A mosquito small RNA genomics resource reveals dynamic evolution and host responses to viruses and transposons

Qicheng Ma[1†] Satyam P. Srivastav[1†], Stephanie Gamez[2†], Gargi Dayama[1], Fabiana Feitosa-Suntheimer[3], Edward I. Patterson[4], Rebecca M. Johnson[5], Erik M. Matson[1], Alexander S. Gold[3], Douglas E. Brackney[6], John H. Connor[3], Tonya M. Colpitts[3], Grant L. Hughes[4], Jason L. Rasgon[5], Tony Nolan[4], Omar S. Akbari[2], and Nelson C. Lau[1,7*]

1. Boston University School of Medicine, Department of Biochemistry
2. University of California San Diego, Division of Biological Sciences, Section of Cell and Developmental Biology, La Jolla, CA 92093-0335, USA.
3. Boston University School of Medicine, Department of Microbiology and the National Emerging Infectious Disease Laboratory
4. Departments of Vector Biology and Tropical Disease Biology, Centre for Neglected Tropical Diseases, Liverpool School of Tropical Medicine, Liverpool L3 5QA, UK
5. Pennsylvania State University, Department of Entomology, Center for Infectious Disease Dynamics, and the Huck Institutes for the Life Sciences
6. Department of Environmental Sciences, The Connecticut Agricultural Experiment Station
7. Boston University Genome Science Institute and the National Emerging Infectious Disease Laboratory

* Corresponding author: NCL: nclau@bu.edu
† These authors contributed equally to this study.

Running title: Mosquito small RNA genomics

Keywords: mosquitoes, small RNAs, piRNAs, viruses, transposons microRNAs, siRNAs

ABSTRACT

Although mosquitoes are major transmission vectors for pathogenic arboviruses, viral infection has little impact on mosquito health. This immunity is due in part to mosquito RNA interference (RNAi) pathways that generate antiviral small interfering RNAs (siRNAs) and Piwi-interacting RNAs (piRNAs). RNAi also maintains genome integrity by potently repressing mosquito transposon activity in the germline and soma. However, viral and transposon small RNA regulatory pathways have not been systematically examined together in mosquitoes. Therefore, we developed an integrated Mosquito Small RNA Genomics (MSRG) resource that analyzes the transposon and virus small RNA profiles in mosquito cell cultures and somatic and gonadal tissues across four medically important mosquito species. Our resource captures both somatic and gonadal small RNA expression profiles within mosquito cell cultures, and we report the evolutionary dynamics of a novel Mosquito-Conserved piRNA Cluster Locus (MCpiRCL) composed of satellite DNA repeats. In the larger culicine mosquito genomes we detected highly regular periodicity in piRNA biogenesis patterns coinciding with the expansion of Piwi pathway genes. Finally, our resource enables detection of crosstalk between piRNA and siRNA populations in mosquito cells during a response to virus infection. The MSRG resource will aid efforts to dissect and combat the capacity of mosquitoes to tolerate and spread arboviruses.

INTRODUCTION

Mosquitoes are one of the most prevalent vectors of human pathogens, yet they have wide variability to vector different pathogens. For example, human malaria parasites are exclusively vectored by anopheline mosquitoes which transmit few viruses other than O'Nyong nyong virus (ONNV) and Mayaro virus (Vanlandingham et al. 2006; Brustolin et al. 2018). In contrast, culicine mosquitoes transmit many human viral pathogens, such as dengue virus (DENV), Zika virus (ZIKV), Chikungunya virus (CHIKV) and yellow fever virus (YFV) in tropical climates where *AeAlbo* and *AeAeg* thrive; whereas eastern equine encephalitis virus (EEEV) and West Nile Virus (WNV) spread mainly in *Culex* mosquitoes inhabiting temperate climates (Olson and Blair 2015; Londono-Renteria and Colpitts 2016; Halbach et al. 2017; Lambrechts and Saleh 2019).

Since vector-pathogen interactions are complex, no dominant theory yet explains why anopheline mosquitoes are less prolific than culicine mosquitoes in spreading arboviruses. Arbovirus infections in humans lead to devastating symptoms including fever, nausea, bleeding, extreme pain, brain damage and death. However, culicine mosquitoes are practically unaffected from active arbovirus replication (Goic and Saleh 2012; Olson and Blair 2015; Lambrechts and Saleh 2019) and therefore are highly competent transmitters of arboviruses to human hosts.

Three main classes of animal small regulatory RNAs are microRNAs (miRNAs) and endogenous small-interfering RNAs (endo-siRNAs), which range in size between 18-23nt long and are typically bound by Argonaute proteins; and Piwi-interacting RNAs (piRNAs) that are bound by Piwi proteins and mainly range in size between 24-32nt in length in most animals. In the model Dipteran, *Drosophila melanogaster* (*Dmel*), the

small RNAs comprise 258 miRNA genes (Kozomara et al. 2019), ~20 large intergenic piRNA cluster loci (Brennecke et al. 2007; Malone et al. 2009; Wen et al. 2014), >1000 genic piRNA cluster loci (Robine et al. 2009; Wen et al. 2014; Chirn et al. 2015), and >1000 endogenous siRNA loci generating either large fold-back transcripts or sense-antisense pairing transcripts (Czech et al. 2008; Ghildiyal et al. 2008; Kawamura et al. 2008; Mirkovic-Hosle and Forstemann 2014; Wen et al. 2014; Wen et al. 2015). Lastly, arbovirus-specific siRNAs and piRNAs persist in *Dmel* cell cultures (Flynt et al. 2009; Wu et al. 2010; Vodovar et al. 2011; Goic et al. 2013; Wen et al. 2014; Palmer et al. 2018).

Culicidae mosquitoes are relatives of Drosophilid fruit flies as members of the Dipteran insect clade (**Figure 1A**, (Wiegmann et al. 2011)), yet ~260 Million Years Ago (MYA) of evolutionary distance between Drosophilids and Culicidae imparts physiological and molecular differences in small RNA compositions. Within mosquito phylogeny, the anopheline subclade represented by *Anopheles gambiae* (*AnGam*) displays stronger chromosome synteny to Drosophilids than the culicine subclade of mosquitoes such as *Culex quinquefasciatus* (*CuQuin*), *Aedes aegypti* (*AeAeg*) and *Aedes albopictus* (*AeAlbo*) (Dudchenko et al. 2017). Indeed, *AnGam*'s genome (~0.28Gb) is as compact as *Dmel*'s genome (~0.18Gb), whereas culicine mosquito genomes are an order of magnitude greater in size due to numerous non-coding and repetitive elements (Fig. 1C) (Rai and Black 1999; Holt et al. 2002; Nene et al. 2007; Arensburger et al. 2010; Chen et al. 2015; Dudchenko et al. 2017; Matthews et al. 2018; Palatini et al. 2020).

Since many viruses replicate their RNA genomes via a double-stranded RNA (dsRNA) intermediate, the conserved RNA interference (RNAi) pathway provides

antiviral activity through Dicer and Argonaute enzymes converting viral dsRNA into siRNAs for repressing viruses (Samuel et al. 2018; Guo et al. 2019). Recently, the piRNA pathway was also implicated in assisting the siRNA pathway with antiviral response in the culicine mosquitoes and cell culture lines (Goic and Saleh 2012; Olson and Blair 2015; Halbach et al. 2017; Lambrechts and Saleh 2019).

A key knowledge gap is to what degree viral siRNAs and piRNAs comprise of or affect mosquito small RNA transcriptomes. Previous mosquito studies have mainly focused on either virus derived small RNAs (Myles et al. 2008; Myles et al. 2009; Sanchez-Vargas et al. 2009; Brackney et al. 2010; Scott et al. 2010; Hess et al. 2011; Morazzani et al. 2012; Saldana et al. 2017; Varjak et al. 2017a; Varjak et al. 2017b; Ruckert et al. 2019); or conducted genomic analyses on earlier incomplete assemblies and preliminary annotations of individual mosquito species (Akbari et al. 2013; Whitfield et al. 2017; Tassetto et al. 2019). In this study, we generated >50 new small RNA libraries from cell cultures, male and female gonads and respective carcasses from four medically important mosquito species (*AnGam, CuQuin, AeAeg, AeAlbo*) to add to the trove of publicly available small RNA libraries. We then implemented our small RNA analysis pipeline to enable cross-species comparisons. Our analysis provides the first comprehensive view of small RNA transcriptomes across mosquito phylogeny, reveals novel evolutionary and host dynamics in viral and somatic piRNA production and uncovers notable periodicity in phased piRNA biogenesis patterns within culicine mosquitoes.

RESULTS

Framework for integrated small RNA analysis across four mosquito species.

We previously built functional annotation pipelines for small RNA libraries generated from the gonads of Drosophilids, mammals and other vertebrates (Chirn et al. 2015). To extend this pipeline to compare small RNAs across mosquito genomes (Fig. 1B), we added a curated list of arboviruses. We queried NCBI GenBank for mosquito arboviruses and viral gene names (Nanfack Minkeu and Vernick 2018; Zakrzewski et al. 2018) and the Virus Pathogen Resource (VIPR)(Pickett et al. 2012), to make a list of 225 mosquito arboviruses in May 2019 that exceeds the 107 Drosophilid viruses listed in (Palmer et al. 2018). We manually inspected entries to reduce redundancy amongst similar entries that are just slight sequence variants of a single virus class.

Our study took advantage of new genome assemblies of various culicine mosquito species and additional genome annotation resources from the legacy VectorBase database (Holt et al. 2002; Nene et al. 2007; Arensburger et al. 2010; Bartholomay et al. 2010; Giraldo-Calderon et al. 2015). *AeAeg* and *AeAlbo* genome assemblies were enhanced with Hi-C information and longer reads sequencing to connect scaffolds into chromosomal assembles (Dudchenko et al. 2017; Matthews et al. 2018; Palatini et al. 2020). From these assemblies, the transposon consensus sequences list were processed to reduce redundancy (**Figure S1 and Supplemental Materials**). Lastly, we curated viruses and transposon consensus lists (**Supplemental Files 1–7**) and the compendium of outputs in a publicly accessible database resource of Mosquito Small RNA Genomics (MSRG, https://laulab.bu.edu/msrg/).

MSRG outputs are organized by the four individual species, with species-specific results described in the Supplementary text and in Supplementary **Figure S2** (*AnGam*),

Figure S3 (*CuQuin*), **Figure S4** (*AeAeg*) and **Figure S5** (*AeAlbo*). These full galleries show complete species-focused analyses of endogenous and arboviral small RNA functional classes and features. The standard culture conditions for the mosquito cells profiled in this study are described in Supplementary **Table S1**, whereas the sequencing statistics of the libraries analyzed per species as well as the curated lists of genic and intergenic piRNA-containing loci are in Supplementary **Table S2** (*AnGam* Metatable), **Table S3** (*DMel* Metatable), **Table S4** (*CuQuin* Metatable), **Table S5** (*AeAeg* Metatable) and **Table S6** (*AeAlbo* Metatable). These outputs enabled comparison between samples and species libraries to derive insights into virus- and transposon-targeting features by the mosquito small RNA transcriptomes.

Multiple common arboviruses persistently infect and generate small RNAs in mosquito cell cultures.

Since many mosquito cell cultures were generated decades ago (Table S1), we expected they would carry viral small RNAs from persistent arbovirus infections (**Figure 2**). However, specific arboviruses could also infect across multiple Dipteran species. For example, consistent with earlier reports (Chandler et al. 2014; Zhang et al. 2016; Maringer et al. 2017; Di Giallonardo et al. 2018; Weger-Lucarelli et al. 2018), there was broad distribution of Phasi Charoen-like virus (PCLV) and Cell Fusing Agent virus (CFAV) viral piRNAs amongst different species of culicine mosquito cell lines (Fig. 2B,C). We also detected viral small RNAs in the *AnGam* Sua5b-JR line and the *AeAeg* CCL-125-JC and Aag2-CB lines from the *Drosophila* American Nodavirus (*Dme*l ANV; related to Flock House virus or FHV, Fig. 2D) that persistently infects *Drosophila* Schneider 2 (S2) line and OSS cells (Aliyari et al. 2008; Flynt et al. 2009; Wu et al.

2010; Han et al. 2011). In addition, abundant viral siRNAs from *Culex* Y virus (CYV) were in *AnGam*, *AeAeg*, and *AeAlbo* cell lines (Fig. 2E). These data support the broadness of these arbovirus tropisms spanning these Dipteran species.

The *AeAeg* densovirus is a small single-stranded virus previously developed for gene transduction of mosquitoes and mosquito cell cultures (Afanasiev et al. 1994; Afanasiev et al. 1999). Our analyses revealed densoviral siRNAs and piRNAs across many cell lines except for the *AeAlbo* C7/10 and U4.4 cells (Fig. 2A). We detected abundant antisense densoviral piRNAs in the *AnGam* Mos55-JR line (-JR from the Rasgon lab) versus no densoviral small RNAs in the Mos55-TC line (-TC from the Colpitts lab), yet both displayed a persistent infection of densovirus (**Figure S6A),** suggesting that densovirus genome integration enables Mos55-JR to generate the densoviral piRNAs. Persistent densovirus infections in C6/36 cells had been proposed to enable stable coinfections with DENV2 (Burivong et al. 2004; Kanthong et al. 2008), suggesting a selective advantage for cells to harbor densovirus.

Recently, persistent infections of mosquito cell cultures by pathogenic arboviruses like flaviviruses and alphaviruses have been re-examined (Avila-Bonilla et al. 2017; Fredericks et al. 2019; Koh et al. 2019; Reyes-Ruiz et al. 2019). Amongst our mosquito cell cultures, we also discovered persistent viral infections reflected by abundant viral siRNAs against ONNV in the Mos55-JR line, and DENV2 siRNAs and piRNAs in the Aag2-TC line (Fig. 2F). Perhaps similarly to how *Dmel* ANV may have passed between *Drosophila* cells to mosquito cells, these infections were most likely inadvertent. Lastly, abundant viral piRNAs from *AeAeg* Anphevirus-1a were detected in the CCL-125-JC line but not in our Aag2 cells which are reported to also be persistently infected (Di Giallonardo et al. 2018; Parry and Asgari 2018), reflecting the similar

dichotomy of persistent densovirus in both Mos55 cell strains but densoviral piRNAs only expressed in one of the strains.

