UNIVERSITÀ
DI PAVIA

Dipartimento di Biologia e Biotecnologie "L. Spallanzani"

Host-pathogen interaction: *Aedes* spp. mosquitoes and arboviruses

Mosquito gambit

Michele Marconcini

Dottorato di Ricerca in
Genetica, Biologia Molecolare e Cellulare
Ciclo XXXII – A.A. 2016-2019

UNIVERSITÀ
DI PAVIA

Dipartimento di Biologia e Biotecnologie "L. Spallanzani"

Host-pathogen interaction: *Aedes spp.* mosquitoes and arboviruses

Michele Marconcini

Supervised by Prof. Mariangela Bonizzoni

Dottorato di Ricerca in
Genetica, Biologia Molecolare e Cellulare
Ciclo XXXII – A.A. 2016-2019

Abstract

Arboviruses such as the flavivirus Zika, Dengue and Usutu viruses, and the alphavirus Chikungunya represent resurging or emerging public health threats worldwide. In Europe, arboviral risk is dependent on the presence of the competent vector *Ae. albopictus*, a mosquito that arrived from South East Asia only 30 years ago.

At the moment, control of vector populations is the primary tool to prevent arboviral transmission because no drugs and limited vaccines are available for these diseases.

Mosquito response to arboviral infection is a complex and evolving phenotype dependent primarily on the activity of RNAi pathways. Among RNAi, the piRNA pathway has been shown to have a significant role in mosquito immunity. However, the function of this pathway is still enigmatic because it was characterized only recently and it has been mostly studied in the model insect *D. melanogaster,* where it does not have antiviral activity, but functions in germ cells through small RNAs, called piRNAs, to silence TEs and preserve genome integrity.

Recently, we discovered that the *Ae. albopictus* genome carries numerous integrations from viral genomes that are integrated mainly in *loci* producing piRNAs. On this basis, we hypothesized a functional link between the piRNA pathway and viral integrations.

During my PhD, I developed a program with three specific aims to test the above-mentioned hypothesis. Briefly, I aimed to 1) characterize the genes of the piRNA pathway in *Ae. albopictus* 2) study the involvement of viral integrations in virus-host interactions in connection with the siRNA and piRNA pathways; 3) analyse *Ae. albopictus* fitness traits, which include the

larval-to-pupal developmental time, pupae sex ratio, wing span and adult lifespan.

The main questions I address in Aim 1 "the piRNA pathway of *Ae. albopictus*" concern the *Piwi* gene copy number, structure, variability, their protein architecture and expression profile, both throughout the mosquito development and following arboviral infection. I also set up experiments at different growth temperatures to analyze the effects on the expression profile of *Piwi* genes caused by temperature variations similar to the ones encountered in the newly invaded temperate regions.

For aim 2 "NIRVS and arboviral infection", I performed a comprehensive analysis of the expression profile of small RNAs whose sequences match the infecting viruses or NIRVS, in an attempt to shed light on the role of small RNAs of both viral and endogenous origin in the establishment of persistent infections. This analysis was done across carcasses and ovaries to be able to contrast the features of somatic and germline tissues.

For aim 3 "Life history trains of *Aedes albopictus*", I carried out comparative analyses of fitness traits of two reference strains and three natural populations of *Ae. albopictus* paving the way for future studies on the analyses of the life history traits of this invasive mosquito species.

Over-all results gained through my PhD activities will enhance our understanding of unexplored functions of the piRNA pathway beyond the model organism *D. melanogaster*, will give new insights into the evolution of the proteins of the piRNA pathway in *Ae. albopictus* and will provide novel insights on mosquito-vector interactions.

As side projects I also contributed to the analyses of the *Ae. albopictus* genome, based on long-sequencing approaches and Hi-C data; I was involved in a project to test the biological function of NIRVS in *Ae. aegypti* and I contributed to population genetics analyses of *Ae. albopictus* populations in

the context of studying the widespread of resistance to pyrethroid insecticides, one of the most-used tools of vector control.

Abbreviations

AMP = antimicrobial peptide

CDS = coding DNA sequence

CHIKV = Chikungunya virus

DENV = Dengue virus

EIP = extrinsic incubation period

FDR = false discovery rate

FG = fast-evolving genes

FPKM = Fragments per kilobase million

HTS = High Throughput Sequencing

IGV = Integrative genomics viewer

Imd = immune deficiency pathway

ISV = insect-specific virus

LoP = level of polymorphism

LTR = long terminal repeats

miRISC = miRNA induced silencing complex

miRNA = micro-RNA

mRNA = messenger RNA

NIRVS = Non-retroviral integrated RNA virus sequence

nt = nucleotide(s)

ORF = open reading frame

PCR = polymerase chain reaction

PIRC = piRNA cluster

piRNA = PIWI-interacting RNA

PTGS = post-transcriptional gene silencing

qPCR = quantitative PCR

qRT-PCR = quantitative reverse transcriptase PCR

RDRP = RNA-directed RNA polymerase

RISC = RNA-induced silencing complex

RNAi = RNA interference

RPKM = Reads per kilobase million

RT-PCR = reverse transcriptase PCR

SFV = Semliki forest virus

SG = slow-evolving genes

SINV = Sindbis virus

siRNA = short interfering RNA

TE = transposable element

TGS = transcriptional gene silencing

UTR = untranslated region

VC = vectorial capacity

vDNA = viral DNA

vpiRNA = viral PIWI-interacting RNA

vsiRNA = viral short interfering RNA

WNV = West Nile virus

YFV = Yellow fever virus

Contents

1. Introduction

Recent decades have witnessed a dramatic expansion of the invasive species *Aedes albopictus*, also known as the Asian tiger mosquito, which is now present in all continents except Antarctica. This mosquito species is a competent vector for arboviruses which are pathogenic for humans, including the Zika, Dengue, yellow fever and the Chikungunya viruses, among the most relevant to public-health. Currently, therapies for these viruses and vaccines are limited, meaning that control of vectors is the only strategy to prevent transmission of arboviruses. Different vector control strategies exist, and most rely on the understanding of the biology and ecology of the target vector species. Due to the recent global expansion of *Ae. albopictus*, we still lack crucial information regarding its life history traits, as well as the molecular and genetic mechanisms that make it a competent arboviral vector.

Expanding our knowledge on all these aspects of the *Ae. albopictus* biology has been the goal of many scientific groups in past and current years, and it has also been the goal of my PhD project. Briefly, my activity has focused on the characterization of genes of the PIWI-interacting RNA pathway, the involvement of sequences of viral origin integrated into the *Ae. albopictus* genome in virus-host interactions in connection with RNA interference pathways, and on the analyses of the fitness traits of *Ae. albopictus*.

2. Review of the literature

2.1 *Aedes albopictus*

Aedes albopictus (Skuse, 1894), also known as the Asian tiger mosquito, belongs to the family Culicidae, subfamily Culicinae, genus *Aedes*, subgenus *Stegomyia*. Native of the tropical and subtropical regions of South East Asia, this mosquito has spread throughout the world in the past three to four decades through passive movements linked to human activities (Gould and Higgs, 2009). More specifically, *Ae. albopictus* has spread to the islands of the Pacific and Indian Oceans in the late '700, with the recent global spread including regions of the Middle East, Europe, North America, Central America, South America and Africa (Bonizzoni *et al.*, 2013). The presence of the mosquito has been also reported in several other countries, such as Germany, Australia and New Zealand (Kraemer *et al.*, 2015) (Figure 1). The life cycle of *Ae. albopictus* consists of four stages: egg, larvae (4 instars), pupa and adult. Eggs are laid on the surface of water containers and successive water inundations stimulate larval hatching (Hawley, 1988). Mosquitoes can use a wide variety of containers as breeding sites, ranging from natural ponds, swamps, bamboo-tree holes to artificial containers such as tyres, bottles and tins. The dramatic dispersion throughout the world seems to have been facilitated especially by commerce of used tyres and lucky-bamboo through international trading (Paupy *et al.*, 2009).
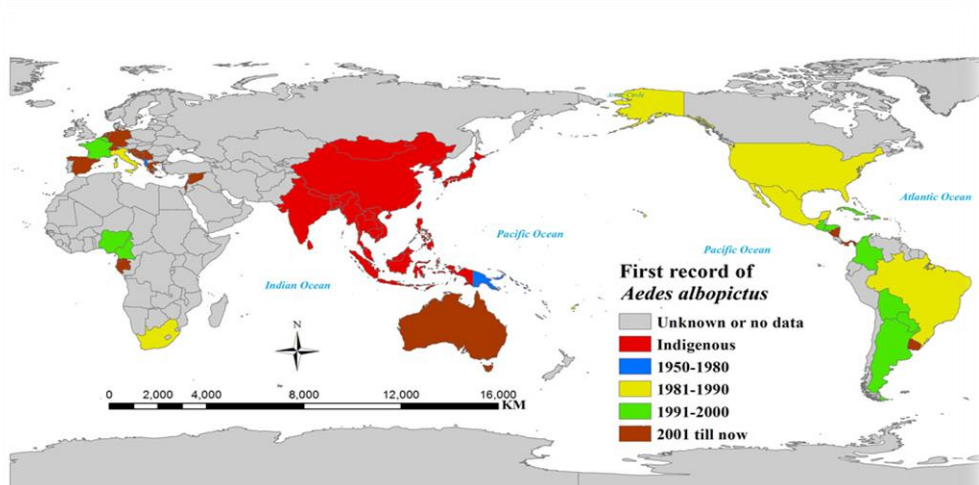
Figure 1. Range distribution of *Aedes albopcitus* (Bonizzoni *et al.*, 2013)*.*

