CPIJ004734
chitin-binding_CPIJ004734
serine_protease_CPIJ006568
serine_protease_CPIJ015103
serine_protease_CPIJ007079
serine_protease_KDR18614
Mal-B2
Mal-A4
Percent Expressed
0
25
50
75
100
Average Expression
2
1
0
-1

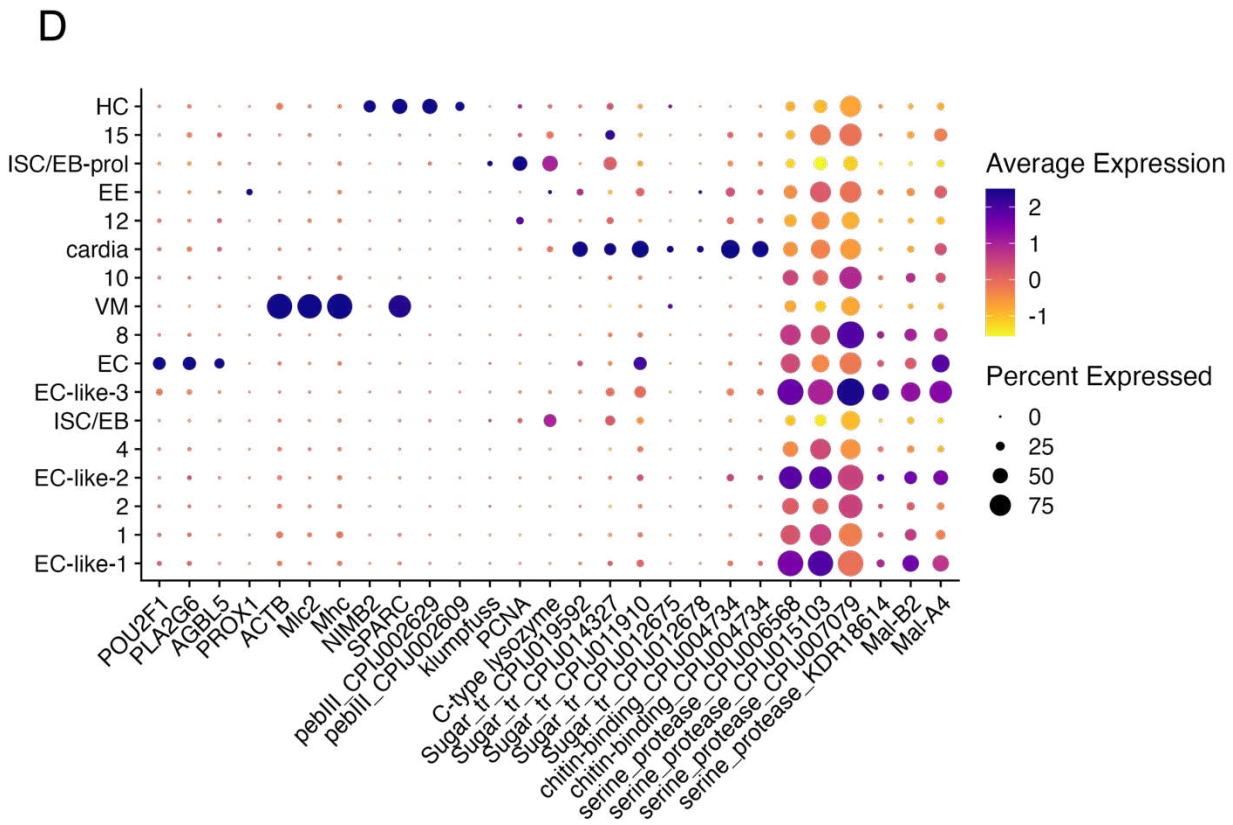



**Figure 3.2. Cell-typing of midgut cell populations.** Uniform Manifold Approximation and Projection (UMAP) reduction was used to visualize midgut cell populations. (**A**) 15 distinct clusters were identified at 4dpi and (**B**) 16 distinct clusters were identified at 12dpi. (**C-D**) Expression of canonical marker genes that were conserved between infection condition (mock and WNV-infected) were used to determine cell type. Dot size denotes percent of cells expressing each gene, color denotes scaled gene expression. Dot plots represent the entire population, not divided by infection condition.

11 (4dpi) was identified as cardia-2 due to expression of C-type lysozyme and a sugar transporter (**Figure 3.2C**). Clusters 2 and 6 (4dpi) were determined to be intestinal stem cells/enteroblasts (ISC/EB) by visualizing klumpfuss expression localized to clusters 2 and 6 via feature expression map (**Figure S3.4A**). Klumpfuss is a canonical marker for EBs not ISCs, however, EBs and ISCs are often indistinguishable by UMAP (66–68). Clusters 5 and 14 (12dpi) did not display obvious klumpfuss expression but shared top features with clusters 2 and 6 (4dpi), and thus were also called as ISC/EBs (**Figure S3.4A-B**). Clusters 6 (4dpi) and 14 (12dpi)

were significantly enriched for PCNA and aurora kinases A and B – markers for cell proliferation and mitosis – and therefore named ISC/EB-prol to reflect this (**Figure 3.2C-D, Figure S3.4C**). Clusters without identifying markers are subsequently referred to by cluster number (e.g., cluster 10). We compared the proportion of each cluster between mock and WNV-infected replicates and found no significant differences (**Figure S3.1A-B**). The percent of the total population comprised by each cluster/cell-type can be found in **Table 3.1**.

**Table 3.1. Cluster proportion.** Percent of the total midgut cell population each distinct cluster comprises.

| | 4dpi | | 12dpi | |
|---|---|---|---|---|
| Cluster | Cell Type | Percent of Population | Cell Type | Percent of Population |
| 0 | EC-like-1 | 19.42 | EC-like-1 | 14.86 |
| 1 | | 19.03 | | 12.02 |
| 2 | ISC/EB | 13.16 | | 9.83 |
| 3 | | 7.68 | EC-like-2 | 9.78 |
| 4 | VM-1 | 6.81 | | 7.88 |
| 5 | | 6.71 | ISC/EB | 7.85 |
| 6 | ISC/EB-prol | 6.29 | EC-like-3 | 7.25 |
| 7 | EC | 5.73 | EC | 6.31 |
| 8 | VM-2 | 5.52 | | 6.25 |
| 9 | cardia-1 | 2.96 | VM | 5.93 |
| 10 | EE | 2.57 | | 3.38 |
| 11 | cardia-2 | 1.56 | cardia | 2.86 |
| 12 | HC-1 | 1.18 | | 2.45 |
| 13 | HC-2 | 1.05 | EE | 1.66 |
| 14 | EC-like-2 | 0.19 | ISC/EB-prol | 0.78 |
| 15 | | 0.14 | | 0.46 |
| 16 | | | HC | 0.45 |

*Characterization of* Cx. tarsalis *midgut secretory and immune cells.* Enteroendocrine (EE) cells are the secretory cells of the midgut and, through the secretion of neuropeptides, regulate behavioral responses (66,68). We identified several *Cx. tarsalis* orthologs for previously

described gut hormones found in EE cells - short neuropeptide (sNPF), bursicon (Burs), ion transport peptide (ITP), and a tachykinin (Tk) receptor (66–68,123,124). Tachykinin receptor, present in 3.3% of 4dpi EEs and 0.65% of 12dpi EEs, was the only detectable neuropeptide/neuropeptide receptor in our EE populations (**Figure 3.3A-B**). At both timepoints EE populations were significantly enriched for IA2, and SCG5 – canonical neuroendocrine genes (**Figure 3.3A-B**) (125,126). EE populations were also significantly enriched for Syt1, and showed expression of Syt4, Syt6, nSyb and Syv1A – all involved in vesicle docking and secretion (**Figure 3.3A-B)** (66,127). Interestingly, both EE populations were enriched for NEUROD6, a neuronal differentiation gene known to be involved in behavioral reinforcement (**Figure 3.3A-B)** (128).

Hemocytes (HC) play a central role in the mosquito immune response – the exact nature of which varies by class of HC (65,72,73,76). We examined the expression of identifying markers in total HC populations to distinguish HC class. At 4dpi, the HC-1 population was determined to be oenocytoids based on expression of SCRB3, and the HC-2 population was determined to be mature granulocytes based on expression of SCRASP1 and enrichment for several c-type lectin, defensin, and cecropin genes (**Figure 3.3C**) (65,72,73,76). The 12dpi HC population is comprised of two classes of hemocyte - mature granulocytes and oenocytoids, based on expression of SCRASP1 and SCRB3 respectively (**Figure 3.3C**) (73). The oenocytoid populations in *Cx. tarsalis* do not appear to express NIMB2, a gene previously identified as a marker present in all hemocyte classes in *Anopheles gambiae* (73).

*COG profiles demonstrate homogeneity between midgut cell populations despite differences in conserved markers.* We next examined the transcriptional profiles of each cluster with a broader perspective to both gain insight into the function of unidentified clusters and compare the transcriptomes of distinct midgut cell populations across timepoints. We used cluster of orthologous genes (COG) categories obtained as part of our ortholog identification efforts, and a

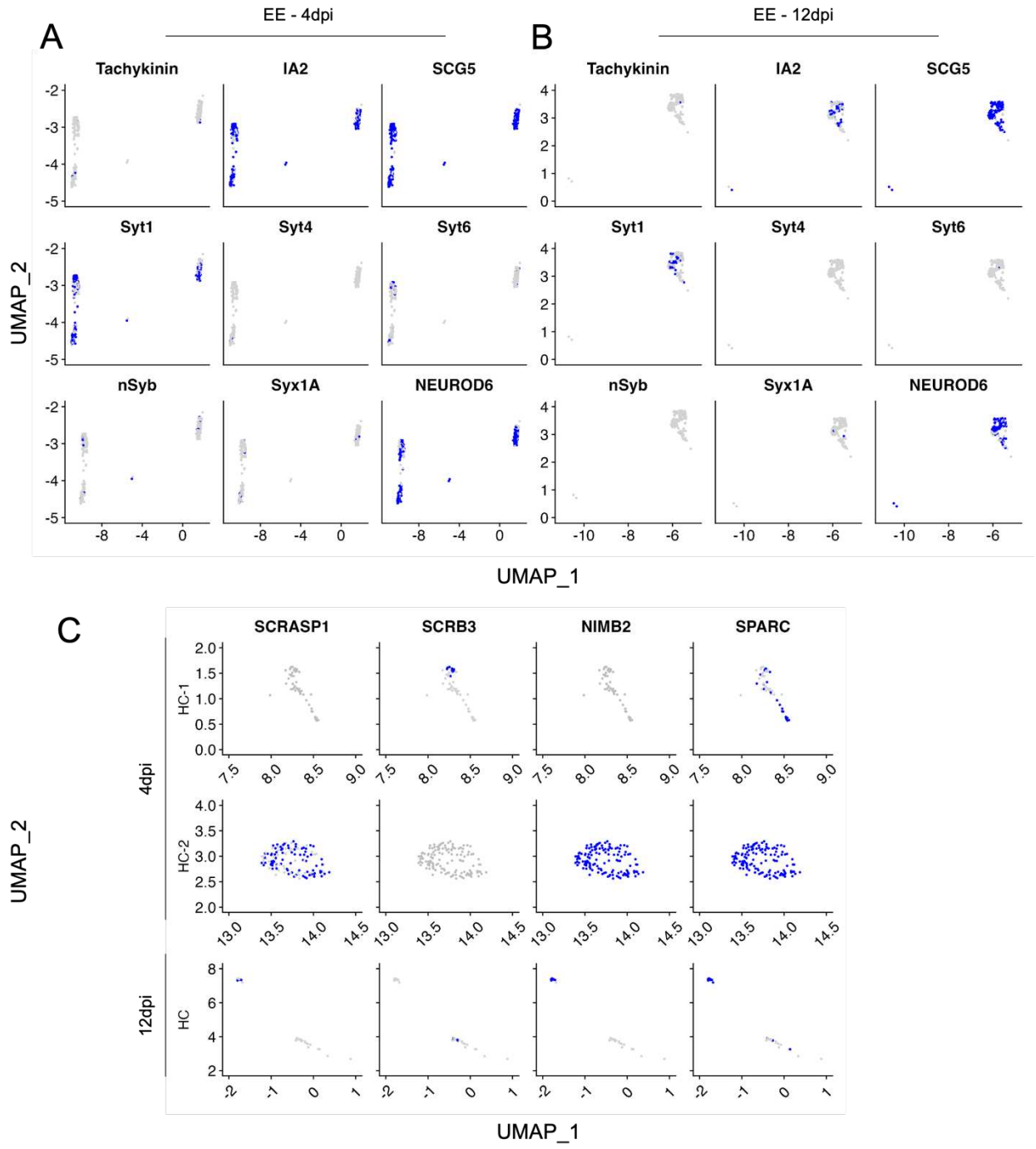



**Figure 3.3. Characterization of enteroendocrine cells and hemocytes.** Gene counts were converted to binary (Blue = 1, Grey = 0) and visualized via UMAP to demonstrate proportion of EE cells expressing neuropeptides, and secretory genes (**A-B**) and proportion of HC cells expressing HC markers, as well as granulocyte and oenocytoid markers (**C**). Feature plots show binary expression for the total population, not divided by infection condition.

two-pronged approach to visualization – COG profiles of all genes expressed in >75% of cells in a given cluster (termed “base genes”) (**Figure 3.4A**)

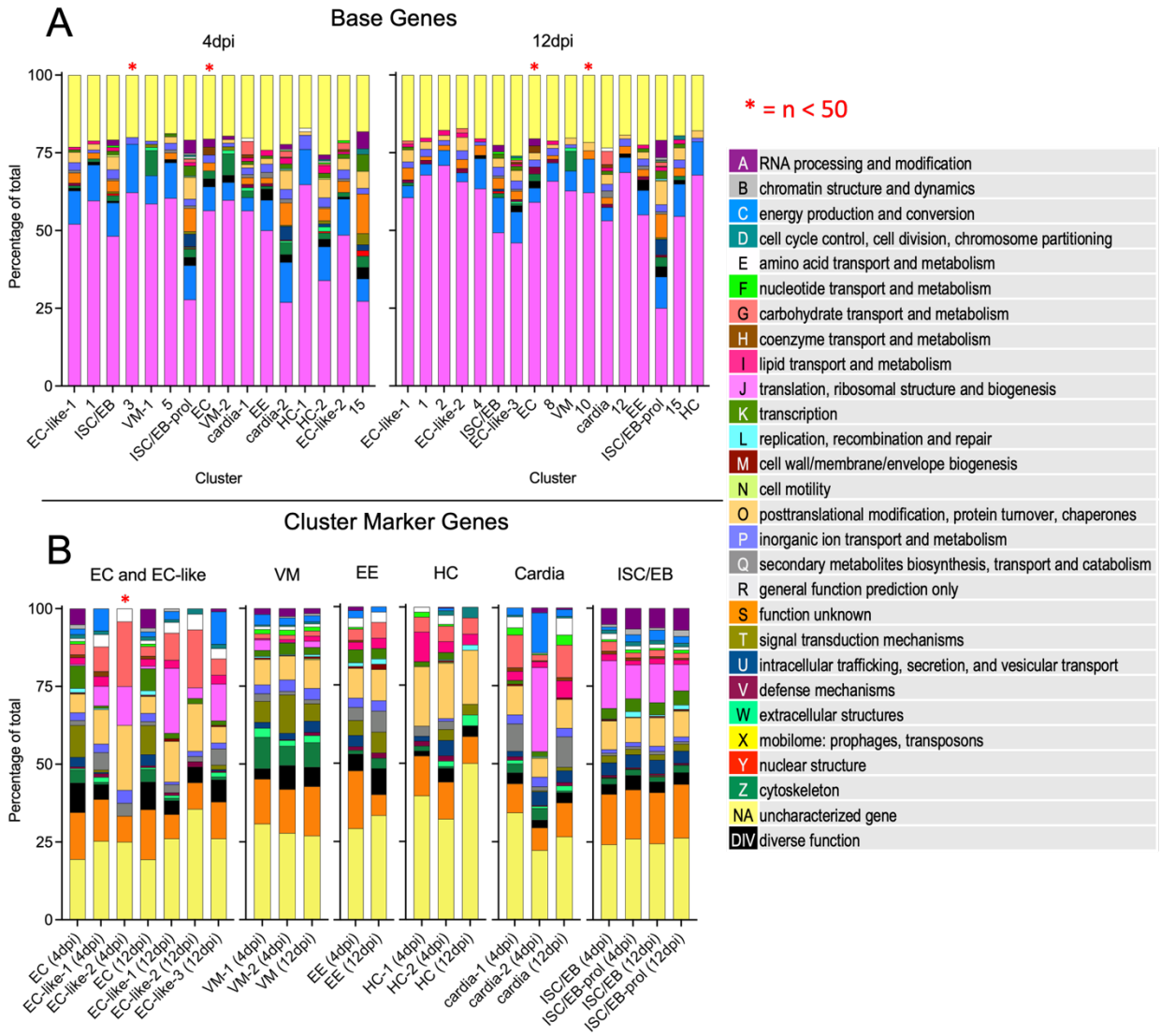



**Figure 3.4. COG profiles of midgut cell populations.** Cluster of orthologous gene (COG) profiles for (**A**) base genes and (**B**) cluster marker genes were visualized as percentage of total for each cluster/population. Colors represent COG notation A-Z, NA and DIV as shown in the legend. Notation key is embedded in the figure. Where applicable, COG profiles were derived from cluster markers that are conserved between infection conditions. Red asterisks indicate n genes < 50.

and COG profiles of all significant ($p<0.05$) cluster markers with positive $\log_2$ fold-changes ($\log_2$FC) relative to the other clusters (**Figure 3.4B**). Base gene and cluster marker gene profiles

were derived from the total population at each timepoint. Despite the varying cell-types identified thus far, we noted a level of homogeneity across base genes for each cluster, with the plurality of each profile comprised of genes involved in translation and ribosomal biogenesis (J), and energy production and conversion (C) (**Figure 3.4A**). The exceptions to this homogeneity are clusters ISC/EB-prol, cardia-2, 15 (4dpi) and ISC/EB-prol (12dpi) which all possessed fewer 'J' COGs than other clusters. Near identical cluster marker profiles reinforced our assessment that the 4dpi and 12dpi ISC/EB and ISC/EB-prol populations were comprised of the same cells (**Figure 3.4B**). EC populations from both timepoints shared near identical profiles of base genes and cluster markers (**Figure 3.4A-B**) while the profiles of EC-like populations suggest variability between their transcriptomes (**Figure 3.4B**). As expected, VM populations contain the highest proportions of cytoskeletal genes (Z) in both base gene and cluster marker profiles compared to other cell types (**Figure 3.4A-B**). Interestingly, base gene and cluster marker profiles differed dramatically between the 4 and 12dpi HC populations which reflects the distinct HC classes (oenocytoid and granulocytes) comprising the 4dpi populations, and the combination of HC classes comprising the 12dpi HC population (**Figure 3.4A-B**). The cardia-2 population possessed a notably different cluster marker profile from the cardia-1 and cardia profiles (**Figure 3.4B**), which aligns with the cardia-2 population's distinction from the other cardia markers by high expression of c-type lysozyme (**Figure 3.2C**).

*WNV vRNA is detected at varying levels in all midgut cell populations.* In addition to characterizing the cellular heterogeneity of the *Cx. tarsalis* midgut we also sought to examine WNV infection dynamics at the single-cell level. The five-prime bias of the scRNA-seq chemistry captured and allowed us to detect the WNV 5' UTR as a feature in our data. Importantly, WNV viral RNA (vRNA) was only detected in our WNV-infected samples, and was broadly detected across cell populations at both time points. (**Figure 3.5A**).

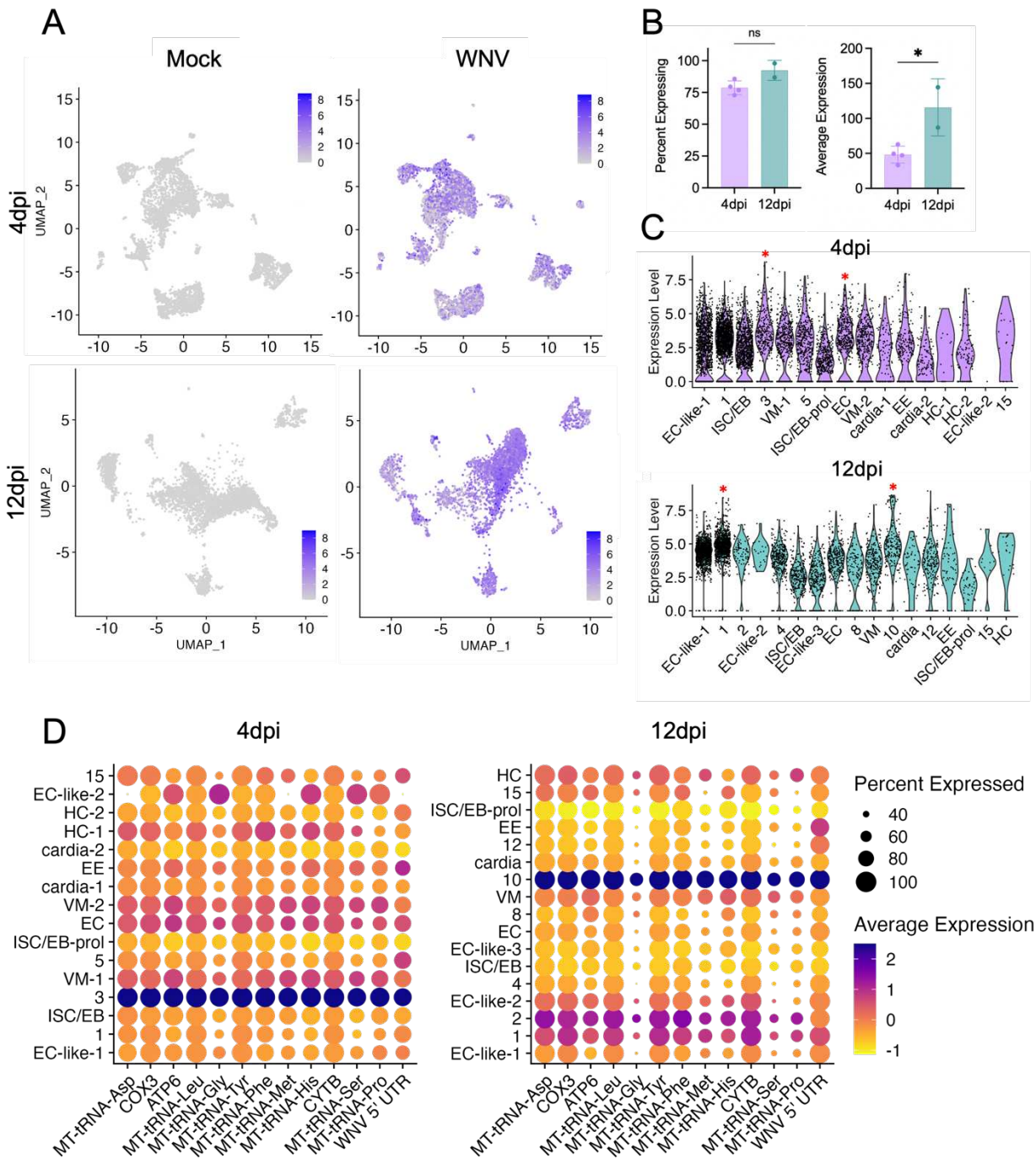



**Figure 3.5. WNV vRNA is detected in all midgut cell populations.** (**A**) Levels of WNV vRNA across the total midgut cell population at 4 and 12dpi. Purple to grey gradient indicates high to low levels of vRNA. (**B**) Average WNV vRNA level in the total population and percent of cells in the total population containing WNV vRNA calculated for each WNV-infected replicate and compared between timepoint. Significance was determined by unpaired t-test, * = $p < 0.05$. (**C**) WNV vRNA level in individual cells of distinct clusters. Red asterisks indicate WNV vRNA was identified as significantly upregulated in a given cluster. (**D**) Enrichment for identical sets of

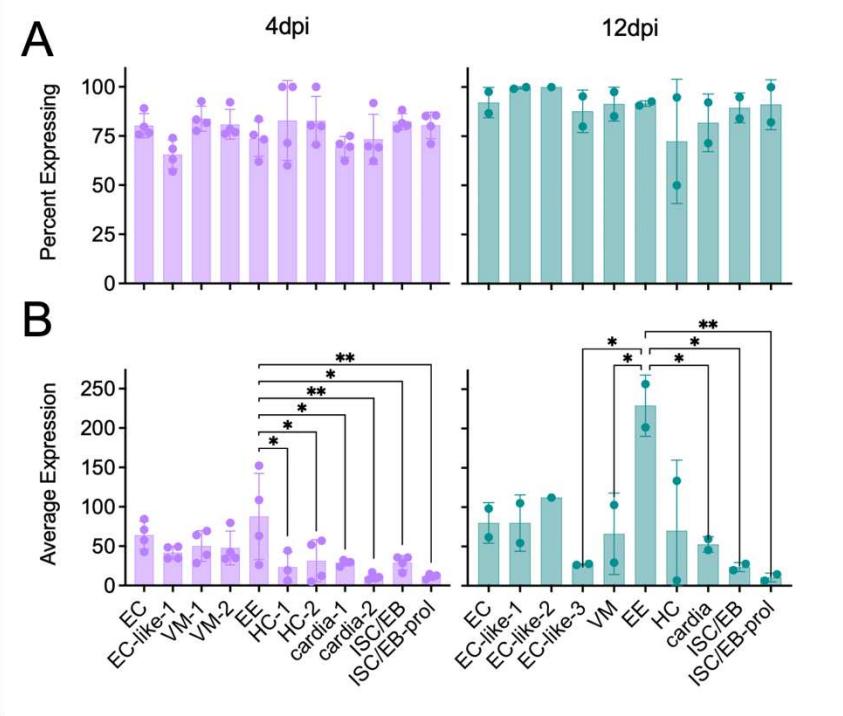
mitochondrial genes and tRNAs specific to Clusters 3 and 10. Dot size denotes percent of cells expressing each gene, color denotes scaled gene expression. “MT-“ prefix was added to tRNA gene names in this figure for clarity. Mitochondrial tRNAs were not included in mitochondrial gene estimation for QC filtering. Plots B, C, D, and E were derived from only WNV-infected samples.

