Radboud Repository

Radboud University Nijmegen

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/175287

Please be advised that this information was generated on 2019-12-04 and may be subject to change.

Untangling the piRNA Pathway in the Arbovirus Vector Mosquito *Aedes aegypti*

Pascal Miesen

» 'What do they live on when they can't get Hobbit?' asked Sam scratching his neck. They spent a miserable day in this lonely and unpleasant country. Their camping-place was damp, cold, and uncomfortable; and the biting insects would not let them sleep. «

J.R.R. Tolkien - The Lord of the Rings

About the cover:	Female *Aedes aegypti* mosquito crawling through 'The One Ring' that symbolizes the ping-pong amplification loop, a central mechanim in the process of piRNA production. The text on the ring says: "In *Aedes aegypti* mosquitoes the ping-pong amplification loop generates PIWI-interacting RNAs from the RNA of arboviruses." Inspired by the review article 'One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing' by B. Czech and G.J. Hannon.
Cover design:	Ellen Wangler & Pascal Miesen
Cover Illustrations:	Images ontained from shutterstock.com, Ring font created by Nancy Lorenz.
Printed by:	Gildeprint, Enschede
ISBN:	978-94-6233-683-4

The reserach described in this thesis was performed at the Department of Medical Microbiology, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen, The Netherlands. This work was financially supported by the Radboud University Medical Center, the Netherlands Organisation for Scientific Research (NWO), the European Research Council (ERC) and the European Molecular Biology Organization (EMBO).

© 2017, Pascal Miesen. All rights reserved. No parts of this publication may be reproduced or transmitted, in any form or by any means, without the permission of the author or, where appropriate, the publisher of the articles.

Untangling the piRNA Pathway in the Arbovirus Vector Mosquito *Aedes aegypti*

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus
prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college van decanen
in het openbaar te verdedigen op vrijdag 15 september 2017
om 10.30 uur precies

door

Pascal Miesen
geboren op 3 juni 1987
te Goch (Duitsland)

Promotor	Prof. dr. Robert W. Sauerwein
Copromotor	Dr. Ronald P. van Rij
Manuscriptcommissie:	Prof. dr. Gert Jan C. Veenstra (voorzitter)
	Prof. dr. Martijn A. Huynen
	Prof. dr. René F. Ketting (Johannes Gutenberg Universiteit Mainz, Duitsland)

Untangling the piRNA Pathway in the Arbovirus Vector Mosquito *Aedes aegypti*

Doctoral Thesis

to obtain the degree of doctor
from Radboud University Nijmegen
on the authority of the Rector Magnificus
prof. dr. J.H.J.M. van Krieken,
according to the decision of the Council of Deans
to be defended in public on Friday, September 15, 2017
at 10.30 hours

by

Pascal Miesen
Born on June 3, 1987
in Goch (Germany)

Supervisor:	Prof. dr. Robert W. Sauerwein
Co-supervisor:	Dr. Ronald P. van Rij
Doctoral Thesis Committee:	Prof. dr. Gert Jan C. Veenstra (chairman)
	Prof. dr. Martijn A. Huynen
	Prof. dr. René F. Ketting (Johannes Gutenberg University Mainz, Germany)

Prelude

Every organism living on this planet, ranging from unicellular bacteria and archaea to multicellular species including plants, fungi, and animals, can be infected by viruses. Viruses are sub-light microscopic infectious agents usually not bigger than a few hundred nanometers. Although variable in their morphology, the principle makeup of viruses usually adheres to a simple blueprint: All viruses have a nucleic acid-based genome surrounded by a protein shell (the capsid) and, in some cases, an additional membranous envelope. The virus genome contains all the information required to produce viral proteins that are needed to copy the genetic material (non-structural proteins) and to build new virus particles (structural proteins). Yet, viruses cannot proliferate autonomously but rely on additional resources that they hijack from infected host cells. The dependency on external means creates two fundamental needs that must be met to allow continuous conveyance of viral genetic information. First, viruses need to enter host cells, reshape the cellular environment, and repurpose available resources to efficiently replicate their own genetic material. Secondly, viruses continuously need to transfer their genetic information to new hosts before their current one succumbs to the potentially harmful infection or dies of other causes.

Host-to-host transmission is especially challenging for viruses that infect the tissues of multicellular organisms that are not directly exposed to the exterior environment. Therefore, viruses have adopted many strategies to cross the barrier from one host to the other. Common transmission routes of human viruses for example include i) respiratory droplets that are expelled from an infected individual and can be inhaled by a naïve host (air-borne transmission), ii) contact with or consumption of contaminated food or water, iii) contact with body fluids of infected individuals such as blood, saliva, semen or mucosal fluids, iv) infection of the unborn offspring with viruses circulating in the mother (vertical transmission), or v) contact with infected invertebrate animals that serve as vectors in the viral life cycle (vector-borne transmission).

Viruses that rely on insect vectors for their transmission are generally known as arthropod-borne viruses or shortly arboviruses. Their life cycle is particularly interesting since these viruses need to replicate in both vertebrate and invertebrate hosts and therefore continuously encounter two entirely different cellular environments. In addition, they are again and again challenged by two distinct types of immune systems. The ability to overcome antiviral defense strategies in both vertebrate hosts and insect vectors is key to successful arbovirus transmission. Currently, however, we know only little about the pathways that shape antiviral immunity in insects. This thesis describes the piRNA pathway as a new component of the immune system of *Aedes aegypti*, an important vector mosquito of many human arboviruses. Personally, I am intrigued by yet another small RNA based pathway that has made it to the battle between viruses and their hosts and I am delighted that I was given the opportunity to help unraveling the molecular mechanisms that underlie its actions.

Pascal Miesen
Nijmegen, March 2017

Table of Contents

Chapter 1	General Introduction and Outline of the Thesis	11
Chapter 2	Distinct Sets of PIWI Proteins Produce Arbovirus and Transposon-Derived piRNAs in *Aedes* Mosquito Cells.	43
Chapter 3	Small RNA Profiling in Dengue Virus 2-Infected *Aedes* Mosquito Cells Reveals Viral piRNAs and Novel Host miRNAs.	77
Chapter 4	Histone-Derived piRNA Biogenesis Depends on the Ping-pong Partners Piwi5 and Ago3 in *Aedes aegypti.*	107
Chapter 5	Comparative Genomics Shows that Viral Integrations are Abundant and Express piRNAs in the Arboviral Vectors *Aedes aegypti* and *Aedes albopictus.*	141
Chapter 6	Post-transcriptional Gene Silencing Mediated by Abundant piRNAs Derived from an Ultra-Conserved Satellite DNA.	169
Chapter 7	The TUDOR Protein AAEL012441 is Required for Efficient Ping-pong Amplification of Viral piRNAs in *Aedes aegypti*.	203
Chapter 8	General Discussion – piRNAs in Virus-Mosquito Interactions	233
Chapter 9	General Discussion – microRNAs in Virus-Host Interactions	259
Appendix	Summary	296
	Nederlandse Samenvatting	300
	Deutsche Zusammenfassung	304
	Acknowledgements, Dankwort	307
	List of Publications and Awards	312
	Curriculum Vitae	314
	PhD Portfolio of the RIMLS Graduate School	315

Chapter 1

General Introduction and Outline of the Thesis

1

INTRODUCTION

The emergence and re-emergence of arthropod-borne viruses (arboviruses) is considered one of the major health concerns of the 21st century (1). Dengue virus, the most prevalent human arbovirus, alone is estimated to cause 100 million infections annually and almost half of the human population worldwide lives at risk of being infected with dengue (2). Outbreaks of chikungunya virus in La Réunion (2005-2006) and the Americas (2013-2014) have caused more than 250,000 and 1 million apparent cases, respectively (3, 4) and the recent 2015-2016 epidemics of Zika virus was estimated to have caused between 440,000 and 1,300,000 infections in Brazil in 2015 alone (5). Recently, arbovirus infections have increased in incidence also in areas outside of tropical and sub-tropical climate, a development that is largely attributed to the invasion of new territories by relevant vector mosquitoes (6). In the last two decades, locally transmitted cases of arboviral diseases have repeatedly been reported in Europe and the United States (7-10).

Currently, specific antivirals or vaccines are not available for most arboviruses including dengue, chikungunya, and Zika and avoiding mosquito bites is considered the most effective strategy to prevent these arboviral diseases. Vector control measures like installation of bed and window nets, use of mosquito repellents, and distribution of insecticides close to human dwellings are currently the most important prevention methods (11). In the future, targeted intervention strategies may include a more direct interference with arbovirus growth in mosquito vectors, thereby limiting its transmission to human hosts. However, we are currently lacking in-depth knowledge about the genetic factors that influence arbovirus replication in mosquitoes.

In this general introduction, I will specifically discuss factors that determine arbovirus transmission. A special emphasis will be given to antiviral immune pathways, in particular small RNA-mediated silencing pathways. I will present in more detail the three major small RNA silencing pathways that exist in animal species: the small interfering RNA, the microRNA, and the PIWI-interacting RNA pathway. I will describe the mechanisms that underlie small RNA biogenesis and summarize the most important functions of the distinct pathways. Special attention will be given to the biogenesis and function of PIWI-interacting RNAs, which will be the major topic of this thesis. For each of the pathways, I will highlight the contributions that they make to the molecular interaction network of arboviruses and their mosquito vectors. Finally, I will provide a short outline of the thesis that summarizes the findings of the work presented.

Arbovirus – mosquito interactions

The term arbovirus refers to the requirement of an insect vector for viral transmission and does not reflect a taxonomical classification. Common insect vectors are ticks, sandflies, and, most importantly, blood-feeding mosquitoes. The most prevalent mosquito-borne

human and livestock arboviruses are members of the *Flaviviridae* (e.g. dengue, yellow fever, Zika, and West Nile virus), *Togaviridae* (e.g. chikungunya and Sindbis virus) and *Bunyaviridae* (e.g. Rift Valley fever and La Crosse encephalitis virus) (12). Transmission of these viruses occurs through biting mosquitoes, almost exclusively from the *Aedes* and *Culex* genera. In contrast, *Anopheles* mosquitoes, which are the major vector for malaria parasites, have hitherto only been reported to transmit a single arbovirus, O'nyong-nyong virus, a member of the *Togaviridae* family (13). Female *Aedes aegypti* (yellow fever mosquito) and/or *Aedes albopictus* (Asian tiger mosquito) mosquitoes are the principal vectors for the most important human arboviruses. Both *Ae. aegypti* and *Ae. albopictus* occur widespread in tropical and subtropical regions of central and Latin America, sub-Saharan Africa, south-east Asia and Oceania. Furthermore, a global rise in average temperatures has allowed their invasion into the southern parts of North America and Europe (6). *Ae. albopictus* in particular is now endemic in more temperate climates largely because of its tolerance towards habitats with cooler ambient temperatures and the ability to produce eggs that can hibernate in a dormancy state (14). In addition, global travel and trade have aided the rapid spread of these important vector mosquitoes worldwide (15). Local transmission of arboviruses is favored by factors that promote a frequent encounter of infected mosquito vectors and vertebrate hosts. High densities of human and domestic animal populations in urban and sub-urban areas facilitate arbovirus transmission (16) and in addition, the availability of uncovered fresh water reservoirs in these neighborhoods serve as attractive breeding grounds for vector mosquitoes (17). Besides the proximity of mosquito habitats to human dwellings, arbovirus transmission is largely affected by the behavior of the vectoring mosquito. Both *Ae. aegypti* and *Ae. albopictus* mosquitoes have an aggressive biting behavior and are active during day time, dusk, and dawn (18, 19). The importance of especially *Ae. aegypti* as arbovirus vector can further be attributed to its strong preference to feed on human (anthropophilia) and to live indoors (20).

Mosquito-borne viruses can be transmitted in an enzootic/sylvatic cycle in which amplification of the virus generally takes place in wild animals (21). Occasionally, blood feeding of infected mosquitoes on humans leads to an infection and the onset of an arboviral disease. Yet, humans are generally dead-end hosts since viral titers in the blood are not sufficiently high to allow infection of a naive mosquito (21, 22). For example, West Nile virus usually circulates between *Culex* mosquitoes and wild birds (23, 24). Occasional infections of human hosts are mostly asymptomatic or lead to a febrile disease (West Nile fever) but can also cause severe encephalitis sometimes with fatal progression (24, 25). However, vector-mediated human-to-human transmission of West Nile virus has not been reported (22). In contrast, major human arboviruses such as dengue, Zika and chikungunya virus are transmitted in an urban epidemic cycle in which anthropophilic *Aedes* mosquitoes transmit the virus directly from an infected individual to a naive human host (21, 22, 26).

A prerequisite for efficient arbovirus transmission by mosquitoes is the presence of high levels of viral particles in the mosquito saliva. To achieve this, arboviruses need to overcome several anatomical and immunological barriers (27). After ingestion in a blood meal, arboviruses need to infect the epithelial cells that line the mosquito midgut and then egress from these at the baso-lateral side. These barriers are referred to as midgut infection and midgut escape barriers. Next, arboviruses disseminate to various secondary tissues outside of the midgut where virus proliferation can occur. Ultimately, the virus needs to infect the salivary glands from which newly produced viral particles are released into the salivary ducts (salivary gland infection and escape barriers) (28). Mosquitoes that efficiently take up viral particles from a blood meal, support virus replication in secondary organs, and horizontally transmit the virus to a naive host are considered *competent* vectors (22, 27). In addition to anatomical barriers, vector competence is shaped by genetic factors and immunological responses to virus infection. For example, a single amino acid substitution in the chikungunya E1 envelope protein was linked to enhanced infectivity in *Ae. albopictus* (29). Whereas normally transmitted by *Ae. aegypti,* enhanced vector competence of *Ae. albopictus* likely caused the onset of the 2005-2006 chikungunya epidemics on La Réunion island where *Ae. albopictus* but not *Ae. aegypti* is endemic (29). Antiviral immune responses impede or prevent the ability of a virus to overcome midgut or salivary gland infection and escape barriers and may therefore reduce the competence of arboviral vector mosquitoes. Signaling pathways such as Toll, IMD and Jak-Stat are triggered by pathogen-associated molecular patterns and induce cellular responses that restrict virus replication or minimize infection-mediated damage (30, 31). Yet, a key role in antiviral immunity in insects has been attributed to small RNA-mediates silencing pathways, which will be discussed in the following section.

Small silencing RNA pathways

The seminal discovery by Andrew Fire and Craig Mello that double-stranded RNA triggers a potent gene silencing response in the nematode *Caenorhabditis elegans* has revolutionized our understanding of post-transcriptional regulation of gene expression (32). Research of the last decades has delineated the underlying mechanism called RNA interference (RNAi) and shown that it acts, in various flavors, in all eukaryotes (33). The key concept of RNAi and related pathways is the association of single stranded small RNAs to a protein of the Argonaute superfamily (34, 35). Based on phylogeny, this protein family can be divided into two major sub-families, the Argonaute (AGO) proteins and the P-element induced wimpy testis (PIWI) proteins (36). To prevent ambiguity between Argonaute superfamily and sub-clade, I will use the capitalized abbreviation 'AGO' specifically for the subfamily and 'Argonaute' for the superfamily including PIWI proteins.

The molecular structure of Argonaute proteins is comprised of, from N to C terminus, an N domain followed by a PAZ domain and a MID and PIWI domain module (Figure 1A) (37-39). The associated small RNA is anchored at its first nucleotide in a pocket formed at the interface of the MID and PIWI domains (40) and at the 3' terminus in the PAZ domain (Figure 1B) (41, 42). Small RNAs are approximately 20 to 30 nt in size and program Argonaute proteins to recognize target RNAs via classical Watson-Crick base pairing. Eventually, the interaction with the Argonaute protein results in silencing of the targeted transcript.

In animals three small RNA-mediated gene silencing pathways exist: the small interfering RNA (siRNA), the microRNA (miRNA) and the PIWI-interacting RNA (piRNA) pathway (33). Whereas all of these adhere to the general concept of small silencing RNA pathways, they diverge in several important details including i) the substrate for production of the small RNAs, ii) the protein complexes involved in small RNA biogenesis and gene silencing, iii) the mode of gene silencing, and iv) the nature of targeted transcripts. Below, these specific aspects of siRNA, miRNA, and piRNA biogenesis and function in insects will be discussed.

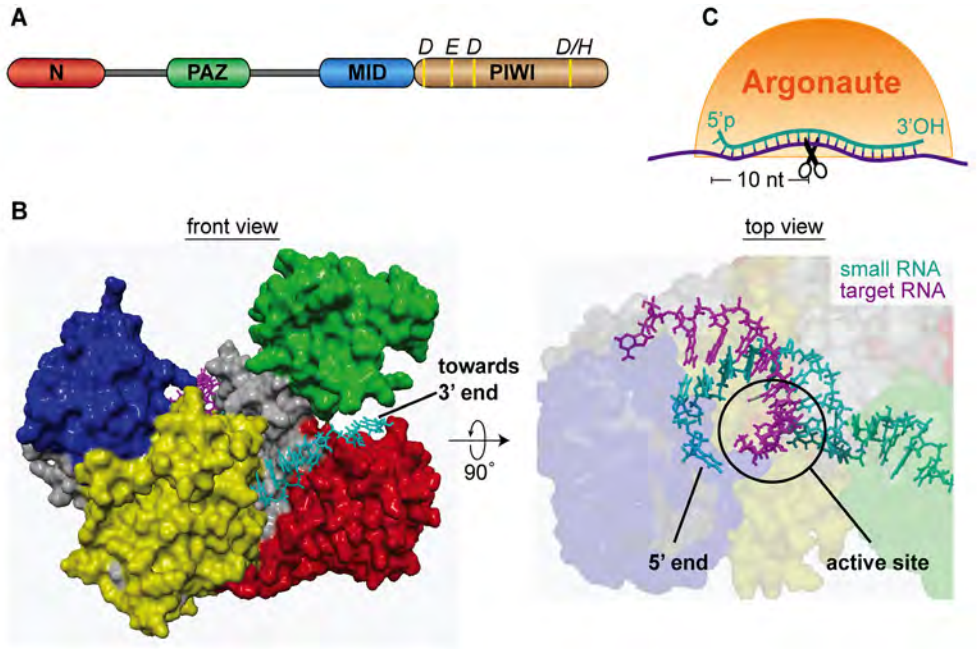


Figure 1. Argonaute proteins are at the heart of small RNA silencing pathways. (A) Schematic representation of the domain organization of eukaryotic Argonaute proteins and the conserved residues required for slicer activity. **(B)** Crystal structure of human Ago2 in association with a guide RNA and a target RNA base pairing from nucleotides 2 to 8. Protein domains are colored in accordance to the scheme in A. The structure was determined by Schirle and colleagues (240) and the published PDB file (4W5Q) was edited in Yasara View (241). **(C)** Schematic representation of target slicing by Argonaute proteins.

The siRNA pathway: biogenesis and functions in insects

The classical RNAi mechanism, uncovered by Fire and Mello (32), is triggered by the presence of double stranded RNA (dsRNA) in the cytoplasm. This initiates a series of processing steps resulting in the production of siRNAs that associate with an Argonaute protein (Figure 2). In *Drosophila melanogaster* (fruit fly), the RNase III enzyme Dicer-2 (Dcr2) recognizes cytoplasmic dsRNA and cleaves it into 21 nucleotides (nt) siRNA duplexes with a characteristic 2 nt overhang at the 3′ ends of both RNA strands (43-46). One of the two strands (the guide strand) is selectively incorporated into the RNA-induced silencing complex (RISC) with at its catalytic core the AGO protein Argonaute-2 (Ago2). The complementary strand (the passenger strand) is degraded in a process that requires Ago2 and the endonuclease Component 3 Promoter of RISC (C3PO) (47-50). Selection of guide and passenger strand is a non-stochastic process and involves the activity of the Dcr2 co-factor R2D2 (51, 52). R2D2 probes the thermodynamic stability of the siRNA duplex and binds the more stable 5′ end eventually defining the passenger strand. Dcr2 selects the opposite strand that will finally be loaded as guide strand into the Ago2-containing RISC complex (53). Dcr2-processing and RISC loading is further promoted by the activity of additional co-factors including the dsRNA binding protein Loquacious isoform PD (Loqs-PD), Arsenic resistance protein2 (Ars2) and heat shock proteins (54-57). These proteins enhance siRNA biogenesis by stabilizing the RNA-protein complexes or facilitating conformational changes during RISC loading. After the guide strand is stably bound by Ago2, it is 2′-O-methylated at the ribose of the 3′terminal nucleotide by the RNA methyl-transferase DmHen1 finalizing the maturation of siRNA-loaded RISC (58).

Ago2-bound siRNAs recognize target RNAs via Watson-Crick base pairing and usually complementarity is required across the entire length of the siRNA/target duplex for efficient target silencing. An exception is the first nucleotide of the siRNA, which is not involved in target recognition as it is locked in a pocket of the Ago2 MID/PIWI domain (40). Upon formation of the siRNA/target RNA duplex, Ago2 cleaves the target RNA between nucleotide ten and eleven counting from the 5′ end of the siRNA (Figure 1C) (44, 45, 47, 59). This small RNA mediated endonuclease activity (slicing) requires the catalytic DEDX (where X is D or H) tetrad present in the PIWI domain of Argonaute proteins (Figure 1A) (60, 61). This motif is conserved amongst slicing-competent Argonaute proteins; nonetheless it is not sufficient for slicing activity since some slicing-incompetent Argonaute proteins still contain the catalytic tetrad (62). After cleavage of the target RNA, the slicing products are generally quickly degraded by cellular ribonucleases (63).

Endogenous sources of dsRNA are either long inverted repeats that fold into perfectly complementary hairpins or transcripts that are derived from convergent transcription. Also gene-pseudogene pairs as well as transposons are potential sources of dsRNA when they express transcripts with full or partial complementarity (Figure 2). These genome-

encoded dsRNA molecules are (endo-siRNA) that have been implicated in transposon control and anecdotally in the regulation of gene expression (64-68). Interestingly, the *Drosophila* Ago2 transcript itself is a prominent source of endo-siRNAs indicating auto-regulation of the core RNAi protein (69). Yet, dsRNA is usually not very abundant in healthy cells and the siRNA pathway has mostly been described as a defense mechanism against foreign dsRNA sources, primarily of viral origin (70). Recognition of dsRNA as danger signal is a remarkable strategy, since the replication of almost all viruses does, at some point, lead to the production of dsRNA (71, 72). Therefore, the antiviral RNAi mechanism is broadly active against a large number of RNA and DNA viruses. The most prominent viral dsRNA sources are i) the genomes of dsRNA viruses, ii) the replication intermediates of positive (+) and negative (-) single stranded RNA viruses, iii) long fold-back structures in viral RNA, and iv) convergent transcripts from the gene-dense genomes of DNA viruses (Figure 2) (73). Akin to endo-siRNA production, Dcr2 recognizes and cleaves these viral dsRNA molecules into 21 nt viral (v)siRNAs, which are then loaded into Ago2-containing RISC complexes (70, 74-76). Interestingly, whereas siRNAs produced from experimentally administered dsRNA rely on both Dcr2 co-factors R2D2 and Loqs-PD (54), viral siRNA biogenesis in *Drosophila* can occur in the absence of Loqs-PD (77). While this differential requirement of Loqs-PD indicates that cells can distinguish the origin of dsRNA molecules, it is currently unclear which molecular mechanisms underlie this phenomenon. Upon loading with vsiRNAs, RISC is programmed to specifically recognize and slice viral RNA present in infected cells (78). Potent antiviral immunity in flies requires a state of systemic antiviral immunity, which is achieved by spread of the RNAi signal to non-infected cells (79).

Insect viruses are not defenseless against the activity of the RNAi pathway; many have developed strategies to antagonize the production or activity of vsiRNAs by expressing viral suppressors of RNAi (VSR) (73, 80, 81). These mostly multifunctional proteins can interfere with the RNAi pathway at various steps for instance by sequestering dsRNA precursors or vsiRNAs to prevent Dcr2 activity or inhibiting Ago2 function through direct interaction with the RISC complex. The fact that viruses, which normally strive to reduce genome size to a minimum, devote genomic space to VSRs only underscores the importance of the RNAi pathway as a potent antiviral mechanism.

The siRNA pathway: implications for arbovirus-mosquito interactions

Similar to *Drosophila*, RNAi acts as a major anti-arboviral immune pathway in mosquitoes (82) and vsiRNA can be readily detected upon infection with arboviruses from different virus families (73). The majority of arboviruses are RNA viruses with a positive (*Flaviviridae, Togaviridae*) or negative single stranded RNA genome (*Bunyaviridae, Rhabdoviridae*) (12). The predominant source of vsiRNAs derived from these viruses is the double-stranded replication intermediate that is invariantly formed when genomic

strands are copied from an antigenomic template and *vice versa.* As a consequence, inactivation of the major RNAi pathway components Dcr2 or Ago2 results in enhanced virus replication of major human arboviruses including dengue (83), chikungunya (84) and yellow fever (85) virus. In addition, viral siRNAs have been sequenced upon infection of mosquitoes and mosquito cells with West-Nile (86), Sindbis (87), O'nyong-nyong (88), Semliki-Forest (89), Rift Valley (90), La Crosse (91) and Bunyamwera virus (92) indicating that RNAi also acts against these arboviruses. In *Anopheles gambiae,* Dicer-2 knockdown and Ago-2 knockdown did not result in higher levels of O'nyong-nyong virus (*Alphavirus* genus, *Togaviridae* family) in the midgut, suggesting that RNAi

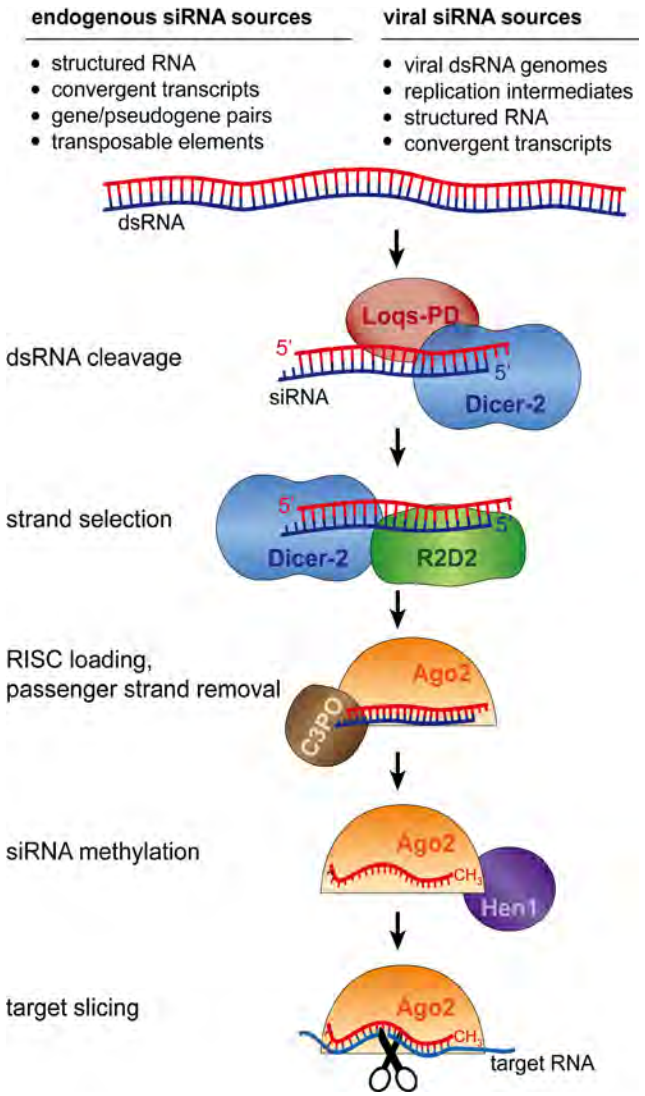


Figure 2. The siRNA pathway in *Drosophila melanogaster.*

acts antiviral mainly after dissemination of the virus to secondary organs (93). In *Ae. aegypti*, however, silencing of RNAi genes did result in elevated levels of Sindbis virus (*Alphavirus* genus, *Togaviridae* family) in the midgut (94, 95). In addition, artificially induced production of dsRNA targeting dengue virus in *Ae. aegypti* mosquitoes is active in midgut epithelial cells (96). Whether these discrepant findings in *An. gambiae* and *Ae. aegypti* are of technical nature or reflect differences in the activity of midgut RNAi in the two species requires additional investigation. Also, it is currently unclear whether initial infection of midgut epithelial cells can trigger a systemic antiviral response and spread of the RNAi signal to secondary tissues like in *Drosophila* (79).

Arboviruses generally cause persistent infections in their mosquito host and do not cause lethality. It has been suggested that targeting by siRNAs contributes to keeping arboviral replication below a pathogenic level. In line with this hypothesis, expression of some VSRs by recombinant Sindbis viruses enhances viral pathogenicity and causes severe mortality in infected *Aedes* mosquitoes (85, 87). Nevertheless, especially for slowly replicating arboviruses interference with RNAi could be less detrimental or even required to allow efficient replication (97). VSRs have been suggested for the NSs proteins of two orthobunya viruses (98, 99), the NS4B protein of dengue (100), and the capsid proteins ofyellow fever and other flaviviruses including Zika, dengue and West Nile virus (85). In addition to these viral proteins acting as VSRs, non-coding RNAs produced from the 3′ untranslated region (UTR) of flaviviruses, called subgenomic flavivirus (sf)RNAs (101), have been attributed to inhibition of RNAi during dengue and West-Nile virus infections (102, 103). However, for most of these putative VSRs the molecular mechanism underlying inhibition of RNAi during an infection in adult mosquitoes awaits further elucidation.

The miRNA pathway: biogenesis and function

miRNAs are an independent class of small RNAs that is present across plant and animal species, protists and even viruses (33). The biogenesis of animal miRNAs (Figure 3) resembles the RNAi pathway in many aspects but there are also clear differences. One of the most striking dissimilarities is the origin of the precursor RNA. miRNAs are processed from genome-encoded hairpins that are transcribed by RNA polymerase II and, far less frequently, by RNA polymerase III (104-106). These so-called pri-miRNA transcripts are located in various genomic contexts including intergenic regions, intronic sequences of protein coding and non-coding transcripts, and, less common, within exons (107-109). They can be transcribed as part of the (non-) coding transcript they reside in or as independent transcription unit (107). Pri-miRNAs are typically a few kilobase (kb) in length (108) and harbor either a single or multiple local stem loop structures that undergo a series of maturation steps to generate an AGO-associated miRNA (107). These stem loops are approximately 90 nt long and consist of two imperfectly base pairing

arms separated by a single-stranded loop region (110). In the nucleus they are released from the pri-miRNA transcript by the Microprocessor complex, which consists of the RNase-III enzyme Drosha and its co-factor Pasha (111-115). Endonucleolytic cleavage by Drosha near the base of the hairpin produces the so-called precursor miRNA (pre-miRNA), a ~70 nt small RNA hairpin with a two-nucleotide overhang at the 3′ end, indicative of RNase-III processing (115). Subsequently, the pre-miRNA is exported from the nucleus via the Ran-GTP dependent nuclear exporter Exportin-5 (116-119).

In the cytoplasm, another RNase-III enzyme, Dicer-1, in complex with the PB isoform of Loqs cleaves off the loop of the pre-miRNA resulting in an RNA duplex with two-nucleotide overhangs at both 3′ ends (43, 120, 121). One of the two strands

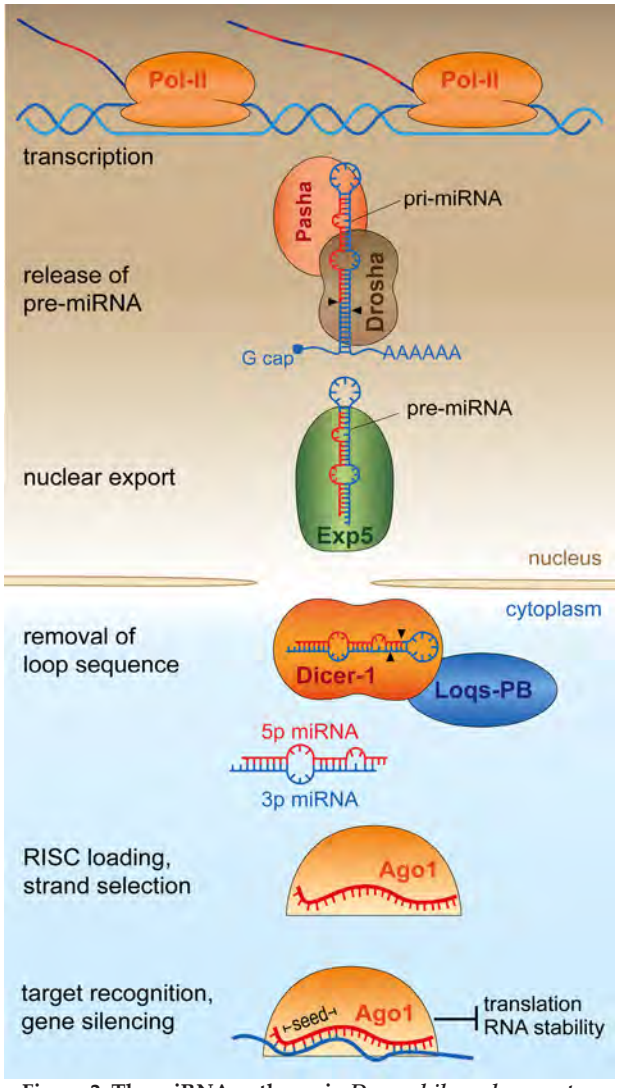


Figure 2. The miRNA pathway in *Drosophila melanogaster*.

is selectively incorporated into the Argonaute-1 containing miRNA induced silencing complex (miRISC) (122, 123). The strand selection is thought to be primarily based on the thermodynamic properties of the duplex; usually the strand with the weaker stability at its 5′ end is incorporated into Ago1 (124, 125). The opposite strand, which in the early miRNA literature was termed miRNA*, is generally not stabilized and quickly degraded. Yet, this dogmatic view on miRNA strand selection has been challenged by small RNA deep-sequencing studies that identified presumable miRNA* sequences associated to Ago1 (126-128). Therefore, miRNAs are currently designated as '5p' or '3p' depending on the arm of the pre-miRNA they are derived from. In some cases the preferential incorporation of either the 5p or 3p miRNA is differentially regulated between different species, distinct cell and tissue types, during development, or as response to infections (129-133). Besides the canonical biogenesis, miRNAs are generated by several alternative pathways that bypass Microprocessor and/or Dicer processing (134-143).

Invariantly, the canonical and alternative pathways result in an miRNA-loaded Ago1 complex. This complex is guided by its associated miRNA to complementary target sites typically located in the 3′ UTR of messenger (m)RNAs (144). In contrastto siRNAs, miRNA targeting does not require base-pairing of the entire small RNA. Instead, partial target recognition by a short nucleotide stretch at the miRNA 5′ end (position 2-8), the so-called seed sequence, accompanied with various degree of base-pairing at the 3′ end is sufficient for gene silencing (144). Whereas siRNAs generally cause slicing of the target RNA, miRNAs exert gene silencing via translational repression, de-adenylation and promotion of enhanced mRNA decay (144-146). The majority of mRNAs is estimated to be regulated by miRNA (147). Therefore, post-transcriptional regulation by miRNAs is implicated in almost all biological processes ranging from cell proliferation, differentiation and apoptosis to regulation of cellular homeostasis and immune responses.

The miRNA pathway: implications for arbovirus-mosquito interactions

Arbovirus infection may interact with the mosquito miRNA machinery in two conceptually distinct ways: i) Arboviruses may encode their own miRNAs that are dependent on canonical or alternative miRNA biogenesis machineries. ii) The host miRNA profile may change as a consequence of arboviral infections, either because viruses actively modulate miRNA levels or because the host induces miRNA expression changes as part of the immune response.

Expression of virus-encoded miRNAs is a common strategy for a large number of viruses with DNA genomes such as herpesviruses (148). In contrast, miRNA expression from RNA viruses has been an issue of debate and to date only few convincing examples have been described. For example, the retrovirus bovine leukemia virus (BVL) encodes a cluster of miRNAs that is transcribed in the nucleus by RNA polymeraseIII from the integrated viral sequence (149). Therefore expression of the miRNAs does not

lead to the destruction of viral genomic RNA. In contrast, miRNA processing from cytoplasmic RNA viruses inevitably consumes viral RNA that will consequentially not be available for replication or translation of viral proteins. In addition, miRNA biogenesis from RNA viruses that do not replicate in the nucleus was doubted, because cytoplasmic RNAs would not have access to the Microprocessor required for release of pre-miRNAs. However, an artificial microRNA is released from recombinant Sindbis virus in mammalian cells by cytoplasmic Microprocessor (150). Interestingly, the core component of the microprocessor, Drosha, was redistributed to the cytoplasm upon Sindbis infection, without affecting cellular miRNA profiles. In addition, artificial introduction of a herpes virus miRNAs into tick-borne encephalitis virus (genus *Flavivirus*, family *Flaviviridae*) did not severely impact virus expression despite efficient production of mature miRNA molecules (151). These studies indicate that miRNA processing from cytoplasmic RNA viruses is principally possible with no or little impairment of virus replication. Infection of insect cells with West Nile virus Kunjin strain results in the production ofa microRNA-like small RNA from the viral 3' UTR, most likely from the subgenomic sfRNA (152). The small RNA binds and up-regulates the expression of host GATA4 transcripts and was proposed to enhance West-Nile virus replication. Similarly, microRNA-like small RNAs were suggested to be generated from the UTR sequences of dengue, but the biological relevance of their production is still an issue of debate (153-155).

Besides encoding their own miRNAs, arboviruses may influence the expression pattern of cellular miRNAs in infected mosquitoes. Host miRNA expression changes have been reported both in *Ae. aegypti* and *Ae. albopictus* mosquitoes after infection with dengue virus (156-158). Similarly, infections of various insect cell lines with major human arboviruses resulted in modulation of cellular miRNA levels (159-162). However, in most of these studies no cellular or viral targets were identified, leaving the question unaddressed if and how differential miRNA expression influences virus replication. Moreover, dengue virus was also reported to alter the processing of miRNA isoforms. Distinct post-transcriptional modifications of mature microRNAs as well as arm-switching events have been observed (132) in adult *Ae. aegypti* as a consequence of dengue infection. Also in this study, the biological relevance of these changes remains to be experimentally established.

The piRNA pathway in insects

piRNAs are the most recently discovered class of animal small RNAs. Their biogenesis mechanism differs from siRNA and miRNA production in three crucial aspects: i) whereas siRNAs and miRNAs are derived from precursors that are, at least partially, double stranded, the piRNA substrate is single stranded RNA. ii) piRNA biogenesis is independent of processing by RNase-III enzymes. iii) Whereas siRNAs and miRNAs associate with AGO proteins, piRNA bind PIWI proteins of the Argonaute superfamily

(163, 164). Given the central role of piRNA biogenesis and function in this thesis, I will discuss this pathway in greater detail below. Since most of our knowledge of this pathway in insects comes from studies in *D. melanogaster* or *Bombyx mori* (silk worm), I will focus this introduction on these two model systems.

The piRNA pathway

The piRNA pathway has initially been discovered as the primary defense mechanism against transposable elements, also shortly known as transposons (165). Transposons are mobile, genetic elements that translocate through the host genome either by a cut-and-paste or by a copy-and-paste mechanism. Mutations as a consequence of random transposon integrations can be detrimental for the integrity of the genome. Moreover, transposition events lead to dsDNA breaks, which result in stalling of cell cycle progression due to the activation of the checkpoint kinase 2 (Chk2) DNA damage response (166). In animal gonads, the piRNA pathway suppresses the activation of transposable elements by posttranscriptional gene silencing and by epigenetic inhibition of transposon transcription (163, 164). As a consequence, loss of piRNA function results in de-repression of transposable elements, Chk2-mediated cell cycle arrest, and defects in germline development (166). Yet, defects in germline development may also be independent from transposon surveillance since certain piRNA pathway mutants cause transposon de-repression without affecting gametogenesis (167). Therefore, how exactly transposon silencing and germ cells maintenance are interconnected is not yet fully understood.

A complex machinery with numerous co-factors is required for the biogenesis of piRNAs. In *Drosophila*, a primary biogenesis pathway produces piRNAs from dedicated genomic loci termed piRNA cluster (168). A secondary pathway, named the ping-pong cycle, specifically amplifies those piRNAs that recognize transcripts of active transposable elements (168, 169).

Definition of piRNA producing loci and piRNA precursor transcription

Mapping of unique piRNA sequences to the *Drosophila* genome identified discrete genomic regions from which most of these small RNAs were derived (168). These piRNA clusters were identified to be rich in remnants of various types of transposable elements (163, 168); hence they were denoted as 'transposon graveyards'. piRNA clusters are often located close to or within heterochromatic regions (170) and although these are typically considered transcriptionally silent, piRNA clusters are actively transcribed (168, 171, 172). Importantly, piRNA clusters come in two flavors: uni-strand clusters and dual-strand clusters (Figure 4). Uni-strand piRNA clusters, such as the *flamenco* locus produce transcripts derived only from one genomic strand. These transcripts largely resemble canonical mRNAs; they are transcribed by RNA polymerase II from loci that contain histone 3 lysine 4 dimethylation (H3K4me2) marks at their promoters (173). In addition,

uni-strand piRNA precursor transcripts are 5′ capped, 3′ poly-adenylated, and spliced (173, 174). In contrast, dual-strand piRNA clusters are transcribed from both genomic strands and lack canonical marks of RNA polymerase II transcription. Their expression depends on the heterochromatin protein 1 (HP1) homolog Rhino that binds to histone 3 lysine 9 trimethylation (H3K9me3) marks (173, 175). Rhino recruits, via the adaptor protein Deadlock, Cutoff (Cuff), which binds to the 5′ end of nascent piRNA precursors and prevents capping and splicing of these transcripts (173, 175). Therefore, Cuff acts as an important marker protein that is thought to specifically license dual-strand piRNA cluster transcripts for piRNA biogenesis. Cuff does not bind to transcripts from uni-strand piRNA clusters and it is currently unknown how these are distinguished from regular mRNAs. Interestingly, introduction of RNA elements from the *flamenco* locus or the piRNA-generating gene *traffic jam* induce *de novo* piRNA biogenesis from an artificial reporter sequence (176, 177). It is, however, unclear whether the primary RNA sequence or secondary structures recruit the piRNA machinery. Nevertheless, these piRNA trigger sequences are strong candidates for the discrimination of uni-strand piRNA cluster transcripts from canonical mRNAs.

Primary piRNA biogenesis

In *Drosophila*, the PIWI proteins Piwi, Aubergine (Aub) and Argoanute 3 (Ago3) are expressed primarily in gonadal cells (164). In the fly ovary, all three proteins are located in the germline compartment; in addition Piwi is expressed in somatic follicle cells surrounding the ovary (171, 178). In these cells, only the primary biogenesis pathway is active and produces piRNAs exclusively from uni-strand clusters such as the *flamenco* locus (Figure 4). The *flamenco* transcript is approximately 180 kb in size and harbors remnants of transposons including the retrotransposons *gypsy*, *idefix* and *ZAM* (171, 172, 178). Most of the transposon insertions are oriented antisense to the direction of transcription . Therefore, piRNAs derived from the *flamenco* transcript are biased to target the mRNA of cognate transposable elements. In germ cells, both uni-strand cluster as well as dual-strand clusters, such as the 42AB locus, give rise to primary piRNAs that are eventually loaded into Piwi and Aub (171). Intriguingly, whereas transposons in dual-stranded clusters are inserted both in sense and antisense orientation, the corresponding primary piRNAs are still largely antisense towards transposon mRNAs (168). The mechanisms that enforce this bias of Piwi and Aub-bound piRNAs are still largely unclear.

In the last years, a number of genetic screens have identified proteins that are essential for primary piRNA biogenesis (179-182), but still the details of this process are largely obscure. The current model proposes that in somatic follicle cells, long piRNA precursor transcripts are exported to the cytoplasm where further piRNA maturation occurs in mitochondria-associated processing sites termed Yb bodies (Figure 4) (180, 183). In these

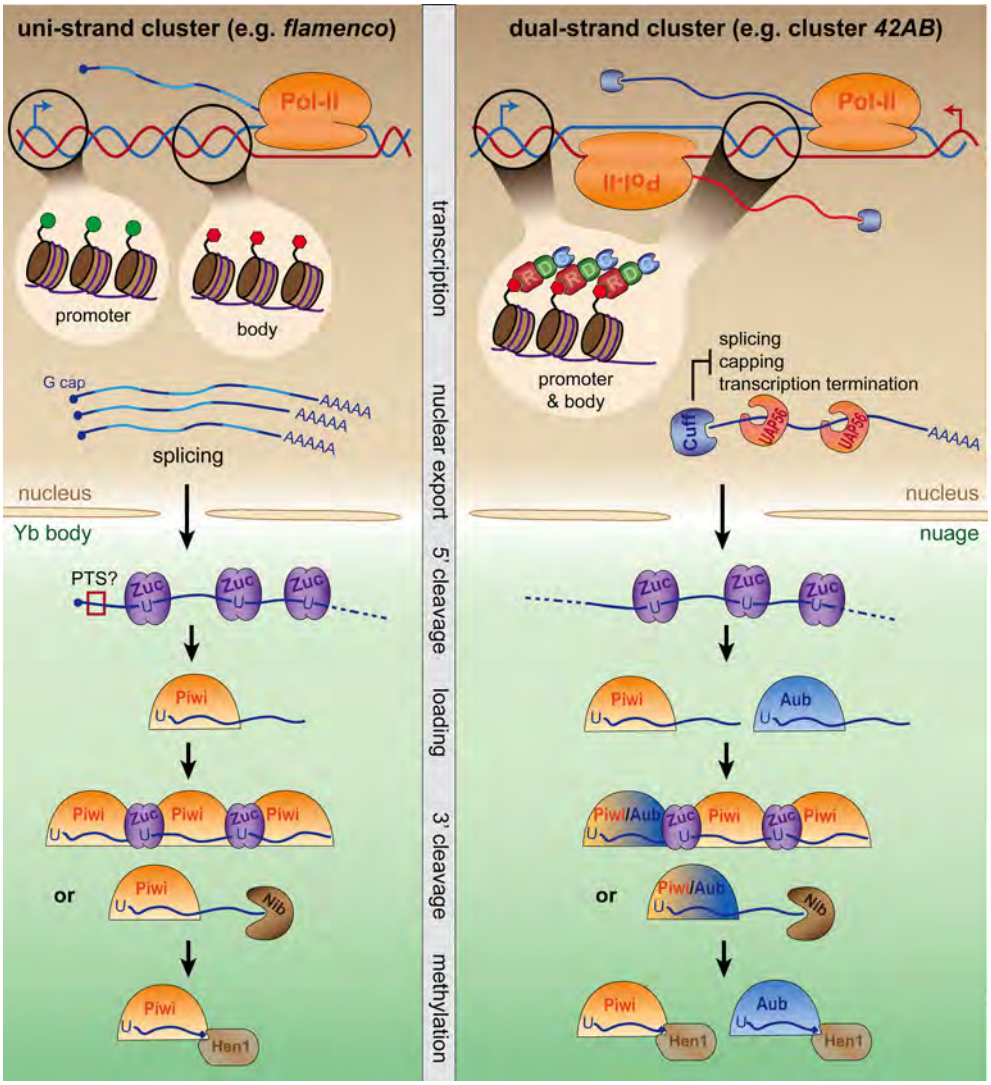


Figure 4. Transcription and primary biogenesis of piRNAs. Schematic illustration of piRNA maturation from uni-strand and dual-strand piRNA clusters in *Drosophila* somatic follicle cells (left) and germline cells (right), respectively. PTS: piRNA trigger sequence, R D C: Rhino-Deadlock-Cutoff complex

non-membranous RNA-protein granules, primary piRNA precursors are cleaved into intermediates by Zucchini (Zuc) (184, 185). In addition, the co-factors Minotaur and GasZ have been implicated in this process (179, 182, 186). The piRNA intermediates are loaded into Piwi and shortened to mature piRNA size either by downstream Zuc cleavage (187-190) or by trimming of the 3′ end by the exonuclease Nibbler (Nib) (191-193). In *Bombyx mori*, the 3′ end of primary piRNAs associated with the PIWI protein Siwi are matured via exonucleolytic trimming by PNLDC1 (38, 194-196). Both in flies and

in silkworms, piRNA maturation is finalized by 3′ end methylation mediated by Hen1 (58, 192, 195-197). Many additional proteins have been identified as important factors for primary piRNA biogenesis in somatic follicle cells, including the Tudor proteins Yb (180, 183, 198) and Vreteno (199, 200), the RNA helicase Armitage (180, 183) and the protein chaperones Heat-shock protein 90 (Hsp90) and its co-chaperones Hsp70/Hsp90 organizing protein (Hop), and Shutdown (201-203). These factors have been proposed to act as molecular scaffolds, to facilitate RNA-protein interactions, or to aid in the loading of PIWI proteins (163, 164).

In *Drosophila* germ cells piRNA biogenesis does not occur in Yb bodies; instead the piRNA machinery is concentrated in a perinuclear, non-membranous structure called 'nuage' (204) (Figure 4). Transfer of piRNA cluster transcripts from the nucleus to the nuage involves the DEAD-box helicase U2AF65-associated protein (UAP56) (175, 205). In the nuage, piRNA maturation depends on similar proteins as in somatic follicle cells. However, some piRNA biogenesis factors are unique to follicle or germline cells, indicating that the molecular mechanisms underlying primary piRNA biogenesis are partly different in these two compartments (164).

Primary piRNAs are largely antisense to transposon sequences and carry a distinctive nucleotide bias – a uridine at the first nucleotide position ('1U bias') (168, 169). This bias is at least partially caused by preferential binding of Piwi/Aub or Siwi to piRNA precursors that begin with a uridine (196, 206). In addition, *Drosophila* Zuc cleaves RNA primarily upstream of uridines, resulting in the production of 1U biased piRNA precursors (187, 188). To which extent binding preferences of PIWI proteins or precursor production by Zuc contribute to the 1U bias of primary piRNAs in different model organisms requires further investigation.

Ping-pong amplification of piRNAs and phased piRNA production

Upon loading with a primary piRNA, Piwi is imported into the nucleus where it exerts transcriptional silencing of transposable elements (207-212) (discussed below). In contrast, Aub resides in the nuage and initiates secondary piRNA biogenesis by the ping-pong loop (168, 169). This sophisticated feed-forward mechanism selectively amplifies from the immense pool of primary piRNAs those that have complementarity to mRNAs of active transposons. Reciprocal cleavage events of transposon sense and antisense RNAs by Aub and Ago3 are required for efficient piRNA amplification (Figure 5A). Aub loaded with a primary piRNA binds and slices complementary transposon mRNAs. As typical for Argonaute proteins, the slicing event occurs between nucleotide ten and eleven counted from the 5′ end of the piRNA (44, 45). Subsequently, the 3′ slicer products become the precursors for secondary piRNAs that are loaded into Ago3 (168, 169). Their 3′ end is defined by Zuc cleavage or Nib trimming (187-190, 193) and methylated by Hen1, akin to primary piRNA biogenesis (58, 192, 197). The mature Ago3 piRNA complex recognizes

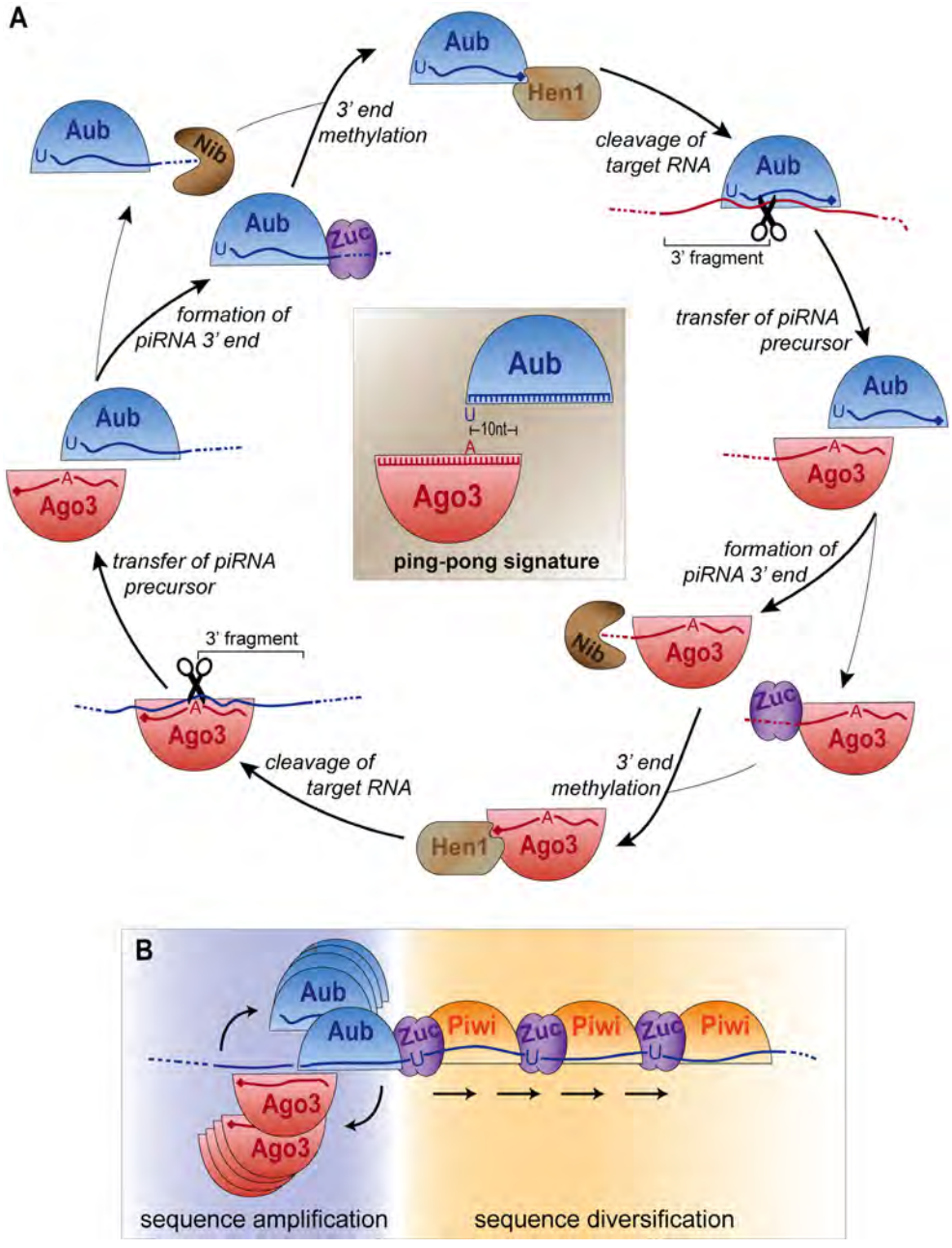


Figure 5. The ping-pong amplification loop and piRNA phasing. (A) Schematic representation of secondary piRNA biogenesis in the *Drosophila* germline. The inner square shows an illustration of the ping-pong signature, hallmark of secondary piRNA production. **(B)** Illustration of piRNA amplification by the ping-pong loop (left) and piRNA phasing by successive Zuc cleavage events (right).

and slices antisense transcripts and this cleavage event will generate the 5′ end of a new piRNA precursor that is identical to the one that initiated the amplification cycle (168, 169). This precursor will be loaded into Aub and matured to complete the ping-pong loop. Ago3-bound piRNAs have a bias for adenine at position 10 ('10A bias'), which is partly due to the 1U bias of Aub-bound piRNAs. In addition, Aub has a preference to bind target RNAs that contain an adenine at the position opposite of the first nucleotide of the associated piRNA thereby enforcing the 10A bias (213). A ten nucleotide overlap between corresponding antisense and sense piRNAs and their distinct sequence biases, 1U and 10A respectively, are the hallmarks of piRNA amplification referred to as the ping-pong signature (Figure 5A).

The piRNA amplification machinery is assisted by a myriad of co-factors that are located in the nuage together with Aub and Ago3. Many of these belong to the class of TUDOR domain containing proteins (214). TUDOR domains interact with PIWI proteins via symmetrical and asymmetrical dimethyl arginines, post translational modifications that are deposited on PIWI proteins by the methyltransferase Capsuleen/PRMT5 (215, 216). Tudor, a protein with eleven TUDOR domains, serves as a molecular scaffold for Aub and Ago3 and is required for efficient ping-pong amplification (216). Qin/Kumo, another Tudor protein, is critical for recruiting Aub and Ago3 to the nuage (217) and loss of function of Qin/Kumo result in homotypic ping-pong amplification of piRNAs mediated by ineffective Aub:Aub interactions (218). In addition, the Tudor proteins Krimper, Tejas, Tapas and Spindle-E (Spn-E) and the DEAD box helicase Vasa are required for ping-pong amplification and defects in these proteins cause failure of secondary piRNA production (194, 204, 219-223).

In the *Drosophila* germline, piRNA amplification by Aub and Ago3 largely defines the population of piRNAs that associates with Piwi by a mechanism called piRNA phasing (187-190). In this process, the definition of 3′ ends of Aub-bound piRNAs by Zuc simultaneously generates the 5′ end of a downstream piRNA precursor that is loaded into Piwi (187, 188). Successive cleavage events mediated by Zuc will generate a trail of Piwi-associated phased piRNAs. Thus, whereas ping-pong amplification by Aub and Ago3 selectively amplifies two specific sequences, phased piRNA results in the diversification of the piRNA repertoire loaded into Piwi (Figure 5B). Qin/Kumo prevents Ago3-bound transcripts from entering piRNA phasing, which enforces the strong antisense bias of Piwi-loaded piRNAs in *Drosophila* germ cells (190). In line with this observation, 3′ ends of Ago3 bound piRNAs are mostly generated by Nib trimming, whereas Aub-bound piRNAs are predominantly matured by Zuc cleavage, potentially triggering piRNA phasing (193). In conclusion, the ping-pong loop selectively amplifies primary piRNAs complementary to active transposon mRNAs in the cytoplasm and it is the major determinant specifying the piRNA repertoire that engages in transcriptional silencing mediated by Piwi.

Transcriptional silencing of transposons

Upon translocation to the nucleus, Piwi induces transcriptional silencing of transposon loci (207-212). Depletion of Piwi enhances transposon transcription as evident by increased RNA polymerase II occupancy and elevated RNA levels. In addition, the heterochromatic H3K9me3 mark is decreased, strongly indicating that Piwi regulates transposon expression by establishing a repressive chromatin environment. A number of recent studies have identified players that are required for deposition of H3K9me3 marks at transposon loci including Asterix/DmGTSF1 (181, 224, 225), Maelstrom (207), and Panoramix/Silencio (226, 227). The coordinated action of these factors is required for the recruitment of the histone methyltransferase eggless and its co-factor Windei (226, 227). Binding of heterochromatin protein 1 (HP1) follows the establishment of H3K9me3 marks and HP1 is essential for transposon repression (211). Yet, H3K9me3 independent silencing of transposon by Piwi has also been reported (209).

Non-transposon-derived piRNAs

Transposable elements are not the exclusive source of piRNAs in the *Drosophila* germline. In fact, the first piRNAs found in *Drosophila* were derived from a repetitive pseudogene locus known as Suppressor of Stellate *Su(Ste)* (228, 229). Su(Ste)-derived piRNAs are essential for silencing of the repetitive Stellate genes in the male germline and failure to establish this piRNA-mediated repression causes male infertility (165). Unbiased deep sequencing of small RNAs in the *Drosophila* germline identified a broad population of genic piRNAs mostly derived from the 3′ UTR (230). Amongst those, *traffic jam* is a prominent source of piRNAs, which have been proposed to regulate the gene expression of Fasciclin (231).

In silkworms, a single genic piRNA lies at the heart of sex-determination (232). *Bombyx mori* females have both a single copy of both a W and a Z chromosome whereas males have two Z chromosomes. Siwi-bound piRNAs are generated from a region called Feminizer (Fem), located on the female specific W chromosome. They target the gene Masculinizer (Masc) located on the Z chromosome. Masc is required for the production of a male-specific splice-isoform of the gene *Bombyx mori* doublesex (Bmdsx), which is a crucial factor for sex determination in silkworm. piRNA-mediated silencing of Masc results in the expression of the female splice isoform of Bmdsx, thus placing genic piRNAs central in the sex determination cascade.

The piRNA pathway – implications for arbovirus-mosquito interactions

Whereas in *Drosophila* siRNAs represent the only class of small RNAs produced from viral RNA, piRNAs with viral sequence (vpiRNAs) accumulate during arbovirus infection of *Aedes* mosquitoes. vpiRNAs were discovered upon infection with dengue (233,

234), chikungunya (235), Sindbis (91, 236) La Crosse (91, 236), Semliki forest (89) and Rift Valley fever virus (90). The presence of a typical ping-pong signature for some of these viruses suggests that a somatic piRNA amplification mechanism exists in mosquitoes. This is in sharp contrast to *Drosophila*, in which the ping-pong cycle is only present in germline cells (168, 169). Indeed, the PIWI gene family is expanded to eight members (Piwi1-7 and Ago3) in *Aedes* mosquitoes and some of these proteins are expressed in so matic tissue (237). Based on phylogeny, the expanded *Aedes* PIWI family can be subdivided into three clades: the Piwi1-4 clade, the Piwi5-7 clade and a clade that only contains Ago3 (238). Besides vpiRNAs, unusual genic piRNAs, not exclusively derived from 3' UTR sequences, are generated in *Ae. aegypti* mosquitoes (239). It is likely, that PIWI gene duplication and expanded expression in somatic tissues has allowed the adaptation to new functions beyond transposon control in gonads. However, which of the *Aedes* PIWI proteins is involved in the biogenesis of piRNAs from different RNA sources (transposons, viruses, protein coding genes, non-coding transcripts) is completely unknown. An intriguing hypothesis is that PIWI protein complexes may have functionally specialized to generate piRNAs from these various substrates.

OUTLINE OF THE THESIS

Small silencing pathways are a crucial component of the interaction network between arboviruses and mosquitoes. The production of vpiRNAs in somatic tissues of *Aedes* mosquitoes provides strong evidence for the gain of additional functions of the piRNA pathway in addition to its canonical function in transposon defense. The aim of this thesis is to shed light on the biogenesis and regulatory functions of virus and host-derived piRNAs in the major arbovirus vector *Ae. aegypti*. In **chapter 2,** the PIWI protein repertoire required for Sindbis virus piRNA production is identified. I show that Piwi5 and Ago3 engage in the ping-pong amplification of vpiRNAs but that Piwi4 and Piwi6 are hardly involved in vpiRNA biogenesis. In contrast, all four PIWI proteins directly or indirectly contribute to the production of canonical transposon-derived piRNAs. This suggests that distinct PIWI protein complexes act in the recognition and/or processing of different sources of piRNA precursors. This idea is further supported by the findings presented in **chapter 3**, in which I show that vpiRNA biogenesis from dengue virus depends on Ago3, Piwi5 and to a lesser extent also on Piwi6. The additional requirement of Piwi6 underscores the functional diversification of PIWI proteins towards distinct classes of precursor RNAs. In this chapter, I also show that in *Ae. aegypti* cells miRNA responses to dengue infection are marginal. Furthermore, new host miRNAs are identified, complementing the repertoire of regulatory RNAs in *Ae. aegypti*. In **chapter 4**, the production of endogenous, non-transposon-derived piRNAs is investigated in greater detail. Numerous protein-coding genes give rise to piRNAs from the coding sequence. These piRNA-producing genes can roughly be divided into two major groups,

one that generates mostly primary piRNAs and one that produces piRNAs via the ping-pong loop in a Piwi5 and Ago3-dependent manner. Amongst the latter, I identify the replication-dependent histones, primarily histone H4 as a dominant source of piRNAs. **Chapter 5** focuses on the production of piRNAs from viral sequences integrated into the host genome, which are widespread in *Ae. aegypti* and *Ae. albopictus* but more scarce in *Culex* and *Anopheles* vector mosquitoes. These elements share sequence homology mostly with insect specific flavi and rhabdoviruses. They are enriched in piRNA clusters and give rise almost exclusively to primary piRNAs. **Chapter 6** describes the biogenesis and regulatory capacity of endogenous piRNA, derived from an ultra-conserved satellite DNA. Individual satellite piRNAs are highly abundant in mosquito cells and *in vivo*, they associate with Piwi4, and have strong targeting potential *in trans*. In **chapter 7**, the Tudor protein AAEL012441 is identified as an important co-factor for the production of secondary Sindbis virus piRNAs. AAEL012441 and the *Ae. aegypti* orthologs of Vasa and Yb assemble in a multi-protein complex around the ping-pong partners Ago3 and Piwi5 and likely facilitate secondary vpiRNA production. **Chapter 8** provides a general discussion on the biogenesis and functions of virus and mosquito-derived piRNAs. Finally, in **chapter 9** I discuss the involvement of miRNAs in virus-host interactions, which seem to play only a minor role during arbovirus infections, but may have profound effects on virus replication in other infection models.

REFERENCES

1. Weaver SC, Reisen WK. Present and future arboviral threats. Antiviral Res. 2010;85(2):328-45.
2. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, Moyes CL, et al. The global distribution and burden of dengue. Nature. 2013;496(7446):504-7.
3. Borgherini G, Poubeau P, Staikowsky F, Lory M, Le Moullec N, Becquart JP, et al. Outbreak of chikungunya on Reunion Island: early clinical and laboratory features in 157 adult patients. Clin Infect Dis. 2007;44(11):1401-7.
4. Johansson MA. Chikungunya on the move. Trends Parasitol. 2015;31(2):43-5.
5. Malone RW, Homan J, Callahan MV, Glasspool-Malone J, Damodaran L, Schneider Ade B, et al. Zika Virus: Medical Countermeasure Development Challenges. PLoS Negl Trop Dis. 2016;10(3):e0004530.
6. Kraemer MU, Sinka ME, Duda KA, Mylne AQ, Shearer FM, Barker CM, et al. The global distribution of the arbovirus vectors Aedes aegypti and Ae. albopictus. Elife. 2015;4:e08347.
7. Rezza G, Nicoletti L, Angelini R, Romi R, Finarelli AC, Panning M, et al. Infection with chikungunya virus in Italy: an outbreak in a temperate region. Lancet. 2007;370(9602):1840-6.
8. La Ruche G, Souares Y, Armengaud A, Peloux-Petiot F, Delaunay P, Despres P, et al. First two autochthonous dengue virus infections in metropolitan France, September 2010. Euro Surveill. 2010;15(39):19676.
9. Gould EA, Gallian P, De Lamballerie X, Charrel RN. First cases of autochthonous dengue fever and chikungunya fever in France: from bad dream to reality! Clin Microbiol Infect. 2010;16(12):1702-4.
10. Murray KO, Rodriguez LF, Herrington E, Kharat V, Vasilakis N, Walker C, et al. Identification of dengue fever cases in Houston, Texas, with evidence of autochthonous transmission between 2003 and 2005. Vector Borne Zoonotic Dis. 2013;13(12):835-45.
11. Staples JE, Fischer M. Chikungunya virus in the Americas--what a vectorborne pathogen can do. N Engl J Med. 2014;371(10):887-9.

12. Gubler DJ. Human arbovirus infections worldwide. Ann N Y Acad Sci. 2001;951:13-24.
13. Keene KM, Foy BD, Sanchez-Vargas I, Beaty BJ, Blair CD, Olson KE. RNA interference acts as a natural antiviral response to O'nyong-nyong virus (Alphavirus; Togaviridae) infection of Anopheles gambiae. Proc Natl Acad Sci U S A. 2004;101(49):17240-5.
14. Nawrocki SJ, Hawley WA. Estimation of the northern limits of distribution of Aedes albopictus in North America. J Am Mosq Control Assoc. 1987;3(2):314-7.
15. Tatem AJ, Hay SI, Rogers DJ. Global traffic and disease vector dispersal. Proc Natl Acad Sci U S A. 2006;103(16):6242-7.
16. Cummings DA, Irizarry RA, Huang NE, Endy TP, Nisalak A, Ungchusak K, et al. Travelling waves in the occurrence of dengue haemorrhagic fever in Thailand. Nature. 2004;427(6972):344-7.
17. Morrison AC, Gray K, Getis A, Astete H, Sihuincha M, Focks D, et al. Temporal and geographic patterns of Aedes aegypti (Diptera: Culicidae) production in Iquitos, Peru. J Med Entomol. 2004;41(6):1123-42.
18. Paupy C, Delatte H, Bagny L, Corbel V, Fontenille D. Aedes albopictus, an arbovirus vector: from the darkness to the light. Microbes Infect. 2009;11(14-15):1177-85.
19. Diallo M, Dia I, Diallo D, Diagne CT, Ba Y, Yactayo S. Perspectives and Challenges in Entomological Risk Assessment and Vector Control of Chikungunya. J Infect Dis. 2016;214(suppl 5):S459-S65.
20. Scott TW, Takken W. Feeding strategies of anthropophilic mosquitoes result in increased risk of pathogen transmission. Trends Parasitol. 2012;28(3):114-21.
21. Weaver SC. Host range, amplification and arboviral disease emergence. Arch Virol Suppl. 2005(19):33-44.
22. Martina BE, Barzon L, Pijlman GP, de la Fuente J, Rizzoli A, Wammes LJ, et al. Human to human transmission of arthropod-borne pathogens. Curr Opin Virol. 2016;22:13-21.
23. Work TH, Hurlbut HS, Taylor RM. Indigenous wild birds of the Nile Delta as potential West Nile virus circulating reservoirs. Am J Trop Med Hyg. 1955;4(5):872-88.
24. Lanciotti RS, Roehrig JT, Deubel V, Smith J, Parker M, Steele K, et al. Origin of the West Nile virus responsible for an outbreak of encephalitis in the northeastern United States. Science. 1999;286(5448):2333-7.
25. Hubalek Z, Halouzka J. West Nile fever--a reemerging mosquito-borne viral disease in Europe. Emerg Infect Dis. 1999;5(5):643-50.
26. Mayer SV, Tesh RB, Vasilakis N. The emergence of arthropod-borne viral diseases: A global prospective on dengue, chikungunya and zika fevers. Acta Trop. 2017;166:155-63.
27. Franz AW, Kantor AM, Passarelli AL, Clem RJ. Tissue Barriers to Arbovirus Infection in Mosquitoes. Viruses. 2015;7(7):3741-67.
28. Hardy JL, Houk EJ, Kramer LD, Reeves WC. Intrinsic factors affecting vector competence of mosquitoes for arboviruses. Annu Rev Entomol. 1983;28:229-62.
29. Tsetsarkin KA, Vanlandingham DL, McGee CE, Higgs S. A single mutation in chikungunya virus affects vector specificity and epidemic potential. PLoS Pathog. 2007;3(12):e201.
30. Merkling SH, van Rij RP. Beyond RNAi: antiviral defense strategies in Drosophila and mosquito. J Insect Physiol. 2013;59(2):159-70.
31. Sim S, Jupatanakul N, Dimopoulos G. Mosquito immunity against arboviruses. Viruses. 2014;6(11):4479-504.
32. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature. 1998;391(6669):806-11.
33. Ghildiyal M, Zamore PD. Small silencing RNAs: an expanding universe. Nat Rev Genet. 2009;10(2):94-108.
34. Schirle NT, Kinberger GA, Murray HF, Lima WF, Prakash TP, MacRae IJ. Structural Analysis of Human Argonaute-2 Bound to a Modified siRNA Guide. J Am Chem Soc. 2016;138(28):8694-7.

35. Elkayam E, Kuhn CD, Tocilj A, Haase AD, Greene EM, Hannon GJ, et al. The structure of human argonaute-2 in complex with miR-20a. Cell. 2012;150(1):100-10.
36. Hock J, Meister G. The Argonaute protein family. Genome Biol. 2008;9(2):210.
37. Song JJ, Smith SK, Hannon GJ, Joshua-Tor L. Crystal structure of Argonaute and its implications for RISC slicer activity. Science. 2004;305(5689):1434-7.
38. Matsumoto N, Nishimasu H, Sakakibara K, Nishida KM, Hirano T, Ishitani R, et al. Crystal Structure of Silkworm PIWI-Clade Argonaute Siwi Bound to piRNA. Cell. 2016;167(2):484-97 e9.
39. Schirle NT, MacRae IJ. The crystal structure of human Argonaute2. Science. 2012;336(6084):1037-40.
40. Ma JB, Yuan YR, Meister G, Pei Y, Tuschl T, Patel DJ. Structural basis for 5'-end-specific recognition of guide RNA by the A. fulgidus Piwi protein. Nature. 2005;434(7033):666-70.
41. Lingel A, Simon B, Izaurralde E, Sattler M. Nucleic acid 3'-end recognition by the Argonaute2 PAZ domain. Nat Struct Mol Biol. 2004;11(6):576-7.
42. Ma JB, Ye K, Patel DJ. Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. Nature. 2004;429(6989):318-22.
43. Bernstein E, Caudy AA, Hammond SM, Hannon GJ. Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature. 2001;409(6818):363-6.
44. Elbashir SM, Lendeckel W, Tuschl T. RNA interference is mediated by 21- and 22-nucleotide RNAs. Genes Dev. 2001;15(2):188-200.
45. Elbashir SM, Martinez J, Patkaniowska A, Lendeckel W, Tuschl T. Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. EMBO J. 2001;20(23):6877-88.
46. Zhang H, Kolb FA, Jaskiewicz L, Westhof E, Filipowicz W. Single processing center models for human Dicer and bacterial RNase III. Cell. 2004;118(1):57-68.
47. Miyoshi K, Tsukumo H, Nagami T, Siomi H, Siomi MC. Slicer function of Drosophila Argonautes and its involvement in RISC formation. Genes Dev. 2005;19(23):2837-48.
48. Matranga C, Tomari Y, Shin C, Bartel DP, Zamore PD. Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. Cell. 2005;123(4):607-20.
49. Rand TA, Petersen S, Du F, Wang X. Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. Cell. 2005;123(4):621-9.
50. Liu Y, Ye X, Jiang F, Liang C, Chen D, Peng J, et al. C3PO, an endoribonuclease that promotes RNAi by facilitating RISC activation. Science. 2009;325(5941):750-3.
51. Liu Q, Rand TA, Kalidas S, Du F, Kim HE, Smith DP, et al. R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway. Science. 2003;301(5641):1921-5.
52. Liu X, Jiang F, Kalidas S, Smith D, Liu Q. Dicer-2 and R2D2 coordinately bind siRNA to promote assembly of the siRISC complexes. RNA. 2006;12(8):1514-20.
53. Tomari Y, Matranga C, Haley B, Martinez N, Zamore PD. A protein sensor for siRNA asymmetry. Science. 2004;306(5700):1377-80.
54. Marques JT, Kim K, Wu PH, Alleyne TM, Jafari N, Carthew RW. Loqs and R2D2 act sequentially in the siRNA pathway in Drosophila. Nat Struct Mol Biol. 2010;17(1):24-30.
55. Sabin LR, Zhou R, Gruber JJ, Lukinova N, Bambina S, Berman A, et al. Ars2 regulates both miRNA- and siRNA- dependent silencing and suppresses RNA virus infection in Drosophila. Cell. 2009;138(2):340-51.
56. Iwasaki S, Kobayashi M, Yoda M, Sakaguchi Y, Katsuma S, Suzuki T, et al. Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. Mol Cell. 2010;39(2):292-9.
57. Miyoshi T, Takeuchi A, Siomi H, Siomi MC. A direct role for Hsp90 in pre-RISC formation in Drosophila. Nat Struct Mol Biol. 2010;17(8):1024-6.
58. Horwich MD, Li C, Matranga C, Vagin V, Farley G, Wang P, et al. The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. Curr Biol. 2007;17(14):1265-72.

59. Rand TA, Ginalski K, Grishin NV, Wang X. Biochemical identification of Argonaute 2 as the sole protein required for RNA-induced silencing complex activity. Proc Natl Acad Sci U S A. 2004;101(40):14385-9.
60. Nakanishi K, Weinberg DE, Bartel DP, Patel DJ. Structure of yeast Argonaute with guide RNA. Nature. 2012;486(7403):368-74.
61. Nowotny M, Gaidamakov SA, Crouch RJ, Yang W. Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. Cell. 2005;121(7):1005-16.
62. Faehnle CR, Elkayam E, Haase AD, Hannon GJ, Joshua-Tor L. The making of a slicer: activation of human Argonaute-1. Cell Rep. 2013;3(6):1901-9.
63. Orban TI, Izaurralde E. Decay of mRNAs targeted by RISC requires XRN1, the Ski complex, and the exosome. RNA. 2005;11(4):459-69.
64. Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, Dus M, et al. An endogenous small interfering RNA pathway in Drosophila. Nature. 2008;453(7196):798-802.
65. Ghildiyal M, Seitz H, Horwich MD, Li C, Du T, Lee S, et al. Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. Science. 2008;320(5879):1077-81.
66. Kawamura Y, Saito K, Kin T, Ono Y, Asai K, Sunohara T, et al. Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. Nature. 2008;453(7196):793-7.
67. Chung WJ, Okamura K, Martin R, Lai EC. Endogenous RNA interference provides a somatic defense against Drosophila transposons. Curr Biol. 2008;18(11):795-802.
68. Okamura K, Chung WJ, Ruby JG, Guo H, Bartel DP, Lai EC. The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs. Nature. 2008;453(7196):803-6.
69. Okamura K, Balla S, Martin R, Liu N, Lai EC. Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in Drosophila melanogaster. Nat Struct Mol Biol. 2008;15(6):581-90.
70. Wang XH, Aliyari R, Li WX, Li HW, Kim K, Carthew R, et al. RNA interference directs innate immunity against viruses in adult Drosophila. Science. 2006;312(5772):452-4.
71. Weber F, Wagner V, Rasmussen SB, Hartmann R, Paludan SR. Double-stranded RNA is produced by positive-strand RNA viruses and DNA viruses but not in detectable amounts by negative-strand RNA viruses. J Virol. 2006;80(10):5059-64.
72. Bronkhorst AW, van Cleef KW, Vodovar N, Ince IA, Blanc H, Vlak JM, et al. The DNA virus Invertebrate iridescent virus 6 is a target of the Drosophila RNAi machinery. Proc Natl Acad Sci U S A. 2012;109(51):E3604-13.
73. Bronkhorst AW, van Rij RP. The long and short of antiviral defense: small RNA-based immunity in insects. Curr Opin Virol. 2014;7:19-28.
74. Deddouche S, Matt N, Budd A, Mueller S, Kemp C, Galiana-Arnoux D, et al. The DExD/H-box helicase Dicer-2 mediates the induction of antiviral activity in drosophila. Nat Immunol. 2008;9(12):1425-32.
75. Galiana-Arnoux D, Dostert C, Schneemann A, Hoffmann JA, Imler JL. Essential function in vivo for Dicer-2 in host defense against RNA viruses in drosophila. Nat Immunol. 2006;7(6):590-7.
76. Sabin LR, Zheng Q, Thekkat P, Yang J, Hannon GJ, Gregory BD, et al. Dicer-2 processes diverse viral RNA species. PLoS One. 2013;8(2):e55458.
77. Marques JT, Wang JP, Wang X, de Oliveira KP, Gao C, Aguiar ER, et al. Functional specialization of the small interfering RNA pathway in response to virus infection. PLoS Pathog. 2013;9(8):e1003579.
78. van Rij RP, Saleh MC, Berry B, Foo C, Houk A, Antoniewski C, et al. The RNA silencing endonuclease Argonaute 2 mediates specific antiviral immunity in Drosophila melanogaster. Genes Dev. 2006;20(21):2985-95.
79. Saleh MC, Tassetto M, van Rij RP, Goic B, Gausson V, Berry B, et al. Antiviral immunity in Drosophila requires systemic RNA interference spread. Nature. 2009;458(7236):346-50.
80. Gammon DB, Mello CC. RNA interference-mediated antiviral defense in insects. Curr Opin Insect Sci. 2015;8:111-20.

81. Nayak A, Tassetto M, Kunitomi M, Andino R. RNA interference-mediated intrinsic antiviral immunity in invertebrates. Curr Top Microbiol Immunol. 2013;371:183-200.
82. Blair CD. Mosquito RNAi is the major innate immune pathway controlling arbovirus infection and transmission. Future Microbiol. 2011;6(3):265-77.
83. Sanchez-Vargas I, Scott JC, Poole-Smith BK, Franz AW, Barbosa-Solomieu V, Wilusz J, et al. Dengue virus type 2 infections of Aedes aegypti are modulated by the mosquito's RNA interference pathway. PLoS Pathog. 2009;5(2):e1000299.
84. McFarlane M, Arias-Goeta C, Martin E, O'Hara Z, Lulla A, Mousson L, et al. Characterization of Aedes aegypti innate-immune pathways that limit Chikungunya virus replication. PLoS Negl Trop Dis. 2014;8(7):e2994.
85. Samuel GH, Wiley MR, Badawi A, Adelman ZN, Myles KM. Yellow fever virus capsid protein is a potent suppressor of RNA silencing that binds double-stranded RNA. Proc Natl Acad Sci U S A. 2016;113(48):13863-8.
86. Brackney DE, Beane JE, Ebel GD. RNAi targeting of West Nile virus in mosquito midguts promotes virus diversification. PLoS Pathog. 2009;5(7):e1000502.
87. Myles KM, Wiley MR, Morazzani EM, Adelman ZN. Alphavirus-derived small RNAs modulate pathogenesis in disease vector mosquitoes. Proc Natl Acad Sci U S A. 2008;105(50):19938-43.
88. Myles KM, Morazzani EM, Adelman ZN. Origins of alphavirus-derived small RNAs in mosquitoes. RNA Biol. 2009;6(4):387-91.
89. Schnettler E, Donald CL, Human S, Watson M, Siu RW, McFarlane M, et al. Knockdown of piRNA pathway proteins results in enhanced Semliki Forest virus production in mosquito cells. J Gen Virol. 2013;94(Pt 7):1680-9.
90. Leger P, Lara E, Jagla B, Sismeiro O, Mansuroglu Z, Coppee JY, et al. Dicer-2- and Piwi-mediated RNA interference in Rift Valley fever virus-infected mosquito cells. J Virol. 2013;87(3):1631-48.
91. Brackney DE, Scott JC, Sagawa F, Woodward JE, Miller NA, Schilkey FD, et al. C6/36 Aedes albopictus cells have a dysfunctional antiviral RNA interference response. PLoS Negl Trop Dis. 2010;4(10):e856.
92. Dietrich I, Shi X, McFarlane M, Watson M, Blomstrom AL, Skelton JK, et al. The Antiviral RNAi Response in Vector and Non-vector Cells against Orthobunyaviruses. PLoS Negl Trop Dis. 2017;11(1):e0005272.
93. Carissimo G, Pondeville E, McFarlane M, Dietrich I, Mitri C, Bischoff E, et al. Antiviral immunity of Anopheles gambiae is highly compartmentalized, with distinct roles for RNA interference and gut microbiota. Proc Natl Acad Sci U S A. 2015;112(2):E176-85.
94. Campbell CL, Keene KM, Brackney DE, Olson KE, Blair CD, Wilusz J, et al. Aedes aegypti uses RNA interference in defense against Sindbis virus infection. BMC Microbiol. 2008;8:47.
95. Khoo CC, Piper J, Sanchez-Vargas I, Olson KE, Franz AW. The RNA interference pathway affects midgut infection- and escape barriers for Sindbis virus in Aedes aegypti. BMC Microbiol. 2010;10:130.
96. Franz AW, Sanchez-Vargas I, Adelman ZN, Blair CD, Beaty BJ, James AA, et al. Engineering RNA interference-based resistance to dengue virus type 2 in genetically modified Aedes aegypti. Proc Natl Acad Sci U S A. 2006;103(11):4198-203.
97. O'Neal ST, Samuel GH, Adelman ZN, Myles KM. Mosquito-borne viruses and suppressors of invertebrate antiviral RNA silencing. Viruses. 2014;6(11):4314-31.
98. Szemiel AM, Failloux AB, Elliott RM. Role of Bunyamwera Orthobunyavirus NSs protein in infection of mosquito cells. PLoS Negl Trop Dis. 2012;6(9):e1823.
99. Soldan SS, Plassmeyer ML, Matukonis MK, Gonzalez-Scarano F. La Crosse virus nonstructural protein NSs counteracts the effects of short interfering RNA. J Virol. 2005;79(1):234-44.
100. Kakumani PK, Ponia SS, S RK, Sood V, Chinnappan M, Banerjea AC, et al. Role of RNA interference (RNAi) in dengue virus replication and identification of NS4B as an RNAi suppressor. J Virol. 2013;87(16):8870-83.

101. Pijlman GP, Funk A, Kondratieva N, Leung J, Torres S, van der Aa L, et al. A highly structured, nuclease-resistant, noncoding RNA produced by flaviviruses is required for pathogenicity. Cell Host Microbe. 2008;4(6):579-91.
102. Moon SL, Dodd BJ, Brackney DE, Wilusz CJ, Ebel GD, Wilusz J. Flavivirus sfRNA suppresses antiviral RNA interference in cultured cells and mosquitoes and directly interacts with the RNAi machinery. Virology. 2015;485:322-9.
103. Schnettler E, Sterken MG, Leung JY, Metz SW, Geertsema C, Goldbach RW, et al. Noncoding flavivirus RNA displays RNA interference suppressor activity in insect and Mammalian cells. J Virol. 2012;86(24):13486-500.
104. Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH, et al. MicroRNA genes are transcribed by RNA polymerase II. EMBO J. 2004;23(20):4051-60.
105. Cai X, Hagedorn CH, Cullen BR. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. RNA. 2004;10(12):1957-66.
106. Borchert GM, Lanier W, Davidson BL. RNA polymerase III transcribes human microRNAs. Nat Struct Mol Biol. 2006;13(12):1097-101.
107. Finnegan EF, Pasquinelli AE. MicroRNA biogenesis: regulating the regulators. Crit Rev Biochem Mol Biol. 2013;48(1):51-68.
108. Saini HK, Griffiths-Jones S, Enright AJ. Genomic analysis of human microRNA transcripts. Proc Natl Acad Sci U S A. 2007;104(45):17719-24.
109. Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A. Identification of mammalian microRNA host genes and transcription units. Genome Res. 2004;14(10A):1902-10.
110. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. 2004;116(2):281-97.
111. Denli AM, Tops BB, Plasterk RH, Ketting RF, Hannon GJ. Processing of primary microRNAs by the Microprocessor complex. Nature. 2004;432(7014):231-5.
112. Gregory RI, Yan KP, Amuthan G, Chendrimada T, Doratotaj B, Cooch N, et al. The Microprocessor complex mediates the genesis of microRNAs. Nature. 2004;432(7014):235-40.
113. Landthaler M, Yalcin A, Tuschl T. The human DiGeorge syndrome critical region gene 8 and Its D. melanogaster homolog are required for miRNA biogenesis. Curr Biol. 2004;14(23):2162-7.
114. Han J, Lee Y, Yeom KH, Kim YK, Jin H, Kim VN. The Drosha-DGCR8 complex in primary microRNA processing. Genes Dev. 2004;18(24):3016-27.
115. Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, et al. The nuclear RNase III Drosha initiates microRNA processing. Nature. 2003;425(6956):415-9.
116. Bohnsack MT, Czaplinski K, Gorlich D. Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. RNA. 2004;10(2):185-91.
117. Lund E, Guttinger S, Calado A, Dahlberg JE, Kutay U. Nuclear export of microRNA precursors. Science. 2004;303(5654):95-8.
118. Yi R, Qin Y, Macara IG, Cullen BR. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev. 2003;17(24):3011-6.
119. Zeng Y, Cullen BR. Structural requirements for pre-microRNA binding and nuclear export by Exportin 5. Nucleic Acids Res. 2004;32(16):4776-85.
120. Hutvagner G, McLachlan J, Pasquinelli AE, Balint E, Tuschl T, Zamore PD. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science. 2001;293(5531):834-8.
121. Lee YS, Nakahara K, Pham JW, Kim K, He Z, Sontheimer EJ, et al. Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways. Cell. 2004;117(1):69-81.
122. Forstemann K, Horwich MD, Wee L, Tomari Y, Zamore PD. Drosophila microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1. Cell. 2007;130(2):287-97.
123. Tomari Y, Du T, Zamore PD. Sorting of Drosophila small silencing RNAs. Cell. 2007;130(2):299-308.

124. Schwarz DS, Hutvagner G, Du T, Xu Z, Aronin N, Zamore PD. Asymmetry in the assembly of the RNAi enzyme complex. Cell. 2003;115(2):199-208.

125. Khvorova A, Reynolds A, Jayasena SD. Functional siRNAs and miRNAs exhibit strand bias. Cell. 2003;115(2):209-16.

126. Okamura K, Phillips MD, Tyler DM, Duan H, Chou YT, Lai EC. The regulatory activity of microRNA* species has substantial influence on microRNA and 3′ UTR evolution. Nat Struct Mol Biol. 2008;15(4):354-63.

127. Czech B, Zhou R, Erlich Y, Brennecke J, Binari R, Villalta C, et al. Hierarchical rules for Argonaute loading in Drosophila. Mol Cell. 2009;36(3):445-56.

128. Marco A, Macpherson JI, Ronshaugen M, Griffiths-Jones S. MicroRNAs from the same precursor have different targeting properties. Silence. 2012;3(1):8.

129. Jagadeeswaran G, Zheng Y, Sumathipala N, Jiang H, Arrese EL, Soulages JL, et al. Deep sequencing of small RNA libraries reveals dynamic regulation of conserved and novel microRNAs and microRNA-stars during silkworm development. BMC Genomics. 2010;11:52.

130. Glazov EA, Cottee PA, Barris WC, Moore RJ, Dalrymple BP, Tizard ML. A microRNA catalog of the developing chicken embryo identified by a deep sequencing approach. Genome Res. 2008;18(6):957-64.

131. Chiang HR, Schoenfeld LW, Ruby JG, Auyeung VC, Spies N, Baek D, et al. Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. Genes Dev. 2010;24(10):992-1009.

132. Etebari K, Osei-Amo S, Blomberg SP, Asgari S. Dengue virus infection alters post-transcriptional modification of microRNAs in the mosquito vector Aedes aegypti. Sci Rep. 2015;5:15968.

133. Griffiths-Jones S, Hui JH, Marco A, Ronshaugen M. MicroRNA evolution by arm switching. EMBO Rep. 2011;12(2):172-7.

134. Ruby JG, Jan CH, Bartel DP. Intronic microRNA precursors that bypass Drosha processing. Nature. 2007;448(7149):83-6.

135. Okamura K, Hagen JW, Duan H, Tyler DM, Lai EC. The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. Cell. 2007;130(1):89-100.

136. Berezikov E, Chung WJ, Willis J, Cuppen E, Lai EC. Mammalian mirtron genes. Mol Cell. 2007;28(2):328-36.

137. Cazalla D, Xie M, Steitz JA. A primate herpesvirus uses the integrator complex to generate viral microRNAs. Mol Cell. 2011;43(6):982-92.

138. Bogerd HP, Karnowski HW, Cai X, Shin J, Pohlers M, Cullen BR. A mammalian herpesvirus uses noncanonical expression and processing mechanisms to generate viral MicroRNAs. Mol Cell. 2010;37(1):135-42.

139. Cheloufi S, Dos Santos CO, Chong MM, Hannon GJ. A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. Nature. 2010;465(7298):584-9.

140. 140. Cifuentes D, Xue H, Taylor DW, Patnode H, Mishima Y, Cheloufi S, et al. A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. Science. 2010;328(5986):1694-8.

141. Yang JS, Maurin T, Robine N, Rasmussen KD, Jeffrey KL, Chandwani R, et al. Conserved vertebrate mir-451 provides a platform for Dicer-independent, Ago2-mediated microRNA biogenesis. Proc Natl Acad Sci U S A. 2010;107(34):15163-8.

142. Yoda M, Cifuentes D, Izumi N, Sakaguchi Y, Suzuki T, Giraldez AJ, et al. Poly(A)-specific ribonuclease mediates 3′-end trimming of Argonaute2-cleaved precursor microRNAs. Cell Rep. 2013;5(3):715-26.

143. Hansen TB, Veno MT, Jensen TI, Schaefer A, Damgaard CK, Kjems J. Argonaute-associated short introns are a novel class of gene regulators. Nat Commun. 2016;7:11538.

144. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell. 2009;136(2):215-33.

145. Eichhorn SW, Guo H, McGeary SE, Rodriguez-Mias RA, Shin C, Baek D, et al. mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues. Mol Cell. 2014;56(1):104-15.

146. Ameres SL, Zamore PD. Diversifying microRNA sequence and function. Nat Rev Mol Cell Biol. 2013;14(8):475-88.

147. Friedman RC, Farh KK, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. 2009;19(1):92-105.

148. Kincaid RP, Sullivan CS. Virus-encoded microRNAs: an overview and a look to the future. PLoS Pathog. 2012;8(12):e1003018.

149. Kincaid RP, Burke JM, Sullivan CS. RNA virus microRNA that mimics a B-cell oncomiR. Proc Natl Acad Sci U S A. 2012;109(8):3077-82.

150. Shapiro JS, Langlois RA, Pham AM, Tenoever BR. Evidence for a cytoplasmic microprocessor of pri-miRNAs. RNA. 2012;18(7):1338-46.

151. Rouha H, Thurner C, Mandl CW. Functional microRNA generated from a cytoplasmic RNA virus. Nucleic Acids Res. 2010;38(22):8328-37.

152. Hussain M, Torres S, Schnettler E, Funk A, Grundhoff A, Pijlman GP, et al. West Nile virus encodes a microRNA-like small RNA in the 3' untranslated region which up-regulates GATA4 mRNA and facilitates virus replication in mosquito cells. Nucleic Acids Res. 2012;40(5):2210-23.

153. Hussain M, Asgari S. MicroRNA-like viral small RNA from Dengue virus 2 autoregulates its replication in mosquito cells. Proc Natl Acad Sci U S A. 2014;111(7):2746-51.

154. Skalsky RL, Olson KE, Blair CD, Garcia-Blanco MA, Cullen BR. A „microRNA-like" small RNA expressed by Dengue virus? Proc Natl Acad Sci U S A. 2014;111(23):E2359.

155. Finol E. Are viral small RNA regulating Dengue virus replication beyond serotype 2? Proc Natl Acad Sci U S A. 2014;111(29):E2915-6.

156. Liu Y, Zhou Y, Wu J, Zheng P, Li Y, Zheng X, et al. The expression profile of Aedes albopictus miRNAs is altered by dengue virus serotype-2 infection. Cell Biosci. 2015;5:16.

157. Liu YX, Li FX, Liu ZZ, Jia ZR, Zhou YH, Zhang H, et al. Integrated analysis of miRNAs and transcriptomes in Aedes albopictus midgut reveals the differential expression profiles of immune-related genes during dengue virus serotype-2 infection. Insect Sci. 2016;23(3):377-85.

158. Campbell CL, Harrison T, Hess AM, Ebel GD. MicroRNA levels are modulated in Aedes aegypti after exposure to Dengue-2. Insect Mol Biol. 2014;23(1):132-9.

159. Shrinet J, Jain S, Jain J, Bhatnagar RK, Sunil S. Next generation sequencing reveals regulation of distinct Aedes microRNAs during chikungunya virus development. PLoS Negl Trop Dis. 2014;8(1):e2616.

160. Du J, Gao S, Tian Z, Xing S, Huang D, Zhang G, et al. MicroRNA expression profiling of primary sheep testicular cells in response to bluetongue virus infection. Infect Genet Evol. 2017;49:256-67.

161. Slonchak A, Hussain M, Torres S, Asgari S, Khromykh AA. Expression of mosquito microRNA Aae-miR-2940-5p is downregulated in response to West Nile virus infection to restrict viral replication. J Virol. 2014;88(15):8457-67.

162. Yan H, Zhou Y, Liu Y, Deng Y, Chen X. miR-252 of the Asian tiger mosquito Aedes albopictus regulates dengue virus replication by suppressing the expression of the dengue virus envelope protein. J Med Virol. 2014;86(8):1428-36.

163. Czech B, Hannon GJ. One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing. Trends Biochem Sci. 2016;41(4):324-37.

164. Iwasaki YW, Siomi MC, Siomi H. PIWI-Interacting RNA: Its Biogenesis and Functions. Annu Rev Biochem. 2015;84:405-33.

165. Siomi MC, Sato K, Pezic D, Aravin AA. PIWI-interacting small RNAs: the vanguard of genome defence. Nat Rev Mol Cell Biol. 2011;12(4):246-58.

166. Klattenhoff C, Bratu DP, McGinnis-Schultz N, Koppetsch BS, Cook HA, Theurkauf WE. Drosophila rasiRNA pathway mutations disrupt embryonic axis specification through activation of an ATR/Chk2 DNA damage response. Dev Cell. 2007;12(1):45-55.

167. Klenov MS, Sokolova OA, Yakushev EY, Stolyarenko AD, Mikhaleva EA, Lavrov SA, et al. Separation of stem cell maintenance and transposon silencing functions of Piwi protein. Proc Natl Acad Sci U S A. 2011;108(46):18760-5.

168. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, et al. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell. 2007;128(6):1089-103.

169. Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, et al. A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in Drosophila. Science. 2007;315(5818):1587-90.

170. Rangan P, Malone CD, Navarro C, Newbold SP, Hayes PS, Sachidanandam R, et al. piRNA production requires heterochromatin formation in Drosophila. Curr Biol. 2011;21(16):1373-9.

171. Malone CD, Brennecke J, Dus M, Stark A, McCombie WR, Sachidanandam R, et al. Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. Cell. 2009;137(3):522-35.

172. Zanni V, Eymery A, Coiffet M, Zytnicki M, Luyten I, Quesneville H, et al. Distribution, evolution, and diversity of retrotransposons at the flamenco locus reflect the regulatory properties of piRNA clusters. Proc Natl Acad Sci U S A. 2013;110(49):19842-7.

173. Mohn F, Sienski G, Handler D, Brennecke J. The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in Drosophila. Cell. 2014;157(6):1364-79.

174. Goriaux C, Desset S, Renaud Y, Vaury C, Brasset E. Transcriptional properties and splicing of the flamenco piRNA cluster. EMBO Rep. 2014;15(4):411-8.

175. Zhang Z, Wang J, Schultz N, Zhang F, Parhad SS, Tu S, et al. The HP1 homolog rhino anchors a nuclear complex that suppresses piRNA precursor splicing. Cell. 2014;157(6):1353-63.

176. Homolka D, Pandey RR, Goriaux C, Brasset E, Vaury C, Sachidanandam R, et al. PIWI Slicing and RNA Elements in Precursors Instruct Directional Primary piRNA Biogenesis. Cell Rep. 2015;12(3):418-28.

177. Ishizu H, Iwasaki YW, Hirakata S, Ozaki H, Iwasaki W, Siomi H, et al. Somatic Primary piRNA Biogenesis Driven by cis-Acting RNA Elements and trans-Acting Yb. Cell Rep. 2015;12(3):429-40.

178. Li C, Vagin VV, Lee S, Xu J, Ma S, Xi H, et al. Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. Cell. 2009;137(3):509-21.

179. Czech B, Preall JB, McGinn J, Hannon GJ. A transcriptome-wide RNAi screen in the Drosophila ovary reveals factors of the germline piRNA pathway. Mol Cell. 2013;50(5):749-61.

180. Olivieri D, Sykora MM, Sachidanandam R, Mechtler K, Brennecke J. An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in Drosophila. EMBO J. 2010;29(19):3301-17.

181. Muerdter F, Guzzardo PM, Gillis J, Luo Y, Yu Y, Chen C, et al. A genome-wide RNAi screen draws a genetic framework for transposon control and primary piRNA biogenesis in Drosophila. Mol Cell. 2013;50(5):736-48.

182. Handler D, Meixner K, Pizka M, Lauss K, Schmied C, Gruber FS, et al. The genetic makeup of the Drosophila piRNA pathway. Mol Cell. 2013;50(5):762-77.

183. Saito K, Ishizu H, Komai M, Kotani H, Kawamura Y, Nishida KM, et al. Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in Drosophila. Genes Dev. 2010;24(22):2493-8.

184. Ipsaro JJ, Haase AD, Knott SR, Joshua-Tor L, Hannon GJ. The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis. Nature. 2012;491(7423):279-83.

185. Nishimasu H, Ishizu H, Saito K, Fukuhara S, Kamatani MK, Bonnefond L, et al. Structure and function of Zucchini endoribonuclease in piRNA biogenesis. Nature. 2012;491(7423):284-7.

186. Vagin VV, Yu Y, Jankowska A, Luo Y, Wasik KA, Malone CD, et al. Minotaur is critical for primary piRNA biogenesis. RNA. 2013;19(8):1064-77.

187. Han BW, Wang W, Li C, Weng Z, Zamore PD. Noncoding RNA. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. Science. 2015;348(6236):817-21.

188. Mohn F, Handler D, Brennecke J. Noncoding RNA. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. Science. 2015;348(6236):812-7.

189. Senti KA, Jurczak D, Sachidanandam R, Brennecke J. piRNA-guided slicing of transposon transcripts enforces their transcriptional silencing via specifying the nuclear piRNA repertoire. Genes Dev. 2015;29(16):1747-62.

190. Wang W, Han BW, Tipping C, Ge DT, Zhang Z, Weng Z, et al. Slicing and Binding by Ago3 or Aub Trigger Piwi-Bound piRNA Production by Distinct Mechanisms. Mol Cell. 2015;59(5):819-30.

191. Feltzin VL, Khaladkar M, Abe M, Parisi M, Hendriks GJ, Kim J, et al. The exonuclease Nibbler regulates age-associated traits and modulates piRNA length in Drosophila. Aging Cell. 2015;14(3):443-52.

192. Wang H, Ma Z, Niu K, Xiao Y, Wu X, Pan C, et al. Antagonistic roles of Nibbler and Hen1 in modulating piRNA 3' ends in Drosophila. Development. 2016;143(3):530-9.

193. Hayashi R, Schnabl J, Handler D, Mohn F, Ameres SL, Brennecke J. Genetic and mechanistic diversity of piRNA 3'-end formation. Nature. 2016;539(7630):588-92.

194. Nishida KM, Iwasaki YW, Murota Y, Nagao A, Mannen T, Kato Y, et al. Respective functions of two distinct Siwi complexes assembled during PIWI-interacting RNA biogenesis in Bombyx germ cells. Cell Rep. 2015;10(2):193-203.

195. Izumi N, Shoji K, Sakaguchi Y, Honda S, Kirino Y, Suzuki T, et al. Identification and Functional Analysis of the Pre-piRNA 3' Trimmer in Silkworms. Cell. 2016;164(5):962-73.

196. Kawaoka S, Izumi N, Katsuma S, Tomari Y. 3' end formation of PIWI-interacting RNAs in vitro. Mol Cell. 2011;43(6):1015-22.

197. Saito K, Sakaguchi Y, Suzuki T, Suzuki T, Siomi H, Siomi MC. Pimet, the Drosophila homolog of HEN1, mediates 2'-O-methylation of Piwi- interacting RNAs at their 3' ends. Genes Dev. 2007;21(13):1603-8.

198. Murota Y, Ishizu H, Nakagawa S, Iwasaki YW, Shibata S, Kamatani MK, et al. Yb integrates piRNA intermediates and processing factors into perinuclear bodies to enhance piRISC assembly. Cell Rep. 2014;8(1):103-13.

199. Handler D, Olivieri D, Novatchkova M, Gruber FS, Meixner K, Mechtler K, et al. A systematic analysis of Drosophila TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. EMBO J. 2011;30(19):3977-93.

200. Zamparini AL, Davis MY, Malone CD, Vieira E, Zavadil J, Sachidanandam R, et al. Vreteno, a gonad-specific protein, is essential for germline development and primary piRNA biogenesis in Drosophila. Development. 2011;138(18):4039-50.

201. Karam JA, Parikh RY, Nayak D, Rosenkranz D, Gangaraju VK. Co-chaperone Hsp70/Hsp90 organizing protein (Hop) is Required for Transposon Silencing and piRNA Biogenesis. J Biol Chem. 2017; 292(15):6039-46.

202. Olivieri D, Senti KA, Subramanian S, Sachidanandam R, Brennecke J. The cochaperone shutdown defines a group of biogenesis factors essential for all piRNA populations in Drosophila. Mol Cell. 2012;47(6):954-69.

203. Preall JB, Czech B, Guzzardo PM, Muerdter F, Hannon GJ. shutdown is a component of the Drosophila piRNA biogenesis machinery. RNA. 2012;18(8):1446-57.

204. Lim AK, Kai T. Unique germ-line organelle, nuage, functions to repress selfish genetic elements in Drosophila melanogaster. Proc Natl Acad Sci U S A. 2007;104(16):6714-9.

205. Zhang F, Wang J, Xu J, Zhang Z, Koppetsch BS, Schultz N, et al. UAP56 couples piRNA clusters to the perinuclear transposon silencing machinery. Cell. 2012;151(4):871-84.

206. Cora E, Pandey RR, Xiol J, Taylor J, Sachidanandam R, McCarthy AA, et al. The MID-PIWI module of Piwi proteins specifies nucleotide- and strand-biases of piRNAs. RNA. 2014;20(6):773-81.

207. Sienski G, Donertas D, Brennecke J. Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. Cell. 2012;151(5):964-80.

208. Le Thomas A, Rogers AK, Webster A, Marinov GK, Liao SE, Perkins EM, et al. Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. Genes Dev. 2013;27(4):390-9.

209. Klenov MS, Lavrov SA, Korbut AP, Stolyarenko AD, Yakushev EY, Reuter M, et al. Impact of nuclear Piwi elimination on chromatin state in Drosophila melanogaster ovaries. Nucleic Acids Res. 2014;42(10):6208-18.

210. Rozhkov NV, Hammell M, Hannon GJ. Multiple roles for Piwi in silencing Drosophila transposons. Genes Dev. 2013;27(4):400-12.

211. Wang SH, Elgin SC. Drosophila Piwi functions downstream of piRNA production mediating a chromatin-based transposon silencing mechanism in female germ line. Proc Natl Acad Sci U S A. 2011;108(52):21164-9.

212. Huang XA, Yin H, Sweeney S, Raha D, Snyder M, Lin H. A major epigenetic programming mechanism guided by piRNAs. Dev Cell. 2013;24(5):502-16.

213. Wang W, Yoshikawa M, Han BW, Izumi N, Tomari Y, Weng Z, et al. The initial uridine of primary piRNAs does not create the tenth adenine that Is the hallmark of secondary piRNAs. Mol Cell. 2014;56(5):708-16.

214. Siomi MC, Mannen T, Siomi H. How does the royal family of Tudor rule the PIWI-interacting RNA pathway? Genes Dev. 2010;24(7):636-46.

215. Kirino Y, Kim N, de Planell-Saguer M, Khandros E, Chiorean S, Klein PS, et al. Arginine methylation of Piwi proteins catalysed by dPRMT5 is required for Ago3 and Aub stability. Nat Cell Biol. 2009;11(5):652-8.

216. Nishida KM, Okada TN, Kawamura T, Mituyama T, Kawamura Y, Inagaki S, et al. Functional involvement of Tudor and dPRMT5 in the piRNA processing pathway in Drosophila germlines. EMBO J. 2009;28(24):3820-31.

217. Anand A, Kai T. The tudor domain protein kumo is required to assemble the nuage and to generate germline piRNAs in Drosophila. EMBO J. 2012;31(4):870-82.

218. Zhang Z, Xu J, Koppetsch BS, Wang J, Tipping C, Ma S, et al. Heterotypic piRNA Ping-Pong requires qin, a protein with both E3 ligase and Tudor domains. Mol Cell. 2011;44(4):572-84.

219. Patil VS, Kai T. Repression of retroelements in Drosophila germline via piRNA pathway by the Tudor domain protein Tejas. Curr Biol. 2010;20(8):724-30.

220. Patil VS, Anand A, Chakrabarti A, Kai T. The Tudor domain protein Tapas, a homolog of the vertebrate Tdrd7, functions in the piRNA pathway to regulate retrotransposons in germline of Drosophila melanogaster. BMC Biol. 2014;12:61.

221. Webster A, Li S, Hur JK, Wachsmuth M, Bois JS, Perkins EM, et al. Aub and Ago3 Are Recruited to Nuage through Two Mechanisms to Form a Ping-Pong Complex Assembled by Krimper. Mol Cell. 2015;59(4):564-75.

222. Sato K, Iwasaki YW, Shibuya A, Carninci P, Tsuchizawa Y, Ishizu H, et al. Krimper Enforces an Antisense Bias on piRNA Pools by Binding AGO3 in the Drosophila Germline. Mol Cell. 2015;59(4):553-63.

223. Xiol J, Spinelli P, Laussmann MA, Homolka D, Yang Z, Cora E, et al. RNA clamping by Vasa assembles a piRNA amplifier complex on transposon transcripts. Cell. 2014;157(7):1698-711.

224. Donertas D, Sienski G, Brennecke J. Drosophila Gtsf1 is an essential component of the Piwi-mediated transcriptional silencing complex. Genes Dev. 2013;27(15):1693-705.

225. Ohtani H, Iwasaki YW, Shibuya A, Siomi H, Siomi MC, Saito K. DmGTSF1 is necessary for Piwi-piRISC-mediated transcriptional transposon silencing in the Drosophila ovary. Genes Dev. 2013;27(15):1656-61.

226. 226. Sienski G, Batki J, Senti KA, Donertas D, Tirian L, Meixner K, et al. Silencio/CG9754 connects the Piwi-piRNA complex to the cellular heterochromatin machinery. Genes Dev. 2015;29(21):2258-71.

227. Yu Y, Gu J, Jin Y, Luo Y, Preall JB, Ma J, et al. Panoramix enforces piRNA-dependent cotranscriptional silencing. Science. 2015;350(6258):339-42.

228. Aravin AA, Naumova NM, Tulin AV, Vagin VV, Rozovsky YM, Gvozdev VA. Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. Curr Biol. 2001;11(13):1017-27.

229. Aravin AA, Lagos-Quintana M, Yalcin A, Zavolan M, Marks D, Snyder B, et al. The small RNA profile during Drosophila melanogaster development. Dev Cell. 2003;5(2):337-50.

230. Robine N, Lau NC, Balla S, Jin Z, Okamura K, Kuramochi-Miyagawa S, et al. A broadly conserved pathway generates 3'UTR-directed primary piRNAs. Curr Biol. 2009;19(24):2066-76.

231. Saito K, Inagaki S, Mituyama T, Kawamura Y, Ono Y, Sakota E, et al. A regulatory circuit for piwi by the large Maf gene traffic jam in Drosophila. Nature. 2009;461(7268):1296-9.

232. Kiuchi T, Koga H, Kawamoto M, Shoji K, Sakai H, Arai Y, et al. A single female-specific piRNA is the primary determiner of sex in the silkworm. Nature. 2014;509(7502):633-6.

233. Scott JC, Brackney DE, Campbell CL, Bondu-Hawkins V, Hjelle B, Ebel GD, et al. Comparison of dengue virus type 2-specific small RNAs from RNA interference-competent and -incompetent mosquito cells. PLoS Negl Trop Dis. 2010;4(10):e848.

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

235. Morazzani EM, Wiley MR, Murreddu MG, Adelman ZN, Myles KM. Production of virus-derived ping-pong-dependent piRNA-like small RNAs in the mosquito soma. PLoS Pathog. 2012;8(1):e1002470.

236. Vodovar N, Bronkhorst AW, van Cleef KW, Miesen P, Blanc H, van Rij RP, et al. Arbovirus-derived piRNAs exhibit a ping-pong signature in mosquito cells. PLoS One. 2012;7(1):e30861.

237. Akbari OS, Antoshechkin I, Amrhein H, Williams B, Diloreto R, Sandler J, et al. The developmental transcriptome of the mosquito Aedes aegypti, an invasive species and major arbovirus vector. G3 (Bethesda). 2013;3(9):1493-509.

238. Campbell CL, Black WCt, Hess AM, Foy BD. Comparative genomics of small RNA regulatory pathway components in vector mosquitoes. BMC Genomics. 2008;9:425.

239. Arensburger P, Hice RH, Wright JA, Craig NL, Atkinson PW. The mosquito Aedes aegypti has a large genome size and high transposable element load but contains a low proportion of transposon-specific piRNAs. BMC Genomics. 2011;12:606.

240. Schirle NT, Sheu-Gruttadauria J, MacRae IJ. Structural basis for microRNA targeting. Science. 2014;346(6209):608-13.

241. Krieger E, Vriend G. YASARA View - molecular graphics for all devices - from smartphones to workstations. Bioinformatics. 2014;30(20):2981-2.

Chapter 2

Distinct Sets of PIWI Proteins Produce Arbovirus and Transposon-Derived piRNAs in *Aedes aegypti* Mosquito Cells.

Pascal Miesen, Erika Girardi, Ronald P. van Rij

Nucleic Acids Res. (2015) 43:6545-56

ABSTRACT

The PIWI-interacting RNA (piRNA) pathway is essential for transposon silencing in many model organisms. Its remarkable efficiency relies on a sophisticated amplification mechanism known as the ping-pong loop. In Alphavirus-infected *Aedes* mosquitoes, piRNAs with sequence features that suggest ping-pong-dependent biogenesis are produced from viral RNA. The PIWI family in *Aedes* mosquitoes is expanded when compared to other model organisms, raising the possibility that individual PIWI proteins have functionally diversified in these insects. Here, we show that Piwi5 and Ago3, but none of the other PIWI family members, are essential for piRNA biogenesis from Sindbis virus RNA in infected *Aedes aegypti* cells. In contrast, the production of piRNAs from transposons relies on a more versatile set of PIWI proteins, some of which do not contribute to viral piRNA biogenesis. These results indicate that functional specialization allows distinct mosquito PIWI proteins to process RNA from different endogenous and exogenous sources.

INTRODUCTION

In the animal kingdom, three major classes of small silencing RNAs exist: microRNAs (miRNAs), small interfering RNAs (siRNAs), and PIWI-interacting RNAs (piRNAs) (1). All of these function in the context of proteins from the Argonaute superfamily. siRNAs and miRNAs associate with the AGO clade, whereas piRNAs are bound by the PIWI clade of Argonaute proteins (2). The small RNAs guide these proteins to complementary RNA molecules, which typically results in sequence-dependent suppression of those targets. Some Argonaute proteins can cleave their target RNAs (slicer activity), which are then susceptible to degradation by cellular exonucleases (3). PIWI proteins, however, are an exception, since their cleavage products can be processed into new piRNAs (4-7).

In animals, the piRNA pathway is key to the protection of the genome against the activity of transposable elements (TEs) (8, 9). Still, our knowledge of piRNA biogenesis is incomplete and remains limited to a few model organisms. In the fruit fly *Drosophila melanogaster*, piRNA biogenesis involves two mechanisms: the primary processing pathway and a secondary amplification pathway, referred to as the ping-pong loop (10). The primary pathway generates from genomically encoded precursors a pool of primary piRNAs, which are loaded into the PIWI proteins Piwi and Aubergine (Aub) (4). From this initial piRNA collection, the ping-pong loop selectively amplifies Aub-bound piRNAs that recognize transcripts of active transposons (4, 5). The PIWI protein Ago3 engages in this sophisticated feed-forward mechanism along with Aub. Both proteins mutually produce the piRNA precursors for each other, since the 3′ cleavage products generated by Aub can be transferred to Ago3 and *vice versa* (4, 5, 11, 12). Once loaded in a PIWI protein, piRNA precursors are further processed into mature piRNAs, which are 25-30 nt in size and contain a 2′-*O*-methyl group at their 3′ terminal nucleotide (8).

Aub-bound piRNAs commonly start with a uridine (1U) and, since target slicing by PIWI proteins occurs between nucleotide ten and eleven, the complementary Ago3-bound piRNAs typically have a 10 nt overlap and contain an adenine at position ten (10A) (4, 5). This specific sequence signature is a hallmark of piRNAs that have been amplified by the ping-pong loop. piRNA amplification was initially thought to occur exclusively in germline tissues, but recently, piRNAs have been detected in somatic cells in several organisms, including various mosquito species (13-16).

Blood-sucking mosquitoes are crucial for the transmission of many arthropod-borne viruses (arboviruses). Intriguingly, infected mosquitoes generally do not show signs of pathology, suggesting that they possess efficient pathways to resist or tolerate virus infection (17). Key to antiviral immunity in insects is the RNA interference (RNAi) pathway with at its core 21 nt viral siRNAs (vsiRNAs) bound to Argonaute 2 (Ago2) (18, 19). These vsiRNAs are processed from viral double-stranded RNA (dsRNA), which accumulates in infected cells during the replication cycle of many viruses (20). Unexpectedly, besides vsiRNAs, we and others have recently cloned and sequenced viral small RNAs with the sequence signature of ping-pong-dependent piRNAs in somatic cells of infected *Aedes* mosquitoes and in cell lines derived from these insects (14, 15, 21-23). Still, the biogenesis and function of these viral piRNAs (vpiRNAs) are not well understood. Neither has their association with a PIWI protein been demonstrated, which would formally classify these viral small RNAs as *PIWI interacting* RNAs. Interestingly, whereas flies encode three PIWI proteins, the PIWI family is expanded to eight members (Piwi1-7 and Ago3) in *Aedes aegypti*. However, with the exception of Ago3, no 1:1 orthology exists between *Aedes* PIWI proteins and known piRNA biogenesis factors (24). Combined knockdown of all *Aedes* PIWI proteins abrogates vpiRNA biogenesis (21), but the contribution of the individual PIWI proteins to vpiRNA biogenesis in mosquitoes remains obscure.

The diversification of PIWI proteins and the accumulation of ping-pong-dependent vpiRNAs suggest that the PIWI pathway in mosquitoes has gained additional functions besides the repression of transposon activity. An exciting possibility is that the PIWI gene expansion has allowed functional specialization in producing piRNAs from different RNA sources. Here, we test this hypothesis making use of the piRNA competent *Aedes aegypti* Aag2 cell line. These cells produce Alphavirus-derived piRNAs with striking similarities to vpiRNAs in the adult mosquito (14). In addition, their PIWI protein repertoire strongly mimics the PIWI expression profile in somatic tissues of adult mosquitoes, as recently determined by RNA sequencing (25). Therefore, the Aag2 cell line is an accessible and relevant model system to investigate the molecular mechanisms of (viral) piRNA biogenesis in *Aedes*. Using this model, we identify Piwi5 and Ago3 as the core proteins of the mosquito ping-pong loop. During infection with Sindbis virus (SINV), the production of piRNAs of viral origin is almost exclusively dependent on ping-pong amplification by Piwi5 and Ago3, whereas the biogenesis of transposon-

derived piRNAs is more versatile and involves additional members of the PIWI protein family. These data suggest that specialized arms of the mosquito PIWI pathway engage in piRNA biogenesis from endogenous or exogenous RNAs.

MATERIALS AND METHODS

Transfection and infection of Aag2 cells

For immunoprecipitation (IP) and immunofluorescence analyses (IFA), Aag2 cells were transfected with expression plasmids encoding individual PIWI proteins and, where indicated, infected with SINV at an MOI of 1 immediately after transfection. For knockdown experiments, Aag2 were transfected with dsRNA and re-transfected 48h after the first transfection to boost the knockdown. Where indicated, cells were then infected with SINV at an MOI of 1. Unless stated differently, samples were harvested 48h post infection. For a detailed description of the experimental procedure, the cloning of expression plasmids, cell culture conditions and virus production, see Supplemental data.

Northern blotting and qPCR

Small RNA northern blotting was performed using 1-ethyl-3- (3-dimethylaminopropyl) carbodiimide (EDC; Sigma) crosslinking after size separation on polyacrylamide gels as detailed in (26). For high molecular weight northern blot, RNA was separated on agarose gels and crosslinked using UV irradiation. For quantitative RT-PCR (RT-qPCR), total RNA was RNase treated, reverse transcribed, and PCR amplified in the presence of SYBR green. For a detailed description of the experimental procedures, the sequences of the northern blot probes, and the qPCR primers, see Supplemental data.

Western blotting and Immunofluorescence analysis

For western blotting, proteins were separated on polyacrylamide gels, blotted to nitrocellulose membranes, and probed with the indicated antibodies. Immunofluorescence analyses were performed on paraformaldehyde-fixed and permeabilized Aag2 cells. For a detailed description of the experimental procedure and the antibodies, see Supplemental data.

Immunoprecipitation

Lysates from Aag2 cells expressing V5-3xFlag tagged PIWI proteins were pre-cleared with protein G agarose beads and then incubated with V5-agarose beads (Sigma). The immunoprecipitates were washed, and RNA was isolated from the beads for subsequent analyses. For a detailed description of the experimental procedure, see Supplemental data.

with protein G agarose beads and then incubated with V5-agarose beads (Sigma). The immunoprecipitates were washed, and RNA was isolated from the beads for subsequent analyses. For a detailed description of the experimental procedure, see Supplemental data.