In nature, adult life lasts about three weeks (Gatt *et al.*, 2009). Both males and females feed on sugar from plant fluids, but females must acquire blood meals in order to obtain the proteins needed for oviposition (Hawley, 1988). Females seek a blood meal every 3-5 days; they are aggressive and opportunistic and bite especially during the day, preferentially outdoor, but also indoor. Infected females can transmit pathogens after acquiring pathogens trough infected saliva while biting; pathogens need to replicate and disseminate in the mosquito body reaching the salivary glands before being transmitted to a new host through a subsequent blood meal (Pingen *et al.*, 2016).

Aedes albopictus is a multivoltine species meaning that, in optimal conditions such as those of tropical regions, it can reach up to 17 generations a year,

13

while in temperate regions this species usually reach 7 generations a year (Hawley, 1988). The establishment in temperate regions is due to the ability of overwintering through photoperiodic egg diapause, allowing *Ae. albopictus* to outlast winter even at temperatures as cold as -2°C to -10°C (Medlock *et al.*, 2006). This biological feature has conferred the ability to this mosquito to establish in regions at northern latitudes, such as Europe and North America.

Aedes albopictus is considered the most invasive mosquito species in the world (Buhagiar, 2009) and it is a competent vector for over 20 pathogenic arboviruses, such as Chikungunya, Dengue and Zika viruses, thus representing a serious public health threat especially in temperate regions of the world (Benedict *et al.*, 2007). Moreover, *Ae. albopictus* can transmit a variety of epizootic viruses (Gratz, 2004; Bonizzoni *et al.*, 2013) and nematodes belonging to the genus Dirofilaria (Giangaspero *et al.*, 2013). The impact of *Ae. albopictus* on public health is shown by the (re)-emergence of arboviral diseases in Europe after its introduction in the early '1990 and establishment (Bonizzoni *et al.*, 2013). For instance, Chikungunya outbreaks caused by *Ae. albopictus* recently occurred in Italy. In 2007, in the region Emilia-Romagna, at least 205 Chikungunya cases were reported (Rezza *et al.*, 2007) and in Southern Italy, especially in Lazio and Calabria, between August and September 2017 another outbreak of Chikungunya took place involving more than 800 cases (Venturi *et al.*, 2017). Dengue virus infections have also been reported in recent years in Europe: 2237 cases were reported in France between 2007 and 2012 (Tomasello and Schlagenhauf, 2013), and in 2010 autochthonous transmission of Dengue through *Ae. albopictus* was reported

in France and Croatia (Gjenero-Margan *et al.*, 2011).

There are currently no therapeutic treatments for viruses vectored by *Ae. albopictus* and vaccines are limited. As a consequence, current strategies to limit arboviral disease transmission are focused on control of vectors. The understanding of *Ae. albopictus* biology and the molecular mechanisms underlying the ability of this mosquito to acquire, maintain and transmit arboviruses, a phenotype called vector competence, is expected to aid in the development of novel vector control strategies. For instance, by possibly identifying "immunity effectors" to be used in genetic-based approaches of transmission-blocking strategies.

The expansion of *Ae. albopictus* is still ongoing: climate changes will most likely move the species range to northern latitude (Caminade *et al.*, 2012). For these reasons and the relatively recent global spread, *Ae. albopictus* is now considered a major threat for outbreaks of arboviruses, especially in temperate regions of the world.

2.2. Arboviral transmission cycle in the mosquito

Arthropod-borne viruses (arboviruses) are viruses spread by arthropods, primarily mosquitoes and ticks. Arboviruses include viruses from different viral groups, including *Flaviviridae* (i.e. Dengue, Yellow fever, West Nile, Zika, Japanese Encephalitis virus)*, Togaviridae* (i.e. Chikungunya), Bunyarivirales (i.e. Rift Valley Fever Virus) and *Reoviridae* (i.e. Blue Tongue Virus)(Weaver and Reisen, 2010). The transmission cycle of arboviruses by mosquitoes generally starts from an enzootic vertebrate reservoir (e.g. primates, rodents) from which female mosquitoes acquire the pathogen when

15

blood-feeding and become infected. Successful spread of the virus within the mosquito body and establishment in the salivary glands makes the mosquito infectious. The time in between acquisition of the pathogen and its possible transmission through the saliva is so-called extrinsic incubation period (EIP). Infectious mosquitoes may then feed on uninfected human hosts, thus infecting them through the injection of saliva containing viral particles (Cox *et al.*, 2012; Moser *et al.*, 2016; Agarwal *et al.*, 2017). For most arboviruses, humans are considered dead-end hosts, contributing little to the maintenance of the arbovirus life cycle. However, for some viruses such as dengue viruses (DENV) and Chikungunya virus (CHIKV), strong evidence suggest that humans are the primary amplifying host for urban arboviral cycles (Hollidge *et al.*, 2010). Some arboviruses also have the capability of being transmitted vertically in the mosquito vector (Lequime and Lambrechts, 2014), which allows viruses to persist in mosquitoes populations. Additionally, sexual and placental transmission have also been observed for Zika virus and it is linked to its pathogenic impact on humans (Beckham *et al.*, 2016) (Figure 2).
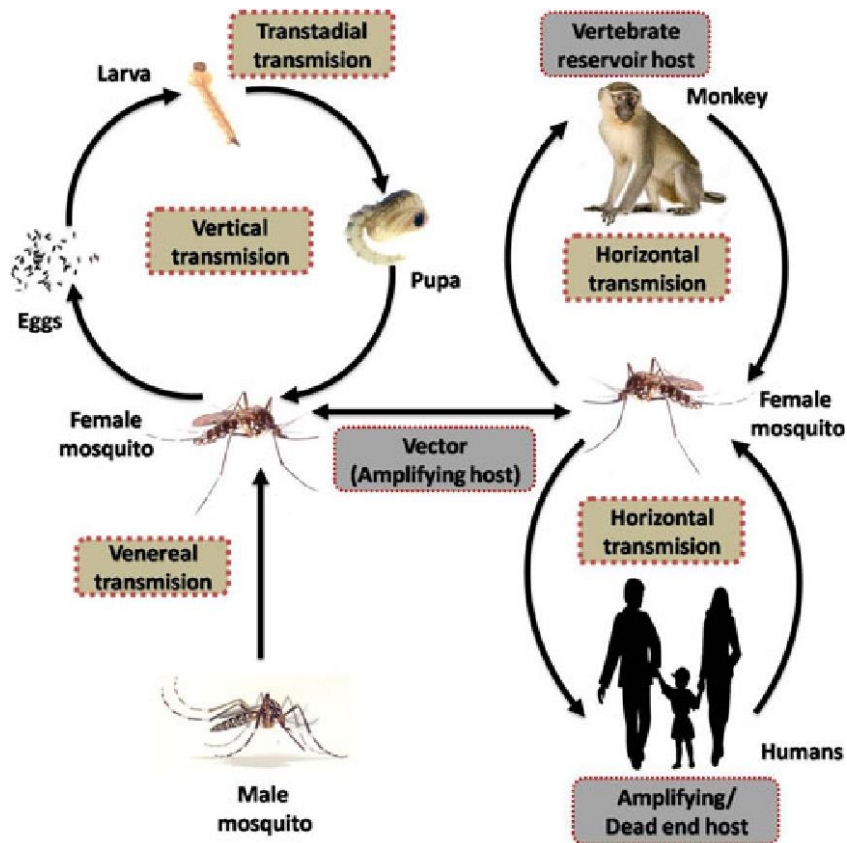
Figure 2. Three important modes of arboviral maintenance and amplification cycles involving horizontal, vertical, and venereal transmission (Agarwal *et al.*, 2017)

The potential of a mosquito population to transmit an arbovirus is defined as "vectorial capacity" (VC) and takes into account the environment, viral, vector and human variables; one formulation of the mathematical model of VC is:

$$VC = \frac{ma^2bp^n}{-ln_p}$$

where m = density of mosquito vectors; a = probability that the vector feeds on the human host; b = vector competence (i.e. proportion of vectors becoming infective); p = daily survival rate of the mosquito vector; n = extrinsic incubation period (EIP), representing the number of days the virus takes to disseminate through the mosquito body after ingestion and reach the salivary glands; 1/ln(p) = the survival rate of the vector during the EIP (Anderson and Rico-Hesse, 2006; Smith *et al.*, 2012; Souza-Neto *et al.*, 2019).