Within the WNV-infected replicates we compared the percent of cells with detectable vRNA (calculated as percent expressing) and the average vRNA level (calculated as average expression) in the total population for each timepoint (**Figure 3.5B**). We saw no significant difference in the total percent of WNV-infected cells between timepoints (~75% and ~90% at 4dpi and 12dpi respectively), but a significant increase in average total vRNA level by 12dpi (**Figure 3.5B**). In our WNV-infected samples (replicates within time points combined), we identified cells containing variable levels of vRNA in all clusters except EC-like-2 at 4dpi (**Figure 3.5C**). However, for clusters 3 and EC (4dpi) and 1 and 10 (12dpi) vRNA was identified as a significant cluster marker with $log_2$FCs (relative to the other clusters) of 1.88, 0.50, 0.83, and 2.35 respectively. Clusters 3 (4dpi) and 10 (12dpi) contained the highest average levels of vRNA at both time points (**Figure 3.5C**). Lack of canonical markers precluded us from identifying their cell-type, however, comparison of the top conserved features between them revealed that they are the same cell type (**Figure 3.5D**). Additionally, enrichment for mitochondrial genes and tRNAs suggests these cells are in states of increased energy demand or stress (**Figure 3.5D**). We noted minimal FADD expression across all clusters – confirming that cell death is neither driving clustering nor causing the upregulation of mitochondrial genes and tRNAS in clusters 3 (4dpi) and 10 (12dpi) (**Figure S3.5**).

*WNV vRNA levels do not significantly differ between populations of known cell types.* Next, we sought to compare the presence of vRNA in our known cell types; EC and EC-like, VM, EE, HC, cardia, ISC/EB, and ISC/EB-prol. The EC-like-2 (4dpi) cluster was comprised of fewer than 5 cells in the WNV-infected condition and thus excluded from this comparison. In WNV-infected

replicates, we compared the percentage of cells within each known cell population containing vRNA (**Figure 3.6A**)

**Figure 3.6. Examining WNV vRNA levels in populations of known cell types.**




(**A**) Comparing the percent of cells in each cell-typed population that contain WNV vRNA for each WNV-infected replicate. (**B**) Comparing average WNV vRNA level in cell-typed populations for each WNV-infected replicate. All plots derived from only WNV-infected samples. Significance determined by one-way ANOVA, * = $p < 0.05$, ** = $p < 0.005$.

and average vRNA levels (**Figure 3.6B**). We observed no significant differences in percent expression between known cell types at either timepoint, with an average of ~60% of cells containing vRNA (**Figure 3.6A**). The EE population had significantly higher average expression of vRNA than the HC-1, HC-2, cardia-1, cardia-2, ISC/EB, and ISC/EB-prol populations at 4dpi, and EC-like-3, VM, cardia, ISC/EB, and ISC/EB-prol populations at 12dpi (**Figure 3.6B**).

*Identification of genes associated with WNV infection at the whole-tissue and single-cell level.* Bulk-RNA sequencing comparing WNV-infected to uninfected *Cx. tarsalis* midguts has not yet been described and, as such, we performed a pseudo-bulk differential expression (DE)

analysis to identify differentially expressed genes associated with mock and WNV-infected midguts at the whole-tissue level. We identified six significant differentially expressed genes (DEGs) between mock and infected conditions at 4dpi. Homocysteine S-methyltransferase, DMAS1 (aldo-keto reductase), and GBE1 (deltamethrin resistance-associated gene) were upregulated in response to WNV infection while BCAN (c-type lectin), uncharacterized gene11056, and a serine protease gene were downregulated (**Figure 3.7A**).

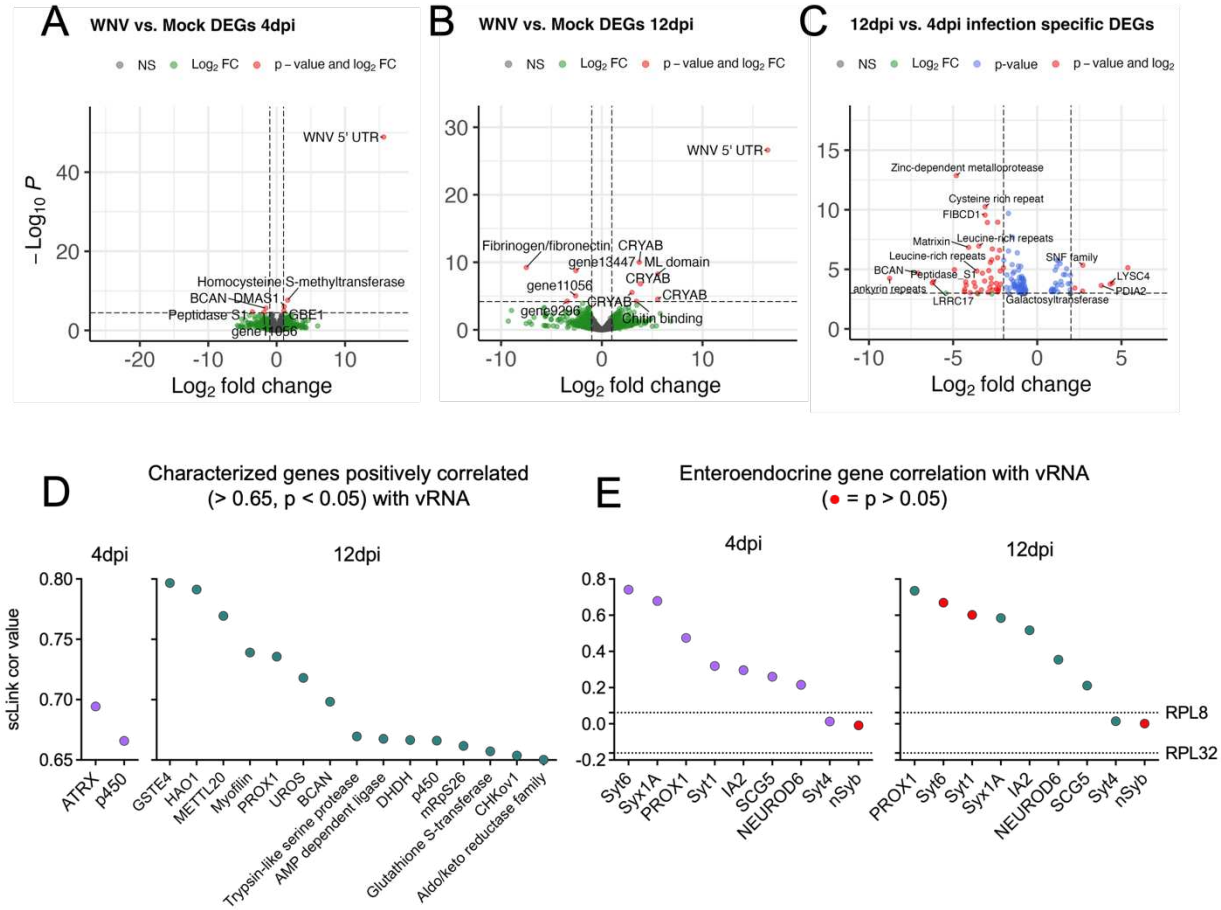



**Figure 3.7. Identifying genes upregulated in response to WNV infection with differential expression and correlation analyses.** DEGs associated with WNV-infected condition when compared to the mock condition at 4dpi (**A**) and 12dpi (**B**) were identified with DESeq2. Negative $log_2FC$ indicates downregulation and positive $log_2FC$ indicates upregulation associated with WNV infection. (**C**) Genes upregulated during WNV infection at 4dpi (negative $log_2FC$) and 12dpi (positive $log_2FC$). (**D**) Characterized genes significantly correlated with WNV vRNA at 4dpi (purple) and 12dpi (teal), with correlation values of >0.65. Only significant ($p < 0.05$) relationships shown. (**E**) Correlation of enteroendocrine specific genes and WNV vRNA at 4dpi (purple) and 12dpi (teal). Red points denote nonsignificant correlation. Dotted lines denote

RPL8 and RPL32 housekeeping reference genes. Panels C-E were derived from only WNV-infected samples.

At 12dpi we identified 10 significant DEGs; an ML (MD-2 related lipid recognition) domain containing gene, four CRYAB (heat shock protein) genes, and a chitin binding domain containing gene were upregulated in response to WNV infection (**Figure 3.7B**). Three uncharacterized genes (gene13447, gene11056, gene9296) and fibrinogen/fibronectin were downregulated in response to WNV infection (**Figure 3.7B**). DEGs between mock and WNV-infection identified by this method differed for each timepoint and, as such, we next examined DEGs between WNV-infection at both timepoints. We found many significantly up and downregulated DEGs, and that several leucine rich repeat containing genes were upregulated at 4dpi, and that host immune gene LYSC4 was upregulated at 12dpi (**Figure 3.7C**).
To further examine genes associated with WNV infection we performed a gene correlation on normalized counts for the top 500 variable genes (genes that have variation in expression across all cells) for each timepoint and determined significance via bootstrapping (**Figure 3.7D**, **Table S3.1**). We then extracted the genes correlated (>0.65) with vRNA and visualized characterized genes (**Figure 3.7D**). Transcription regulator ATRX was strongly correlated with vRNA at 4dpi along with a cytochrome p450 gene and several uncharacterized genes (**Figure 3.7D, Table S3.1**). At 12dpi, GSTE4, HAO1, METTLE20, PROX1, UROS, BCAN, DHDH, and CHKov1 were strongly correlated with vRNA along with serine protease, AMP dependent ligase, cytochrome p450, mitochondrial ribosomal S26, glutathione S-transferase, and aldo/keto reductase family genes (**Figure 3.7D, Table S3.1**).

Upon observing that PROX1, the canonical marker for EE cells, was significantly positively correlated with vRNA at 12dpi, we examined the correlation between vRNA and several previously described (**Figure 3.3A-B**) neuroendocrine genes (**Figure 3.7E, Figure 3.3A-B**). Two previously described housekeeping genes, RPL8 and RPL32, were validated as

having broad expression throughout the total population and used to both confirm that the high prevalence of vRNA in these populations was not confounding the results, and provide a visual reference for a biologically insignificant correlation value (**Figure 3.7E**) (129). At 4dpi PROX1, Syt6, and Syx1A, and at 12dpi PROX1, IA2, and Syx1A have strong positive correlations with vRNA (**Figure 3.7E**).

*Characterization of the midgut immune response to WNV infection at the whole-tissue and single-cell level.* While previous work demonstrated an increase in hemocyte proliferation upon bloodmeal ingestion and infection, we did not see significant increases in the proportion of hemocyte populations associated with infection at either time point (**Figure 3.8A-B**) (72). Upon observing that no mosquito immune genes were identified as significantly upregulated in response to WNV infection by pseudo-bulk DEG and correlation analyses, we manually compared percent of cells expressing and expression level of key immune genes, as identified in other studies (65,121,130–132). We identified orthologs for DOME, NANOS1, MYD88, IMD, AGO2, R2D2, STAT5B, Cactus, PIAS1, SUMO2, LYSC4, MARCH8, PIWIL1, PIWIL2, DICER2, and NFKB1 in the *Cx. tarsalis* genome and found no significant differences in the percent of cells expressing and average expression and of these genes at either time point (**Figure 3.8C-D**). Next, we examined the relationships between expression of these immune genes and vRNA at the single-cell level in the WNV-infected population (**Figure 3.8E**). Interestingly, almost all genes were significantly positively correlated with vRNA at both timepoints (**Figure 3.8E**). To further confirm these findings, we visualized the relationship of the four most highly correlated immune genes (IMD, PIWIL1, PIAS1, and DOME) with vRNA in individual cells for both timepoints combined, and compared the expression level of each immune gene in both mock and WNV-infected conditions for both timepoints combined (**Figure 3.8F-G, Figure S3.7**). We observed visual evidence of each correlation between these genes and vRNA, but comparable expression levels of each gene across infection conditions, confirming that while vRNA load is

correlated with specific immune genes at the individual cell level, it does not induce significant population level immune gene enrichment (**Figure 3.8F-G, Figure S3.7**).

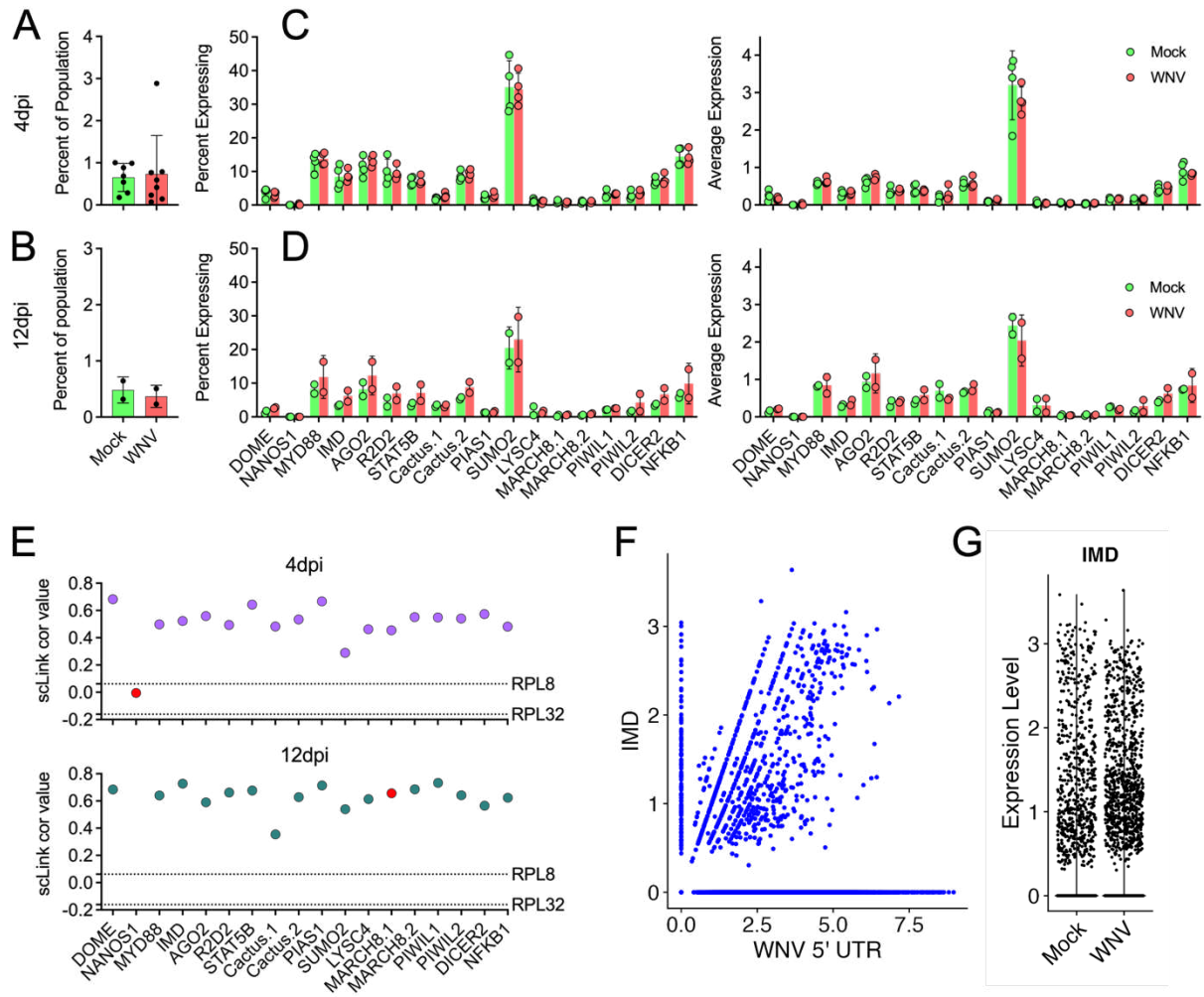



**Figure 3.8. Lack of key immune gene upregulation in response to WNV infection.** (**A-B**) Percent of the total population comprised by hemocyte populations compared between mock and WNV-infected conditions at each timepoint. Significance was determined by unpaired t-test. (**C-D**) Average expression of, and percent of cells expressing, mosquito innate immune genes compared between mock and WNV-infected replicates for each timepoint. Significance was determined by multiple unpaired t-tests. (**E**) Correlation of immune genes and WNV vRNA at 4dpi (purple) and 12dpi (teal). Red points denote non significant correlation. Dotted lines denote RPL8 and RPL32 housekeeping control/reference genes. Panel E was derived from only WNV-infected samples. (**F**) Feature scatter of IMD vs. WNV 5' UTR for both timepoints combined. (**G**) Expression level of IMD in mock and WNV-infected conditions for both timepoints combined.

### 3.3 Discussion

In this study, we sought to generate a midgut cell atlas (i.e., map of cell type and function) for *Cx. tarsalis* and characterize WNV infection of the midgut at single-cell resolution. We performed scRNA-seq on mock and WNV-infected midguts, collected at days 4 and 12 post infection. We identified and described nutrient absorptive (enterocyte), secretory (enteroendocrine), peritrophic matrix secreting (cardia), undifferentiated progenitor (intestinal stem cell/enteroblast), visceral muscle, and immune (hemocyte) cell populations (64,68,73). The distribution and proportion of each cell-type in the total population varied between timepoints, however we identified at least one cluster comprised of each cell-type at each timepoint. Several clusters were precluded from identification due to either lack of canonical markers/enrichment patterns, or origination from a single replicate. Nonetheless, we have demonstrated that single-cell sequencing of *Cx. tarsalis* midguts is feasible and that distinct cell populations can be identified and characterized using previously described canonical cell-type markers and enrichment patterns (66–68,73,76).

We detected WNV RNA (vRNA) in the majority of midgut cells at both timepoints and observed non-zero levels in all distinct cell populations. We observed a significant increase in vRNA in the total midgut by 12dpi, however the percent of infected cells did not increase significantly between timepoints. This demonstrates that the majority of midgut cells that will become infected are infected by 4dpi, while vRNA load increases as infection progresses. We observed highly similar profiles of “base genes” (i.e. genes present in ≥ 75% of a cells in a given cluster) across cell populations, suggesting that WNV is capable of productively replicating in all cell types, potentially due to the heterogeneity of base genes. The high percentage of WNV-infected cells and permissibility of all cell populations to infection supports previous work demonstrating the extreme competence of *Cx. tarsalis* as a WNV vector (10,15,118,133).

Interestingly, while WNV infected almost all midgut cell populations, cluster 3 and enterocytes (EC) at 4dpi, and clusters 1 and 10 at 12dpi were significantly enriched for vRNA. Of these, clusters 3 and 10 had the most biologically interesting enrichment for vRNA. These high WNV-expressing clusters are associated with very few (<15) defining cluster markers, precluding us from identifying their cell-type. However, they clearly possess identical sets of cluster markers which implies they are the same cell-type. These shared cluster markers are comprised entirely of mitochondrial genes and mitochondrial tRNAs, suggesting they are in a heightened state of cell stress and/or energy production. Two possible interpretations of this can be made; either WNV was able to replicate to higher levels in cells enriched for mitochondrial genes, or high levels of viral replication induced significant stress and/or energy production responses in these populations (134,135). However, the existence of these clusters, and their enrichment for mitochondrial genes, in the mock condition does not align with the latter interpretation. Several previous studies have demonstrated that positive sense single stranded RNA viruses like DENV, and SARS-CoV-2 modulate mitochondrial dynamics to facilitate replication and/or immune evasion (135–137). High levels of vRNA observed in cell populations enriched for mitochondrial genes could potentially be the result of beneficial interactions between WNV and the mitochondria. Additionally, a study in *Lepidoptera* (moths and butterflies) purported that enrichment of mitochondrial genes is associated with insect stress resistance (134). Stress resistance responses modulate cell viability and it is known that the maintenance of cell viability is central to productive WNV infection (138,139). Moreover, we demonstrated that several heat shock genes (known to be protective against cell stress) were significantly upregulated in WNV infected midguts at 12dpi, further suggesting a relationship between WNV infection and the stress response (140–142). Therefore, we hypothesize that cell stress impedes WNV replication while enrichment for factors that are protective against cell stress facilitates WNV replication.

Our enteroendocrine (EE) cell populations only contained one of the previously described mosquito neuropeptides/neuropeptide receptors, tachykinin receptor, and only in a small subset of cells (66–68,123,124). This could be due to the known bias of single-cell RNA-seq towards highly expressed genes, or due to *Cx. tarsalis* EE cell secretion of yet uncharacterized neuropeptides. High expression of NEUROD6 – a neurogenic differentiation factor frequently found in neurons involved in behavioral reinforcement – in EE populations at both timepoints supports the hypothesis that additional/uncharacterized neuropeptides may be present in *Cx. tarsalis* EE cells (128). Interestingly, we found that EE populations contained more vRNA than all other cell-typed populations, significantly so for more than half of the cell-typed populations at each timepoint. The lack of any significant differences between the percent of cells in each population containing vRNA confirm that increased vRNA in EE cells is due to enhanced replication and not increased susceptibility to infection. Further, PROX1 (the canonical marker for EE cells) and select neuroendocrine and vesicle docking genes enriched in EE cells, were strongly positively correlated with vRNA at both timepoints, supporting our hypothesis that EE cells serve as sites of enhanced WNV replication during midgut infection in *Cx. tarsalis*. A previous study characterizing ISC dynamics in response to DENV in the *Ae. aegypti* midgut found that ISC proliferation increased refractoriness to infection, suggesting that cell renewal is an important part of the midgut immune response (143). Interestingly, we observed that proliferating ISC/EB populations in *Cx. tarsalis* consistently had the lowest levels of vRNA, along with the cardia-2 population at 4dpi. These three cell populations expressed notably fewer base genes involved in translation, ribosomal structure, and biogenesis. These findings taken together suggest that the transcriptional state of proliferating ISC/EBs impedes virus replication. The factors that determine cell-type specific enhancement or suppression of WNV replication are not currently known, but could include more efficient evasion of antiviral pathways, more efficient mechanisms of midgut escape/dissemination, or abundance of proviral genes. Moreover, while most/all cells within the *Cx. tarsalis* midgut are susceptible to WNV

infection, we observed modest differences in virus replication efficiency that appeared to occur in a cell-type specific manner, with EE cells being the most permissive and proliferating ISC/EB cells being the most refractory.