Higher levels of somatic piRNAs in mosquitoes with persistent arboviral small RNAs.

Animal piRNAs mainly silence transposons in gonads to ensure fertility, with less evidence for somatic functions in mammals where somatic piRNAs are lowly expressed. However, mosquitoes are like most other insects expressing significant somatic piRNAs, and only *Drosophila* is the outlier for low levels of somatic piRNAs (Lewis et al. 2018; Genzor et al. 2019). In spite of this, some mosquito carcasses had subdued amounts of somatic piRNAs (**Figure 3A**: *AeAeg* Female and male carcasses from BH; *AnGam* Male and Female carcasses from TN, and *CuQuin* Male and female carcasses). This contrasted other mosquito carcasses containing abundant somatic piRNAs (Fig. 3B: *AeAeg* Female carcasses from FJ, TC and GH; and *AeAlbo* Male and Female carcasses from OA).

What could explain this variation of somatic piRNA levels amongst different isolates of the same species of *AeAeg*? We ruled out unintended detection bias like residual gonads contaminating carcass, since there were no contaminating germline transcripts like *vasa*. We then hypothesized that three *AeAeg* isolates with abundant somatic piRNAs may be due to persistent arbovirus infection as reflected by viral small RNAs. This hypothesis was supported by the absence of viral small RNAs in the *CuQuin* samples we analyzed, the *AeAeg* isolate from the Hay lab (Akbari et al. 2013), and the *AnGam* isolate from the Nolan lab (this study and (Castellano et al. 2015)).

Indeed, our analysis showed that *AeAeg* isolates with abundant somatic piRNAs also carried persistent arbovirus infections reflected by viral small RNAs (Fig. 3B). The

FJ *AeAeg* strain from Miami, USA (Lewis et al. 2018) expressed *AeAeg* Anphevirus strain-1a piRNAs, and viral siRNAs from the Humaita-Tubiacanga virus (HTV), similar to HTV siRNAs detected in *AeAeg* strains from Rio de Janeiro, Brazil (Aguiar et al. 2015). In the *AeAeg* TC isolate of the ROCK strain, we detected *Dmel* ANV siRNAs and densovirus small RNAs. Lastly, the GH *AeAeg* strain from Galveston, USA, harbored persistent CFAV (Kim et al. 2009) and both CFAV siRNAs and piRNAs in the ovary and carcass (Fig. 3B).

Somatic piRNA levels were also high in the OA *AeAlbo* strain from Los Angeles, USA (Gamez et al. 2020), which correlated with persistent *Dmel* ANV (Fig. 3B). Other reports have described *AeAlbo* viral small RNAs from densovirus (Morazzani et al. 2012) and ONNV (Wang et al. 2018), which are circulating in wild mosquito populations. We speculate the *Drosophila* lab stocks, a reservoir for nodaviruses (Goic et al. 2013; Kandul et al. 2019) could explain these Drosophilid arboviruses persisting in *AeAlbo* strains (Fig. S5E).

Potential crosstalk between flavivirus infection and endogenous small RNA levels.

Despite wide competency of *AeAeg* cells and mosquitoes to support arbovirus replication, viral piRNAs are a minor fraction of total small RNAs, even with ectopic infections of CHIKV, DENV or ZIKV (<~6%, Fig. 3A, 3B, S4A, S4B). *AeAeg* mosquitoes and cell cultures appear unaffected by arbovirus infection presumably because antiviral RNAi pathways are generating viral siRNAs and piRNAs (Aliyari and Ding 2009; Karlikow et al. 2014; Blair and Olson 2015; Samuel et al. 2018). However, new infections from pathogenic viruses can be affected by persistent infections of other arboviruses, perhaps through small RNA crosstalk (Burivong et al. 2004; Kanthong et

al. 2008; Myles et al. 2008; Kanthong et al. 2010; Goic and Saleh 2012; Parry and Asgari 2018; Reyes-Ruiz et al. 2019).

To see if flavivirus infection affected endogenous small RNA levels in mosquitoes and cell cultures, we reanalyzed small RNAs from female *AeAeg* mosquitoes fed blood that lacked or contained ZIKV. We reconfirmed that both ZIKV siRNAs and piRNAs were only detectable 7- and 14-days post-infection (Saldana et al. 2017). Whereas bulk overall small RNAs were the same whether the mosquitoes harbored ZIKV or not (**Figure S7A**), our analysis revealed new piRNAs from a specific region of CFAV only stimulated after ZIKV replication (blue arrows in **Figure 4A**). This region did not have specific homology to ZIKV piRNAs but generated both plus and minus strand piRNAs indicative of the “ping-pong” mode of piRNA interactions. Despite clear signals of ZIKV and CFAV small RNAs, these viral small RNAs were only a tiny fraction of the total small RNAs samples in these libraries (Fig. S7A).

Next, we tested if flavivirus infections of Aag2 cells with DENV and ZIKV might also affect CFAV small RNA patterns. Therefore, we performed DENV and ZIKV infections at two different multiplicities of infection (MOI, 0.1 and 0.01) of Aag2 cells, including two strains of the DENV2 serotype (NGC-a high passage and K0048-low passage) (Troupin et al. 2016); as well as the Old World (OW) and Puerto Rico (PR) isolates of ZIKV (Araujo et al. 2020) (Fig. 4B). Because the Aag2 cells were incubated for 7 days post inoculation, the higher MOI=0.1 left fewer cells and viral RNAs remaining compared to the lower MOI=0.01 (Fig. S7B).

Flavivirus small RNAs correlated with viral genomic RNA levels measured by qRT-PCR, but there was variability in the proportions of flavivirus siRNAs and piRNAs (Fig. S7C). The unusual patterns of abundant singular DENV piRNAs from the plus

strand that we observed was consistent with other studies (Scott et al. 2010; Hess et al. 2011; Miesen et al. 2016). Whereas DENV and ZIKV siRNAs were generated from both plus and minus strands indicative of a dsRNA precursor, the viral piRNAs were biased from the plus strand and predominantly arose from a few very abundant reads (Fig. S6F bottom plots), recapitulating the same confounding patterns observed by others (Goic et al. 2016; Miesen et al. 2016; Whitfield et al. 2017; Merkling et al. 2020). This pattern of viral piRNA accumulation defies the generalized biogenesis patterns of phased piRNAs (Han et al. 2015; Mohn et al. 2015; Pandey et al. 2017; Gainetdinov et al. 2018; Izumi et al. 2020).

Although both batches of Mock Control Aag2 cells had expected bimodal distributions of 18-23nt siRNAs and miRNAs versus 24-32nt piRNAs, we observed instances were these distributions were greatly affected by viral infection. In both replicates, DENV2K0048 distorted these two distributions, in one case greatly enhancing endogenous siRNAs while depressing piRNAs, and in another case a vice versa response (Fig. 4B, red arrows). Also, in both replicates, the ZIKV_OW infections enhanced endogenous siRNAs while depressing piRNAs, while this was vice versa in one ZIKV_PR infection. Although DENV2NGC, the high passage strain, repeatedly lacked impact on small RNA populations, there was marked variability in one of the experiments but not in the other for when DENV1, DENV3 and DENV4 infections greatly affected the bimodal distribution of piRNAs versus siRNAs and miRNAs.

Future studies will dissect this variability in Aag2 cells' small RNA populations during arbovirus infection. However, two batches of Mock Control Aag2 cells already displayed enhanced minus-strand piRNAs similar to the region of CFAV piRNAs amplified in the ZIKV-infected mosquitoes (Fig. 4C). Because Aag2 cells are already

persistently infected by multiple arboviruses, the DENV and ZIKV infections did not affect these CFAV piRNAs corresponding to the NS2A gene (Fig. 4D). What specifies the NS2A gene as a piRNA precursor and CFAV 3'UTR as a stronger initiator of siRNA biogenesis remains unclear (Fig. 4A), although other flavivirus 3'UTRs have been described to have an antiviral role (Moon et al. 2015).

Repetitive element targeting by endogenous piRNAs

Mosquito genomic insertions called Endogenous Viral Elements (EVEs) were proposed to have an antiviral role by generating endogenous piRNAs complementary to flavivirus sequences (Katzourakis and Gifford 2010; Lequime and Lambrechts 2017; Suzuki et al. 2017; Whitfield et al. 2017; Houe et al. 2019; Tassetto et al. 2019; Blair et al. 2020)). The most active EVE in our dataset, the AEFE1/AY347953 EVE has homology to the NS5 gene of flaviviruses like Kamiti River virus and CFAV (Crochu et al. 2004), and predominantly generated piRNAs with fewer siRNAs in the gonads, soma and cell lines (**Figure 5A**). In contrast, antisense piRNAs to PCLV, largely from the S-fragment of the PCLV genome (Fig. 2B) suggests this is also an EVE signature (Whitfield et al. 2017; Tassetto et al. 2019). *AeAeg* mosquitoes *and* cell cultures produced significant CFAV small RNAs from the CFAV-like EVE which should theoretically target CFAV (Fig. 2C, Fig. 3B,) (Suzuki et al. 2017; Whitfield et al. 2017), yet there is persisting replication of CFAV RNAs in the GH *AeAeg* isolate (Fig. 4A). We were unable to cross-reference other analyses of *AeAeg* EVEs (Whitfield et al. 2017; Tassetto et al. 2019) because this was performed on an incomplete genome assembly from their isolate of the Aag2 cell line. In summary, EVEs may be contemporary

versions of the more ancient LTR-containing transposons that are templates for abundantly generating small RNAs.

Amongst the other most prominent mosquito transposons to generate piRNAs in both cell cultures and animals were LTR-containing transposons, along with notable LINE-like retrotransposons and the Tc1 DNA-type transposon in *AeAlbo* and *AnGam*, respectively (Fig. 5A). There were also cell line-specific and soma-versus-germline differences in small RNA targeting of transposons, with the greatest number of transposons with small RNA targeting evident in the germline tissues (clustering heatmaps and coverage plots in Fig. S2E,F; S3E,F; S4F,H; and S5E,I).

Piwi proteins require antisense piRNAs to target transposon sense transcripts (Post et al. 2014; Batki et al. 2019), so we expected *Drosophila* small RNAs to have a biased ratio of ~3.8:1, antisense:sense mapping to transposons (Fig. 5B). Although *AnGam* had a lower fraction of small RNAs mapping to transposons than *Drosophila* (~6% versus ~18%), the culicine mosquitoes had the lowest proportion of small RNAs mapping antisense to transposons. In fact, *CuQuin* small RNAs were slightly biased for sense mapping reads to repeats such as the top examples of an LTR-Gypsy transposon and rDNA repeats small RNAs (Fig. 5A, B). Although we cannot explain this *CuQuin* discrepancy, other differences in our transposon piRNA quantitation, such as *AeAlbo* piRNAs measured in (Liu et al. 2016), can be attributed to using the newer *AeAlbo* assembly (Palatini et al. 2020) and reducing the redundancy in repeats lists (Fig. S1).

For *Drosophila* to generate piRNAs antisense to transposons, the transposon sequences in major piRNA cluster loci (piRCL) are oriented antisense to the single plus strand precursor transcript like in the *flamenco* locus (Li et al. 2009; Malone et al. 2009). Although *flamenco* homologs are only conserved in the closest relatives of *D.*

melanogaster (Chirn et al. 2015), *flamenco* is notable for its high uni-strand expression of piRNAs in the somatic compartment of *Drosophila* follicle cells and dense insertions of transposons and repeats. Only a few instances of the largest piRCLs in mosquitoes display similar features of uni-strand piRNA expression both in the germline and soma proper (**Figure 6A,** Fig. S2-S5**,** Tables S2-S6). However, in contrast to *Drosophila flamenco*, the transposon density in these "*flamenco*-like" mosquito piRCLs appears lower and with fewer piRNAs directly overlapping transposon sequences (Fig. 6A). One of our determinations was also confirmed by the Marques lab annotation of a "*flamenco*-like" cluster in *AeAeg* (Aguiar et al. 2020), and through genome synteny, we found a homologous piRCL in *AeAlbo* but it is half the size of its counterpart in *AeAeg* (~72kb versus ~142kb, Fig 6A). These observations underlie the dynamic evolution of these piRCLs amongst mosquitoes.

A major genic piRCL is dynamically evolving yet syntenically conserved through mosquito phylogeny.

To define other genic and intergenic piRCLs in mosquitoes (Table S2, S4-S6), we combined automated genome scanning with manual curation. The six top major *AeAlbo* piRCLs exist on three super-scaffolds, with mostly single-stranded biases in the small RNA expression patterns (Fig. S5J,K). Two of these AeAlbo genic piRCLs displayed patterns of satellite DNA repeats. (Fig. S5J, rightmost windows), which we also observed in other *CuQuin* and *AeAeg* piRCLs with satellite DNA repeats generating very abundant amounts of piRNAs (Fig. S3H, S4J) but no such satellite DNA repeats in *AnGam*. In addition, the lack of synteny around these piRCLs made it challenging to compare these particular piRCL across the mosquito species.

However, one *AeAlbo* piRCL with satellite DNA repeats enabled comparative genomics because it was linked to protein-coding genes (Fig. 6B-ii). Expressed very highly in *AeAlbo* gonads, somatic tissues, and cell cultures, this genic piRCL generates on average >10,000 reads per million (rpm) from mainly two major piRNAs which have 33 and 27 alternating repeats spread out in a ~5.6kb region (Fig. 6B-iii). The *AeAeg* orthologous gene also contained a genic piRCL with satellite DNA repeats and identical piRNA sequences, but a different arrangement of 21 and 19 alternating repeats (Fig. 6B-iii, second row).