Cytoplasmic and nuclear fractionation

Aag2 cells were lysed in cytoplasmic lysis buffer (25mM Tris HCl, pH 7.5, 50mM NaCl, 2mM EDTA, 0.5% NP40, 1x protease inhibitors) and the cytoplasmic fraction was separated from the nuclear pellet by centrifugation. The nuclear pellet was washed in cytoplasmic lysis buffer and lysed in 1x SDS PAGE loading buffer for protein analysis or Isol-RNA lysis reagent (5 PRIME) for RNA isolation. Similarly, 5x SDS PAGE loading buffer or Isol-RNA lysis reagent was added to the cytoplasmic fraction for further processing. Protein or RNA fractions representing an equal number of cells were loaded on gel for western or northern blot analyses, respectively.

Preparation of small RNA libraries and bioinformatic analyses

For the analysis of small RNAs in PIWI protein knockdown samples, small RNA libraries were prepared as previously described (27) and sequenced on an Illumina HiSeq 2500. The sequence data were analyzed with Galaxy (galaxyproject.org) (28). Reads were clipped from the adapter sequence and mapped with Bowtie, version 1.1.2 (29), to the SINV genome (pTE2J-3'GFP) or to the *Aedes aegypti* transposon database (http://tefam.biochem.vt.edu; sequences downloaded on April 10th, 2014). Size profiles of the small RNAs were obtained from all reads that mapped to these sequences with a maximum of one mismatch. Read counts were normalized to the size of the corresponding library and expressed as '% of library'. To analyze the genome distribution of vpiRNAs or vsiRNAs, the 5' ends of the 25-30 nt or 21 nt SINV-mapping reads were plotted onto the viral genome. For plotting the genome distribution of vpiRNA reads from the PIWI IPs, the number of reads in the GFP-IP was subtracted from the PIWI-protein IP, to correct for background binding. When this corrected normalized read count was a negative value, it was set to zero. The overlap probability of viral piRNAs has been determined using the approach detailed in (30) using the small RNA signature tool available at the Mississippi Galaxy instance (mississippi.fr). Sequence logos were generated using WebLogo3.3 (31, 32) using the tool available at the Galaxy main server. For analyzing the number of piRNAs that map to individual transposons, only uniquely-mapping reads were taken into consideration. For each transposons, the piRNA enrichment upon PIWI knockdowns relative to the luciferase control knockdown was calculated and hierarchical clustering of the transposons was performed using Multiple experiment viewer (MEV version 4.8, Tm4) (33). Sequence data have been deposited in the NCBI Sequence Read Archive under accession number SRA188616.

RESULTS

Individual vpiRNAs are highly abundant in SINV-infected Aag2 cells

Previously, deep-sequencing of small RNAs in infected Aag2 cells identified vpiRNAs derived from SINV, a positive (+) strand RNA virus of the genus Alphavirus within the *Togaviridae* family (15). During SINV replication, the viral (+) RNA strand serves as a template for the production of negative (-) strand RNA, which in turn is a template for the production of full-length genomic RNA as well as for a subgenomic RNA species. The vast majority of vpiRNAs is derived from the viral (+) strand and has a 10A nucleotide bias, suggesting that their production requires ping-pong amplification. An approximately 200 nt large hotspot region for vpiRNA biogenesis is located in the capsid gene, 300 nt downstream of the SINV subgenomic promoter (Figure 1A). Read counts of several vpiRNAs within this hotspot are similar to those of average to highly expressed miRNAs, suggesting that they are efficiently produced and stably retained in Aag2 cells.

We selected four highly abundant vpiRNA sequences from the subgenomic hotspot region for small RNA northern blotting, all of which derive from the viral (+) strand. Indeed, these vpiRNAs were readily detected by northern blot in SINV-infected Aag2 cells (Figure 1B). These analyses were performed with recombinant SINV that expresses GFP from a second subgenomic promoter, which permits simple assessment of infections (Figure 1A). However, the same vpiRNA sequences were found in Aag2 cells infected with the parental virus, indicating that vpiRNAs are not an artefact of transgene expression from the second subgenomic promoter (Figure 1C). During the course of infection, vpiRNAs were visible as soon as 24 hours post infection (hpi), when infection was fully established (Figure 1D). In addition, northern blotting detected vpiRNAs in SINV-infected *Aedes albopictus* U4.4 and C6/36 mosquito cells, in line with previous observations using deep-sequencing technology (Figure 1E) (15). The higher accumulation of vpiRNAs in C6/36 cells is likely caused by elevated viral RNA replication, due to a defect in the antiviral RNAi response in these cells (34). As expected, mammalian BHK21 cells, which allow SINV replication to similarly high levels but are devoid of an active piRNA pathway, did not produce SINV-derived piRNAs (Figure 1E). To analyze whether the detected viral small RNAs were mature vpiRNAs, we performed sodium periodate ($NaIO_4$) oxidation followed by β-elimination. This reaction uncovers potential modifications of the ribose at the 3′ end of RNAs as it removes the terminal nucleoside of unmodified RNAs, leaving a 3′ monophosphate behind (35). Mature piRNAs are 2′-*O*-methylated at their 3′ end, and are therefore protected against this treatment (36, 37). This distinguishes them from animal miRNAs, which have no 3′ end modification and are therefore shortened by β-elimination. Northern blot of individual vpiRNAs showed that their electrophoretic mobility is unaffected by β-elimination, indicating that their 3′ end is 2′-*O*-methylated. Likewise, piRNAs derived from a Ty3/Gypsy transposon were equally insensitive to the treatment. As expected, a miRNA was shortened by the reaction and its electrophoretic

mobility clearly changed after treatment (Figure 1F). Taken together, these data indicate that individual, 2′-*O*-methylated vpiRNAs accumulate to high levels in infected Aag2 cells.

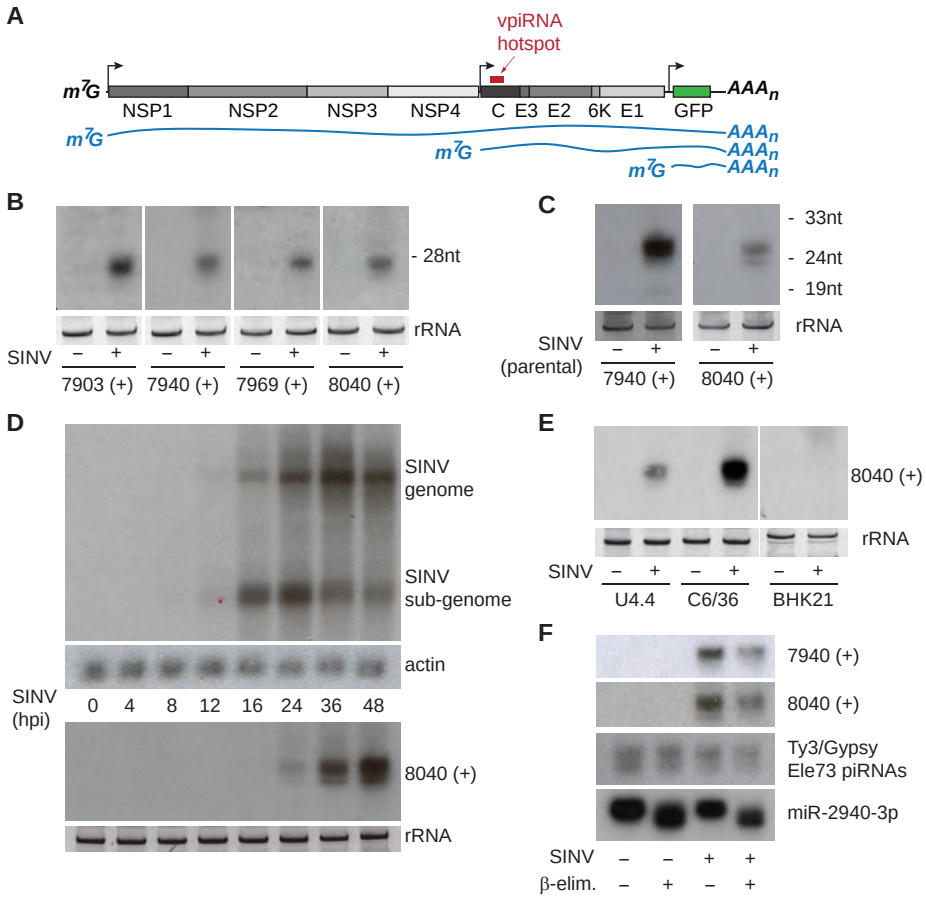


Figure 1. Selected mature vpiRNAs are abundant in Aag2 cells. (A) Schematic representation of the SINV-GFP genome. The individual viral proteins are indicated in grey and the position of the piRNA hotspot is marked by the red bar. The blue lines show the three (+) strand RNA species that can be found in infected cells. **(B)** Small RNA northern blot of four vpiRNAs in uninfected or SINV infected Aag2 cells. Probe names indicate the 5′ end position of the detected vpiRNAs, which are all derived from the SINV (+) strand. **(C)** Small RNA northern blot for vpiRNA in uninfected or SINV (parental virus) infected Aag2 cells. **(D)** Northern blot analysis of viral genomic and subgenomic RNA (upper panel) or vpiRNAs (lower panel) using a probe against vpiRNA 8040 (+). Probing for actin mRNA serves as loading control. **(E)** Northern blot analysis of vpiRNA in uninfected or SINV-infected *Aedes albopictus* mosquito cells (U4.4 and C6/36) and baby-hamster kidney cells (BHK21). For small RNA northern blots in panels B to E, ethidium bromide staining of ribosomal RNA (rRNA) serves as loading control. In panel B the loading controls for 7903 (+) and 7940 (+) are identical, since the same membrane was subsequently hybridized to these probes after harsh stripping in hot 0.1% SDS. **(F)** Northern blot detection of vpiRNAs, Ty3/Gypsy element 73 transposon piRNAs, or miR2940-3p. Before blotting, β-elimination was performed on total RNA as indicated.

Knockdown of Piwi5 and Ago3 abolishes secondary vpiRNA biogenesis

In Aag2 cells, transcripts of Piwi4, Piwi5, Piwi6, and Ago3 are readily detected; the abundance of Piwi1, Piwi2, Piwi3 and Piwi7, however, is considerably lower (15). This expression pattern mimics the PIWI expression profile in somatic tissue of adult mosquitoes, since Piwi1-3 are largely germline specific and Piwi7 is highly expressed only in the early embryo. (25). To investigate whether SINV infection alters PIWI mRNA abundance, we performed RT-qPCR for the individual PIWI transcripts, as well as for Ago1 and Ago2, which are involved in the biogenesis of miRNAs and siRNAs, respectively (38, 39). Expression of Piwi1, Piwi2, Piwi3 and Piwi7 was close to, or below the detection limit of our quantification method, both in uninfected and SINV-infected Aag2 cells. These genes were therefore excluded from qPCR analyses. With the exception of Piwi6, for which we noticed a mild reduction, infection with SINV did not substantially change mRNA expression of the remaining PIWI/AGO transcripts (Figure 2A). Next, we investigated which of the PIWI protein family members are involved in vpiRNA biogenesis. To this end, Aag2 cells were transfected with dsRNAs targeting the eight

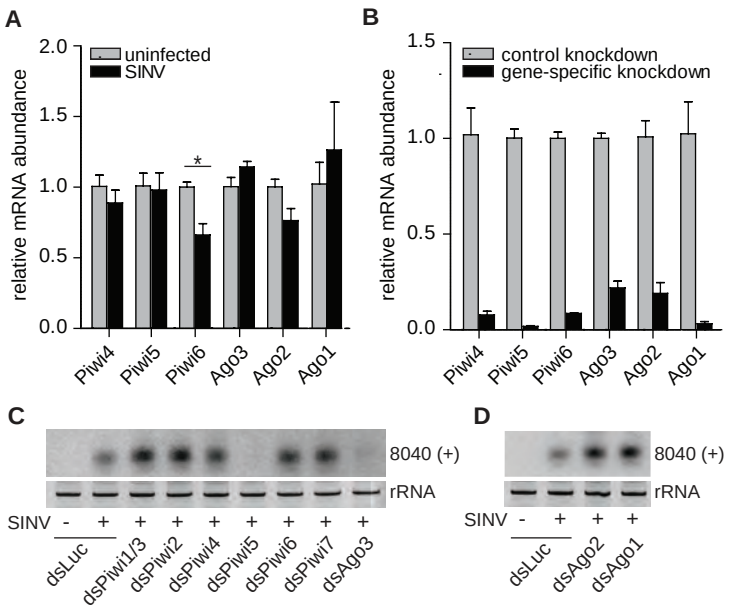


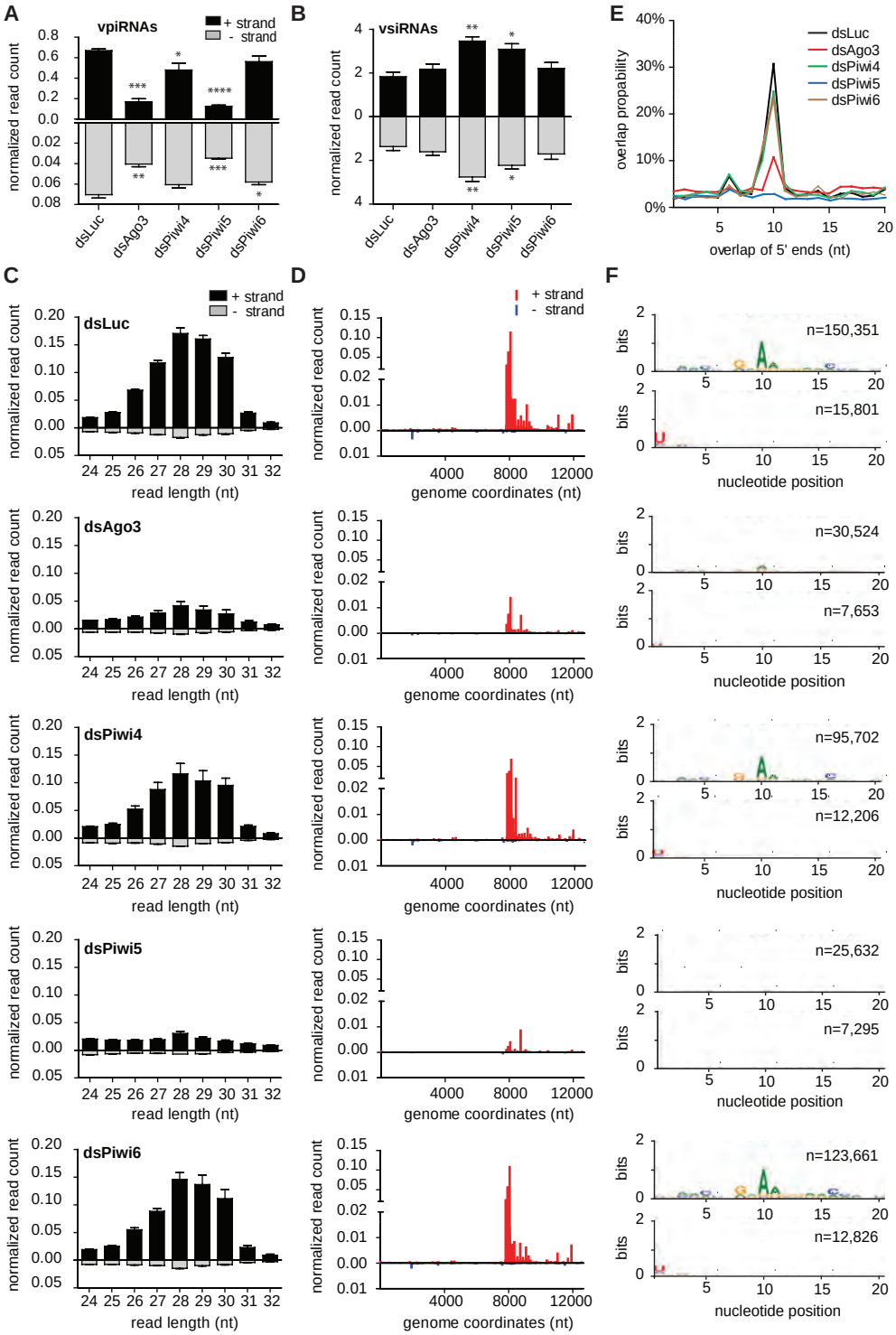
Figure 2. Piwi5 and Ago3 are required for secondary vpiRNA biogenesis. (A) qPCR analysis of the indicated PIWI/AGO transcripts in uninfected or SINV-infected Aag2 cells. Bars are the mean +/- SEM of three independent experiments. Student's t-test was used to determine statistical significance (*, $P<0.05$). **(B)** qPCR of PIWI/AGO transcripts in Aag2 cells at 48h after transfection of control dsRNA (dsLuc) or dsRNA targeting the corresponding gene. Expression levels were normalized to the control knockdown. Bars are the mean +/- SEM of three independent experiments. All changes in mRNA abundance shown are statistically significant with $P<0.005$. Expression of Piwi1-3 and Piwi7 were close to or below the detection limit and excluded from the analyses in panel A and B. **(C,D)** Northern blot for vpiRNA 8040 (+) upon knockdown of the indicated PIWI/AGO genes. Piwi1 and Piwi3 mRNA sequences are highly similar and are targeted by the same dsRNA. Staining of rRNA serves as loading control.

individual PIWI proteins (Piwi1-7/Ago3) prior to infection with SINV. Knockdown of Ago1 and Ago2 served as negative control. Using qPCR, we verified specific and efficient knockdown of at least 78% for all PIWI/AGO proteins (Figure 2B, Figure S1A-D). We then analyzed the levels of vpiRNAs by small RNA northern blot. Knockdown of Piwi5 and Ago3 resulted in substantial loss of vpiRNAs, while knockdown of the other PIWI proteins did not lead to apparent reduction of vpiRNA levels (Figure 2C, Figure S1E). As expected, knockdown of Ago1 or Ago2 likewise did not cause reduced vpiRNA accumulation (Figure 2D). These data identify Piwi5 and Ago3 as the first biogenesis factors for vpiRNA biogenesis in *Aedes aegypti*.

Piwi5 and Ago3 are required for vpiRNA biogenesis

Small RNA northern blotting is only suitable for the detection of highly abundant vpiRNAs which are, without exception, secondary piRNAs derived from the SINV (+) strand. To analyze the full repertoire of vpiRNAs, we prepared small RNA deep-sequencing libraries from SINV-infected Aag2 cells individually depleted of those PIWI proteins that are expressed in somatic tissues of adult mosquitoes and readily detectable in Aag2 cells (Piwi4, Piwi5, Piwi6, and Ago3). Knockdown of luciferase served as negative control. For each of these five conditions, three independent libraries were prepared and sequenced (Table S1). Confirming our northern blot results, knockdown of Piwi5 and Ago3 resulted in considerable reduction of vpiRNAs, whereas knockdown of Piwi4 or Piwi6 only mildly affected vpiRNA levels (Figure 3A, 3C). In general, the vast majority of (+) strand vpiRNAs mapped to the subgenomic region of SINV, suggesting that the viral subgenome is the predominant source of secondary vpiRNAs. In contrast, the low number of (-) strand vpiRNAs mapped across the viral genome without enrichment at specific hotspot regions, suggesting that the entire (-) strand serves as a source for vpiRNAs. While the number of vpiRNAs was reduced upon Piwi5 and Ago3 knockdown, the genomic distribution of vpiRNAs did not change upon knockdown of any of the PIWI proteins (Figure 3D).

Loss of vpiRNAs could be explained by a reduced biogenesis rate or by suppressed virus replication, which would limit the amount of substrate RNA. However, the number of viral siRNAs produced in the different knockdown conditions remained stable or was even slightly elevated, arguing against the second option (Figure 3B, Figure S2). To further confirm that the biogenesis of mature vpiRNAs is impaired in the absence of Piwi5 and Ago3 proteins, we analyzed the ping-pong signature of the remaining 25-30 nt small RNAs in the different knockdown conditions. Probing for 5′ end overlaps of sense and antisense small RNAs showed a strong reduction of read pairs with 10nt overlaps upon knockdown of Ago3 and Piwi5 (Figure 3E). In addition, the characteristic 1U and 10A nucleotide bias of respectively antisense and sense piRNAs was lost upon Ago3 and Piwi5 knockdown (Figure 3F). In contrast, these hallmarks of ping-pong amplification


A
vpiRNAs
+ strand
- strand
normalized read count
dsLuc
dsAgo3
dsPiwi4
dsPiwi5
dsPiwi6
B
vsiRNAs
C
dsLuc
dsAgo3
dsPiwi4
dsPiwi5
dsPiwi6
read length (nt)
D
genome coordinates (nt)
E
overlap propability
overlap of 5' ends (nt)
F
n=150,351
n=15,801
n=30,524
n=7,653
n=95,702
n=12,206
n=25,632
n=7,295
n=123,661
n=12,826
bits
nucleotide position

were retained in the absence of Piwi4 and Piwi6 (Figure 3E, 3F). Collectively, these data underscore the pivotal role of Piwi5 and Ago3 in ping-pong-dependent biogenesis of SINV-derived piRNAs.

Piwi5 and Ago3 bind piRNAs from opposite viral strands

We hypothesized that Piwi5 and Ago3 act as complementary partners of a ping-pong loop in *Aedes* mosquitoes. Such a model predicts that 1U-biased piRNAs derived from viral (-) strand would predominantly bind to one of the two PIWI proteins, whereas 10A-biased piRNAs from the (+) strand would associate with its counterpart (4, 5). To test this hypothesis, we designed expression vectors for Piwi4, Piwi5, Piwi6 and Ago3 N-terminally fused to V5-3xFlag tags. As a control, we generated a V5-3xFlag-tagged GFP vector. Of note, multiple attempts to clone the Piwi5 cDNA failed, and using rapid amplification of cDNA ends (5′ RACE) we revised the current gene-annotation (Figure S3).

We expressed the individual PIWI proteins in SINV-infected Aag2 cells and performed V5-ribonucleoprotein immunoprecipitation (IP) followed by vpiRNA northern blot. In line with our hypothesis, the 10A-biased vpiRNA sequences were enriched in Ago3 IP, but not in Piwi4-6 IPs (Figure 4A). These findings suggest that only Ago3 efficiently binds the highly abundant, (+) strand-derived vpiRNAs and that Piwi5, although required for their biogenesis, does not directly associate with this population of vpiRNAs. To analyze the PIWI association in more detail, we cloned and sequenced the small RNA fraction from Piwi4, Piwi5, Piwi6 and Ago3 IPs. As a control for non-specific binding, we sequenced small RNAs from a GFP-IP (Table S1). Efficient IP was shown by the depletion of the transgenic proteins in the supernatant after IP (Figure S4A). Confirming the northern blot analyses, (+) strand-derived vpiRNAs were strongly enriched in Ago3-IP only (Figure 4B, Figure S4B). Similar to vpiRNAs sequenced from total RNA, Ago3-bound piRNAs were predominantly derived from the hotspot region downstream of the SINV subgenomic promoter (Figure 4C). In line with our hypothesis, Piwi5-IP exclusively enriched piRNAs derived from the SINV (-) strand (Figure 4B, S4D), which mapped across the entire length of the viral antigenome (Figure 4C). The Piwi4-

◀ **Figure 3. Piwi5 and Ago3 are required for vpiRNA biogenesis. (A,B)** Number of 25-30 nt piRNA reads (A) and 21-nt siRNA reads (B) derived from the SINV (+) strand (black bars) and (-) strand (grey bars) in the indicated PIWI-protein knockdown libraries. Two-tailed student's t-test was used to determine statistical significance (*, $P<0.05$; **, $P<0.01$; ***, $P<0.001$: ****, $P<0.0001$). **(C)** Size profile of small RNAs mapping to the (+) strand (black bars) or the (-) strand (grey bars) of SINV. Bars in A-C are the mean +/- SEM of the three independent libraries. **(D)** Genome distribution of 25-30 nt small RNAs across the (+) strand (red) or (-) strand (blue) of the SINV genome. The average counts (three experiments) of the 5′ ends of the small RNA reads at each nucleotide position are shown. **(E)** The mean probability (n=3) for 5′ overlaps between viral piRNAs from opposite strands in the indicated knockdown libraries. **(F)** Nucleotide bias at each position in the 25-30 nt small RNA reads mapping to the SINV (+) strand (upper panels) and (-) strand (lower panels). All reads of three independent experiments were combined to generate the sequence logo; n, number of reads.

IP was not enriched for vpiRNAs (Figure 4B, Figure S4C) and Piwi6-IP was only mildly enriched for vpiRNAs from viral (-) strand (Figure 4B, 4C, Figure S4E).

Next, we analyzed the nucleotide bias of PIWI-protein associated vpiRNAs. To this end, we determined the fraction of 10A and 1U-containing vpiRNA reads in the PIWI-IPs that were enriched for vpiRNAs (Ago3, Piwi5 and Piwi6; Figure 4A, 4B). In the GFP

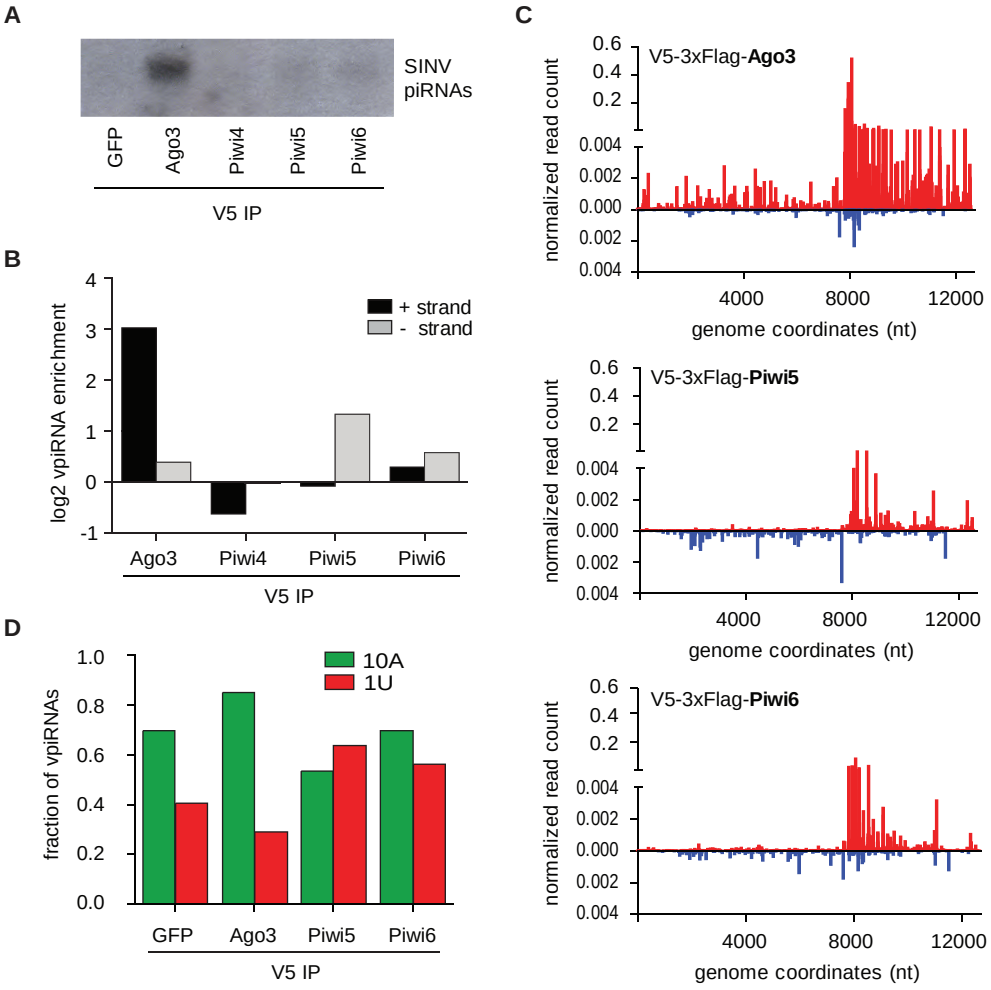


Figure 4. Association of vpiRNAs with individual PIWI proteins. (A) Northern blot analysis of vpiRNAs in RNA isolated from IPs of the indicated V5-epitope tagged proteins. Viral piRNAs were detected using a pool of the four probes presented in Figure 1A. **(B)** Enrichment of 25-30 nt small RNAs from the SINV (+) strand (black bars) or (-) strand (grey bars) in the IP of the indicated V5-epitope tagged PIWI proteins compared to the V5-tagged GFP-IP. **(C)** Distribution of 25-30 nt small RNAs in the indicated PIWI IPs across the (+) strand (red) or (-) strand (blue) of the SINV genome. Every data-point shows the number of reads at each nucleotide position normalized against the size of the libray (% of library). To account for background binding, the normalized read counts of the GFP-IP at each position were subtracted. **(D)** Fraction of 25-30 nt SINV-derived small RNA reads from the indicated deep-sequencing libraries that have an adenine at position 10 (10A; green bars) or uridine at position 1 (1U; red bars), respectively. No data for Piwi4 is shown in panel C and D since the V5-IP for this protein was not enriched for vpiRNAs.

control precipitation, 70% of the vpiRNA sequences had an adenine at position ten. This fraction increased to 85% in the Ago3-IP, but in none of the other PIWI-IPs (Figure 4D). Furthermore, the fraction of 1U-containing vpiRNAs declined from 40% in the GFP-IP to 29% in the Ago3-IP. Thus, parallel to raising the absolute number of (+) strand-derived vpiRNAs more than 8-fold (Figure 4B), Ago3-IP purified this population towards a stronger 10A nucleotide bias. In contrast, the Piwi5-IP was enriched for vpiRNAs with a uridine at position one (63%) and was depleted of 10A-containing sequences (53%), when compared to the control GFP-IP (Figure 4D). Piwi6-IP resulted in an enrichment of 1U-containing vpiRNAs (56%), which likely reflects the mild enrichment for (-) strand-derived vpiRNAs (Figure 4B). Altogether, these data formally classify the 25-30 nt SINV-derived small RNAs in Aag2 cells as *PIWI-interacting* RNAs. In addition, our findings show that in *Aedes aegypti*, Ago3 and Piwi5 are the complementary core proteins of the ping-pong loop, which is the dominant mechanism for vpiRNA synthesis in response to SINV infection. Piwi4 and Piwi6, if at all, only have a minor contribution to vpiRNA biogenesis.

Ago3 and Piwi5 co-localize with vpiRNAs in the cytoplasm

In the *Drosophila* germline, ping-pong amplification of piRNAs occurs in a non-membranous perinuclear structure in the cytoplasm, termed *nuage*. In mutant flies with defects in Aub and Ago3 localization to this region, piRNA amplification is disrupted (40, 41). Therefore, we analyzed the sub-cellular localization of 3xHA-tagged Piwi5 or Ago3 in Aag2 cells. Both proteins were diffusely expressed in the cytoplasm with only little expression in the nucleus (Figure 5A, 5C, Figure S5A, S5C). In some instances, we found perinuclear enrichment for both proteins, but this was minor compared to the clear, ring-like localization of Aub and Ago3 in the *Drosophila* germline (4). SINV infection did not alter the subcellular localization of Piwi5 and Ago3. Furthermore, Piwi5 and Ago3 did not accumulate at sites of dsRNA production in infected cells (Figure 5B, 5D, Figure S5B, S5D). The predominant expression of both Piwi5 and Ago3 in the cytoplasm was confirmed by western blotting after cytoplasmic and nuclear fractionation (Figure 5E). Thus, since SINV RNA replication occurs in the cytoplasm, viral RNAs and the vpiRNA core biogenesis factors are co-expressed in the cytoplasm. Indeed, the vast majority of vpiRNAs was also present in the cytoplasmic fraction (Figure 5F), suggesting that vpiRNA biogenesis occurs in the cytoplasm of infected Aag2 cells.

Differential association of virus and TE-derived piRNAs with *Aedes* PIWI proteins

The expansion of the PIWI protein family in *Aedes aegypti* may have allowed functional specialization of PIWI proteins in the biogenesis of piRNAs from different sources, such as viral or transposon RNA. To test this hypothesis, we catalogued the requirement for individual PIWI proteins in the production of TE-derived piRNAs. We analyzed the

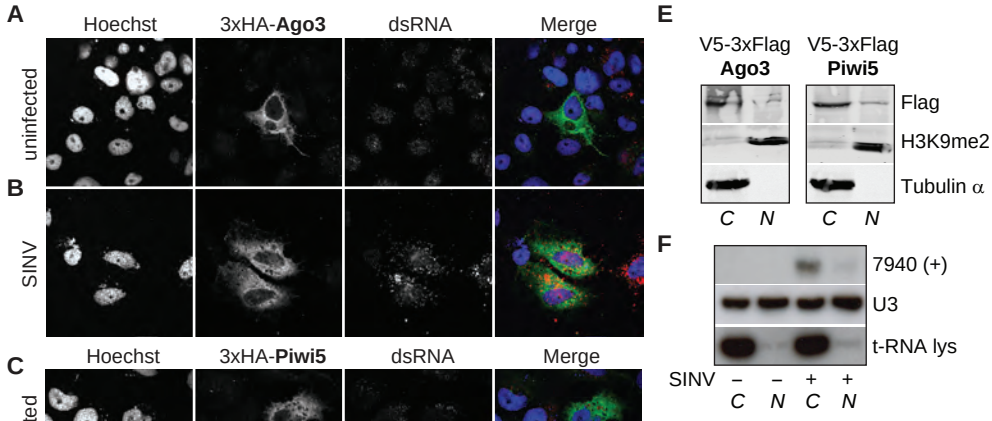


Figure 5. vpiRNA biogenesis occurs in the cytoplasm. Localization of 3xHA-tagged Ago3 **(A,B)** and Piwi5 **(C,D)** in uninfected (A,C) and SINV (parental virus) infected Aag2 cells (B,D) as determined by confocal microscopy. Hoechst staining indicates the nuclei. Infected cells were identified by a strong cytoplasmic dsRNA staining, which can be clearly distinguished from low-level background staining in non-infected cells. **(E)** Western blot analysis of V5-3xFlag-tagged Piwi5 and Ago3 in the cytoplasmic (*C*) and nuclear (*N*) fraction of Aag2 cell lysates. Tubulin and Histone 3 dimethylated at lysine 9 (H3K9me2) were used as cytoplasmic or nuclear markers, respectively. **(F)** Northern blot of vpiRNA 7940 (+) in nuclear and cytoplasmic fractions of Aag2 cells. U3 RNA and t-RNA lysine serve as nuclear and cytoplasmic marker, respectively.

repertoire of piRNAs that map to the annotated *Aedes aegypti* TE database (TEfam) upon PIWI protein knockdown. In line with previous observations (15), the vast majority of piRNAs was antisense to annotated TE sequences (Figure 6A, Figure S6A). Furthermore, antisense TE-derived piRNAs had a strong 1U bias, whereas sense piRNAs showed a 10A bias, indicating the existence of a ping-pong-dependent piRNA population (Figure S6B). However, whereas piRNA production from viral RNA was almost exclusively dependent on Ago3 and Piwi5, TE-derived piRNA levels were also decreased after Piwi4 depletion. Both upon knockdown of Piwi4 and, even more pronounced, upon knockdown of Piwi5 the number of antisense piRNAs was reduced. In contrast, Ago3 knockdown only mildly affected the levels of antisense TE-derived piRNAs, but caused the strongest reduction of sense strand piRNAs (Figure 6A). This suggests that, similar to the biogenesis of vpiRNAs, Ago3 might be directly involved in the production of (+) strand, 10A-biased TE-derived piRNAs. Indeed, when we analyzed the TE-derived piRNA populations in the different PIWI IPs, only the Ago3-IP was enriched for sense strand piRNAs. Strongest enrichment for antisense piRNAs, on the other hand, was observed in the Piwi5-IP (Figure 6B). Unexpectedly, although Piwi4 knockdown resulted

in a decline of TE-derived piRNAs, the Piwi4-IP was depleted of, rather than enriched for transposon piRNAs (Figure 6B). This indicates that Piwi4 binds to neither viral nor TE-derived piRNAs, suggesting that the observed reduction of transposon piRNAs upon Piwi4 knockdown is likely to be an indirect effect that requires further investigation. Interestingly, although Piwi6 knockdown did not reduce transposon piRNA levels, Piwi6-IP was enriched for transposon piRNAs, albeit to a lower extent than the Piwi5-IP. It is currently unclear why knockdown of Piwi6 did not alter global transposon piRNA levels. Taken together, these data suggest that the requirement for different PIWI proteins is broader for TE-derived piRNAs than for SINV-derived piRNAs, production of which is solely dependent on Piwi5 and Ago3.

piRNAs from individual TEs require different PIWI proteins for their biogenesis

Next, we analyzed the changes in piRNA levels for individual transposons upon knockdown of Piwi4-6 and Ago3. To classify transposons based on the PIWI proteins that medicate their piRNA biogenesis, we performed hierarchical clustering of the top 50 piRNA producing transposons. We identified four groups of transposons, based on the changes in piRNA abundance upon PIWI protein knockdown (Figure 6C). Group I and Group II transposons were characterized by a similar decrease of piRNAs upon knockdown of Piwi4 and Piwi5, but they differed in their dependence on Ago3 and Piwi6. Whereas piRNA biogenesis for group I transposons was reduced upon Piwi6 knockdown and not influenced by Ago3, group II transposons showed the opposite trend (Figure 6C, Figure S6C, S6D). Group III transposons clustered with SINV, suggesting that piRNA biogenesis from these TE sequences depends on a similar set of PIWI proteins as vpiRNAs. Indeed, group III transposon piRNAs were reduced to a similar extent upon Piwi5 and Ago3 knockdown, but they were less affected by Piwi4 and Piwi6 knockdown (Figure 6C, Figure S6E). This suggests that group III transposon piRNAs are, like vpiRNAs, produced in a ping-pong-dependent manner. Group IV is comprised of two transposons, which predominantly require Ago3 and Piwi4 for piRNA biogenesis (Figure 6C, Figure S6F).

We next analyzed the association of the piRNAs from the selected 50 transposons with the four PIWI proteins. Reflecting our analyses of the total TE-derived piRNA population, Piwi4-IP was depleted of piRNAs from all individual transposons, indicating that it does not directly bind mature piRNAs. Piwi5 and Piwi6 were enriched for piRNAs from all groups of transposons. Yet, piRNA enrichment is strongest for group I and group II transposons and only weak for group III and group IV transposons (Figure 6C, Figure S6G-S6J). Ago3-IP was enriched for piRNAs from group III and group IV transposons and an individual group II transposon (Ty3/Gypsy element 123). We noted that the piRNA population of group I and II TEs shows a strong antisense bias, whereas the piRNA population of group III and IV has a weaker antisense bias or

even a slight sense bias. To further analyze this correlation, we sorted the transposons according to their antisense bias and performed a sliding window analysis on this ranking. Confirming our previous observations, Ago3 knockdown resulted in the strongest reduction of piRNA levels for transposons that have a sense or weak antisense bias and Ago3 dependence decreased with increasing antisense bias (Figure 6D). In line with these observations, Ago3-IP was only enriched for piRNAs from transposons that have strong sense bias (Figure 6E). Piwi5 knockdown generally had the biggest impact on piRNA levels, except for the transposons with the strongest sense bias (Figure 6D). Piwi6 knockdown primarily reduced piRNA levels of transposons with strong antisense bias, although the effect was minor compared to Piwi4 and Piwi5 knockdown (Figure 6D). Yet, Piwi6-IP was enriched for transposon piRNAs to a similar extent as Piwi5-IP, and both IPs tended to be more enriched for piRNAs from transposons with a strong antisense bias (Figure 6E). Since Piwi6-IP was almost not enriched for piRNAs of viral origin (Figure 4B), these data suggest that Piwi6 binds more specifically to piRNAs derived from selected transposons. Thus, whereas SINV piRNAs are almost exclusively produced via ping-pong amplification by Piwi5 and Ago3, TE-derived piRNA biogenesis directly or indirectly requires the activity of all analyzed PIWI proteins.

DISCUSSION

Like in other invertebrates, recognition of viral dsRNA and its processing into vsiRNAs is key to antiviral immunity in mosquitoes (42). Yet, the recent discovery of vpiRNAs has challenged the idea that vsiRNAs are the sole small RNA species produced from viral RNA. Whereas the biogenesis of vsiRNAs is well-characterized in mosquitoes and fruit flies, little is known about the molecular mechanisms of vpiRNA production. The only cues come from the typical piRNA sequence signature that suggests a biogenesis pathway that includes ping-pong amplification (14, 15, 21-23).

Ping-pong amplification has previously been postulated for the production of TE-derived piRNAs in the fly (4, 5). However, ping-pong-dependent piRNAs of viral origin have hitherto only been detected in mosquitoes and mosquito cells. In the fly, piRNA-sized viral small RNAs have been described in persistently infected ovarian somatic sheet (OSS) cells. These cells, however, are deficient of the secondary piRNA biogenesis factors Aub and Ago3 (11) and therefore vpiRNAs from OSS lack the ping-pong signature (43).

▶ **(C)** Relative changes of the top 50 transposons upon PIWI protein knockdown and IP. Left panel: Heat map showing the relative piRNA abundance in the indicated knockdown libraries compared to the control knockdown (dsLuc). These data were used to generate the hierarchical clustering. Middle panel: Heat map showing the relative piRNA abundance in the indicated IP libraries over the control IP (GFP). Right panel: antisense bias, defined as the percentage of antisense 25-30 nt reads that uniquely map to the individual transposon sequences. **(D,E)** The 50 transposons from panel C were ranked according to their antisense bias. The mean relative piRNA abundance (log2-transformed) for five consecutive transposons is plotted with an offset of one rank number for the indicated knockdown libraries (D) or IP libraries (E). The corresponding antisense bias is indicated with the dashed line.

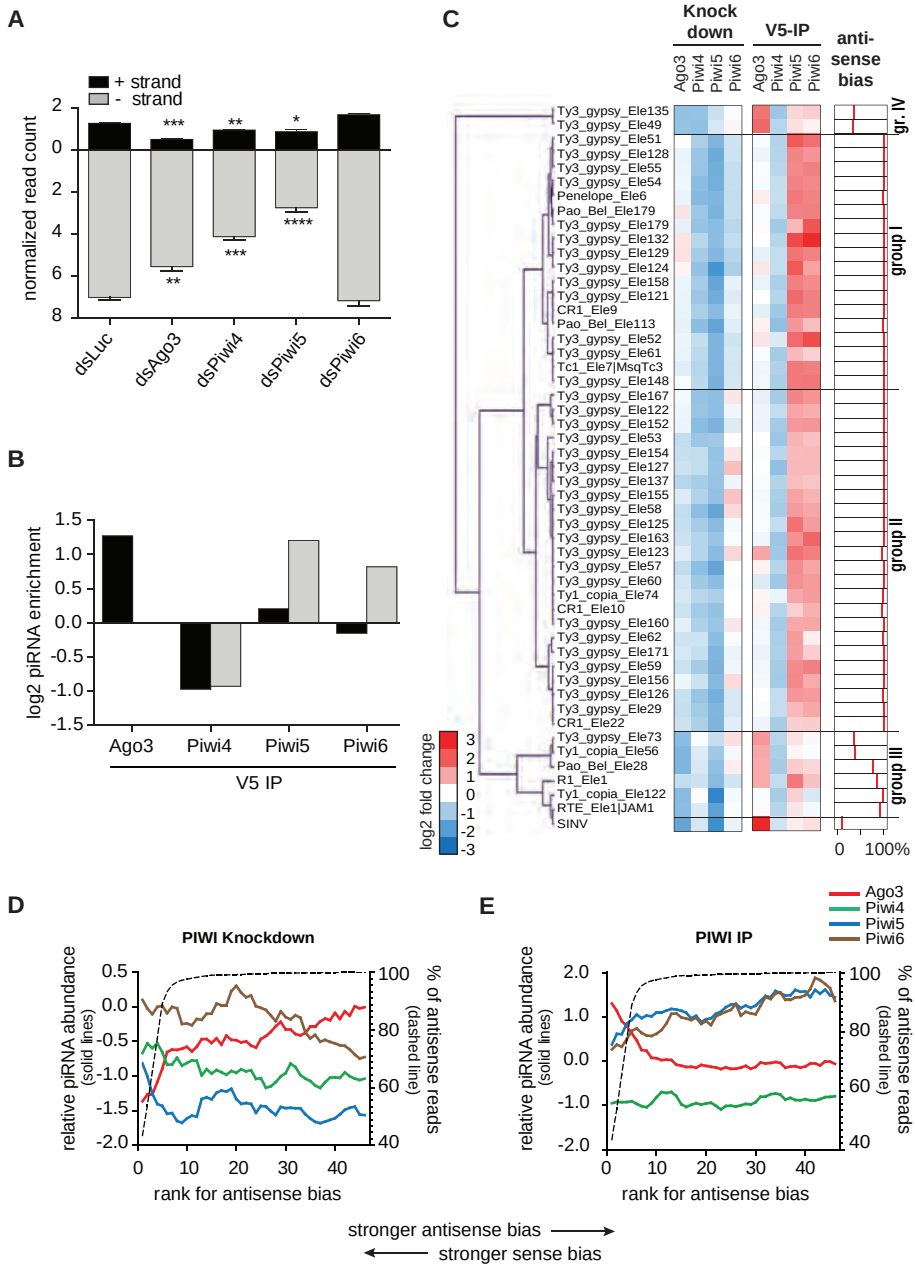


Figure 6. Association of TE-derived piRNAs with different PIWI-proteins. (A) Normalized read counts of 25-30 nt reads from the different knockdown libraries mapping to the TEfam transposon database. The mean +/- SEM of three independent libraries are shown. Two-tailed student's t-test was used to determine statistical significance (*, $P<0.05$; **, $P<0.01$; ***, $P<0.001$: ****, $P<0.0001$). **(B)** Enrichment of 25-30 nt reads in the V5-IP of the indicated PIWI proteins compared to the GFP-IP. The number of reads from the (+) strand (black bars) or (-) strand (grey bars) in panel A and B was normalized to the corresponding library size.

In adult flies, PIWI proteins do not appear to be expressed in somatic tissues (4, 5) and thus far no vpiRNA-like molecules have been identified in small RNA libraries of virus-infected flies. In sharp contrast, PIWI proteins are expressed in somatic cells of *Aedes aegypti* mosquitoes and secondary piRNAs can readily be detected outside the germline (14). Since most arboviruses exclusively infect somatic tissues and are not transmitted through the germline, it is likely that somatic PIWI expression has favoured viral RNA as a new substrate for piRNA biogenesis.

Aedes aegypti Aag2 cells are competent in producing ping-pong-dependent vpiRNAs that have strikingly similar sequence features as vpiRNAs found in adult mosquitoes (14). Using this cell culture model we show that Ago3 and Piwi5 engage in a ping-pong amplification loop in which each of them binds vpiRNAs derived from opposite viral strands. Piwi5 predominantly binds 1U-biased, antisense piRNAs, whereas Ago3 preferentially associates with 10A-biased sense piRNAs, reflecting the nucleotide signature found for TE-derived piRNAs bound to *Drosophila* Aub and Ago3, respectively (4, 5). These findings formally classify vpiRNAs as *PIWI-interacting* RNAs. Somatic cells in adult *Aedes* mosquitoes express a strikingly similar set of PIWI proteins as Aag2 cells with only low expression of Piwi1-3 and Piwi7 (25). Piwi1 and Piwi3 are highly expressed specifically in the ovaries, a tissue that is generally not infected by SINV (44, 45). Piwi7 is only expressed in the early embryo (25) and is therefore unlikely to contribute to the biogenesis of arbovirus-derived piRNAs. Thus, it is very likely that similar mechanisms are responsible for the production of SINV-derived piRNAs in Aag2 cells and adult mosquitoes.

The vast majority of vpiRNAs derives from the SINV (+) strand, has a 10A nucleotide bias, and is associated with Ago3. Yet, the number of (+) strand, 10A-biased vpiRNAs is also strongly reduced upon Piwi5 knockdown. This is in line with the ping-pong model in which one PIWI protein generates the piRNA precursor for the other one. In the *Drosophila* germline, loss-of-function of Ago3 similarly eliminates the Aub-bound, antisense transposon-derived piRNA population (12). During SINV infection, vpiRNAs derived from the viral (-) strand accumulate to much lower levels, most likely because antigenomic RNA itself is scarce. Nevertheless, upon knockdown of Ago3 the number of antisense vpiRNAs declines even further, suggesting that in Aag2 cells the ping-pong loop is a full circle with both Ago3 and Piwi5 producing the piRNA precursors for each other.

It remains to be explained what determines the strand bias of Ago3-bound and Piwi5-bound vpiRNAs. In *Bombyx mori* Bmn4 cells, the MID-PIWI module of the PIWI proteins Siwi and Ago3 determine the strand bias of the associated piRNAs (46). The authors propose that the primary piRNA transcripts contain features that mark their nuclear origin and sort these precursors into Siwi based on the structure of the MID-PIWI domains. Since these transcripts tend to be antisense to transposon mRNAs, the

nuclear origin of the piRNA precursor would dictate the strand bias of Siwi-associated piRNAs (46). Although this is an attractive model for transposon-derived piRNAs, it is unlikely to explain the strand bias of vpiRNAs, as it demands a nuclear component of the biogenesis pathway. We envision that vpiRNA production is a purely cytoplasmic event because SINV RNAs generally do not enter the nucleus. Thus, additional features must exist that sort piRNAs from the viral sense and antisense strands into Ago3 and Piwi5, respectively. The nature of such features is currently unknown. Likewise, it is not understood what discriminates the viral single-stranded RNA, which serves as piRNA precursor, from other abundant cellular mRNAs. Whereas dsRNA serves as an explicit non-self signal for the siRNA pathway, no such signal is known for the piRNA pathway.

Aedes aegypti is not a natural host for SINV, which is transmitted by *Culex* mosquitoes in the wild. To date, there is no conclusive data on whether *Culex* mosquitoes or cells derived from these animals produce Alphavirus-derived piRNAs. Yet, *Aedes* mosquitoes transmit chikungunya virus (CHIKV), which belongs to the same virus family as SINV. Interestingly SINV and CHIKV produce ping-pong-dependent vpiRNAs with strikingly similar sequence features and genome distribution (14). The same is true for Semliki Forest virus, another member of the Alphavirus family (21) and probably CHIKV and SFV piRNA biogenesis relies on a similar, if not identical, molecular machinery as SINV. It is likely that specific features, common to Alphaviruses, are recognized by the piRNA biogenesis machinery and make the viral RNA a favourable piRNA substrate. These features must be independent of primary nucleotide sequence, since SINV, CHIKV, and SFV only share little sequence similarity. Outside of the Alphaviruses, vpiRNAs with ping-pong signature have been shown for La Crosse virus (15), Rift Valley fever virus (22) and Schmallenberg virus (23), all of which belong to the *Bunyaviridae* family. In RNAi-deficient C6/36 cells, vpiRNAs from dengue virus, a Flavivirus, have been proposed based on the small RNA size range and a 10A bias, but no 1U was detected (47). Additional studies did not detect dengue virus-derived piRNA-sized small RNAs with the characteristic ping-pong signature (48, 49). Future research will have to establish which viruses produce vpiRNAs and if the piRNA biogenesis mechanism is similar to the one described here.

The *Aedes aegypti* genome is remarkably rich in transposons (50), which are the dominant substrate for piRNAs in all studied model organisms. In *Aedes* mosquitoes, the diversification of the PIWI family may have facilitated the recognition of novel RNA substrates and even functional specialization of PIWI proteins in producing piRNAs from various RNA sources. Indeed, in Aag2 cells the biogenesis of SINV-derived piRNAs is abrogated specifically upon knockdown of Piwi5 or Ago3, but not Piwi4 or Piwi6. Knockdown of Piwi5 also causes a reduction in TE-derived piRNA levels for the vast majority of transposons, suggesting that it is essential for the biogenesis of both virus- and TE-derived piRNAs. Ago3 however, whereas crucial for vpiRNA biogenesis, is only

relevant for piRNA production of transposons whose piRNAs are weakly antisense or sense biased. Thus, Ago3 may be dispensable for the biogenesis of primary piRNAs, an observation that needs validation in a full genetic Ago3 knockout. Interestingly, although nonessential for vpiRNA biogenesis, Piwi4 and Piwi6 do play a role in the production of piRNAs derived from a number of different TEs, suggesting functional specialization of PIWI proteins. Similar to Piwi5, Piwi6 associates with antisense piRNAs derived from a large number of transposon. Yet, Piwi6 knockdown does not greatly affect TE-piRNA levels. Thus far, the reason for this apparent contradiction is unknown. It may be explained by a dominant role of Piwi5 in binding (-) strand piRNAs, thereby veiling the effect of Piwi6 knockdown.

Amongst all the PIWI family members analyzed, Piwi4 did not directly bind piRNAs of either viral or transposon origin. In line with this observation, knockdown of Piwi4 results in a negligible decrease of SINV piRNA levels, which has previously been noted for a related virus (21). Interestingly, although devoid of piRNA binding capacity, knockdown of Piwi4 results in decreased TE-derived piRNA levels. This suggests that Piwi4 indirectly influences the production of transposon, but not SINV-derived piRNAs, by either modulating the activity of piRNA biogenesis factors or by influencing the amount of available substrate that could feed into the piRNA pathway. To our knowledge, the data presented here is the first example of functional specialization of PIWI proteins in producing piRNAs from endogenous or exogenous sources.

ACKNOWLEDGEMENTS

We thank the members of the Van Rij lab for fruitful discussions, Inge C.E. Smeets and Joep Joosten for technical support and Bas E. Dutilh for bioinformatic advice. Furthermore, we thank Christophe Antoniewski, the ARTbio bioinformatics analyses platform of the Institut de Biologie Paris Seine and the http://mississippi.fr Galaxy server for sharing their bioinformatics tools.

This work was supported by an ECHO project grant from the Netherlands organization for scientific research [NWO; grant number 711.013.001]; and an European Research Council consolidator grant under the European Union's Seventh Framework Programme [ERC grant number 615680] to RPvR and by a PhD fellowship from the Radboud University Medical Centre to PM. Funding for open access charge: NWO.

REFERENCES

1. Ghildiyal M, Zamore PD. Small silencing RNAs: an expanding universe. Nat Rev Genet. 2009;10(2):94-108.
2. Cenik ES, Zamore PD. Argonaute proteins. Curr Biol. 2011;21(12):R446-9.
3. Orban TI, Izaurralde E. Decay of mRNAs targeted by RISC requires XRN1, the Ski complex, and the exosome. RNA. 2005;11(4):459-69.

4. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, et al. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell. 2007;128(6):1089-103.
5. Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, et al. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. Science. 2007;315(5818):1587-90.
6. Houwing S, Berezikov E, Ketting RF. Zili is required for germ cell differentiation and meiosis in zebrafish. EMBO J. 2008;27(20):2702-11.
7. Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ. Developmentally regulated piRNA clusters implicate MILI in transposon control. Science. 2007;316(5825):744-7.
8. Luteijn MJ, Ketting RF. PIWI-interacting RNAs: from generation to transgenerational epigenetics. Nat Rev Genet. 2013;14(8):523-34.
9. Iwasaki YW, Siomi MC, Siomi H. PIWI-Interacting RNA: Its Biogenesis and Functions. Annu Rev Biochem. 2015;84:405-33.
10. Ishizu H, Siomi H, Siomi MC. Biology of PIWI-interacting RNAs: new insights into biogenesis and function inside and outside of germlines. Genes Dev. 2012;26(21):2361-73.
11. Malone CD, Brennecke J, Dus M, Stark A, McCombie WR, Sachidanandam R, et al. Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. Cell. 2009;137(3):522-35.
12. Li C, Vagin VV, Lee S, Xu J, Ma S, Xi H, et al. Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. Cell. 2009;137(3):509-21.
13. Ross RJ, Weiner MM, Lin H. PIWI proteins and PIWI-interacting RNAs in the soma. Nature. 2014;505(7483):353-9.
14. Morazzani EM, Wiley MR, Murreddu MG, Adelman ZN, Myles KM. Production of virus-derived ping-pong-dependent piRNA-like small RNAs in the mosquito soma. PLoS Pathog. 2012;8(1):e1002470.
15. Vodovar N, Bronkhorst AW, van Cleef KW, Miesen P, Blanc H, van Rij RP, et al. Arbovirus-derived piRNAs exhibit a ping-pong signature in mosquito cells. PLoS One. 2012;7(1):e30861.
16. Yan Z, Hu HY, Jiang X, Maierhofer V, Neb E, He L, et al. Widespread expression of piRNA-like molecules in somatic tissues. Nucleic Acids Res. 2011;39(15):6596-607.
17. Lambrechts L, Scott TW. Mode of transmission and the evolution of arbovirus virulence in mosquito vectors. Proc Biol Sci. 2009;276(1660):1369-78.
18. Bronkhorst AW, van Rij RP. The long and short of antiviral defense: small RNA-based immunity in insects. Curr Opin Virol. 2014;7C:19-28.
19. Ding SW, Voinnet O. Antiviral immunity directed by small RNAs. Cell. 2007;130(3):413-26.
20. Weber F, Wagner V, Rasmussen SB, Hartmann R, Paludan SR. Double-stranded RNA is produced by positive-strand RNA viruses and DNA viruses but not in detectable amounts by negative-strand RNA viruses. J Virol. 2006;80(10):5059-64.
21. Schnettler E, Donald CL, Human S, Watson M, Siu RW, McFarlane M, et al. Knockdown of piRNA pathway proteins results in enhanced Semliki Forest virus production in mosquito cells. J Gen Virol. 2013;94(Pt 7):1680-9.
22. Leger P, Lara E, Jagla B, Sismeiro O, Mansuroglu Z, Coppee JY, et al. Dicer-2- and Piwi-mediated RNA interference in Rift Valley fever virus-infected mosquito cells. J Virol. 2013;87(3):1631-48.
23. Schnettler E, Ratinier M, Watson M, Shaw AE, McFarlane M, Varela M, et al. RNA interference targets arbovirus replication in Culicoides cells. J Virol. 2013;87(5):2441-54.
24. Campbell CL, Black WCt, Hess AM, Foy BD. Comparative genomics of small RNA regulatory pathway components in vector mosquitoes. BMC Genomics. 2008;9:425.
25. Akbari OS, Antoshechkin I, Amrhein H, Williams B, Diloreto R, Sandler J, et al. The developmental transcriptome of the mosquito Aedes aegypti, an invasive species and major arbovirus vector. G3 (Bethesda). 2013;3(9):1493-509.

26. Pall GS, Hamilton AJ. Improved northern blot method for enhanced detection of small RNA. Nat Protoc. 2008;3(6):1077-84.
27. van Cleef KW, van Mierlo JT, Miesen P, Overheul GJ, Fros JJ, Schuster S, et al. Mosquito and Drosophila entomobirnaviruses suppress dsRNA- and siRNA-induced RNAi. Nucleic Acids Res. 2014;42(13):8732-44.
28. Blankenberg D, Gordon A, Von Kuster G, Coraor N, Taylor J, Nekrutenko A, et al. Manipulation of FASTQ data with Galaxy. Bioinformatics. 2010;26(14):1783-5.
29. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25.
30. Brennecke J, Malone CD, Aravin AA, Sachidanandam R, Stark A, Hannon GJ. An epigenetic role for maternally inherited piRNAs in transposon silencing. Science. 2008;322(5906):1387-92.
31. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004;14(6):1188-90.
32. Schneider TD, Stephens RM. Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. 1990;18(20):6097-100.
33. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A. 1998;95(25):14863-8.
34. Brackney DE, Scott JC, Sagawa F, Woodward JE, Miller NA, Schilkey FD, et al. C6/36 Aedes albopictus cells have a dysfunctional antiviral RNA interference response. PLoS Negl Trop Dis. 2010;4(10):e856.
35. Kawaoka S, Katsuma S, Tomari Y. Making piRNAs in vitro. Methods Mol Biol. 2014;1093:35-46.
36. Horwich MD, Li C, Matranga C, Vagin V, Farley G, Wang P, et al. The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. Curr Biol. 2007;17(14):1265-72.
37. Saito K, Sakaguchi Y, Suzuki T, Suzuki T, Siomi H, Siomi MC. Pimet, the Drosophila homolog of HEN1, mediates 2'-O-methylation of Piwi- interacting RNAs at their 3' ends. Genes Dev. 2007;21(13):1603-8.
38. Tomari Y, Du T, Zamore PD. Sorting of Drosophila small silencing RNAs. Cell. 2007;130(2):299-308.
39. Okamura K, Ishizuka A, Siomi H, Siomi MC. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. Genes Dev. 2004;18(14):1655-66.
40. Anand A, Kai T. The tudor domain protein kumo is required to assemble the nuage and to generate germline piRNAs in Drosophila. EMBO J. 2012;31(4):870-82.
41. Zhang Z, Xu J, Koppetsch BS, Wang J, Tipping C, Ma S, et al. Heterotypic piRNA Ping-Pong requires qin, a protein with both E3 ligase and Tudor domains. Mol Cell. 2011;44(4):572-84.
42. Blair CD, Olson KE. The role of RNA interference (RNAi) in arbovirus-vector interactions. Viruses. 2015;7(2):820-43.
43. Wu Q, Luo Y, Lu R, Lau N, Lai EC, Li WX, et al. Virus discovery by deep sequencing and assembly of virus-derived small silencing RNAs. Proc Natl Acad Sci U S A. 2010;107(4):1606-11.
44. Jackson AC, Bowen JC, Downe AE. Experimental infection of Aedes aegypti (Diptera: Culicidae) by the oral route with Sindbis virus. J Med Entomol. 1993;30(2):332-7.
45. Bowers DF, Abell BA, Brown DT. Replication and tissue tropism of the alphavirus Sindbis in the mosquito Aedes albopictus. Virology. 1995;212(1):1-12.
46. Cora E, Pandey RR, Xiol J, Taylor J, Sachidanandam R, McCarthy AA, et al. The MID-PIWI module of Piwi proteins specifies nucleotide- and strand-biases of piRNAs. RNA. 2014;20(6):773-81.
47. Scott JC, Brackney DE, Campbell CL, Bondu-Hawkins V, Hjelle B, Ebel GD, et al. Comparison of dengue virus type 2-specific small RNAs from RNA interference-competent and -incompetent mosquito cells. PLoS Negl Trop Dis. 2010;4(10):e848.
48. Hess AM, Prasad AN, Ptitsyn A, Ebel GD, Olson KE, Barbacioru C, et al. Small RNA profiling of Dengue virus-mosquito interactions implicates the PIWI RNA pathway in anti-viral defense. BMC Microbiol. 2011;11:45.

49. Schirtzinger EE, Andrade CC, Devitt N, Ramaraj T, Jacobi JL, Schilkey F, et al. Repertoire of virus-derived small RNAs produced by mosquito and mammalian cells in response to dengue virus infection. Virology. 2014;476C:54-60.

50. Nene V, Wortman JR, Lawson D, Haas B, Kodira C, Tu ZJ, et al. Genome sequence of Aedes aegypti, a major arbovirus vector. Science. 2007;316(5832):1718-23.

SUPPPLEMENTAL FIGURES

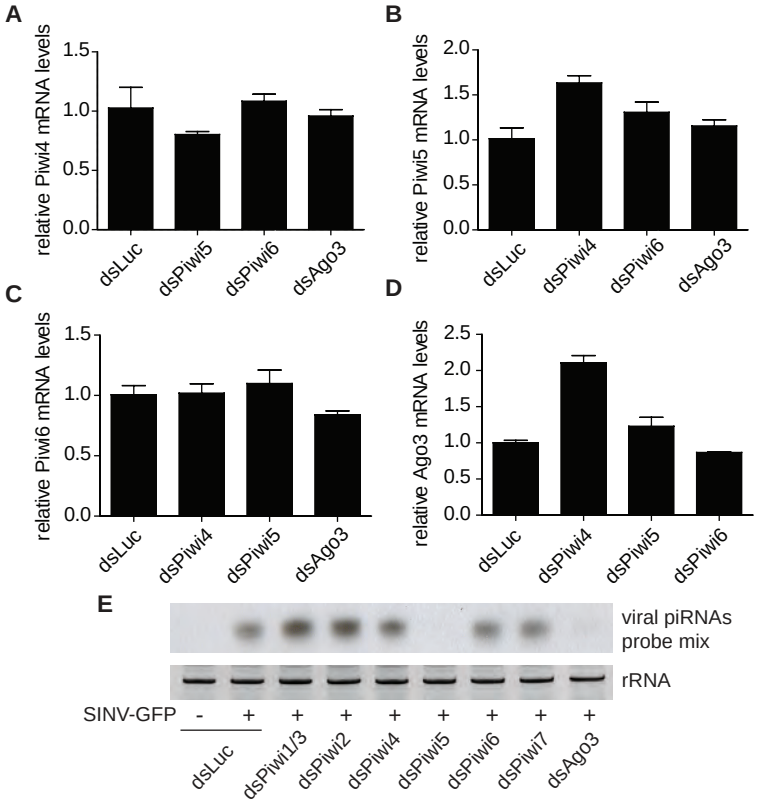


Figure S1. Specificity of PIWI knockdown. (A-D) qPCR for Piwi4, Piwi5, Piwi6 and Ago3 transcripts, respectively, in Aag2 cells transfected with the indicated dsRNA. mRNA levels were normalized to the dsLuc control knockdown; bars represent the mean +/- SEM of three independent experiments. **(E)** Small RNA northern blot for SINV piRNAs in Aag2 cells transfected with the indicated dsRNA. A pool of four probes (individual probes shown in Figure 1A) was used to detect vpiRNAs. The rRNA loading control is identical to the one shown in Figure 2C, since the same membrane was subsequently hybridized to the individual 8040 (+) probe and the vpiRNA probe mix after harsh stripping.

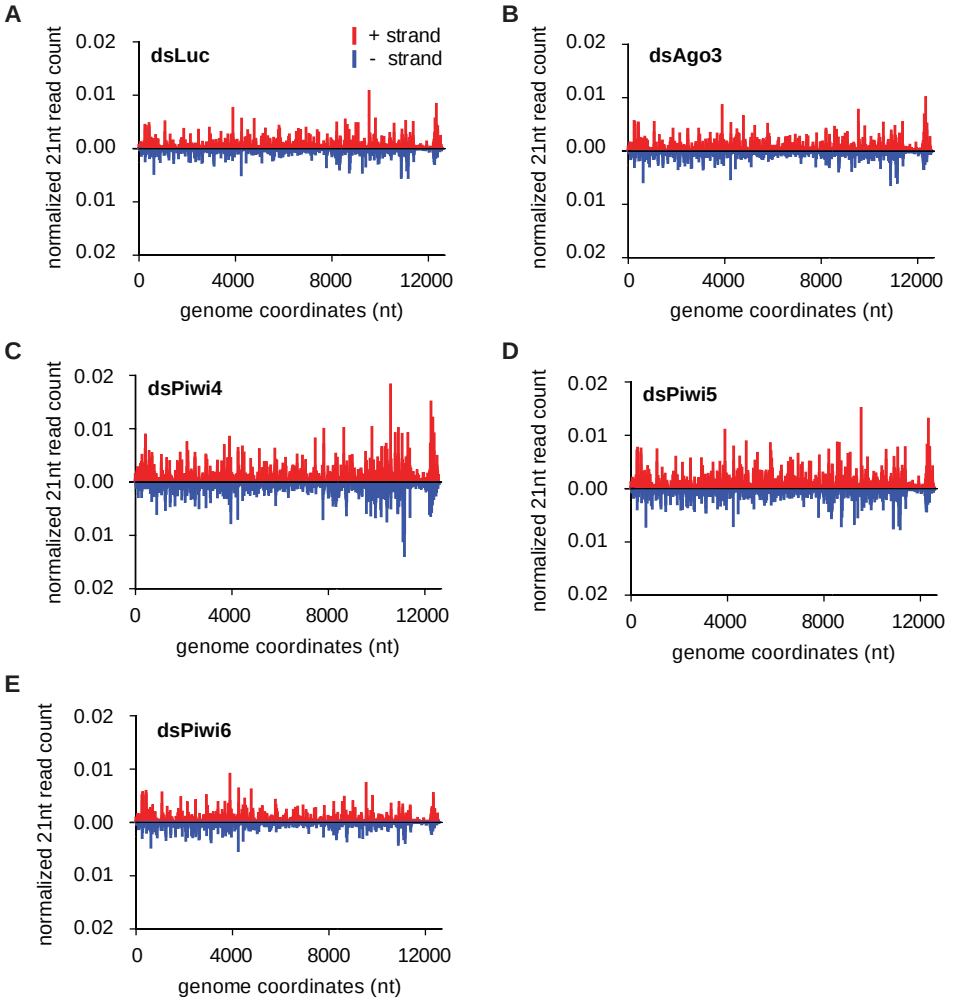


Figure S2. Levels of vsiRNAs remain stable upon PIWI knockdown. (A-E) Genome distribution of 21 nt vsiRNAs across the (+) strand (red) or (-) strand (blue) of the SINV genome. The average counts (*n*=3) of the 5′ ends of the small RNA reads at each nucleotide position are shown. The read number was normalized to the corresponding library size. The low read counts around position 12,000 coincides with the position of the GFP transgene, which is occasionally lost from the recombinant virus genome during virus replication.

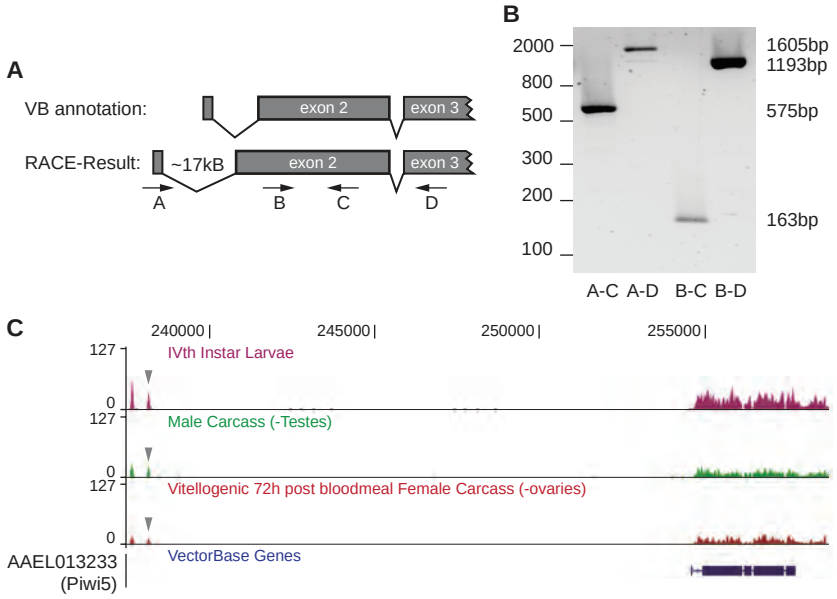


Figure S3. Revised gene annotation of Piwi5. (A) Schematic representation of the currently annotated gene topology of Piwi5 exon 1 to exon 3 published in Vectorbase (VB; release April 2015) and the revised gene structure as determined by 5′ RACE on RNA isolated from Aag2 cells (exons and introns are not drawn to scale). The entire first annotated exon is not included in the Piwi5 mRNA. Instead, a previously un-annotated 108nt exon was found approximately 17 kB upstream of exon 2. Furthermore, the second exon includes 223 additional nucleotides at its 5′ end. **(B)** Reverse-transcriptase PCR on RNA prepared from Aag2 cells. The positions of the primers are indicated by the arrows in panel A. A size marker is indicated on the left-hand side of the gel picture. Numbers on the right indicate the expected size of the PCR products. **(C)** Genome browser shots from *Aedes aegypti* supercontig 1.809 (www.vector.caltech.edu/). Numbers on top indicate the nucleotide position of the genomic scaffold. RNA sequencing data from the indicated tissue/developmental stage, as well as the annotated Piwi5 gene topology from Vectorbase are shown. The grey arrowheads indicate a putative, additional exon that we did not detect in our RACE analyses and subsequent cloning experiments.

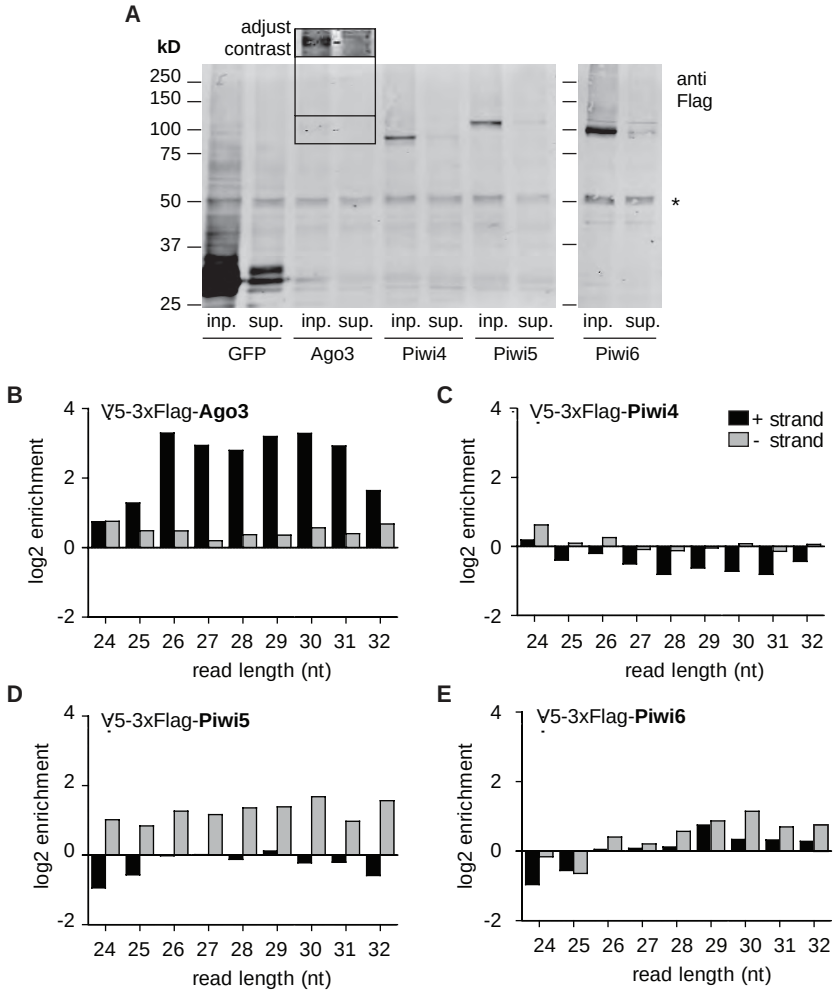


Figure S4. Fold enrichment of vpiRNAs in PIWI IPs. (A) Anti-Flag western blot for V5-3xFlag tagged PIWI proteins (size range 104kD-112kD) or V5-3xFlag-GFP (32.5kD). Protein expression is analyzed in 5% of the input (inp.) samples before V5-IP and 5% of the supernatant (sup.) after IP. * indicates a non-specific protein band. Linear contrast adjustment was used in Adobe Photoshop to enhance the signal for V5-3xFlag-Ago3, which was only lowly expressed. **(B-E)** Enrichment of small RNA reads in the V5-IPs for the indicated PIWI proteins over the control IP (V5-3xFlag-GFP IP). Log2-transformed enrichment scores of reads that map the SINV (+) strand and (-) strand are indicated in black and grey, respectively.

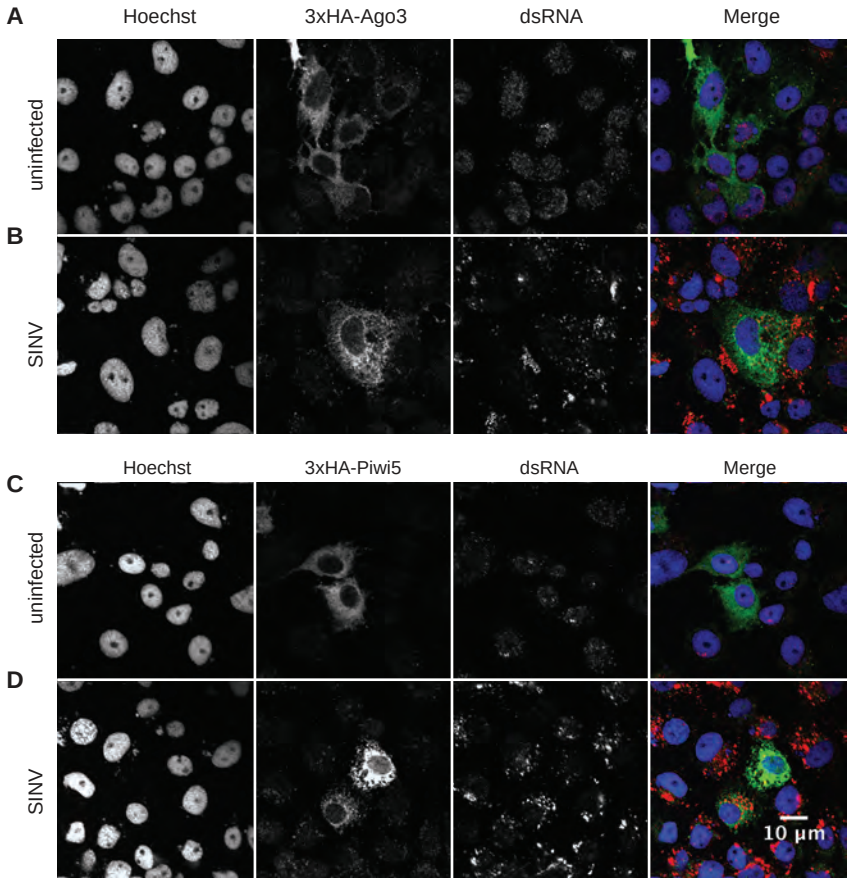


Figure S5. Ago3 and Piwi5 sub-cellular localization. Additional images of the experiment shown in Figure 5. The localization of 3xHA-tagged Ago3 **(A,B)** or Piwi5 **(C,D)** in uninfected (A,C) or SINV (parental virus) infected (B,D) Aag2 cells was determined by confocal immunofluorescence analysis. Hoechst staining was used to stain the nuclei. Intracellular dsRNA staining was used to identify infected cells as described in the legend to Figure 5. Scale bar indicates 10µm.

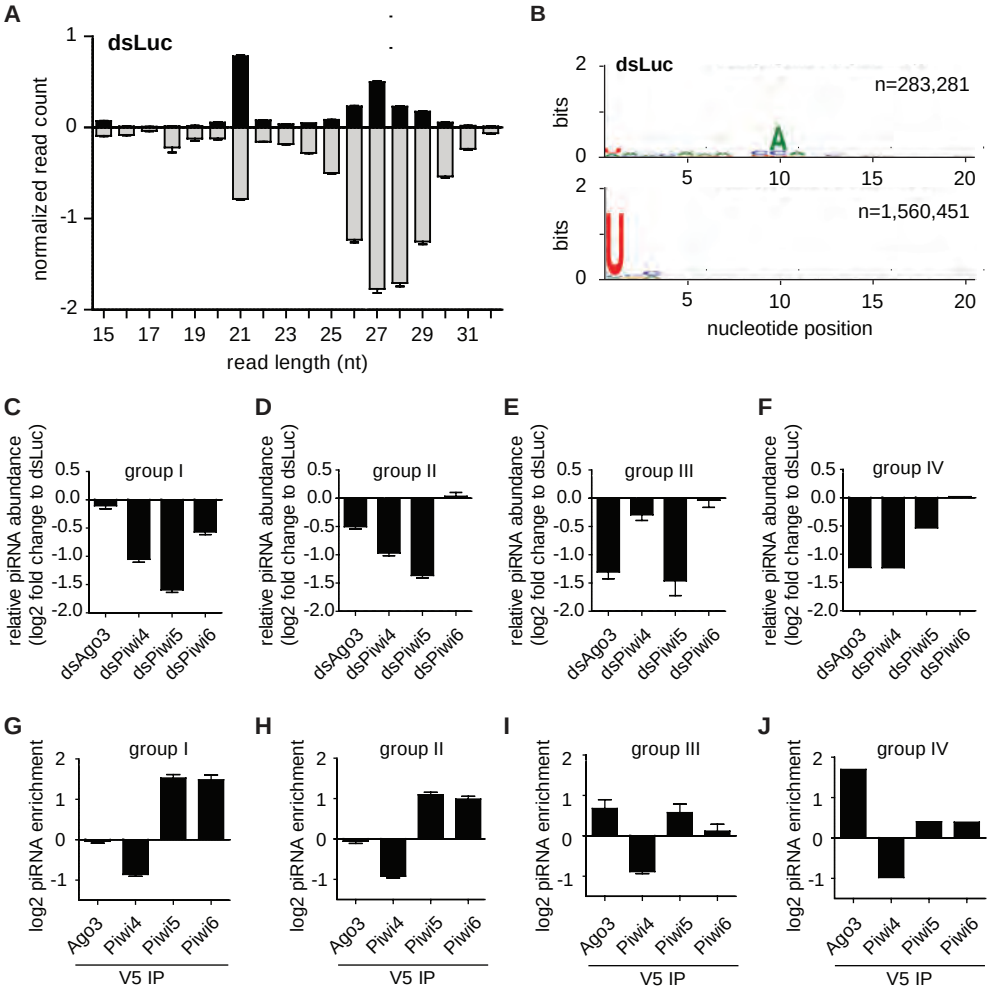


Figure S6. Groups of transposons can be classified based on their dependence on different PIWI proteins for piRNA biogenesis. (A) Size profile of small RNA reads from the dsLuc libraries that map to the collection of *Aedes aegypti* TE sequences published in the TEfam transposon database. Read counts were normalized to the corresponding library sizes. The average and SEM of the three independent libraries are shown. **(B)** Nucleotide bias at each position in the 25-30nt small RNA reads from the dsLuc libraries that map to TE sense strands (upper panel) or antisense strands (lower panel). All reads of the three independent libraries were combined to generate the sequence logo; *n*, number of reads. **(C-F)** Relative abundance of 25-30nt reads in the indicated knockdown libraries compared to the dsLuc control libraries for group I (C), group II (D), group III (E) and group IV (F) transposons. **(G-J)** Enrichment of 25-30nt reads in the indicated IPs over the GFP control IP (for group I (G), group II (H), group III (I) and group IV (J) transposons. Panels C-J represent the mean of all transposons belonging to the corresponding group; error bars indicate the SEM. For a definition of group I to IV transposons, please consult Figure 6.

Table S1. Basic characteristics of the small RNA deep sequencing libraries.
For each category the number of reads is indicated. The numbers in brackets indicate the percentage of the total library size.

Library name	Total Library Size	SINV-GFP mappers	SINV-GFP mappers (25-30 nt)	TEfam mappers	TEfam mappers (25-30 nt)
Knockdown libraries					
dsLuc-1	8,037,253	411,612 (5.1%)	62,296 (0.8%)	913,974 (11.4%)	638,343 (7.9%)
dsLuc-2	8,250,209	442,710 (5.4%)	60,718 (0.7%)	995,007 (12.1%)	687,766 (8.3%)
dsLuc-3	6,071,409	229,474 (3.8%)	43,138 (0.7%)	729,563 (12.0%)	517,623 (8.5%)
dsAgo3-1	7,068,908	394,776 (5.6%)	12,497 (0.2%)	669,293 (9.5%)	403,034 (5.7%)
dsAgo3-2	7,904,410	437,717 (5.5%)	13,786 (0.2%)	767,388 (9.7%)	463,596 (5.8%)
dsAgo3-3	4,330,827	173,706 (4.0%)	11,894 (0.3%)	450,046 (10.4%)	283,516 (6.5%)
dsPiwi4-1	7,015,754	581,753 (8.3%)	44,936 (0.6%)	755,696 (10.8%)	366,409 (5.2%)
dsPiwi4-2	5,380,559	522,716 (9.7%)	30,809 (0.6%)	588,755 (10.9%)	248,328 (4.6%)
dsPiwi4-3	7,935,701	607,975 (7.7%)	32,163 (0.4%)	898,310 (11.3%)	423,123 (5.3%)
dsPiwi5-1	6,544,842	498,693 (7.6%)	9,524 (0.1%)	498,407 (7.6%)	206,713 (3.2%)
dsPiwi5-2	8,641,706	649,885 (7.5%)	12,565 (0.1%)	687,957 (8.0%)	301,385 (3.5%)
dsPiwi5-3	5,886,880	341,206 (5.8%)	10,838 (0.2%)	496,328 (8.4%)	245,686 (4.2%)
dsPiwi6-1	6,825,076	425,580 (6.2%)	43,645 (0.6%)	858,151 (12.6%)	586,311 (8.6%)
dsPiwi6-2	8,139,535	521,903 (6.4%)	41,781 (0.5%)	1,036,124 (12.7%)	688,963 (8.5%)
dsPiwi6-3	7,222,447	317,259 (4.3%)	51,061 (0.7%)	988,822 (13.7%)	682,842 (9.5%)
IP libraries					
GFP	1,374,646	19,542 (1.4%)	5,641 (0.4%)	137,252 (10.0%)	63,755 (4.6%)
Ago3	2,525,542	103,906 (4.1%)	74,696 (3.0%)	219,556 (8.7%)	138,592 (5.5%)
Piwi4	3,443,700	49,602 (1.4%)	9,828 (0.3%)	182,848 (5.3%)	83,766 (2.4%)
Piwi5	3,113,402	30,158 (1.0%)	14,803 (0.5%)	401,394 (12.9%)	310,650 (10.0%)
Piwi6	2,857,742	34,396 (1.2%)	14,760 (0.5%)	344,484 (12.1%)	263,350 (9.2%)

SUPPLEMENTAL MATERIALS AND METHODS

Cells and viruses

Aag2 cells were cultured at 25°C in Leibovitz's L-15 medium (Invitrogen) supplemented with 10% heat inactivated Fetal calf serum (PAA), 2% Tryptose Phosphate Broth Solution (Sigma), 1x MEM Non-Essential Amino Acids (Invitrogen) and 50 U/ml penicillin and 50 μg/ml streptomycin (Invitrogen). U4.4 and C6/36 cells were cultured in the same medium at 28°C. BHK-21 cells were cultured at 37°C, 5% CO_2 in Dulbecco's modified Eagles Medium (DMEM) supplemented with 10% FCS and 50 U/ml penicillin and 50 μg/ml streptomycin. The virus used throughout this study, with the exception of Figure 1C, 5A-5D and S5, is a recombinant, double-subgenomic Sindbis virus expressing GFP from the second subgenomic promoter (pTE-3'2J-GFP). For Figure 1C, 5A-5D and S5, the parental virus was used (pTE-3'2J). Viruses were produced in BHK-21 cells as previously described (1). Unless stated differently, Aag2 cells were infected with SINV at a multiplicity of infection (MOI) of 1 for 48h hours.

2

Generation of plasmids and dsRNA production

Insect expression vectors, based on pAc5.1 (Invitrogen), were constructed for N-terminal tagging of proteins with V5-3xFlag or 3xHA tags. The full-length coding sequence of Piwi4, Piwi5, Piwi6 and Ago3 was amplified from Aag2 complementary DNA (cDNA) and cloned downstream of the tag sequences.

For dsRNA production, *in vitro* transcription using T7 RNA polymerase was performed on T7-promoter-flanked PCR products. To allow the formation of double-stranded RNA, the reaction products were heated to 80°C and then gradually cooled to room temperature. Subsequently, the RNA was purified using the GenElute Mammalian Total RNA Miniprep Kit (Sigma) following the manufacturer's instructions. Primers used for plasmid or dsRNA production were:

XbaI-Piwi4	cagtctagaATGTCTGACCGTTACTCTC
NotI-Piwi4	catgcggccgcTTACAAGAAGTACAGCTTC
XbaI-Piwi5	cagtctagaATGGCGGATAGACAGCAAG
NotI-Piwi5	catgcggccgcTTACAGATAATAGAGTTTC
XbaI-Piwi6	cagtctagaATGGCTGATAATCCACAGG
NotI-Piwi6	catgcggccgcCTACAAAAAGTAAAGTTTC
XbaI-Ago3	cagtctagaATGTCCTCGCGGTTGAATTTAG
NotI-Ago3	catgcggccgcTCACAGGTAGAACAGTTT
T7Fw-Piwi1/3	taatacgactcactatagggagaCCACGCCCATCGTTTCAA
T7Re-Piwi1/3	taatacgactcactatagggagaCCTCAGTTTGTTCACCATA
T7Fw-Piwi2	taatacgactcactatagggagaCCGTCCTACTTTCCAGCAC
T7Re-Piwi2	taatacgactcactatagggagaGCGGCACTCCAGGGACAAT
T7Fw-Piwi4	taatacgactcactatagggagaCGTGGAAGTCCTTCTTCTCG
T7Re-Piwi4	taatacgactcactatagggagaTGTCAGTTGATCGCTTCTCAA
T7Fw-Piwi5	taatacgactcactatagggagaGCCATACATCGGGTCAAAAT
T7Re-Piwi5	taatacgactcactatagggagaCTCTCCACCGAAGGATTGAA
T7Fw-Piwi6	taatacgactcactatagggagaCAACGGAGGATCTTCACGAG
T7Re-Piwi6	taatacgactcactatagggagaAATCGATGGCTTGATTTGGA
T7Fw-Piwi7	taatacgactcactatagggagaGTGGAGGTCGTGGAGGTAAC
T7Re-Piwi7	taatacgactcactatagggagaGTTTGCGGTGTTTCCGTACT
T7Fw-Ago3	taatacgactcactatagggagaTGCTTACTCGTGTCGCGTAG
T7Re-Ago3	taatacgactcactatagggagaGGCATGGCAGATCCAATACT
T7Fw-Ago2	taatacgactcactatagggagaCTACGAGCAGGAGGTCAAGG
T7Re-Ago2	taatacgactcactatagggagaTCCATGCCTTTGAGGAAATC
T7Fw-Ago1	taatacgactcactatagggagaCCGGTCATCGAGTTCATGT
T7Re-Ago1	taatacgactcactatagggagaCGTGGCTTTGATCATGGTT
T7Fw-Luc	taatacgactcactatagggagaTATGAAGAGATACGCCCTGGTT
T7Re-Luc	taatacgactcactatagggagaTAAAACCGGGAGGTAGATGAGA

Transfection and infection of Aag2 cells

For IP or transgenic PIWI protein analysis, Aag2 cells were transfected with PIWI protein expression plasmids using X-tremeGENE HP (Roche) according to the manufacturer's instructions. Three hours post transfection, the medium was refreshed with supplemented Leibovitz's medium and, where indicated, infected with SINV. For knockdown experiments, Aag2 were transfected with dsRNA using X-tremeGENE HP. To increase knockdown efficiency, Aag2 cells were re-transfected at 48h after the first transfection. Three hours post transfection the medium was refreshed with supplemented medium and, where indicated, cells were infected with SINV. Unless stated differently, samples were harvested at 48 hours post infection.

Northern blot and β-elimination

Small RNA northern blot was performed as described previously (2). Briefly, total RNA was isolated using Isol-RNA Lysis Reagent (5 PRIME), size separated on a 15% PAGE gel, blotted to a nylon membrane (Hybond NX; Amersham) and cross-linked using 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC; Sigma). For $NaIO_4$ oxidation and β-elimination, total RNA (in 13.5µl water) was added to 4µl borate buffer (148mM borax, 148mM boric acid, pH 8.6) and 2.5µl 200mM $NaIO_4$. After 10min incubation at room temperature, unreacted $NaIO_4$ was quenched by adding 2µl glycerol. After an additional 10min incubation at room temperature, samples were dried by centrifugation under vacuum and resuspended in 50µl of borax buffer (30mM borax, 30mM boric acid, 50mM NaOH, pH 9.5). Samples were incubated at 45°C for 1.5h. Finally, RNA was ethanol-precipitated and reconstituted in 15µl water for further northern blot analysis. Hybridization with [^{32}P] labeled DNA oligonucleotides was performed overnight at 42°C. The membrane was washed in 0.1% SDS, 2x SSC, followed by two washing steps in 0.1% SDS, 1x SSC and 0.1% SDS, 0.1x SSC, respectively. All washes were performed at 42°C. For detection of the radioactive signal, the membrane was exposed to an X-ray film (Kodak). For northern blotting of high molecular weight RNA, total RNA was separated by denaturing agarose gel electrophoresis. The RNA was transferred to a nylon membrane using the turboblotter system (Whatman) according to the manufacturer's recommendations and crosslinked to the membrane using UV irradiation. Probe hybridization, washing of the membrane, and detection of the radioactive signal was carried out as described for the small RNA northern blot. Sequences of northern blot probes were:

nSINV-7903+	GGTTGCTTCTTCTTCTTCTCCTGCGTTT
nSINV-7940+	AGTGCCATGCGCTGTCTCTTTCCGGGTTTG
nSINV-7969+	TCGAACAATCTGTCGGCCTCCAACTTAA
nSINV-8040+	GCAGAGGTTTCATTACCTTTCCTTCCAT
nMir2940-3p	AGTGATTTATCTCCCTGTCGAC

nTF345-1570+	CTTGCTGGATTTCCGTCCCTTCGCTGATC
nTF345-1630+	AACGAAACGACGTTTGCTTTGGACACTGT
nTF345-1058-	AGAAGCGCTATTCCACGCTCACCATCGCA
nActin	ATGGGCACGGTGTGGGAGACACCA
nRPL5	GCTTCTGCAGGATGCGGCGGGCAA
nTRNA-lys	AAAAGTCCAACGCTCTACCGACTGAGCTACCCGGGC
nU3	AAACTTGCCTACAGAAATGATCCTGTGAAGCACAGT

qRT-PCR for AGO/PIWI proteins

Total RNA was DNaseI-treated (Ambion) and reverse-transcribed using the Taqman Reverse transcription kit (Roche) and random primers following the manufacturer's instructions. qPCR reactions were prepared using GoTaq qPCR SYBR Mastermix (Promega) and measured on a Light Cycler 480 (Roche). Expression of AGO/PIWI genes was internally normalized against the expression of Lysosomal Aspartic Protease (LAP) and the relative mRNA abundance was determined using to the ΔΔCt method (3). The following primers were used for qPCR.

qFw-Piwi1/3	GGCCGTTAGCGAGTCTCAT
qRe-Piwi1/3	GGCAGAACCTTCGTGGTAAG
qFw-Piwi2	CCGCGGGTACACCGCCGTCAACTT
qRe-Piwi2	CGCTGGTCGAACTCGATGCCCCGC
qFw-Piwi4	TCTTCTTCTCCACCACAGCC
qRe-Piwi4	ATGGTGACCACCTCACAGTTAC
qFw-Piwi5	ACGGCATCACATCGAGACTC
qRe-Piwi5	CGACCTCCACGCTGTCCTC
qFw-Piwi6	TTTTCTTCCACCCCGAGCAG
qRe-Piwi6	AATACATTTGCGATGCGGCC
qFw-Piwi7	ATGCGACGAAACTTCAACTTG
qRe-Piwi7	CCAGCAGCAACCGCATAATT
qFw-Ago3	CTCCAGACGACGGTTTTGGA
qRe-Ago3	GCAGGTACGAAATTGGCTGC
qFw-Ago2	ATTTGGCTCAAGATCAACGC
qRe-Ago2	GAGATCGTATGAAGCGGCCA
qFw-Ago1	CGAACAGCATGATGGAAGTG
qRe-Ago1	AAATTGTTTGCCTCGCATGT
qFw-LAP	GTGCTCATTCACCAACATCG
qRe-LAP	AACTTGGCCGCAACAAATAC

Immunoprecipitation

Aag2 cells expressing V5-3xFlag-tagged PIWI proteins were lysed in lysis buffer (50mM Tris-HCl (pH7.8), 150mM NaCl, 1mM EDTA, 1mM DTT, 0.5% NP40, 1x Protease

inhibitors). The lysates were cleared by incubation with washed Protein G agarose beads (Thermo Scientific) at 4˚C under constant rotation for 6 hours. For V5-IP, the cleared lysates were incubated with washed V5 agarose beads (Sigma) overnight under the same conditions. The beads were then washed 5 times in wash buffer I (50mM Tris-HCl (pH7.8), 150mM NaCl, 1mM EDTA), followed by two wash steps in wash buffer II (25mM Tris-HCl (pH7.8) 150mM). Finally, the bound RNA was isolated from the beads using Isol-RNA Lysis Reagent.

Immunofluorescence and western blot analyses

For subcellular localization of PIWI proteins, uninfected or SINV-infected Aag2 cells expressing 3xHA-tagged Piwi5 or Ago3 were fixed on coverslips using 4% paraformaldehyde. Cells were permeabilized in PBS/0.1% Triton and incubated with rabbit anti-HA (1:200 dilution; Abcam ab 9110), or mouse anti-dsRNA (1:1000 dilution; English & Scientific consulting J2 mAb) antibodies. Subsequently, cells were washed in PBS/0.1% Triton and incubated with secondary antibodies (1:400 dilution, goat anti mouse-AlexaFluor594, or goat anti rabbit-AlexaFluor488; Life technologies). After washing, the nuclei were stained with Hoechst reagent and cover slips were fixed to microscope slides using Mowiol. Pictures were taken on an Olympus FV1000 confocal microscope.

For western blot, proteins were size separated on SDS-PAGE gels and transferred to a nitrocellulose membrane. Antibodies used to detect the proteins were: mouse anti-Flag M2 (1:1000 dilution, Sigma), rat anti-tubulin alpha (1:1000 dilution, Sanbio), and mouse anti-H3K9me2 (1:1000 dilution, Abcam ab1220). Secondary antibodies were IRdye680 or IRdye800 conjugated goat anti mouse or goat anti rat, respectively (both 1:15000 dilution; Li-cor).

5′ RACE and Piwi5 RT-PCR

5′RACE for Piwi5 was performed using the First Choice RNA Ligase Mediated RACE kit (Ambion) according to the manufacturer's instructions. Briefly, total RNA was treated with Calf Intestine Alkaline Phosphatase (CIP) to remove free 5′ phosphate groups and Tobacco Acid Pyrophosphatase (TAP) to remove the cap of mRNAs. The RNA was then ligated to the 5′RACE RNA adapter, reverse transcribed and PCR amplified. The PCR product was sequenced by Sanger technology. Piwi5 RT-PCR was performed on Aag2 cDNA using Thermoperfect DNA Polymerase. The following primers were used:

pFw-Piwi5A	ACGGCATCACATCGAGACTC (=qFw-Piwi5)
pFw-Piwi5B	CAGCAACCGCAACAGCCAGCGCCT
pRe-Piwi5C	GTGCTTCTCCGCCAGTGGCACCCC
pRe-Piwi5D	AAATGCTCCAAGCGCTGTAT

SUPPLEMENTAL REFERENCES

1. Vodovar N, Bronkhorst AW, van Cleef KW, Miesen P, Blanc H, van Rij RP, et al. Arbovirus-derived piRNAs exhibit a ping-pong signature in mosquito cells. PLoS One. 2012;7(1):e30861.
2. Pall GS, Hamilton AJ. Improved northern blot method for enhanced detection of small RNA. Nat Protoc. 2008;3(6):1077-84.
3. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001;25(4):402-8.

Chapter 3

Small RNA Profiling in Dengue Virus 2-Infected *Aedes* Mosquito Cells Reveals Viral piRNAs and Novel Host miRNAs.

Pascal Miesen, Alasdair Ivens, Amy H. Buck, Ronald P. van Rij

PLoS negl. Trop. Dis. (2016) 10:e0004452

ABSTRACT

In *Aedes* mosquitoes, infections with arthropod-borne viruses (arboviruses) trigger or modulate the expression of various classes of viral and host-derived small RNAs, including small interfering RNAs (siRNAs), PIWI interacting RNAs (piRNAs), and microRNAs (miRNAs). Viral siRNAs are at the core of the antiviral RNA interference machinery, one of the key pathways that limit virus replication in invertebrates. Besides siRNAs, *Aedes* mosquitoes and cells derived from these insects produce arbovirus-derived piRNAs, the best studied examples being viruses from the *Togaviridae* or *Bunyaviridae* families. Host miRNAs modulate the expression of a large number of genes and their levels may change in response to viral infections. In addition, some viruses, mostly with a DNA genome, express their own miRNAs to regulate host and viral gene expression. Here, we perform a comprehensive analysis of both viral and host-derived small RNAs in *Aedes aegypti* Aag2 cells infected with dengue virus 2 (DENV), a member of the *Flaviviridae* family. Aag2 cells are competent in producing all three types of small RNAs and provide a powerful tool to explore the crosstalk between arboviral infection and the distinct RNA silencing pathways. Interestingly, besides the well-characterized DENV-derived siRNAs, a specific population of viral piRNAs was identified in infected Aag2 cells. Knockdown of Piwi5, Ago3 and, to a lesser extent, Piwi6 results in reduction of vpiRNA levels, providing the first genetic evidence that *Aedes* PIWI proteins produce DENV-derived small RNAs. In contrast, we do not find convincing evidence for the production of virus-derived miRNAs. Neither do we find that host miRNA expression is strongly changed upon DENV2 infection. Finally, our deep-sequencing analyses detect 30 novel *Aedes* miRNAs, complementing the repertoire of regulatory small RNAs in this important vector species.

AUTHOR SUMMARY

Mosquitoes of the *Aedes* family transmit many important viruses including dengue virus between their vertebrate hosts. In the mosquito, the growth of these viruses is limited by the antiviral RNA interference pathway. Key to this pathway is a class of small non-coding RNAs known as small interfering RNAs (siRNAs). In addition, two related but distinct small RNA pathways known as the microRNA (miRNA) and the PIWI-interacting RNA (piRNA) pathway are implicated in regulating virus replication in mosquitoes. Thus, since small RNAs may critically influence the transmission of dengue virus, we set out to analyze the populations of viral and mosquito small RNAs that are produced in infected *Aedes* mosquito cells. We found that besides the well-known viral siRNAs, dengue virus-derived piRNAs were produced in these cells and we identified the PIWI proteins that these small RNAs rely on. In addition, we found that viral miRNAs were not expressed from the dengue virus genome and that the levels of mosquito miRNAs were barely changed upon infection. Finally, our data allowed for the identification of novel *Aedes* miRNAs, complementing the repertoire of these important regulatory RNAs in vector mosquitoes.

INTRODUCTION

Aedes mosquitoes are essential vectors for the transmission of important arthropod-borne viruses (arboviruses), including dengue virus (DENV), yellow fever virus, and chikungunya virus (1).While several of these arboviral infections cause disease in humans, virus replication generally does not lead to severe pathology in vector mosquitoes. Infected mosquitoes thus serve as a persistent reservoir for arboviruses in the wild and they may transmit these viruses to vertebrate hosts throughout their entire lives (2).

After ingestion in a mosquito's blood meal, arboviruses need to overcome a number of anatomical and immunological barriers to reach sufficiently high titres in the saliva. Only then can transmission to a naive vertebrate host efficiently occur. One of the most important immune responses to arboviral infection is antiviral RNA interference (RNAi) (3-5). This pathway is triggered by the presence of double stranded RNA (dsRNA), which is produced during the replication of RNA and DNA viruses (6, 7). The dsRNA is recognized and cleaved by the RNase-III enzyme Dicer-2 (Dcr2) into 21 nucleotide (nt) small interfering RNA duplexes (viral siRNA; vsiRNA) (8, 9). One of the siRNA strands is incorporated in Argonaute-2 (Ago2), the core protein of the RNA induced silencing complex (RISC) (10). The siRNA-loaded RISC complex is guided to complementary viral RNA molecules and cleaves these target RNAs using the endonuclease (slicer) activity of Ago2 (11).

MicroRNAs (miRNAs) are a distinct class of small RNAs that are produced from genome-encoded stem loop-containing transcripts known as primary miRNA (pri-miRNAs). During the canonical miRNA biogenesis pathway, the stem loop structures, known as precursor miRNA (pre-miRNA), are released from the pri-miRNA by the microprocessor complex with at its core the RNase-III enzyme Drosha. After translocation into the cytoplasm, pre-miRNAs are cleaved by Dicer-1 (Dcr1) to produce a small RNA duplex comprised of the two mature miRNA strands. Usually, one of these strands is then preferentially incorporated into the Argonaute-1 (Ago1) containing miRNA-induced silencing complex (miRISC), whereas the other strand (the passenger or miRNA* strand) is usually discarded (12). Loaded miRISC complexes are able to bind to specific target sites within mRNAs. This miRNA-mRNA interaction is initiated by nucleotide two to seven of the miRNA, the so-called seed sequence (13). Stable binding of miRISC to an mRNA target, generally causes down-regulation of gene expression via translational inhibition and mRNA destabilization (14). Importantly, infecting viruses can directly or indirectly, as a consequence of the immune response, reshape the host miRNA expression landscape. While quite a number of studies have reported on this matter in mammalian systems (15), little is known about virus-induced changes in miRNA levels in mosquito vectors. In *Aedes* mosquitoes, miRNA levels or modifications have been reported to be changed upon infections with DENV, West Nile virus, and chikungunya virus (16-20). For most of these differentially expressed miRNAs, the biological relevance as well as the targeted mRNAs still await experimental validation.

Besides modulation of host miRNAs, some DNA and retroviruses encode their own miRNAs to regulate viral and host mRNAs (21). The expression of miRNAs from cytoplasmic RNA viruses has been controversial. However, functional introduction of artificial miRNAs into the genomes of Sindbis virus (SINV) and tick-borne encephalitis virus provides evidence that miRNA production from cytoplasmic RNA viruses may in principle be possible (22, 23). Yet, the presence and biological relevance of miRNAs encoded in the genomes of flaviviruses such as DENV is still an issue of debate (24-26).

The third, most enigmatic class of small RNAs are PIWI interacting RNAs (piRNAs). These are processed from long RNA precursors that are transcribed from genomic loci known as piRNA clusters. In sharp contrast to siRNAs and miRNAs, their biogenesis into mature piRNAs is Dicer-independent. In *Drosophila*, piRNA maturation involves endonucleolytic cleavage of precursor transcripts by the Zucchini nuclease and the three PIWI proteins Piwi, Aubergine (Aub) and Argonaute-3 (Ago3) (27-30). The primary function of the piRNA pathway in this model organism is the defence against transposable elements, mainly in germ-line tissues. Interestingly, piRNAs of viral origin (vpiRNA) have been found in somatic tissue of *Aedes* mosquitoes, suggesting that they contribute to the regulation of virus replication (31). At present, vpiRNAs have been discovered upon infection with a number of Alphaviruses, Bunyaviruses and Flaviviruses, including DENV (31-39). However, with the exception of SINV (Alphavirus), their molecular biogenesis has not been investigated (37).

Here, we make use of small RNA deep-sequencing in the siRNA, miRNA, and piRNA competent *Aedes aegypti* Aag2 cell line to investigate the production of small RNAs during DENV infection. We find that in addition to the well-characterized vsiRNAs, specific vpiRNAs are produced from DENV, which for their biogenesis in Aag2 cells rely on Piwi5 and Ago3 and, to a lesser extent, on Piwi6. We do not detect DENV-derived miRNAs, or prominent changes in host miRNA levels upon infection. Finally, we identify novel host miRNAs in our small RNA deep-sequencing libraries, complementing the currently annotated miRNA repertoire in *Aedes aegypti* vector mosquitoes.

MATERIALS AND METHODS

Cells and viruses

Aag2 cells were cultured at 25°C in Leibovitz L-15 medium (Gibco) supplemented with 10% heat inactivated fetal calf serum (FCS; PAA), 2% tryptose phosphate broth solution (Sigma), 1x MEM non-essential amino acids (Gibco), and 50 U/ml penicillin and 50 µg/ml streptomycin (pen/strep; Gibco). U4.4 and C6/36 were kept in the same culture medium at 28°C. BHK-21 cells were cultured at 37°C, 5% CO_2 in Dulbecco's modified Eagles medium (DMEM) supplemented with 10% FCS and pen/strep. Stocks of DENV serotype 2 (DENV2), New Guinea C (NGC) and 16681 strains were grown on C6/36 cells and titred on BHK-15 cells as detailed in (40).

Infection of Aag2 cells with DENV2

Aag2 cells were seeded one day prior to infection and infected with DENV2 at a multiplicity of infection (MOI) of 0.5 by directly adding the virus to the culture medium. Three days post infection the culture medium was removed and cells were harvested for RNA and protein isolation as detailed below.

Western blot

For the detection of the DENV NS1 protein in samples used for small RNA deep-sequencing, 5% of the cells were harvested in 50 µl lysis buffer (50 mM Tris-HCl pH 7.8; 150 mM NaCl; 1 mM EDTA; 0.5% NP-40; 1x Protease inhibitor cocktail (Roche); 1 mM DTT). 12.5 µl of 5x Laemmli buffer was added to each sample, incubated at 95°C for 5 min, and 30 µl of each sample was loaded on a 12.5% polyacrylamide gel. After gel electrophoresis, proteins were transferred to a nitrocellulose membrane (Bio-Rad) using a semi-dry blotting system (Bio-Rad). The membrane was blocked in 5% non-fat dry milk (Bio-Rad) in 0.1% Tween20 in PBS (PBS-T) for 30 min at room temperature. Mouse anti DENV NS1 antibody was kindly provided by Dr. Peter Mason (41). The antibody was added to the membrane in a 1:1,000 dilution in 5% blocking buffer. After an incubation for 1.5 hours at room temperature, the membrane was washed three times in PBS-T. IRdye680 conjugated goat anti mouse antibody (1:15,000 dilution in PBS-T; Licor) was then added to the membrane and incubated at room temperature for 1.5 hours. After three washing steps, the membrane was imaged on an Odyssey infrared image system (Licor).

dsRNA production and transfection of Aag2 cells

dsRNAs targeting PIWI/AGO transcripts or luciferase as a negative control were produced by *in vitro* transcription from T7-promoter flanked PCR products as detailed in (37). Primers to produce T7-flanked PCR products are indicated in S1 Table.
For dsRNA transfection, $7.5x10^5$ Aag2 cells were seeded in one well of a 24-well plate. For each condition, three wells were plated. The following day, transfection mixes containing 300 µl non-supplemented L-15 medium, 450 ng dsRNA and 1.8 µl X-tremeGENE HP (Roche) were prepared according to the manufacturer's recommendations. 100 µl of the mix was added dropwise to one well. After 2-3 hours the medium was replaced with fully supplemented L-15 medium. 48 hours later, the transfection was repeated to enhance knockdown efficiencies. Where indicated, the cells were infected with DENV2 which was added to the L-15 medium used to replace the transfection medium as specified above.

RNA isolation

Aag2 cells were lysed in Isol-RNA Lysis reagent (5 PRIME) as described in the manufacturer's instructions. Briefly, 200 μl of chloroform was added to 1 ml of Lysis reagent and mixed well. After centrifugation, the aqueous phase was collected and total RNA was purified using isopropanol precipitation. RNA was quantified on a Nanodrop spectrophotometer and RNA integrity was checked by ethidium bromide staining of ribosomal RNA bands after agarose gel electrophoresis.

DNaseI treatment, reverse transcription and (quantitative) PCR

For RT-(q)PCR, 1 μg of total RNA was DNaseI (Ambion) treated according to the manufacturer's instructions. The RNA was subsequently reverse transcribed in a 20 μl reaction using the Taqman reverse transcription kit (Applied Biosystems). Complementary DNA (cDNA) was diluted 5-10 times before PCR amplification. Endpoint PCR was performed using Thermoperfect DNA Polymerase. Quantitative PCR (qPCR) analysis was performed using the GoTaq qPCR SYBR mastermix (Promega) on a LightCycler 480 instrument (Roche). The relative changes in gene expression were calculated using the ΔΔCt method (42) using lysosomal aspartic protease (LAP) as an internal normalization control. Sequences of the PCR primers are indicated in S1 Table.

β-elimination

Sodium periodate ($NaIO_4$) oxidation and β-elimination of total RNA was performed as described previously (43). Total RNA (10 μg in 47.5 μl nuclease-free water) was mixed with 12.5 μl 200 mM $NaIO_4$ and 40 μl 5x borate buffer. As a control, RNA was treated with water instead of $NaIO_4$. The reaction was incubated at room temperature for 30 min and 10 μl glycerol was added to the reaction. The reaction was incubated for another 10 min before 10 μl of 500 mM sodium hydroxide (NaOH) was added to induce β-elimination. The reaction was incubated at 45°C for 90 min. After these treatments, total RNA was purified by ethanol precipitation in the presence of 300 mM NaCl and 5 μg of glycogen. Electrophoretic mobility of *Aedes aegypti* miR-2940-3p and DENV2 piRNAs was then analyzed by small RNA northern blotting as detailed below.

Small RNA northern blotting

Small RNA northern blot was performed as described in (44). Briefly, total RNA was size-separated on 0.5x TBE, 7 M Urea, 15% Polyacrylamide gels, transferred to Hybond NX nylon membranes (Amersham), and cross-linked using 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC; Sigma). Individual small RNAs were detected with DNA oligonucleotides that were 5′ end-labelled with [^{32}P] γ-adenosine-triphosphate (Perking Elmer) using T4 Polynucleotide kinase (Roche). Hybridization

to the oligo-probes was performed overnight at 42°C in Ultrahyb Oligo hybridization buffer (Ambion). Membranes were then washed three times at 42°C in 0.1% SDS with decreasing concentrations of SSC (2x, 1x, 0.1x). Membranes were exposed to X-ray films (Carestream) or Phosphorimager screens (BioRad). Sequences of DNA oligonucleotide probes are indicated in S1 Table. Quantification of northern blot panels was performed using ImageJ software. Bands were defined using the rectangular selection tool and the pixel density (area under the curve) was measured and normalized to uninfected dsLuc samples.

Preparation of small RNA libraries

Small RNA libraries were prepared as described previously (37, 45). Briefly, three 25 cm^2 flasks of Aag2 cells were infected in parallel with DENV2 NGC. Three additional flasks were left uninfected. Total RNA was then isolated from these six flasks as specified above and 30 μg of total RNA was size separated on a 15% Polyacrylamide, 7 M urea, 0.5x TBE gel. Subsequently, the small RNAs in the size range from 18 nt to 33 nt were excised from gel using radioactively-labelled RNA oligos, loaded in the adjacent lanes of the gel, as rulers. The gel was crushed and the small RNAs were eluted in 300 mM sodium acetate overnight at 4 °C under constant rotation. The RNA was recovered from the elution buffer using ethanol precipitation and eluted in 10 μl of nuclease-free water. 5 μl of the sample was directly used as input for small RNA deep-sequencing library preparation using the TruSeq small RNA library preparation kit (Illumina) following the manufacturer's recommendations. The RNA was ligated to 3′ and 5′ adapters, reverse transcribed, and PCR amplified. The small RNA libraries were then size-purified from 1x TBE, 6% polyacrylamide gel using overnight elution in 300 mM sodium acetate followed by ethanol precipitation. The individual small RNA libraries were pooled and sequenced on a single sequencing lane on a HighSeq2500 by Baseclear (Leiden, The Netherlands).

Viral small RNA profiling

FASTQ sequence reads were generated using the Casava pipeline (v.1.8.3) and initial quality analysis was performed using the Illumina Chastity filter and an in-house filtering protocol by Baseclear. Subsequent quality assessment was based on the FASTQC quality control tool (v.1.10.0). The individual small RNA sequencing libraries were separated based on the TruSeq indices (no. 1 to 6) that were introduced during PCR amplification. The individual libraries were subsequently analyzed using the Galaxy bioinformatics tool shed (46, 47). For the analysis of viral small RNAs, reads were mapped to the DENV2 NGC genome (GenBank accession: KM204118.1) using Bowtie (v.1.1.2) (48). Size profiles were obtained from all reads that align to this reference sequence with a maximum of one mismatch. The genome distribution of 21 nt siRNAs, 22-24 nt small RNAs, or 25-30 nt piRNAs was obtained by plotting the number of 5′ ends of these reads at each

position of the genome. For the pileup plots of UTR-derived miRNA-like small RNAs, the 22-24 nt small RNAs were selected from the initial FASTQ files and mapped to the DENV-NGC genome. From the resulting SAM files the reads mapping to the (+) strand of the virus genome were selected and used as input for the 'Generate pileup from BAM dataset' tool (v.1.1.2). The values at the nucleotide positions of the DENV2 5'UTR (1-96) and 3' UTR (10273-10723) were selected for display.

Re-analysis of the data published by Hess *et al*. (36) was performed on the dataset with the accession number SRR921363. The SOLiD-formatted dataset was groomed to fit the requirements for manipulation in Galaxy. After adapter clipping, reads were mapped to the DENV2-JAM1409 genome (GenBank accession: M20558) using Bowtie2 (49). Size and genome profiles were obtained as described above. Unless specified differently, all small RNA read counts were normalized against the size of the corresponding sequencing library and are expressed as '% of the library' (i.e. reads per hundred).

miRNA analysis and prediction

Analysis of miRNA expression levels was performed using the miRDeep2 tool. Raw data were assessed for quality using FASTQC. Subsequently, adapters were removed from the raw reads, and the reads were quality trimmed using cutadapt software (http://dx.doi.org/10.14806/ej.17.1.200) with parameters -O 6 -m 17 -n 5 -q 20. Within each library, the resulting reads were collapsed to generate a non-redundant set of FASTA sequences, subsequently processed to the format required for miRNA prediction with miRDeep2 software (50). Collapsed reads longer than 17 nucleotides were aligned to the *Aedes* genome (assembly AaegL3, downloaded from vectorbase) and the DENV2 NGC genome using the mapper.pl component of miRDeep2 (parameters: -o 20 -l 19 -r 100 -c). The resulting outputs were parsed to remove alignments that were not full length and perfect match (FLPM). miRDeep2 predictions were generated from the FLPM aligned sequences, with miRBase v21 Arthropoda mature miRNA and pre-miRNA sequences as templates (51). The 'miRNAs_expressed' output from miRDeep2, which comprises tallies for each known miRNA in each sample, was further processed in the R/Bioconductor environment. Briefly, miRNA read counts were normalized to the number of *Aedes*-specific genome reads within each sample group, using the lowest number of reads aligning as the baseline. Subsequently, the counts were converted to abundances within each sample, converted to log2 equivalent counts, and all samples quantile normalized prior to linear model fitting with the limma package (52). MiRNA predictions from the miRDeep2 output were manually curated using the following criteria: i) high-confidence miRNA predictions have reads mapping to both a predicted mature and star sequence, ii) the mature sequences have a homogenous 5' end (80% of reads start at same position), and iii) miRNA-miRNA* duplex should resemble a Dicer product on a genomically-encoded hairpin, having a two nt (+/-1 nt) overhang at the 3' end. miRNA predictions

supported by >1000 reads as well as miRNA predictions with a seed match to known insect miRNAs were also kept. To be retained, these predictions required a homogenous 5′ end, but did not require the presence of reads mapping to the expected star strand. miRNA names and accession numbers were assigned by the miRBase repository.

All deep sequencing libraries have been submitted to NCBI Sequence Read Archive under the accession number SRA303329. All source data are available online in Dataset S1.

RESULTS

DENV2-derived small RNAs in infected Aag2 cells

DENV is a positive (+) strand RNA virus belonging to the *Flavivirus* genus in the *Flaviviridae* family. Its RNA genome is approximately 10.7 kilobases in size and encodes a single polypeptide that is processed by proteolytic cleavage events into three structural proteins and seven non-structural proteins (Figure 1A). Since various classes of small RNAs have been implicated in modulating DENV infections in its mosquito vectors, we aimed to characterize the repertoire of virus and host-derived small RNAs in *Aedes aegypti* Aag2 cells. To this end, we prepared three independent small RNA deep-sequencing libraries from uninfected and DENV2 (NGC strain) infected cells, each. The efficiency of the three infections was comparable as assessed by western blot for the DENV2 NS1 protein (Figure 1B). As expected, DENV2-derived viral small RNAs showed a clear population of 21 nt vsiRNAs mapping to both the viral positive (+) strand and the negative (-) strand in roughly equal numbers (Figure 1C). Interestingly, besides siRNAs, a second population of viral small RNAs was produced that resembled vpiRNAs. These were 25-30 nt in length and almost exclusively derived from the viral (+) strand (Figure 1C). In contrast to vsiRNA, which were distributed along the entire length of the viral genome, these putative vpiRNAs were produced only from few specific positions (Figure 1D). In fact, 85% of all the 25-30 nt reads were derived from four individual vpiRNA sequences, present in the NS5 gene at positions 9180 and 9985, 9989 and 9990 of the DENV2 NGC genome. To test whether these small RNA profiles reflect those from adult mosquitoes, we re-analyzed deep sequencing data from DENV2 infected *Aedes aegypti* mosquitoes published by Hess *et al.* (36). We analyzed the 9 days post infection sample, which showed the highest number of viral siRNAs and piRNAs. Whereas normalized vsiRNA levels were only 2.2-fold lower in these libraries than in our Aag2 data, vpiRNAs were about forty times lower. Yet, the viral small RNA profiles were strikingly similar, with 21 nt reads being scattered throughout the entire viral genome and piRNA-sized reads being predominantly produced from few positions located towards the 3′ end of the viral genome (Figure S1). These data suggest that similar mechanisms might produce viral piRNAs in Aag2 cells and adult mosquitoes.