We identified granulocyte and oenocytoid hemocyte (HC) populations - known to play important roles in the mosquito innate immune response - at both timepoints, allowing us to characterize the cellular immune component of mosquito midguts (64,72). We found no obvious evidence of immune gene upregulation or immune cell proliferation in the total infected population compared to mock – suggesting little to no immune activation in the midgut upon WNV infection. This was surprising given the importance of the midgut as a site of innate immune activation (10,15,64,116). Although scRNA-seq of *Cx. tarsalis* after WNV infection has not been described, several previous studies in *Ae. aegypti* and *Cx. pipiens* have noted significant upregulation of IMD and Toll pathway genes in response to viral infection and highlighted that the innate immune response in mosquitoes is a strong determinant of vector competence (140,144,145). The absence of a notable immune response to WNV infection in *Cx. tarsalis* could be a determinant of the vector's extreme susceptibility and competence (10,15,118,133). However, we saw that in individual cells, most immune genes had some degree of significant positive correlation with vRNA suggesting that, while WNV infection does not cause significant enrichment of these genes in the total population, WNV infection and replication influences the expression of these genes at the single-cell level. This finding highlights that scRNA-seq is a powerful tool for characterizing infection dynamics that are not apparent when looking at the population average.

Our inability to detect certain genes (i.e. neuropeptide genes and additional canonical markers we would expect to see) could be due to the low percentage (~30%) of reads mapped to the *Cx. tarsalis* genome. The low mapping percentage was due to a predominance of reads that were too short to map. Future scRNA-seq studies in *Culex* mosquitoes could potentially

benefit from adjusting the fragmentation time recommended by 10X Genomics. Additionally, a multitude of genes detected in our dataset remain uncharacterized due to a lack of appropriate orthologs. This could be explained by the evolutionary divergence between *Cx. tarsalis* and the species from which most gene orthologs were derived; *Cx. quinquefasciatus* and *Ae. aegypti* which diverged 15-22 million years ago (MYA) and 148-216 MYA respectively (120). High levels of vRNA in specific cell-types imply that replication is occurring/has occurred but it is important to note that the presence of vRNA is not analogous to active viral replication (e.g., the presence of vRNA could be the result of phagocytosis of an infected cell). Early WNV infection in the *Cx. tarsalis* midgut could be further studied via immunofluorescence assays using cell-type specific RNA probes in whole midguts, putting the findings described here into a spatial context. Additionally, future studies could use qRT-PCR, focusing on our top genes of interest, to measure expression kinetics and levels following WNV infection in *Cx. tarsalis* and other relevant vectors. Finally, RNA trajectory analysis could be applied to this dataset to shed light on the functional states of the cells that comprise each distinct population.

The work presented here demonstrates that WNV is capable of infecting all midgut cell-types in *Cx. tarsalis*, but replicates to higher levels in enteroendocrine cells – implying a relationship between WNV and the enteric endocrine system that should be further explored. Our findings strongly suggest interplay between WNV infection and the cell stress response, and we have provided evidence that WNV infection of the *Cx. tarsalis* midgut results in the upregulation of cell stress associated genes. We observed mild to no upregulation of key mosquito immune genes in the midgut as a whole, however, we show that immune gene expression is significantly correlated with WNV vRNA level within individual cells. Additionally, we have generated a midgut cell atlas for *Cx. tarsalis* and, in doing so, contributed to the field's understanding of how WNV establishes infection in this highly competent vector.

### 3.4 Methods

*Virus.* All infections were performed with a recombinant barcoded WNV (bcWNV) passage 2 stock (epidemic lineage I strain, 3356) grown on Vero cells. Titer for the stock was determined by standard Vero cell plaque assay (50).

*Mosquito infection.* Mosquito studies were conducted using laboratory colony-derived *Cx. tarsalis* mosquitoes (>50 passages) WNV infections in mosquitoes were performed under A-BSL3 conditions. Larvae were raised on a diet of powdered fish food. Mosquitoes were maintained at 26°C with a 16:8 light:dark cycle and maintained at 70–80% relative humidity, with water and sucrose provided ad libitum. *Cx. tarsalis* mosquitoes were transferred to A-BSL3 conditions 48 hours prior to blood feeding, and dry starved 20-24 hours before blood feeding. Seven days after pupation (6-7 days after emergence) mosquitoes were exposed to an infectious bloodmeal containing a 1:1 dilution of defibrinated calf's blood and bcWNV stock diluted in infection media (Dulbecco's Modified Eagle's Medium, 5% penicillin-streptomycin, 2% amphotericin B, and 1% fetal bovine serum (FBS)) for a final concentration of 3-6e$^7$ PFU/mL, or a mock bloodmeal containing a 1:1 dilution of defibrinated calf's blood and infection media. All bloodmeals were provided in a hog's gut glass membrane feeder, warmed by circulating 37°C water. Following 50-60 minutes of feeding, mosquitoes were cold-anesthetized, and engorged females were separated into cartons and maintained on sucrose.

*Collection of mosquito tissues.* At indicated time points, mosquitoes were cold-anesthetized and transferred to a dish containing Sf900III insect cell culture media (Gibco) with 5% FBS. Midguts were dissected, and transferred to tubes containing 500uL Sf900III media + 5% FBS and kept on ice for the duration of dissections. Ten pooled midguts per sample/tube were collected for dissociation and sequencing.

*Midgut dissociation and single-cell suspension preparation.* A dissociation buffer containing *Bacillus licheniformis* protease (10mg/mL) and DNAse I (25U/mL) was prepared in Sf900III

media (Gibco). Pooled midguts were resuspended in dissociation media, transferred to a 96-well culture dish, and triturated with a p1000 pipet at 15-20 minute intervals for 105 minutes. At each interval 100-125µL containing dissociated single-cells was collected (with replacement) from the top of the dissociation reaction and transferred to 25mL of Sf900III + 5% FBS on ice. Dissociation reactions were kept covered at 4°C between trituration. Upon complete tissue dissociation the entire remaining volume of each reaction was transferred to Sf900III + 5% FBS on ice. Collection tubes with dissociated cells were centrifuged at 700xg for 10 minutes at 4°C, resuspended in 500uL of Sf900III + 5% FBS, and passed through a 40um small volume filter (PluriSelect). Immediately prior to loading on the Chromium Controller, cell suspensions were spun down at 700xg for 10 minutes and washed twice in 1mL PBS + 0.04% bovine serum albumin (BSA), and resuspended in 50µL of PBS + 0.04% BSA. Cell concentration was determined using the Countess II Automated Cell Counter (Thermo Fisher Scientific) and the appropriate cell suspension volume (target recovery of 10,000 cells) was loaded on the Chromium controller.

*Gel Bead-In Emulsions (GEM) generation and cDNA synthesis.* GEM generation and cDNA synthesis were performed using the Next GEM Single Cell 5' GEM kit v2 (PN-1000266) and Next GEM Chip K Single Cell Kit (PN-1000286) (10X Genomics). Reactions for GEM generation were prepared according to the Chromium Next GEM Single Cell 5' Reagent Kit v2 (dual index) user guide with one alteration; a primer specific to the WNV envelope region of the genome was added at a concentration of 10nM (7.5µl – displacing 7.5µl of the total $H_2O$ added to each reaction). GEMs were generated using the 10X Chromium controller X series and cDNA was synthesized according to the Chromium Next GEM Single Cell 5' Reagent Kit v2 (dual index) user guide.

*Library preparation and sequencing.* Libraries were prepared using the Next GEM single cell 5' v2 library construction kit and Dual Index Kit TT Set A (10X Genomics, PN-1000190 and PN-

1000215 respectively). Library construction was carried out according to the Chromium Next GEM Single Cell 5' Reagent Kit v2 (dual index) user guide. Library concentration was determined by KAPA Library Quantification Kit (Roche). Libraries were then diluted to 15nM, pooled by volume, and sequenced at the CU Anschutz Genomics and Microarray core on the NovaSeq 6000 (150x10x10x150) (Illumina) at a target coverage of $4.0e^8$ read pairs per sample, equating to 40,000 read pairs per cell. Average sample coverage and cell recovery per sample was $5.3e^8$ read pairs and 2417.9 cells respectively (**Supplemental Tables S3.2, S3.3**).

*Reference generation and sample processing with Cell Ranger.* The gene feature files associated with the *Cx. tarsalis* and WNV genomes were converted to gene transfer format (gtf) using AGAT (v1.0.0) prior to being filtered with the CellRanger (v7.0.1) mkgtf function (146). An 'MT-' prefix was manually added to all non-tRNA features located in the mitochondrial chromosome of the *Cx. tarsalis* genome. The contents of the filtered WNV genome feature file, and fasta file, were appended to the *Cx. tarsalis* feature and fasta files respectively, and run through CellRanger::mkref. All sequencing data were processed and mapped to the aforementioned *Cx. tarsalis* reference genome using CellRanger::count with the following parameters: --include-introns=true \ --expect-cells=10000.

*Quality control and Seurat workflow.* Cell Ranger output files were individually read into RStudio (RStudio - v2023.09.0+463, R – v4.3.2) as SingleCellExperiment objects using the singleCellTK package (v2.12.0). Doublet identification and ambient RNA estimation were performed with singleCellTK::runCellQC using the algorithms "scDblFinder" and "DecontX" respectively. Samples were filtered for doublets and ambient RNA contamination by keeping cells with the following metrics: decontX_contamination < 0.6, scDblFinder_doublet_score < 0.9. Samples were then converted to Seurat objects and merged into one Seurat object for each timepoint, with columns pertaining to sample of origin and infection condition added to the object metadata prior to processing as described in the Seurat (v4.3.0.1) guided clustering tutorial (147). Briefly,

mitochondrial gene percentage for each cell was calculated, and cells with the following metrics were retained: nFeature_RNA > 100, nFeature_RNA < 2500, percent_mt < 25. Percent of cells retained after QC for each sample can be found in **Supplemental Table S3.4**. Features were log normalized, variable features were identified, the top 2000 variable features were scaled, a principle component (PC) analysis dimensionality reduction was run, and the number of PCs needed to adequately capture variation in the data was determined via elbow plot. Nearest neighbors were computed, and appropriate clustering granularity was determined with Clustree (v0.5.0). Uniform Manifold Approximation and Projection (UMAP) dimensional reduction was performed, and clusters visualized with UMAP reduction. Cluster markers were identified with Seurat::FindConservedMarkers using default parameters and infection condition as the grouping variable. For clusters that had very few conserved markers we split the dataset by infection condition and used Seurat::FindMarkers on clusters in either the mock or WNV-infected condition. WNV vRNA as a cluster marker was identified by splitting the dataset by infection condition and using Seurat::FindAllMarkers on the WNV infected samples. In all cases where calculations were performed on individual replicates or individual conditions, the merged Seurat object was split by sample or condition using Seurat::SplitObject so that calculations performed on subsets of the data or individual replicates were derived from a dataset that had been normalized as one. Percent expression and average expression were calculated with scCustomize::Percent_Expressing (v1.1.3) and Seurat::AverageExpression respectively. Feature expression levels were visualized with Seurat::FeaturePlot and Seurat::VlnPlot.

*Pseudo-bulk differential expression analysis with DESeq2.* Pseudo-bulk dataset was generated and differential expression analysis performed as described by Khushbu Patel's (aka bioinformagician) pseudo-bulk analysis for single-cell RNA-Seq data workflow tutorial (148). Briefly, raw counts were aggregated at the sample level using Seurat::AggregateExpression, and the aggregated counts matrix extracted and used to create a DESeq2 object (dds). The dds

object was filtered to retain genes with counts >=10 prior to running DESeq2 and extracting results for the appropriate contrast. DESeq2 results were visualized via EnhancedVolcano (v1.20.0). We identified infection associated DEGs between timepoints by performing a pseudo-bulk DE analysis between 4dpi and 12dpi WNV-infected samples and mock samples separately. We then filtered out DEGs associated only with bloodmeal consumption (genes that came up as significantly differentially expressed between timepoints in our mock condition) leaving only DEGs associated with infection.

*Gene correlation with vRNA level.* Raw gene counts for each timepoint were extracted and normalized using scLink::sclink_norm (v1.0.1) (149). Correlation matrices were then generated for the top 500 variable genes (identified during the Seurat guided clustering workflow) using scLink::sclink_corr (149). For specific neuroendocrine and immune gene correlations a vector of specific gene names was supplied in lieu of the top 500 variable genes. P-values associated with the correlation values were determined via bootstrapping.

*Ortholog identification with EggNOG Mapper.* Coding sequence (CDS) genome coordinates were extracted from the *Cx. tarsalis* gene transfer format file and corresponding genome sequences were extracted from the available fasta file using Bedtools (v2.26.0). CDS were then assigned orthologs via eggNOG-mapper v2 (web version) using default parameters (150,151). gtf file gene IDs were merged with the eggNOG output file using custom R scripts. Gene ID, ortholog seed, preferred gene name, COG category notation, PFAM information, and gene description were retained in a gene name and ID file within which information corresponding to identical gene IDs was aggregated using a custom R script. This aggregated gene name and ID file was used to assign gene names to the marker files used for cell-typing, DESeq2 results files, and scLink correlation matrices.

*Statistical analyses.* Statistical analyses were performed in GraphPad Prism version 10.0.3. Differences in vRNA level in known clusters were measured by one-way ANOVA with Tukey's

multiple comparisons test. Differences in average expression and percent expression of mosquito immune genes between mock and WNV-infected conditions were measured by multiple unpaired t-tests. Differences in hemocyte population proportion were measured by unpaired t-test. The average expression value obtained from one HC-1 replicate was identified as an outlier via zscore (4.35 standard deviations above the mean), interquartile range (>3 IQR above Q3), and chi-squared outlier test ($p<0.0001$). This value was also found to be derived from one single cell. For these reasons that replicate was excluded from the HC-1 average expression calculation in **Figure 6B**. The population proportion of one mock HC replicate (4dpi) was identified as an outlier (>3 standard deviations from the mean, >3 IQR from Q3, chi square outlier test $p < 0.001$) and removed from the comparison in **Figure 3.8A**. The HC population in said replicate comprised a notably higher proportion of the total replicate population due to an additional, replicate specific, cluster called as HC in sample Mg5c (**Figure S3.1C**). All scripts used for data processing, analysis, and visualization are available on GitHub: https://github.com/fitz-meyer/scRNA_seq_fitz.

# CHAPTER 4: SUMMARY AND FUTURE CONSIDERATIONS

## 4.1 Summary

West Nile virus (WNV) is one of the most epidemically important arthropod-borne viruses (arboviruses) in North America (1). WNV can be transmitted by several mosquito vectors of varying vector competence (11). However, it remains unknown whether vector competence influences stochastic mechanisms of virus evolution (i.e., population bottlenecks) within the mosquito. In the first study described in this dissertation we assessed the relationship between bottleneck stringency and vector competence, quantified the impact of bottlenecks on a synthetically diverse WNV population at unprecedented resolution, and examined the influence of vector competence with respect to the transmission of low frequency/rare virus variants. In this study we demonstrate that WNV diversity in the mosquito midgut increases with increasing vector competence, vector competence does not impact the diversity of transmissible virus populations, and highly competent vectors are more likely to transmit rare virus variants.

The mosquito midgut serves as a critical interface between vector and pathogen and plays a central role in host immune response, intrahost virus population dynamics, and vector competence (20,33,64). *Cx. tarsalis* is a significant vector of WNV, however, despite the public health burden imposed by this virus:vector pairing, studies of *Cx. tarsalis* physiology, and associated WNV infection dynamics, remain scarce (7,118,121). In the second study described in this dissertation we used scRNA-seq to characterize WNV infection of the *Cx. tarsalis* midgut at single-cell resolution, and identified the distinct cell populations that comprise the midgut. In this study we found that WNV infects almost all midgut cell-types by four days post-infection, and exhibits enhanced replication in enteroendocrine (EE) cells and suppressed replication in proliferating intestinal stem cells/enteroblasts (ISC/EB-prol). Further, we showed that WNV appears to elicit an immune response at the individual-cell level but does not cause tissue-wide

enrichment of immune genes. Finally, we uncovered an association between WNV infection and the cell stress response.

Collectively, the work described in this dissertation has advanced the field's understanding of how vector competence impacts virus evolution and transmission dynamics, characterized how WNV establishes infection in a primary vector, and generated a midgut cell atlas for *Cx. tarsalis* – a highly relevant vector of human viruses.

### 4.2 Future considerations

The single-cell work described herein proposes a novel relationship between WNV and enteroendocrine/neuroendocrine system in the mosquito midgut. A previous study in mice demonstrated that EE cells are capable of direct connection with neurons via cytoplasmic processes termed neuropods (152). This study also determined that infection of EE cells with monosynaptic rabies led to the infection of mucosal nerves, indicating that this neuroepithelial circuit can serve as a conduit for viruses to enter the central nervous system (CNS) (152). While many mechanisms of WNV entry to the CNS in vertebrates have been proposed, the exact mechanism remains unknown (153). Future work should explore whether enhanced WNV replication in EE cells persists in vertebrate models, and explore infection of EE cells in the gut as a potential mechanism of CNS entry.

We pursued several additional research questions that arose during the studies described in this dissertation, and the preliminary results of these experiments highlight the need for further exploration in these areas. As mentioned in section 2.3 of this dissertation, we wanted to assess whether the number of midgut cells involved in WNV infection contributed to the differences in barcode population diversity observed between *Culex* and *Ae. aegypti* mosquitoes. To do this we infected *Cx. tarsalis* (high competence for WNV) and *Ae. aegypti* (low competence for WNV) with WNV (Fort Collins 3699), collected midguts at 8 days post-infection (dpi), stained dissociated midgut cells for WNV capsid protein, and determined percent

of infected cells by flow cytometry. We found no difference in the percent of infected cells between species, suggesting that cell involvement is not a determinant of WNV population diversity in the midguts of these species (**Figure 4.1A, C, E**).

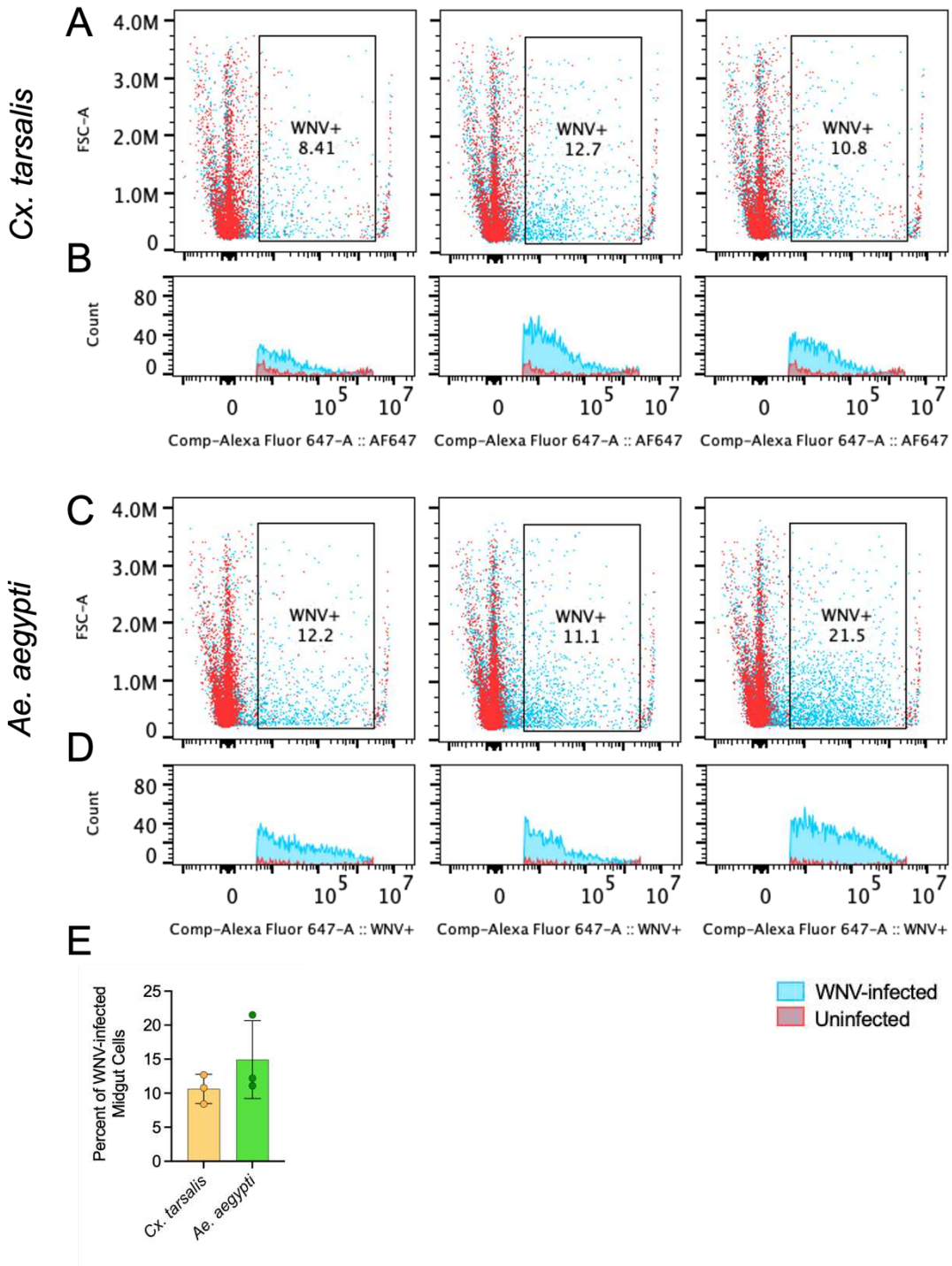


**Figure 4.1. Percent of cells infected by WNV at 8 days post-infection in *Cx. tarsalis* and *Ae. aegypti* midguts. (A, C)** Percent of cells positive for WNV capsid protein denoted within gate. Red = uninfected stained cells, blue = infected stained cells. (**B, D**) Histograms showing mean fluorescent intensity of true WNV infected cells (blue) and uninfected cells within the infected cell gate (red). (**E**) Comparison of percent infected midgut cells between *Cx. tarsalis* (orange) and *Ae. aegypti* (green). Only significant comparisons shown. Significance determined by unpaired t-test.

However, examining the presence of WNV capsid protein alone does not indicate active infection of a cell. Further, the anti-WNV-capsid antibody employed in this experiment was only validated for use in western blots and exhibited high levels of nonspecific binding and background fluorescence, making it difficult to distinguish true infection versus background (**Figure 4.1B, D**). In the interest of reducing the background fluorescence and measuring active intracellular viral replication, replications of this experiment should be performed with an anti-double-stranded RNA (a by-product of viral replication) antibody such as J2, that has been validated for use in flow cytometry.

A previous study with bcWNV in avian hosts in our lab highlighted the importance of polyinfection (i.e., infection of a single cell with multiple unique genomes) in contributing to virus population diversity in a host-dependent manner (62). Additional work in our lab demonstrated, using a barcoded Zika virus, that individual infectious units can be composed of multi-genome aggregates, elucidating a potential mechanism through which polyinfection could occur. Further, it has been hypothesized that infected individual cells, which serve as sites of cellular bottlenecks, are potentially the most important compartment with respect to virus population dynamics (154). For these reasons we sought to both examine the occurrence of polyinfection in WNV-infected mosquito cells and assess whether WNV forms multi-genome aggregates.

