Theses and Dissertations   Theses and Dissertations

5-12-2023

Host-acquired virus genes support an ancient antiviral role of the piRNA pathway in dipterans

Rebecca Christian

Mississippi State University, rchristian1210@gmail.com

Host-acquired virus genes support an ancient antiviral role of the piRNA pathway in dipterans

By

Rebecca Christian

Approved by:

Matthew Ballinger (Major Professor)
Jean Francois Gout
Ying Wang
Justin A. Thornton (Graduate Coordinator)
Rick Travis (Dean, College of Arts & Sciences)

A Thesis
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in Biology
in the College of Arts and Sciences

Mississippi State, Mississippi

May 2023

Name: Rebecca Christian

Date of Degree: May 12, 2023

Institution: Mississippi State University

Major Field: Biology

Major Professor: Matthew Ballinger

Title of Study: Host-acquired virus genes support an ancient antiviral role of the piRNA pathway in dipterans

Pages in Study 46

Candidate for Degree of Master of Science

Endogenous viral elements (EVEs) have been recently investigated as a source of transgenerational immune memory. These "viral fossils" are abundant in *Aedes* mosquitoes and partner with the host's primary antiviral defense system, the RNA interference (RNAi) pathways. This partnership appears unique to mosquitoes, which encode an expansion of the Piwi endoribonucleases. To interrogate EVE-Piwi partnerships and their role in antiviral defense, I performed a comparative small RNA analysis of two naturally occurring EVE-virus pairs – one in the mosquito *Aedes albopictus*, and one in the midge *Chaoborus americanus*. Both express an EVE related to the nucleoprotein of their respective bunyavirus. My results show that Piwis generally do not have antiviral functions in *Chaoborus*, however EVEs are associated with Piwi recruitment to matched viral RNAs. These findings raise the possibility that RNAi-mediated EVE-virus interactions may be more common among insects than currently recognized.

ACKNOWLEDGMENTS

Foremost, I would like to express my deepest appreciation for my advisor, Dr. Matthew Ballinger, for the continuous support of my study and research. His patience, enthusiasm, infinite knowledge, and humor have kept me motivated throughout my entire time as a student. His guidance has been irreplaceable, and I could not imagine having a better advisor and mentor for my thesis project.

Besides my advisor, I would like to extend many thanks to the rest of my committee members: Dr. Ying Wang and Dr. Jean Francois Gout. They have given me expert advice and have pushed me to think beyond the scope of my studies. I appreciate all their help in making me both a better writer and a better scientist.

I would also like to recognize my fellow lab mates in the Ballinger Lab: Logan Moore and Tolu Chris-Amuwa for always encouraging me throughout the entire process. I could not thank them enough for making my research experience so enjoyable. I consider them not only lab mates, but also incredible friends.

Lastly, I would like to give a special thanks to my mom, dad and two brothers who have always believed in me. A special thanks to my dearest friends, Hallie Troell and Kensey Thomas, and my partner Jordan Besson for keeping me motivated and loving me through all the great times and the hard times.

TABLE OF CONTENTS

CHAPTER

APPENDIX

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Adaptive and innate immunity

All organisms utilize their immune system to combat invading pathogens. Immune

systems can be generally classified into two fundamental lines of defense: adaptive immunity

and innate immunity. Both types offer their own unique mechanisms of defense, although not all

organisms utilize both types. Adaptive immunity exists exclusively within vertebrate hosts and is

based on the production of lymphocytes containing antigen-specific receptors (Janeway et al.

2001). Adaptive immunity has the unique ability to specifically recognize invading pathogens

and "remember" them during reinfection to provide enhanced protection (Cooper and

Eleftherianos 2017). Innate immunity can be found in vertebrates, invertebrates, and plants and is

considered to be broad and non-specific (Brubaker et al. 2015). In invertebrate organisms, the

innate immune system is capable of fighting against many pathogens, but it is restricted to

relaying on germ-line encoded receptors to recognize invading microorganisms (Janeway et al.

2001). Traditionally, it does not have the same immunological memory capabilities that the

adaptive immune system has. However, it has recently been found that some insects have

evolved "immune memory"-like processes in antiviral defense by transmitting repurposed virus

genome sequences to their offspring. These repurposed viral genomes are known as endogenous

viral elements (EVEs). EVEs are sequences of viral DNA or RNA that are integrated into a

host's genome during infection (Tassetto et al. 2019). They are found in a wide range of both vertebrate and invertebrate organisms and can originate from many different types of viruses.

Endogenous viral elements in vertebrate and non-vertebrate hosts

In vertebrate hosts, most EVEs are integrated during infection by a reverse transcribing RNA virus (i.e. retroviruses) (Tristem 2000). Retroviruses use the host's genome and polymerases as part of their own replication process and are therefore guaranteed to leave their own genetic material behind (Lower et al. 1996). When this occurs within a germline cell, that genetic material is able to be vertically transmitted to offspring (Tristem 2000). These sequences are often eliminated from the host population after a few generations. However, some have become fixed in host populations and have been passed down for millions of years (Mager and Freeman 1995).

In contrast, nonretroviral EVEs have been unexpectedly identified across the tree of life, including in at least eight orders of insects and have origins from at least six families of DNA viruses and 22 families of RNA viruses (Gilbert and Belliardo 2022). In insects, most EVEs are reverse-transcribed and integrated into the host's genome after infection by a negative sense single stranded (-ss) RNA virus (ter Horst et al. 2019). The exact reason is unknown, but one possibility is that reverse transcription occurs more efficiently on the shorter mRNAs produced by -ssRNA viruses in comparison to other viruses (Horie et al. 2010 and ter Horst et al. 2019).

EVEs bring a unique and important contribution to the study of virus evolution. They serve an invaluable role in paleovirology, or the study of ancient, extinct viruses. EVEs are often described as "molecular fossils" of ancient viruses that once infected organisms (Emerman and Malik 2010). They are used to infer information about these viruses that we would not otherwise be able to access. For example, orthologous EVEs in closely related organisms can be used to

estimate the date of viral insertion into a host's genome (Etienne and Emerman 2013). This can

lead to an improved understanding of age, geographical distribution, genetic information,

replication strategies, and host distribution (Horie and Tomonaga 2019). Understanding these

ancient viruses and their evolution to modern day viruses can have implications for modern

human pandemics. For example, a multi-organism study of EVEs derived from Hepatitis B

showed that this virus existed within passerine bird populations at least 19 million years before

humans (Etienne and Emerman 2013). Continuing the study of EVEs will further our

understanding of these high-risk viruses that affect us today.

RNA interference pathways

In recent years, increasing evidence suggests nonretroviral EVEs are not just

informational relics, but functional elements. For example, mosquitoes incorporate EVEs into

their RNA interference (RNAi) machinery to identify and destroy exogenous viral RNAs

(Tassetto et al. 2019). There are three main pathways involved in invertebrate RNA interference:

the microRNA (miRNA) pathway, the small interfering RNA (siRNA) pathway, and the Piwi

interacting RNA (piRNA) pathway. All are biological process within a cell where RNA

molecules are used to inhibit gene translation or expression. However, the siRNA pathway is the

major antiviral defense mechanism in insects (Petit et al. 2016).

In the siRNA pathway, exogenously derived double-stranded viral RNAs are processed

into 19 base pair (bp) siRNA duplexes by the endonuclease Dicer-2. These duplexes are

composed of two 21 nucleotide (nt) single-stranded RNA molecules and contain characteristic 2

nt overhangs at each 3' end of the duplex, a recognition signal that facilitates loading onto the

endonuclease Argonaute-2 (Ago2)-containing RNA-induced silencing complex (RISC)

(Gammon and Mello 2015). An activated RISC is able to use the guide strand of the siRNA

duplex as a sequencing guide to search for complementary viral RNA genomes and transcripts (Liu et al. 2011). It then cleaves one strand of the viral RNA and thus interferes with viral replication. This process occurs in infected somatic cells; however, some dipteran insects have antiviral signals that travel to uninfected, healthy cells. Virus infection induces dsRNA release from infected cells, which are then taken up by uninfected cells in order to generate virus-specific immunity to prevent virus spread thus creating a system wide antiviral response (Saleh et al. 2009).

In *Drosophila*, the piRNA pathway is Dicer-2-independent, non-antiviral, and occurs primarily in germline cells where its central role is to prevent the proliferation of transposable genetic elements (Brenneke et al. 2007). It is split into two parts: the primary piRNA biogenesis step and secondary piRNA amplification step (Petit et al. 2016) (Figure 1). Primary piRNA biogenesis begins with piRNA precursor transcripts, called piRNA clusters, located in the nucleus. These clusters are typically made up of transposons and are described as either dual-stranded or uni-stranded. Dual-stranded clusters map to both genomic strands while uni-stranded clusters only map to one (Mohn et al. 2014). However, it appears that the majority of piRNAs come from uni-stranded transcripts which are derived from antisense single-stranded (ss) RNAs (Brennecke et al. 2007). Uni-stranded clusters begin transcription at promoter regions that are marked by peaks of RNA polymerase II Ser5P and histone 3 lysine 4 di-methylation (H3K4me2). Once transcription begins, the transcript is processed by 5' capping, 3' end polyadenylation, and occasionally alternative splicing (Mohn et al. 2014; Huang et al. 2017) After transcription, these piRNA precursor transcripts are transported out of the nucleus and into the cytoplasm. This is often accomplished by cleaving introns located within the piRNA precursors which leads to the creation of the exon-junction complex. It is the binding of this

complex to the exportin complex that allows the transcripts to efficiently exit the nucleus (Goriaux et al. 2014). After entering the cytoplasm, these pre piRNAs are cleaved into 24-31 nt RNA fragments by the endonuclease Zucchini (Zuc) (Rogers et al. 2017). These 24-31 nt RNAs are referred to as primary RNAs. *In vitro*, Zuc has no preference towards the piRNA transcripts (Nishimasu et al. 2022). However, primary piRNAs show an overwhelming bias for uridine at their 5' end suggesting that there is an unknown co-factor involved (Ipsaro et al. 2012).

Primary RNAs can be further processed in what is referred to as the "ping-pong amplification cycle," also known as secondary piRNA biogenesis. In this cycle, primary piRNAs with a 5' uridine are preferentially loaded onto the Piwi protein Aubergine (Aub) (Saleh et al. 2009). These 5' uridine piRNA/Aub complexes will then seek and bind to complementary target transcripts (Wang et al. 2014). These are subsequentially cleaved ten positions away from the 5' uridine binding site, yielding piRNAs in the sense orientation with an adenine at position 10 (Saito and Siomi 2010). These piRNAs are then preferentially loaded onto Piwi protein Argonaute-3 (Ago3) and these piRNA/Ago3 complexes then seek and bind to additional complementary target transcripts (Wang et al. 2014). The transcripts are subsequentially cleaved, yielding piRNAs in the antisense orientation with a 5' uridine and the cycle continues (Tassetto et al. 2019). This cycle creates an amplification of piRNAs which allows for more piRNAs to seek and cleave target RNA molecules found in the host.
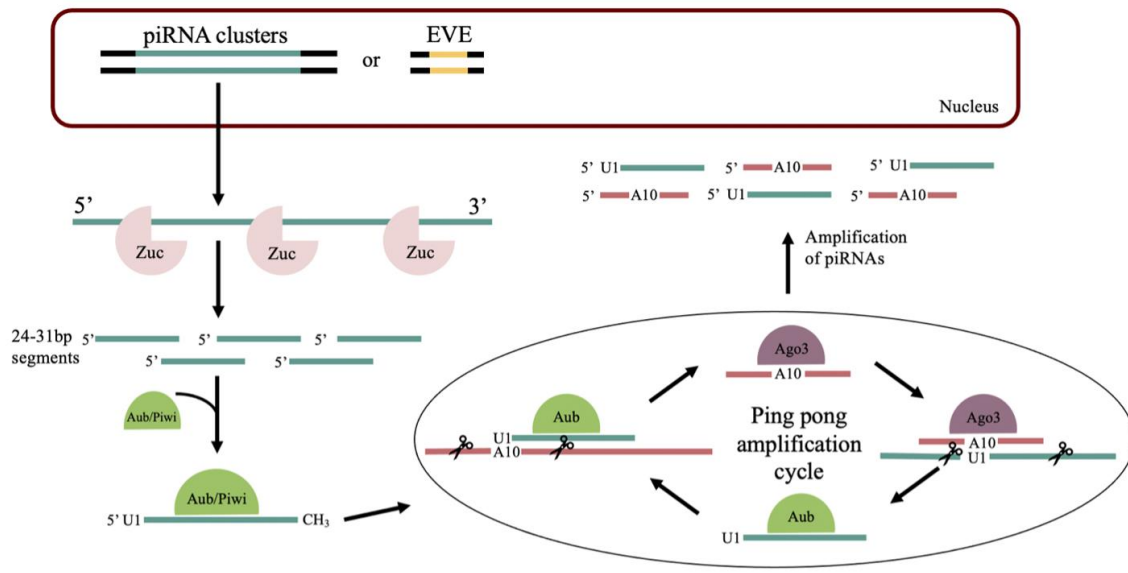
Figure 1   An overview of the piRNA pathway

This overview of the piRNA pathway includes the primary piRNA biogenesis step and the secondary piRNA amplification step referred to as the "ping-pong amplification cycle." EVEs can be used instead of transposons in piRNA clusters to create primary piRNAs.