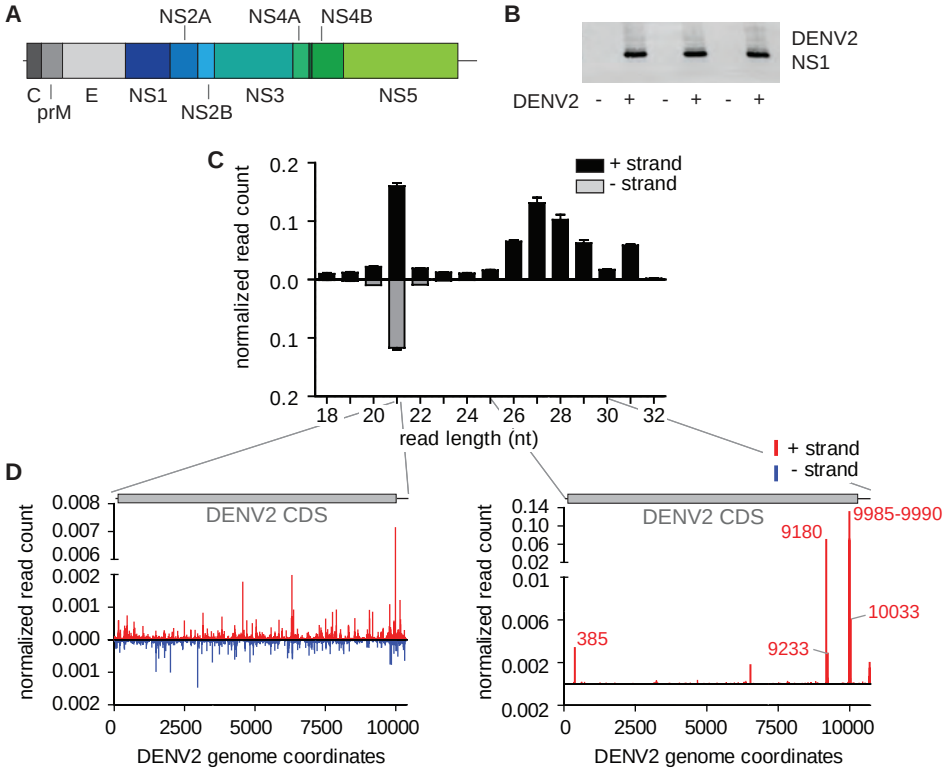


Figure 1. Small RNA production in DENV infected Aag2 cells. (A) Schematic representation of the DENV NGC genome (accession KM204118; 10723 bp). Structural proteins are indicated in grey scale, non-structural proteins are displayed in blue to green scale. **(B)** Western blot against the DENV2 NS1 protein in the three infected and uninfected samples used for small RNA library preparation. **(C)** Size profile of small RNAs mapping to the DENV2 genome with a maximum of one mismatch. Black bars represent reads mapping to the (+) strand of the genome, grey bars depict reads from the (-) strand. The read counts have been normalized to the size of the small RNA library and the mean +/- standard error of the mean (SEM) is presented (*n*=3). **(D)** Distribution of 21 nt vsiRNAs (left panel) or 25-30 nt small RNAs (right panel) across the DENV genome. Reads from the (+) and (-) strands are depicted in red and blue, respectively. The read counts have been normalized as described in C, the mean read count of the three libraries is shown. Numbers in red indicate genome positions of the vpiRNA spikes.

vpiRNA production from DENV2 RNA

To exclude the possibility that the piRNA-like molecules are sequencing artefacts and to characterize this small RNA population in more detail, we performed small RNA northern blotting for the highly-abundant small RNAs starting at DENV2 genome positions 9180 or 9985-9990. Indeed, small RNAs in the expected size range could readily be detected specifically in DENV2 infected Aag2 cells (Figure 2A). In addition, these sequences were also present in DENV2-infected U4.4 and C6/36 cells derived from *Aedes albopictus* mosquitoes, which we have previously shown to be competent in producing SINV-derived vpiRNAs (32, 37) (Figure 2B). The levels of vpiRNAs did not correlate with the expression of viral genomic RNA. Whereas viral RNA levels were roughly

eighteen to nineteen fold higher in U4.4 and C6/36 cells than in Aag2 cells, vpiRNAs were most abundant in Aag2 cells (Figure 2B). These data suggest that the composition of host factors required for their biogenesis is most favourable in Aag2 cells. As expected, mammalian BHK21 cells, which lack an active piRNA pathway, did not produce vpiRNAs

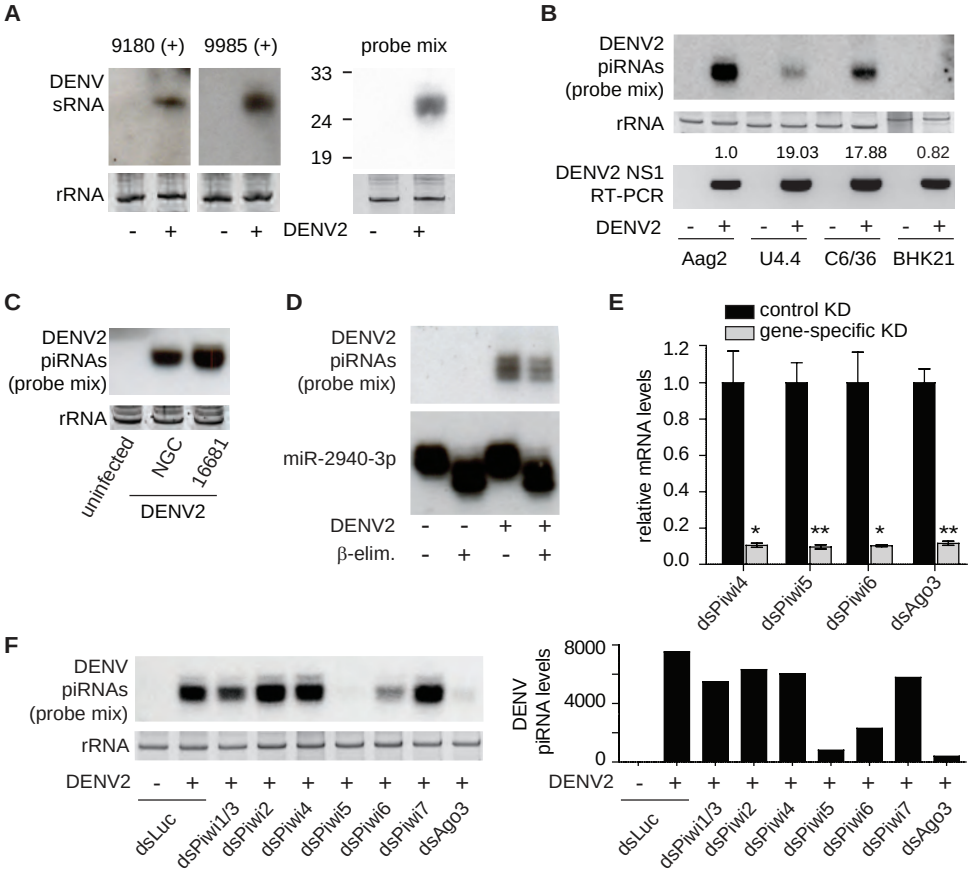


Figure 2. vpiRNA production in Aag2 cells. (A) Northern blot of highly abundant vpiRNAs. Two individual DNA oligonucleotide probes (left panels), or a combination of these probes (right panel) were used to detect the small RNAs. The combination of probes was used in all subsequent small RNA blots. **(B)** Upper panel: Small RNA northern blot of vpiRNAs in the indicated cell lines after infection with DENV2. Lower panel: RT-PCR for DENV genomic RNA in the same samples used for the northern blot. Numbers on top indicate DENV genomic RNA levels (relative to Aag2 cells) as determined by RT-qPCR (n=1). **(C)** Northern blot of DENV2 piRNAs in Aag2 cells infected with the DENV2 NGC or 16681 strain, both at an MOI of 0.5. **(D)** Northern blot of DENV2 piRNAs and *Aedes* miR-2940-3p in uninfected or DENV2 infected Aag2 cells. Where indicated, total RNA was subjected to β-elimination. **(E)** RT-qPCR for the indicated PIWI proteins after gene-specific knockdown (KD) in Aag2 cells normalized to a control KD (dsLuc). Bars represent the mean of three experiments +/- SEM. Statistical significance was determined using two tailed, unpaired student t-test. * $p<0.05$; ** $p<0.01$. **(F)** Upper panel: Small RNA northern blot of vpiRNAs upon KD of the indicated PIWI proteins. RNA samples analyzed in E were pooled for this blot. Lower panel: Quantification of two independent blots including the one shown in the upper panel using ImageJ software. For the other blot, see S1 Dataset. Ethidium bromide staining of ribosomal RNA was used as loading control in panel A, B, C, and F.

(Figure 2B). To exclude that viral piRNA production is an artefact of the use of the specific DENV2 strain, we analyzed piRNA accumulation in Aag2 cells infected with either the DENV NGC or DENV 16681 strain. Small RNA northern blotting revealed that infection with either strain resulted in the production of those viral piRNA sequences that we had found by deep-sequencing (Figure 2C). Next, we aimed to test whether DENV2-derived piRNAs are methylated at their 3′ end. To this end, we performed sodium periodate oxidation followed by beta-elimination, which reveals modifications of the 3′ terminal nucleotide of RNA molecules (43). Unmodified small RNAs, such as animal miRNAs, are susceptible to this treatment and will be shortened by one nucleoside resulting in increased electrophoretic mobility. In contrast to miRNAs, piRNAs are protected against this treatment by methylation of the 2′OH on the ribose of the 3′ terminal nucleotide. Indeed, beta-elimination resulted in increased electric mobility of miR-2940-3p. Yet, the migration of DENV2 piRNA bands was not affected by beta-elimination, indicating that their 3′ terminal nucleotides are modified, most likely methylated (Figure 2D). Since piRNA methylation occurs after loading into PIWI protein complexes, these data suggest that the identified DENV2 piRNAs are mature piRNAs associated with a PIWI protein.

To identify which PIWI proteins are required for the biogenesis of DENV2 piRNAs in Aag2 cells, we individually knocked down expression of all the eight *Aedes* PIWI proteins and analyzed the production of vpiRNAs by northern blot. We confirmed knockdown efficiency of roughly 90% for the four PIWI proteins that are detectable by RT-qPCR in Aag2 cells (i.e. Piwi4, Piwi5, Piwi6, Ago3; Figure 2E). Expression levels of Piwi1-3 and Piwi7 were too low to allow reliable quantification. DENV2 piRNAs were almost undetectable upon knockdown of Piwi5 and Ago3 and clearly reduced upon knockdown of Piwi6 (Figure 2F). These data confirm that the 25-30 nt population of DENV2-derived small RNAs are *bona fide* piRNAs that require host PIWI proteins for their biogenesis. To test whether knockdown of PIWI expression results in enhanced DENV2 replication, we performed RT-qPCR to compare viral RNA levels in the different knockdown conditions. We found that none of the knockdowns resulted in a significant change in viral RNA levels (Figure S2A). Yet, also knockdown of the well-established antiviral factor Ago2 (53) only resulted in a minor, statistically not-significant, increase of viral RNA replication although knockdown efficiency was higher than 90% (Figure S2B). This suggests that, in our hands, knockdown of small silencing pathway components in Aag2 cells is not suited to uncover robust antiviral activity against DENV2.

DENV2 miRNA-like small RNAs are not expressed in Aag2 cells

The significance of viral miRNA production from DENV genomic RNA is heavily debated (24-26, 54). To investigate whether viral miRNA-like molecules are produced in DENV2 infected Aag2 cells, we filtered 22-24 nt small RNA reads that map to the DENV2 genome with a maximum of one mismatch. In general, the number of 22 to 24 nt reads

was rather low (~5% of all DENV2 mapping reads) when compared to 21 nt siRNAs (~28%) and 25-30nt piRNAs (~40%). Furthermore, there were only four outstanding peaks that gave rise to a somewhat higher number of small RNAs (Figure 3A). All of them coincided with the position of a vpiRNA peaks (Figure 1D), suggesting that these small RNAs were by-products of vpiRNA production. Parallel analysis of the virus-derived reads using miRDeep2 did not identify convincing miRNA-like candidates: some reads mapped to two predicted hairpin sequences (genome positions: 9542 (+) strand; 4888 (-) strand), however the mapping patterns showed heterogeneity of the 5′ start sites and did not suggest Dicer processing (Figure S3A).

Recently, eight miRNA-like small RNAs were computationally predicted based on hairpin structures in the DENV2 genome, but they were not experimentally validated (54). We specifically looked for small RNA reads in our sequencing data mapping in the proximity of these predicted viral miRNAs, allowing a margin of 3nt around the start site. For each of the predicted miRNAs, we identified only very few (<20) reads in the combined set of DENV2-infected small RNA libraries (total of >3.7x10^7 reads of which >3.6x10^5 are DENV specific). In another publication, several 'miRNA-like' RNAs

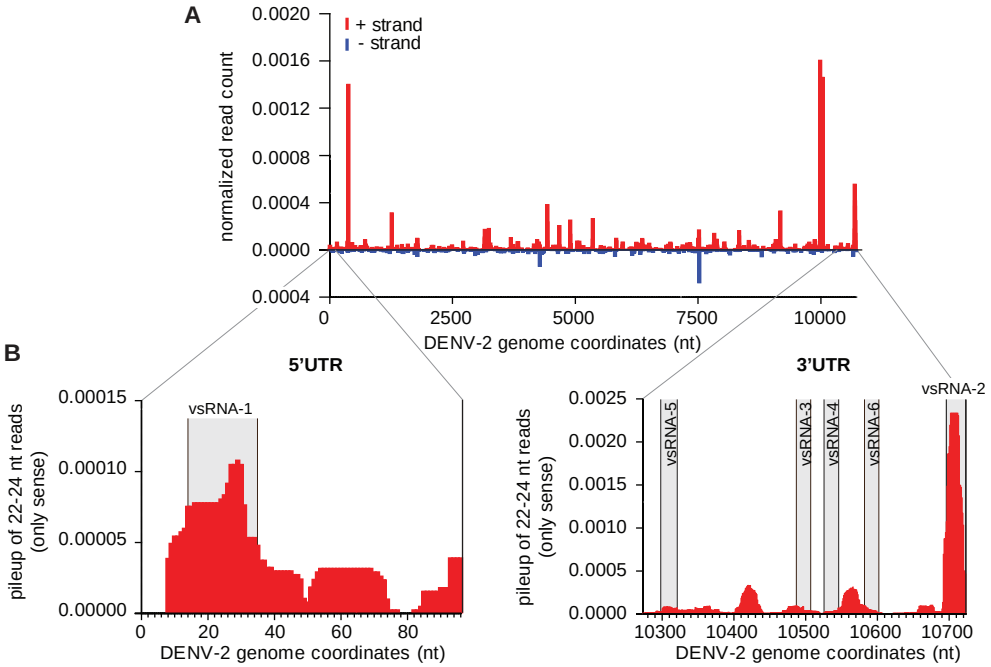


Figure 3. DENV miRNA-like small RNAs are not produced in infected Aag2 cells. (A) Distribution of 22-24 nt RNA reads across the DENV genome. Red bars indicate the number of 5′ ends of small RNAs that map to the (+) strand of the genome, blue bars represent the small RNAs mapping to the (-) strand. Read counts were normalized to the size of the corresponding library and the mean of the three libraries is plotted. **(B)** Pile-up of 22-24 nt small RNAs mapping to the (+) strand of DENV2 5′ (left panel) and 3′ UTRs (right panel), respectively. The mean of the three libraries is shown. Grey shadings highlight the boundaries of the mature DENV2 vsRNA sequences as reported in (24).

(termed vsRNA-1 to 6) were proposed to be produced from the 5′ and 3′ UTRs of the DENV2 RNA, based on the analysis of small RNA sequencing data (24). The DENV UTRs are indeed prone to form RNA structures and hairpins, which were suggested to be processed by the miRNA machinery into specific small RNA species (24). We specifically looked for the proposed vsRNA sequences in our dataset but could only identify a small RNA population that resembled vsRNA-2 located at the terminal hairpin of the DENV2 genome (Figure 3B). However, small RNAs mapping in that region showed a broad size distribution with the majority ranging in size from 26 to 28 nt, arguing against vsRNA-2 small RNAs being *bona fide* Dicer products (Figure S3B and S3C). These data suggest that the proposed DENV2 vsRNAs are not an abundant class of small RNAs.

Host miRNA levels are only mildly affected by DENV2 infection

Host miRNAs function as key regulators of gene expression and changes in miRNA expression have been reported during virus infections in various animal hosts, including mosquitoes. To assess host miRNA expression in response to DENV2 infection, we made use of the miRDeep2 toolkit to quantify miRNAs in the uninfected and DENV2-infected Aag2 small RNA libraries. DENV2 infection caused only minimal changes in miRNA levels (Figure 4). The expression of three and seven miRNAs was changed more than 2-fold up or down, respectively in response to DENV infection. Yet, the majority of differentially regulated miRNAs, including all up-regulated miRNAs, were poorly expressed (mean expression levels below twenty reads) making it hard to discriminate these expression changes from experimental noise due to low read counts. Collectively, these data suggest that miRNA expression in Aag2 cells is not heavily affected by DENV infection.

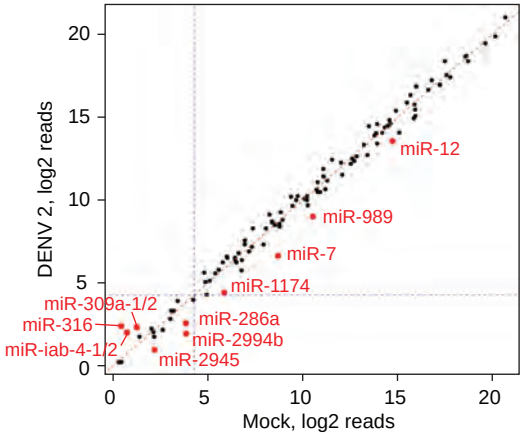


Figure 4. Host miRNA levels are not affected by DENV2 infection. Correlation of expression levels (log2-transformed) of known *Aedes aegypti* miRNAs in uninfected (x-axis) and DENV2 infected (y-axis) Aag2 cells. Mean expression of miRNAs across the three small RNA library replicates was determined using miRDeep2. Highlighted miRNAs are changed >2 fold. After p-value correction for multiple testing, no change in miRNA expression is statistically significant.

Novel *Aedes aegypti* miRNAs

The most recent version of miRBase (version 21) contains 101 *Aedes aegypti* miRNAs, which is considerably less than for other insect species including *Drosophila melanogaster* (fruitfly; 256 miRNAs), *Apis mellifera* (honey bee; 254 miRNAs) or *Bombyx mori* (silkworm; 487 miRNAs). We therefore suspected that the repertoire of published mosquito miRNAs is not yet complete and we set out to identify novel *Aedes aegypti* miRNAs in our small RNA sequencing data using miRDeep2. We obtained a list of 399 miRNA predictions (S2 Table), 73 of which were known miRNAs annotated in miRBase (51, 55, 56). We also confirmed 16 miRNAs that were recently reported by Akbari *et al*. or Hu *et al*., but were not yet available in miRBase (57, 58) (Figure S4). The remaining 310 predictions of miRNA hairpins were manually inspected for novel miRNAs using a similar approach as described in (59, 60), based on three criteria. First, only hairpin predictions that were supported by at least 1000 mature miRNA reads or those with at least one predicted miRNA* strand were retained. If a miRNA had an identical seed to a known insect miRNA, it was also retained irrespectively of read count or the presence of a miRNA*. In total 140 predictions met these criteria. Second, mature miRNAs were inspected for a homogeneous 5′ end of the supporting small RNA reads, defined as having at least 80% of the miRNA reads starting at the same nucleotide. 68 miRNA hairpin predictions met this requirement. Third, the predicted miRNA duplex was checked to resemble a (near) perfect Dicer product, defined as mapping to the stem of a stem-loop structure with one, two or three nucleotide overhangs at the 3′ end. Using this approach, we identified 31 unique mature miRNAs sequences mapping to 39 predicted miRNA hairpins. Nine of the 31 mature miRNA predictions did not have reads mapping to the star strand, but were supported by a homogeneous 5′ end in combination with a seed-match to an insect miRNA or having >1000 reads (Table 1, Figure 5). Further inspection identified one of these predictions to be derived from a tRNA which has therefore been removed from the list of predicted miRNAs.

miRNA offset RNAs in Aag2 cells

During the analysis of miRNA predictions, we noted the expression of specific small RNAs adjacent to the mature miRNA and miRNA* star strands. These miRNA offset RNAs (moRs) have been detected in small RNA deep-sequencing data from invertebrates, simple chordates, vertebrates and even viruses (61-68). In total, we identified moRs for 27% (24/89) of previously reported miRNA hairpins. In many of these cases (nine out of 24), the number of moRs per hairpin is below one thousandth of the number of mature miRNA reads. Others were more abundant, with moR reads accumulating up to 3.6% of the number of mature miRNAs. In a single instance, miR-11894b-1, the number of moRs reached 10.8% of the number of mature miRNA reads. In agreement with previous findings in *Drosophila* (63) 5′ moRs were more abundant

Figure 5. Novel *Aedes aegypti* pre-miRNAs. Hairpin structures of novel pre-miRNAs, as predicted using the RNAfold algorithm. Red letters indicate the position of the predicted mature miRNA sequence. Blue letters indicate the miRNA* strand. No miR* strand was found for ten predictions, representing eight mature miRNAs; these predictions are supported either by a seed-sequence known in insects or by high read counts. The mature miRNA sequences of aae-miR11908, aae-miR-11909 and aae-miR-11926 map to multiple different hairpins in the *Aedes* genome. For aae-miR-11914 and aae-miR-11920, the entire hairpin is encoded at multiple locations in the *Aedes* genome, as specified in table 1.

Table 1. Novel *Aedes aegypti* microRNAs

Name	ID	Position of the hairpin(s)	Mature miRNA mean read count*		Sequence of the mature miRNA	length
			Uninf.	DENV2		
aae-miR-11900†	MI0037941	supercont1.245 [-]: 1561005-1561086	79.7	31.9	auuuuuuugacuguaauuuuauag	24
aae-miR-11901	MI0037942	supercont1.885 [+]: 342921-342978	6.5	4.1	caucacagaauuguuuuuacug	22
aae-miR-11902†	MI0037943	supercont1.71 [+]: 1610981-1611037	2.6	5.0	aacuucaacgaagcguucggcuu	23
aae-miR-11903a	MI0037944	supercont1.484 [-]: 464269-464333	2.4	1.8	aacguuacaaaucguaaggcgag	23
aae-miR-11904†	MI0037945	supercont1.379 [+]: 580338-580393	0.6	1.4	ugaacaagaaugcugagaggau	22
aae-miR-11905	MI0037946	supercont1.14 [-]: 906240-906306	1.3	0.6	uaucgcgaguacuaaacaccuc	22
aae-miR-11906†	MI0037947	supercont1.517 [+]: 749300-749352	1.0	0.7	agagauugcaaggcaggcaggc	22
aae-miR-11907	MI0037948	supercont1.160 [+]: 1335655-1335712	0.5	0.4	uuuuaggaggucaaaaucgcgu	22
aae-miR-11908-1/2†	MI0037949	supercont1.70 [-]: 443560-443626	0.7	0.4	agcauuuuuacaccucaggac	21
	MI0037950	supercont1.369 [-]: 1123285-1123351				
aae-miR-11909-1/2†	MI0037951	supercont1.8 [-]:934642-934718	0.5	0.3	accguuacgcgcauauaauuug	22
	MI0037952	supercont1.41 [-]: 1367904-1367977				
aae-miR-11910†	MI0037953	supercont1.5 [+]: 2047651-2047726	0.3	0.3	aucgaucgcauccgucugaccu	22
aae-miR-11911	MI0037954	supercont1.89 [+]: 2726022-2726075	2.1	2.2	aguugauccaaguagucuugccu	23
aae-miR-11912	MI0037955	supercont1.48 [-]: 2566398-2566441	2.6	3.1	gugugugaaccguuggcggc	20
aae-miR-11913	MI0037956	supercont1.1336 [+]: 85845-85922	2.0	2.9	aauguuggacaacugcaaggu	21
aae-miR-11914-1/2/3	MI0037957	supercont1.1893 [-]: 19601-19653	1.5	2.3	ucacuuuuuagugacuugguc	21
	MI0037958	supercont1.224 [+]: 1335173-1335225				
	MI0037959	supercont1.222 [-]: 1778752-1778804				
aae-miR-11903b	MI0037960	supercont1.701 [+]: 403301-403361	1.8	0.7	cgccauaugauuuguaacucu	21
aae-miR-11915	MI0037961	supercont1.123 [-]: 1296986-1297043	0.5	0.6	cgagauacggagagauugcgaca	23
aae-miR-11916†	MI0037962	supercont1.916 [-]: 285767-285846	22.0	28.1	gaugccucguaaagcaacuggac	23
aae-miR-11917	MI0037963	supercont1.135 [+]: 1523348-1523407	3.1	4.0	cugaaaacuuaaucgauugguc	22
aae-miR-11918	MI0037964	supercont1. 151[-]: 1291075-1291138	3.1	2.4	caugaacgacgacgugacgccg	22
aae-miR-11919	MI0037965	supercont1.235 [-]: 964754-964810	3.3	3.0	uagcuagguugcggugcacugcu	23
aae-miR-11920-1/2/3/4	MI0037966	supercont1.441 [-]: 390672-390730	0.9	2.0	cccaucaacugcugaacuguuuu	23
	MI0037967	supercont1.49 [+]: 557677-557735				
	MI0037968	supercont1.16 [+]: 2615847-2615905				
	MI0037969	supercont1. 496 [+]: 473980-474038				
aae-miR-11921	MI0037970	supercont1.339 [+]: 1278980-1279041	2.5	2.0	aaaugggacugauaugcgaguau	23
aae-miR-11922	MI0037971	supercont1.551 [+]: 468137-468200	1.6	1.7	uucaggagaucaucgagguagc	22
aae-miR-11923	MI0037972	supercont1.220 [-]: 213093-213151	1.6	2.1	acaacggcagccggaacgaucu	22
aae-miR-11924	MI0037973	supercont1.657 [-]: 493003-493076	1.2	0.6	uagaaccugguagaauucggca	22
aae-miR-11925	MI0037974	supercont1.164 [-]: 201696-201781	0.4	0.9	cugucgagccgguugaaccac	21
aae-miR-11926-1/2	MI0037975	supercont1.607 [-]: 541794-541859	0.7	0.6	uuggacuggcaugcguuuaugg	22
	MI0037976	supercont1.21 [-]: 2803780-2803844				
aae-miR-11927	MI0037977	supercont1.97 [-]: 2388289-2388362	0.5	0.5	caaaagaucuggcuacacuga	21
aae-miR-11928	MI0037978	supercont1.215 [-]: 177863-177948	0.3	0.2	uuccgaaguugggugucucgc	21

miRNA predictions above the first thick line have a seed sequence that is present in an insect species, miRNAs between the two thick lines have a seed sequence that is present in a metazoan species, and miRNAs below the second thick line have seed sequences that are not present in any metazoan miRNA.

* The mean read count is normalized to the size of the corresponding small RNA sequencing library and presented as reads per million.

† miRNA prediction is supported by a seed match to known insect miRNAs or >1000 reads (equivalent to appr. 80 rpm), but not by the presence of a star strand.

than 3′ moRs (Figure 6A). moRs have been proposed to be the by-products of Drosha cleavage and in line with this suggestion, we found the 3′ end of 5′ moRs and the 5′ end of 3′ moRs to be fixed, reflecting potential Drosha cleavage sites (Figure 6B and 6C). In contrast, the ends of moRs facing the termini of the miRNA stemloop were less well defined, suggesting that they are processed by exonuclease activity (Figure 6B). Why certain miRNA hairpins are prone to accumulation of moRs remains unclear.

DISCUSSION

Small RNA pathways critically influence the outcome of virus infections in many host organisms, including plants, fungi, invertebrates, and vertebrates (3-5, 15, 69). In plants and invertebrates, siRNA-mediated antiviral immunity is key to the defence against a broad range of virus infections. In *Aedes* mosquitoes, viral siRNAs were detected from several virus families, including *Togaviridae*, *Flaviviridae*, *Bunyaviridae*, and *Reoviridae* (5). In line with previous reports, our analysis identified viral siRNAs derived from the entire genomic RNA of DENV2 in *Aedes aegypti* cells (35, 36). These vsiRNAs are produced in roughly equal amounts from the (+) strand and the (-) strand of the virus, indicating that the dsRNA replication intermediates serve as substrate for Dcr2. Upon knockdown of either Dcr2 or Ago2 in whole *Aedes* mosquitoes, DENV2 titres and transmission are enhanced, underlining the pivotal role of RNA interference in limiting DENV2 replication (53).

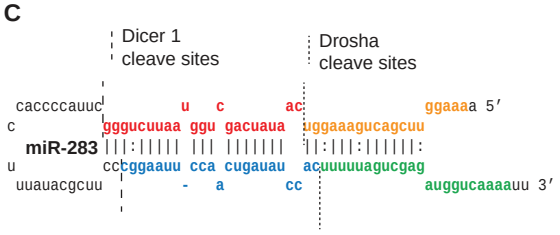
Besides siRNAs, our DENV2 infected small RNA libraries contained a substantial number of virus-mapping reads in the size range of piRNAs. Their expression was confirmed using small RNA northern blotting, validating that these small RNA reads were no sequencing artefacts. Only very few DENV2 genomic locations near the 3′ end of the DENV genome give rise to these vpiRNAs but the origin of this spiky pattern remains obscure. We hypothesized that perhaps endogenous, transposon-derived piRNAs would loosely bind the DENV2 genome at these positions triggering the production of secondary vpiRNAs. A similar mechanism has been suggested to initiate piRNA production from specific mRNAs in *Drosophila* (28). However, various mapping strategies allowing small RNA alignment with up to six mismatches did not uncover endogenous piRNAs that could trigger vpiRNA production at the observed positions.

▶ **Figure 6. miRNA offset RNAs in Aag2 cells. (A)** List of miRNA hairpins that give rise to moRs (sorted by the total moR count). The combined miRNA and moR read count from all six deep-sequencing libraries is shown. Highlighted miRNAs are described in more detail in panel B. **(B)** Three examples of mature miRNAs and moRs mapping to miRNA hairpins. The height of the bar (on log scale) reflects the number of reads covering the corresponding nucleotide position. The total amount of miRNA/moR reads is indicated below each bar stack. The most abundant miRNA/moR sequence is highlighted using the following color coding: orange, 5′ moR; red, mature miRNA; blue, miRNA*; green, 3′ moR. The dashed vertical line marked with a 'd' reflects the putative Drosha cleavage site. **(C)** miR-283 hairpin with 5′ and 3′ miRNA/moR sequences highlighted with colored nucleotide letters (see panel B). The sites of Dcr1 and Drosha cleavage are indicated by the dashed lines.

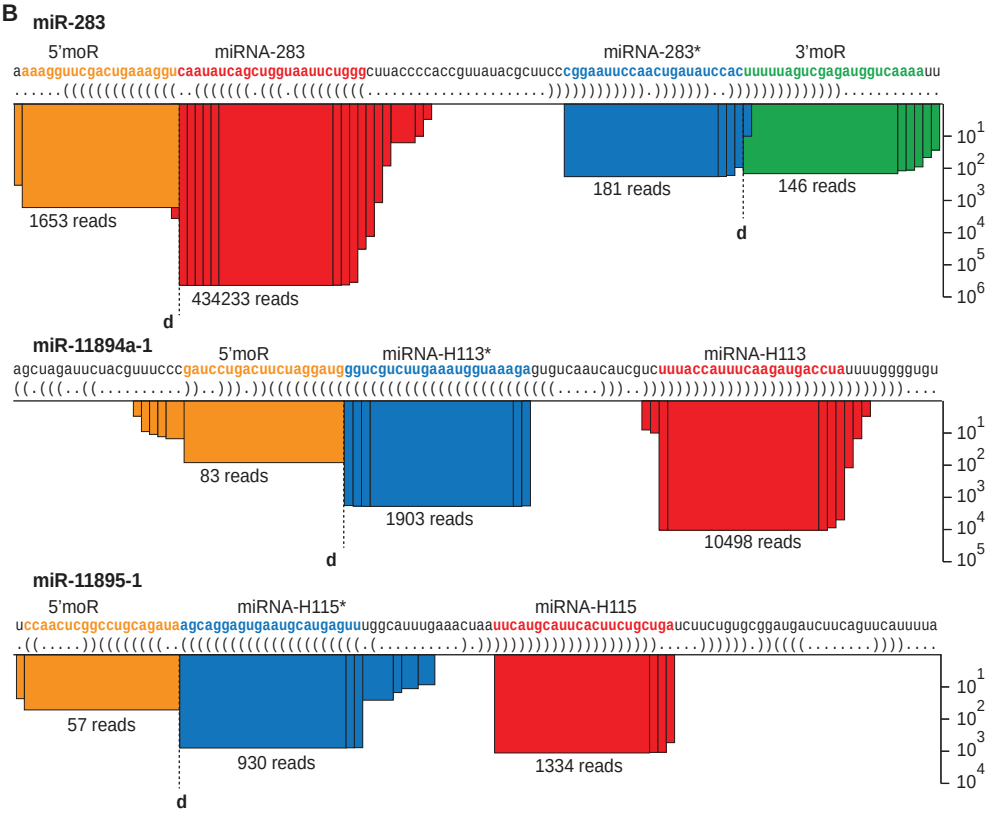
A

miRNA name	5' moRs	5' miRs	3' miRs	3' moRs
miR-283	1653	434233	181	146
miR-11894b-1	356	152	3134	0
miR-13	295	3762	38861	0
miR-11894a-2	256	1903	10495	0
miR-87	146	1483	4218	0
miR-998	114	635	26078	0
miR-H86	91	538	10498	22
miR-9c	89	36430	12633	0
miR-11894a-1	83	1903	10498	0
miR-317	20	11691	220396	35
miR-11895-1	57	930	1334	0
miR-12	36	41049	14355	0
miR-79	14	74	472	0
miR-275	10	39207	277074	0
miR-279	7	103	125836	0
miR-307	7	0	194	0
miR-184	4	5366	23700590	0
miR-306	4	908231	0	0
miR-988	4	156	4809	0
miR-1890	4	2	956	0
miR-2946	3	0	207	0
miR-7	2	752	133	0
miR-11	2	59893	874761	0
miR-92b	2	562	852126	0

C



B



Previous analyses of DENV2-derived small RNAs identified vpiRNAs in *Aedes aegypti* mosquitoes or Aag2 cells (35, 36). Also in the viral small RNA population reported by Scott and colleagues (35), a major small RNA spike is located near the 3′ end of the DENV genome. Since the entire population of viral RNA is analyzed in this study, it is however hard to assess which type of small RNA contributes to the spike. Interestingly, the spiky genome distribution of vpiRNAs is also recapitulated in adult *Aedes* mosquitoes. Although in this study the exact location of vpiRNA spikes differs from the positions we found in Aag2 cells, these data suggest that similar mechanisms may be responsible for piRNA biogenesis in Aag2 cells and adult mosquitoes. Yet, genetic evidence for the PIWI protein dependency of vpiRNA production *in vivo* is lacking. Of note, piRNAs are far less abundant in the small RNA libraries reported by Scott *et al*. (35) and Hess *et al.* (36) compared to our data from Aag2 cells. These differences might be due to different experimental conditions, including the chosen MOI, the time point of sampling, or to differences in small RNA library preparation and sequencing methodology. In addition, the specific viral strains may critically influence the accumulation of vpiRNAs. We have tested two laboratory-adapted DENV2 strains which both give rise to vpiRNAs, and it would be interesting to test if pathogenic strains from DENV2 endemic areas would show similar phenotypes.

Using knockdown of PIWI proteins, we identified Piwi5, Ago3 and, to a lesser extent, Piwi6 as responsible for the production of vpiRNAs in Aag2 cells. Therefore, DENV2 vpiRNA biogenesis in Aag2 cells relies on a similar set of PIWI proteins as SINV vpiRNAs, which also depend on Piwi5 and Ago3 (37). We have recently proposed that *Aedes* PIWI proteins are specialized in producing piRNAs from various sources. Whereas in mosquito cells piRNA biogenesis from transposons directly and indirectly depends on Piwi4-6 and Ago3, piRNA biogenesis from SINV predominantly requires Piwi5 and Ago3 only (37). The additional involvement of Piwi6 for piRNA production from DENV2 suggests that *Aedes* PIWI proteins are even further specialized towards RNA substrates from different viruses. This may be caused by virus-specific sequence elements or structures that are preferentially recognized by certain PIWI proteins. Alternatively, but not mutually exclusive, differences in replication strategies or replication sites might favour recognition of viral RNA by distinct sets of PIWI proteins.

The almost complete loss of DENV2 piRNAs upon knockdown of Piwi5 and Ago3 indicates that both proteins are equally important for vpiRNA biogenesis in Aag2 cells, similar to piRNA biogenesis during SINV infection (37). SINV piRNAs are produced by a two-step amplification mechanism that resembles ping-pong amplification of transposon piRNAs in *Drosophila* (27, 29). During this process, a piRNA-loaded PIWI protein (Piwi5 in *Aedes* or Aubergine in *Drosophila*) slices a complementary target RNA and transfers the 3′ slicer products as the new piRNA precursor to a second PIWI protein (Ago3 in *Aedes* and *Drosophila*). From this precursor, an Ago3-bound secondary piRNA is

produced that in turn is able to slice a target RNA, giving rise to a new piRNA precursor. This precursor will be matured to generate the same primary piRNA sequence that initiated the amplification. Therefore, this model predicts the presence of piRNAs derived from both strands. Although also during SINV infection (-) strand derived piRNAs are only a minor fraction, they can be identified as the primary piRNAs by a nucleotide bias that is characteristic for Piwi5/Aub bound piRNAs (uridine at position one). During DENV2 infection, 25-30 nt reads from the (-) strand are extremely scarce and they do not have the nucleotide bias that would classify them as primary piRNAs. Therefore, exactly how the production of the secondary, (+) strand vpiRNAs is triggered or whether a different, amplification-independent mechanism is responsible for their production, remains unclear.

The expression of miRNAs from DENV2 genomic RNA is still debated. Based on small RNA sequencing data, Hussain and Asgari have described a set of six viral small RNAs that have miRNA-like properties (24). Inhibition of one of them, vsRNA-5, by complementary RNA molecules strongly enhances DENV virus replication (24). However, expression of these viral small RNAs is generally not high and the relevance of such a lowly abundant small RNA during the exponential growth of a virus was therefore questioned (25). In our dataset, we find a high number of specific vsRNA reads for only vsRNA-2. For all the other predicted vsRNAs we find no or very low numbers of reads. vsRNA-2 is located on a hairpin at the very end of the DENV2 genome. This strongly resembles KUN-miRNA1, a viral small RNA expressed in mosquito cells infected with West Nile virus, a related flavivirus (70). KUN-miR1 is 21nt in size and its expression is Dcr1-dependent. In contrast, our data demonstrate that vsRNA-2 has a broad size distribution of primarily 26-28 nt, arguing against it being a canonical Dicer-dependent miRNA. Altogether, these data support the notion that miRNA-like small RNAs from DENV2 are extremely lowly abundant, with a questionable role in the regulation of viral replication.

Modulation of host miRNAs after virus infection may be a mechanism that coordinates gene expression during the course of the immune response. Alternatively, it may be a consequence of a viral strategy to manipulate host gene expression. Comparing uninfected with DENV2-infected Aag2 cells showed that the expression of almost all miRNAs was unchanged upon infection. Only a handful of miRNAs were up or down-regulated after exposure to DENV2. The fold changes ranged from approximately 4 fold up to 4 fold down. In whole *Aedes aegypti* mosquitoes, a total of 31 miRNAs were recently shown to be differentially regulated following infection at three different time points (17). The set of differentially expressed miRNAs is inconsistent between the different analyzed time points (2, 4 and 9 days post infection, dpi), but the number of differentially expressed miRNAs was higher at nine dpi than at two or four dpi. These data suggest that changes in miRNA expression may be more prominent in a long-term

infection and we thus cannot exclude the possibility that prolonged infection of Aag2 cells may result in more pronounced changes in miRNA expression. Alternatively, the observed miRNA changes in adult mosquitoes might not directly happen in infected cells *per se*, but could reflect an indirect effect of homeostatic or metabolic responses during the infection. It should be remembered that miRNA expression can be highly cell-type specific; if miRNA levels are responsive to DENV2-infection only in selected cell types within the entire mosquito, we may miss those in our Aag2 cell-based assays.

In summary, here we provide an in-depth analysis of small RNAs in DENV2 infected Aag2 cells in comparison to uninfected cells. Aag2 cells provide a powerful model system for studying biochemical details of small RNA biogenesis pathways, as they are fully competent in producing siRNAs, miRNAs and piRNAs originating from both virus and host. Our analyses add both DENV-derived piRNAs and novel *Aedes aegypti* miRNAs to the small RNA repertoire in this medically important virus-host interaction.

ACKNOWLEDGEMENTS

We thank members of the Van Rij laboratory for fruitful discussions. Furthermore, we would like to thank Dr. Peter Mason (Department of Pathology, University of Texas Medical Branch, Galveston, TX, USA) for kindly providing the DENV NS1 antibody, Dr. Jolanda Smit (Department of Medical Microbiology, University Medical Center Groningen, The Netherlands) for providing DENV2 NGC and BHK-15 cells, and Dr. Claire Huang (Centers for Disease Control and Prevention, Fort Collins, CO, USA) for providing DENV2 16681. We would also like to thank Sam Griffith-Jones and the team from the miRBase repository for assistance with assigning miRNA names.

REFERENCES

1. Weaver SC, Reisen WK. Present and future arboviral threats. Antiviral Res. 2010;85(2):328-45.
2. Lambrechts L, Scott TW. Mode of transmission and the evolution of arbovirus virulence in mosquito vectors. Proc Biol Sci. 2009;276(1660):1369-78.
3. Gammon DB, Mello CC. RNA interference-mediated antiviral defense in insects. Curr Opin Insect Sci. 2015;8:111-20.
4. Blair CD, Olson KE. The role of RNA interference (RNAi) in arbovirus-vector interactions. Viruses. 2015;7(2):820-43.
5. Bronkhorst AW, van Rij RP. The long and short of antiviral defense: small RNA-based immunity in insects. Curr Opin Virol. 2014;7:19-28.
6. Weber F, Wagner V, Rasmussen SB, Hartmann R, Paludan SR. Double-stranded RNA is produced by positive-strand RNA viruses and DNA viruses but not in detectable amounts by negative-strand RNA viruses. J Virol. 2006;80(10):5059-64.
7. Son KN, Liang Z, Lipton HL. Double-Stranded RNA Is Detected by Immunofluorescence Analysis in RNA and DNA Virus Infections, Including Those by Negative-Stranded RNA Viruses. J Virol. 2015;89(18):9383-92.
8. Sabin LR, Zheng Q, Thekkat P, Yang J, Hannon GJ, Gregory BD, et al. Dicer-2 processes diverse viral RNA species. PLoS One. 2013;8(2):e55458.

9. Galiana-Arnoux D, Dostert C, Schneemann A, Hoffmann JA, Imler JL. Essential function in vivo for Dicer-2 in host defense against RNA viruses in drosophila. Nat Immunol. 2006;7(6):590-7.
10. van Rij RP, Saleh MC, Berry B, Foo C, Houk A, Antoniewski C, et al. The RNA silencing endonuclease Argonaute 2 mediates specific antiviral immunity in Drosophila melanogaster. Genes Dev. 2006;20(21):2985-95.
11. van Mierlo JT, Bronkhorst AW, Overheul GJ, Sadanandan SA, Ekstrom JO, Heestermans M, et al. Convergent evolution of argonaute-2 slicer antagonism in two distinct insect RNA viruses. PLoS Pathog. 2012;8(8):e1002872.
12. Ha M, Kim VN. Regulation of microRNA biogenesis. Nat Rev Mol Cell Biol. 2014;15(8):509-24.
13. Bartel DP. MicroRNAs: Target Recognition and Regulatory Functions. Cell. 2009;136(2):215-33.
14. Jonas S, Izaurralde E. Towards a molecular understanding of microRNA-mediated gene silencing. Nat Rev Genet. 2015;16(7):421-33.
15. Libri V, Miesen P, van Rij RP, Buck AH. Regulation of microRNA biogenesis and turnover by animals and their viruses. Cell Mol Life Sci. 2013;70(19):3525-44.
16. Yan H, Zhou Y, Liu Y, Deng Y, Chen X. miR-252 of the Asian tiger mosquito Aedes albopictus regulates dengue virus replication by suppressing the expression of the dengue virus envelope protein. J Med Virol. 2014;86(8):1428-36.
17. Campbell CL, Harrison T, Hess AM, Ebel GD. MicroRNA levels are modulated in Aedes aegypti after exposure to Dengue-2. Insect Mol Biol. 2014;23(1):132-9.
18. Shrinet J, Jain S, Jain J, Bhatnagar RK, Sunil S. Next generation sequencing reveals regulation of distinct Aedes microRNAs during chikungunya virus development. PLoS Negl Trop Dis. 2014;8(1):e2616.
19. Slonchak A, Hussain M, Torres S, Asgari S, Khromykh AA. Expression of mosquito microRNA Aae-miR-2940-5p is downregulated in response to West Nile virus infection to restrict viral replication. J Virol. 2014;88(15):8457-67.
20. Etebari K, Osei-Amo S, Blomberg SP, Asgari S. Dengue virus infection alters post-transcriptional modification of microRNAs in the mosquito vector Aedes aegypti. Sci Rep. 2015;5:15968.
21. Tycowski KT, Guo YE, Lee N, Moss WN, Vallery TK, Xie M, et al. Viral noncoding RNAs: more surprises. Genes Dev. 2015;29(6):567-84.
22. Rouha H, Thurner C, Mandl CW. Functional microRNA generated from a cytoplasmic RNA virus. Nucleic Acids Res. 2010;38(22):8328-37.
23. Shapiro JS, Varble A, Pham AM, Tenoever BR. Noncanonical cytoplasmic processing of viral microRNAs. RNA. 2010;16(11):2068-74.
24. Hussain M, Asgari S. MicroRNA-like viral small RNA from Dengue virus 2 autoregulates its replication in mosquito cells. Proc Natl Acad Sci U S A. 2014;111(7):2746-51.
25. Skalsky RL, Olson KE, Blair CD, Garcia-Blanco MA, Cullen BR. A „microRNA-like" small RNA expressed by Dengue virus? Proc Natl Acad Sci U S A. 2014;111(23):E2359.
26. Finol E. Are viral small RNA regulating Dengue virus replication beyond serotype 2? Proc Natl Acad Sci U S A. 2014;111(29):E2915-6.
27. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, et al. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell. 2007;128(6):1089-103.
28. Mohn F, Handler D, Brennecke J. Noncoding RNA. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. Science. 2015;348(6236):812-7.
29. Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, et al. A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in Drosophila. Science. 2007;315(5818):1587-90.
30. Han BW, Wang W, Li C, Weng Z, Zamore PD. Noncoding RNA. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. Science. 2015;348(6236):817-21.

31. Morazzani EM, Wiley MR, Murreddu MG, Adelman ZN, Myles KM. Production of virus-derived ping-pong-dependent piRNA-like small RNAs in the mosquito soma. PLoS Pathog. 2012;8(1):e1002470.
32. Vodovar N, Bronkhorst AW, van Cleef KW, Miesen P, Blanc H, van Rij RP, et al. Arbovirus-derived piRNAs exhibit a ping-pong signature in mosquito cells. PLoS One. 2012;7(1):e30861.
33. Schnettler E, Ratinier M, Watson M, Shaw AE, McFarlane M, Varela M, et al. RNA Interference Targets Arbovirus Replication in Culicoides Cells. Journal of Virology. 2013;87(5):2441-54.
34. Schnettler E, Donald CL, Human S, Watson M, Siu RW, McFarlane M, et al. Knockdown of piRNA pathway proteins results in enhanced Semliki Forest virus production in mosquito cells. J Gen Virol. 2013;94(Pt 7):1680-9.
35. Scott JC, Brackney DE, Campbell CL, Bondu-Hawkins V, Hjelle B, Ebel GD, et al. Comparison of Dengue Virus Type 2-Specific Small RNAs from RNA Interference-Competent and -Incompetent Mosquito Cells. Plos Neglected Tropical Diseases. 2010;4(10).
36. Hess AM, Prasad AN, Ptitsyn A, Ebel GD, Olson KE, Barbacioru C, et al. Small RNA profiling of Dengue virus-mosquito interactions implicates the PIWI RNA pathway in anti-viral defense. BMC Microbiol. 2011;11:45.
37. Miesen P, Girardi E, van Rij RP. Distinct sets of PIWI proteins produce arbovirus and transposon-derived piRNAs in Aedes aegypti mosquito cells. Nucleic Acids Res. 2015;43(13):6545-56.
38. Leger P, Lara E, Jagla B, Sismeiro O, Mansuroglu Z, Coppee JY, et al. Dicer-2- and Piwi-mediated RNA interference in Rift Valley fever virus-infected mosquito cells. J Virol. 2013;87(3):1631-48.
39. Brackney DE, Scott JC, Sagawa F, Woodward JE, Miller NA, Schilkey FD, et al. C6/36 Aedes albopictus cells have a dysfunctional antiviral RNA interference response. PLoS Negl Trop Dis. 2010;4(10):e856.
40. van Cleef KWR, Overheul GJ, Thomassen MC, Kaptein SJF, Davidson AD, Jacobs M, et al. Identification of a new dengue virus inhibitor that targets the viral NS4B protein and restricts genomic RNA replication. Antiviral Research. 2013;99(2):165-71.
41. Mason PW, Zugel MU, Semproni AR, Fournier MJ, Mason TL. The antigenic structure of dengue type 1 virus envelope and NS1 proteins expressed in Escherichia coli. J Gen Virol. 1990;71 (Pt 9):2107-14.
42. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001;25(4):402-8.
43. Kawaoka S, Katsuma S, Tomari Y. Making piRNAs in vitro. Methods Mol Biol. 2014;1093:35-46.
44. Pall GS, Hamilton AJ. Improved northern blot method for enhanced detection of small RNA. Nat Protoc. 2008;3(6):1077-84.
45. van Cleef KW, van Mierlo JT, Miesen P, Overheul GJ, Fros JJ, Schuster S, et al. Mosquito and Drosophila entomobirnaviruses suppress dsRNA- and siRNA-induced RNAi. Nucleic Acids Res. 2014;42(13):8732-44.
46. Blankenberg D, Gordon A, Von Kuster G, Coraor N, Taylor J, Nekrutenko A, et al. Manipulation of FASTQ data with Galaxy. Bioinformatics. 2010;26(14):1783-5.
47. Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, et al. Galaxy: a web-based genome analysis tool for experimentalists. Curr Protoc Mol Biol. 2010;Chapter 19:Unit 19 0 1-21.
48. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25.
49. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357-9.
50. Friedlander MR, Mackowiak SD, Li N, Chen W, Rajewsky N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res. 2012;40(1):37-52.
51. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 2014;42(Database issue):D68-73.
52. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.

53. Sanchez-Vargas I, Scott JC, Poole-Smith BK, Franz AW, Barbosa-Solomieu V, Wilusz J, et al. Dengue virus type 2 infections of Aedes aegypti are modulated by the mosquito's RNA interference pathway. PLoS Pathog. 2009;5(2):e1000299.

54. Ospina-Bedoya M, Campillo-Pedroza N, Franco-Salazar JP, Gallego-Gomez JC. Computational Identification of Dengue Virus MicroRNA-Like Structures and their Cellular Targets. Bioinform Biol Insights. 2014;8:169-76.

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

56. Li S, Mead EA, Liang S, Tu Z. Direct sequencing and expression analysis of a large number of miRNAs in Aedes aegypti and a multi-species survey of novel mosquito miRNAs. BMC Genomics. 2009;10:581.

57. Akbari OS, Antoshechkin I, Amrhein H, Williams B, Diloreto R, Sandler J, et al. The developmental transcriptome of the mosquito Aedes aegypti, an invasive species and major arbovirus vector. G3 (Bethesda). 2013;3(9):1493-509.

58. Hu W, Criscione F, Liang S, Tu Z. MicroRNAs of two medically important mosquito species: Aedes aegypti and Anopheles stephensi. Insect Mol Biol. 2015;24(2):240-52.

59. Tarver JE, Sperling EA, Nailor A, Heimberg AM, Robinson JM, King BL, et al. miRNAs: small genes with big potential in metazoan phylogenetics. Mol Biol Evol. 2013;30(11):2369-82.

60. Buck AH, Coakley G, Simbari F, McSorley HJ, Quintana JF, Le Bihan T, et al. Exosomes secreted by nematode parasites transfer small RNAs to mammalian cells and modulate innate immunity. Nat Commun. 2014;5:5488.

61. Ruby JG, Stark A, Johnston WK, Kellis M, Bartel DP, Lai EC. Evolution, biogenesis, expression, and target predictions of a substantially expanded set of Drosophila microRNAs. Genome Res. 2007;17(12):1850-64.

62. Shi W, Hendrix D, Levine M, Haley B. A distinct class of small RNAs arises from pre-miRNA-proximal regions in a simple chordate. Nat Struct Mol Biol. 2009;16(2):183-9.

63. Berezikov E, Robine N, Samsonova A, Westholm JO, Naqvi A, Hung JH, et al. Deep annotation of Drosophila melanogaster microRNAs yields insights into their processing, modification, and emergence. Genome Res. 2011;21(2):203-15.

64. Asikainen S, Heikkinen L, Juhila J, Holm F, Weltner J, Trokovic R, et al. Selective microRNA-Offset RNA expression in human embryonic stem cells. PLoS One. 2015;10(3):e0116668.

65. Zhou H, Arcila ML, Li Z, Lee EJ, Henzler C, Liu J, et al. Deep annotation of mouse iso-miR and iso-moR variation. Nucleic Acids Res. 2012;40(13):5864-75.

66. Ma H, Wu Y, Choi JG, Wu H. Lower and upper stem-single-stranded RNA junctions together determine the Drosha cleavage site. Proc Natl Acad Sci U S A. 2013;110(51):20687-92.

67. Donohoe OH, Henshilwood K, Way K, Hakimjavadi R, Stone DM, Walls D. Identification and Characterization of Cyprinid Herpesvirus-3 (CyHV-3) Encoded MicroRNAs. PLoS One. 2015;10(4):e0125434.

68. Meshesha MK, Veksler-Lublinsky I, Isakov O, Reichenstein I, Shomron N, Kedem K, et al. The microRNA Transcriptome of Human Cytomegalovirus (HCMV). Open Virol J. 2012;6:38-48.

69. Ding SW. RNA-based antiviral immunity. Nat Rev Immunol. 2010;10(9):632-44.

70. Hussain M, Torres S, Schnettler E, Funk A, Grundhoff A, Pijlman GP, et al. West Nile virus encodes a microRNA-like small RNA in the 3' untranslated region which up-regulates GATA4 mRNA and facilitates virus replication in mosquito cells. Nucleic Acids Res. 2012;40(5):2210-23.

SUPPLEMENTAL FIGURES

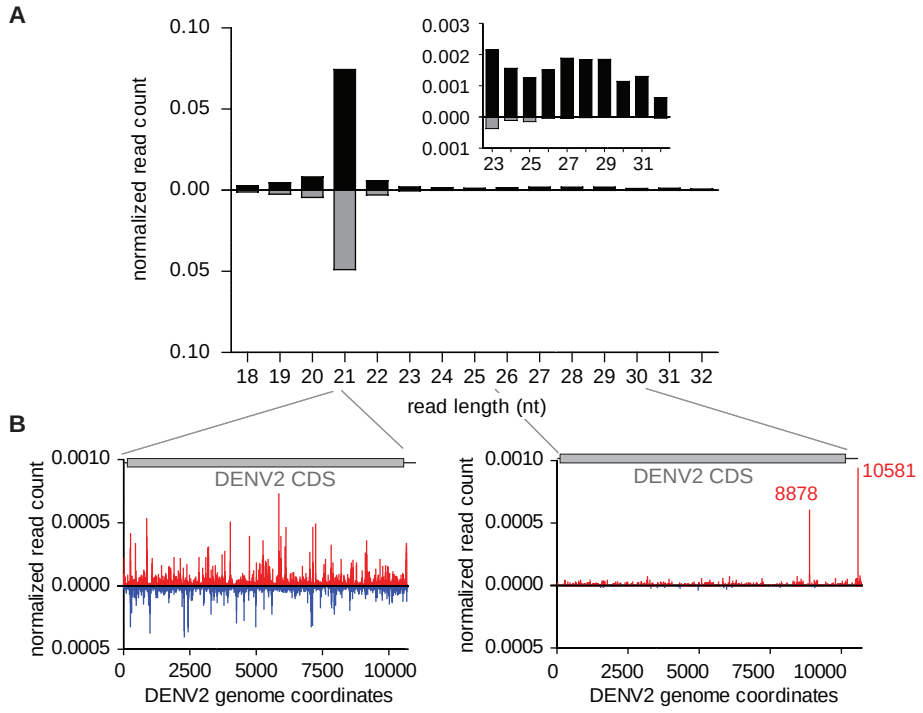


Figure S1. DENV small RNAs in *Aedes* mosquitoes. Re-analysis of small RNA sequencing data from DENV2-JAM1409 infected *Aedes* mosquitoes (9 days post infection) published by Hess *et al*. (1). **(A)** Size profile of small RNA reads mapping to the sense strand (black) or antisense strand (grey) of the DENV2 JAM1409 genome. Inlay shows reads of 23 to 32 nt in size, with a different scale for the y-axis. **(B)** Peaks indicate the number of 5'ends of small RNAs of 21 nt (left panel) or 25-30 nt (right panel) across the sense (red) or antisense (blue) strand of the viral genome. Read counts have been normalized to the depth of the library. Red numbers in B indicate the genome position of the 5' end of the small RNA peaks.

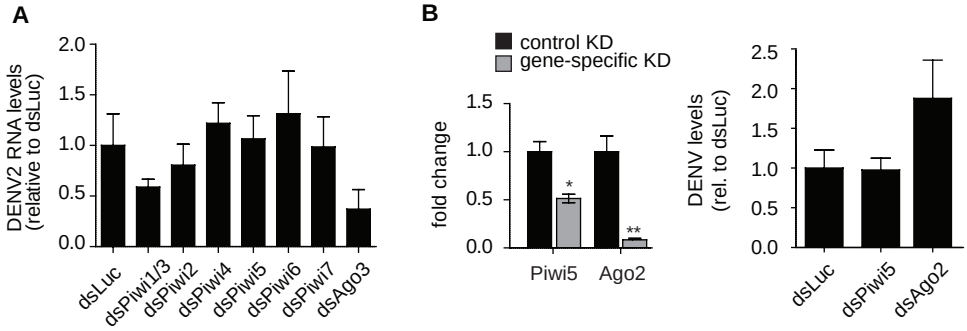


Figure S2. Effect of PIWI knockdown on DENV2 RNA levels. (A) Quantification of DENV2 RNA levels by RT-qPCR in the samples used for figure 2E and 2F. **(B)** Knockdown (KD) efficiency of Piwi5 and Ago2 (left panel) and the corresponding levels of DENV2 RNA (right panel) as assessed by RT-qPCR. Bars indicate mean +/- SEM of three independent experiments. Statistical significance was determined using two tailed, unpaired student t-test. * $P<0.05$; ** $P<0.01$.

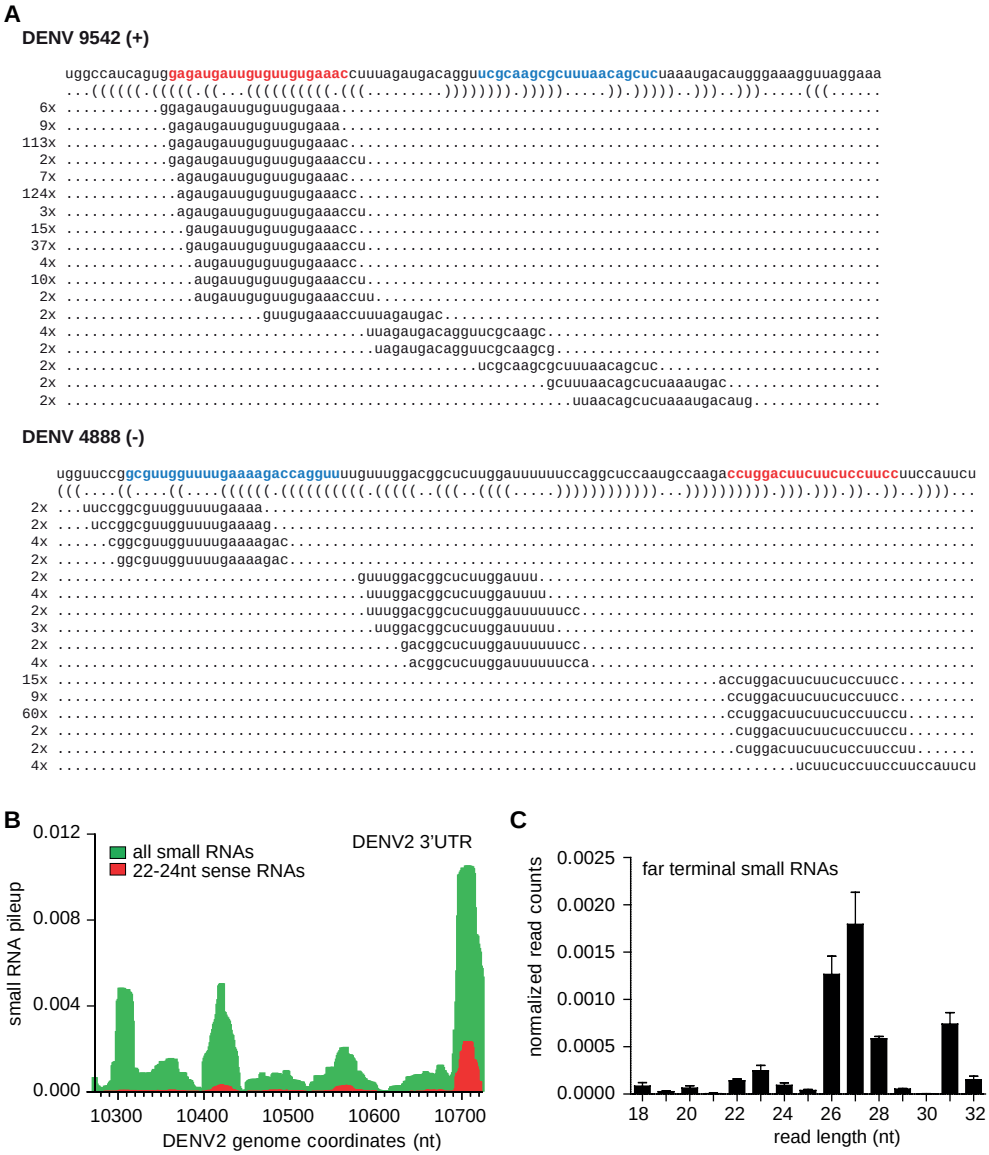


Figure S3. DENV miRNAs are not expressed in Aag2 cells. (A) miRNA predictions using miRDeep2 on the DENV2-NGC genome. The read counts of the individual sequences, combined from all the six sequencing libraries, that support the predictions are shown to the left. The predicted mature miRNA and the expected miRNA* sequences are indicated in red and blue, respectively. **(B)** Genome coverage of small RNAs mapping to the DENV2-NGC 3′ UTR. Bars represent the mean of the normalized coverage at each nucleotide position (*n*=3). Green bars show all small RNA reads mapping to the region. Red bars show the 22-24 nt reads derived from the viral (+) strand only. **(C)** Size distribution of (+) strand-derived small RNA reads that end at the 3′ terminus of the DENV2-NGC genome. Bars represent the average read count and SEM of the three deep-sequencing libraries normalized to their corresponding size (displayed as % of library).

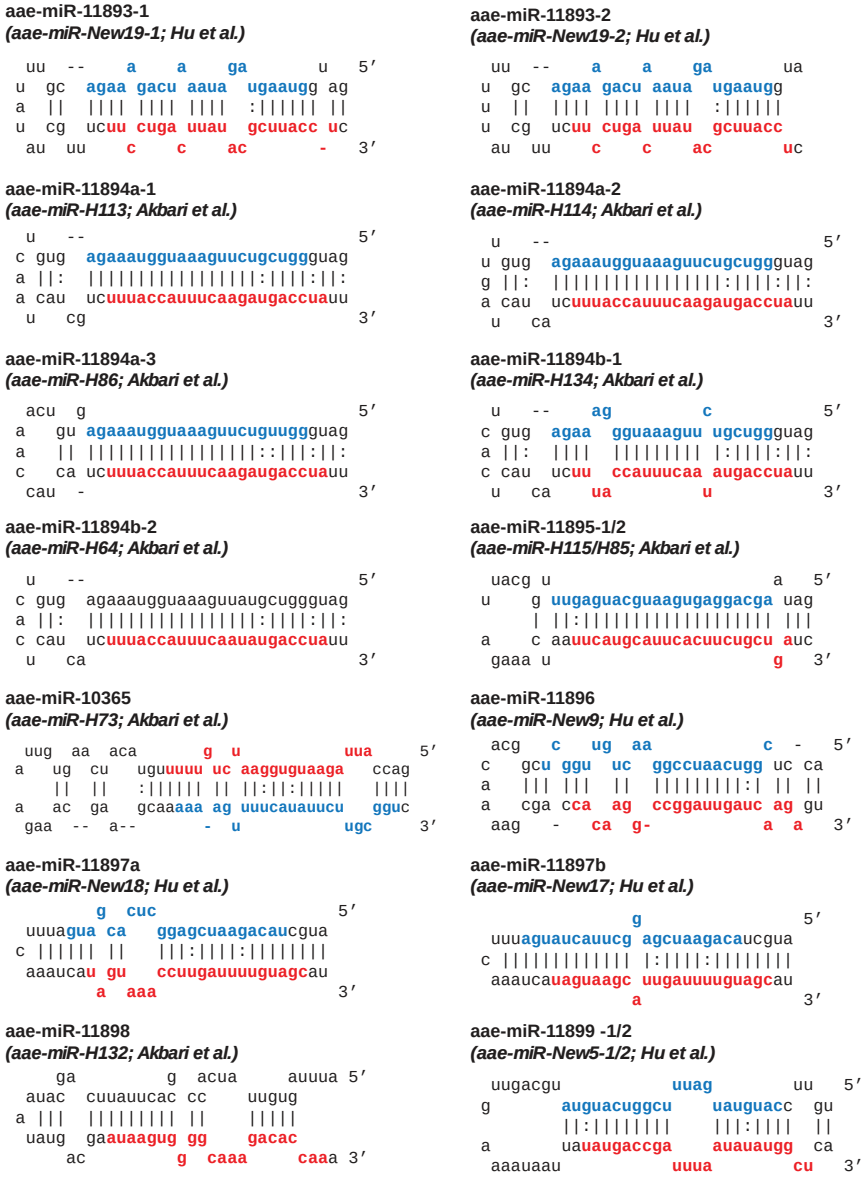


Figure S4. New *Aedes* miRNAs reported by Hu *et al.* or Akbari *et al.* Folding of miRNA hairpins that have been published by Hu *et al.* or Akbari *et al.* (2, 3), and have also been recovered by the miRDeep2 analyses of our libraries. These miRNAs were not yet published in the most recent version of miRBase. The folding of the hairpin was predicted using RNAfold. Red and blue letters indicate the position of the predicted, mature miRNA and miRNA* sequences, respectively. The shown miRNA names have been assigned by miRBase; the previous names given by the authors of the indicated studies are presented between brackets. The mature miRNA sequences of aae-miR-11893-1 maps to similar hairpins at two locations in the *Aedes* genome. Also aae-miR-11894a-1 and aae-miR-11894a-2 are identical sequences that map to two very similar hairpins. Aae-miR-11895-1 and aae-miR-11895-2 are two identical hairpins that map to two locations in the *Aedes* genome. Similarly, the mature aae-miR-11899 sequence maps to two identical hairpins in the genome.

Table S1. Oligonucleotides used in this study.

Name	Sequence
Primers used for the production of T7-flanked PCR products	
T7F-Luc	taatacgactcactatagggagaTATGAAGAGATACGCCCTGGTT
T7R-Luc	taatacgactcactatagggagaTAAAACCGGGAGGTAGATGAGA
T7F-Piwi1/3	taatacgactcactatagggagaCCACGCCCATCGTTTCAA
T7R-Piwi1/3	taatacgactcactatagggagaCCTCAGTTTGTTCACCATA
T7F-Piwi2	taatacgactcactatagggagaCCGTCCTACTTTCCAGCAC
T7R-Piwi2	taatacgactcactatagggagaGCGGCACTCCAGGGACAAT
T7F-Piwi4	taatacgactcactatagggagaCGTGGAAGTCCTTCTTCTCG
T7R-Piwi4	taatacgactcactatagggagaTGTCAGTTGATCGCTTCTCAA
T7F-Piwi5	taatacgactcactatagggagaGCCATACATCGGGTCAAAAT
T7R-Piwi5	taatacgactcactatagggagaCTCTCCACCGAAGGATTGAA
T7F-Piwi6	taatacgactcactatagggagaCAACGGAGGATCTTCACGAG
T7R-Piwi6	taatacgactcactatagggagaAATCGATGGCTTGATTTGGA
T7F-Piwi7	taatacgactcactatagggagaGTGGAGGTCGTGGAGGTAAC
T7R-Piwi7	taatacgactcactatagggagaGTTTGCGGTGTTTCCGTACT
T7F-Ago3	taatacgactcactatagggagaTGCTTACTCGTGTCGCGTAG
T7R-Ago3	taatacgactcactatagggagaGGCATGGCAGATCCAATACT
(quantitative) PCR primers	
F-Piwi4	TCTTCTTCTCCACCACAGCC
R-Piwi4	ATGGTGACCACCTCACAGTTAC
F-Piwi5	ACGGCATCACATCGAGACTC
R-Piwi5	CGACCTCCACGCTGTCCTC
F-Piwi6	TTTTCTTCCACCCCGAGCAG
R-Piwi6	AATACATTTGCGATGCGGCC
F-Ago3	CTCCAGACGACGGTTTTGGA
R-Ago3	GCAGGTACGAAATTGGCTGC
F-Ago2	ATTTGGCTCAAGATCAACGC
R-Ago2	GAGATCGTATGAAGCGGCCA
F-LAP	GTGCTCATTCACCAACATCG
R-LAP	AACTTGGCCGCAACAAATAC
F-DV2-NS1	AGAACTGAAGTGTGGCAGTGGGAT
R-DV2-NS1	TGCCCTCTTCATGAGCTTTCTGGA
Northern blot probes	
nDV2-9180+	GGTCTTCTAGTGTGATTCTTGTGTCCCAT
nDV2-9985+	CCCTGTTCCAGACTGTCAGCATGTCTTCCGT
nMiR-2940-3p	AGTGATTTATCTCCCTGTCGAC

Available online:

- **Table S2.** Spreadsheet of raw miRDeep2 output. Sheet 1 provides an overview of the selection procedure to reach the final miRNA predictions. Sheets 2 to 6 show the unfiltered miRNA predictions and the filtered miRNA predictions after the individual rounds of selection.
- **Dataset S1.** Data related to Figure 1C, 1D, 2E, 2F, 3, 4, S1, S2. S3B and S3C.

The data can be downloaded from:
http://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0004452#sec022

SUPPLEMENTAL REFERENCES

1. Hess AM, Prasad AN, Ptitsyn A, Ebel GD, Olson KE, Barbacioru C, et al. Small RNA profiling of Dengue virus-mosquito interactions implicates the PIWI RNA pathway in anti-viral defense. BMC Microbiol. 2011;11:45.
2. Akbari OS, Antoshechkin I, Amrhein H, Williams B, Diloreto R, Sandler J, et al. The developmental transcriptome of the mosquito Aedes aegypti, an invasive species and major arbovirus vector. G3 (Bethesda). 2013;3(9):1493-509.
3. Hu W, Criscione F, Liang S, Tu Z. MicroRNAs of two medically important mosquito species: Aedes aegypti and Anopheles stephensi. Insect Mol Biol. 2015;24(2):240-52.

Chapter 4

Histone-Derived piRNA Biogenesis Depends on the Ping-pong Partners Piwi5 and Ago3 in *Aedes aegypti*.

Erika Girardi*, Pascal Miesen*, Bas Pennings,
Lionel Frangeul, Maria-Carla Saleh, Ronald P. van Rij

Nucleic Acids Res. (2017) 45:4881-92

*equal contribution

ABSTRACT

The piRNA pathway is of key importance in controlling transposable elements in most animal species. In the vector mosquito *Aedes aegypti*, the presence of eight PIWI proteins and the accumulation of viral piRNAs upon arbovirus infection suggest additional functions of the piRNA pathway beyond genome defense. To better understand the regulatory potential of this pathway, we analyzed in detail host-derived piRNAs in *Ae. aegypti* Aag2 cells. We show that a large repertoire of protein-coding genes and non-retroviral integrated RNA virus elements are processed into genic piRNAs by different combinations of PIWI proteins. Among these, we identify a class of genes that produces piRNAs from coding sequences in an Ago3- and Piwi5-dependent fashion. We demonstrate that the replication-dependent histone gene family is a genic source of ping-pong dependent piRNAs and that histone-derived piRNAs are dynamically expressed throughout the cycle, suggesting a role for the piRNA pathway in the regulation of histone gene expression. Moreover, our results establish the Aag2 cell line as an accessible experimental model to study gene-derived piRNAs.

INTRODUCTION

Small RNA-guided gene regulation has come to light as a major, widely conserved mechanism across almost all eukaryotes (1). PIWI-interacting RNAs (piRNAs) are a class of ~25 to 30 nt small RNAs that associate with the PIWI subclass of Argonaute proteins to form gene regulatory piRNA induced silencing complexes (piRISCs). Specificity of piRISC is mediated by base-pairing between the piRNA sequence and a target RNA (2). In flies, primary piRNAs are processed from long single-stranded RNA precursors and are loaded into the PIWI proteins Piwi and Aubergine. These piRNAs show a strong bias for uridine at their 5′ end (1U). In the presence of target RNA, piRNA amplification by the the so-called ping-pong loop is initiated: antisense primary piRNAs mediate cleavage of the target and the 3′ cleavage fragment is processed into secondary piRNAs, which are loaded into a different PIWI protein, called Argonaute 3 (Ago3). This mechanism is at the origin of the ping-pong signature of piRNAs: a 1U bias for antisense piRNAs and a bias for adenosine at position 10 (10A) for sense piRNAs (2-4).

The piRNA pathway has been defined as an RNA-based defense system against transposon activity and many studies have addressed its role in maintaining genome stability in the germline (5). Despite this conserved function across species, increasing evidence suggests the presence of piRNAs and PIWI proteins in non-germline tissues (6,7). Expression of PIWI proteins in somatic tissues has been linked to stem cell renewal, maintenance, and regeneration in several primitive organisms (1, 8-12). Somatic piRNAs have been cloned from *Drosophila* (2, 13, 14), rhesus macaque and mouse tissues (2-4, 15, 16). Moreover, accumulation of piRNAs and PIWI proteins has been detected in human

somatic and cancer cells (5, 17, 18). However, the biogenesis and functions of somatic PIWI proteins and their associated piRNAs remain largely unexplored.

Previously, we and others identified an additional class of somatic piRNAs: virus-derived piRNAs in *Aedes* mosquitoes (6, 7,19-24). As the vector for several pathogenic human viruses, including dengue virus, Zika virus, and chikungunya virus, *Ae. aegypti* is one of the most medically important mosquito species. Its genome encodes eight PIWI family members (Piwi1–7 and Ago3), of which Ago3 clusters with *Drosophila* Ago3, whereas Piwi1-7 form clades distinct from *Drosophila* Piwi and Aubergine (25). We recently used the piRNA competent Aag2 cell line to investigate viral and transposon-derived piRNA biogenesis. Aag2 cells express the same PIWI genes present in adult mosquitoes and are fully competent in producing piRNAs via the ping-pong amplification mechanism (18). We demonstrated that Piwi5 and Ago3 are the core proteins of the viral piRNA ping-pong amplification loop, whereas additional PIWI proteins are involved in transposon-derived piRNA biogenesis (20).

Ae. aegypti has a large genome size (~1.4 Gb) of which half is composed of transposable elements (26). However, only 19% of the sequenced piRNAs map to transposable elements (TEs), suggesting that the remainder of small RNAs could arise from other genomic loci (27). Indeed, an increasing number of studies indicate that piRNAs may also arise from cellular non-coding and protein-coding genes studies in different animal species (5, 27-30). Gene-derived piRNAs generally derive from the 3′ untranslated regions (UTRs) and are produced in a ping-pong independent fashion in *Drosophila* ovaries, murine testes, *Xenopus* eggs, and *Anopheles gambiae* germline (3, 31-33).

Core histone proteins (H2A, H2B, H3, H4) are highly conserved proteins that play essential structural and functional roles in genome packaging and gene regulation in all eukaryotes. The organization of histone genes in one or more clusters is conserved from yeast to human. Although clustering is important for their transcriptional activation at the G1/S phase transition, histone transcripts are also regulated post-transcriptionally (34). Devoid of a poly(A) tail, replication-dependent histone mRNAs end in a highly conserved stem-loop (SL) structure that is responsible for their cell cycle regulated degradation at the end of the S phase. The conserved cis regulatory elements in mRNA, such as the SL motif and the purine-rich histone downstream element (HDE), and the machinery for histone mRNA 3′end processing are conserved in *Ae. aegypti* (35).

Here, we present evidence that specific PIWI proteins produce genic piRNAs in *Ae. aegypti* Aag2 cells. We show that coding sequences of replication-dependent histone genes are a major source of 3′end-modifed piRNAs, which accumulate in an Ago3-Piwi5 and ping-pong dependent fashion. Our results imply a new link between the piRNA pathway and histone gene expression and establish the Aag2 cell line as an experimental model to study genic piRNAs.

MATERIALS AND METHODS

Transfection of Aag2 cells

For immunoprecipitation (IP) analyses, Aag2 cells were transfected with expression plasmids encoding individual PIWI proteins (see Supplemental data). For knockdown experiments, Aag2 were transfected with dsRNA using X-tremeGENE HP (Roche) according to the manufacturer's instructions. To increase knockdown efficiency, Aag2 cells were re-transfected at 48 h after the first transfection. Where indicated, cells were infected with a Sindbis virus recombinant expressing GFP from a second subgenomic promoter (SINV-GFP). Unless stated differently, samples were harvested at 48 h post infection.

RNA and protein detection

For a detailed description of the experimental procedures for northern blot, RT-qPCR, strand specific RT-PCR, immunoprecipitation, western blot and and small RNA deep-sequencing, see Supplemental data. Oligonucleotide sequences are presented in Table S1.

Cell cycle analysis by flow cytometry

Sub-confluent Aag2 cells were treated with 0.2 mM Hydroxyurea (HU, Sigma) for 24 h and subsequently released by changing the medium. At 0, 2, 4, 6 and 8 hours post release (hpr), cells were harvested, washed with PBS, fixed overnight with ice-cold 70% ethanol at 4°C and stained in Staining buffer (50 μg/ml propidium iodide, 50 μg/ml Ribonuclease A, 3.8 mM tri-sodium citrate dehydrate, 0.1% Triton X-100) overnight at 4°C in the dark. Intracellular DNA content was then analyzed by flow cytometry on a BD FACSCalibur. FlowJo software was used for the analyses.

Statistical analysis

Experiments were repeated as biological replicate as indicated in the figure legends and data are expressed as mean ± standard deviation (SD). Unpaired two-tailed student's t-tests were used to determine statistical significance. A P-value of <0.05 was considered statistically significant. Graphs were plotted and statistical analyses were performed using GraphPad Prism (version 6.00 for Mac, GraphPad Software, La Jolla CA, USA).

RESULTS

Ae. aegypti coding genes are a source of piRNA-sized small RNAs

We have previously shown that different combinations of PIWI proteins can generate either transposon-derived or viral piRNAs in the *Ae. aegypti* Aag2 cell line (20, 21). Although it has been proposed that protein-coding genes may also be sources of piRNAs

in *Ae. aegypti* (27), the biogenesis mechanism and the PIWI proteins involved in this pathway remained to be defined.

To characterize the piRNA repertoire in Aag2 cells and explore its dependency on specific PIWI proteins, we analyzed our previous small RNA sequencing data from Sindbis virus (SINV)-infected Aag2 cells, upon RNAi-mediated knockdown or immunoprecipitation (IP) of Ago3, Piwi4, 5 and 6 proteins (20). In addition to the expected siRNA and miRNA populations (21-23 nt), a piRNA-sized population of small RNAs (25-30 nt) accumulated in control knockdown libraries (Figure S1A).

We mapped small RNA reads to the SINV genome and to the genomes of viruses known to persistently infect Aag2 cells and to the *Ae. aegypti* genome (AaegL3) and 25-30 nt reads were assigned to different categories of transcripts (Table 1). Most reads mapped to unannotated regions of the *Ae. aegypti* genome (40%). As previously indicated (27), TE-derived piRNAs represented only a modest fraction (~24%) of the piRNA-like reads. Approximately 26% of reads mapped to other repeated regions in the *Ae. aegypti* genome. Among the remaining 25-30 nt piRNA-like reads, ~8% mapped to annotated *Ae aegypti* coding genes (Table 1) and were selected for further analyses. Mapping to individual transcripts, a total of 339 protein-coding genes produced at least ten piRNA-sized reads per million mapped reads.

Table 1. Annotation of 25-30 nt small RNA reads in Aag2 cells. Small RNA reads from control SINV-GFP infected Aag2 cells (luciferase dsRNA treated) were mapped to the indicated viral genomes and the *Ae. aegypti* genome. *Ae. aegypti* specific 25-30 nt RNAs were assigned to different categories of transcripts (Transposable Elements, TEs; other repeats; protein-coding or non-coding genes). SINV, Sindbis virus; AaDV2, Aedes aegypti densovirus; MXV, mosquito X virus; CFAV, cell fusing agent virus.

Mapping	Annotation		Number of reads (three libraries)	Percentage of total reads	Percentage of AaegL3 mapping reads
Total	--		11,717,519	100	--
SINV	--		166,152	1.42	--
AaDV2	--		12	0.0001	--
MXV	--		5,839	0.05	--
CFAV	--		5,284	0.05	--
Ae. aegypti genome (AaegL3)	Total		8,945,062	76.34	100
	TEfam		2,177,498	18.58	24.34
	Other repeats		2,299,843	19.63	25.71
	Non repeated	Protein-coding genes	701,882	5.99	7.85
		Non-coding genes	153,705	1.31	1.72
		Shared	34,075	0.29	0.38
	Unannotated		3,578,059	30.54	40.00
Unmapped	--		2,595,170	22.15	--

Different combinations of PIWI proteins produce distinct sets of genic piRNAs

We next investigated genic piRNA dependence on and association with specific PIWI proteins. To this end, fold changes of piRNA-sized reads in sense or antisense orientation from individual PIWI knockdown or IP libraries were compared to the control libraries (dsLuc or GFP-IP, respectively). The global levels of sense, genic 25-30 nt RNA reads were specifically reduced in Ago3 knockdown (Figure S1B) and enriched in Ago3 IP (Figure S1C), whereas the antisense reads were reduced by both Ago3 and Piwi5 knockdowns (Figure S1B) and enriched in Piwi5 and Piwi6 IP (Figure S1C).

To better characterize genic piRNA-like reads based on their PIWI protein dependency, we analyzed the changes in small RNA levels for individual coding genes upon PIWI knockdown and immunoprecipitation. We performed hierarchical clustering of the top 50 piRNA producing coding genes and we classified them into six groups based on their dependency on specific PIWI proteins (Figure 1A, left panel and 1B). In addition, we calculated the sense or antisense bias of the reads relative to the annotated transcriptional orientation of the locus (Figure 1A, central panel). Furthermore, we evaluated the nucleotide bias at each position of small RNA sequences within the six defined groups to determine if genic piRNA-like reads contained the characteristic 1U/10A ping-pong signature (Figure 1A, right panel).

A large number of genes fall in group II and III, which contained piRNA-like reads that were dependent on and enriched in Piwi5 and Piwi6, indicating that these PIWI proteins are directly responsible for their production, classifying these small RNAs as *bona fide* genic piRNAs. BLAST analyses of the predicted gene products identified eight loci that seem to be of viral origin (group II: AAEL007844, AAEL007866, AAEL009873, AAEL017001; group III: AAEL000976, AAEL00991, AAEL00997, AAEL001003, and AAEL009870). Such non-retroviral integrated RNA virus elements (NIRVS) are likely integrated into the host genome by spurious reverse transcription and integration events by retrotransposons and have been proposed to be a catalogue of previous viral encounters in the mosquito genome (36, 37). These Piwi5/Piwi6 dependent viral-like piRNAs were 1U-biased and in antisense orientation to the annotated gene. The viral-like piRNAs from group II have previously been identified as a source of piRNAs in adult *Ae. aegypti* (27). Strikingly, four out of five viral-like loci of group III (AAEL000976, AAEL00991, AAEL00997, and AAEL001003) are clustered on the *Ae. aegypti* genome in a 6,761 bp window.

▶ genes of viral origin. Middle panel: antisense bias, defined as the percentage of antisense 25–30 nt reads that map to genic sequences. Right panel: nucleotide bias at each position of the 25–30 nt small RNA reads mapping to the sense (upper panels) and antisense coding gene sequences (lower panels). All reads of three independent experiments were combined to generate sequence logos using the Weblogo3 program. *n*, number of reads; *u*, number of unique sequences. **(B)** Heat map showing the relative piRNA abundance in the indicated Piwi IP libraries over the control IP (GFP).

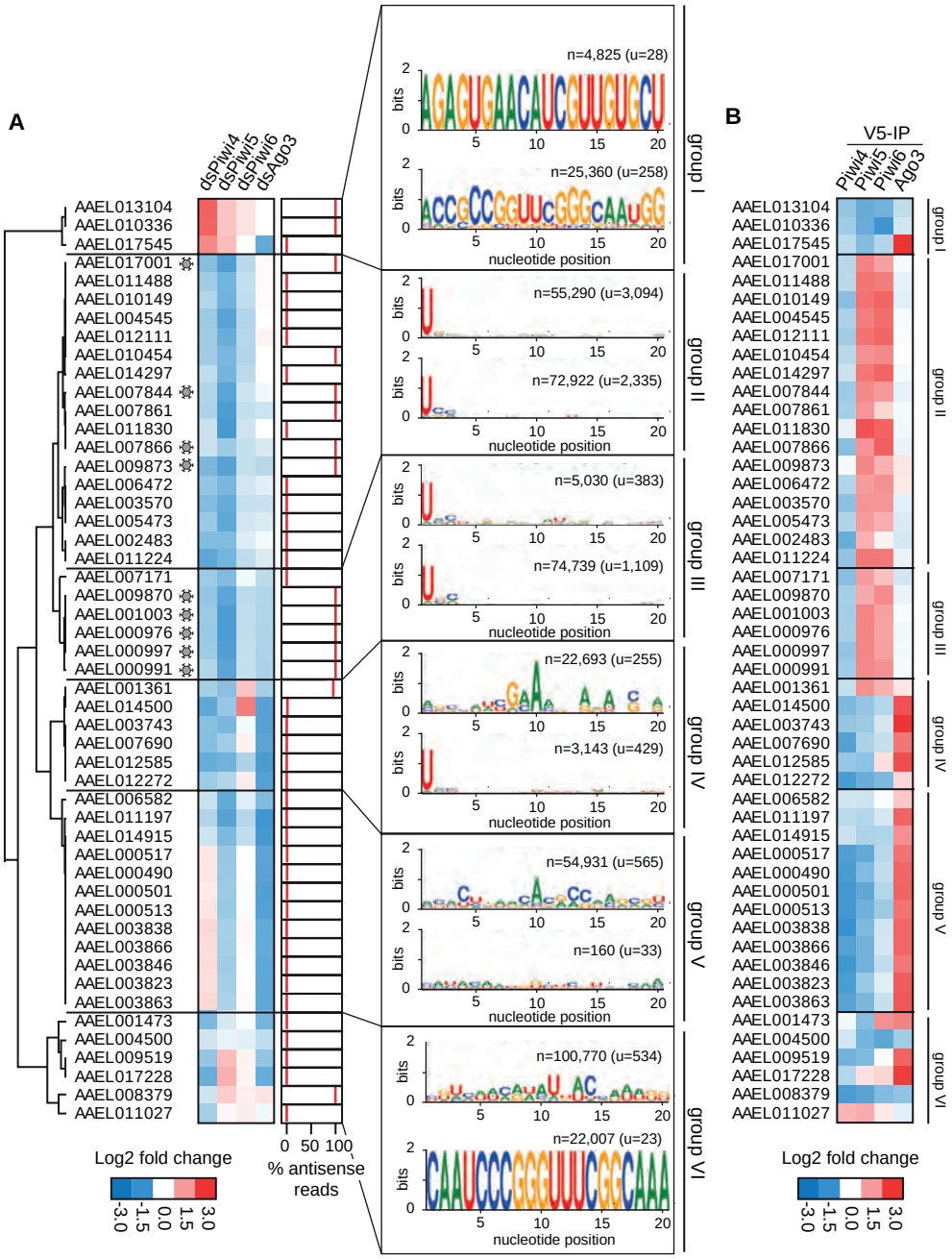


Figure 1. Genic piRNAs depend on different combinations of *Ae. aegypti* PIWI proteins. (A) Relative changes in piRNA abundance of the top-50 piRNA producing protein-coding genes upon PIWI knockdown. Left panel: heat map showing the relative piRNA abundance in the indicated knockdown libraries compared to the control knockdown (dsLuc). These data were used to generate the hierarchical clustering and assign individual genes to specific groups (I to VI). The virus symbol indicate annotated

Group IV and V included genic piRNA-like reads that depend on the ping-pong partners Ago3 and Piwi5. Accordingly, these genes predominantly produced piRNAs from the sense strand, which were enriched in Ago3 and showed a 10A bias (Figure 1A-B). These features resemble *Ae. aegypti* viral piRNAs that are produced in a ping-pong dependent manner (20, 21). Like group II and III, group IV piRNA levels were reduced upon Piwi4 knockdown, but they were depleted in Piwi4 IP. This pattern resembled our previous observations for TE-derived piRNAs (20), suggesting an indirect role for Piwi4 in piRNA biogenesis also in genic piRNA production.

Group I comprised three genes characterized by an increase in piRNA expression upon Piwi4 knockdown and a depletion in Piwi4 IP, suggesting an indirect effect of Piwi4 on piRNA accumulation. Group VI consisted of a heterogeneous group of six genes, overall distinguished by a loss of piRNAs upon Piwi4 and Ago3 knockdown and an association with Ago3. The strong nucleotide bias for group I (sense and antisense) and VI (antisense) members was caused by the predominance of individual sequences. Together, these results suggest that different combinations of PIWI proteins mediate the biogenesis of genic piRNAs from subsets of protein-coding genes.

Sense and antisense histone 4-derived piRNAs accumulate in an Ago3/Piwi5 ping-pong dependent fashion

We have previously shown that Ago3 and Piwi5 process viral RNA substrates in a ping-pong dependent manner to generate antisense, U1 biased, Piwi5-bound piRNAs and sense, 10A biased, Ago3-bound piRNAs (20). To further study the role of the ping-pong partners Ago3 and Piwi5 in control of host gene expression, we first analyzed representative genes from group IV (AAEL012272, AAEL007690, AAEL003743) and group V (AAEL011197, AAEL14915, AAEL006582) (Figure S2). Interestingly, for all genes a similar small RNA distribution was observed: piRNA-sized reads accumulated as hotspots in exons, almost exclusively in a sense orientation to the host transcript (Figure S2A and B). Despite the absence of antisense piRNA sized reads accumulating from those loci, low levels of minus strand-derived 21 nt reads were detected, suggesting the existence of antisense transcripts (Figure S2B). Nonetheless, RT-qPCR analyses indicate that mRNA steady-state levels remained largely unchanged in Ago3 and Piwi5 knockdown, despite the associated decrease in piRNA levels (Figure 1A and S2C).

Among the genes that produce Ago3/Piwi5-dependent piRNAs in group V, we found nine members of the histone 4 (H4) gene family (AAEL000517, AAEL000490, AAEL000501, AAEL000513, AAEL003838, AAEL003866, AAEL003846, AAEL003823 and AAEL003863). H4 forms the central core nucleosome with histone 3 (H3) and interacts with DNA and all other core histones (34). In addition to coding an evolutionary conserved protein, H4 genes display a strikingly high conservation at the nucleotide level (38). The *Ae. aegypti* genome encodes fifteen almost identical H4 genes. Among

those, thirteen display the unique features of replication-dependent histone genes in metazoans (Figure S3A). They encode mRNAs ending in a conserved stem-loop sequence, rather than a poly(A) tail, their 3′ end formation is directed by a purine-rich sequence known as histone downstream element (HDE), they lack introns, and they are clustered with the other core histone genes (H2A, 2B and 3) in the genome (Figure S3B). All replication-dependent H4 genes can be a source of genic piRNAs (Figure S3A). The remaining two H4 genes (AAEL011999 and AAEL013709) are not clustered and possess

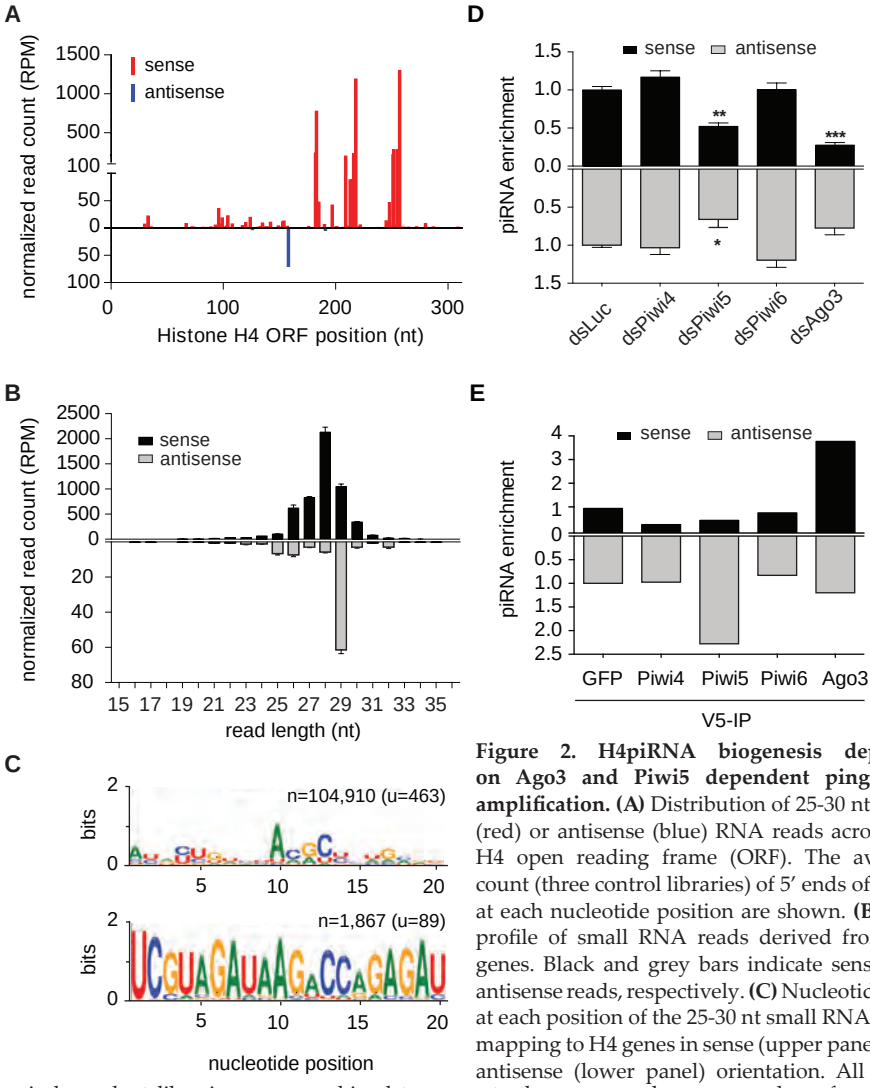


Figure 2. H4piRNA biogenesis depends on Ago3 and Piwi5 dependent ping-pong amplification. (A) Distribution of 25-30 nt sense (red) or antisense (blue) RNA reads across the H4 open reading frame (ORF). The average count (three control libraries) of 5′ ends of reads at each nucleotide position are shown. **(B)** Size profile of small RNA reads derived from H4 genes. Black and grey bars indicate sense and antisense reads, respectively. **(C)** Nucleotide bias at each position of the 25-30 nt small RNA reads mapping to H4 genes in sense (upper panel) and antisense (lower panel) orientation. All reads of three independent libraries were combined to generate the sequence logo; *n*, number of reads; *u*, number of unique sequences. **(D,E)** Relative abundance of 25-30 nt sense (black) and antisense (grey) H4piRNA reads in the indicated (D) PIWI knockdown or (E) PIWI IP libraries. Bars are the mean +/- SD of three independent small RNA libraries. Two-tailed student's t-test was used to determine statistical significance (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$).

a canonical polyadenylation signal at the 3'end, typical of replication-independent H4 replacement variants (39).

We mapped small RNA reads to the H4 genes and analyzed their size profile and distribution across the open reading frame (ORF) (Figure 2A-B). Both sense and antisense small RNAs ranged in size from 25 to 30 nt, resembling the size distribution of piRNAs. The vast majority of H4 small RNA reads derived from the sense strand of the gene (Figure 2A-B) and displayed a 10A nucleotide bias (Figure 2C).

To confirm our data, we compared our dataset to a publicly available small RNA sequence dataset from dsGFP transfected Aag2 cells (40) (Figure S4) and found a strong correlation in abundance of gene-derived piRNAs in both datasets (r_s=0.75; $P < 0.001$, Figure S4A). Moreover, in the Haac *et al*. dataset, H4-derived piRNAs accumulate with a similar size profile and ping-pong signature as in our dataset (Figure S4B and C).

H4 piRNAs predominantly derived from the sense strand in the second half of the H4 ORF and were dominated by a few, highly abundant sequences (Figure 2A, in red). Nonetheless, antisense 29 nt small RNA reads were also detectable (Figure 2A, in blue). Moreover, the 10-nt overlap between small RNAs that mapped to opposite strands suggests that the ping-pong amplification loop mediates biogenesis of these small RNAs (Figure S5A). In agreement, both sense and antisense piRNAs were reduced in Ago3 or Piwi5 depleted cells (Figure 2D). Moreover, sense piRNAs were specifically enriched in Ago3, whereas antisense reads were preferentially bound by Piwi5 (Figure 2E). These results indicate that H4-derived 25-30 nt small RNAs are ping-pong dependent piRNAs (H4piRNAs).

Sense histone 4-derived piRNAs are 3'end-modifed and associate with Ago3

To confirm these findings in a sequencing-independent manner, we extracted total RNA from mock and SINV-infected Aag2 cells in different PIWI knockdown conditions and analyzed H4 sense piRNA accumulation levels by northern blot using a mix of four DNA probes (Figure 3A). The presence of individually cloned H4 sense piRNAs was verified using each of the four DNA probes separately (Figure S5B). Consistent with the sequencing results (Figure 1A, group V and Figure 2D), Ago3 and Piwi5 knockdown specifically reduced H4piRNA levels (Figure 3A, S5B and C). As expected, H4piRNAs are independent of Ago1 and Ago2, which are the effector proteins in the microRNA and siRNA pathways, respectively (Figure S5C). Our results showed that H4piRNAs accumulate as abundant and discrete RNA molecules independently of virus infection. For this reason, further analyses have been performed in uninfected Aag2 cells. In agreement with the sequencing data (Figure 1B, group V and Figure 3B), we confirmed that sense H4piRNAs are specifically enriched in Ago3 upon GFP-Trap immunoprecipitation of a GFP-Ago3 fusion protein (Figure 3B-C and S5D).

The PAZ domain of PIWI proteins is known to recognize the 3'-end of piRNAs (41, 42),

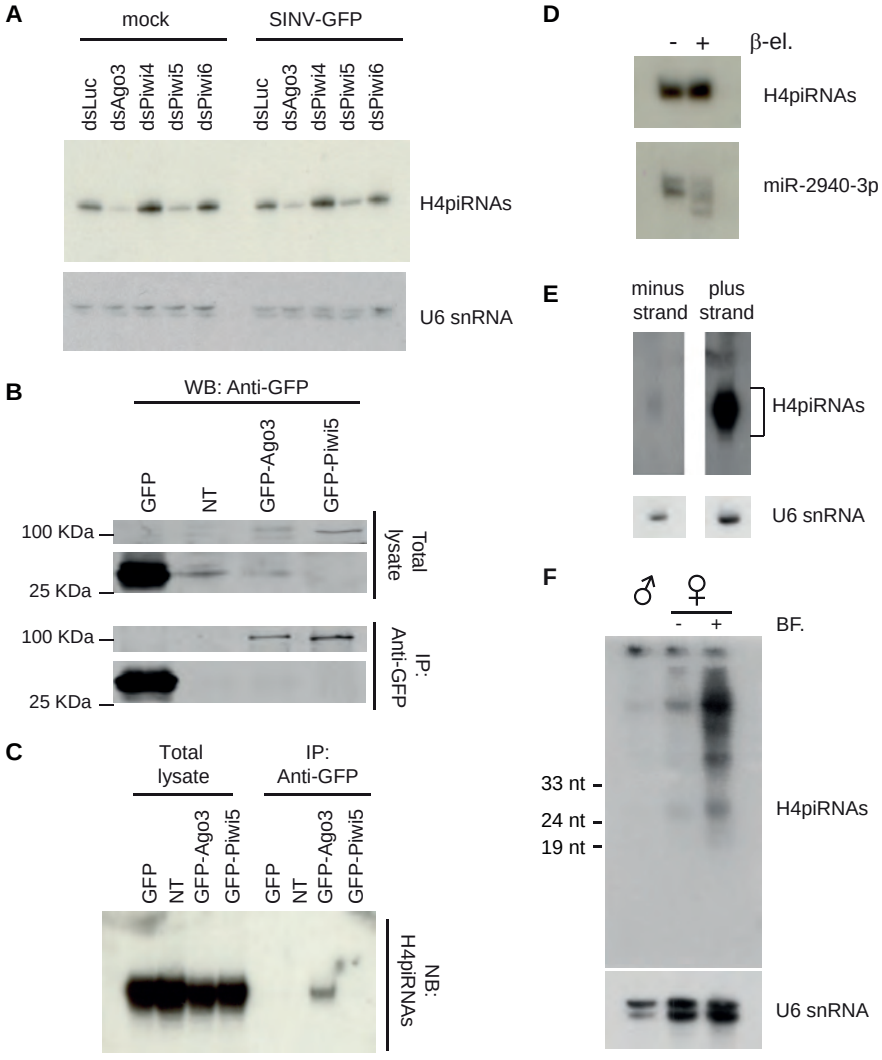


Figure 3. H4-derived sense piRNAs associate with Ago3. (A) Northern blot analysis of sense H4piRNAs upon knockdown of the indicated PIWI/AGO transcripts in mock and SINV-GFP infected Aag2 cells. H4piRNAs were detected using a pool of four probes. U6 snRNA serves as loading control. **(B)** GFP-trap immunoprecipitation (IP) on Aag2 cells expressing GFP-Ago3, GFP-Piwi5, or, as a negative control, GFP. Immunoprecipitates and total lysates were analyzed by western blot using anti-GFP antibodies. The complete western blot is shown in Figure S5. NT, non transfected. **(C)** Northern blot analysis of sense H4piRNAs in RNA isolated from IPs of the indicated GFP-tagged proteins, using a strand-specific riboprobe complementary to the H4 mRNA. **(D)** Beta-elimination on total RNA of Aag2 cells followed by northern blot of H4piRNAs and miR-2940-3p. H4piRNAs were detected using a pool of four DNA oligonucleotide probes. **(E)** Northern blot for antisense (left) and sense (right) H4piRNAs in total RNA extracted from Aag2 cells. Strand-specific riboprobes complementary to either the sense or antisense strand of H4 gene sequence were used. U6 snRNA serves as loading control. All RNA samples were analyzed on high resolution 17.5% polyacrylamide gel. **(F)** Northern blot analyses of H4-piRNAs in male, female, and blood-fed (BF) female *Ae. aegypti* mosquitoes. A pool of four DNA oligonucleotide probes was used to detect H4-piRNAs. U6 snRNA serves as a loading control.

which invariantly carries a 2′-*O*-methyl group (1). Also, *in vitro* studies in silkworm BmN4 cells established that the 3′end modification takes place on piRNA precursors that are loaded into PIWI proteins (43). We thus performed a beta elimination assay on total RNA extracted from Aag2 cells and confirmed the presence of a modification at the H4piRNA 3′ terminus, consistent with 2′*O*-methylation (Figure 3D).

The ping-pong cycle predicts that antisense piRNAs are produced from antisense transcripts. Although their accumulation was low compared to sense piRNAs, we could also detect antisense H4piRNAs by northern blot (Figure 3E). We reasoned that a transcript that is antisense to the H4 mRNA could serve as a potential precursor for these antisense H4piRNA. We therefore established a strand-specific RT-PCR to specifically detect sense and antisense RNAs (Figure S6A). As expected, we readily amplified cDNA from the sense strand, which corresponds to the H4 mRNA. In addition, we detected specific RT-PCR products from the antisense strand, indicating that antisense H4 transcripts accumulate in Aag2 cells (Figure S6B).

Although not ranking among the top-50 piRNA producing genes, the other core histone genes, H2A, H2B, and H3, were also a source of Ago3 and Piwi5-dependent piRNAs in Aag2 cells (Figure S7A-C). In addition to piRNAs, these core histone genes produced low levels of 21 nt sense and antisense reads, indicating that these genes also produced antisense transcripts.

To analyze whether histone-derived piRNAs are produced *in vivo*, we analyzed H4-derived small RNAs from adult mosquitoes by northern blot (Figure 3F) and observed that H4piRNA accumulate in adult mosquitoes and that their levels were higher in blood-fed female mosquitoes than in non-blood-fed ones (Figure 3F). Moreover, small RNA sequence data from adult mosquitoes contained histone-derived small RNAs in the size range of 26-29 nt, which were predominantly sense to histone transcripts and derived from the second half of the ORF (Figure S8A-D).

Together, our results show that sense H4piRNAs are highly abundant molecules and that they are 3′ end modified, 10A biased, loaded in Ago3 and expressed in an Ago3 and Piwi5-dependent manner. An H4 antisense transcript is the likely source of Piwi5-dependent, 1U biased, antisense H4piRNAs that initiates the ping-pong amplification loop.

Histone-derived piRNAs accumulate during the cell cycle

High expression of the intronless, non-polyadenylated, replication-dependent histone genes is specifically required when DNA is being synthesized. Highly cell cycle-regulated activation of transcription, coupled with tight control of mRNA stability, causes a rapid increase in histone mRNA abundance as cells enter S phase and a rapid decrease at the end of the S phase (34). H4piRNAs specifically derive from the replication-dependent H4 genes in class I and II (Figure S3A) rather than from the constitutively expressed

H4 replacement variants (class IV). For this reason, we hypothesized that H4piRNAs production and function was linked with DNA replication and the cell cycle.

Cell cycle progression can be arrested at the transition step between G1 and S phase (G1/S) using hydroxyurea (HU), which reversibly induces replication stalling by nucleotide depletion and inhibition of DNA synthesis. To test our hypothesis, Aag2 cells were synchronized with 0.2 mM HU for 24 hours and the cell cycle was subsequently reinitiated by the removal of the drug. The cells were harvested at different time points post-release and their cell cycle distribution was assessed by flow cytometry based on DNA content (Figure 4A-B). As expected, H4 mRNA levels increased upon entry into S phase at 2 hours post

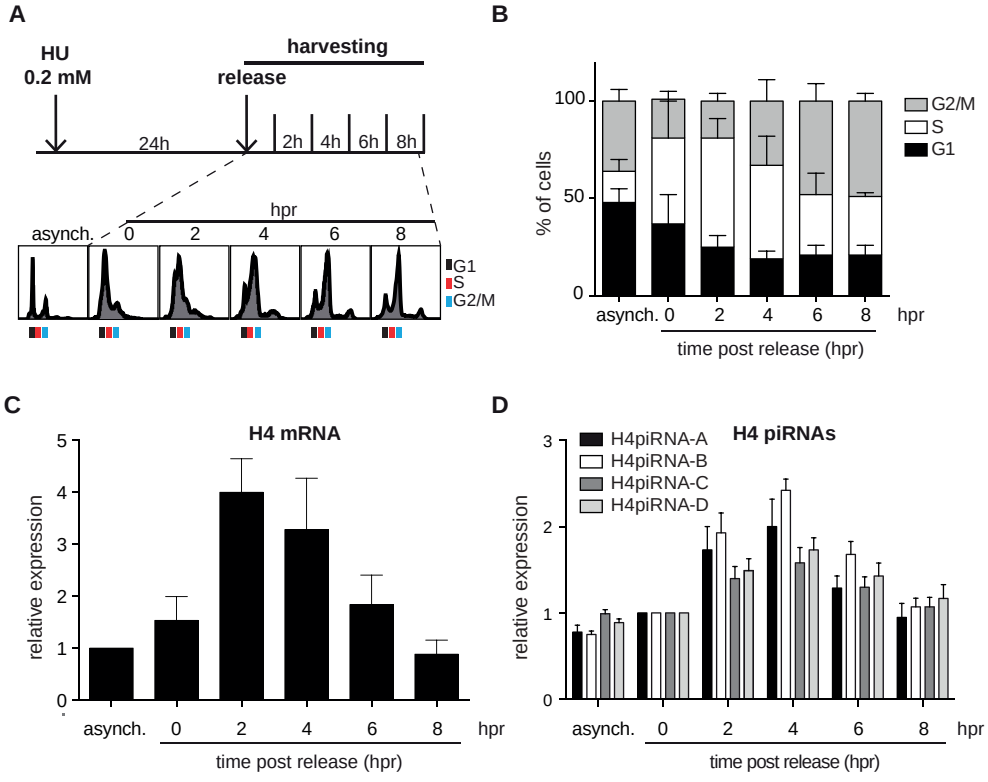


Figure 4. H4piRNAs accumulate during the cell cycle in synchronized Aag2 cells. (A) Schematic representation of the synchronization protocol. Aag2 cells at 50% confluency were treated with 0.2 mM hydroxyurea (HU) for 24 hours to induce cell cycle arrest at the G1/S phase. Once released, cells were collected every 2 hours, stained with propidium iodide (PI) and quantified for cell cycle distribution by flow cytometry. An example is shown (lower panel). **(B)** The relative percentage of cells in G1, S or G2/M phase cell cycle was assessed in three independent experiments. Data are presented as means +/- SD. **(C)** RT-qPCR analysis of H4 mRNA levels in synchronized cells. Expression is normalized to Lysosomal Aspartic Protease (LAP) levels and is expressed relative to asynchronous cells. Bars represent means +/- SD of three biological replicates. **(D)** Stem-loop (SL) RT-qPCR analyses of individual H4-piRNAs (shown in Figure S5B) in non-synchonized cells and in HU treated, synchronized cells at the indicate time points. Expression is normalized to aae-bantam-3p and presented relative to 0 hpr. Bars represent means +/- SD of three biological replicates. hpr, hours post release.

release (hpr) and rapidly dropped when cells progressed through the G2/M phase (Figure 4C). Likewise, the other core histone mRNAs showed the same dynamics (Figure S9A).

To quantify h4piRNA accumulation, we set up stem loop (SL) qPCR assays for four individual H4piRNAs (A-D, Figure S5B). The previously observed H4piRNA reduction upon Ago3 and Piwi5 knockdown (Figure 3A) was recapitulated by the SL-qPCR assay (Figure S9B), thus validating the approach. We then analyzed H4piRNA accumulation in asynchronous and synchronized cells and found that H4piRNAs are dynamically expressed throughout the cycle with a peak in expression at 4 hpr, with a slight delay compared to the peak of H4 mRNA expression (Figure 4C-D). Together, our results indicate that Piwi5 and Ago3 produce histone-derived piRNAs and suggest that piRNAs contribute to histone mRNA turnover during the cell cycle.

DISCUSSION

Ten years have passed since the first identification of piRNAs by several independent laboratories (44-48). In addition to the canonical function in protecting germline integrity, recent evidence imply broader roles for PIWI/piRNAs as regulators of gene expression in both germline and somatic tissues (6,7). *Aedes* mosquitoes display an expansion of the PIWI gene family, several of which are expressed in the soma (25, 49). In this study, we characterized the biogenesis of gene-derived piRNAs in the *Ae. aegypti* Aag2 cell line, which express the same PIWI genes as are expressed in somatic tissues of adult mosquitoes. We find that replication-dependent histone genes produce piRNAs in a ping-pong dependent manner, suggesting that piRNAs can be involved in dynamic regulation of mRNA expression in the soma. Moreover, since cell culture has a higher experimental amenability than adult mosquitoes, our work establishes Aag2 cells as an accessible and relevant model to study gene-derived piRNAs.

We identified several classes of genic piRNAs that depend on different combinations of PIWI proteins. Among these, class IV and V genic piRNAs are dependent on the ping-pong proteins Ago3 and Piwi5 and accumulate as hotspots in exonic sequences along the gene body. This distinguishes them from gene-derived piRNAs reported before in other species, which generally derive from the 3′ UTRs and are generated in a ping-pong independent fashion (3, 31-33). Class IV and V genes do not seem to share biological and molecular functions, nor do they share structural similarities. For example, group IV and V included canonical spliced host genes that are expressed as polyadenylated transcripts, as well as replication-dependent histone genes that produce unspliced, non-polyadenylated mRNAs.

Among the core histone genes, H4 produced the most abundant piRNAs. Sense and antisense H4piRNAs are specifically enriched in Ago3 and Piwi5, respectively. This reflects their nucleotide bias: 10A for sense H4piRNAs and 1U for the antisense ones. As a consequence, sense H4piRNA biogenesis seems to rely on the feed-forward ping-pong

amplification loop, most likely initiated by cleavage of H4 mRNAs by antisense piRNAs. Although ping-pong-dependent production of genic piRNAs initiated by transposon-derived piRNAs has recently been reported (50, 51), to our knowledge H4piRNAs are the first example of genic piRNAs that are produced in an autonomous ping-pong amplification loop.

In metazoans, expression of replication-dependent histone gene is tightly controlled to ensure their massive production as cells enter S phase and their reduction to baseline between the end of the S phase and mitosis. An intriguing possibility would thus be that piRNA-PIWI complexes are regulated by or have a role in the cell cycle. We found that histone mRNAs and histone piRNAs are dynamically expressed during the cell cycle, with the peak of H4piRNA abundance lagging ~2 h behind the peak of histone mRNA expression, suggesting that histone-derived piRNAs are processing products of histone mRNAs. The factors involved in cell cycle-dependent histone mRNA expression and degradation are conserved during evolution, and it is likely that the canonical mechanisms for histone mRNA metabolism are responsible for the major changes in histone mRNA levels in *Ae. aegypti*. However, our results indicate that the piRNA pathway may add an additional layer of dynamic histone mRNA regulation in a narrow window of the cell cycle.

Previous studies have indirectly shown a connection between piRNAs and histone genes in different organisms. For instance, Hiwi2 IP analyses in human somatic cells identified several genes involved in cell growth and proliferation among the piRNA-producing loci (30). Moreover, sense 1U biased piRNA-like molecules from histone mRNAs have been identified by deep sequencing in the cnidarian Hydra magnipapillata. During regeneration, down-regulation of these piRNA-like molecules leads to a small, but consistent, up-regulation of histone transcripts (52). Support for a role of PIWI proteins in the regulation of histone mRNA levels have also been suggested in other species. For example, expression of the histone variant H3.3 during macronuclear differentiation is impaired after PIWI knockdown in the ciliate *Stylonychia lemnae* (53). In the parasitic protozoan *Leishmania*, histone transcripts are upregulated in PIWI null mutants. As PIWI in this species is unable to bind piRNAs due to lack of a typical PAZ domain, this result suggests a role in the stability of histone transcripts independently of piRNAs (54). In fly ovaries, H2B mRNA expression levels are upregulated upon nuclear PIWI elimination (55).

Our analyses indicate that, although all core histone genes produce piRNAs, the majority derive from H4 genes. This is unexpected given that replication dependent histone mRNAs are produced at the same time during the cell cycle. We hypothesize that H4piRNA production not only depends on relative transcript abundance, but also on other specific features. Compared to other histone genes, H4 genes show the highest level of conservation at the nucleotide level during evolution, suggesting an important

role for sequence or structure of their mRNA (38). Indeed, the secondary structure of murine H4 mRNA is crucial for its non-canonical translational initiation mechanism (56). Of note, *Ae. aegypti* H4piRNAs originate from a region in the mRNA that corresponds to the structural element that is critical for internal translation initiation on the murine transcript. This raises the intriguing possibility that histone mRNA structure or sequence enhances recognition or processing by the piRNA pathway.

Among the most abundant genic piRNAs, we retrieved piRNAs from annotated genes consisting of sequences of RNA viruses that are integrated in the *Ae. aegypti* genome. Virus-like genic piRNAs are in antisense orientation to the annotated host gene, show a clear 1U nucleotide bias, and associate with Piwi5 and Piwi6. We noticed that four of the virus-like loci in group III (AAEL001003, AAEL000976, AAEL00997 and AAEL00991 which are reminiscent of rhabdoviral nucleoprotein sequences) are clustered in the *Ae. aegypti* genome and may represent something akin to a piRNA cluster.

Sequences with similarity to non-retroviral RNA viruses have been detected in the genome of both *Ae. aegypti* (27) and *Ae. albopictus* and have been referred to as non-retroviral integrated RNA virus sequences (NIRVS) (36, 37). Interestingly, vertebrate genomes also contain sequences corresponding to viral fragments. For instance, Endogenous Bornavirus-Like Nucleoprotein elements (EBLNs) are the result of reverse transcription and integration of ancient bornaviral nucleoprotein mRNA in the genome of primates and rodents. A recent study proposed that EBLN-derived piRNAs explains the resistance to bornaviral infections in these host species (57). The majority of these piRNAs are antisense to viral sequences, which would render the primary piRNA pathway capable of slicing viral gene transcripts. Similarly, as virus-like piRNAs are antisense and Piwi5 and Piwi6 associated in *Aedes*, one can envision that piRNAs confer heritable immunity to infection. An invading cognate virus may thus be targeted directly by the host-encoded piRNA, initiating a ping-pong amplification cycle, and inducing the production of phased piRNAs to diversify the viral piRNA population, as was recently proposed for transposon-derived piRNAs in *Drosophila* (50, 58).

Even though PIWI proteins are well conserved across different organisms, piRNA sequences are generally not conserved during evolution (2). However, given recent examples of piRNA that regulate coding genes in the soma (29, 51, 59-61), sequence constraints between piRNAs and their RNA targets may be expected. For example, a recent study has identified Eutherian-Conserved piRNA cluster (ECpiC) loci, which most likely have a functional relevance (62). The high conservation and essential function of the histone genes in all eukaryotes suggest that histone piRNAs have key functions not only in *Ae. aegypti* but also in other species.

ACKNOWLEDGEMENTS

We thank members of the Van Rij laboratory for fruitful discussions. Furthermore, we thank Franck Martin for sharing his expertise in histone biology and providing experimental advice and Patryk Ngondo for critical reading of the manuscript. We also thank Rob Andriessen and Bart Knols (In2Care) for kindly providing *Ae. aegypti* mosquitoes used for northern blot analysis.

This work was financially supported by the Netherlands Organization for Scientific Research (NWO, ECHO project grant) [grant number 711.013.001]; European Research Council Consolidator Grant under the European Union's Seventh Framework Programme [grant number ERC CoG 615680] to R.P.v.R.; PhD fellowship from the Radboud University Medical Center to P.M.; European Research Council grants [grant number ERC StG 242703, ERC CoG 615220); the French Government's Investissements d'avenir program Laboratoire d'Excellence „Integrative Biology of Emerging Infectious Diseases" [grant number ANR-10-LABX-62-IBEID] to M.C.S. Funding for open access charge: ERC CoG 615680.