To examine polyinfection in mosquitoes we infected CT cells (*Cx. tarsalis* derived cultured cells) with bcWNV, stained cells for WNV capsid protein, sorted cells into samples containing one, three, or five individual cells, and performed amplicon sequencing of the

barcode region to examine barcode diversity as previously described in this dissertation. We saw that several samples comprised of individual cells contained two unique barcodes, suggesting that polyinfection can occur in cultured mosquito cells (**Figure 4.2**).

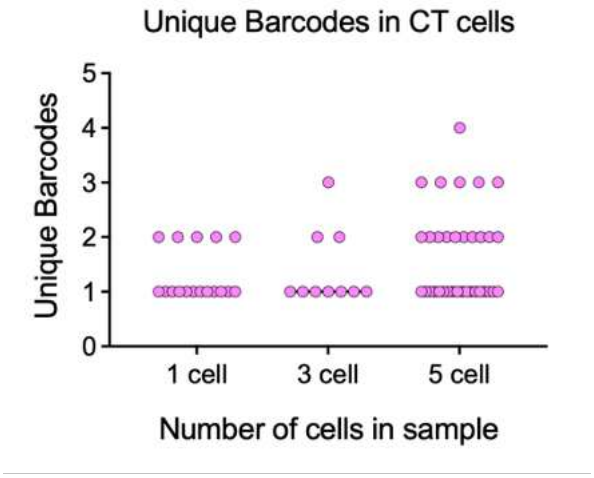



**Figure 4.2. Number of unique barcodes detected in samples containing one, three, or five individual cells.**

This preliminary finding has important implications for mechanisms of WNV evolution in mosquitoes but requires significant reproduction and troubleshooting in both *in vitro* and *in vivo* systems, as we only performed one iteration of this experiment and were unable to successfully measure polyinfection in cells derived from dissociated mosquito tissues.

All mosquito infections for the scRNA-seq experiments described herein were performed with barcoded WNV. Additionally, the 10X Genomics platform employs cell-specific barcodes that allow the Cell Ranger analysis software to bin transcripts by cell of origin. Therefore, future work in this area could confirm the occurrence of polyinfection in *Cx. tarsalis* by computationally determining whether multiple unique barcode-containing reads originated from a single cell.

To assess whether WNV forms multi-genome aggregates we performed standard Vero cell plaque assay on the tissues and secretions obtained from each of the three bcWNV-

infected mosquito species previously described herein. We then picked individual plaques and characterized plaque barcode diversity via amplicon sequencing of the barcode region. We found that the majority of picked plaques from all species contained >1 unique barcode and that, in *Ae. aegypti*, plaques derived from midgut samples were significantly more complex than plaques derived from salivary gland and saliva samples (**Figure 4.3**). These findings suggest

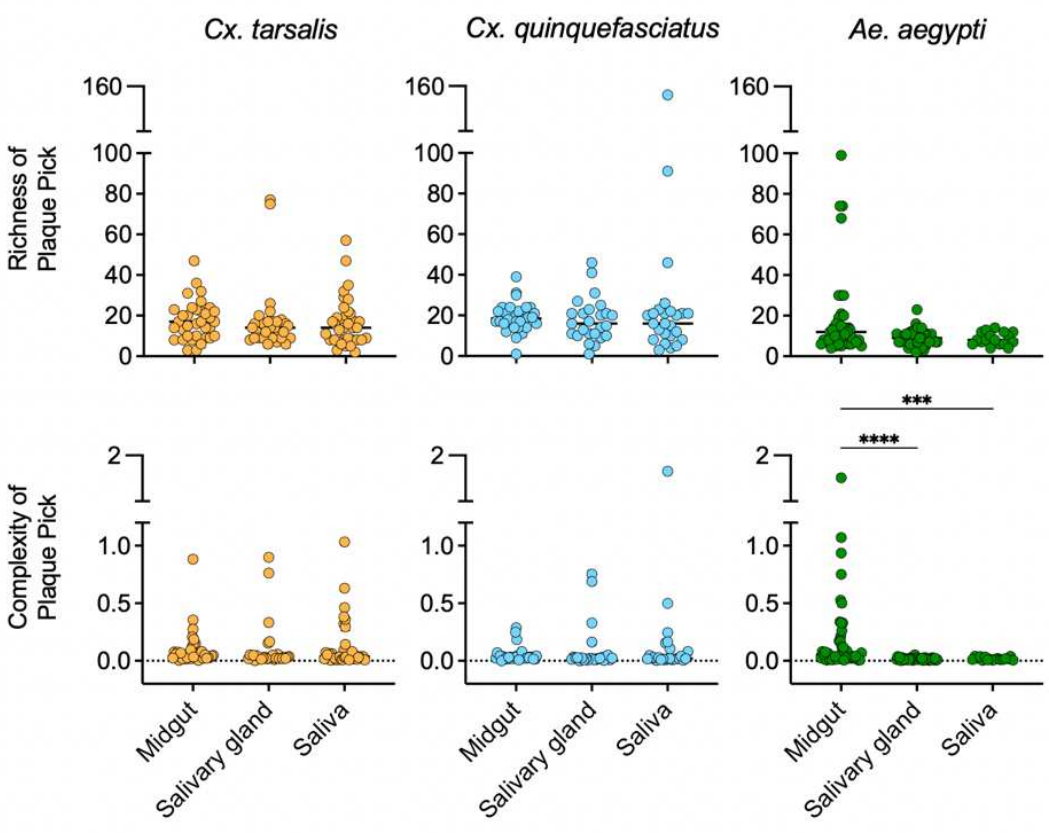



**Figure 4.3. Barcode population richness and complexity in individual plaques.** Barcode population diversity in plaques derived from midgut, salivary gland, and saliva samples obtained from *Cx. tarsalis* (orange), *Cx. quinquefasciatus* (blue), and *Ae. aegypti* (green). Only significant comparisons shown. Significance determined by Kruskal-Wallis test. *** = $p<0.0005$, **** = $p<0.0001$. Only significant comparisons shown.

that individual infectious WNV units can be composed of multiple unique genomes, with implications for tissue-specific increases in infectious unit diversity. However, several picked plaque samples were found to contain greater numbers of unique barcodes than were detected

in their parent whole-tissue samples. Two possible interpretations of this finding can be made. First, barcode diversity in individual plaques could appear inflated compared to whole-tissue parent samples due to a higher ratio of sequencing coverage to barcodes. This interpretation aligns with what we know about sequencing saturation of the whole-tissue samples (**Figure 2.1B**). Second, the presence of significant background virus levels in the plaque assay could artificially increase plaque diversity. Future work in this area could benefit from including background controls picked from non-plaque areas, and re-sequencing whole-tissue parent samples at increased coverage to confirm the diversity of individual plaques.

## REFERENCES


1. Artsob H, Gubler DJ, Enria DA, Morales MA, Pupo M, Bunning ML, et al. West Nile Virus in the New World: Trends in the Spread and Proliferation of West Nile Virus in the Western Hemisphere. Zoonoses Public Health. 2009 Aug 9;56(6–7):357–69.
2. Gubler DJ. The Continuing Spread of West Nile Virus in the Western Hemisphere. Clinical Infectious Diseases. 2007 Oct 15;45(8):1039–46.
3. Moudy RM, Meola MA, Morin LLL, Ebel GD, Kramer LD. A Newly Emergent Genotype of West Nile Virus Is Transmitted Earlier and More Efficiently by Culex Mosquitoes. 2007.
4. Duggal NK, Langwig KE, Ebel GD, Brault AC. On the Fly: Interactions between Birds, Mosquitoes, and Environment That Have Molded West Nile Virus Genomic Structure over Two Decades. Vol. 56, Journal of Medical Entomology. Oxford University Press; 2019. p. 1467–74.
5. BERNARD KA, KRAMER LD, EBEL GD, CARRICABURU J, YOUNG D. GENETIC AND PHENOTYPIC VARIATION OF WEST NILE VIRUS IN NEW YORK, 2000–2003. Am J Trop Med Hyg. 2004 Oct 1;71(4):493–500.
6. Ronca SE, Ruff JC, Murray KO. A 20-year historical review of West Nile virus since its initial emergence in North America: Has West Nile virus become a neglected tropical disease? PLoS Negl Trop Dis. 2021 May 6;15(5):e0009190.
7. https://www.cdc.gov/westnile/statsmaps/data-and-maps.html [Internet]. 2023. Center for Disease Control and Prevention, West Nile virus.
8. Sejvar JJ. West Nile Virus Infection. Microbiol Spectr. 2016 May 6;4(3).
9. Busch MP, Wright DJ, Custer B, Tobler LH, Stramer SL, Kleinman SH, et al. West Nile Virus Infections Projected from Blood Donor Screening Data, United States, 2003. Emerg Infect Dis. 2006 Mar;12(3):395–402.
10. Fitzmeyer EA, Gallichotte EN, Weger-Lucarelli J, Kapuscinski ML, Abdo Z, Pyron K, et al. Loss of West Nile virus genetic diversity during mosquito infection due to species-dependent population bottlenecks. iScience. 2023 Oct;26(10):107711.
11. Turell MJ, O'guinn ML, Dohm DJ, Jones JW. Vector Competence of North American Mosquitoes (Diptera: Culicidae) for West Nile Virus [Internet]. Vol. 38, J. Med. Entomol. 2001. Available from: https://academic.oup.com/jme/article/38/2/130/876536
12. Weger-Lucarelli J, Rückert C, Chotiwan N, Nguyen C, Garcia Luna SM, Fauver JR, et al. Vector Competence of American Mosquitoes for Three Strains of Zika Virus. PLoS Negl Trop Dis. 2016 Oct 26;10(10):e0005101.
13. Brault AC. Changing patterns of west nile virus transmission: Altered vector competence and host susceptibility. Vol. 40, Veterinary Research. 2009.
14. Ciota AT, Ehrbar DJ, Van Slyke GA, Payne AF, Willsey GG, Viscio RE, et al. Quantification of intrahost bottlenecks of West Nile virus in Culex pipiens mosquitoes using an artificial mutant swarm. Infection, Genetics and Evolution. 2012 Apr;12(3):557–64.
15. Grubaugh ND, Weger-Lucarelli J, Murrieta RA, Fauver JR, Garcia-Luna SM, Prasad AN, et al. Genetic Drift during Systemic Arbovirus Infection of Mosquito Vectors Leads to Decreased Relative Fitness during Host Switching. Cell Host Microbe. 2016 Apr 13;19(4):481–92.
16. Kramer LD, Ciota AT. Dissecting vectorial capacity for mosquito-borne viruses. Vol. 15, Current Opinion in Virology. Elsevier B.V.; 2015. p. 112–8.

17. Mayton EH, Tramonte AR, Wearing HJ, Christofferson RC. Age-structured vectorial capacity reveals timing, not magnitude of within-mosquito dynamics is critical for arbovirus fitness assessment. Parasit Vectors. 2020 Dec 15;13(1):310.
18. Miller B, Monath T, Tabachnick W, Ezike V. Epidemic Yellow fever caused by an incompetent mosquito vector . Trop Med Parasitol. 1989;
19. Bialosuknia SM, Dupuis II AP, Zink SD, Koetzner CA, Maffei JG, Owen JC, et al. Adaptive evolution of West Nile virus facilitated increased transmissibility and prevalence in New York State. Emerg Microbes Infect. 2022 Dec 31;11(1):988–99.
20. Weaver SC, Forrester NL, Liu J, Vasilakis N. Population bottlenecks and founder effects: implications for mosquito-borne arboviral emergence. Vol. 19, Nature Reviews Microbiology. Nature Research; 2021. p. 184–95.
21. Domingo E, Holland JJ. RNA VIRUS MUTATIONS AND FITNESS FOR SURVIVAL [Internet]. 1997. Available from: www.annualreviews.org
22. Brackney DE, Beane JE, Ebel GD. RNAi targeting of West Nile virus in mosquito midguts promotes virus diversification. PLoS Pathog. 2009 Jul;5(7).
23. Fitzpatrick KA, Deardorff ER, Pesko K, Brackney DE, Zhang B, Bedrick E, et al. Population variation of West Nile virus confers a host-specific fitness benefit in mosquitoes.
24. Coffey LL, Vasilakis N, Brault AC, Powers AM, dé ric Tripet F, Weaver SC. Arbovirus evolution in vivo is constrained by host alternation [Internet]. 2008. Available from: www.pnas.orgcgidoi10.1073pnas.0712130105
25. Forrester NL, Coffey LL, Weaver SC. Arboviral bottlenecks and challenges to maintaining diversity and fitness during mosquito transmission. Vol. 6, Viruses. MDPI AG; 2014. p. 3991–4004.
26. Sardelis MR, Turell MJ, Dohm DJ, O'guinn ML. Vector Competence of Selected North American Culex and Coquillettidia Mosquitoes for West Nile Virus. Vol. 7.
27. Kent R, Juliusson L, Weissmann M, Evans S, Komar N. Seasonal Blood-Feeding Behavior of Culex tarsalis (Diptera: Culicidae) in Weld County, Colorado, 2007 [Internet]. Vol. 46, J. Med. Entomol. 2009. Available from: www.barcodinglife.org
28. Apperson CS, Hassan HK, Harrison BA, Savage HM, Aspen SE, Farajollahi A, et al. Host feeding patterns of established and potential mosquito vectors of West Nile virus in the Eastern United States. Vector-Borne and Zoonotic Diseases. 2004 Mar;4(1):71–82.
29. Grubaugh ND, Fauver JR, Rückert C, Weger-Lucarelli J, Garcia-Luna S, Murrieta RA, et al. Mosquitoes Transmit Unique West Nile Virus Populations during Each Feeding Episode. Cell Rep. 2017 Apr 25;19(4):709–18.
30. Jerzak G, Bernard KA, Kramer LD, Ebel GD. Genetic variation in West Nile virus from naturally infected mosquitoes and birds suggests quasispecies structure and strong purifying selection.
31. Jerzak GVS, Brown I, Shi PY, Kramer LD, Ebel GD. Genetic diversity and purifying selection in West Nile virus populations are maintained during host switching. Virology. 2008 May;374(2):256–60.
32. Gutiérrez S, Michalakis Y, Blanc S. Virus population bottlenecks during within-host progression and host-to-host transmission. Vol. 2, Current Opinion in Virology. Elsevier B.V.; 2012. p. 546–55.
33. Franz AWE, Kantor AM, Passarelli AL, Clem RJ. Tissue barriers to arbovirus infection in mosquitoes. Vol. 7, Viruses. MDPI AG; 2015. p. 3741–67.

34. Sanchez-Vargas I, Olson KE, Black WC. The genetic basis for salivary gland barriers to arboviral transmission. Insects. 2021 Jan 1;12(1):1–22.
35. Forrester NL, Guerbois M, Seymour RL, Spratt H, Weaver SC. Vector-Borne Transmission Imposes a Severe Bottleneck on an RNA Virus Population. PLoS Pathog. 2012 Sep;8(9).
36. Patterson EI, Khanipov K, Rojas MM, Kautz TF, Rockx-Brouwer D, Golovko G, et al. Mosquito bottlenecks alter viral mutant swarm in a tissue and time-dependent manner with contraction and expansion of variant positions and diversity. Virus Evol. 2018 Jan 1;4(1).
37. Weger-Lucarelli J, Garcia SM, Rückert C, Byas A, O'Connor SL, Aliota MT, et al. Using barcoded Zika virus to assess virus population structure in vitro and in Aedes aegypti mosquitoes. Virology. 2018 Aug 1;521:138–48.
38. McCrone JT, Woods RJ, Martin ET, Malosh RE, Monto AS, Lauring AS. Stochastic processes constrain the within and between host evolution of influenza virus. Elife. 2018 May 3;7.
39. Muñoz-Moreno R, Martínez-Romero C, Blanco-Melo D, Forst C V., Nachbagauer R, Benitez AA, et al. Viral Fitness Landscapes in Diverse Host Species Reveal Multiple Evolutionary Lines for the NS1 Gene of Influenza A Viruses. Cell Rep. 2019 Dec 17;29(12):3997-4009.e5.
40. Khanal S, Fennessey CM, O'Brien SP, Thorpe A, Reid C, Immonen TT, et al. In Vivo Validation of the Viral Barcoding of Simian Immunodeficiency Virus SIVmac239 and the Development of New Barcoded SIV and Subtype B and C Simian-Human Immunodeficiency Viruses . J Virol. 2019 Dec 12;94(1).
41. Immonen TT, Camus C, Reid C, Fennessey CM, Del Prete GQ, Davenport MP, et al. Genetically barcoded SIV reveals the emergence of escape mutations in multiple viral lineages during immune escape. Available from: www.pnas.org/cgi/doi/10.1073/pnas.1914967117
42. Fennessey CM, Pinkevych M, Immonen TT, Reynaldi A, Venturi V, Nadella P, et al. Genetically-barcoded SIV facilitates enumeration of rebound variants and estimation of reactivation rates in nonhuman primates following interruption of suppressive antiretroviral therapy. PLoS Pathog. 2017 May 1;13(5).
43. Varble A, Albrecht RA, Backes S, Crumiller M, Bouvier NM, Sachs D, et al. Influenza a virus transmission bottlenecks are defined by infection route and recipient host. Cell Host Microbe. 2014 Nov 12;16(5):691–700.
44. Amato KA, Haddock LA, Braun KM, Meliopoulos V, Livingston B, Honce R, et al. Influenza A virus undergoes compartmentalized replication in vivo dominated by stochastic bottlenecks. Nat Commun. 2022 Dec 1;13(1).
45. Pfeiffer JK, Kirkegaard K. Bottleneck-mediated quasispecies restriction during spread of an RNA virus from inoculation site to brain. Proc Natl Acad Sci U S A. 2006 Apr 4;103(14):5520–5.
46. McCrone JT, Lauring AS. Measurements of Intrahost Viral Diversity Are Extremely Sensitive to Systematic Errors in Variant Calling. J Virol. 2016 Aug;90(15):6884–95.
47. Aliota MT, Dudley DM, Newman CM, Weger-Lucarelli J, Stewart LM, Koenig MR, et al. Molecularly barcoded Zika virus libraries to probe in vivo evolutionary dynamics. PLoS Pathog. 2018 Mar 1;14(3).

48. Marsden MD, Zhang T hao, Du Y, Dimapasoc M, Soliman MSA, Wu X, et al. Tracking HIV Rebound following Latency Reversal Using Barcoded HIV. Cell Rep Med. 2020 Dec 22;1(9).
49. Riemersma KK, Jaeger AS, Crooks CM, Braun KM, Weger-Lucarelli J, Ebel GD, et al. Rapid Evolution of Enhanced Zika Virus Virulence during Direct Vertebrate Transmission Chains GENETIC DIVERSITY AND EVOLUTION. 2021; Available from: https://doi.org/10.1128/JVI.02218
50. Sexton NR, Bellis ED, Murrieta RA, Spangler MC, Cline PJ, Weger-Lucarelli J, et al. Genome Number and Size Polymorphism in Zika Virus Infectious Units [Internet]. 2021. Available from: http://jvi.asm.org/
51. Xie C, Su W, Sia SF, Choy KT, Morrell S, Zhou J, et al. A(H1N1)pdm09 Influenza Viruses Replicating in Ferret Upper or Lower Respiratory Tract Differed in Onward Transmission Potential by Air. J Infect Dis. 2022 Jan 5;225(1):65–74.
52. Phipps KL, Ganti K, Jacobs NT, Lee CY, Carnaccini S, White MC, et al. Collective interactions augment influenza A virus replication in a host-dependent manner. Nat Microbiol. 2020 Sep 1;5(9):1158–69.
53. Mccune BT, Lanahan MR, Tenoever BR, Pfeiffer JK. Rapid Dissemination and Monopolization of Viral Populations in Mice Revealed Using a Panel of Barcoded Viruses. 2020; Available from: https://doi.org/10.1128/JVI
54. Immonen TT, Fennessey CM, Lipkey L, Thorpe A, Del Prete GQ, Lifson JD, et al. Transient viral replication during analytical treatment interruptions in SIV infected macaques can alter the rebound-competent viral reservoir. PLoS Pathog. 2021 Jun 1;17(6).
55. Pinkevych M, Fennessey CM, Cromer D, Tolstrup M, Søgaard OS, Rasmussen TA, et al. Estimating Initial Viral Levels during Simian Immunodeficiency Virus/Human Immunodeficiency Virus Reactivation from Latency. J Virol. 2018 Jan 15;92(2).
56. Chen HC, Zorita E, Filion GJ. Using Barcoded HIV Ensembles (B-HIVE) for Single Provirus Transcriptomics. Curr Protoc Mol Biol. 2018 Apr 1;122(1).
57. Moriarty R V., Golfinos AE, Gellerup DD, Schweigert H, Mathiaparanam J, Balgeman AJ, et al. The mucosal barrier and anti-viral immune responses can eliminate portions of the viral population during transmission and early viral growth. PLoS One. 2021 Dec 1;16(12 December).
58. Pekrun K, De Alencastro G, Luo QJ, Liu J, Kim Y, Nygaard S, et al. Using a barcoded AAV capsid library to select for clinically relevant gene therapy vectors. JCI Insight. 2019 Nov 14;4(22).
59. Weinmann J, Weis S, Sippel J, Tulalamba W, Remes A, El Andari J, et al. Identification of a myotropic AAV by massively parallel in vivo evaluation of barcoded capsid variants. Nat Commun. 2020 Dec 1;11(1).
60. Brown D, Altermatt M, Dobreva T, Chen S, Wang A, Thomson M, et al. Deep Parallel Characterization of AAV Tropism and AAV-Mediated Transcriptional Changes via Single-Cell RNA Sequencing. Front Immunol. 2021 Oct 21;12.
61. Xu M, Li J, Xie J, He R, Su Q, Gao G, et al. High-Throughput Quantification of In Vivo Adeno-Associated Virus Transduction with Barcoded Non-Coding RNAs. Hum Gene Ther. 2019 Aug 1;30(8):946–56.
62. Frank DT, Byas AD, Murrieta R, Weger-Lucarelli J, Rückert C, Gallichotte E, et al. Intracellular diversity of WNV within circulating avian peripheral blood mononuclear

cells reveals host-dependent patterns of polyinfection. Available from: https://doi.org/10.1101/2023.01.27.525959
63. Bacsik DJ, Dadonaite B, Butler A, Greaney AJ, Heaton NS, Bloom JD, et al. Influenza virus transcription and progeny production are poorly correlated in single cells. Available from: https://doi.org/10.1101/2022.08.30.505828
64. Hixson B, Taracena ML, Buchon N. Midgut Epithelial Dynamics Are Central to Mosquitoes' Physiology and Fitness, and to the Transmission of Vector-Borne Disease. Front Cell Infect Microbiol. 2021 Mar 25;11.
65. Kumar A, Srivastava P, Sirisena P, Dubey S, Kumar R, Shrinet J, et al. Mosquito Innate Immunity. Insects. 2018 Aug 8;9(3):95.
66. Hung RJ, Hu Y, Kirchner R, Liu Y, Xu C, Comjean A, et al. A cell atlas of the adult *Drosophila* midgut. Proceedings of the National Academy of Sciences. 2020 Jan 21;117(3):1514–23.
67. Hung RJ, Li JSS, Liu Y, Perrimon N. Defining cell types and lineage in the Drosophila midgut using single cell transcriptomics. Curr Opin Insect Sci. 2021 Oct;47:12–7.
68. Cui Y, Franz AWE. Heterogeneity of midgut cells and their differential responses to blood meal ingestion by the mosquito, Aedes aegypti. Insect Biochem Mol Biol. 2020 Dec;127:103496.
69. Demerec M. The internal anatomy and histology of the imago of Drosophila melanogaster. In: Biology of Drosophila. Facsim. ed. Plainview, N.Y. : Cold Spring Harbor Laboratory Press; 1994. p. 435.
70. Xia J, Fei S, Huang Y, Lai W, Yu Y, Liang L, et al. Single-nucleus sequencing of silkworm larval midgut reveals the immune escape strategy of BmNPV in the midgut during the late stage of infection. Insect Biochem Mol Biol. 2024 Jan;164:104043.
71. Merkling SH, Lambrechts L. Taking Insect Immunity to the Single-Cell Level. Trends Immunol. 2020 Mar;41(3):190–9.
72. Severo MS, Landry JJM, Lindquist RL, Goosmann C, Brinkmann V, Collier P, et al. Unbiased classification of mosquito blood cells by single-cell genomics and high-content imaging. Proceedings of the National Academy of Sciences. 2018 Aug 7;115(32).
73. Kwon H, Mohammed M, Franzén O, Ankarklev J, Smith RC. Single-cell analysis of mosquito hemocytes identifies signatures of immune cell subtypes and cell differentiation. Elife. 2021 Jul 28;10.
74. Cho B, Yoon SH, Lee D, Koranteng F, Tattikota SG, Cha N, et al. Single-cell transcriptome maps of myeloid blood cell lineages in Drosophila. Nat Commun. 2020 Sep 8;11(1):4483.
75. Tattikota SG, Cho B, Liu Y, Hu Y, Barrera V, Steinbaugh MJ, et al. A single-cell survey of Drosophila blood. Elife. 2020 May 12;9.
76. Raddi G, Barletta ABF, Efremova M, Ramirez JL, Cantera R, Teichmann SA, et al. Mosquito cellular immunity at single-cell resolution. Science (1979). 2020 Aug 28;369(6507):1128–32.
77. Li H, Horns F, Wu B, Xie Q, Li J, Li T, et al. Classifying Drosophila Olfactory Projection Neuron Subtypes by Single-Cell RNA Sequencing. Cell. 2017 Nov;171(5):1206-1220.e22.
78. Davie K, Janssens J, Koldere D, De Waegeneer M, Pech U, Kreft Ł, et al. A Single-Cell Transcriptome Atlas of the Aging Drosophila Brain. Cell. 2018 Aug;174(4):982-998.e20.