Interestingly, EVEs play an unsuspected role in the piRNA pathway in *Aedes* mosquitoes. EVEs can be used instead of transposons in piRNA clusters to create primary piRNAs (Tassetto et al. 2019). In *Aedes* mosquitoes, EVE-derived piRNAs are specifically loaded onto Piwi4, one of many duplicated Piwi proteins unique to mosquitoes, to act as guide RNAs to inhibit virus replication (Tassetto et al. 2019). However, there is still little understanding of the mechanisms behind EVE-virus interactions and the role of EVEs in antiviral defense in other dipteran species. This is mainly because there are few natural insect systems that contain both an EVE and its corresponding infecting virus at the same time.

Virus and EVE systems

Our lab has identified two parallel insect-virus systems in which a nonretroviral RNA virus infects hosts with a closely related EVE. The viruses belong to the family Phasmaviridae within the order Bunyavirales. Bunyaviruses consist of segmented -ssRNA viruses (Neriya et al. 2022). They comprise a large order of over 560 viruses classified into twelve different families (Teng et al. 2022). Most families package a tripartite genome and share the same genetic organization consisting of the S, M, and L segment, based on their relative size (Neriya et al. 2022). Each segment encodes a different protein. The S segment encodes the nucleocapsid protein (NP), the M segment encodes a glycosylated polyprotein precursor (GnGc), and the L segment encodes an RNA-dependent RNA polymerase (RdRp) (Neriya et al. 2022). Phasmaviridae is a recently discovered family of Bunyaviruses. It is diverse, widespread, and insect specific (Ballinger and Taylor 2019). Like all Bunyaviruses, they contain segmented, negative sense RNA genomes (Ballinger and Taylor 2019). Bunyaviruses can infect both livestock and humans. Those that cause clinical symptoms usually cause hemorrhagic fevers and cardiopulmonary syndromes (Teng et al. 2022). Many of these viruses pose a threat to public health. Well known examples of Bunyaviruses include: the Lassa virus (LASV), La Crosse virus (LACV), Hantaan virus (HTNV), Andes virus (ANDV), and Rift Valley fever virus (RVFV) (Mehand et al. 2018).

In this work, I analyze Bunyaviruses within two different insect hosts: *Chaoborus americanus* (Order Diptera, Family Chaoboridae) and *Aedes albopictus* (Order Diptera, Family Culicidae). *C. americanus*, also known the phantom midge, is a key predator in fishless freshwater systems across temperate North America (von Ende 1979). *Ae. albopictus*, also known as the Asian tiger mosquito, is an important vector of diseases and is located on every

continent except Antarctica. The widespread expansion of these mosquitoes has caused an increased concern for public health. Among the viral diseases vectored by *Ae. albopictus* are globally important arboviruses including dengue virus, chikungunya virus, and yellow fever virus (Kraemer et al. 2015).

In my comparative system, *Ae. albopictus* is infected with Barstukas virus (BARV) and C. americanus is infected with Niukluk phantom virus (NUKV). Both of these viruses are found within the same clade of Phasmaviridae (Figure 2). BARV has been detected in *Ae. albopictus* mosquitoes in only two locations around the world: Guangzhou, China, and California, USA (Batson et al. 2020; Shi et al. 2020). Like most Bunyavirus, the BARV and NUKV genomes consist of three different segments: the S, M, and L segment. A recent ancestor of BARV is the source of the S segment EVE as they match at 74.7% at the amino acid level. NUKV has been detected in *C. americanus* across North America (Ballinger et al. 2022), and like BARV, its S segment is the source of an EVE that shares 78.5% at the amino acid level.
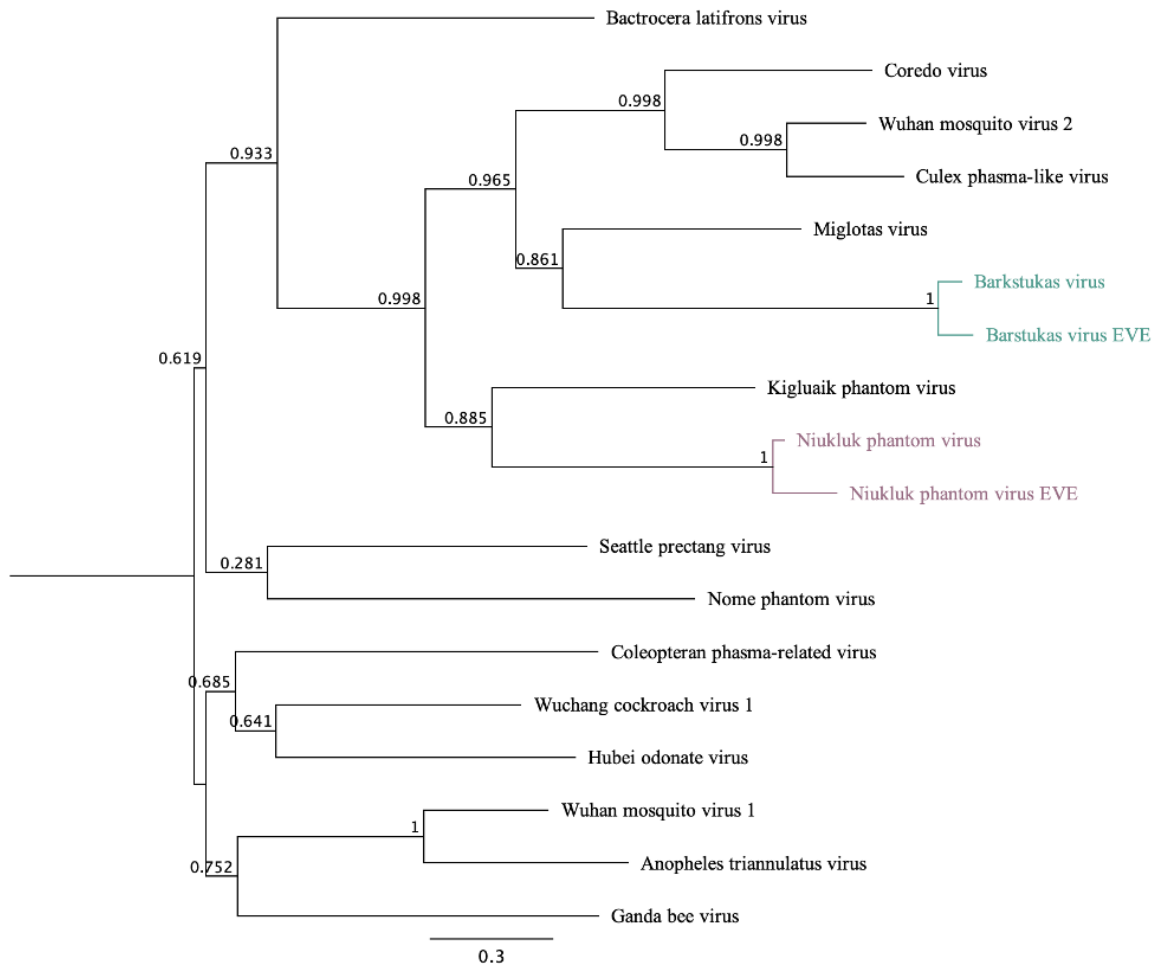
Figure 2   Evolutionary relationships of Phasmaviridae

Phylograms were built from full length nucleoprotein amino acid sequence alignments. Branch numbers indicate FastTree maximum likelihood-like support values. Phylogenetic analyses show that Barstukas virus and Niukluk virus belong to a well-supported clade of Diptera-infecting phasmavirids. It also shows that the EVEs of both viruses originated from recent ancestors of their nucleoproteins.

My interest in these two systems comes from their differences in two areas: their differences in RNAi responses and their differences in antiviral defense machinery. *C. americanus* midges' antiviral response is directed heavily by the siRNA pathway. They show an enrichment of 21 nt RNAs when mapped to multiple sets of endogenous RNA viruses (Ballinger et al. 2022). In contrast, *Ae. albopictus* mosquitoes show an enriched piRNA response. They

9

show an enrichment of 24-31 nt RNAs after mapping to multiple RNA viruses (Wang et al. 2018 and Morazzani et al. 2012). Their differences in RNAi machinery lies within the Piwi clade of endonucleases. Phylogenetic analysis of dipteran Ago proteins indicate a large expansion of the Piwi clade in *Ae. albopictus* mosquitoes (Campbell et al. 2008 and Morazzani et al. 2012). They encode and express Argonaute-3 (Ago3) as well as seven different Piwi proteins in their somatic cells (Morazzani et al. 2012 and Schnettler et al. 2013). This is in contrast to most other dipteran species which only encode and express one Piwi, one Aub, and one Ago3 protein, and these tend to be limited to germline cells (Schnettler et al. 2013 and Campbell et al. 2008).

My research aims to identify differences in EVE-virus interactions between these two systems. Using small RNA data sets, I studied EVE-Piwi partnerships in the two separate hosts: one with an expanded Piwi protein family and one with a conserved set of Piwi proteins. Specifically, I ask whether expanded Piwis are required to mediate interactions between EVEs and exogenous viruses. I expect this research to yield a better understanding of a novel, transgenerational mechanism of antiviral defense used by arthropods. First, I determine whether each host is producing piRNAs in response to viral infection and EVE expression. My results suggest that both species produce piRNAs in response to viral infection, but to a much greater extent in *Aedes* mosquitoes. Second, I analyze nucleotide compositions of all piRNAs produced to determine if there is a ping-pong signature present, indicating that the secondary piRNA amplification step was used. I identify clear ping-pong signatures present in virus-derived piRNAs as well as EVE-derived piRNAs in both species. However, in C. americanus I assess ping-pong signature for seven RNA viruses, including NUKV, and find the S segment is the only viral RNA being targeted by this piRNA pathway. My findings suggest the role of Piwi proteins as mediators of EVE-virus interactions is ancestral to the duplication and specialization of this

gene family in *Aedes* and other mosquitoes, raising the possibility that RNAi-mediated EVE-

virus interactions may be more common among insects than currently recognized.

CHAPTER II

MATERIALS AND METHODS

Tissue Sources

 Aedes albopictus small RNA data sets were collected from available public data on NCBI (Table 1). *Chaoborus americanus* small RNA datasets were generated by the Ballinger lab. In brief, *C. americanus* larvae were collected in freshwater ponds in Washington, United States and reared to adulthood. Ovaries or eggs were dissected from two-day old adults and total RNA was extracted using the miRNeasy Mini Kit (Qiagen). Small RNA sequencing libraries were built using the NEBNext Small RNA Library Prep Set for Illumina (New England Biolabs) and size-selected from a 5% polyacrylamide gel to retain siRNAs, miRNAs, and piRNAs. Four libraries (SNOA1 and A2) were sequenced on an Illumina NextSeq as 75 bp single end reads at HudsonAlpha. An additional four libraries (SNOA3 and A4) were sequenced on an Illumina NovaSeq as 50 bp single end reads at Novogene. Tissue sources and sequencing yields for *Chaoborus* samples are reported in Table 2.

Table 1   *Ae. albopictus* small RNA data sets used in this study.