The orthologous *CuQuin* genic piRCL also displayed satellite DNA repeats with two alternating piRNA sequences from 17 and 26 repeats abundantly expressed in gonads, somatic tissues, and the Hsu cell line (Fig. 6B-ii, third row). One satellite piRNA's primary sequence, "UUUCGGAUAUGUUUUAGAAAUUCGUUUUU", is perfectly conserved across mosquito evolution (Fig. 1A) but its repeat number has evolved from 17 sites in *Culex* to 21 and 33 sites in *Aedes* species. Notably, the other *Culex* satellite piRNA sequence differs from the *Aedes* sequence only by the first nucleotide of 5'-"C" in *Culex* and 5'-"G" in *Aedes* in each of 26 repeats in *CuQuin* versus the 19 and 27 sites in *AeAeg* and *AeAlbo*, respectively (Fig. 6B-iii). The most parsimonious explanation for this type of sequence evolution is a base change first in the early divergence of their ancestors and then parallel evolutionary expansion of the mutated piRNA sequence to form these satellite DNA repeats.

In accordance with the long divergence between culicine and anopheline mosquitoes, *AnGam* appears to lack piRCLs containing satellite DNA repeats, however the orthologous genic piRCL extends to the *AnGam* gene AGAP003387 (Fig. 6B, fourth row). In contrast to the culicine genic piRCL, this *AnGam* piRCL is very compact at

~500bp long within the 3'UTR of AGAP003387 with no tandem repeats but has four main piRNAs comprising >~1500 rpm. Two of these *AnGam* piRNAs were perfectly conserved at the primary sequence level as one of the culicine satellite DNA piRNAs (Fig. 6B-iii), and this *AnGam* piRCL was also abundantly expressed in *AnGam* gonads and cell cultures. The gene AGAP003387 only has homologs within other mosquitoes, whereas a neighboring gene AGAP003388 is homologous to the *Dmel* gene CG5746 that does generate some 3'UTR piRNAs (Chirn et al. 2015). Therefore, we have named this a Mosquito-Conserved piRNA Cluster Locus (MCpiRCL).

The *AnGam* piRCL may represent the ancestral mosquito locus > ~200 MYA that began as genic piRCL region already primed to express important piRNAs. As the culicine branch expanded their genomes with transposon repeats, the MCpiRCL also gained satellite DNA repeat perhaps to amplify piRNA expression. This satellite DNA piRCL was also discovered in *AeAeg* by (Halbach et al. 2020), and was proposed to cause maternally-deposited transcripts to turnover during embryogenesis, similar to the vertebrate tandem repeat cluster of miRNAs miR-430 and miR-427 (Giraldez et al. 2006; Lund et al. 2009). However, whereas miR-430 and miR-427 expression is restricted to the embryo, the MCpiRCL in all four of these mosquitoes is expressed throughout the gonads, somatic tissues, and cell culture lines (Fig 6B-iii), suggesting the targeting capacity of these piRNAs may be broader than maternally-deposited transcripts. We predicted many hundreds of transcripts and highlight the top two mRNA, transposon, and virus targets in **Figure S8**. Although the incomplete draft CpipJ2 genome assembly and annotation (Arensburger et al. 2010) may be limiting the number of predicted *CuQuin* targets, there is an expanded repertoire of potential gene and transposon targets for the *AeAeg* and *AeAlbo* piRNAs from this MCpiRCL.

Culicine mosquitoes exhibit periodicity to the patterns of piRNA biogenesis.

Only culicine mosquitoes contained piRCL with satellite DNA repeats (Fig. 6B, S3H, S4J, S5J), and these single abundant piRNAs were biased on one strand and spaced out from each other by a >29nt gap. This piRCL configuration challenges the prototypical phasing pattern of primary piRNA biogenesis first described in *Dmel* (Han et al. 2015; Mohn et al. 2015; Pandey et al. 2017; Gainetdinov et al. 2018; Izumi et al. 2020). Indeed, a previous study applying piRNA phasing algorithms across piRNA datasets from a phylogenetic spectrum of hydra to insects to mammals showed that *AeAeg* piRNAs stood out with the most periodic of 5' to 5' piRNA distance peaks (Gainetdinov et al. 2018).

We applied the same algorithm of a LOWESS non-parametric regression and auto-correlation smoothing (Gainetdinov et al. 2018) to a wide number of *Dmel*, *AnGam*, *CuQuin*, *AeAeg* and *AeAlbo* libraries. We confirmed the strong conservation throughout Dipterans of the one piRNA phasing mechanism that juxtaposes the 3' terminus of the upstream piRNA to the 5' start of the downstream piRNA (**Figure 7A, Figure S9**). There was also a very periodic 5'-to-5' phasing pattern for the *CuQuin*, *AeAeg*, and *AeAlbo* samples, both in mosquito tissues and cell cultures (Fig. 7A). However, this periodic pattern was dampened in *AnGam* and *Dmel*, with perhaps only *Dmel* ovarian small RNAs subjected to beta-elimination showing the enhanced periodic signal (Song et al. 2014).

We speculate the expansion of Piwi pathway genes in culicine mosquitoes (Lewis et al. 2016) may promote periodicity in piRNA phasing biogenesis patterns while also enabling the innovation of satellite DNA repeats in piRCL. To re-examine the

evolutionary relationships of Dipteran Piwi pathway genes, we took *Dmel* Piwi pathway genes and conducted BLASTP and manual curation between NCBI GenBank and VectorBase to better define the mosquito homologs (**Table S7**). Ten core Piwi pathway genes in *Dmel* had single orthologs in *AnGam* that were then expanded into multiple homologs in culicine lineages (**Figure S10A, Table S8**). *AeAlbo* stands out from *AeAeg* and *CuQuin* with the most expanded Piwi pathway gene families including two *Ago3* homologs and three homologs of *valois* and *vreteno* (Fig. S10A). Another fifteen Piwi pathway genes from *Dmel* had single orthologs in mosquitoes (Fig. S10B). Perhaps the expansion of *piwi*/*aub* homologs in culicine mosquitoes explains piRCL innovation such as *AeAeg* PIWI4 being required for the satellite repeat MCpiRCL (Halbach et al. 2020). Although seven *Dmel* genes in *Drosophila's* piRNA-mediated transcriptional silencing pathways (i.e. *panx, rhi, del,* and *cuff* (Le Thomas et al. 2014; Mohn et al. 2014; Zhang et al. 2014)) were completely absent in mosquito genomes, this may foretell potential mosquito-specific factors required for its unique repertoire of Piwi pathway genes.

Lastly, to examine whether more Piwi pathway genes in culicine mosquitoes might impact piRNA 'ping-pong' biogenesis mechanisms, we adapted the autocorrelation algorithm to count the frequencies of 5'-to-5' distances of piRNA reads mapping on the opposite strand, and then noted the Z_{10} scores > 2 as a signal that piRNA ping-pong signatures were significant (Fig. 7B). We also analyzed siRNA reads with this same algorithm but noting Z_{21} scores > 2 as a signal of siRNA duplexes processed by Dicer. The piRNA ping-pong signatures were strong in all mosquito cell culture lines and gonads, but the ping-pong signature present in the carcasses of *AnGam*, *AeAeg* and *AeAlbo* were absent in *CuQuin* carcasses. In most of the mosquito carcasses and some of the cell lines, an siRNA duplex signature was evident. From

these results, we interpret that piRNA ping-pong mechanisms and Dicer-generation of siRNA duplexes generally remain the same amongst these Dipterans.

DISCUSSION

Cell cultures are invaluable for genomic studies as demonstrated by the important genomic, transcriptomic and epigenetic datasets for model organism and human cell lines in the modENCODE and ENCODE projects, respectively (Graveley et al. 2011; Kharchenko et al. 2011; Negre et al. 2011; The ENCODE Project Consortium 2012; Djebali et al. 2012; Thurman et al. 2012). Mosquito cell cultures from various species (Fig. 1C) also facilitate virology studies, and our study can place cell lines in better context to the tissues of the animal. For example, our Principal Component Analysis (PCA) plots (**Figure S11**) and hierarchical clustering of miRNA and transposon small RNA profiles show that cell cultures have a distinct transcriptomes from gonads and somatic tissues (Fig. S2D,E; S3D,E; S4E,J; and S5G,H). However, the PCA plots also suggest that different labs' isolates of *AnGam*, *CuQuin* and *AeAeg* cell cultures showed a higher degree of clustering together than the cell lines from *AeAlbo*.

Mosquitoes have a major translational impact on human health, yet genomic characterizations of the culicine mosquitoes have lagged because their significantly larger genomes are inflated by repetitive elements. New genomic approaches such as high-throughput long-read and Hi-C sequencing may bridge scaffolding gaps to bring about major improvements in the *AeAeg* and *AeAlbo* genome assemblies (Dudchenko et al. 2017; Matthews et al. 2018; Palatini et al. 2020). However, functional annotations such as improving gene models with better transcriptome data is still needed for

mosquito genomics advancement including this study in which we opted to analyze the CpipJ2 assembly that had genes and repeats tables (Arensburger et al. 2010) but was still more fragmented than the newer CpipJ3 assembly which lacked annotation (Dudchenko et al. 2017). Our study also demonstrates the need for better repetitive elements annotations including refinement of transposons beyond the automated programs like RepeatModeler (Wheeler et al. 2013; Flynn et al. 2020) which generate comprehensive but redundant repeats list. Strangely, the majority of mosquito piRNAs across species do not appear to target transposons and may ultimately have a wide range of other targets yet to be determined.

As the diversity of *Dmel* cell culture lines has greatly expanded just in the last decade, only 4 *Dmel* lines are known to express piRNAs (fGS/OSS, OSS-OSCs-OSC-delta-MBT, WRR1 and Kc cells, (Lau et al. 2009; Saito et al. 2009; Fagegaltier et al. 2016; Sumiyoshi et al. 2016; Vrettos et al. 2017)), while the vast majority of *Dmel* cell lines only express miRNAs and siRNAs (Wen et al. 2014). Such few piRNA-expressing *Dmel* cell lines may reflect the exceptional nature of *Dmel* to restrict Piwi pathway gene expression to the gonads, whereas most other insects robustly express piRNAs in the soma (Lewis et al. 2018). The smaller selection of mosquito cell lines (Table S1) coupled with their long history would contribute to their gene expression profiles diverging greatly from mosquito tissues. Yet every mosquito cell line in this study expressed piRNAs, including our culture of C7/10 cells (Fig. 2) that may differ from a previous report of C7/10 cells that lacked piRNAs (Skalsky et al. 2010).

With this initial survey of cell cultures and wild-caught versus domesticated lab mosquitoes, our data suggests that somatic piRNAs and siRNAs may be an insect vector response to a persistent arbovirus infection. Our future effort is to profile more

wild mosquito isolates as additions to the MSRG resource. In addition to mosquito field studies, the MSRG resource will enhance future virology and biochemistry of mosquito cell cultures. Lastly, the MSRG resource provides a reference list of curated mosquitoes genic and intergenic piRCLs (Fig. S11C, Tables S2, S4, S5, S6) and reference lists of mosquito arboviruses and transposons with abundant small RNAs from both cell cultures and colonies, which will aid the direction of future functional genomics studies.

MATERIALS AND METHODS

Mosquito strains, cell cultures and virus infections.

The *AnGam* isolate from Imperial College, UK was kept in standard rearing conditions as in (Castellano et al. 2015). The *AeAeg* isolates from Colpitts lab were maintained in the insectary of the National Emerging Infectious Disease Laboratory (NEIDL) as described in (Araujo et al. 2020). The *AeAeg* isolate from the Hughes lab were maintained in the insectary at the University of Texas Medical Branch as described in (Saldana et al. 2017). The *AeAlbo* isolates from the Akbari lab were described in (Gamez et al. 2020). The *CuQuin* isolates were purchased from Benzon Research.

All mosquito cell culture media are described in Table S1, and all cultures were established in the Lau lab for months before cells were used for total RNA extraction and multiple live aliquots were cryopreserved. Cells were all kind gifts: Sua5b and Mos55 cells from the Rasgon lab; C6/36 and Mos55 cells from the Colpitts lab; Aag2 cells from the Blair lab and Colpitts lab, CCL-125 from the Connor lab; C7/10 cells from the Fallon lab; and U4.4 and Hsu cells from the Brackney lab. All cells were maintained in a humidified incubator at 28C with 5% CO2 atmosphere. The DENV and ZIKV infections were performed on Aag2 cells that were ~80% confluent in T25 flasks grown in Shield & Sang Media (Table S1) using viral supernatants from previous C6/36 infections. The infections were conducted under two different multiplicities of infection (MOI=0.1 and 0.01) in the BSL2+ facility in the NEIDL and were cultured for 7 days before cells were neutralized in the TRI-reagent for total RNA extraction. Viral infection status was confirmed by the qRT-PCR assay detailed in (Araujo et al. 2020).

Small RNA library preparation and deep sequencing.

Most small RNA libraries were constructed from small RNAs size fractionated from Urea-Polyacrylamide Gel Electrophoresis as in (Chirn et al. 2015), while only new *Dmel* libraries were

subjected to a process Q-sepharose matrix enrichment of small RNAs (Srivastav et al. 2019). For size-fractionation of small RNAs, 1-5 µg of total RNA from mosquito tissues and ~10µg of total RNA from cell lines was extracted with TRI-reagent. Size fractionation was performed on a urea-denaturing 15% polyacrylamide gel with TBE buffer and 18-nt and 32-nt fluorescent oligos were used as markers. 18-32nt sized RNA portion of gel was excised under UV and eluted in 500µL 0.3M NaCl overnight with mild agitation at RT. Small RNA containing eluate was saved and supplemented with 2 volumes of ethanol and 1µL of 20mg/mL glycogen for precipitation at -20°C overnight. Small RNAs were precipitated by centrifuging at 15,000rpm at 4°C for 20 mins. Small RNA containing glycogen pellet was next washed with chilled 75% ethanol and eluted in 12µL of freshly made 50% (w/v) PEG-8000 to enhance 3' end ligation efficiency. 6µL of the small RNAs in PEG-8000 were used for library construction using NEBNext Small RNA Library Construction kit (E7330S) as per manufacturer's protocol.

All small RNA libraries were purified with the Monarch PCR & DNA Cleanup Kit (5 µg), quantified using Qubit 2.0 and analyzed on Agilent Bioanalyzer 2100 before sequencing on the BUSM Microarray and Sequencing Resource. For total RNA from *Drosophila* OSS and WRR1 cells and AnGam Sua5b and Mos55 cells, we subjected this to beta-eliminition treatment as in (Song et al. 2014).

RT-PCR analysis of *AnGam* densovirus in Mos55 cells.