79. Croset V, Treiber CD, Waddell S. Cellular diversity in the Drosophila midbrain revealed by single-cell transcriptomics. Elife. 2018 Apr 19;7.
80. Suomalainen M, Greber UF. Virus Infection Variability by Single-Cell Profiling. Viruses. 2021 Aug 9;13(8):1568.
81. Dulbecco R. Production of Plaques in Monolayer Tissue Cultures by Single Particles of an Animal Virus. Proceedings of the National Academy of Sciences. 1952 Aug;38(8):747–52.
82. Combe M, Garijo R, Geller R, Cuevas JM, Sanjuán R. Single-Cell Analysis of RNA Virus Infection Identifies Multiple Genetically Diverse Viral Genomes within Single Infectious Units. Cell Host Microbe. 2015 Oct;18(4):424–32.
83. Zanini F, Pu SY, Bekerman E, Einav S, Quake SR. Single-cell transcriptional dynamics of flavivirus infection. Elife. 2018 Feb 16;7.
84. McDonald E, Martin SW, Landry K, Gould C V., Lehman J, Fischer M, et al. West Nile Virus and Other Domestic Nationally Notifiable Arboviral Diseases — United States, 2018. MMWR Morb Mortal Wkly Rep. 2019 Aug 9;68(31):673–8.
85. Murray KO, Mertens E, Desprès P. West Nile virus and its emergence in the United States of America. Vet Res. 2010 Nov 24;41(6):67.
86. Ciota AT, Lovelace AO, Jia Y, Davis LJ, Young DS, Kramer LD. Characterization of mosquito-adapted West Nile virus. Journal of General Virology. 2008 Jul 1;89(7):1633–42.
87. Tsetsarkin KA, Vanlandingham DL, McGee CE, Higgs S. A Single Mutation in Chikungunya Virus Affects Vector Specificity and Epidemic Potential. PLoS Pathog. 2007 Dec 7;3(12):e201.
88. Carpenter A, Clem RJ. Factors Affecting Arbovirus Midgut Escape in Mosquitoes. Vol. 12, Pathogens. MDPI; 2023.
89. Brackney DE, Pesko KN, Brown IK, Deardorff ER, Kawatachi J, Ebel GD. West Nile Virus Genetic Diversity is Maintained during Transmission by Culex pipiens quinquefasciatus Mosquitoes. PLoS One. 2011 Sep 12;6(9):e24466.
90. Fitzmeyer EA, Gallichotte EN, Ebel GD. Scanning barcodes: A way to explore viral populations. PLoS Pathog. 2023 Apr 20;19(4):e1011291.
91. Novella IS, Presloid JB, Smith SD, Wilke CO. Specific and Nonspecific Host Adaptation during Arboviral Experimental Evolution. Microb Physiol. 2011;21(1–2):71–81.
92. Pybus OG, Tatem AJ, Lemey P. Virus evolution and transmission in an ever more connected world. Proceedings of the Royal Society B: Biological Sciences. 2015 Dec 22;282(1821):20142878.
93. Franklinos LH V, Jones KE, Redding DW, Abubakar I. The effect of global change on mosquito-borne disease. Lancet Infect Dis. 2019 Sep;19(9):e302–12.
94. Coffey LL, Forrester N, Tsetsarkin K, Vasilakis N, Weaver SC. Factors shaping the adaptive landscape for arboviruses: Implications for the emergence of disease. Vol. 8, Future Microbiology. 2013. p. 155–76.
95. Sene NM, Diouf B, Gaye A, Ndiaye EH, Ngom EHM, Gueye A, et al. Blood Feeding Patterns of Aedes aegypti Populations in Senegal. American Journal of Tropical Medicine and Hygiene. 2022 May 1;106(5):1402–5.
96. Ponlawat A, Harrington LC. Blood Feeding Patterns of Aedes aegypti and Aedes albopictus in Thailand [Internet]. Vol. 42, J. Med. Entomol. 2005. Available from: https://academic.oup.com/jme/article/42/5/844/863877

97. Zhou S, Jones C, Mieczkowski P, Swanstrom R. Primer ID Validates Template Sampling Depth and Greatly Reduces the Error Rate of Next-Generation Sequencing of HIV-1 Genomic RNA Populations. J Virol. 2015 Aug 15;89(16):8540–55.
98. Jabara CB, Jones CD, Roach J, Anderson JA, Swanstrom R. Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. Proceedings of the National Academy of Sciences. 2011 Dec 13;108(50):20166–71.
99. Sabi R, Tuller T. Modelling the Efficiency of Codon–tRNA Interactions Based on Codon Usage Bias. DNA Research. 2014 Oct 1;21(5):511–26.
100. Fernández-Sanlés A, Ríos-Marco P, Romero-López C, Berzal-Herranz A. Functional Information Stored in the Conserved Structural RNA Domains of Flavivirus Genomes. Front Microbiol. 2017 Apr 3;08.
101. Scroggs SLP, Grubaugh ND, Sena JA, Sundararajan A, Schilkey FD, Smith DR, et al. Endless Forms: Within-Host Variation in the Structure of the West Nile Virus RNA Genome during Serial Passage in Bird Hosts. 2019;
102. Smith DR, Adams AP, Kenney JL, Wang E, Weaver SC. Venezuelan equine encephalitis virus in the mosquito vector Aedes taeniorhynchus: Infection initiated by a small number of susceptible epithelial cells and a population bottleneck. Virology. 2008 Mar;372(1):176–86.
103. Cansado-Utrilla C, Zhao SY, McCall PJ, Coon KL, Hughes GL. The microbiome and mosquito vectorial capacity: rich potential for discovery and translation. Vol. 9, Microbiome. BioMed Central Ltd; 2021.
104. Cirimotich CM, Ramirez JL, Dimopoulos G. Native microbiota shape insect vector competence for human pathogens. Vol. 10, Cell Host and Microbe. 2011. p. 307–10.
105. Ramirez JL, Souza-Neto J, Cosme RT, Rovira J, Ortiz A, Pascale JM, et al. Reciprocal tripartite interactions between the Aedes aegypti midgut microbiota, innate immune system and dengue virus influences vector competence. PLoS Negl Trop Dis. 2012 Mar;6(3).
106. Hegde S, Khanipov K, Albayrak L, Golovko G, Pimenova M, Saldaña MA, et al. Microbiome interaction networks and community structure from laboratory-reared and field-collected Aedes aegypti, Aedes albopictus, and Culex quinquefasciatus mosquito vectors. Front Microbiol. 2018 Sep 10;9(SEP).
107. Duguma D, Hall MW, Rugman-Jones P, Stouthamer R, Terenius O, Neufeld JD, et al. Developmental succession of the microbiome of Culex mosquitoes Ecological and evolutionary microbiology. BMC Microbiol. 2015 Jul 24;15(1).
108. Ross PA, Endersby-Harshman NM, Hoffmann AA. A comprehensive assessment of inbreeding and laboratory adaptation in Aedes aegypti mosquitoes. Evol Appl. 2019 Mar 1;12(3):572–86.
109. Schirtzinger EE, Jasperson DC, Swanson DA, Mitzel D, Drolet BS, Richt JA, et al. Establishment of a *Culex tarsalis* (Diptera: Culicidae) Cell Line and its Permissiveness to Arbovirus Infection. J Med Entomol. 2023 Jan 12;60(1):239–44.
110. Shi PY, Tilgner M, Lo MK, Kent KA, Bernard KA. Infectious cDNA Clone of the Epidemic West Nile Virus from New York City. J Virol. 2002 Jun 15;76(12):5847–56.
111. Weger-Lucarelli J, Duggal NK, Bullard-Feibelman K, Veselinovic M, Romo H, Nguyen C, et al. Development and Characterization of Recombinant Virus Generated from a New World Zika Virus Infectious Clone. J Virol. 2017 Jan;91(1).

112. Gregori J, Perales C, Rodriguez-Frias F, Esteban JI, Quer J, Domingo E. Viral quasispecies complexity measures. Vol. 493, Virology. Academic Press Inc.; 2016. p. 227–37.
113. Girard M, Nelson CB, Picot V, Gubler DJ. Arboviruses: A global public health threat. Vaccine. 2020 May;38(24):3989–94.
114. Pierson TC, Diamond MS. The continued threat of emerging flaviviruses. Nat Microbiol. 2020 May 4;5(6):796–812.
115. Carpenter A, Bryant WB, Santos SR, Clem RJ. Infection of Aedes aegypti Mosquitoes with Midgut-Attenuated Sindbis Virus Reduces, but Does Not Eliminate, Disseminated Infection. J Virol. 2021 Jun 10;95(13).
116. Caccia S, Casartelli M, Tettamanti G. The amazing complexity of insect midgut cells: types, peculiarities, and functions. Cell Tissue Res. 2019 Sep 29;377(3):505–25.
117. Lan H, Chen H, Liu Y, Jiang C, Mao Q, Jia D, et al. Small Interfering RNA Pathway Modulates Initial Viral Infection in Midgut Epithelium of Insect after Ingestion of Virus. J Virol. 2016 Jan 15;90(2):917–29.
118. Dunphy BM, Kovach KB, Gehrke EJ, Field EN, Rowley WA, Bartholomay LC, et al. Long-term surveillance defines spatial and temporal patterns implicating Culex tarsalis as the primary vector of West Nile virus. Sci Rep. 2019 Apr 29;9(1):6637.
119. Turell MJ, O'Guinn ML, Dohm DJ, Webb JP, Sardelis MR. Vector Competence of *Culex tarsalis* from Orange County, California, for West Nile Virus. Vector-Borne and Zoonotic Diseases. 2002 Sep;2(3):193–6.
120. Main BJ, Marcantonio M, Johnston JS, Rasgon JL, Brown CT, Barker CM. Whole-genome assembly of *Culex tarsalis*. G3 Genes|Genomes|Genetics. 2021 Apr 12;11(2).
121. Lee WS, Webster JA, Madzokere ET, Stephenson EB, Herrero LJ. Mosquito antiviral defense mechanisms: a delicate balance between innate immunity and persistent viral infection. Parasit Vectors. 2019 Dec 11;12(1):165.
122. Sanborn MA, Li T, Victor K, Siegfried H, Fung C, Rothman AL, et al. Analysis of cell-associated DENV RNA by oligo(dT) primed 5' capture scRNAseq. Sci Rep. 2020 Jun 3;10(1):9047.
123. Christ P, Hill SR, Schachtner J, Hauser F, Ignell R. Functional characterization of mosquito short neuropeptide F receptors. Peptides (NY). 2018 May;103:31–9.
124. Predel R, Neupert S, Garczynski SF, Crim JW, Brown MR, Russell WK, et al. Neuropeptidomics of the Mosquito *Aedes aegypti*. J Proteome Res. 2010 Apr 5;9(4):2006–15.
125. TORII S. Expression and Function of IA-2 Family Proteins, Unique Neuroendocrine-specific Protein-tyrosine Phosphatases. Endocr J. 2009;56(5):639–48.
126. MBIKAY M, SEIDAH NG, CHRÉTIEN M. Neuroendocrine secretory protein 7B2: structure, expression and functions. Biochemical Journal. 2001 Jul 15;357(2):329.
127. Shao X, Li C, Fernandez I, Zhang X, Südhof TC, Rizo J. Synaptotagmin–Syntaxin Interaction: The C2 Domain as a Ca2+-Dependent Electrostatic Switch. Neuron. 1997 Jan;18(1):133–42.
128. Bimpisidis Z, König N, Stagkourakis S, Zell V, Vlcek B, Dumas S, et al. The NeuroD6 Subtype of VTA Neurons Contributes to Psychostimulant Sensitization and Behavioral Reinforcement. eNeuro. 2019 May;6(3):ENEURO.0066-19.2019.

129. Dzaki N, Ramli KN, Azlan A, Ishak IH, Azzam G. Evaluation of reference genes at different developmental stages for quantitative real-time PCR in Aedes aegypti. Sci Rep. 2017 Mar 16;7(1):43618.
130. Cai D, Liu L, Tian B, Fu X, Yang Q, Chen J, et al. Dual-Role Ubiquitination Regulation Shuttling the Entire Life Cycle of the Flaviviridae. Front Microbiol. 2022 May 6;13.
131. Pujhari S, Hughes GL, Pakpour N, Suzuki Y, Rasgon JL. Wolbachia-induced inhibition of O'nyong nyong virus in Anopheles mosquitoes is mediated by Toll signaling and modulated by cholesterol. bioRxiv. 2023 Jun 1;
132. Stokes S, Almire F, Tatham MH, McFarlane S, Mertens P, Pondeville E, et al. The SUMOylation pathway suppresses arbovirus replication in Aedes aegypti cells. PLoS Pathog. 2020 Dec 22;16(12):e1009134.
133. Fikrig E, Main AJ, Cheng G, Anderson JF, Ferrandino FJ. Horizontal and Vertical Transmission of West Nile Virus Genotype NY99 by Culex salinarius and Genotypes NY99 and WN02 by Culex tarsalis. Am J Trop Med Hyg. 2012 Jan 1;86(1):134–9.
134. Pandey A, Suman S, Chandna S. Predictive role of mitochondrial genome in the stress resistance of insects and nematodes. Bioinformation. 2010 Jun 24;5(1):21–7.
135. Gatti P, Ilamathi HS, Todkar K, Germain M. Mitochondria Targeted Viral Replication and Survival Strategies—Prospective on SARS-CoV-2. Front Pharmacol. 2020 Aug 28;11.
136. Barbier V, Lang D, Valois S, Rothman AL, Medin CL. Dengue virus induces mitochondrial elongation through impairment of Drp1-triggered mitochondrial fission. Virology. 2017 Jan;500:149–60.
137. Chatel-Chaix L, Cortese M, Romero-Brey I, Bender S, Neufeldt CJ, Fischl W, et al. Dengue Virus Perturbs Mitochondrial Morphodynamics to Dampen Innate Immune Responses. Cell Host Microbe. 2016 Sep;20(3):342–56.
138. Muralidharan S, Mandrekar P. Cellular stress response and innate immune signaling: integrating pathways in host defense and inflammation. J Leukoc Biol. 2013 Aug 29;94(6):1167–84.
139. Ambrose RL, Mackenzie JM. ATF6 Signaling Is Required for Efficient West Nile Virus Replication by Promoting Cell Survival and Inhibition of Innate Immune Responses. J Virol. 2013 Feb 15;87(4):2206–14.
140. Evans CG, Chang L, Gestwicki JE. Heat Shock Protein 70 (Hsp70) as an Emerging Drug Target. J Med Chem. 2010 Jun 24;53(12):4585–602.
141. Li DC, Yang F, Lu B, Chen DF, Yang WJ. Thermotolerance and molecular chaperone function of the small heat shock protein HSP20 from hyperthermophilic archaeon, Sulfolobus solfataricus P2. Cell Stress Chaperones. 2012 Jan;17(1):103–8.
142. Qian J, Vafiadaki E, Florea SM, Singh VP, Song W, Lam CK, et al. Small Heat Shock Protein 20 Interacts With Protein Phosphatase-1 and Enhances Sarcoplasmic Reticulum Calcium Cycling. Circ Res. 2011 Jun 10;108(12):1429–38.
143. Taracena ML, Bottino-Rojas V, Talyuli OAC, Walter-Nuno AB, Oliveira JHM, Angleró-Rodriguez YI, et al. Regulation of midgut cell proliferation impacts Aedes aegypti susceptibility to dengue virus. PLoS Negl Trop Dis. 2018 May 21;12(5):e0006498.
144. Yang C, Wang F, Huang D, Ma H, Zhao L, Zhang G, et al. Vector competence and immune response of Aedes aegypti for Ebinur Lake virus, a newly classified mosquito-borne orthobunyavirus. PLoS Negl Trop Dis. 2022 Jul 18;16(7):e0010642.

145. Chen TY, Bozic J, Mathias D, Smartt CT. Immune-related transcripts, microbiota and vector competence differ in dengue-2 virus-infected geographically distinct Aedes aegypti populations. Parasit Vectors. 2023 May 19;16(1):166.
146. Dainat J. AGAT: Another Gff Analysis Toolkit to handle annotations in any GTF/GFF format.
147. Satija R. Seurat - Guided Clustering Tutorial. 2023.
148. Patel K. https://www.youtube.com/watch?v=04gB2owLKus. 2022. DESeq2 workflow tutorial | Differential Gene Expression Analysis | Bioinformatics 101.
149. Li WV, Li Y. scLink: Inferring Sparse Gene Co-expression Networks from Single-cell Expression Data. Genomics Proteomics Bioinformatics. 2021 Jun;19(3):475–92.
150. Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. 2021.
151. Huerta-Cepas J, Szklarczyk D, Heller D, Herńandez-Plaza A, Forslund SK, Cook H, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. 2019.
152. Bohórquez D V., Shahid RA, Erdmann A, Kreger AM, Wang Y, Calakos N, et al. Neuroepithelial circuit formed by innervation of sensory enteroendocrine cells. Journal of Clinical Investigation. 2015 Feb 2;125(2):782–6.
153. Lim SM, Koraka P, Osterhaus ADME, Martina BEE. West Nile Virus: Immunity and Pathogenesis. Viruses. 2011 Jun 15;3(6):811–28.
154. Pybus OG, Rambaut A. Evolutionary analysis of the dynamics of viral infectious disease. Nat Rev Genet. 2009 Aug;10(8):540–50.

# APPENDIX A: ADDITIONAL ACADEMIC ACCOMPLISHMENTS

I began my doctorate in fall 2019. CSU closed and halted all research activities, in response to the SARS-CoV-2 pandemic, around March 2020. Upon being repatriated from Peru that same month I officially joined the Ebel lab and immediately went to work on the various SARS-CoV-2 surveillance projects the lab had undertaken. These efforts resulted in 5 co-author publications detailed below.

Gallichotte EN, Nehring M, Young MC, Pugh S, Sexton NR, **Fitzmeyer E**, Quicke KM, Richardson M, Pabilonia KL, Ehrhart N, Fosdick BK, VandeWoude S, Ebel GD. Durable Antibody Responses in Staff at Two Long-Term Care Facilities, during and Post SARS-CoV-2 Outbreaks. Microbiol Spectr. 2021 Sep 3;9(1):e0022421. PubMed PMID: 34287058.

Nelson TL, Fosdick BK, Biela LM, Schoenberg H, Mast S, McGinnis E, Young MC, Lynn L, Fahrner S, Nolt L, Dihle T, Quicke K, Gallichotte EN, **Fitzmeyer E**, Ebel GD, Pabilonia K, Ehrhart N, VandeWoude S. Association Between COVID-19 Exposure and Self-reported Compliance With Public Health Guidelines Among Essential Employees at an Institution of Higher Education in the US. JAMA Netw Open. 2021 Jul 1;4(7):e2116543. PubMed Central PMCID: PMC8295736.

Gallichotte EN, Quicke KM, Sexton NR, **Fitzmeyer E**, Young MC, Janich AJ, Dobos K, Pabilonia KL, Gahm G, Carlton EJ, Ebel GD. Early adoption of longitudinal surveillance for SARS-CoV-2 among staff in long-term care facilities: prevalence, virologic and sequence analysis. Microbiology Spectrum. 2021 Dec 22;9(3):e01003-21.

Quicke KM, Baxter BA, Stromberg S, Gallichotte EN, **Fitzmeyer E**, Young MC, ... & Ryan EP. (2022). The BioFire® RP2. 1 Panel Did Not Identify Concurrent Respiratory Virus Infection in Adults with Variable SARS-CoV-2 Disease Severity and Infection Duration. Advances in Virology, 2022.

Sexton, NR, Cline, PJ, Gallichotte, EN, **Fitzmeyer, E**, Young, MC, Janich, AJ, Pabilonia, KL, Ehrhart, N, & Ebel, GD (2023). SARS-CoV-2 entry into and evolution within a skilled nursing facility. *Scientific reports*, *13*(1), 11657.

In 2021 I dedicated extensive time to sequencing and identifying significant variants in a series of Cedar virus-infected *Artibeus jamaicensis* bat cell libraries for the Schountz lab.

I began my thesis work in 2020 and became a doctoral candidate June 2021. In 2021 I was awarded a T32 fellowship (T32-AI162691) through the microbiology department, as well as an “Accelerating Innovations in Pandemic Disease” initiative fellowship through The Anschutz Foundation. In 2023 I was awarded the department’s Graduate Student Excellence Award for research, and 1$^{st}$ place graduate student talk at the Rocky Mountain Virology Conference.