SRA	Location	Tissue	Total Reads	NCBI Accession Number
SRR11252296	California, USA	Female carcass	5,537,655	SRX7862422
SRR10390736	Vietnam	Female carcass	21,120,438	SRX7091108
SRR2182496	Vietnam	Female carcass	76,013,673	SRX1163802
SRR2182437	Vietnam	Female carcass	20,148,844	SRX1163649
SRR2182471	Vietnam	Female carcass	13,039,329	SRX1163797
SRR2182500	Vietnam	Female carcass	31,230,578	SRX1163811
SRR11213090	Japan	Female carcass	84,804,848	SRX7825916
SRR11252298	California, USA	Male carcass	30,400,921	SRX7862424
SRR10390734	Vietnam	Male carcass	28,579,332	SRX7091110
SRR11252299	California, USA	Ovary	48,606,869	SRX7862425
SRR11213089	Japan	Ovary	10,540,551	SRX7825917
SRR10390735	Vietnam	Ovary	20,804,930	SRX7091109
SRR11252300	California, USA	Testis	26,522,056	SRX7862426
SRR10390733	California, USA	Testis	33,414,052	SRX7091111
SRR2182526	Laboratory	C6/36 cells	55,582,997	SRX1163812
SRR2182527	Laboratory	C6/36 cells	59,576,278	SRX1163813
SRR11252301	Laboratory	U4.4 cells	19,839,975	SRX7862427
SRR2182528	Laboratory	U4.4 cells	25,020,872	SRX1163814
SRR2182529	Laboratory	U4.4 cells	2,842,713	SRX1163816

Table 2   *C. americanus* small RNA data sets used in this study

Sample	SRA	NUKV status	Tissue	Total Reads	NCBI Accession
SNOA1C	SRR18143948	Infected	adult carcass	20,155,818	SRX14291809
SNOA1O	SRR18143947	Infected	eggs	21,279,218	SRX14291810
SNOA2C	SRR18143971	Infected	adult carcass	23,165,465	SRX14291786
SNOA2O	SRR18143970	Infected	ovaries	28,444,662	SRX14291787
SNOA3C	Not available	Uninfected	adult carcass	20,253,182	Not available
SNOA3O	Not available	Uninfected	ovaries	18,856,286	Not available
SNOA4C	Not available	Uninfected	adult carcass	10,817,285	Not available
SNOA4O	Not available	Infected	ovaries	34,336,864	Not available

Sequence read trimming, assembly, and mapping

Reads were trimmed for adapter sequences and low-quality bases with the BBDuk function of BBMap 38.35 using Illumina adaptor sequences (Bushnell 2014). Trimmed small RNA reads were then mapped to virus reference sequences using BBMap (Table 3). Small RNA size profiles were generated using the lhist function of BBMap on these mapped read sets. Excel (Microsoft Corporation 2018) was used to generate histogram plots. Exact parameters are shown in Table A6. All data was normalized using the reads per kilobase per million reads mapped (RPKM) normalization method (Li et al. 2015) shown below:

$$\text{RPKM} = \frac{\text{reads mapped to the transcript}}{(\text{total reads}) \cdot (\text{transcript length})} . \tag{1}$$

Table 3   RNA virus genomes used to map against small RNA profiles generated from both *Ae. albopictus* and *C. americanus* tissue samples.

Virus	Segment	Length (bp)	NCBI Accession
Barstukas virus	S segment	1578	MW434649.1
	M segment	1935	MW434634.1
	L segment	6462	MW434662.1
	EVE	1601	Not available
Niukluk virus	S segment	2470	ON059790.1
	M segment	2835	ON059787.1
	L segment	6669	ON059784.1
	EVE	2755	Not available
Tolvoit virus	Non-segmented	12750	ON059796.1
Rovykyts virus SNO1 Strain	Non-segmented	16119	ON059791.1
Rovykyts virus SNO2 Strain	Non-segmented	16105	ON059792.1
Lantra virus	L segment	6588	ON059778.1
	NPGM segment	5481	ON059780.1
Giez virus	Non-segmented	7954	ON059777.1
Ezimos virus	Chaq segment	1541	ON059771.1
	RdRp segment	1760	ON059772.1
	CP segment	1575	ON059773.1

Positional Nucleotide Frequencies

Putative piRNAs (25-30 nt) were extracted from virus-mapped small RNA reads and re-mapped to virus segments using Geneious Prime 2022.0.1 in order to obtain more precise readings. From these reads, positional nucleotide frequencies were calculated using Python (Van Rossum and Drake 1995). Exact parameters are located in Table A5.

Statistical analysis

In order to identify significant difference in the abundance of small RNA populations I used the Tukey HSD test at a confidence level of 95% in R (Abdi 2010 and R Core Team 2022). To test whether the nucleotide frequency biases at positions 1 and 10 were significantly different from those of other positions, I used the Dixon-Q test for outliers (Rorabacher 1991). Before I performed the Dixon-Q test, I first checked my data for normality using the qqnorm function (Bloomfield 2009). All reports of significance were determined by using the following p-value cutoffs: $p > .05$, NS; $p = .05-.01$, significant (*); $p = .01-.001$, highly significant (**); $p < .001$, very highly significant (***).

Phylogenetic analysis

The phylogenetic tree was built using the FastTree 2.1.1 plugin on Geneious Prime 2022.01. Accession numbers of virus sequences included in this phylogeny are present in Table 4.

Table 4   Sequences used for virus phylogenetics

Virus	NCBI Accession
Wuhan mosquito virus 1	YP_009305134.1
Wuhan mosquito virus 2	YP_009305133.1
Wuchang cockroach virus 1	YP_009304998.1
Seattle prectang virus	YP_009666960.1
Nome phantom virus	YP_009507890.1
Miglotas virus	QRW41748.1
Kigluaik phantom virus	YP_009362030.1
Hubei odonate virus	YP_009329888.1
Ganda bee virus	YP_009666983.1
Culex phasma-like virus	QRD99883.1
Coredo virus	QHA33848.1
Coleopteran phasma-related virus	QMP82274.1
Bactrocera latifrons virus	UPT53728.1
Anopheles triannulatus virus	YP_010086187.1

CHAPTER III

RESULTS

BARV small RNA populations are dominated by piRNAs in *Aedes*

To explore the RNAi response to BARV in *Aedes albopictus*, small RNA data sets generated from *Aedes albopictus* tissues were collected from NCBI and mapped against all three segments of Barstukas virus (BARV). All mapped small RNAs were quantified and characterized. Small RNA profiles showed an enrichment of 24-31 nt piRNAs across all samples (Figure 3). Female carcass tissues and ovary tissues had a higher amount of piRNAs produced in comparison to any other tissue types. It appears that the S segment may be targeted at a higher rate in the piRNA pathway than other segments. This trend appears most heavily in female carcasses and ovaries. However, statistical analyses show that there is no significant difference (Tukey test, $p > .05$) though, this could be due to limited sample size for each tissue.
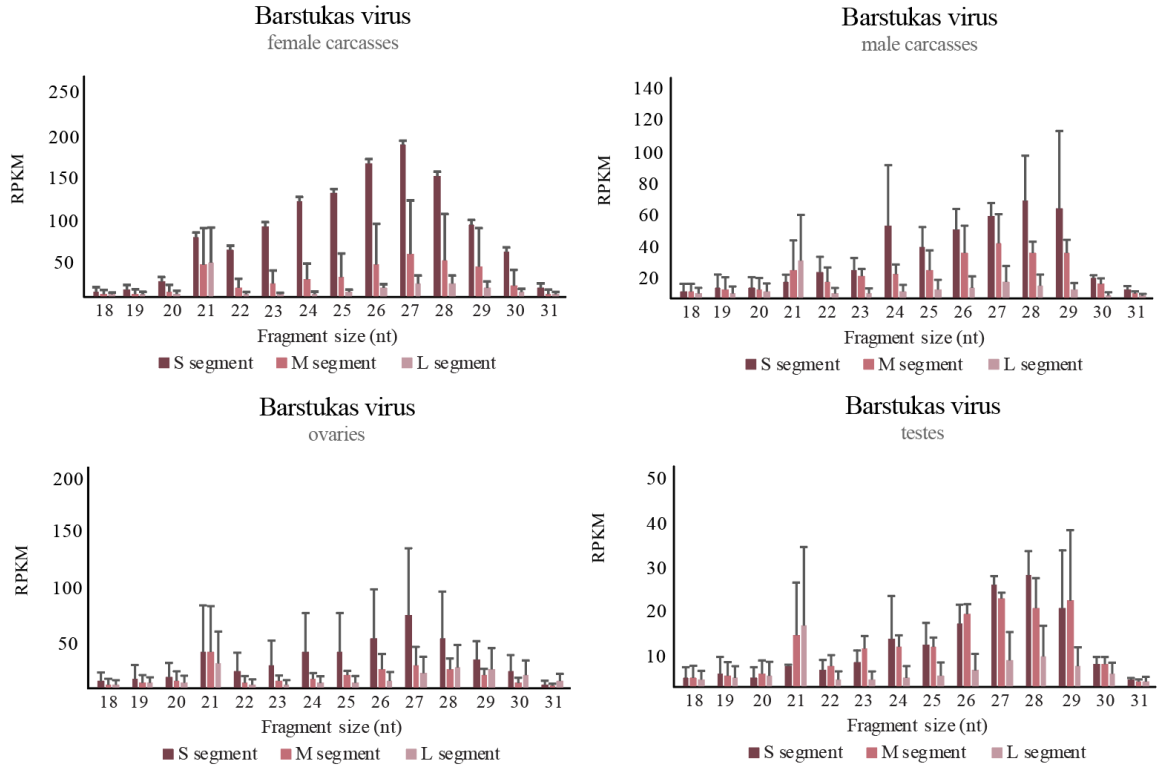
Figure 3　　　Size profiles of BARV-derived small RNAs in *Ae. albopictus* tissues

Ae. albopictus tissues: female carcasses (n=5), male carcasses (n=2), ovaries (n=2), and testes (n=2). All show an enrichment of 24-31 nt piRNAs, especially female carcasses and ovaries.

NUKV small RNA populations are dominated by siRNAs in *Chaoborus*

To explore the RNAi response to NUKV in *C. americanus,* I mapped small RNAs generated previously by the Ballinger lab against all three segments of Niukluk virus (NUKV). I had three adult female carcasses with two sets of extracted ovaries and one set of eggs. One female carcass sample yielded limited small RNAs and was excluded from the analysis. All mapped small RNAs were quantified and characterized. Small RNA profiles showed an enrichment of 21 nt siRNAs across all samples when mapped to the viral segments (Figure 4). Putative piRNAs were detected, but in small quantities. However, statistical analyses show that the S derived piRNAs are more abundant in comparison to the M and L derived piRNAs,

19

respectively, in female carcasses (Tukey test, S-M p=.002, S-L p=.003). S derived piRNAs are

also more abundant than L piRNAs in ovaries (Tukey test, p= .030). No corresponding

enrichment of S-derived piRNAs was observed in the eggs.
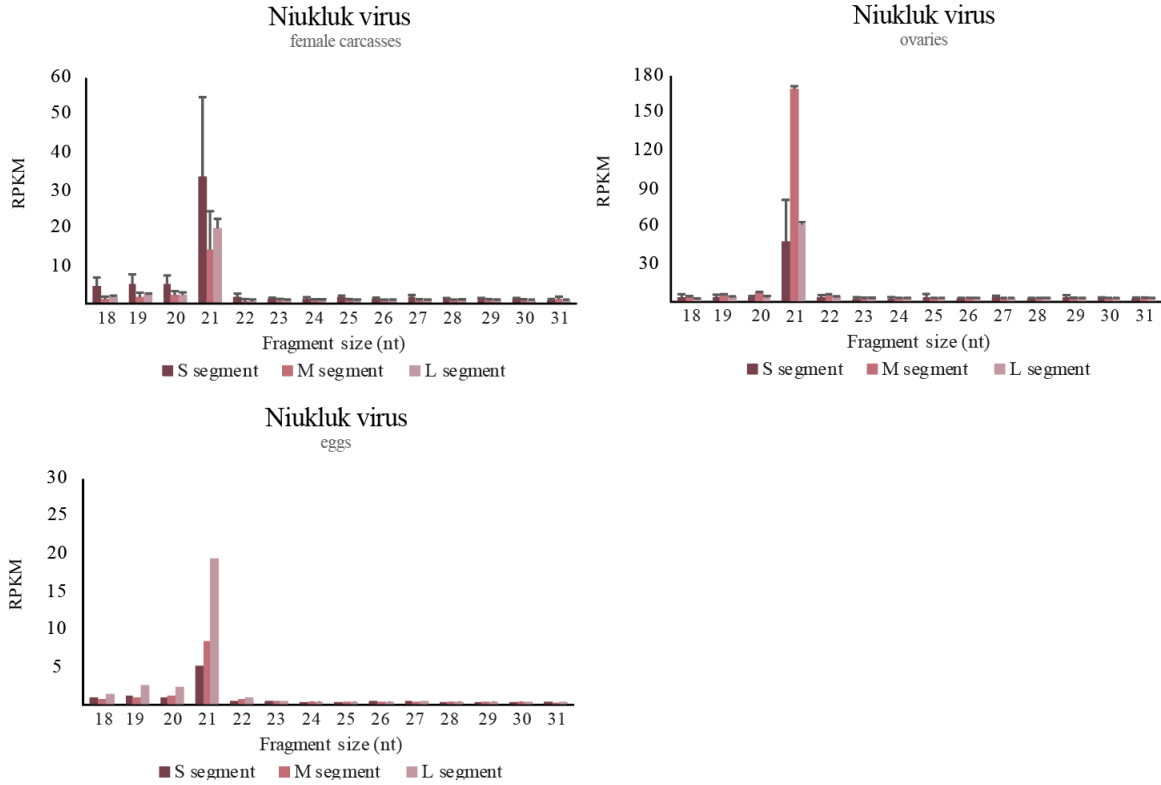
Figure 4   Size profiles of NUKV-derived small RNAs in *C. americanus* tissues

C. americanus tissues: adult carcasses (n=2), ovaries (n=2), and eggs (n=1). All show an
enrichment of 21 nt siRNAs.

BARV EVE and NUKV EVE small RNA populations are dominated by piRNAs in both *Aedes* and *Chaoborus*