These parameters can be influenced by environmental and genetic factors (Liu-Helmersson *et al.*, 2014; Souza-Neto *et al.*, 2019). The molecular and genetic mechanisms underlying vector competence are still poorly understood. Deeper understanding of vector competence and vectorial capacity are required to properly assess the risk for arboviral outbreak entailed by the recent spread on global scale of *Ae. albopictus*, and to possibly obtain useful knowledge for the development of novel genetic-based strategies of vector control.

2.3 Life history traits of *Aedes albopictus*

The understanding of variation in life-history traits and its molecular basis are major goals of evolutionary studies, which are traditionally pursued using model organisms such as *Drosophila melanogaster* (Stern and Orgogozo, 2008; Fonseca *et al.*, 2013; Klepsatel *et al.*, 2013). The life history traits of *D. melanogaster* are well understood and, in several cases, the genetic bases

of phenotypic variation in fitness traits have been identified (Pool and Aquadro, 2007; Bono *et al.*, 2008; Collet *et al.*, 2016). In non-model organisms, this type of studies is limited (Shapiro *et al.*, 2017). Invasive species are particularly interesting because they represent natural ecological and evolutionary experiments unfolding in a recent historical time frame, thereby providing a window on ecological and evolutionary processes. *Aedes albopictus* is an example of an organism for which the understanding of phenotypic variation in life history traits and of the genetic components of these traits has both biological and applicative motivations, due to its rapid spread out of its native range and its role as vector for human pathogens. The importance of quantifying mosquito fitness traits such as adult longevity to predict epidemics of mosquito-borne diseases and to organize effective control programs was observed since the original formulation of the theory of mosquito-borne pathogen transmission by Ross and Macdonald (Ross, 1911; Macdonald, 1956). However, despite progresses in mathematical modelling of mosquito-borne diseases towards the inclusion of parameters of heterogeneity (i.e. variable mosquito biting rate, spatial and temporal variation in mosquito distribution) and local adaptation, modelers are often faced with uncertainties due to the paucity of biological data with which to evaluate the performance of their models (Reiner *et al.*, 2013; Legros *et al.*, 2016). Thus, a deeper understanding of *Ae. albopictus* fitness traits, their variability across populations and their genetic regulation would aid for the design of mathematical models, which are often required by policy makers before allocating resources for vector control programs.

Moreover, the responses to abiotic environmental conditions are of particular

interest due to the diverse geographic areas recently colonized by *Ae. albopictus*. The response to different temperatures is most relevant, since it has been shown to affect multiple traits at every state of mosquito development, including egg viability (Parker, 1986), larval development (Patel *et al.*, 1990), blood-feeding behaviour (Crans *et al.*, 1996), female fecundity (Hurlbut, 1973), adult longevity (Day *et al.*, 2015) and vector competence (Adelman *et al.*, 2013). For instance, exposing the closely related *Ae. aegypti* to large fluctuations at low temperatures exerted an increased potential for DENV transmission because EIP is shorter at this temperature (Carrington *et al.*, 2013). Also, when larvae are reared at 18ºC, a decreased expression of RNAi genes is observed in emerging adults, and this results in an increased susceptibility for infection with the yellow fever virus (YFV) and CHIKV (Adelman *et al.*, 2013).

2.4 Mosquito innate immunity pathways

Several mosquito organs are exposed to the viruses but, differently from vertebrate hosts, mosquitoes seem not to suffer from fitness consequences from viral infection: studies have demonstrated that viral persistence in their body affects neither mosquito behaviour nor its lifespan (Xiao *et al.*, 2015). Distinct innate immune pathways are active during arbovirus infection in insects, primarily the RNA interference (RNAi) pathways, the JAK-STAT, Toll and Imd pathways (Kingsolver *et al.*, 2013; Sim *et al.*, 2014) (Figure 3).
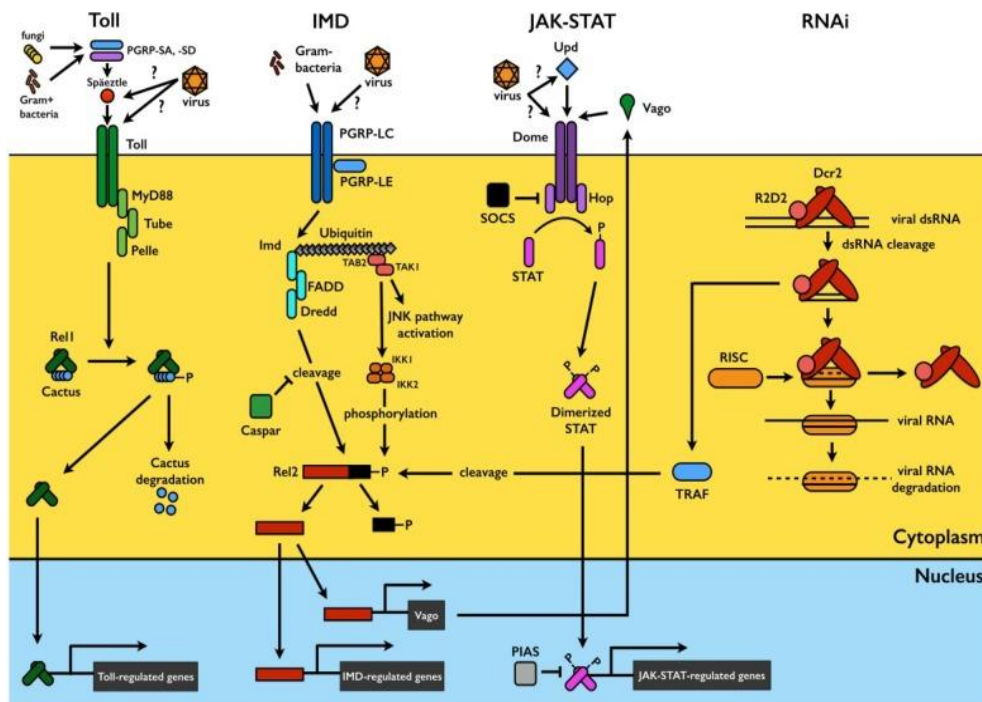
Figure 3. Mosquito immune signaling and RNAi pathways (Sim *et al.*, 2014).

The JAK-STAT and the immune deficiency (Imd) pathways play an important role in the antiviral defence in mosquitoes. Their relevance is due to the high expression of their genes in the midgut, the entry site for arboviruses into mosquito bodies and the first physical barrier to prevent viral dissemination (Souza-Neto *et al.*, 2009).

2.4.1 JAK-STAT, Toll and Imd pathways

These innate immune pathways are signalling cascade pathways that have been shown to limit viral replication in infected mosquitoes primarily by activating the transcription of antimicrobial peptides (AMPs) (Fragkoudis *et*

al., 2008; Xi *et al.*, 2008; Angleró-Rodríguez *et al.*, 2017). AMPs are active against a broad range of pathogens, especially bacteria and parasites (Bartholomay *et al.*, 2004). However, in *Ae. aegypti* cecropins, a class of AMPs, were shown to also exert antiviral effects against DENV and CHIKV (Luplertlop *et al.*, 2011). Similarly, in *Drosophila* RNA synthesis of the Sindbis virus (SINV) is controlled by two AMPs regulated by both the Imd and the JAK-STAT pathways (Huang *et al.*, 2013). Another piece of evidence regarding the antiviral function of these pathways comes from West Nile virus (WNV) infection in *Culex* mosquitoes, where overexpression of the secreted peptide Vago under the regulation of the JAK-STAT pathway reduces viral load (Paradkar *et al.*, 2012).

The JAK-STAT pathway in mosquitoes has been mostly studied for its antiviral role during DENV infection (Souza-Neto *et al.*, 2009). However, interesting results have been obtained with other viruses: for instance, CHIKV is able to inhibit the JAK-STAT pathway (Fros *et al.*, 2010) and Semliki Forest Virus (SFV) downregulates transcription of the JAK-STAT pathway genes (Fragkoudis *et al.*, 2008), indicating that this pathway needs to be avoided by these two viruses in order to establish infection.