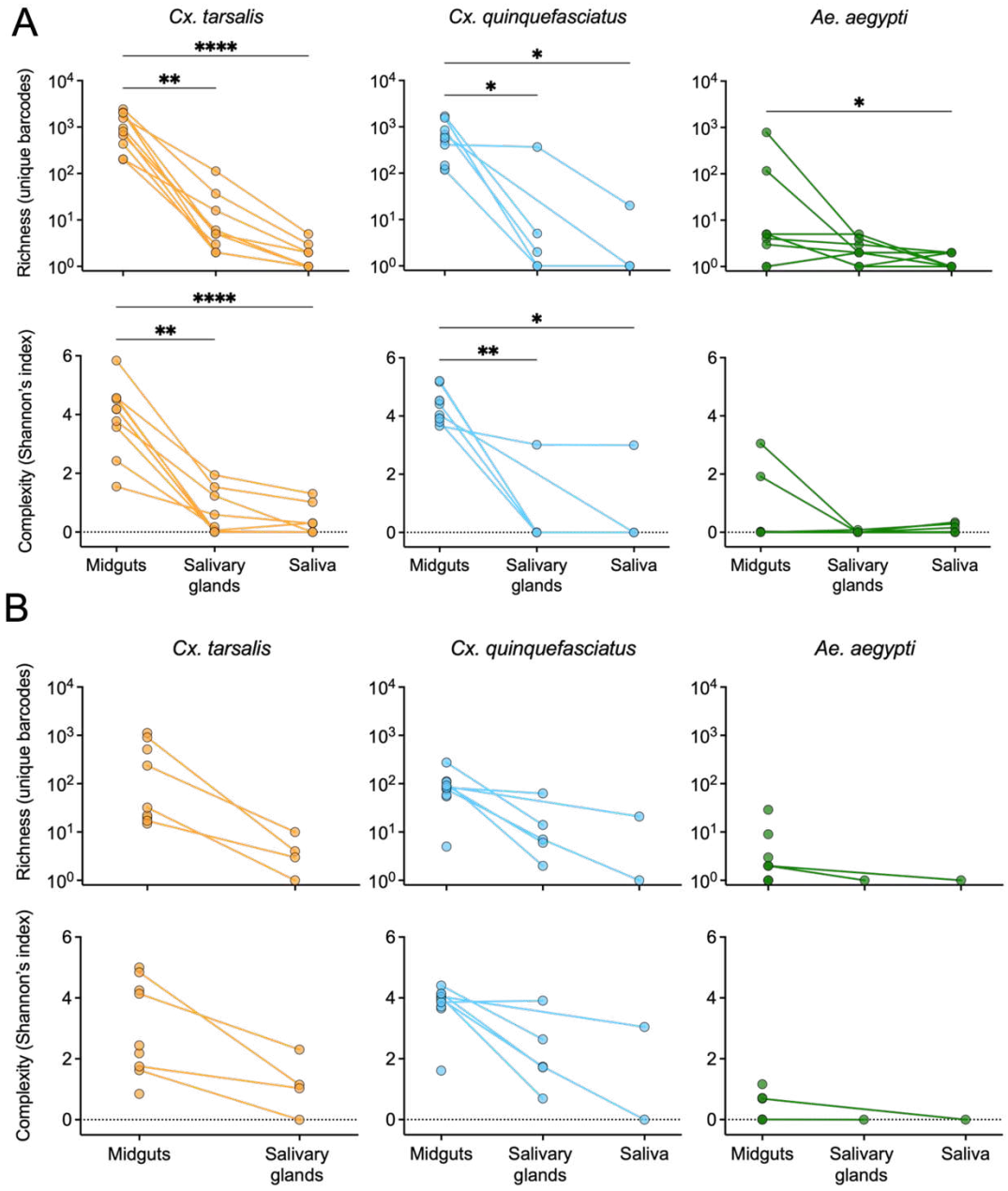



**Figure S2.1. Impact of intrahost bottlenecks on bcWNV populations in *Cx. tarsalis*, *Cx. quinquefasciatus*, and *Ae. aegypti* at days 4 and 12 post infection.** bcWNV population richness and complexity in midgut, salivary gland, and saliva samples from *Cx. tarsalis*, *Cx. quinquefasciatus*, and *Ae. aegypti* mosquitoes at days 12 (A) and 4 (B) post-infection. Lines connecting points indicate that samples were collected from the same mosquito. Diversity indices were generated by applying the Shannon function from the QSutils package in R to barcode abundance vectors. Dashed line represents 0. Statistical significance determined using Kruskal-Wallis with Dunn's multiple comparisons test, * = $p<0.05$, ** = $p<0.005$, *** = $p<0.0005$, **** = $p<0.0001$. Only significant comparisons shown.

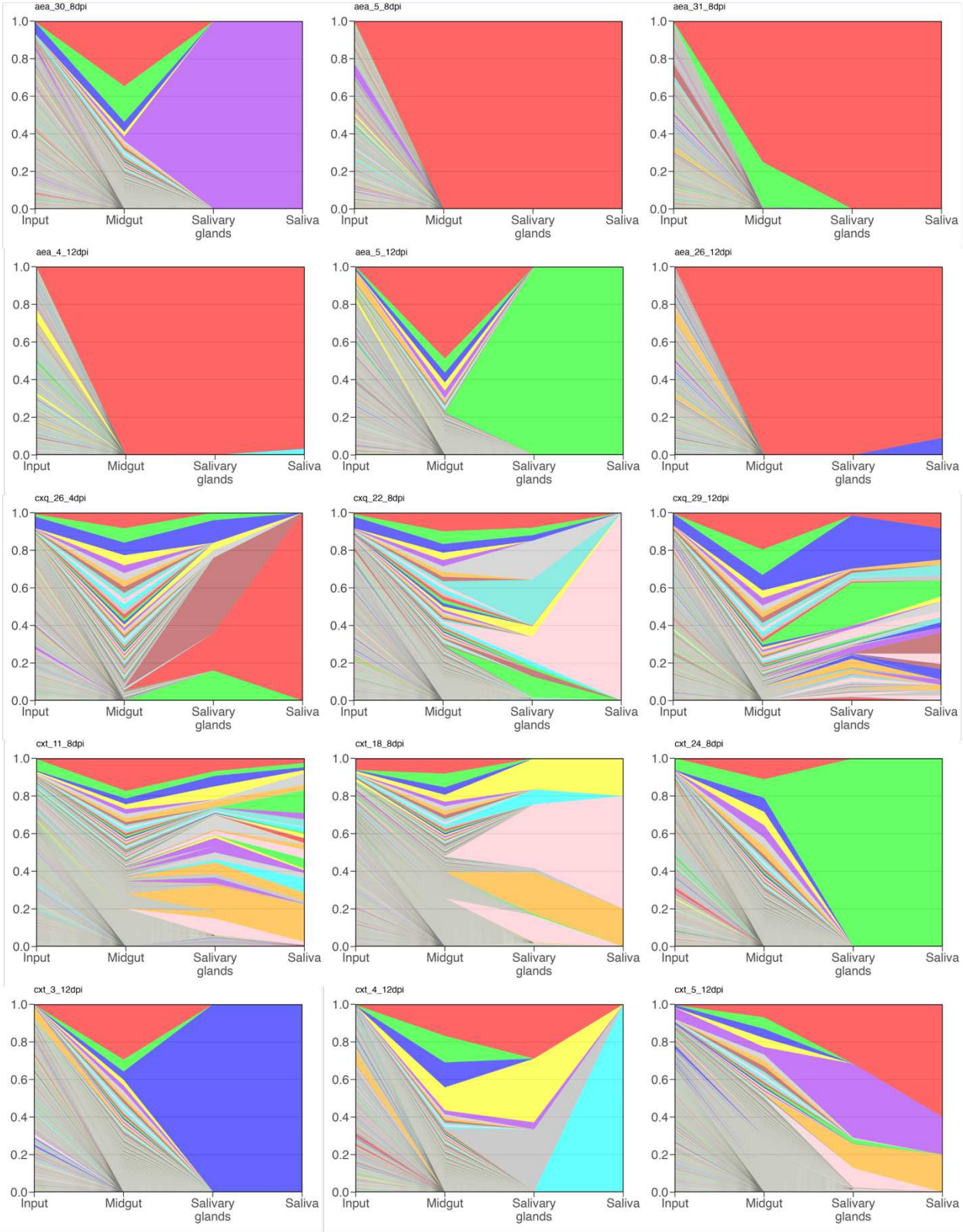

aea_30_8dpi
aea_5_8dpi
aea_31_8dpi
aea_4_12dpi
aea_5_12dpi
aea_26_12dpi
cxq_26_4dpi
cxq_22_8dpi
cxq_29_12dpi
cxt_11_8dpi
cxt_18_8dpi
cxt_24_8dpi
cxt_3_12dpi
cxt_4_12dpi
cxt_5_12dpi
1.0
0.8
0.6
0.4
0.2
0.0
Input
Midgut
Salivary glands
Saliva

**Figure S2.2. Visualization of bcWNV dynamics in individual mosquitoes not selected for Figure 2.4A.** bcWNV barcode dynamics during mosquito infection in individual *Cx. tarsalis*, *Cx. quinquefasciatus*, and *Ae. aegypti* mosquitoes at days 4, 8 and 12 post-infection. Species and timepoint are indicated in plot header; aea = *Ae. aegypti*, cxq = *Cx. quinquefasciatus*, cxt = *Cx. tarsalis.* Barcodes were ranked from most to least frequent by each midgut sample. Color represents barcode rank and is not consistent across graphs.

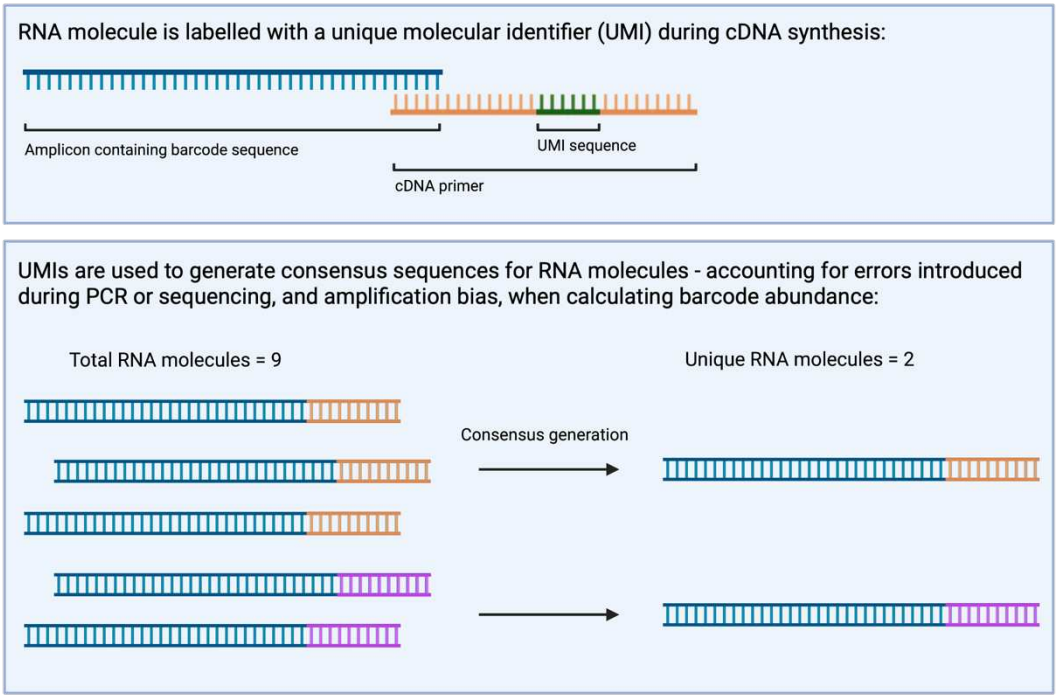



**Figure S2.3. Inclusion of a unique molecular identifier accounts for amplification bias and sequencing error.** (created with BioRender.com)

**Table S2.1. Infection rates of relevant compartments in *Cx. tarsalis*, *Cx. quinquefasciatus*, and *Ae. aegypti*.**

| Species: | DPI: | Rate of infection in tissue/secretion: | | |
|---|---|---|---|---|
| | | Midguts | Salivary Glands | Saliva |
| *Cx. tarsalis* | 4 | 29/36 (80.5%) | 27/36 (75.0%) | 2/36 (5.5%) |
| | 8 | 34/36 (94.4%) | 34/36 (94.4%) | 14/36 (38.9%) |
| | 12 | 35/36 (97.2%) | 35/36 (97.2%) | 35/36 (97.2%) |
| | Total | 98/108 (90.7%) | 96/108 (88.9%) | 51/108 (47.2%) |
| *Cx. quinquefasciatus* | 4 | 36/36 (100%) | 20/36 (55.6%) | 5/36 (13.9%) |
| | 8 | 36/36 (100%) | 15/36 (41.7%) | 3/36 (8.3%) |
| | 12 | 36/36 (100%) | 24/36 (66.7%) | 9/36 (25.0%) |
| | Total | 108/108 (100%) | 59/108 (54.6%) | 17/108 (15.7%) |
| *Ae. aegypti* | 4 | 10/40 (25.0%) | 2/40 (5.0%) | 2/40 (5.0%) |
| | 8 | 24/39 (61.5%) | 22/39 (56.4%) | 10/39 (25.6%) |
| | 12 | 15/39 (38.5%) | 19/39 (48.7%) | 12/39 (30.8%) |
| | Total | 49/118 (41.5%) | 43/118 (36.4%) | 24/108 (22.2%) |

**Table S2.2. Primers and probes used for qRT-PCR and library preparation.**

| **Primer:** | **Sequence:** |
|---|---|
| **WNV_ENV_F70** | TCAGCGATCTCTCCACCAAAG |
| **WNV_ENV_R70** | GGGTCAGCACGTTTGTCATTG |
| **WNENV-prob** | FAM/TGCCCGACC/ZEN/ATGGGAGAAGCTC/3\|ABkFQ/ |
| **ID_cDNAWNV_7374_Rev** | GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNNNNNcagtGCCATCCACTACAGCGTTCT |
| **5'_ID_Primer_Rev** | GTCTCGTGGGCTCGGAGATGTGTAT |
| **5'_ID_WNV_7171_For** | TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNTTCCCCTTCGTCGATGTTGG |
| **2nd round FP (N5XX)** | AATGATACGGCGACCACCGAGATCTACAC[i5]TCGTCGGCAGCGTC |
| **2nd round RP (N7XXX)** | CAAGCAGAAGACGGCATACGAGAT[i7]GTCTCGTGGGCTCGG |

*Nextera sequencing primer binding sites

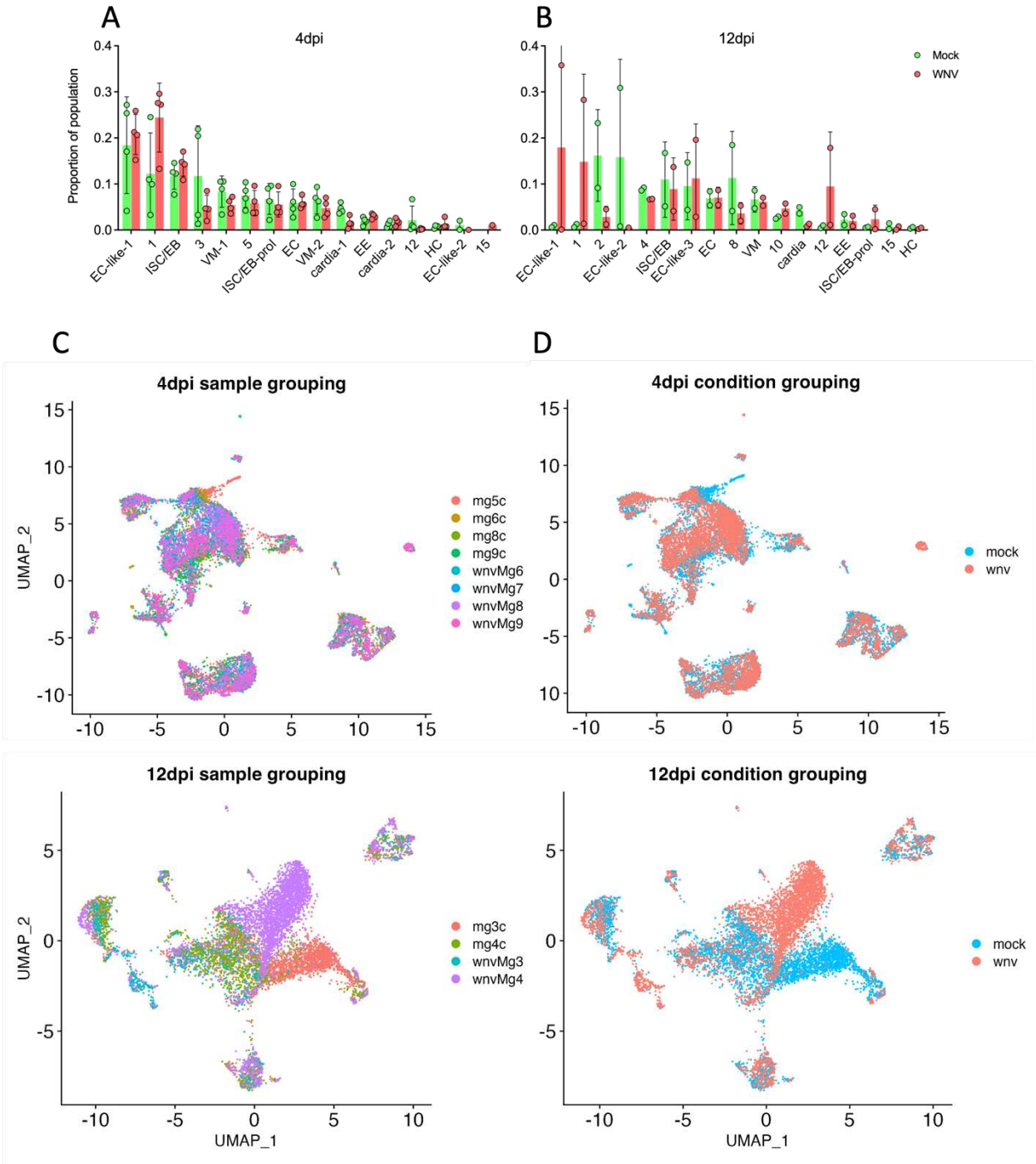



**Figure S3.1. Cluster proportion and grouping by condition and replicate.** Proportion of the total population comprised by each cluster compared between mock and WNV-infected conditions at 4dpi (**A**) and 12dpi (**B**). Only significant comparisons shown. Significance determined by multiple unpaired t-tests. (**C**) Cluster grouping and composition by sample – mock and infected samples combined. Different colors denote different samples. (**D**) Cluster grouping and composition by infection condition. Salmon = mock, blue = WNV-infected.

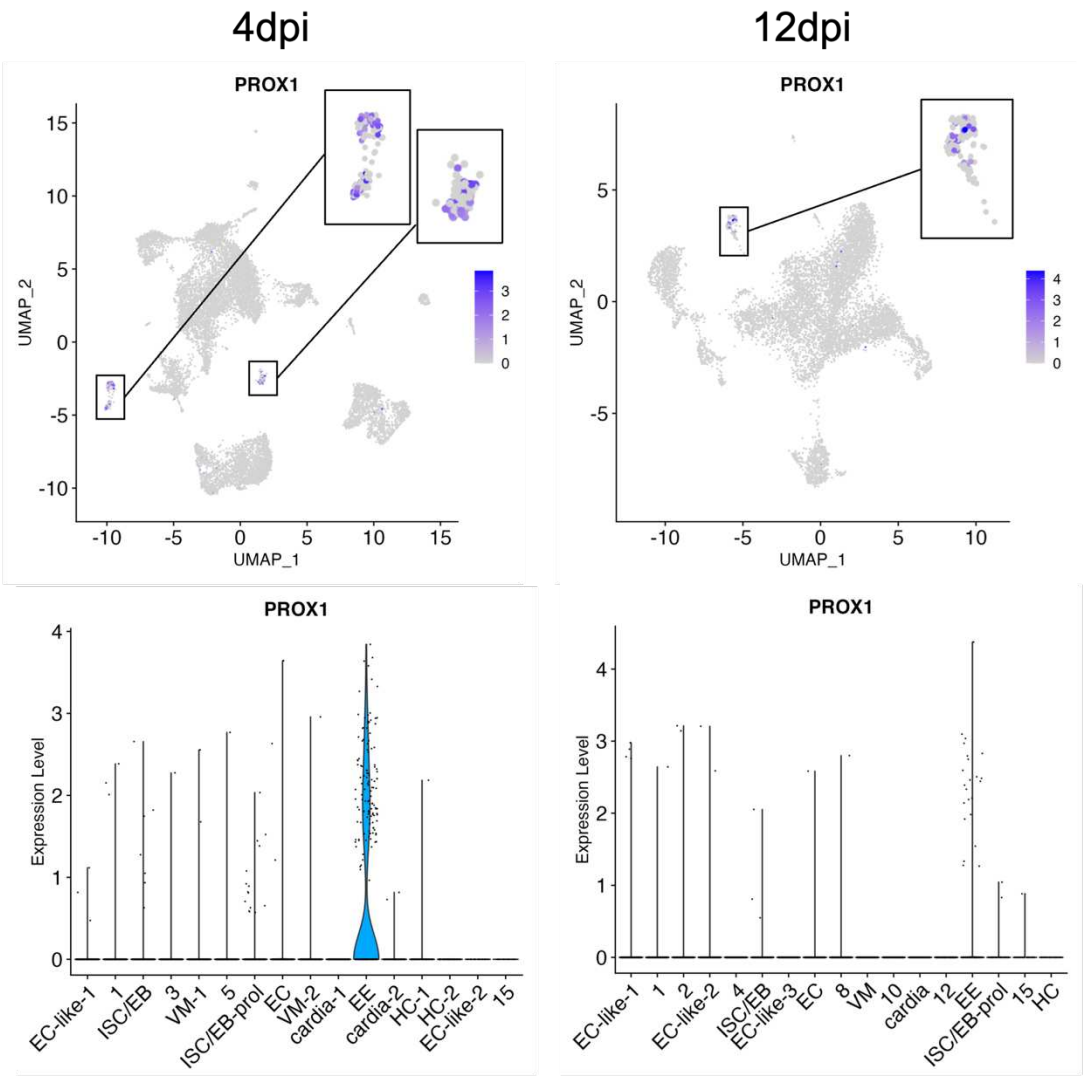



**Figure S3.2. Identifying enteroendocrine cells by visualizing PROX1 expression.** PROX1 expression visualized in the total population at both timepoints via uniform manifold approximation and projection (UMAP) feature map and violin plot. Color in feature map denotes expression level.

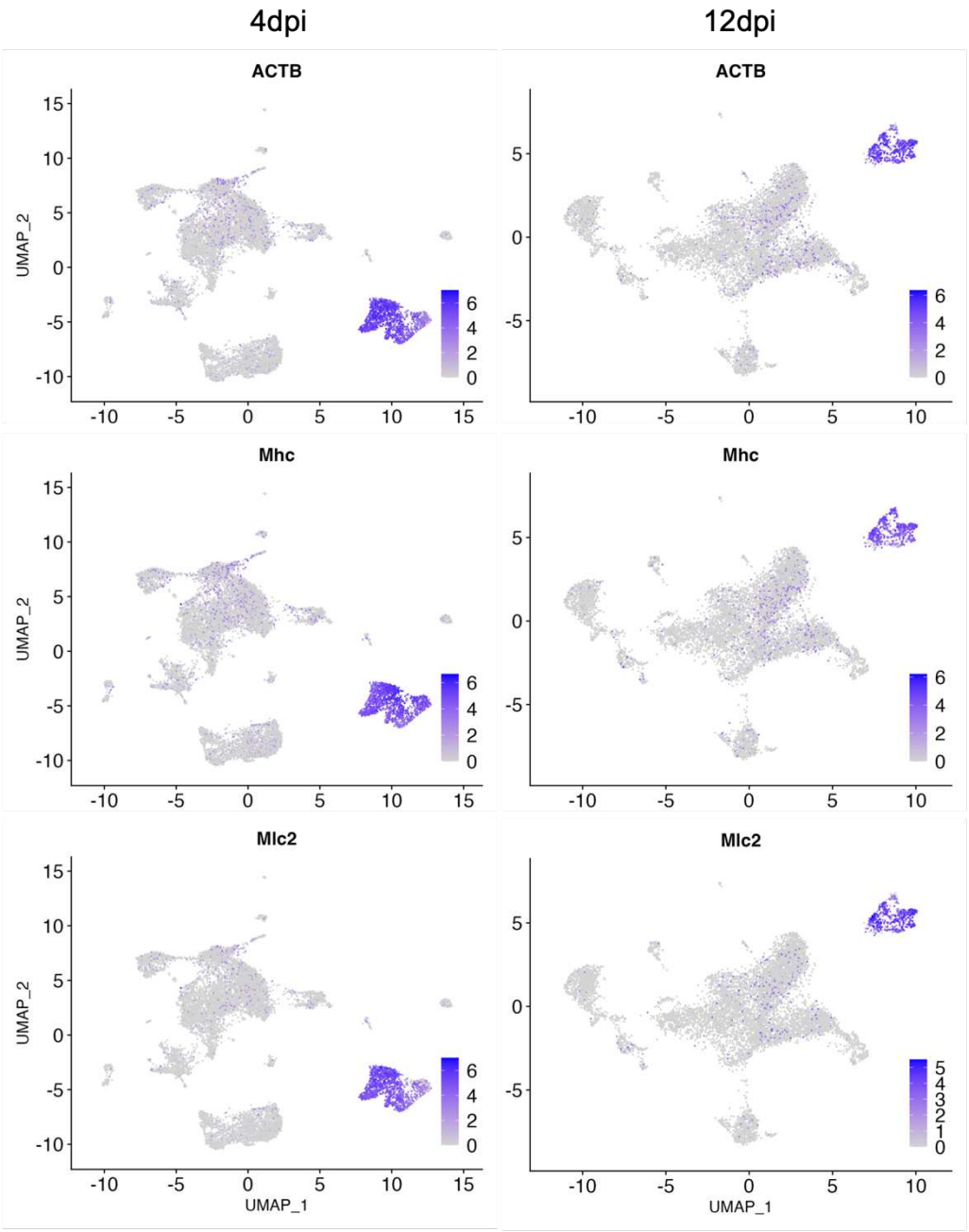



**Figure S3.3. Confirming visceral muscle cell-type by visualizing cytoskeletal gene expression.** Actin (ACTB), myosin (Mhc), and light chain (Mlc2) expression visualized in the total population at both timepoints via uniform manifold approximation and projection (UMAP) feature map. Color in feature map denotes expression level.