To explore the RNAi response to BARV EVE in *Ae. albopictus* and NUKV EVE in *C.*

americanus, I mapped small RNA data sets derived from *Aedes* and *Chaoborus* tissues to their

respective S segment EVEs. All EVE-derived small RNA profiles showed an enrichment of 24-

31nt piRNAs across all tissue samples (Figure 5). Adult carcasses harbor the most EVE derived

piRNAs in *Aedes*, while in *Chaoborus* they are most abundant in ovaries. Interestingly, NUKV

EVE-derived piRNAs are rare in the eggs, mirroring observations from S-segment derived
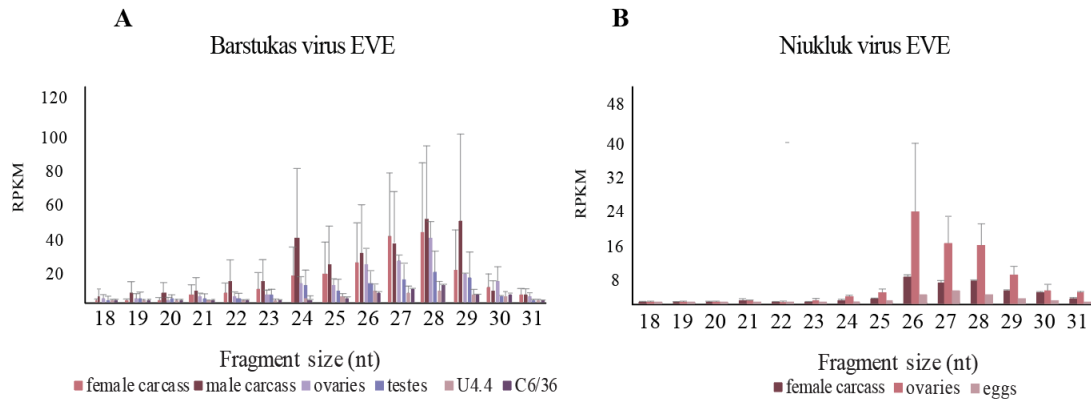
piRNAs.

Figure 5   *Ae. albopictus* tissue samples mapped against BARV EVE and *C. americanus*
tissues mapped against NUKV EVE

Ae. albopictus tissue samples include female carcasses (n=7), male carcasses (n=2), ovaries
(n=3), testes (2), C6/36 cell lines (n=2), and U4.4 cell lines (n=3). All small RNAs derived from
BARV EVE show an enrichment of 24-31nt piRNAs (A). *C. americanus* tissue samples were
mapped against NUKV EVE. Tissue samples include adult carcasses (n=3), ovaries (n=4), and
eggs (n=1). All small RNAs derived from NUKV EVE show an enrichment of 24-31nt piRNAs
(B).

Evidence for secondary piRNA biogenesis in both *Aedes* and *Chaoborus*

In order to investigate the presence of secondary piRNA biogenesis in *Ae. albopictus* and

C. americanus, I extracted and analyzed 25-30 nt piRNAs from all small RNAs derived from

BARV and NUKV in *Aedes* and *Chaoborus* respectively. Enrichment of uridine at nucleotide 1

(U1) in genome-derived (antisense) piRNAs and adenine at nucleotide 10 (A10) in

antigenome/transcript-derived (sense) piRNAs was interpreted as the signature of secondary

piRNA biogenesis. In *Aedes*, there is U1 and A10 enrichment present in genome- and

antigenome-derived piRNAs respectively for the S, M, and L segments of BARV in the female carcasses, the male carcasses, the ovaries, and the testes (Dixons Q test, $p < .05$) (Figure 6).

In *Chaoborus*, there is U1 enrichment in NUKV S genome-derived piRNAs in female carcasses and ovaries (Dixons Q test, $p = .003$, $p = .05$) and a nonsignificant A10 enrichment in NUKV S antigenome-derived piRNAs in ovaries (Dixons Q test, $p > .05$). In eggs, there is no enrichment of U1 in genome-derived piRNAs. There is a trend showing an enrichment of A10 in NUKV S antigenome-derived piRNAs, but this is not supported by statistical analysis (Dixons Q test, $p = .06$), though this could be due to limited sample size (Figure 7).

In *Aedes*, only genome-derived piRNAs were generated from infection with BARV EVE (Figure 8). There is a U1 bias present in BARV EVE genome-derived piRNAs in female carcasses, ovaries, testes, C6/36 cell lines, and U4.4 cell lines (Dixons Q test, $p < .05$). There is a trend showing U1 bias in genome-derived piRNAs in the male carcasses, but statistical analyses show no significant difference (Dixons Q test, $p = .12$).

In *Chaoborus*, mainly genome-derived piRNAs were generated from infection with NUKV EVE (Figure 9). Few antigenome-derived piRNAs were generated and they were excluded from this study as they contained less than 100 piRNAs. There is a U1 bias present in NUKV EVE genome-derived piRNAs in female carcasses, ovaries, and eggs (Dixons Q test, $p = $ 2.2E-16).
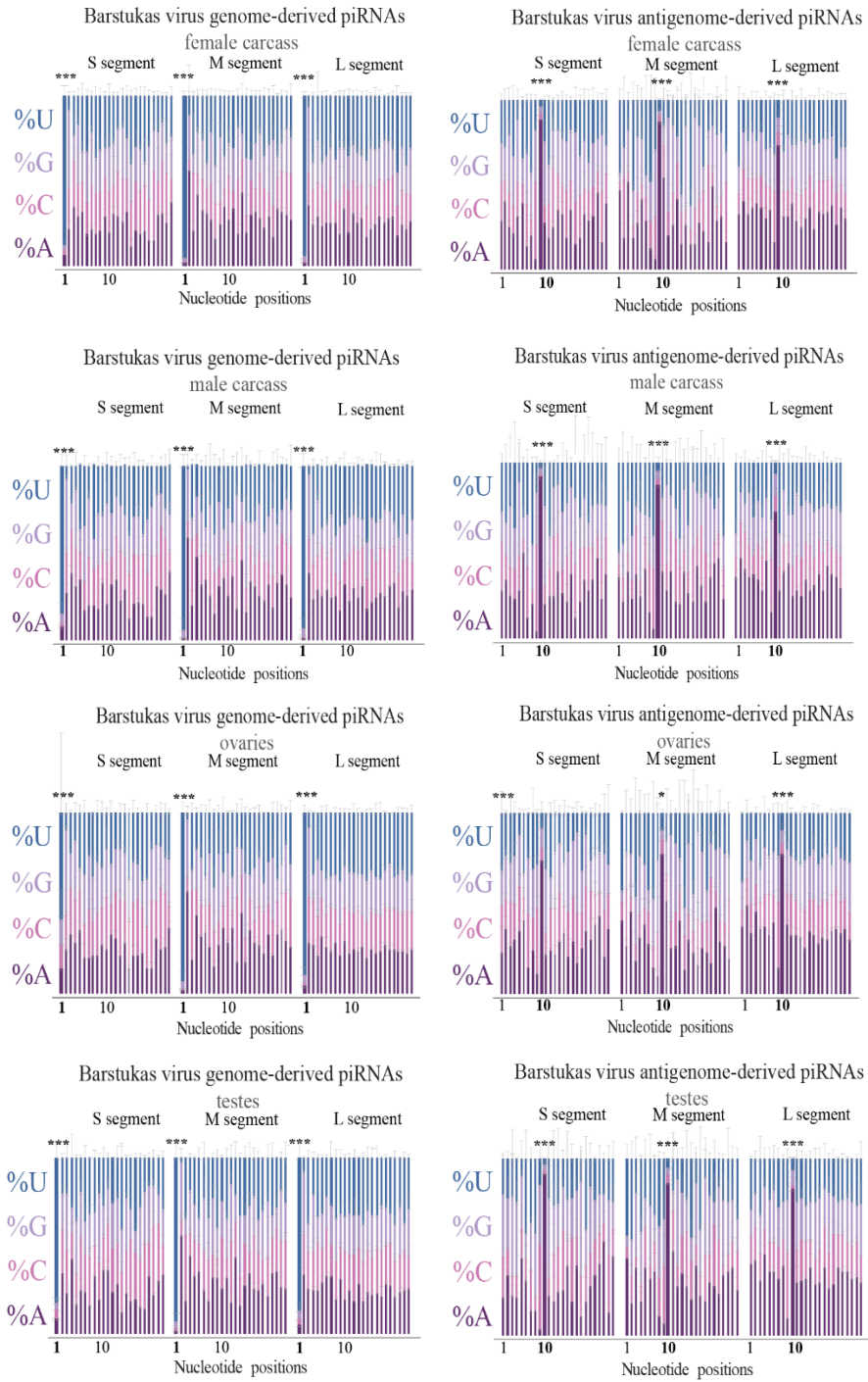
Figure 6   Positional nucleotide frequencies of BARV-derived small RNA profiles in *Ae. albopictus*

There is evidence of secondary piRNA biogenesis in female carcasses, male carcasses, ovaries, and testes. All statistical analyses were done using the Dixons Q test (*$p<=.05$, **$p<=.01$, ***$p<=.001$).

Figure 7   Positional nucleotide frequencies of NUKV-derived small RNA profiles in *C. americanus*

There is a U1 bias present in NUKV S genome-derived piRNAs in female carcasses and ovaries. There is no U1 or A10 enrichment in eggs. All statistical analyses were done using the Dixons Q test (*$p<=.05$, **$p<=.01$, ***$p<=.001$).

Figure 8 Positional nucleotide frequencies of BARV EVE-derived small RNA profiles in *Ae. albopictus*