On the other hand, the role of the Toll and Imd pathways in mosquito antiviral response is still poorly characterized. Genes in the Toll pathway were found to be upregulated during DENV infection in the salivary glands of *Ae. aegypti* (Luplertlop *et al.*, 2011), in the midgut during Sindbis infection (Sanders *et al.*, 2005) and the secretion of antiviral peptides was found to be activated by the Imd pathway in *Culex* mosquitoes upon WNV infection (Paradkar *et al.*, 2012). However, it was recently suggested that the antiviral activity of the

22

Toll and Imd pathways may not primarily depend on the expression of antiviral effectors. In flies, the Toll pathway is involved in differentiation of hemocytes, phagocytic cells of the hemolymph (Qiu *et al.*, 1998), and the Imd pathway can be pro-apoptotic and activate other antiviral-pathways (Georgel *et al.*, 2001; Sansone *et al.*, 2015).

In conclusion, the JAK-STAT, Toll and Imd pathways are general immune responses active against a variety of pathogens which, to a certain extent, directly or indirectly, are part of the mechanisms employed by mosquitoes and flies against infecting viruses, but they are not specific (reviewed in: Palmer, Varghese and Van Rij, 2018; Lee *et al.*, 2019).

2.4.2 RNAi

RNA interference (RNAi) comprises three different pathways: the microRNA (miRNA) pathway, the short-interfering RNA (siRNA) pathway and the PIWI-interacting RNA (piRNA) pathway (Gangaraju and Lin, 2009) (Figure 4). The miRNA pathway starts with the expression of genome-encoded hairpins called primary-miRNAs (pri-miRNAs) which harbour one or more stem-loops (Bartel, 2004). Stem-loops are recognized and cleaved by the microprocessor complex (Drosha-Pasha) producing precursor miRNAs (pre-miRNA), RNA hairpins about 60 nucleotides in length with a two-nucleotide overhang at the 3' end (Denli *et al.*, 2004). Once exported from the nucleus into the cytoplasm, the pre-miRNAs loop is cleaved by Dicer-1, leaving an RNA heteroduplex with 3' overhangs on both strands. Only one of the strands
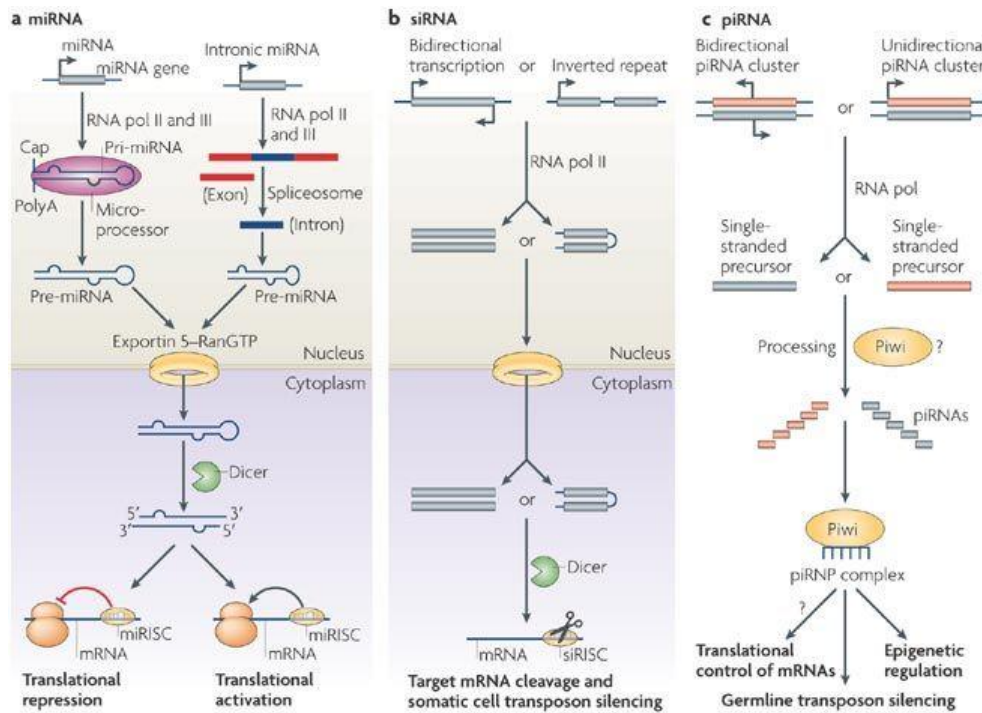
Figure 4. Biogenesis and regulatory features of the miRNA, siRNA and piRNA pathways (Gangaraju and Lin, 2009).

is then loaded onto Argonaute-1 via a not fully understood selection mechanism, thus forming the miRNA induced silencing complex (miRISC) (Förstemann *et al.*, 2007). The miRNA guides this complex via base-pairing and canonically targets endogenous messenger RNAs, mainly in the 3' UTR regions, and regulates targets expression by either cleaving the mRNA or hampering its translation (Bartel, 2009). Recently, it has been suggested that this pathway may be also involved in antiviral defences (reviewed in Monsanto-Hearne and Johnson, 2018).

Among the three RNA-interference mechanisms, the siRNA pathway is

considered the main antiviral defence in mosquitoes and has been widely characterized. This pathway is triggered by the detection of double-stranded RNA molecules in the cytoplasm which are then cleaved by Dicer enzymes into ~21 nucleotide long duplexes with two-nucleotide 3' overhangs on both strands as for microRNAs (Elbashir *et al.*, 2001). The thermodynamically more stable strand is then complexed with the protein Argonaute-2 (Ago2) into the RNA-induced silencing complex (RISC) (Tomari *et al.*, 2004) where final maturation of the siRNA occurs via 2'-O-methylation at the 3' terminal nucleotide (Horwich *et al.*, 2007). As for miRNAs, the RISC-loaded siRNA guides the complex via Watson-Crick base pairing to cleave target RNA molecules (Sinha *et al.*, 2018). Interestingly, this mechanism was shown to occur also for genome-encoded double-stranded RNA molecules (endo-siRNAs) that have been implicated in transposon control and gene expression regulation (Ghildiyal *et al.*, 2008). However, dsRNA in healthy cells are usually not abundant, suggesting that this is a secondary function of the siRNA pathway with respect to its antiviral role (Wang *et al.*, 2006).

Knock-down of the core components of the siRNA pathway (Dcr2, r2d2 and Ago2) in *Ae. aegypti* results in increase of DENVs replication and reduction of the EIP (Sánchez-Vargas *et al.*, 2009). Similarly, silencing Ago2 and Dcr2 in *Ae. aegypti* increases the replication of SINV, as virus-specific siRNAs cannot be produced to fight infection (Campbell *et al.*, 2008b).

The siRNA pathway genes are among the fastest evolving 3% of all the fly genes, suggesting a positive selection during coevolution with RNA viruses (Obbard *et al.*, 2006). Rapid and positive selection of these genes was confirmed in *Ae. aegypti* (Bernhardt *et al.*, 2012), where specific Dcr2

genotypes were associated with resistance to specific DENV2 genotypes (Lambrechts *et al.*, 2013).

The piRNA pathway has been mostly studied in *D. melanogaster*, where it regulates transposon movement during germ-line formation (Brennecke *et al.*, 2007a). Three proteins are the active components of the piRNA pathway in *D. melanogaster*: Aubergine (AUB), PIWI and Argonaute 3 (AGO3). PIWI and AUB associate with piRNAs that have an uracil at their 5' ends, whereas AGO3 associates with piRNAs that have an adenine as 10th nucleotide (Brennecke *et al.*, 2007b). Furthermore, AUB- and AGO3-associated piRNAs are complementary for the first 10 nt. The single-stranded piRNA precursors are transcribed from specialized genomic regions called piRNA clusters (PIRCs) and exported from the nucleus into the cytoplasm to be further processed. A candidate for the first cleavage of piRNA precursors is the endonuclease Zucchini (ZUC) (Stein *et al.*, 2019) that generates the uracil-biased 5' end, which is required to load the piRNA precursor into PIWI and AUB proteins (Ipsaro *et al.*, 2012). Some other factors are probably involved in this step, like Armitage (ARMI) and the Tudor-domain protein, Yb (Olivieri *et al.*, 2010). Once loaded, an unknown 3' > 5' exonuclease trims the precursor's 3' end. The final product undergoes a 2' O-methylation process at the 3' end by the piRNA methyl-transferase HEN1, which stabilizes the piRNA (Kawaoka *et al.*, 2011). PIWI-piRNA complexes re-enter the nucleus where target repression is performed by target gene silencing (TGS), based on sequence complementarity. On the contrary, AUB-piRNA complexes perform post-transcriptional gene silencing (PTGS) by targeting transposon mRNAs in the perinuclear cytoplasm (Brennecke *et al.*,

26

2007b). AUB cleavage of the target mRNAs leads to the formation of sense secondary piRNAs, which are further processed and loaded onto AGO3. AGO3-piRNA complexes then recognize cluster transcripts and cleave them, thus amplifying the silencing mechanism by generating even more antisense piRNAs to be loaded onto PIWI and AUB (Brennecke *et al.*, 2007b).