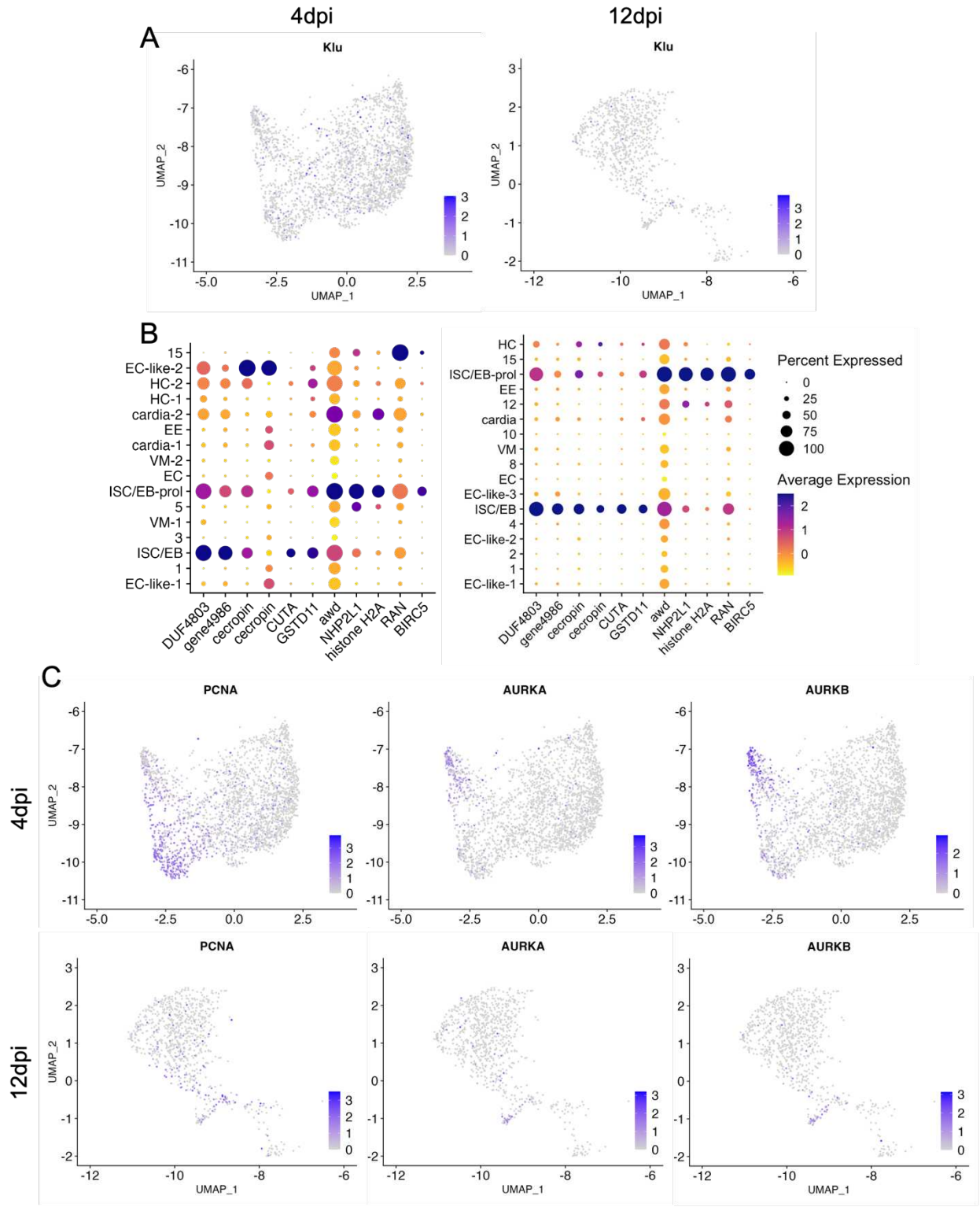



**Figure S3.4. Identifying intestinal stem cell/enteroblast (ISC/EB) and proliferating ISC/EB cell-types by visualizing canonical marker and shared marker gene expression. (A)** Klumpfuss (klu) expression visualized in the total population at both timepoints via UMAP

feature map. (**B**) Demonstrating that ISC/EB and ISC/EB-prol populations across timepoints share identical top cluster markers via dotplot visualization. Dot size denotes percent of cells expressing each gene, color denotes scaled gene expression. (**C**) Identifying proliferation in ISC/EB-prol populations by visualizing PCNA, and aurora kinases A and B in the total population at both timepoints via uniform manifold approximation and projection (UMAP) feature map. Color in feature maps denotes expression level.

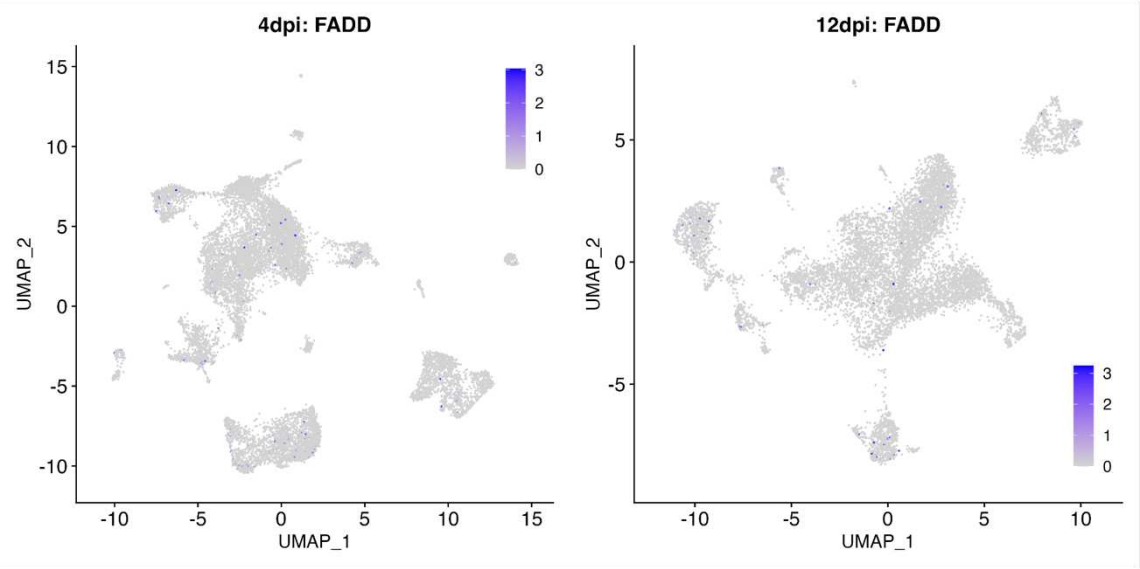



**Figure S3.5. Confirming that cell death does not drive clustering.** FADD (death domain) expression visualized in the total population at each timepoint via UMAP feature plot. Color in feature map denotes expression level.

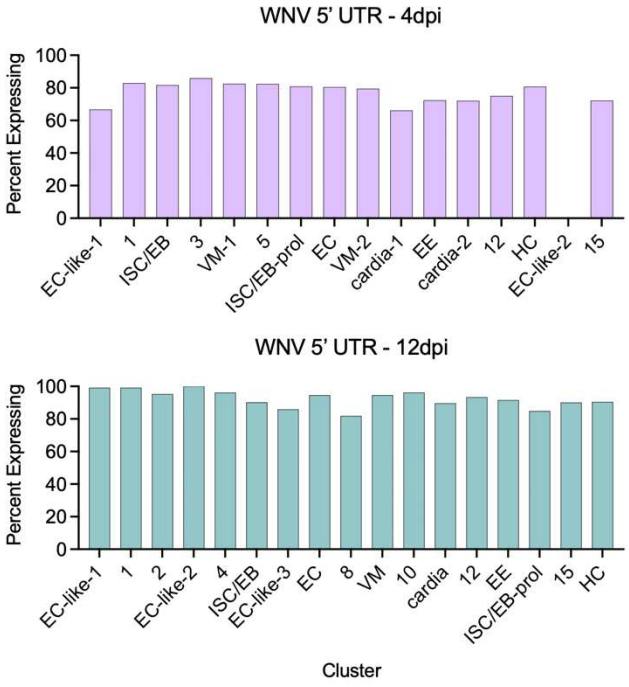



**Figure S3.6. Percent of cells in each cluster containing WNV vRNA.**

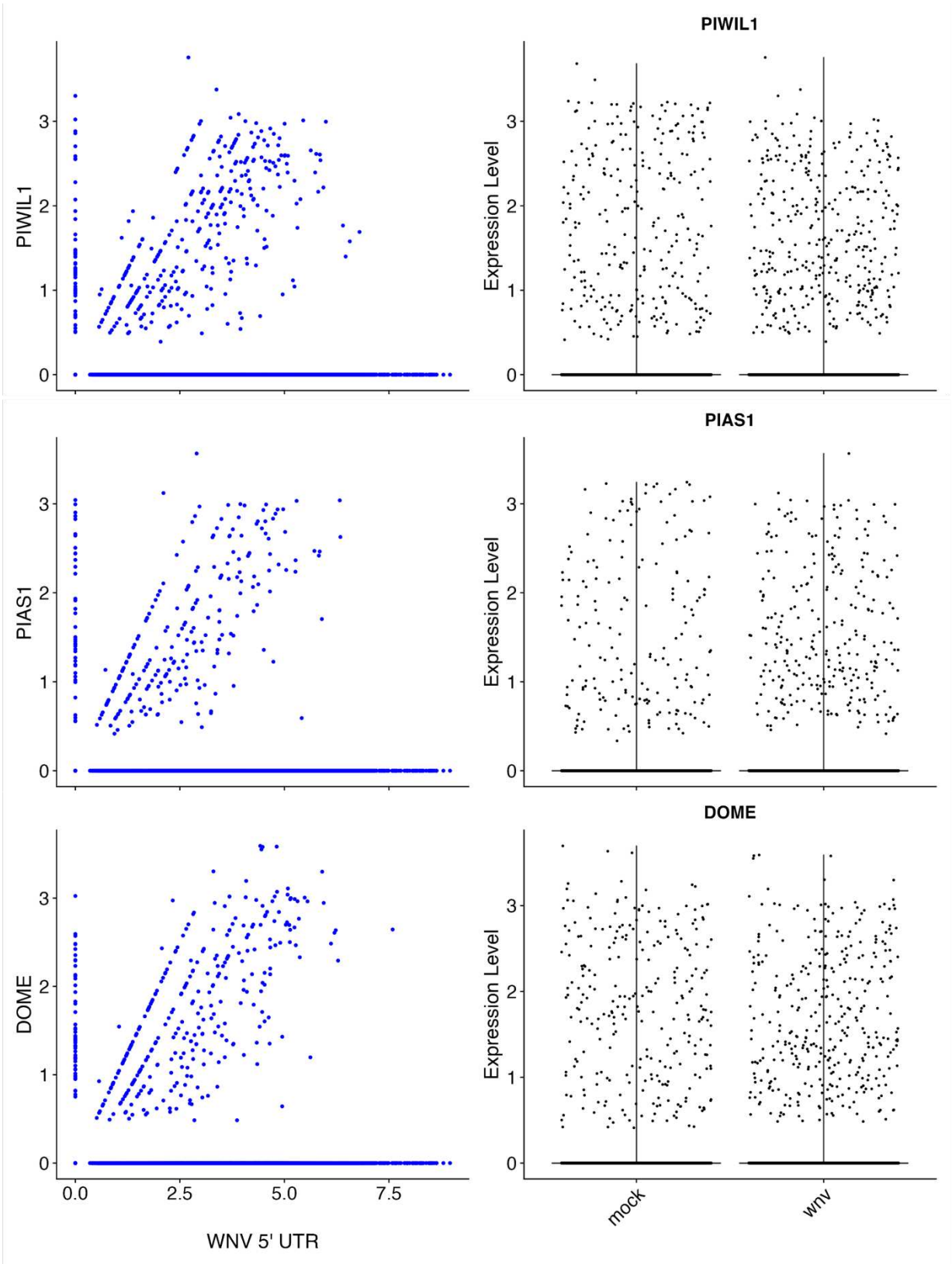



**Figure S3.7. Visually confirming vRNA level is correlated with select immune gene expression without significantly increasing expression in the total population.** Correlation

between vRNA and 3 of the most highly correlated immune genes (as determined by scLink) confirmed by feature scatter. Equivalent expression levels of immune genes between mock and WNV-infected conditions confirmed by violin plot. Timepoints combined for all plots.

**Table S3.1. Complete list of genes identified as significantly positively correlated with WNV vRNA.** Gene name, ID, correlation value, p-value, and description for all genes significantly ($p < 0.05$) positively correlated (> 0.65) with vRNA. Significance was determined by bootstrapping.

| DPI | Preferred Name | Gene ID | cor | p-value | Description |
|---|---|---|---|---|---|
| 4dpi | | nbis-gene-2-utr | 1.00 | 0 | WNV 5' UTR |
| 4dpi | ATRX | gene14505 | 0.69 | 0 | Transcriptional regulator |
| 4dpi | | gene9014 | 0.67 | 0 | Cytochrome p450 |
| 4dpi | | gene7420 | 0.66 | 0 | |
| 12dpi | | nbis-gene-2-utr | 1.00 | 0 | WNV 5’ UTR |
| 12dpi | GSTE4 | gene10812 | 0.80 | 0 | Glutathione S-transferase |
| 12dpi | HAO1 | gene13987 | 0.79 | 0.015 | FMN-dependent dehydrogenase |
| 12dpi | METTL20 | gene5405 | 0.77 | 0 | Lysine methyltransferase |
| 12dpi | | gene870 | 0.75 | 0 | |
| 12dpi | | gene2408 | 0.74 | 0.005 | Myofilin |
| 12dpi | PROX1 | gene10632 | 0.74 | 0.01 | Homeo-prospero domain |
| 12dpi | | gene3582 | 0.72 | 0 | |
| 12dpi | | gene9007 | 0.72 | 0 | |
| 12dpi | UROS | gene5208 | 0.72 | 0 | Uroporphyrinogen-III synthase |
| 12dpi | | gene2263 | 0.71 | 0 | |
| 12dpi | | gene11056 | 0.71 | 0 | |
| 12dpi | BCAN | gene8287 | 0.70 | 0 | C-Type lectin |
| 12dpi | | gene10708 | 0.68 | 0 | |
| 12dpi | | gene13467 | 0.67 | 0 | |
| 12dpi | | gene5114 | 0.67 | 0 | Trypsin-like serine protease |
| 12dpi | | gene6473 | 0.67 | 0 | AMP dependent ligase |

| | | | | | |
|---|---|---|---|---|---|
| 12dpi | DHDH | gene9019 | 0.67 | 0 | dihydrodiol dehydrogenase |
| 12dpi | | gene5757 | 0.67 | 0 | cytochrome P450 |
| 12dpi | mRpS26 | gene10824 | 0.66 | 0 | Mitochondrial ribosomal protein |
| 12dpi | | gene1254 | 0.66 | 0 | Glutathione S-transferase |
| 12dpi | CHKov1 | gene1164 | 0.65 | 0 | |
| 12dpi | | gene5412 | 0.65 | 0 | |
| 12dpi | | gene9874 | 0.65 | 0 | Aldo/keto reductase family |

**Table S3.2. Sequencing metadata.**

| DPI | Sample ID | Coverage | %Q30 | Pooling method |
|---|---|---|---|---|
| 12 | wnv_mg3 | 530,994,115 | 87.26 | KAPA_quant |
| 12 | wnv_mg4 | 518,982,659 | 85.87 | KAPA_quant |
| 12 | mg3_c | 525,148,879 | 87.04 | KAPA_quant |
| 12 | mg4_c | 488,374,119 | 87.23 | KAPA_quant |
| 4 | wnv_mg5 | 625,343,922 | 82.57 | KAPA_quant |
| 4 | wnv_mg6 | 570,441,976 | 81.9 | KAPA_quant |
| 4 | wnv_mg7 | 551,765,711 | 80.42 | KAPA_quant |
| 4 | wnv_mg8 | 485,435,325 | 81.02 | KAPA_quant |
| 4 | wnv_mg9 | 468,776,342 | 83.11 | KAPA_quant |
| 4 | mg5_c | 537,529,562 | 80.88 | KAPA_quant |
| 4 | mg6_c | 537,560,364 | 83.45 | KAPA_quant |
| 4 | mg7_c | 583,093,824 | 82.62 | KAPA_quant |
| 4 | mg8_c | 491,293,995 | 84.01 | KAPA_quant |
| 4 | mg9_c | 510,863,819 | 84.27 | KAPA_quant |
| Average: | | 529,473,672 | 84.00 | |

**Table S3.3. Cell recovery and read mapping metadata.**

| DPI | Sample ID | Recovered cells | Median genes | Mean reads/cell | Reads mapped to genome | Reads mapped to transcriptome | Fraction reads in cells |
|---|---|---|---|---|---|---|---|
| 12 | wnv_mg3 | 996 | 513 | 553127 | 25.40% | 9.80% | 20.20% |
| 12 | wnv_mg4 | 4077 | 322 | 127295 | 28.90% | 12.40% | 25.40% |
| 12 | mg3_c | 3060 | 250 | 171617 | 27.70% | 11.80% | 20.20% |
| 12 | mg4_c | 2010 | 342 | 242972 | 33.00% | 14.40% | 25.30% |
| 4 | mg5_c | 2204 | 312 | 243888 | 28.80% | 12.30% | 19.50% |
| 4 | mg6_c | 2154 | 326 | 249564 | 26.90% | 12.70% | 22.10% |
| 4 | mg7_c | 2532 | 184 | 230290 | 27.90% | 12.20% | 21.70% |
| 4 | mg8_c | 1555 | 400 | 315945 | 32.00% | 14.20% | 35.20% |
| 4 | mg9_c | 1639 | 500 | 311692 | 36.10% | 15.50% | 39.10% |
| 4 | wnv_mg5 | 5081 | 155 | 123075 | 31.00% | 13.70% | 22.50% |
| 4 | wnv_mg6 | 1992 | 382 | 286366 | 26.60% | 10.90% | 28.10% |
| 4 | wnv_mg7 | 1691 | 374 | 326296 | 19.50% | 7.40% | 25.50% |
| 4 | wnv_mg8 | 2471 | 423 | 196453 | 21.80% | 8.80% | 33.50% |
| 4 | wnv_mg9 | 2374 | 524 | 197463 | 28.50% | 11.30% | 38.80% |
| Average: | | 2417 | 358 | 255,432 | 28.15% | 11.96% | 26.94% |

**Table S3.4. Quality control (QC) metadata.**

| Sample ID | Pre-QC cell count | Post-QC cell count | Post-QC retention | Percent infected |
|---|---|---|---|---|
| wnv_mg3 | 996 | 868 | 87.10% | 86.75 |
| wnv_mg4 | 4077 | 3741 | 91.80% | 97.89 |
| mg3_c | 3060 | 2834 | 92.60% | 0.00% |
| mg4_c | 2010 | 1858 | 92.40% | 0.00% |
| wnv_mg5 | 5081 | 3292 | 64.80% | 85.45 |
| wnv_mg6 | 1992 | 1683 | 84.50% | 76.83 |
| wnv_mg7 | 1691 | 1429 | 84.50% | 85.65 |
| wnv_mg8 | 2471 | 2160 | 87.40% | 72.82 |
| wnv_mg9 | 2374 | 2114 | 89.00% | 79.52 |
| mg5_c | 2204 | 1660 | 75.30% | 0.00% |
| mg6_c | 2145 | 1132 | 52.60% | 0.00% |
| mg7_c | 2532 | 1809 | 71.40% | 0.00% |
| mg8_c | 1555 | 1395 | 89.70% | 0.00% |
| mg9_c | 1639 | 1313 | 80.10% | 0.00% |

* Excluded from analyses due to high mitochondrial gene percentages in a large proportion of post QC cells