REFERENCES

1. Ghildiyal M, Zamore PD. Small silencing RNAs: an expanding universe. Nat Rev Genet. 2009;10(2):94-108.
2. Ishizu H, Siomi H, Siomi MC. Biology of PIWI-interacting RNAs: new insights into biogenesis and function inside and outside of germlines. Genes Dev. 2012;26(21):2361-73.
3. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, et al. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell. 2007;128(6):1089-103.
4. Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, et al. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. Science. 2007;315(5818):1587-90.
5. Siomi MC, Sato K, Pezic D, Aravin AA. PIWI-interacting small RNAs: the vanguard of genome defence. Nat Rev Mol Cell Biol. 2011;12(4):246-58.
6. Ross RJ, Weiner MM, Lin H. PIWI proteins and PIWI-interacting RNAs in the soma. Nature. 2014;505(7483):353-9.
7. Miesen P, Joosten J, van Rij RP. PIWIs go viral: Arbovirus-Derived piRNAs in Vector Mosquitoes. PloS Pathog. 2016; 12(12):e1006017
8. Palakodeti D, Smielewska M, Lu YC, Yeo GW, Graveley BR. The PIWI proteins SMEDWI-2 and SMEDWI-3 are required for stem cell function and piRNA expression in planarians. RNA. 2008;14(6):1174-86.
9. Reddien PW, Oviedo NJ, Jennings JR, Jenkin JC, Sanchez Alvarado A. SMEDWI-2 is a PIWI-like protein that regulates planarian stem cells. Science. 2005;310(5752):1327-30.
10. Zhou X, Battistoni G, El Demerdash O, Gurtowski J, Wunderer J, Falciatori I, et al. Dual functions of Macpiwi1 in transposon silencing and stem cell maintenance in the flatworm Macrostomum lignano. RNA. 2015;21(11):1885-97.
11. Rinkevich Y, Rosner A, Rabinowitz C, Lapidot Z, Moiseeva E, Rinkevich B. Piwi positive cells that line the vasculature epithelium, underlie whole body regeneration in a basal chordate. Dev Biol. 2010;345(1):94-104.
12. Seipel K, Yanze N, Schmid V. The germ line and somatic stem cell gene Cniwi in the jellyfish Podocoryne carnea. Int J Dev Biol. 2004;48(1):1-7.

13. Malone CD, Brennecke J, Dus M, Stark A, McCombie WR, Sachidanandam R, et al. Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. Cell. 2009;137(3):522-35.
14. Li C, Vagin VV, Lee S, Xu J, Ma S, Xi H, et al. Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. Cell. 2009;137(3):509-21.
15. Yan Z, Hu HY, Jiang X, Maierhofer V, Neb E, He L, et al. Widespread expression of piRNA-like molecules in somatic tissues. Nucleic Acids Res. 2011;39(15):6596-607.
16. Lee EJ, Banerjee S, Zhou H, Jammalamadaka A, Arcila M, Manjunath BS, et al. Identification of piRNAs in the central nervous system. RNA. 2011;17(6):1090-9.
17. Martinez VD, Vucic EA, Thu KL, Hubaux R, Enfield KS, Pikor LA, et al. Unique somatic and malignant expression patterns implicate PIWI-interacting RNAs in cancer-type specific biology. Sci Rep. 2015;5:10423.
18. Cichocki F, Lenvik T, Sharma N, Yun G, Anderson SK, Miller JS. Cutting edge: KIR antisense transcripts are processed into a 28-base PIWI-like RNA in human NK cells. J Immunol. 2010;185(4):2009-12.
19. Vodovar N, Bronkhorst AW, van Cleef KW, Miesen P, Blanc H, van Rij RP, et al. Arbovirus-derived piRNAs exhibit a ping-pong signature in mosquito cells. PLoS One. 2012;7(1):e30861.
20. Miesen P, Girardi E, van Rij RP. Distinct sets of PIWI proteins produce arbovirus and transposon-derived piRNAs in Aedes aegypti mosquito cells. Nucleic Acids Res. 2015;43(13):6545-56.
21. Miesen P, Ivens A, Buck AH, van Rij RP. Small RNA Profiling in Dengue Virus 2-Infected Aedes Mosquito Cells Reveals Viral piRNAs and Novel Host miRNAs. PLoS Negl Trop Dis. 2016;10(2):e0004452.
22. Schnettler E, Donald CL, Human S, Watson M, Siu RW, McFarlane M, et al. Knockdown of piRNA pathway proteins results in enhanced Semliki Forest virus production in mosquito cells. J Gen Virol. 2013;94(Pt 7):1680-9.
23. Leger P, Lara E, Jagla B, Sismeiro O, Mansuroglu Z, Coppee JY, et al. Dicer-2- and Piwi-mediated RNA interference in Rift Valley fever virus-infected mosquito cells. J Virol. 2013;87(3):1631-48.
24. Morazzani EM, Wiley MR, Murreddu MG, Adelman ZN, Myles KM. Production of virus-derived ping-pong-dependent piRNA-like small RNAs in the mosquito soma. PLoS Pathog. 2012;8(1):e1002470.
25. Campbell CL, Black WCt, Hess AM, Foy BD. Comparative genomics of small RNA regulatory pathway components in vector mosquitoes. BMC Genomics. 2008;9:425.
26. Nene V, Wortman JR, Lawson D, Haas B, Kodira C, Tu ZJ, et al. Genome sequence of Aedes aegypti, a major arbovirus vector. Science. 2007;316(5832):1718-23.
27. Arensburger P, Hice RH, Wright JA, Craig NL, Atkinson PW. The mosquito Aedes aegypti has a large genome size and high transposable element load but contains a low proportion of transposon-specific piRNAs. BMC Genomics. 2011;12:606.
28. Biryukova I, Ye T. Endogenous siRNAs and piRNAs derived from transposable elements and genes in the malaria vector mosquito Anopheles gambiae. BMC Genomics. 2015;16:278.
29. Zhong F, Zhou N, Wu K, Guo Y, Tan W, Zhang H, et al. A SnoRNA-derived piRNA interacts with human interleukin-4 pre-mRNA and induces its decay in nuclear exosomes. Nucleic Acids Res. 2015;43(21):10474-91.
30. Keam SP, Young PE, McCorkindale AL, Dang TH, Clancy JL, Humphreys DT, et al. The human Piwi protein Hiwi2 associates with tRNA-derived piRNAs in somatic cells. Nucleic Acids Res. 2014;42(14):8984-95.
31. Castellano L, Rizzi E, Krell J, Di Cristina M, Galizi R, Mori A, et al. The germline of the malaria mosquito produces abundant miRNAs, endo-siRNAs, piRNAs and 29-nt small RNAs. BMC Genomics. 2015;16:100.
32. Robine N, Lau NC, Balla S, Jin Z, Okamura K, Kuramochi-Miyagawa S, et al. A broadly conserved pathway generates 3'UTR-directed primary piRNAs. Curr Biol. 2009;19(24):2066-76.
33. Yamtich J, Heo SJ, Dhahbi J, Martin DI, Boffelli D. piRNA-like small RNAs mark extended 3'UTRs present in germ and somatic cells. BMC Genomics. 2015;16:462.

34. Marzluff WF, Wagner EJ, Duronio RJ. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. Nat Rev Genet. 2008;9(11):843-54.
35. Davila Lopez M, Samuelsson T. Early evolution of histone mRNA 3′ end processing. RNA. 2008;14(1):1-10.
36. Chen XG, Jiang X, Gu J, Xu M, Wu Y, Deng Y, et al. Genome sequence of the Asian Tiger mosquito, Aedes albopictus, reveals insights into its biology, genetics, and evolution. Proc Natl Acad Sci U S A. 2015;112(44):E5907-15.
37. Crochu S, Cook S, Attoui H, Charrel RN, De Chesse R, Belhouchet M, et al. Sequences of flavivirus-related RNA viruses persist in DNA form integrated in the genome of Aedes spp. mosquitoes. J Gen Virol. 2004;85(Pt 7):1971-80.
38. Baxevanis AD, Landsman D. Histone Sequence Database: a compilation of highly-conserved nucleoprotein sequences. Nucleic Acids Res. 1996;24(1):245-7.
39. Akhmanova A, Miedema K, Hennig W. Identification and characterization of the Drosophila histone H4 replacement gene. FEBS Lett. 1996;388(2-3):219-22.
40. Haac ME, Anderson MA, Eggleston H, Myles KM, Adelman ZN. The hub protein loquacious connects the microRNA and short interfering RNA pathways in mosquitoes. Nucleic Acids Res. 2015;43(7):3688-700.
41. Simon B, Kirkpatrick JP, Eckhardt S, Reuter M, Rocha EA, Andrade-Navarro MA, et al. Recognition of 2′-O-methylated 3′-end of piRNA by the PAZ domain of a Piwi protein. Structure. 2011;19(2):172-80.
42. Tian Y, Simanshu DK, Ma JB, Patel DJ. Structural basis for piRNA 2′-O-methylated 3′-end recognition by Piwi PAZ (Piwi/Argonaute/Zwille) domains. Proc Natl Acad Sci U S A. 2011;108(3):903-10.
43. Kawaoka S, Izumi N, Katsuma S, Tomari Y. 3′ end formation of PIWI-interacting RNAs in vitro. Mol Cell. 2011;43(6):1015-22.
44. Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, Iovino N, et al. A novel class of small RNAs bind to MILI protein in mouse testes. Nature. 2006;442(7099):203-7.
45. Girard A, Sachidanandam R, Hannon GJ, Carmell MA. A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature. 2006;442(7099):199-202.
46. Grivna ST, Beyret E, Wang Z, Lin H. A novel class of small RNAs in mouse spermatogenic cells. Genes Dev. 2006;20(13):1709-14.
47. Watanabe T, Takeda A, Tsukiyama T, Mise K, Okuno T, Sasaki H, et al. Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. Genes Dev. 2006;20(13):1732-43.
48. Vagin VV, Sigova A, Li C, Seitz H, Gvozdev V, Zamore PD. A distinct small RNA pathway silences selfish genetic elements in the germline. Science. 2006;313(5785):320-4.
49. Akbari OS, Antoshechkin I, Amrhein H, Williams B, Diloreto R, Sandler J, et al. The developmental transcriptome of the mosquito Aedes aegypti, an invasive species and major arbovirus vector. G3 (Bethesda). 2013;3(9):1493-509.
50. Mohn F, Handler D, Brennecke J. Noncoding RNA. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. Science. 2015;348(6236):812-7.
51. Barckmann B, Pierson S, Dufourt J, Papin C, Armenise C, Port F, et al. Aubergine iCLIP Reveals piRNA-Dependent Decay of mRNAs Involved in Germ Cell Development in the Early Embryo. Cell Rep. 2015;12(7):1205-16.
52. Krishna S, Nair A, Cheedipudi S, Poduval D, Dhawan J, Palakodeti D, et al. Deep sequencing reveals unique small RNA repertoire that is regulated during head regeneration in Hydra magnipapillata. Nucleic Acids Res. 2013;41(1):599-616.
53. Forcob S, Bulic A, Jonsson F, Lipps HJ, Postberg J. Differential expression of histone H3 genes and selective association of the variant H3.7 with a specific sequence class in Stylonychia macronuclear development. Epigenetics Chromatin. 2014;7(1):4.
54. Padmanabhan PK, Dumas C, Samant M, Rochette A, Simard MJ, Papadopoulou B. Novel features of a PIWI-like protein homolog in the parasitic protozoan Leishmania. PLoS One. 2012;7(12):e52612.

55. Klenov MS, Lavrov SA, Korbut AP, Stolyarenko AD, Yakushev EY, Reuter M, et al. Impact of nuclear Piwi elimination on chromatin state in Drosophila melanogaster ovaries. Nucleic Acids Res. 2014;42(10):6208-18.

56. Martin F, Barends S, Jaeger S, Schaeffer L, Prongidi-Fix L, Eriani G. Cap-assisted internal initiation of translation of histone H4. Mol Cell. 2011;41(2):197-209.

57. Parrish NF, Fujino K, Shiromoto Y, Iwasaki YW, Ha H, Xing J, et al. piRNAs derived from ancient viral processed pseudogenes as transgenerational sequence-specific immune memory in mammals. RNA. 2015;21(10):1691-703.

58. Han BW, Wang W, Li C, Weng Z, Zamore PD. Noncoding RNA. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. Science. 2015;348(6236):817-21.

59. Gonzalez J, Qi H, Liu N, Lin H. Piwi Is a Key Regulator of Both Somatic and Germline Stem Cells in the Drosophila Testis. Cell Rep. 2015;12(1):150-61.

60. Kiuchi T, Koga H, Kawamoto M, Shoji K, Sakai H, Arai Y, et al. A single female-specific piRNA is the primary determiner of sex in the silkworm. Nature. 2014;509(7502):633-6.

61. Rouget C, Papin C, Boureux A, Meunier AC, Franco B, Robine N, et al. Maternal mRNA deadenylation and decay by the piRNA pathway in the early Drosophila embryo. Nature. 2010;467(7319):1128-32.

62. Chirn GW, Rahman R, Sytnikova YA, Matts JA, Zeng M, Gerlach D, et al. Conserved piRNA Expression from a Distinct Set of piRNA Cluster Loci in Eutherian Mammals. PLoS Genet. 2015;11(11):e1005652.

SUPPLEMENTAL FIGURES

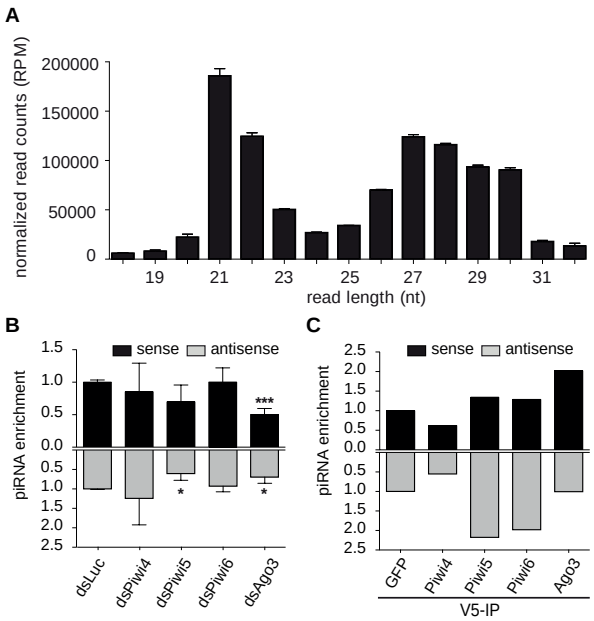


Figure S1. Accumulation of genic PIWI-dependent small RNAs in Aag2 cells. (A) Length distribution of small RNA reads in control libraries (luciferase dsRNA treated, SINV-GFP infected Aag2 cells) before mapping. Reads were normalized to total library size. Bars are the means +/- SEM of three independent small RNA libraries. **(B,C)** Relative abundance of 25-30 nt reads derived from *Ae. aegypti* protein-coding genes in the indicated (B) PIWI knockdown and (C) PIWI IP libraries compared to control libraries (dsLuc and GFP IP, respectively). Black and grey bars indicate, respectively, sense and antisense reads relative to the annotated host transcript. Bars are the means +/- SD of three independent small RNA libraries (IP libraries represent one single experiment). Two-tailed student's t-test was used to determine statistical significance (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$).

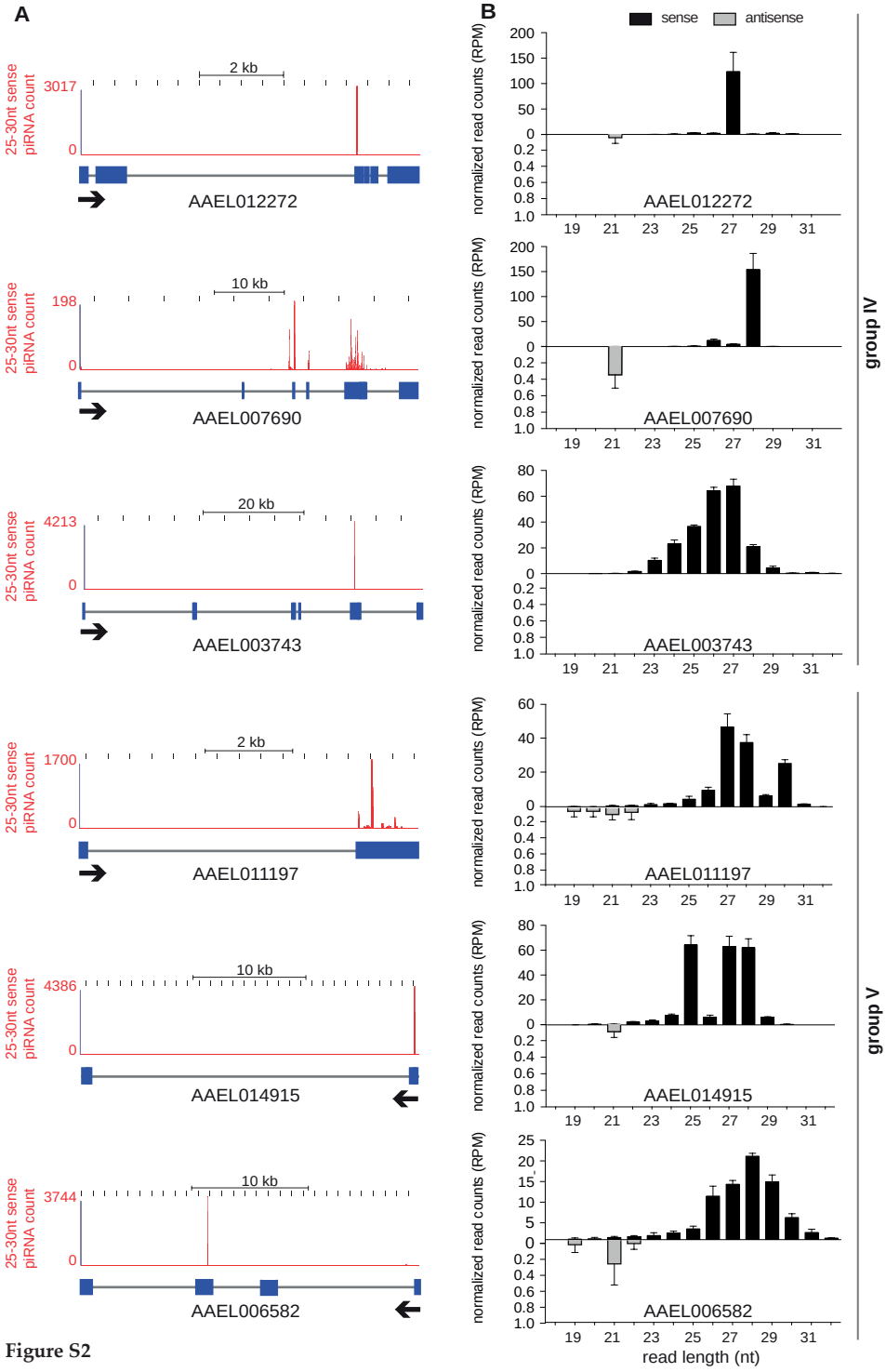

A
B
2 kb
3017
0
25-30nt sense piRNA count
AAEL012272
10 kb
198
AAEL007690
20 kb
4213
AAEL003743
2 kb
1700
AAEL011197
10 kb
4386
AAEL014915
10 kb
3744
AAEL006582
sense
antisense
normalized read counts (RPM)
group IV
group V
read length (nt)

Figure S2

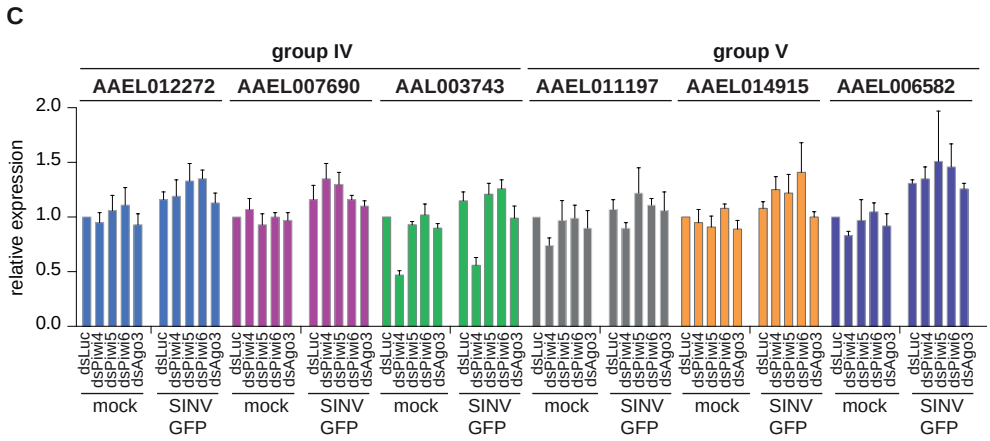
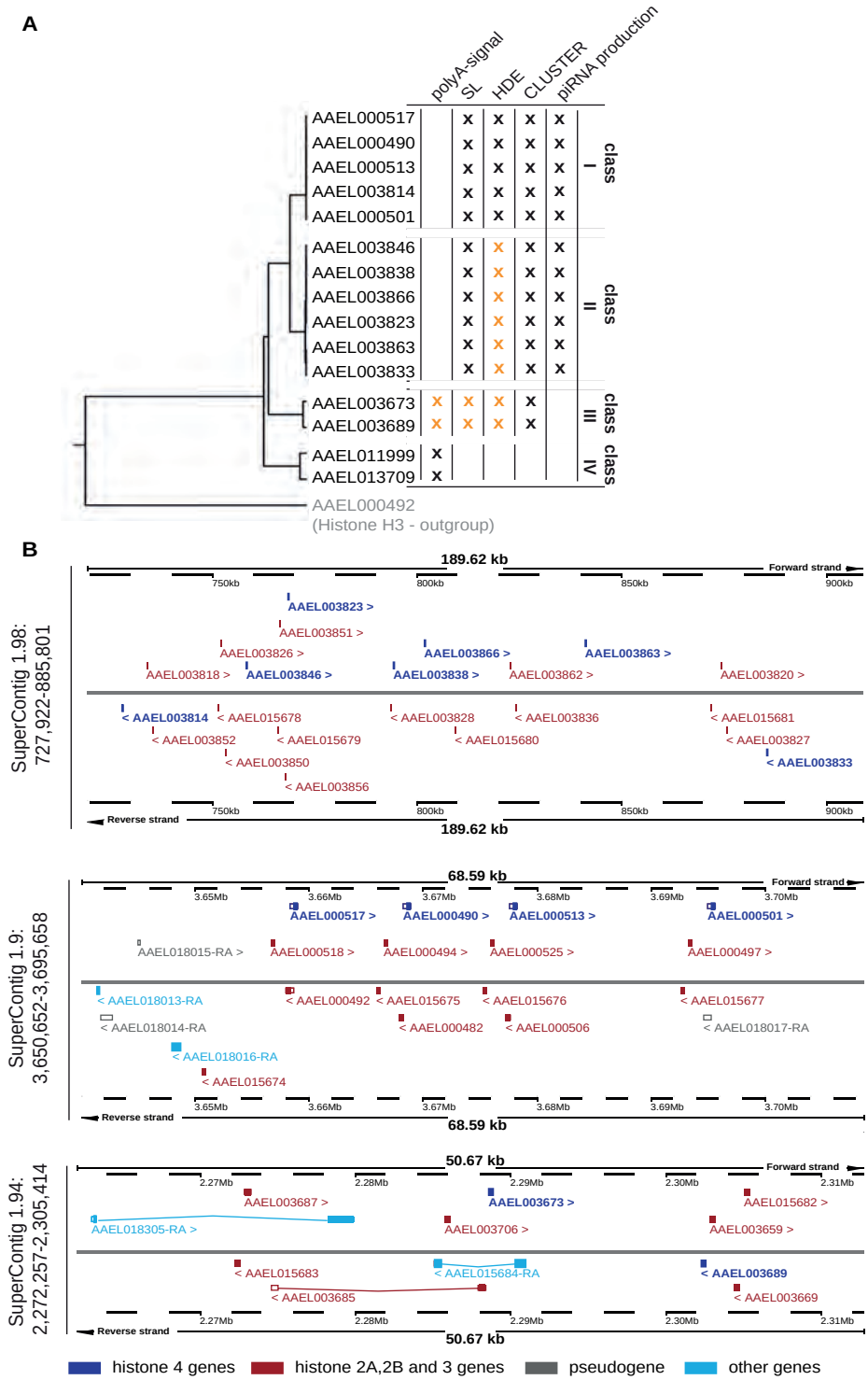


Figure S2. Examples of Ago3 and Piwi5 dependent piRNAs from class IV and V genes. (A) UCSC genome browser views of piRNA producing genes with the distribution of sense 25-30 nt RNAs across the indicated host transcript. The relative position of exons (blue boxes) and introns (grey lines) is schematically represented. Black arrows indicate the direction of transcription. **(B)** Size profiles of all small RNA reads derived from sense (black) or antisense (grey) transcripts. Bars are the mean +/- SD of three independent small RNA libraries. AAEL012272, fk506 binding protein; AAEL007690, RPTOR-like protein; AAEL003743, vacuolar protein ATPase; AAEL011197, actin; AAEL14915, 26S proteasome; AAEL006582, calcium transporting ATPase sarcoplasmic/endoplasmic reticulum type. Group IV and V are defined in Figure 1. **(C)** RT-qPCR analysis on group IV and V mRNAs upon transfection of the indicated dsRNAs in mock and SINV-GFP infected Aag2 cells. Expression is normalized to Lysosomal Aspartic Protease (LAP) levels and presented relative to dsLuc. Bars represent means of two biological replicates +/- SD.

▶ **Figure S3. piRNAs originate from replication-dependent, clustered histone 4 genes in *Ae. aegypti*. (A)** Phylogeny of annotated Histone 4 genes. The H3 gene AAEL000492 was used as outgroup. Different classes (I to IV) were defined based on nucleic acid sequence similarity. For each H4 gene, the presence of a canonical polyadenylation signal (AAUAAA), conserved stem loop (SL) (AANGGNNNNNNNNNGNGCC), purine-rich Histone downstream element (HDE), localization in genomic histone clusters, and processing into piRNAs is indicated. Sub-canonical polyadenylation signals, stem loop and HDE sequences are indicated in orange. **(B)** VectorBase genome view of the three major *Ae. aegypti* histone clusters. The relative position and direction of transcription of the annotated histone genes in these regions are indicated. H4 genes are indicated in dark blue. H2A, H2B and H3 are indicated in red. Pseudogenes and other protein-coding genes are indicated in grey and light-blue, respectively.


A
polyA-signal
SL
HDE
CLUSTER
piRNA production
AAEL000517
AAEL000490
AAEL000513
AAEL003814
AAEL000501
class I
AAEL003846
AAEL003838
AAEL003866
AAEL003823
AAEL003863
AAEL003833
class II
AAEL003673
AAEL003689
class III
AAEL011999
AAEL013709
class IV
AAEL000492
(Histone H3 - outgroup)
B
SuperContig 1.98:
727,922-885,801
189.62 kb
Forward strand
AAEL003823 >
AAEL003851 >
AAEL003826 >
AAEL003866 >
AAEL003863 >
AAEL003818 >
AAEL003846 >
AAEL003838 >
AAEL003862 >
AAEL003820 >
< AAEL003814
< AAEL015678
< AAEL003828
< AAEL003836
< AAEL015681
< AAEL003852
< AAEL015679
< AAEL015680
< AAEL003827
< AAEL003850
< AAEL003833
< AAEL003856
Reverse strand
189.62 kb
SuperContig 1.9:
3,650,652-3,695,658
68.59 kb
Forward strand
AAEL000517 >
AAEL000490 >
AAEL000513 >
AAEL000501 >
AAEL018015-RA >
AAEL000518 >
AAEL000494 >
AAEL000525 >
AAEL000497 >
< AAEL018013-RA
< AAEL000492
< AAEL015675
< AAEL015676
< AAEL015677
< AAEL018014-RA
< AAEL000482
< AAEL000506
< AAEL018017-RA
< AAEL018016-RA
< AAEL015674
Reverse strand
68.59 kb
SuperContig 1.94:
2,272,257-2,305,414
50.67 kb
Forward strand
AAEL003687 >
AAEL003673 >
AAEL015682 >
AAEL018305-RA >
AAEL003706 >
AAEL003659 >
< AAEL015683
< AAEL015684-RA
< AAEL003689
< AAEL003685
< AAEL003669
Reverse strand
50.67 kb
histone 4 genes
histone 2A,2B and 3 genes
pseudogene
other genes

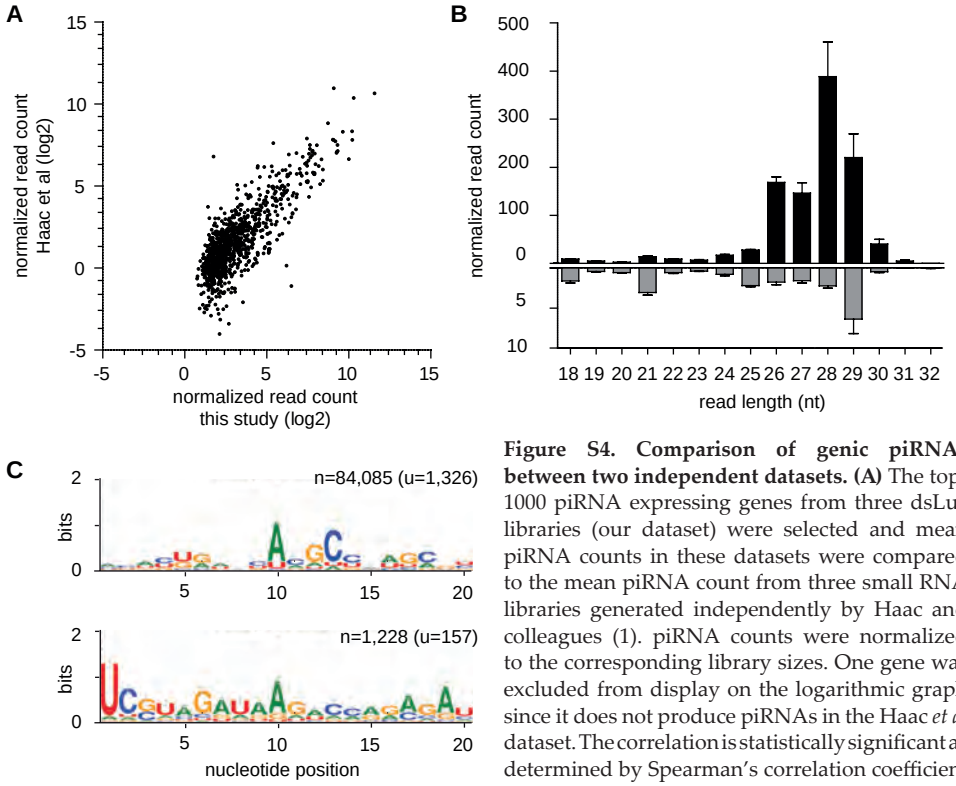


Figure S4. Comparison of genic piRNAs between two independent datasets. (A) The top-1000 piRNA expressing genes from three dsLuc libraries (our dataset) were selected and mean piRNA counts in these datasets were compared to the mean piRNA count from three small RNA libraries generated independently by Haac and colleagues (1). piRNA counts were normalized to the corresponding library sizes. One gene was excluded from display on the logarithmic graph since it does not produce piRNAs in the Haac *et al* dataset. The correlation is statistically significant as determined by Spearman's correlation coefficient (r_s=0.75; $P < 0.001$; two-tailed). **(B)** Size distribution of small RNAs mapping to *Ae. aegypti* histone 4 genes. Bars show the mean and SEM of three libraries generated by Haac *et al*. **(C)** Sequence logo of 25-30 nt reads mapping in sense (upper panel) or antisense orientation (lower panel) to histone 4. The reads of three libraries were combined. *n* indicates the number of reads used to generate the logo, *u* indicates the number of unique sequences.

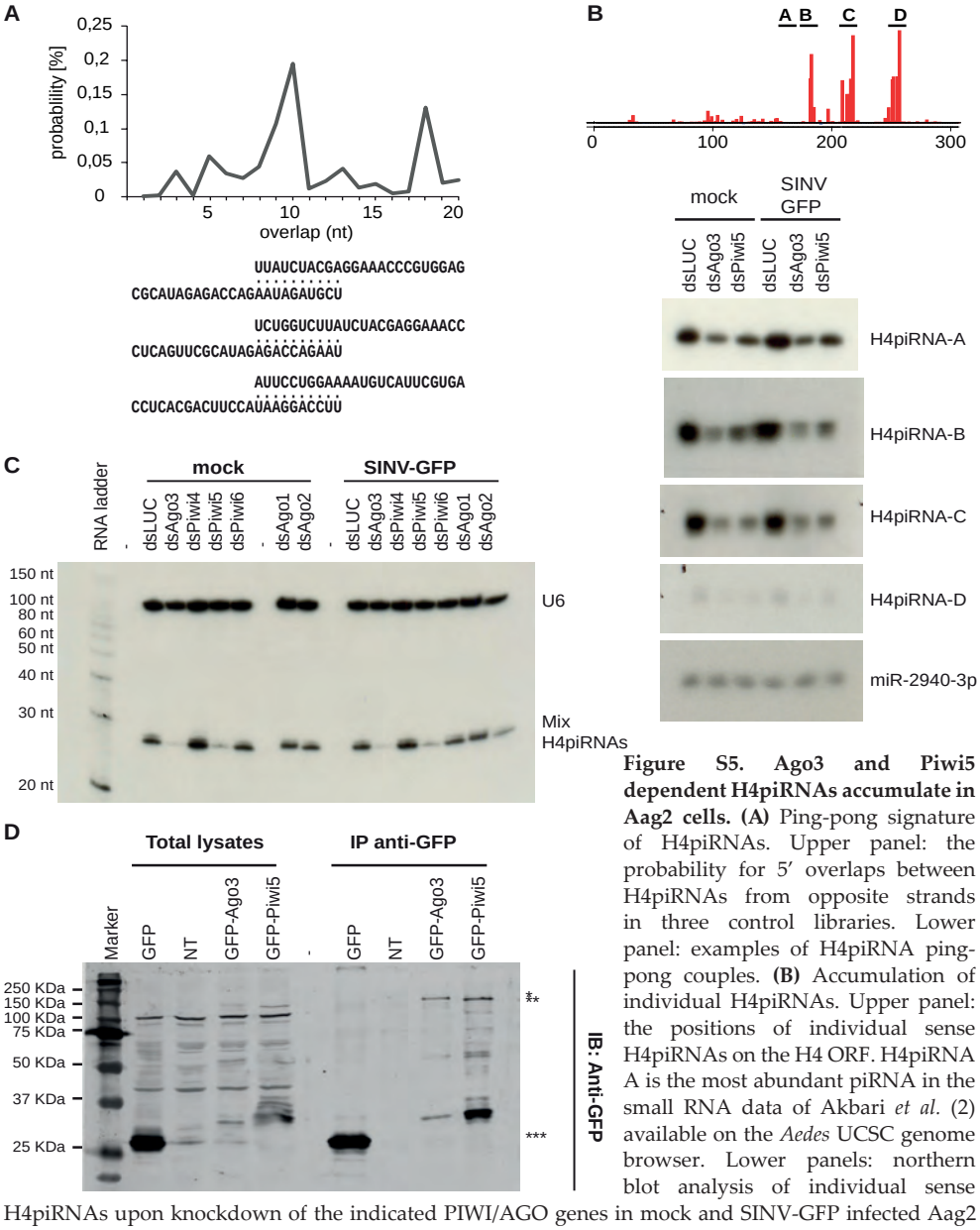


Figure S5. Ago3 and Piwi5 dependent H4piRNAs accumulate in Aag2 cells. (A) Ping-pong signature of H4piRNAs. Upper panel: the probability for 5′ overlaps between H4piRNAs from opposite strands in three control libraries. Lower panel: examples of H4piRNA ping-pong couples. **(B)** Accumulation of individual H4piRNAs. Upper panel: the positions of individual sense H4piRNAs on the H4 ORF. H4piRNA A is the most abundant piRNA in the small RNA data of Akbari *et al.* (2) available on the *Aedes* UCSC genome browser. Lower panels: northern blot analysis of individual sense H4piRNAs upon knockdown of the indicated PIWI/AGO genes in mock and SINV-GFP infected Aag2 cells. **(C)** Northern blot analysis of sense H4piRNAs upon knockdown of the indicated PIWI/AGO genes in SINV-GFP infected and mock-infected Aag2 cells. An RNA marker (10 to 150 nt) was loaded to define H4piRNA size. H4 piRNAs were detected using a pool of the four DNA oligonucleotide probes. U6 snRNA serves as loading control. All RNA samples have been analyzed on high resolution 17.5% polyacrylamide gel. **(D)** Full image of Ago3 and Piwi5 immunoprecipitation (IP) shown in Figure 3. Aag2 cells were transfected with expression plasmids for GFP (negative control), GFP-Ago3, or GFP-Piwi5 for 48 hours, harvested, and subjected to immunoprecipitation using GFP-trap beads. Immunoprecipitates and total lysates were analyzed by western blot using anti-GFP antibodies. NT, non transfected. Asterisks indicate the position of GFP-Ago3 (*), GFP-Piwi5 (**), and GFP (***).

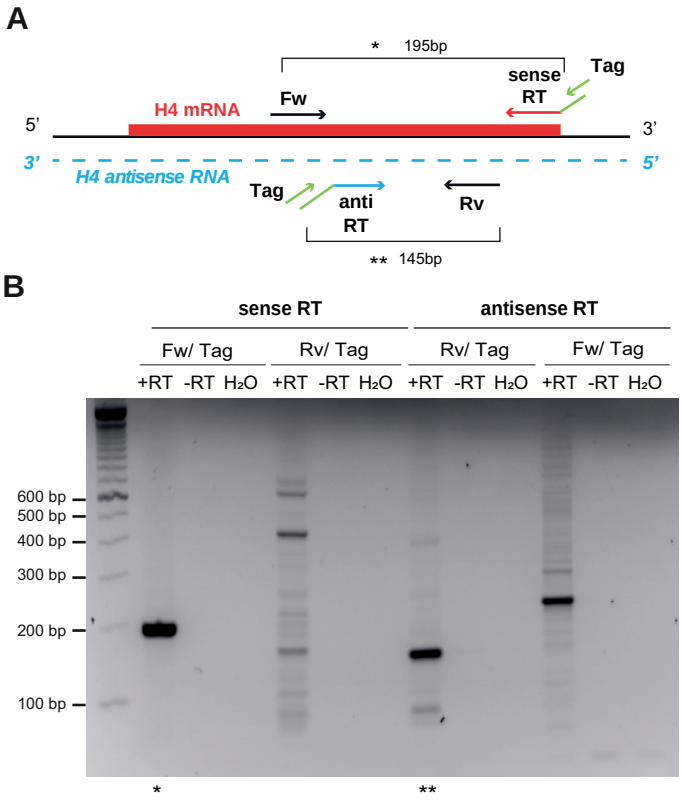


Figure S6. Detection of H4 antisense transcripts in Aag2 cells. (A) Schematic representation of strand-specific RT-PCR assay for the detection of sense and antisense histone 4 transcripts. Sense (red arrow) or antisense (blue arrow) strand-specific RT primer with a tag sequence (green line) were used for cDNA synthesis from Aag2 total RNA. Following cDNA synthesis, PCR was performed using a combination of a H4 and a tag-specific primer. As control, a 195 bp region on H4 cDNA (red) was amplified using primer Fw (black arrow) and primer Tag (green arrow). The presence of an antisense transcript (light blue) was analyzed by PCR amplification with primer Rv (black arrow) and primer Tag (green arrow). **(B)** PCR was performed using the indicated primer combinations on sense and antisense RTs. +RT correspond to the reverse transcribed sample. Reaction without reverse transcriptase (-RT) and PCR amplification without cDNA template (H2O) were performed to verify the absence of contaminating DNA in RNA preparations and PCR reagents, respectively. PCR amplicons were separated on a 2% agarose gel, stained with ethidium bromide. (*) and (**) indicate the expected amplicons.

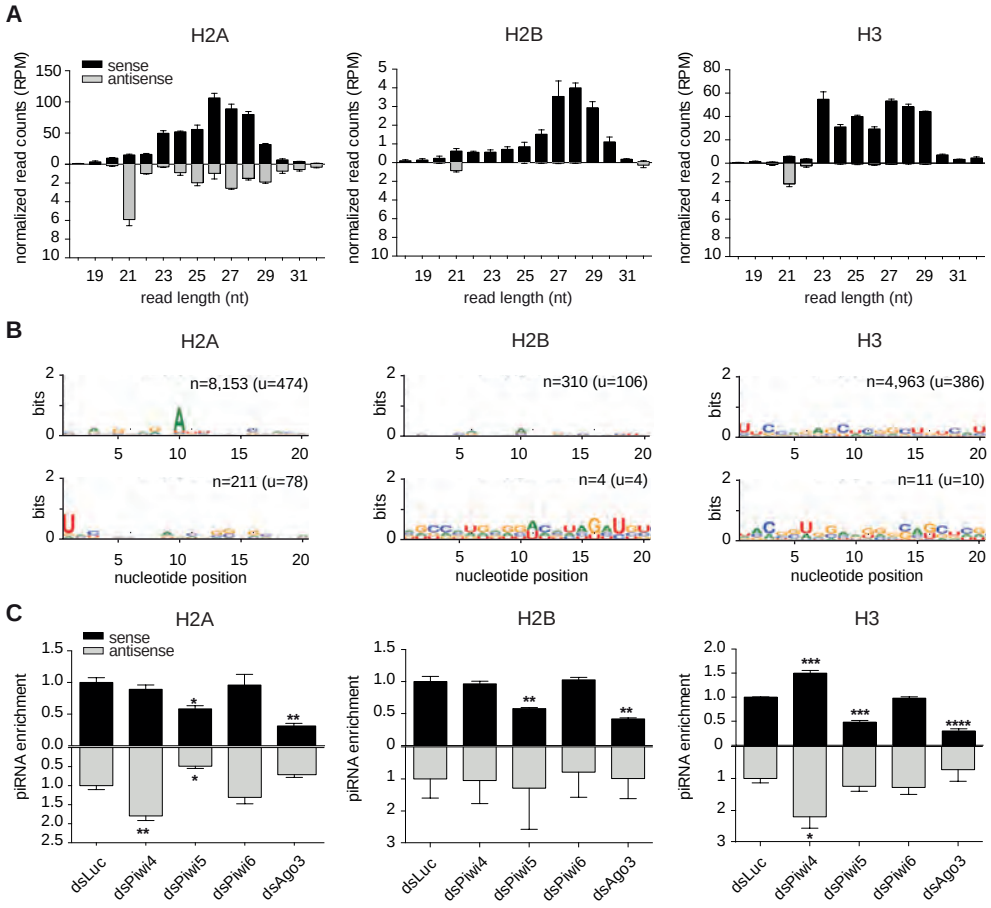


Figure S7. Histone 2A, 2B and 3-derived piRNAs in Aag2 cells. (A) Size profile of small RNA reads derived from H2A, H2B and H3 histone genes in Aag2 cells. Black and grey bars indicate sense and antisense reads, respectively. **(B)** Nucleotide bias at each position of the 25-30 nt small RNA reads mapping to the sense (upper panels) and antisense histone sequence (lower panels). All reads of three independent experiments were combined to generate the sequence logo. *n*, number of reads; *u*, number of unique sequences. **(C)** Relative abundance of the 25-30 nt sense (black) and antisense (grey) histone reads in the indicated PIWI knockdown libraries. Bars are the means +/- SD of three independent small RNA libraries. Two-tailed student's t-test was used to determine statistical significance (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$).

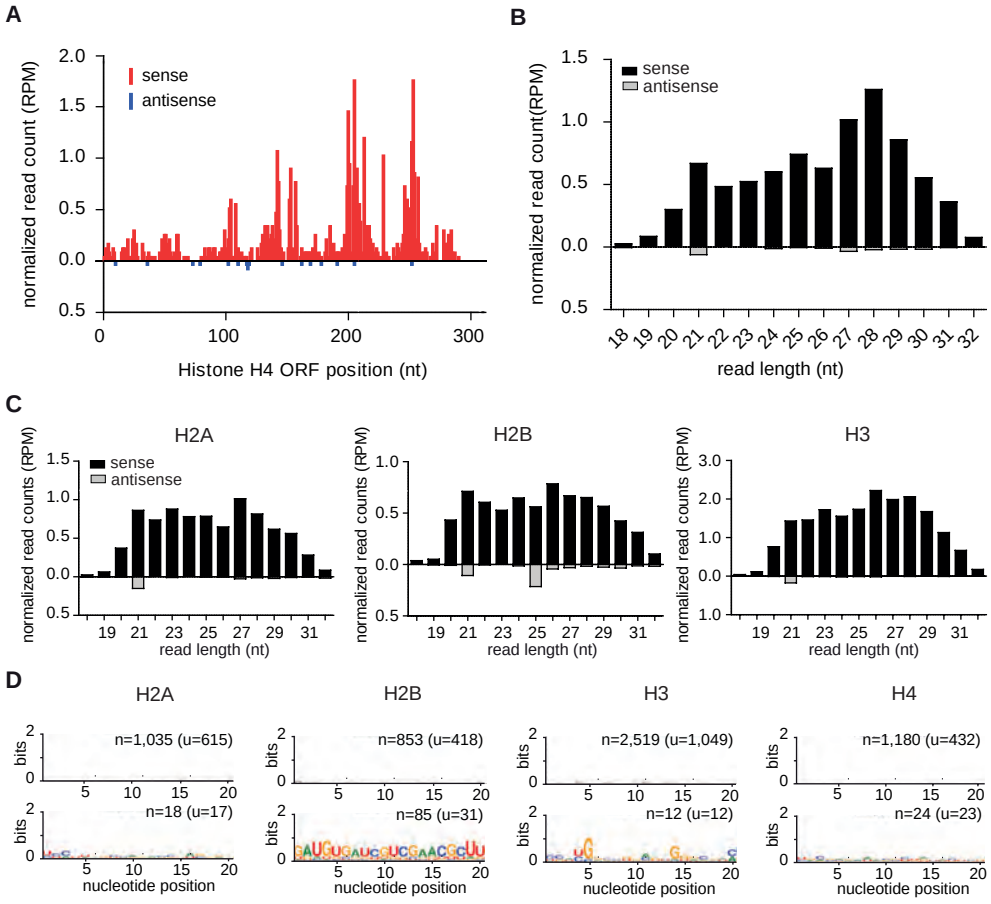


Figure S8. Histone-derived piRNAs in adult *Ae. aegypti* mosquitoes. (A) Distribution of 25-30 nt sense (red) or antisense (blue) RNA reads across the H4 open reading frame (ORF). The counts of 5′ ends of small RNA reads at each nucleotide position are shown. **(B)** Size profile of small RNA reads derived from the H4 gene. Black and grey bars indicate sense and antisense reads, respectively. **(C)** Size profile of small RNA reads derived from H2A, H2B and H3 histone genes in *Ae. aegypti* mosquitoes. Black and grey bars indicate sense and antisense reads, respectively. **(D)** Nucleotide bias at each position of the 25–30 nt small RNA reads mapping to the sense (upper panels) and antisense histone sequence (lower panels). *n*, number of reads; *u*, number of unique sequences. The size distribution of histone piRNAs in adult mosquitoes is broader than in Aag2 cells, likely reflecting the accumulation of non-specific degradation products, which were also seen in the northern blot (Figure 3F). Maybe for this reason, the piRNA-sized reads do not show a strong nucleotide bias.

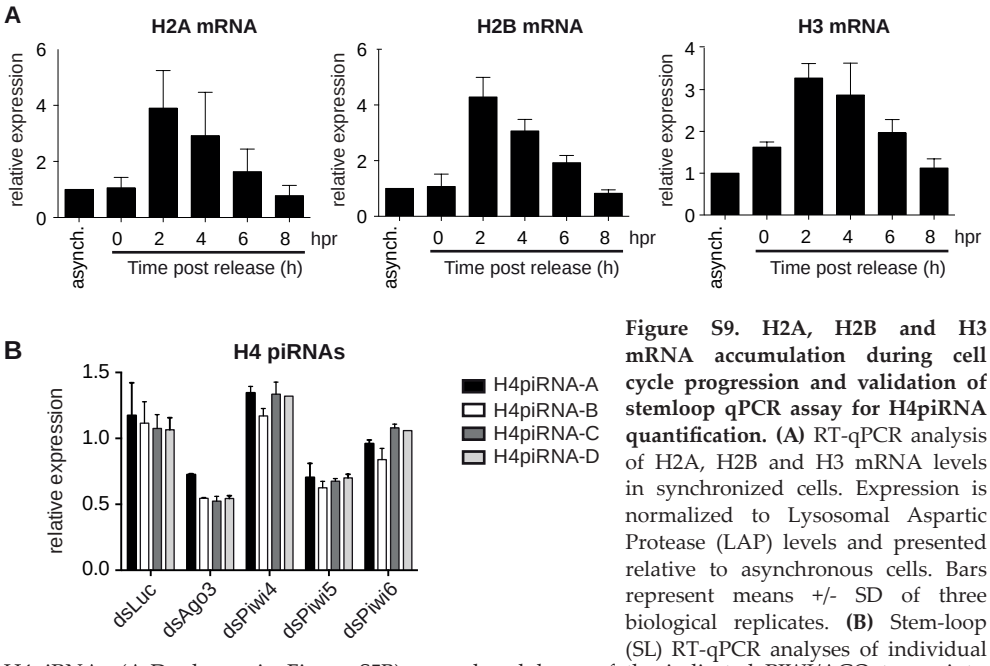


Figure S9. H2A, H2B and H3 mRNA accumulation during cell cycle progression and validation of stemloop qPCR assay for H4piRNA quantification. (A) RT-qPCR analysis of H2A, H2B and H3 mRNA levels in synchronized cells. Expression is normalized to Lysosomal Aspartic Protease (LAP) levels and presented relative to asynchronous cells. Bars represent means +/- SD of three biological replicates. **(B)** Stem-loop (SL) RT-qPCR analyses of individual H4piRNAs (A-D, shown in Figure S5B) upon knockdown of the indicated PIWI/AGO transcripts. Expression is normalized to aae-bantam-3p levels and presented relative to dsLuc. Bars represent means +/- SD of two biological replicates.

Available online:

- **Table S1.** Oligonucleotides used in this study
- **Table S2.** VectorBase accession numbers for annotated *Ae. aegypti* histone genes analyzed in this study

The data can be downloaded from:
https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw1368

SUPPLEMENTAL MATERIALS AND METHODS

Cells and virus infection

Aag2 cells were cultured at 25°C in Leibovitz's L-15 medium (Invitrogen) supplemented with 10% heat inactivated fetal calf serum (PAA), 2% tryptose Phosphate Broth Solution (Sigma), 1x MEM Non-Essential Amino Acids (Invitrogen) and 50 U/ml penicillin and 50 μg/ml streptomycin (Invitrogen). The virus used throughout this study is a Sindbis virus recombinant expressing GFP from a duplicated sub-genomic promoter (pTE-3'2J-GFP, SINV-GFP), which was produced in BHK-21 cells as previously described (3). Aag2 cells were infected with SINV-GFP at a multiplicity of infection (MOI) of 1 for 48 hours.

Mosquito manipulation for small RNA libraries

For small RNA library preparation, field-derived *Aedes aegypti* mosquitoes originally collected in Nakhon Chum, Muang District, Kamphaeng Phet, Thailand were used within 3 generations of laboratory colonization. Seven-day-old female mosquitoes were allowed to feed on pre-washed rabbit blood meals for 30 minutes at 37°C. After blood feeding, engorged females were incubated at 28°C with 70% humidity for 7 days. Total RNA from a pool of five mosquitoes was isolated with TRIzol (Invitrogen). Size fractionation of small RNAs of 19-33 nt in length was performed as described in (4). Purified RNA was used for library preparation using the NEBNext Multiplex Small RNA Library Prep kit for Illumina (E7300L). Libraries were diluted to 4 nM and sequenced using NextSeq 500 High Output Kit v2 (75 cycles) on a NextSeq 500 (Illumina, San Diego, CA, USA).

Generation of plasmids and dsRNA production

Insect expression vectors based on the *Drosophila* Gateway Vector pAGW (kindly provided by the Carnegie Institution for Science) were constructed for N-terminal tagging of proteins with GFP. The full-length coding sequence of Ago3 and Piwi5 was amplified from Aag2 complementary DNA (cDNA) and cloned by recombination downstream of the tag sequences according to the Gateway manufacturer's instructions (Invitrogen).

For dsRNA production, *in vitro* transcription using T7 RNA polymerase was performed on T7 promoter-flanked PCR products. To allow the formation of double-stranded RNA, the reaction products were heated to 80°C and then gradually cooled to room temperature. Subsequently, the RNA was purified using the GenElute Mammalian Total RNA Miniprep Kit (Sigma) following the manufacturer's instructions. Primers used for construction of plasmids and dsRNA production are indicated in Table S1.

Northern blotting and qPCR

Small RNA northern blot was performed as described previously (5). Briefly, total RNA was isolated using Isol-RNA Lysis Reagent (5 PRIME) and 5 μg of RNA was separated on a 17.5% PAGE gel, blotted to a nylon membrane (Hybond NX; Amersham) and cross-linked using 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC; Sigma). For detection of antisense H4piRNAs, 15 μg of total RNA was used. NaIO4 oxidation and β-elimination were performed as described in (5). For northern blot analyses on adult *Ae. aegypti*, total RNA from ten male, female, or blood-fed female mosquitoes was used (kindly provided by In2Care, Wageningen, The Netherlands). Hybridization with ^{32}P-labeled DNA oligonucleotides or *in vitro*-transcribed riboprobes was performed overnight at 42°C. The membrane was washed in 0.1% SDS, 2x SSC, followed by two washing steps in 0.1% SDS, 1x SSC and 0.1% SDS, 0.1x SSC, respectively. All washes were performed at 42°C. For detection of the radioactive signal, the membrane was exposed to a Carestream Kodak Biomax XAR film (Sigma Aldrich). Sequences of northern blot

probes are indicated in Table S1.

For quantitative RT-PCR (RT-qPCR), 1 μg of total RNA was DNaseI-treated (Ambion) and reverse-transcribed using the Taqman Reverse transcription kit (Roche) with random primers following the manufacturer's instructions. qPCR reactions were prepared using GoTaq qPCR SYBR Mastermix (Promega) and measured on a LightCycler 480 (Roche). Expression was internally normalized against the expression of Lysosomal Aspartic Protease (LAP) and the relative mRNA abundance was determined using to the ΔΔCt method (3). The primers used for qPCR are indicated in Table S1.

Stem-Loop RT-qPCR for piRNA quantification

For Stem-Loop RT-qPCR assays, 100 ng of total RNA was reverse transcribed in a final volume of 7.5 μl, in the presence of 0.5 μl of each Stem-Loop oligonucleotide (SL_Aae_H4piRNA_A to D, and SL-bantam-3p; 0.75 μM), 1.5 μl 5xFirst Strand buffer (Invitrogen), 1 μl dNTPs (0.25 mM; Qiagen), 0.125 μl Superscript II Reverse Transcriptase (200 U/μl; Invitrogen) and 0.1 μl RNase Inhibitor (20 U/μl; Applied Biosystems). Reactions were incubated for 30 minutes at 60°C, for 30 minutes at 42°C, and for 5 minutes at 85°C. Samples were placed on ice and were adjusted to 40 μl. Quantification was done by qPCR on a LightCycler 480 (Roche). Briefly, 0.6 μl forward primer (Fw_Aae_H4piRNA_A to D, F-bantam-3p; 10 μM) and 0.6 μl universal reverse primer R-univ-sRNAqPCR (10 μM), 3.8 μl MilliQ water and 10 μl GoTaq qPCR Master Mix (2x; Promega) were added to 5 μl RT reaction mix. After an incubation of 5 minutes at 95°C, 40 amplification cycles were performed (10 seconds at 95°C, 20 seconds at 60°C, and 10 seconds at 72°C). Primer sequences are provided in Table S1.

Strand specific RT-PCR

For strand-specific RT-PCR assays, cDNA synthesis was performed on 1 μg of DNase I-treated RNA using TaqMan Reverse Transcription Reagents in a 20 μl reaction according to the manufacturer's instructions (Applied Biosystems), using strand-specific primers tagged with a 5′ T7 promoter sequence (Table S1). Following cDNA synthesis, PCR analysis was performed using a combination of a H4-specific primer and a primer specific for the T7 promoter sequence (Table S1). The following control reactions were run in parallel to each sample: cDNA synthesis without reverse transcriptase was performed to verify the absence of contaminating DNA in RNA preparations; PCR amplification without cDNA template was used to exclude contaminations in PCR reagents.

Immunoprecipitation

Lysates from Aag2 cells expressing GFP-tagged PIWI proteins were incubated with GFP-Trap magnetic beads (Chromotek) according to the manufacturer's instructions. The immunoprecipitates were washed twice in wash buffer (10 mM Tris-HCl pH 7.5, 150

mM NaCl, 0.5 mM EDTA) and split for either RNA or protein analyses. The bound RNA was isolated from the beads using Isol-RNA Lysis Reagent, extracted, and analyzed by small RNA northern blot. The bound proteins were isolated from the beads using 2x SDS sample buffer (120 mM Tris-HCl pH 6.8, 20% glycerol, 4% SDS, 0.04% bromophenol blue, 10% beta-mercaptoethanol) and analyzed by western blot. Rabbit anti-GFP antibody (1:10,000) and secondary IRdye800 conjugated goat anti-rabbit (1:10,000, LI-COR) were used to detect the proteins of interest. Odyssey CLx Imaging System was used to acquire images.

Bioinformatic analyses of small RNA libraries

Characterization of control (dsLuc) small RNA libraries

The small RNA libraries from SINV-infected Aag2 cells have been characterized previously (5). Small RNA sequences from *Ae. aegypti* mosquitoes have been deposited in NCBI Sequence Reads Archive (accession SRA291268).

Small RNA reads were mapped to the *Ae. aegypti* genome (AaegL3, downloaded from VectorBase) using Bowtie (Galaxy tool version 1.1.2), allowing no mismatches in the first 28nt of each read. For mapping to the genomes of persistently infecting viruses (Aedes aegypti densovirus, GenBank accession M37899.1; mosquito X virus segment A and B, GenBank JX403941.1 and JX403942; cell fusing agent virus, GenBank NC_001564.1) one mismatch in the first 28 nt was permitted. Before *Ae. aegypti* genome-derived piRNAs were analyzed, reads mapping to SINV-GFP or to the persistently infecting viruses were removed from the libraries. Subsequently, reads in the size range of 25-30 nt were selected.

The genome positions of the piRNA-sized reads were overlapped with the genome locations of repetitive elements present in the *Ae. aegypti* genome ('AaegL3 repeatfeatures' downloaded from VectorBase). To determine the number of piRNAs that derive from transposable elements, reads that overlap 'TEfam elements' within the repeatfeatures library were counted. All piRNA reads that overlapped repeat features other than TEfam elements were designated as 'other repeats'.

For the subsequent analysis of piRNA reads that overlap (non-)coding genes, reads that intersected with any type of repetitive element were excluded. To identify piRNA-sized reads that map to coding RNAs, the 'mRNA' elements from the 'AaegL3 basefeatures library' (downloaded from VectorBase) were extracted and the transcript IDs from VectorBase were replaced with the corresponding gene ID. Next, the genomic positions were overlapped with the small RNAs using the 'intersect genomic intervals' tool in Galaxy. From these intersected datasets, the number of overlapping piRNAs was determined. Non-coding RNAs were defined as the collection of tRNAs, miRNAs, rRNA, snRNAs, snoRNAs, misc RNA, pseudogenes, RNase MRP RNA, RNase P RNA, SRP RNA, and antisense RNAs. Their genomic positions were extracted from the basefeatures

library, intersected with the position of piRNAs, and the number of overlapping piRNAs was determined.

Comparison of piRNA levels in PIWI knockdown or IP libraries

25-30 nt piRNA reads that mapped to the *Ae. aegypti* genome and did not overlap with repeated elements were selected from the individual PIWI knockdown or IP libraries described in Miesen et al (5). The genomic positions of piRNAs were joined to mRNA positions using the 'join' tool of the 'operate on genomic intervals' section in Galaxy. In the joined datasets the occurrence of individual mRNA names was counted to obtain piRNA counts, which were subsequently normalized to the total number of reads in the corresponding library. Finally, the fold change of normalized piRNA levels was calculated for every PIWI knockdown or IP library compared to the control libraries (dsLuc or GFP-IP, respectively).

Characterization of piRNAs mapping to individual/small groups of transcripts

Small RNA libraries were mapped to FASTA-formatted *Ae. aegypti* transcripts available from VectorBase. Small RNA sequencing data for individual or groups of transcripts (for instance all H2A, H2B, H3 or H4 genes) were selected from the mapped reads file. Small RNA size profiles were generated from all reads that map to the transcript in sense or antisense orientation with a maximum of one nucleotide mismatch in the first 28 nt. The small RNA distribution along the transcript was plotted as the number of 5′ ends starting at the individual nucleotide position of the transcripts. Nucleotide biases were determined with the WebLogo3 program (Sequence logo generator Galaxy tool version 0.4). For presentation of the genome distribution of H4 piRNAs on the total collection of histone 4 genes, the ORFs of the individual histone transcripts were aligned from start codon to stop codon irrespective of few single nucleotide polymorphisms. Subsequently, the combined count of small RNA 5′ ends was plotted for every nucleotide position on the ORF. The VectorBase accession numbers for the histone H2A, H2B H3 and H4 families analyzed in this study are shown in Table S2.

Comparison of Girardi et al. and Haac et al. datasets

To compare our dataset with independently generated small RNA libraries from Aag2 cells, publically available data (1) were imported into Galaxy (SRA submissions SRR1765315, SRR1765316 and SRR1765317). These libraries have been generated from size-purified small RNAs from Aag2 cells transfected with dsRNA targeting EGFP using Illumina's small RNA Truseq sample prep kit. They have been sequenced on a HiSeq2500 and have a combined sequencing depth of more than 84 million reads. Gene-derived piRNAs were analyzed as described above and the correlation between our data and the Haac *et al.* data was analyzed using a Spearman's correlation analysis.

SUPPLEMENTAL REFERENCES

1. Haac ME, Anderson MA, Eggleston H, Myles KM, Adelman ZN. The hub protein loquacious connects the microRNA and short interfering RNA pathways in mosquitoes. Nucleic Acids Res. 2015;43(7):3688-700.
2. Akbari OS, Antoshechkin I, Amrhein H, Williams B, Diloreto R, Sandler J, et al. The developmental transcriptome of the mosquito Aedes aegypti, an invasive species and major arbovirus vector. G3 (Bethesda). 2013;3(9):1493-509.
3. Vodovar N, Bronkhorst AW, van Cleef KW, Miesen P, Blanc H, van Rij RP, et al. Arbovirus-derived piRNAs exhibit a ping-pong signature in mosquito cells. PLoS One. 2012;7(1):e30861.
4. Gausson V, Saleh M.-C. Viral small RNA cloning and sequencing. Methods Mol. Biol. 2011;721: 107-22.
5. Miesen P, Girardi E, van Rij RP. Distinct sets of PIWI proteins produce arbovirus and transposon-derived piRNAs in Aedes aegypti mosquito cells. Nucleic Acids Res. 2015;43(13):6545-56.

Chapter 5

Comparative Genomics Shows that Viral Integrations are Abundant and Express piRNAs in the Arboviral Vectors *Aedes aegypti* and *Aedes albopictus*.

adapted from:

Umberto Palatini*, Pascal Miesen*, Rebeca Carballar-Lejarazu, Lino Ometto, Ettore Rizzo, Zhijian Tu, Ronald P. van Rij, and Mariangela Bonizzoni

BMC Genomics (2017) 18:512

*equal contribution

ABSTRACT

Background

Arthropod-borne viruses (arboviruses) transmitted by mosquito vectors cause many important emerging or resurging infectious diseases in humans including dengue, chikungunya and Zika. Understanding the co-evolutionary processes among viruses and vectors is essential for the development of novel transmission-blocking strategies. Arboviruses form episomal viral DNA fragments upon infection of mosquito cells and adults. Additionally, sequences from insect-specific viruses and arboviruses have been found integrated into mosquito genomes.

Results

We used a bioinformatic approach to analyze the presence, abundance, distribution, and transcriptional activity of integrations from 425 non-retroviral viruses, including 133 arboviruses, across the presently available 22 mosquito genome sequences. Large differences in abundance and types of viral integrations were observed in mosquito species from the same region. Viral integrations are unexpectedly abundant in the arboviral vector species *Aedes aegypti* and *Ae. albopictus*, but are ~10-fold less abundant in all other mosquitoes analyzed. Additionally, viral integrations are enriched in piRNA clusters of both the *Ae. aegypti* and *Ae. albopictus* genomes and, accordingly, they express piRNAs, but not siRNAs.

Conclusions

Differences in number of viral integrations in the genomes of mosquito species from the same geographic area support the conclusion that integrations of viral sequences is not dependent on viral exposure, but that lineage-specific interactions exits. Viral integrations are abundant in *Ae. aegypti* and *Ae. albopictus*, and represent a thus far unappreciated component of their genomes. Additionally, the genome locations of viral integrations and their production of piRNAs indicate a functional link between viral integrations and the piRNA pathway. These results greatly expand the breadth and complexity of small RNA-mediated regulation and suggest a role for viral integrations in antiviral defense in these two mosquito species.

INTRODUCTION

Nearly one-quarter of emerging or resurging infectious diseases in humans are vector-borne (1). Hematophagous mosquitoes of the *Culicidae* family are the most serious vectors in terms of their worldwide geographic distribution and the public health impact of the pathogens they transmit. The *Culicidae* is a large family whose members separated between 180 to 257 million years ago into the *Culicinae* and *Anophelinae* subfamilies (2). Mosquitoes of the *Aedes* and *Culex* genera within the *Culicinae* subfamily are the primary vectors of RNA viruses. These viruses include taxa with different RNA genomic structures and replication strategies, but all are non-retroviral viruses (3). Collectively,

these viruses are referred to as arthropod-borne (arbo-) viruses. Within the *Aedes* genus, *Aedes aegypti* and *Aedes albopictus* are the main arboviral vectors due to their broad geographic distribution, adaptation to breed in human habitats, and the wide number of viral species from different genera that they can vector (4,5). These two mosquito species are able to efficiently transmit arboviruses of the genera Flavivirus (e.g. dengue viruses [DENV], Zika virus [ZKV], Usutu, Japanese encephalitis and yellow fever viruses), Alphavirus (e.g. chikungunya virus [CHIKV]), viruses of the Venezuelan equine encephalitis [VEE] and eastern equine encephalitis [EEE] complexes), Orthobunyavirus (e.g. Potosi, Cache Valley and La Crosse virus [LACV]), Phlebovirus (e.g. Rift Valley fever virus [RVFV]) and Orbivirus (e.g. Orungo Virus) (5-7). Mosquitoes of the *Culex pipiens* complex, such as *Cx. pipiens pipiens* and *Cx. quinquefasciatus*, are the most prominent *Culex* vectors because of their wide distribution and close association with humans (7). These mosquito species are primary vectors of encephalitic flaviviruses, such as West Nile virus (WNV) and Japanese encephalitis virus, and they can also vector RVFV (7,8). The only arbovirus know to be transmitted by *Anophelinae* is the alphavirus O'nyong-nyong (9). Recently, additional RNA viruses have been identified from wild mosquitoes, but their virulence to humans and their impact on vector competence is still uncertain (9-11).

Mosquito competence for arboviruses is a complex and evolving phenotype because it depends on the interaction of genetic factors from both mutation-prone RNA viruses and mosquito vectors with environmental variables (12-15). Not surprisingly, large variation exists in vector competence not only among mosquito species, but also across geographic populations within a species (16,17). Understanding the genetic components of vector competence and how these genetic elements are distributed in natural populations and interact with environmental factors is essential for predicting the risk of arboviral diseases and for developing new transmission-blocking strategies (12). Genomic and functional studies, primarily in *Drosophila melanogaster* and *Aedes* mosquitoes, have shown that RNA interference (RNAi) is the main antiviral mechanism in insects (18-20). In this pathway, small RNAs are used to guide a protein-effector complex to target RNA based on sequence-complementarity. Three RNA silencing mechanisms exist: the microRNA, small interfering RNA (siRNA) and PIWI-interacting RNA (piRNA) pathways. They can be distinguished based on the mechanism of small RNA biogenesis and the effector protein complex to which these small RNAs associate (18,19). While the role of the siRNA pathway in restricting arboviral infection has been widely studied and appears universal across mosquitoes, recent studies highlight the contribution of the piRNA pathway in antiviral immunity of *Aedes* mosquitoes (21). Although important aspects of piRNA biogenesis and function in mosquitoes remains to be elucidated, it is clear that endogenous piRNAs arise from specific genomic loci called piRNA clusters, as was originally observed in *D. melanogaster* (22). These piRNA clusters contain repetitive sequences, remnants of transposable elements and, in *Ae. aegypti*, virus-derived sequences (23).

Recent studies have shown that the genomes of some eukaryotic species, including mosquitoes, carry integrations from non-retroviral RNA viruses (24-32). Viral integrations are generally referred to as Endogenous Viral Elements (EVEs) (33) or, if they derive from non-retroviral RNA viruses, as Non-Retroviral Integrated RNA Viruses Sequences (NIRVS) (29,34). Integration of non-retroviral sequences into host genomes is considered a rare event because it requires reverse transcription by an endogenous reverse transcriptase, nuclear import and genomic insertion of virus-derived DNA (vDNA) (35). During infection with DENV, WNV, Sindbis virus, CHIKV and LACV, fragments of RNA virus genomes are converted into vDNA by the reverse transcriptase activity of endogenous transposable elements (TEs) in cell lines derived from *D. melanogaster, Culex tarsalis*, *Ae. aegypti,* and *Ae. albopictus,* as well as in adult mosquitoes. The episomal vDNA forms produced by this mechanism reside in the nucleus and it has been proposed that they contribute to the establishment of persistent infections through the RNAi machinery (20,36,37). These recent studies not only show that reverse transcription of RNA viruses occurs in *Culicinae*, they also suggest the functional involvement of RNAi.

Here we used a bioinformatics approach to analyze the presence, abundance, distribution, and transcriptional activity of NIRVS across the currently available 22 mosquito genome sequences. We probed these genomes for integrations from 425 non-retroviral viruses, including 133 arboviruses. We observed a ten-fold difference in the number of NIRVS between *Aedes* and the other tested mosquitoes. NIRVS were not evenly distributed across *Aedes* genomes, but occurred preferentially in piRNA clusters and, accordingly, they produced piRNAs. Among the viral species tested, integrations had the highest similarities to rhabdoviruses, flaviviruses and bunyaviruses, viruses that share the same evolutionary origin (38). The larger number of NIRVS identified in *Ae. aegypti* and *Ae. albopictus*, their genome locations and their production of piRNAs show that in these species genomic integrations of viral sequences is a more pervasive process than previously thought and we propose that viral integrations contributes to shape vector competence.

MATERIALS AND METHODS

In silico screening of viral integrations

Genome assemblies of *D. melanogaster* and 22 currently available mosquito species were screened in silico using tBLASTx and a library consisting of genome sequences of 424 non-retroviral RNA viruses and one DNA arbovirus (Table S1).

Tested mosquito species were classified in arboviral (*Aedes aegypti, Aedes albopictus*, *Culex quinquefasciatus*) and protozoan (*Anopheles gambiae, Anopheles albimanus, Anopheles arabiensis, Anopheles darling, Anopheles stephensi, Anopheles funestus, Anopheles atroparvus, Anopheles coluzzii, Anopheles culicifacies, Anopheles dirus, Anopheles epiroticus, Anopheles farauti, Anopheles maculatus, Anopheles melas, Anopheles merus, Anopheles minimus, Anopheles*

sinensis) vectors depending on whether they most efficiently transmit arboviruses or protozoans to humans, respectively (Table S2). The non-vector *Anopheles christiy* and *Anopheles quadriannulatus* were also included in the analyses (59).

Host genome sequences of at least 100 bp and with high identity (e-values <0.0001) to viral queries were extracted from the respective insect genomes using custom scripts. When several queries mapped to the same genomic region, only the query with the highest score was retained. Blast hits were considered different when they mapped to genomic positions at least 100 bp apart from each other, otherwise they were included in the same NIRVS-locus.

All putative viral integrations were subjected to a three-step filtering process before being retained for further analyses to reduce the chance of false positives and ensure that the identified sequences are from non-retroviral RNA viruses (25). Filtering steps included 1) a reverse-search against all nucleotide sequences in the NCBI database using the BLAST algorithm, 2) a search for ORFs encompassing viral proteins based on NCBI ORFfinder and 3) a functional annotation based on Argot^2 (60).

Although our search expanded the range of viral integrations identified in *Ae. albopictus* and *Ae. aegypti* (2,26,28), we cannot exclude that refinements of the current genome annotations of the species analyzed, especially in repeat regions, the application of alternative bioinformatic pipelines and the identification of novel viral species could lead to the characterization of additional integrations. Additionally, to reduce chance of false positives, our bioinformatics pipeline focused on sequences in which we could unambiguously identify viral ORFs, thus excluding viral sequences coming from UTRs or sub-genomic regions.

Genomic data from 16 *Ae. albopictus* mosquitoes

Mosquitoes of the *Ae. albopictus* Foshan strain were used in this study. The strain was received from Dr. Chen of the Southern Medical University of Guangzhou (China) in 2013. Since 2013, the Foshan strain has been reared in an insectary of the University of Pavia at 70-80% relative humidity, 28°C and with a 12-12 h light–dark photoperiod. Larvae are fed on a finely ground fish food (Tetramin, Tetra Werke, Germany). A membrane feeding apparatus and commercially available mutton blood is used for blood-feeding females.

DNA was extracted from single mosquitoes using the DNeasy Blood and Tissue Kit (Qiagen, Hilden Germany) following manufacturer's protocol. DNA was shipped to the Polo D'Innovazione Genomica, Genetica e Biologia (Siena, Italy) for quality control, DNA-seq library preparation and sequencing on an Illumina HiSeq 2500. After quality control, retrieved sequences were aligned to the genome of *Ae. albopictus* reference Foshan strain (AaloF1 assembly) using the Burrows-Wheeler Aligner (BWA) (61) and marking identical read copies. The resulting indexed BAM files were used to calculate the counts of alignments, with mapping quality score above 10, which overlapped

intervals of *Ae. albopictus* NIRVS using BEDTools (62). Alignment files were visualized using the Integrative Genomics Viewer (63).

Phylogenetic analyses

Deduced NIRVS protein sequences were aligned with subsets of corresponding proteins from *Flavivirus, Rhabdovirus, Reovirus* and *Bunyavirus* genomes using MUSCLE. Maximum likelihood (ML) phylogenies were estimated in MEGA6 (64), implementing in each case the best fitting substitution model. Statistical support for inferred tree nodes was assessed with 1000 bootstrap replicates. Figures were generated using FIGTREE (v.1.4) (http://tree.bio.ed.ac.uk/software/figtree/).

Bioinformatic analyses of integration sites

Clustering of viral integrations in piRNA loci was estimated using cumulative binomial distribution, where the probability of integration was assumed to equal the fraction of the genome occupied by the respective genomic region. Genomic regions considered were piRNA clusters, coding regions and intergenic regions as previously defined (2,23,44). A value of $P < 0.05$ suggests a statistically significant enrichment of these events in the corresponding genomic region (Table 2).

Analyses of TE enrichment in all non-retroviral integration sites as well as region 1 and region 2 of *Ae. aegypti* were based on RepeatMasker (version open-4.0.3, default parameters) using *Ae. aegypti* TEs retrieved from TEfam (http://tefam.biochem.vt.edu/tefam/), which was manually annotated. We used percent TE occupancy (percent of bases in the genomic sequence that match TEs) as an indication for possible enrichment of certain TEs. We did not use TE copy number as indications for TE enrichment because it is likely that some TEs can be broken into multiple fragments and be counted multiple times. We retrieved sequences of the viral integration sites plus 5 kb sequences flanking each side of the integration for the analysis. In addition, to identify potentially full-length TEs, 10 kb sequences flanking each side of the viral integration were analyzed by RepeatMasker (version open-4.0.3). Presence of full-length TEs was verified by comparing the length of masked sequences with the length of the annotated TEs.

Analyses of piRNAs production from NIRVS

Small RNA deep-sequencing data of female *Ae. aegypti* (methoprene treated; SRX397102) (65) or *Ae. albopictus* mosquitoes (sugar-fed; SRX201600) (66) as well as PIWIs knockdown and IP libraries in Aag2 cells (SRA188616) (45) were downloaded from the European Nucleotide Archive. Subsequently, small RNA datasets were manipulated using the programs available in the Galaxy toolshed (67). After removal of the 3′ adapter sequences, small RNAs were mapped to NIRVS sequences that were oriented in the direction of the predicted ORF, using Bowtie permitting one mismatch in a 32 nt seed (68). From the

mapped reads, size profiles were generated. For the analysis of nucleotide biases, the 25-30 nt reads were selected and separated based on the strand. The FASTA-converted sequences of small RNA reads were then trimmed to 25 nt and used as input for the Sequence-Logo generator (Galaxy version 0.4 based on Weblogo 3.3 (69). piRNA counts on individual NIRVS were generated by mapping to NIRVS sequence after collapsing (near-) identical sequences (Table S5). Bowtie was used to map the small RNAs allowing one mismatch in a 32 nt seed. Only uniquely mapping reads were considered and the --best and the -- strata options were enabled. From the mapped reads, 25-30 nt small RNAs were selected. To identify secondary piRNAs, reads in sense orientation to viral ORFs that had an adenine at position 10 were selected. To avoid taking piRNAs into consideration that coincidentally contain a 10A, the population of 10A sense piRNAs was required to make up at least 50% of all sense piRNAs derived from the NIRVS. If this criterion was not met, sense reads from the corresponding NIRVS did not qualify as secondary piRNAs. Total piRNA counts and secondary piRNA counts were determined for F-NIRVS, R-NIRVS and Reovirus NIRVS and normalized to the corresponding library size. The size of the Ring-graph was scaled to reflect the total normalized read counts. piRNA counts on individual NIRVS was also determined from acetone treated female *Ae. aegypti*, male *Ae. aegypti*, blood-fed *Ae. albopictus* and male *Ae. albopictus* mosquitoes. The data was obtained from the same studies as described above.

To identify the PIWI dependency of NIRVS-derived piRNAs, we analyzed libraries from Aag2 cells transfected with double stranded RNA (dsRNA) targeting the somatic PIWI genes (Piwi4-6, Ago3) and a non-targeting control (dsRNA targeting luciferase, dsLuc) (45). These datasets were mapped against the collapsed NIRVS dataset as described above. Since small RNA profiles were dominated by piRNA-sized reads, no further size selection was performed. The mean fold change in small RNA read counts was calculated for each PIWI knockdown condition compared to the negative control. To identify the PIWI proteins that NIRVS piRNAs associate with, we analyzed the IP libraries of PIWI proteins in Aag2 cells previously published in the same study. For the different PIWI IPs the enrichment of small RNA counts compared to a control GFP-IP was determined. Hierarchical clustering of NIRVS based on the combined fold changes of PIWI knockdowns and IPs was performed using multiple experiment viewer (Tm4). Clustering was based on Pearson correlation and performed independently for F-NIRVS and R-NIRVS.

NIRVS transcriptional activity

RNA deep-sequencing data of *Ae. albopictus* and A*e. aegypti* mosquitoes, including both DENV-infected and non-infected mosquitoes were downloaded from NCBI's SRA. Libraries analyzed correspond to data SRA438038 for *Ae. albopictus*, and SRA058076, SRX253218, SRX253219 and SRX253220 for *Ae. aegypti.* RNA-seq reads were mapped

using BWA (61) to NIRVS, after collapsing identical sequences (Table S5), and read counts were converted into RPKM using custom scripts.

To analyze AlbFlavi34 expression in different *Ae. albopictus* developmental stages, total RNA was extracted using Trizol (Life Technologies) from 3 pools of 5 entities for each condition (eggs, larvae, adult males, blood-fed and sugar-fed females). From each pool, a total of 100 ng of RNA was used for reverse transcription using the qScript cDNA SuperMix following manufacturer's protocol (Quanta Biosciences). AlbFlavi34 expression was quantified in a 20 µL final reaction volume containing 10 µL of QuantiNova SYBR Green PCR Master Mix (Qiagen), 700 nM each forward (5'-CTTGCGACCCATGGTCTTCT-3') and reverse (5'-GTCCTCGGCGCTGAATCATA-3') primers and 5.0 µL cDNA sample on an Eppendorf RealPlex Real-Time PCR Detection System (Eppendorf). We used a two-step amplification protocol consisting of 40 cycles of amplification (95°C for 5 s, 60°C for 10 s) after an initial denaturation of 2 minutes at 95°C. AlbFlavi34 expression values were normalized to mRNA abundance levels of the *Ae. albopictus* Ribosomal Protein L34 (RPL34) gene (70). QBASE+ software was used to visualise data and compare expression profiles across samples. Absence of *Flavivirus* infection was verified using a published protocol (24) on all samples before qRT-PCR.

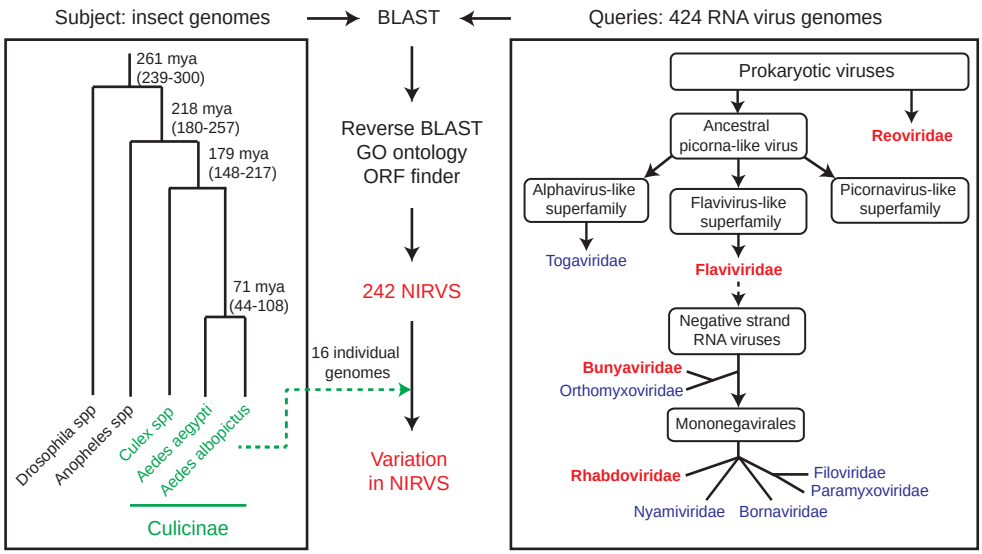


Figure 1. Pipeline for NIRVS identification. The currently available 22 mosquito genomes and the genome of *Drosophila melanogaster* were probed bioinformatically using tblastx and 425 viral species (424 non-retroviral RNA viruses and 1 DNA arbovirus). Tested insect and viral RNA genomes are shown in the context of their phylogeny (2, 38). Identified blast hits were parsed based on gene ontology and the presence of partial or complete viral ORFs. In *Ae. albopictus*, bioinformatic analyses was extended to whole-genome sequencing data from 16 individual mosquitoes of the Foshan strain. This stringent pipeline led to the characterization of 242 loci with NIRVS. Viral families for which NIRVS were characterized are shown in red.

RESULTS

NIRVS are unevenly distributed across mosquito species

25 genome assemblies currently available for 22 *Culicinae* species, along with the genome of *D. melanogaster*, were searched bioinformatically for sequence integrations derived from all 424 non-retroviral RNA viruses for which a complete genome sequence is currently available. Additionally, we tested the genome of African Swine Fever Virus, the only known DNA arbovirus (3), giving a total of 133 arboviruses (Figure 1, Table S1,S2). The genomes of 16 individual *Ae. albopictus* Foshan mosquitoes were sequenced to further validate NIRVS in this species. Retrieved sequences longer than 100 base pairs (bp) were filtered based on gene ontology and the presence of partial or complete open reading frames (ORFs) of viral proteins. This stringent pipeline led to the characterization of a total of 242 loci harboring NIRVS across the genome of 15

Table 1. Number of viral integrations (NIRVS) detected for each of the viral families tested across the 22 mosquito genomes. A total of 424 non-retroviral RNA viruses with complete genomes were analyzed. The genome of African swine fever virus, the only known DNA arbovirus was also included in the analyses, but no NIRVS were found for this virus.

Mosquito species	Families of tested non-retroviral RNA viruses (N. species)									
	Toga* (24)	Flavi* (92)	Bunya* (59)	Reo* (70)	Ortho-myxo* (4)	Rhabdo* (93)	Borna (6)	Filo (8)	Nyami (4)	Para-myxo (64)
Aedes aegypti	-	32	1	1	-	88	-	-	-	-
Aedes albopictus	-	30	-	-	-	42	-	-	-	-
Culex quinquefasciatus	-	-	-	-	-	1	-	-	-	-
Anopheles christy	-	-	-	-	-	-	-	-	-	-
Anopheles gambiae	-	-	-	-	-	-	-	-	-	-
Anopheles coluzzi	-	-	-	-	-	-	-	-	-	-
Anopheles arabiensis	-	-	1	-	-	4	-	-	-	-
Anopheles melas	-	-	-	-	-	-	-	-	-	-
Anopheles merus	-	-	-	-	-	2	-	-	-	-
Anopheles quadrianulatus	-	-	-	-	-	2	-	-	-	-
Anopheles epiroticus	-	-	-	-	-	7	-	-	-	-
Anopheles stephensi	-	-	-	-	-	-	-	-	-	-
Anopheles maculatus	-	-	-	-	-	2	-	-	-	-
Anopheles culicifacies	-	-	-	-	-	-	-	-	-	-
Anopheles minimus	-	1	-	-	-	1	-	-	-	-
Anopheles funestus	-	-	-	1	-	7	-	-	-	-
Anopheles dirus	-	-	-	-	-	4	-	-	-	-
Anopheles farauti	-	-	-	-	-	7	-	-	-	-
Anopheles atroparvus	-	-	-	-	-	3	-	-	-	-
Anopheles sinensis	-	1	-	-	-	1	-	-	-	-
Anopheles albimanus	-	-	-	-	-	-	-	-	-	-
Anopheles darlingi	-	-	-	-	-	-	-	-	-	-

*Virus families that contain arboviruses.

mosquitoes (Table 1, Figure 2). NIRVS loci were unevenly distributed across species. *Anophelinae* species had a maximum of 7 NIRVS-loci, one NIRVS-locus was found in *Cx. quinquefasciatus*, 122 NIRVS were detected in *Ae. aegypti*, and 72 were found in *Ae. albopictus*. The NIRVS landscape was highly variable across the 16 *Ae. albopictus* sequenced genomes with extensive differences in the number of NIRVS and in their length, suggesting that NIRVS are frequently rearranged (Figure 3). No read coverage was observed in any of the 16 sequenced genomes for a total of 10 integrations that had been identified bioinformatically from the genome assembly of the Foshan strain (Table S3). The percentage of mapped reads and coverage was comparable across libraries excluding insufficient sequence depth as an explanation for the differential presence of NIRVS (Table S4). It is currently unclear if these 10 NIRVS are rare integrations or result from mis-assembly of the reference genome. Among the 11 viral families tested, NIRVS had sequence similarities exclusively with viruses of the *Rhabdoviridae*, *Flaviviridae*, *Bunyaviridae* and *Reoviridae* families, including currently circulating viruses (Table 1). *Reoviridae*- and *Bunyaviridae*-like integrations were similar to recently characterized viruses (39,40) and were rare, with no more than one integration per species (Figure 2). Phylogenetic analyses showed that viral integrations from *Reoviridae* were separated from currently known viral species in this family (Figure 4A,B). Integrations from *Bunyaviridae* were at the base of the phylogenetic tree and clustered with newly identified viruses such as Imjin virus and Wutai mosquito virus (40,41) (Figure 4C).

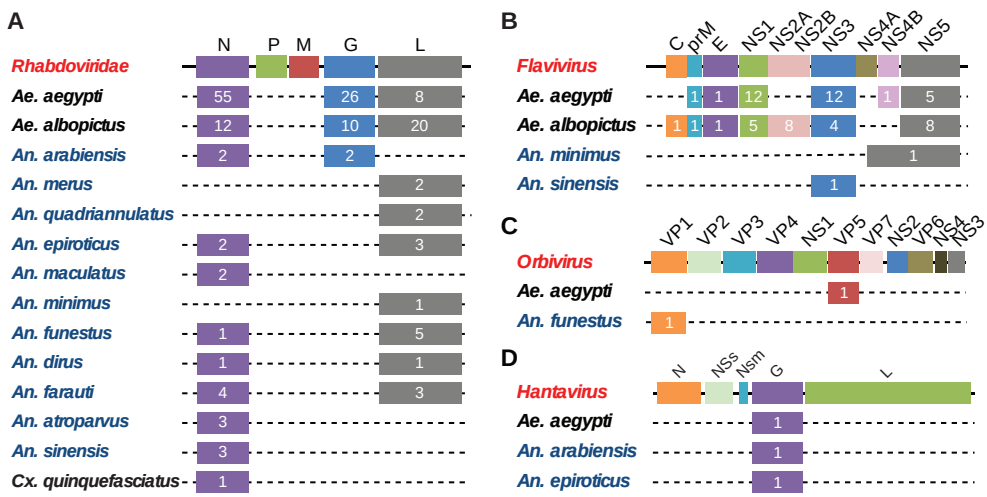


Figure 2. Different abundance of NIRVS across virus genera, genes and host species. (A-D) Schematic representation of the genome structures of *Rhabdoviridae* (A) and the genera *Flavivirus* (family *Flaviviridae*) (B), *Orbivirus* (family *Reoviridae*) (C) and *Hantavirus* (Family *Bunyaviridae*) (D). Numbers within each box represent the number of NIRVS loci spanning the corresponding viral gene per mosquito species. When a NIRVS locus encompassed more than one viral gene, the viral gene with the longest support was considered. Mosquitoes of the *Culicinae* and *Anophelinae* subfamilies are in black and blue, respectively. Dotted lines indicate viral integrations were not contiguous in the host genomes.

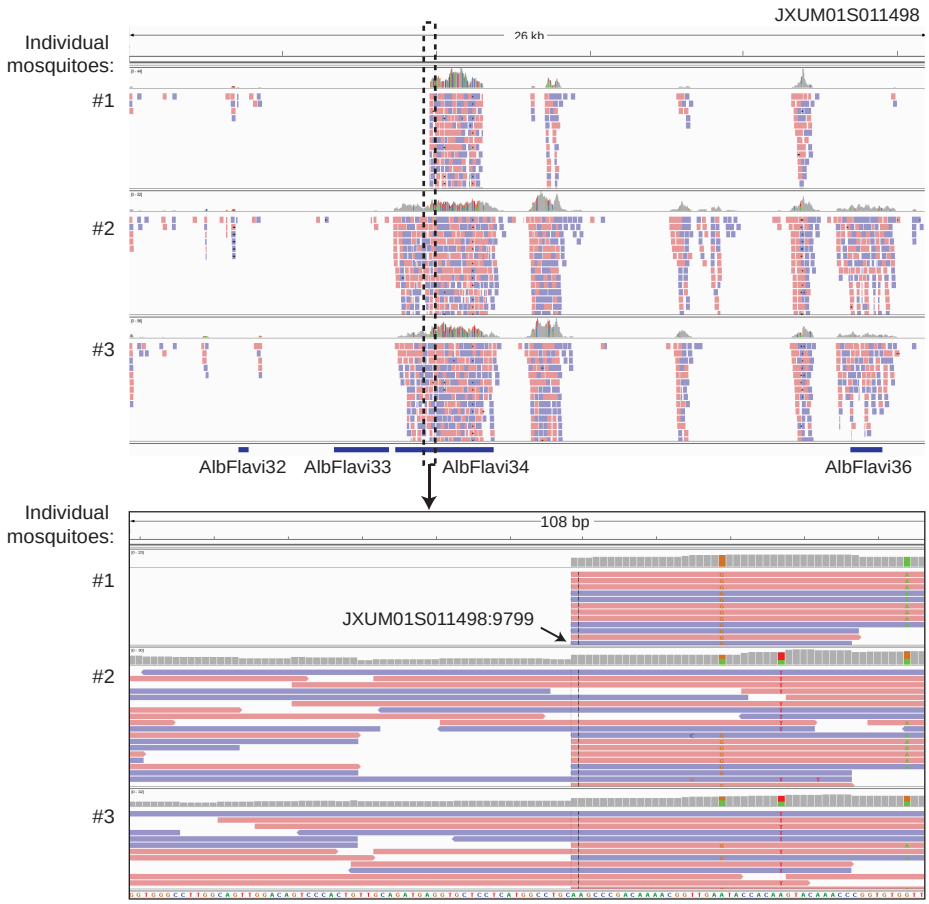
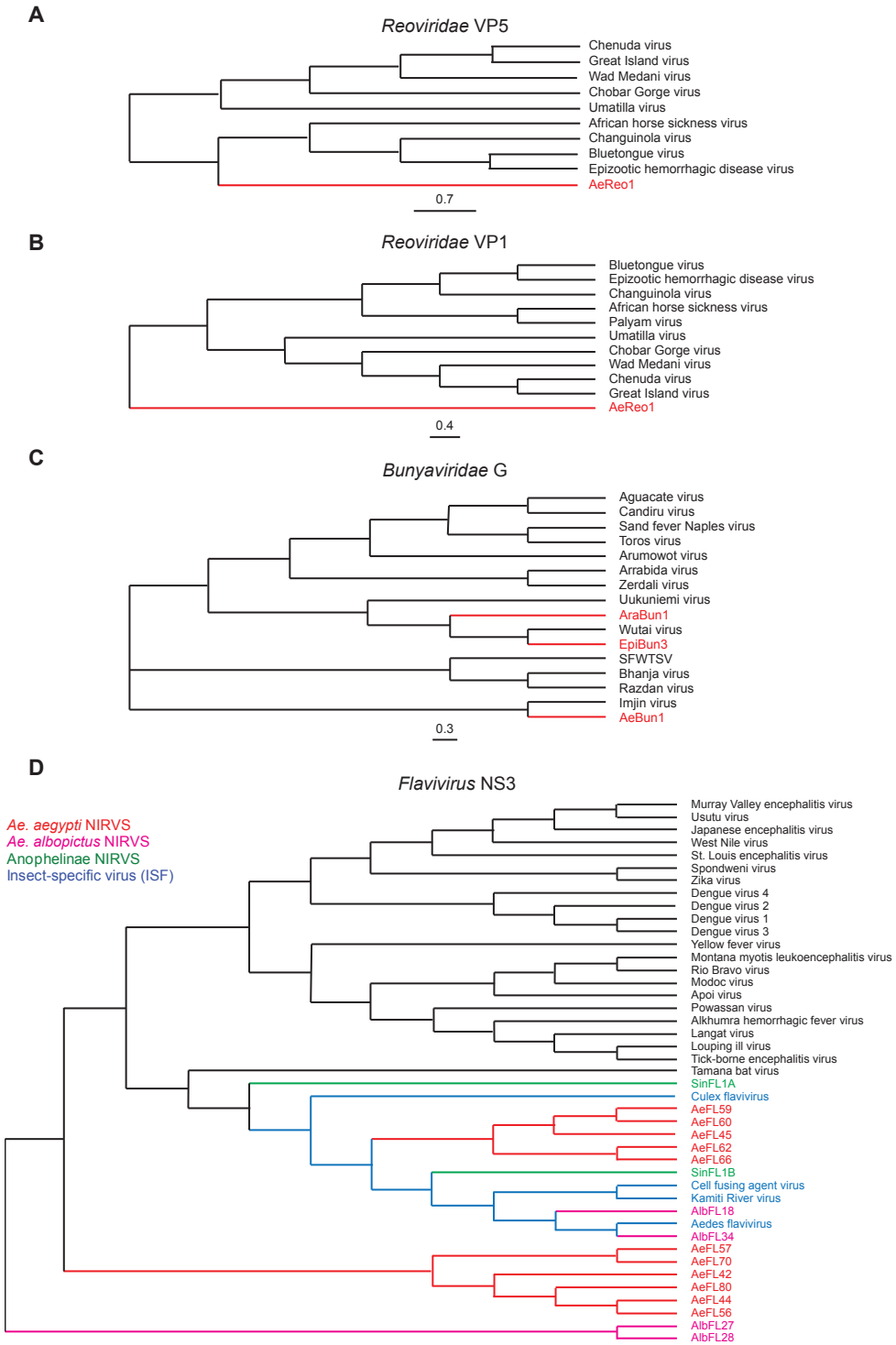


Figure 3. Variability of NIRVS within the *Ae. albopictus* Foshan strain. Bioinformatic analyses of the *Ae. albopictus* genome identified 4 NIRVS on scaffold JXUM01S011498: AlbFlavi32, AlbFlavi33, AlbFlavi34 and AlbFlavi36. No read coverage was seen for AlbFlavi32 and AlbFlavi33 in any of the 16 sequenced genomes. AlbFlavi36 had read coverage in 13 of the 16 tested mosquitoes, whereas AlbFlavi34 showed length variability.

In contrast, we observed numerous integrations from viruses of different genera within the *Rhabdoviridae* family and from viruses of the *Flavivirus* genus in multiple mosquito species, predominantly in *Ae. aegypti* and *Ae. albopictus* (Figure 2). *Rhabdoviridae*-like NIRVS (R-NIRVS) had similarities to genes encoding Nucleoprotein (N), Glycoprotein (G) and the RNA-dependent RNA polymerase (L), the relative abundance of which differed across mosquito species. We did not detect integrations corresponding to the matrix (M) or phosphoprotein (P) genes, consistent with observations in other arthropods (28). R-NIRVS from *Culicinae* and *Anophelinae* formed separate clades in phylogenetic trees, supporting the conclusion that independent integrations occurred in the two mosquito lineages (Figure 4E). *Flavivirus*-like NIRVS (F-NIRVS) with similarities to structural genes (envelope [E], membrane [prM] and capsid [C]) were less frequent


A
Reoviridae VP5
Chenuda virus
Great Island virus
Wad Medani virus
Chobar Gorge virus
Umatilla virus
African horse sickness virus
Changuinola virus
Bluetongue virus
Epizootic hemorrhagic disease virus
AeReo1
0.7
B
Reoviridae VP1
Bluetongue virus
Epizootic hemorrhagic disease virus
Changuinola virus
African horse sickness virus
Palyam virus
Umatilla virus
Chobar Gorge virus
Wad Medani virus
Chenuda virus
Great Island virus
AeReo1
0.4
C
Bunyaviridae G
Aguacate virus
Candiru virus
Sand fever Naples virus
Toros virus
Arumowot virus
Arrabida virus
Zerdali virus
Uukuniemi virus
AraBun1
Wutai virus
EpiBun3
SFWTSV
Bhanja virus
Razdan virus
Imjin virus
AeBun1
0.3
D
Flavivirus NS3
Ae. aegypti NIRVS
Ae. albopictus NIRVS
Anophelinae NIRVS
Insect-specific virus (ISF)
Murray Valley encephalitis virus
Usutu virus
Japanese encephalitis virus
West Nile virus
St. Louis encephalitis virus
Spondweni virus
Zika virus
Dengue virus 4
Dengue virus 2
Dengue virus 1
Dengue virus 3
Yellow fever virus
Montana myotis leukoencephalitis virus
Rio Bravo virus
Modoc virus
Apoi virus
Powassan virus
Alkhumra hemorrhagic fever virus
Langat virus
Louping ill virus
Tick-borne encephalitis virus
Tamana bat virus
SinFL1A
Culex flavivirus
AeFL59
AeFL60
AeFL45
AeFL62
AeFL66
SinFL1B
Cell fusing agent virus
Kamiti River virus
AlbFL18
Aedes flavivirus
AlbFL34
AeFL57
AeFL70
AeFL42
AeFL80
AeFL44
AeFL56
AlbFL27
AlbFL28

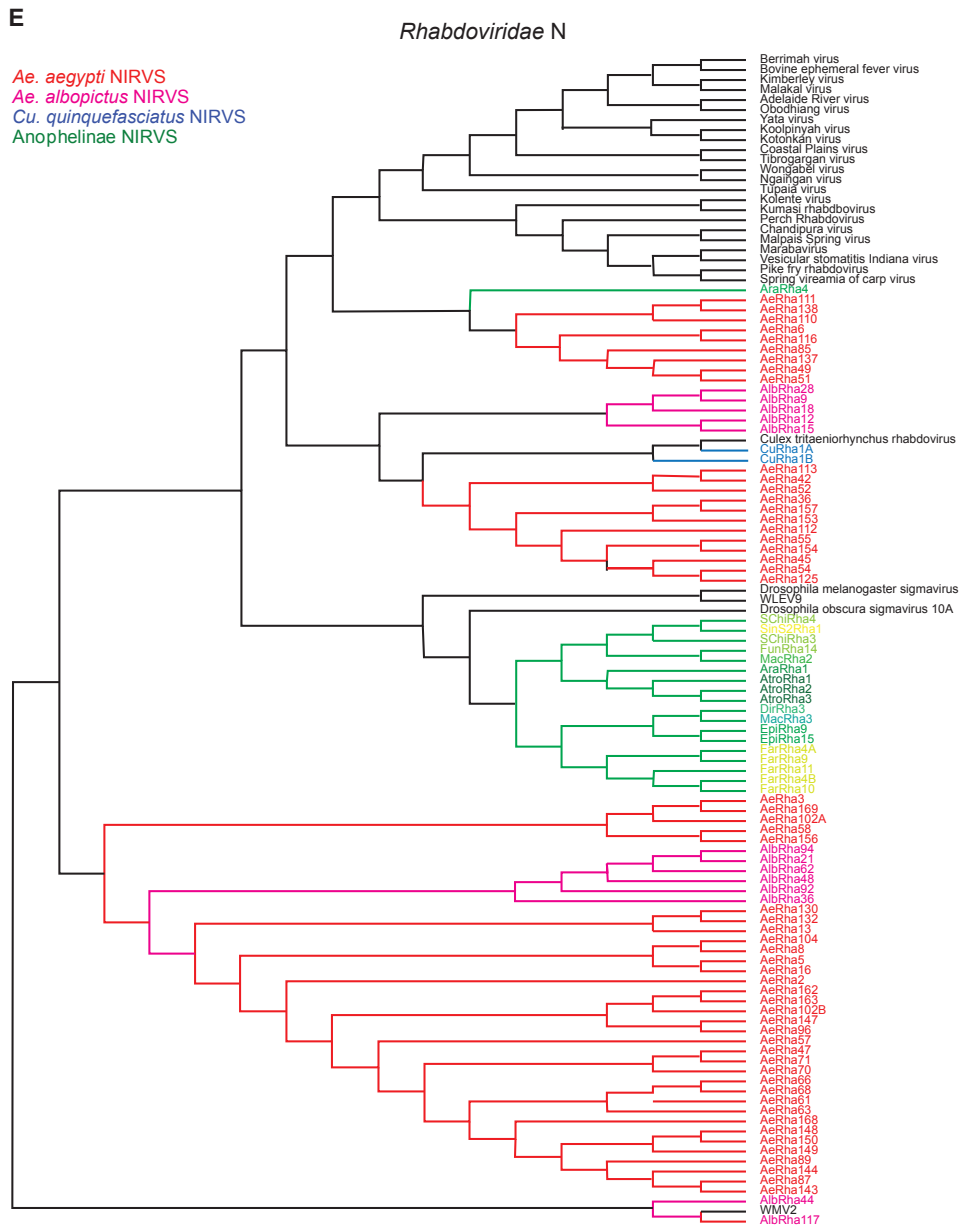


Figure 4. Phylogenetic analyses of *Reoviridae*, *Bunyaviridae*, Flavivirus, and *Rhabdoviridae*-like integrations. (A-E) Phylogenetic relationships of NIRVS with similarity to the *Reoviridae* VP5 (A), *Reoviridae* VP1 (B), *Bunyaviridae* G (C),Flavivirus NS3 (D), and *Rhabdoviridae* N (E) genes. The evolutionary history was inferred using the Maximum Likelihood method. The trees with the highest log likelihood are shown. Support for tree nodes was established after 1000 bootstraps.

than integrations corresponding to non-structural genes (Figure 2). Some R-NIRVS or F-NIRVS sequences within one mosquito genome were highly similar to each other (nucleotide identity > 90%, (Table S5), which suggest that these were duplicated in the genome after a single integration event. This interpretation is also supported by the genomic proximity of several of these NIRVS (Figure 5). Surprisingly, identical NIRVS in *Ae. aegypti* were found not only adjacent to one another, but also at locations that are physically unlinked (i.e. AeRha138, AeRha110 and AeRha111). Thus, we cannot determine whether these identical NIRVS represent recent independent integration events or arose from duplication or ectopic recombination after integration.

Generally, NIRVS were most similar to insect-specific viruses (ISVs), which replicate exclusively in arthropods, but are phylogenetically-related to arboviruses (10,42) (Figure 4D). However, we observed integrations that were most similar to arboviruses of the *Vesiculovirus* genus (*Rhabdoviridae*) in both *Ae. aegypti* and *Ae. albopictus* (Table S3).

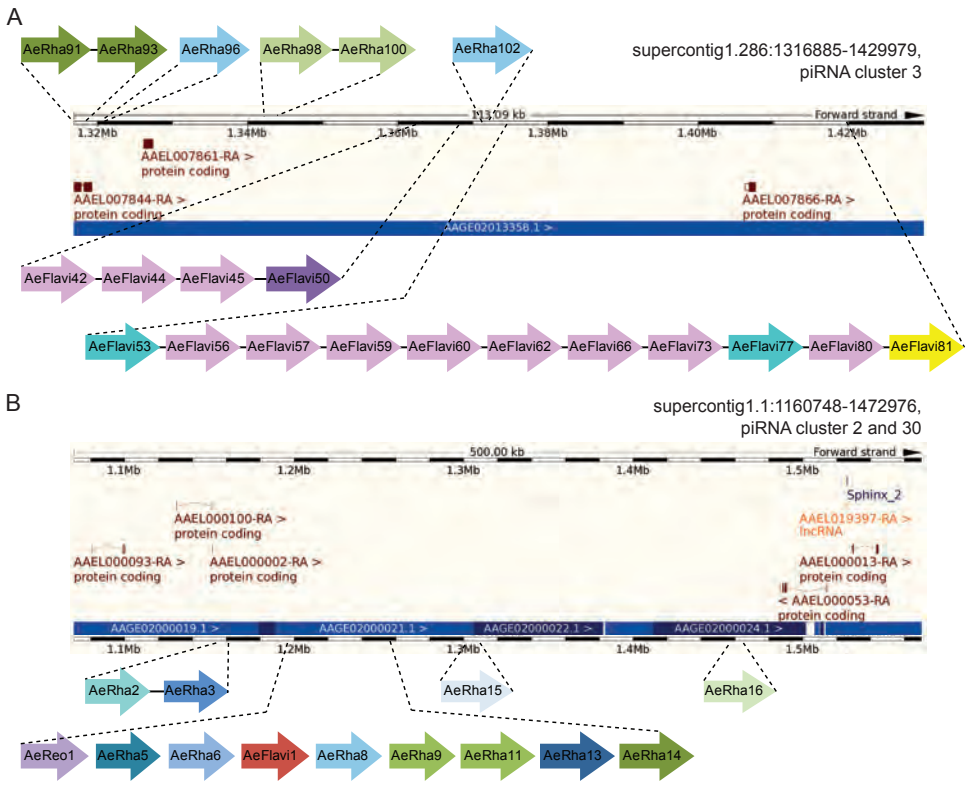


Figure 5. Enrichment of NIRVS in two regions of the *Ae. aegypti* genome. One fourth of the identified NIRVS in *Ae. aegypti* map to two genomic regions. **(A)** Region 1 (supercont1.286:1316885-1429979) includes piRNA cluster 3 (23) and is enriched in the LTR transposons LTR/Pao_Bel and LTR/Ty3_gypsy, which occupy 16.33 and 14.98% of the region, respectively. **(B)** Region 2 (supercont1.1:1160748-1472976) includes piRNA clusters 2 and 30 and is also enriched for LTR transposons. LTR/Ty3_gypsy occupancy in region 2 is 24.18%. NIRVS are color-coded based on their sequence identity (Table S5).

NIRVS produce piRNAs and map in piRNA clusters more frequently than expected by chance

To better understand the mechanisms of integration, we analyzed in greater detail the genomic context of NIRVS in *Ae. aegypti* and *Ae. albopictus*, the mosquitoes with the largest number of identified NIRVS. Previously, uncharacterized viral sequences were identified as piRNA producing loci in *Ae. aegypti* (23,43), and these observations prompted us to analyze whether NIRVS are enriched in piRNA clusters. Currently annotated piRNA clusters represent 1.24% and 0.61% of the *Ae. aegypti* and *Ae. albopictus* genomes, respectively (23,44). Remarkably, 44% and 12.5% of all NIRVS map to these genomic loci, and these frequencies are significantly higher than expected by chance (Table 2). Enrichment of NIRVS in piRNA clusters in *Ae. aegypti* was driven by two regions that harbored one fourth of all NIRVS loci (region1: scaffold 1.286: 1316885-1429979; region 2: scaffold 1.1: 1160748-1472976), which includes piRNA cluster 3 and piRNA clusters 2 and 30, respectively (23). In these two regions, NIRVS span partial ORFs with similarities to different *Rhabdovirus* and *Flavivirus* genes, with instances of duplications as well as unique viral integrations (Figure 5). NIRVS also were enriched in regions annotated as exons in *Ae. albopictus*, but not in *Ae. aegypti* (Table 2).

The presence of NIRVS in piRNA clusters prompted us to analyze the expression of NIRVS-derived small RNAs. Therefore we used deep-sequencing data from published resources and mapped small RNAs on NIRVS sequences after collapsing those elements that shared identical sequences (Table S5). Small RNAs in the size range of piRNAs (25-30 nucleotides), but not siRNAs (21-nucleotides) mapped to NIRVS in both *Ae. aegypti* and *Ae. albopictus*, independently of genomic localization and corresponding viral ORFs, (Figure 6A,B). Generally, piRNAs derived from individual NIRVS sequences are not highly abundant. Of all tested NIRVS, 43% (n=33) and 11% (n=6) had at least 10 piRNA reads per million genome-mapped reads in *Ae. aegypti* and *Ae. albopictus,* respectively. In *Ae. aegypti*, the highest piRNA counts were a few hundred reads per million genome-mapped reads. In *Ae. albopictus* the maximum piRNA counts per NIRVS were about 10 fold lower, suggesting that NIRVS piRNA are less efficiently produced or retained in this species. In both mosquito species, R-NIRVS showed higher coverage than F-NIRVS (Figure 6E). These piRNAs were biased for uridine at position 1 and primarily in antisense orientation to the predicted viral ORF, establishing the potential to target viral mRNA (Figure 6A-D). Yet, a 10A bias of sense piRNAs, particularly in *Ae. albopictus* indicates some NIRVS produce piRNAs through ping-pong amplification. Interestingly, ping-pong dependent secondary piRNAs seem to be almost exclusively (100% in *Ae. aegypti* and >99.5% in *Ae. albopictus*) derived from R-NIRVS (Figure 6E). The nature of this specific induction of secondary piRNAs biogenesis from rhabdoviral sequences is currently unknown.

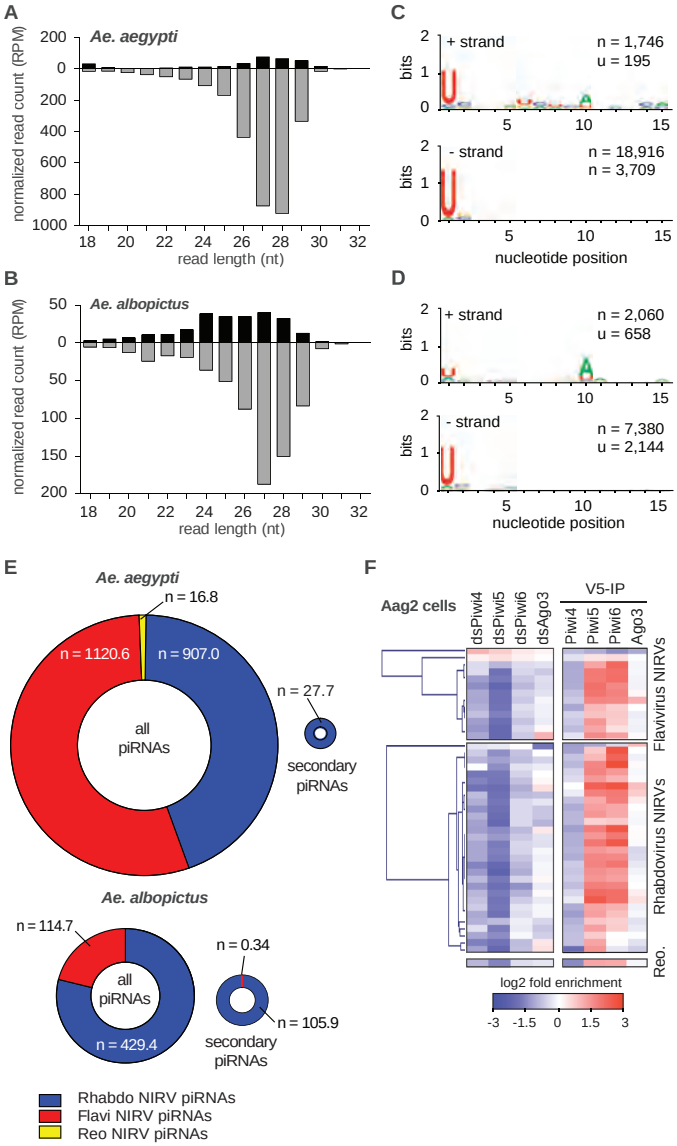


Figure 6. NIRVS produce 25-30 nt piRNAs, but not 21-nt siRNAs. (A,B) Size distribution of small RNAs from published resources mapping to NIRVS in the *Ae. aegypti* (A) and *Ae. albopictus* (B) genomes. NIRVS-derived piRNAs are biased for sequences that are antisense to viral mRNAs, suggesting potential to target viral RNA. NIRVS-derived piRNAs are biased for uridine at position 1, in both *Ae. aegypti* **(C)** and *Ae. albopictus* **(D)**. **(E)** Number of all piRNAs and secondary piRNAs expressed in *Ae. aegypti* or *Ae albopictus*. Ring charts are drawn to scale and numbers indicate the normalized piRNA counts of F-NIRVS (red), R-NIRVS (blue), and NIRVS from *Reovirus* that have been found only in *Ae. aegypti* (yellow). **(F)** Left panel, heat map of the relative abundance of NIRVS-derived small RNAs in Aag2 cells in which PIWI expression was silenced using RNAi (dsPiwi4-6, and dsAgo3), compared to control dsRNA treatment. Right panel, heatmap of small RNA enrichment in immunoprecipitations (IP) of the indicated PIWI proteins over a control GFP IP.

We next analyzed the dependency on and association with PIWI proteins of NIRVS-derived small RNAs in Aag2 cells (45) and found that small RNA expression was reduced by knockdown of Piwi5 and, to a lesser extent, Piwi4 and Piwi6 (Figure 6F), with only few exceptions. Consistent with this finding, NIRVS-derived small RNAs were most enriched in immunoprecipitations (IP) of Piwi5 and Piwi6 (Figure 6F). Together, these data indicate that NIRVS produce piRNAs, the majority of which have the characteristics of primary piRNAs. Yet, secondary piRNA biogenesis as indicated by a 10A bias and association with Ago3, seems to occur specifically from R-NIRVS.

Table 2. Clustering of viral integrations (NIRVS) in piRNA loci of the *Ae. aegypti* (A) and *Ae. albopictus* (B) genomes. The probability (*P*) of observing k NIRVS loci in piRNA clusters, coding genes and intergenic regions. *P* was estimated using cumulative binomial distribution; a value of $P < 0.05$ indicates a statistically significant enrichment of NIRVS in the corresponding genomic region.

Host	Genomic region	Length (bp)	% genome	k integrations*	*P*
Ae. aegypti	piRNA cluster	17,000,000	1.24	54	$< 10^{-10}$
	Coding genes	286,538,182	20.82	24	0.66
	Intergenic regions	1,072,461,818	77.94	44	1
Ae. albopictus	piRNA cluster	1,926,670	0.61	9	$< 10^{-10}$
	Coding genes	163,407,667	8.26	14	$2.08\ 10^{-3}$
	Intergenic regions	1,803,592,333	91.14	49	1

*Six integrations in the *Ae. aegypti* genome were in exons of genes within piRNA clusters; in this analyses they were attributed to piRNA clusters. Statistical significance did not change when these integrations were assigned to coding genes (*P* changed from 0.66 to 0.180).

NIRVS and transposable elements

piRNA clusters in *D. melanogaster* are enriched for remnants of TE sequences, and it is likely that vDNA is produced by the reverse transcriptase activity of TEs (36,37). Moreover, NIRVS-derived piRNAs resembled the characteristics of TE-derived piRNAs in their antisense 1U bias and enrichment in Piwi5 and Piwi6 protein complexes. We analyzed the transposon landscape of NIRVS loci by systematically identifying all annotated TEs in the 5 and 10 kb genomic regions flanking each side of the NIRVS integration. We observed that NIRVS were predominantly associated with long terminal repeat (LTR) retrotransposons. Within LTR-retrotransposons, we observed enrichment of members of the Ty3_gypsy families (Table 3). Such enrichments were even more pronounced in the two regions in *Ae. aegypti* where 40% of NIRVS reside (Figure 5). While LTR retrotransposon occupancy was 12.34% across the entire *Ae. aegypti* genome, it reached 23.60-25.88%, 31.35%, and 30.55% in regions flanking all NIRVS-loci, region 1, and region 2, respectively. More strikingly, while the Ty3_gypsy families of LTR retrotransposon occupancy was 2.58% across the entire *Ae. aegypti* genome, it reached 14.7-17.5%, 14.98% and 24.18% in regions flanking all NIRVS-loci, region 1, and region 2, respectively (Table 3). Nine full-length TEs were found flanking NIRVS-loci, seven of which are Ty3_gypsy retrotransposons. For example, 3 copies and 1 copy of the full-length Ty_gypsy_Ele58 (TF000321) were found in regions 1 and 2, respectively. Moreover, one viral integration in *Ae. aegypti* (i.e. AeBunya1) was found embedded within a full-length TE of the Pao-Bel family.

NIRVS transcriptional activity

All NIRVS encompassed partial viral ORFs, with the exception of AlbFlavi34. AlbFlavi34 corresponds to a portion of the first *Flavivirus*-like sequence characterized in mosquitoes

and includes a complete ORF for NS3 (24). Two alleles of different lengths were seen for AlbFlavi34 in the 16 sequenced *Ae. albopictus* genomes (Figure 3). The short allele, which interrupts the NS3 ORF, had a frequency of 53% (Supplemental file 1). Based on recent experimental data showing that NIRVS are transcriptionally active even if they do not encode a complete ORF (24,28,32,46) we analyzed NIRVS expression using published RNA-seq data from poly(A) selection protocols. Expression levels were < 5 reads per kilobase per million mapped reads (RPKM) for > 92% of all tested NIRVS, including NIRVS that produce piRNAs (Table S6). Similar to small RNA profiles, expression levels of R-NIRVS were higher than those of F-NIRVS (Table S6).

Despite RNA-seq data showing limited transcriptional activity for AlbFlavi34 (RPKM values ranging from 0.009 to 0.013), we analyzed its expression in different developmental stages by RT-qPCR using primers that amplify both the short and long alleles. Cycle threshold (Ct) values ranged from 27 (found in pupae) to 39.34 (detected

Table 3. NIRVS and transposable elements (TEs). Analyses of TE enrichment througout the *Ae. aegypti* genome (AaegL3), in regions harboring NIRVS (NIRVS), in region1 and in region 2, respectively.

	TE Occupancy (%)			
TE group[1]	AaegL3[2]	NIRVS[3]	Region 1[4]	Region 2[5]
LTR retrotransposons	12.34	23.06 (25.88)	31.35	30.56
LTR/Pao_Bel	4.42	6.9 (6.49)	16.33	4.15
LTR/Ty1_copia	5.34	1.46 (1.90)	0.04	2.22
LTR/Ty3_gypsy	2.58	14.7 (17.50)	14.98	24.18
non-LTR retrotransposons	12.81	3.91 (5.54)	0	3.43
SINEs	1.14	0.16 (0.22)	0	0
DNA transposons	6.96	3.29 (3.49)	3.38	2.72
MITEs	12.81	8.03 (7.34)	0.26	2.34
Helitrons	1.2	2.01 (2.32)	0	2.04
Penelope	0.42	0.2 (0.28)	0.26	0.8

[1] For consistency with previous publications and for clear classification, only TEs annotated in TEfam are used. We used TE occupancy (number of bases in the genomic sequence that match TEs) as an indication for possible TE enrichment. TE copy number was not used to prevent that TEs that are broken into multiple fragments are counted multiple times.