Total RNA was extracted from Mos55 cells by TRI-reagent RT, and 10 µg RNA was subjected to DNase I and RNase A digestion for 30 minutes at 37 ˚C, heat-inactivated at 65 ˚C, and then subjected to standard phenol-chloroform:IAA extraction and isopropanol precipitation. First strand cDNA synthesis was performed using 1.0 µg untreated RNA, 0.78 µg DNase I-treated RNA, and 0.25 µg RNaseA-treated RNA using the NEB Random Primer Mix and Protoscript. PCR was performed on 1 uL of Mos55 cDNA in 50 µL reactions using the specified

Amp1, Amp2, and *AnGam* Rps7 primer pairs with Phusion DNA Polymerase. Amp1 primers: TACAAGAACAAGGCAGTTCCAGC; CCAATAAGTTATCCAATATTAGTG. Amp2 primers: TGGACTTATATCAAATTCCTATATGG; ACGGGGATCCCGGACTAATGTTGGC. *AnGam* Rps7 primers: GGTGCACCTGGATAAGAACCA; CGGCCAGTCAGCTTCTTGTAC.

Reducing redundancy in transposon family consensus sequences lists.

Since most mosquito transposon annotations were derived automatically with bioinformatic prediction scripts such as the RepeatModeler package that consists of RepeatMasker, RepeatScout/TEFam, RECON and TRF program tools (Bao and Eddy 2002; Price et al. 2005; Gelfand et al. 2007; Wheeler et al. 2013), the heuristic issue is that its efficient process generates lists of transposon families that are very redundant. Therefore, we developed different strategies for each species to mitigate over-counting of small RNAs that are elaborated in the Supplementary document and Table S1.

From these consolidated lists, we applied the RepeatMasker program (Wheeler et al. 2013) to identify the genome copy numbers and genome coverages for each transposon from four organism, and then applied small RNA counts for the benchmarking results in Fig. S1. Different merging methods were required to accommodate the different genome sizes and transposable element (TE) type compositions amongst the mosquito species. We treated manually curated Repbase entries as the prime standard keeping as a representative TE family consensus sequence, which was only extensive for *AnGam* and enabled quick merging just with BLAT. However, in *CuQuin*, *AeAeg* and *AeAlbo*, Repbase entries were very few while all other prediction entries were numerous, so for *CuQuin* and *AeAeg* we used the more specific MeShClust program to cluster TE entries and pick centroid entries we kept as representative of the merged TE family consensus sequences at the 55% similarity cutoff. But in *AeAlbo*, a nearly doubling of the number of TE species predictions, primarily from a huge expansion of

LTR elements, repeatedly caused the MeShClust program to crash. Therefore, we used the less-specific CD-HIT program, also at 55% similarity cutoff and additional repeat lengths and small RNA mapping cutoffs to reduce the redundancy in the list of *AeAlbo* TE family consensus sequences.

Bioinformatics analysis of small RNA datasets.

For these mosquito species, we adapted our bioinformatics analysis pipelines for analyzing genic/intergenic small RNA counts and analyzing transposons/virus counts (Chirn et al. 2015). Our original pipeline consisted of a series of shell, Perl and C scripts coupled with various short read mapping packages like Bowtie as well as BLAST and BLAT (Altschul et al. 1990; Kent 2002; Langmead et al. 2009; Langmead and Salzberg 2012). Together, the pipeline determines read length distributions, assigns reads to defined lists of miRNAs and structural RNAs such as transfer and ribosomal RNAs; then maps remaining reads to the genome with annotation overlays that allow for binning and counting of reads mapping to genes and predicted gene models, transposon consensus sequences, and intergenic regions.

We first indexed the genome assembly file by running BWA version 1 (Li and Durbin 2010) and formatdb from NCBI. Within the genic/intergenic small RNA pipeline, small RNA reads were first trimmed by Cutadapt program (Didion et al. 2017) to remove the adaptor sequences in the 3' end. Trimmed reads were then mapped to a collection of virus sequences using Bowtie with 2 mismatches (Langmead et al. 2009). Reads which were mapped to the virus were removed. Next, reads were mapped to miRNAs and structure RNAs, e.g. snRNAs, tRNAs, rRNAs, snoRNAs using Bowtie with 2 mismatches. Reads which were mapped to miRNAs and structure RNAs were removed. Finally, reads were mapped to genomes using Bowtie with 2 mismatches to get the genic/intergenic counts using the genome GTF file. Genic counts were further categorized into 5'UTR counts, CDS counts, 3'UTR counts.

The fixed step Wig file was generated by recording the normalized read counts within every window of 25 bases for positive strand and negative strand respectively. The wigToBigWig program was used to covert the fixed step wig file to the bigWig file which was loaded to the Broad Institute Integrative Genomics Viewer (IGV(Robinson et al. 2011)) together with the genome assembly and GTF files. Reads mapped to the intergenic regions were progressively clustered together if normalized read counts is over 0.02 within a sliding window of 25 base. To reduce the redundancy in the genic table caused by different isoforms of a gene, the mergeBed program (Quinlan 2014) was used to consolidate different isoforms by providing the genomic location of each isoform. The isoform with the highest read counts was chosen as the representative of the gene.

Within transposons/virus sRNA pipeline, reads were first trimmed by Cutadapt program to remove the adaptor sequences in the 3'' end. Then trimmed reads were mapped to miRNAs with BLAST (Altschul et al. 1990). Reads which were mapped to miRNAs were removed. Then reads were mapped to transposons using Bowtie with 2 mismatches and virus using Bowtie with 1 mismatch. Finally, the mapping patterns with respect to transposons/viruses were plotted with an R script. Hierarchical clustering was performed by calling Python Seaborn Clustermap function using Euclidean distance and average linkage clustering method. Principal Component Analysis (PCA) was carried out by R prcomp function, with plots generated by the ggplot function. Methods for curating genic and intergenic piRNA Cluster Loci (piRCL) and predicting the piRNA targets are elaborated in the Supplemental Materials document.

piRNA Ping-pong and Phasing analysis

Reads were first trimmed by Cutadapt program to remove the adaptor sequences in the 3' end. Then, trimmed reads longer than 23 nucleotides were aligned to the genome using Bowtie with no mismatch. The genomic location and the number of times of mapped reads

were recorded. Using this information, we carried out autocorrelation analysis to identify periodic peaks based on a previous script from (Gainetdinov et al. 2018). For 3' to 5' phasing analysis, autocorrelation analysis of 3' to 5' distance on the same genomic strands were carried out and Z score at distance 0 was calculated, and a significant Z score over 2 was observed in most cases. For 5' to 5' phasing analysis, autocorrelation analysis of 5' to 5' distance on the same genomic strands were carried out and periodic peaks were observed on the autocorrelation scores. For piRNA ping-pong analysis, autocorrelation analysis of 5' to 5' distance on the opposite genomic strands were carried out and Z score at distance 10 was calculated, noting Z scores over 2 as significant. The siRNA duplex analysis was similar except that Z score at distance 21 was calculated.

DATA ACCESS

All new deep-sequencing data from this study was submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE146545 and Study SRP251875. Additional curated outputs and source file details can be found at https://laulab.bu.edu/msrg/ and computational scripts at https://github.com/laulabbumc/MosquitoSmallRNA and as Supplemental Code.

ACKNOWLEDGEMENTS

We thank Ildar Gainetdinov and Phil Zamore for assistance with the piRNA phasing code, Jullien Flynn for analyzing the *AeAlbo* genome with RepeatModeler v2; Dianne Schwarz and Mohsan Saeed for comments on the manuscript. SPS, SG, FF, TC, EIP, GH, TN, AG, and EM conducted the experiments; RJ, JR, JHC and DEB provided mosquito cultures, QM developed and executed the bulk of the bioinformatics analyses,

GD provided follow bioinformatics assistance and SG also provided extensive help with analysis and curation. NCL conceived and conducted experiments and wrote the paper with comments from all authors. All mosquito images in Figure 1 were public domain images from the NIH and CDC government websites. This work was supported by NIH grants R01-AG052465, R21-HD088792 and R01-GM135215 to NCL; NIH grants R01-AI128201, R01-AI116636, and R01-AI150251 to JLR; Defense Advanced Research Project Agency (DARPA) Safe Genes Program Grant (HR0011-17-2- 0047) and an NIH New Innovator Award (1DP2AI152071-01) to OSA; NEIDL Pilot funds and NIH grant R21AI129881 to TMC; NIH grants R21NS101151 and R21AI121933-01 to JHC; the LSTM Director's Catalyst Fund award to EIP; NIH grants R21AI129507 and R21AI138074), the BBSRC (BB/T001240/1), and a Royal Society Wolfson Fellowship (RSWF\R1\180013) to GLH; grants from BBSRC Network Grant "ANTI-VeC" (AV/PP0020/1) and the Bill & Melinda Gates Foundation to TN. DB was also supported by the Cooperative Agreement Number U01CK000509, funded by the Centers for Disease Control and Prevention.

REFERENCES

Afanasiev BN, Kozlov YV, Carlson JO, Beaty BJ. 1994. Densovirus of Aedes aegypti as an expression vector in mosquito cells. *Exp Parasitol* **79**: 322-339.

Afanasiev BN, Ward TW, Beaty BJ, Carlson JO. 1999. Transduction of Aedes aegypti mosquitoes with vectors derived from Aedes densovirus. *Virology* **257**: 62-72.

Aguiar E, de Almeida JPP, Queiroz LR, Oliveira LS, Olmo RP, de Faria I, Imler JL, Gruber A, Matthews BJ, Marques JT. 2020. A single unidirectional piRNA cluster similar to the flamenco locus is the major source of EVE-derived transcription and small RNAs in Aedes aegypti mosquitoes. *Rna* **26**: 581-594.

Aguiar ER, Olmo RP, Paro S, Ferreira FV, de Faria IJ, Todjro YM, Lobo FP, Kroon EG, Meignin C, Gatherer D et al. 2015. Sequence-independent characterization of viruses based on the pattern of viral small RNAs produced by the host. *Nucleic acids research* **43**: 6191-6206.

Akbari OS, Antoshechkin I, Amrhein H, Williams B, Diloreto R, Sandler J, Hay BA. 2013. The developmental transcriptome of the mosquito Aedes aegypti, an invasive species and major arbovirus vector. *G3 (Bethesda)* **3**: 1493-1509.

Aliyari R, Ding SW. 2009. RNA-based viral immunity initiated by the Dicer family of host immune receptors. *Immunol Rev* **227**: 176-188.

Aliyari R, Wu Q, Li HW, Wang XH, Li F, Green LD, Han CS, Li WX, Ding SW. 2008. Mechanism of induction and suppression of antiviral immunity directed by virus-derived small RNAs in Drosophila. *Cell Host Microbe* **4**: 387-397.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* **215**: 403-410.

Araujo RV, Feitosa-Suntheimer F, Gold AS, Londono-Renteria B, Colpitts TM. 2020. One-step RT-qPCR assay for ZIKV RNA detection in Aedes aegypti samples: a protocol to study infection and gene expression during ZIKV infection. *Parasit Vectors* **13**: 128.

Arensburger P, Megy K, Waterhouse RM, Abrudan J, Amedeo P, Antelo B, Bartholomay L, Bidwell S, Caler E, Camara F et al. 2010. Sequencing of Culex quinquefasciatus establishes a platform for mosquito comparative genomics. *Science* **330**: 86-88.

Avila-Bonilla RG, Yocupicio-Monroy M, Marchat LA, De Nova-Ocampo MA, Del Angel RM, Salas-Benito JS. 2017. Analysis of the miRNA profile in C6/36 cells persistently infected with dengue virus type 2. *Virus Res* **232**: 139-151.

Bao Z, Eddy SR. 2002. Automated de novo identification of repeat sequence families in sequenced genomes. *Genome research* **12**: 1269-1276.

Bartholomay LC, Waterhouse RM, Mayhew GF, Campbell CL, Michel K, Zou Z, Ramirez JL, Das S, Alvarez K, Arensburger P et al. 2010. Pathogenomics of Culex quinquefasciatus and meta-analysis of infection responses to diverse pathogens. *Science* **330**: 88-90.

Batki J, Schnabl J, Wang J, Handler D, Andreev VI, Stieger CE, Novatchkova M, Lampersberger L, Kauneckaite K, Xie W et al. 2019. The nascent RNA binding complex SFiNX licenses piRNA-guided heterochromatin formation. *Nat Struct Mol Biol* **26**: 720-731.

Blair CD, Olson KE. 2014. Mosquito immune responses to arbovirus infections. *Current opinion in insect science* **3**: 22-29.

-. 2015. The role of RNA interference (RNAi) in arbovirus-vector interactions. *Viruses* **7**: 820-843.

Blair CD, Olson KE, Bonizzoni M. 2020. The Widespread Occurrence and Potential Biological Roles of Endogenous Viral Elements in Insect Genomes. *Curr Issues Mol Biol* **34**: 13-30.

Brackney DE, Scott JC, Sagawa F, Woodward JE, Miller NA, Schilkey FD, Mudge J, Wilusz J, Olson KE, Blair CD et al. 2010. C6/36 Aedes albopictus cells have a dysfunctional antiviral RNA interference response. *PLoS Negl Trop Dis* **4**: e856.

Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* **128**: 1089-1103.

Brustolin M, Pujhari S, Henderson CA, Rasgon JL. 2018. Anopheles mosquitoes may drive invasion and transmission of Mayaro virus across geographically diverse regions. *PLoS Negl Trop Dis* **12**: e0006895.

Burivong P, Pattanakitsakul SN, Thongrungkiat S, Malasit P, Flegel TW. 2004. Markedly reduced severity of Dengue virus infection in mosquito cell cultures persistently infected with Aedes albopictus densovirus (AalDNV). *Virology* **329**: 261-269.

Castellano L, Rizzi E, Krell J, Di Cristina M, Galizi R, Mori A, Tam J, De Bellis G, Stebbing J, Crisanti A et al. 2015. The germline of the malaria mosquito produces abundant miRNAs, endo-siRNAs, piRNAs and 29-nt small RNAs. *BMC genomics* **16**: 100.