**Table S3.5. Cluster markers.**

| | | | | | | Mock | | | | WNV | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cell type | Timepoint | Cluster | Marker_ID_method | Gene ID | Gene name/description | avg log2FC | pct.1 | pct.2 | p-value adj | avg log2FC | pct.1 | pct.2 | p-value adj |
| EC | 4 | 7 | CNSRV | gene11957 | POU2F1 | 2.549 | 0.468 | 0.049 | 1.47E-171 | 2.439 | 0.433 | 0.051 | 1.75E-197 |
| EC | 4 | 7 | CNSRV | gene316 | PLA2G6 | 2.035 | 0.325 | 0.05 | 1.36E-78 | 2.503 | 0.438 | 0.071 | 5.05E-151 |
| EC | 4 | 7 | CNSRV | gene769 | AGBL5 | 1.833 | 0.331 | 0.038 | 3.99E-106 | 1.931 | 0.304 | 0.051 | 1.22E-94 |
| EE | 4 | 10 | CNSRV | gene10632 | PROX1 | 1.913 | 0.291 | 0.002 | 2.10E-236 | 2.237 | 0.421 | 0.004 | 0 |
| VM-1 | 4 | 4 | CNSRV | gene2985 | ACTB | 4.782 | 0.993 | 0.209 | 0 | 4.989 | 0.993 | 0.136 | 0 |
| VM-1 | 4 | 4 | CNSRV | gene9804 | Mlc2 | 4.726 | 0.965 | 0.146 | 0 | 4.979 | 0.988 | 0.1 | 0 |
| VM-1 | 4 | 4 | CNSRV | gene11104 | Mhc/Mhc1 | 3.677 | 0.972 | 0.233 | 0 | 4.318 | 0.986 | 0.183 | 0 |
| VM-2 | 4 | 8 | CNSRV | gene11104 | Mhc/Mhc1 | 2.910 | 0.968 | 0.25 | 1.41E-230 | 3.241 | 0.978 | 0.189 | 0 |
| VM-2 | 4 | 8 | CNSRV | gene2985 | ACTB | 1.926 | 0.9 | 0.234 | 3.51E-178 | 2.874 | 0.919 | 0.146 | 0 |
| VM-2 | 4 | 8 | CNSRV | gene9804 | Mlc2 | 1.799 | 0.783 | 0.177 | 1.70E-155 | 2.676 | 0.832 | 0.114 | 0 |
| HC-2 | 4 | 13 | CNSRV | gene9384 | SCRASP1 | 2.464 | 0.528 | 0.003 | 0 | 2.644 | 0.646 | 0.004 | 0 |
| HC-2 | 4 | 13 | CNSRV | gene452 | SPARC | 4.169 | 1 | 0.15 | 1.51E-51 | 4.631 | 1 | 0.119 | 6.60E-176 |
| HC-1 | 4 | 12 | CNSRV | gene7512 | PebIII | 5.743 | 0.485 | 0.012 | 2.73E-278 | 5.510 | 0.75 | 0.012 | 4.66E-144 |
| HC-1 | 4 | 12 | CNSRV | gene1416 | PebIII | 4.406 | 0.39 | 0.005 | 1.53E-294 | 7.798 | 0.562 | 0.001 | 0 |
| Cardia-1 | 4 | 9 | CNSRV | gene7832 | Trypsin-like serine protease | 5.910 | 0.961 | 0.153 | 1.01E-297 | 6.971 | 1 | 0.063 | 0 |
| Cardia-1 | 4 | 9 | CNSRV | gene1865 | peptidase S1 family | 3.657 | 0.232 | 0.02 | 4.02E-78 | 1.861 | 0.252 | 0.005 | 1.85E-177 |
| Cardia-1 | 4 | 9 | CNSRV | gene2767 | peptidase S1 family | 3.118 | 0.78 | 0.177 | 2.44E-150 | 4.244 | 0.772 | 0.093 | 3.81E-147 |
| Cardia-1 | 4 | 9 | CNSRV | gene12143 | peptidase S1 family | 1.582 | 0.339 | 0.199 | 3.00E-04 | 1.245 | 0.378 | 0.22 | 0.736 |
| Cardia-1 | 4 | 9 | CNSRV | gene5114 | Trypsin-like serine protease | 1.101 | 0.268 | 0.058 | 1.08E-33 | 1.219 | 0.354 | 0.067 | 2.53E-30 |
| Cardia-1 | 4 | 9 | CNSRV | gene5111 | peptidase S1 family | 0.809 | 0.343 | 0.169 | 3.43E-07 | 0.937 | 0.433 | 0.21 | 0.0019956 |
| Cardia-1 | 4 | 9 | CNSRV | gene11373 | Mal-A4 | 5.760 | 0.972 | 0.298 | 1.43E-197 | 7.377 | 0.961 | 0.079 | 3.03E-284 |
| Cardia-1 | 4 | 9 | CNSRV | gene14075 | Mal-A1 | 5.545 | 0.894 | 0.138 | 1.86E-269 | 7.027 | 0.953 | 0.03 | 0 |
| Cardia-1 | 4 | 9 | CNSRV | gene5056 | Mal-A4 | 4.085 | 0.646 | 0.029 | 0 | 3.960 | 0.622 | 0.013 | 0 |
| Cardia-1 | 4 | 9 | CNSRV | gene11375 | Mal-A4 | 3.305 | 0.524 | 0.017 | 0 | 4.124 | 0.685 | 0.006 | 0 |
| Cardia-1 | 4 | 9 | CNSRV | gene11376 | Mal-A4 | 3.063 | 0.52 | 0.028 | 6.69E-270 | 3.599 | 0.614 | 0.011 | 0 |
| Cardia-1 | 4 | 9 | CNSRV | gene927 | Mal-A4 | 1.366 | 0.563 | 0.266 | 2.87E-28 | 1.418 | 0.559 | 0.323 | 4.27E-06 |
| Cardia-1 | 4 | 9 | CNSRV | gene9049 | Sugar_tr_CPIJ019592 | 1.834 | 0.327 | 0.038 | 8.94E-87 | 2.521 | 0.551 | 0.047 | 5.84E-132 |
| Cardia-1 | 4 | 9 | CNSRV | gene3360 | Sugar_tr_CPIJ014327 | 1.248 | 0.354 | 0.159 | 4.46E-13 | 1.316 | 0.496 | 0.193 | 7.80E-15 |
| Cardia-1 | 4 | 9 | CNSRV | gene3283 | Sugar_tr_CPIJ011910 | 1.215 | 0.35 | 0.154 | 2.23E-12 | 1.507 | 0.433 | 0.186 | 3.43E-09 |
| Cardia-1 | 4 | 9 | CNSRV | gene11820 | Sugar_tr_CPIJ012675 | 0.871 | 0.205 | 0.02 | 2.57E-59 | 0.839 | 0.213 | 0.015 | 1.01E-54 |
| Cardia-1 | 4 | 9 | CNSRV | gene9132 | transmembrane transport | 0.498 | 0.213 | 0.075 | 5.00E-08 | 1.098 | 0.228 | 0.045 | 1.74E-15 |
| Cardia-1 | 4 | 9 | CNSRV | gene6861 | chitin-binding_CPIJ004734 | 3.555 | 0.709 | 0.096 | 4.45E-205 | 2.776 | 0.63 | 0.103 | 1.05E-81 |
| Cardia-1 | 4 | 9 | CNSRV | gene13589 | chitin-binding_CPIJ004734 | 3.246 | 0.524 | 0.084 | 1.84E-119 | 2.368 | 0.472 | 0.087 | 2.91E-47 |
| Cardia-1 | 4 | 9 | CNSRV | gene4758 | chitin binding | 1.977 | 0.567 | 0.234 | 4.82E-37 | 1.283 | 0.512 | 0.245 | 1.26E-06 |
| Cardia-2 | 4 | 11 | CNSRV | gene12446 | C-type lysozyme | 1.704 | 0.677 | 0.132 | 8.20E-30 | 2.184 | 0.713 | 0.148 | 2.67E-65 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cardia-2 | 4 | 11 | CNSRV | gene3360 | Sugar_tr_CPIJ014327 | 1.406 | 0.738 | 0.161 | 1.76E-28 | 1.317 | 0.757 | 0.187 | 2.84E-55 |
| Cardia-2 | 4 | 11 | CNSRV | gene6862 | chitin-binding | 3.552 | 0.846 | 0.277 | 4.09E-30 | 3.077 | 0.853 | 0.224 | 1.15E-76 |
| EC-like-1 | 4 | 0 | CNSRV | gene14180 | serine_peptidase_CPIJ006568 | 2.497 | 0.882 | 0.374 | 0 | 2.574 | 0.898 | 0.372 | 0 |
| EC-like-1 | 4 | 0 | CNSRV | gene4641 | peptidase S1 family | 2.775 | 0.864 | 0.415 | 2.25E-272 | 3.028 | 0.945 | 0.526 | 0 |
| EC-like-1 | 4 | 0 | CNSRV | gene3328 | peptidase S1 family | 1.614 | 0.577 | 0.178 | 3.00E-153 | 1.339 | 0.572 | 0.207 | 1.84E-178 |
| EC-like-1 | 4 | 0 | CNSRV | gene5110 | serine_peptidase_CPIJ015103 | 1.492 | 0.898 | 0.508 | 1.29E-186 | 1.141 | 0.879 | 0.571 | 5.68E-189 |
| EC-like-1 | 4 | 0 | CNSRV | gene12764 | KLK7 | 1.395 | 0.367 | 0.063 | 1.83E-151 | 1.056 | 0.29 | 0.049 | 2.47E-165 |
| EC-like-1 | 4 | 0 | CNSRV | gene2767 | serine_peptidase_CPIJ004984 | 1.258 | 0.302 | 0.183 | 2.00E-09 | 0.848 | 0.177 | 0.086 | 2.47E-19 |
| EC-like-1 | 4 | 0 | CNSRV | gene5109 | peptidase S1 family | 1.087 | 0.493 | 0.181 | 1.41E-85 | 0.799 | 0.485 | 0.197 | 2.91E-100 |
| EC-like-1 | 4 | 0 | CNSRV | gene2591 | peptidase S1 family | 0.920 | 0.172 | 0.04 | 1.70E-46 | 0.505 | 0.116 | 0.032 | 6.17E-34 |
| EC-like-1 | 4 | 0 | CNSRV | gene12143 | peptidase S1 family | 0.810 | 0.418 | 0.16 | 5.18E-61 | 0.831 | 0.421 | 0.171 | 1.60E-85 |
| EC-like-1 | 4 | 0 | CNSRV | gene5111 | peptidase S1 family | 0.446 | 0.295 | 0.151 | 1.19E-18 | 0.458 | 0.349 | 0.179 | 7.85E-36 |
| EC-like-1 | 4 | 0 | CNSRV | gene5114 | peptidase S1 family | 0.362 | 0.128 | 0.054 | 1.95E-10 | 0.307 | 0.12 | 0.06 | 7.76E-10 |
| EC-like-1 | 4 | 0 | CNSRV | gene10257 | Mal-B2_CPIJ013170 | 2.369 | 0.644 | 0.15 | 7.84E-242 | 2.266 | 0.645 | 0.136 | 0 |
| EC-like-1 | 4 | 0 | CNSRV | gene927 | Mal-A4_CPIJ010128 | 1.872 | 0.655 | 0.2 | 4.39E-178 | 1.818 | 0.701 | 0.229 | 9.04E-285 |
| EC-like-1 | 4 | 0 | CNSRV | gene9012 | atgp-2 | 0.417 | 0.112 | 0.021 | 2.91E-34 | 0.502 | 0.152 | 0.029 | 2.11E-70 |
| EC-like-1 | 4 | 0 | CNSRV | gene12593 | Mal-B1_CPIJ013172 | 1.512 | 0.157 | 0.026 | 5.94E-60 | 1.148 | 0.184 | 0.036 | 5.43E-89 |
| EC-like-2 | 4 | 14 | FindMarkers | gene14180 | serine_peptidase_CPIJ006568 | 1.314 | 0.917 | 0.462 | 0.008 | - | - | - | - |
| EC-like-2 | 4 | 14 | FindMarkers | gene3328 | peptidase S1 family | 1.710 | 0.708 | 0.247 | 0.007 | - | - | - | - |
| EC-like-2 | 4 | 14 | FindMarkers | gene10257 | Mal-B2_CPIJ013170 | 1.900 | 0.833 | 0.235 | 3.23E-08 | - | - | - | - |
| ISC/EB-prol | 4 | 6 | CNSRV | gene2167 | PCNA | 1.619 | 0.609 | 0.049 | 4.00E-301 | 1.383 | 0.585 | 0.047 | 0 |
| ISC/EB-prol | 4 | 6 | FeaturePlot | gene13413 | AURKA | - | - | - | - | - | - | - | - |
| ISC/EB-prol | 4 | 6 | FeaturePlot | gene12926 | AURKB | - | - | - | - | - | - | - | - |
| ISC/EB-prol | 4 | 6 | FeaturePlot | gene13104 | klumpfuss | - | - | - | - | - | - | - | - |
| ISC/EB | 4 | 2 | FeaturePlot | gene13104 | klumpfuss | - | - | - | - | - | - | - | - |
| EC | 12 | 7 | CNSRV | gene11957 | POU2F1 | 2.118 | 0.32 | 0.039 | 6.75E-88 | 2.278 | 0.496 | 0.076 | 9.56E-111 |
| EC | 12 | 7 | CNSRV | gene316 | PLA2G6 | 2.136 | 0.379 | 0.063 | 1.62E-79 | 2.067 | 0.482 | 0.069 | 1.44E-110 |
| EC | 12 | 7 | CNSRV | gene769 | AGBL5 | 1.931 | 0.259 | 0.022 | 1.66E-93 | 1.844 | 0.345 | 0.029 | 8.28E-121 |
| EE | 12 | 13 | FeaturePlot | gene10632 | PROX1 | - | - | - | - | - | - | - | - |
| VM | 12 | 9 | CNSRV | gene2985 | ACTB | 6.111 | 0.928 | 0.061 | 0 | 5.722 | 0.88 | 0.102 | 0 |
| VM | 12 | 9 | CNSRV | gene11104 | Mhc/Mhc1 | 5.130 | 0.887 | 0.064 | 0 | 4.962 | 0.934 | 0.105 | 0 |
| VM | 12 | 9 | CNSRV | gene9804 | Mlc2 | 5.241 | 0.894 | 0.028 | 0 | 5.229 | 0.846 | 0.058 | 0 |
| HC | 12 | 16 | CNSRV | gene452 | SPARC | 4.128 | 0.571 | 0.067 | 2.55E-15 | 3.033 | 0.429 | 0.083 | 0.006 |
| HC | 12 | 16 | CNSRV | gene7512 | PebIII | 6.151 | 0.524 | 0.004 | 5.65E-178 | 4.189 | 0.476 | 0.011 | 6.08E-71 |
| HC | 12 | 16 | CNSRV | gene1416 | PebIII | 5.625 | 0.333 | 0 | 4.56E-292 | 5.976 | 0.19 | 0.001 | 1.75E-79 |
| EC-like-1 | 12 | 0 | FindMarkers | gene14180 | serine_peptidase_CPIJ006568 | - | - | - | - | 1.070 | 0.922 | 0.599 | 9.31E-177 |
| EC-like-1 | 12 | 0 | FindMarkers | gene13667 | Trypsin-like serine protease | - | - | - | - | 1.497 | 0.554 | 0.196 | 2.30E-134 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| EC-like-1 | 12 | 0 | FindMarkers | gene5110 | serine_peptidase_CPIJ015103 | - | - | - | - | 1.007 | 0.914 | 0.678 | 1.18E-145 |
| EC-like-1 | 12 | 0 | FindMarkers | gene5109 | peptidase S1 family | - | - | - | - | 0.917 | 0.333 | 0.195 | 3.69E-24 |
| EC-like-1 | 12 | 0 | FindMarkers | gene3328 | peptidase S1 family | - | - | - | - | 0.796 | 0.436 | 0.278 | 3.75E-26 |
| EC-like-1 | 12 | 0 | FindMarkers | gene12143 | peptidase S1 family | - | - | - | - | 0.755 | 0.213 | 0.122 | 6.45E-11 |
| EC-like-1 | 12 | 0 | FindMarkers | gene2984 | Trypsin-like serine protease | - | - | - | - | 0.635 | 0.11 | 0.048 | 4.70E-09 |
| EC-like-1 | 12 | 0 | FindMarkers | gene7832 | Trypsin-like serine protease | - | - | - | - | 0.607 | 0.451 | 0.197 | 1.05E-62 |
| EC-like-1 | 12 | 0 | FindMarkers | gene2767 | serine_peptidase_CPIJ004984 | - | - | - | - | 0.274 | 0.133 | 0.075 | 0.001 |
| EC-like-1 | 12 | 0 | FindMarkers | gene4641 | peptidase S1 family | - | - | - | - | 0.252 | 0.924 | 0.748 | 5.26E-56 |
| EC-like-1 | 12 | 0 | FindMarkers | gene12593 | Mal-B1_CPIJ013172 | - | - | - | - | 0.901 | 0.227 | 0.097 | 2.25E-26 |
| EC-like-1 | 12 | 0 | FindMarkers | gene10257 | Mal-B2_CPIJ013170 | - | - | - | - | 1.009 | 0.561 | 0.31 | 2.02E-65 |
| EC-like-1 | 12 | 0 | FindMarkers | gene11373 | Mal-A4 | - | - | - | - | 0.975 | 0.79 | 0.33 | 4.18E-195 |
| EC-like-1 | 12 | 0 | FindMarkers | gene927 | Mal-A4_CPIJ010128 | - | - | - | - | 0.720 | 0.553 | 0.328 | 6.72E-44 |
| EC-like-1 | 12 | 0 | FindMarkers | gene14075 | Mal-A1 | - | - | - | - | 0.586 | 0.337 | 0.146 | 1.35E-42 |
| EC-like-2 | 12 | 3 | FindMarkers | gene14180 | serine_peptidase_CPIJ006568 | 1.186 | 0.811 | 0.554 | 3.33E-99 | - | - | - | - |
| EC-like-2 | 12 | 3 | FindMarkers | gene1866 | peptidase S1 family | 2.351 | 0.349 | 0.053 | 6.12E-144 | - | - | - | - |
| EC-like-2 | 12 | 3 | FindMarkers | gene2767 | serine_peptidase_CPIJ004984 | 1.549 | 0.543 | 0.154 | 8.17E-143 | - | - | - | - |
| EC-like-2 | 12 | 3 | FindMarkers | gene13667 | Trypsin-like serine protease | 1.250 | 0.291 | 0.108 | 8.51E-42 | - | - | - | - |
| EC-like-2 | 12 | 3 | FindMarkers | gene5110 | serine_peptidase_CPIJ015103 | 1.085 | 0.799 | 0.571 | 1.56E-82 | - | - | - | - |
| EC-like-2 | 12 | 3 | FindMarkers | gene7832 | Trypsin-like serine protease | 0.818 | 0.501 | 0.179 | 3.48E-87 | - | - | - | - |
| EC-like-2 | 12 | 3 | FindMarkers | gene12593 | Mal-B1_CPIJ013172 | 1.169 | 0.154 | 0.06 | 6.63E-16 | - | - | - | - |
| EC-like-2 | 12 | 3 | FindMarkers | gene11373 | Mal-A4 | 1.326 | 0.929 | 0.354 | 1.75E-284 | - | - | - | - |
| EC-like-2 | 12 | 3 | FindMarkers | gene927 | Mal-A4_CPIJ010128 | 0.935 | 0.487 | 0.323 | 1.25E-27 | - | - | - | - |
| EC-like-2 | 12 | 3 | FindMarkers | gene10257 | Mal-B2_CPIJ013170 | 0.894 | 0.41 | 0.278 | 2.20E-17 | - | - | - | - |
| EC-like-2 | 12 | 3 | FindMarkers | gene14075 | Mal-A1 | 0.770 | 0.453 | 0.158 | 5.07E-78 | - | - | - | - |
| EC-like-2 | 12 | 3 | FindMarkers | gene11375 | Mal-A4 | 0.745 | 0.113 | 0.05 | 1.82E-06 | - | - | - | - |
| EC-like-3 | 12 | 6 | CNSRV | gene14180 | peptidase S1 family | 0.900 | 0.949 | 0.677 | 8.64E-28 | 0.800 | 0.927 | 0.573 | 3.82E-36 |
| EC-like-3 | 12 | 6 | CNSRV | gene12764 | peptidase S1 family | 0.842 | 0.606 | 0.148 | 8.90E-60 | 0.837 | 0.537 | 0.119 | 5.28E-80 |
| EC-like-3 | 12 | 6 | CNSRV | gene4641 | peptidase S1 family | 1.661 | 0.982 | 0.787 | 2.89E-65 | 1.445 | 0.992 | 0.832 | 7.16E-76 |
| EC-like-3 | 12 | 6 | CNSRV | gene5111 | peptidase S1 family | 0.314 | 0.368 | 0.128 | 6.91E-16 | 0.276 | 0.365 | 0.118 | 5.60E-27 |
| EC-like-3 | 12 | 6 | CNSRV | gene5114 | Trypsin-like serine protease | 0.336 | 0.217 | 0.044 | 6.63E-26 | 0.373 | 0.232 | 0.054 | 2.92E-31 |
| EC-like-3 | 12 | 6 | CNSRV | gene9741 | GBE1 | 0.798 | 0.39 | 0.083 | 5.98E-48 | 0.843 | 0.363 | 0.074 | 1.00E-61 |
| EC-like-3 | 12 | 6 | CNSRV | gene927 | Mal-A4 | 0.838 | 0.78 | 0.369 | 1.03E-24 | 0.767 | 0.811 | 0.312 | 2.70E-47 |
| Cardia | 12 | 11 | CNSRV | gene7832 | Trypsin-like serine protease | 4.195 | 0.952 | 0.207 | 5.61E-184 | 4.367 | 0.948 | 0.262 | 3.40E-45 |
| Cardia | 12 | 11 | CNSRV | gene2767 | peptidase S1 family | 3.281 | 0.808 | 0.2 | 8.06E-106 | 3.479 | 0.569 | 0.086 | 3.46E-34 |
| Cardia | 12 | 11 | CNSRV | gene1865 | peptidase S1 family | 2.181 | 0.389 | 0.028 | 6.39E-129 | 1.849 | 0.379 | 0.019 | 3.28E-66 |
| Cardia | 12 | 11 | CNSRV | gene5114 | Trypsin-like serine protease | 1.400 | 0.351 | 0.056 | 5.63E-53 | 1.493 | 0.345 | 0.051 | 2.00E-16 |
| Cardia | 12 | 11 | CNSRV | gene5111 | peptidase S1 family | 0.273 | 0.312 | 0.131 | 9.20E-05 | 0.815 | 0.431 | 0.139 | 0.005 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cardia | 12 | 11 | CNSRV | gene14075 | Mal-A1 | 4.174 | 0.889 | 0.182 | 2.86E-169 | 4.376 | 0.81 | 0.194 | 2.99E-37 |
| Cardia | 12 | 11 | CNSRV | gene11373 | Mal-A4 | 3.973 | 0.981 | 0.439 | 6.58E-120 | 4.275 | 0.931 | 0.458 | 3.62E-26 |
| Cardia | 12 | 11 | CNSRV | gene11375 | Mal-A4 | 3.213 | 0.615 | 0.037 | 1.74E-244 | 3.164 | 0.603 | 0.043 | 3.23E-81 |
| Cardia | 12 | 11 | CNSRV | gene5056 | Mal-A4 | 3.038 | 0.404 | 0.034 | 1.04E-121 | 3.145 | 0.379 | 0.029 | 7.40E-45 |
| Cardia | 12 | 11 | CNSRV | gene11376 | Mal-A4 | 2.787 | 0.514 | 0.015 | 3.50E-291 | 2.333 | 0.448 | 0.015 | 1.11E-113 |
| Cardia | 12 | 11 | CNSRV | gene9049 | Sugar_tr_CPIJ019592 | 2.346 | 0.471 | 0.029 | 1.57E-174 | 2.441 | 0.638 | 0.039 | 3.61E-96 |
| Cardia | 12 | 11 | CNSRV | gene3283 | Sugar_tr_CPIJ011910 | 1.691 | 0.548 | 0.156 | 1.38E-42 | 1.678 | 0.638 | 0.169 | 7.29E-15 |
| Cardia | 12 | 11 | CNSRV | gene3360 | Sugar_tr_CPIJ014327 | 1.079 | 0.346 | 0.083 | 2.52E-29 | 1.391 | 0.483 | 0.123 | 1.11E-09 |
| Cardia | 12 | 11 | CNSRV | gene11820 | Sugar_tr_CPIJ012675 | 0.776 | 0.159 | 0.007 | 8.13E-68 | 0.816 | 0.172 | 0.009 | 1.97E-26 |
| Cardia | 12 | 11 | CNSRV | gene6403 | Sugar_tr_CPIJ012678 | 0.760 | 0.154 | 0.002 | 2.48E-104 | 0.848 | 0.19 | 0.003 | 1.96E-72 |
| Cardia | 12 | 11 | CNSRV | gene7250 | Sugar transporter | 0.583 | 0.12 | 0.002 | 1.41E-76 | 0.717 | 0.103 | 0.002 | 4.06E-38 |
| Cardia | 12 | 11 | CNSRV | gene6861 | chitin-binding_CPIJ004734 | 2.964 | 0.673 | 0.081 | 3.56E-160 | 2.368 | 0.5 | 0.089 | 3.94E-21 |
| Cardia | 12 | 11 | CNSRV | gene13589 | chitin-binding_CPIJ004734 | 2.941 | 0.591 | 0.058 | 1.01E-163 | 1.605 | 0.362 | 0.07 | 1.08E-10 |
| Cardia | 12 | 11 | CNSRV | gene4758 | chitin binding | 1.911 | 0.663 | 0.233 | 3.63E-43 | 2.090 | 0.552 | 0.183 | 1.80E-07 |
| ISC/EB-prol | 12 | 14 | CNSRV | gene2167 | PCNA | 1.282 | 0.587 | 0.027 | 3.62E-93 | 0.837 | 0.296 | 0.023 | 1.04E-13 |
| ISC/EB-prol | 12 | 14 | FeaturePlot | gene13413 | AURKA | - | - | - | - | - | - | - | - |
| ISC/EB-prol | 12 | 14 | FeaturePlot | gene12926 | AURKB | - | - | - | - | - | - | - | - |
| ISC/EB-prol | 12 | 14 | FeaturePlot | gene13104 | klumpfuss | - | - | - | - | - | - | - | - |
| ISC/EB | 12 | 5 | FeaturePlot | gene13104 | klumpfuss | - | - | - | - | - | - | - | - |