Sequencing of small RNAs following infection of *Ae. aegypti* and *Ae. albopictus* with SINV and CHIKV revealed the presence of piRNAs derived from viral sequences and showing the typical ping-pong amplification features (Morazzani *et al.*, 2012; Vodovar *et al.*, 2012). These viral-derived piRNAs were called vpiRNAs. Furthermore, the *Piwi* gene family has undergone expansion in *Aedes* and *Culex* mosquitoes, leading to the intriguing possibility of functional specialization of different Piwi proteins.

Recent studies showed that viral DNA fragments are synthetized in both circular and linear forms, in complexes comprising sequences from retrotransposons, but the mechanisms underlying their mode of action are still under investigation (Palatini *et al.*, 2017; Poirier *et al.*, 2018). We and others recently reported the presence of fragmented viral sequences inserted in the genomes of *Aedes spp.* mosquitoes. These sequences were found in close proximity to transposon sequence; moreover, they are enriched in PIRCs and produce PIWI-interacting RNAs (piRNAs) (Palatini *et al.*, 2017; Whitfield *et al.*, 2017). The similarities observed between vDNAs and viral integrations, and the evidence that piRNAs of viral origin (vpiRNAs) are produced during arboviral infection led to the hypothesis that the piRNA pathway cooperates with the siRNA pathway in the acquisition of tolerance to infection (Miesen *et al.*, 2016b; Olson and Bonizzoni, 2017; Palatini *et al.*, 2017).

2.5 A closer look into the pi-RNA pathway of *Aedes* spp. mosquitoes

The PIWI-interacting RNA pathway is the most-recently identified of the three RNAi pathways and it has been extensively studied in *D. melanogaster* where it contributes in silencing transposable elements (TEs) in gonadal tissues (Brennecke *et al.*, 2007b). piRNA clusters are programmable in that they are made of fragments of TE sequences from previous infections and they can acquire novel TE sequences upon novel TE invasions (Zanni *et al.*, 2013; Guida *et al.*, 2016). As such, the piRNA pathway is considered a *bona fide* adaptive immunity system functionally analogous to the prokaryotic Clustered Regulatory Interspaces Palindromic Repeats (CRISPR) and CRISPR-associated proteins (CAS) (Koonin, 2017). Differently than *D. melanogaster*, besides being active in the germline to suppress TE movement, the piRNA pathway of *Aedes* spp. mosquitoes has evolved to exert antiviral activity against different arboviruses (Miesen *et al.*, 2016b) (Figure 5). In *Aedes* spp. mosquitoes, piRNA clusters were shown to contain sequences of viral origin (Olson and Bonizzoni, 2017) and the *Piwi* gene family has undergone expansion: in *Ae. aegypti*, seven proteins (i.e. Ago3, Piwi2, Piwi3, Piwi4, Piwi5, Piwi6 and Piwi7) are alternatively expressed in somatic and germline cells and interact with both endogenous piRNAs and vpiRNAs (Akbari *et al.*, 2013; Miesen *et al.*, 2016b; Matthews *et al.*, 2018).
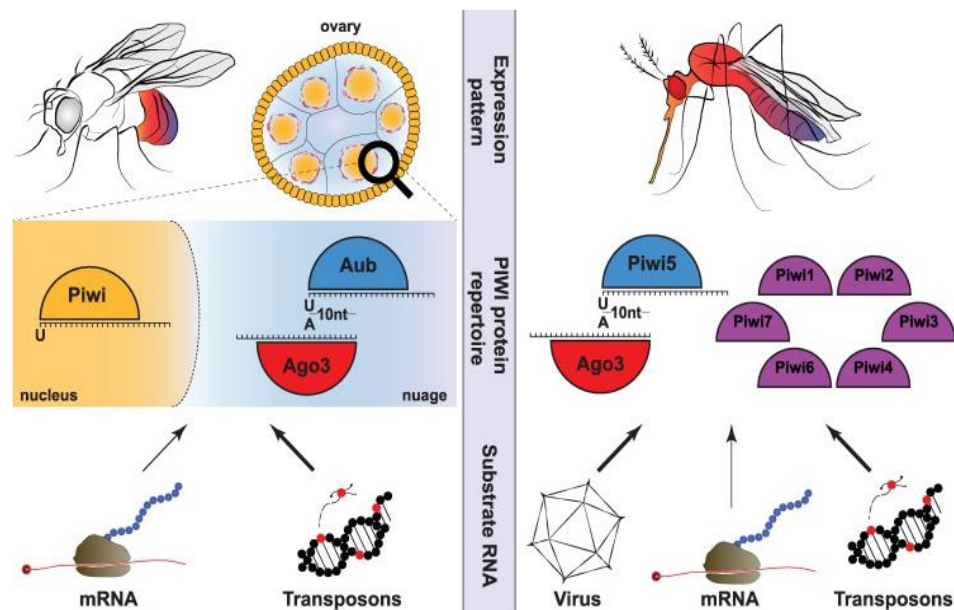
Figure 5. Differences between the composition and activity of the piRNA pathway in *Drosophila* and *Aedes* spp. (Miesen *et al*., 2016)

In the *Ae. aegypti* cell line Aag2, *Piwi1/3* is mainly expressed in gonadal tissues, *Piwi7* appears to be embryonic (Akbari *et al.*, 2013), while *Ago3, Piwi4, Piwi5 and Piwi6* are primarily expressed in somatic tissues and are related to the production of transposon-derived piRNAs (Akbari *et al.*, 2013; Miesen *et al.*, 2016a). *Ago3* and *Piwi5* also regulate biogenesis of piRNAs from the replication-dependent histone gene family (Girardi *et al.*, 2017). In Aag2 cells infected with the alphaviruses CHIKV, SINV and SFV, *Piwi5* and *Ago3* were shown to be the main PIWI proteins involved in vpiRNAs production, and Piwi6 was also elicited following infection with the DENV2 (Miesen *et al.*, 2015, 2016a; Varjak *et al.*, 2017, 2018).

Recent evidence also indicate that Piwi4 bridges between the siRNA and

piRNA pathways: Piwi4 does not bind piRNAs and its knock-down does not alter vpiRNA production upon infection of Aag2 cells with either SFV or DENV2 (Schnettler *et al.*, 2013; Miesen *et al.*, 2016a). However, PIWI4 coimmunoprecipitates with components of the siRNA pathway such as Ago2, Dcr2, and with Piwi5, Piwi6 and Ago3 (Varjak *et al.*, 2017). Despite these studies support an antiviral role for the piRNA pathway in *Aedes* spp. mosquitoes, the individual functions of each PIWI protein still needs to be investigated. Sites under positive selection in genes that have undergone duplication is usually a sign of the acquisition of novel functions (Hahn, 2009) and, if found, may indicate that the *Piwi* genes have undergone or may still be undergoing functional specialization. Additionally, under the "arm-race theory", rapid evolution is expected for genes with immunity functions because their products should act against fast evolving viruses (Obbard *et al.*, 2006).

2.6 Nonretroviral integrated RNA virus sequences (NIRVS)

All arboviruses, with the exception of the Swine fever virus, are nonretroviral RNA viruses, meaning they have an RNA-based genome and do not code for reverse transcriptase and integrase**.** Another group of RNA viruses named insect-specific viruses (ISVs) have been recently identified (Bolling *et al.*, 2015). ISVs have been shown to be phylogenetically related to arboviruses belonging to the families *Flaviviridae, Togaviridae, Rhabdoviridae* and *Bunyaviridae* and recent analyses suggest that in some viral families, ISVs may be precursors of arboviruses (Marklewitz *et al.*, 2015; Moureau *et al.*, 2015). ISVs infect mosquitoes persistently and are likely to be vertically