Chandler JA, Thongsripong P, Green A, Kittayapong P, Wilcox BA, Schroth GP, Kapan DD, Bennett SN. 2014. Metagenomic shotgun sequencing of a Bunyavirus in wild-caught Aedes aegypti from Thailand informs the evolutionary and genomic history of the Phleboviruses. *Virology* **464-465**: 312-319.

Chen XG, Jiang X, Gu J, Xu M, Wu Y, Deng Y, Zhang C, Bonizzoni M, Dermauw W, Vontas J et al. 2015. Genome sequence of the Asian Tiger mosquito, Aedes albopictus, reveals insights into its biology, genetics, and evolution. *Proceedings of the National Academy of Sciences of the United States of America* **112**: E5907-5915.

Chirn GW, Rahman R, Sytnikova YA, Matts JA, Zeng M, Gerlach D, Yu M, Berger B, Naramura M, Kile BT et al. 2015. Conserved piRNA Expression from a Distinct Set of piRNA Cluster Loci in Eutherian Mammals. *PLoS genetics* **11**: e1005652.

The ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57-74.

Conway MJ, Colpitts TM, Fikrig E. 2014. Role of the Vector in Arbovirus Transmission. *Annu Rev Virol* **1**: 71-88.

Crochu S, Cook S, Attoui H, Charrel RN, De Chesse R, Belhouchet M, Lemasson JJ, de Micco P, de Lamballerie X. 2004. Sequences of flavivirus-related RNA viruses persist in DNA form integrated in the genome of Aedes spp. mosquitoes. *J Gen Virol* **85**: 1971-1980.

Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, Dus M, Perrimon N, Kellis M, Wohlschlegel JA, Sachidanandam R et al. 2008. An endogenous small interfering RNA pathway in Drosophila. *Nature* **453**: 798-802.

Di Giallonardo F, Audsley MD, Shi M, Young PR, McGraw EA, Holmes EC. 2018. Complete genome of Aedes aegypti anphevirus in the Aag2 mosquito cell line. *J Gen Virol* **99**: 832-836.

Didion JP, Martin M, Collins FS. 2017. Atropos: specific, sensitive, and speedy trimming of sequencing reads. *PeerJ* **5**: e3720.

Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F et al. 2012. Landscape of transcription in human cells. *Nature* **489**: 101-108.

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP et al. 2017. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* **356**: 92-95.

Fagegaltier D, Falciatori I, Czech B, Castel S, Perrimon N, Simcox A, Hannon GJ. 2016. Oncogenic transformation of Drosophila somatic cells induces a functional piRNA pathway. *Genes & development* **30**: 1623-1635.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences of the United States of America* **117**: 9451-9457.

Flynt A, Liu N, Martin R, Lai EC. 2009. Dicing of viral replication intermediates during silencing of latent Drosophila viruses. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 5270-5275.

Fredericks AC, Russell TA, Wallace LE, Davidson AD, Fernandez-Sesma A, Maringer K. 2019. Aedes aegypti (Aag2)-derived clonal mosquito cell lines reveal the effects of pre-existing persistent infection with the insect-specific bunyavirus Phasi Charoen-like virus on arbovirus replication. *PLoS Negl Trop Dis* **13**: e0007346.

Gainetdinov I, Colpan C, Arif A, Cecchini K, Zamore PD. 2018. A Single Mechanism of Biogenesis, Initiated and Directed by PIWI Proteins, Explains piRNA Production in Most Animals. *Molecular cell* **71**: 775-790 e775.

Gamez S, Antoshechkin I, Mendez-Sanchez SC, Akbari OS. 2020. The Developmental Transcriptome of Aedes albopictus, a Major Worldwide Human Disease Vector. *G3 (Bethesda)* **10**: 1051-1062.

Gelfand Y, Rodriguez A, Benson G. 2007. TRDB--the Tandem Repeats Database. *Nucleic acids research* **35**: D80-87.

Genzor P, Cordts SC, Bokil NV, Haase AD. 2019. Aberrant expression of select piRNA-pathway genes does not reactivate piRNA silencing in cancer cells. *Proceedings of the National Academy of Sciences of the United States of America* **116**: 11111-11112.

Ghildiyal M, Seitz H, Horwich MD, Li C, Du T, Lee S, Xu J, Kittler EL, Zapp ML, Weng Z et al. 2008. Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. *Science* **320**: 1077-1081.

Giraldez AJ, Mishima Y, Rihel J, Grocock RJ, Van Dongen S, Inoue K, Enright AJ, Schier AF. 2006. Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* **312**: 75-79.

Giraldo-Calderon GI, Emrich SJ, MacCallum RM, Maslen G, Dialynas E, Topalis P, Ho N, Gesing S, VectorBase C, Madey G et al. 2015. VectorBase: an updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases. *Nucleic acids research* **43**: D707-713.

Goic B, Saleh MC. 2012. Living with the enemy: viral persistent infections from a friendly viewpoint. *Curr Opin Microbiol* **15**: 531-537.

Goic B, Stapleford KA, Frangeul L, Doucet AJ, Gausson V, Blanc H, Schemmel-Jofre N, Cristofari G, Lambrechts L, Vignuzzi M et al. 2016. Virus-derived DNA drives mosquito vector tolerance to arboviral infection. *Nature communications* **7**: 12410.

Goic B, Vodovar N, Mondotte JA, Monot C, Frangeul L, Blanc H, Gausson V, Vera-Otarola J, Cristofari G, Saleh MC. 2013. RNA-mediated interference and reverse transcription control the persistence of RNA viruses in the insect model Drosophila. *Nature immunology* **14**: 396-403.

Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, Artieri CG, van Baren MJ, Boley N, Booth BW et al. 2011. The developmental transcriptome of Drosophila melanogaster. *Nature* **471**: 473-479.

Guo Z, Li Y, Ding SW. 2019. Small RNA-based antimicrobial immunity. *Nat Rev Immunol* **19**: 31-44.

Halbach R, Junglen S, van Rij RP. 2017. Mosquito-specific and mosquito-borne viruses: evolution, infection, and host defense. *Current opinion in insect science* **22**: 16-27.

Halbach R, Miesen P, Joosten J, Taşköprü E, Rondeel I, Pennings B, Vogels CBF, Merkling SH, Koenraadt CJ, Lambrechts L et al. 2020. A satellite repeat-derived piRNA controls embryonic development of Aedes. *Nature*.

Han BW, Wang W, Li C, Weng Z, Zamore PD. 2015. Noncoding RNA. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. *Science* **348**: 817-821.

Han YH, Luo YJ, Wu Q, Jovel J, Wang XH, Aliyari R, Han C, Li WX, Ding SW. 2011. RNA-based immunity terminates viral infection in adult Drosophila in the absence of viral suppression of RNA

interference: characterization of viral small interfering RNA populations in wild-type and mutant flies. *J Virol* **85**: 13153-13163.

Hess AM, Prasad AN, Ptitsyn A, Ebel GD, Olson KE, Barbacioru C, Monighetti C, Campbell CL. 2011. Small RNA profiling of Dengue virus-mosquito interactions implicates the PIWI RNA pathway in anti-viral defense. *BMC Microbiol* **11**: 45.

Holt RA Subramanian GM Halpern A Sutton GG Charlab R Nusskern DR Wincker P Clark AG Ribeiro JM Wides R et al. 2002. The genome sequence of the malaria mosquito Anopheles gambiae. *Science* **298**: 129-149.

Horwich MD, Li C, Matranga C, Vagin V, Farley G, Wang P, Zamore PD. 2007. The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Curr Biol* **17**: 1265-1272.

Houe V, Bonizzoni M, Failloux AB. 2019. Endogenous non-retroviral elements in genomes of Aedes mosquitoes and vector competence. *Emerg Microbes Infect* **8**: 542-555.

Izumi N, Shoji K, Suzuki Y, Katsuma S, Tomari Y. 2020. Zucchini consensus motifs determine the mechanism of pre-piRNA production. *Nature* **578**: 311-316.

Kamminga LM, Luteijn MJ, den Broeder MJ, Redl S, Kaaij LJ, Roovers EF, Ladurner P, Berezikov E, Ketting RF. 2010. Hen1 is required for oocyte development and piRNA stability in zebrafish. *The EMBO journal* **29**: 3688-3700.

Kandul NP, Liu J, Sanchez CH, Wu SL, Marshall JM, Akbari OS. 2019. Transforming insect population control with precision guided sterile males with demonstration in flies. *Nature communications* **10**: 84.

Kanthong N, Khemnu N, Pattanakitsakul SN, Malasit P, Flegel TW. 2010. Persistent, triple-virus co-infections in mosquito cells. *BMC Microbiol* **10**: 14.

Kanthong N, Khemnu N, Sriurairatana S, Pattanakitsakul SN, Malasit P, Flegel TW. 2008. Mosquito cells accommodate balanced, persistent co-infections with a densovirus and Dengue virus. *Dev Comp Immunol* **32**: 1063-1075.

Karlikow M, Goic B, Saleh MC. 2014. RNAi and antiviral defense in Drosophila: setting up a systemic immune response. *Dev Comp Immunol* **42**: 85-92.

Katzourakis A, Gifford RJ. 2010. Endogenous viral elements in animal genomes. *PLoS genetics* **6**: e1001191.

Kawamura Y, Saito K, Kin T, Ono Y, Asai K, Sunohara T, Okada TN, Siomi MC, Siomi H. 2008. Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* **453**: 793-797.

Kharchenko PV, Alekseyenko AA, Schwartz YB, Minoda A, Riddle NC, Ernst J, Sabo PJ, Larschan E, Gorchakov AA, Gu T et al. 2011. Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature* **471**: 480-485.

Kim DY, Guzman H, Bueno R, Jr., Dennett JA, Auguste AJ, Carrington CV, Popov VL, Weaver SC, Beasley DW, Tesh RB. 2009. Characterization of Culex Flavivirus (Flaviviridae) strains isolated from mosquitoes in the United States and Trinidad. *Virology* **386**: 154-159.

Kirino Y, Mourelatos Z. 2007. The mouse homolog of HEN1 is a potential methylase for Piwi-interacting RNAs. *Rna* **13**: 1397-1401.

Koh C, Audsley MD, Di Giallonardo F, Kerton EJ, Young PR, Holmes EC, McGraw EA. 2019. Sustained Wolbachia-mediated blocking of dengue virus isolates following serial passage in Aedes aegypti cell culture. *Virus Evol* **5**: vez012.

Kozomara A, Birgaoanu M, Griffiths-Jones S. 2019. miRBase: from microRNA sequences to function. *Nucleic acids research* **47**: D155-D162.

Lambrechts L, Saleh MC. 2019. Manipulating Mosquito Tolerance for Arbovirus Control. *Cell Host Microbe* **26**: 309-313.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10**: R25.

Lau NC, Robine N, Martin R, Chung WJ, Niki Y, Berezikov E, Lai EC. 2009. Abundant primary piRNAs, endo-siRNAs, and microRNAs in a Drosophila ovary cell line. *Genome research* **19**: 1776-1785.

Le Thomas A, Stuwe E, Li S, Du J, Marinov G, Rozhkov N, Chen YC, Luo Y, Sachidanandam R, Toth KF et al. 2014. Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. *Genes & development* **28**: 1667-1680.

Lequime S, Lambrechts L. 2017. Discovery of flavivirus-derived endogenous viral elements in Anopheles mosquito genomes supports the existence of Anopheles-associated insect-specific flaviviruses. *Virus Evol* **3**: vew035.

Lewis SH, Quarles KA, Yang Y, Tanguy M, Frezal L, Smith SA, Sharma PP, Cordaux R, Gilbert C, Giraud I et al. 2018. Pan-arthropod analysis reveals somatic piRNAs as an ancestral defence against transposable elements. *Nat Ecol Evol* **2**: 174-181.

Lewis SH, Salmela H, Obbard DJ. 2016. Duplication and Diversification of Dipteran Argonaute Genes, and the Evolutionary Divergence of Piwi and Aubergine. *Genome Biol Evol* **8**: 507-518.

Li C, Vagin VV, Lee S, Xu J, Ma S, Xi H, Seitz H, Horwich MD, Syrzycka M, Honda BM et al. 2009. Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell* **137**: 509-521.

Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**: 589-595.

Liu P, Dong Y, Gu J, Puthiyakunnon S, Wu Y, Chen XG. 2016. Developmental piRNA profiles of the invasive vector mosquito Aedes albopictus. *Parasit Vectors* **9**: 524.

Londono-Renteria B, Colpitts TM. 2016. A Brief Review of West Nile Virus Biology. *Methods in molecular biology* **1435**: 1-13.

Lund E, Liu M, Hartley RS, Sheets MD, Dahlberg JE. 2009. Deadenylation of maternal mRNAs mediated by miR-427 in Xenopus laevis embryos. *Rna* **15**: 2351-2363.

Malone CD, Brennecke J, Dus M, Stark A, McCombie WR, Sachidanandam R, Hannon GJ. 2009. Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. *Cell* **137**: 522-535.

Maringer K, Yousuf A, Heesom KJ, Fan J, Lee D, Fernandez-Sesma A, Bessant C, Matthews DA, Davidson AD. 2017. Proteomics informed by transcriptomics for characterising active transposable elements and genome annotation in Aedes aegypti. *BMC genomics* **18**: 101.

Matthews BJ, Dudchenko O, Kingan SB, Koren S, Antoshechkin I, Crawford JE, Glassford WJ, Herre M, Redmond SN, Rose NH et al. 2018. Improved reference genome of Aedes aegypti informs arbovirus vector control. *Nature* **563**: 501-507.

Merkling SH, Raquin V, Dabo S, Henrion-Lacritick A, Blanc H, Moltini-Conclois I, Frangeul L, Varet H, Saleh MC, Lambrechts L. 2020. Tudor-SN Promotes Early Replication of Dengue Virus in the Aedes aegypti Midgut. *iScience* **23**: 100870.

Miesen P, Ivens A, Buck AH, van Rij RP. 2016. Small RNA Profiling in Dengue Virus 2-Infected Aedes Mosquito Cells Reveals Viral piRNAs and Novel Host miRNAs. *PLoS Negl Trop Dis* **10**: e0004452.