There is a U1 bias present in BARV EVE genome-derived piRNAs in female carcasses (n=7), ovaries (n=3), testes (n=2), C6/36 cell lines (n=2), and U4.4 cell lines (n=3). All statistical analyses were done using the Dixons Q test (*$p<=.05$, **$p<=.01$, ***$p<=.001$).
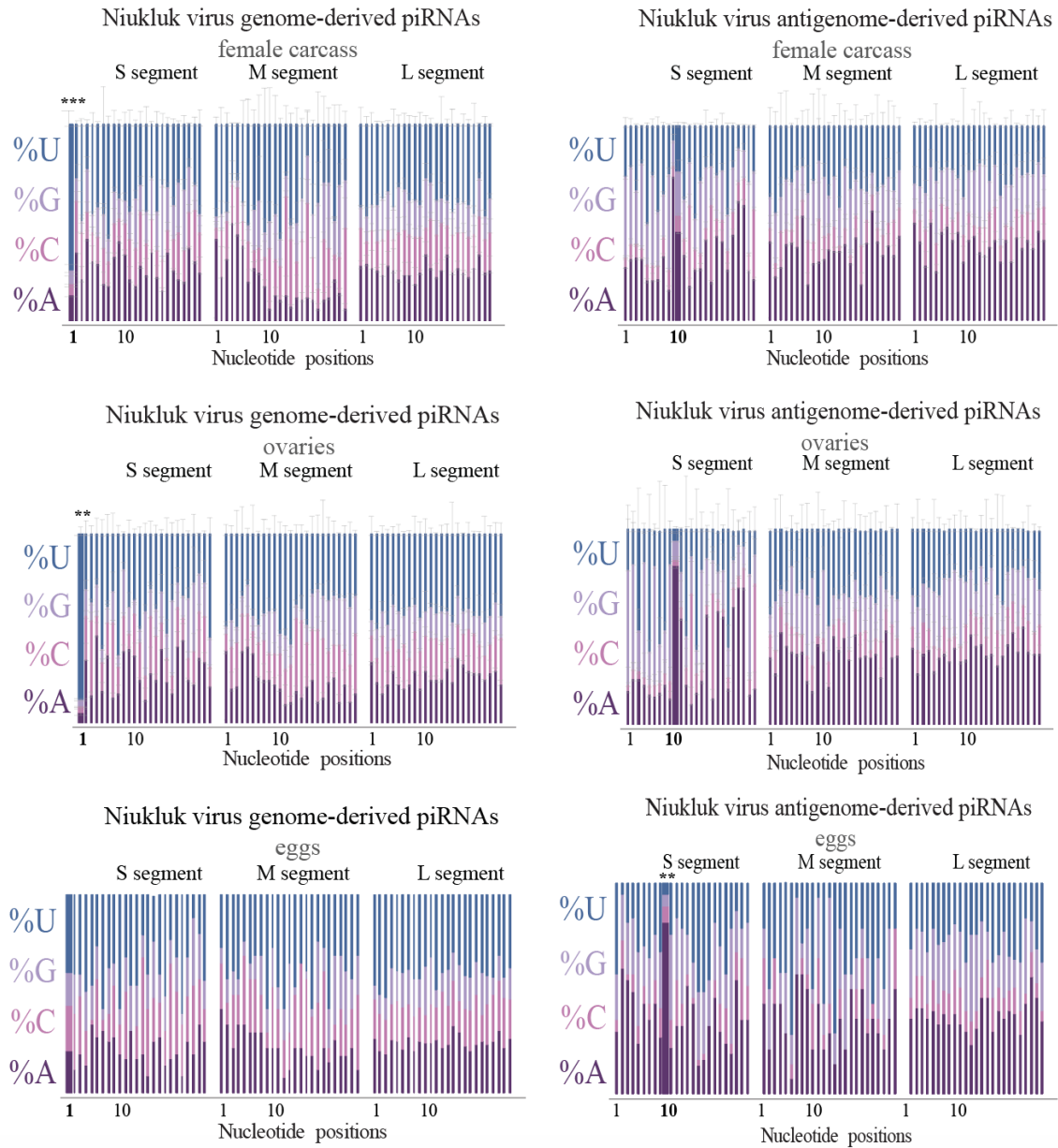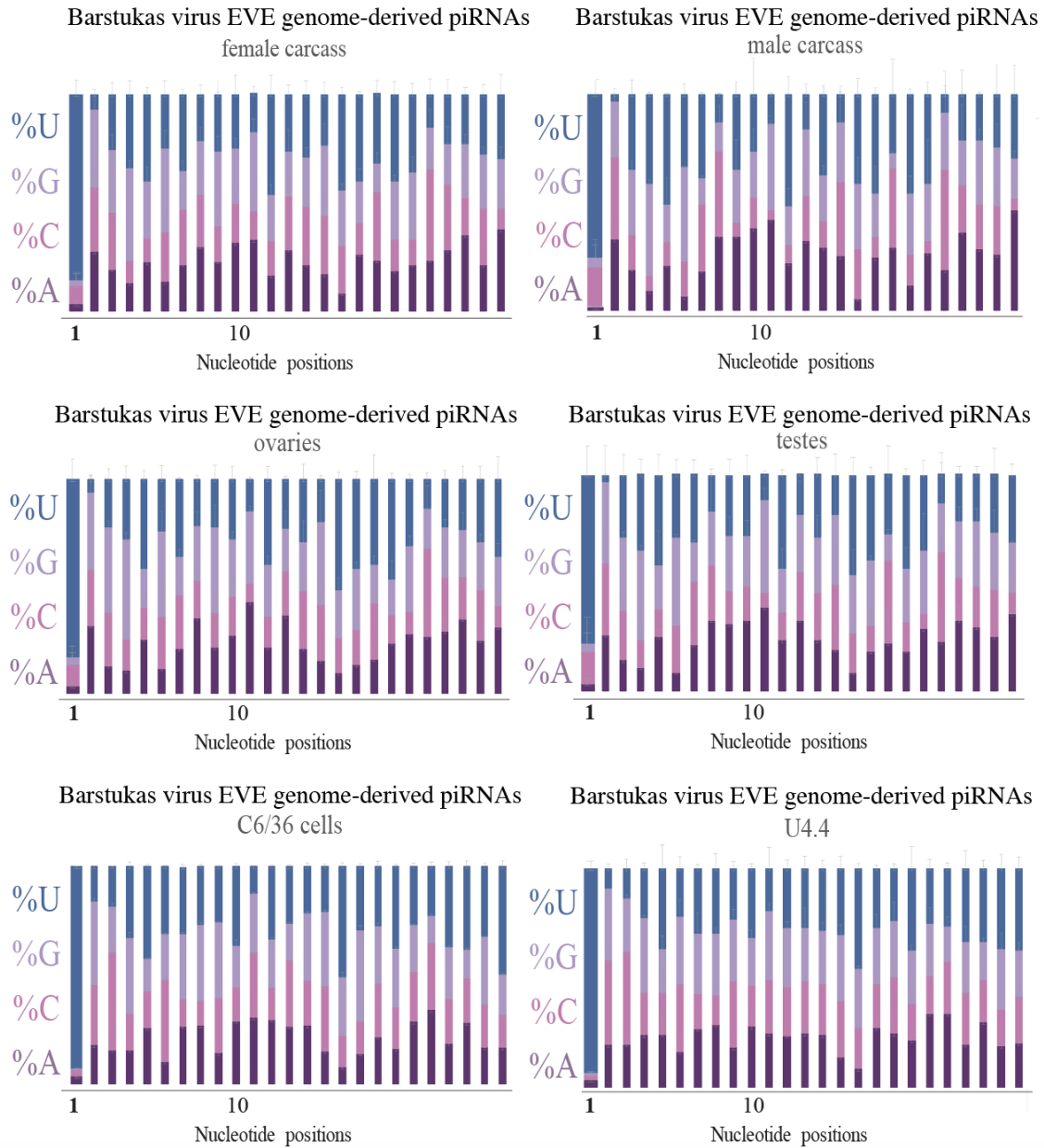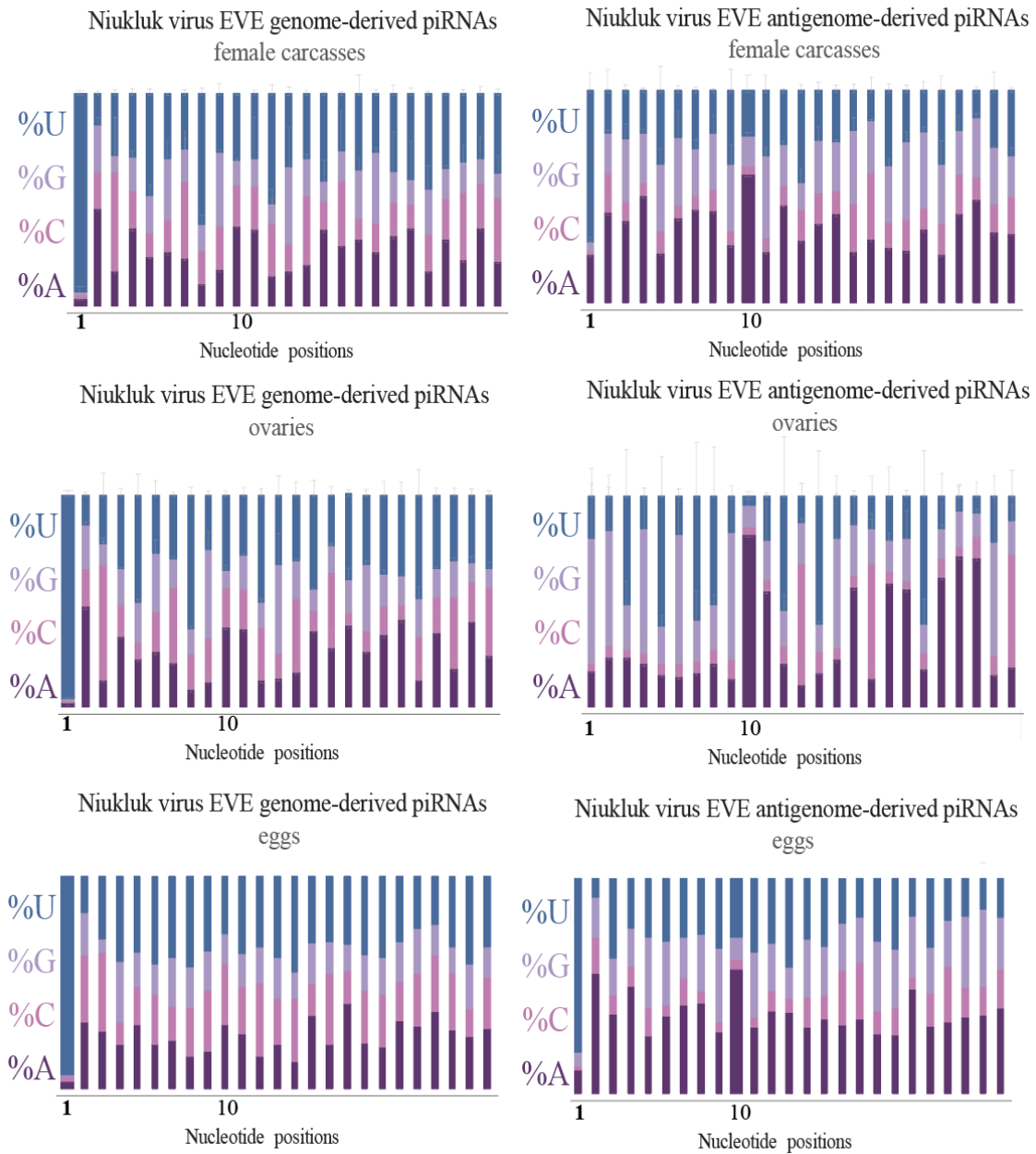
25

Figure 9   Positional nucleotide frequencies of NUKV EVE-derived small RNA profiles in *C. americanus*

There is a U1 bias present in NUKV EVE genome-derived piRNAs female carcasses, ovaries, and eggs. All statistical analyses were done using the Dixons Q test (*$p<=.05$, **$p<=.01$, ***$p<=.001$).

Small RNA populations derived from viruses lacking EVEs are dominated by siRNAs in
Chaoborus

To put context on the observed U1 and A10 enrichment in NUKV S-derived piRNAs, small RNA data sets of *C. americanus* tissues previously generated in the Ballinger lab were mapped against six exogenous viruses. All of these viruses are found naturally occurring in North American *Chaoborus* populations. Unlike NUKV, none of these six viruses have a known corresponding EVE within the host genome. All mapped small RNA profiles had an enrichment of 21 nt siRNAs across all tissue samples (Figure 10), while piRNA abundance was limited. However, the few 25-30 nt fragments (candidate piRNAs) that were mapped were extracted and their positional nucleotide frequencies analyzed (Figure 11 and 12). All small RNA profiles which contained less than 100 piRNAs were excluded. U1 or A10 enrichment was not observed in any of these small RNA populations, suggesting they are not bona fide piRNAs processed by Piwis.
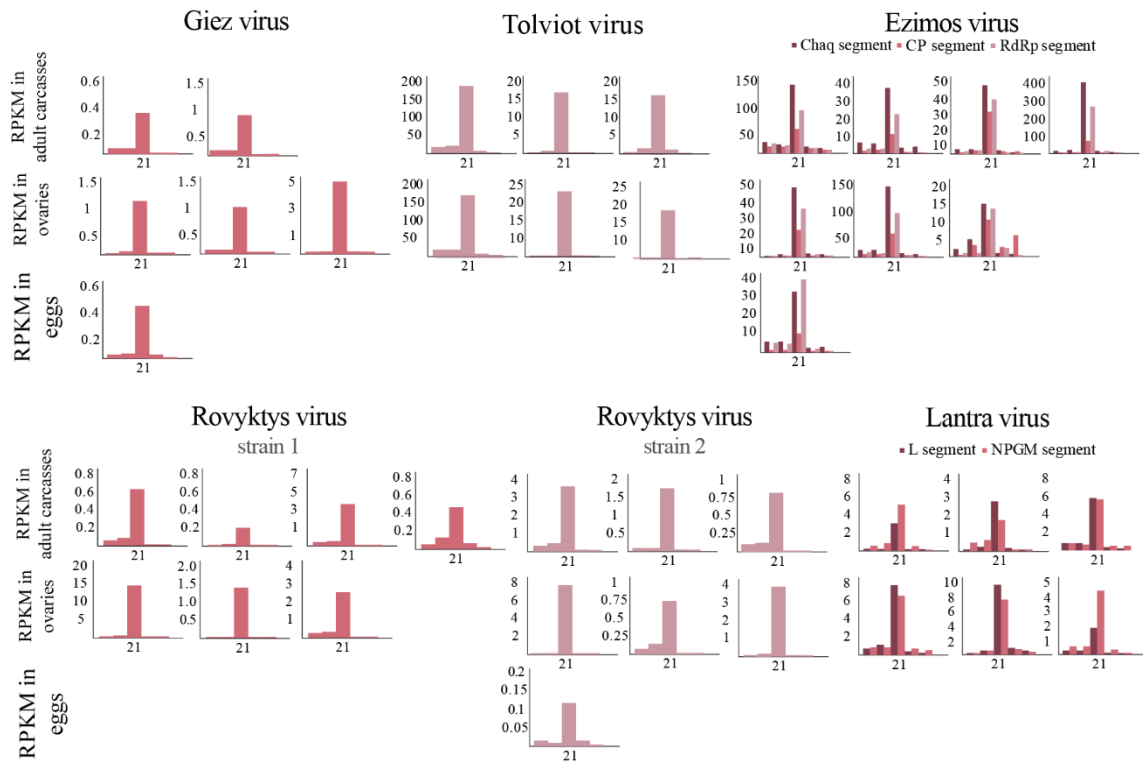
Figure 10   Size profiles of small RNAs derived from exogenous RNA viruses found in
Chaoborus tissues

None of these viruses have an EVE. All size profiles show an enrichment of 21 nt siRNAs across all tissue samples.