transmitted by infecting the ovaries (Bolling *et al.*, 2015; Marklewitz *et al.*, 2015; Moureau *et al.*, 2015). Despite the fact that arboviruses and ISVs are nonretroviral RNA viruses and therefore lack the machinery necessary for integration, recent studies have found evidence of sequences from these viruses integrated into eukaryotic genomes, including those of mosquitoes (Blair *et al.*, 2019). These viral sequences are called Non Retroviral Integrated RNA Virus Sequences (NIRVS). NIRVS presence in mosquitoes have been evaluated by different research groups with different pipelines and queries, and they all consistently found that the arboviral vectors *Ae. aegypti* and *Ae. albopictus* harbor a larger amount of NIRVS when compared to *Cx. quinquefasciatus* and *Anophelinae* mosquitoes (Fort *et al.*, 2012; Chen *et al.*, 2015; Lequime and Lambrechts, 2017). We contributed to this result by surveying all the mosquito genomes for the presence of integrations form 424 nonretroviral RNA viruses. We observed that NIRVS are enriched in *Ae. aegypti* and *Ae. albopictus,* but are rare in *Anophelinae*, which are primarily parasitic vectors (Palatini *et al.*, 2017) (Figure 6). *In silico* analyses also revealed the presence of ISVs integrated sequences in *Ae. aeygypti, Ae. albopictus, An. sinensis* and *An. mininums* which encompass viral ORFs actively transcribed (Lequime and Lambrechts, 2017; Palatini *et al.*, 2017). Little is currently known about the molecular mechanisms that lead to NIRVS integrations, but since NIRVS are found next to transposable elements, it has been suggested that the integration of viral sequences depends on the activity
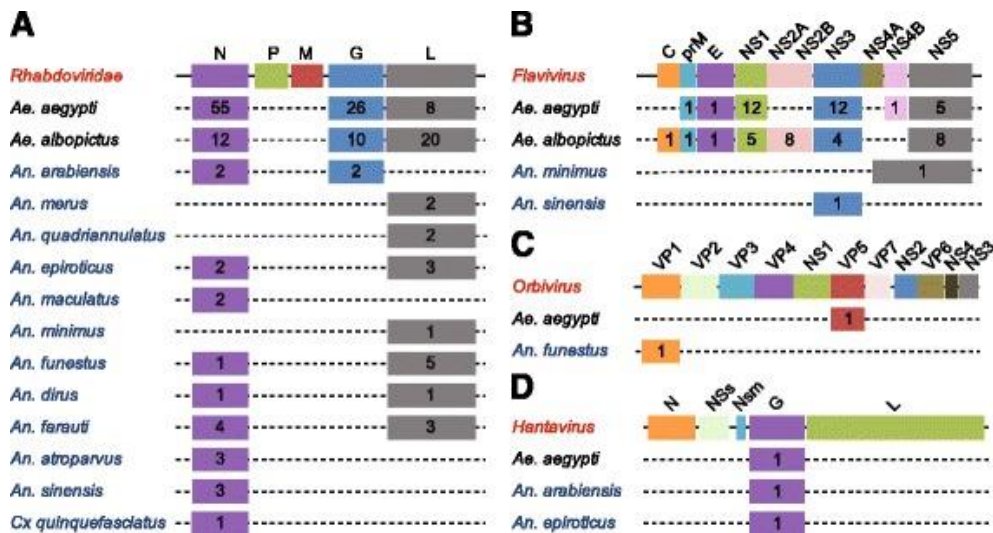
Figure 6. Different abundance of NIRVS across virus genera, genes and host species (Palatini *et al*., 2017)

of the machinery encoded by retrotransposons (Geuking *et al.*, 2009).

Recent studies showed that viral DNA fragments are synthetized in both circular and linear forms, in complexes comprising sequences from retrotransposons, upon infection of cells and mosquitoes with nonretroviral RNA viruses (Goic *et al.*, 2016; Nag *et al.*, 2016). vDNAs are produced by the DExD/H helicase domain of Dcr2, a protein of the siRNA pathway, in complex arrangements with TE sequences (Goic *et al.*, 2016; Poirier *et al.*, 2018). vDNA fragments were shown to be present in several body parts in adult mosquitoes, including wings and legs, hinting that either vDNA can be formed in any infected cell, or that vDNA can be released from the cells to disseminate in the mosquito body as a form of inter-cellular communication (Yáñez-Mó *et al.*, 2015; Goic *et al.*, 2016). These observations and the

similarity in the organization between vDNAs and NIRVS led to the hypothesis that the piRNA pathway cooperates with the siRNA pathway in the acquisition of tolerance to infection (Miesen *et al.*, 2016b; Olson and Bonizzoni, 2017; Palatini *et al.*, 2017). For instance, we observed that NIRVS are enriched in piRNA clusters in both the *Ae. aegypti* and *Ae. albopictus* genomes and produce piRNAs with a typical bias for uridine at position 1 and predominantly in antisense orientation to the predicted viral open reading frame (ORF) (Palatini *et al.*, 2017). Additionally, production of viral derived piRNAs (vpiRNAs) and siRNAs (vsiRNAs) from vDNA favoring the establishment of persistent infection was shown in *Ae. aegypti* CHIKV-infected cells and adult mosquitoes (Goic *et al.*, 2016), and vpiRNAs are found in *Ae. aegypti* following both flavivirus and alphavirus infection (Miesen *et al.*, 2016b). Overall, these elements establish the potentials for NIRVS derived piRNAs and siRNAs to target viral mRNA contributing to mosquito immunity to viruses and/or to the establishment of persistent infection, the so-called "viral accommodation" (Flegel, 2007; Flegel and Sritunyalucksana, 2011). Alternatively, NIRVS encompassing a complete viral ORF and being transcriptionally active NIRVS could produce mRNAs that compete in *in trans* with infecting viruses, preventing viral replication, transcription or assembly (Park *et al.*, 2013; Palatini *et al.*, 2017).

Interestingly, NIRVS presence has been associated with immunity to subsequent infection with cognate viruses in squirrels (Fujino *et al.*, 2014) and bees (Maori *et al.*, 2007) and to viral persistence in shrimps (Flegel, 2007). In the context of mosquitoes, an effect of NIRVS on mosquito immunity would be significant because it could alter its vector competence in

opposite direction depending on whether it leads to immunity to subsequent infections of cognate viruses or viral tolerance. Thus, it is important to understand under which circumstances NIRVS are formed and if they impact mosquito biology. Determining whether such differences in viral origin of NIRVS may lead to different physiological or genetic roles is also an intriguing subject for future studies.

2.7 The genome sequence of *Ae. albopictus*

The first genome sequence and assembly of the Asian tiger mosquitoes was made public in 2015 (Chen *et al.*, 2015). This analysis was based solely on a deep collection of short DNA sequence reads and the genome of *Ae. albopictus* appears to be the largest mosquito genome sequenced to date. However, due to very high levels of repetitive DNA and reliance on short-read sequencing, the initial assembly of *Ae. albopictus* from the Foshan strain remains highly fragmented. The laboratory in which I conduct my PhD program, in collaboration with other groups world-wide, have recently engaged in the re-sequencing of the genome of *Ae. albopictus* using long-read and Hi-C technologies (under preparation). Briefly, starting from 2 pools of 40 mosquitoes of an inbred strain derived from the Foshan strain, we generated 82 Gb of PacBio and 135 Gb of Hi-C sequence. A first draft assembly of this data increased the median contig size 60-fold versus the prior short-read assembly, yielding a contig NG50 size over 1 Mbp and chromosome-sized scaffolds, representing an enormous leap in assembly quality for this important vector. In this assembly, we re-annotated NIRVS

and found 485 NIRVS homogeneously distributed across scaffolds (Figure 7). These NIRVS were assigned to eight different taxonomic categories based on sequence similarity to viral species, including Flavi-NIRVS, Rhabdo-NIRVS and Chuviridae-Like NIRVS. NIRVS length vary from hundreds to thousands of base pairs, with a median size of 1881bp. Flavi-NIRVS and Rhabdo-NIRVS include the longer NIRVS, up to 6593bp. The majority (74.6%) of the identified NIRVS are similar to viral structural proteins, while only a few (6.4%) show similarity to viral non-structural proteins. Flavi-NIRVS show similarity to different insect-specific flaviviruses and, when they are all mapped to a representative *Flavivirus* genome, they span the
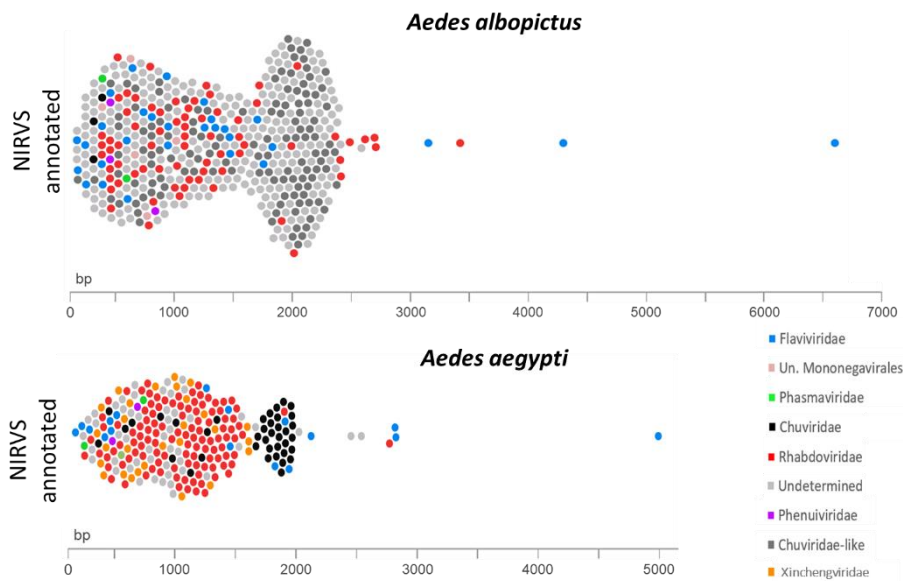
Figure 7. The landscape of NIRVS in the genome of *Aedes albopictus* (new Canu genome assembly) and *Aedes aegypti* (AaegL5). Each dot represents a viral integration according to its length and color-coded based on its viral origin.