[2] The genome assembly described in Nene *et al.* (2007) is slightly different from AaegL3 (*Aedes-aegypti*-Liverpool_SCAFFOLDS_AaegL3.fa), which is used in this analysis. For better comparison with viral integration sites, a new RepeatMasker analysis was performed using the AaegL3 assembly under the same default parameters.

[3] In addition to the integration site, 5 kb or 10 kb (in brackets) sequences flanking each side of the NIRVS were retrieved for the analysis. Because the NIRVS sequences are also included in the analyses, these results may be an under-estimate of the actual TE occupancy.

[4] TE analysis of viral integration sites (plus 5kb flanking each side) in Supercontig 1.286 between positions 1316885 bp and 1429979 bp.

[5] TE analysis of viral integration sites (plus 5kb flanking each side) in Supercontig 1.1 between positions 1160748 bp and 1472976 bp.

in ovaries of blood fed-females) and 60% of the samples having Ct> 30, confirming low AlbFlavi34 expression. AlbFlavi34 expression was highest in the pupae and adult males in comparison to expression in larvae (Figure S1). These data support the conclusions that steady-state RNA levels of most NIRVS are rather low or even undetectable. Yet, the production of piRNAs indicates that they must be transcriptionally active. Whether their precursor transcripts are non-polyadenylated or rapidly processed into piRNAs remains to be established.

DISCUSSION

The genomes of mosquitoes and several eukaryotic species carry integrations from non-retroviral RNA viruses, including arboviruses. To shed light on the widespread and biological significance of this phenomenon, we analyzed the presence, distribution and transcriptional activity of integrations from 424 non-retroviral RNA viruses, and one DNA arbovirus, in 22 mosquito genomes, in the context of both their phylogeny and mosquito vector competence. We showed that the arboviral vector species *Ae. aegypti* and *Ae. albopictus* have ten-fold more integrations than all other tested mosquitoes. Moreover, we found that viral integrations produce piRNAs and occur predominantly in piRNA clusters. Our results support the conclusion that the abundance of viral integrations is not dependent on viral exposure, but seems to correlate with the TE landscape and piRNA pathway of the mosquito.

NIRVS viral origin

Across all 425 viral species tested, viral integrations had similarities primarily to ISVs of the *Bunyaviridae*, *Reoviridae* and, predominantly, *Rhabdoviridae* and *Flaviviridae* families. Notably, although the *Togaviridae* family contains mosquito-borne members as well as insect specific viruses, we identified no integrations from viruses of in this family. Further studies are required to clarify whether this result is due to a sampling bias or to the different evolutionary history of *Alphavirus*-like versus *Flavivirus*-like viruses (38). For instance, Eilat virus and the Taï Forest alphavirus are the only insect-specific alphaviruses (family *Togaviridae*) identified and a large screen suggests that mosquito-specific viruses may not be abundant among alphaviruses (47) unlike the *Bunyaviridae*, *Reoviridae*, *Rhabdoviridae* and *Flaviviridae* families in which many ISVs have been identified (42). An alternative explanation may be based on the interactions of these viruses with the piRNA machinery. For example, while both alphaviruses and flaviviruses produce vpiRNAs in *Aedes*, the distribution of piRNAs on the viral genomes are not comparable between these genera, suggesting that piRNA biogenesis might be different (21). Both alphaviruses (Sindbis and CHIK viruses) and flaviviruses (DENV, WNV) have been shown to produce episomal vDNA forms that locate to the nucleus after infection of mosquitoes (20,36,37). These vDNA forms do not arise uniformly from

the whole viral genome and their profile may be different between alphaviruses and flaviviruses (37). If these episomal vDNA are integrated into the genome, a different vDNA profile will result in a different NIRVS landscape.

ISVs of the families *Bunyaviridae*, *Flaviviridae* and *Rhabdoviridae* families are ancient and diversified within their hosts, and they seem to be maintained in mosquitoes through transovarial transmission (10,42). Additionally, mounting phylogenetic evidence implicate ISVs as precursors of arboviruses (48), for which vertical transmission occurs at a lower frequency than horizontal transmission through a vertebrate host (49). Vertical transmission provides access to the mosquito germ-line, a mechanism through which NIRVS could be maintained within vector populations. Thus, the observed higher incidence of NIRVS from ISVs than arbovirus may be linked to differences in the frequency of their transovarial transmission.

NIRVS from *Bunyaviridae* and *Rhabdoviridae* have been identified in insects other than mosquitoes, including different *Drosophila* species and the tick Ixodes scapularis (26-28). In contrast, NIRVS from Flaviviruses have been found only in mosquitoes, predominantly in *Ae. aegypti* and *Ae. albopictus* (2,26,32) . Interestingly, vertebrates that may be part of the arbovirus transmission cycle do not have integrations from arboviruses, but a low number (<10) of integrations from *Bornaviruses* and/or *Filoviruses* have been identified in humans, squirrel, microbat, opossum, lemur, wallaby and medaka (25-27). Finally, several *Anophelinae* mosquitoes analyzed here were sampled in the same geographic area as *Ae. albopictus*, but showed 10 times fewer NIRVS than *Ae. albopictus* and *Ae. aegypti.* Overall, these data indicate that viral exposure is not a determinant of NIRVS, but that virus-host lineage-specific interactions play a crucial role in how their genomes co-evolve. Additionally, our comparative analysis shows that *Aedes* mosquitoes acquire and retain fragments of infecting non-retroviral RNA viruses primarily the *Flaviviridae* and *Rhabdoviridae* families, more frequently than other tested arthropods and vertebrates. A deeper understanding of the evolution of viruses within these large and diverse families, especially their recently characterized ISVs, along with insights into the variability of the genomes of mosquito populations are warranted to elucidate the dynamic species-specific interactions between RNA viruses and *Aedes* mosquitoes.

NIRVS genomic context

NIRVS are significantly enriched in piRNA clusters in both *Ae. aegypti* and *Ae. albopictus*, which could be the result of positive selection favoring the retention of those NIRVS that integrated by chance in these genomic loci (50). However, we also observed NIRVS in intergenic and coding sequences and found that NIRVS expressed piRNAs independently of their genomic localization. These observations suggest that additional piRNA clusters exist (23,44) or that other features in these NIRVS loci prime piRNAs production. For example, a piRNA trigger sequence (PTS) was recently

found to drive piRNA production from a major piRNA cluster (named *Flamenco*) in *Drosophila* (51). We analyzed the mosquito genome sequences, but we did not find PTS orthologous sequences in either *Ae. aegypti* nor *Ae. albopictus*. It remains to be established whether other PTS sequences exist that may explain piRNA production from non-cluster associated NIRVS.

Analyses of the integration sites showed that NIRVS are primarily associated with LTR transposons of the Gypsy and Pao families, which are the most abundant TE families in both *Ae. aegypti* and *Ae. albopictus* genomes (2). Additionally, full-length TEs, primarily Ty3_gypsy retrotransposons, were found to flank NIRVS-loci. This organization is compatible with recent experimental data showing that vDNA forms are produced by retrotransposon-derived reverse transcriptase, likely by template switching (20,37). This arrangement also is favorable for ectopic recombination, a mechanism proposed for both NIRVS biogenesis and piRNA cluster evolution (52). Ectopic recombination would be a more parsimonious explanation than independent integrations from the same viral source for our finding of several not physically-linked, but identical *Ae. aegypti* NIRVS. Despite many remaining uncertainties due to the highly repetitive and complex structure of the regions in which NIRVS map, these data confirm a functional link among NIRVS, TEs, and the piRNA pathway.

NIRVS and mosquito immunity

Our data indicate that in *Ae. aegypti* and *Ae. albopictus* NIRVS do not encode proteins that interfere in trans with viral products as was observed in bornavirus-derived NIRVS in vertebrates (53). Rather our data suggest that NIRVS may be part of a piRNA-based antiviral response. Only one of the characterized NIRVS had a complete viral ORF, which showed two alleles of different length within the 16 individuals of the Foshan strain that we sequenced. The short variant interrupted the NS3 ORF. We cannot exclude that this is due to lack of purifying selective pressure as the *Ae. albopictus* Foshan strain has been reared under standard laboratory conditions without infection challenges for more than 30 years (2). However, the enrichment of NIRVS within piRNA clusters and their small RNA profile suggest that their transcriptional activity is geared to produce piRNA precursors. Our results show a basal expression of NIRVS-derived primary piRNAs that are antisense to viral mRNA. These piRNAs could block novel infections with cognate viruses or they could interact with RNAi mechanisms to contain replication of incoming viruses at a level that does not become detrimental to mosquitoes. Albeit leading to opposite effects on vector competence, both mechanisms display functional similarities to the CRISPR-Cas system of prokaryotic adaptive immunity. Even if further studies are essential to clarify the effect of NIRVS-derived piRNAs on mosquitommunity, our study clearly demonstrates that *Ae. aegypti* and *Ae. albopictus* have a high number of NIRVS in their genome, which confers heritable immune signals.

The higher number of NIRVS in *Aedeine* than in *Anophelinae* mosquitoes correlates with competence for a larger number of arboviruses of *Aedeine* mosquitoes. In this regard, *Cx. quinquefasciatus* shows an interesting intermediate phenotype because it is phylogenetically closer to *Aedeine* mosquitoes, but vectors a smaller range of arboviruses than *Aedeine* mosquitoes and, like *Anophelinae*, it can vector more protozoans and nematodes than *Aedeine* (54). Additionally, *Cx. quinquefasciatus* has a number of NIRVS and TE load comparable to *Anophelinae*, but an expanded gene family like *Ae. aegypti* (2,55-56).

CONCLUSIONS

NIRVS are regarded as viral fossils, occurring as occasional events during to the long co-evolutionary history of viruses and their hosts (33,35). The high abundance and diversity of NIRVS in the genomes of *Ae. aegypti* and *Ae. albopictus* and the observation that NIRVS produce piRNAs and reside in piRNA clusters support the intriguing hypothesis that the formation and maintenance of NIRVS are coupled with the evolution of the PIWI pathway in these two species. This may have led to functional specialization of the expanded PIWI gene family, PIWI expression in the soma, and a role for the piRNA pathway in antiviral immunity (21,45). This hypothesis is compatible with two scenarios. First, NIRVS formation is an occasional event, which occurs more frequently in *Aedeine* than *Culicinae* and *Anopheline* because of the higher abundance of retrotransposons in the genome of *Aedeine* mosquitoes (2). NIRVS that have integrated by chance into piRNA clusters produce transcripts that are shuttled into the piRNA pathway. PIWI proteins loaded with viral sequences may target incoming viruses, possibly conferring selective advantage. Thus, an occasional event linked to a particular TE landscape may be the trigger for the functional specialization of PIWI proteins. This scenario remains compatible with the possibility that NIRVS outside of piRNA loci encode protein products that compete *in trans* with virus replication, thereby affecting vector competence (57). Second, it has been hypothesized that PIWI proteins actively interact with incoming viruses and that they are loaded with episomal vDNAs and integrate them into piRNA clusters (58). Under this scenario, the selective pressure favoring PIWI protein specialization would come primarily from viruses. Taken together our data show that the interaction between viruses and mosquitoes is a more dynamic process than previously thought and that this interplay can lead to heritable changes in mosquito genomes.

ACKNOWLEDGEMENTS

We thank Professor Anthony A. James (University of California at Irvine) for helpful comments on the manuscript. We thank Patrizia Chiari for insectary work. The project was supported by a European Research Council Consolidator Grant (ERC-CoG) under the European Union's Seventh Framework Programme (grant number ERC CoG 615680) to R.P.v.R and an ERC-CoG under the European Union's Horizon 2020 Programme (Grant Number ERC CoG 682394) to M.B. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

REFERENCES

1. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. Nature. 2008;451(7181):990-3.
2. Chen XG, Jiang X, Gu J, Xu M, Wu Y, Deng Y, et al. Genome sequence of the Asian Tiger mosquito, Aedes albopictus, reveals insights into its biology, genetics, and evolution. Proc Natl Acad Sci U S A. 2015;112(44):E5907-15.
3. Weaver SC, Reisen WK. Present and future arboviral threats. Antiviral Res. 2010;85(2):328-45.
4. Powell JR, Tabachnick WJ. History of domestication and spread of Aedes aegypti--a review. Mem Inst Oswaldo Cruz. 2013;108 Suppl 1:11-7.
5. Paupy C, Delatte H, Bagny L, Corbel V, Fontenille D. Aedes albopictus, an arbovirus vector: from the darkness to the light. Microbes Infect. 2009;11(14-15):1177-85.
6. Bonizzoni M, Gasperi G, Chen X, James AA. The invasive mosquito species Aedes albopictus: current knowledge and future perspectives. Trends Parasitol. 2013;29(9):460-8.
7. Farajollahi A, Fonseca DM, Kramer LD, Marm Kilpatrick A. "Bird biting" mosquitoes and human disease: a review of the role of Culex pipiens complex mosquitoes in epidemiology. Infect Genet Evol. 2011;11(7):1577-85.
8. Linthicum KJ, Britch SC, Anyamba A. Rift Valley Fever: An Emerging Mosquito-Borne Disease. Annu Rev Entomol. 2016;61:395-415.
9. Carissimo G, Eiglmeier K, Reveillaud J, Holm I, Diallo M, Diallo D, et al. Identification and Characterization of Two Novel RNA Viruses from Anopheles gambiae Species Complex Mosquitoes. PLoS One. 2016;11(5):e0153881.
10. Bolling BG, Weaver SC, Tesh RB, Vasilakis N. Insect-Specific Virus Discovery: Significance for the Arbovirus Community. Viruses. 2015;7(9):4911-28.
11. Fauver JR, Grubaugh ND, Krajacich BJ, Weger-Lucarelli J, Lakin SM, Fakoli LS, 3rd, et al. West African Anopheles gambiae mosquitoes harbor a taxonomically diverse virome including new insect-specific flaviviruses, mononegaviruses, and totiviruses. Virology. 2016;498:288-99.
12. Tabachnick WJ. Nature, nurture and evolution of intra-species variation in mosquito arbovirus transmission competence. Int J Environ Res Public Health. 2013;10(1):249-77.
13. Stapleford KA, Coffey LL, Lay S, Borderia AV, Duong V, Isakov O, et al. Emergence and transmission of arbovirus evolutionary intermediates with epidemic potential. Cell Host Microbe. 2014;15(6):706-16.
14. Dennison NJ, Jupatanakul N, Dimopoulos G. The mosquito microbiota influences vector competence for human pathogens. Curr Opin Insect Sci. 2014;3:6-13.
15. Cheng G, Liu Y, Wang P, Xiao X. Mosquito Defense Strategies against Viral Infection. Trends Parasitol. 2016;32(3):177-86.
16. Chouin-Carneiro T, Vega-Rua A, Vazeille M, Yebakima A, Girod R, Goindin D, et al. Differential Susceptibilities of Aedes aegypti and Aedes albopictus from the Americas to Zika Virus. PLoS Negl Trop Dis. 2016;10(3):e0004543.

17. Weger-Lucarelli J, Ruckert C, Chotiwan N, Nguyen C, Garcia Luna SM, Fauver JR, et al. Vector Competence of American Mosquitoes for Three Strains of Zika Virus. PLoS Negl Trop Dis. 2016;10(10):e0005101.
18. Bronkhorst AW, van Rij RP. The long and short of antiviral defense: small RNA-based immunity in insects. Curr Opin Virol. 2014;7:19-28.
19. Olson KE, Blair CD. Arbovirus-mosquito interactions: RNAi pathway. Curr Opin Virol. 2015;15:119-26.
20. Goic B, Stapleford KA, Frangeul L, Doucet AJ, Gausson V, Blanc H, et al. Virus-derived DNA drives mosquito vector tolerance to arboviral infection. Nat Commun. 2016;7:12410.
21. Miesen P, Joosten J, van Rij RP. PIWIs Go Viral: Arbovirus-Derived piRNAs in Vector Mosquitoes. PLoS Pathog. 2016;12(12):e1006017.
22. Handler D, Meixner K, Pizka M, Lauss K, Schmied C, Gruber FS, et al. The genetic makeup of the Drosophila piRNA pathway. Mol Cell. 2013;50(5):762-77.
23. Arensburger P, Hice RH, Wright JA, Craig NL, Atkinson PW. The mosquito Aedes aegypti has a large genome size and high transposable element load but contains a low proportion of transposon-specific piRNAs. BMC Genomics. 2011;12:606.
24. Crochu S, Cook S, Attoui H, Charrel RN, De Chesse R, Belhouchet M, et al. Sequences of flavivirus-related RNA viruses persist in DNA form integrated in the genome of Aedes spp. mosquitoes. J Gen Virol. 2004;85(Pt 7):1971-80.
25. Belyi VA, Levine AJ, Skalka AM. Unexpected inheritance: multiple integrations of ancient bornavirus and ebolavirus/marburgvirus sequences in vertebrate genomes. PLoS Pathog. 2010;6(7):e1001030.
26. Katzourakis A, Gifford RJ. Endogenous viral elements in animal genomes. PLoS Genet. 2010;6(11):e1001191.
27. Horie M, Honda T, Suzuki Y, Kobayashi Y, Daito T, Oshida T, et al. Endogenous non-retroviral RNA virus elements in mammalian genomes. Nature. 2010;463(7277):84-7.
28. Fort P, Albertini A, Van-Hua A, Berthomieu A, Roche S, Delsuc F, et al. Fossil rhabdoviral sequences integrated into arthropod genomes: ontogeny, evolution, and potential functionality. Mol Biol Evol. 2012;29(1):381-90.
29. Ballinger MJ, Bruenn JA, Taylor DJ. Phylogeny, integration and expression of sigma virus-like genes in Drosophila. Mol Phylogenet Evol. 2012;65(1):251-8.
30. Ballinger MJ, Bruenn JA, Hay J, Czechowski D, Taylor DJ. Discovery and evolution of bunyavirids in arctic phantom midges and ancient bunyavirid-like sequences in insect genomes. J Virol. 2014;88(16):8783-94.
31. Geisler C, Jarvis DL. Rhabdovirus-like endogenous viral elements in the genome of Spodoptera frugiperda insect cells are actively transcribed: Implications for adventitious virus detection. Biologicals. 2016;44(4):219-25.
32. Lequime S, Lambrechts L. Discovery of flavivirus-derived endogenous viral elements in Anopheles mosquito genomes supports the existence of Anopheles-associated insect-specific flaviviruses. Virus Evol. 2017;3(1):vew035.
33. Aiewsakun P, Katzourakis A. Endogenous viruses: Connecting recent and ancient viral evolution. Virology. 2015;479-480:26-37.
34. Tromas N, Zwart MP, Forment J, Elena SF. Shrinkage of genome size in a plant RNA virus upon transfer of an essential viral gene into the host genome. Genome Biol Evol. 2014;6(3):538-50.
35. Aswad A, Katzourakis A. Paleovirology and virally derived immunity. Trends Ecol Evol. 2012;27(11):627-36.
36. Goic B, Vodovar N, Mondotte JA, Monot C, Frangeul L, Blanc H, et al. RNA-mediated interference and reverse transcription control the persistence of RNA viruses in the insect model Drosophila. Nat Immunol. 2013;14(4):396-403.
37. Nag DK, Brecher M, Kramer LD. DNA forms of arboviral RNA genomes are generated following infection in mosquito cell cultures. Virology. 2016;498:164-71.

38. Koonin EV, Dolja VV, Krupovic M. Origins and evolution of viruses of eukaryotes: The ultimate modularity. Virology. 2015;479-480:2-25.

39. Attoui H, Mohd Jaafar F, Belhouchet M, Tao S, Chen B, Liang G, et al. Liao ning virus, a new Chinese seadornavirus that replicates in transformed and embryonic mammalian cells. J Gen Virol. 2006;87(Pt 1):199-208.

40. Song JW, Kang HJ, Gu SH, Moon SS, Bennett SN, Song KJ, et al. Characterization of Imjin virus, a newly isolated hantavirus from the Ussuri white-toothed shrew (Crocidura lasiura). J Virol. 2009;83(12):6184-91.

41. Li CX, Shi M, Tian JH, Lin XD, Kang YJ, Chen LJ, et al. Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. Elife. 2015;4.

42. Vasilakis N, Tesh RB. Insect-specific viruses and their potential impact on arbovirus transmission. Curr Opin Virol. 2015;15:69-74.

43. Girardi E, Miesen P, Pennings B, Frangeul L, Saleh MC, van Rij RP. Histone-derived piRNA biogenesis depends on the ping-pong partners Piwi5 and Ago3 in Aedes aegypti. Nucleic Acids Res. 2017;45(8):4881-92.

44. Liu P, Dong Y, Gu J, Puthiyakunnon S, Wu Y, Chen XG. Developmental piRNA profiles of the invasive vector mosquito Aedes albopictus. Parasit Vectors. 2016;9(1):524.

45. Miesen P, Girardi E, van Rij RP. Distinct sets of PIWI proteins produce arbovirus and transposon-derived piRNAs in Aedes aegypti mosquito cells. Nucleic Acids Res. 2015;43(13):6545-56.

46. Maori E, Tanne E, Sela I. Reciprocal sequence exchange between non-retro viruses and hosts leading to the appearance of new host phenotypes. Virology. 2007;362(2):342-9.

47. Hermanns K, Zirkel F, Kopp A, Marklewitz M, Rwego IB, Estrada A, et al. Discovery of a novel alphavirus related to Eilat virus. J Gen Virol. 2017;98(1):43-9.

48. Marklewitz M, Zirkel F, Kurth A, Drosten C, Junglen S. Evolutionary and phenotypic analysis of live virus isolates suggests arthropod origin of a pathogenic RNA virus family. Proc Natl Acad Sci U S A. 2015;112(24):7536-41.

49. Grunnill M, Boots M. How Important is Vertical Transmission of Dengue Viruses by Mosquitoes (Diptera: Culicidae)? J Med Entomol. 2016;53(1):1-19.

50. Parrish NF, Fujino K, Shiromoto Y, Iwasaki YW, Ha H, Xing J, et al. piRNAs derived from ancient viral processed pseudogenes as transgenerational sequence-specific immune memory in mammals. RNA. 2015;21(10):1691-703.

51. Homolka D, Pandey RR, Goriaux C, Brasset E, Vaury C, Sachidanandam R, et al. PIWI Slicing and RNA Elements in Precursors Instruct Directional Primary piRNA Biogenesis. Cell Rep. 2015;12(3):418-28.

52. Assis R, Kondrashov AS. Rapid repetitive element-mediated expansion of piRNA clusters in mammalian evolution. Proc Natl Acad Sci U S A. 2009;106(17):7079-82.

53. Fujino K, Horie M, Honda T, Merriman DK, Tomonaga K. Inhibition of Borna disease virus replication by an endogenous bornavirus-like element in the ground squirrel genome. Proc Natl Acad Sci U S A. 2014;111(36):13175-80.

54. Bhattacharya S, Basu P. The Southern house mosquito, Culex quinquefasciatus: profile of a smart vector. J Entomol Zool Studies 2016;4(2):73-81

55. Bartholomay LC, Waterhouse RM, Mayhew GF, Campbell CL, Michel K, Zou Z, et al. Pathogenomics of Culex quinquefasciatus and meta-analysis of infection responses to diverse pathogens. Science. 2010;330(6000):88-90.

56. Lewis SH, Salmela H, Obbard DJ. Duplication and Diversification of Dipteran Argonaute Genes, and the Evolutionary Divergence of Piwi and Aubergine. Genome Biol Evol. 2016;8(3):507-18.

57. Honda T, Tomonaga K. Endogenous non-retroviral RNA virus elements evidence a novel type of antiviral immunity. Mob Genet Elements. 2016;6(3):e1165785.

58. Blair CD, Olson KE. Targeting Dengue virus replication in Mosquitoes. In: Aksoy S, Wikel S, Dimopoulos G, editors. Arthropod Vector: The Controller of Transmission Vol. 1. Elsevier. Pages 201 - 217

59. Neafsey DE, Waterhouse RM, Abai MR, Aganezov SS, Alekseyev MA, Allen JE, et al. Mosquito genomics. Highly evolvable malaria vectors: the genomes of 16 Anopheles mosquitoes. Science. 2015;347(6217):1258522.

60. Falda M, Toppo S, Pescarolo A, Lavezzo E, Di Camillo B, Facchinetti A, et al. Argot2: a large scale function prediction tool relying on semantic similarity of weighted Gene Ontology terms. BMC Bioinformatics. 2012;13 Suppl 4:S14.

61. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26(5):589-95.

62. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841-2.

63. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24-6.

64. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 2013;30(12):2725-9.

65. Hu W, Criscione F, Liang S, Tu Z. MicroRNAs of two medically important mosquito species: Aedes aegypti and Anopheles stephensi. Insect Mol Biol. 2015;24(2):240-52.

66. Gu J, Hu W, Wu J, Zheng P, Chen M, James AA, et al. miRNA genes of an invasive vector mosquito, Aedes albopictus. PLoS One. 2013;8(7):e67638.

67. Blankenberg D, Gordon A, Von Kuster G, Coraor N, Taylor J, Nekrutenko A, et al. Manipulation of FASTQ data with Galaxy. Bioinformatics. 2010;26(14):1783-5.

68. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25.

69. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004;14(6):1188-90.

70. Reynolds JA, Poelchau MF, Rahman Z, Armbruster PA, Denlinger DL. Transcript profiling reveals mechanisms for lipid conservation during diapause in the mosquito, Aedes albopictus. J Insect Physiol. 2012;58(7):966-73.

SUPPLEMENTAL FIGURE

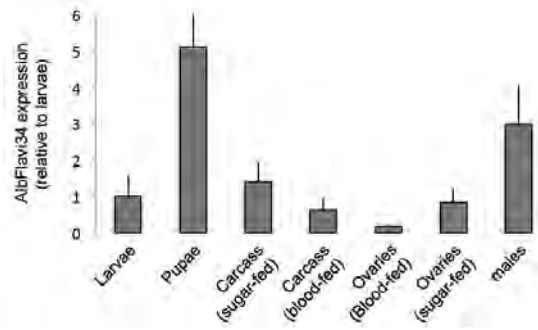


Figure S1. Expression of AlbFlavi34 in different developemental stages of *Ae. albopictus*. Expression was analyzed by qRT-PCR and expressed relative to expression in larvae. Data represent means and standard error from 3 biological replicates. Expression was analyzed in pupae, carcasses of sugar-fed (SF) or blood-fed (BF) females, eggs from BF females and males.

Available online:

Table S1	List of Dipteran species analysed
Table S2	List of viral genomes used for in silico screening of Diperan genomes. The only DNA arbovirus is in yellow.
Table S3	Description of all identified NIRVS
Table S4	DNA-seq library statistics. Total number of reads, percentage of mapped reads and percentage of properly paired reads from DNA-seq of 16 individual *Ae. albopictus* mosquitoes (Foshan strain).
Table S5	Viral integrations (NIRVS) in the *Ae. aegypti* and *Ae. albopictus* genomes. Classification based on their sequence identity. Integrations with sequence identity > 90% were grouped together.
Table S6	(A) piRNAs coverage at NIRVS loci; (B) NIRVS transcriptional activity
Supplemental File 1	Details of AlbFlavi34

The data can be downloaded from:
https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-017-3903-3

Chapter 6

Post-transcriptional Gene Silencing Mediated by Abundant piRNAs Derived from an Ultra-Conserved Satellite DNA.

Pascal Miesen, Rebecca Halbach, Bas Pennings, Chantal Vogels, Hervé Blanc, Maria-Carla Saleh, Constantianus J. Koenraadt, Ronald P. van Rij

manuscript in preparation

ABSTRACT

In animals, the PIWI interacting RNA (piRNA) pathway is an ingenious system to protect genomes against the detrimental effects of transposon mobilization. Whereas piRNA biogenesis primarily occurs in gonadal tissue in most model organisms, a somatic piRNA pathway has emerged in vector mosquitoes of the *Aedes* family. *Aedes* piRNAs are generated from various additional RNA sources other than transposon sequences, including mRNA of protein coding genes and viral RNA. Here we report the production of piRNAs from a tandem repeat satellite DNA (*satDNA1*). Strikingly, the sequences of two abundant piRNAs within the satellite are highly conserved and their expression was detected in mosquito species from the *Aedes*, *Culex*, *Culiseta* and *Coquillettidia* lineages within the *Culicinae* mosquito subfamily. These species diverged about 200 million years ago placing *satDNA1* amongst the most conserved tandem repeats known to date. piRNA production from this satellite DNA relies on the ubiquitously expressed PIWI protein Piwi4 both in *Aedes aegypti* Aag2 cells and in adult mosquitoes. To test the targeting capacity of the satellite-derived piRNAs, we designed a luciferase reporter harboring a single target site for the most abundant piRNA. This reporter is heavily silenced in Aag2 cells. Target site mutagenesis revealed that silencing is particularly sensitive to mismatches in a short nucleotide stretch near the piRNA 5′ end, analogous to the seed-sequence of microRNAs. To our knowledge, this study is the first to demonstrate post-transcriptional gene silencing by piRNAs derived from a canonical satellite DNA. The strong conservation of *satDNA1* suggests that the produced piRNAs may target genes that are active in regulatory networks conserved across many lineages of mosquitoes.

INTRODUCTION

Since their discovery about a decade ago, PIWI interacting RNAs (piRNAs) have been established as a powerful defense mechanism that protect animal genomes against the harm of transposon activity (1, 2). Transposons can randomly integrate into host genomes, thereby posing a threat to the integrity of gene sequences or regulatory elements (3, 4). In *Drosophila melanogaster*, piRNAs were found to be crucial for minimizing such mutational damage by destroying messenger RNA (mRNA) of active transposable elements and preventing their transcription by establishing a repressive chromatin environment at transposon loci (1, 2). At the heart of the pathway, PIWI proteins associated with piRNAs recognize transposon sequences and mediate gene silencing. The piRNA pathway is endowed with specificity for transposon sequences by the individual piRNAs, which are produced from dedicated genomic loci, termed piRNA clusters (5, 6). These regions are heavily enriched for remnants of transposable elements and hence, cluster-derived piRNAs are biased towards recognizing transposable elements (7). piRNA clusters give rise to an extremely variable population of primary piRNAs;

for instance, in *Drosophila* hundreds of thousands of unique piRNA sequences are produced in the female germline (5). From this pool of primary piRNAs, those recognizing complementary transcripts, for example mRNA of active transposons, are amplified by a sophisticated feed-forward amplification mechanism called the 'ping-pong loop' (5, 8).

The sequences present in the pool of primary piRNAs in each species largely depend on the repertoire of transposons that have been archived in piRNA clusters. Furthermore, the biogenesis machinery is not guided by specific sequences or structural elements, and therefore produces primary piRNAs almost at random. Only a preference for a uridine residue at the first position has been noted both in vertebrate and invertebrate species (5, 9-12). Thus, whereas the general concepts of the piRNA pathway are conserved between animal species, the individual piRNA sequences are not (5, 10, 13, 14). This makes them intrinsically different from genome-encoded microRNAs, some of which are conserved to the nucleotide level between worms and human (15, 16).

The biogenesis and functions of piRNAs have been primarily established in genetic model organisms such as flies and mice. However, recent evidence indicates that not all insights gained in these animal models can be directly extrapolated to other animal species. A striking example is the piRNA pathway of *Aedes* mosquitoes that deviates in many aspects from the paradigms established in the prototype insect model *D. melanogaster*. Most importantly, the PIWI family underwent multiple gene duplications and is represented by eight members in *Aedes (Ae.) aegypti* (Piwi1-7 and Ago3) in contrast to three PIWI genes (Piwi, Aubergine, Ago3) in the fly (17, 18). Four of the *Aedes* PIWI genes (Piwi4-6, Ago3) are abundantly expressed in somatic tissue (19), whereas in *Drosophila* piRNA pathway components are highly enriched in gonadal tissue (5, 9, 10). Accordingly, piRNAs can readily be detected in somatic tissues of *Aedes* mosquitoes (20). In addition, the *Aedes* piRNA pathway processes a broader range of substrate RNAs including protein-coding mRNAs (21, 22) and, most remarkably, RNA from cytoplasmic RNA viruses including dengue and chikungunya (20, 23-25).

Whereas transposons have been recognized as prominent source of piRNAs, the production of piRNAs from other repetitive elements has not been studied extensively. Tandem repeats include arrays of paralogous genes, DNA repeats coding for ribosomal RNAs, and a heterogeneous class of mostly direct (head-to-tail) repeats known as satellite DNA (26). Although the exact definitions vary, satellite DNAs are generally categorized based on the length of their repeat unit as either microsatellite (<10 bp) or minisatellites (10 bp or more). Repeats with a monomer size greater than 135 bp are sometimes referred to as megasatellite (27). Satellite DNAs are the major component of heterochromatin in eukaryotic genomes and serve important structural roles in the regulation of chromatin condensation and centromere formation (28). They are generally very unstable and the fastest evolving satellites have mutation rates up to 10-3 or even 10-2 per cell division (29-31). Most of these mutations are variations in the number of repeat units (27). Nonetheless,

point-mutations randomly acquired in one repeat unit can also rapidly spread through the satellite array due to unequal crossing over and gene conversion (32). The obtained mutations can then be fixed in the population by a stochastic process called molecular drive (33). Therefore, satellite DNAs generally lack strong sequence conservation, even between closely related species (32). Interestingly, some satellite DNAs show a non-uniform mutation rate across the repeat monomer with certain regions being more conserved than others. It has been inferred from this skewed rate of evolution that more conserved regions are likely to harbor functional elements (32). However, there are currently only few examples where such functional elements have been molecularly characterized.

Transcription of satellite DNA has been reported in a number of insect species, but although small RNAs derived from satellite transcripts have been detected, only little is known about their targeting capacity (34). Here, we identify an ultra-conserved satellite DNA in *Ae. aegypti* that gives rise to piRNAs, two of which are expressed at high levels both in cultured cells and in adult mosquitoes. These piRNAs rely on the ubiquitously expressed PIWI protein Piwi4 for their biogenesis. Furthermore, we show that they directly associate with Piwi4, formally classifying them as *PIWI interacting* RNAs. A luciferase reporter harboring a target site for one of the piRNAs is strongly suppressed *in trans*. In summary, this study identifies highly conserved satellite DNA-derived piRNAs as potent regulators of gene expression via post-transcriptional gene silencing.

MATERIALS AND METHODS

Cells and mosquitoes

Aag2 cells were cultured in Leibovitz's L-15 medium supplemented with 10% fetal calf serum, 1x non-essential amino acids, 2% Tryptose phosphate broth and 1% penicillin/streptomycin at 25°C without humidity and CO_2 control.

Ae. aegypti, *Culex* (*Cx.*) *pipiens pipiens*, *Cx. pipiens molestus*, *Anopheles* (*An.*) *stephensi*, *An. coluzzi*, and *An. quadriannulatus* mosquitoes used to analyze sapiR1/2 expression by northern blot, were laboratory adapted strains reared at the Department of Entomology at Wageningen University, the Netherlands, or the Department of Medical Microbiology at Radboud University Medical Center, Nijmegen, the Netherlands. *Ae. albopictus*, *Ae. intrudens*, *Ae. cantans*, *Ae. pullatus*, *Culiseta morsitans*, and *Coquillettidia richardii* were collected from the field in Sweden (Linköping), the Netherlands (Wageningen), and Italy (San Benedetto del Tronto) from June 2014 to July 2015 and long-term stored at -20°C (35). Specimens were identified to the species at the laboratory of entomology of Wageningen University.

Larvae and adult *Ae. aegypti* mosquitoes (Rockefeller strain, obtained from Bayer AG, Monheim, Germany) were maintained at 27±1°C with 12h:12h light:dark cycle and 70% relative humidity. *Adults* were provided with 6% glucose solution, and human blood

(Sanquin Blood Supply Foundation, Nijmegen, The Netherlands) was provided using the Hemotek PS5 feeder (Discovery Workshops, Lancashire, United Kingdom) for egg production. Larvae were fed with Liquifry No. 1 (Interpet, Dorking, United Kingdom) and Tetramin Baby fish food (Tetra, Melle, Germany). Naive females and males, as well as specimens from all larval stages (L1-4) and the pupal stage were selected from the rearing for further analysis of sapiR1/2 expression. In addition, naive females were offered a blood meal and engorged individuals were selected for further processing at the indicated time points. For RNA extraction, mosquitoes, larvae or pupae were deep-frozen and homogenized in RNA-Solv reagent (Omega) and RNA was isolated as described below.

For dsRNA-mediated knockdown of PIWI proteins, an iso-female mosquito line derived from *Ae. aegypti* mosquitoes originally collected in Muang District, Kamphaeng Phet, Thailand was used. Mosquitoes were reared at 28°C with 70% humidity and 12h:12h light:dark cycle. Adults were provided a 10% sucrose solution. Two to three-day-old female mosquitoes were injected with approximately 600 ng dsRNA in a volume of 207 nl using a NanojectII nanoliter injector (Drummond Scientific). 5 days post injection mosquitoes were collected and homogenized in 300 µl TRIzol reagent (Thermo Scientific) in a Precellys 24 homogenizer (Bertin technologies). After homogenization, 700 µl TRIzol was added to the homogenate and total RNA was isolated as described below.

Plasmids, dsRNA production, and antisense oligonucleotides

The sapiR1 wild type and mutant target sites were cloned into the 3′ UTR of the firefly luciferase gene in the pMT-GL3 plasmid (36). Inserts harboring the target site were generated by annealing of two complementary DNA oligonucleotides creating overhangs suitable for ligation into *Pme*I and *Sac*II digested pMT-GL3 vector. For the 5′UTR reporters, the oligonucleotides were ligated between the *Not*I and *Xho*I restriction sites. The coding sequence reporters were constructed by ligating annealed oligonucleotides directly upstream of the luciferase start codon into the *Xho*I and *Nco*I restriction sites. The oligonucleotides contained a duplication of the first 45 nucleotides of the luciferase open reading frame followed by the target sequences, thereby extending the 5′ end of luciferase by 25 amino acids but leaving the nucleotide context of the start codon intact. All sequences of the DNA oligonucleotides used for cloning are indicated in table S1. For N-terminal tagging with a GFP-tag, PIWI cDNAs were sub-cloned into the gateway system and ultimately recombined into the pAGW destination vectors, which drive protein expression by a *Drosophila* actin promotor.

Double stranded RNA targeting *Aedes* PIWI/AGO genes has been produced as previously described (37). RNase-resistant RNA oligonucleotides used to interfere with sapiR1 silencing were fully 2′*O*-methylated and antisense to sapiR1 or sense to a 28 nt region of the luciferase open reading frame as control (Sigma Aldrich). The sequences of these RNA oligos as well as the primers to generate dsRNA are provided in table S1.

Transfection of Aag2 cells

Aag2 cells were transfected using X-treme GENE HP (Roche) as described previously (37). Briefly, cells were transfected with a ratio of 2µl transfection reagent for each microgram of plasmid DNA or 4µl transfection reagent for each microgram of RNA. pMT constructs were induced with 0.5 mM copper sulfate three hours post transfection for 24 hours.

RNA isolation and beta elimination

Aag2 cells were directly lysed in RNA-solv reagent and mosquitoes were homogenized in RNA lysis reagents as described above. Total RNA was extracted following the standard procedure for phenol-chloroform based extraction methods. Briefly 200 µl chloroform was mixed with 1 ml of RNA lysis reagent. Then, the mixture was separated into an aqueous and an organic by centrifugation and the aqueous phase was recovered. RNA was precipitated using 1 volume of isopropanol and washed in 80% ethanol. The RNA pellets were resolved in nuclease-free water. Beta elimination was performed as described previously (25).

Northern blotting

Small RNA northern blotting was performed according to the protocol published in (38). Briefly, for sapiR1/2 detection 1-5 µg of total RNA were size separated on a 15% urea polyacrylamide gel. Ribosomal RNA was stained in the gel using ethidium bromide. Then, RNA was transferred to nylon membranes by semi-dry blotting and cross-linked using 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimid (EDC) . Pre-hybridization as well as hybridization with ^{32}P end-labeled DNA oligo nucleotides was performed in Ultrahyb oligo hybridization buffer (Ambion) in a hybridization oven under constant slow rotation at 42°C. Three rounds of washing were performed with 42°C warm 0.1% SDS and decreasing concentration of SSC buffer as described earlier (37). Membranes were exposed to X-ray films and developed on a table-top X-ray developer. Probe sequences are indicated in table S1. Stripping of membranes was performed in 0.1% SDS heated to 95 °C.

(Stem-loop) reverse transcription and quantitative PCR.

For analysis of transcript expression levels by reverse transcription (RT) and quantitative (q)PCR, up to 1µg of total RNA was DNaseI treated (Ambion for Aag2 cells or Promega for adult mosquitoes) and reverse transcribed (Taqman RT kit (Thermo Scientific) for Aag2 cells or Maxima H minus (Thermo Scientific) for adult mosquitoes). Stemloop RT for sapiR1 was performed according to the protocol outlined in (39) and as described earlier (21). To analyze gene expression in Aag2 cells, 5 µl of cDNA was mixed with 0.6 µl of forward and reverse primers (10 µM), 3.8 µl nuclease-free water and 10 µl 2x

GoTaq qPCR mastermix (Promega) and qPCRs were performed on a Roche Light Cycler 480 machine. For RT-qPCR of PIWI expression in adult mosquitoes, the Roche FastStart SYBR green master mix was used on a StepOnePlus real time PCR system (Applied Biosystems). Primers used for stem-loop RT and qPCR are indicated in table S1.

GFP-TRAP immunoprecipitation

Plasmids encoding GFP-fused PIWI proteins were transfected into two wells of a six wells plate. After 48 hours, cells were harvested in 500µl RIPA lysis buffer. 16% of the lysate were retained for RNA (8%) and protein (8%) analysis of the input material. The remaining lysate was subjected to GFP-immunoprecipitation using high-affinity magnetic GFP-TRAP beads (Chromotek). After an overnight incubation at 4°C under constant rotation, GFP-TRAP beads were recovered from the lysate using magnetic separation. The bead fractions were washed three times with lysis buffer and then split in a 1:2 ratio for separate analysis of protein and sapiR1 expression by western and northern blotting, respectively.

Western blotting

Protein lysates were separated on SDS polyacrylamide gels and blotted to 0.45 µm nitrocellulose membranes using a wet-transfer system (BioRad). Protein transfer was assessed using Ponceau red staining. The membranes were blocked in blocking buffer (5% dry milk in PBS-Tween) and subsequently incubated with primary antibodies (rabbit anti GFP 1:5000 or rat anti tubulin-α (Sanbio) 1:1000) in blocking buffer for 1.5 hours at room temperature or overnight at 4°C. After washing with PBS-Tween, the membranes were incubated with secondary antibodies (Licor IRdye goat-anti-rabbit or IRdye anti rat) diluted 1:10,000 in PBS-Tween. After washing with PBS-Tween and PBS, protein signal was detected on an Odyssey CLx imaginger (Li-Cor).

Luciferase reporter assays

For luciferase assays, transfected cells were harvested in Passive Lysis Buffer (Promega) 24 hours after induction of the promoter. Cell lysates were either used directly or stored at -20°C until further analyses using the Dual Luiferase assay. For this assay, 10 µl of cell lysate was mixed with 25 µl LAR-II reagent (Promega) and firefly luciferase levels were measured on a Modulus single tube luminumeter (Turner biosystems). Subsequently, 25 µl Stop and Glo solution (Promega) was added to quench firefly luciferase signal and provide a substrate for *Renilla* luciferase. Samples were vortexed and *Renilla* luciferase levels were measured on the luminometer.

3′ Rapid amplification of cDNA ends (RACE)

3′ RACE was performed using the First Choice RLM RACE kit (Ambion) according to the manufacturer's instructions. 1 μg of total RNA was reverse transcribed with 3′ RACE adapter primer for 1 hour at 42°C. Outer PCR was performed with 3′ RACE outer primer and an AAEL017385 gene-specific primer located in a region common to both splice isoforms A and B of the gene. The PCR program was: 3 min 94°C; 35 cycles of 30 s 94°C, 30 s 60°C, 45 s 72°C; 7 min 72°C. A nested PCR was performed on the PCR product with 3′ RACE inner primer and an AAEL017385 gene specific primer located in RB-exon 4, which is unique to AAEL017385 isoform RB. The PCR program was identical to the outer PCR. PCR products were size separated by agarose gel electrophoresis, purified and Sanger sequenced. Sequences of gene-specific primers are indicated in table S1.

Preparation of small RNA libraries

Small RNA libraries from pools of 5-6 adult female mosquitoes were essentially performed as described in (40). 18 to 27 μg of total RNA was size separated on a 15% acrylamide, 7M urea, 0.5x TBE gel and the small RNAs were excised from gel using 19 nt and 33 nt radioactive size markers as rulers. The RNA was retrieved into 0.3 M NaCl from the gel pieces in an overnight incubation and then precipitated in isopropanol. After washing in 70% ethanol, the RNA was dissolved in 11 μl nuclease-free water. 5μl were used as input for Illumina's TruSeq small RNA library preparation kit, which was used according to the manufacturer's recommendations. After the PCR step, reactions were loaded on a 3 % agarose gel, and the band corresponding to the expected size of the amplified small RNA libraries was excised and purified using hot-phenol extraction. The small RNA libraries were then quantified on a Qbit fluorometer, diluted to 4 nM, pooled, and sequenced using the NextSeq 500 High Output Kit v2 (75 cycles) on a NextSeq 500 (Illumina).

Bioinformatics

Basic processing of the deep-sequencing libraries

Small RNA deep-sequencing libraries of PIWI knockdowns and IPs were described previously and published in the NCBI Sequence read archive under the accession number SRA188616. FASTQ files from adult mosquito libraries were generated using the bcl2fastq script (Illumina) and subsequent manipulation of deep-sequencing data was performed using the Galaxy bioinformatics toolshed (41). The TruSeq 3′ adapter sequences were removed from the raw FastQ reads using the Clip adapter tool and standard settings (Galaxy version 1.0.1).

Analysis of sapiR coverage on satDNA1

The consensus sequence of *satDNA1* was retrieved from the AaegL3 repeat features database available on VectorBase. The small RNA libraries were mapped against this

sequence using Bowtie for Illumina (Galaxy version 1.1.2) allowing two mismatches (42). The resulting SAM file was converted into BAM and the base coverage on the locus was determined using the Genome coverage tool (Galaxy version 2.24.0). Only reads that mapped in sense orientation were taken into consideration, which accounts for more than 99.9% of the reads. As output, data suitable for histogram and genome position with 1-based coordinates were selected. The base-coverage data were exported into Microsoft Excel, log10 transformed where indicated, and imported into Prism6 for plotting.

sapiR1 and sapiR2 levels in PIWI knockdown and IP data

The SAM files obtained from mapping small RNA deep-sequencing libraries to the *satDNA1* sequence with no mismatches were converted into interval files. In these datasets, the frequencies of small RNA 5' ends at each nucleotide position were counted and the counts for position 19 (sapiR1) and 78 (sapiR2) were exported into Microsoft Excel. The counts were normalized against the size of the corresponding deep-sequencing libraries. For plotting and statistical analyses, data were imported into Prism6.

Genome-wide analysis of satDNA derived piRNAs

From the repeat features database (AaegL3) available from VectorBase all satDNAs (predicted with tandem repeats finder; TRF) were selected (43). To obtain a non-redundant set of satDNA loci, the genomic intervals of these repeats were overlapped with piRNA-sized reads from the three combined dsLuc Aag2 small RNA libraries that mapped to the *Ae. aegypti* genome (AaegL3). All piRNAs that overlapped with at least one satDNA were used to define a set of non-overlapping tandem repeat loci *de novo* using the cluster genomic intervals tool (Galaxy version 1.0.0). The maximum distance between two small RNAs to fall into one cluster was defined as 100 and the minimal number of small RNAs in each cluster was defined as 2. As output, a single interval for each cluster was obtained representing a new set of satDNAs. These loci were subsequently overlapped with the piRNA-sized reads from each PIWI knockdown and IP dataset using the join genomic intervals tool (Galaxy version 1.0.0). In the obtained datasets the frequency of piRNAs was counted for all satDNAs. These counts were exported into Microsoft Excel and normalized against the size of the corresponding deep-sequencing libraries. Then, for each satDNA the fold change was calculated of piRNA counts in the PIWI knockdown and IP libraries compared to respectively the dsLuc (average of three libraries) or GFP-IP libraries. The top 50 piRNA producing satDNAs, defined as those that had the highest average count in the dsLuc datasets, were selected for plotting. The heat-map and the hierarchical clustering, based on Pearson correlation, were created using Multiple Experiment Viewer (Tm4).

Analysis of satDNA1 nucleotide conservation

satDNA1 repeat monomers were split into parts that contained the sapiR1 sequences and parts that contained sapiR2 sequences. A multiple sequence alignment was created manually (Figure S2A,B) and used to generate sequence logos using WebLogo 3.5.0 (44).

RESULTS

Identification of a piRNA generating satellite DNA in *Aedes aegypti*

Previously, we analyzed the population of endogenous piRNAs that we had sequenced in *Ae. aegypti* Aag2 cells and found that, as expected, repetitive elements were the major source of piRNAs and accounted for about 50% of all piRNA reads that mapped to the *Aedes* genome (21). Of these repeat-derived piRNAs 50% mapped to transposon sequences published in the TE-fam database, whereas the remaining 50% aligned to other repetitive elements annotated in the repeat features database available on VectorBase (21). Careful inspection of these other repeat classes revealed that within the tandem repeats (*n*=279,076, identified by the tandem repeats finder algorithm (43)) more than 80% of all mapping piRNA reads were derived from a single locus on supercontig 1.320 (Figure 1A). This genomic region was approximately 2.5 kb in size and represented a satellite DNA uniquely present at that genomic location in the current assembly of the

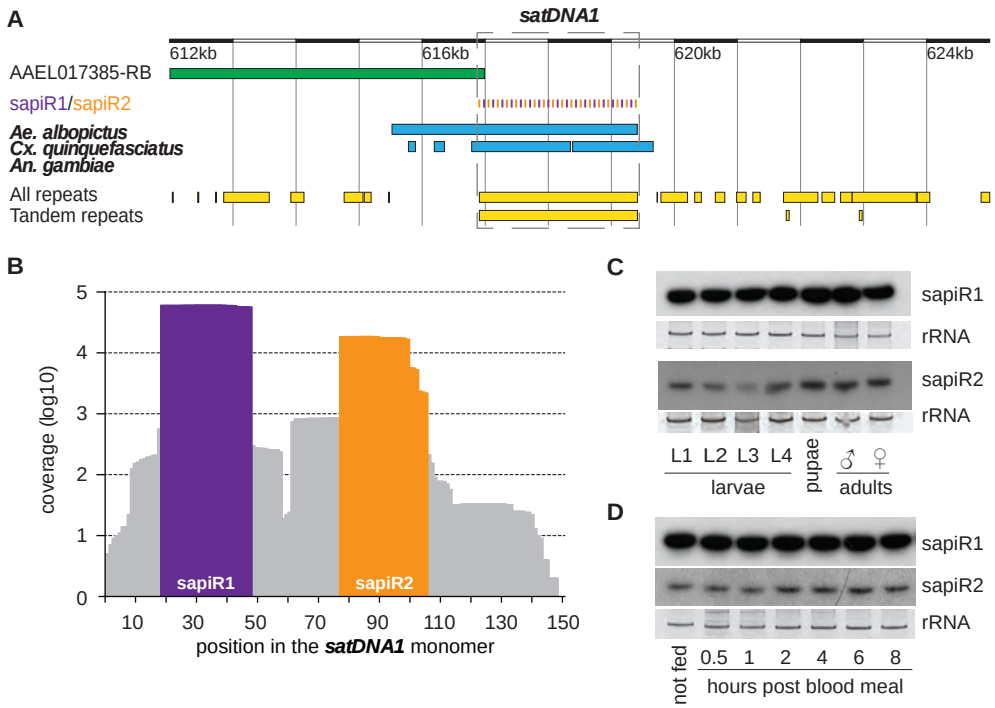


Figure 1. Expression of highly abundant direct repeat-derived piRNAs in *Aedes*. (A) Schematic illustration of the *satDNA1* locus obtained from VectorBase. A 13 kb window on supercontig1.320 is shown indicating annotated genes (green), sapiR1/2 small RNAs (purple/orange), regions conserved in other vector mosquitoes (blue) as well as repeat features (yellow). **(B)** The coverage of small RNAs from female *Ae. aegypti* (intrathoracally injected with dsRNA targeting the luciferase gene) on each position of the *satDNA1* repeat monomer (150bp). The *satDNA1* sequence was extracted from the repeat features database (AaegL3) available on VectorBase. The positions of sapiR1 and sapiR2 are indicated in purple and orange, respectively. **(C-D)** Small RNA northern blots for sapiR1/2 in (C) the indicated developmental stages, and (D) adult *Ae. aegypti* females after blood meal. Ethidium bromide staining of ribosomal RNA (rRNA) serves as loading control.

Ae. aegypti genome (45, 46). We named this tandem repeat locus *satellite DNA 1* (*satDNA1*). The repeat monomer is a consensus sequence of 150 nt in size. The *satDNA1* array contained 17 full repeats preceded by one incomplete unit (Figure S1A). To assess the distribution of small RNAs on *satDNA1*, we re-analyzed small RNA deep-sequencing libraries from Aag2 cells (37). Intriguingly, one small RNA, which we named satellite piRNA 1 (sapiR1), accounts for more than 95 % of reads that map to the *satDNA1* monomer. sapiR1 expression exceeded that of the second-most abundant small RNA mapping to the same tandem repeat (sapiR2) by about 50 fold (Figure S1B). In fact, sapiR1 was the small RNA with the highest read count in the combined Aag2 deep-sequencing libraries; its expression was similar as or slightly higher than miR-2940-3p, which is the most abundant miRNA in Aag2 cells.

The 5′ end of the *satDNA1* locus overlaps with the 3′ UTR of the annotated but uncharacterized *Aedes* gene AAEL017385 isoform B for approximately 150 bp (Figure 1A, S1A, S1C). Transcription of some satDNAs occurs via read-through from neighboring genes (34); hence we asked whether transcription or processing of *satDNA1* and AAEL017385-RB mRNA were linked. Knockdown of AAEL017385 transcripts did not influence sapiR1 levels suggesting that the sapiR locus and AAEL017385-RB are not transcriptionally coupled (Figure S1D, S1E). Rapid amplification of cDNA 3′ ends (3′ RACE) identified two transcription termination sites upstream of the currently annotated mRNA 3′ end. Both of these sites were in very close proximity to the annotated 5′ ends of sapiR1 or sapiR2. (Figure S1F). Thus whereas sapiR1 does not appear to be derived from the protein coding mRNA AAEL017385-RB, transcription termination of this gene may be linked to the production of a sapiR precursor transcript.

sapiR1 and sapiR2 are expressed *in vivo*

To investigate whether sapiR1 and sapiR2 were expressed *in vivo*, we analyzed small RNAs from adult female *Ae. aegypti* mosquitoes that have been intrathoracally injected with dsRNA. Similar to Aag2 cells, sapiR1 and sapiR2 were by far the most abundant small RNAs produced from *satDNA1*, yet sapiR1 expression exceeded that of sapiR2 only by ~3 fold (Figure 1B). sapiR1 was approximately 8 times less abundant than the most highly expressed microRNA (miR-281-5p), placing it amongst the top 60 most abundant small RNAs in our deep-sequencing libraries. In most organisms, transcription of satellite DNA occurs in a temporally or spatially regulated fashion (32). We therefore asked whether sapiR expression was dynamically expressed during the course of mosquito development from larvae to adult mosquitoes. sapiR1 and sapiR2 could readily be detected in the four larval stages, pupae, and male and female adult mosquitoes, but we did not reproducibly detect differential expression during development (Figure 1C). Besides the early developmental stages from larvae to adult, the first blood meal is an essential developmental process in female mosquitoes. For mosquitoes, blood feeding

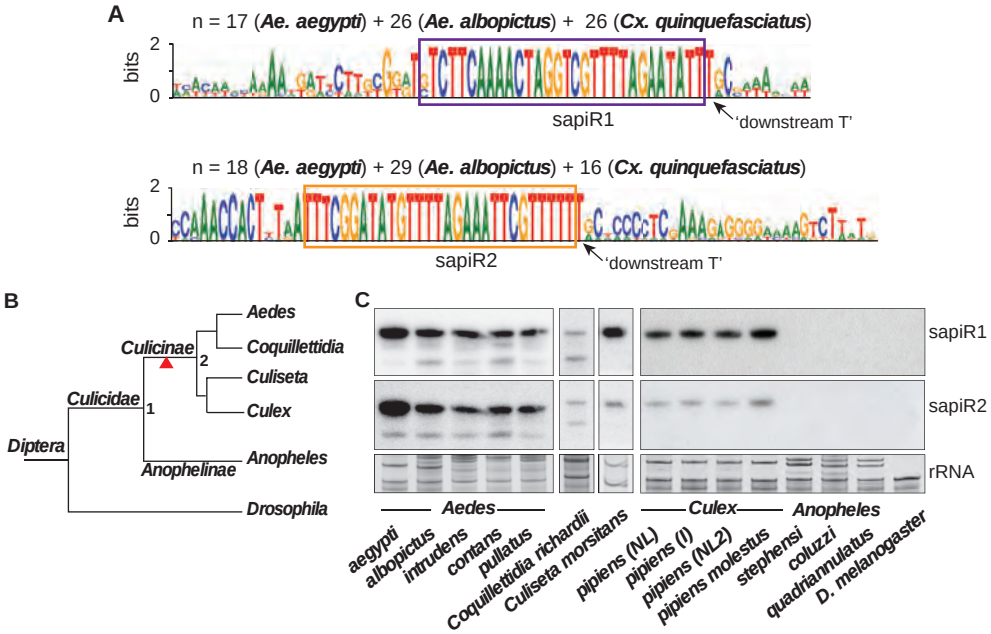


Figure 2. ***satDNA1*** **and sapiR1/2 expression are conserved amongst** ***Culicinae*** **mosquitoes.** **(A)** Conservation of *satDNA1* in *Ae. aegypti, Ae. albopictus* and *Cx. quinquefasciatus* at the nucleotide level. Boxes indicate the positions of sapiR1 and sapiR2, respectively, which guided the multiple sequence alignment. The number of repeat monomers from each species that contributed to the generation of the sequence logo is indicated above each panel. **(B)** Schematic representation of phylogenetic relationships among mosquito (*Culicidae*) sub-families based on (51). *Drosophila* serves as out-group. Branch lengths are arbitrary and do not reflect evolutionary distances. Branchpoint 1: 216.9 mya (229.5-192.2); branchpoint 2: 204.5 mya (226.2-172.3) (51). The presumable origin of *satDNA1* is marked with the red triangle. **(C)** Small RNA northern blot of sapiR1/2 in the indicated species. Ribosomal RNA (rRNA) serves as loading control.

is required for oogenesis and oviposition and numerous physiological processes are modified after a blood meal (19, 47-49). In addition, blood feeding is the most important route for infection with pathogens that circulate between mosquitoes and vertebrate hosts. Given the major implication of a blood meal on mosquito physiology or immunity we asked whether sapiR accumulation was differentially regulated as an immediate response to blood feeding. We collected female *Aedes* mosquitoes before and up to 8 hours post blood feeding and analyzed sapiR1 and sapiR2 levels by northern blotting. The accumulation of both small RNAs did not change during this period of time (Figure 1D). In conclusion, sapiR1/2 is expressed in adult *Ae. aegypti* mosquitoes without a clear differential regulation during major developmental transitions.

satDNA1 is conserved in *Culicinae* mosquitoes

Comparative genomics data available on VectorBase indicated that the *satDNA1* locus is conserved between *Ae. aegypti, Ae. albopictus* and *Cx. quinquefasciatus* but not *An. gambiae* (Figure 1A). A more detailed analysis of this locus in *Aedes* and *Culex*

revealed differences in copy number, size and intrinsic organization of the repeat monomers. In *Ae. albopictus,* the size of the consensus repeat unit is 154, only marginally larger than in *Ae. aegypti*. With exception of the 5′ terminus of the *satDNA1* array in *Ae. albopictus,* the regular alternating pattern of sapiR1 and sapiR2 was preserved between the two *Aedes* species. The most pronounced difference was the gain in copy numbers of *satDNA1* repeat-units in *Ae. albopictus* (*n*=27; Figure S1A).

In contrast to the regular organization of the repeat array in *Aedes*, *satDNA1* in *Cx. quinquefasciatus* was much more heterogeneous. Some repeat units contain stretches of unique sequences and the regularly alternating pattern of sapiR1 and sapiR2 was disturbed in large parts of the repeat (Figure S1A). In the more regular parts of *Cx. quinquefasciatus satDNA1,* the consensus monomer sequence was reduced to 120-130 bp in size. To analyze *satDNA1* sequence conservation at the nucleotide level, we split all repeat monomer into parts that contained the sapiR1 sequence and parts that contained the sapiR2 sequence. This strategy allowed us to generate local sequence alignments containing the sapiR sites and flanking sequences and to also include the more irregular *satDNA1* monomers from *Cx. quinquefasciatus* in this analysis (Figure S2A, S2B). To assess the conservation of each nucleotide in these alignments, sequence logos were generated. Strikingly, the sequences coding for sapiR1 and sapiR2 showed the highest degree of conservation, whereas flanking sequences were more variable (Figure 2A), indicating the sequence of sapiR1 and sapiR2 are under positive selection. In addition, the downstream T of both sapiR1 and sapiR2 belonged to the most conserved residues of *satDNA1*, perhaps because the nucleotide identity at this position is crucial for defining the 3′ ends of sapiR1/2. To assess whether also the expression of sapiR1 and sapiR2 was conserved, we performed small RNA northern blots on RNA isolated from 14 wild-caught and laboratory-adapted mosquito strains representing a total of five different mosquito genera (*Aedes, Culex, Culiseta, Coquillettidia* and *Anopheles*) and *D. melanogaster* (Figure 2B). Expression of sapiR1 and sapiR2 was detected in all *Aedes* species and *Culex* species analyzed. In addition the small RNAs were expressed in *Culiseta morsitans* and *Coquillettidia richardii.* In contrast, sapiR1 and sapiR2 were not found in any of five *Anopheles* species tested (Figure 2C). In addition, the analysis of publically available small RNA deep-sequencing libraries of *Ae. albopictus* (50) and *Cx. pipiens* (23) confirmed that sapiR1 and sapiR2 were the most abundant piRNAs in the *satDNA1* loci of these species (Figure S2C, S2D). The comparative genomics and small RNA expression data suggest that *satDNA1* arose in the common ancestor of *Aedes, Culex, Culiseta* and *Coquilletidia* mosquitoes after the divergence of the *Culicinae* and *Anophelinae* subfamilies of mosquitoes and before the separation of *Culex* and *Aedes* genera (Figure 2B). Therefore, the age of *satDNA1* is estimated to be approximately 200 million years (51), making it one of the most highly conserved satellite DNAs known to date. This strongly suggests that *satDNA1*-derived piRNAs exert a conserved function in *Culicinae* mosquitoes.

sapiR biogenesis is dependent on Piwi4 in cells and adult mosquitoes

We next set out to investigate the mechanism of sapiR biogenesis. The size of sapiR1 and sapiR2, which is similar to piRNAs, prompted us to investigate whether their abundance changed upon knockdown of PIWI proteins in Aag2 cells. Quantification of sapiR levels by small RNA northern blotting or stem-loop RT followed by quantitative PCR, indeed revealed that both sapiR1 and sapiR2 levels declined upon knockdown of Piwi4 (Figure 3A, Figure S3A and S3B). Of note, both methods identified a slight increase in sapiR1 levels upon Piwi5 knockdown, suggesting interplay between different branches of the *Aedes* piRNA pathway that warrants further investigation. Silencing of microRNA as well as siRNA pathway components did not result in reduced levels of sapiR1 (Figure S3C, S3D). The dependency of sapiR1 and sapiR2 expression on Piwi4 was confirmed in our previously published PIWI knockdown small RNA deep-sequencing libraries (Figure 3B). These datasets were generated from Aag2 cells after infection with Sindbis virus, an arthropod borne virus from the genus *Alphavirus* in the *Togaviridae* family. Accordingly, Piwi4 dependency of sapiR1 in SINV-infected Aag2 cells was verified by small RNA northern blotting (Figure S3E).

The final step of piRNA maturation is methylation of the 2′ hydroxyl group at the 3′ terminal ribose, which protects small RNAs against 3′-5′ exonuclease activity (52, 53). To uncover this chemical modification we performed oxidation followed by beta-elimination on total RNA isolated from Aag2 cells. This treatment shortens unmodified RNAs by one nucleoside, but leaves RNAs that bear a 3′ end modification unaffected (54). Animal microRNAs are not methylated at the 3′ terminal nucleotide and hence the observed increase in electrophoretic mobility of an abundant *Aedes* microRNA proved the effectiveness of the treatment (Figure 3C). In contrast, both sapiR1 and sapiR2 were insensitive to beta-elimination indicating that they are protected at their 3′ end, most likely by methylation (Figure 3C). Since, 3′ end methylation generally occurs after loading of the piRNA precursor into the PIWI protein, these data indicate that the majority of sapiR1 and sapiR2 are PIWI bound.

To directly assess PIWI association, we expressed GFP-tagged PIWI proteins in Aag2 cells and performed GFP-trap immunoprecipitation (IP; Figure S3F) experiments followed by small RNA northern blotting. Consistent with the results from the knockdown experiments, sapiR1 was exclusively enriched in a Piwi4 IP (Figure 3D). Also in Sindbis virus infected cells, sapiR1 and sapiR2 were only enriched upon V5-IP of V5-3xFLAG tagged Piwi4 (Figure 3E, S3G). Finally, we aimed to investigate whether Piwi4 dependency could be verified in adult mosquitoes. Therefore, we intrathoracally injected adult *Ae. aegypti* mosquitoes with dsRNA against Piwi4-6 and Ago3, which are expressed in somatic tissue (Figure S3H). As a control, we injected dsRNA against either luciferase or GFP. We assayed silencing of PIWI transcript expression in 12 mosquitoes and selected pools of 5 or 6 mosquitoes (Figure S3I) to prepare deep sequencing libraries. Consistent with results

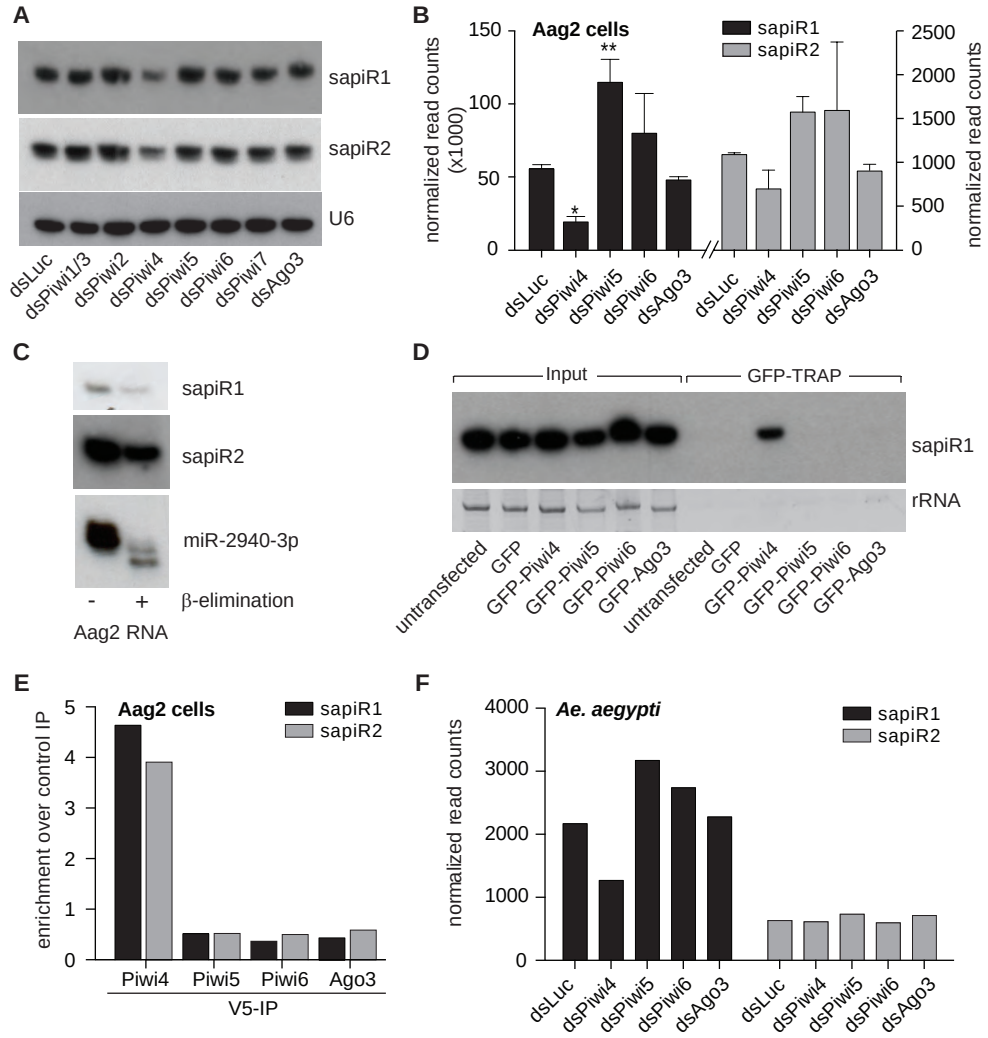


Figure 3. Biogenesis of sapiR1 and sapiR2 requires Piwi4. (A) Small RNA northern blot of sapiR1 and sapiR2 in Aag2 cells 48 hours after transfection of dsRNA targeting the indicated PIWI gene. Blotting for the U6 snRNA serves as loading control. **(B)** sapiR1 (black) and sapiR2 (grey) read counts in deep sequencing libraries from Aag2 cells transfected with dsRNA to silence the indicated PIWI gene. Libraries have been descried in more detail previously (37). Statistical significance was determined by one-way ANOVA comparing all PIWI knockdowns to the dsLuc control. *P*-values (* $P<0.05$; ** $P<0.01$) have been Bonferroni-corrected for multiple testing. **(C)** Small RNA northern blot for sapiR1/2 in Aag2 cells subjected to beta-elimination or control treatment. Efficiency of the reaction was confirmed by enhanced electrophoretic mobility of miR-2940-3p. **(D)** Small RNA northern blot for sapiR1 in Aag2 cells after GFP-trap IP of the indicated PIWI proteins. RNA for the input samples has been extracted from cell lysate prior to the IP. rRNA serves as loading control for input samples; as expected, rRNA is depleted from IP samples. **(E)** Enrichment of sapiR1 (black) and sapiR2 (grey) sequences in deep sequencing libraries from Aag2 cells after IP of the indicated PIWI protein compared to a control GFP IP. Libraries have been described in more detail previously (37). **(F)** sapiR1/2 read counts in deep-sequencing libraries prepared from adult *Ae. aegypti* females 5 days post injection of the indicated dsRNAs.

obtained in Aag2 cells, *in vivo* knockdown of only Piwi4 resulted in a decrease of sapiR1 levels but no effect on sapiR2 was detected in this experiment (Figure 3F). Reduction of sapiR1 levels was confirmed in a second knockdown experiment using independent dsRNA sequences targeting Piwi4 and, as negative controls, GFP and Ago3 (Figure S3J, S3K). In this experiment also a slight reduction of sapiR2 was observed. Noteworthy, Piwi5 knockdown modestly enhanced the expression of sapiR1 *in vivo* comparable to Aag2 cells (Figure 3B, 3F). Collectively, these data formally classify sapiR1/2 as *PIWI-interacting* RNAs, which depend on the ubiquitously expressed PIWI protein Piwi4.

Piwi4 association is not a common feature of direct repeat derived piRNAs

Next we asked whether dependency on Piwi4 was common to all piRNAs derived from direct repeats. Therefore, we re-analyzed our PIWI knockdown and IP small RNA deep-sequencing data (37), focusing on those small RNAs that overlapped with direct repeats. In total, 77 direct repeats gave rise to more than 10 piRNA reads per million mapped reads in the dsLuc control libraries. Assessing PIWI dependency and association of the top 100 piRNA producing repeat loci revealed that only three additional tandem repeats showed a similar pattern as *satDNA1* (Figure S4). Most other loci were dependent on Piwi4 in combination with other PIWI proteins, mostly Piwi5. This profile of PIWI dependency was very similar to the one we had previously described for piRNAs from the majority of transposable elements (37). Also the enrichment of many tandem repeat-derived piRNAs in Piwi5 and Piwi6 IPs reflected the PIWI association of transposon piRNAs (Figure S4). We thus concluded that Piwi4 dependency is not a common feature of all direct repeat derived piRNAs. It is currently unclear which features are responsible for the selective association of sapiR1, sapiR2 and a handful of additional, repeat-derived piRNAs with Piwi4.

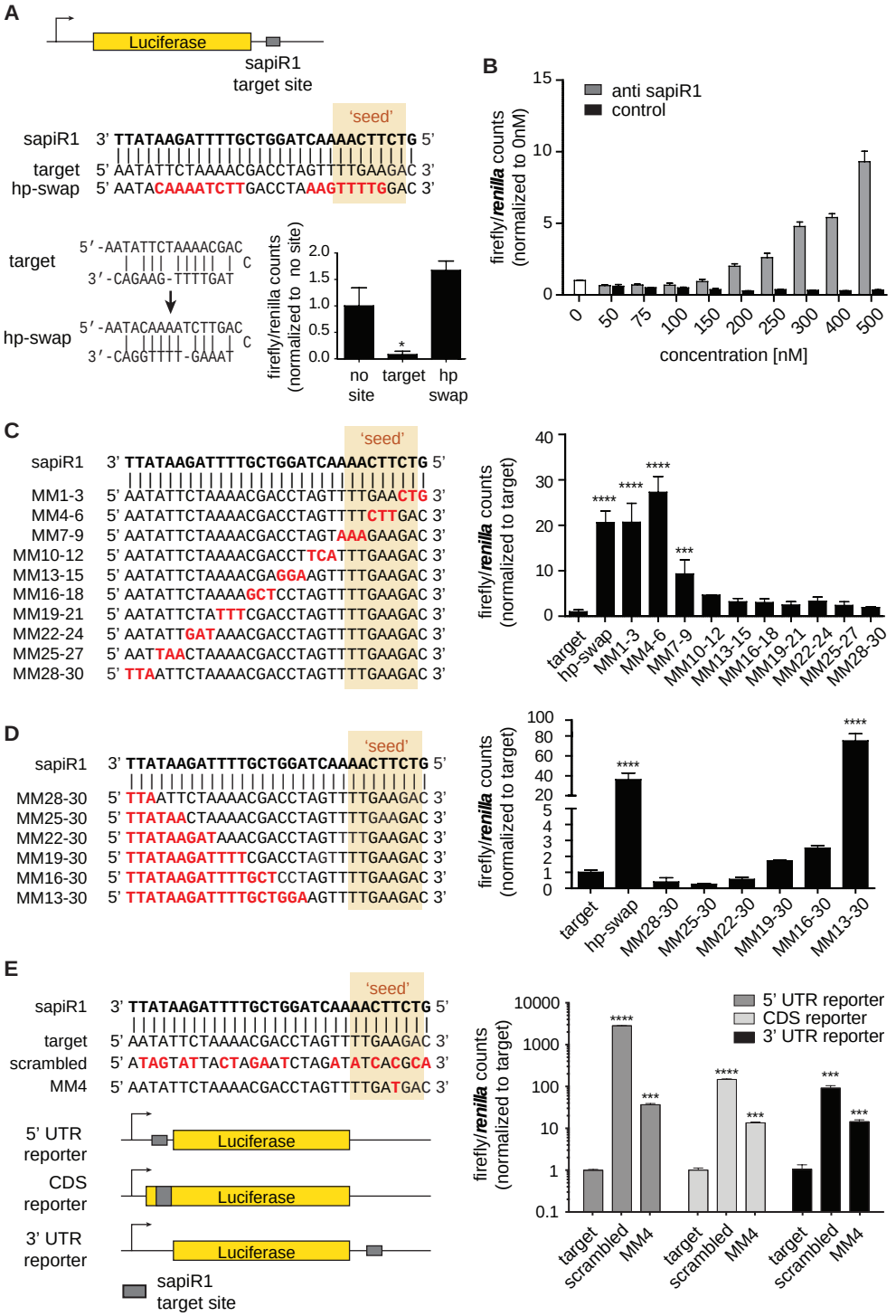
Post-transcriptional silencing by sapiR1

The association of sapiR1 and sapiR2 with a PIWI protein prompted us to investigate whether *satDNA1*-derived piRNAs were able to target RNAs that contained complementary sequences. To this end, we designed a reporter system, analogous to established microRNA reporters, in which a single sapiR1 target site was inserted in the 3' UTR of firefly luciferase (Figure 4A). As control, we used a luciferase plasmid without a target site. Since the sapiR1 site has potential to form an intramolecular hairpin structure (Figure 4A) that may modulate the stability of the luciferase reporter, we also constructed a mutated reporter in which we swapped the sequence orientation on both arms of the hairpin. This construct was expected to disrupt target binding of sapiR1 *in trans*, yet it retained the potential of hairpin formation. We transfected these reporters into *Ae. aegypti* Aag2 cells together with a Renilla luciferase expression plasmid to normalize the firefly luciferase counts. Introduction of a sapiR1 target site caused 14-fold lower luciferase

activity compared to the control pMt-GL3 construct (Figure 4A, bar chart). The hairpin mutant was expressed to similar levels as the control, indicating that repression of the sapiR1 reporter was not caused by internal destabilization of the luciferase transcript but via sapiR1 binding *in trans*. To further validate this observation, we co-transfected the sapiR1 reporter together with a nuclease-resistant, antisense oligonucleotide that was 2′*O*-methylated at each ribose. Transfection of a sapiR1 antisense oligonucleotide but not a control oligonucleotide de-silenced the sapiR1 reporter in a concentration-dependent manner, confirming that reporter silencing was due to targeting by sapiR1 (Figure 4B). Expression of sapiR1 itself was unchanged by co-transfection of the antisense oligo, indicating that de-silencing was caused by sequestration of sapiR1 complexes rather than degradation of the small RNA (Figure S5A).

Next, we aimed to investigate the requirements needed for efficient reporter silencing by sapiR1 in greater detail. We generated ten reporters, in which we mutated three consecutive nucleotides in the sapiR1 target region, keeping the GC percentage of each construct identical (Figure 4C). Reporter silencing was significantly impaired when the nucleotides base pairing to sapiR1 position 1-9 were mutated (Figure 4C). These data suggest that sapiR1-mediated silencing requires recognition of target RNA via a short stretch of nucleotides near the 5′ end of the small RNA, reminiscent of a microRNA seed sequence (55). Targeted mutagenesis of single nucleotides opposite of the presumable sapiR1 seed resulted in de-silencing of the sapiR1 reporter when the nucleotides opposite sapiR1 position 2-7 were altered (Figure S5B). Yet, de-repression was weaker than when triple mismatches were introduced into the sapiR1 target, suggesting that single mismatches could at least partially be compensated for. Next, we aimed to investigate how many mismatched were tolerated at the 3′ end to still allow efficient target silencing. Therefore, we constructed reporter mutants with an increasing number of mismatches opposite of the sapiR1 3′ end (Figure 4D). The reporter construct was still silenced to similar levels as the wildtype sapiR1 reporter when 15 consecutive mismatches were introduced at the 3′ end, leaving 15 nt complementarity at the 5′ end of draP1. Three additional mismatches de-silenced the reporter about 40 fold (Figure 4D). Interactions with the seed sequence was not affected in this construct suggesting that seed binding alone is not sufficient to induce target silencing.

To assess if targeting of sapiR1 was dependent of the location of the target site on the mRNA, we constructed reporters in which we placed the sapiR1 target site or a control sequence in the 5′ UTR or the coding sequence of the luciferase gene. (Figure 4E). Introduction of a sapiR1 target site in the 5′ UTR and coding sequence resulted in profound repression of luciferase activity of approximately 2700 fold or 150 fold, respectively (Figure 4E). These data suggest that sapiR1 can exert its silencing function independent of the location of the target site in a transcript. Altogether, our reporter experiments indicate that sapiR1 mediated silencing resembles microRNA-mediated


A
Luciferase
sapiR1
target site
'seed'
sapiR1 3' TTATAAGATTTTGCTGGATCAAAACTTCTG 5'
target 5' AATATTCTAAAACGACCTAGTTTTGAAGAC 3'
hp-swap 5' AATACAAAATCTTGACCTAAAGTTTTGGAC 3'
target
hp-swap
firefly/renilla counts (normalized to no site)
no site
target
hp swap
B
anti sapiR1
control
firefly/renilla counts (normalized to 0nM)
concentration [nM]
C
'seed'
MM1-3
MM4-6
MM7-9
MM10-12
MM13-15
MM16-18
MM19-21
MM22-24
MM25-27
MM28-30
firefly/renilla counts (normalized to target)
D
'seed'
MM28-30
MM25-30
MM22-30
MM19-30
MM16-30
MM13-30
E
'seed'
target
scrambled
MM4
5' UTR reporter
CDS reporter
3' UTR reporter
Luciferase
sapiR1 target site
5' UTR reporter
CDS reporter
3' UTR reporter

gene repression which is largely influenced by seed-binding (55). Yet, a seed-target interaction alone is insufficient for repression of a target RNA carrying a single target site.

DISCUSSION

The population of primary piRNAs targeting transposons in the germ line cells of animals comprises an enormous variety of individual sequences that are generally not conserved between species (5, 10, 13, 14). It has been suggested that adaption to a rapidly changing transposon environments within and between species underlie this enormous piRNA diversity (13, 56). The ping-pong loop selectively amplifies only those piRNAs that have recognized a target RNA (5, 8), a concept that somewhat resembles the clonal expansion of lymphocytes upon recognition of foreign antigens. In sharp contrast, sapiR1 and sapiR2 are highly abundant in mosquito cells without an apparent amplification step. In fact, more than 99.99% of *satDNA1*-derived small RNAs map in the same orientation suggesting that they are produced exclusively in a primary biogenesis pathway directly from a *satDNA1* transcript. Yet, the nature of the transcript as well the regulatory elements that underlie its expression are currently unknown. Our data argue against *satDNA1* being expressed by transcriptional read-through from the upstream gene AAEL017385-RB since its suppression did not alter sapiR1 levels. In addition, although the *satDNA1* locus is well conserved between *Ae. aegypti* and *Ae. albopictus* the proximity to AAEL017385 is not. Rather, the *Ae. albopictus* ortholog AALF011179 is located on a different supercontig as *satDNA1* and the sequence directly upstream of satDNA is not conserved between *Ae. aegypti* and *Ae. albopictus*. Alternatively, *satDNA1* may represent an autonomous transcription unit with its own promoter sequence. Such internal promoter elements and transcription factor binding sites have been mapped within satellite DNAs in other insect species (57-59).

Downstream biogenesis steps that mediate piRNA maturation from a putative *satDNA1* transcript also remain to be investigated, but a few conclusions can be drawn from the small RNA deep-sequencing data. A putative *satDNA1* transcript is not predicted to form double stranded RNA or local hairpins that would allow processing by an RNaseIII enzyme. Accordingly, neither Drosha nor Dicer1 and its ortholog Dicer2 are involved in sapiR biogenesis. This suggests that, similar to transposon piRNAs in *Drosophila*, sapiRs

◀ **Figure 4. Post-transcriptional gene silencing mediated by sapiR1. (A)** Schematic representation of the sapiR1 luciferase reporter and sequences of the sapiR1 target site as well as the hairpin swap control (hp-swap). The putative hairpin formed within the target sites is shown. The bar chart shows firefly/*Renilla* luciferase counts for the indicated reporter constructs. **(B)** Firefly/*Renilla* luciferase counts in the lysate of Aag2 cells transfected with a sapiR1 target site reporter in combination with increasing amounts of a nuclease resistant RNA oligo antisense to sapiR1 or a control oligo (sense to the luciferase transcript). **(C,D)** Left panel: illustration of the mutagenesis strategy for the sapiR1 reporter; the 'seed' sequence of sapiR1 is indicated. Right panel: firefly/*Renilla* liciferase counts in the lysate of Aag2 cells transfected with the indicated reporter plasmids. **(E)** Left panel: schematic representation of sapiR1 reporter plasmids with target sites located either in the 5′ UTR, coding sequence (CDS), or 3′ UTR of the luciferase gene. Right panel: firefly/*Renilla* luciferase counts in the lysate of Aag2 cells transfected with the indicated reporter plasmids.

are excised from single-stranded RNA precursors. Interestingly, sapiR1 is a near-perfect 30 mer both in adult mosquitoes and in cells. Also, the vast majority of sapiR2 in Aag2 cells is 29 nt in size, but *in vivo* the 3′ end is more heterogeneous with the major species being 23, 26 and 29 nucleotides in size. While this pattern does not fully exclude the activity of an exonuclease shaping the 3′ end of sapiRs, our data suggest that both the 5′ end and the 3′ end are defied by an, as yet unidentified, endonuclease. A candidate is Zucchini, a nuclease that creates both 5′ and 3′ end of primary piRNAs and thus is central in piRNA biogenesis in *Drosophila* (11, 12, 60-62). *In vivo*, Zucchini cleavages are biased to occur directly upstream of uridine residues (60-62), which partly explains the 1U bias of primary piRNAs in *Drosophila*. Strikingly, the T residue downstream of both sapiR1 and sapiR2 is amongst the most highly conserved nucleotides in *satDNA1* outside of the small RNA sequences themselves, suggesting that it might be important for 3` end formation. Also the first nucleotide of sapiR2 is a U, in line with a putative U-preference of the sapiR1/2 generating nuclease. However, sapiR1 begins either with a G (dominant nucleotide in *Aedes*) or a C (in *Culex*) and despite U residues in neighboring positions, this small RNA is very efficiently processed both in Aag2 cells as well as *in vivo* upstream of the 1G/C position. This suggests that either a putative nucleotide bias of the responsible nuclease is not 100% penetrant or that other structural or sequence elements determine the termini of sapiR1 and/or sapiR2.

Both sapiR1 and sapiR2 associated with the PIWI protein Piwi4, formally classifying them as piRNAs. Piwi4 is ubiquitously expressed in somatic and germline tissues and through all developmental stages (19). It was recently reported to be at least 2-fold up-regulated 5 hours after blood feeding of female *Ae. aegypti* mosquitoes (49). This increase in Piwi4 expression is not reflected in increased levels of sapiR1/2 after blood feeding, perhaps indicating that Piwi4 availability is not a restricting factor for intracellular sapiR1/2 expression. We have previously shown that Piwi4 knockdown interferes with piRNA formation from a majority of transposons in Aag2 cells. Yet, an IP of transgenic Piwi4 expressed in Aag2 cells was depleted of transposon piRNAs suggesting an indirect mode of action (37). In addition, Piwi4 was suggested to confer antiviral activity against Semliki Forest virus and Orthobunyaviruses in mosquito cells (63, 64). At the moment it is unclear whether any of these reported effects can be attributed to sapiR1/2 associated to Piwi4. Since, at least in Aag2 cells, sapiR1 is amongst the most abundant small RNAs, it is perceivable that this small RNA largely dictates Piwi4 activity.

Small RNAs derived from satellite DNA have been reported to act *in cis* and aid in establishing heterochromatin at the locus they were originally derived from (65-68). In this scenario, small RNA-loaded protein complexes in the nucleus may bind to nascent transcripts and/or RNA Polymerase-II derived at the repeated loci (66, 67). In a second step, heterochromatin-modifying factors are recruited to create a repressive chromatin environment (68), a process reminiscent of transcriptional silencing of transposon loci

mediated by Piwi in *Drosophila* (69-71). Much more rarely, satellite DNA-derived piRNAs were reported to guide suppression of target RNAs *in trans*. In *Drosophila*, the repetitive testis expressed *Stellate* genes are silenced by small RNAs derived from the cognate Suppressor of Stellate *Su(Ste)* direct repeat (65, 72) and failure to maintain this silencing results in male sterility (73). Yet, the *Su(Ste)* repeat is a rather complex repeat of 2.8 kb that contains intronic sequences as well as a transposon insertion (72). In contrast, the *satDNA1* repeat reported here is a stereotypical tandem repeat of less than 200 nucleotides in size. To our knowledge, this study is the first to report post-transcriptional gene silencing mediated by piRNAs derived from a simple satellite DNA sequence. Whereas we cannot exclude that sapiRs bind a putative nascent transcript *in cis* at the *satDNA1* locus, our data strongly support the notion that *trans* targets harboring a complementary site can be efficiently silenced by sapiR1. Our reporter experiments suggest that targeting is location independent and target sites both in the coding sequence as well as in the 5′ and 3′ UTR are recognized. Gene silencing by sapiR1 strongly depends on target binding via a short nucleotide sequence stretch near the 5′ end of the piRNA, reminiscent of microRNA seed sequences (55). Target recognition is not impaired when the nucleotide facing the first position of the small RNA is mutated. This is typical for small RNAs bound to Argonaute proteins since the first residue is usually locked in a pocket within the MID-PIWI domain module and therefore inaccessible for binding complementary RNAs (74). In line with this notion, the first nucleotide of sapiR1 is not conserved between *Aedes* and *Culex* mosquitoes. Interestingly, also mutations of the target site at position ten and eleven of the piRNA do not strongly inhibit gene silencing. Usually, full complementarity at these positions is required for slicing activity of Argonaute proteins (75). Hence, these data suggest that target repression by sapiR1 is not mediated by RNA cleavage but via alternative mechanisms such as translational repression and RNA destabilization through de-capping or de-adenylation.

sapiR1 and sapiR2 sequences are the most conserved parts of the *satDNA1* locus and expression of these piRNAs could be detected in eight species representing four sub-families in the family of *Culicinae* mosquitoes. The *satDNA1* locus and sapiR1/2 expression was absent from *Anopheles* mosquitoes. This dates the origin of *satDNA1* back to about 200 mya making it one of the oldest and most conserved satellite DNAs known to date (34). The ultra-high conservation of especially the small RNA sequences strongly supports the idea that these are functional elements that execute a function that requires conservation at the nucleotide level. This is in sharp contrast to canonical piRNAs, which are typically not conserved, even between species that are more closely related than *Culex* and *Aedes* mosquitoes (5, 10, 13, 14). This example once more illustrates that the mosquito piRNA pathway does not fully adhere to the paradigms established in other model organisms. It would be of great interest to establish if piRNA production from similarly ancient satellite DNAs occurs in other animal species as well.

ACKNOWLEDGEMENTS

We thank the members of the van Rij laboratory for fruitful discussions on the manuscript, Bas Dutilh for bioinformatic advice and Tim Möhlmann for providing samples from field-collected mosquitoes. Thanks also to Louis Lambrechts for providing *Ae. aegypti* mosquitoes used for dsRNA injections and to Yasutsugu Suzuki and Catherine Lallemand for their assistance with mosquito manipulations and rearing.

This work was financially supported by a European Research Council Consolidator Grant under the European Union's Seventh Framework Programme (ERC CoG 615680) to RvR, by a PhD fellowship from Radboud University Medical Center and an EMBO Short-Term Fellowship (EMBO-ASTF 449-2016) to PM.

REFERENCES

1. Czech B, Hannon GJ. One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing. Trends Biochem Sci. 2016;41(4):324-37.
2. Iwasaki YW, Siomi MC, Siomi H. PIWI-Interacting RNA: Its Biogenesis and Functions. Annu Rev Biochem. 2015;84:405-33.
3. Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the genome. Nat Rev Genet. 2007;8(4):272-85.
4. Siomi MC, Sato K, Pezic D, Aravin AA. PIWI-interacting small RNAs: the vanguard of genome defence. Nat Rev Mol Cell Biol. 2011;12(4):246-58.
5. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, et al. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell. 2007;128(6):1089-103.
6. Le Thomas A, Stuwe E, Li S, Du J, Marinov G, Rozhkov N, et al. Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. Genes Dev. 2014;28(15):1667-80.
7. Le Thomas A, Toth KF, Aravin AA. To be or not to be a piRNA: genomic origin and processing of piRNAs. Genome Biol. 2014;15(1):204.
8. Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, et al. A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in Drosophila. Science. 2007;315(5818):1587-90.
9. Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, Iovino N, et al. A novel class of small RNAs bind to MILI protein in mouse testes. Nature. 2006;442(7099):203-7.
10. Girard A, Sachidanandam R, Hannon GJ, Carmell MA. A germline-specific class of small RNAs binds mammalian Piwi proteins. Nature. 2006;442(7099):199-202.
11. Ipsaro JJ, Haase AD, Knott SR, Joshua-Tor L, Hannon GJ. The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis. Nature. 2012;491(7423):279-83.
12. Nishimasu H, Ishizu H, Saito K, Fukuhara S, Kamatani MK, Bonnefond L, et al. Structure and function of Zucchini endoribonuclease in piRNA biogenesis. Nature. 2012;491(7423):284-7.
13. Chirn GW, Rahman R, Sytnikova YA, Matts JA, Zeng M, Gerlach D, et al. Conserved piRNA Expression from a Distinct Set of piRNA Cluster Loci in Eutherian Mammals. PLoS Genet. 2015;11(11):e1005652.
14. Lau NC, Seto AG, Kim J, Kuramochi-Miyagawa S, Nakano T, Bartel DP, et al. Characterization of the piRNA complex from rat testes. Science. 2006;313(5785):363-7.
15. Pasquinelli AE, Reinhart BJ, Slack F, Martindale MQ, Kuroda MI, Maller B, et al. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. Nature. 2000;408(6808):86-9.

16. Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. Identification of novel genes coding for small expressed RNAs. Science. 2001;294(5543):853-8.
17. Lewis SH, Salmela H, Obbard DJ. Duplication and Diversification of Dipteran Argonaute Genes, and the Evolutionary Divergence of Piwi and Aubergine. Genome Biol Evol. 2016;8(3):507-18.
18. Campbell CL, Black WCt, Hess AM, Foy BD. Comparative genomics of small RNA regulatory pathway components in vector mosquitoes. BMC Genomics. 2008;9:425.
19. Akbari OS, Antoshechkin I, Amrhein H, Williams B, Diloreto R, Sandler J, et al. The developmental transcriptome of the mosquito Aedes aegypti, an invasive species and major arbovirus vector. G3 (Bethesda). 2013;3(9):1493-509.
20. Morazzani EM, Wiley MR, Murreddu MG, Adelman ZN, Myles KM. Production of virus-derived ping-pong-dependent piRNA-like small RNAs in the mosquito soma. PLoS Pathog. 2012;8(1):e1002470.
21. Girardi E, Miesen P, Pennings B, Frangeul L, Saleh MC, van Rij RP. Histone-derived piRNA biogenesis depends on the ping-pong partners Piwi5 and Ago3 in Aedes aegypti. Nucleic Acids Res. 2017;45(8):4881-92.
22. Arensburger P, Hice RH, Wright JA, Craig NL, Atkinson PW. The mosquito Aedes aegypti has a large genome size and high transposable element load but contains a low proportion of transposon-specific piRNAs. BMC Genomics. 2011;12:606.
23. Miesen P, Joosten J, van Rij RP. PIWIs Go Viral: Arbovirus-Derived piRNAs in Vector Mosquitoes. PLoS Pathog. 2016;12(12):e1006017.
24. Hess AM, Prasad AN, Ptitsyn A, Ebel GD, Olson KE, Barbacioru C, et al. Small RNA profiling of Dengue virus-mosquito interactions implicates the PIWI RNA pathway in anti-viral defense. BMC Microbiol. 2011;11:45.
25. Miesen P, Ivens A, Buck AH, van Rij RP. Small RNA Profiling in Dengue Virus 2-Infected Aedes Mosquito Cells Reveals Viral piRNAs and Novel Host miRNAs. PLoS Negl Trop Dis. 2016;10(2):e0004452.
26. Richard GF, Kerrest A, Dujon B. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. Microbiol Mol Biol Rev. 2008;72(4):686-727.
27. Gemayel R, Vinces MD, Legendre M, Verstrepen KJ. Variable tandem repeats accelerate evolution of coding and regulatory sequences. Annu Rev Genet. 2010;44:445-77.
28. Palomeque T, Lorite P. Satellite DNA in insects: a review. Heredity (Edinb). 2008;100(6):564-73.
29. Brinkmann B, Klintschar M, Neuhuber F, Huhne J, Rolf B. Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. Am J Hum Genet. 1998;62(6):1408-15.
30. Legendre M, Pochet N, Pak T, Verstrepen KJ. Sequence-based estimation of minisatellite and microsatellite repeat variability. Genome Res. 2007;17(12):1787-96.
31. Verstrepen KJ, Jansen A, Lewitter F, Fink GR. Intragenic tandem repeats generate functional variability. Nat Genet. 2005;37(9):986-90.
32. Ugarkovic D. Functional elements residing within satellite DNAs. EMBO Rep. 2005;6(11):1035-9.
33. Dover G. Molecular drive. Trends Genet. 2002;18(11):587-9.
34. Pezer Z, Brajkovic J, Feliciello I, Ugarkovic D. Transcription of Satellite DNAs in Insects. Prog Mol Subcell Biol. 2011;51:161-78.
35. Vogels CB, Mohlmann TW, Melsen D, Favia G, Wennergren U, Koenraadt CJ. Latitudinal Diversity of Culex pipiens Biotypes and Hybrids in Farm, Peri-Urban, and Wetland Habitats in Europe. PLoS One. 2016;11(11):e0166959.
36. van Rij RP, Saleh MC, Berry B, Foo C, Houk A, Antoniewski C, et al. The RNA silencing endonuclease Argonaute 2 mediates specific antiviral immunity in Drosophila melanogaster. Genes Dev. 2006;20(21):2985-95.
37. Miesen P, Girardi E, van Rij RP. Distinct sets of PIWI proteins produce arbovirus and transposon-derived piRNAs in Aedes aegypti mosquito cells. Nucleic Acids Res. 2015;43(13):6545-56.

38. Pall GS, Hamilton AJ. Improved northern blot method for enhanced detection of small RNA. Nat Protoc. 2008;3(6):1077-84.

39. Chen C, Ridzon DA, Broomer AJ, Zhou Z, Lee DH, Nguyen JT, et al. Real-time quantification of microRNAs by stem-loop RT-PCR. Nucleic Acids Res. 2005;33(20):e179.

40. van Cleef KW, van Mierlo JT, Miesen P, Overheul GJ, Fros JJ, Schuster S, et al. Mosquito and Drosophila entomobirnaviruses suppress dsRNA- and siRNA-induced RNAi. Nucleic Acids Res. 2014;42(13):8732-44.

41. Blankenberg D, Gordon A, Von Kuster G, Coraor N, Taylor J, Nekrutenko A, et al. Manipulation of FASTQ data with Galaxy. Bioinformatics. 2010;26(14):1783-5.

42. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25.

43. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27(2):573-80.

44. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004;14(6):1188-90.

45. Nene V, Wortman JR, Lawson D, Haas B, Kodira C, Tu ZJ, et al. Genome sequence of Aedes aegypti, a major arbovirus vector. Science. 2007;316(5832):1718-23.

46. Giraldo-Calderon GI, Emrich SJ, MacCallum RM, Maslen G, Dialynas E, Topalis P, et al. VectorBase: an updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases. Nucleic Acids Res. 2015;43(Database issue):D707-13.

47. Hocking B. Blood-sucking behavior of terrestrial arthropods. Annu Rev Entomol. 1971;16:1-26.

48. Sanders HR, Evans AM, Ross LS, Gill SS. Blood meal induces global changes in midgut gene expression in the disease vector, Aedes aegypti. Insect Biochem Mol Biol. 2003;33(11):1105-22.

49. Bonizzoni M, Dunn WA, Campbell CL, Olson KE, Dimon MT, Marinotti O, et al. RNA-seq analyses of blood-induced changes in gene expression in the mosquito vector species, Aedes aegypti. BMC Genomics. 2011;12:82.

50. Gu J, Hu W, Wu J, Zheng P, Chen M, James AA, et al. miRNA genes of an invasive vector mosquito, Aedes albopictus. PLoS One. 2013;8(7):e67638.

51. Reidenbach KR, Cook S, Bertone MA, Harbach RE, Wiegmann BM, Besansky NJ. Phylogenetic analysis and temporal diversification of mosquitoes (Diptera: Culicidae) based on nuclear genes and morphology. BMC Evol Biol. 2009;9:298.

52. Horwich MD, Li C, Matranga C, Vagin V, Farley G, Wang P, et al. The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. Curr Biol. 2007;17(14):1265-72.

53. Saito K, Sakaguchi Y, Suzuki T, Suzuki T, Siomi H, Siomi MC. Pimet, the Drosophila homolog of HEN1, mediates 2'-O-methylation of Piwi- interacting RNAs at their 3' ends. Genes Dev. 2007;21(13):1603-8.

54. Kawaoka S, Katsuma S, Tomari Y. Making piRNAs in vitro. Methods Mol Biol. 2014;1093:35-46.

55. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell. 2009;136(2):215-33.

56. Song J, Liu J, Schnakenberg SL, Ha H, Xing J, Chen KC. Variation in piRNA and transposable element content in strains of Drosophila melanogaster. Genome Biol Evol. 2014;6(10):2786-98.

57. Pezer Z, Ugarkovic D. Satellite DNA-associated siRNAs as mediators of heat shock response in insects. RNA Biol. 2012;9(5):587-95.

58. Renault S, Rouleux-Bonnin F, Periquet G, Bigot Y. Satellite DNA transcription in Diadromus pulchellus (Hymenoptera). Insect Biochem Mol Biol. 1999;29(2):103-11.

59. Raff JW, Kellum R, Alberts B. The Drosophila GAGA transcription factor is associated with specific regions of heterochromatin throughout the cell cycle. EMBO J. 1994;13(24):5977-83.

60. Han BW, Wang W, Li C, Weng Z, Zamore PD. Noncoding RNA. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. Science. 2015;348(6236):817-21.

61. Hayashi R, Schnabl J, Handler D, Mohn F, Ameres SL, Brennecke J. Genetic and mechanistic diversity of piRNA 3'-end formation. Nature. 2016;539(7630):588-92.

62. Mohn F, Handler D, Brennecke J. Noncoding RNA. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. Science. 2015;348(6236):812-7.

63. Schnettler E, Donald CL, Human S, Watson M, Siu RW, McFarlane M, et al. Knockdown of piRNA pathway proteins results in enhanced Semliki Forest virus production in mosquito cells. J Gen Virol. 2013;94(Pt 7):1680-9.

64. Dietrich I, Shi X, McFarlane M, Watson M, Blomstrom AL, Skelton JK, et al. The Antiviral RNAi Response in Vector and Non-vector Cells against Orthobunyaviruses. PLoS Negl Trop Dis. 2017;11(1):e0005272.

65. Aravin AA, Lagos-Quintana M, Yalcin A, Zavolan M, Marks D, Snyder B, et al. The small RNA profile during Drosophila melanogaster development. Dev Cell. 2003;5(2):337-50.

66. Verdel A, Jia S, Gerber S, Sugiyama T, Gygi S, Grewal SI, et al. RNAi-mediated targeting of heterochromatin by the RITS complex. Science. 2004;303(5658):672-6.

67. Kato H, Goto DB, Martienssen RA, Urano T, Furukawa K, Murakami Y. RNA polymerase II is required for RNAi-dependent heterochromatin assembly. Science. 2005;309(5733):467-9.

68. Volpe TA, Kidner C, Hall IM, Teng G, Grewal SI, Martienssen RA. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science. 2002;297(5588):1833-7.

69. Pal-Bhadra M, Leibovitch BA, Gandhi SG, Chikka MR, Bhadra U, Birchler JA, et al. Heterochromatic silencing and HP1 localization in Drosophila are dependent on the RNAi machinery. Science. 2004;303(5658):669-72.

70. Le Thomas A, Rogers AK, Webster A, Marinov GK, Liao SE, Perkins EM, et al. Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. Genes Dev. 2013;27(4):390-9.

71. Sienski G, Donertas D, Brennecke J. Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. Cell. 2012;151(5):964-80.

72. Aravin AA, Naumova NM, Tulin AV, Vagin VV, Rozovsky YM, Gvozdev VA. Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the D. melanogaster germline. Curr Biol. 2001;11(13):1017-27.

73. Livak KJ. Organization and mapping of a sequence on the Drosophila melanogaster X and Y chromosomes that is transcribed during spermatogenesis. Genetics. 1984;107(4):611-34.

74. Jinek M, Doudna JA. A three-dimensional view of the molecular machinery of RNA interference. Nature. 2009;457(7228):405-12.

75. Wang Y, Juranek S, Li H, Sheng G, Tuschl T, Patel DJ. Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. Nature. 2008;456(7224):921-6.

SUPPLEMENTAL FIGURES

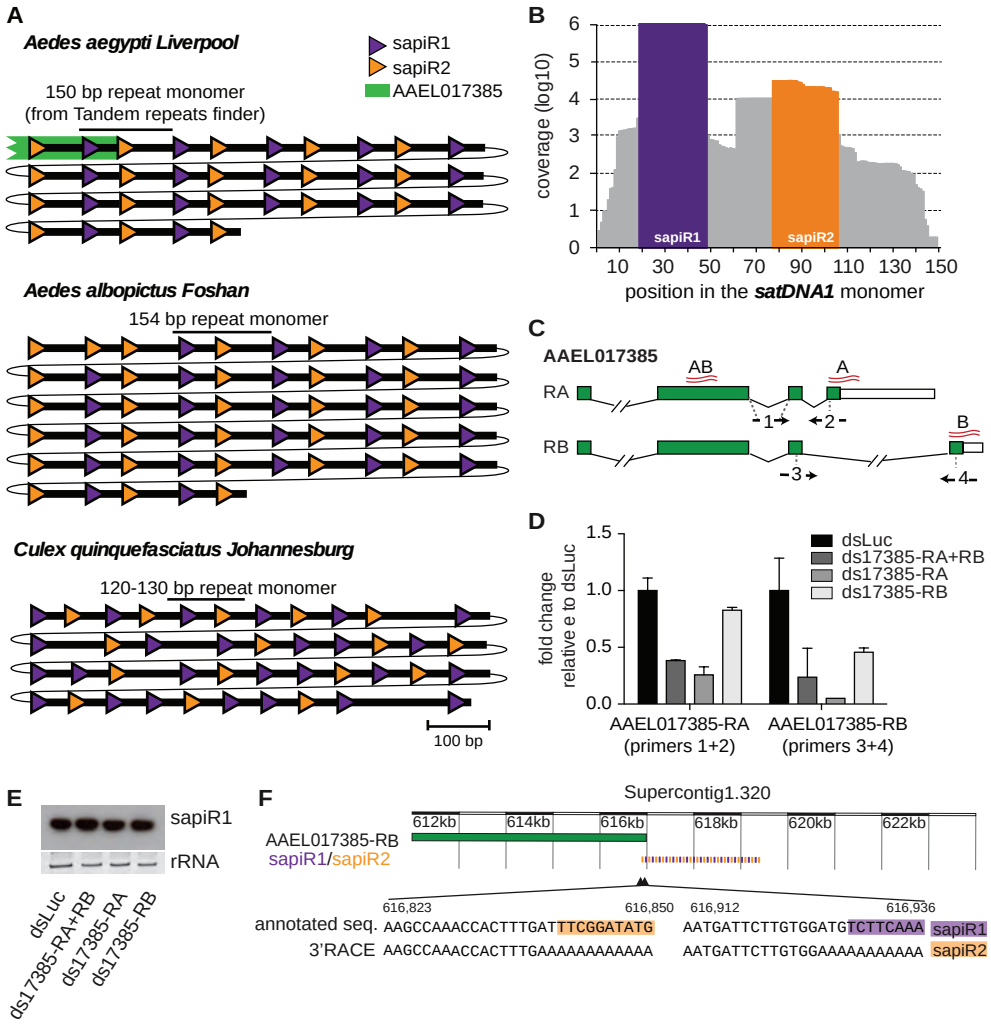


Figure S1. sapiR1 expression is independent from AAEL017385. (A) Relative positions of sapiR1 and sapiR2 sites within the *satDNA1* array of *Ae. aegypti*, *Ae. albopictus* and *Cx. quinquefasciatus*. The position of the repeat monomer is indicated. For *Ae. aegypti*, the sequence was extracted from the repeat features database on VectorBase, for *Ae. albopictus* and *Cx. quinquefasciatus*, it was deduced from the *Ae. aegypti* monomer. The position of the last part of the AAEL017385-RB 3′ UTR is indicated. **(B)** Coverage of small RNAs from Aag2 cells (combined datasets from (1)) on each position of the *satDNA1* repeat monomer. The positions of sapiR1 and sapiR2 are indicated. **(C)** Scheme of the AAEL017385 splice isoforms. The location of primers and dsRNA used in D is indicated. Primer 1 spans an exon-exon boundary. **(D)** Real-time PCR of AAEL017385 transcript isoforms after dsRNA-induced knockdown in Aag2 cells. Three different preparations of dsRNA targeting different common or unique regions of the gene were used for silencing. The dsRNA designed to only target AAEL017385-RA also effectively reduced AAEL017385-RB levels, suggesting that the annotated gene topology on VectorBase may differ from the transcript that is expressed in Aag2 cells. **(E)** Northern blot for sapiR1 in Aag2 cells 48 h post transfection of dsRNA targeting AAEL0017385. EtBr staining of ribosomal RNA (rRNA) serves as loading control. **(F)** Results of 3′ RACE. The annotated genomic sequence and the sequence obtained from Sanger-sequenced 3′ RACE products are indicated. In addition, the position of the 5′ ends of sapiR1 and sapiR2 are highlighted.

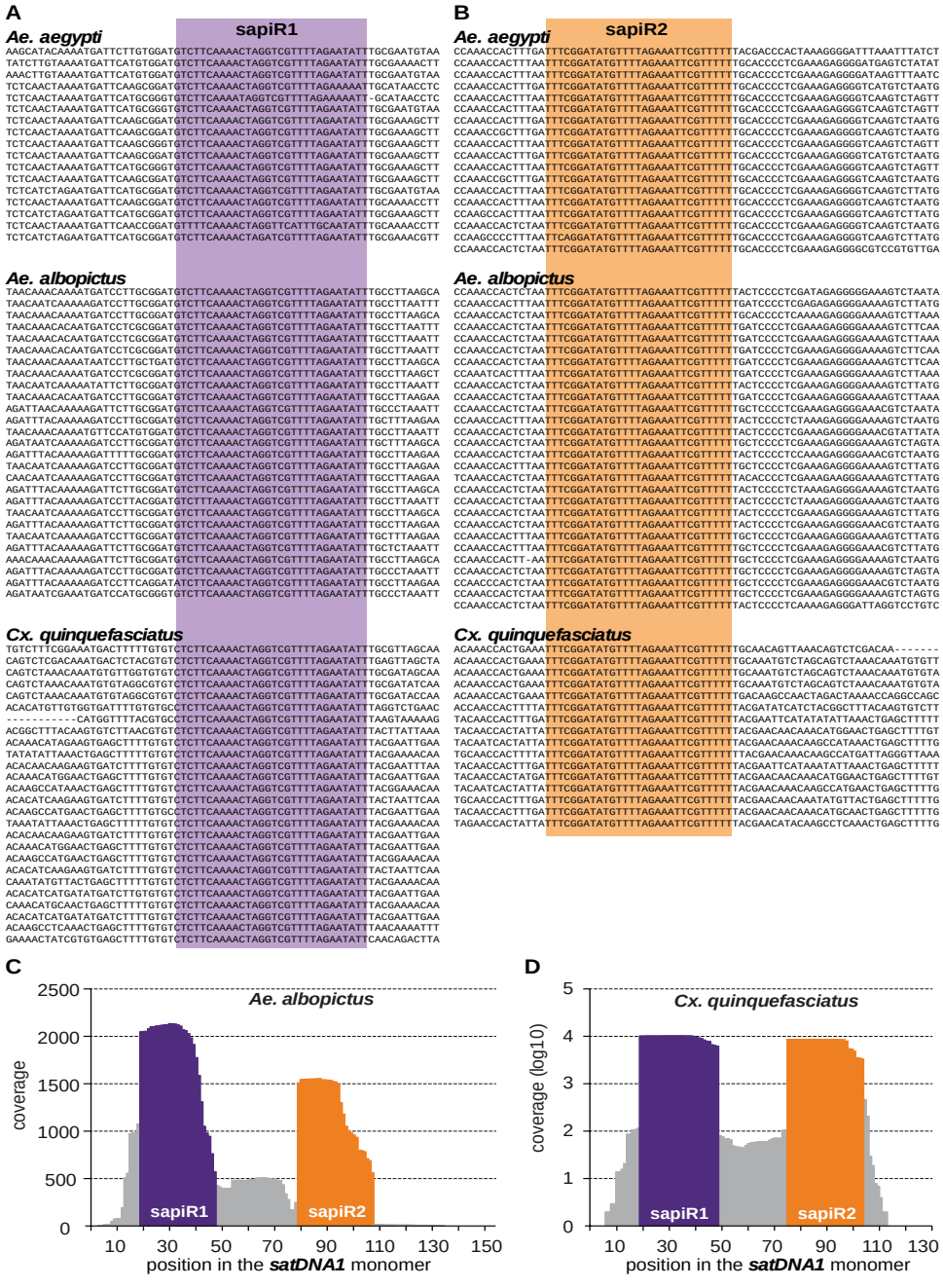


Figure S2. Ultra-Conservation of sapiR1 and sapiR2 within the *satDNA1* repeat. (A,B) Alignment of *satDNA1* repeat units flanking (A) sapiR1 and (B) sapiR2 used as input to generate the sequence logo in Figure 2A. **(C,D)** The coverage of small RNAs from (C) *Ae. albopictus* (data retrieved from SRR609263 (2)) or (D) *Cx. pipiens* (data retrieved from SRA486748 (3)) on each position of the *satDNA1* repeat monomer. It is unclear whether the heterogeneity of 3′ end of *Ae. albopictus* sapiR1/2 reflects a distinct biogenesis mechanism or is due to differences in library preparation.

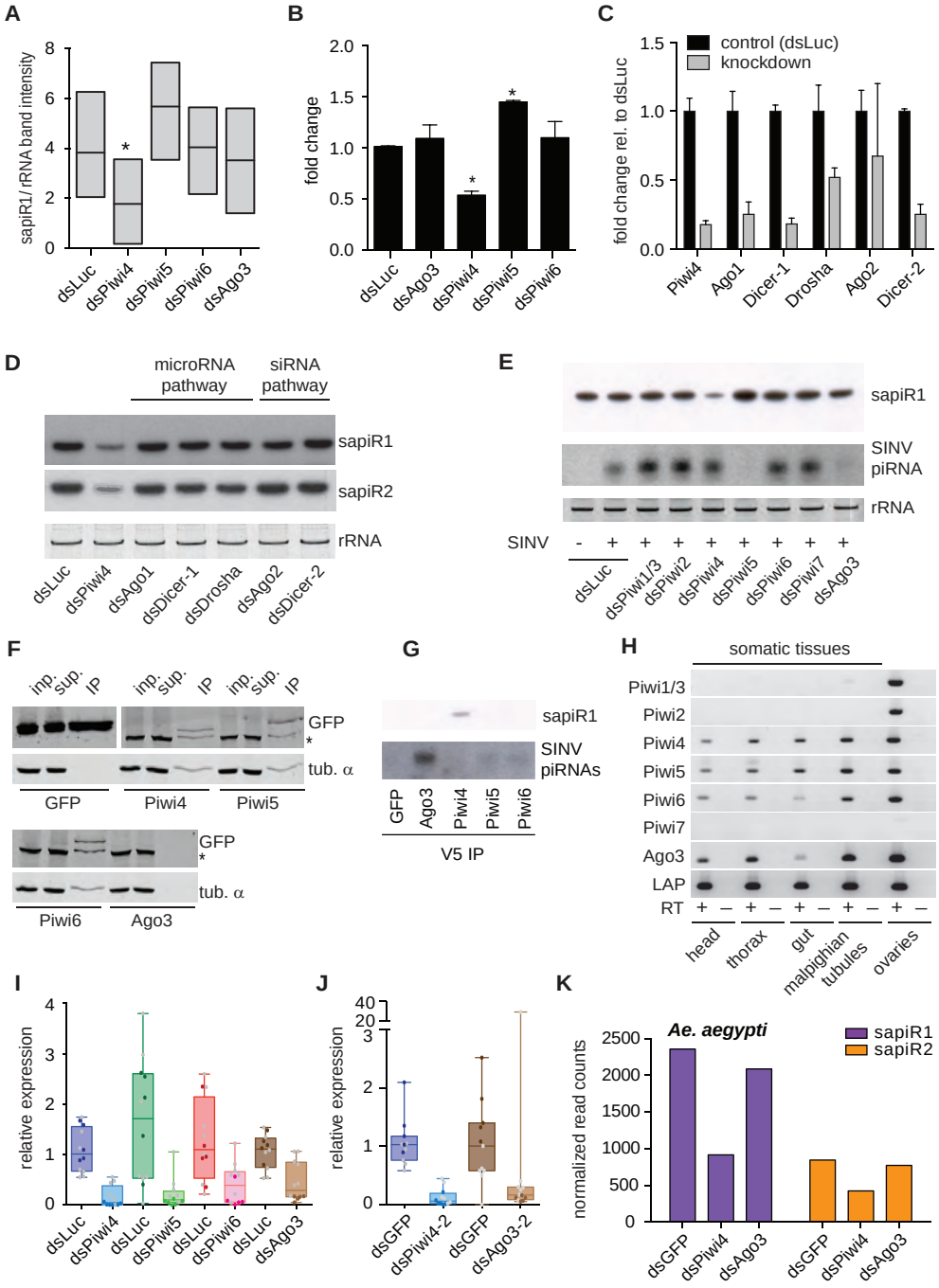


Figure S3. sapiR1/2 expression depends on Piwi4. (A) Quantification of four independent northern blots for sapiR1 including the one shown in Figure 3A. The pixel intensity (area under the curve) of the bands was determined using FIJI image analysis software. The band intensity of sapiR1 was normalized against the corresponding rRNA band. The box plot depicts the mean, and minimum

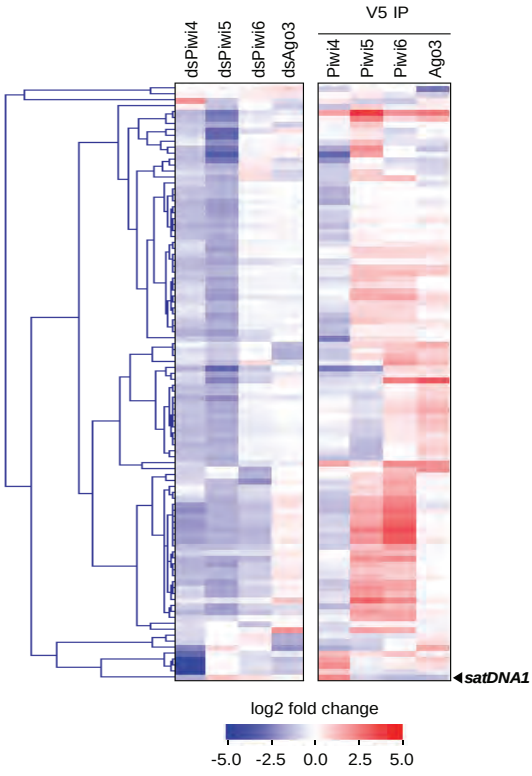


Figure S4. Piwi4-association is not a general feature of satellite DNA-derived piRNAs. For the top 100 piRNA producing tandem repeats, the heat map shows the depletion (blue) or enrichment (red) of piRNA-sized reads (25-30 nt) in the indicated PIWI knockdown libraries compared to dsLuc control libraries and in PIWI IP libraries compared to a control GFP IP library. Log2-transformed fold changes in piRNA abundance are shown.

to maximum range. Statistical significance was determined using repeated measurements one-way ANOVA with Bonferroni correction for multiple testing. * $P<0.05$ **(B)** Relative abundance of sapiR1 in Aag2 cells transfected with dsRNA targeting the indicated PIWI gene for 48 h. Expression levels were determined using stem-loop reverse transcription followed by real-time PCR. Statistical significance was determined using a One-way ANOVA comparing all PIWI knockdowns against the dsLuc control. *P*-values (* $P<0.05$) were Bonferroni corrected for multiple testing. **(C)** Knockdown efficiencies relative to dsLuc control as measured by RT-qPCR for the samples used in D. **(D)** Small RNA northern blot for sapiR1 and sapiR2 in Aag2 cells transfected with the indicated dsRNA for 48 h. **(E)** Small RNA northern blot for sapiR1 in Sindbis virus (SINV) infected Aag2 cells transfected with the indicated dsRNA. EtBr straining of rRNA serves as loading control for blots shown in panels D-E. **(F)** Western blot for GFP and tubulin α for the GFP-trap IP analyzed by small RNA northern in Figure 3C. The input (inp.) and supernatant (sup.) samples have been taken from the cell lysate prior to and after incubation with GFP-Trap beads, respectively. The IP lanes represent the protein fraction bound to the beads. The asterisk indicates a non-specific band. **(G)** Small RNA northern blot for sapiR1 in SINV-infected Aag2 cells upon V5-IP of the indicated PIWI protein. The blots in (E) and (G) are re-probed after stripping of a membrane used in a previous publication (1). The panels showing SINV piRNAs and ribosomal RNA have been displayed in this publication in Figures 2C and 4A, respectively, and they are used here as controls. **(H)** RT-PCR of PIWI transcripts in the indicated organs/parts of female *Ae. aegypti* mosquitoes. The housekeeping gene lysosomal aspartic protease (LAP) served as positive control. PCR on a no-RT control serves as negative control. Piwi1 and Piwi3 are highly similar and are detected by the same primer pairs. Piwi7 is exclusively expressed in early embryos (4) and accordingly, was not expressed in the analyzed samples. **(I)** RT-qPCR of PIWI transcripts in adult female *Ae. aegypti* mosquitoes 5 days after intrathoracal injection of dsRNA to induce gene silencing of the indicated PIWI genes. The expression levels in 5 or 6 mosquitoes selected to prepare tsmall RNA deep-sequencing libraries analyzed in Figure 3F are indicated by the colored dots. **(J)** Same as I for mosquitoes injected with a different set of dsRNAs against the indicated transcripts. **(K)** sapiR1/2 read counts in deep-sequencing libraries prepared from adult *Ae. aegypti* females shown in J.