Mirkovic-Hosle M, Forstemann K. 2014. Transposon defense by endo-siRNAs, piRNAs and somatic pilRNAs in Drosophila: contributions of Loqs-PD and R2D2. *PloS one* **9**: e84994.

Mohn F, Handler D, Brennecke J. 2015. Noncoding RNA. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. *Science* **348**: 812-817.

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**: 1364-1379.

Moon SL, Dodd BJ, Brackney DE, Wilusz CJ, Ebel GD, Wilusz J. 2015. Flavivirus sfRNA suppresses antiviral RNA interference in cultured cells and mosquitoes and directly interacts with the RNAi machinery. *Virology* **485**: 322-329.

Morazzani EM, Wiley MR, Murreddu MG, Adelman ZN, Myles KM. 2012. Production of virus-derived ping-pong-dependent piRNA-like small RNAs in the mosquito soma. *PLoS Pathog* **8**: e1002470.

Myles KM, Morazzani EM, Adelman ZN. 2009. Origins of alphavirus-derived small RNAs in mosquitoes. *RNA biology* **6**: 387-391.

Myles KM, Wiley MR, Morazzani EM, Adelman ZN. 2008. Alphavirus-derived small RNAs modulate pathogenesis in disease vector mosquitoes. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 19938-19943.

Nanfack Minkeu F, Vernick KD. 2018. A Systematic Review of the Natural Virome of Anopheles Mosquitoes. *Viruses* **10**.

Negre N, Brown CD, Ma L, Bristow CA, Miller SW, Wagner U, Kheradpour P, Eaton ML, Loriaux P, Sealfon R et al. 2011. A cis-regulatory map of the Drosophila genome. *Nature* **471**: 527-531.

Nene V, Wortman JR, Lawson D, Haas B, Kodira C, Tu ZJ, Loftus B, Xi Z, Megy K, Grabherr M et al. 2007. Genome sequence of Aedes aegypti, a major arbovirus vector. *Science* **316**: 1718-1723.

Olson KE, Blair CD. 2015. Arbovirus-mosquito interactions: RNAi pathway. *Curr Opin Virol* **15**: 119-126.

Palatini U, Masri RA, Cosme LV, Koren S, Thibaud-Nissen F, Biedler JK, Krsticevic F, Johnston JS, Halbach R, Crawford JE et al. 2020. Improved reference genome of the arboviral vector Aedes albopictus. *Genome biology* **21**: 215.

Palmer WH, Medd NC, Beard PM, Obbard DJ. 2018. Isolation of a natural DNA virus of Drosophila melanogaster, and characterisation of host resistance and immune responses. *PLoS Pathog* **14**: e1007050.

Pandey RR, Homolka D, Chen KM, Sachidanandam R, Fauvarque MO, Pillai RS. 2017. Recruitment of Armitage and Yb to a transcript triggers its phased processing into primary piRNAs in Drosophila ovaries. *PLoS genetics* **13**: e1006956.

Parry R, Asgari S. 2018. Aedes Anphevirus: an Insect-Specific Virus Distributed Worldwide in Aedes aegypti Mosquitoes That Has Complex Interplays with Wolbachia and Dengue Virus Infection in Cells. *J Virol* **92**.

Pickett BE, Sadat EL, Zhang Y, Noronha JM, Squires RB, Hunt V, Liu M, Kumar S, Zaremba S, Gu Z et al. 2012. ViPR: an open bioinformatics database and analysis resource for virology research. *Nucleic acids research* **40**: D593-598.

Post C, Clark JP, Sytnikova YA, Chirn GW, Lau NC. 2014. The capacity of target silencing by Drosophila PIWI and piRNAs. *Rna* **20**: 1977-1986.

Price AL, Jones NC, Pevzner PA. 2005. De novo identification of repeat families in large genomes. *Bioinformatics* **21 Suppl 1**: i351-358.

Quinlan AR. 2014. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics* **47**: 11 12 11-11 12 34.

Rai KS, Black WCt. 1999. Mosquito genomes: structure, organization, and evolution. *Adv Genet* **41**: 1-33.

Reyes-Ruiz JM, Osuna-Ramos JF, Bautista-Carbajal P, Jaworski E, Soto-Acosta R, Cervantes-Salazar M, Angel-Ambrocio AH, Castillo-Munguia JP, Chavez-Munguia B, De Nova-Ocampo M et al. 2019. Mosquito cells persistently infected with dengue virus produce viral particles with host-dependent replication. *Virology* **531**: 1-18.

Robine N, Lau NC, Balla S, Jin Z, Okamura K, Kuramochi-Miyagawa S, Blower MD, Lai EC. 2009. A broadly conserved pathway generates 3'UTR-directed primary piRNAs. *Curr Biol* **19**: 2066-2076.

Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nature biotechnology* **29**: 24-26.

Ruckert C, Prasad AN, Garcia-Luna SM, Robison A, Grubaugh ND, Weger-Lucarelli J, Ebel GD. 2019. Small RNA responses of Culex mosquitoes and cell lines during acute and persistent virus infection. *Insect Biochem Mol Biol* **109**: 13-23.

Saito K, Inagaki S, Mituyama T, Kawamura Y, Ono Y, Sakota E, Kotani H, Asai K, Siomi H, Siomi MC. 2009. A regulatory circuit for piwi by the large Maf gene traffic jam in Drosophila. *Nature* **461**: 1296-1299.

Saito K, Sakaguchi Y, Suzuki T, Siomi H, Siomi MC. 2007. Pimet, the Drosophila homolog of HEN1, mediates 2'-O-methylation of Piwi- interacting RNAs at their 3' ends. *Genes & development* **21**: 1603-1608.

Saldana MA, Etebari K, Hart CE, Widen SG, Wood TG, Thangamani S, Asgari S, Hughes GL. 2017. Zika virus alters the microRNA expression profile and elicits an RNAi response in Aedes aegypti mosquitoes. *PLoS Negl Trop Dis* **11**: e0005760.

Samuel GH, Adelman ZN, Myles KM. 2018. Antiviral Immunity and Virus-Mediated Antagonism in Disease Vector Mosquitoes. *Trends Microbiol*.

Sanchez-Vargas I, Scott JC, Poole-Smith BK, Franz AW, Barbosa-Solomieu V, Wilusz J, Olson KE, Blair CD. 2009. Dengue virus type 2 infections of Aedes aegypti are modulated by the mosquito's RNA interference pathway. *PLoS Pathog* **5**: e1000299.

Scott JC, Brackney DE, Campbell CL, Bondu-Hawkins V, Hjelle B, Ebel GD, Olson KE, Blair CD. 2010. Comparison of dengue virus type 2-specific small RNAs from RNA interference-competent and -incompetent mosquito cells. *PLoS Negl Trop Dis* **4**: e848.

Skalsky RL, Vanlandingham DL, Scholle F, Higgs S, Cullen BR. 2010. Identification of microRNAs expressed in two mosquito vectors, Aedes albopictus and Culex quinquefasciatus. *BMC genomics* **11**: 119.

Song J, Liu J, Schnakenberg SL, Ha H, Xing J, Chen KC. 2014. Variation in piRNA and transposable element content in strains of Drosophila melanogaster. *Genome Biol Evol* **6**: 2786-2798.

Srivastav SP, Rahman R, Ma Q, Pierre J, Bandyopadhyay S, Lau NC. 2019. Har-P, a short P-element variant, weaponizes P-transposase to severely impair Drosophila development. *eLife* **8**.

Sumiyoshi T, Sato K, Yamamoto H, Iwasaki YW, Siomi H, Siomi MC. 2016. Loss of l(3)mbt leads to acquisition of the ping-pong cycle in Drosophila ovarian somatic cells. *Genes & development* **30**: 1617-1622.

Suzuki Y, Frangeul L, Dickson LB, Blanc H, Verdier Y, Vinh J, Lambrechts L, Saleh MC. 2017. Uncovering the Repertoire of Endogenous Flaviviral Elements in Aedes Mosquito Genomes. *J Virol* **91**.

Tassetto M, Kunitomi M, Whitfield ZJ, Dolan PT, Sanchez-Vargas I, Garcia-Knight M, Ribiero I, Chen T, Olson KE, Andino R. 2019. Control of RNA viruses in mosquito cells through the acquisition of vDNA and endogenous viral elements. *eLife* **8**.

Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B et al. 2012. The accessible chromatin landscape of the human genome. *Nature* **489**: 75-82.

Troupin A, Shirley D, Londono-Renteria B, Watson AM, McHale C, Hall A, Hartstone-Rose A, Klimstra WB, Gomez G, Colpitts TM. 2016. A Role for Human Skin Mast Cells in Dengue Virus Infection and Systemic Spread. *J Immunol* **197**: 4382-4391.

Vanlandingham DL, Tsetsarkin K, Klingler KA, Hong C, McElroy KL, Lehane MJ, Higgs S. 2006. Determinants of vector specificity of o'nyong nyong and chikungunya viruses in Anopheles and Aedes mosquitoes. *Am J Trop Med Hyg* **74**: 663-669.

Varjak M, Donald CL, Mottram TJ, Sreenu VB, Merits A, Maringer K, Schnettler E, Kohl A. 2017a. Characterization of the Zika virus induced small RNA response in Aedes aegypti cells. *PLoS Negl Trop Dis* **11**: e0006010.

Varjak M, Maringer K, Watson M, Sreenu VB, Fredericks AC, Pondeville E, Donald CL, Sterk J, Kean J, Vazeille M et al. 2017b. Aedes aegypti Piwi4 Is a Noncanonical PIWI Protein Involved in Antiviral Responses. *mSphere* **2**.

Vodovar N, Goic B, Blanc H, Saleh MC. 2011. In silico reconstruction of viral genomes from small RNAs improves virus-derived small interfering RNA profiling. *J Virol* **85**: 11016-11021.

Vrettos N, Maragkakis M, Alexiou P, Mourelatos Z. 2017. Kc167, a widely used Drosophila cell line, contains an active primary piRNA pathway. *Rna* **23**: 108-118.

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**.

Weger-Lucarelli J, Ruckert C, Grubaugh ND, Misencik MJ, Armstrong PM, Stenglein MD, Ebel GD, Brackney DE. 2018. Adventitious viruses persistently infect three commonly used mosquito cell lines. *Virology* **521**: 175-180.

Wen J, Duan H, Bejarano F, Okamura K, Fabian L, Brill JA, Bortolamiol-Becet D, Martin R, Ruby JG, Lai EC. 2015. Adaptive regulation of testis gene expression and control of male fertility by the Drosophila hairpin RNA pathway. [Corrected]. *Molecular cell* **57**: 165-178.

Wen J, Mohammed J, Bortolamiol-Becet D, Tsai H, Robine N, Westholm JO, Ladewig E, Dai Q, Okamura K, Flynt AS et al. 2014. Diversity of miRNAs, siRNAs, and piRNAs across 25 Drosophila cell lines. *Genome research* **24**: 1236-1250.

Wheeler TJ, Clements J, Eddy SR, Hubley R, Jones TA, Jurka J, Smit AF, Finn RD. 2013. Dfam: a database of repetitive DNA based on profile hidden Markov models. *Nucleic acids research* **41**: D70-82.

Whitfield ZJ, Dolan PT, Kunitomi M, Tassetto M, Seetin MG, Oh S, Heiner C, Paxinos E, Andino R. 2017. The Diversity, Structure, and Function of Heritable Adaptive Immunity Sequences in the Aedes aegypti Genome. *Curr Biol* **27**: 3511-3519 e3517.

Wiegmann BM, Trautwein MD, Winkler IS, Barr NB, Kim JW, Lambkin C, Bertone MA, Cassel BK, Bayless KM, Heimberg AM et al. 2011. Episodic radiations in the fly tree of life. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 5690-5695.

Wu Q, Luo Y, Lu R, Lau N, Lai EC, Li WX, Ding SW. 2010. Virus discovery by deep sequencing and assembly of virus-derived small silencing RNAs. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 1606-1611.

Zakrzewski M, Rasic G, Darbro J, Krause L, Poo YS, Filipovic I, Parry R, Asgari S, Devine G, Suhrbier A. 2018. Mapping the virome in wild-caught Aedes aegypti from Cairns and Bangkok. *Sci Rep* **8**: 4690.

Zhang G, Etebari K, Asgari S. 2016. Wolbachia suppresses cell fusing agent virus in mosquito cells. *J Gen Virol* **97**: 3427-3432.

Zhang Z, Wang J, Schultz N, Zhang F, Parhad SS, Tu S, Vreven T, Zamore PD, Weng Z, Theurkauf WE. 2014. The HP1 Homolog Rhino Anchors a Nuclear Complex that Suppresses piRNA Precursor Splicing. *Cell* **157**: 1353-1363.

FIGURE LEGENDS

Figure 1. Overview of the mosquito small RNA genomics resource.

(A) Phylogenetic tree of Dipteran insects in this study, with evolutionary distance measured by Million Years Ago (MYA). Blue and red color denote the anopheline and culicine lineages. (B) Organization of this resource that compares mosquito cell cultures to tissue types via determining the small RNA types and their genomic profiles. (C) Overview of the four mosquito species genomes and eight cell culture lines subjected to the small RNA genomics analysis pipeline. The specific genome assembly names are noted with genome configuration statistics below. The asterisk by the *AeAlbo* AalbF2 assembly indicates the early-stage assembly annotation has a redundant list of gene models.

Figure 2. Multiple arboviruses persistently infect mosquito cell cultures and generate arboviral small RNAs.

Profiles of viral small RNAs in cell culture lines from *AnGam, CuQuin, AeAeg* and *AeAlbo.* Reads per million (rpm) numbers are totals of the siRNA-length and piRNA-length small RNAs that come from the plus strand in red and minus strand in blue. The X-axis is the coordinates of the virus sequence, the Y-axis is the autoscaled read frequency. The total small RNA normalized counts are below each plot. The suffix to sample names is the initials of the laboratory investigator where the sample was originally obtained (i.e. JR/NL: Jason Rasgon to Nelson Lau, DB: Doug Brackney, JC: John Connor, CB: Carol Blair, TC: Tonya Colpitts, JS: Juan Salas-Benito). The S, M and L segments of the Phasi Charoen-Like virus (PCLV) are marked on these coverage

plots. (A) Various species densoviruses. (B) Phasi Charoen-like virus. (C) Cell Fusing Agent virus. (D) *Drosophila* American Nodavirus and two other cells with PCLV and Merida virus. (E) *Culex* Y virus. (F) Alphaviruses and flaviviruses.