35

entire length of the sequence. As previously observed, NIRVS are enriched in piRNA clusters and, for the most part, associate with TEs (Palatini *et al*., 2017). A total of 55% of NIRVS have a TE both upstream and downstream, while 26% have only one TE on either end. During my PhD program, I used the newly generated *Ae. albopictus* assembly to assess piRNAs and siRNAs production from NIRVS after viral infection with CHIKV and DENV1.

3. Aims of the research

The Asian tiger mosquito *Ae. albopictus* is the main arboviral vector in temperate regions of the world. For instance, its presence in Europe has been associated with autochthonous cases of Dengue and Chikungunya in France, Croatia and Italy since the early 2010.

Because there are no effective, approved vaccines or specific treatments for most arboviruses, the major manner in which arboviruses are controlled is through control of the vectors. As traditional methods such as insecticides have become less effective (resistance) or acceptable (environmental harm), alternative methods of control are being pursued, the most promising of which involve genetic manipulation of mosquitoes. This may be through reducing population sizes or genetically modifying mosquitoes to make them less competent to transmit pathogens such as viruses. The understanding of the genetic factors related to viral acquisition, dissemination and transmission by the vector, all of which contribute to the so-called vector competence, is a prerequisite to develop genetic-based strategies of vector control.

In my PhD program I developed three aims to further our understanding of mosquito immunity, with a focus on the piRNA pathway and sequences of viral origin integrated into mosquito genomes.

The piRNA pathway is the most-recently identified of the small RNA-based pathways within RNAi and it has been mostly studied in the model insect *D. melanogaster* where it functions through small RNAs called piRNAs in germ cells to silence TEs and preserve genome integrity. Recent literature supports the involvement of the piRNA pathway in antiviral immunity in *Aedes* spp. mosquitoes, differently, then in *D. melanogaster*. However, the genes and the

function of this pathway in *Ae. albopictus* is still enigmatic. As a consequence, my first aim is to characterise the genes and the function of this pathway in *Ae. albopictus*.

piRNAs are produced from genomic loci called piRNA clusters. Previous work in my laboratory showed that piRNA clusters of *Aedes* spp. mosquitoes are statistically-significantly enriched in sequences of viral origin embedded within transposable element (TE) sequences and producing piRNAs. This observation led to the hypothesis that viral integrations may constitute an archive of past viral infection, like fragments of TEs in piRNA clusters, and contribute to immunity towards subsequent infections with cognate viruses. The second aim of my PhD program addresses this question by studying the small RNA expression profile of viral integrations and viruses after an infection with either Dengue virus or Chikungunya.

Finally, because immunity pathways are related to mosquito fitness, I want to start understanding differences in life history trains of different *Ae. albopictus* strains derived from geographically different populations.

4. Materials and methods

4.1 The piRNA pathway of *Ae. albopictus*

4.1.1 Mosquitoes

Aedes albopictus mosquitoes of the Foshan strain were reared under constant conditions, at 28 °C and 70-80% relative humidity with a 12/12 h light/dark cycle. Larvae were reared in plastic containers at a controlled density to avoid competition for food. Larvae were fed daily with fish food (Tetra Goldfish Gold Colour). Emerged adults were reared in 30 cm^3 cages and fed with cotton soaked in a solution of 0.2 g/ml sucrose. Adult females were fed with defibrinated mutton blood (Biolife Italiana) using a Hemotek blood feeding apparatus. Mosquitoes from Mexico and La Reunion island were field collected in 2017, maintained in 70% ethanol and then shipped to Italy after having established appropriate Material Transfer Agreements. All samples were processed at the University of Pavia.

4.1.2 Mosquito infections

Infection experiments were carried out at the Institut Pasteur, France. I infected mosquitoes with either DENV serotype 1, genotype 1806, or CHIKV 06.21. These viruses were provided by the French National Reference Center for Arboviruses at the Institut Pasteur. DENV-1 (1806) was isolated from an autochthonous case from Nice, France in 2010 (La Ruche *et al.*, 2010). CHIKV 06-21 was isolated from a patient on La Reunion Island in 2005

(Schuffenecker *et al.*, 2006).

Four boxes containing 60 one-week-old females were given infectious blood-meal composed by 2 mL of washed rabbit red blood cells, 1 mL of viral suspension and 5 mM of ATP, for a titer of 10^7 PFU/mL for CHIKV and $10^{6.8}$ PFU/mL for DENV. Engorged females were selected and kept under standard conditions until analysis. Control mosquitoes were fed with uninfected blood-meal or kept on a sugar-diet and grown in the same conditions. Thirty mosquitoes from infected and non-infected batches were dissected at days 4 and 14 post-infection (pi) for CHIKV, and at days 4 and 21 dpi for DENV. 14 and 21 dpi represent late infection times for CHIKV and DENV, respectively, based on their different infection kinetics (Vazeille *et al.*, 2007; Zhang *et al.*, 2010). Ovaries and carcasses were pooled from 15 individuals each (Figure 8).

To estimate transmission, saliva was collected from individual infected mosquitoes as previously described (Dubrulle *et al.*, 2009). Saliva was retrieved by inserting the proboscis into a 20 µL tip containing 5 µL of Fetal Bovine Serum (FBS) (Gibco, MA, USA) for 30 minutes. FBS containing saliva was expelled in 45 µL of Leibovitz L15 medium (Invitrogen, CA, USA) and used for the following titration. The number of infectious particles in saliva was estimated by focus fluorescent assay on C6/36 *Ae. albopictus* cells. Serial dilutions were incubated at 28°C for 3 days (CHIKV) and 5 days (DENV), plates were stained using hyperimmune ascetic fluid specific to CHIKV or DENV-1 as primary antibody. A Fluorescein-conjugated goat anti-mouse was used as the second antibody (Biorad). Viral titers were 16,266±50,446 FFU and 155±125 FFU for CHIKV at 14 dpi and DENV at

21 dpi, respectively.
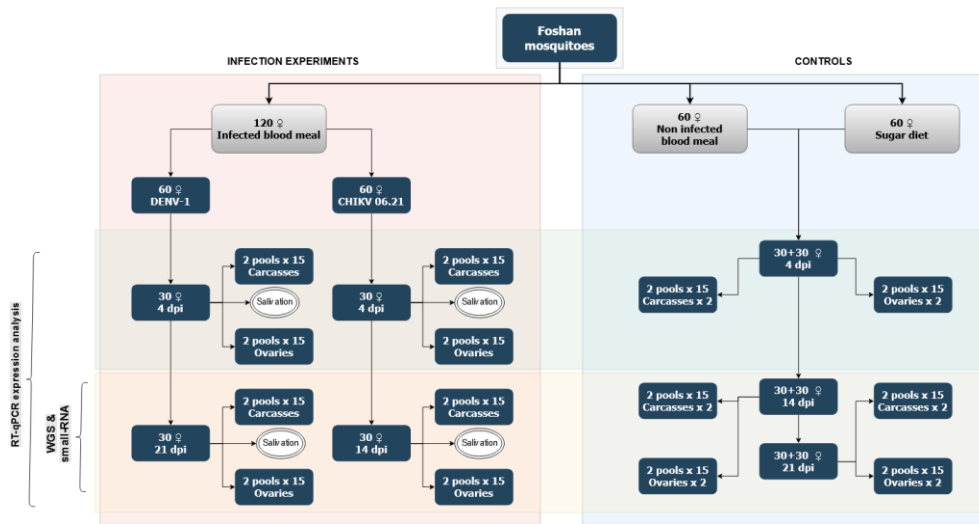
Figure 8. Scheme of the infection experiments. Colour code: red = infection experiments; blue = controls; green = early infection; yellow = late infection. RT-qPCR analysis on the *Piwi* genes expression patterns were performed for both early and late infection data, while WGS and small-RNA data analyses were performed only for the late infection stages.