Figure 11   Positional nucleotide frequencies of small RNAs derived from exogenous RNA
viruses found in *Chaoborus* ovaries

None of these viruses have an EVE. There is no U1 bias in genome-derived strands and there is
no A10 bias in antigenome-derived strands.
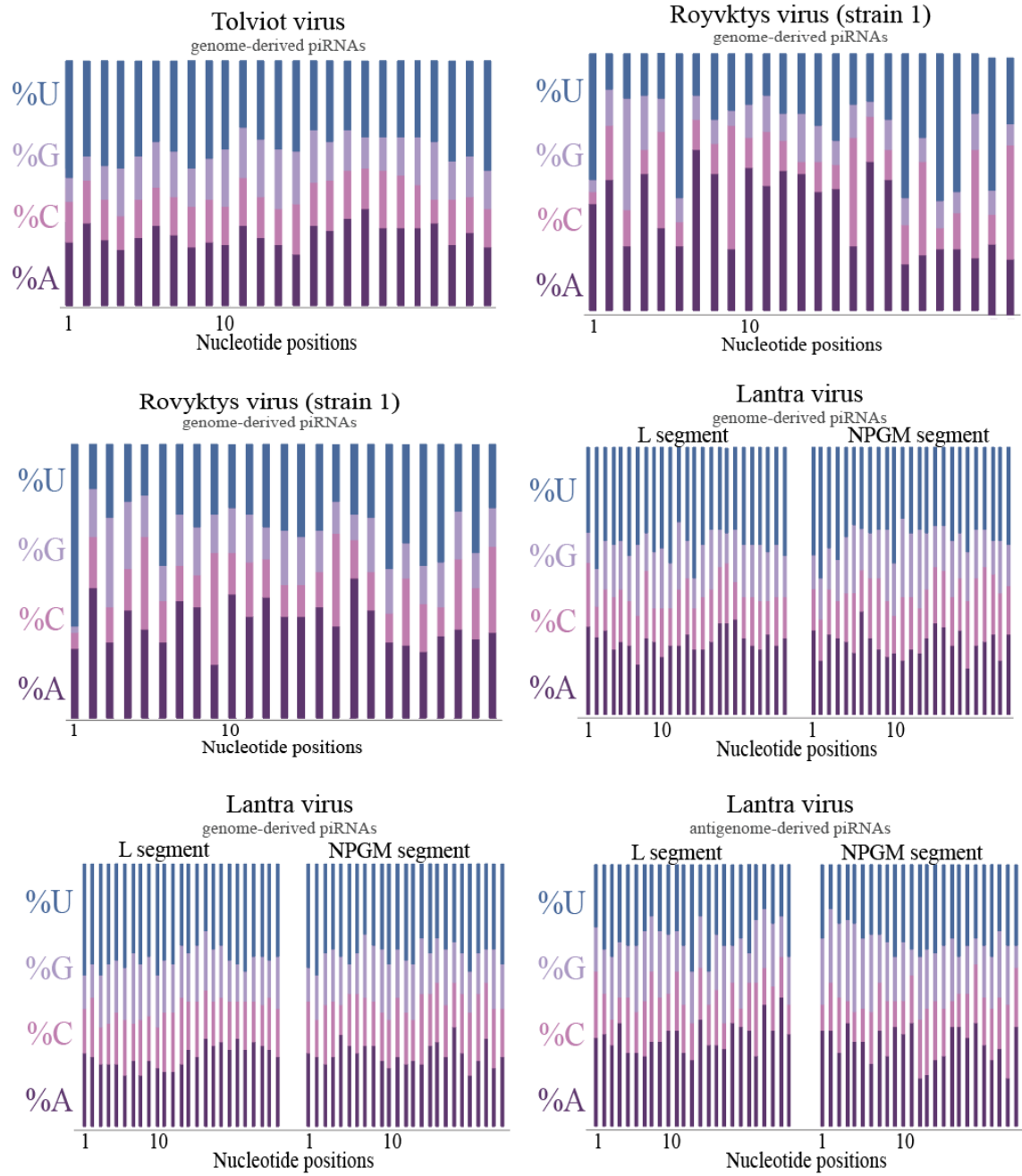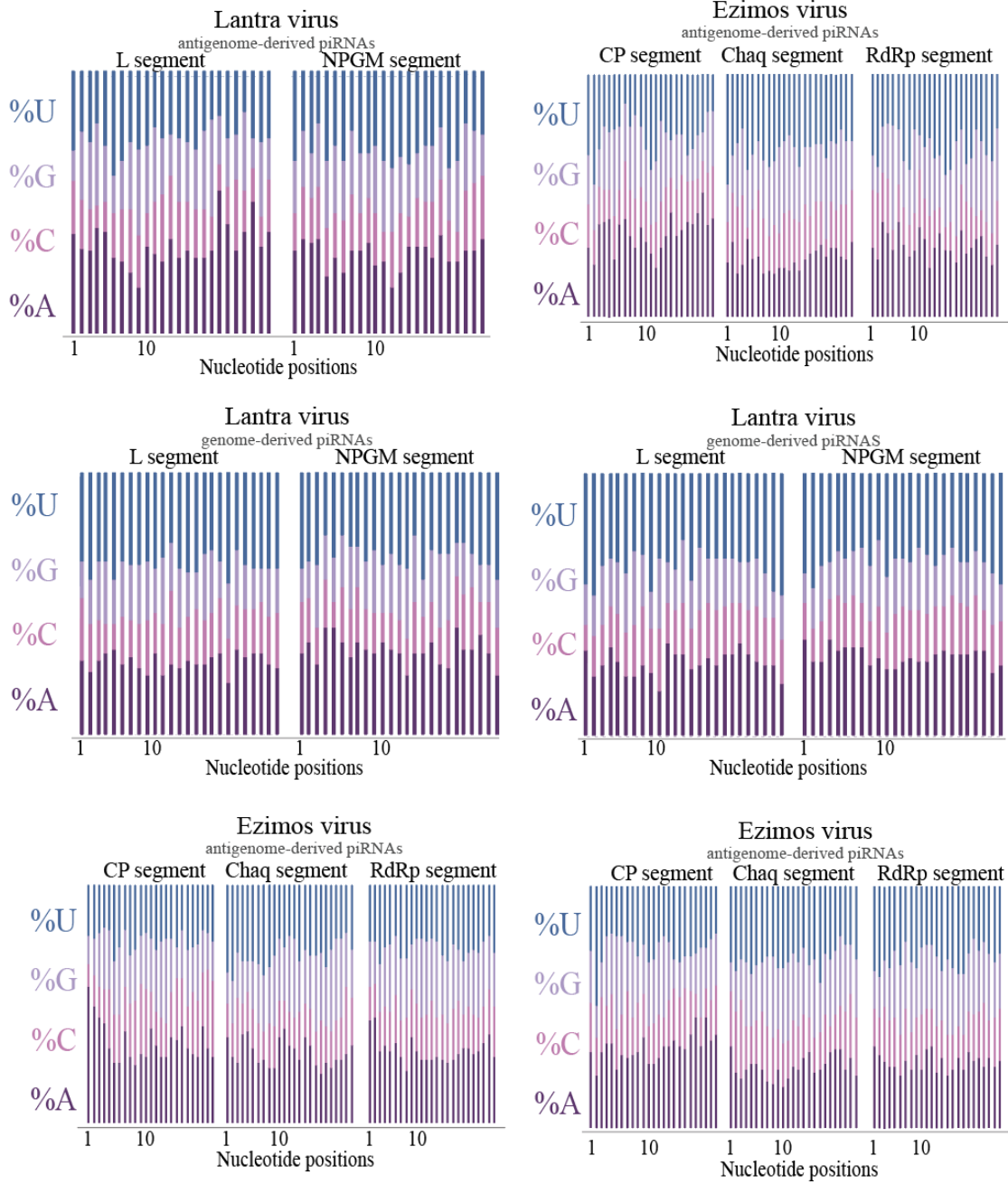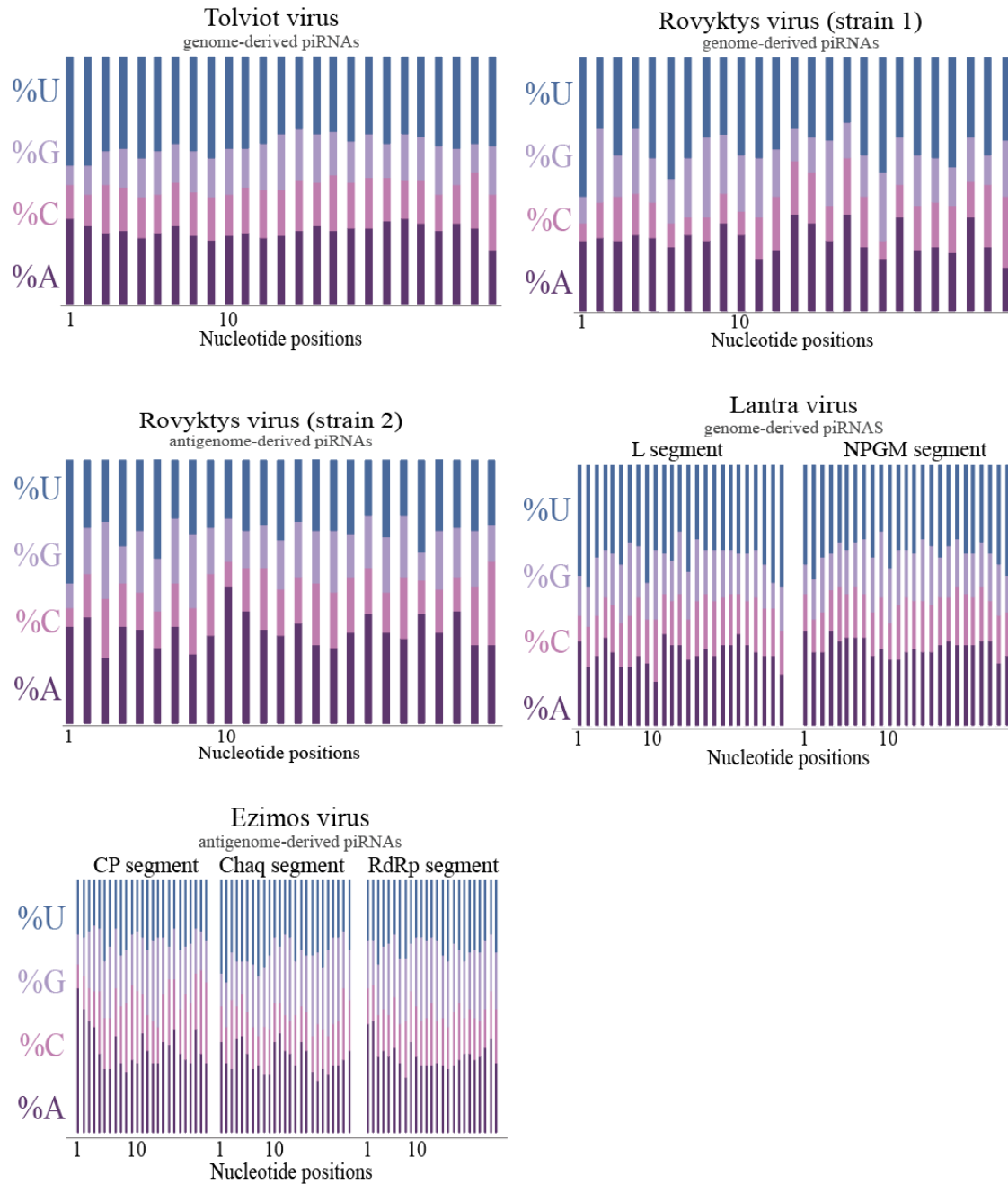
Figure 11 (continued)

30

Figure 12　Positional nucleotide frequencies of small RNAs derived from exogenous RNA viruses found in *Chaoborus* female carcasses

None of these viruses have an EVE. There is no U1 bias in genome-derived strands and there is no A10 bias in antigenome-derived strands.

CHAPTER IV

DISCUSSION

Endogenous viral elements (EVEs) have recently been investigated for their role in antiviral immunity in insects. Nonretroviral EVEs are frequently identified in arthropod genomes, and they have been found to partner with the principal antiviral defense response in insects, the RNA interference pathway (Tassetto et al. 2019). Specifically, EVEs have been implicated in the piRNA pathway of the RNAi response in *Aedes* mosquitoes (Tassetto et al. 2019 and Suzuki et al. 2020). However, little is known about the role of EVEs in antiviral immunity in other dipterans. In order to further interrogate the EVE-Piwi partnership and its potential role in antiviral defense, I used small RNA profiling to perform a comparative analysis of two naturally occurring EVE-virus pairs – one in a mosquito *Aedes albopictus*, and one in *Chaoborus americanus,* which belongs a sister family of mosquitoes, the phantom midges. Both hosts harbor a bunyavirus in the family Phasmaviridae and encode an EVE derived from the viral nucleoprotein gene. However, *Aedes* mosquitoes also encode an expansion of the Piwi family endoribonucleases while *Chaoborus* midges only encode for the three ancestral Piwi proteins conserved across all other dipteran insects (Morazzani et al. 2012 and Schnettler et al. 2013). Using these two systems allows me to ask whether expanded Piwis are required to mediate interactions between EVEs and exogenous viruses. My results suggest that EVEs can in fact facilitate recruitment of ancestral dipteran Piwi pathway machinery to exogenous viral targets.