Figure 3. Variation in proportion of somatic piRNAs in mosquito strains correlates with persistent arboviral small RNAs.

(A) Small RNA size distributions from mosquito samples where the somatic piRNA levels are much lower in comparison to the gonads, and these samples lack other arbovirus small RNAs. Colored lines at bottom mark the siRNAs and miRNAs ranging between 19-23nt, while piRNAs are between 24-30nt. The inset charts magnify the distribution of transposon and virus sRNAs under a different Y-axis range, and the red arrow points to low levels of somatic piRNAs. (B) Additional small RNA size distributions (left) of mosquito samples with high levels of somatic piRNAs along with the detection of other persistent arbovirus small RNAs in the pattern plots (right). The X-axis is the coordinates of the virus sequence, the Y-axis is the autoscaled read frequency. The total small RNA normalized counts are below each plot.

Figure 4. Small RNA crosstalk in *Aedes aegypti* (*AeAeg*) during flavivirus infections.

(A) Re-analysis of ZIKV and CFAV small RNAs from *AeAeg* females as sequenced from (Saldana et al. 2017). Blue arrow notes emergence of new piRNAs from CFAV after active replication of ZIKV small RNAs. The X-axis is the coordinates of the virus sequence, the Y-axis is the autoscaled read frequency. The total small RNA normalized

counts are below each plot. (B) Small RNA length distributions as a proportion of the small RNA library. Inset graph zooms in on the modest proportions of viral and transposons small RNAs. Red arrows point to the significant change from the normal proportion of small RNAs in control cells. (C) Counts and small RNA profiles from CFAV in control and infected Aag2-NL cells. Blue arrows point to pre-existing group of negative strand piRNAs potentially because of multiple pre-existing viruses replicating and generating small RNAs in Aag2-NL cells. (D) The regions generating notable piRNAs and siRNAs from CFAV in mosquitoes and Aag2 cells are the NS2A gene and 3'UTR.

Figure 5. Transposons and repeats are targeted by common small RNAs in mosquito cells and tissues.

(A) Profiles of the transposons and repeats with most abundant small RNAs both in cell cultures and mosquito tissues. Positive strand reads are in read, minus strand reads are in blue. The X-axis is the coordinates of the transposon and repeats sequence, the Y-axis is the autoscaled read frequency. The total small RNA normalized counts are below each plot. (B) Comparisons of Dipteran genome sizes, fraction of the genome as repeats, average percentage of the small RNAs targeting mosquito transposons and repeats, and the average ratios of the repeats-targeting small RNAs being antisense or the same sense as the repeats.

Figure 6. Prominent mosquito piRNA cluster loci.

(A) Genome browser snapshots of notably large piRNA Cluster Loci (piRCL) in mosquitoes. Genes and repeats (TEs) tracks are at the bottom of each snapshot. (B) A

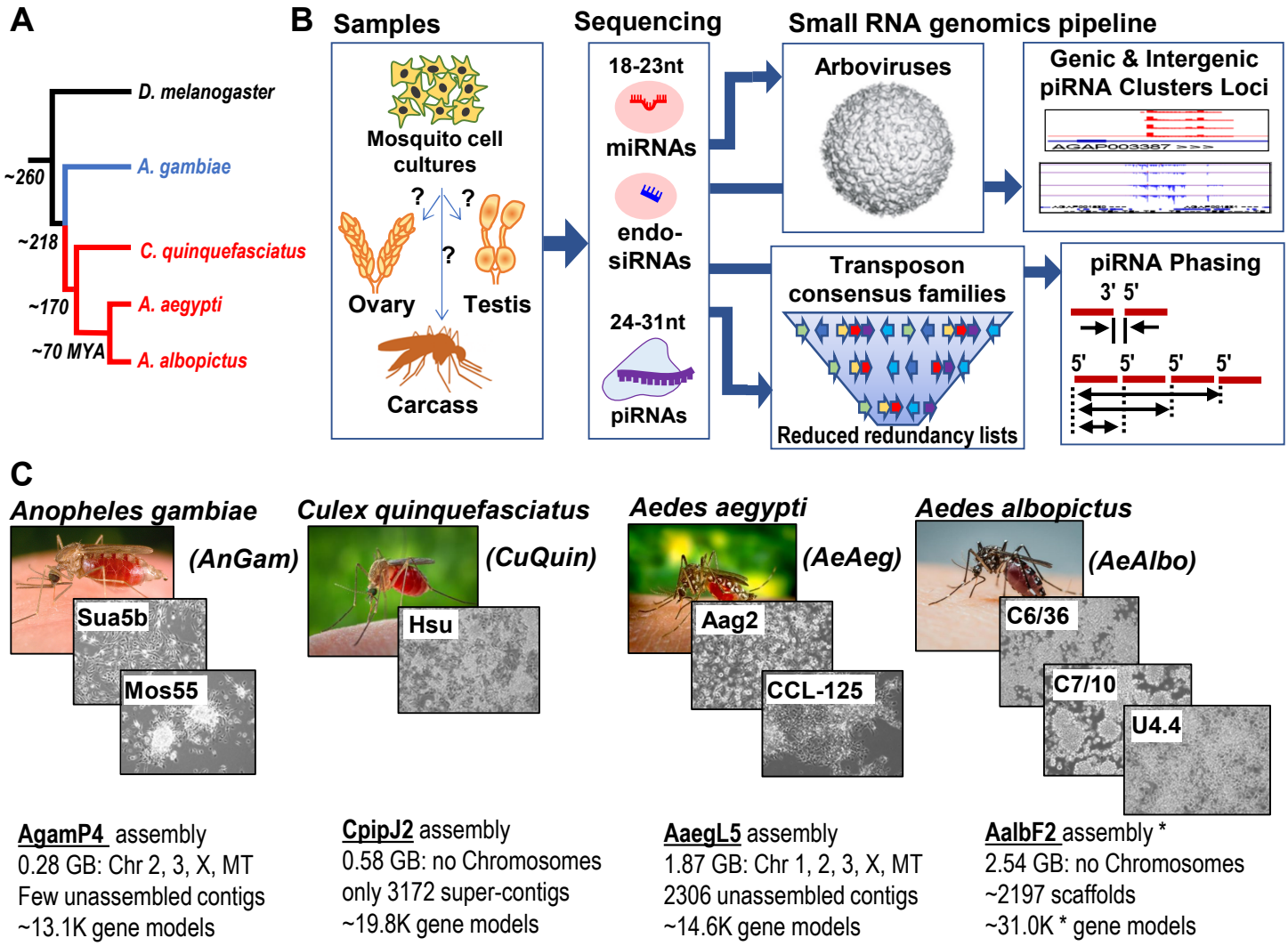
dynamically evolving Mosquito-Conserved piRNA Cluster locus (MCpiRCL) expressed throughout gonads, soma, and cell cultures. (i) Zoomed out genome browser snapshots at the kilobase level of the MCpiRCL. (ii) Zoomed in view of the MCpiRCL from the dashed box in (i). The descriptions of the nearest transcript are listed at the top of the browser window. (iii) Microscopic view of the MCpiRCL from the dashed box in (ii). The peaks are color-coded according to the specific reads as DNA in the sequence below each diagram, derived from the region highlighted by the dashed box above the sequence. Reads per million (rpm) and how many occurrences of the read in the satellite tandem repeats within this MCpiRCL.

Figure 7. Mosquitoes with expanded Piwi-pathway gene numbers display periodic piRNA biogenesis phasing patterns.

(A) Autocorrelation analysis of the 3'-to-5' and 5'-to-5' piRNA phasing patterns from various small RNA libraries in the MSRG. Red arrows mark the periodicity of the 5'-to-5' phasing in samples from independent labs supports a biological process rather than a technical feature in the detection of this periodic pattern. (B) Autocorrelation analysis of the piRNA ping-pong and overlapping siRNA patterns from various small RNA libraries in the MSRG, with Z_{10} and Z_{21} scores >1.0 as denoting a significant ping-pong piRNA or fully-duplexed siRNA signature, respectively. The full gallery of additional pattern diagrams is in Figure S10. X-axis is the base coordinates from the autocorrelation analysis, whereas the Y-axis are arbitrary units that vary for each individual library.

Ma, Srivastav, Gamez et al.

Figure 1.



AgamP4 assembly
0.28 GB: Chr 2, 3, X, MT
Few unassembled contigs
~13.1K gene models

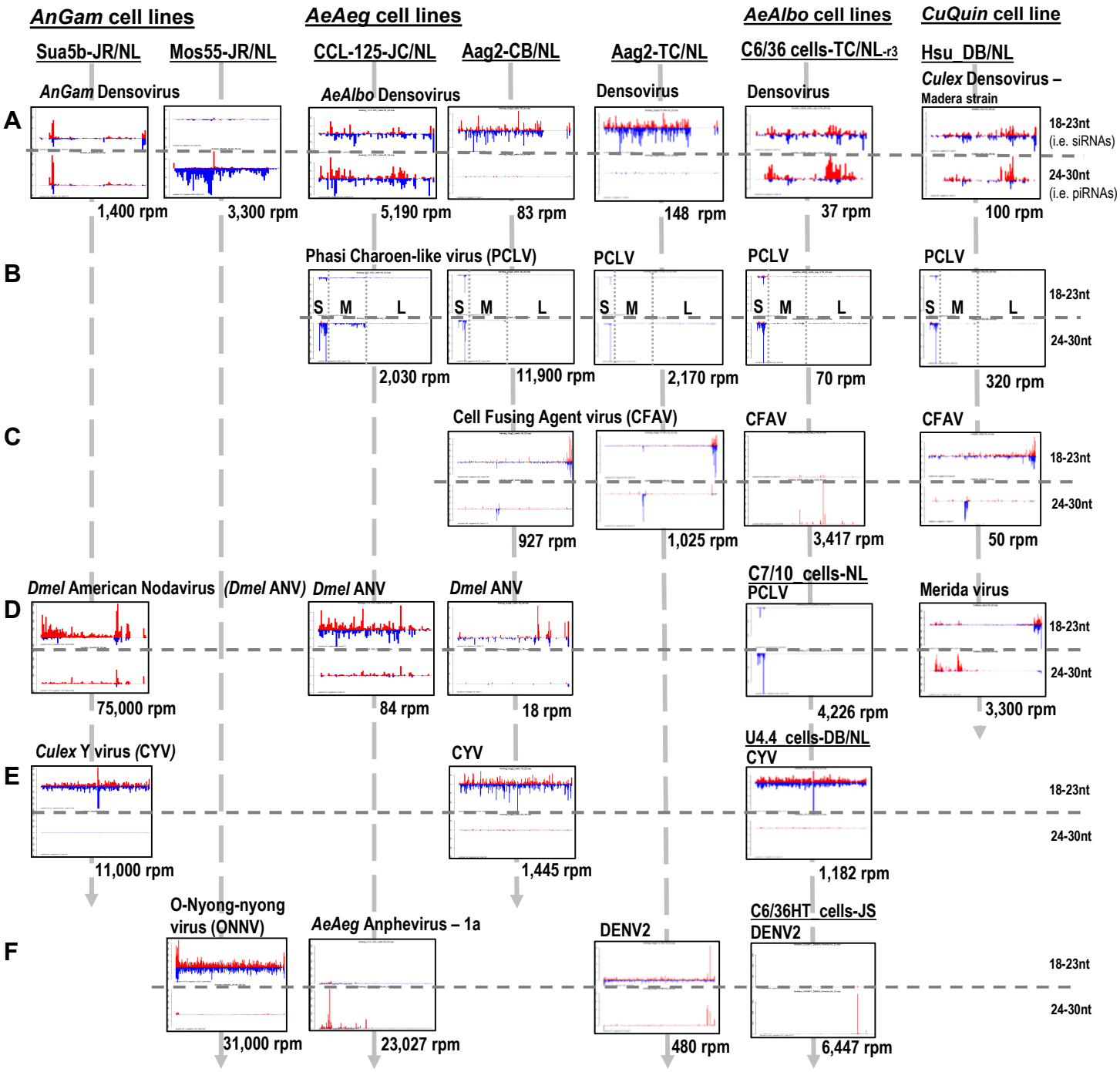
CpipJ2 assembly
0.58 GB: no Chromosomes
only 3172 super-contigs
~19.8K gene models

AaegL5 assembly
1.87 GB: Chr 1, 2, 3, X, MT
2306 unassembled contigs
~14.6K gene models