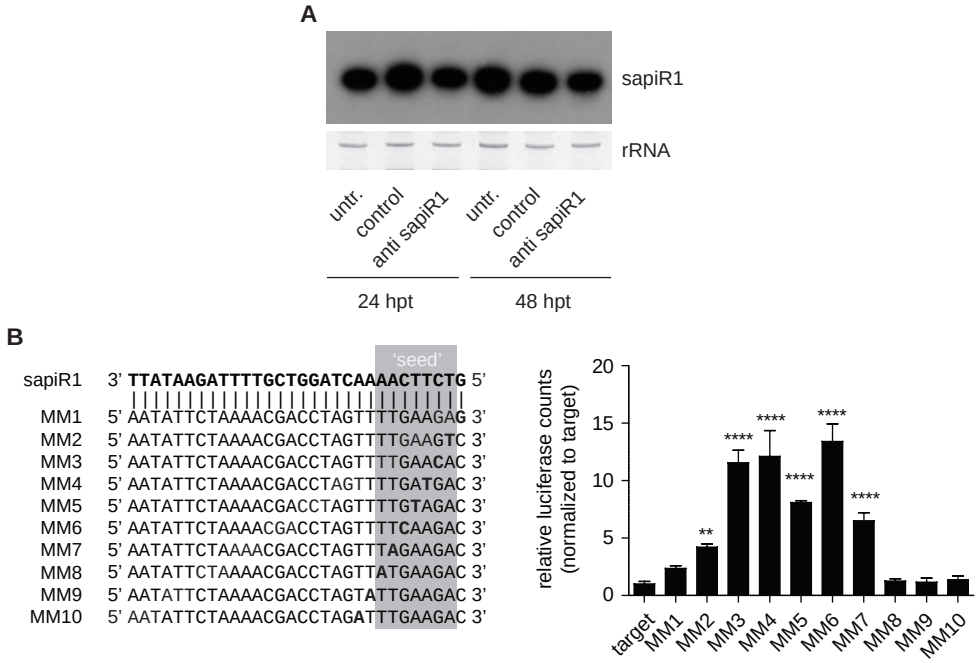


Figure S5. sapiR1 gene silencing is mediated by a seed-sequence. (A) Small RNA northern blot of sapiR1 in Aag2 cells left untreated or transfected with a nuclease-resistant RNA oligo antisense to sapiR or with a control sequence. Expression of sapiR has been analyzed at 24 h and 48 h post transfection. EtBr staining of ribosomal RNA (rRNA) served as loading control. **(B)** Left panel: illustration of the mutagenesis strategy for the sapiR1 reporter; the 'seed' sequence of sapiR1 is indicated. Right panel: firefly/*Renilla* luciferase counts in the lysate of Aag2 cells transfected with the indicated reporter plasmids. Statistical significance of expression changes compared to the target construct was determined using one-way ANOVA with Bonferroni correction for multiple testing. ** $P<0.01$, **** $P<0,001$.

Table S1. Oligonucleotides used in this study

Cloning	
s-sapiR1-3'target*	AATGACCAATATTCTAAAACGACCTAGTTTTGAAGACCCGCGGCTGCGTTT
as-sapiR1-3'target*	AAACGCAGCCGCGGGTCTTCAAAACTAGGTCGTTTTAGAATATTGGTCATTGC
s-hp-swap-3'	AATGACCAATACAAAATCTTGACCTAAAGTTTTGGACCCGCGGCTGCGTTT
as-hp-swap-3'	AAACGCAGCCGCGGGTCCAAAACTTTAGGTCAAGATTTTGTATTGGTCATTGC
s-scambled-3'	AATGACCATAGTATTACTAGAATCTAGATATCACGCACCGCGGCTGCGTTT
as-scrambled-3'	AAACGCAGCCGCGGTGCGTGATATCTAGATTCTAGTAATACTATGGTCATTGC
s-sapiR1-5'target*	GGCCGCACGGATCCGCAATATTCTAAAACGACCTAGTTTTGAAGACC
as-sapiR1-5'target*	TCGAGGTCTTCAAAACTAGGTCGTTTTAGAATATTGCGGATCCGTGC
s-scambled-5'	GGCCGCACGGATCCGCATAGTATTACTAGAATCTAGATATCACGCAC
as-scrambled-5'	TCGAGTGCGTGATATCTAGATTCTAGTAATACTATGCGGATCCGTGC
s-sapiR1-5'MM4	GGCCGCACGGATCCGCAATATTCTAAAACGACCTAGTTTTGATGACC
as-sapiR1-5'MM4	TCGAGGTCATCAAAACTAGGTCGTTTTAGAATATTGCGGATCCGTGC
s-sapiR1-ORFtarget*	TCGAGACCATGGAAGACGCCAAAAACATAAAGAAAGGCCCGGCGCCATTCTATAATATTCTAAAACGACCTAG TTTTGAAGACCT
as-sapiR1-ORFtarget*	CATGAGGTCTTCAAAACTAGGTCGTTTTAGAATATTATAGAATGGCGCCGGGCCTTTCTTTATGTTTTTGGCG TCTTCCATGGTC
s-scrambled-ORF*	TCGAGACCATGGAAGACGCCAAAAACATAAAGAAAGGCCCGGCGCCATTCTATATAGTATTACTAGAATCTAG ATATCACGCACT
as-scrambled-ORF*	CATGAGTGCGTGATATCTAGATTCTAGTAATACTATATAGAATGGCGCCGGGCCTTTCTTTATGTTTTTGGCG TCTTCCATGGTC
s-sapiR1-ORF-MM4	TCGAGACCATGGAAGACGCCAAAAACATAAAGAAAGGCCCGGCGCCATTCTATAATATTCTAAAACGACCTAG TTTTGATGACCT
as-sapiR1-ORF-MM4	CATGAGGTCATCAAAACTAGGTCGTTTTAGAATATTATAGAATGGCGCCGGGCCTTTCTTTATGTTTTTGGCG TCTTCCATGGTC
Fw_attB1_AaePIWI4_N	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCTCTGACCGTTACTCTCAAGG
Rv_attB2_AaePIWI4_unt-N	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTATTATTACAAGAAGTACAGCTTC
Fw_attB1_AaePIWI5_N	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGCGGATAGACAGCAAGGAGG
Rv_attB2_AaePIWI5_unt-N	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTATTATTACAGATAATAGAGTTTCTTTTCC
Fw_attB1_AaePIWI6_N	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGCTGATAATCCACAGGAAGG
Rv_attB2_AaePIWI6_unt-N	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTATTACTACAAAAAGTAAAGTTTCTTCTCC
Fw_attB1_AaeAGO3_N	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCTCCTCGCGGTTGAATTTAGTTCG
Rv_attB2_AaeAGO3_unt-N	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTATTATCACAGGTAGAACAGTTTGTCG
*Sequences of mutant sapiR1 target site in the 3' UTR are derived from these sequences	
dsRNA production	
F-T7-Luc	TAATACGACTCACTATAGGGAGATATGAAGAGATACGCCCTGGTT
R-T7-Luc	TAATACGACTCACTATAGGGAGATAAAACCGGGAGGTAGATGAGA
F-T7-GFP	TAATACGACTCACTATAGGGAGAAGCTGACCCTGAAGTTCATCTG
R-T7-GFP	TAATACGACTCACTATAGGGAGAGGTGTTCTGCTGGTAGTGGTC
F-T7-Piwi1/3	TAATACGACTCACTATAGGGAGACCACGCCCATCGTTTCAA
R-T7-Piwi1/3	TAATACGACTCACTATAGGGAGACCTCAGTTTGTTCACCATA
F-T7-Piwi2	TAATACGACTCACTATAGGGAGACCGTCCTACTTTCCAGCAC
R-T7-Piwi2	TAATACGACTCACTATAGGGAGAGCGGCACTCCAGGGACAAT
F-T7-Piwi4	TAATACGACTCACTATAGGGAGACGTGGAAGTCCTTCTTCTCG
R-T7-Piwi4	TAATACGACTCACTATAGGGAGATGTCAGTTGATCGCTTCTCAA
F-T7-Piwi5	TAATACGACTCACTATAGGGAGAGCCATACATCGGGTCAAAAT
R-T7-Piwi5	TAATACGACTCACTATAGGGAGACTCTCCACCGAAGGATTGAA
F-T7-Piwi6	TAATACGACTCACTATAGGGAGACAACGGAGGATCTTCACGAG
R-T7-Piwi6	TAATACGACTCACTATAGGGAGAAATCGATGGCTTGATTTGGA
F-T7-Piwi7	TAATACGACTCACTATAGGGAGAGTGGAGGTCGTGGAGGTAAC
R-T7-Piwi7	TAATACGACTCACTATAGGGAGAGTTTGCGGTGTTTCCGTACT
F-T7-Ago3	TAATACGACTCACTATAGGGAGATGCTTACTCGTGTCGCGTAG
R-T7-Ago3	TAATACGACTCACTATAGGGAGAGGCATGGCAGATCCAATACT
F-UT-Piwi4-set2**	GCCCGACGCTACCAGCTGCGCATTGTC
R-UT-Piwi4-set2**	CGCCTCGGCGACGTTTTACCGGCCTTG

F-T7-Ago3-set2	TAATACGACTCACTATAGGGAGAGACCGACGAGATGCGAAG
R-T7-Ago3-set2	TAATACGACTCACTATAGGGAGAACAGCTCGGTTTTGTCGG
F-T7-Ago1	TAATACGACTCACTATAGGGAGACCGGTCATCGAGTTCATGT
R-T7-Ago1	TAATACGACTCACTATAGGGAGACGTGGCTTTGATCATGGTT
F-T7-Dicer1	TAATACGACTCACTATAGGGAGAGCAGTTGAAATGCC
R-T7-Dicer1	TAATACGACTCACTATAGGGAGATCCGAAACGTACGA
F-T7-Drosha	TAATACGACTCACTATAGGGAGATGTCTCGGTTCGGCAAGG
R-T7-Drosha	TAATACGACTCACTATAGGGAGACAACGGGTGCGGTGGA
F-T7-Ago2	TAATACGACTCACTATAGGGAGACTACGAGCAGGAGGTCAAGG
R-T7-Ago2	TAATACGACTCACTATAGGGAGATCCATGCCTTTGAGGAAATC
F-T7-Dicer2	TAATACGACTCACTATAGGGAGATCCTATGCACGGGATTATGG
R-T7-Dicer2	TAATACGACTCACTATAGGGAGAATTGATCCCCCAAAAAGACC
F-UT-AAEL017385-AB**	GCCCGACGCTGAAAACGGCAGACACCA
F-UT-AAEL017385-AB**	CGCCTCGGCTTGGAGCACCTCCGTAGC
F-UT-AAEL017385-A**	GCCCGACGCGCAAGCCTACTCGCAAGG
F-UT-AAEL017385-A**	CGCCTCGGCTTCTCGGGATGCTTTTGG
F-UT-AAEL017385-B**	GCCCGACGCTTGGAATCCCGTCGGATA
F-UT-AAEL017385-B**	CGCCTCGGCAATCCCCTTTAGTGGGTCGT
**PCR products generated with primers harboring a universal tag (UT) were used as inputs for a second PCR with universal primers introducing the T7 sequence: F-T7-UT: TAATACGACTCACTATAGGGAGAGCCCGACGC; R-T7-UT: TAATACGACTCACTATAGGGAGACGCCTCGGC	
Nuclease resistant antisense oligonucleotides	
Anti-sapiR1	[mA][mA][mU][mA][mU][mU][mC][mU][mA][mA][mA][mA][mC][mG][mA][mC][mC][mU][mA][mG][mU][mU][mU][mU][mG][mA][mA][mG][mA][mC]
Anti-Luc	[mU][mG][mG][mA][mC][mA][mU][mC][mA][mC][mU][mU][mA][mC][mG][mC][mU][mG][mA][mG][mU][mA][mC][mU][mU][mC][mG][mA]
Northern blotting	
n-sapiR1	AATATTCTAAAACGACCTAGTTTTGAAGAC
n-sapiR2	AAAAAACGAATTTCTAAAACATATCCGAAA
n-miR-2940-3p	AGTGATTTATCTCCCTGTCGAC
n-U6	GATTTTGCGTGTCATCCTTGTGCAGGGGCCATGCTAA
(q)PCR	
F-LAP	GTGCTCATTCACCAACATCG
R-LAP	AACTTGGCCGCAACAAATAC
F-RPL5	TCACCTGCCAGATTGCGTACGCCCG
R-RPL5	GCTTCTGCAGGATGCGGCGGGCAA
F-Piwi1/3***	GGCCGTTAGCGAGTCTCAT
R-Piwi1/3	GGCAGAACCTTCGTGGTAAG
R-Piwi1/3***	AACTTGATAGACCTCCCCGG
F-Piwi2***	CCGCGGGTACACCGCCGTCAACTT
R-Piwi2	CGCTGGTCGAACTCGATGCCCCGC
R-Piwi2***	TGATCTGGTACACCTCCCCG
F-Piwi4	TCTTCTTCTCCACCACAGCC
R-Piwi4	ATGGTGACCACCTCACAGTTAC
F-Piwi4***	CGTTACTCTCAAGGGCGCTA
R-Piwi4***	ACGTTCGTGTCTGGATTGCA
F-Piwi5	ACGGCATCACATCGAGACTC
R-Piwi5	CGACCTCCACGCTGTCCTC
F-Piwi5***	GGAAATGGTGGAAATGGCGG
R-Piwi5***	ACCAATGGTGCCCTTCAGAT
F-Piwi6	TTTTCTTCCACCCCGAGCAG
R-Piwi6	AATACATTTGCGATGCGGCC
F-Piwi6***	TAATCCACAGGAAGGCTCCA
R-Piwi6***	TCCATCGAACACATACCCGC
F-Piwi7	ATGCGACGAAACTTCAACTTG
R-Piwi7	CCAGCAGCAACCGCATAATT
F-Ago3	CTCCAGACGACGGTTTTGGA

R-Ago3	GCAGGTACGAAATTGGCTGC
F-Ago3***	TCTGCTTACTCGTGTCGCGTAGTGCTCGT
R-Ago3***	ACGCGCGAACTAAATTCAACCGCGAGGA
F-Ago1	CGAACAGCATGATGGAAGTG
R-Ago1	AAATTGTTTGCCTCGCATGT
F-Dicer1	AATGCGATGAGGCGGAAGAT
R-Dicer1	ACTGTTCGATTTCAGGTCCCA
F-Drosha	TGCCGATGTTCAGCAAGATC
R-Drosha	GTCGGTGGTGAAATGGTTATCC
F-Ago2	CAGTTGCAAGCGCTGACTTA
R-Ago2	ATCTCGGTCCTGATCTGCAT
F-Dicer2	TGAAGAGGAGTTGCGAAGGT
R-Dicer2	AAATGCATCTCTCGGCATTC
F-AAEL017385-RA	TACAGCACCGGAACATACGA
R-AAEL017385-RA	CACCATGAGCGTACTGATCG
F-AAEL017385-RB	ACCTATCCCAGCCAGCAGTA
R-AAEL017385-RB	CCGGTAAAATCGCAAAGTGT
*** used for evaluation of knockdown efficiencies in adult mosquitoes	
Stemloop RT-qPCR	
sapiR1 stemloop RT primer	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACAATATT
F-sapiR1-sRNAqPCR	GCCCGCGTCTTCAAAACTAGGTC
bantam stemloop RT primer	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGCACTGGATACGACATCAGC
F-bantam-sRNAqPCR	GCCCGCTGAGATCATTTTGAAAG
R-univ-sRNAqPCR	GTGCAGGGTCCGAGGT
3' RACE	
AAEL017385 outer	TACAGCACCGGAACATACGA (same as F-AAEL017385-RA)
AAEL017385 inner	TGAGGATACCACTTTCGCCG

SUPPLEMENTAL REFERENCES

1. Miesen P, Girardi E, van Rij RP. Distinct sets of PIWI proteins produce arbovirus and transposon-derived piRNAs in Aedes aegypti mosquito cells. Nucleic acids research. 2015;43(13):6545-56.
2. Gu J, Hu W, Wu J, Zheng P, Chen M, James AA, et al. miRNA genes of an invasive vector mosquito, Aedes albopictus. PloS one. 2013;8(7):e67638.
3. Miesen P, Joosten J, van Rij RP. PIWIs Go Viral: Arbovirus-Derived piRNAs in Vector Mosquitoes. PLoS pathogens. 2016;12(12):e1006017.
4. Akbari OS, Antoshechkin I, Amrhein H, Williams B, Diloreto R, Sandler J, et al. The developmental transcriptome of the mosquito Aedes aegypti, an invasive species and major arbovirus vector. G3. 2013;3(9):1493-509.

Chapter 7

The TUDOR Protein AAEL012441 is Required for Efficient Ping-pong Amplification of Viral piRNAs in *Aedes aegypti*.

Joep Joosten, Pascal Miesen, Bas Pennings, Pascal Jansen,
Martijn Huynen, Michiel Vermeulen, Ronald P. Van Rij

manuscript in preparation

ABSTRACT

Biogenesis of PIWI interacting RNA (piRNAs) in *Drosophila* and other model organisms relies on a myriad of co-factors, many of which belong to the Tudor protein family. These auxiliary proteins prevent non-specific degradation of piRNA substrates, serve as scaffold for PIWI proteins, and aid in loading small RNAs onto specific PIWI proteins. In the vector mosquito *Aedes aegypti*, a somatic piRNA pathway is active which generates piRNAs *de novo* from cytoplasmic RNA viruses during an acute infection. In *Ae. aegypti* Aag2 cells, piRNA biogenesis from Sindbis virus requires ping-pong amplification by the PIWI proteins Piwi5 and Ago3. Yet, accessory proteins in the somatic viral piRNA (vpiRNA) pathway in *Aedes* are unknown. We hypothesized that Tudor proteins are involved in this process and performed an RNAi screen targeting all *Aedes* TUDOR-domain containing genes in Aag2 cells. Knockdown of several Tudor proteins resulted in reduced (vpiRNA) accumulation with silencing of AAEL012441 having the most robust effect. This Tudor protein localizes in peri-nuclear foci, similar to piRNA processing granules of *Drosophila* and it directly interacts with Ago3. Interestingly, mass-spectrometry analysis of AAEL012441-interacting proteins reveals a network of additional co-factors including orthologs of the *Drosophila* piRNA pathway components Vasa (AAEL004978) and Yb (AAEL001939). Yb in turn interacts with Piwi5 and we propose that this multi-protein complex provides a molecular scaffold that allows efficient ping-pong amplification of vpiRNAs.

INTRODUCTION

In animals, three distinct small RNA-mediated silencing pathways exist: the micro (mi) RNA, small interfering (si)RNA and PIWI-interacting (pi)RNA pathway (1). In all three, a small RNA molecule provides sequence specificity to guide a member of the Argonaute protein family to target RNAs. Whereas miRNAs and siRNAs associate with proteins of the AGO clade of this family, piRNAs are loaded onto PIWI clade proteins exclusively, forming piRNA induced silencing complexes (piRISCs) (2).

The piRNA pathway is known for its role in transgenerational protection of genome integrity by silencing transposable elements in the germline (3, 4). Despite ubiquitous expression of piRNAs across metazoans, our knowledge on the molecular mechanisms that govern the piRNA pathway is limited to only a small number of model organisms (5). In the *Drosophila melanogaster* germline, single stranded piRNA precursors are produced from genomic piRNA clusters that contain transposon remnants (6). These precursors leave the nucleus and are processed to give rise to a pool of primary piRNAs. Preferentially piRNAs that are antisense towards transposon mRNAs and that bear a uridine at the first nucleotide position (1U) are loaded onto the PIWI proteins Piwi and Aubergine (Aub) (6-8). Piwi enter the nucleus to enforce transcriptional silencing, whereas

Aub piRISCs target and cleave cognate transposon RNA in an electron-dense perinuclear structure termed *nuage* (3, 9). The resulting 3′-cleavage fragment is subsequently loaded onto the PIWI protein Ago3 and processed further into a mature secondary piRNA, primarily of sense orientation (6, 7). The resulting Ago3-piRISC can target and cleave antisense precursor transcripts to produce a new Aub-bound piRNA, completing the so-called ping-pong amplification cycle. Aub cleaves target RNA between nucleotides 10 and 11. Therefore, the 5′ ends of corresponding Aub-bound antisense piRNAs and Ago3-bound sense piRNAs overlap for 10 nucleotides. In addition, Ago3-bound piRNAs predominantly have adenosine residues at the tenth position (10A). Collectively, the overlap of 5′ ends and the 1U/10A nucleotide biases are the hallmarks of piRNA amplification by the ping-pong cycle and are referred to as ping-pong signature (6, 7).

Ping-pong amplification of piRNAs was previously thought to be restricted to germline tissues, but recently, ping-pong dependent piRNA production has been demonstrated in somatic tissues of hematophagous mosquitoes of the *Aedes* family (10). These anthropophilic vector mosquitoes, primarily *Ae. aegypti* and *Ae. albopictus*, are crucial for the distribution of several arthropod-borne (arbo)viruses that can cause debilitating diseases such as dengue, chikungunya and Zika (11). Since arboviral infectivity is greatly affected by the ability of the virus to replicate in the vector, mosquito antiviral immunity is a key determinant for virus transmission. Intriguingly, while causing severe disease in vertebrate hosts, arboviruses are able to replicate to high levels in the mosquito without causing apparent pathology (12). An efficient immune response based on small silencing RNA pathways at least party contributes to this tolerance as genetic interference with production of viral siRNAs cause elevated virus replication accompanied by enhanced mosquito mortality (13-16) The siRNA pathway processes viral double-stranded (ds) RNA into 21 nt long viral siRNAs. These siRNAs guide the endonuclease Argonaute 2 (Ago2) to cognate viral mRNA, which is subsequently degraded (17).

Besides siRNAs, arbovirus infection results in *de novo* production of virus-derived piRNAs (vpiRNAs) in aedine mosquitoes and cell lines, indicating that two independent small RNA pathways may contribute to antiviral immunity (10). In *Aedes aegypti* cells, vpiRNAs from the model alphavirus Sindbis virus (SINV, family *Togaviridae*) are predominantly produced in a ping-pong amplification loop involving the PIWI proteins Piwi5 and Ago3 (18). These proteins associate directly with vpiRNAs, which bear the distinct 1U/10A nucleotide signature indicative of ping-pong amplification. The further composition of the protein complexes that mediate vpiRNA biogenesis is currently unknown. Moreover, it is unclear whether vpiRNA require dedicated complexes that differ from those that mediate biogenesis of transposon-derived piRNAs.

Studies in *D. melanogaster* and other model organisms have shown that TUDOR domain-containing proteins (Tudor proteins) are essential for multiple steps of piRNA biogenesis (3, 4, 19, 20). TUDOR domains contain conserved motifs organized in an

aromatic cage that is known to interact with symmetrically dimethylated arginines (sDMAs), a common post-translational modification on PIWI proteins (21-23). Consequently, Tudor proteins may serve as adapter molecules facilitating the assembly of multi-molecular complexes involved in vpiRNA biogenesis.

In order to further our understanding of the molecular mechanisms underlying vpiRNA biogenesis, we performed a functional RNAi screen of all predicted *Ae. aegypti* Tudor proteins. This screen demonstrates that several Tudor proteins are involved in biogenesis of vpiRNAs in *Ae. aegypti.* Of these, knockdown of the hitherto uncharacterized Tudor protein AAEL012441 shows the most prominent vpiRNA depletion. The protein resides in perinuclear foci, reminiscent of piRNA processing granules in *Drosophila*. Intriguingly, mainly sense (+) strand derived vpiRNAs are depleted upon knockdown of AAEL012441, which is suggestive of an impaired ping-pong amplification loop. In line with this, AAEL012441 forms a complex with Ago3 through canonical TUDOR-domain mediated recognition. In addition, the RNA helicase AAEL004978 and the Tudor protein AAEL001939, orthologs of the piRNA pathway components Vasa and Yb (9, 24-29), respectively, directly bind to AAEL012441. AAEL001939 in turn associates with Piwi5 and we propose that this complex provides a molecular platform that supports efficient ping-pong amplification.

MATERIALS AND METHODS

Tudor protein identification and ortholog detection

To allow a comprehensive identification of all *Ae. aegypti* Tudor proteins, we combined HHpred homology detection with Jackhmmer iterative searches (30, 31). Subsequently, we used T-Coffee to generate multiple sequence alignments to determine orthologous relations between *Ae. aegypti* and *D. melanogaster* Tudor proteins (32). See Supplemental Materials and Methods for a detailed description of our approach.

Transfection and infection of Aag2-cells

For immunoprecipitation (IP) and immunofluorescence (IFA) experiments, Aag2 cells were transfected with expression plasmids encoding peptide-tagged transgenes as indicated. Samples were harvested 48 hours after transfection. In knockdown experiments, cells were transfected with dsRNA and re-transfected 48 hours later to ensure prolonged knockdown. Where indicated, cells were infected with a recombinant Sindbis virus (SINV; pTE3'2J-GFP (33)) at a multiplicity of infection (MOI) of 1 and harvested 48 hours post infection. For a more detailed description of cell culture conditions, generation of expression vectors, and virus production, see Supplemental Materials and Methods.

Small RNA northern blotting and RT-qPCR

For small RNA northern blotting, RNA was size separated on polyacrylamide gels and cross-linked to nylon membranes using 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (34). Small RNAs were detected with ^{32}P-labelled DNA oligonucleotides. For quantitative RT-PCR analyses, DNaseI-treated RNA was reverse transcribed and PCR amplified in the presence of SYBR green. See Supplemental Materials and Methods for a detailed description of the experimental procedures, sequences of probes used for northern blotting and qPCR primers.

Preparation of small RNA libraries and bioinformatic analyses

Total RNA from Aag2 cells transfected with dsRNA targeting either AAEL012441 or luciferase was used to generate small RNA deep sequencing libraries. For each condition, three transfections and library preparations were performed in parallel using Illumina's Truseq technology, as described in (35). Deep sequencing libraries were sequenced on an Illumina Hiseq2500 machine by Baseclear (Leiden, The Netherlands). FASTQ files were generated by Casava1.8.2 and initial quality control was performed using in-house filter programs (Baseclear) and FastQC version 0.10.0. Subsequent manipulations were performed in Galaxy (36). First, the 3′ adapters were clipped from the small RNA sequence reads using the FASTX Clip adapter software. Reads were then mapped to the SINV-GFP virus genome, to transposable elements sequences available on TEfam (http://tefam.biochem.vt.edu; downloaded 12/16/2016) or *Ae. aegypti* transcripts using Bowtie allowing one mismatch in a 28 bp seed (37). Size profiles of indicated small RNA populations were generated and the counts of siRNAs (21 nt reads) or piRNAs (25-30 nt reads) derived from either the sense or antisense strand of the genome were plotted.

Fluorescence and confocal imaging

Fluorescent imaging was performed on paraformaldehyde-fixed Aag2-cells that were permeabilized and counterstained using Hoechst-solution. All slides were imaged using the Olympus FV1000 confocal microscope and images were processed using FIJI. See Supplemental Materials and Methods for more details on the experimental approach.

Immunoprecipitation, western blotting and mass spectrometry analyses

GFP- and RFP-tagged transgenes were immunoprecipitated using GFP- and RFP-TRAP beads (Chromotek), respectively, according to manufacturer's instructions. V5-transgenes were purified using V5-agarose beads (Sigma). Anti-Ago3 antibodies were raised against a mix of two selected peptides (TSGADSSESDDKQSS and IIYKRKQRMSENIQF) by immunization of two rabbits (both rabbits received both peptides). After an initial immunization and 3 boost immunizations at t = 14 days, t = 28 days and t = 56 days final bleeds were collected at t = 87 days and pooled. Antibodies were purified against each

peptide separately and specificity was determined by Western blot analysis. Anti-PIWI5 antibodies were obtained using the same protocol (DIVRSRPLDSKVVKQ and CANQGGNWRDNYKRAI as immunizing peptides). These antibodies were chemically crosslinked to Protein A/G PLUS agarose beads (Santa Cruz) using dimethyl pimelimidate (DMP; Sigma D8388) for purification of Ago3 and Piwi5 complexes. Protein extracts were resolved on polyacrylamide gels, blotted to nitrocellulose membranes and probed with indicated antibodies.

For mass spectrometry analysis, precipitated proteins were washed extensively and subjected to on-bead trypsin digestion as described in (38). Subsequently, tryptic peptides were acidified and desalted using Stagetips (39) before elution onto a NanoLC-MS/MS. Mass spectra were recorded on a QExactive mass spectrometer (Thermo Scientific). For details on the experimental procedures and the analyses of mass spectra, see Supplemental Materials and Methods.

RESULTS

Comprehensive identification of Tudor proteins in *Aedes aegypti*

Tudor proteins play fundamental roles in the biogenesis of piRNAs in both vertebrate and invertebrate species (19, 20). We therefore hypothesized that vpiRNA production in *Ae. aegypti* also relies on members of this protein family. To faithfully identify all *Ae. aegypti* Tudor proteins and their corresponding fly orthologs, we used a homology-based prediction approach by combining HHPred and Jackhmmer algorithms (30, 31). Ultimately, a neighbor joining clustering was generated (32), which enabled identification of orthologous relationships between *Ae. aegypti* and *D. melanogaster* Tudor proteins (Figure 1).

Our analyses reveal that the majority of *Ae. aegypti* Tudor proteins cluster with a single *D. melanogaster* ortholog. Some however (AAEL008834, AAEL008112, AAEL008101, AAEL009987) have no clear one-to-one ortholog, suggesting that these proteins emerged as a result of duplication events that occurred after divergence of the *Drosophilidae* and *Culicidae* families. Similarly, CG15042 and Krimper are paralogous in *D. melanogaster* proteins that lack clear *Ae. aegypti* counterparts. Furthermore, the *Ae. aegypti* genome encodes only one ortholog, AAEL001939, for the *D. melanogaster* Yb protein subfamily (Yb, SoYb and BoYb). Intriguingly, TUDOR domain sequences alone are sufficient to predict clusters of orthologous proteins with remarkably similar domain compositions.

Several Tudor proteins are involved in vpiRNA biogenesis

We included all *Ae. aegypti* Tudor proteins in a functional RNAi screen and also added AAEL004290, which is the ortholog of *Drosophila* dSETDB1/eggless. This methyltransferase is involved in the piRNA pathway in *D. melanogaster* and is predicted to contain TUDOR domains (40, 41), although it did not surface in our HHpred-based homology detection.

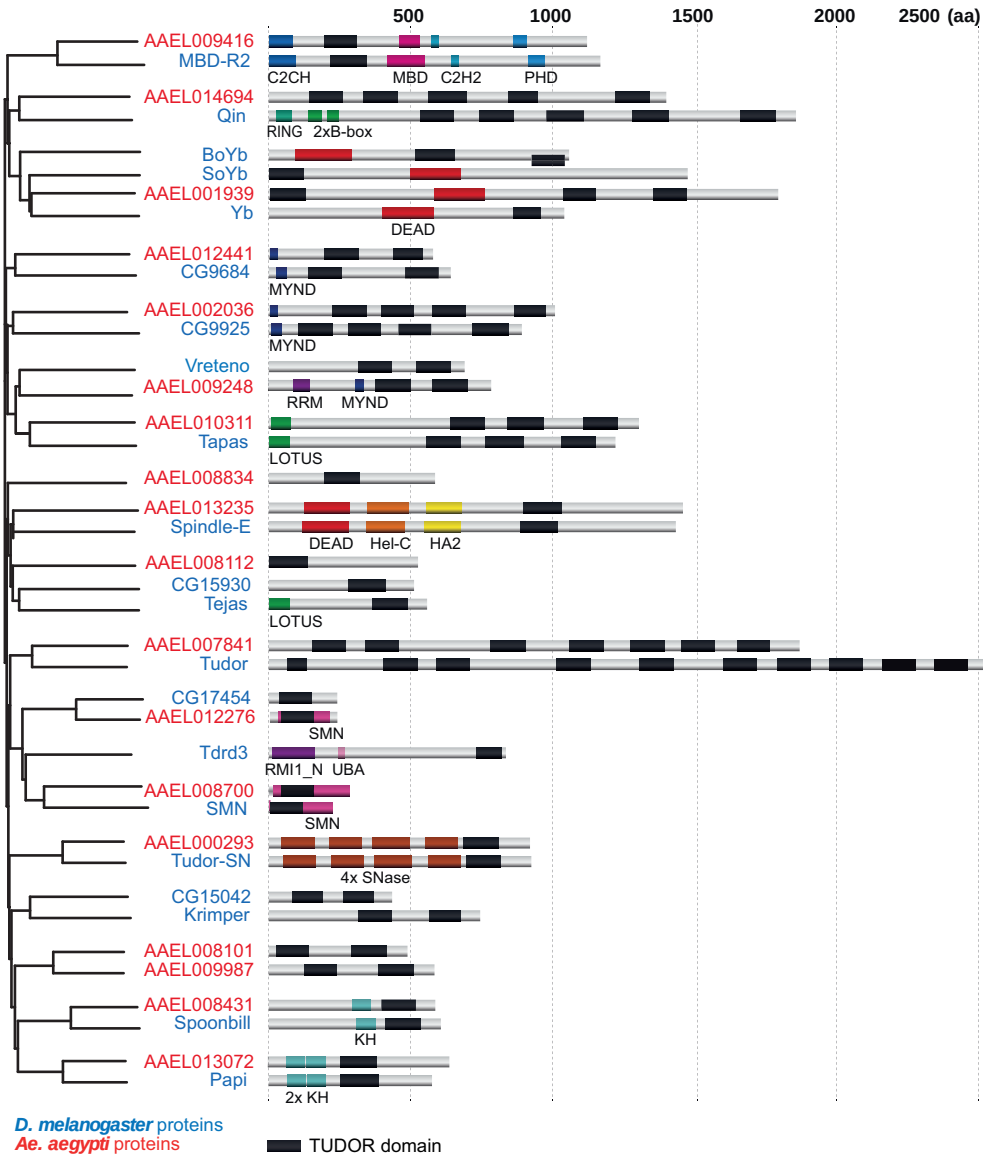


Figure 1. Orthologous Tudor proteins in *D. melanogaster* and *Ae. Aegypti* largely have a conserved domain structure. Ortholog detection using all predicted *Ae. aegypti* (red) and *D. melanogaster* (blue) Tudor proteins. On the left, the neighbor-joining tree generated from the multiple sequence alignment of the TUDOR domains is depicted. On the right, domain structures of analyzed Tudor proteins are drawn schematically. Numbers at the top indicate protein length in amino acids (aa). B-box, B-box type zinc finger (Zf) domain; C2CH, C2CH-type Zf domain; C2H2, C2H2-type Zf domain; DEAD, DEAD box domain; HA2, Helicase-associated domain; Hel-C, helicase C domain; KH, K homology RNA-binding domain; LOTUS, OST-HTH/LOTUS domain; MBD, Methyl-CpG-binding domain; MYND, MYND (myeloid, Nervy, DEAF-1)-type Zf domain; PHD, PHD-type Zf domain; RING, RING-type Zf domain; RMI1_N, RecQ mediated genome instability domain; RRM, RNA recognition motif; SMN, survival motor neuron domain; SNase¸ Staphylococcal nuclease homologue domain; UBA, ubiquitin associated domain.

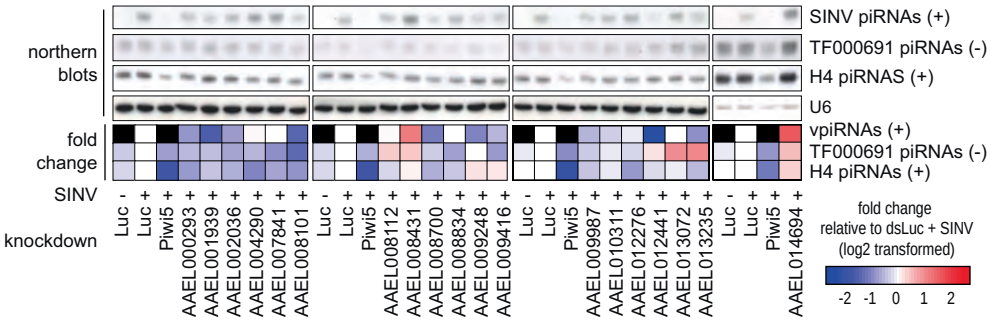


Figure 2. Knockdown of selected Tudor proteins in Aag2 results in vpiRNA depletion. Individual Tudor proteins were knocked down in Aag2 cells by transfection of dsRNA. Small RNA northern blot was then used to assess the accumulation of (+) strand-derived SINV piRNAs, (-) strand piRNAs from the Copia retro-transposon TF000691, and Histone 4 mRNA-derived piRNAs. The heatmap depicts relative changes in quantified signals from the small RNA northern blots above. Band intensities were normalized to the U6 snRNA loading control and log2-transformed fold changes were calculated relative to dsLuc + SINV. The vpiRNA signal was not quantified for uninfected and dsPiwi5 samples (shown in black), as vpiRNAs are absent or undetectable in these samples.

In a previous study, we used small RNA deep sequencing to analyze SINV derived-piRNAs in infected Aag2 cells (18). The majority of these vpiRNAs derive from a ~200nt hotspot in the capsid gene. We selected four highly abundant sense (+) strand derived vpiRNA sequences from this hotspot region for small RNA northern blotting after knockdown of Tudor proteins. Knockdown of several Tudor proteins lead to reduced vpiRNA production in Aag2 cells (Figure 2). This reduction cannot be explained by changes in viral replication, as only minor differences were seen in viral RNA levels across knockdowns (Figure S1A-C). Knockdown was efficient, resulting in 50 to 80% reduction in mRNA abundance for most genes (Figure S1D). Production of (-) strand primary piRNAs derived from the Copia transposon TF000691 was less strongly affected by Tudor knockdown. Similarly, histone H4 derived secondary piRNAs, which we had previously shown to be dependent on piRNA amplification by Piwi5 and Ago3 (42), were slightly reduced upon AAEL012441 knockdown (Figure 2), suggesting that AAEL012441 is mainly active in a PIWI complex that preferentially processes piRNAs from SINV transcripts. Knockdown of AAEL012441 resulted in the most prominent reduction of vpiRNA levels in repeated experiments (Figure S2), hence we continued with a more detailed characterization of this Tudor protein:

Knockdown of AAEL012441 predominantly affects production of secondary vpiRNAs

Small RNA northern blotting is suitable for the detection of only a handful of highly abundant secondary (+) strand-derived vpiRNAs. To enable a more comprehensive analysis of small RNA populations upon knockdown of AAEL012441, we prepared small RNA deep sequencing libraries from SINV infected Aag2 cells. Knockdown of

AAEL012441 in these cells resulted in a considerable reduction of (+) strand-derived vpiRNAs, while those derived from the (-) strand were not affected (Figure 3A). In line with northern blot data for TF000691, global levels of transposon-derived piRNAs, both from the (+) strand and the (-) strand were slightly reduced by AAEL012441 silencing (Figure 3B). Also, histone H4 piRNA levels were only marginally changed, in accordance with the northern blot results (Figure 3C). As expected, siRNA production was not changed by depletion of AAEL012441 regardless of the substrate RNA (virus and transposon; Figures 3D and 3E).

The Tudor protein AAEL012441 localizes to perinuclear foci

To further characterize the molecular function of AAEL012441 during vpiRNA biogenesis, we expressed GFP-tagged AAEL012441 and three domain mutants (Figure 4A) in Aag2 cells. When cloning these constructs, we noticed that the VectorBase

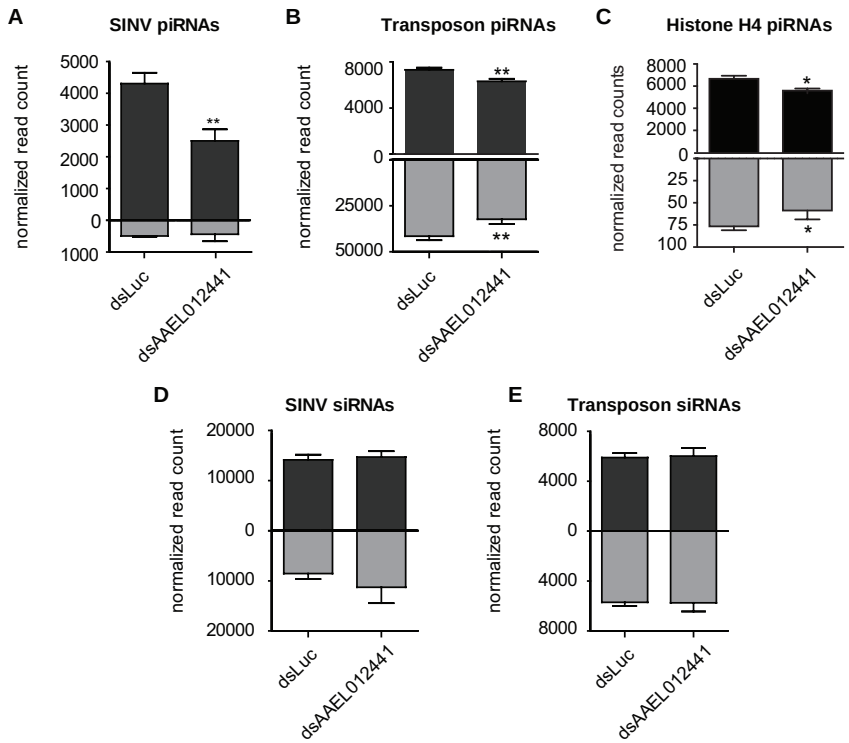


Figure 3. AAEL012441 is required for efficient biogenesis of (+) strand-derived vpiRNAs. (A-C) The sum of normalized read counts of 25-30 nt piRNAs mapping to the SINV genome (A), TEfam transposons (B) or histone H4 genes (C). Read counts of three libraries were normalized to the corresponding library size and analyzed separately for the (+) strand (black) and (-) strand (grey). Bars indicate mean and standard deviation. **(D-E)** Normalized read counts of 21 nt siRNAs mapping to the SINV genome (D) or TEfam transposons (E) analyzed as described for A and B. As previously reported, no siRNAs are produced from histone H4 mRNA (42). Two-tailed student's T-test was used to determine statistical significance (* $P<0.05$; ** $P< 0.01$)

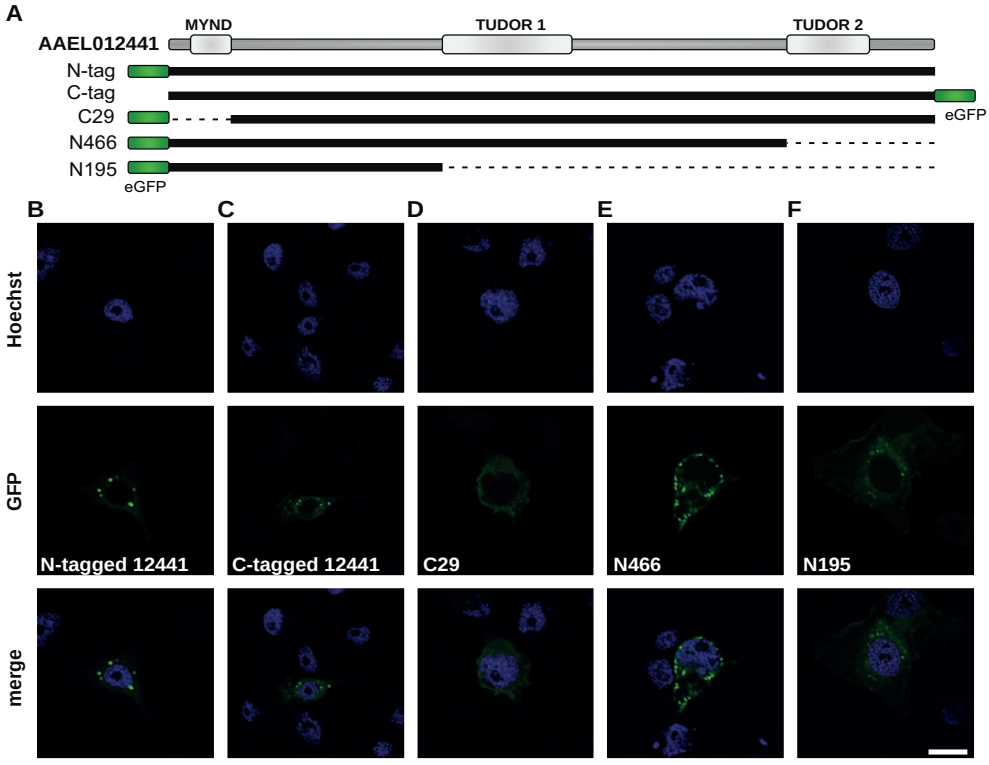


Figure 4. AAEL012441 accumulates in perinuclear foci. (A) Schematic representation of AAEL012441 transgenes. Dashed lines depict deleted segments. **(B-F)** Representative confocal images of cells expressing transgenes shown in (A). Scale bar represents 10μm.

annotation of the complementary DNA (cDNA) sequence for AAEL012441 is erroneous. By Sanger sequencing of PCR products, we revised the current gene annotation (Figure S3). Expression of GFP-tagged AAEL012441 reveals accumulation of the protein in perinuclear foci reminiscent of piRNA processing granules such as Yb bodies or the nuage in *D. melanogaster* (9, 27). No difference was observed between N- and C-terminal tagged AAEL012441 (Figure 4B and C). AAEL012441 contains an N-terminal MYND-type zinc finger, a domain that functions primarily in protein-protein interactions. Removal of this MYND-domain (C29), abolishes perinuclear accumulation of AAEL012441 (Figure 4D). Removal of the C-terminal TUDOR domain (N466) does not affect subcellular localization (Figure 4E), whereas removal of both TUDOR domains (N195) results in an intermediate phenotype (Figure 4F).

AAEL012441 associates with Ago3 through its TUDOR domain

As AAEL012441 is essential for efficient ping-pong amplification of vpiRNAs, we hypothesized that the protein may serve as a scaffold to facilitate the interaction between the ping-pong partners Ago3 and Piwi5. To investigate this possibility,

we immunoprecipitated GFP-tagged AAEL012441 and domain mutants (Figure 5A) and probed lysates using endogenous antibodies recognizing Ago3 and Piwi5 (Figure S4). We found that AAEL012441 interacts with Ago3, but not Piwi5 (Figure 5B). We confirm these findings by reciprocal detection of AAEL012441 in isolated Ago3-complexes specifically (Figure 5C). To further map the domains required for this interaction with Ago3, we evaluated the interaction of AAEL012441 mutants with Ago3. The MYND-domain mutant (C29), in which granular subcellular localization is distorted, still aptly binds Ago3, suggesting that granular localization is not required for interaction with Ago3. The C terminus, which contains two TUDOR domains (C195) is sufficient for interaction with Ago3, whereas Ago3-binding is lost upon deletion of the C-terminal or both Tudor domains (N466, N195; Figure 5D). However, this loss of binding, especially for N466, may be caused by reduced expression of this mutant (Figure 5D). Therefore, to further specify if interaction with Ago3 is TUDOR-mediated, we generated AAEL012441 transgenes carrying point mutations in aromatic cage residues (2Δ and 4Δ; Figure 5A). We found that the C-terminal TUDOR domain is atypical in that only one of the residues, which are predicted to be required for sDMA recognition is conserved (Figure S5). We therefore only analyzed binding of Ago3 to AAEL012441 that carries point mutations in the first TUDOR domain. Interaction with Ago3 was lost in these mutants, suggesting that the first TUDOR domain of AAEL012441 binds Ago3 in a canonical sDMA-dependent manner. It is likely that the C-terminal TUDOR domain is not involved in Ago3 binding since critical residues are not conserved (Figure S5). However, we cannot fully exclude that cooperative binding of both TUDOR domains by Ago3 is required for efficient association with AAEL012441. The interaction with Ago3 is not required for localization of AAEL012441 in perinuclear foci, as this subcellular localization pattern is also seen for the 2Δ-mutant (Figure 5E).

To further dissect the multi-molecular complexes containing AAEL012441, we employed quantitative mass spectrometry. These data confirm association with Ago3 and reveal interesting additional interaction partners of AAEL012441 (Figure 5F). Specifically, the DEAD-box Helicase AAEL004978, which is the *Ae. aegypti* ortholog of Vasa, is highly enriched in AAEL012441-IP. Similarly, the ortholog of the Yb-subfamily, AAEL001939 associates with AAEL012441. We verified these interactions by co-purifying its constituents in reciprocal IPs followed by western blot (Figure 5G). Interestingly, we also detected Piwi5 in AAEL001939 IPs. Together, these findings suggest the presence of a multi-molecular complex involving the ping-pong partners Ago3 and Piwi5, Tudor proteins AAEL012441 and AAEL001939 and the helicase AAEL004978 (Figure 5H).

DISCUSSION

Mosquito antiviral immunity largely relies on the processing of viral dsRNA into vsiRNAs that direct post-transcriptional degradation of viral RNAs. The discovery of *de novo* production of vpiRNAs from arboviral RNA however, uncovered the intriguing

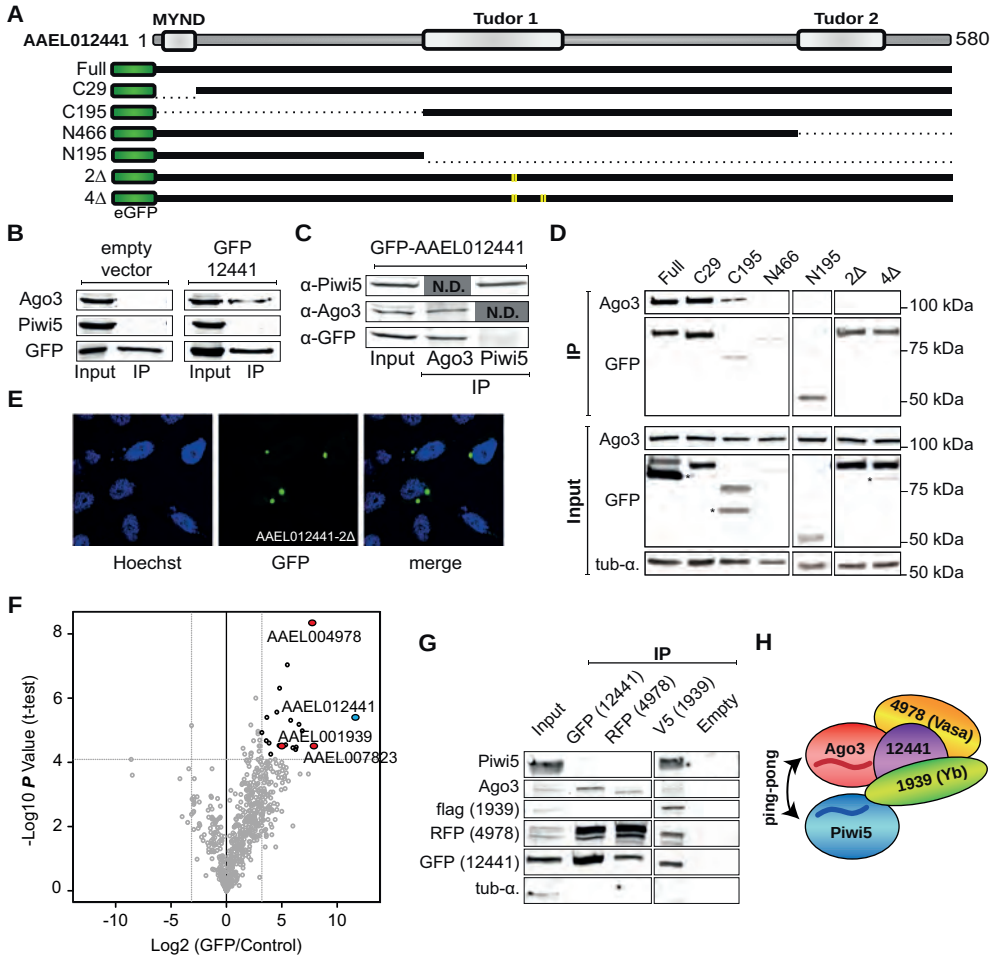


Figure 5. Identification of a multi-protein complex around *Ae. aegypti* ping-pong partners. (A) Schematic representation of AAEL012441 transgenes used in immunoprecipitation experiments. (C29: aa 30-580; C195: aa 196-580; N466: aa1-466; N195: aa 1-195; 2Δ: G259A, Y261A; 4Δ: G259A, Y261A, D279A, N283A) **(B)** Protein lysates of Aag2 cells transfected with an expression plasmids encoding GFP or GFP-AAEL012441 before (input) and after GFP-IP were analyzed for expression of endogenous Ago3 and Piwi5 as well as the GFP transgene by western blot. **(C)** Protein lysates of Aag2 cells transfected with GFP-AAEL012441 before (input) and after IP with endogenous Ago3 (A3-IP) and Piwi5 (P5-IP) antibodies were analyzed for expression of Ago3 and Piwi5 as well as GFP-AAEL012441. **(D)** IPs as described in B including AAEL012441 deletion and point mutants. Blotting of α-tubulin serves as loading control. Asterisks indicate non-specific bands. **(E)** Representative confocal image of Aag2 cells expressing GFP-tagged AAEL012441-2Δ-mutant. **(F)** Identification of AAEL012441 interacting proteins by label-free quantitative mass spectrometry (LFQ). Permutation-based FDR-corrected t-test was used to determine proteins that are statistically enriched in the AAEL012441-IP. The LFQ-intensity of GFP-AAEL012441 IP over a non-transfected control IP (log2-transformed) is plotted against the -log10 *P* value. The proteins in the top right corner represent the bait in blue (AAEL012441) and its interactors. Ago3 (AAEL007823) and orthologs of known piRNA biogenesis factors in *D. melanogaster* are indicated in red. (fold change > 3.7; *P* value > 4.097) **(F)** Reciprocal IPs of GFP-AAEL012441, RFP-AAEL004978 and V5-3xflag-AAEL001939. Samples were probed with antibodies against GFP, RFP, FLAG, Ago3, Piwi5 and α-tubulin. **(G)** Schematic model of the identified multi-protein complex involving the *Ae. aegypti* ping-pong partners.

possibility of an additional small RNA-based line of defense against arboviruses. Processing of viral dsRNA into vsiRNAs by the siRNA pathway has been thoroughly characterized in mosquitoes (43, 44). As of yet, it is unclear how viral RNA produced in the cytoplasm is entered into the piRNA pathway, especially as the canonical substrates for the piRNA pathway are genomically encoded single stranded precursors (3, 4). To better understand how viral RNA is processed by the mosquito piRNA pathway, more insight in the underlying mechanisms of vpiRNA biogenesis is needed.

In *Ae. aegypti,* vpiRNAs are amplified by the ping-pong partners Ago3 and Piwi5, but co-factors involved in this process remain enigmatic. As a tightly regulated network of Tudor proteins promotes production of piRNAs in *Drosophila* (19, 20), we performed a comprehensive RNAi screen to evaluate the role of *Ae. aegypti* Tudor proteins in vpiRNA biogenesis. Knockdown of several Tudor proteins affects vpiRNA biogenesis, with AAEL012441 knockdown resulting in the strongest depletion of vpiRNAs. Additional candidates that reproducibly show depletion in vpiRNA levels are AAEL001939 the only ortholog of the *Drosophila* Yb-subfamily and AAEL008101, which does not contain a one to one homologue in the fly. Whereas AAEL001939 involvement in vpiRNA biogenesis is likely explained by direct interaction with Piwi5 and Ago3 in the multi-protein complex discovered in this study (see below), the function of AAEL008101 remains to be discovered. Involvement of additional Tudor proteins may be masked by redundancy of paralogous proteins or sub-optimal knockdown efficiency.

Thus far, the direct ortholog of AAEL012441 in *D. melanogaster* (CG9684) has not been studied extensively. In a systematic analysis of all *Drosophila* Tudor proteins, germline-specific knockdown of CG9684 did not affect steady-state levels of transposon transcripts or female fertility rate (26). This study, however, did not evaluate the effect of CG9684 knockdown on small RNA populations.

AAEL012441 is essential for efficient production of (+) strand vpiRNAs, suggesting a role in ping-pong amplification. The protein accumulates in perinuclear foci similar to piRNA processing bodies in the fly. In *Drosophila* somatic follicle cells, which surround the germ cells, primary piRNA biogenesis takes place in Yb bodies. One of the core factors present in these structures is their eponym Yb (26-28). Yet, no piRNA amplification takes place in Yb bodies, since the ping-pong partners Aub and Ago3 are not expressed in follicle cells (45, 46) In contrast, in *Drosophila* germ cells piRNA amplification takes place in the nuage and one of the core proteins of this perinuclear structure is the helicase Vasa (9, 24, 25). In *Drosophila* and silkworm, Vasa is directly implicated in secondary piRNA amplification by preventing non-specific degradation of piRNA precursors and facilitating their transfer to PIWI proteins (24). Yb is not present in *nuage* but it has been suggested that its function may be taken over by its close paralogs brother of Yb (BoYb) and sister of Yb (SoYb) (26). In *Ae. aegypti* only one paralog of Yb is encoded (AAEL001939), which directly and/or indirectly associates

with AAEL012441, AAEL004978, the *Ae. aegypti* ortholog of Vasa, and the ping-pong partners Ago3 and Piwi5. The presence of a multi-protein complex including homologues of Vasa and Yb supports the idea that AAEL012441 foci resemble *nuage*-like piRNA processing bodies. However, we previously did not detect pronounced perinuclear localization of transgenic Ago3 and Piwi5 (18) and formation of perinuclear foci was not required for AAEL012441 interaction with Ago3. Therefore, the relevance of the AAEL012441 bodies and their link to piRNA biogenesis remains enigmatic.

Similar to AAEL012441, the *Drosophila* Tudor protein Krimper localizes in perinuclear granules, which are lost upon deletion of the amino terminus of the protein (47, 48). While Krimper interacts with both partners in the ping-pong loop in flies (Ago3 and Aub), AAEL012441 associates exclusively with Ago3. Intriguingly, Krimper-Ago3 interaction is retained when using an arginine-methylation-deficient mutant of Ago3, whereas an sDMA-recognition-deficient mutant of AAEL012441 is unable to bind Ago3. Thus, sDMA recognition seems to be required for AAEL012441-Ago3 interaction in mosquitoes, but dispensable for Krimper-Ago3 association in flies.

AAEL012441 knockdown primarily diminishes the production of vpiRNAs from the viral (+) strand, which is puzzling since silencing of the secondary piRNA binding protein Ago3 results in detectable reduction also of primary piRNAs. Similarly, interfering with piRNA amplification in the fly affects the primary piRNA population (45, 46). In our experiments, AAEL012441 knockdown only resulted in about 50% reduction of (+) strand vpiRNAs and it is conceivable that remaining Ago3-loaded secondary piRNAs are sufficiently abundant to target the viral (-) strand, which by itself is a rather scarce molecule in infected cells. In contrast to (+) strand vpiRNAs, accumulation of endogenous transposon and histone H4 derived piRNAs were only modestly affected by AAEL012441 knockdown. The piRNA population from transposable elements is dominated by primary piRNAs (18, 49) and only few individual transposons engage in the ping-pong loop in Aag2 cells. Hence, it may not be surprising that AAEL012441, which seems to be primarily implicated in piRNA amplification is not strongly involved in their production. Much more surprising however is the stability of histone H4 derived piRNAs, a large proportion of which depends on ping-pong amplification (42). These data suggest that AAEL012441 is involved in specifying the substrate for piRNA production and may preferentially shuttle viral RNA into the ping-pong loop. It would be interesting to assess whether viral RNA from other virus arbovirus families are similarly affected by AAEL012441 knockdown, which would point towards a more general role of AAEL012441 in self-nonself discrimination. Specific requirement of small RNA pathway co-factors for the recognition of different RNA sources been reported. For example, the siRNA pathway co-factor Loqs-PD is required for processing of endogenous-siRNA precursors but is dispensable for siRNA production from exogenous dsRNA or viral RNA (50, 51). The TUDOR protein Qin/Kumo specifically prevents (+) strand transposon

RNAs from becoming PIWI-bound piRNAs during the process of piRNA phasing (52). Tdrd1, the mouse ortholog of AAEL012441, ensures processing of the correct transcripts by the piRNA pathway and in *Tdrd1* knockout mice, the PIWI protein Mili contains a disproportionally large population of piRNAs derived from cellular mRNA and ribosomal RNA (53). In a similar fashion, AAEL012441 could promote processing specifically of viral RNA by the mosquito piRNA pathway. Yet, we expect the molecular mechanism underlying this TUDOR-guided sorting to be different as Tdrd1 interacts with Mili, the PIWI protein that predominantly binds 1U biased primary piRNAs, whereas AAEL012441 associates with Ago3, which mainly binds 10A biased secondary piRNAs.

A sophisticated network of auxiliary proteins that guide diverse RNA substrates into distinct piRISC complexes may be of particular importance in *Ae. Aegypti* as this mosquito species encodes an expanded PIWI gene family consisting of eight members (54, 55), of which Piwi 4-6 and Ago3 are expressed in somatic tissues (56). Moreover, the repertoire of RNA molecules that are processed into piRNAs is extended and includes mRNA and viral RNA (10). Tudor proteins like AAEL012441 may therefore aid in streamlining piRNA processing and perhaps even allow flexible adaptation of the piRNA pathway in response to internal and external stimuli such as arbovirus infection.

ACKNOWLEDGEMENTS

We thank members of the van Rij laboratory for fruitful discussions and the Microscopic Imaging Centre of Radboudumc for access to, and support for confocal microscopy. We would also like to thank Gorben Pijlman, (University of Wageningen, the Netherlands) for kindly providing the pPUbB-GW vector generated in his lab. This work was financially supported by a PhD fellowship from Radboud University Medical Center to PM, and a European Research Council Consolidator Grant under the European Union's Seventh Framework Programme (grant number ERC CoG 615680) to RPvR.

REFERENCES

1. Ghildiyal M, Zamore PD. Small silencing RNAs: an expanding universe. Nat Rev Genet. 2009;10(2):94-108.
2. Meister G. Argonaute proteins: functional insights and emerging roles. Nat Rev Genet. 2013;14(7):447-59.
3. Czech B, Hannon GJ. One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing. Trends Biochem Sci. 2016;41(4):324-37.
4. Hirakata S, Siomi MC. piRNA biogenesis in the germline: From transcription of piRNA genomic sources to piRNA maturation. Biochim Biophys Acta. 2016;1859(1):82-92.
5. Lim RS, Kai T. A piece of the pi(e): The diverse roles of animal piRNAs and their PIWI partners. Semin Cell Dev Biol. 2015;47-48:17-31.
6. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, et al. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell. 2007;128(6):1089-103.

7. Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, et al. A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in Drosophila. Science. 2007;315(5818):1587-90.
8. Cora E, Pandey RR, Xiol J, Taylor J, Sachidanandam R, McCarthy AA, et al. The MID-PIWI module of Piwi proteins specifies nucleotide- and strand-biases of piRNAs. RNA. 2014;20(6):773-81.
9. Lim AK, Kai T. Unique germ-line organelle, nuage, functions to repress selfish genetic elements in Drosophila melanogaster. Proc Natl Acad Sci U S A. 2007;104(16):6714-9.
10. Miesen P, Joosten J, van Rij RP. PIWIs Go Viral: Arbovirus-Derived piRNAs in Vector Mosquitoes. PLoS Pathog. 2016;12(12):e1006017.
11. Kraemer MU, Sinka ME, Duda KA, Mylne AQ, Shearer FM, Barker CM, et al. The global distribution of the arbovirus vectors Aedes aegypti and Ae. albopictus. Elife. 2015;4:e08347.
12. Lambrechts L, Scott TW. Mode of transmission and the evolution of arbovirus virulence in mosquito vectors. Proc Biol Sci. 2009;276(1660):1369-78.
13. Myles KM, Wiley MR, Morazzani EM, Adelman ZN. Alphavirus-derived small RNAs modulate pathogenesis in disease vector mosquitoes. Proc Natl Acad Sci U S A. 2008;105(50):19938-43.
14. Samuel GH, Wiley MR, Badawi A, Adelman ZN, Myles KM. Yellow fever virus capsid protein is a potent suppressor of RNA silencing that binds double-stranded RNA. Proc Natl Acad Sci U S A. 2016;113(48):13863-8.
15. Sanchez-Vargas I, Scott JC, Poole-Smith BK, Franz AW, Barbosa-Solomieu V, Wilusz J, et al. Dengue virus type 2 infections of Aedes aegypti are modulated by the mosquito's RNA interference pathway. PLoS Pathog. 2009;5(2):e1000299.
16. Cirimotich CM, Scott JC, Phillips AT, Geiss BJ, Olson KE. Suppression of RNA interference increases alphavirus replication and virus-associated mortality in Aedes aegypti mosquitoes. BMC Microbiol. 2009;9:49.
17. Bronkhorst AW, van Rij RP. The long and short of antiviral defense: small RNA-based immunity in insects. Curr Opin Virol. 2014;7:19-28.
18. Miesen P, Girardi E, van Rij RP. Distinct sets of PIWI proteins produce arbovirus and transposon-derived piRNAs in Aedes aegypti mosquito cells. Nucleic Acids Res. 2015;43(13):6545-56.
19. Siomi MC, Mannen T, Siomi H. How does the royal family of Tudor rule the PIWI-interacting RNA pathway? Genes Dev. 2010;24(7):636-46.
20. Pek JW, Anand A, Kai T. Tudor domain proteins in development. Development. 2012;139(13):2255-66.
21. Maurer-Stroh S, Dickens NJ, Hughes-Davies L, Kouzarides T, Eisenhaber F, Ponting CP. The Tudor domain ‚Royal Family': Tudor, plant Agenet, Chromo, PWWP and MBT domains. Trends Biochem Sci. 2003;28(2):69-74.
22. Kirino Y, Kim N, de Planell-Saguer M, Khandros E, Chiorean S, Klein PS, et al. Arginine methylation of Piwi proteins catalysed by dPRMT5 is required for Ago3 and Aub stability. Nat Cell Biol. 2009;11(5):652-8.
23. Liu H, Wang JY, Huang Y, Li Z, Gong W, Lehmann R, et al. Structural basis for methylarginine-dependent recognition of Aubergine by Tudor. Genes Dev. 2010;24(17):1876-81.
24. Xiol J, Spinelli P, Laussmann MA, Homolka D, Yang Z, Cora E, et al. RNA clamping by Vasa assembles a piRNA amplifier complex on transposon transcripts. Cell. 2014;157(7):1698-711.
25. Patil VS, Kai T. Repression of retroelements in Drosophila germline via piRNA pathway by the Tudor domain protein Tejas. Curr Biol. 2010;20(8):724-30.
26. Handler D, Olivieri D, Novatchkova M, Gruber FS, Meixner K, Mechtler K, et al. A systematic analysis of Drosophila TUDOR domain-containing proteins identifies Vreteno and the Tdrd12 family as essential primary piRNA pathway factors. EMBO J. 2011;30(19):3977-93.
27. Saito K, Ishizu H, Komai M, Kotani H, Kawamura Y, Nishida KM, et al. Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in Drosophila. Genes Dev. 2010;24(22):2493-8.

28. Olivieri D, Sykora MM, Sachidanandam R, Mechtler K, Brennecke J. An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in Drosophila. EMBO J. 2010;29(19):3301-17.
29. Zamparini AL, Davis MY, Malone CD, Vieira E, Zavadil J, Sachidanandam R, et al. Vreteno, a gonad-specific protein, is essential for germline development and primary piRNA biogenesis in Drosophila. Development. 2011;138(18):4039-50.
30. Soding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. 2005;33(Web Server issue):W244-8.
31. Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, et al. HMMER web server: 2015 update. Nucleic Acids Res. 2015;43(W1):W30-8.
32. Di Tommaso P, Moretti S, Xenarios I, Orobitg M, Montanyola A, Chang JM, et al. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. Nucleic Acids Res. 2011;39(Web Server issue):W13-7.
33. Saleh MC, Tassetto M, van Rij RP, Goic B, Gausson V, Berry B, et al. Antiviral immunity in Drosophila requires systemic RNA interference spread. Nature. 2009;458(7236):346-50.
34. Pall GS, Hamilton AJ. Improved northern blot method for enhanced detection of small RNA. Nat Protoc. 2008;3(6):1077-84.
35. van Cleef KW, van Mierlo JT, Miesen P, Overheul GJ, Fros JJ, Schuster S, et al. Mosquito and Drosophila entomobirnaviruses suppress dsRNA- and siRNA-induced RNAi. Nucleic Acids Res. 2014;42(13):8732-44.
36. Blankenberg D, Gordon A, Von Kuster G, Coraor N, Taylor J, Nekrutenko A, et al. Manipulation of FASTQ data with Galaxy. Bioinformatics. 2010;26(14):1783-5.
37. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10(3):R25.
38. Smits AH, Jansen PW, Poser I, Hyman AA, Vermeulen M. Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. Nucleic Acids Res. 2013;41(1):e28.
39. Rappsilber J, Mann M, Ishihama Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat Protoc. 2007;2(8):1896-906.
40. Ishizu H, Siomi H, Siomi MC. Biology of PIWI-interacting RNAs: new insights into biogenesis and function inside and outside of germlines. Genes Dev. 2012;26(21):2361-73.
41. Clough E, Moon W, Wang S, Smith K, Hazelrigg T. Histone methylation is required for oogenesis in Drosophila. Development. 2007;134(1):157-65.
42. Girardi E, Miesen P, Pennings B, Frangeul L, Saleh MC, van Rij RP. Histone-derived piRNA biogenesis depends on the ping-pong partners Piwi5 and Ago3 in Aedes aegypti. Nucleic Acids Res. 2017; 45(8):4881-92.
43. Gammon DB, Mello CC. RNA interference-mediated antiviral defense in insects. Curr Opin Insect Sci. 2015;8:111-20.
44. Blair CD, Olson KE. The role of RNA interference (RNAi) in arbovirus-vector interactions. Viruses. 2015;7(2):820-43.
45. Malone CD, Brennecke J, Dus M, Stark A, McCombie WR, Sachidanandam R, et al. Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. Cell. 2009;137(3):522-35.
46. Li C, Vagin VV, Lee S, Xu J, Ma S, Xi H, et al. Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. Cell. 2009;137(3):509-21.
47. Sato K, Iwasaki YW, Shibuya A, Carninci P, Tsuchizawa Y, Ishizu H, et al. Krimper Enforces an Antisense Bias on piRNA Pools by Binding AGO3 in the Drosophila Germline. Mol Cell. 2015;59(4):553-63.
48. Webster A, Li S, Hur JK, Wachsmuth M, Bois JS, Perkins EM, et al. Aub and Ago3 Are Recruited to Nuage through Two Mechanisms to Form a Ping-Pong Complex Assembled by Krimper. Mol Cell. 2015;59(4):564-75.

49. Vodovar N, Bronkhorst AW, van Cleef KW, Miesen P, Blanc H, van Rij RP, et al. Arbovirus-derived piRNAs exhibit a ping-pong signature in mosquito cells. PLoS One. 2012;7(1):e30861.

50. Marques JT, Wang JP, Wang X, de Oliveira KP, Gao C, Aguiar ER, et al. Functional specialization of the small interfering RNA pathway in response to virus infection. PLoS Pathog. 2013;9(8):e1003579.

51. Hartig JV, Forstemann K. Loqs-PD and R2D2 define independent pathways for RISC generation in Drosophila. Nucleic Acids Res. 2011;39(9):3836-51.

52. Wang W, Han BW, Tipping C, Ge DT, Zhang Z, Weng Z, et al. Slicing and Binding by Ago3 or Aub Trigger Piwi-Bound piRNA Production by Distinct Mechanisms. Mol Cell. 2015;59(5):819-30.

53. Reuter M, Chuma S, Tanaka T, Franz T, Stark A, Pillai RS. Loss of the Mili-interacting Tudor domain-containing protein-1 activates transposons and alters the Mili-associated small RNA profile. Nat Struct Mol Biol. 2009;16(6):639-46.

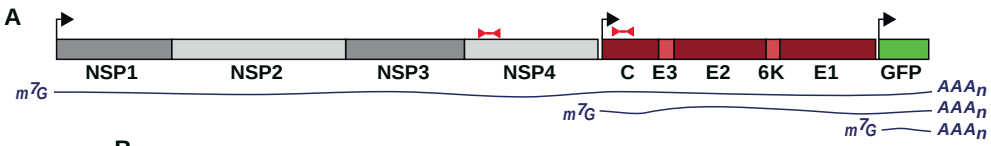
54. Campbell CL, Black WCt, Hess AM, Foy BD. Comparative genomics of small RNA regulatory pathway components in vector mosquitoes. BMC Genomics. 2008;9:425.

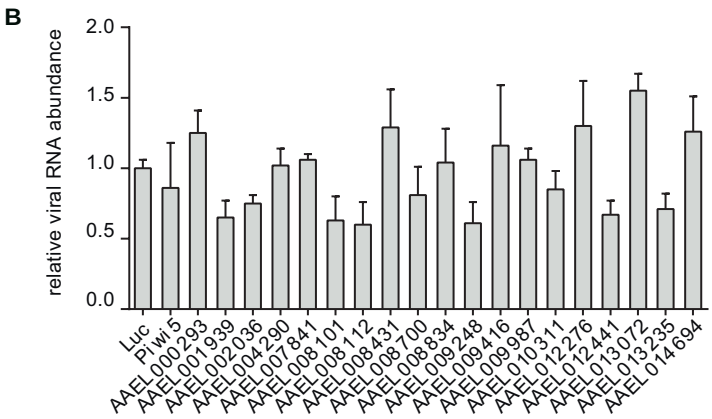
55. Lewis SH, Salmela H, Obbard DJ. Duplication and Diversification of Dipteran Argonaute Genes, and the Evolutionary Divergence of Piwi and Aubergine. Genome Biol Evol. 2016;8(3):507-18.

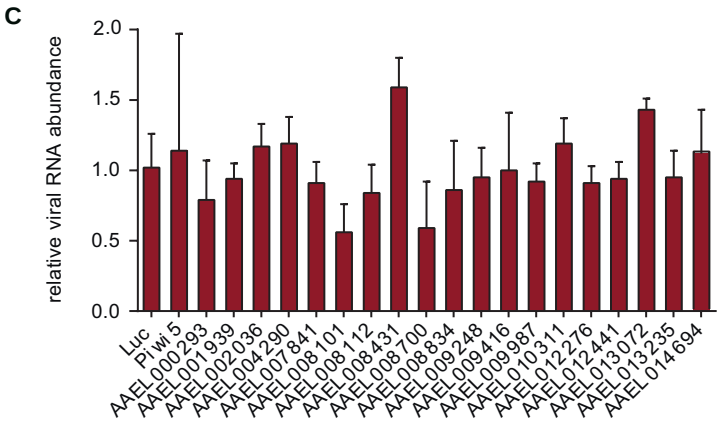
56. Akbari OS, Antoshechkin I, Amrhein H, Williams B, Diloreto R, Sandler J, et al. The developmental transcriptome of the mosquito Aedes aegypti, an invasive species and major arbovirus vector. G3 (Bethesda). 2013;3(9):1493-509.

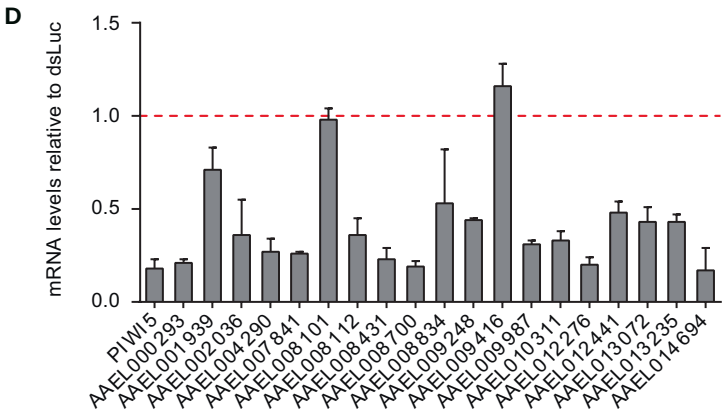
▶ **Figure S1. Viral RNA levels are unaltered upon knockdown of Tudor proteins. (A)** Schematic representation of the SINV-GFP (pTE-3'2J-GFP) genome. Individual non-structural proteins are depicted in grey, structural proteins in red and the GFP expressed under a duplicated sub-genomic promoter in green. The blue lines represent the (+) strand RNA species that are produced during the course of an infection. Red bars flanked by arrowheads indicate the amplicons that were used to quantify viral RNA. **(B,C)** Relative quantification of viral RNAs by RT-qPCR with primers located in the NSP4 (B) and Capsid (C) genes. Note that the Capsid sequence is present in both genomic and subgenomic RNA, whereas NSP4 is only found on genomic RNA. **(D)** Relative abundance of Tudor transcripts in the RNAi knockdown screen shown in Figure 2. Expression levels were normalized to a control dsRNA (dsLuc). The dashed red line depicts the expression level in the dsLuc control. Except for AAEL001939, AAEL008101, AAEL008834, and AAEL009416 knockdown efficiencies are greater than 50%. Bars are the mean +/- standard deviation of three independent experiments.

SUPPLEMENTAL FIGURES









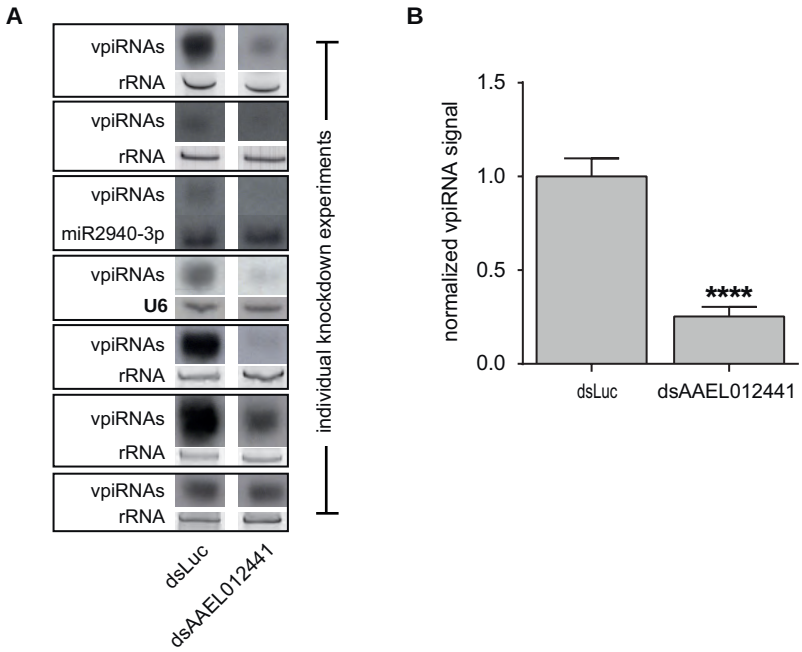


Figure S2. Robust depletion of vpiRNAs in AAEL012441 knockdown experiments. (A) Images used for signal quantification; the vpiRNA signal is shown for each blot together with the signal used for normalization. Probing for miR2940-3p or U6 was used for normalization of two blots, whereas EtBr stained rRNA served as loading control for the remainder. **(B)** Bars are the mean +/- SEM of quantified northern blot signals shown in (A) Two-tailed student's T-test was used to determine statistical significance (**** $P = 8.82 \times 10\text{-}6$).

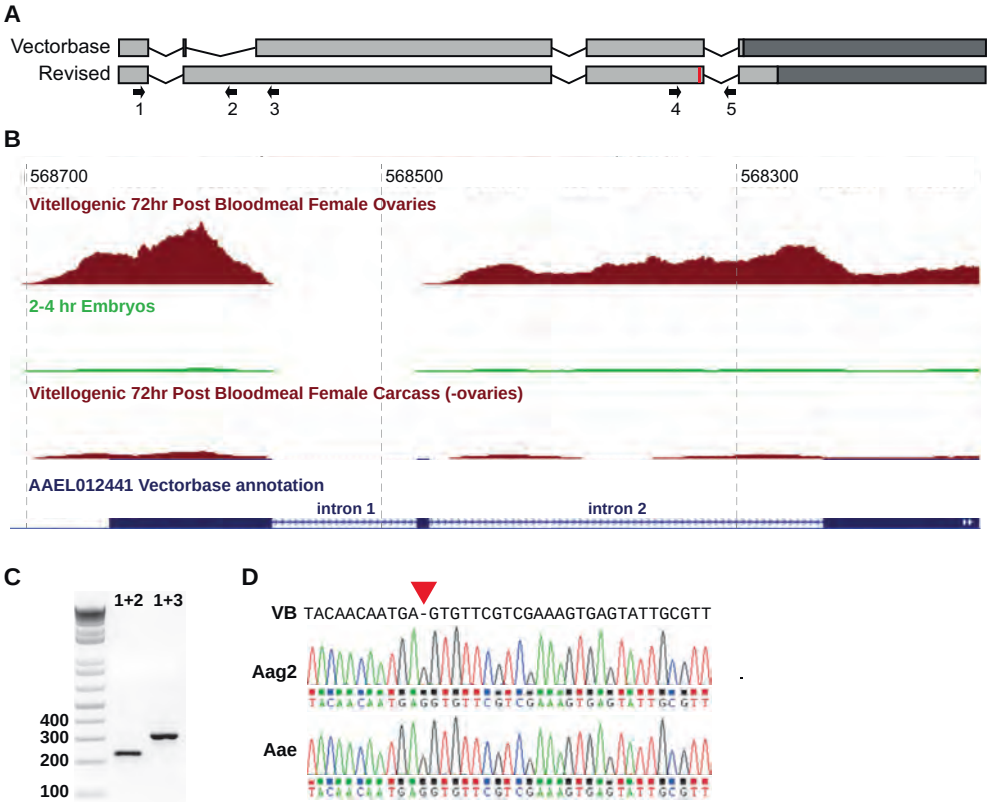


Figure S3. Revised gene annotation of AAEL012441. (A) Schematic representation of the gene annotation for AAEL012441 as published by VectorBase (VB, December 2016) and our revision based on sequencing of PCR fragments. The sequence annotated as the second intron by VB is part of exon 2 in the revised annotation. Furthermore, in the 3' terminus of the coding sequence, an additional Guanosine is present that is not annotated in VB (indicated in red). Together, these revisions result in an increase in protein length from 470 to 580 amino acids. Of note, the sequence that was annotated by VB as the second intron translates into a Glutamine (Q)-rich stretch, which may serve as a trans-activating domain. Arrows 1-5 indicate the position and orientation of primers used to generate RT-PCR products used in C and D. **(B)** RNA sequencing data from indicated tissues taken from the genome browser for *Ae. aegypti* supercontig 1.697 (http://aedes.caltech.edu/) reveal transcription from the sequence annotated as intron 2. Numbers on top show nucleotide position on the scaffold and the annotation by VB is depicted in blue (1). **(C)** RT-PCR products amplified from Aag2 cDNA using the indicated primer combinations. The size of these products fits our revised annotation. The presence of the intron 2 sequence as part of the mature cDNA was also confirmed by Sanger sequencing. **(D)** Sanger sequencing of PCR-products generated from both Aag2 and *Ae. aegypti* genomic DNA using primers 4 and 5 reveal the presence of an non-annotated Guanosine-residue. This additional residue causes a frameshift with respect to annotation, extending the coding sequence at the C-terminus.

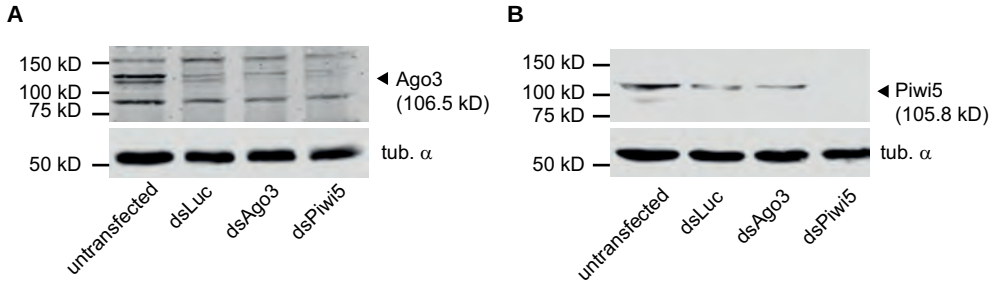


Figure S4. Generation of Ago3 and Piwi5 antibodies. (A-B) Western blot of Ago3 (A) and Piwi5 (B) in Aag2 cells 48 hours post transfection of dsRNA against the indicated genes. Antibodies were used in a 1:100 and 1:500 dilution, respectively. Tubulin alpha serves as loading control.

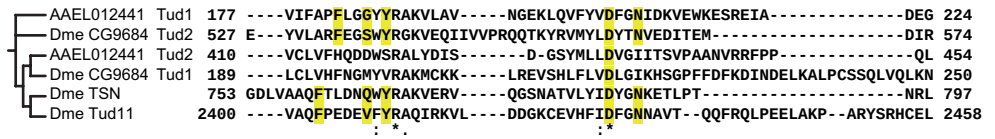


Figure S5. Analysis of AAEL012441 TUDOR domains. Multiple sequence alignment of TUDOR domains from AAEL012441, its *Drosophila* (Dme) orthologue CG9684, and *Drosophila* TSN and Tudor, of which crystal structures have been resolved (2, 3). Residues predicted to be involved in sDMA recognition sites are marked in yellow; asterisks (*) mark identical residues; colon (:) marks conserved substitutions and a period (.) marks a substitution by weakly similar residues. The second TUDOR domain of AAEL012441 and the first TUDOR domain of CG9684 lack aromatic cage residues required for sDMA recognition. T-Coffee (http://www.ebi.ac.uk/Tools/msa/tcoffee/) was used to align the sequences.

SUPPLEMENTAL MATERIALS AND METHODS

Detection of TUDOR orthologues from *Drosophila* and *Ae. aegypti*

The *Drosophila* proteome was scanned using the conserved TUDOR multi-domain sequence (pfam00567 – LPEGSYIDVVVSHIESPSTFYIQPVSDDKKLEKLTEELQEY YASKPPESLPPAVGDGCVAAFSEDGKWYRAKITESLDDGLVEVLFIDYGNTETVPLSD LRPLPPEFESLPPQAIKCQLAG) in HHpred 2.18 (cutoff E ≤ 0.01) (4). Homologous sequences were used as input for iterative searches using Jackhmmer 2.7 to extract all predicted *D. melanogaster* and *Ae. Aegypti* TUDOR domains (5). Subsequently, we used T-Coffee to align the extracted TUDOR-sequences and generate a neighbor joining clustering tree (6), based on sequence identity without correcting for multiple substitutions. For Tudor proteins containing multiple TUDOR domains (eg. Tudor and AAEL007841), we used only the most similar orthologous domains. Using the remaining TUDOR domain sequences, we generated a new neighbor clustering tree using T-coffee, which is shown in Figure 1. A combination of SMART-, Pfam- and Hmmscan-mediated domain prediction was subsequently used to generate protein domain composition maps (5, 7, 8).

Cells and viruses

Aag2 cells were cultured at 25°C in Leibovitz's L-15 medium (Invitrogen) supplemented with 10% fetal bovine serum (Gibco), 50 U/ml Penicillin-Streptomycin (Invitrogen), 1x Non-essential Amino Acids (Invitrogen) and 2% Tryptose phosphate broth solution (Sigma). The virus used in this study is a recombinant Sindbis virus expressing eGFP from a second subgenomic promoter, located downstream of the structural genes. Virus was produced in BHK-21 cells as previously described (9).

Generation of plasmids

AAEL012441 wildtype or mutant as well as AAEL004978 sequences were sub-cloned into the pAGW, pARW (Carnegie Gateway vector collection) or pUGW expression vectors, using a Gateway cloning strategy (Invitrogen). AttB1 and AttB2 recombination sites were added during PCR amplification from Aag2 complementary DNA (cDNA). Donor vectors were generated through BP-recombination of the produced PCR products with pDONR/Zeo (Invitrogen). Subsequently, gene fragments were cloned into the expression vectors by LR-recombination (Invitrogen). The pUGW expression vector was derived from the pPUbB-GW vector (kindly provided by Gorben Pijlman, University of Wageningen), which was generated by exchanging the OpIE2 promoter from pIB-GW by the poly-ubiquitin promoter from pGL3 pUb MCS. A PCR product of pPUb was created with *BspH*I and *Sac*I sites, which was then ligated into the *BspH*I and *Sac*I-digested pIB-GW vector. To generate the pUGW vector, the GFP sequence for N-terminal tagging of proteins was ligated into

the pPUbB-GW vector using the *Sac*I restriction sites. The pAc5.1-V5-3x flag-AAEL001939 plasmid was produced using Infusion (Clontech). Primers used for cloning were:

FW-SacI-EGFP	CCACCGAGCTCATGGTGAGCAAGGGCG
RV-SacI-EGFP	CGGTGGAGCTCCCTTGTACAGCTCGTCCATGC
FW-AttB1-AAEL012441_N	ggggacaagtttgtacaaaaaagcaggcttcCCTAGCTATGAGTGCCGCTGC
RV-AttB2-AAEL012441_N	tggtgaccactttgtacaagaaagctgggttttattattaGGCATCCACGAGAGAAATGGA
FW-AttB1-AAEL012441_C	ggggacaagtttgtacaaaaaagcaggcttcgaaggagatagaaacaccATGCCTAGCTATG-AGTGCCGCTGC
RV-AttB2-AAEL012441_C	ggggaccactttgtacaagaaagctgggtcGGCATCCACGAGAGAAATGGA
FW-AttB1-C29_N	ggggacaagtttgtacaaaaaagcaggcttcATGCCTCGACTGGTACCCATT
FW-AttB1-C195_N	ggggacaagtttgtacaaaaaagcaggcttcGGTGCAATGGTCAAGATTACCG
RV-AttB2-N466_N	ggggaccactttgtacaagaaagctgggttttattattaGCTACAATGCTCCAGGATGGC
RV-AttB2-N195_N	ggggaccactttgtacaagaaagctgggttttattattaTTCCCGAGGGAACGGTCC
FW-AAEL001939_Inf	gataagctttctagaCTCGAGATGTTCGAAGACGACACCAT
RV-AAEL001939_Inf	gagctcgcggccgctCTCGAGTTAACCATCCCGCAGGAAATA
FW-AttB1-AAEL004978	ggggacaagtttgtacaaaaaagcaggcttcTGCGATGAATGGGAGGATAATGA
RV-AttB2-AAEL004978	ggggaccactttgtacaagaaagctgggttttattattaATCCCAGTCTTCTTCTGG-TTC

dsRNA production

dsRNA was produced from *in vitro* transcription from PCR products flanked on both sites by the T7 promoter sequence. The T7 sequence was either directly coupled to the primers used to generate these PCR products or introduced in a second PCR using universal primers that hybridize to a short GC rich universal tag (UT). T7 RNA polymerase was used to *in vitro* transcribe these PCR products. Finally, dsRNAs were heated to 80°C and allowed to gradually cool to room temperature to promote dsRNA formation. Subsequently, the RNA was purified using the GenElute Mammalian Total RNA Miniprep Kit (Sigma). The primers used for dsRNA production were:

Fw-T7-Luc	taatacgactcactatagggagaTATGAAGAGATACGCCCTGGTT
Rv-T7-Luc	taatacgactcactatagggagaTAAAACCGGGAGGTAGATGAGA
Fw-T7-Piwi5	taatacgactcactatagggagaGCCATACATCGGGTCAAAAT
Rv-T7-Piwi5	taatacgactcactatagggagaCTCTCCACCGAAGGATTGAA
Fw-T7-AAEL007841	taatacgactcactatagggagaGCTACCAGAGCCAGAGCAAC
Rv-T7- AAEL007841	taatacgactcactatagggagaTCGGTCAACGCGTAATCATA
Fw-T7-AAEL014694	taatacgactcactatagggagaTCGGATGCGTATCATTACGA
Rv-T7-AAEL014694	taatacgactcactatagggagaAATTCCTTCGTGCTGTTTGG
Fw-UT-AAEL000293	gcccgacgcTCTATTCCGAACGGCCGC
Rv-UT-AAEL000293	cgcctcggcCAGCCGTCGTGTCTGGTT
Fw-UT-AAEL001939	gcccgacgcCGGCATTTGTTGGACCCG
Rv-UT-AAEL001939	cgcctcggcGACAGTCCACGCACCTCA
Fw-UT-AAEL002036	gcccgacgcGCCCTGCCGGATGAGTAC
Rv-UT-AAEL002036	cgcctcggcCGTCGTCCAAGGCCACAA
Fw-UT-AAEL004290	gcccgacgcGCTCACAGAGGAAGCGGG
Rv-UT-AAEL004290	cgcctcggcTATGGCAGGGCTAGGAGC
Fw-UT-AAEL008101	gcccgacgcTGTGCTTAGCGAGGCGAC
Rv-UT-AAEL008101	cgcctcggcCCAGCGGTGGCAGATTCT
Fw-UT-AAEL008112	gcccgacgcTCCACAACGGGGGCATTC
Rv-UT-AAEL008112	cgcctcggcCTTGTGTAGGGCAGGGGC

Fw-UT-AAEL008431	gcccgacgcTGGACGAAAAGCCGGCTT
Rv-UT-AAEL008431	cgcctcggcCAGGTAGCTGTGGCGCTT
Fw-UT-AAEL008700	gcccgacgcAACAGACGGTGGCCATCG
Rv-UT-AAEL008700	cgcctcggcTCTAGGACGGTCGGGCTC
Fw-UT-AAEL008834	gcccgacgcGCGGATACAGCTGCCCAA
Rv-UT-AAEL008834	cgcctcggcGGACGGCTTGACACACCA
Fw-UT-AAEL009248	gcccgacgcCAGCCGGAATCAGCGTCA
Rv-UT-AAEL009248	cgcctcggcTGTCGTCTAGTCGGGCCA
Fw-UT-AAEL009416	gcccgacgcAGCACCTCTGCAGCAGTG
Rv-UT-AAEL009416	cgcctcggcAGGTATGAGGCAACGCGG
Fw-UT-AAEL009987	gcccgacgcGAGCGCGACCGGTATCAA
Rv-UT-AAEL009987	cgcctcggcGGTTTTTCCACACAGGCCA
Fw-UT-AAEL010311	gcccgacgcCGCATACAGTGTCGCGTG
Rv-UT-AAEL010311	cgcctcggcGCCGTCATGCACTTTGCC
Fw-UT-AAEL012276	gcccgacgcAGCAGCTCTTCTGACGGA
Rv-UT-AAEL012276	cgcctcggcTGATTGGGTCGCATGCGT
Fw-UT-AAEL013072	gcccgacgcTCGGGCTGAAGTGATCGC
Rv-UT-AAEL013072	cgcctcggcCCTTGGCATGACCCTCGG
Fw-UT-AAEL013235	gcccgacgcCTGCCATGTCCATCGCGA
Rv-UT-AAEL013235	cgcctcggcATCGCAAAGTCCAGCCGG
F-T7-universal primer	TAATACGACTCACTATAGGGAGAgcccgacgc
R-T7-universal primer	TAATACGACTCACTATAGGGAGAcgcctcggc

Transfection of plasmids and dsRNA

For immunoprecipitation and immunofluorescent analyses, Aag2 cells were transfected with expression vectors using X-tremeGENE HP (Roche) according to manufacturer's instructions. For knockdown experiments, dsRNA was transfected into Aag2 cells using X-tremeGENE and transfected again 48 hours after the initial transfection to ensure efficient knockdown. Three hours after each transfection, medium was refreshed with supplemented Leibovitz medium, and infected with SINV in case of knockdown experiments. Samples were harvested 48 hours post infection.

RNA isolation

Total RNA was isolated using Isol-RNA Lysis reagent (5 PRIME) or RNA-solv reagent (Omega) following the manufacturers instructions. Briefly, 200 µl of chloroform was assed to 1 ml of RNA lysate. After harsh mixing and centrifugation, the aqueous phase was collected and RNA was precipitated by adding one volume of isopropanol. RNA pellets were washed in 80 % ethanol and dissolved in milliQ.

Small RNA Northern Blotting

3-5 µg of total RNA was diluted in 2x loading buffer and size separated on 7M urea 15 % polyacrylamide 0.5x TBE gels by gel-electrophorese. RNAs were then transferred to Hybond Nx nylon membranes (Amersham) and cross-linked using 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride as described in (10).

Membranes were then pre-hypridized in Ultrahyb Oligo hybridization buffer (Ambion) for at least 30 min at 42°C under constant rotation. Then ^{32}P end-labelled DNA oligonucleotides were added directly into the hybridization buffer and hybridization was performed overnight at 42°C under constant rotation. Afterwards, membranes were washed three times in wash buffers containing SDS and decreasing concentrations of SSC as described in (11). Probes used for northern blotting were:

SINV-7903 (+)	GGTTGCTTCTTCTTCTTCTCCTGCGTTT
SINV-7940 (+)	AGTGCCATGCGCTGTCTCTTTCCGGGTTTG
SINV-7969 (+)	TCGAACAATCTGTCGGCCTCCAACTTAA
SINV-8040 (+)	GCAGAGGTTTCATTACCTTTCCTTCCAT
miR-2940 3p	AGTGATTTATCTCCCTGTCGAC
H4-149 (+)	GCACTCCACGGGTTTCCTCGTAGATAA
H4-183 (+)	AGCATCACGAATGACATTTTCCAGGAAT
H4-218 (+)	ACGGTTTTACGCTTGGCGTGTTCAGTGT
H4-257 (+)	CCCTGACGCTTCAGAGCGTAGACAACAT
TF000691-2718 (-)	GGCTCCAATCGTCCGTAAATGTTACGAAA
TF000691-2788 (-)	TACGTCCGAATGGACTAACTCTAGCACTC
TF000691-3251 (-)	ATCAACTTCCCGTTAACGTTAGACACAAG
TF000691-4144 (-)	CACCTTTAGCGCCTGCACGTCATCAACCT
U6 snRNA	GATTTTGCGTGTCATCCTTGTGCAGGGGCCATGCTAA

Reverse transcription and (quantitative) PCR

Total RNA was DNaseI treated (Ambion) and reverse transcribed using Taqman reverse transcriptase (Life Technologies) as per the manufacturers recommendations. SYBR-green qPCR was performed using the GoTaq qPCR system (Promega) according to the manufacturers instructions. Expression levels of target genes were internally normalized against the expression of the house keeping gene lysosomal aspartic protease (LAP) and fold changes were calculated using the 2^(-delta delta CT) method (12). RT-PCR of AAEL012441 to amplify genomic or cDNA for Sanger sequencing was performed using Phusion polymerase (NEB). The following primers were used for qPCR:

Fw-SINV-NSP4	AACTCTGCCACAGATCAGCC
Rv-SINV-NSP4	GGGGCAGAAGGTTGCAGTAT
Fw-SINV-Capsid	CTGGCCATGGAAGGAAAGGT
Rv-SINV-Capsid	CCACTATACTGCACCGCTCC
qFw-LAP	GTGCTCATTCACCAACATCG
qRv-LAP	AACTTGGCCGCAACAAATAC
qFw-Piwi5	ACGGCATCACATCGAGACTC
qRv-Piwi5	CGACCTCCACGCTGTCCTC
qFw-AAEL000293	ACAAGAAGGACCGCAGACTG
qRv-AAEL000293	TCGATTAGTTGGTGGCCGAG
qFw-AAEL001939	GTTGCCGGATTGTCAGCATC
qRv-AAEL001939	GGCAATCGGCGGAATTCTTC
qFw-AAEL002036	GCCTGGAGGTGTACTGTTCC
qRv-AAEL002036	ATTCGACTTGAGGCCTGCTC
qFw-AAEL004290	CGATGATTCACTGCTTGGCG
qRv-AAEL004290	ATCGTCCTCGCAGTCACATC

qFw-AAEL007841	CGTCGTCGAAGAAATCCAAT
qRv-AAEL007841	TGAACTTGCTCTGCAGGATG
qFw-AAEL008101	TTCCAGGCCGTCCACTGT
qRv-AAEL008101	GGATGTCTAGCACTTCGACG
qFw-AAEL008112	AGGTTCGGTAACATGCCCAG
qRv-AAEL008112	TCTGGAAAACACGGACCACC
qFw-AAEL008431	ACTCCTGAAGCATCGGAAGC
qRv-AAEL008431	TATTCCTCCCACACTGCCAG
qFw-AAEL008700	TTGTAGCAAGGCGTCCAACT
qRv-AAEL008700	GATCCAATTCCGCCGGTTTG
qFw-AAEL008834	AAACTGTCTGGATGTGGTTCTG
qRv-AAEL008834	AAATGATTTCGTACGCTCGCG
qFw-AAEL009248	CTCCGTTCTATGCGAGCAGA
qRv-AAEL009248	CCGTTGATTTGGCCTTTGGG
qFw-AAEL009416	AAAGAAGTGGGAGAGCAGCC
qRv-AAEL009416	TTGACATTCGGCCGGATCAA
qFw-AAEL009987	GCTTGATCGAGCTGCAAGTC
qRv-AAEL009987	GTGCCCGGTACCATAGATCG
qFw-AAEL010311	CGTTCCGAAAGACAGCGTTG
qRv-AAEL010311	GCAGGTAGGATTCGCAGTGA
qFw-AAEL012276	GGTGGTGAAGTCAGTGTCGT
qRv-AAEL012276	TCTTCCAGCTCCTTGAAGCG
qFw-AAEL012441	TGGGCATCATTACCAGCGTT
qRv-AAEL012441	TCGATCAACGCTCCGTGAAT
qFw-AAEL013072	TTGAGCAGCGTTGAAAACCG
qRv-AAEL013072	GGCTGGGATGCTGACTCATC
qFw-AAEL013235	GTGGTCGCGATCCCTGTAAT
qRv-AAEL013235	TTGCTACCCAGGAACGTCAC
qFw-AAEL014694	CGGGTTGCTTATTCTCCTCA
qRv-AAEL014694	GCCAAGAATTGTTTCGCAAT
1-Fw-AAEL012441ex1	GCGGTGCCTTCTACTGC
2-Rv-AAEL012441in2	TTGTCTTTGTTGTTGCTGC
3-Rv-AAEL012441ex3	TCTGACCGCGTTGACG
4-Fw-AAEL012441ex4	ACTCTAACGTCTACAAACCGG
5-Rv-AAEL012441in4	AGTTATACAATCAAGCCAAACAC

Immunoprecipitation, western blotting and mass spectrometry analyses

Aag2 cells expressing the transgene of interest were lysed in the following lysis buffer (50mM Tris-Cl (pH 7.4) containing 150mM NaCl, 0.5 mM DTT, 1% Igepal, 10% glycerol, 1x protease inhibitors). This lysate was incubated under rotation at 4°C for 1 hour and centrifuged at 4°C for 30 min at 13000 rpm. Subsequently, the supernatant was taken and snap frozen in liquid nitrogen for later analyses. Lysates from cells expressing GFP-tagged transgenes were subjected to GFP-affinity enrichment using GFP-TRAP beads (Chromotek). For RFP-tagged transgene purification, RFP-TRAP beads (Chromotek) were used. For the Ago3 and Piwi5 IPs, we used protein A/G PLUS Agarose beads (Santa Cruz) to which endogenous Ago3 or Piwi5 antibodies, generated in our laboratory, were cross-linked using dimethyl pimelimidate. Beads were equilibrated using lysis buffer

and incubated with protein lysate under rotation at 4°C for 3 hours. Subsequently, beads were washed three times in wash buffer (10mM Tris-Cl (pH 7.5), 150mM NaCl, 0.5mM EDTA, 1x complete protease inhibitors) and heated at 95°C for 5 minutes in 2xSDS sample buffer (120mM Tris-Cl (pH 6.8), 4% SDS, 10% β-mercaptoethanol, 20% glycerol, 0.04% bromophenol blue) to dissociate protein from the beads. Before V5-immunopurification, lysates were pre-cleared using empty protein A/G PLUS Agarose beads (Santa Cruz) under rotation at 4°C for 5 hours. Subsequently, V5-tagged transgenes were precipitated overnight using V5-agarose beads under rotation at 4°C. Samples were washed and boiled as described above (11). Before western blot analyses, samples were diluted in 1x SDS sample buffer by adding an equal volume of wash buffer, resolved on 7.5% polyacrylamide gels, blotted to nitrocellulose and stained with the following antibodies: rabbit-anti-GFP (1:10,000), rabbit-anti-Ago3 (1:500), rabbit-anti-Piwi5 (1:500) and rat-anti-α-tubulin (1:1000, Sanbio, MCA77G), mouse anti-flag (1:1000, Sigma, F1804), mouse anti-RFP (1:1000, Chromotek, 6G6). Secondary antibodies were goat-anti-rabbit-IRdye 680 and goat-anti-rat-IRdye800 (both: 1:10,000; Licor).

For mass spectrometry analysis, precipitated proteins were washed using lysis buffer and subjected to on-bead trypsin digestion as described in (13). Subsequently, tryptic peptides were acidified and desalted using Stagetips before elution onto a NanoLC-MS/MS. Mass spectra were recorded on a QExactive mass spectrometer (Thermo Scientific). MS and MS2 data were recorded using TOP10 data-dependent acquisition. Raw ata files were analyzed using Maxquant version 1.5.1.0. using standard recommended settings (14). LFQ, IBAQ and match between runs were enabled. Data was searched against the *Ae. Aegypti* Uniprot database (downloaded April 2016). Data was further analyzed and visualized using Perseus version 1.3.0.4 (15) and R. In short, identified proteins were filtered for contaminants and reverse hits. LFQ values were subsequently log2 transformed and missing values were imputed assuming a normal distribution. A t-test was then performed to calculate significantly enriched proteins. Volcanoplots were generated with R.

Immunofluorescence

To evaluate the subcellular localization of AAEL012441 transgenes, Aag2 cells expressing GFP-tagged constructs were fixed on coverslips using 4% paraformaldehyde. After permeabilization with 0.1% Triton in PBS, nuclei were stained using Hoechst reagent, washed and mounted onto microscope slides using Mowiol. All images were taking using the Olympus FV1000 confocal microscope and were processed using ImageJ.

SUPPLEMENTAL REFERENCES

1. Akbari OS, Antoshechkin I, Amrhein H, Williams B, Diloreto R, Sandler J, et al. The developmental transcriptome of the mosquito Aedes aegypti, an invasive species and major arbovirus vector. G3 (Bethesda). 2013;3(9):1493-509.
2. Liu H, Wang JY, Huang Y, Li Z, Gong W, Lehmann R, et al. Structural basis for methylarginine-dependent recognition of Aubergine by Tudor. Genes Dev. 2010;24(17):1876-81.
3. Friberg A, Corsini L, Mourao A, Sattler M. Structure and ligand binding of the extended Tudor domain of D. melanogaster Tudor-SN. J Mol Biol. 2009;387(4):921-34.
4. Soding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. 2005;33(Web Server issue):W244-8.
5. Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F, et al. HMMER web server: 2015 update. Nucleic Acids Res. 2015;43(W1):W30-8.
6. Di Tommaso P, Moretti S, Xenarios I, Orobitg M, Montanyola A, Chang JM, et al. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. Nucleic Acids Res. 2011;39(Web Server issue):W13-7.
7. Letunic I, Doerks T, Bork P. SMART: recent updates, new developments and status in 2015. Nucleic Acids Research. 2015;43(D1):D257-D60.
8. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. Nucleic Acids Research. 2014;42(D1):D222-D30.
9. Vodovar N, Bronkhorst AW, van Cleef KW, Miesen P, Blanc H, van Rij RP, et al. Arbovirus-derived piRNAs exhibit a ping-pong signature in mosquito cells. PLoS One. 2012;7(1):e30861.
10. Pall GS, Hamilton AJ. Improved northern blot method for enhanced detection of small RNA. Nat Protoc. 2008;3(6):1077-84.
11. Miesen P, Girardi E, van Rij RP. Distinct sets of PIWI proteins produce arbovirus and transposon-derived piRNAs in Aedes aegypti mosquito cells. Nucleic Acids Res. 2015;43(13):6545-56.
12. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001;25(4):402-8.
13. Smits AH, Jansen PW, Poser I, Hyman AA, Vermeulen M. Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. Nucleic Acids Res. 2013;41(1):e28.
14. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol. 2008;26(12):1367-72.
15. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Methods. 2016;13(9):731-40.

Chapter 8

General Discussion
piRNAs in Virus-Mosquito Interactions

adapted and expanded from:

Pascal Miesen*, Joep Joosten*, Ronald P. van Rij
PIWIs Go Viral: Arbovirus-Derived piRNAs in Vector Mosquitoes.
PLoS Pathog. (2016) 12:e1006017.

*equal contribution

Vector mosquitoes are responsible for transmitting the majority of human and livestock arthropod-borne (arbo-) viruses. Virus replication in these vectors needs to be sufficiently high to permit efficient virus transfer to vertebrate hosts. The antiviral immunity is therefore an important determinant for the ability of a mosquito to transmit specific arboviruses (vector competence). Immunity to arboviruses is primarily mediated by the small interfering RNA pathway in mosquitoes. Besides this well-established antiviral machinery, the PIWI-interacting (pi)RNA pathway processes viral RNA into piRNAs in vector mosquitoes of the *Aedes* genus (**Chapters 2 and 3**). Originally, the piRNA pathway was studied for its role in transposon defense in animal gonads and clearly, this function is conserved in *Aedes* mosquitoes, since transposon RNA is a prominent source of piRNAs (**Chapter 2**). Yet, beyond transposable elements, additional endogenous RNA substrates are processed into piRNAs in *Aedes aegypti*. These include sequences from Non-retroviral Integrated RNA Virus Sequences (NIRVS), which are versatile genetic elements in the genomes of *Ae. aegypti* and *Ae. albopictus* (**Chapter 4 and 5**). Furthermore, canonical protein-coding mRNAs give rise to abundant populations of endogenous piRNAs (**Chapter 4**) and in addition, piRNA production from an ancient satellite DNA locus has been discovered (**Chapter 6**). I have hypothesized that distinct PIWI protein complexes assemble to allow regulated production of piRNAs from these distinct piRNA precursors. Indeed, I found that different sets of PIWI proteins are involved in the biogenesis of piRNAs from the various exogenous and endogenous RNA sources (**Chapters 2 to 6**). Moreover, auxiliary proteins from the Tudor protein family may be involved in sorting specific RNA to dedicated PIWI complexes (**Chapter 7**). In this chapter, I will discuss the findings presented in this thesis, identify knowledge gaps, and suggest directions for future research.

Small RNAs in Arboviral Infections

Mosquitoes and other hematophagous (blood-feeding) arthropods transmit important human and animal viruses, some of which are responsible for debilitating diseases such as dengue, chikungunya, and Zika (1). Collectively, this non-taxonomical group of viruses is termed arthropod-borne viruses (arboviruses). Most arboviruses are RNA viruses with either double-stranded RNA (dsRNA) genomes or single-stranded RNA (ssRNA) genomes of positive (+) or negative (-) polarity. The majority can be assigned to the families *Bunyaviridae* (-ssRNA), *Flaviviridae* (+ssRNA), *Reoviridae* (dsRNA), *Rhabdoviridae* (-ssRNA), and *Togaviridae* (+ssRNA) (2). The global threat of arboviruses is increasing due to the expansion of the geographical range of vector mosquitoes that prefer to feed on humans (anthropophilia) (1, 3). Interestingly, while potentially causing severe disease in vertebrates, arboviruses replicate to high levels in their mosquito vectors without causing apparent pathology (4, 5). This suggests that vector mosquitoes possess efficient mechanisms to resist or tolerate virus infection, despite lacking the adaptive immune system and interferon-mediated antiviral responses of vertebrate species (6).

Whereas the evolutionary conserved Toll, Imd, and Jak-Stat signaling pathways are implied in antiviral defense (7), the cornerstone of antiviral immunity in insects is the small interfering RNA (siRNA) pathway (**Chapter 1**) (8, 9). This pathway is initiated by cleavage of viral dsRNA into 21-nucleotides (nt)-long siRNAs by the RNase-III endonuclease Dicer-2 (10, 11). These siRNAs associate with Argonaute-2 (Ago2) in an RNA-induced silencing complex (RISC) and serve as a guide for Ago2-mediated cleavage of viral target sequences (10, 12). Accordingly, experimental inactivation of siRNA pathway components in mosquitoes results in increased arbovirus replication (13-18). The fact that several insect viruses have evolved suppressors of the siRNA pathway underlines its importance in antiviral immunity (8, 19). Likewise, arboviral gene products have been proposed to act as antagonists of the siRNA pathway in mosquitoes (20-22).

MicroRNAs comprise an independent class of small RNAs that may be involved in the cellular response to arboviral infections by regulation of host immune genes (23). They are produced from genome-encoded stem-loop RNA structures in a Dicer-1- and Ago1-dependent manner, akin to siRNA biogenesis (**Chapter 1**). The role of siRNAs and microRNAs in mosquito–arbovirus interactions will not be reviewed in this chapter. A separate discussion on microRNAs in host-virus interactions is provided in **chapter 9**.

The focus of this thesis has been on viral and host-derived piRNAs, the most enigmatic class of small silencing RNAs in the context of arbovirus–vector interactions. piRNAs associate with the PIWI clade of the Argonaute protein superfamily, display a broad size range (24–30 nt), and are produced independently of Dicer (**Chapter 1**) (24). The canonical function of the piRNA pathway is protection of genome integrity in animal germ cells by silencing transposons, selfish genetic elements with the ability to randomly integrate into the host genome (25). Recently, however, several studies, including **chapters 2 and 3** of this thesis, have reported *de novo* production of piRNAs derived from viral sequences in the vector mosquitoes *Ae. aegypti* and *Ae. albopictus* and in cell lines derived from these animals (26-35). Biogenesis of viral piRNAs (vpiRNAs) occurs independent of siRNA production, which raises the exciting possibility that vpiRNAs may constitute an additional line of defense against arboviruses in vector mosquitoes.

Our understanding of the piRNA pathway in insects is incomplete and largely biased towards studies in the genetic model insect *Drosophila melanogaster* (Box 1). Yet, piRNA pathways in vector mosquitoes differ considerably from *Drosophila* and other model organisms. This becomes apparent in several aspects: (i) The composition of piRNA pathway components differs between *Drosophila* and mosquitoes (Figure 1). Notably, the PIWI gene family, which lies at the heart of the piRNA pathway, has undergone expansion in both *Aedes* and *Culex* mosquitoes (36, 37). In addition, the recent annotations of mosquito genomes do not contain orthologs for all the established factors involved in *Drosophila* piRNA biogenesis and function (38). (ii) Mosquito PIWI proteins have an extended expression pattern (Figure 1). For instance, some of the members of the expanded *Aedes* PIWI family are expressed in somatic tissue (**Chapter 6**) (39), whereas expression

of PIWI proteins in *Drosophila* is largely restricted to gonadal tissues (40-43). (iii) The piRNA pathway in *Aedes* processes a broader repertoire of substrates (Figure 1). Despite the large transposon content of the *Ae. aegypti* genome (44), relatively few piRNAs are derived from these mobile elements (45). Instead, a considerable proportion of piRNAs are derived from non-repetitive genomic areas, including the open reading frames of protein-coding genes (**Chapter 4**) (45). Yet, the most prominent gain of function is the production of piRNAs from viral RNA during the course of an acute infection (**Chapters 2 and 3**).

Box 1. piRNA Biogenesis in *Drosophila*

In the *Drosophila* germline, the mobilization of transposable elements is efficiently suppressed by transcriptional and post-transcriptional gene silencing by the piRNA pathway. piRNA biogenesis involves the primary processing pathway and ping-pong amplification, which is capable of triggering phased piRNA production. Below, I provide a brief description of the *Drosophila* piRNA pathway; for a more comprehensive description I refer to **chapter 1** or to recent literature reviews (24, 46).

During primary processing, single-stranded piRNA precursors are generated from genomically encoded piRNA clusters that are rich in transposon remnants. The endonuclease Zucchini (Zuc) cleaves these precursors directly upstream of uridine residues, thus producing piRNA intermediates with a bias for a uridine at the first nucleotide position (1U) (47-49). In an electron-dense peri-nuclear structure termed nuage, these piRNA intermediates are loaded onto the PIWI proteins Piwi and Aubergine (Aub). Once bound, piRNA intermediates are trimmed and 2′*O*-methylated at their 3′ end, forming mature piRNAs (50-53). Mature piRNA-loaded Piwi translocates to the nucleus and associates with Asterix and Panoramix/Silencio for transcriptional silencing of transposons through deposition of repressive chromatin marks (54-59).

piRNA-loaded Aub remains in the nuage where it initiates the secondary ping-pong amplification cycle by recognition and cleavage of cognate transposon mRNA (40, 41, 60). The resulting cleavage product forms the precursor of a secondary sense piRNA that associates with Ago3. piRNA-loaded Ago3 can target and cleave antisense piRNA precursors generating the 5′ end of new sense piRNAs that can be loaded onto Aub, completing the ping-pong amplification cycle (40, 41).

Recent work has demonstrated a preference for uridine at the 5′ position in the binding pocket of the MID (middle) domain of PIWI proteins (61, 62). In combination with the predisposition of Zuc to cleave directly 5′ of uridine residues, this causes Aub to associate predominantly with 1U antisense piRNAs. A subset of PIWI proteins, including Aub and silkworm Siwi, have an additional preference for target RNAs carrying an adenosine directly opposite of the first position of the piRNA (62, 63). As PIWI-mediated cleavage occurs specifically between nucleotide 10 and 11, Ago3-associated sense piRNAs are enriched for adenosine residues at their tenth position (10A). The resulting 1U/10A signature is a characteristic hallmark of secondary ping-pong amplification of piRNAs. Secondary amplification endows the piRNA pathway with specificity, as from a diverse pool of primary piRNAs, only those recognizing active transposons are amplified.

Recent studies have proposed that secondary piRNAs initiate Zuc-dependent production of phased piRNAs (48, 49). Cleavage by Zuc determines the 3′ termini of Aub-associated piRNAs, while the downstream fragment is processed further into Piwi-associated piRNAs by successive Zuc-mediated cleavage events (64). These piRNAs show ~27 nt phasing and a strong 1U bias because of the preference of Zuc to cleave upstream of uridine residues. Phased piRNA production increases the diversity of the piRNA pool and allows adaptation of the piRNA pathway to changes in transposon sequence.