In order to investigate the role of EVEs in this context, I first had to explore the RNAi response in both hosts when infected with this virus. Small RNA data sets generated from *Ae. albopictus* tissues were mapped against all three segments of the Barstukas virus (BARV) and its related EVE. These BARV small RNA populations showed an enrichment of 24-31 nt piRNAs across all tissue samples including female carcasses, male carcasses, ovaries, and testes. This aligns with previous research which also demonstrate that the piRNA pathway is dominate in *Aedes* mosquitoes (Wang et al. 2018).

Small RNA data sets generated from *C. americanus* tissue samples were also mapped against all three segments of their bunyavirus, Niukluk virus (NUKV) and its related EVE. In contrast to BARV in *Aedes*, these NUKV small RNA populations showed an enrichment of 21 nt siRNAs across all tissue samples, which included female carcasses, ovaries, and eggs. This demonstrates that the siRNA pathway is dominant in *Chaoborus*, consist with the antiviral RNAi response across many dipterans, including *Drosophila* (Marques et al. 2010). Here, statistical analyses do support that S segment derived piRNAs, while low in absolute abundance, are more abundant than the M and L segment derived piRNAs in female carcasses and more than the L segment derived piRNAs in ovaries. This is intriguing since both of the EVEs in my study systems originate from recent ancestors of their nucleoproteins, encoded by the S segment. This could suggest that the presence of EVEs influence the abundance of piRNAs in these hosts.

To further interrogate this idea, I explored whether and how the BARV EVE in *Aedes* tissues and the NUKV EVE in *Chaoborus* tissues are processed by RNAi pathway machinery. Interestingly, all tissue types showed an enrichment for 24-31 nt EVE-derived small RNAs showing that the piRNA pathway is dominate here in both hosts. However, adult carcasses harbor the most EVE derived piRNAs in *Aedes*, while in *Chaoborus* they are most abundant in

33

ovaries. This aligns with previous research. In *Aedes* mosquitoes, the expanded Piwi family genes are expressed in somatic cells and as a result secondary piRNAs have been detected in abundance throughout the body (Miesen et al. 2015), while *Drosophila* express their Piwi proteins mainly in germline cells to protect the nuclear genome from transposable elements (Schnettler et al. 2013).

I then investigated signatures of secondary piRNA biogenesis in both hosts. Secondary piRNA biogenesis is characterized by having an enrichment of adenine at the tenth nucleotide position (A10) of sense-derived piRNAs (Wang et al. 2014 and Tassetto et al. 2019). This is also referred to as the ping-pong signature as these transcripts are in a continuous cycle to produce an amplification of piRNAs used to seek and cleave target RNA molecules. My results show that there is a U1 and A10 enrichment present in both genome-derived (antisense) and antigenome-derived (sense) piRNAs, respectively, for all segments of BARV in all *Aedes* tissues. This establishes the presence of primary piRNAs as well as a ping-pong signature; however, this is also exhibited in other *Aedes* mosquitoes infected with an RNA virus without a related EVE as well (Schnettler et al. 2013). In *Chaoborus*, I identified unambiguous U1 enrichment, but only in piRNAs derived from the NUKV S segment. I also identified an apparent enrichment of adenine at position 10, but this is not yet supported by statistical analyses, possibly due to limited samples, which could be improved in future work. Eggs do not show an enrichment of U1 or A10 in either genome-derived or antigenome-derived piRNAs. The presence of virus-derived secondary piRNAs in a host without an expanded Piwi protein set suggests the ancestral dipteran piRNA machinery is capable of targeting the ping-pong cycle to viral RNAs, even if it is typically not implicated as such. One possibility is that an additional cofactor is required to amplify piRNA biogenesis in this context. Because the L or M segments are neither targeted in

34

this way nor the source of EVEs in the phantom midge, the NUKV S segment EVE is a strong candidate for such a cofactor.

I also investigated the presence of secondary piRNA biogenesis in BARV EVE derived piRNAs and NUKV EVE derived piRNAs. In *Aedes*, only antisense piRNAs were generated and all show a U1 enrichment in female carcasses, ovaries, testes, C6/36 cell lines, and U4.4 cell lines. Male carcasses show a trend for U1 enrichment, but it was not supported by statistical analyses. In *Chaoborus*, only antisense piRNAs were generated from the NUKV EVE as well. There is a U1 enrichment present in genome-derived piRNAs in female carcasses, ovaries, and eggs. The presence of EVE derived piRNAs illustrates that EVEs are being processed by the piRNA pathway. However, this data alone does not support a clear functional link between EVEs and viruses.

To put context on the observed U1 and A10 enrichment in NUKV S-derived piRNAs, I mapped the *Chaoborus* small RNA data sets against six exogenous RNA viruses found naturally occurring in this host (Ballinger et al. 2022). None of these viruses have an EVE. Results show that these viruses are processed into 21 nt siRNAs but few fragments of 24-31 nt. Of the piRNA-sized fragments produced, none showed an enrichment of U1 or A10 in either sense or antisense transcripts. So not only do they not get cycled into the ping-pong amplification cycle, but most of the time they do not seem to get targeted by Piwi proteins at all. Together, this and previous results suggest that EVEs modestly recruit Piwi proteins to viral RNAs. Only viruses expressing an EVE are targeted by Piwi proteins.

Investigating the EVE-Piwi partnership is fundamental to antiviral research in insect hosts. Previous research found that EVE inactivation results in increased viral titer in *Aedes* ovaries (Suzuki et al. 2020), suggesting these elements are associated not only with increased

35

piRNA pathway targeting, but potentially with meaningful reductions to fitness costs on the host via reduced viral load. While it appears this "heritable immune memory" does not rapidly eliminate infection, it may suppress viral transmission dose from mother to offspring, resulting in reduced fitness cost of infection for offspring of EVE-expressing mothers. This could lead to unexpected consequences for long-term virus persistence. For example, if EVE-free hosts suffer elevated virus-induced mortality compared to counterparts, selection favoring EVEs could promote rising frequency of low titer virus infection as well.

My research suggests that EVEs may recruit Piwi proteins to viral RNA in somatic cells, a phenomenon that was previously not thought to occur in non-culicid dipterans. In the case of insect-specific viruses, somatic recruitment of Piwi proteins may play a role in suppressing viral replication and reducing the fitness costs of infection. Because viral load is an important feature of arbovirus vector competence (Elrefaey et a. 2020), it is interesting to speculate that somatic antiviral recruitment of Piwis could also reduce transmission success in non-mosquito arbovirus vectors.

This study was limited in two important ways. First, I analyzed a small sample of infected *Chaoborus* tissues. Although many of my analyses resulted in strong effect sizes, this is likely to have limited the scope of variation captured and reduced statistical power in some cases. Second, while I was able to identify a clear link between piRNA pathway recruitment and EVE expression in *Chaoborus*, my data do not establish a causal connection. In order to further examine the role of EVEs in antiviral immunity, more experimentally tractable EVE-virus systems are needed. Creating an EVE-virus system in laboratory-maintained *Drosophila* would allow for antiviral predictions to be formally tested in an organism with a conserved Piwi protein family.

My thesis highlights the potential for future research focused on the EVE-Piwi partnership in insect antiviral immunity. Mosquitoes and other arthropods are important vectors of many critical human and livestock RNA viruses. Given the broad distribution of nonretroviral EVEs in arthropod genomes (ter Horst et al. 2019), my findings suggest that Piwi-virus interactions could have greater implications for insect health and disease transmission than is currently recognized.

REFERENCES

Abdi, H., & Williams, L. J. (2010). Tukey's honestly significant difference (HSD) test. *Encyclopedia of research design*, *3*(1), 1-5. https://doi.org/10.4135/9781412961288.n478

Abudurexiti, A., Adkins, S., Alioto, D., Alkhovsky, S. V., Avšič-Županc, T., Ballinger, M. J., & Kuhn, J. H. (2019). Taxonomy of the order Bunyavirales: update 2019. *Archives of virology*, *164*(7), 1949-1965. https://doi.org/10.1007/s00705-019-04253-6

Asgari, S. (2011). Role of microRNAs in insect host–microorganism interactions. *Frontiers in physiology*, *2*, 48 https://doi.org/10.1093/ve/veac018

Ballinger, M. J., Christian, R. C., Moore, L. D., Taylor, D. J., & Sabet, A. (2022). Evolution and diversity of inherited viruses in the Nearctic phantom midge, *Chaoborus americanus*. *Virus Evolution*, *8*(1), veac018. https://doi.org/10.1093/ve/veac018

Ballinger, M. J., & Taylor, D. J. (2019). Evolutionary persistence of insect bunyavirus infection despite host acquisition and expression of the viral nucleoprotein gene. *Virus Evolution*, *5*(2), vez017. https://doi.org/10.1093/ve/vez017

Batson, J., Dudas, G., Haas-Stapleton, E., Kistler, A. L., Li, L. M., Logan, P., Ratnasiri, K., & Retallack, H. (2021). Single mosquito metatranscriptomics identifies vectors, emerging pathogens and reservoirs in one assay. *ELife*, *10*, e68353. https://doi.org/10.7554/eLife.68353

Bézier, A., Annaheim, M., Herbinière, J., Wetterwald, C., Gyapay, G., Bernard-Samain, S., Wincker, P., Roditi, I., Heller, M., Belghazi, M., Pfister-Wilhem, R., Periquet, G., Dupuy, C., Huguet, E., Volkoff, A.-N., Lanzrein, B., & Drezen, J.-M. (2009). Polydnaviruses of Braconid Wasps Derive from an Ancestral Nudivirus. *Science*, *323*(5916), 926–930. https://doi.org/10.1126/science.1166788

Bloomfield, V. (2009). *Computer simulation and data analysis in molecular biology and biophysics: an introduction using R*. Springer Science & Business Media. https://doi.org/10.1007/978-3-642-01973-9_4

Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., & Hannon, G. J. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell*, *128*(6), 1089-1103. https://doi.org/10.1016/j.cell.2007.01.043

Brubaker, S. W., Bonham, K. S., Zanoni, I., & Kagan, J. C. (2015). Innate Immune Pattern Recognition: A Cell Biological Perspective. *Annual Review of Immunology*, *33*(1), 257–290. https://doi.org/10.1146/annurev-immunol-032414-112240

Bushnell, B. (2014). *BBMap: a fast, accurate, splice-aware aligner* (No. LBNL-7065E). Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States). URL https://www.osti.gov/biblio/1241166

Campbell, C. L., Black, W. C., Hess, A. M., & Foy, B. D. (2008). Comparative genomics of small RNA regulatory pathway components in vector mosquitoes. *BMC genomics*, *9*(1), 1-16. https://doi.org/10.1186/1471-2164-9-425.

Cooper, D., & Eleftherianos, I. (2017). Memory and Specificity in the Insect Immune System: Current Perspectives and Future Challenges. *Frontiers in Immunology*, *8*, 539. https://doi.org/10.3389/fimmu.2017.00539

Diamond, M. S., & Kanneganti, T.-D. (2022). Innate immunity: The first line of defense against SARS-CoV-2. *Nature Immunology*, *23*(2), 165–176. https://doi.org/10.1038/s41590-021-01091-0

Elrefaey, A. M., Abdelnabi, R., Rosales Rosas, A. L., Wang, L., Basu, S., & Delang, L. (2020). Understanding the Mechanisms Underlying Host Restriction of Insect-Specific Viruses. *Viruses*, 12(9), 964. https://doi.org/10.3390/v12090964

Emerman, M., & Malik, H. S. (2010). Paleovirology—Modern Consequences of Ancient Viruses. *PLoS Biology*, *8*(2), e1000301. https://doi.org/10.1371/journal.pbio.1000301

Etienne, L., & Emerman, M. (2013). The Mongoose, the Pheasant, the Pox, and the Retrovirus. *PLoS Biology*, *11*(8), e1001641. https://doi.org/10.1371/journal.pbio.1001641

Gammon, D. B., & Mello, C. C. (2015). RNA interference-mediated antiviral defense in insects. *Current Opinion in Insect Science*, *8*, 111–120. https://doi.org/10.1016/j.cois.2015.01.006

Gilbert, C., & Belliardo, C. (2022). The diversity of endogenous viral elements in insects. *Current Opinion in Insect Science*, *49*, 48–55. https://doi.org/10.1016/j.cois.2021.11.007

Goriaux, C., Desset, S., Renaud, Y., Vaury, C., & Brasset, E. (2014). Transcriptional properties and splicing of the *flamenco* pi RNA cluster. *EMBO Reports*, *15*(4), 411–418. https://doi.org/10.1002/embr.201337898

Holmes, E. C. (2011). The Evolution of Endogenous Viral Elements. *Cell Host & Microbe*, *10*(4), 368–377. https://doi.org/10.1016/j.chom.2011.09.002

Horie, M., Honda, T., Suzuki, Y., Kobayashi, Y., Daito, T., Oshida, T., Ikuta, K., Jern, P., Gojobori, T., Coffin, J. M., & Tomonaga, K. (2010). Endogenous non-retroviral RNA virus elements in mammalian genomes. *Nature*, 463(7277), 84–87. https://doi.org/10.1038/nature08695

Horie, M., & Tomonaga, K. (2019). Paleovirology of bornaviruses: What can be learned from molecular fossils of bornaviruses. *Virus Research*, *262*, 2–9. https://doi.org/10.1016/j.virusres.2018.04.006

Huang, X., Fejes Tóth, K., & Aravin, A. A. (2017). PiRNA Biogenesis in Drosophila melanogaster. *Trends in Genetics*, *33*(11), 882–894. https://doi.org/10.1016/j.tig.2017.09.002

Ipsaro, J. J., Haase, A. D., Knott, S. R., Joshua-Tor, L., & Hannon, G. J. (2012). The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis. *Nature*, *491*(7423), 279-283. https://doi.org/10.1038/nature11502.