AalbF2 assembly *
2.54 GB: no Chromosomes
~2197 scaffolds
~31.0K * gene models

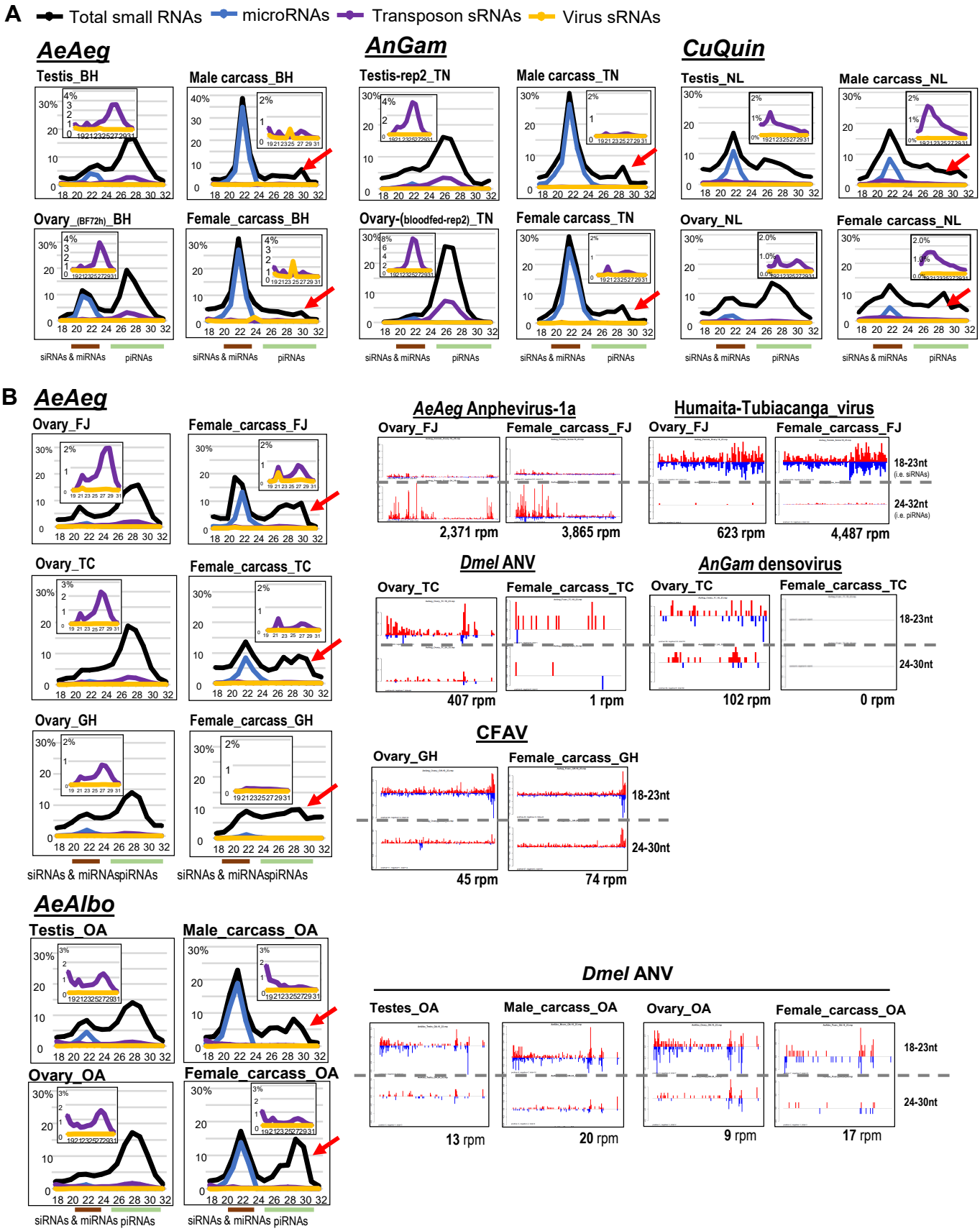
Ma, Srivastav, Gamez et al.

Figure 2.



Ma, Srivastav, Gamez et al.

Figure 3.



Ma, Srivastav, Gamez et al.

Figure 4.

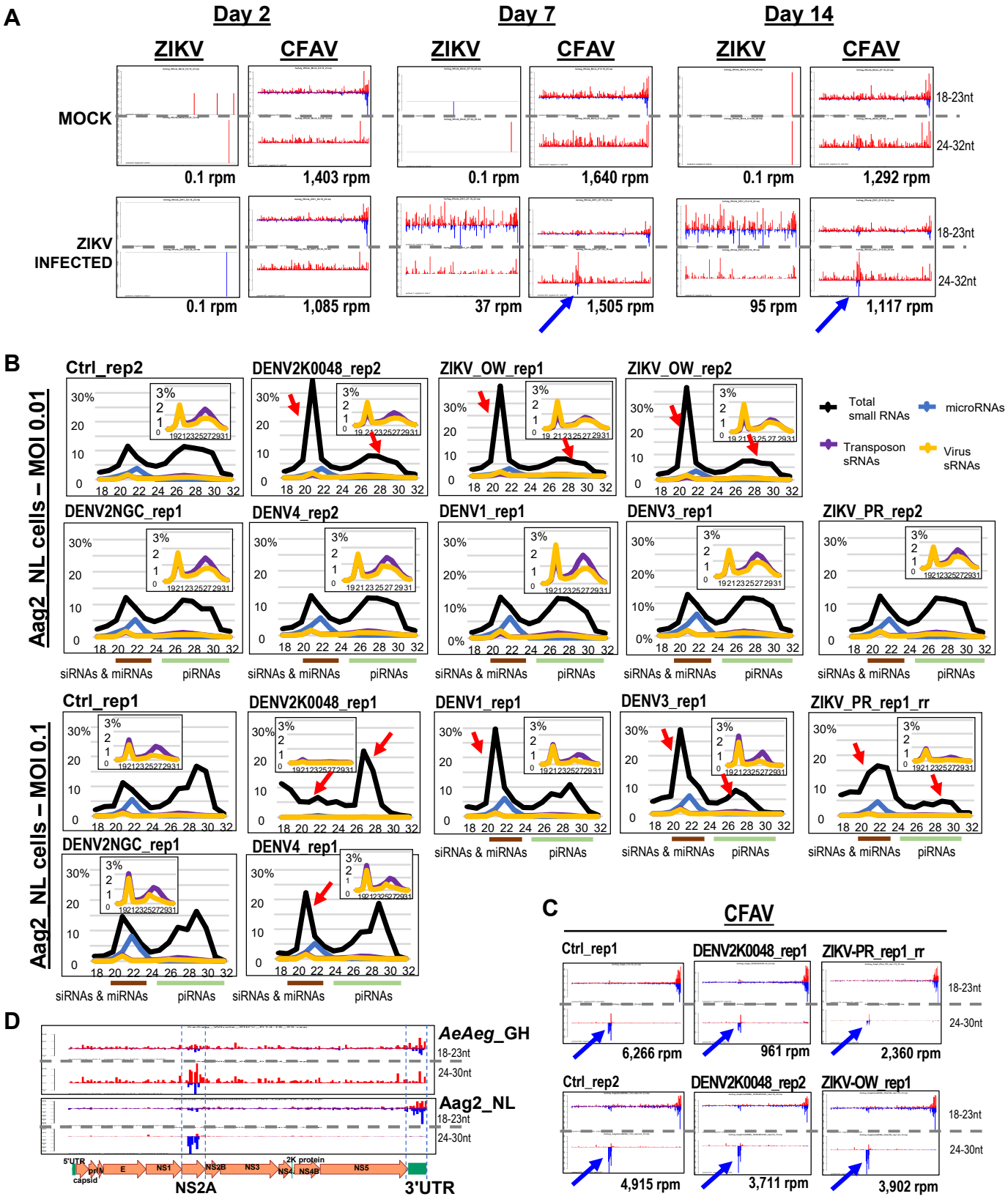
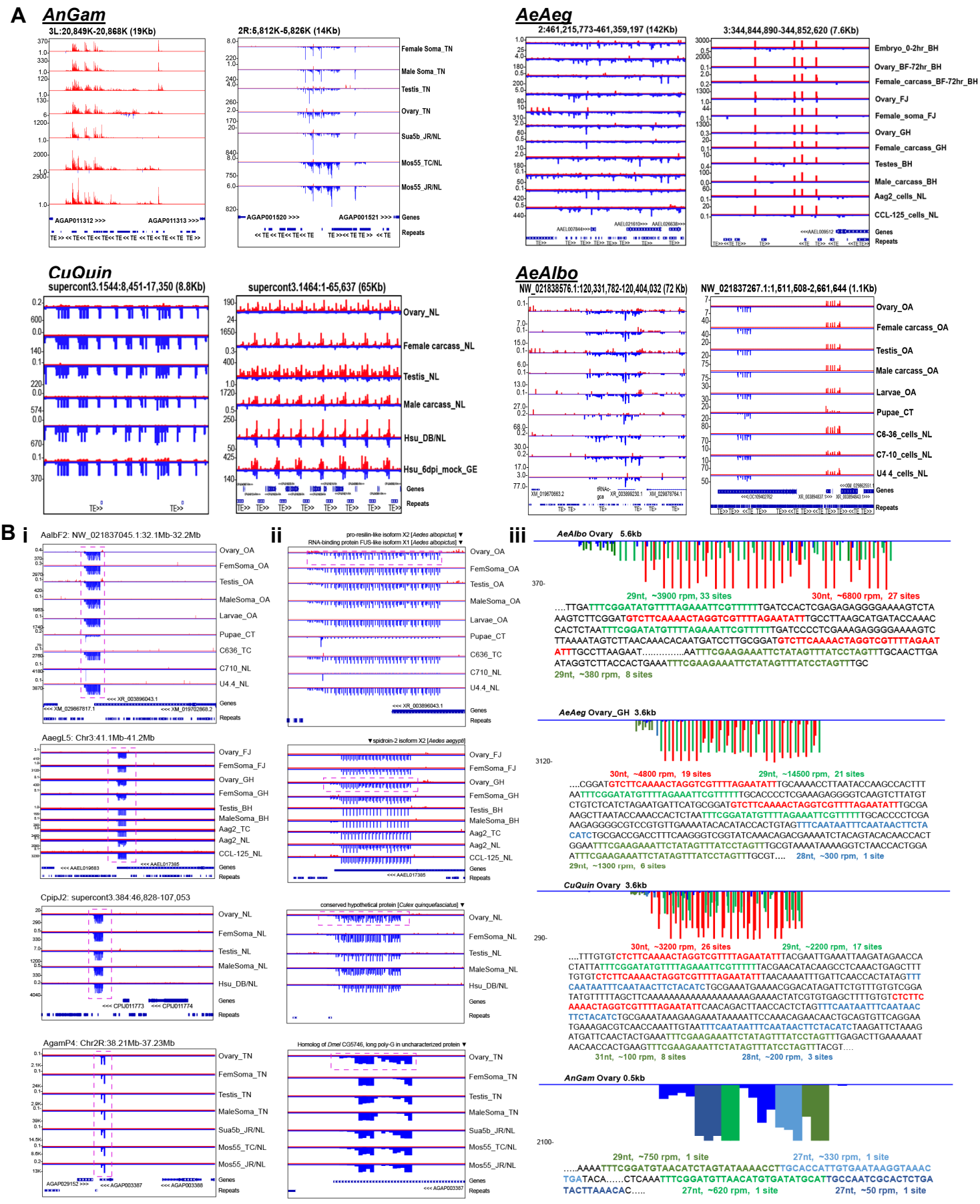


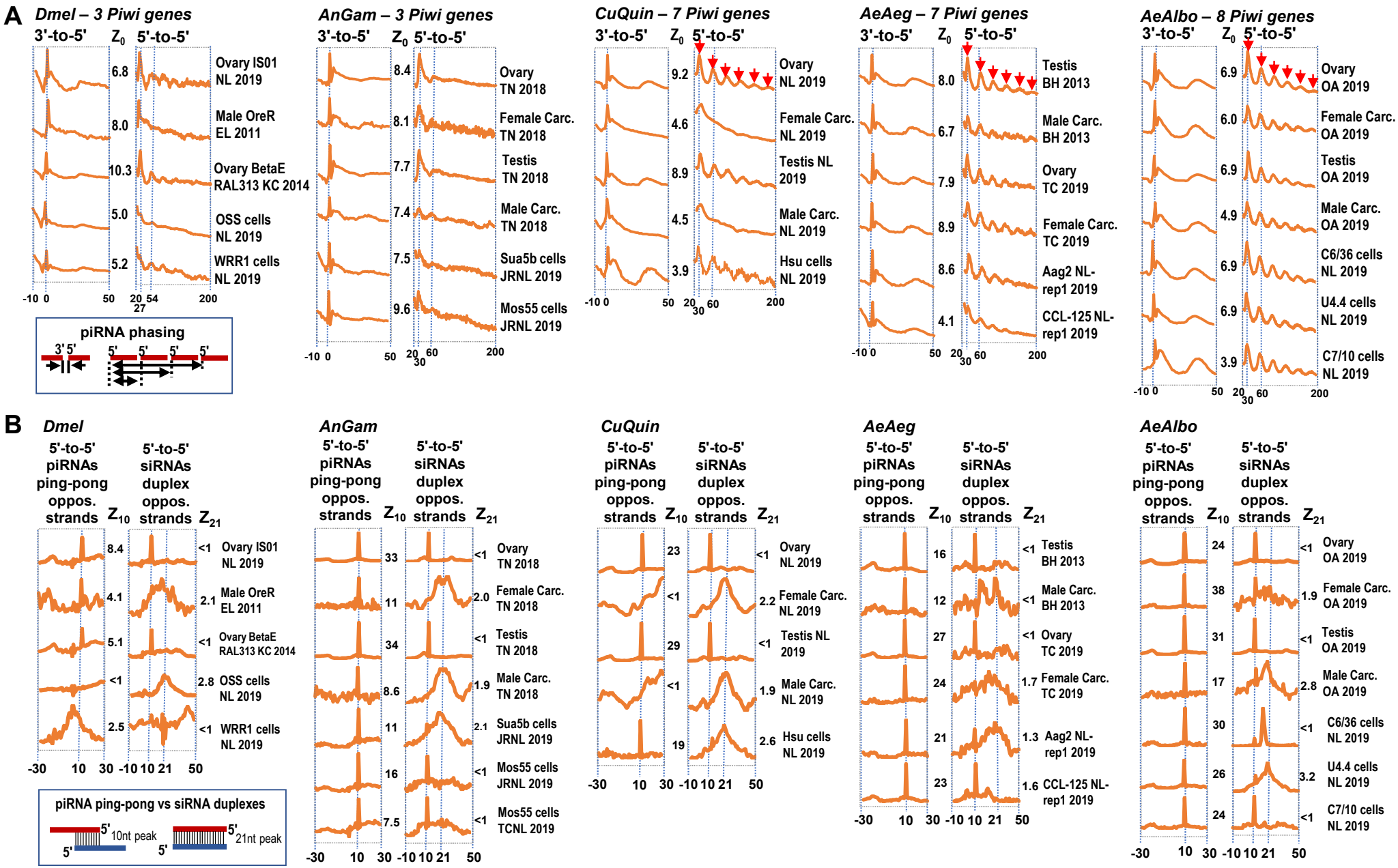
Figure 5.

Ma, Srivastav, Gamez et al.

Figure 6.



Ma, Srivastav, Gamez et al.
Figure 7.



A mosquito small RNA genomics resource reveals dynamic evolution and host responses to viruses and transposons

Qicheng Ma, Satyam P. Srivastav, Stephanie Gamez, et al.

Genome Res. published online January 8, 2021
Access the most recent version at doi:10.1101/gr.265157.120