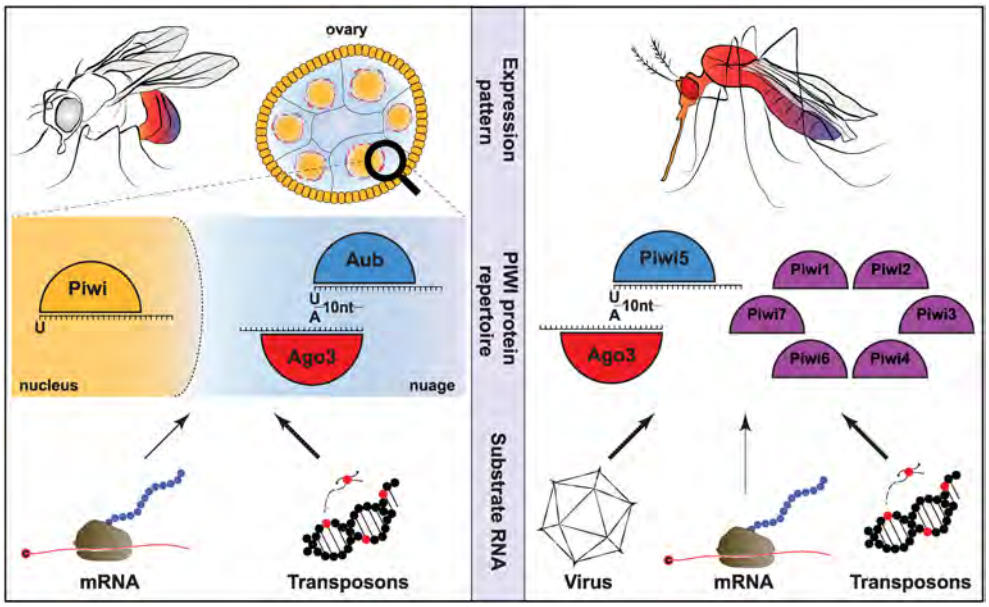


Figure 1. Divergence of piRNA pathways in *Drosophila melanogaster* and *Aedes aegypti*. In *Drosophila* (left panel), PIWI proteins are almost exclusively expressed in gonadal tissues. Nuclear Piwi is expressed in both germ cells and ovarian somatic cells, whereas Aub and Ago3 expression is limited to germ cells specifically. In the *nuage* surrounding the nucleus of these cells, Aub and Ago3 form the ping-pong amplification complex, which is responsible for secondary piRNA production with the characteristic 1U/10A nucleotide biases (Box 1). *Drosophila* piRNAs are mainly derived from transposon sequences and to a lesser extent from mRNA.
In *Ae. aegypti* (right panel), the PIWI gene family is expanded to eight members (Piwi 1–7 and Ago3), some of which are expressed in somatic tissues. Of these PIWI proteins, Piwi5 and Ago3 interact to produce piRNAs with the 1U/10A nucleotide biases indicative of secondary piRNA production through ping-pong amplification. In *Aedes*, piRNAs are produced from viral RNA, in addition to transposon sequences and mRNA.

vpiRNAs in *Aedes* mosquitoes

Initial evidence for vpiRNA production came from the analysis of small RNA deep-sequencing data of the *Drosophila* ovarian somatic sheet (OSS) cells persistently infected with several RNA viruses (65). OSS cells exclusively express Piwi but lack the PIWI proteins that act in the ping-pong amplification machinery. Since Piwi preferentially associates with piRNAs containing a uridine at the first nucleotide position, both sense and antisense vpiRNAs produced in these cells bear a 1U bias (Table 1). However, to date, vpiRNAs have never been found in adult flies. Even infection with Sigma virus, which naturally infects *Drosophila* germ cells, does not give rise to vpiRNAs (66), despite ample expression of PIWI proteins in these cells. In sharp contrast, vpiRNAs are readily detected both in *Aedes* cell lines and in somatic tissues of adult *Aedes* mosquitoes upon infection with several arboviruses, including members of the *Togaviridae* (**Chapter 2**) (26-30), *Flaviviridae* (**Chapter 3**) (31, 32), *Bunyaviridae* (26, 27, 33-35), and *Reoviridae* (33) (Table 1). Besides a typical size distribution of small RNAs around 24–30 nt, piRNAs

from several viruses display the characteristic nucleotide bias indicative of ping-pong amplification (Box 1). Across all virus families, the secondary 10A-biased piRNAs are enriched for the strand with coding capacity, yet the mechanisms responsible for this sorting remain elusive. In addition, vpiRNAs from dengue virus (*Flavivirus* genus, *Flaviviridae* family) and Sindbis virus (*Alphavirus* genus, *Togaviridae* family) have been verified to be 2′*O*-methylated at the 3′ terminal nucleotide (Table 1), a modification that is present on all PIWI-loaded mature piRNAs (Box 1). PIWI-dependence of vpiRNAs has been established for Sindbis (**Chapter 2**), dengue (**Chapter 3**), and Semliki Forest virus (*Alphavirus* genus, *Togaviridae* family) (29) and direct association with PIWI proteins has been demonstrated for Sindbis virus–derived piRNAs (**Chapter 2**).

Determinants of vpiRNA biogenesis

The substrate for the antiviral siRNA pathway, double-stranded RNA, is not abundant in the cytoplasm of healthy, uninfected cells and therefore serves as a danger signal indicating ongoing virus infection (67). In contrast, the substrate for vpiRNA biogenesis is single-stranded viral RNA. It is unknown how PIWI proteins distinguish viral from host RNA and how they determine which of these transcripts are used for piRNA biogenesis. Like cellular mRNAs, single-stranded (+) RNAs of major arbovirus families carry a 5′ cap, which is produced by a virus-encoded capping machinery (flaviviruses and alphaviruses) or obtained through a mechanism termed cap-snatching (bunyaviruses) (68). In contrast to the eukaryotic and flavivirus capping machineries, that of alphaviruses does not deposit 2′*O*-methylation marks at the first two nucleotides downstream of the cap structure (68, 69). Additionally, genomic RNAs of flaviviruses lack the poly-A tail normally present on cellular mRNAs (70). In analogy to innate immune sensors of vertebrates, it is conceivable that the mosquito PIWI proteins specifically recognize such non-self RNA features or that they are recruited to these features by adaptor proteins.

A clue that may help understanding the mechanisms of target selection lies in the genomic distribution of vpiRNAs. While approximately equal levels of viral siRNAs (vsiRNAs) are produced along arbovirus genomes, vpiRNA production is mostly confined to specific hotspot regions. In alphaviruses such as Sindbis, chikungunya, and Semliki Forest virus, vpiRNAs are predominantly produced from a subgenomic RNA that is transcribed from an internal promoter sequence (Figure 2A). This may be due to higher expression of subgenomic relative to genomic RNA. For example, for Sindbis virus it has been shown that the subgenomic promoter yields an excess of subgenomic RNA compared to full length genomic RNA (71, 72). Furthermore, subgenomic ssRNA may be more accessible for the piRNA machinery because it is required for translation of the structural proteins at later stages of the infection. However, these hypotheses do not explain why alphavirus-derived piRNAs are not uniformly distributed over the length of the subgenomic RNA but rather display very discrete hotspots in the 5′ region of the capsid gene (Figure 2A; see also **chapter 2**). One mechanism that could

Table 1: vpiRNA production in insect viruses

Virus family	Name	Genus	Genome	Host/cells*	Nucleotide and (strand) biases**	3' end modification	PIWI protein dependent
Togaviridae	**Sindbis virus**	*Alphavirus*	+ssRNA	Aag2, U4.4, C6/36	1U (-), 10A (+)	yes	Piwi5/Ago3 in Aag2 cells***
	chikungunya virus	*Alphavirus*	+ssRNA	*Ae. aegypti; Ae. albopictus* (soma); U4.4, C6/36, C7-10	1U (-), 10A (+)	n.a.	n.a.
	Semliki Forest virus	*Alphavirus*	+ssRNA	Aag2, U4.4	1U (-), 10A (+)	n.a.	Loss of vpiRNAs upon combined knockdown of Piwi1-7 and Ago3 in Aag2 cells
Flaviviridae	**dengue virus, serotype 2**	*Flavivirus*	+ssRNA	*Ae. aegypti;* Aag2, C6/36	10A (+)	yes	Piwi5, Ago3 and to a lesser extent Piwi6 in Aag2 cells
	cell fusing agent virus	*Flavivirus*	+ssRNA	Aag2, C6/36	10A (+)	n.a	n.a.
Bunyaviridae	**La Crosse Virus**	*Orthobunya-virus*	-ssRNA3 segments	C6/36	1U (-), 10A (+)	n.a.	n.a.
	Schmallen-berg virus	*Orthobunya-virus*	-ssRNA3 segments	KC, Aag2	1U (-), 10A (+)	n.a.	n.a.
	Rift Valley fever virus	*Phlebovirus*	-ssRNA3 segments	Aag2, U4.4, C6/36	1U (-), 10A (+)	n.a.	n.a.
	Phasi Charoen-like virus	*unclassified*	-ssRNA3 segments	*Ae. aegypti*	1U (-), 10A (+)	n.a.	n.a.
Reoviridae	**bluetongue virus**	*Orbivirus*	dsRNA10 segments	KC, Aag2	n.a.	n.a.	n.a.
Dicistroviridae	**Drosophila C virus**	*Cripavirus*	+ssRNA	OSS	1U	n.a.	n.a.
Nodaviridae	**American nodavirus**	*Alphanoda-virus*	+ssRNA2 segments	OSS	1U	n.a	n.a.

n.a., not analyzed

*Aag2 cells are derived from *Ae. aegypti* mosquitoes; U4.4, C6/36 and C7-10 cells are derived from *Ae. albopictus* mosquitoes; KC cells are derived from *Culicoides sonorensis;* OSS cells are derived from the ovarian somatic sheet of *Drosophila melanogaster*.

** The strand orientation is defined in relation to translation; (+) refers to the sense strand with coding potential, (-) refers to the antisense strand. For –ssRNA viruses this reflects the antigenome and genome, respectively.

*** (+) strand piRNAs associate with Ago3 and (-) strand piRNAs associate with Piwi5.

underlie this pattern is processing of abortive viral RNA transcripts by the piRNA machinery. Incomplete viral transcripts are not protected by RNA replication or translation machineries and may therefore represent easily accessible substrates for vpiRNA production. Alternatively, RNA sequences or structural elements may recruit piRNA biogenesis factors to specific regions of the viral genomes. Recently, Homolka *et al.* described such a piRNA-trigger sequence (PTS) in the *Drosophila flamenco* locus, which evokes piRNA biogenesis independent of its genomic context. However, whether this PTS is a structural motif or harbors a small, as-yet unrecognized sequence motif remains to be unraveled (73). Similarly, Ishizu *et al.* identified a *cis*-

acting, 100-nt fragment in the 3′UTR of the piRNA-producing gene *traffic jam* that triggers piRNA production when expressed from unintegrated plasmid DNA. These plasmid-derived piRNAs were efficient in transcriptional silencing of endogenous genes (74). In light of these data, it would be interesting to test whether vpiRNA hotspot sequences promote piRNA production when placed outside their viral context.

piRNA hotspots in flavivirus genomes, including dengue and cell fusing agent virus, differ considerably from those in alphaviruses. Flavivirus piRNAs mostly derive from few very discrete hotspots, sometimes representing single sequences (Figure 2B). The nature of these piRNA spikes remains obscure but this difference strongly suggests that the mechanisms underlying alphavirus and flavivirus piRNA biogenesis are fundamentally different.

Common to alphavirus and flavivirus piRNAs is their extreme strand bias towards sequences from the viral sense strands. In sharp contrast, bunyavirus piRNAs are produced from both antigenomic and genomic strands at a more equal ratio (Figure 2C). It is currently unclear whether this reflects the differences in the replication strategies of alphaviruses and flaviviruses (both +ssRNA viruses) compared to bunyaviruses (-ssRNA

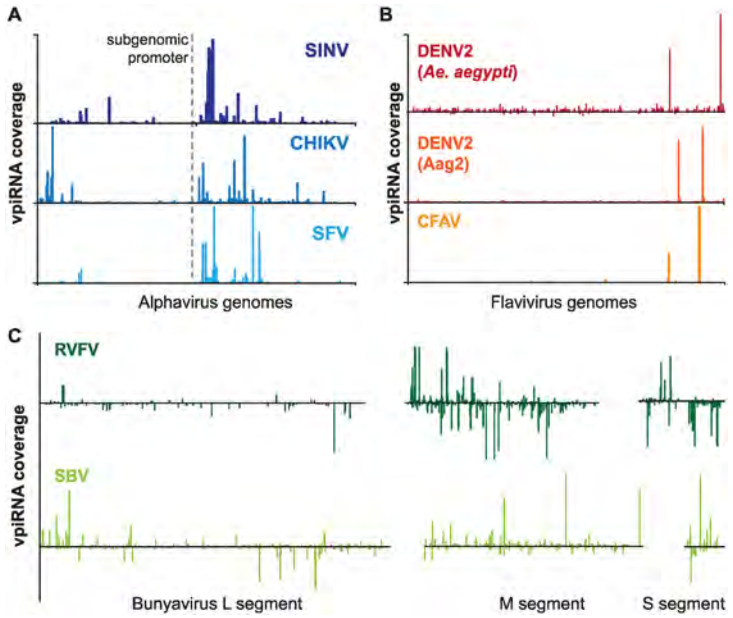


Figure 2. Viral piRNA profiles. piRNA distributions across the genomes of selected **(A)** alphaviruses, **(B)** flaviviruses, and (**C)** bunyaviruses. The plots depict published genome profiles of Sindbis virus (SINV) (27), chikungunya virus (CHIKV) (28), Semliki Forest virus (SFV) (29), dengue virus serotype 2 (DENV2) (**Chapter 3**) and (32), cell fusing agent virus (CFAV) (31), Rift Valley fever virus (RVFV) (34), and Schmallenberg virus (SBV) (33). For alphaviruses, the position of the subgenomic promoter is depicted. The piRNA coverage on the sense or antisense strand is shown as peaks above or below the x-axis, respectively. Please note that the plots are representations of piRNA profiles from multiple studies that used different ways of normalizing and presenting read counts. Therefore, the heights of the bars are arbitrary and do not allow a quantitative comparison between the different viruses.

virus) or if this is due to variations in the piRNA machinery acting on RNAs of distinct viruses. These observations clearly underscore the need for a comprehensive analysis of *cis*- and *trans*-acting factors required for the piRNA biogenesis from arboviruses of all families.

Biogenesis of vpiRNAs

Functional diversification of *Aedes* PIWI proteins after gene duplication in combination with somatic expression are likely the main drivers of the expansion of piRNA substrates, including viral RNA. *Ae. aegypti* Piwi4, Piwi5, Piwi6, and Ago3 are abundantly expressed in somatic tissue of adult mosquitoes (**Chapter 6**) (39) and *Ae. aegypti* Aag2 cells (27). In an RNAi screen targeting individual PIWI proteins in Aag2 cells, Piwi5 and Ago3 were identified as the main players for vpiRNA production from Sindbis virus (**Chapter 2**). Piwi5 and Ago3 bind vpiRNAs from opposite strands and with distinct nucleotide biases. Whereas Piwi5 binds 1U-biased antisense piRNAs, Ago3 binds 10A-biased piRNAs derived from the viral sense strand. These observations suggest a model in which ping-pong amplification is initiated by Piwi5-bound primary piRNAs from the Sindbis virus antisense strand. Cleavage of the sense strand by Piwi5 results in the production of secondary sense strand piRNA precursors that are loaded into Ago3 (Figure 3). Knockdown of Piwi5 and Ago3 and to a lesser extent, Piwi6 results in reduced vpiRNA production from dengue virus serotype 2 in Aag2 cells (**Chapter 3**). The additional requirement of Piwi6 specifically for dengue virus piRNA biogenesis suggests that *Aedes* PIWI proteins have specialized in processing distinct RNA sources. This is further supported by the differential requirement of PIWI proteins for the processing of transposon-derived piRNAs that, in contrast to Sindbis virus–derived piRNAs, directly or indirectly relies on all somatic *Aedes* PIWI proteins (**Chapter 2**) (Figure 3). Future research should define to what extent vpiRNA production relies on similar or distinct PIWI family members for viruses within the same virus family and between different virus families. Of special interest are bunyaviruses, for which PIWI dependency thus far has not been studied, despite the fact that these viruses represent the largest arbovirus family (2).

The piRNA pathway acts in resistance and tolerance to virus infections

Arboviruses establish persistent infections in mosquitoes and replicate to high levels without causing apparent fitness loss in their vectors. Such a defense strategy in which high pathogen levels are tolerated and the focus lies on preventing infection-induced damage has been termed tolerance. In contrast, actively restricting virus growth and potentially clearing the infection is a defense strategy called resistance (6). Although a comprehensive model for vpiRNA function is still lacking, there is good evidence that the piRNA pathway is implicated in both strategies. For example, it was shown that upon

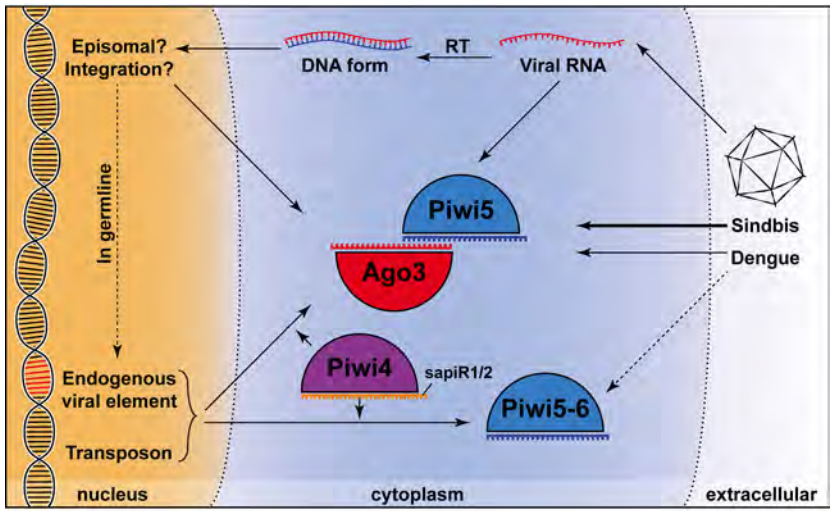


Figure 3. Model for piRNA biogenesis in *Aedes aegypti*. RNA molecules from various sources are processed differently by the piRNA machinery in *Ae. aegypti*. Upon acute infection, Sindbis virus RNA is processed into ping-pong–dependent piRNAs by the PIWI proteins Piwi5 and Ago3. In contrast, dengue virus RNA can also be processed into piRNAs by Piwi6. Transposon-derived piRNAs associate primarily with Piwi5 and Piwi6; however, some transposon RNAs feed into the ping-pong loop and give rise to Ago3-bound secondary piRNAs. Additionally, the production of transposon piRNAs is dependent on Piwi4 in an indirect manner, as transposon-derived piRNAs are not loaded in Piwi4, but knockdown of Piwi4 does reduce their numbers. Instead, Piwi4 associates with abundant piRNAs derived from a conserved satellite DNA locus (sapiR1 and sapiR2)
Viral RNA may directly enter the piRNA machinery; additionally, viral RNA is reverse transcribed to produce a DNA form of the virus (vDNA). The vDNA may either remain episomal or integrate into the host genome. Putative vDNA-derived transcripts may serve as additional precursors for vpiRNA production. Moreover, when genome integration occurs in the germline, the vDNA fragment forms a novel endogenous viral element (EVE) that may lead to the production of EVE-derived piRNAs.

knockdown of Piwi4 in *Ae. aegypti* Aag2 cells, replication of Semliki Forest virus is strongly enhanced (29). Yet, this resistance seems to be independent of vpiRNA production, as Piwi4 depletion does not cause reduction of vpiRNA levels (29). In line with this observation, immunoprecipitation of Piwi4 in Aag2 cells infected with a related alphavirus (Sindbis virus) is depleted of vpiRNAs (**Chapter 2**). Therefore, the molecular mechanism by which Piwi4 exerts its antiviral activity remains to be investigated. Knockdown of Piwi5 and Ago3 in Aag2 cells results in profound decline in vpiRNA expression from dengue virus, but viral replication is not strongly affected (**Chapter 3**). Whether PIWI depletion in adult mosquitoes causes enhanced arbovirus replication remains to be shown.

Interestingly, in mosquito cells infected with Rift Valley fever virus (*Phlebovirus* genus, *Bunyaviridae* family), vpiRNAs are primarily detected late in infection following a first wave of vsiRNAs. The vpiRNAs vastly outnumber vsiRNAs at 72 hours postinfection (34). These data suggest that vpiRNAs may exert their function primarily late during Rift Valley fever virus infection or during the establishment of a persistent infection.

Similarly, Goic *et al.* show that ping-pong–amplified piRNAs are present nine days after infection of *Ae. albopictus* mosquitoes with chikungunya virus, yet that population is not seen at three days postinfection (30). In contrast, mosquitoes infected with dengue virus type 2 show the highest accumulation of vsiRNAs at nine days postinfection, whereas piRNA-sized reads are the dominant population at two days postinfection (32). On the whole, it is currently unclear how differential accumulation of vsiRNAs and vpiRNAs during the course of infection shapes the immune response in mosquitoes. An intriguing possibility is that the ratio of these two classes of small RNAs is important for the transition from an acute defense mechanism to the establishment of a persistent infection.

In line with this idea, Goic *et al.* have proposed a model through which the mosquito piRNA pathway may regulate tolerance against dengue and chikungunya virus in *Aedes* mosquitoes during persistent infections (30). Central to the proposed mechanism is the production of piRNAs from a viral DNA form (vDNA) of these cytoplasmic RNA viruses (Figure 3). Unlike retroviruses, these viruses do not encode their own reverse transcriptase necessary for the generation of a DNA form. Instead, it is thought that cDNA production depends on the reverse transcription activity of endogenous retrotransposons, a mechanism that has been demonstrated previously in *Drosophila* (75). Administration of a reverse transcriptase inhibitor causes reduction of both vsiRNA and vpiRNA levels, suggesting that a viral cDNA form is required for the establishment of effective small RNA responses. Mosquitoes treated with reverse transcriptase inhibitors die faster after virus inoculation without a strong increase in viral loads. Therefore, the authors conclude that the production of viral cDNA is important for tolerance to virus infection (30). Yet, the molecular mechanisms linking vpiRNA production and this tolerance phenotype require further investigation. It is possible that vDNA, either integrated in the host genome or existing as episomal sequences, give rise to aberrant transcripts that are processed into piRNAs. Additionally, genomic integration of vDNA close to transposable elements may favor recognition of vDNA-derived transcripts by the piRNA machinery. Supporting the latter, Non-retroviral Integrated RNA Virus Sequences (NIRVS) present in the genome of *Ae. aegypti* and *Ae. albopictus* are mostly proximal to transposable elements (**Chapter 5**).

Many viruses have developed strategies to evade or interfere with antiviral pathways. For instance, several insect viruses have evolved mechanisms to suppress various steps of the antiviral siRNA pathway (8, 76). Likewise, if the piRNA pathway exerts strong antiviral activity, it is likely that arboviruses have evolved suppressors of piRNA biogenesis and function. Introduction of the gene encoding the Flock House virus B2 protein, an established suppressor of the siRNA pathway, into the chikungunya genome results in a slight decrease of vpiRNA levels (28). Whether this is due to direct interference with the piRNA pathway or to indirect effects (for example, by affecting RNA abundance or accessibility) remains unclear.

piRNAs and endogenous viral elements: heritable immune memory?

The canonical function of piRNAs is to provide heritable immunity against transposable elements. The piRNA machinery is able to adapt to newly acquired transposable elements when these integrate into genomic piRNA clusters (77). In germ cells, these integration events are heritable and therefore provide an evolutionary benefit. It is an intriguing hypothesis that the piRNA pathway in mosquitoes, besides providing memory of transposon encounters, may establish heritable immunity against circulating viruses.

Strikingly, remnants of cytoplasmic RNA virus genomes are frequently integrated in genomes of host species, thus providing a record of previous virus encounters (78, 79). These endogenous viral elements (EVE) may contribute to antiviral immunity in both invertebrates and vertebrates. For example, the genome of the ground squirrel accommodates a large number of endogenous bornavirus-like N elements (EBLN), which are commonly integrated in mammalian genomes (80). Some EBLNs contain intact open reading frames, and expression of the encoded proteins interferes with infection with a related virus (81). Besides the expression of viral proteins from EBLNs, piRNAs have recently been hypothesized to contribute to the EBLN-mediated immunity in the mammalian germline (82).

Strikingly, *Aedes* genomes contain a large number of EVEs, mostly NIRVS, some of which are annotated as protein-coding ORFs in the published genome assembly (**Chapters 4 and 5**) (45, 83-86). PCR-based surveys show that mosquito populations differ in NIRVS content, indicating that NIRVS may be dynamically acquired and stably inherited to the next generation (83-85). Intriguingly, mosquito NIRVS are a prominent source of piRNAs (45). These piRNAs are mostly antisense to the orientation of the putative viral ORFs, suggesting an evolutionary benefit in retaining NIRVS that produce piRNAs with the potential to target cognate viral protein-coding RNA (**Chapters 4 and 5**). Yet, the extent to which these NIRVS-derived piRNAs represent a heritable antiviral immune memory needs to be explored.

Interestingly, RNA-mediated antiviral resistance had previously been demonstrated in adult mosquitoes and cells. Expression of genome segments of dengue or La Crosse virus prior to infection with the same viruses interfered with virus replication (72, 87-89). Mutagenesis of in-frame start codons in the expressed viral sequence did not alter this resistance phenotype, indicating that it was mediated by RNA (88). Moreover, the expression of viral sequences provided partial cross-protection, since replication of related viruses but not viruses from a distinct family was inhibited (87, 88). Similarly, in an attempt to gain siRNA-mediated immunity against dengue virus, Adelman *et al*. generated clonal C6/36 cell lines harboring a plasmid-encoded inverted repeat to produce dsRNA targeting the dengue prM gene. A highly resistant cell line was obtained, and the authors attributed this resistance phenotype to the production of viral siRNAs. Indeed, production of small RNAs with dengue sequences was shown by

northern blotting (90). However, later studies found that C6/36 cells are Dicer-2 deficient and therefore incapable of producing siRNAs (26). It is tempting to speculate that the observed dengue resistance was in fact mediated by piRNAs.

Another small RNA-mediated pathway that provides immune memory through integration of foreign genetic information into the genome is the CRISPR-Cas system. In the prokaryotic CRISPR system, short spacer sequences derived from foreign genetic material are incorporated in designated genomic loci. These spacer sequences guide CRISPR-associated (Cas) proteins to exogenous target sequences and as such provide heritable immunity against viruses and plasmids (91). The piRNA pathway has many similarities with the CRISPR system; in both systems, exogenous nucleic acid sequences are found in specific clusters, which produce small RNAs that guide proteins with endonucleic activity to cognate target sequences (24, 45, 92). Despite their obvious similarities, there are also major differences between the two RNA-guided silencing pathways. While in the CRISPR system newly acquired spacers are incorporated in an orderly fashion, incorporation of novel sequences into piRNA clusters depends on retrotransposon activity and appears to be random. Hence, adaptation to new threats is thought to be less efficient in piRNA clusters than in CRISPR loci (77). Nonetheless, the possibility that piRNA clusters may encode a heritable immune memory in vector mosquitoes similar to the prokaryotic CRISPR system is intriguing and solicits further investigation.

vpiRNAs in other host species

Whereas vpiRNAs can be readily detected in *Aedes* mosquitoes and cell lines, vpiRNAs have thus far not been reported in important blood-feeding mosquito vectors from the *Anopheles* and *Culex* genera. The *Anopheles* gambiae genome encodes, like *Drosophila*, two orthologs of Piwi/Aub and one copy of Ago3. The *Cx. quinquefasciatus* PIWI gene family, however, has undergone expansion to seven members (36, 37).

Infection of *An. gambiae* with o'nyong-nyong virus (*Alphavirus* genus, *Togaviridae* family) does not give rise to an abundant population of piRNA-sized small RNAs (93). Yet, in this study, the authors did not analyze additional piRNA features of the small amount of piRNA-sized reads in the sequencing libraries, making it hard to conclusively exclude low-level vpiRNA production. Since related viruses give rise to ping-pong amplified vpiRNAs in *Aedes* mosquitoes, it would be interesting to investigate whether a ping-pong signature is also present for o'nyong-nyong piRNA-sized reads. This may also provide an explanation for the observed increase of o'nyong-nyong virus upon depletion of Ago3 in *An. gambiae* mosquitoes (15).

Small RNA deep-sequencing in *Cx. pipiens* mosquitoes infected with West Nile virus (WNV) or Usutu virus (*Flavivirus* genus, *Flaviviridae* family) did not uncover vpiRNAs, whereas vsiRNAs were readily detected (94, 95). Whether this is due to *Cx. pipiens* being

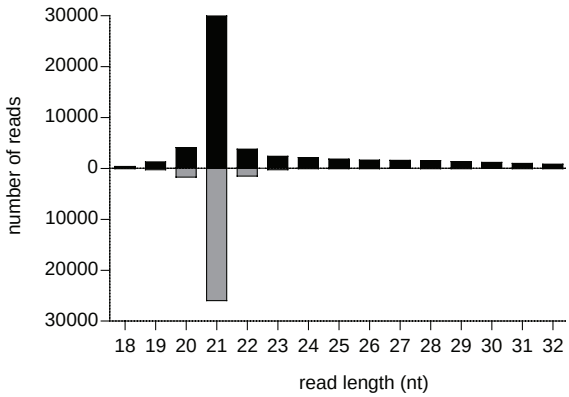


Figure 4. Size profile of Sindbis virus small RNAs in *Culex pipiens*. *Cx. pipiens* mosquitoes were infected with 9,660 TCID50 Sindbis virus (pTE 3'2J-GFP) by intrathoracic injection. Two days postinfection, RNA was extracted from the mosquitoes using Isol-RNA lysis reagent. Small RNAs were isolated by gel-electrophoresis, and deep-sequencing libraries were prepared using Illumina's Truseq small RNA preparation kit. Small RNA libraries were then sequenced on an Illumina Hiseq2500 system and mapped to the Sindbis virus genome. The size distribution of viral small RNAs derived from the sense strand (black) or antisense strand (grey) is depicted for sequencing reads that align to the genome with a maximum of one mismatch in the first 28 nt. The size profile suggests that *Cx. pipiens* does not produce vpiRNAs, but it cannot be excluded that vpiRNAs are found when using a different route of inoculation, at other time points, or in infections with other viruses. Deep-sequencing data have been deposited in the NCBI Sequence Read Archive under accession number SRA486748.

unable to produce vpiRNAs or the inability of WNV to trigger vpiRNA production is unclear, especially as WNV also failed to induce vpiRNA production in *Ae. albopictus* C6/36 cells, which are competent in producing vpiRNAs from other flaviviruses. In contrast, Sindbis virus infection of *Aedes* cells gives rise to an abundant population of vpiRNAs (**Chapter 2**) (27) yet fails to induce vpiRNA production in *Culex* mosquitoes (Figure 4). Thus, although PIWI gene duplications have occurred both in *Aedes* and *Culex*, only *Aedes* PIWI proteins seem to support efficient vpiRNA biogenesis. A possible explanation for this discrepancy would be that *Culex* PIWI genes are not co-expressed with viral RNA in somatic cells. Alternatively, viral RNA might not be a favorable substrate for *Culex* PIWI proteins. Future research will have to characterize to what extent vpiRNA production is supported in different hematophagous mosquito species.

The piRNA pathway is not frequently studied in insects other than mosquitoes, the silkworm and fruit flies. Nevertheless, PIWI gene duplication and somatic expression of PIWI proteins has been observed in the pea aphid *Acyrthosiphon pisum* (96). This indicates that there is potential for functional innovation and perhaps viral piRNA biogenesis beyond mosquitoes. Likewise, although PIWI proteins are generally highly expressed in germline tissues in vertebrates, emerging evidence suggests that PIWI proteins may also be expressed in somatic cells including neurons, cancer cells, and stem cells (97, 98). However, it is not yet known whether these somatically expressed PIWI proteins are capable of targeting viral RNA.

Determinants for the biogenesis of endogenous piRNAs

Besides producing piRNAs from arbovirus RNA, **chapters 4 to 6** describe a versatile collection of endogenous, non transposon-derived piRNAs both in adult *Ae. aegypti*

mosquitoes and in mosquito cells, amongst which piRNAs from protein coding mRNAs (genic piRNAs). In *Drosophila* genic piRNAs are predominantly derived from 3′ UTR sequences and are dependent on primary biogenesis and rarely undergo ping-pong amplification (99, 100). In contrast, *Ae. aegypti* genic piRNAs are not biased towards the 3′ UTR and are frequently produced from the protein coding sequence. Moreover, several mRNAs generate piRNAs in a ping-pong dependent manner, amongst which the replication-dependent histones, primarily histone H4 (**Chapter 4**). A lowly expressed antisense transcript is the likely source of primary antisense piRNAs that initiate ping-pong amplification.

It is currently unclear why certain mRNAs are specifically licensed or precluded from piRNA biogenesis. Similar to what has been discussed above for vpiRNAs, it is perceivable that special RNA elements akin to piRNA trigger sequences in *Drosophila* direct the piRNA machinery to selected transcripts (73, 74). Alternatively, piRNA-producing genes may be defined by a specific chromatin environment, analogous to dual-strand piRNA clusters in *Drosophila*. These germline specific piRNA loci are characterized by the presence of the Rhino-Deadlock-Cutoff complex (**Chapter 1**), which is responsible for marking precursor transcripts for piRNA production (101-103). Currently, there are no data on the chromatin state or the presence of specific histone modifications at loci of endogenous piRNA production in *Ae. aegypti*. Since formation of heterochromatin is essential for piRNA biogenesis in the fly (101-104), it would be important to investigate the relationship between the chromatin landscape and the piRNA biogenesis machinery in greater detail in vector mosquitoes.

RNA elements or specific features at the chromatin level may also underlie the selective production of a highly abundant piRNA population from the *satDNA1* locus (**Chapter 6**). Our data suggest that transcriptional read-through from a neighboring gene is unlikely to generate the piRNA precursor transcript. Which alternative mechanism recruits the transcription and piRNA biogenesis machineries to *satDNA1*, remains to be shown. Satellite DNA is the major constituent of heterochromatin (105) and perhaps a specific heterochromatin-binding protein, analogous to Rhino in flies (101-103), is required for *satDNA1* expression and the subsequent loading of the piRNA precursor specifically into Piwi4. The function *of satDNA1* is currently unknown but its evolutionary conservation over 200 million years suggests an important role in *Culicinae* mosquitoes, which includes major vector mosquitoes of the *Aedes* and *Culex* families (**Chapter 6**) (85, 106). A biological function exerted by a unique piRNA sequence may be unexpected, considering that piRNAs normally act as a population of sequences that adapt to and target transposable elements (40, 77, 107). Yet this paradigm has previously been challenged when a single piRNA sequence was discovered to be crucial for sex determination in the silkworm (108).

Dissecting the *Aedes aegypti* piRNA pathway – more protein families, more mechanisms

The powerful genetics toolbox available for *D. melanogaster* has facilitated the discovery of numerous PIWI co-factors that are indispensible for piRNA biogenesis and function (24, 46, 109). Many of these belong to the family of Tudor proteins and accordingly, several mosquito Tudor proteins have been identified that are involved in piRNA biogenesis (**Chapter 7**). Auxiliary proteins aid in stabilizing RNA and protein interactions, prevent non-specific degradation of RNAs, and help funneling specific RNA precursors into the different braches of the piRNA pathway. The latter is probably best exemplified by the dynamic association of silkworm Siwi with its co-factors Vasa, Spindle-E and Qin (110). In a complex with Spindle-E and Qin, Siwi functions in primary piRNA biogenesis whereas binding to Vasa is required for efficient piRNA amplification in the ping-pong loop. Qin/Kumo also prevents Ago3 bound piRNA precursors from becoming Piwi-bound piRNAs during piRNA phasing in *Drosophila* (64), providing another example of a PIWI co-factor that aids in selecting piRNA precursors for specific PIWI proteins. Interestingly, the *Ae. aegypti* Tudor protein AAEL012441 is required for vpiRNA biogenesis but has almost no effect on the production of transposon and histone H4 piRNAs, suggesting that sorting mechanisms exist that can discriminate the origin of the piRNA precursor (**Chapter 7**).

Besides PIWI and Tudor proteins, a number of additional protein families such as RNA helicases, nucleases or protein chaperones are active in the *Drosophila* piRNA pathway (24, 46, 109) and almost nothing is known about their involvement in piRNA production in mosquitoes. Zucchini, a core enzyme in the primary biogenesis pathway and during piRNA phasing (47-49) also awaits genetic identification in *Aedes*. Recent analyses of transposon and viral piRNAs in Aag2 cells indicated that their 3' end is generated, at least partially, by an endonuclease with a preference for cleaving upstream of Us, comparable to Zucchini (unpublished observations). Whether piRNA 3' end cleavage by a putative Zucchini ortholog also triggers piRNA phasing, as in flies, is thus far unclear. Similarly, it is currently not known whether a nuclear branch of the piRNA pathway exists in mosquitoes that could silence expression of transposons, NIRVS, or genes targeted by piRNAs at the transcriptional level.

For the genetic and biochemical dissection of the piRNA pathway, mosquito cells are a powerful tool (111). *Ae. aegypti* Aag2 cells fully recapitulate primary and secondary piRNA biogenesis and they express the PIWI proteins that are also expressed in somatic tissues of adult mosquitoes (39). Mosquito cells are experimentally more amendable than adult mosquitoes, especially for screening purposes. In particular, loss-of-function mutants of piRNA pathway components may cause sterility in mosquitoes, significantly impeding genetic studies *in vivo*. Also for investigating the *Drosophila* piRNA pathway, the value of cell lines that support both primary and secondary biogenesis pathways

has been appreciated. Recently, an ovarian somatic sheet cell line, which normally only produces Piwi-dependent primary piRNAs was genetically engineered to also support ping-pong amplification (112). Besides being fully piRNA competent, *Aedes* cell lines can be infected with important human arboviruses including dengue and chikungunya (111), making them excellent model systems to dissect the mosquito piRNA pathway and its interactions with arboviruses. The advent of genome editing technologies for vector mosquitoes makes it possible, at least for mutants that do not cause lethality or sterility, to investigate the relevance of findings obtained in cell culture *in vivo* (113, 114).

The piRNAs pathway – a potential target for transmission control strategies?

In order to qualify as competent vectors for arboviruses, mosquitoes need to take up virus in a blood meal, support virus amplification in secondary organs, and horizontally transmit virus particles to naive hosts (**Chapter 1**) (115, 116). Insights into the immune pathways that modulate vector competence are crucial to understand or predict arbovirus transmission by mosquitoes and to develop novel transmission-blocking strategies (117). Genetic approaches that seek to introduce a heritable trait that reduces vector competence into a mosquito population are currently developed (117, 118). Most commonly, genetically modified insects that do not sufficiently support virus uptake, growth or transmission are released into a target area to spread this modification while reproducing. An alternative strategy aims at reducing the size of mosquito populations near human dwellings by introducing genetic traits that cause lethality or sterility (118). A prominent example is the release of sterile males carrying a genetic modification that causes death of the offspring, thereby affecting the number of competent vectors in the target area (119). Over time, these traits, especially those that cause lethality and sterility, are prone to disappear from the population since no progeny will emerge that carry the modification. Also heritable traits that aim at reducing vector competence are likely outcompeted in wildtype populations, since only a minority of mosquitoes is infected with arboviruses and infections usually have little or no impact on mosquito fitness (4). Therefore, on the population level, the fitness costs likely associated with carrying the genetic modification outweigh the benefit of enhanced virus resistance or lower tolerance (120, 121). Repeated releases of modified insects are thus required to retain the new genetic trait at high frequency in the population. In contrast, strategies that employ the autonomous spread of 'selfish' transgenes aim for indefinite persistence of a genetic modification. In its most invasive form, these so called gene-drive systems favor the replacement of the wildtype allele by the new genetic trait in the mosquito germline, creating homozygous transgenic insects at a frequency that is higher than predicted by Mendelian inheritance (122, 123) Their autonomous and potentially uncontrolled spread makes gene-drive strategies more controversial and risky to implement (118, 124).

Whether the *Aedes* piRNA pathway is a relevant target for genetic transmission blocking strategies remains to be carefully investigated. We are currently lacking fundamental knowledge on how the different classes of endogenous and viral piRNAs influence vector competence. It remains an intriguing observation that the largest expansion of the piRNA pathway has happened in the major arboviral vector mosquitoes of the *Aedes* and *Culex* genera (36, 37). Moreover, the genomes of *Ae. aegypti* and *Ae. albopictus* have the highest number of piRNA producing NIRVS (**Chapter 5**) and both *Aedes* and *Culex* genera encode the piRNA producing *satDNA1* locus (**Chapter 6**). In contrast, most *Anopheles* mosquitoes have three, in some species four, PIWI orthologs (36), NIRVS are rare, and *satDNA1* is absent from the genome. Currently, only one arbovirus, O 'nyong-nyong virus, is known to be transmitted by *Anopheles* mosquitoes (15, 125), whereas *Aedes* and *Culex* mosquitoes are competent vectors for a much larger number of arboviruses including the most important human vector-borne pathogens (1-3). It is an intriguing possibility that the expansion of the piRNA pathway and/or the accumulation of specific viral or endogenous piRNAs have contributed to the increase in vector competence in *Culex* and *Aedes* mosquitoes. Whether the piRNA pathway modulates vector competence in these species is an important question to address in the future. The work presented in this thesis has shed the first light on the mechanisms that underlie the production of host and viral piRNAs in the major arboviral vector *Ae. aegypti.* These insights will be beneficial for future studies aiming at elucidating the role of the piRNA pathway in vector competence and arbovirus transmission.

Open questions

Despite the progress in our understanding of vpiRNA biogenesis and function, many important questions remain: (i) Which mosquito species are capable of producing vpiRNAs and which viruses elicit a piRNA response? In relation to these questions, future research should investigate to what extent the piRNA pathway determines vector competence and the specificity of arboviruses for certain mosquito species. (ii) What is the composition of macromolecular complexes required for piRNA production from various RNA sources? It is of particular importance to investigate which PIWI proteins are required for piRNA production from different arboviruses as well as from transposons and other endogenous sources. Also, the contribution of additional protein families to piRNA biogenesis and function warrants investigation. (iii) What is the role of the mosquito piRNA pathway in mediating resistance to and tolerance for arbovirus infections? (iv) What is the contribution of NIRVS to antiviral immunity and immune memory in mosquitoes? (v) Have arboviruses developed strategies to evade or interfere with the piRNA pathway? (vi) What are the determinants at the RNA and chromatin level that define vpiRNA hotspots and endogenous loci of piRNA production? (vii) How widely does somatic piRNA expression occur across the tree of life, and has piRNA-

mediated gene silencing acquired additional functions beyond transposon control in other animal species? Somatic piRNA biogenesis in *Aedes* mosquitoes, in particular the production of vpiRNAs, shows that the piRNA pathway is much more versatile than previously anticipated. It remains to be seen how many more surprises PIWI proteins have in store when we take a closer look at this fascinating pathway in other species.

ACKNOWLEDGEMENTS

Parts of this discussion have been published as *Miesen P., Joosten J. and van Rij R.P. (2016) PLoS Pathog. 12(12): e1006017.* The authors would like to thank members of the Van Rij lab for fruitful discussions. A special thanks to Erika Girardi and Finny Varghese for their valuable input on this manuscript. Sindbis virus infections of *Cx. pipiens* mosquitoes were performed by Jelke Fros in the laboratory of Gorben Pijlman, Wageningen University, The Netherlands. Sarah Merkling (Pasteur Institute, Paris, France) kindly provided the drawings of the mosquito and fruit fly in Figure 1.

REFERENCES

1. Weaver SC, Reisen WK. Present and future arboviral threats. Antiviral Res. 2010;85(2):328-45.
2. Gubler DJ. Human arbovirus infections worldwide. Ann N Y Acad Sci. 2001;951:13-24.
3. Kraemer MU, Sinka ME, Duda KA, Mylne AQ, Shearer FM, Barker CM, et al. The global distribution of the arbovirus vectors Aedes aegypti and Ae. albopictus. Elife. 2015;4:e08347.
4. Lambrechts L, Scott TW. Mode of transmission and the evolution of arbovirus virulence in mosquito vectors. Proc Biol Sci. 2009;276(1660):1369-78.
5. Putnam JL, Scott TW. Blood-feeding behavior of dengue-2 virus-infected Aedes aegypti. Am J Trop Med Hyg. 1995;52(3):225-7.
6. Cheng G, Liu Y, Wang P, Xiao X. Mosquito Defense Strategies against Viral Infection. Trends Parasitol. 2016;32(3):177-86.
7. Merkling SH, van Rij RP. Beyond RNAi: antiviral defense strategies in Drosophila and mosquito. J Insect Physiol. 2013;59(2):159-70.
8. Bronkhorst AW, van Rij RP. The long and short of antiviral defense: small RNA-based immunity in insects. Curr Opin Virol. 2014;7:19-28.
9. Gammon DB, Mello CC. RNA interference-mediated antiviral defense in insects. Curr Opin Insect Sci. 2015;8:111-20.
10. Wang XH, Aliyari R, Li WX, Li HW, Kim K, Carthew R, et al. RNA interference directs innate immunity against viruses in adult Drosophila. Science. 2006;312(5772):452-4.
11. Galiana-Arnoux D, Dostert C, Schneemann A, Hoffmann JA, Imler JL. Essential function in vivo for Dicer-2 in host defense against RNA viruses in drosophila. Nat Immunol. 2006;7(6):590-7.
12. van Rij RP, Saleh MC, Berry B, Foo C, Houk A, Antoniewski C, et al. The RNA silencing endonuclease Argonaute 2 mediates specific antiviral immunity in Drosophila melanogaster. Genes Dev. 2006;20(21):2985-95.
13. Cirimotich CM, Scott JC, Phillips AT, Geiss BJ, Olson KE. Suppression of RNA interference increases alphavirus replication and virus-associated mortality in Aedes aegypti mosquitoes. BMC Microbiol. 2009;9:49.
14. Campbell CL, Keene KM, Brackney DE, Olson KE, Blair CD, Wilusz J, et al. Aedes aegypti uses RNA interference in defense against Sindbis virus infection. BMC Microbiol. 2008;8:47.

15. Keene KM, Foy BD, Sanchez-Vargas I, Beaty BJ, Blair CD, Olson KE. RNA interference acts as a natural antiviral response to O'nyong-nyong virus (Alphavirus; Togaviridae) infection of Anopheles gambiae. Proc Natl Acad Sci U S A. 2004;101(49):17240-5.
16. Myles KM, Wiley MR, Morazzani EM, Adelman ZN. Alphavirus-derived small RNAs modulate pathogenesis in disease vector mosquitoes. Proc Natl Acad Sci U S A. 2008;105(50):19938-43.
17. Sanchez-Vargas I, Scott JC, Poole-Smith BK, Franz AW, Barbosa-Solomieu V, Wilusz J, et al. Dengue virus type 2 infections of Aedes aegypti are modulated by the mosquito's RNA interference pathway. PLoS Pathog. 2009;5(2):e1000299.
18. Franz AWE, Sanchez-Vargas I, Adelman ZN, Blair CD, Beaty BJ, James AA, et al. Engineering RNA interference-based resistance to dengue virus type 2 in genetically modified Aedes aegypti. Proc Natl Acad Sci U S A. 2006;103(11):4198-203.
19. Marques JT, Imler JL. The diversity of insect antiviral immunity: insights from viruses. Curr Opin Microbiol. 2016;32:71-6.
20. Kakumani PK, Ponia SS, S RK, Sood V, Chinnappan M, Banerjea AC, et al. Role of RNA interference (RNAi) in dengue virus replication and identification of NS4B as an RNAi suppressor. J Virol. 2013;87(16):8870-83.
21. Schnettler E, Sterken MG, Leung JY, Metz SW, Geertsema C, Goldbach RW, et al. Noncoding flavivirus RNA displays RNA interference suppressor activity in insect and Mammalian cells. J Virol. 2012;86(24):13486-500.
22. Moon SL, Dodd BJ, Brackney DE, Wilusz CJ, Ebel GD, Wilusz J. Flavivirus sfRNA suppresses antiviral RNA interference in cultured cells and mosquitoes and directly interacts with the RNAi machinery. Virology. 2015;485:322-9.
23. Asgari S. Role of microRNAs in arbovirus/vector interactions. Viruses. 2014;6(9):3514-34.
24. Hirakata S, Siomi MC. piRNA biogenesis in the germline: From transcription of piRNA genomic sources to piRNA maturation. Biochim Biophys Acta. 2016;1859(1):82-92.
25. Siomi MC, Sato K, Pezic D, Aravin AA. PIWI-interacting small RNAs: the vanguard of genome defence. Nat Rev Mol Cell Biol. 2011;12(4):246-58.
26. Brackney DE, Scott JC, Sagawa F, Woodward JE, Miller NA, Schilkey FD, et al. C6/36 Aedes albopictus cells have a dysfunctional antiviral RNA interference response. PLoS Negl Trop Dis. 2010;4(10):e856.
27. Vodovar N, Bronkhorst AW, van Cleef KW, Miesen P, Blanc H, van Rij RP, et al. Arbovirus-derived piRNAs exhibit a ping-pong signature in mosquito cells. PLoS One. 2012;7(1):e30861.
28. Morazzani EM, Wiley MR, Murreddu MG, Adelman ZN, Myles KM. Production of virus-derived ping-pong-dependent piRNA-like small RNAs in the mosquito soma. PLoS Pathog. 2012;8(1):e1002470.
29. Schnettler E, Donald CL, Human S, Watson M, Siu RW, McFarlane M, et al. Knockdown of piRNA pathway proteins results in enhanced Semliki Forest virus production in mosquito cells. J Gen Virol. 2013;94(Pt 7):1680-9.
30. Goic B, Stapleford KA, Frangeul L, Doucet AJ, Gausson V, Blanc H, et al. Virus-derived DNA drives mosquito vector tolerance to arboviral infection. Nat Commun. 2016;7:12410.
31. Scott JC, Brackney DE, Campbell CL, Bondu-Hawkins V, Hjelle B, Ebel GD, et al. Comparison of dengue virus type 2-specific small RNAs from RNA interference-competent and -incompetent mosquito cells. PLoS Negl Trop Dis. 2010;4(10):e848.
32. Hess AM, Prasad AN, Ptitsyn A, Ebel GD, Olson KE, Barbacioru C, et al. Small RNA profiling of Dengue virus-mosquito interactions implicates the PIWI RNA pathway in anti-viral defense. BMC Microbiol. 2011;11:45.
33. Schnettler E, Ratinier M, Watson M, Shaw AE, McFarlane M, Varela M, et al. RNA interference targets arbovirus replication in Culicoides cells. J Virol. 2013;87(5):2441-54.
34. Leger P, Lara E, Jagla B, Sismeiro O, Mansuroglu Z, Coppee JY, et al. Dicer-2- and Piwi-mediated RNA interference in Rift Valley fever virus-infected mosquito cells. J Virol. 2013;87(3):1631-48.

35. Aguiar ER, Olmo RP, Paro S, Ferreira FV, de Faria IJ, Todjro YM, et al. Sequence-independent characterization of viruses based on the pattern of viral small RNAs produced by the host. Nucleic Acids Res. 2015;43(13):6191-206.

36. Lewis SH, Salmela H, Obbard DJ. Duplication and Diversification of Dipteran Argonaute Genes, and the Evolutionary Divergence of Piwi and Aubergine. Genome Biol Evol. 2016;8(3):507-18.

37. Campbell CL, Black WCt, Hess AM, Foy BD. Comparative genomics of small RNA regulatory pathway components in vector mosquitoes. BMC Genomics. 2008;9:425.

38. Giraldo-Calderon GI, Emrich SJ, MacCallum RM, Maslen G, Dialynas E, Topalis P, et al. VectorBase: an updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases. Nucleic Acids Res. 2015;43(Database issue):D707-13.

39. Akbari OS, Antoshechkin I, Amrhein H, Williams B, Diloreto R, Sandler J, et al. The developmental transcriptome of the mosquito Aedes aegypti, an invasive species and major arbovirus vector. G3 (Bethesda). 2013;3(9):1493-509.

40. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, et al. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell. 2007;128(6):1089-103.

41. Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, et al. A slicer-mediated mechanism for repeat-associated siRNA 5′ end formation in Drosophila. Science. 2007;315(5818):1587-90.

42. Li C, Vagin VV, Lee S, Xu J, Ma S, Xi H, et al. Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. Cell. 2009;137(3):509-21.

43. Cox DN, Chao A, Baker J, Chang L, Qiao D, Lin H. A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. Genes Dev. 1998;12(23):3715-27.

44. Nene V, Wortman JR, Lawson D, Haas B, Kodira C, Tu ZJ, et al. Genome sequence of Aedes aegypti, a major arbovirus vector. Science. 2007;316(5832):1718-23.

45. Arensburger P, Hice RH, Wright JA, Craig NL, Atkinson PW. The mosquito Aedes aegypti has a large genome size and high transposable element load but contains a low proportion of transposon-specific piRNAs. BMC Genomics. 2011;12:606.

46. Czech B, Hannon GJ. One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing. Trends Biochem Sci. 2016;41(4):324-37.

47. Pane A, Wehr K, Schupbach T. zucchini and squash encode two putative nucleases required for rasiRNA production in the Drosophila germline. Dev Cell. 2007;12(6):851-62.

48. Han BW, Wang W, Li C, Weng Z, Zamore PD. Noncoding RNA. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. Science. 2015;348(6236):817-21.

49. Mohn F, Handler D, Brennecke J. Noncoding RNA. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. Science. 2015;348(6236):812-7.

50. Kawaoka S, Izumi N, Katsuma S, Tomari Y. 3′ end formation of PIWI-interacting RNAs in vitro. Mol Cell. 2011;43(6):1015-22.

51. Feltzin VL, Khaladkar M, Abe M, Parisi M, Hendriks GJ, Kim J, et al. The exonuclease Nibbler regulates age-associated traits and modulates piRNA length in Drosophila. Aging Cell. 2015;14(3):443-52.

52. Horwich MD, Li C, Matranga C, Vagin V, Farley G, Wang P, et al. The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. Curr Biol. 2007;17(14):1265-72.

53. Saito K, Sakaguchi Y, Suzuki T, Suzuki T, Siomi H, Siomi MC. Pimet, the Drosophila homolog of HEN1, mediates 2′-O-methylation of Piwi- interacting RNAs at their 3′ ends. Genes Dev. 2007;21(13):1603-8.

54. Donertas D, Sienski G, Brennecke J. Drosophila Gtsf1 is an essential component of the Piwi-mediated transcriptional silencing complex. Genes Dev. 2013;27(15):1693-705.

55. Ohtani H, Iwasaki YW, Shibuya A, Siomi H, Siomi MC, Saito K. DmGTSF1 is necessary for Piwi-piRISC-mediated transcriptional transposon silencing in the Drosophila ovary. Genes Dev. 2013;27(15):1656-61.

56. Yu Y, Gu J, Jin Y, Luo Y, Preall JB, Ma J, et al. Panoramix enforces piRNA-dependent cotranscriptional silencing. Science. 2015;350(6258):339-42.
57. Sienski G, Batki J, Senti KA, Donertas D, Tirian L, Meixner K, et al. Silencio/CG9754 connects the Piwi-piRNA complex to the cellular heterochromatin machinery. Genes Dev. 2015;29(21):2258-71.
58. Le Thomas A, Rogers AK, Webster A, Marinov GK, Liao SE, Perkins EM, et al. Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. Genes Dev. 2013;27(4):390-9.
59. Sienski G, Donertas D, Brennecke J. Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. Cell. 2012;151(5):964-80.
60. Lim AK, Kai T. Unique germ-line organelle, nuage, functions to repress selfish genetic elements in Drosophila melanogaster. Proc Natl Acad Sci U S A. 2007;104(16):6714-9.
61. Cora E, Pandey RR, Xiol J, Taylor J, Sachidanandam R, McCarthy AA, et al. The MID-PIWI module of Piwi proteins specifies nucleotide- and strand-biases of piRNAs. RNA. 2014;20(6):773-81.
62. Matsumoto N, Nishimasu H, Sakakibara K, Nishida KM, Hirano T, Ishitani R, et al. Crystal Structure of Silkworm PIWI-Clade Argonaute Siwi Bound to piRNA. Cell. 2016;167(2):484-97 e9.
63. Wang W, Yoshikawa M, Han BW, Izumi N, Tomari Y, Weng Z, et al. The initial uridine of primary piRNAs does not create the tenth adenine that Is the hallmark of secondary piRNAs. Mol Cell. 2014;56(5):708-16.
64. Wang W, Han BW, Tipping C, Ge DT, Zhang Z, Weng Z, et al. Slicing and Binding by Ago3 or Aub Trigger Piwi-Bound piRNA Production by Distinct Mechanisms. Mol Cell. 2015;59(5):819-30.
65. Wu Q, Luo Y, Lu R, Lau N, Lai EC, Li WX, et al. Virus discovery by deep sequencing and assembly of virus-derived small silencing RNAs. Proc Natl Acad Sci U S A. 2010;107(4):1606-11.
66. Petit M, Mongelli V, Frangeul L, Blanc H, Jiggins F, Saleh MC. piRNA pathway is not required for antiviral defense in Drosophila melanogaster. Proc Natl Acad Sci U S A. 2016;113(29):E4218-27.
67. Weber F, Wagner V, Rasmussen SB, Hartmann R, Paludan SR. Double-stranded RNA is produced by positive-strand RNA viruses and DNA viruses but not in detectable amounts by negative-strand RNA viruses. J Virol. 2006;80(10):5059-64.
68. Decroly E, Ferron F, Lescar J, Canard B. Conventional and unconventional mechanisms for capping viral mRNA. Nat Rev Microbiol. 2011;10(1):51-65.
69. Ramanathan A, Robb GB, Chan SH. mRNA capping: biological functions and applications. Nucleic Acids Res. 2016;44(16):7511-26.
70. Bidet K, Garcia-Blanco MA. Flaviviral RNAs: weapons and targets in the war between virus and host. Biochem J. 2014;462(2):215-30.
71. Raju R, Huang HV. Analysis of Sindbis virus promoter recognition in vivo, using novel vectors with two subgenomic mRNA promoters. J Virol. 1991;65(5):2501-10.
72. Adelman ZN, Blair CD, Carlson JO, Beaty BJ, Olson KE. Sindbis virus-induced silencing of dengue viruses in mosquitoes. Insect Mol Biol. 2001;10(3):265-73.
73. Homolka D, Pandey RR, Goriaux C, Brasset E, Vaury C, Sachidanandam R, et al. PIWI Slicing and RNA Elements in Precursors Instruct Directional Primary piRNA Biogenesis. Cell Rep. 2015;12(3):418-28.
74. Ishizu H, Iwasaki YW, Hirakata S, Ozaki H, Iwasaki W, Siomi H, et al. Somatic Primary piRNA Biogenesis Driven by cis-Acting RNA Elements and trans-Acting Yb. Cell Rep. 2015;12(3):429-40.
75. Goic B, Vodovar N, Mondotte JA, Monot C, Frangeul L, Blanc H, et al. RNA-mediated interference and reverse transcription control the persistence of RNA viruses in the insect model Drosophila. Nat Immunol. 2013;14(4):396-403.
76. van Mierlo JT, van Cleef KW, van Rij RP. Defense and counterdefense in the RNAi-based antiviral immune system in insects. Methods Mol Biol. 2011;721:3-22.
77. Khurana JS, Wang J, Xu J, Koppetsch BS, Thomson TC, Nowosielska A, et al. Adaptation to P Element Transposon Invasion in Drosophila melanogaster. Cell. 2011;147(7):1551-63.

78. Feschotte C, Gilbert C. Endogenous viruses: insights into viral evolution and impact on host biology. Nat Rev Genet. 2012;13(4):283-96.

79. Katzourakis A, Gifford RJ. Endogenous viral elements in animal genomes. PLoS Genet. 2010;6(11):e1001191.

80. Horie M, Kobayashi Y, Suzuki Y, Tomonaga K. Comprehensive analysis of endogenous bornavirus-like elements in eukaryote genomes. Philos Trans R Soc Lond B Biol Sci. 2013;368(1626):20120499.

81. Fujino K, Horie M, Honda T, Merriman DK, Tomonaga K. Inhibition of Borna disease virus replication by an endogenous bornavirus-like element in the ground squirrel genome. Proc Natl Acad Sci U S A. 2014;111(36):13175-80.

82. Parrish NF, Fujino K, Shiromoto Y, Iwasaki YW, Ha H, Xing J, et al. piRNAs derived from ancient viral processed pseudogenes as transgenerational sequence-specific immune memory in mammals. RNA. 2015;21(10):1691-703.

83. Crochu S, Cook S, Attoui H, Charrel RN, De Chesse R, Belhouchet M, et al. Sequences of flavivirus-related RNA viruses persist in DNA form integrated in the genome of Aedes spp. mosquitoes. J Gen Virol. 2004;85(Pt 7):1971-80.

84. Roiz D, Vazquez A, Seco MP, Tenorio A, Rizzoli A. Detection of novel insect flavivirus sequences integrated in Aedes albopictus (Diptera: Culicidae) in Northern Italy. Virol J. 2009;6:93.

85. Chen XG, Jiang X, Gu J, Xu M, Wu Y, Deng Y, et al. Genome sequence of the Asian Tiger mosquito, Aedes albopictus, reveals insights into its biology, genetics, and evolution. Proc Natl Acad Sci U S A. 2015;112(44):E5907-15.

86. Cook S, Bennett SN, Holmes EC, De Chesse R, Moureau G, de Lamballerie X. Isolation of a new strain of the flavivirus cell fusing agent virus in a natural mosquito population from Puerto Rico. J Gen Virol. 2006;87(Pt 4):735-48.

87. Blair CD, Adelman ZN, Olson KE. Molecular strategies for interrupting arthropod-borne virus transmission by mosquitoes. Clin Microbiol Rev. 2000;13(4):651-61.

88. Powers AM, Kamrud KI, Olson KE, Higgs S, Carlson JO, Beaty BJ. Molecularly engineered resistance to California serogroup virus replication in mosquito cells and mosquitoes. Proc Natl Acad Sci U S A. 1996;93(9):4187-91.

89. Gaines PJ, Olson KE, Higgs S, Powers AM, Beaty BJ, Blair CD. Pathogen-derived resistance to dengue type 2 virus in mosquito cells by expression of the premembrane coding region of the viral genome. J Virol. 1996;70(4):2132-7.

90. Adelman ZN, Sanchez-Vargas I, Travanty EA, Carlson JO, Beaty BJ, Blair CD, et al. RNA silencing of dengue virus type 2 replication in transformed C6/36 mosquito cells transcribing an inverted-repeat RNA derived from the virus genome. J Virol. 2002;76(24):12925-33.

91. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, et al. CRISPR provides acquired resistance against viruses in prokaryotes. Science. 2007;315(5819):1709-12.

92. Kobayashi H, Tomari Y. RISC assembly: Coordination between small RNAs and Argonaute proteins. Biochim Biophys Acta. 2016;1859(1):71-81.

93. Carissimo G, Pondeville E, McFarlane M, Dietrich I, Mitri C, Bischoff E, et al. Antiviral immunity of Anopheles gambiae is highly compartmentalized, with distinct roles for RNA interference and gut microbiota. Proc Natl Acad Sci U S A. 2015;112(2):E176-85.

94. Brackney DE, Beane JE, Ebel GD. RNAi targeting of West Nile virus in mosquito midguts promotes virus diversification. PLoS Pathog. 2009;5(7):e1000502.

95. Fros JJ, Miesen P, Vogels CB, Gaibani P, Sambri V, Martina BE, et al. Comparative Usutu and West Nile virus transmission potential by local Culex pipiens mosquitoes in north-western Europe. One Health. 2015;1:31-6.

96. Lu HL, Tanguy S, Rispe C, Gauthier JP, Walsh T, Gordon K, et al. Expansion of genes encoding piRNA-associated argonaute proteins in the pea aphid: diversification of expression profiles in different plastic morphs. PLoS One. 2011;6(12):e28051.

97. Lim RS, Kai T. A piece of the pi(e): The diverse roles of animal piRNAs and their PIWI partners. Semin Cell Dev Biol. 2015;47-48:17-31.
98. Ross RJ, Weiner MM, Lin H. PIWI proteins and PIWI-interacting RNAs in the soma. Nature. 2014;505(7483):353-9.
99. Robine N, Lau NC, Balla S, Jin Z, Okamura K, Kuramochi-Miyagawa S, et al. A broadly conserved pathway generates 3'UTR-directed primary piRNAs. Curr Biol. 2009;19(24):2066-76.
100. Saito K, Inagaki S, Mituyama T, Kawamura Y, Ono Y, Sakota E, et al. A regulatory circuit for piwi by the large Maf gene traffic jam in Drosophila. Nature. 2009;461(7268):1296-9.
101. Mohn F, Sienski G, Handler D, Brennecke J. The rhino-deadlock-cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in Drosophila. Cell. 2014;157(6):1364-79.
102. Zhang Z, Wang J, Schultz N, Zhang F, Parhad SS, Tu S, et al. The HP1 homolog rhino anchors a nuclear complex that suppresses piRNA precursor splicing. Cell. 2014;157(6):1353-63.
103. Klattenhoff C, Xi H, Li C, Lee S, Xu J, Khurana JS, et al. The Drosophila HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. Cell. 2009;138(6):1137-49.
104. Rangan P, Malone CD, Navarro C, Newbold SP, Hayes PS, Sachidanandam R, et al. piRNA production requires heterochromatin formation in Drosophila. Curr Biol. 2011;21(16):1373-9.
105. Lippman Z, Martienssen R. The role of RNA interference in heterochromatic silencing. Nature. 2004;431(7006):364-70.
106. Reidenbach KR, Cook S, Bertone MA, Harbach RE, Wiegmann BM, Besansky NJ. Phylogenetic analysis and temporal diversification of mosquitoes (Diptera: Culicidae) based on nuclear genes and morphology. BMC Evol Biol. 2009;9:298.
107. Chirn GW, Rahman R, Sytnikova YA, Matts JA, Zeng M, Gerlach D, et al. Conserved piRNA Expression from a Distinct Set of piRNA Cluster Loci in Eutherian Mammals. PLoS Genet. 2015;11(11):e1005652.
108. Kiuchi T, Koga H, Kawamoto M, Shoji K, Sakai H, Arai Y, et al. A single female-specific piRNA is the primary determiner of sex in the silkworm. Nature. 2014;509(7502):633-6.
109. Iwasaki YW, Siomi MC, Siomi H. PIWI-Interacting RNA: Its Biogenesis and Functions. Annu Rev Biochem. 2015;84:405-33.
110. Nishida KM, Iwasaki YW, Murota Y, Nagao A, Mannen T, Kato Y, et al. Respective functions of two distinct Siwi complexes assembled during PIWI-interacting RNA biogenesis in Bombyx germ cells. Cell Rep. 2015;10(2):193-203.
111. Walker T, Jeffries CL, Mansfield KL, Johnson N. Mosquito cell lines: history, isolation, availability and application to assess the threat of arboviral transmission in the United Kingdom. Parasit Vectors. 2014;7:382.
112. Sumiyoshi T, Sato K, Yamamoto H, Iwasaki YW, Siomi H, Siomi MC. Loss of l(3)mbt leads to acquisition of the ping-pong cycle in Drosophila ovarian somatic cells. Genes Dev. 2016;30(14):1617-22.
113. Kistler KE, Vosshall LB, Matthews BJ. Genome engineering with CRISPR-Cas9 in the mosquito Aedes aegypti. Cell Rep. 2015;11(1):51-60.
114. Basu S, Aryan A, Overcash JM, Samuel GH, Anderson MA, Dahlem TJ, et al. Silencing of end-joining repair for efficient site-specific gene insertion after TALEN/CRISPR mutagenesis in Aedes aegypti. Proc Natl Acad Sci U S A. 2015;112(13):4038-43.
115. Franz AW, Kantor AM, Passarelli AL, Clem RJ. Tissue Barriers to Arbovirus Infection in Mosquitoes. Viruses. 2015;7(7):3741-67.
116. Martina BE, Barzon L, Pijlman GP, de la Fuente J, Rizzoli A, Wammes LJ, et al. Human to human transmission of arthropod-borne pathogens. Curr Opin Virol. 2017;22:13-21.
117. Kean J, Rainey SM, McFarlane M, Donald CL, Schnettler E, Kohl A, et al. Fighting Arbovirus Transmission: Natural and Engineered Control of Vector Competence in Aedes Mosquitoes. Insects. 2015;6(1):236-78.

118. Alphey L. Genetic control of mosquitoes. Annu Rev Entomol. 2014;59:205-24.
119. Alphey L, Benedict M, Bellini R, Clark GG, Dame DA, Service MW, et al. Sterile-insect methods for control of mosquito-borne diseases: an analysis. Vector Borne Zoonotic Dis. 2010;10(3):295-311.
120. Lambrechts L, Koella JC, Boete C. Can transgenic mosquitoes afford the fitness cost? Trends Parasitol. 2008;24(1):4-7.
121. Marrelli MT, Moreira CK, Kelly D, Alphey L, Jacobs-Lorena M. Mosquito transgenesis: what is the fitness cost? Trends Parasitol. 2006;22(5):197-202.
122. Hammond A, Galizi R, Kyrou K, Simoni A, Siniscalchi C, Katsanos D, et al. A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector Anopheles gambiae. Nat Biotechnol. 2016;34(1):78-83.
123. Gantz VM, Jasinskiene N, Tatarenkova O, Fazekas A, Macias VM, Bier E, et al. Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi. Proc Natl Acad Sci U S A. 2015;112(49):E6736-43.
124. Oye KA, Esvelt K, Appleton E, Catteruccia F, Church G, Kuiken T, et al. Biotechnology. Regulating gene drives. Science. 2014;345(6197):626-8.
125. Johnson BK, Gichogo A, Gitau G, Patel N, Ademba G, Kirui R, et al. Recovery of o'nyong-nyong virus from Anopheles funestus in Western Kenya. Trans R Soc Trop Med Hyg. 1981;75(2):239-41.

Chapter 9

General Discussion microRNAs in Virus-Host Interactions

Valentina Libri*, Pascal Miesen*, Ronald P. van Rij, Amy H. Buck
Regulation of microRNA biogenesis and turnover by animals and their viruses.

Cell Mol Life Sci. (2013) 70:3525-44

*equal contribution

ABSTRACT

MicroRNAs (miRNAs) are a ubiquitous component of gene regulatory networks that modulate the precise amounts of proteins expressed in a cell. Despite their small size, miRNA genes contain various recognition elements that enable specificity in when, where and to what extent they are expressed. The importance of precise control of miRNA expression is underscored by functional studies in model organisms and the association between miRNA mis-expression and disease. In the last decade, identification of the pathways by which miRNAs are produced, matured and turned-over has revealed many aspects of their biogenesis that are subject to regulation. Studies in viral systems have revealed a range of mechanisms by which viruses target these pathways through viral proteins or non-coding RNAs in order to regulate cellular gene expression. In parallel, a field of study has evolved around the activation and suppression of antiviral RNA interference (RNAi) by viruses. Virus encoded suppressors of RNAi can impact miRNA biogenesis in cases where miRNA and small interfering RNA (siRNA) pathways converge. Here we review the literature on the mechanisms by which miRNA biogenesis and turnover are regulated in animals and the diverse strategies that viruses use to subvert or inhibit these processes.

INTRODUCTION

Small RNA classification

The specific recognition of nucleic acid sequences by ribonucleic-protein complexes (RNPs) is central to transcriptional and post-transcriptional gene regulation. Small RNAs are incorporated into many RNPs in order to mediate the specific recognition of target nucleic acids through Watson-Crick base-pairing. Different classes of small RNA continue to be discovered, including some that are specific to plants or animal lineages, reviewed in (1, 2). There are three major classes in animals: microRNAs (miRNAs), short interfering RNAs (siRNAs), and piwi-interacting RNAs (piRNAs). These classes differ in their origin and biogenesis, the proteins with which they interact, the mechanism of action of the RNP in which they are contained, and the nature of their targets. MiRNAs are derived from single-stranded (ss) RNAs that fold back on themselves into stem-loop structures. Endogenous siRNAs originate from double-stranded (ds) RNA precursors that result from convergent bi-directional transcription, inverted repeat regions in structured RNA, or base-pairing between protein-coding genes and pseudogene-derived antisense transcripts. The detailed mechanism(s) of piRNA biogenesis remains somewhat elusive, but the primary piRNAs originate from single-stranded precursor RNAs and are only found in animals, and specifically in the germline (3). Each class of small RNA binds to a member of the Argonaute (Ago) family of proteins: siRNAs and miRNAs associate with the Ago clade, whereas piRNAs associate with the Piwi clade, reviewed in (4). The Ago protein bound to the small RNA comprises the RNA-induced silencing

complex (RISC). There is increasing diversity in the mechanisms by which RISCs function and in the genes they target (5). RISCs containing miRNAs are found throughout the eukaryal domain and primarily target messenger RNAs (mRNAs), causing the inhibition of translation and/or de-adenylation and degradation of the mRNAs, reviewed in (6). Recognition of the mRNA target does not require perfect complementarity with the miRNA and is generally dictated by the "seed region" within the 5' terminal region of the miRNA (nucleotides 2-8), reviewed in (7). Based on this low sequence requirement for recognition, each miRNA is predicted to target several hundred genes. The majority of human protein-coding genes have miRNA binding sites that are maintained under selective pressure (8).

miRNAs in hosts and viruses

Based on the large number of genes targeted by miRNAs, together with the ability of miRNAs to operate synergistically with one another, these small RNAs are involved in regulating numerous aspects of cellular biology including proliferation, tumorigenesis, metabolism, differentiation, development, apoptosis, and innate and adaptive immune responses, reviewed in (9-14). Viruses have evolved to exploit and manipulate these same cellular pathways. Therefore, it is not surprising that they use the miRNA pathway to do this, either by encoding their own miRNAs, or encoding molecules that activate or inhibit cellular miRNA expression. Seven different virus families have been reported to encode miRNAs or miRNA-like molecules: herpesviruses, polyomaviruses, adenoviruses, baculoviruses, an ascovirus, a retrovirus, reviewed in (15), and recently a falvivirus (16). Analysis of a wide range of RNA viruses failed to identify viral miRNAs (17), apart from the identification of miRNAs in bovine leukemia virus (BLV), a retrovirus that replicates in the nucleus (18) and the identification of a miRNA-like species in West Nile virus, a cytoplasmic RNA virus that encodes a stem-loop structure in its 3'UTR (16). In the latter study the small RNA was detected in infected mosquito cells but not infected mammalian cells, raising the question of how biogenesis factors differ in the two animals. There have been several reports, some controversial, suggesting that additional retroviruses may encode miRNAs (19-21) but it remains unclear if this strategy would be advantageous to cytoplasmic RNA viruses (17). However, both DNA and RNA viruses can modulate the expression of host miRNAs to enhance replication or facilitate the progression through their life cycles, reviewed in (22).

Given the intricate role of miRNAs in regulating cell biology, it is not surprising that miRNA expression is subject to various levels of regulation, which viruses can also exploit. miRNA biogenesis encompasses a series of sequential processing steps to convert the primary miRNA (pri-miRNA) transcript into the biologically active, mature miRNA (Figure 1), reviewed in (1, 5). Following transcription, the pri-miRNA is cleaved by the RNase III-like enzyme Drosha in the nucleus (23) to generate a ~60-70 nt precursor

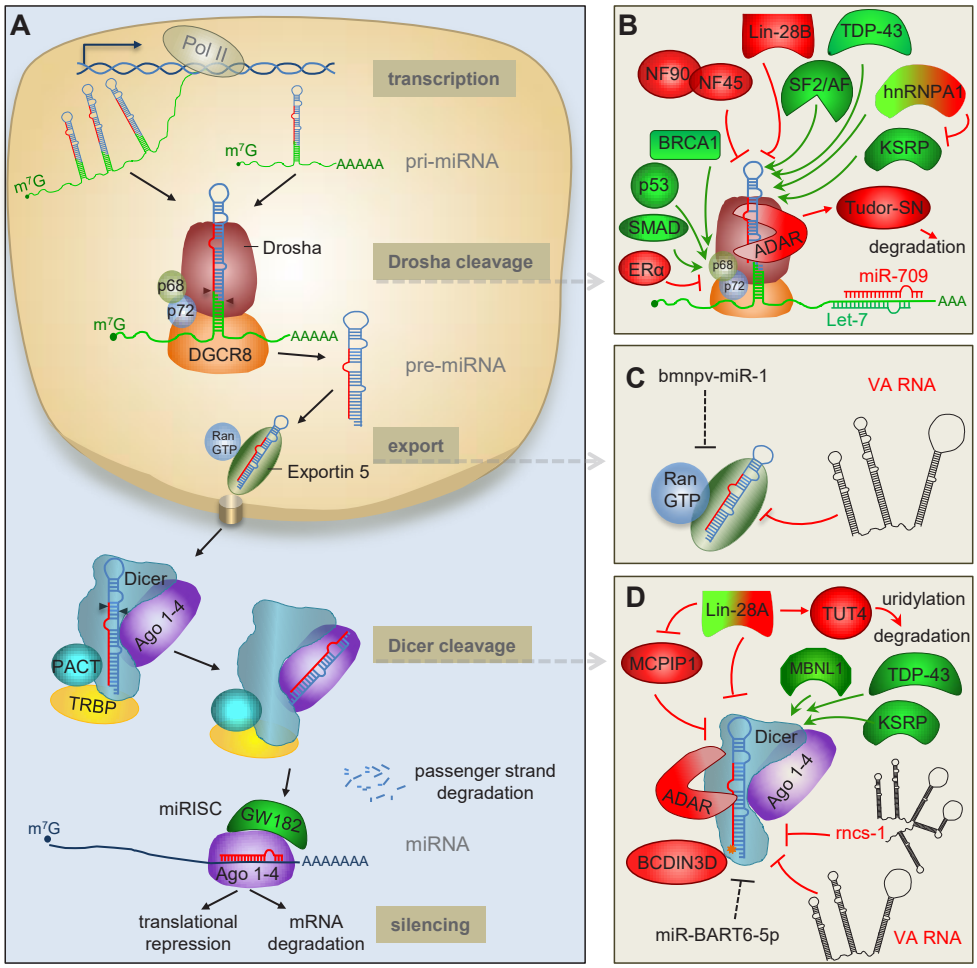


Figure 1. Schematic overview of microRNA biogenesis and regulation in animals. (A) The canonical biogenesis pathway. Pri-miRNAs are transcribed in the nucleus by Polymerase II with a cap (m7G, 7-methylguanosine-cap) and poly A tail. The pri-miRNA can harbour a single pre-miRNA or a cluster of pre-miRNAs; the mature miRNA sequence is depicted in red. Cleavage of the pri-miRNA occurs in the nucleus by the Microprocessor complex, composed minimally of Drosha and DGCR8, which interact with helicases p68 and p72. The pre-miRNA is then exported through the nuclear pore complex into the cytoplasm where the stem is cleaved by Dicer, supported by TRBP or PACT. The miRNA/miRNA* duplexes are loaded into the Ago protein within RISC, where one of the strand is preferentially retained; this complex contains an Ago protein and GW182, which is required for gene silencing. **(B)** Regulation of pri-miRNA cleavage. Proteins can either positively (green) or negatively (red) influence cleavage of pri-miRNAs by Drosha, based on direct interactions with the pri-miRNA or interactions with auxiliary proteins p68/p72 (indicated by arrows). Factors depicted in both green and red can behave as positive or negative regulators depending on the identity of the miRNA and the presence of other factors. Mature miRNAs can also regulate pri-miRNA processing through interactions downstream of the stem-loop: let-7 promotes processing of pri-let-7 whereas miR-709 inhibits processing of pri-miR 15/16. **(C)** Regulation of pre-miRNA export. Two viral non-coding RNAs inhibit miRNA translocation to the cytoplasm: VA1 competes with endogenous pre-miRNAs for binding to Exportin-5 whereas the viral miRNA, Bmnp-miR-1, regulates export indirectly (dotted line) by targeting RanGTP. **(D)** Regulation of pre-miRNA

miRNA (pre-miRNA). The pre-miRNA is then exported into the cytoplasm (24) and processed into a ~22 nt duplex by the RNase III-like enzyme Dicer (25-29). One strand of this duplex is then loaded into RISC which is comprised of at least one Ago protein (30, 31) and GW182, a glycine-tryptophan repeat containing protein required for gene silencing (32). Each stage in the miRNA biogenesis pathway is subject to regulation. Here we summarise the current literature on the regulation of miRNA biogenesis and turnover and detail the mechanisms by which viruses exploit or manipulate these processes. We focus primarily on animal miRNAs, but highlight some common and distinct properties of plant miRNAs, which evolved separately (33).

Regulation of miRNA transcription

The first regulatory layer governing miRNA abundance occurs at the stage of transcription of the pri-miRNA. The stem-loop structures from which miRNAs are derived are disseminated throughout the genome, either within intronic sequences of protein-coding genes, within intronic or exonic regions of noncoding RNAs, or set between independent transcription units (intergenic). The majority of intronic miRNAs are transcribed from the same promoter as the host gene leading to a strong correlation of mRNA and miRNA expression. However, approximately one third of intronic miRNAs are transcribed from independent promoters, enabling separate control of their transcription (34-36). Most pri-miRNAs are transcribed by RNA polymerase II (Pol II) (37), however, a subset of miRNAs, including viral miRNAs, are transcribed by Pol III (35, 38-40). Like mRNAs, Pol II-derived pri-miRNAs are poly-adenylated at their 3′ end and bear 7-methyl-guanosine caps at their 5′ end (37). The promoters of pri-miRNAs also contain CpG islands, TATA box sequences, initiation elements and certain histone modifications, indicating potential for regulation by transcription factors (TFs), enhancers, silencing elements and chromatin modifications (9, 35). Therefore, many of the properties dictating the transcriptional regulation of miRNAs are the same as those regulating protein-coding genes. Following transcription, the stem-loop sequence of the pri-miRNA is recognized by a series of enzymes that orchestrate a tightly controlled maturation process.

Pri-miRNA cleavage by the Microprocessor

In the canonical pathway, the pri-miRNA is cleaved in the nucleus by the RNase III enzyme Drosha into a ~ 60-70 nt pre-miRNA. Cleavage by Drosha requires the co-factor DGCR8 (DiGeorge critical region 8), also known as Pasha (41). Together these two proteins

◀ cleavage by Dicer. Proteins that regulate Dicer processing include: 1) Lin28 (Lin28A), which recruits TUT4 that oligo-uridylates pre-miRNAs leading to degradation, 2) MCPIP1 which cleaves the loop, 3) TDP-43 and KSRP, which bind to the loops of both pri-miRNAs and pre-miRNAs and 4) BCDIN3D, which can add methyl groups to the 5′end of pre-miRNA and inhibit recognition by Dicer. RNA factors that are known to inhibit Dicer processing include an ~ 800 non-coding RNA termed rnc-1, VA RNAs from Vaccinia virus (black) and a viral miRNA regulates Dicer indirectly (dotted line).

comprise the minimum components of the Microprocessor complex (Figure 1B). DGCR8 functions at least in part by binding to the junction between single-stranded and double-stranded regions of the pri-miRNA and directing Drosha to cleave approximately 11 bp downstream of this junction (42), generating products with 2 nt 3′ overhangs. It is thought that cleavage of the pri-miRNA by Drosha occurs co-transcriptionally along with splicing (43, 44), supported by the fact that Drosha co-localizes to sites of active transcription (45). Processing of a pri-miRNA into a pre-miRNA can be regulated by a variety of protein co-factors that are either recruited to the Microprocessor through protein-protein interactions or through direct interactions with the pri-miRNAs.

Regulation of pri-miRNA processing by proteins that interact with the Microprocessor

Many proteins have been identified that interact with Drosha, including the DEAD-box helicase proteins p68 (also known as DDX5) and p72 (DDX17) (41). These helicases facilitate processing of nearly one-third of pri-miRNAs, according to studies with p68/p72 knock-out mice (46). In some cases they do this by mediating interactions of TFs with the Microprocessor. A well-characterized example is the stimulation of maturation of specific pri-miRNAs by SMAD proteins, which are TFs induced upon stimulation with tumour growth factor β (TGF-β). SMAD proteins associate with p68 to enhance processing through binding a consensus sequence in pri-miRNAs that strongly resembles the DNA SMAD-binding element (Figure 2) (47-49). Other TFs that regulate processing include the tumour supressor p53, which promotes pri-miRNA processing via interaction with p68 (50) and ERα (estrogen receptor α), which the processing of specific pri-miRNAs via interactions with p68/p72 (51). Another tumour suppressor, BRCA1 (breast cancer susceptibility gene 1), also associates with Drosha, p68, SMAD3 and p53 to accelerate processing of specific pri-miRNAs associated with cancer (52). In contrast to the SMAD-regulated miRNAs, no consensus sequence has been identified within the miRNAs regulated by these TFs and the mechanisms underlying specificity in their regulatory functions are unknown. In addition to p68/p72, NF90 and NF45 (nuclear factor 90 and 45) also associate with the Microprocessor (41) and can inhibit processing of several miRNAs, including let-7 family members (53). Other proteins that associate with Drosha and positively regulate processing include the multifunctional protein SNIP1 (SMAD-interacting protein) (54) and ARS2 (Arsenite- resistance protein 2) (55, 56). However the precise mechanisms by which these multi-functional proteins influence biogenesis are unclear.

Regulation of pri-miRNA processing by recognition of the stem-loop sequence or structure

Comparative analysis of pri-miRNA sequences suggests that 14% of human pri-miRNAs have conserved nucleotides in their terminal loops, which may relate to interactions with regulatory proteins (57). One of the first proteins identified to operate in this way was hnRNP-A1 (heterogeneous nuclear ribonucleoprotein A1), which binds to the terminal loop and stem of pri-miR-18a and facilitates processing by alteration of the stem structure

(57, 58) (Figure 1 and Figure 2). Interestingly, this protein can also interact with pri-let-7a, but in this case it negatively regulates processing (59). The inhibitory effect appears to result from competition between hnRNP-A1 and KSRP (KH-type splicing regulatory protein), which both bind to the loop of pri-let-7a. KSRP regulates only a subset of miRNAs and recognition has been proposed to derive from 2 or 3 sequential guanidines in the loop sequences (60) (Figure 1B and Figure 2). Interestingly, KSRP activity is modulated through its phosphorylation state in response to different stimuli and provides a link between PI3K/AKT signalling and miRNA processing (61, 62) (Figure 1B and Figure 2). Other RNA-binding proteins that interact with pri-miRNAs and promote their biogenesis include TDP-43 (TAR DNA-binding protein-43) (63) and the serine/arginine-rich SR protein SF2/ASF. The SF2/ASF protein binds to a motif in the stem of pri-miR-7 and has been proposed to alter the structure as observed for HnRNP-A1 (64). Interestingly, miR-7 targets the 3'UTR of SF2/ASF, providing a negative feedback loop that may be important for controlling the steady-state expression level of this miRNA (64).

A key protein involved in regulating multiple aspects of miRNA biogenesis is Lin28 (abnormal cell lineage factor 28), which was originally discovered as a heterochronic gene regulating developmental timing in worms (65). Lin28 can inhibit both pri-let-7 processing (66-68) and pre-let-7 processing (69-74) and recognition is mediated by the primary sequence and structure of the terminal loop (Figure 2) (75). Two Lin28 paralogs are present in mammals, Lin28A and Lin28B. Lin28A is predominantly cytoplasmic whereas Lin28B contains nuclear localisation signals and accumulates in the nucleolus. It has been proposed that Lin28B blocks let-7 processing by sequestering pri-let-7 miRNAs in the nucleoli away from the Microprocessor (68), suggesting a new mechanism by which other RNA-binding proteins might inhibit pri-miRNA biogenesis.

Regulation of pri-miRNAs by other miRNAs

A recent study by Zisoulis and colleagues demonstrates the pri-let-7 processing is also regulated by mature miR-let-7 (76), providing the first example of a direct auto-regulatory loop for let-7 biogenesis. In *C. elegans*, the ALG-1 (Argonaute-like protein-1) binds to a specific site at the 3' end of the pri-let-7 and thereby promotes processing of the pri-miRNA. The interaction between ALG-1 and pri-let-7 is mediated by mature let-7 through a conserved site in the pri-miRNA transcript (Fig. 1B, Fig. 2). Immunoprecipitation of Ago proteins in human cells also suggests an interaction with pri-let-7, though it is not clear if this is mediated by a miRNA (76). Interaction between a mature miRNAs and a pri-miRNA can also have inhibitory effects on processing (Fig. 1B, Fig. 2). For example, miR-709 binds to a stretch of 19 nt in the sequence of pri-miR-15a/16-1, preventing pri-miRNA processing, leading to reduced levels of mature miR-15a/16-1 (77). The factors underlying nuclear localisation of miR-709 remain unknown but this appears to be associated with apoptotic stimuli, and may be a dynamic mechanism for altering

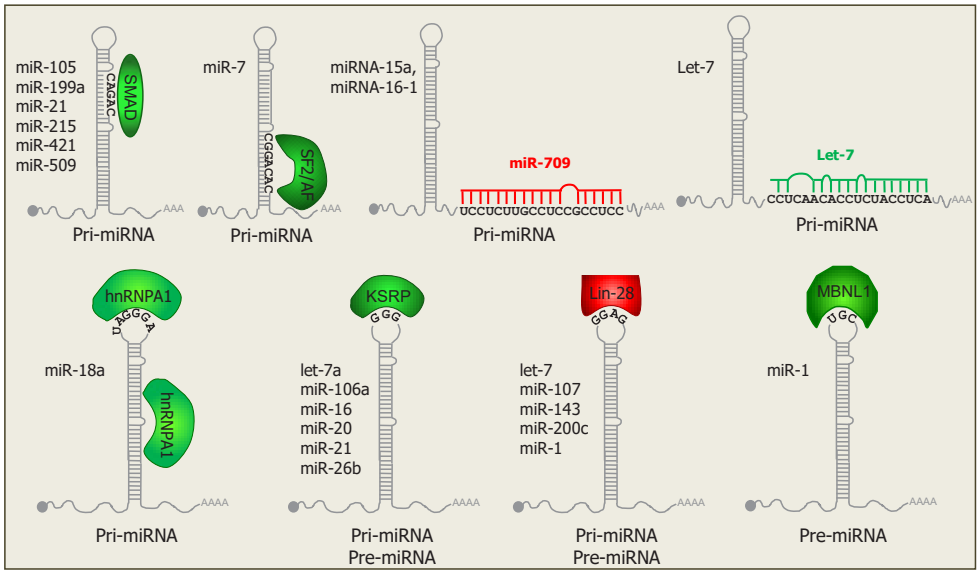


Figure 2. RNA motifs that mediate regulation of pri-miRNA or pre-miRNA processing. Proteins that positively (green) or negatively (red) regulate biogenesis associate with specific motifs in the stem-loop structures; depending on localization of the proteins, these either regulate the pri-miRNA or the pre-miRNA as listed below the hairpin; Lin28 and KSRP can regulate both forms. The identity of the miRNAs that contain the recognition motifs and have been validated to be regulated by each protein are listed to the left of the hairpin structure. A mature miRNA-binding sites for miR-709 in pri-miR-15/16 inhibits its processing whereas a binding site for let-7 in pri-let-7 stimulates its processing.

miR-15a/16 levels in response to external signals. Transfection of a miR-709 mimic into cells results in nuclear localisation of the synthetic RNA, indicating that the localisation signal is contained within the mature miRNA sequence. Nuclear localisation of miRNAs was first reported in a study showing that a hexanucleotide element within the mature miRNA sequence of miR-29b directs its nuclear transport (78). However, this element is not present in miR-709 and the mechanism of nuclear transport is unknown. It appears that miR-709 and its binding site in pri-miR-15a/16 have co-evolved recently, as they are both only present in mouse (77). Further analyses are required to understand the breadth of regulation of pri-miRNAs by mature miRNAs and whether this relates to the nuclear localisation of Ago proteins that has been reported previously (79).

The Drosha-DGCR8 regulatory loop and additional substrates of the Microprocessor

Regulatory feedback loops are thought to be a key feature of how miRNAs function in biological systems; for example, miRNAs that are induced by Toll-like receptor signalling target genes in this pathway, thereby dampening the inflammatory response (80). The miRNA biogenesis machinery is also subject to regulation by feedback loops, as observed for the Drosha-DGCR8 complex (81-83). DGCR8 stabilizes Drosha protein in the microprocessor complex and the Microprocessor complex in turn cleaves hairpin structures embedded in the 5′ UTR of DGCR8 mRNA, leading to degradation of the

DGCR8 transcript. This auto-regulatory loop is postulated to be critical to maintain the appropriate balance between the levels of the Drosha-DGCR8 complex and its substrates: when the Drosha-DGCR8 complex expression level is too low there is suboptimal miRNA processing; when the Drosha-DGCR8 complex expression level is too high, cleavage of non-miRNA substrates such as mRNAs may occur. Barad *et al*. propose that efficient miRNA processing and minimal off-target cleavage is obtained only for a narrow range of microprocessor concentration values (84). These studies also suggest that, apart from miRNA processing, the microprocessor might play roles in mRNA stability control (83). Consistent with this, HITS-CLIP analysis identified hundreds of mRNAs bound to DGCR8, including DGCR8 mRNA (85). This study further demonstrated that cleavage within exonic cassettes can influence ratios of alternative spliced isoforms, suggesting complex roles of the Microprocessor in various modes of gene regulation. A viral mRNA was also shown to be regulated by Drosha in KSHV: the KapB (Kaposin B) mRNA includes two pre-miRNAs in its 3'UTR and excision of these by Drosha alters the stability of the mRNA, thereby reducing KapB protein expression (86). This mode of regulating viral gene expression during lytic or latent infection could represent an alternative function of viral miRNAs, where their processing serves a purpose, rather than (or in addition to) their activities in gene silencing.

Regulation of pre-miRNA export

Once produced, the pre-miRNA is translocated to the cytoplasm through the nuclear pore complex by Exportin-5, which requires the co-factor RanGTP (Fig. 1) (24, 87, 88). Structural analyses suggest that the length of the double-stranded stem and presence of 3' overhangs are important for Exportin-5 recognition (1, 89). Interestingly, Exportin-5 interacts with the RNA-binding protein NF90, also known as ILF-3 (interleukin enhancer-binding factor 3), (90), which is found in the Microprocessor complex (41). It is possible that there is coordination between pri-miRNA cleavage and export but this has not been examined. Exportin-5 also shuttles tRNAs and other abundant RNAs to the cytoplasm and several studies suggest that export of pre-miRNAs can be regulated by these RNAs through competition. For example, Adenovirus produces a ~160 nt hairpin RNA (VA1 in Figure 1C) that binds to Exportin-5 and inhibits nuclear export of pre-miRNAs (91). Over-expression of short hairpin RNAs (shRNAs) in animals can also be toxic due to saturation of Exportin-5 and subsequent inhibition of pre-miRNA export (92). Interestingly, Exportin-5 was also reported to interact with Dicer mRNA and high levels of pre-miRNAs or other Exportin-5 substrates can lead to accumulation of Dicer mRNA in the nucleus, providing another feedback loop for regulating the miRNA biogenesis factors (93). The insect virus *Bombyx mori* nucleopolyhedrosis virus (BmNPV) also negatively regulates nucleocytoplasmic transport of miRNAs by encoding a viral miRNA that targets RanGTP (94), although the functional relevance of this is not yet known.

Regulation of pre-miRNA processing: proteins and RNA motifs involved

Dicer-mediated processing of pre-miRNAs is subject to regulation by co-factors that interact with Dicer and RNA-binding proteins that recognize RNA elements within the pre-miRNAs. The Dicer protein alone can catalyse the cleavage of pre-miRNA, however, the specificity of cleavage is enhanced by TRBP and PACT (104). Binding of TRBP and PACT also stabilizes Dicer and knockdown of TRBP and PACT reduces mature miRNA levels (99, 101). TRBP also provides a link between MAPK (mitogen-activated protein kinase) signalling and miRNA processing since it is phosphorylated by Erk (extracellular signal regulated protein) (105). The phosphorylated form of TRBP is more stable and leads to increased levels of many growth-promoting miRNAs in HEK293 cells and also causes a decrease in let-7 members. The mechanism for differential effects of phosphorylated TRBP on individual miRNAs is not yet clear (105).

The best-studied regulator of pre-miRNA processing by Dicer is Lin28 (Figure. 1D). Lin28A, the cytoplasmic isoform, binds a tetra-nucleotide sequence motif (GGAG) in the terminal loop of let-7 precursors and recruits TUT4 (terminal uridylyltransferase-4, also known as ZCCHC11), which adds an oligo U-tail to pre-let-7. This U tail blocks Dicer processing and mediates decay of pre-let-7, presumably through recruitment of 3' to 5' exonucleases (73, 106). Lin-28A-dependent uridylation has also been observed for several other pre-miRNAs that contain the GGAG motif in their terminal loops, including miR-107, miR-143 and miR-200c (106, 107). Kim and colleagues have recently shown that TUT4, as well as TUT2 and TUT7, can also add a single uridine to the 3'end of a specific set of pre-miRNAs (termed "group 2" pre-miRNAs), which is independent of Lin28A. Up to 30% of pre-let-7 family members have an untemplated uridine at the 3'end in cells not expressing Lin28A (107, 108). The pre-miRNAs that are modified lack a classical 2nt 3'end overhang, such that monouridylation results in the 2'nt overhang and thereby improves processing by Dicer (108).

Like Lin28, KSRP and TDP-43 are also involved in both pri- and pre-miRNA processing but they serve to promote, rather than inhibit, processing (Figure. 1B, 1D) (60, 63). These findings suggest that the terminal loop is an important platform for both ''activators'' (for example, hnRNP A1, KSRP and TDP-43) and ''repressors''(for example, Lin28) to modulate miRNA levels and thereby gene regulation, reviewed in (109). There also appears to be some interplay between the activators and repressors. For example, the RNA binding protein MBNL1 (muscleblind-like splicing regulatory protein 1) binds to pre-miR-1 through recognition of a UGC motif that overlaps with a binding site for Lin28 (Figure. 2), such that MBNL1 binding blocks Lin28-mediated oligouridylation and subsequent degradation of pre-miRNA-1 (110). Similar competition is seen with the mammalian immune regulator MCPIP1 (monocyte chemoattractant protein induced protein-1) and Lin-28: MCPIP1 is a ribonuclease that inhibits miRNA biogenesis by competing with Dicer for the cleavage of the terminal loop of pre-miRNAs. Addition of

Lin28 abolishes MCPIP1-mediated cleavage *in vitro*, presumably through competition for binding to the terminal loop (111). Other negative regulators of processing might also stabilize pre-miRNAs against degradation, but it is not clear if this is one of their functions *in vivo*. Recently Kouzarides's group showed that Dicer processing can also be regulated by methylation of the 5′ end of the pre-miRNA by the human RNA-methyltransferase, BCDIN3D (112). BCDIN3D adds two methyl groups to the 5′ phosphate of pre-miR-145 *in vitro* and *in vivo*; since Dicer specifically recognizes the 5′ monophosphate (113), this modification inhibits processing (Figure. 1D). A noncoding RNA in *C. elegans* was also shown to inhibit pre-miRNA processing: the ~ 800 nt noncoding RNA, rncs-1 (RNA noncoding, starvation up-regulated), competes with endogenous dsRNAs for binding to Dicer or accessory dsRBD proteins (114) (Figure. 1D). The VA RNAs in Adenovirus have also been shown to operate as competitive inhibitors for Dicer processing of pre-miRNA (91, 115), in addition to their inhibitory effects on Exportin 5.

Other viruses also inhibit this step in miRNA biogenesis. For example, Vaccinia Virus (VACV) infection leads to a drastic reduction in Dicer protein expression and a concomitant defect in pre-miRNA processing. The mechanism by which the virus abrogates Dicer expression remains unclear (116). The human herpesvirus Epstein-Barr virus (EBV) influences Dicer processing through a more subtle mechanism: the viral-encoded miRNA miR-BART6-5p targets human Dicer mRNA (117); it is expected that this could form a feedback loop to regulate the level of viral miRNAs. The host-encoded let-7 also regulates Dicer levels through target sites in the coding sequence, suggesting that feedback loops for controlling miRNA biogenesis may be inherent to miRNA homeostasis (118), which viruses can exploit.

Regulation of miRNA expression by Argonaute proteins

MiRNAs function in partnership with Ago proteins, and a number of studies suggest that expression levels of miRNAs are tied to the expression levels of Agos. For example, ectopically expressed Ago proteins (Ago1–4) enhance expression of miRNAs under conditions where the miRNAs saturate the endogenous machinery (119), and endogenous miRNAs are reduced in mouse embryonic fibroblasts from Ago2-knockout mice (120). Ago proteins are also subject to various levels of transcriptional and post-transcriptional regulation that might therefore influence miRNA expression. For example, the expression level of the Ago2 protein is specifically up-regulated in breast cancer cells lacking ERα which is dependent on EGFR/MAPK signalling pathway and leads enhanced miRNA activity (121). Ago2 can also be phosphorylated within the RNA binding pocket, which inhibits small RNA binding and is expected to thereby influence miRNA stability (122). In addition to its role in miRNA stabilization, Ago2 has also been shown to catalyse an alternative pre-miRNA processing event (120). Cleavage occurs within the 3′ arm of a pre-miRNA such that only the small RNA generated from the 5′ arm can be functional.

The relevance of this alternative processing pathway remains elusive, but it may play a role in passenger strand dissociation for hairpins with a high degree of complementarity, where this might otherwise be inefficient (120).

Non-canonical pathways of biogenesis: breaking the rules

In addition to the canonical biogenesis pathway, some miRNAs are processed by Drosha-independent and Dicer-independent pathways (Figure 3) (123). Studies of viral-encoded miRNAs in particular illuminate a range of non-canonical possibilities. For example, murine γ-herpesvirus 68 (MHV68) expresses its miRNAs in the same Pol III primary transcripts as the viral-encoded tRNAs (39, 40). The pre-miRNAs are generated following cleavage by RNase Z and are subsequently processed by Dicer, thus bypassing the Microprocessor complex (124). The retrovirus BLV also encodes Pol III-dependent pre-miRNA-like species that bypass Drosha cleavage and are subsequently processed by Dicer. Importantly, this mechanism provides a route for viral miRNA biogenesis that does not result in cleavage of the retroviral genomic RNA (18). A miRNA-like species was also recently reported in West Nile virus (a cytoplasmic RNA virus) (16) and several reports have shown that artificial miRNAs engineered into RNA viruses are processed to a detectable level (125-127). However, the mechanism(s) for biogenesis of these viral RNAs are not reported. Another alternative processing pathway has been described for miRNAs encoded by Herpes Virus Saimiri (HVS). These miRNAs are derived from the same Pol II transcripts that encode another class of viral noncoding RNA, HSURs (H. saimiri U-rich RNAs), which resemble small nuclear RNAs (snRNAs). The pre-miRNAs are located directly downstream of the 3′ end processing signals of HSURs and processing of the viral miRNAs does not require the Microprocessor (128). Rather, the 5′ ends of the viral pre-miRNAs are produced by the Integrator, a nuclear complex of 12 proteins that associates with Pol II and is required for HSUR biogenesis. As in the canonical miRNA biogenesis pathway, HVS pre-miRNAs require Exportin-5 for transit to the cytoplasm, where they are processed by Dicer. An Integrator-dependent mechanism has not been reported for biogenesis of endogenous miRNAs. However a range of reports suggest other mechanisms by which RNAs can be processed into miRNA-like species without a requirement for Drosha. For example, some miRNAs are derived from "mirtrons", which are generated by splicing and debranching of short hairpin introns (Fig. 3) (129, 130). The 5′ and 3′ ends are defined by donor and acceptor splice sites, but in some cases include additional unstructured tails (131, 132). The biogenesis of 3′-tailed mirtrons in *Drosophila* was recently reported to utilize the RNA exosome, the major 3′-5′ exoribonuclease in eukaryotes (133). Indeed, there is increasing overlap in the factors involved in miRNAs biogenesis and other RNA processing pathways. The list of RNAs that feed into the miRNA biogenesis pathway is also increasing: snoRNAs (small nucleolar RNAs), tRNAs and endogenous shRNAs can be processed by Dicer into small RNA fragments that then mediate gene silencing (131, 134-136).

Dicer is generally considered essential for the biogenesis of miRNAs, but at least one highly conserved miRNA, miR-451, is produced by a Dicer-independent mechanism in human, mouse and zebrafish (137-139). The mature miRNA maps to the stem as well as loop sequence of the pre-miRNA and directly binds to Ago proteins (Figure. 3B). Ago1 and Ago3 can actively load pre-mir-451 but only Ago2 can process the miRNA since this requires the endonuclease activity (140). To date, no other Dicer-independent miRNAs have been identified and the specific features that dictate routing to Dicer versus Ago are under investigation (140). A recent report showed that pre-miRNAs could be designed to be processed by Ago2 as well as Integrator, eliminating the need for either Drosha or Dicer and opening up the possibility that such pathways could exist naturally (141).

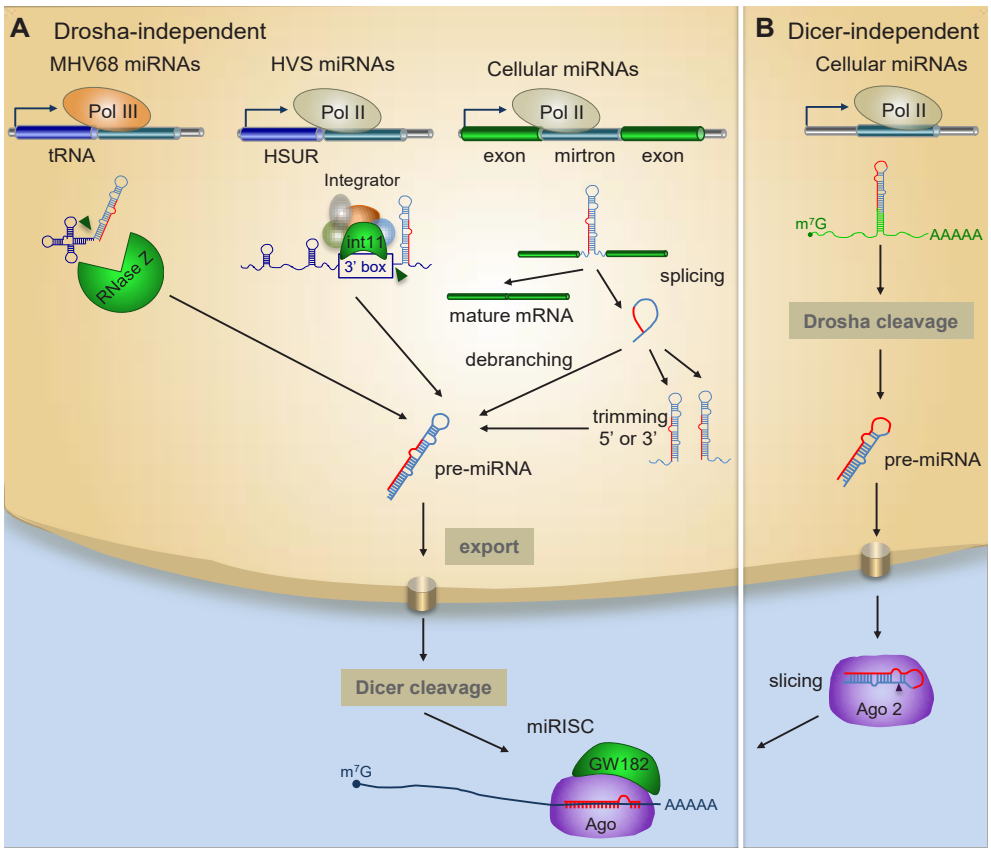


Figure 3. Alternative miRNA biogenesis pathways in animals and viruses (A) Drosha-independent biogenesis. Pre-miRNAs are co-transcribed with tRNAs in Pol III transcripts in MHV68 and bypass processing by Drosha. Pre-miRNA like miRNAs in HVS are derived from the same Pol II transcripts as HSURs and require the Integrator for generation of their 5′ ends. Cellular miRNAs termed mirtrons also do not require Drosha: they are Pol II transcripts that are excised by splicing and linearized by lariat debranching; tailed mirtrons require further 5′ or 3′ trimming by nucleases and then they are directly processed by Dicer. **(B)** Dicer-independent biogenesis. The highly conserved miRNA, miR-451 is produced in a Dicer-independent mechanism involving cleavage by Ago. The mature miRNA (red) derives from the stem as well as loop sequence of the pre-miRNA.

Regulation of miRNA biogenesis by Single Nucleotide Polymorphisms and RNA editing

Natural sequence variations in pri-miRNAs, pre-miRNAs or mature miRNAs can influence their processing, stability and target selection. These sequence variations originate from changes in DNA-coding sequence or from post-transcriptional modifications to the RNA (142-147). In humans, differences in processing by Drosha were observed for alleles of miR-125a, miR-126, miR-146a, miR-502, miR-510, miR-890, and miR-892b (143-145, 147), while alteration of processing by Dicer was postulated for SNPs (single nucleotide polymorphisms) in miR-196a (146). A natural variant of miR-934 was found to contain a mutation in the first nucleotide of the pre-miRNA, which affects strand selection for incorporation into RISC (145).

MiRNAs can also be post-transcriptionally modified by ADAR family members (Adenosine Deaminase acting on RNA proteins) which convert adenosines to inosine reviewed in (148). The hairpin structures of pre-miRNAs are favourable substrates for ADARs (149), which recognize dsRNA. Blow *et al.* sequenced 99 miRNAs from 10 human tissues and identified 6% of pri-miRNA transcripts with A to I conversions in at least one of the analysed tissues (150). Another survey reported that 16% of pri-miRNAs are edited in the brain, where there is generally a higher frequency of RNA editing (151). Editing can affect pri-miRNA and pre-miRNA processing and can also alter the target repertoire of the miRNA when editing occurs in the mature sequence (152-155). For example, editing of pri-miR-142 substantially reduces processing by Drosha and leads to cleavage by Tudor-SN (Tudor staphylococcal nuclease), a component of RISC with ribonuclease activity specific for inosine-containing dsRNAs (154, 155). In contrast, editing of pri-miR-151 by ADAR1 does not affect pri-miRNA processing but interferes with pre-miRNA cleavage by Dicer, as seen by accumulation of edited pre-miR-151 (Figure 1D) (153). The A to I conversion within the mature miRNA can retarget the miRNA to a new set of mRNAs since inosine base pairs with cytosine rather than uridine. For example, editing of sites within the miR-376 seed alters its target repertoire both *in vitro* and *in vivo* (152). Interestingly, Haele *et al.* reported that ADAR enzymes can also influence miRNA processing independently of their catalytic activity, suggesting that in some cases binding of the ADAR proteins alone might be sufficient to interfere with miRNA processing (156).

Some viral miRNAs have also been found to be edited, for example KSHV miR-K12-10 (40), Marek's disease virus miR-M7 (157) and EBV miR-BART6 (117). To date the functional relevance of this editing has only been suggested for the latter. In HEK-293 cells, editing of EBV miR-BART6-3p decreases the efficiency with which the miRNA encoded on the opposite strand, miR-BART6-5p, is loaded into RISC. Strikingly, miR-BART6-5p targets human Dicer via 4 binding sites in its 3'UTR. Therefore, editing of miR-BART6-3p relieves Dicer from posttranscriptional gene silencing. Dicer levels

affect the expression levels of multiple genes that regulate the infectious and lytic states of EBV and it is postulated that editing of miR-BART6-3p could be an indirect way to modulate miRNA biogenesis and thereby the viral life cycle (117).

Regulation of miRNA stability

Once a mature miRNA is incorporated into RISC it is generally considered to be extraordinary stable (158, 159). Indeed, upon inactivation of miRNA transcription or processing the majority of mature miRNAs in human and rodent cell lines have half-lives in the range of many hours to days (160, 161). However, recent reports from various model systems have demonstrated differences in the stabilities of individual miRNAs, suggesting that regulated degradation of specific miRNAs is a physiologically relevant way to modulate their expression, reviewed further in (162). In particular, active miRNA decay seems to play a prominent role in neurons. In mouse retinal cells the sensory neuron-specific miR-183/96/182 cluster and miR-204 and miR-211 are differentially expressed in response to light. The mature miRNAs are rapidly down regulated upon dark-adaptation due to active degradation by a yet unidentified enzyme (163). Several other brain-enriched miRNAs have short half lives both in primary human neuronal cell culture and post mortem brain tissue (164). The fast turnover is recapitulated in primary neurons outside the retina as well as in neurons derived from mouse embryonic stem cells. Strikingly, blocking of action potentials by inhibition of sodium channels prevented the degradation of selected miRNAs, indicating that activation of neurons is required for the regulated decay of some neuronal miRNAs (163). In line with this observation, a small RNA deep sequencing approach identified several brain-enriched miRNAs that also were rapidly down regulated upon transient exposure to the neurotransmitter serotonin in the marine snail *Aplysia* (165). Active miRNA decay represents an elegant way to re-activate neuronal transcripts, which might be important for a rapid response to various external stimuli (166-169). Regulated miRNA turnover also occurs during viral infection (described below), although to date the mechanisms of miRNA turnover in neurons or during infection in mammals remain unknown. However, studies from other model systems have identified molecular determinants of regulated miRNA decay and here we will summarize the current knowledge on these determinants and their modes of action.

Modifications to the 3'end of miRNAs

Chemical modifications of mature miRNAs plays a crucial role in regulating their stabilities. The first appreciation for miRNA stability factors came from studies in plants, where the methyltransferase HEN1 (Hua Enhancer 1) methylates the 2' hydroxyl group of the 3' terminal nucleotide of a miRNA (170-172). Methylation of plant miRNAs protects their 3' ends from terminal uridylation by the nucleotidyl transferase HESO1 (HEN1 Suppressor 1), which triggers their degradation (173-175). Uridylation at the 3'ends of

RNAs is also associated with reduced stability of piRNAs, siRNA and mRNAs (176-178). Similarly, a nucleotidyl transferase in the unicellular alga *Chlamydomonas reinhardtii*, MUT68, uridylates small RNAs leading to their degradation by the peripheral exosomal subunit RRP6 (ribosomal binding protein 6) (179).

Animal miRNAs generally lack a protective 2′*O*-methyl group at their 3′ terminus and display template-independent nucleotide addition, mostly adenylation or uridylation that may regulate miRNA stability (180-182). Several enzymes, including MTPAP, PAPD4/GLD2, PAPD5, ZCCHC6, TUT4/ZCCHC11, and PAPD2/TUT1 display terminal nucleotidyl transferase activity and knockdown experiments indicate that these proteins are responsible for miRNA 3′end variation to various extents (183, 184). However, functional implications have thus far been described for only a few of these enzymes. For example, TUT4, the nucleotidyl transferase implicated in the degradation of histone mRNA and several pre-miRNAs (73, 110, 178), regulates cytokine levels by uridylation of mature miR-26 family members (185). In the human A549 cell line, miR-26b targets the IL6 (interleukin 6) transcript but terminal uridylation of this miRNA interferes with its function. Knockdown of TUT4 results in reduced miR-26a uridylation along with decreased expression of a reporter containing the IL6 3′UTR. Conversely, overexpression of TUT4 leads to enhanced levels of the same reporter, indicating that uridylated miR-26a is less effective in targeting IL6. Notably, knockdown of TUT4 does not increase miR-26 expression levels, indicating that uridylation of the miRNA affects its activity without affecting its expression in this case (185).

Adenylation at the 3′ends of miRNAs is associated with both enhanced and decreased miRNA stability (186-189). For example, the most highly expressed miRNA in the liver, miR-122, is monoadenylated by the cytoplasmic poly(A) polymerase GLD2 (germline development defective-2). In GLD2 knockout mice, miR-122 is selectively destabilized whereas the levels of 10 other miRNAs remain unchanged. The stability of the miR-122 precursors is not affected by GLD2 knockout, suggesting a role for adenylation in modulating stability of the mature form (186).

Recently it was demonstrated that VACV induces polyadenylation of endogenous miRNAs during infection. The viral poly(A) polymerase is responsible for the non-templated adenylation that results in a ~30-fold reduction of endogenous miRNA levels in infected mouse embryonic fibroblasts; other small RNAs such as tRNAs and snRNAs remain largely unaffected by VACV infection. It was suggested that viral poly(A) polymerase operates only on Ago-bound small RNAs, but the mechanism is unknown. Whereas polyadenylation of miRNAs is mediated by a viral gene product, the actual degrading activity is postulated to stem from a yet undefined cellular protein (189). It is not clear if and how the modification of miRNAs by VACV is linked to the reduction in Dicer expression that was described previously (116); it may be that this virus uses two different mechanisms to shut-off cellular miRNA expression. Poxviruses infect a wide

range of vertebrate and invertebrate hosts. Infection of *Drosophila* cells with VACV leads to global reduction in miRNA expression whereas the levels of endogenous siRNAs are unaffected. Like plant miRNAs, insect siRNAs are methylated, which protects them from polyadenylation by the virus. Indeed 3' methylation of a transfected miRNA prevents it from being polyadenylated and degraded during infection (189). The advent of deep sequencing technology has enabled a much greater appreciation for the extent of heterogeneity and modifications at the 3'ends of miRNAs (182, 183, 190). In the coming years it will be important to further characterise the enzymes that write and read these modifications and to understand their impact on miRNA stability and function.

Sequence motifs regulating miRNA stability

Several reports have demonstrated altered kinetics in the turnover of individual miRNAs under conditions where the expression levels of most miRNAs are unchanged (160, 161). This suggests that cis acting elements in the mature miRNA sequence provide specificity to the miRNA degradation process. In a survey to characterise the role of miRNA turnover during the cell cycle, Rissland and colleagues (191) found that miR-503 and other members of the extended miR-16 family are constitutively unstable in NIH-3T3 cells. The high turnover rate allows dynamic transcriptional regulation of these miRNAs during the cell cycle. For example, miR-503 is rapidly down regulated upon cell cycle re-entry but accumulates during cell cycle arrest by serum starvation. Sequence elements within the seed and 3' end of the miRNA appear to be required for the degradation. Similarly, miR-382 is selectively unstable in HEK293 cells and an element in the 3' end of the miRNA is required for its enhanced turnover *in vitro* (160). Optimal paradigms to study cis acting elements with a role in miRNA decay are miRNAs that are co-transcribed and highly similar on a sequence level, yet differ in their decay rates. The miR-29 family provides such an example: miR-29b is unstable in cycling cells and only accumulates during mitosis whereas miR-29a is stable throughout the cell cycle (78). miR-29a and miR-29b share the same seed sequence but are distinguished by a C to U substitution at position 10 and miR-29b contains a hexanucleotide motif (AGUGUU) at its 3' end that is responsible its nuclear localisation. However, the motif does not account for the accelerated miRNA decay. Instead, uridines at position 9-11 in miR-29b seem to enhance destabilisation and many, but not all, miRNAs that contain a uridine stretch at this position are reported to display faster turnover rates (192). Therefore, additional factors must dictate the differential stability of miRNAs. Altogether these studies show that miRNAs, though limited in coding space, contain sequence elements outside the classical seed that may critically influence miRNA abundance and function. To date, no viral miRNAs have been reported to contain such motifs, but this could provide another strategy for viruses to diversify miRNA function and regulation during their life cycles. Identification of the *trans*-acting factors that recognise these motifs is important for further investigations in this area.

Trans-acting factors regulating miRNA stability

The first report of enzymes that are capable of degrading single-stranded small RNAs came from a candidate gene approach in plants. In *Arabidopsis*, SDN1 (Small RNA degrading nuclease 1) possesses 3′-5′ exonuclease activity on small RNAs including miRNAs. In a cell free assay system, SDN1 specifically degrades ssRNA but not dsRNA. The 2′*O*-methylation present on the 3′ terminal nucleotide of plant miRNAs is protective against SDN1 activity (193). The enzyme belongs to a family of exoribonucleases with partially overlapping functions *in vivo* that are responsible for miRNA turnover in plants. Interestingly, members of this protein family are conserved in all eukaryotes and it seems likely that animal homologues of SDNs have similar functions but these have not yet been reported (194). The XRN family of enzymes play various roles in miRNA stability in different organisms: in *Arabidopsis* XRN2 and XRN3 are involved in degrading the loop sequence of pre-miRNAs (195), in mammalian cells XRN2 degrades the pri-miRNA following processing by Drosha (43, 196). In *C. elegans*, XRN2 degrades mature miRNAs once released from the RISC complex and may also influence the rate at which they are released (197). Interestingly, the presence of target RNA counteracts the decay of miRNAs by XRN2 both *in vitro* and *in vivo* (197, 198). Whether this is due to direct competition between the target and XRN2 for miRNA binding or through another molecular mechanism is not yet known. The exoribonuclease XRN1 and the exosome core subunit Rrp42 (ribosomal RNA-processing protein-42) are proposed to be involved in turnover of miR-382 in HEK293 cells, as knock-down of these factors selectively increases miR-382 expression levels (160).

In a human melanoma cell line, ectopic expression of hPNPase$^{old\text{-}35}$ (human polynucleotide phosphorylase protein) leads to the selective down regulation of several miRNAs (miR-221, miR-222 and miR-106b). Immunoprecipitation studies show that this 3′-5′ exoribonuclease directly associates with these miRNAs and causes their degradation *in vitro*. However, it remains unclear whether hPNPase$^{old\text{-}35}$ is also able to actively dislodge them from the RISC complex. Interestingly, hPNPase$^{old\text{-}35}$ is an interferon-stimulated gene and mediates IFN-β-induced down regulation of miR-221. One of the direct targets of miR-221 is the cell-cycle suppressor p27^{kip1}. Consequently, both miR-221 overexpression and knockdown of hPNPase$^{old\text{-}35}$ protect human melanoma cells from INF-β- induced growth arrest, indicating a pivotal role of controlled miRNA decay in tuning cell proliferation (199). The 3′-5′ exoribonuclease Eri1 was recently implicated in regulating miRNA stability in mouse lymphocytes, based on the global increase in miRNA levels observed in NK and T cells from Eri1 knockout mice (200). The regulation of miRNA levels by Eri1 appears to be required for NK-cell development and antiviral immunity, but its mechanism of action remains to be established.