Janeway Jr, C. A., Travers, P., Walport, M., & Shlomchik, M. J. (2001). Principles of innate and adaptive immunity. In *Immunobiology: The Immune System in Health and Disease. 5th edition*. Garland Science. URL https://www.ncbi.nlm.nih.gov/books/NBK27090/

Jiang, L., Wang, G., Cheng, T., Yang, Q., Jin, S., Lu, G., Wu, F., Xiao, Y., Xu, H., & Xia, Q. (2012). Resistance to Bombyx mori nucleopolyhedrovirus via overexpression of an endogenous antiviral gene in transgenic silkworms. *Archives of Virology*, *157*(7), 1323–1328. https://doi.org/10.1007/s00705-012-1309-8

Kraemer, M. U., Sinka, M. E., Duda, K. A., Mylne, A. Q., Shearer, F. M., Barker, C. M., ... & Hay, S. I. (2015). The global distribution of the arbovirus vectors Aedes aegypti and Ae. albopictus. *elife*, *4*, e08347. https://doi.org/10.7554/eLife.08347.

Li, P., Piao, Y., Shon, H. S., & Ryu, K. H. (2015). Comparing the normalization methods for the differential analysis of Illumina high-throughput RNA-Seq data. *BMC bioinformatics*, *16*(1), 1-9. https://doi.org/10.1186/s12859-015-0778-7

Liu, Y., Tan, H., Tian, H., Liang, C., Chen, S., & Liu, Q. (2011). Autoantigen La Promotes Efficient RNAi, Antiviral Response, and Transposon Silencing by Facilitating Multiple-Turnover RISC Catalysis. *Molecular Cell*, *44*(3), 502–508. https://doi.org/10.1016/j.molcel.2011.09.011

Löwer, R., Löwer, J., & Kurth, R. (1996). The viruses in all of us: characteristics and biological significance of human endogenous retrovirus sequences. *Proceedings of the National Academy of Sciences*, *93*(11), 5177-5184. https://doi.org/10.1073/pnas.93.11.5177

Mager, D. L., & Freeman, J. D. (1995). HERV-H endogenous retroviruses: presence in the New World branch but amplification in the Old World primate lineage. *Virology,* 213(2), 395-404. https://doi.org/10.1006/viro.1995.0012

Marques, J. T., Kim, K., Wu, P. H., Alleyne, T. M., Jafari, N., & Carthew, R. W. (2010). Loqs and R2D2 act sequentially in the siRNA pathway in Drosophila. *Nature structural & molecular biology*, *17*(1), 24-30.. https://doi.org/10.1038/nsmb.1735

Mehand, M. S., Al-Shorbaji, F., Millett, P., & Murgue, B. (2018). The WHO R&D Blueprint: 2018 review of emerging infectious diseases requiring urgent research and development efforts. *Antiviral Research*, *159*, 63–67. https://doi.org/10.1016/j.antiviral.2018.09.009

Miesen, P., Girardi, E., & van Rij, R. P. (2015). Distinct sets of PIWI proteins produce arbovirus and transposon-derived piRNAs in Aedes aegypti mosquito cells. *Nucleic acids research*, *43*(13), 6545-6556.

Mohn, F., Sienski, G., Handler, D., & Brennecke, J. (2014). The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in Drosophila. *Cell*, *157*(6), 1364-1379. https://doi.org/10.1016/j.cell.2014.04.031

Morazzani, E. M., Wiley, M. R., Murreddu, M. G., Adelman, Z. N., & Myles, K. M. (2012). Production of virus-derived ping-pong-dependent piRNA-like small RNAs in the mosquito soma. *PLoS pathogens*, *8*(1), e1002470. https://doi.org/10.1371/journal.ppat.1002470

Neriya, Y., Kojima, S., Sakiyama, A., Kishimoto, M., Iketani, T., Watanabe, T., ... & Matsumoto, Y. (2022). A comprehensive list of the Bunyavirales replication promoters reveals a unique promoter structure in Nairoviridae differing from other virus families. *Scientific reports*, *12*(1), 1-10. https://doi.org/10.1038/s41598-022-17758-z

Nishimasu, H., Ishizu, H., Saito, K., Fukuhara, S., Kamatani, M. K., Bonnefond, L., ... & Nureki, O. (2012). Structure and function of Zucchini endoribonuclease in piRNA biogenesis. *Nature*, *491*(7423), 284-287. https://doi.org/10.1038/nature11509

Petit, M., Mongelli, V., Frangeul, L., Blanc, H., Jiggins, F., & Saleh, M.-C. (2016). PiRNA pathway is not required for antiviral defense in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences*, *113*(29). https://doi.org/10.1073/pnas.1607952113

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ribet, D., Harper, F., Dupressoir, A., Dewannieux, M., Pierron, G., & Heidmann, T. (2008). An infectious progenitor for the murine IAP retrotransposon: Emergence of an intracellular genetic parasite from an ancient retrovirus. *Genome Research*, *18*(4), 597–609. https://doi.org/10.1101/gr.073486.107

Rorabacher, D. B. (1991). Statistical treatment for rejection of deviant values: critical values of Dixon's" Q" parameter and related subrange ratios at the 95% confidence level. *Analytical Chemistry*, *63*(2), 139-146. https://doi.org/10.1021/ac00002a010

Saito, K., & Siomi, M. C. (2010). Small RNA-Mediated Quiescence of Transposable Elements in Animals. *Developmental Cell*, *19*(5), 687–697. https://doi.org/10.1016/j.devcel.2010.10.011

Saleh, M.-C., Tassetto, M., van Rij, R. P., Goic, B., Gausson, V., Berry, B., Jacquier, C., Antoniewski, C., & Andino, R. (2009). Antiviral immunity in Drosophila requires systemic RNA interference spread. *Nature*, *458*(7236), 346–350. https://doi.org/10.1038/nature07712

Schnettler, E., Donald, C. L., Human, S., Watson, M., Siu, R. W. C., McFarlane, M., Fazakerley, J. K., Kohl, A., & Fragkoudis, R. (2013). Knockdown of piRNA pathway proteins results in enhanced Semliki Forest virus production in mosquito cells. *Journal of General Virology*, *94*(7), 1680–1689. https://doi.org/10.1099/vir.0.053850-0

Shi, C., Zhao, L., Atoni, E., Zeng, W., Hu, X., Matthijnssens, J., Yuan, Z., & Xia, H. (2020). The conservation of a core virome in Aedes mosquitoes across different developmental stages and continent*s* [Preprint]. *Microbiology*. https://doi.org/10.1101/2020.04.23.058701

Soares, Z. G., Gonçalves, A. N. A., de Oliveira, K. P. V., & Marques, J. T. (2014). Viral RNA recognition by the Drosophila small interfering RNA pathway. *Microbes and Infection*, *16*(12), 1013–1021. https://doi.org/10.1016/j.micinf.2014.09.001

Suzuki, Y., Baidaliuk, A., Miesen, P., Frangeul, L., Crist, A. B., Merkling, S. H., Fontaine, A., Lequime, S., Moltini-Conclois, I., Blanc, H., van Rij, R. P., Lambrechts, L., & Saleh, M. C. (2020). Non-retroviral Endogenous Viral Element Limits Cognate Virus Replication in Aedes aegypti Ovaries. *Current biology* : CB, 30(18), 3495–3506.e6. https://doi.org/10.1016/j.cub.2020.06.057

Tassetto, M., Kunitomi, M., Whitfield, Z. J., Dolan, P. T., Sánchez-Vargas, I., Garcia-Knight, M., Ribiero, I., Chen, T., Olson, K. E., & Andino, R. (2019). Control of RNA viruses in mosquito cells through the acquisition of vDNA and endogenous viral elements. *ELife*, *8*, e41244. https://doi.org/10.7554/eLife.41244

ter Horst, A. M., Nigg, J. C., Dekker, F. M., & Falk, B. W. (2019). Endogenous Viral Elements Are Widespread in Arthropod Genomes and Commonly Give Rise to PIWI-Interacting RNAs. *Journal of Virology*, *93*(6), e02124-18. https://doi.org/10.1128/JVI.02124-18

Teng, A. Y., Che, T. L., Zhang, A. R., Zhang, Y. Y., Xu, Q., Wang, T., ... & Fang, L. Q. (2022). Mapping the viruses belonging to the order Bunyavirales in China. *Infectious diseases of poverty*, *11*(04), 43-61. https://doi.org/10.1186/s40249-022-00993-x

Tristem, M. (2000). Identification and Characterization of Novel Human Endogenous Retrovirus Families by Phylogenetic Screening of the Human Genome Mapping Project Database. *Journal of Virology*, *74*(8), 3715–3730. https://doi.org/10.1128/JVI.74.8.3715-3730.2000

Van Rossum, G., & Drake Jr, F. L. (1995). *Python tutorial* (Vol. 620). Amsterdam, The Netherlands: Centrum voor Wiskunde en Informatica. URL https://ir.cwi.nl/pub/5008

von Ende, C. N. (1979). Fish Predation, Interspecific Predation, and the Distribution of Two Chaoborus Species. *Ecology*, *60*(1), 119–128. https://doi.org/10.2307/1936474

Wang, W., Yoshikawa, M., Han, B. W., Izumi, N., Tomari, Y., Weng, Z., & Zamore, P. D. (2014). The Initial Uridine of Primary piRNAs Does Not Create the Tenth Adenine that Is the Hallmark of Secondary piRNAs. *Molecular Cell*, *56*(5), 708–716. https://doi.org/10.1016/j.molcel.2014.10.016

Wang, Y., Jin, B., Liu, P., Li, J., Chen, X., & Gu, J. (2018). PiRNA Profiling of Dengue Virus Type 2-Infected Asian Tiger Mosquito and Midgut Tissues. *Viruses*, *10*(4), 213. https://doi.org/10.3390/v10040213

APPENDIX A

SUPPLEMENTAL INFORMATION

Supplemental Text

Table 5   Command line software parameters used in this study

---

```
#Small RNA trimming
module load BBMap/38.35
bbduk.sh in=SRR11035226.fastq out=clean_SRR11035226.fastq
ref=adapters.fa ktrim=r k=23 mink=11 hdist=1 maq=10 ftm=5 qtrim=lr trimq=10


#Small RNA mapping
module load BBMap/38.35
bbmap.sh in=clean_ SRR11035226.fastq outm=mapped_reads_BARV_S_
SRR11035226.fastq ref=BARV_S.fasta covstats=covstats_BARV_S_ SRR11035226.txt
ambig=all fast minratio=0.90 qtrim=r trimq=10 untrim


#Small RNA length histograms
module load BBMap/38.35
bbmap.sh in= mapped_reads_BARV_S_ SRR11035226.fastq
lhist=lhist_mapped_reads_BARV_S_ SRR11035226.fastq.txt


#Postional nucleotide frequencies found using Python
hl = []
for i in range(30):
    hl.append({'A': 0, 'C': 0, 'G': 0, 'T': 0, 'N':0})


h = open("../Hoo.txt", "w")
with open("foo2.txt", "r") as f:
    for line in f.read().split("\n")[1::2]:
        h.write(line + "\n")
h.close()
```

Table 5 (Continued)

```python
h = open("../Hoo.txt")
nLines = 0
count = 0
for line in h:
    for idx, c in enumerate(line.strip()):
        hl[idx][c] += 1
    nLines += 1

nLines = float(nLines)
for char in ['A', 'C', 'G', 'T', 'N']:
    print("{}\t{}".format(char, "\t".join(["{:0.4f}".format(x[char]/nLines) for x in hl])))
h.close()
```