Besides promoting miRNA degradation, RNA binding proteins can also enhance the stability of mature miRNAs. For example, Quaking, a member of the STAR (signal transduction and activation of RNA) family of RNA binding proteins, is up regulated

in response to p53 signalling and stabilises mature miR-20a (201). The identification of proteins that stabilise and de-stabilise mature miRNAs supports the idea that regulation of miRNA decay is important in controlling the miRNA repertoire of the cell. Yet, there are still major gaps in understanding how specificity in degradation or stabilization is mediated.

Target mediated miRNA turnover

In contrast to target-mediated stabilization of miRNA in *C. elegans*, binding of miRNAs to RNAs can promote miRNA degradation in *Drosophila* and mammals. In flies, most miRNAs are incorporated in Ago1-containing RISC complexes whereas siRNAs, usually derived from dsRNA from viruses and transposons, are loaded into Ago2 (202) and are 3' methylated by the *Drosophila* homolog of HEN-1 (203). Intriguingly, binding of Ago1 associated miRNAs to target sites with extensive complementarity results in destabilization of the miRNAs (204). Deep sequencing the small RNAs revealed that a large proportion of these miRNAs are either shortened or have non-templated nucleotide additions at their 3'ends (mostly adenines and uridines). This mechanism of trimming and tailing, mediated by as yet unknown enzymes, seems to precede miRNA decay (Figure 4). In contrast, miRNAs that associate with Ago2 and thus are methylated appear to be protected from degradation. In human cells, miRNAs are also subject to this target-directed destabilisation, as evidenced by trimming and tailing in Hela cells *in vitro* (204). Baccarini and colleagues examined in more detail the fate of a miRNA molecule after target recognition and demonstrated that miRNAs generally out-live their targets, whether the target is perfectly complementary or contains a central bulge. However, target recognition promotes post-transcriptional modification of miRNAs (mostly 3' uridylation) which is postulated to induce their degradation, thereby limiting miRNA recycling (205). It is not yet known what features in the target RNA direct the posttranscriptional modification of a miRNA but this may extensive pairing as proposed in flies (204).

Two distinct mammalian herpesviruses, a gamma herpesvirus infecting new world primates and a beta herpesvirus infecting mice, exploit the mechanism of target-directed miRNA degradation (Figure 4). Several HSURs are expressed in Herpesvirus saimiri (HVS)-transformed T-cells and one of these, HSUR1, contains an interaction site for the endogenous miRNA, miR-27 (206). Cazalla *et al.* showed that binding of HSUR1 to miR-27 accelerates its rate of turnover and replacing the miR-27 interaction site with a binding site for miR-20 re-targets HSUR1 to the other miRNA (206). Similarly, miR-27 is also rapidly down regulated in murine cytomegalovirus (MCMV) infection of several mouse cell lines as well as primary macrophages. Yet, the expression levels of miR-27 precursors remain stable, indicating that the mature form is subject to enhanced degradation, presumably by a viral inhibitor (207). Indeed, the MCMV *m169* gene contains a binding site for miR-27 in its 3' UTR and miR-27 levels are rescued if the *m169* gene is knocked

down or deleted from the virus (208, 209). During lytic MCMV infection, *m169* is among the most highly transcribed genes (209) and it represents the most frequent non-miRNA segment sequenced in Ago2 immunoprecipitations (208). Down regulation of miR-27 is linked to its 3' end tailing and trimming, indicating that a similar mechanism as suggested in flies and human cells might underlie the degradation process (204, 209). As reported for HSUR1, replacing the miR-27 binding site with an interaction site for an unrelated miRNA is able to redirect *m169* to target that specific miRNA (206, 208, 209). The degradation of miR-27 by two distinct herpesviruses might suggest that this miRNA plays an important role in the viral life cycles. Indeed, miR-27 represses MCMV replication when over-expressed in cell culture experiments (207) and MCMV mutants incapable of down regulating miR-27 display attenuated viral growth in mice (209). So far, however, it remains unclear which cellular miR-27 target(s) are responsible for modulating MCMV replication and whether it plays the same role in both MCMV and HVS infections. In summary, the pairing patterns of miRNAs with their targets as well as the relative amounts of each seem to be crucial factors that determine the extent of target-mediated miRNA decay (205). A range of reports suggest that endogenous mRNAs, noncoding RNAs and pseudogenes also play a role in regulating miRNA activity and/or stability, reviewed in (210).

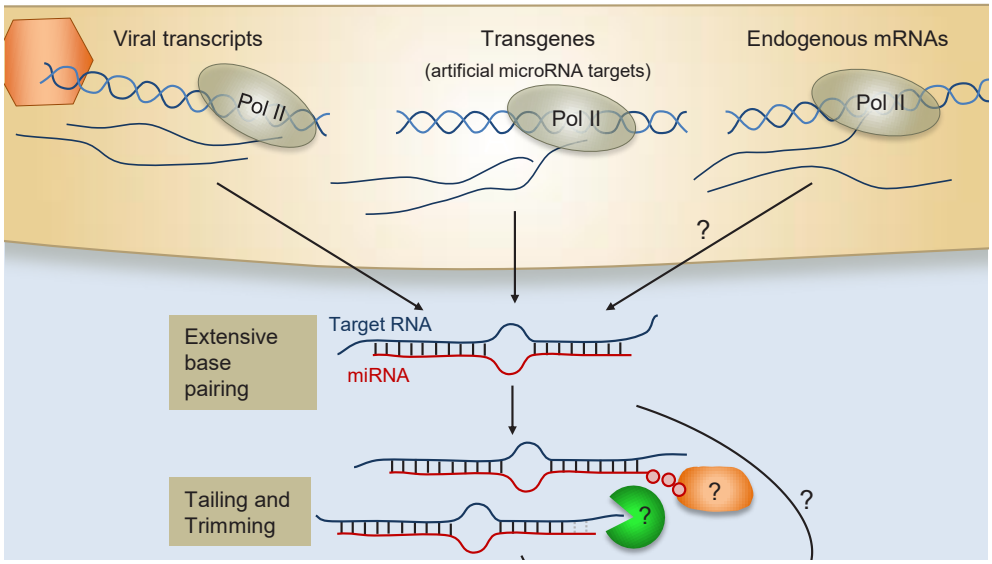


Figure 4: Target-mediated miRNA degradation. Different sources of target RNA can induce miRNA decay including two herpesviral transcripts (Herpesvirus saimiri HSUR1 and murine Cytomegalovirus m169) and transgenic miRNA targets with extensive basepairing. Whether there are endogenous mRNAs that induce miRNA degradation remains to be investigated. Both in vertebrates and invertebrates target-mediated miRNA degradation has been associated with tailing and trimming of miRNAs. The relationship between tailing and trimming is still unclear, and the factors involved in mediating these effects and subsequent degradation remain to be determined.

Viral suppressors of RNA interference may modulate miRNA expression

In insects and plants RNA silencing pathways mediate a potent antiviral response. For efficient replication, viruses that infect these hosts therefore rely on virus-encoded suppressors of RNAi (VSRs) (211). Also in mammalian viruses, proteins with RNAi suppressive activity have been identified, although the importance of this suppressive activity *in vivo* remains to be established (212, 213). In the following section we will discuss how the expression of these VSR affects miRNA biosynthesis in insects and plants and we will further speculate about their possible influence on miRNA expression in mammals.

The RNA interference machinery in insects recognizes viral dsRNA in the cytoplasm and processes it into vsiRNA (viral siRNAs) (211). These vsiRNAs associate with Ago2-containing RISC complexes, which then act as antiviral effectors by cleaving viral RNA in the cytoplasm (214). Whereas the production of siRNA and miRNA molecules in mammals largely rely on the same biogenesis factors, the miRNA and antiviral RNAi pathways in insects are governed by a distinct set of processing and effector complexes. Specifically, pre-miRNAs are processed by Dicer-1 to be loaded into Ago1-containing RISC complexes. In contrast, cytoplasmic long dsRNA is sensed and cleaved by Dicer-2 and the resulting 21 nt siRNAs are predominantly loaded into Ago2-containing RISC (202, 215, 216). Insect VSRs interfere with the RNAi machinery at different stages of the pathway. Drosophila C virus 1A for example binds long dsRNA, thereby preventing its efficient processing into siRNA (214). Flock house virus B2 binds both long dsRNA and siRNAs (217-220). Cricket Paralysis virus 1A and Noravirus VP1 directly interact with the small RNA-loaded Ago2 effector complex and prevent its target RNA cleavage activity (221, 222) (and unpublished observations).

Although the siRNA and miRNA biogenesis machineries are distinct in insects and plants, many VSRs have dsRNA binding properties, and it might be expected that they could affect miRNA processing too. However, this does not seem to be the case in flies. VSR expression in transgenic *Drosophila* does not alter levels of mature miRNAs nor does it affect the activities of miRNA reporters. Furthermore, in contrast to Ago1 loss-of-function mutants, transgenic animals expressing VSRs do not display developmental defects, suggesting that VSRs do not affect global miRNA biogenesis and function (214, 219, 221-223). In contrast, transgenic expression of VSRs in plants leads to pleiotropic, developmental defects due to alterations in miRNA-mediated gene regulation (224-226). This is likely based on the convergence of the plant siRNA and miRNA biogenesis pathways, which use the same processing factors. For instance, both miRNAs and antiviral siRNAs can be loaded into Ago1 effector complexes in plants (227-229). Yet, for many plant VSRs it remains elusive how they manipulate the miRNA machinery *in vivo.* A number of VSRs have dsRNA binding activity *in vitro*, which has been hypothesised to explain their interference with miRNA biogenesis

(230-235). For instance, Tombusvirus P19 directly binds siRNA duplexes preventing their efficient loading into effector RISC complexes *in vitro* (224, 225, 230, 236-238). In transgenic *Arabidopsis*, P19 also prevents miRNA loading into Ago1-containing RISC. However, this seems to be a rather exceptional property as three other VSRs tested, Turnip crinckle virus P38, Peanut Clump virus P15, Turnip mosaic virus Hc-Pro, blocked siRNA loading into Ago1 but did not disturb its association with miRNAs (238).

A number of plant VSRs may act on the miRNA machinery in other ways than by small RNA sequestration. Turnip crinckle virus (TCV) P38 and Sweet potato mild mottle virus (SPMMV) P1 directly interact with the siRNA/miRNA effector Ago1 by mimicking the glycine/tryptophane (GW)/WG repeats normally found in host proteins that associate with Ago proteins (239, 240). Indeed, host miRNA levels were reduced in TCV infections (240) and P1 expression suppresses silencing of a miRNA sensor (239). However, in a study using transgenic Arabidopsis, P38 did not suppress accumulation of miRNAs in Ago1-containing RISC complexes (238), which might reflect the differences between the two model systems (TCV infection versus P38 transgenic plants). Beet western yellow virus P0 has been suggested to target Ago1 for degradation by acting as a F-box protein (241-244). F-box proteins are components of E3 ubiquitin ligase complexes, which target proteins for ubiquitination and subsequent degradation (245). Interestingly, the VSR activity of P0 is insensitive to proteasome inhibition, indicating that P0 induces Ago1 degradation via a non-canonical pathway (241). Besides suppression of dsRNA-induced RNAi, transgenic expression of P0 in *Arabidopsis* causes developmental defects reminiscent of miRNA pathway-defective plants. Indeed, six out of twelve analysed miRNA target genes have elevated expression levels suggesting that P0 also affects the miRNA pathway (242). The indications that P38, P1 and P0 inhibit both (v)siRNA and miRNA biogenesis may reflect the convergence of these two pathways on Ago1 (227-229).

In mammalian cells, virus infection triggers a potent protein-based immune response and it remains unclear to what extent RNAi-based mechanisms contribute to antiviral immunity. Yet, three lines of evidence support the idea that vsiRNAs could contribute to antiviral immune defence in mammals. First, in a broad small RNA deep-sequencing survey of six different RNA virus infecting multiple hosts virus-derived small RNAs were discovered in 4 positive (+) strand RNA viruses and 1 negative (-) strand RNA virus (246). However, the origin, Dicer-dependence, and functional importance of these small RNAs remains to be established. Second, siRNAs engineered to target viruses restrict virus growth in several mammalian model systems (247, 248). This suggests that the RNAi pathway could have intrinsic antiviral activity, provided that vsiRNAs are naturally generated at sufficient levels. Third, several viruses were suggested to encode proteins that suppress RNAi in mammalian cells, including Influenza virus NS1, Vaccinia virus E3L, Nodamura virus B2, La Crosse virus NSs, HIV Tat and Ebola virus VP30, VP35 and VP40 (216, 249-253). Many of these VSRs, including NS1, E3, VP30 and VP35, have dsRNA binding activity. Influenza NS1 protein has been demonstrated to

function as VSR only in heterologous plant and *Drosophila* cell systems (216, 254, 255). In mammalian cells this protein fails to suppress RNAi induced by exogenous shRNA or siRNAs (256). VSR activity of Nodamura virus B2 has also been attributed to its RNA binding properties. B2 binds both siRNAs and shRNAs and interferes with Dicer processing in mammalian cells *in vitro* (249). Since pre-miRNAs are structurally similar to shRNAs it is expected that this VSR could bind pre-miRNAs and thereby hinder their processing. Indeed, human cells stably expressing NoV B2 display elevated levels of pre-let-7d, suggesting that efficient Dicer processing of this pre-miRNA is inhibited (249). However, this effect was not observed for two other endogenous miRNAs and the mechanism has not been examined further (249). Nonetheless, these results demonstrate that viral RNA binding proteins have the potential to interfere with miRNA biogenesis through RNA-protein interactions.

In contrast to RNA binding, VSRs may also function through direct interaction with protein components of the mammalian RNAi machine. Ebola virus VP30 and VP35 can directly interact with Dicer or with Dicer-associated factors TRBP and PACT, and thereby inhibit the production of functional siRNAs (252, 253). Unlike the small RNA biogenesis machinery in insects, mammalian cells only express one Dicer that is responsible for both the production of siRNAs and miRNAs (5). Inhibition of Dicer processing by VP30 and VP35 is therefore expected to interfere with pre-miRNA processing but this requires further experimental validation. Similarly, the HIV Tat protein has been suggested to interfere with Dicer processing of shRNAs *in vitro* (251). Tat associates with Dicer in an RNA-dependent manner but the molecular identity of the required RNA is still unknown (257). Furthermore, it remains elusive if the Tat-Dicer interaction is necessary for the VSR activity of Tat. A retrovirus, Primate foamy virus (PFV) type 1 encodes the Tas protein, which has been suggested to be a non-specific suppressor of miRNA-mediated silencing with an as yet unknown mode of action (258). Interestingly, PFV is efficiently targeted by the host miR-32 and inhibiting this cellular miRNAs with locked nucleic acid miRNA antagonists enhances PFV replication. Blocking the miRNA-virus interaction may thus represent a major function of Tas VSR activity. However, the antiviral activity of miR-32 remains an item of debate (259), as does the functional importance of retroviral VSRs. For example, Qian *et al*. suggest that HIV Tat protein suppresses RNAi by inhibiting a step downstream of siRNA processing (260). In another study, overexpression of both HIV tat and PFV Tas failed to suppress shRNA-induced RNAi in human cells (261).

To conclude, a number of mammalian VSRs have the potential to actively manipulate host miRNA biogenesis either through interactions with RNA or protein components of the small RNA processing machinery. Yet, for most candidate VSRs, firm support for a global change of miRNA levels or activity in the context of an authentic infection is lacking. Making use of high throughput sequencing and screening approaches it will be possible to assess to what extent VSRs contribute to changes in miRNA expression or activity in infected mammalian cells.

CONCLUSIONS

Since their initial discovery nearly 20 years ago miRNAs have been shown to play fundamental roles in virtually all cell-biological processes. Therefore it is not surprising that their expression is tightly regulated in a spatio-temporal fashion. There are many mechanisms by which miRNAs can be produced and subsequently regulated in mammalian cells. Studies of viral systems have revealed diversity in the origin of miRNAs, the factors required for their synthesis, and the factors that can control their turnover. In some cases, viruses influence global expression levels of miRNAs, in-line with their mode of action in targeting RNAi pathways in plants and insects. However, as reviewed here, miRNAs play diverse functional roles in a cell and there are numerous mechanisms for regulating specific subsets of miRNAs, or individual miRNAs, rather than the global machinery. It appears that some viruses such as HVS and MCMV have tapped into these modes of regulation, most likely in order to precisely control specific pathways in the host cell. With the advancement of RNA-protein mapping techniques and sequencing technologies it is likely that many more viral-host interactions targeting miRNA regulation will emerge.

ACKNOWLEDGEMENTS

We thank G. Michlewski and D. Santhakumar for comments on the manuscript. This work was financially supported by VIDI fellowship (project number 864.08.003) from the Netherlands Organization for Scientific Research to RvR, by a PhD fellowship from Radboud University Nijmegen Medical Centre to P.M. Research in AHB's lab is supported by the BBSRC (BB/J001279) and a Wellcome Trust RCDF (WT097394A1A).

REFERENCES

1. Kim VN, Han J, Siomi MC. Biogenesis of small RNAs in animals. Nat Rev Mol Cell Biol. 2009;10(2):126-39.
2. Ghildiyal M, Zamore PD. Small silencing RNAs: an expanding universe. Nat Rev Genet. 2009;10(2):94-108.
3. Malone CD, Hannon GJ. Small RNAs as Guardians of the Genome. Cell. 2009;136(4):656-68.
4. Cenik ES, Zamore PD. Argonaute proteins. Curr Biol. 2011;21(12):R446-9.
5. Carthew RW, Sontheimer EJ. Origins and Mechanisms of miRNAs and siRNAs. Cell. 2009;136(4):642-55.
6. Fabian MR, Sonenberg N. The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. Nat Struct Mol Biol. 2012;19(6):586-93.
7. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell. 2009;136(2):215-33.
8. Friedman RC, Farh KKH, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. Genome research. 2009;19(1):92-105.
9. Lee YS, Dutta A. MicroRNAs in cancer. Annu Rev Pathol. 2009;4:199-227.
10. Mendell JT, Olson EN. MicroRNAs in stress signaling and human disease. Cell. 2012;148(6):1172-87.

11. Xiao C, Rajewsky K. MicroRNA control in the immune system: basic principles. Cell. 2009;136(1):26-36.
12. Chang TC, Mendell JT. microRNAs in vertebrate physiology and human disease. Annu Rev Genomics Hum Genet. 2007;8:215-39.
13. Schickel R, Boyerinas B, Park SM, Peter ME. MicroRNAs: key players in the immune system, differentiation, tumorigenesis and cell death. Oncogene. 2008;27(45):5959-74.
14. Ambros V. The functions of animal microRNAs. Nature. 2004;431(7006):350-5.
15. Grundhoff A, Sullivan CS. Virus-encoded microRNAs. Virology. 2011;411(2):325-43.
16. Hussain M, Torres S, Schnettler E, Funk A, Grundhoff A, Pijlman GP, et al. West Nile virus encodes a microRNA-like small RNA in the 3' untranslated region which up-regulates GATA4 mRNA and facilitates virus replication in mosquito cells. Nucleic Acids Res. 2012;40(5):2210-23.
17. Cullen BR. Viruses and microRNAs: RISCy interactions with serious consequences. Genes & development. 2011;25(18):1881-94.
18. Kincaid RP, Burke JM, Sullivan CS. RNA virus microRNA that mimics a B-cell oncomiR. Proceedings of the National Academy of Sciences of the United States of America. 2012;109(8):3077-82.
19. Ouellet DL, Plante I, Landry P, Barat C, Janelle ME, Flamand L, et al. Identification of functional microRNAs released through asymmetrical processing of HIV-1 TAR element. Nucleic Acids Res. 2008;36(7):2353-65.
20. Omoto S, Fujii YR. Regulation of human immunodeficiency virus 1 transcription by nef microRNA. J Gen Virol. 2005;86(Pt 3):751-5.
21. Omoto S, Ito M, Tsutsumi Y, Ichikawa Y, Okuyama H, Brisibe EA, et al. HIV-1 nef suppression by virally encoded microRNA. Retrovirology. 2004;1:44.
22. 22. Skalsky RL, Cullen BR. Viruses, microRNAs, and host interactions. Annu Rev Microbiol. 2010;64:123-41.
23. Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, et al. The nuclear RNase III Drosha initiates microRNA processing. Nature. 2003;425(6956):415-9.
24. Yi R, Qin Y, Macara IG, Cullen BR. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev. 2003;17(24):3011-6.
25. Bernstein E, Caudy AA, Hammond SM, Hannon GJ. Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature. 2001;409(6818):363-6.
26. Knight SW, Bass BL. A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in Caenorhabditis elegans. Science. 2001;293(5538):2269-71.
27. Grishok A, Pasquinelli AE, Conte D, Li N, Parrish S, Ha I, et al. Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. Cell. 2001;106(1):23-34.
28. Hutvagner G, McLachlan J, Pasquinelli AE, Balint E, Tuschl T, Zamore PD. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science. 2001;293(5531):834-8.
29. Ketting RF, Fischer SE, Bernstein E, Sijen T, Hannon GJ, Plasterk RH. Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. Genes Dev. 2001;15(20):2654-9.
30. Mourelatos Z, Dostie J, Paushkin S, Sharma A, Charroux B, Abel L, et al. miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. Genes Dev. 2002;16(6):720-8.
31. Hammond SM, Boettcher S, Caudy AA, Kobayashi R, Hannon GJ. Argonaute2, a link between genetic and biochemical analyses of RNAi. Science. 2001;293(5532):1146-50.
32. Chen CY, Zheng D, Xia Z, Shyu AB. Ago-TNRC6 triggers microRNA-mediated decay by promoting two deadenylation steps. Nature structural & molecular biology. 2009;16(11):1160-6.
33. Millar AA, Waterhouse PM. Plant and animal microRNAs: similarities and differences. Funct Integr Genomics. 2005;5(3):129-35.

34. Monteys AM, Spengler RM, Wan J, Tecedor L, Lennox KA, Xing Y, et al. Structure and activity of putative intronic miRNA promoters. Rna. 2010;16(3):495-505.
35. Ozsolak F, Poling LL, Wang Z, Liu H, Liu XS, Roeder RG, et al. Chromatin structure analyses identify miRNA promoters. Genes Dev. 2008;22(22):3172-83.
36. Wang X, Xuan Z, Zhao X, Li Y, Zhang MQ. High-resolution human core-promoter prediction with CoreBoost_HM. Genome Res. 2009;19(2):266-75.
37. Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH, et al. MicroRNA genes are transcribed by RNA polymerase II. EMBO J. 2004;23(20):4051-60.
38. Borchert GM, Lanier W, Davidson BL. RNA polymerase III transcribes human microRNAs. Nat Struct Mol Biol. 2006;13(12):1097-101.
39. Diebel KW, Smith AL, van Dyk LF. Mature and functional viral miRNAs transcribed from novel RNA polymerase III promoters. Rna. 2010;16(1):170-85.
40. Pfeffer S, Sewer A, Lagos-Quintana M, Sheridan R, Sander C, Grasser FA, et al. Identification of microRNAs of the herpesvirus family. Nat Methods. 2005;2(4):269-76.
41. Gregory RI, Yan KP, Amuthan G, Chendrimada T, Doratotaj B, Cooch N, et al. The Microprocessor complex mediates the genesis of microRNAs. Nature. 2004;432(7014):235-40.
42. Han J, Lee Y, Yeom KH, Nam JW, Heo I, Rhee JK, et al. Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell. 2006;125(5):887-901.
43. Morlando M, Ballarino M, Gromak N, Pagano F, Bozzoni I, Proudfoot NJ. Primary microRNA transcripts are processed co-transcriptionally. Nat Struct Mol Biol. 2008;15(9):902-9.
44. Kim YK, Kim VN. Processing of intronic microRNAs. EMBO J. 2007;26(3):775-83.
45. Pawlicki JM, Steitz JA. Primary microRNA transcript retention at sites of transcription leads to enhanced microRNA production. J Cell Biol. 2008;182(1):61-76.
46. Fukuda T, Yamagata K, Fujiyama S, Matsumoto T, Koshida I, Yoshimura K, et al. DEAD-box RNA helicase subunits of the Drosha complex are required for processing of rRNA and a subset of microRNAs. Nat Cell Biol. 2007;9(5):604-11.
47. Davis BN, Hilyard AC, Lagna G, Hata A. SMAD proteins control DROSHA-mediated microRNA maturation. Nature. 2008;454(7200):56-61.
48. Davis BN, Hilyard AC, Nguyen PH, Lagna G, Hata A. Smad proteins bind a conserved RNA sequence to promote microRNA maturation by Drosha. Mol Cell. 2010;39(3):373-84.
49. Blahna MT, Hata A. Smad-mediated regulation of microRNA biosynthesis. FEBS Lett. 2012;586(14):1906-12.
50. Suzuki HI, Yamagata K, Sugimoto K, Iwamoto T, Kato S, Miyazono K. Modulation of microRNA processing by p53. Nature. 2009;460(7254):529-33.
51. Yamagata K, Fujiyama S, Ito S, Ueda T, Murata T, Naitou M, et al. Maturation of microRNA is hormonally regulated by a nuclear receptor. Mol Cell. 2009;36(2):340-7.
52. Kawai S, Amano A. BRCA1 regulates microRNA biogenesis via the DROSHA microprocessor complex. J Cell Biol. 2012;197(2):201-8.
53. Sakamoto S, Aoki K, Higuchi T, Todaka H, Morisawa K, Tamaki N, et al. The NF90-NF45 complex functions as a negative regulator in the microRNA processing pathway. Mol Cell Biol. 2009;29(13):3754-69.
54. Yu B, Bi L, Zheng B, Ji L, Chevalier D, Agarwal M, et al. The FHA domain proteins DAWDLE in Arabidopsis and SNIP1 in humans act in small RNA biogenesis. Proc Natl Acad Sci U S A. 2008;105(29):10073-8.
55. Gruber JJ, Zatechka DS, Sabin LR, Yong J, Lum JJ, Kong M, et al. Ars2 links the nuclear cap-binding complex to RNA interference and cell proliferation. Cell. 2009;138(2):328-39.
56. Sabin LR, Zhou R, Gruber JJ, Lukinova N, Bambina S, Berman A, et al. Ars2 regulates both miRNA- and siRNA- dependent silencing and suppresses RNA virus infection in Drosophila. Cell. 2009;138(2):340-51.

57. Michlewski G, Guil S, Semple CA, Caceres JF. Posttranscriptional regulation of miRNAs harboring conserved terminal loops. Mol Cell. 2008;32(3):383-93.
58. Guil S, Caceres JF. The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. Nat Struct Mol Biol. 2007;14(7):591-6.
59. Michlewski G, Caceres JF. Antagonistic role of hnRNP A1 and KSRP in the regulation of let-7a biogenesis. Nat Struct Mol Biol. 2010;17(8):1011-8.
60. Trabucchi M, Briata P, Garcia-Mayoral M, Haase AD, Filipowicz W, Ramos A, et al. The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs. Nature. 2009;459(7249):1010-4.
61. Zhang X, Wan G, Berger FG, He X, Lu X. The ATM kinase induces microRNA biogenesis in the DNA damage response. Mol Cell. 2011;41(4):371-83.
62. Briata P, Lin WJ, Giovarelli M, Pasero M, Chou CF, Trabucchi M, et al. PI3K/AKT signaling determines a dynamic switch between distinct KSRP functions favoring skeletal myogenesis. Cell Death Differ. 2012;19(3):478-87.
63. Kawahara Y, Mieda-Sato A. TDP-43 promotes microRNA biogenesis as a component of the Drosha and Dicer complexes. Proc Natl Acad Sci U S A. 2012;109(9):3347-52.
64. Wu H, Sun S, Tu K, Gao Y, Xie B, Krainer AR, et al. A splicing-independent function of SF2/ASF in microRNA processing. Mol Cell. 2010;38(1):67-77.
65. Moss EG, Lee RC, Ambros V. The cold shock domain protein LIN-28 controls developmental timing in C. elegans and is regulated by the lin-4 RNA. Cell. 1997;88(5):637-46.
66. Newman MA, Thomson JM, Hammond SM. Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. Rna. 2008;14(8):1539-49.
67. Viswanathan SR, Daley GQ, Gregory RI. Selective blockade of microRNA processing by Lin28. Science. 2008;320(5872):97-100.
68. Piskounova E, Polytarchou C, Thornton JE, LaPierre RJ, Pothoulakis C, Hagan JP, et al. Lin28A and Lin28B inhibit let-7 microRNA biogenesis by distinct mechanisms. Cell. 2011;147(5):1066-79.
69. Heo I, Joo C, Cho J, Ha M, Han J, Kim VN. Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. Mol Cell. 2008;32(2):276-84.
70. Lehrbach NJ, Armisen J, Lightfoot HL, Murfitt KJ, Bugaut A, Balasubramanian S, et al. LIN-28 and the poly(U) polymerase PUP-2 regulate let-7 microRNA processing in Caenorhabditis elegans. Nat Struct Mol Biol. 2009;16(10):1016-20.
71. Rybak A, Fuchs H, Hadian K, Smirnova L, Wulczyn EA, Michel G, et al. The let-7 target gene mouse lin-41 is a stem cell specific E3 ubiquitin ligase for the miRNA pathway protein Ago2. Nat Cell Biol. 2009;11(12):1411-20.
72. Rybak A, Fuchs H, Smirnova L, Brandt C, Pohl EE, Nitsch R, et al. A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment. Nat Cell Biol. 2008;10(8):987-93.
73. Hagan JP, Piskounova E, Gregory RI. Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells. Nat Struct Mol Biol. 2009;16(10):1021-5.
74. Piskounova E, Viswanathan SR, Janas M, LaPierre RJ, Daley GQ, Sliz P, et al. Determinants of microRNA processing inhibition by the developmentally regulated RNA-binding protein Lin28. J Biol Chem. 2008;283(31):21310-4.
75. Nam Y, Chen C, Gregory RI, Chou JJ, Sliz P. Molecular basis for interaction of let-7 microRNAs with Lin28. Cell. 2011;147(5):1080-91.
76. Zisoulis DG, Kai ZS, Chang RK, Pasquinelli AE. Autoregulation of microRNA biogenesis by let-7 and Argonaute. Nature. 2012;486(7404):541-4.
77. Tang R, Li L, Zhu D, Hou D, Cao T, Gu H, et al. Mouse miRNA-709 directly regulates miRNA-15a/16-1 biogenesis at the posttranscriptional level in the nucleus: evidence for a microRNA hierarchy system. Cell Res. 2012;22(3):504-15.
78. Hwang HW, Wentzel EA, Mendell JT. A hexanucleotide element directs microRNA nuclear import. Science. 2007;315(5808):97-100.

79. Robb GB, Brown KM, Khurana J, Rana TM. Specific and potent RNAi in the nucleus of human cells. Nature structural & molecular biology. 2005;12(2):133-7.

80. O'Neill LA, Sheedy FJ, McCoy CE. MicroRNAs: the fine-tuners of Toll-like receptor signalling. Nat Rev Immunol. 2011;11(3):163-75.

81. Triboulet R, Chang HM, Lapierre RJ, Gregory RI. Post-transcriptional control of DGCR8 expression by the Microprocessor. Rna. 2009;15(6):1005-11.

82. Han J, Pedersen JS, Kwon SC, Belair CD, Kim YK, Yeom KH, et al. Posttranscriptional crossregulation between Drosha and DGCR8. Cell. 2009;136(1):75-84.

83. Kadener S, Rodriguez J, Abruzzi KC, Khodor YL, Sugino K, Marr MT, 2nd, et al. Genome-wide identification of targets of the drosha-pasha/DGCR8 complex. Rna. 2009;15(4):537-45.

84. Barad O, Mann M, Chapnik E, Shenoy A, Blelloch R, Barkai N, et al. Efficiency and specificity in microRNA biogenesis. Nat Struct Mol Biol. 2012;19(6):650-2.

85. Macias S, Plass M, Stajuda A, Michlewski G, Eyras E, Caceres JF. DGCR8 HITS-CLIP reveals novel functions for the Microprocessor. Nature structural & molecular biology. 2011;19(8):760-6.

86. Lin YT, Sullivan CS. Expanding the role of Drosha to the regulation of viral gene expression. Proc Natl Acad Sci U S A. 2011;108(27):11229-34.

87. Bohnsack MT, Czaplinski K, Gorlich D. Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. Rna. 2004;10(2):185-91.

88. Lund E, Guttinger S, Calado A, Dahlberg JE, Kutay U. Nuclear export of microRNA precursors. Science. 2004;303(5654):95-8.

89. Okada C, Yamashita E, Lee SJ, Shibata S, Katahira J, Nakagawa A, et al. A high-resolution structure of the pre-microRNA nuclear export machinery. Science. 2009;326(5957):1275-9.

90. Brownawell AM, Macara IG. Exportin-5, a novel karyopherin, mediates nuclear export of double-stranded RNA binding proteins. J Cell Biol. 2002;156(1):53-64.

91. Lu S, Cullen BR. Adenovirus VA1 noncoding RNA can inhibit small interfering RNA and MicroRNA biogenesis. J Virol. 2004;78(23):12868-76.

92. Grimm D, Streetz KL, Jopling CL, Storm TA, Pandey K, Davis CR, et al. Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. Nature. 2006;441(7092):537-41.

93. Bennasser Y, Chable-Bessia C, Triboulet R, Gibbings D, Gwizdek C, Dargemont C, et al. Competition for XPO5 binding between Dicer mRNA, pre-miRNA and viral RNA regulates human Dicer levels. Nat Struct Mol Biol. 2011;18(3):323-7.

94. Singh CP, Singh J, Nagaraju J. A Baculovirus-Encoded MicroRNA (miRNA) Suppresses Its Host miRNA Biogenesis by Regulating the Exportin-5 Cofactor Ran. J Virol. 2012;86(15):7867-79.

95. Gregory RI, Chendrimada TP, Cooch N, Shiekhattar R. Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. Cell. 2005;123(4):631-40.

96. Forstemann K, Tomari Y, Du T, Vagin VV, Denli AM, Bratu DP, et al. Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. PLoS Biol. 2005;3(7):e236.

97. Jiang F, Ye X, Liu X, Fincher L, McKearin D, Liu Q. Dicer-1 and R3D1-L catalyze microRNA maturation in Drosophila. Genes Dev. 2005;19(14):1674-9.

98. Saito K, Ishizuka A, Siomi H, Siomi MC. Processing of pre-microRNAs by the Dicer-1-Loquacious complex in Drosophila cells. PLoS Biol. 2005;3(7):e235.

99. Chendrimada TP, Gregory RI, Kumaraswamy E, Norman J, Cooch N, Nishikura K, et al. TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. Nature. 2005;436(7051):740-4.

100. Haase AD, Jaskiewicz L, Zhang H, Laine S, Sack R, Gatignol A, et al. TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. EMBO Rep. 2005;6(10):961-7.

101. Lee Y, Hur I, Park SY, Kim YK, Suh MR, Kim VN. The role of PACT in the RNA silencing pathway. EMBO J. 2006;25(3):522-32.

102. Khvorova A, Reynolds A, Jayasena SD. Functional siRNAs and miRNAs exhibit strand bias. Cell. 2003;115(2):209-16.

103. Schwarz DS, Hutvagner G, Du T, Xu Z, Aronin N, Zamore PD. Asymmetry in the assembly of the RNAi enzyme complex. Cell. 2003;115(2):199-208.

104. Koscianska E, Starega-Roslan J, Krzyzosiak WJ. The role of Dicer protein partners in the processing of microRNA precursors. PloS one. 2011;6(12):e28548.

105. Paroo Z, Ye X, Chen S, Liu Q. Phosphorylation of the human microRNA-generating complex mediates MAPK/Erk signaling. Cell. 2009;139(1):112-22.

106. Heo I, Joo C, Kim YK, Ha M, Yoon MJ, Cho J, et al. TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. Cell. 2009;138(4):696-708.

107. Newman MA, Mani V, Hammond SM. Deep sequencing of microRNA precursors reveals extensive 3' end modification. Rna. 2011;17(10):1795-803.

108. Heo I, Ha M, Lim J, Yoon MJ, Park JE, Kwon SC, et al. Mono-Uridylation of Pre-MicroRNA as a Key Step in the Biogenesis of Group II let-7 MicroRNAs. Cell. 2012;151(3):521-32.

109. Choudhury NR, Michlewski G. Terminal loop-mediated control of microRNA biogenesis. Biochem Soc Trans. 2012;40(4):789-93.

110. Rau F, Freyermuth F, Fugier C, Villemin JP, Fischer MC, Jost B, et al. Misregulation of miR-1 processing is associated with heart defects in myotonic dystrophy. Nat Struct Mol Biol. 2011;18(7):840-5.

111. Suzuki HI, Arase M, Matsuyama H, Choi YL, Ueno T, Mano H, et al. MCPIP1 ribonuclease antagonizes dicer and terminates microRNA biogenesis through precursor microRNA degradation. Mol Cell. 2011;44(3):424-36.

112. Xhemalce B, Robson SC, Kouzarides T. Human RNA Methyltransferase BCDIN3D Regulates MicroRNA Processing. Cell. 2012;151(2):278-88.

113. Park JE, Heo I, Tian Y, Simanshu DK, Chang H, Jee D, et al. Dicer recognizes the 5' end of RNA for efficient and accurate processing. Nature. 2011;475(7355):201-5.

114. Hellwig S, Bass BL. A starvation-induced noncoding RNA modulates expression of Dicer-regulated genes. Proc Natl Acad Sci U S A. 2008;105(35):12897-902.

115. Andersson MG, Haasnoot PC, Xu N, Berenjian S, Berkhout B, Akusjarvi G. Suppression of RNA interference by adenovirus virus-associated RNA. J Virol. 2005;79(15):9556-65.

116. Grinberg M, Gilad S, Meiri E, Levy A, Isakov O, Ronen R, et al. Vaccinia virus infection suppresses the cell microRNA machinery. Arch Virol. 2012;157(9):1719-27.

117. Iizasa H, Wulff BE, Alla NR, Maragkakis M, Megraw M, Hatzigeorgiou A, et al. Editing of Epstein-Barr virus-encoded BART6 microRNAs controls their dicer targeting and consequently affects viral latency. J Biol Chem. 2010;285(43):33358-70.

118. Forman JJ, Legesse-Miller A, Coller HA. A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence. Proc Natl Acad Sci U S A. 2008;105(39):14879-84.

119. O'Carroll D, Mecklenbrauker I, Das PP, Santana A, Koenig U, Enright AJ, et al. A Slicer-independent role for Argonaute 2 in hematopoiesis and the microRNA pathway. Genes Dev. 2007;21(16):1999-2004.

120. Diederichs S, Haber DA. Dual role for argonautes in microRNA processing and posttranscriptional regulation of microRNA expression. Cell. 2007;131(6):1097-108.

121. Adams BD, Claffey KP, White BA. Argonaute-2 expression is regulated by epidermal growth factor receptor and mitogen-activated protein kinase signaling and correlates with a transformed phenotype in breast cancer cells. Endocrinology. 2009;150(1):14-23.

122. Rudel S, Wang Y, Lenobel R, Korner R, Hsiao HH, Urlaub H, et al. Phosphorylation of human Argonaute proteins affects small RNA binding. Nucleic Acids Res. 2012;39(6):2330-43.

123. Yang JS, Lai EC. Alternative miRNA biogenesis pathways and the interpretation of core miRNA pathway mutants. Mol Cell. 2011;43(6):892-903.

124. Bogerd HP, Karnowski HW, Cai X, Shin J, Pohlers M, Cullen BR. A mammalian herpesvirus uses noncanonical expression and processing mechanisms to generate viral MicroRNAs. Mol Cell. 2010;37(1):135-42.

125. Shapiro JS, Varble A, Pham AM, Tenoever BR. Noncanonical cytoplasmic processing of viral microRNAs. RNA. 2010;16(11):2068-74.

126. Rouha H, Thurner C, Mandl CW. Functional microRNA generated from a cytoplasmic RNA virus. Nucleic Acids Res. 2010;38(22):8328-37.

127. Varble A, Chua MA, Perez JT, Manicassamy B, Garcia-Sastre A, tenOever BR. Engineered RNA viral synthesis of microRNAs. Proc Natl Acad Sci U S A. 2010;107(25):11519-24.

128. Cazalla D, Xie M, Steitz JA. A primate herpesvirus uses the integrator complex to generate viral microRNAs. Mol Cell. 2011;43(6):982-92.

129. Okamura K, Hagen JW, Duan H, Tyler DM, Lai EC. The mirtron pathway generates microRNA-class regulatory RNAs in Drosophila. Cell. 2007;130(1):89-100.

130. Ruby JG, Jan CH, Bartel DP. Intronic microRNA precursors that bypass Drosha processing. Nature. 2007;448(7149):83-6.

131. Babiarz JE, Ruby JG, Wang Y, Bartel DP, Blelloch R. Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. Genes Dev. 2008;22(20):2773-85.

132. Glazov EA, Cottee PA, Barris WC, Moore RJ, Dalrymple BP, Tizard ML. A microRNA catalog of the developing chicken embryo identified by a deep sequencing approach. Genome Res. 2008;18(6):957-64.

133. Flynt AS, Greimann JC, Chung WJ, Lima CD, Lai EC. MicroRNA biogenesis via splicing and exosome-mediated trimming in Drosophila. Mol Cell. 2010;38(6):900-7.

134. Ender C, Krek A, Friedlander MR, Beitzinger M, Weinmann L, Chen W, et al. A human snoRNA with microRNA-like functions. Mol Cell. 2008;32(4):519-28.

135. Saraiya AA, Wang CC. snoRNA, a novel precursor of microRNA in Giardia lamblia. PLoS Pathog. 2008;4(11):e1000224.

136. Cole C, Sobala A, Lu C, Thatcher SR, Bowman A, Brown JW, et al. Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs. Rna. 2009;15(12):2147-60.

137. Cheloufi S, Dos Santos CO, Chong MM, Hannon GJ. A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. Nature. 2010;465(7298):584-9.

138. Cifuentes D, Xue H, Taylor DW, Patnode H, Mishima Y, Cheloufi S, et al. A novel miRNA processing pathway independent of Dicer requires Argonaute2 catalytic activity. Science. 2010;328(5986):1694-8.

139. Yang JS, Maurin T, Robine N, Rasmussen KD, Jeffrey KL, Chandwani R, et al. Conserved vertebrate mir-451 provides a platform for Dicer-independent, Ago2-mediated microRNA biogenesis. Proc Natl Acad Sci U S A. 2010;107(34):15163-8.

140. Yang JS, Maurin T, Lai EC. Functional parameters of Dicer-independent microRNA biogenesis. Rna. 2012;18(5):945-57.

141. Maurin T, Cazalla D, Yang JS, Bortolamiol-Becet D, Lai EC. RNase III-independent microRNA biogenesis in mammalian cells. RNA. 2012.

142. Iwai N, Naraba H. Polymorphisms in human pre-miRNAs. Biochem Biophys Res Commun. 2005;331(4):1439-44.

143. Duan R, Pak C, Jin P. Single nucleotide polymorphism associated with mature miR-125a alters the processing of pri-miRNA. Hum Mol Genet. 2007;16(9):1124-31.

144. Jazdzewski K, Murray EL, Franssila K, Jarzab B, Schoenberg DR, de la Chapelle A. Common SNP in pre-miR-146a decreases mature miR expression and predisposes to papillary thyroid carcinoma. Proc Natl Acad Sci U S A. 2008;105(20):7269-74.

145. Sun G, Yan J, Noltner K, Feng J, Li H, Sarkis DA, et al. SNPs in human miRNA genes affect biogenesis and function. Rna. 2009;15(9):1640-51.

146. Hu Z, Chen J, Tian T, Zhou X, Gu H, Xu L, et al. Genetic variants of miRNA sequences and non-small cell lung cancer survival. J Clin Invest. 2008;118(7):2600-8.

147. Harnprasopwat R, Ha D, Toyoshima T, Lodish H, Tojo A, Kotani A. Alteration of processing induced by a single nucleotide polymorphism in pri-miR-126. Biochem Biophys Res Commun. 2010;399(2):117-22.

148. Nishikura K. Functions and regulation of RNA editing by ADAR deaminases. Annu Rev Biochem. 2010;79:321-49.

149. Luciano DJ, Mirsky H, Vendetti NJ, Maas S. RNA editing of a miRNA precursor. Rna. 2004;10(8):1174-7.

150. Blow MJ, Grocock RJ, van Dongen S, Enright AJ, Dicks E, Futreal PA, et al. RNA editing of human microRNAs. Genome Biol. 2006;7(4):R27.

151. Kawahara Y, Megraw M, Kreider E, Iizasa H, Valente L, Hatzigeorgiou AG, et al. Frequency and fate of microRNA editing in human brain. Nucleic Acids Res. 2008;36(16):5270-80.

152. Kawahara Y, Zinshteyn B, Sethupathy P, Iizasa H, Hatzigeorgiou AG, Nishikura K. Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. Science. 2007;315(5815):1137-40.

153. Kawahara Y, Zinshteyn B, Chendrimada TP, Shiekhattar R, Nishikura K. RNA editing of the microRNA-151 precursor blocks cleavage by the Dicer-TRBP complex. EMBO Rep. 2007;8(8):763-9.

154. Yang W, Chendrimada TP, Wang Q, Higuchi M, Seeburg PH, Shiekhattar R, et al. Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. Nat Struct Mol Biol. 2006;13(1):13-21.

155. Scadden AD. The RISC subunit Tudor-SN binds to hyper-edited double-stranded RNA and promotes its cleavage. Nat Struct Mol Biol. 2005;12(6):489-96.

156. Heale BS, Keegan LP, McGurk L, Michlewski G, Brindle J, Stanton CM, et al. Editing independent effects of ADARs on the miRNA/siRNA pathways. EMBO J. 2009;28(20):3145-56.

157. Yao Y, Zhao Y, Xu H, Smith LP, Lawrie CH, Watson M, et al. MicroRNA profile of Marek's disease virus-transformed T-cell line MSB-1: predominance of virus-encoded microRNAs. J Virol. 2008;82(8):4007-15.

158. Bhattacharyya SN, Habermacher R, Martine U, Closs EI, Filipowicz W. Relief of microRNA-mediated translational repression in human cells subjected to stress. Cell. 2006;125(6):1111-24.

159. van Rooij E, Sutherland LB, Qi X, Richardson JA, Hill J, Olson EN. Control of stress-dependent cardiac growth and gene expression by a microRNA. Science. 2007;316(5824):575-9.

160. Bail S, Swerdel M, Liu H, Jiao X, Goff LA, Hart RP, et al. Differential regulation of microRNA stability. Rna. 2010;16(5):1032-9.

161. Gantier MP, McCoy CE, Rusinova I, Saulep D, Wang D, Xu D, et al. Analysis of microRNA turnover in mammalian cells following Dicer1 ablation. Nucleic Acids Res. 2011;39(13):5692-703.

162. Ruegger S, Grosshans H. MicroRNA turnover: when, how, and why. Trends in biochemical sciences. 2012;37(10):436-46.

163. Krol J, Busskamp V, Markiewicz I, Stadler MB, Ribi S, Richter J, et al. Characterizing light-regulated retinal microRNAs reveals rapid turnover as a common property of neuronal microRNAs. Cell. 2010;141(4):618-31.

164. Sethi P, Lukiw WJ. Micro-RNA abundance and stability in human brain: specific alterations in Alzheimer's disease temporal lobe neocortex. Neurosci Lett. 2009;459(2):100-4.

165. Rajasethupathy P, Fiumara F, Sheridan R, Betel D, Puthanveettil SV, Russo JJ, et al. Characterization of small RNAs in Aplysia reveals a role for miR-124 in constraining synaptic plasticity through CREB. Neuron. 2009;63(6):803-17.

166. Ashraf SI, McLoon AL, Sclarsic SM, Kunes S. Synaptic protein synthesis associated with memory is regulated by the RISC pathway in Drosophila. Cell. 2006;124(1):191-205.

167. Banerjee S, Neveu P, Kosik KS. A coordinated local translational control point at the synapse involving relief from silencing and MOV10 degradation. Neuron. 2009;64(6):871-84.

168. Schratt GM, Tuebing F, Nigh EA, Kane CG, Sabatini ME, Kiebler M, et al. A brain-specific microRNA regulates dendritic spine development. Nature. 2006;439(7074):283-9.

169. Siegel G, Obernosterer G, Fiore R, Oehmen M, Bicker S, Christensen M, et al. A functional screen implicates microRNA-138-dependent regulation of the depalmitoylation enzyme APT1 in dendritic spine morphogenesis. Nat Cell Biol. 2009;11(6):705-16.

170. Yu B, Yang Z, Li J, Minakhina S, Yang M, Padgett RW, et al. Methylation as a crucial step in plant microRNA biogenesis. Science. 2005;307(5711):932-5.

171. Yang Z, Ebright YW, Yu B, Chen X. HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2′ OH of the 3′ terminal nucleotide. Nucleic Acids Res. 2006;34(2):667-75.

172. Huang Y, Ji L, Huang Q, Vassylyev DG, Chen X, Ma JB. Structural insights into mechanisms of the small RNA methyltransferase HEN1. Nature. 2009;461(7265):823-7.

173. Li J, Yang Z, Yu B, Liu J, Chen X. Methylation protects miRNAs and siRNAs from a 3′-end uridylation activity in Arabidopsis. Curr Biol. 2005;15(16):1501-7.

174. Ren G, Chen X, Yu B. Uridylation of miRNAs by hen1 suppressor1 in Arabidopsis. Curr Biol. 2012;22(8):695-700.

175. Zhao Y, Yu Y, Zhai J, Ramachandran V, Dinh TT, Meyers BC, et al. The Arabidopsis nucleotidyl transferase HESO1 uridylates unmethylated small RNAs to trigger their degradation. Curr Biol. 2012;22(8):689-94.

176. Kamminga LM, Luteijn MJ, den Broeder MJ, Redl S, Kaaij LJ, Roovers EF, et al. Hen1 is required for oocyte development and piRNA stability in zebrafish. EMBO J. 2010;29(21):3688-700.

177. van Wolfswinkel JC, Claycomb JM, Batista PJ, Mello CC, Berezikov E, Ketting RF. CDE-1 affects chromosome segregation through uridylation of CSR-1-bound siRNAs. Cell. 2009;139(1):135-48.

178. Schmidt MJ, West S, Norbury CJ. The human cytoplasmic RNA terminal U-transferase ZCCHC11 targets histone mRNAs for degradation. Rna. 2011;17(1):39-44.

179. Ibrahim F, Rymarquis LA, Kim EJ, Becker J, Balassa E, Green PJ, et al. Uridylation of mature miRNAs and siRNAs by the MUT68 nucleotidyltransferase promotes their degradation in Chlamydomonas. Proc Natl Acad Sci U S A. 2010;107(8):3906-11.

180. Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, et al. A mammalian microRNA expression atlas based on small RNA library sequencing. Cell. 2007;129(7):1401-14.

181. Chiang HR, Schoenfeld LW, Ruby JG, Auyeung VC, Spies N, Baek D, et al. Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. Genes Dev. 2010;24(10):992-1009.

182. Westholm JO, Ladewig E, Okamura K, Robine N, Lai EC. Common and distinct patterns of terminal modifications to mirtrons and canonical microRNAs. Rna. 2012;18(2):177-92.

183. Burroughs AM, Ando Y, de Hoon MJ, Tomaru Y, Nishibu T, Ukekawa R, et al. A comprehensive survey of 3′ animal miRNA modification events and a possible role for 3′ adenylation in modulating miRNA targeting effectiveness. Genome Res. 2010;20(10):1398-410.

184. Wyman SK, Knouf EC, Parkin RK, Fritz BR, Lin DW, Dennis LM, et al. Post-transcriptional generation of miRNA variants by multiple nucleotidyl transferases contributes to miRNA transcriptome complexity. Genome Res. 2011;21(9):1450-61.

185. Jones MR, Quinton LJ, Blahna MT, Neilson JR, Fu S, Ivanov AR, et al. Zcchc11-dependent uridylation of microRNA directs cytokine expression. Nat Cell Biol. 2009;11(9):1157-63.

186. Katoh T, Sakaguchi Y, Miyauchi K, Suzuki T, Kashiwabara S, Baba T. Selective stabilization of mammalian microRNAs by 3′ adenylation mediated by the cytoplasmic poly(A) polymerase GLD-2. Genes Dev. 2009;23(4):433-8.

187. Lu S, Sun YH, Chiang VL. Adenylation of plant miRNAs. Nucleic Acids Res. 2009;37(6):1878-85.

188. Burns DM, D'Ambrogio A, Nottrott S, Richter JD. CPEB and two poly(A) polymerases control miR-122 stability and p53 mRNA translation. Nature. 2011;473(7345):105-8.

189. Backes S, Shapiro JS, Sabin LR, Pham AM, Reyes I, Moss B, et al. Degradation of Host MicroRNAs by Poxvirus Poly(A) Polymerase Reveals Terminal RNA Methylation as a Protective Antiviral Mechanism. Cell Host Microbe. 2012;12(2):200-10.

190. Cerutti H, Ibrahim F. Turnover of Mature miRNAs and siRNAs in Plants and Algae. Adv Exp Med Biol. 2011;700:124-39.

191. Rissland OS, Hong SJ, Bartel DP. MicroRNA destabilization enables dynamic regulation of the miR-16 family in response to cell-cycle changes. Mol Cell. 2011;43(6):993-1004.

192. Zhang Z, Zou J, Wang GK, Zhang JT, Huang S, Qin YW, et al. Uracils at nucleotide position 9-11 are required for the rapid turnover of miR-29 family. Nucleic Acids Res. 2011;39(10):4387-95.

193. Ramachandran V, Chen X. Degradation of microRNAs by a family of exoribonucleases in Arabidopsis. Science. 2008;321(5895):1490-2.

194. Kai ZS, Pasquinelli AE. MicroRNA assassins: factors that regulate the disappearance of miRNAs. Nature structural & molecular biology. 2010;17(1):5-10.

195. Gy I, Gasciolli V, Lauressergues D, Morel JB, Gombert J, Proux F, et al. Arabidopsis FIERY1, XRN2, and XRN3 are endogenous RNA silencing suppressors. Plant Cell. 2007;19(11):3451-61.

196. Ballarino M, Pagano F, Girardi E, Morlando M, Cacchiarelli D, Marchioni M, et al. Coupled RNA processing and transcription of intergenic primary microRNAs. Molecular and cellular biology. 2009;29(20):5632-8.

197. Chatterjee S, Grosshans H. Active turnover modulates mature microRNA activity in Caenorhabditis elegans. Nature. 2009;461(7263):546-9.

198. Chatterjee S, Fasler M, Bussing I, Grosshans H. Target-mediated protection of endogenous microRNAs in C. elegans. Dev Cell. 2011;20(3):388-96.

199. Das SK, Sokhi UK, Bhutia SK, Azab B, Su ZZ, Sarkar D, et al. Human polynucleotide phosphorylase selectively and preferentially degrades microRNA-221 in human melanoma cells. Proc Natl Acad Sci U S A. 2010;107(26):11948-53.

200. Thomas MF, Abdul-Wajid S, Panduro M, Babiarz JE, Rajaram M, Woodruff P, et al. Eri1 regulates microRNA homeostasis and mouse lymphocyte development and antiviral function. Blood. 2012;120(1):130-42.

201. Chen AJ, Paik JH, Zhang H, Shukla SA, Mortensen R, Hu J, et al. STAR RNA-binding protein Quaking suppresses cancer via stabilization of specific miRNA. Genes Dev. 2012;26(13):1459-72.

202. Tomari Y, Du T, Zamore PD. Sorting of Drosophila small silencing RNAs. Cell. 2007;130(2):299-308.

203. Horwich MD, Li C, Matranga C, Vagin V, Farley G, Wang P, et al. The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. Curr Biol. 2007;17(14):1265-72.

204. Ameres SL, Horwich MD, Hung JH, Xu J, Ghildiyal M, Weng Z, et al. Target RNA-directed trimming and tailing of small silencing RNAs. Science. 2010;328(5985):1534-9.

205. Baccarini A, Chauhan H, Gardner TJ, Jayaprakash AD, Sachidanandam R, Brown BD. Kinetic analysis reveals the fate of a microRNA following target regulation in mammalian cells. Curr Biol. 2011;21(5):369-76.

206. Cazalla D, Yario T, Steitz JA. Down-regulation of a host microRNA by a Herpesvirus saimiri noncoding RNA. Science. 2010;328(5985):1563-6.

207. Buck AH, Perot J, Chisholm MA, Kumar DS, Tuddenham L, Cognat V, et al. Post-transcriptional regulation of miR-27 in murine cytomegalovirus infection. Rna. 2010;16(2):307-15.

208. Libri V, Helwak A, Miesen P, Santhakumar D, Borger JG, Kudla G, et al. Murine cytomegalovirus encodes a miR-27 inhibitor disguised as a target. Proc Natl Acad Sci U S A. 2012;109(1):279-84.

209. Marcinowski L, Tanguy M, Krmpotic A, Radle B, Lisnic VJ, Tuddenham L, et al. Degradation of cellular mir-27 by a novel, highly abundant viral transcript is important for efficient virus replication in vivo. PLoS Pathog. 2012;8(2):e1002510.

210. Pasquinelli AE. MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. Nature reviews Genetics. 2012;13(4):271-82.

211. Ding SW, Voinnet O. Antiviral immunity directed by small RNAs. Cell. 2007;130(3):413-26.

212. Haasnoot J, Westerhout EM, Berkhout B. RNA interference against viruses: strike and counterstrike. Nat Biotechnol. 2007;25(12):1435-43.

213. Umbach JL, Cullen BR. The role of RNAi and microRNAs in animal virus replication and antiviral immunity. Genes Dev. 2009;23(10):1151-64.

214. van Rij RP, Saleh MC, Berry B, Foo C, Houk A, Antoniewski C, et al. The RNA silencing endonuclease Argonaute 2 mediates specific antiviral immunity in Drosophila melanogaster. Genes Dev. 2006;20(21):2985-95.

215. Okamura K, Ishizuka A, Siomi H, Siomi MC. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. Genes Dev. 2004;18(14):1655-66.

216. Li WX, Li H, Lu R, Li F, Dus M, Atkinson P, et al. Interferon antagonist proteins of influenza and vaccinia viruses are suppressors of RNA silencing. Proc Natl Acad Sci U S A. 2004;101(5):1350-5.

217. Li H, Li WX, Ding SW. Induction and suppression of RNA silencing by an animal virus. Science. 2002;296(5571):1319-21.

218. Lu R, Maduro M, Li F, Li HW, Broitman-Maduro G, Li WX, et al. Animal virus replication and RNAi-mediated antiviral silencing in Caenorhabditis elegans. Nature. 2005;436(7053):1040-3.

219. Chao JA, Lee JH, Chapados BR, Debler EW, Schneemann A, Williamson JR. Dual modes of RNA-silencing suppression by Flock House virus protein B2. Nat Struct Mol Biol. 2005;12(11):952-7.

220. Aliyari R, Wu Q, Li HW, Wang XH, Li F, Green LD, et al. Mechanism of induction and suppression of antiviral immunity directed by virus-derived small RNAs in Drosophila. Cell Host Microbe. 2008;4(4):387-97.

221. Nayak A, Berry B, Tassetto M, Kunitomi M, Acevedo A, Deng C, et al. Cricket paralysis virus antagonizes Argonaute 2 to modulate antiviral defense in Drosophila. Nat Struct Mol Biol. 2010;17(5):547-54.

222. van Mierlo JT, Bronkhorst AW, Overheul GJ, Sadanandan SA, Ekstrom JO, Heestermans M, et al. Convergent evolution of argonaute-2 slicer antagonism in two distinct insect RNA viruses. PLoS Pathog. 2012;8(8):e1002872.

223. Berry B, Deddouche S, Kirschner D, Imler JL, Antoniewski C. Viral suppressors of RNA silencing hinder exogenous and endogenous small RNA pathways in Drosophila. PloS one. 2009;4(6):e5866.

224. Chapman EJ, Prokhnevsky AI, Gopinath K, Dolja VV, Carrington JC. Viral RNA silencing suppressors inhibit the microRNA pathway at an intermediate step. Genes Dev. 2004;18(10):1179-86.

225. Dunoyer P, Lecellier CH, Parizotto EA, Himber C, Voinnet O. Probing the microRNA and small interfering RNA pathways with virus-encoded suppressors of RNA silencing. Plant Cell. 2004;16(5):1235-50.

226. Kasschau KD, Xie Z, Allen E, Llave C, Chapman EJ, Krizan KA, et al. P1/HC-Pro, a viral suppressor of RNA silencing, interferes with Arabidopsis development and miRNA unction. Dev Cell. 2003;4(2):205-17.

227. Voinnet O. Origin, biogenesis, and activity of plant microRNAs. Cell. 2009;136(4):669-87.

228. Morel JB, Godon C, Mourrain P, Beclin C, Boutet S, Feuerbach F, et al. Fertile hypomorphic ARGONAUTE (ago1) mutants impaired in post-transcriptional gene silencing and virus resistance. Plant Cell. 2002;14(3):629-39.

229. Zhang X, Yuan YR, Pei Y, Lin SS, Tuschl T, Patel DJ, et al. Cucumber mosaic virus-encoded 2b suppressor inhibits Arabidopsis Argonaute1 cleavage activity to counter plant defense. Genes Dev. 2006;20(23):3255-68.

230. Lakatos L, Szittya G, Silhavy D, Burgyan J. Molecular mechanism of RNA silencing suppression mediated by p19 protein of tombusviruses. EMBO J. 2004;23(4):876-84.

231. Merai Z, Kerenyi Z, Kertesz S, Magna M, Lakatos L, Silhavy D. Double-stranded RNA binding may be a general plant RNA viral strategy to suppress RNA silencing. J Virol. 2006;80(12):5747-56.

232. Merai Z, Kerenyi Z, Molnar A, Barta E, Valoczi A, Bisztray G, et al. Aureusvirus P14 is an efficient RNA silencing suppressor that binds double-stranded RNAs without size specificity. J Virol. 2005;79(11):7217-26.

233. Lakatos L, Csorba T, Pantaleo V, Chapman EJ, Carrington JC, Liu YP, et al. Small RNA binding is a common strategy to suppress RNA silencing by several viral suppressors. EMBO J. 2006;25(12):2768-80.

234. Csorba T, Bovi A, Dalmay T, Burgyan J. The p122 subunit of Tobacco Mosaic Virus replicase is a potent silencing suppressor and compromises both small interfering RNA- and microRNA-mediated pathways. J Virol. 2007;81(21):11768-80.

235. Hemmes H, Lakatos L, Goldbach R, Burgyan J, Prins M. The NS3 protein of Rice hoja blanca tenuivirus suppresses RNA silencing in plant and insect hosts by efficiently binding both siRNAs and miRNAs. Rna. 2007;13(7):1079-89.

236. Vargason JM, Szittya G, Burgyan J, Hall TM. Size selective recognition of siRNA by an RNA silencing suppressor. Cell. 2003;115(7):799-811.

237. Ye K, Malinina L, Patel DJ. Recognition of small interfering RNA by a viral suppressor of RNA silencing. Nature. 2003;426(6968):874-8.

238. Schott G, Mari-Ordonez A, Himber C, Alioua A, Voinnet O, Dunoyer P. Differential effects of viral silencing suppressors on siRNA and miRNA loading support the existence of two distinct cellular pools of ARGONAUTE1. EMBO J. 2012;31(11):2553-65.

239. Giner A, Lakatos L, Garcia-Chapa M, Lopez-Moya JJ, Burgyan J. Viral protein inhibits RISC activity by argonaute binding through conserved WG/GW motifs. PLoS Pathog. 2010;6(7):e1000996.

240. Azevedo J, Garcia D, Pontier D, Ohnesorge S, Yu A, Garcia S, et al. Argonaute quenching and global changes in Dicer homeostasis caused by a pathogen-encoded GW repeat protein. Genes Dev. 2010;24(9):904-15.

241. Baumberger N, Tsai CH, Lie M, Havecker E, Baulcombe DC. The Polerovirus silencing suppressor P0 targets ARGONAUTE proteins for degradation. Curr Biol. 2007;17(18):1609-14.

242. Bortolamiol D, Pazhouhandeh M, Marrocco K, Genschik P, Ziegler-Graff V. The Polerovirus F box protein P0 targets ARGONAUTE1 to suppress RNA silencing. Curr Biol. 2007;17(18):1615-21.

243. Csorba T, Lozsa R, Hutvagner G, Burgyan J. Polerovirus protein P0 prevents the assembly of small RNA-containing RISC complexes and leads to degradation of ARGONAUTE1. Plant J. 2010;62(3):463-72.

244. Pazhouhandeh M, Dieterle M, Marrocco K, Lechner E, Berry B, Brault V, et al. F-box-like domain in the polerovirus protein P0 is required for silencing suppressor function. Proc Natl Acad Sci U S A. 2006;103(6):1994-9.

245. Ho MS, Ou C, Chan YR, Chien CT, Pi H. The utility F-box for protein destruction. Cell Mol Life Sci. 2008;65(13):1977-2000.

246. Parameswaran P, Sklan E, Wilkins C, Burgon T, Samuel MA, Lu R, et al. Six RNA viruses and forty-one hosts: viral small RNAs and modulation of small RNA repertoires in vertebrate and invertebrate systems. PLoS Pathog. 2010;6(2):e1000764.

247. van Rij RP, Andino R. The silent treatment: RNAi as a defense against virus infection in mammals. Trends Biotechnol. 2006;24(4):186-93.

248. Shah PS, Schaffer DV. Antiviral RNAi: translating science towards therapeutic success. Pharm Res. 2011;28(12):2966-82.

249. Sullivan CS, Ganem D. A virus-encoded inhibitor that blocks RNA interference in mammalian cells. J Virol. 2005;79(12):7371-9.

250. Soldan SS, Plassmeyer ML, Matukonis MK, Gonzalez-Scarano F. La Crosse virus nonstructural protein NSs counteracts the effects of short interfering RNA. J Virol. 2005;79(1):234-44.

251. Bennasser Y, Le SY, Benkirane M, Jeang KT. Evidence that HIV-1 encodes an siRNA and a suppressor of RNA silencing. Immunity. 2005;22(5):607-19.

252. Fabozzi G, Nabel CS, Dolan MA, Sullivan NJ. Ebolavirus proteins suppress the effects of small interfering RNA by direct interaction with the mammalian RNA interference pathway. J Virol. 2011;85(6):2512-23.

253. Haasnoot J, de Vries W, Geutjes EJ, Prins M, de Haan P, Berkhout B. The Ebola virus VP35 protein is a suppressor of RNA silencing. PLoS Pathog. 2007;3(6):e86.

254. Bucher E, Hemmes H, de Haan P, Goldbach R, Prins M. The influenza A virus NS1 protein binds small interfering RNAs and suppresses RNA silencing in plants. J Gen Virol. 2004;85(Pt 4):983-91.

255. Delgadillo MO, Saenz P, Salvador B, Garcia JA, Simon-Mateo C. Human influenza virus NS1 protein enhances viral pathogenicity and acts as an RNA silencing suppressor in plants. J Gen Virol. 2004;85(Pt 4):993-9.

256. Kok KH, Jin DY. Influenza A virus NS1 protein does not suppress RNA interference in mammalian cells. J Gen Virol. 2006;87(Pt 9):2639-44.

257. Bennasser Y, Jeang KT. HIV-1 Tat interaction with Dicer: requirement for RNA. Retrovirology. 2006;3:95.

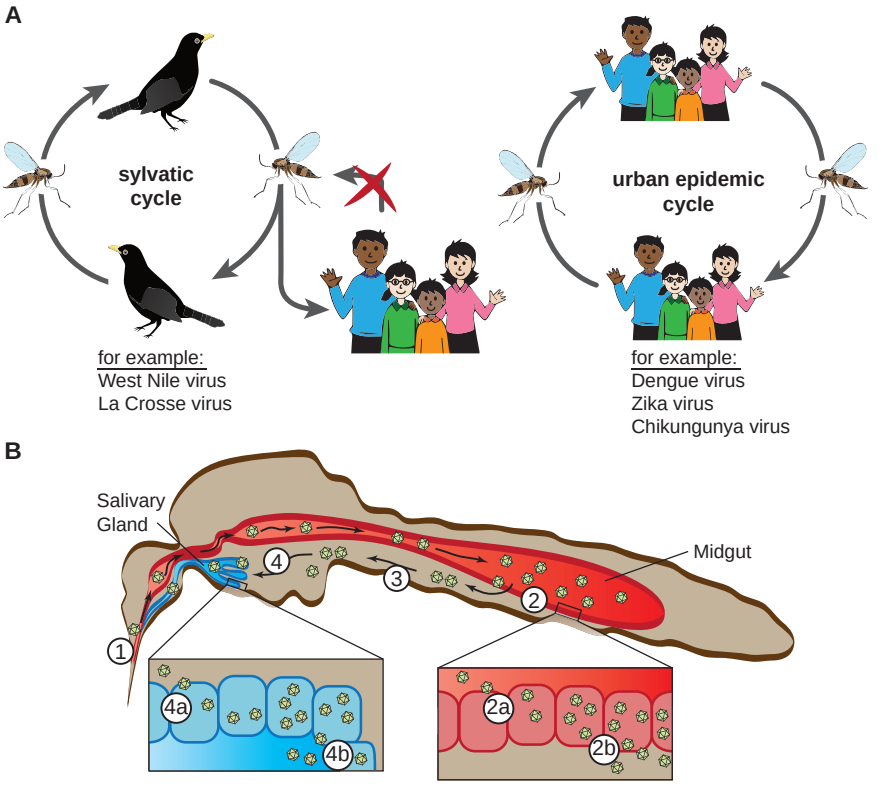
258. Lecellier CH, Dunoyer P, Arar K, Lehmann-Che J, Eyquem S, Himber C, et al. A cellular microRNA mediates antiviral defense in human cells. Science. 2005;308(5721):557-60.

259. Mahajan VS, Drake A, Chen J. Virus-specific host miRNAs: antiviral defenses or promoters of persistent infection? Trends in immunology. 2009;30(1):1-7.

260. Qian S, Zhong X, Yu L, Ding B, de Haan P, Boris-Lawrie K. HIV-1 Tat RNA silencing suppressor activity is conserved across kingdoms and counteracts translational repression of HIV-1. Proc Natl Acad Sci U S A. 2009;106(2):605-10.

261. Lin J, Cullen BR. Analysis of the interaction of primate retroviruses with the human RNA interference machinery. J Virol. 2007;81(22):12218-26.

Appendix



salivary gland: speekselklier, Speicheldrüse
midgut: middendarm, Mitteldarm

Figure 1. Arbovirus transmission by mosquitoes.
Afbeelding 1. Overdracht van arbovirussen door muggen.
Abbildung 1. Übertragung von Arboviren durch Stechmücken.

SUMMARY

Blood-feeding insects, in particular mosquitoes are important vectors for numerous viruses especially in the tropical and subtropical areas of Africa, Latin America and Southeast Asia. Well-known examples of mosquito-transmitted viral diseases include dengue fever, yellow fever, chikungunya fever or Zika fever. In general, viruses that are transmitted by mosquitoes are referred to as arboviruses, an acronym for arthropod-borne viruses. Infections with these viruses can cause febrile illnesses, which last for few days but sometimes a more serious disease course can occur, especially in young children, elderly or individuals with a weak immune system. Dependent on the virus, severe symptoms may involve shock symptoms, hemorrhagic fever, seizures, or encephalitis. In rare cases the disease is fatal. In the past years, a number of large arbovirus epidemics occurred, such as the 2013-2014 chikungunya and the 2015-2016 Zika outbreaks in Latin America with hundred thousands of estimated cases. The most important human arbovirus,

dengue virus, is present in virtually all areas with tropical and subtropical climates and about half of the human population lives at risk of being infected with dengue. For most arboviral diseases, specific treatments or prophylactic vaccines are lacking.

Arbovirus transmission occurs either in a sylvatic (forest dwelling) or urban cycle. In the sylvatic cycle, arboviruses are transmitted primarily between mosquitoes and wild animals. Occasional infections of humans may lead to the onset of a disease, but virus uptake and transmission from the infected individual by a naïve mosquito is not possible. In contrast, in the urban cycle mosquitoes can directly transmit arboviruses from human to human which increases the risk of large epidemics (Figure 1A). Arboviruses in an urban, epidemic cycle are transmitted by mosquitoes of the *Aedes* genus, in particular *Aedes aegypti* (Yellow fever mosquito) and *Aedes albopictus* (Asian tiger mosquito). They have adapted to city habitats and prefer to feed on humans. Because these mosquito species are an essential component of the viral life cycle they represent a suitable target for the development of a strategy to reduce arbovirus transmission. Yet, to achieve this, more insights into the factors that determine virus transmission by mosquitoes are needed.

For effective human-to-human transmission, arboviruses that have been taken up in a blood meal (step 1 in Figure 1B) need to infect the epithelial cells of the mosquito midgut (step 2a) prior to complete inactivation of virus particles in the digestive tract. Subsequently, viral particles egress these cells at the baso-lateral side (step 2b in Figure 1B) and disseminate to secondary tissues (step 3). Finally, the virus infects the epithelial cells of the salivary gland (step 4a) and replicates in the cells that line the salivary duct. When a sufficiently high number of virus particles has been shed into the saliva (step 4b) the mosquito is able to transmit the virus to non-infected hosts with every subsequent bite.

Besides these anatomical barriers, arboviruses need to overcome the immune responses that are raised in the mosquito. The most important branch of the antiviral immune system of insects is the so-called RNA-interference system. In this pathway, small RNA molecules, which are called small interfering RNAs (siRNAs), are produced from replicating viral RNA. Eventually, these siRNAs are responsible for the destruction of the viral genetic material and therefore suppress arbovirus replication. Very much to our surprise we and others found that in *Aedes* mosquitoes, besides siRNAs, a second class of small RNAs, called PIWI-interacting RNAs (piRNAs) is produced from viral RNA. These findings suggested that two independent small RNA pathways contribute to antiviral immunity in mosquitoes.

The insect piRNA pathway is studied most extensively in the fruit fly Drosophila melanogaster in which it is primarily responsible for the defense against transposons in the female germline. Transposons are genetic elements that randomly integrate into the genome of their host species with detrimental consequences for the integrity of the

genetic material. In *Drosophila*, transposon-derived piRNAs associated to PIWI proteins suppress transposon mobilization and therefore protect the genome from potentially harmful mutations.

Interestingly, in *Aedes* mosquitoes the piRNA pathway seems to have gained additional functions beyond repression of transposons in germ cells. The PIWI protein family is greatly expanded in these mosquitoes and some of its members are expressed outside of the germline in somatic tissue. In addition, the repertoire of the associated piRNAs is broader than in fruit flies and includes viral piRNAs, which have not been observed in any other model organism before. However, the molecular determinants of piRNA biogenesis and function in *Aedes* mosquitoes are almost completely unknown.

The increased complexity of the *Aedes* piRNA pathway makes it difficult, if not impossible, to directly translate findings from other model organisms to the mosquito. In this doctoral thesis, we have therefore used *Aedes aegypti* Aag2 cells to biochemically and genetically dissect the molecular mechanisms that underlie the production of viral and non-viral piRNAs.

The second and third chapters focus on the production of piRNAs from viral RNA: In **chapter 2**, we studied the biogenesis of arboviral piRNAs from Sindbis virus. We knocked down individual PIWI proteins in Aag2 cells and using small RNA northern blotting and deep-sequencing we found that Piwi5 and Ago3 depletion strongly reduced viral piRNA levels. By immunoprecipitation of PIWI proteins followed by small RNA deep sequencing we could verify that these two proteins directly bind viral piRNAs. Interestingly, canonical transposon-derived piRNAs were dependent on all somatic PIWI proteins Piwi4, Piwi5, Piwi6 and Ago3, indicating that distinct PIWI proteins have specialized in recognizing and processing piRNAs from different sources. In **chapter 3**, this finding was further supported by our observation that piRNA production from another arbovirus, dengue virus, was dependent on Piwi5 and Ago3 and in addition, Piwi6, although to a lesser extent. Using small RNA deep-sequencing of uninfected and dengue-infected mosquito cells, we also found that in Aag2 cells, dengue infection did not largely influence the expression profile of microRNAs, another class of small RNAs. Yet, mining our deep sequencing data, we could identify novel host microRNAs thereby complementing the repertoire of these regulatory small RNAs in *Aedes aegypti.*

The chapters four to six focus on the characterization of piRNAs from various endogenous sources encoded in the *Aedes aegypti* genome. In **chapter 4**, we analyzed the production of piRNAs from protein-coding genes and we identified different classes of genes for which piRNA biogenesis was dependent on distinct sets of PIWI proteins. Amongst the group that, like viral piRNAs, relied on Ago3 and Piwi5 we identified the replication-dependent histones and showed that their piRNAs accumulated dynamically during the cell cycle. Interestingly, amongst the piRNA-producing genes were several viral sequence elements that in the course of evolution had integrated into the mosquito

genome. In **chapter 5**, a broad comparative-genomics analysis of these so-called endogenous viral elements revealed that these were particularly abundant in *Aedes aegypti* and *Aedes albopictus* and that they gave rise to piRNAs in these mosquitoes. In **chapter 6** our analysis of non-viral piRNAs unveiled a rather unexpected source of piRNAs in the *Aedes aegypti* genome: a satellite DNA locus that we named *satDNA1*. Satellite DNA is special type of repetitive DNA elements, consisting of directly repeated DNA sequences of very few to a couple of hundred nucleotides in length. These elements evolve quite rapidly in the course of evolution and are barely conserved even between closely related species. *satDNA1* however was outstanding since it was conserved for approximately 180-200 million years. Two highly abundant piRNAs were derived from this locus and they exclusively associated with Piwi4. Strikingly, a *satDNA1* piRNA was capable of silencing the expression of a reporter gene indicating that it has the capacity to regulate the expression of mosquito genes.

Collectively, our results indicated that in *Aedes* mosquitoes different PIWI-protein – piRNA complexes are generated which serve different roles in the production of viral and non-viral piRNA. We hypothesized that, besides PIWI proteins, additional protein families were involved in the assembly, stabilization, and function of these complexes. Prime candidates were TUDOR-domain containing proteins, which are known PIWI-protein interaction partners in other model organisms. Indeed, in **chapter 7**, using a knockdown screen, we identified the Tudor protein AAEL012441 to be required for piRNA biogenesis from Sindbis virus. Immunoprecipitation of AAEL012441 followed by mass-spectrometry identified additional proteins that, in cooperation with Piwi5 and Ago3, form a macro-molecular protein complex responsible for the production of viral piRNAs.

In **chapter 8**, the findings of this doctoral thesis are discussed in the broader context of the current literature and suggestions for future studies are made. In **chapter 9**, the role of microRNAs in virus-host interactions is reviewed.

In summary, this doctoral thesis provides important fundamental knowledge about the mechanisms of piRNA production from viral and non-viral sources in *Aedes aegypti* mosquitoes. It establishes mosquito cell culture as a powerful tool to delineate somatic piRNA production and identifies interesting, novel classes of small RNAs that expand the ever-growing repertoire of regulatory RNAs. Based on these results future studies can investigate the role of viral and *Aedes* genome-derived piRNAs in the mosquito immune system and determine their impact on the transmission of arboviruses.

NEDERLANDSE SAMENVATTING

Insecten die zich met bloed voeden, steekmuggen in het bijzonder, zijn belangrijke overdragers van talrijke virusziekten met name in tropische en subtropische gebieden van Afrika, Latijns Amerika en Zuidoost Azië. Bekende voorbeelden van dergelijke ziekten zijn dengue ofwel knokkelkoorts, gele koorts, chikungunya koorts en Zika koorts. Door muggen overgedragen virussen worden arbovirussen genoemd, gebaseerd op de Engelse term *arthropod-borne,* wat 'overgedragen door geleedpotigen' betekent. Infecties met arbovirussen kunnen leiden tot koortsachtige ziektes die meestal na enkele dagen afzwakt. In sommige gevallen, vooral bij jonge kinderen, ouderen of mensen met een verzwakt immuunsysteem kan de ziekte een ernstiger verloop hebben. Hierbij kan er, afhankelijk van het virus, sprake zijn van shocksymptomen, hemorragische (=gepaard gaande met bloedingen) koorts, krampaanvallen of hersenontsteking. Een dergelijke infectie is echter zelden dodelijk. In de afgelopen jaren is er een aantal grote arbovirus epidemieën geweest, waaronder de chikungunya uitbraak in 2013-2014 en de Zika uitbraak in 2015-2016 in Zuid-Amerika, met honderdduizenden geschatte ziektegevallen. Het belangrijkste humane arbovirus, dengue, komt voor in bijna alle streken met subtropisch en tropisch klimaat en ruim de helft van de wereldbevolking leeft tegenwoordig in een risicogebied voor dengue infectie. Voor de meeste arbovirusziektes zijn geen specifieke medicijnen of profylactische vaccins beschikbaar.

Bij de overdracht van arbovirussen maakt men onderscheid tussen een sylvatische (in de bos voorkomend) en een stedelijke cyclus. In de sylvatische cyclus dragen steekmuggen het arbovirus voornamelijk over tussen wilde dieren. Mensen die door een geïnfecteerde mug worden gestoken kunnen weliswaar ziekteverschijnselen ontwikkelen, maar verdere overdracht van het virus naar een niet-besmette mug is niet mogelijk. In de stedelijke cyclus daarentegen worden arbovirussen door muggen rechtstreeks van mens naar mens overgedragen en dit gaat gepaard met een verhoogd risico op grote epidemieën (Afbeelding 1A op pagina 296). Arbovirussen in een stedelijke, epidemische cyclus worden overgedragen door muggen van het genus *Aedes*, met name *Aedes aegypti* (denguemug) en *Aedes albopictus* (tijgermug). Deze soorten zijn goed aangepast aan het leven in de stad en hebben een voorkeur om zich op mensen te voeden. Omdat deze muggensoorten een essentieel onderdeel zijn van de levenscyclus van arbovirussen, zijn zij een veelbelovend doelwit voor de ontwikkeling van nieuwe strategieën om virustransmissie te voorkomen. Hiervoor is het echter noodzakelijk om goed te begrijpen welke factoren invloed hebben op de overdracht van arbovirussen door muggen.

Voor een efficiënte overdracht van mens naar mens moeten arbovirussen, nadat ze door de mug zijn opgenomen in een bloedmaaltijd (Stap 1 in Afbeelding 1B op pagina 296), de cellen van het epitheel weefsel van de middendarm infecteren (Stap 2a) voordat alle virusdeeltjes zijn geïnactiveerd door het verteringsproces. Vervolgens verlaten de

virusdeeltjes deze cellen aan de baso-laterale (=van de middendarm afgekeerde) kant (Stap 2b) om zich naar andere organen te verspreiden (Stap 3). Uiteindelijk infecteren de virussen het epitheelweefsel van de speekselklier (Stap 4a) om zich te vermeerderen in de cellen langs het speekselkanaal. Wanneer er een groot aantal virusdeeltjes in het speeksel van de mug aanwezig is (Stap 4b), kan de mug met elke steek virusdeeltjes overbrengen naar een niet geïnfecteerd individu.

Naast deze anatomische barrières, moeten arbovirussen ook de immuunrespons van de mug zien te overleven. Het belangrijkste onderdeel van het immuunsysteem van insecten is RNA-interferentie. In dit proces wordt replicerend viraal RNA omgezet in kleine RNA moleculen, die men 'small interfering RNAs (siRNAs)' noemt. Deze siRNAs zijn verantwoordelijk voor de afbraak van het genetische materiaal van virussen, waardoor de vermenigvuldiging van een arbovirus geremd wordt. Tot onze grote verassing hebben wij en anderen ontdekt, dat er naast siRNAs in *Aedes* muggen nog een tweede soort kleine virale RNAs gemaakt wordt, die 'PIWI-interacterende RNAs (piRNAs)' heten. Deze bevindingen suggereren dat er twee onafhankelijke kleine RNA systemen bijdragen aan de antivirale immuunafweer in muggen.

Het piRNA systeem is voornamelijk bestudeerd in de fruitvlieg *Drosophila melanogaster*, waar het een belangrijke functie heeft in de bescherming tegen zogenaamde transposons in de vrouwelijke kiembaan. Transposons zijn genetische elementen die zich willekeurig in het genoom van hun gasteer kunnen verplaatsen en daardoor de integriteit van het erfelijk materiaal kunnen aantasten. In de vlieg onderdrukken piRNAs de mobiliteit van deze transposons en beschermen zodoende het genoom tegen potentieel schadelijke mutaties.

In *Aedes* muggen vervult het piRNA mechanisme extra functies naast het onderdrukken van transposons in de geslachtscellen. Deze muggensoorten hebben meerdere PIWI genen waarvan sommigen buiten de kiembaan in somatische cellen tot expressie komen. Bovendien is het repertoire aan piRNAs uitgebreider dan in de fruitvlieg en omvat het onder andere virale piRNAs die tot nog toe in geen ander model organisme zijn beschreven. De moleculaire mechanismes die ten grondslag liggen aan de productie en functie van piRNAs in *Aedes* muggen zijn echter vrijwel onbekend.

De verhoogde complexiteit van het piRNA systeem maakt het lastig, zo niet onmogelijk, om bevindingen die in andere model organismen zijn gedaan rechtstreeks te vertalen naar de mug. Daarom is er in dit proefschrift gebruik gemaakt van *Aedes aegypti* Aag2 cellen om de moleculaire mechanismes van virale en niet-virale piRNA productie to ontrafelen.

In het tweede en derde hoofdstuk ligt de focus op de biogenese van arbovirale piRNAs: In **hoofdstuk 2**, hebben wij de productie van piRNAs bestudeerd die gemaakt werden van Sindbisvirus RNA. Door middel van knockdown van individuele PIWI eiwitten en met behulp van northern blot en deep-sequencing analyses hebben we kunnen aantonen

dat aantallen virale piRNAs sterk omlaag gingen na depletie van Piwi5 en Ago3. Met behulp van immuno-precipitatie van PIWI eiwitten gevolgd van deep-sequencing hebben wij kunnen bevestigen dat virale piRNAs direct gebonden zijn aan deze PIWI eiwitten. De productie van transposon piRNAs was afhankelijk van alle somatische PIWI eiwitten, Piwi4, Piwi5, Piwi6 en Ago3. Dit suggereert dat de verschillende PIWI eiwitten gespecialiseerd zijn in het herkennen en verwerken van diverse piRNA substraten. In **hoofdstuk 3** werd deze bevinding verder ondersteund door onze observatie dat piRNA productie van een ander arbovirus, denguevirus, naast Piwi5 en Ago3 ook afhankelijk was van Piwi6, hetzij in mindere mate. Door kleine RNAs in dengue geïnfecteerde en niet-geïnfecteerde muggencellen te deep-sequencen, hebben wij verder aangetoond dat denguevirus infectie geen groot effect heeft op de expressie van cellulaire microRNAs, een onafhankelijke klasse van kleine RNAs. Wel hebben we nieuwe microRNAs in onze data kunnen identificeren en daarmee hebben we het repertoire van deze regulatoire RNAs in *Aedes aegypti* verder uitgebreid.

De hoofdstukken vier tot en met zes richten zich op de karakterisering van piRNAs die gecodeerd zijn in het genetisch materiaal van de mug zelf. In **hoofdstuk 4** vonden wij dat eiwit-coderende genen een bron zijn van piRNAs en wij hebben verschillende groepen genen geïdentificeerd die voor hun piRNA biogenese afhankelijk zijn van verschillende combinaties van PIWI eiwitten. In de groep van genen die, vergelijkbaar met virale piRNAs, afhankelijk waren van Piwi5 en Ago3 zaten de replicatie-afhankelijke histonen. We hebben kunnen aantonen dat hun piRNAs dynamisch accumuleren tijdens de celcyclus. Sommige piRNA-producerende genen bleken virale elementen te zijn die in de loop van evolutie zijn geïntegreerd in het genoom van de mug. In **hoofdstuk 5** hebben wij met een vergelijkende genoomanalyse laten zien dat deze zogenaamde endogene virale elementen vooral voorkomen in *Aedes aegypti* en *Aedes albopictus* en dat zij een bron van piRNAs zijn. In **hoofdstuk 6** hebben wij een onverwachte bron van piRNAs in het *Aedes aegypti* genoom aan het licht gebracht: een satelliet DNA locus dat wij *satDNA1* hebben genoemd. Satelliet DNA hoort tot de repetitive DNA elementen en bestaat uit directe herhalingen van DNA sequenties die enkele tot een paar honderd nucleotides groot zijn. Dit soort elementen evolueren nornaal heel snel en zijn zelfs tussen evolutionair nauw verwante soorten nauwelijks geconserveerd. *satDNA1* daarentegen bleek al ruim 180 miljoen jaar geconserveerd. Vanuit de *satDNA1* locus kwamen twee piRNAs hoog tot expressie die uitsluitend gebonden waren aan Piwi4. Een *satDNA1* piRNA was in staat om een reportergen uit te schakelen, wat suggereert dat deze piRNAs betrokken zijn bij de regulatie van genexpressie in muggen

Onze resultaten laten zien dat in *Aedes* muggen diverse PIWI eiwit – piRNA complexen gemaakt worden die verschillende functies hebben in de productie van diverse klassen virale en niet-virale piRNAs. Wij veronderstelden dat, naast PIWI eiwitten, andere eiwitfamilies betrokken zijn bij de opbouw, het stabiliseren, en functie van deze complexen.

Uitgelezen kandidaten hiervoor zijn TUDOR eiwitten, waarvan bekend is dat ze in andere modelorganismen belangrijke cofactoren zijn van PIWI eiwitten. Dit vermoeden bevestigen we in **hoofdstuk 7** waar we met behulp van een knockdown screen het Tudor eiwit AAEL012441 geïdentificeerd hebben als biogenese factor voor Sindbisvirus piRNAs. Immuno-precipitatie van AAEL012441 gevolgd door massaspectrometrie heeft nog verdere eiwitten aan het licht gebracht, die samen met Piwi5 en Ago3 een groot eiwitcomplex vormen dat verantwoordelijk is voor de productie van virale piRNAs.
In **hoofdstuk 8** worden de resultaten van dit proefschrift bediscussieerd in relatie met de recente literatuur en worden mogelijke richtingen voor vervolgstudies besproken. In **hoofdstuk 9** wordt de rol van microRNAs in virus-gasteer interacties beschreven.
Samenvattend heeft dit proefschrift belangrijke fundamentele kennis opgeleverd over virale en niet-virale piRNAs in *Aedes aegypti*. Bovendien heeft het aangetoond dat muggencellen een uitstekend systeem zijn om somatische piRNAs te bestuderen en zijn er interessante nieuwe klassen van kleine RNAs geïdentificeerd. Deze resultaten vormen de basis voor toekomstige studies naar de functie van virale en niet-virale piRNAs in het immuunsysteem van de mug en hun rol in de transmissie van arbovirussen.

DEUTSCHE ZUSAMMENFASSUNG

Blutsaugende Insekten, insbesondere Stechmücken, sind vor allem in den tropischen und subtropischen Regionen Afrikas, Lateinamerikas und Südostasiens die Hauptüberträger zahlreicher Viruserkrankungen. Zu trauriger Berühmtheit sind Denguefieber, Gelbfieber, Chikungunyafieber oder Zikafieber gelangt. Die durch Mücken übertragenen Viren werden im Allgemeinen als Arboviren (basierend auf der englischen Abkürzung *Arthropod-borne* = übertragen durch Gliederfüßer wie Mücken, Sandfliegen oder Zecken) bezeichnet. Infektionen mit diesen Viren können zu fiebrigen Erkrankungen führen, die in der Regel nach wenigen Tagen abklingen. In einigen Fällen, vor allem bei Kleinkindern, Senioren und Menschen mit geschwächtem Immunsystem kann sich ein komplikationsreicherer Krankheitsverlauf einstellen, bei dem es, je nach Virus, zu Schocksymptomen, hämorrhagischen (=mit Blutungen einhergehend) Fieberschüben, Krampfanfällen und Hirnentzündung kommen kann. In seltenen Fällen kann so eine Viruserkrankung tödlich enden. In den vergangenen Jahren kam es immer wieder zu fatalen Arbovirus Epidemien, einige, sowie der Chikungunya-Ausbruch (2013-2014) oder der Zika-Ausbruch (2015-2016) in Südamerika, mit mehreren hunderttausenden geschätzten Krankheitsfällen. Das weitverbreitetste humane Arbovirus, Dengue, kommt in nahezu allen tropischen und subtropischen Regionen vor und rund die Hälfte der menschlichen Bevölkerung lebt in einem Risikogebiet für Dengue Infektionen. Für die meisten Arbovirus Erkrankungen gibt es keine spezifischen Behandlungsmethoden oder vorbeugende Impfungen.

Man unterscheidet bei der Übertragung von Arboviren zwischen einem sylvatischen (=im Wald auftretenden) und einem städtischen Zyklus. Im sylvatischen Zyklus übertragen Stechmücken das Arbovirus in erster Linie zwischen wilden Tieren. Menschen die von einer infizierten Mücke gestochen werden, können zwar eine Viruserkrankung entwickeln, aber eine weitere Übertragung des Virus auf eine nicht-infizierte Mücke ist nicht möglich. Im städtischen Zyklus hingegen werden Arboviren durch Mücken direkt von Mensch zu Mensch übertragen, was ein höheres Risiko auf größere Epidemien nach sich zieht (Abbildung 1A auf Seite 296).

Arboviren im städtischen, epidemischen Zyklus werden von *Aedes* Mücken, vor allem *Aedes aegypti* (Gelbfiebermücke) oder *Aedes albopictus* (Asiatische Tigermücke) übertragen. Diese Spezies haben sich an städtische Lebensräume angepasst und bevorzugen den Menschen als Wirt. Da diese Mücken einen essentiellen Teil des Arbovirus-Lebenszyklus darstellen, sind sie ein vielversprechendes Zielobjekt für die Entwicklung einer Strategie zur Reduzierung von Arbovirusübertragungen. Hierfür ist es jedoch wichtig zu verstehen, welche Faktoren die Übertragungsrate von Arboviren beeinflussen:

Für eine effiziente Übertragung von Mensch zu Mensch, müssen Arboviren nach der Aufnahme in einer Blutmahlzeit (Schritt 1 in Abbildung 1B) die Zellen des

Mitteldarmdeckgewebes der Mücke infizieren (Schritt 2a). Im nächsten Schritt verlassen die Virusteilchen diese Zellen auf der Mitteldarm abgewandten Seite wieder (Schritt 2b) und verbreiten sich in weiteren Organen (Schritt 3). Letztendlich, infiziert das Virus das Speicheldrüsendeckgewebe (Schritt 4a) und vermehrt sich in den Zellen, die den Speichelkanal umgeben. Wenn sich genügend Virusteilchen im Speichel der Mücke angesammelt haben (Schritt 4b), kann die Mücke mit jedem Stich den Viruserreger auf eine nicht-infizierte Person übertragen.

Neben diesen anatomischen Barrieren, müssen Viren die Immunantwort der Mücke überwinden. Das wichtigste Element des antiviralen Immunsystems von Insekten ist die sogenannte „RNA-interferenz". Hierbei werden aus der RNA des Arbovirus kleine RNA Moleküle produziert, die man „small interfering RNAs (siRNAs)" nennt. Diese siRNAs sind in letzter Konsequenz verantwortlich für die Zerstörung des genetischen Materials von Arboviren und unterdrücken somit deren Vermehrung. Überaschenderweise haben wir und andere Gruppen entdeckt, dass in *Aedes* Mücken außer siRNAs ein zweiter Typ kleiner RNAs, die man „PIWI-interagierende RNAs (piRNAs)" nennt aus viraler RNA produziert wird. Dieses Ergebnis lässt vermuten, dass zwei unabhängige kleine RNA Systeme ihren Beitrag zur antiviralen Immunität in Mücken liefern.

Die Produktion und die Funktionsweise dieser piRNA Moleküle ist bis dato aber weitgehend unbekannt. Im Rahmen dieser Doktorarbeit haben wir mithilfe biochemischer und genetischer Experimente in *Aedes* Mückenzellen den Produktionsmechanismus viraler und nicht-viraler piRNAs erforscht. Das Hauptaugenmerk lag dabei auf PIWI-Eiweißen, die für die Funktion von piRNAs notwendig sind. Wir konnten ermitteln, dass die PIWI-Eiweiße „Piwi5" und „Ago3" hauptverantwortlich für die Herstellung von viralen piRNAs zweier Arboviren, Sindbisvirus und Denguevirus, waren (**Kapitel 2 und 3**). Die Produktion von Denguevirus piRNAs war darüber hinaus auch in geringerem Maße abhängig von Piwi6 (**Kapitel 3**). Desweiteren haben wir die Produktion nicht-viraler piRNAs beschrieben: Im genetischen Material von *Aedes* Mücken selbst gibt es nämlich auch unterschiedliche Quellen von piRNAs. Wir konnten zeigen, dass aus sogenannte „Transposon-Sequenzen" piRNAs produziert werden, und dass die PIWI-Eiweiße Piwi4, Piwi5, Piwi6 und Ago3 hierfür verantwortlich sind (**Kapitel 2**). Ferner haben wir festgestellt, dass einige Mückengene eine Quelle von piRNAs waren, und dass, ähnlich wie bei Transposon-Sequenzen, Piwi4-6 und Ago3 maßgeblich für ihre Herstellung waren (**Kapitel 4**). Interessanterweise stellten sich in dieser Analyse einige dieser Gene als virale Elemente heraus, die im Laufe der Evolution in das genetische Material der Mücke integriert worden waren. Eine ausführlichere Studie dieser sogenannten „endogenen, viralen Elemente" ergab, dass insbesondere in *Aedes aegypti* und *Aedes albopictus* zahlreiche Virus Elemente im genetischen Material der Mücken vorhanden sind (**Kapitel 5**). Als eine weitere Quelle von piRNAs entdeckten wir ein genetisches Element, das wir *satDNA1* getauft haben. Die piRNAs von *satDNA1* waren

ausschließlich an Piwi4 gebunden. (**Kapitel 6**). Diese Ergebnisse deuteten darauf hin, dass in *Aedes* Mückenzellen diverse PIWI-Eiweiß – piRNA Komplexe gebildet werden, die unterschiedliche Rollen in der Produktion von viralen und nicht-viralen piRNAs ausüben. Wir nahmen an, dass außer PIWI-Eiweißen noch weitere Eiweiße beteiligt waren, diese Komplexe aufzubauen und zu stabilisieren. Tatsächlich, konnten wir zusätzliche Eiweiße identifizieren, die zusammen mit Piwi5 und Ago3 an der Produktion viraler piRNAs beteiligt waren (**Kapitel 7**).

Zusammenfassend, liefert diese Doktorarbeit fundamentale Erkenntnisse über die Produktionsmechanismen viraler und nicht-viraler piRNAs in *Aedes* Mücken. Sie etabliert darüber hinaus Mückenzellen als relevantes System zur Studie des piRNA Systems und sie identifiziert neue, interessante Klassen kleiner RNAs und erweitert damit das Spektrum von regulierenden RNA Molekülen. Auf Basis dieser Erkenntnisse können weiterführende Studien die Rolle von PIWI-Eiweißen und piRNAs im Immunsystem von *Aedes* Mücken analysieren und somit erforschen, ob und wie durch sie die Übertragung von Arboviren beeinflusst wird.

ACKNOWLEDGEMENTS, DANKWORT

After more than five years of thinking in terms of *Aubergine* and *ping-pong* my time as a PhD student is approaching its end. The road ahead is already promising new exciting challenges but now is the time to stand still and look back. The last years have been an amazing journey with ups and downs and surprising turns. I consider myself incredibly lucky that so many great people have supported me during this time and I want to use the next paragraphs for some words of gratitude.

"Small RNAs – yes, but please not piRNAs." **Ronald** - I guess these were my thoughts when we were brainstorming about the content of my PhD project. I wanted to stick to the topic that I was familiar with – microRNAs – and in the first year of my PhD you let me further explore this comfort zone of mine. At the same time, however, with a PCR here and some cloning there, you slowly allured me to the pathway with the vegetable names…
Thank you so much for encouraging but never forcing me to leave the safe harbor. You have given me all the freedom to follow my scientific interests and provided guidance when I needed it. Your office door was always open when I wanted to ask for advice or discuss the latest data. I particularly enjoyed our long discussions in which we planned how to further explore the untouched scientific territories of mosquito piRNAs. I feel truly honored that you have given me the opportunity to pursue on this interesting topic and to hopefully take it to the next level in the new mosquito facility. I'm very much looking forward to more excitement in the coming years.

Robert, I want to thank you for having accepted to guide me through the process of graduating as my promoter. With the mosquito as common interest, I am hopeful that in the future our scientific paths will cross frequently and we can share ideas about how to tackle mosquito-transmitted diseases – either of viral or parasitic origin.

Christian, du bist für mich ein Vorbild-Wissenschaftler. Grundsätzlich neugierig und ständig begeistert von den Wundern und Rätseln des Lebens. Es ist faszinierend dir zuzuhören wenn du über deine Forschungsarbeit berichtest, ob das während eines Vortrages ist oder gemütlich abends bei einem Kaffee („also bei nem Bier" – „KAF-FEE" – „BIE-ER"). Danke, dass du seit Kindestagen immer an meiner Seite stehst – ob als Freund, Astronom, Trauzeuge oder Paranymph. Ich habe keine Zweifel, dass, egal wo es dich im Laufe deiner weiteren Karriere noch hinschlägt: You're gonna rock that place!

Joep, both as a student and PhD fellow I have experienced you as talented researcher. It is a great pleasure to see how you took your internship project and really made it *your* work. I'm confident that also in the future you will provide new exciting twists to the piRNA project. I'm also looking forward to some unforgettable nights out in Nijmegen or maybe Edinburgh where we can enjoy good-good, bad-bad, and, most importantly, good-bad music.

Rebecca, der Grat zwischen Genie und Wahnsinn ist dünn, manchmal sehr dünn, und in deinem Fall mikroskopisch klein. Mit anderen Worten – du bist aus dem Holz aus dem große Forscher/innen gemacht sind. Sowohl als Studentin als auch als PhD Kollegin habe ich dich als extrem motivierte Wissenschaftlerin erlebt, die dem Flippi Projekt zu einem echten Höhenflug verholfen hat. Und natürlich noch einmal Danke für die verrückte 4-daagse Erfahrung. "You better get out of our way now...".

Bas has golden hands – so they said. However, that can only be considered a bold understatement. You perform each experiment with the greatest precision and your results almost come with a guarantee for being conclusive. Thank you so much for all the great assistance in the different piRNA sub-projects.

Finny, you brought great virology expertise to the molecular biology-biased piRNA club, which I really enjoy a lot. Yet, I still have a hard time seeing, how you can honestly enjoy doing that many plaque assays :-). I'm also happy to have somebody around who so passionately defends the greatest book of the 20th century against the ignorant – my dear fellow of the ring.

Susan, du hast das Phänomen CRISPR zu uns ins Labor gebracht und deine Expertise auf diesem Gebiet wurde nicht nur in unserer Gruppe sondern auch im ganzen Institut (à la ‚meet the expert meeting') wertgeschätzt. An dich darf ich auch den Promotions-Staffelstab weitergeben. Ich wünsche dir alles Gute auf den letzten Metern und viel Erfolg beim Schreiben deiner Doktorarbeit.

Febi, you came to our group all the way from Indonesia and brought your little family to Nijmegen – I have great respect for that decision. Thank you for trying to make us believe that on Java it is actually too hot and that the Dutch summers are just perfect.

Bodine, you are the latest addition to the club of van Rij lab PhDs. I wish you good luck for your Zika project. And also important, I wish you a fun time and a lot of exciting memories to look back at when you are about to graduate.

Gijs, you are the good soul of the lab. You know where everything is, how everything works and how to help a country getting over its fear for cell-fusing agent virus. The mosquito scientific community of the Netherlands owes you a lot! I'm already looking forward to celebrate that during the happy hardcore session of the next 'Het foute uur' party.

Erika, you were a great colleague and you pushed the genic piRNA project so much. You also established so many valuable techniques like RNA-Seq, GFP-Trap or Cell-cycle manipulation in the lab. It is a pity that we never had the chance to put your Café-Jos-under-the-table-claim to the test. I hope that there will be time for this at some point.

Sarah, for a long time we have shared an (I-)U-tje and our work spaces were just separated by the paper river Rhine. Who would have thought that, after you left the lab, we would once be living in the same neighborhood only a walking distance away from Pasteur institute and a cozy Café opposite of métro Convention. Thank you for great scientific and non-scientific discussions with food and drinks served from lab glassware. Also I want to say thank you for sharing your expertise in mosquito handling and P3 training (what doesn't the P stand for).

Koen, you taught me a lot about the general lab-habits when I just started in the group and you also introduced me to the BSL-3 lab. I still remember your warning in the old BSL3 lab –" in summer it's going to be warm in here". And then later in July, we were practically running out of sweat. Luckily there is always Café Jos to re-stabilize the water-balance.

Walter, you always showed great interest in the progress of the piRNA project and stayed involved, which I really appreciated. In Keystone it was great to meet you again and I will remember the good times we had at this conference, both in the plenary room and on the slopes.

Joël, as the alpha PhD student, I was happy that you convinced the pack to accept me although I was the first one not to work with flies. You were always there to help even a lost mosquito person. Thanks also for the gezellige evenings with some bottles of Schneider's on the table (I'm not talking about the medium here).

Rob, during the time you stayed in our group you were the master of funny, unbelievable stories – I mean they were literally unbelievable. I'm still surprised that many/most/all of them seemed to be true in the end except of course for your masterpiece – the 6th finger. Thank you also for introducing us to the Discodel!!

Inge and **Jeroen**, you were my first students and both of you have done great work that pushed the piRNA project further. I wish you all the best for your future careers.

Thanks to the lab mates from the **Parasitology group** and to the **colleagues from "the other side"** especially in the BSL-3 lab for your regular support. I'm looking forward to more memorable activities at our joined lab outings or Christmas parties.

Carla, you entered the stage when my PhD studies were supposed to slowly finish and I was mentally preparing for finalizing some last experiments and writing things up. But the wind changed, and instead I got the great opportunity to join your group during the last months of my PhD and beyond. I want to thank you again for having made this experience possible. I learned a lot about mosquito work during my stay and hope to stay in close contact also in the future.
Thanks also to all the members of the group who have given me such a warmhearted

welcome making it easy to feel home in Paris: **Annabelle, Brigitte, Irena, Lorena, Valerie, Vanesa, Virginia, Hervé, Juan, Lionel** and **Yasu**. Thanks for all the support and assistance that I got in the lab – from identifying the locations of enzymes, over preparing and analyzing deep-seq libraries, to learning how to dissect mosquito ovaries. I enjoyed the friendly atmosphere with ice-breaking questions, Ponyo Ponyo songs, cheese and wine tastings or Whisky Wednesdays. Thanks also to **members of the Vignuzzi and Lambrechts labs** who have made my stay in Paris such a memorable time. I'm looking forward to seeing many of you in the future.

I was lucky to still be in the lab when the **van Kuppeveld groep** was around. When you moved to Utrecht, not only did we lose a lot of valuable expertise - most importantly we lost a bunch of great lab mates. Therefore, it is always nice to meet some of you guys at the DAVS or at a graduation party.

Bert, we know already since my first master's internship and I was happy when you agreed to be my PhD mentor. I have enjoyed our meetings a lot and will remember many of your valuable advices on how to shape a scientific career.

Science does not occur in a bubble and I'm happy to be part of so many inspiring collaborations. First of all, I want to thank **Amy** and **Valentina**. You were my supervisors during the second Master internship at the University of Edinburgh. I'm happy that also during my PhD, together with **Al**, we have had the chance to work together on some projects. Also Valentina, it was great to spend (too little) time together in Paris. Thanks to **Gorben**, **Giel** and **Jelke** from the virology group and **Sander** and **Chantal** from the entomology group in Wageningen for the fruitful collaborations. Finally, thanks to **Umberto** and **Mariangela** for the exciting work on endogenous viral elements in mosquito genomes. I'm certain we will stay in close contact in the future.

The co-editors of The Microbial TIMES **Amelieke** and **Tobias**, it has been and still is a great pleasure to prepare the newsletter together with you. Rumors say that the MT is currently the most-read newsletter in the department – a great achievement :-).

Sometimes I can barely believe that the rotation of the earth is independent of the interaction between Piwi5 and Ago3. Therefore, I'm glad to have friends who regularly remind me that, *apparently*, there is a reality outside of the piRNA world: **B, Bowo, Jockel, Monty, Murvin, Yogi** – Sehr geehrte Vereinsmitglieder des Doppelkopfklubs Kalkar ®. Einer Nachricht an euch, für die ich keinen Anwalt brauche: Jungs, ihr seid einfach die Besten. **Christel** und **Floppi**: Ihr marschiert seit vielen Jahren mit Gitarre und Wasserflasche in der Hand mit mir auf dem Weg des Lebens. Möge er noch lange voller Freude so weitergehen. **Ellen**, **Helena**, **Ina**, **Vanessa**, **Vera**, **Marcel**, **René**, **Andrea**, **Öppi**, **Sarah**, **Sina**, **Markus** ob ruhiger Spieleabend, ein leckeres Abendessen, oder eine ausgelassene Schlagerparty – mit euch sind diese Events gleich noch mal so schön. Und Ellen – ganz lieben Dank für

die Nachhilfe in Photoshop und Indesign und deine Unterstützung beim Coverdesign. **Fieke**, **Laurien**, **Steef** and **Abby** – You have made my time in Edinburgh unforgettable. I'm looking forward to our next reunion to share a wee dram of Monkey Shoulder.

Opa, wer weiß – vielleicht haben die Spaziergänge mit dir über die Dünen, damals als kleiner Junge, schon mein Interesse an den Naturwissenschaften geweckt – genug Mücken gab es da ja im Sommer auf jeden Fall. Danke auch, dass ich während meines Studiums in der kleinen Wohnung hausen durfte.

Antje und **Laura**, gefühlt gerade noch wohlbehütet Kamilla und von Steinfurt im Sandkasten geerntet, hat sich doch gerade in den letzten Jahren so viel Spannendes getan, sowohl beruflich als privat. Ihr wart in all den Jahren immer an meiner Seite und dafür möchte ich aus tiefstem Herzen danke sagen.

Mama und **Papa**. So, ich denke mal euer innigster Traum ist wahr geworden. Ihr dürft in den kommenden Jahren weiterhin regelmäßig sagen „Er ist in der Forschung und macht was mit Mücken". Aber jedes Mal, wenn ihr diesen Satz benutzt, dürft ihr euch bewusst sein, dass ihr einen großen Beitrag dazu geleistet habt, dass ich *was mit Mücken mache*. Ihr standet mir bei all meinen noch so ungewöhnlichen Karriere-Entscheidungen zur Seite, und das bedingungslos. Immer konnte ich mir eures Rückhalts bewusst sein und das tat und tut auch heute noch sehr gut. Ich danke euch dafür.

Sonja, seit 2004 gehen wir gemeinsam durchs Leben und heute möchte ich noch einmal Danke dafür sagen. Ich kann mir vorstellen, dass es nicht immer einfach ist, einen Wissenschaftler an seiner Seite zu haben, mit einem merkwürdigen Verständnis von Arbeitszeiten und immer mal wieder kurzen oder längeren Auslandsaufenthalten. Ich möchte dir danken, dass du mich trotz aller zu erwartenden Schwierigkeiten unterstützt hast, als ich mich dazu entschieden habe für jeweils mehrere Monate nach Edinburgh und Paris zu gehen. Ich weiß das wirklich sehr zu schätzen.
Im Sommer 2015 hast du mir den glücklichsten Moment beschert, als du „ja" zu mir gesagt hast. Ich danke dir dafür, dass du mich so annimmst wie ich bin, und freue mich auf weitere spannende Jahre mit dir an meiner Seite.

LIST OF PUBLICATIONS

Publications featured in this thesis

1. Palatini U*, **Miesen P***, Carballar-Lejarazu R, Ometto L, Rizzo E, Tu Z, van Rij RP, Bonizzoni M. (2017) Comparative genomics shows that viral integrations are abundant and express piRNAs in the arboviral vectors Aedes aegypti and Aedes albopictus. *BMC Genomics.* 18: 512

2. Girardi E*, **Miesen P***, Pennings B, Frangeul L, Saleh MC, van Rij RP. (2017) Histone-derived piRNA biogenesis depends on the ping-pong partners Piwi5 and Ago3 in Aedes aegypti. *Nucleic Acids Res*. 45: 4881-92

3. **Miesen P***, Joosten J*, van Rij RP. (2016) PIWIs Go Viral: Arbovirus-Derived piRNAs in Vector Mosquitoes. *PLoS Pathog*. 12:e1006017

4. **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

5. **Miesen P**, Girardi E, van Rij RP. (2015) Distinct sets of PIWI proteins produce arbovirus and transposon-derived piRNAs in Aedes aegypti mosquito cells. *Nucleic Acids Res*. 43: 6545-56

6. Libri V*, **Miesen P***, van Rij RP, Buck AH. (2013) Regulation of microRNA biogenesis and turnover by animals and their viruses. Cell Mol Life Sci. 70: 3525-44

Additional publications

7. Göertz GP, Fros JJ, **Miesen P**, Vogels CB, van der Bent ML, Geertsema C, Koenraadt CJ, van Rij RP, van Oers MM, Pijlman GP (2016) Non-coding subgenomic flavivirus RNA is processed by the mosquito RNAi machinery and determines West Nile virus transmission by Culex pipiens mosquitoes. *J Virol*. 90: 10145-59

8. Fros JJ, **Miesen P**, Vogels CBF, Gaibani P, Sambri V, Koenraadt CJ, van Rij RP, Vlak JM, Takken W, Pijlman GP. (2015) Comparative Usutu and West Nile virus transmission potential by local Culex pipiens mosquitoes in north-western Europe. *One Health.* 1:31-36

9. van Cleef KW*, van Mierlo JT*, **Miesen P**, Overheul GJ, Fros JJ, Schuster S, Marklewitz M, Pijlman GP, Junglen S, van Rij RP. (2014) Mosquito and Drosophila entomobirnaviruses suppress dsRNA- and siRNA-induced RNAi. *Nucleic Acids Res*. 42: 8732-44

10. Bronkhorst AW, **Miesen P,** van Rij RP. (2013) Small RNAs tackle large viruses: RNA interference-based antiviral defense against DNA viruses in insects. Fly (Austin). 7: 216-23

11. Vodovar N, Bronkhorst AW, van Cleef KW, **Miesen P**, Blanc H, van Rij RP, Saleh MC. (2012) Arbovirus-derived piRNAs exhibit a ping-pong signature in mosquito cells. *PLoS One*. 7: e30861

12. Libri V*, Helwak A*, **Miesen P,** Santhakumar D, Borger JG, Kudla G, Grey F, Tollervey D, Buck AH. (2011) Murine cytomegalovirus encodes a miR-27 inhibitor disguised as a target. *Proc Natl Acad Sci U S A*. 109: 279-84

*equal contribution

LIST OF AWARDS

January 2007	Study scholarship from the German National Academic Foundation
January 2011	Scholarship from the German National Academic Foundation to support scientific projects abroad
January 2011	PhD fellowship from Radboud University Medical center
January 2012	RIMLS 'Best master thesis' award
December 2013	Travel grant by the Nijmegen Institute for Infection, Inflammation and Immunity to attend EMBL conference
January 2016	Travel grant by the Radboud Institute for Molecular Life Sciences to attend Keystone conference
May 2016	Travel grant by the Federation of European Microbiological Societies to attend the 11th Microsymposium on small RNAs
December 2016	Short term fellowship by the European Molecular Biology Organization to visit Pasteur institute

CURRICULUM VITAE

Pascal Miesen was born on 3rd June 1987 in Goch, Germany as son of Petra-Maria Miesen and Wilhelm-Johannes Miesen. Together with his two sisters Antje and Laura, he was raised in Kalkar-Wissel near Kleve in the Lower Rhine area. He visited the Städtisches Gymnasium Kalkar where he obtained his general qualification for university entrance (Abitur) in 2006. He was selected as a member of the German National Academic Foundation (Studienstiftung des deutschen Volkes e.V.) from which he obtained a study scholarship for his Bachelor and Master studies. In 2006, he started his Bachelor studies 'Medical Biology' at Radboud University Nijmegen. Besides his studies, Pascal participated in the interdisciplinary Honours programme of Radboud University. In 2009, he performed his Bachelor internship in the group of Prof. Dr. Gert Jan Veenstra, at the Department of Molecular Biology. During this internship he investigated the expression of microRNAs in *Xenopus tropicalis* embryos. This research project was his first encounter with the field of small RNAs, which immediately fascinated him. In August 2009, he obtained his Bachelor diploma with the distinction *cum laude*. To pursue his interest in molecular biology, Pascal applied to the research master program 'Molecular Mechanisms of Disease' at Radboud University Medical Center. In the course of his master studies he performed his first internship at the Department of Laboratory Medicine in the group of Dr. Bert van der Reijden where he studied the role of ubiquitylation in normal and malignant blood development. For his second internship, Pascal joined the group of Dr. Amy Buck at the University of Edinburgh. In this internship he returned to the field of small RNAs and studied the interactions between murine cytomegalovirus and the microRNA machinery. In August 2011, he graduated *summa cum laude* and his Master thesis was awarded as best Master thesis of the year by the Radboud Institute for Molecular Life Sciences.

In 2011, Pascal was awarded a personal PhD scholarship from Radboud University Medical Center that allowed him to do his PhD studies in the research group of Dr. Ronald van Rij at the Department of Medical Microbiology. Following his enthusiasm for small RNA pathways, his project focused on the role of small RNAs, primarily piRNAs, in arbovirus infections. In 2016, Pascal obtained a short-term fellowship from the European Molecular Biology Organization (EMBO) to join the laboratory of Dr. Maria-Carla Saleh at Pasteur Institute, Paris as a visiting scientist. During this period, he was trained to rear and manipulate *Aedes aegypti* mosquitoes and to infect them with human arboviruses under strict safety conditions. In June 2017, Pascal returned as a post-doctoral researcher to the group of Ronald van Rij to establish a new research line that aims to dissect molecular determinants of arbovirus transmission in *Aedes aegypti* mosquitoes.

In July 2015, Pascal got married to Sonja Miesen, née van Bernum. They live together in Sonsbeck, a small town in the Lower Rhine area.

PHD PORTFOLIO

Name PhD candidate: P. Miesen	***PhD period: 01-10-2011 – 31-03-2017***	
Department: Medical Microbiology	***Promotor: Prof. RW Sauerwein***	
Graduate School: Radboud Institute for Molecular Life Sciences	***Co-promotor: Dr. RP van Rij***	
	Year(s)	ECTS
TRAINING ACTIVITIES		
a) Courses & Workshops		
- ***Radboud Introduction day***	2011	0.5
- ***Technical Forums***	2011-2016	1.6
- ***Course: The Art of presenting science***	2014	1.5
- ***Course: Academic Writing***	2014	3.0
- ***Course: Education in a Nutshell#***	2015	3.0
b) Seminars & lectures		
- ***RIMLS Seminars***	2011-2016	1.8
- ***Spotlights and Radboud Research Rounds***	2011-2016	1.3
- ***RIMLS Lecture series***	2012-2014	0.5
c) Symposia & congresses		
- ***Radboud New Frontiers Symposia***	2011-2013	1.5
- ***RIMLS PhD retreat***	2012-2015	3.0
- ***Dutch Annual Virology Symposia (oral presentation in 2015)***	2012-2017	1.75
- ***Microsymposia on small RNAs (poster presentation in 2016)***	2012,2014,2016	2.75
- ***Radboud Center for Infectious diseases Science days and summer symposia (oral presentation in 2016)***	2012, 2016	1.25
- ***RNA Society meeting: Biogenesis and Turnover of small RNAs***	2012	0.75
- ***VECTORIE meeting***	2014	0.25
- ***EMBO|EMBL symposium: The complex life of mRNA (poster presentation)***	2014	1.5
- ***Chemistry as innovating Science (CHAINS) meeting (oral presentation)***	2015	0.75
- ***Keystone Meeting: Small RNA silencing – Little guides big biology (poster presentation)***	2016	1.5
- ***Dutch Arbovirus Network (oral presentation)***	2016	0.5
d) Other		
- ***Training and supervision plan committee member***	2014-2016	1.0
TEACHING ACTIVITIES		
e) Lecturing		
- ***Masterclass lecture: microRNAs***	2012	0.1
f) Supervision of internships / other		
- ***Supervision of internship Master student, Medical Biology***	2012-2013	2.7
- ***Supervision of internship HLO student, Saxion Enschede***	2014	1.3
- ***Supervision of internship Master student, Biomedical Sciences***	2014-2015	2.7
- ***Supervision of internship Master student, Hamburg University***	2015-2016	4.0
- ***Assistance of Practical course: Viral respiratory diseases***	2015	1.2
TOTAL		**41.7**