4.1.3 Bioinformatic identification of *Piwi* genes in the genome of *Ae. albopictus*

The sequences of the *Ae. aegypti Piwi* genes (Campbell *et al.*, 2008a) were used as query to find orthologs in the reference genome of the *Ae. albopictus* Foshan strain (AaloF1 assembly) and in the genome of the *Ae. albopictus* C6/36 cell line (canu_80X_arrow2.2 assembly) using the BLAST tool in Vectorbase (www.vectorbase.org).

Inferred coding sequences (CDS) where analysed in Prosite (Prosite.expasy.org/prosite.html) to confirm the presence of hallmark PAZ and PIWI domains of Argonaute proteins (Arensburger *et al.*, 2011).

4.1.4 Copy number of *Piwi* genes

qPCR reactions were performed using the QuantiNova SYBR Green PCR Kit (Qiagen) following the manufacturer's instructions on an Eppendorf Mastercycler RealPlex4. Genomic DNA was extracted from four adult mosquitoes using DNA Isolation DNeasy Blood & Tissue Kit (Qiagen) and used for the amplification reactions with gene-specific primers whose efficiency had been previously verified. Estimates of gene copy number were performed based on the $2^{-\Delta CT}$ method using the inferred *Piwi6* and the para sodium channel gene (AALF000723) as references (Yuan *et al.*, 2007).

4.1.5 Structure of *Piwi* genes

The genomic sequence of each bioinformatically-identified *Piwi* gene was confirmed by PCR using DNA extracted from whole mosquitoes and dissected ovaries (Baruffi *et al.*, 1995) as templates. Primers were designed to amplify each exon and distinguish between paralogous *Piwi* genes. The DreamTaq Green PCR Master Mix (Thermo Scientific) was used for PCR reactions with the following parameter: 94 °C for 3 minutes, 40 cycles at 94 °C for 30 sec, 55 °C-62 °C for 40 sec, 72 °C for 1-2 minutes and final extension step of 72 °C for 10 minutes. PCR products were confirmed by gel electrophoresis using 1% agarose gels stained with ethidium bromide and visualized under UV light. Positive results were cloned using the TOPO® TA Cloning® Kit strategy (Invitrogen) following the manufacturer's instructions, DNA plasmids were purified using the QIAprep Spin Miniprep Kit and sequenced by Sanger sequencing.

4.1.6 *Piwi* gene transcript sequences and phylogeny

The transcript sequence of each *Piwi* gene was confirmed by RT-PCR

reactions using a High Fidelity taq-polymerase (Platinum SuperFi DNA Polymerase, Invitrogen) following manufacturer's instructions. RNA was extracted from pools of 5 adult female mosquitoes using a standard TRIzol protocol. Sets of primers were designed for each gene to amplify its entire transcript sequence. PCR products were confirmed, transformed, purified and sequenced as previously described. Rapid amplification of cDNA ends (RACE) PCRs were also performed using FirstChoice RLM-RACE Kit (Thermo Fisher Scientific) to analyse 5' and 3' ends of the transcript sequences following manufacturer's instructions.

The obtained transcript sequences of the identified *Ae. albopictus Piwi* genes were aligned with MUSCLE (Edgar, 2004) to sequences of transcripts from the *Piwi* genes annotated in *Culicidae* species and *D. melanogaster*, as downloaded from VectorBase (www.vectorbase.org). Maximum-likelihood based phylogenetic inference was based on RAxML after 1000 bootstrap resampling. The resulting tree was visualised using FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

4.1.7 Northern Blot analysis

For each analysis, 10µg of total RNA were extracted from a pool of 10 sugar-fed females and run in a 1% x 2% agarose/formaldehyde gel (1 g agarose, 10 ml 10x MOPS buffer, 5.4 ml 37% formaldehyde, 84.6 ml DEPC water). Gel blotting was performed following two washes in 20x SSC for 15 minutes and the RNA transferred to an Amersham Hybond-N+ nylon membrane (GE healthcare) using 20x SSC. The RNA was then cross-linked to the membrane using UV light exposure for 1 minute. Biotin-High Prime (Roche) was used for probe labelling. Hybridization and detection of biotinylated probes was

43

performed using the North2South™ Chemiluminescent Hybridization and Detection Kit (Thermo Fisher Scientific) following manufacturer's instructions.

4.1.8 Polymorphisms of *Piwi* genes

High Throughput Sequencing (HTS) data of 56 individual mosquitoes were obtained from Whole genome sequencing libraries sequenced on the Illumina HiSeqX platform (PE 150) at the Genomics Laboratory of Verily in South San Francisco. Of these 56 mosquitoes, 24 were from Mexico, 16 were from the island of La Reunion island and 16 from the reference Foshan strain. We investigated *Piwi* gene polymorphism by looking at the distribution of single nucleotide polymorphism, assessed sign of adaptive evolution and estimated their level of polymorphism in comparison to fast and slow evolving genes.

Illumina reads were mapped to *Piwi* gene transcript sequences using Burrows-Wheeler Aligner (BWA-MEM) (Li, 2013). Polymorphisms was tested by Freebayes (Garrison and Marth, 2012) and annotated with snpEff (Cingolani *et al.*, 2012). Plots of frameshift and non-synonymous variants were performed with muts-needle-plot (Schroeder and Lopez-Bigas, 2015). Venn diagrams of positions displaying polymorphisms in the three strains were built using Venny 2.1 (Oliveros, 2016). Haplotype reconstruction was performed using seqPHASE (Flot, 2010) and PHASE (Stephens *et al.*, 2001; Stephens and Scheet, 2005). Haplotypes were then analysed in DnaSP (Librado and Rozas, 2009), estimating the number of segregating sites and the level of nucleotide diversity π (Tajima, 1983) in both synonymous and non-synonymous sites. We manually computed the nucleotide diversity estimator θ (Watterson, 1975) and Tajima's *D* statistic (Tajima, 1989).

Signatures of adaptive evolution were assessed using the McDonald-Kreitman test (McDonald and Kreitman, 1991), including the orthologous sequences of the Piwi genes from *Ae. aegypti*.

Haplotypes from each individual were also aligned in TranslatorX (Abascal *et al.*, 2010) using Clustalw (Thompson *et al.*, 1994) and used for Maximum-likelihood based phylogenetic inference based on RAxML after 1000 bootstrap under the GTRGAMMA model. Signs of selective pressure among populations (Bernhardt *et al.*, 2012) were investigated with Codeml in PAML v. 4.9 under codon-based models (Yang, 2007). I opted for the M1a (nearly-neutral) model to compare to the M2a (positive selection) model and inferred ω estimations and posterior probabilities under the Bayes empirical Bayes (BEB) approach (Yang, 2007).

The overall level of polymorphism (LoP) for slow-evolving genes (SGs) (AALF008224, AALF005886, AALF020750, AALF026109, AALF014156 , AALF018476, AALF014287, AALF004102, AALF003606, AALF019476, AALF028431, AALF018378, AALF027761, AALF014448), fast-evolving genes (FGs) (AALF010748, AALF022019, AALF024551, AALF017064, AALF004733, AALF018679, AALF028390, AALF026991, AALF014993, AALF009493, AALF010877, AALF012271, AALF009839, AALF019413) and the *Piwi* genes was calculated for each population following the pipeline as described in Pischedda *et al.*, (2019). Briefly, SNPs and INDELs were inferred using four Variant callers (i.e. Freebayes (Garrison and Marth, 2012), Platypus (Rimmer *et al.*, 2014), Vardict (Lai *et al.*, 2016) and GATK UnifiedGenotyper (McKenna *et al.*, 2010)) and the results from each were merged with custom scripts. The LoP for each individual was defined as the

number of variants averaged over the region length, and we used the median value of each population for subsequent statistical analyses in R studio (Rstudio Team, 2016). Fold-change differences were calculated as the ratio of the median LoP for each *Piwi* gene and each FG gene over the median LoP of the SG genes. Statistical differences in LoP distribution was assessed using the Kolmogorov-Smirnov test after Bonferroni correction for multiple testing.

4.1.9 Homology modelling

In collaboration with the laboratory of prof. Federico Forneris (Department of Biology and Biotechnology, University of Pavia) computational structural investigations were carried out initially through the identification of the closest homologs based on sequence identity (using *NCBI Blast* (Altschul *et al.*, 1990)) and secondary structure matching (using *HHPRED* (Zimmermann *et al.*, 2018)). Homology model were then generated *MODELLER* (Webb and Sali, 2017) using on the structures *Kluyveromyces polysporus* Argonaute with a guide RNA (PDB ID 4F1N), Human Argonaute2 Bound to t1-G Target RNA (PDB ID 4z4d) (Schirle *et al.*, 2015), *T. thermophilus* Argonaute complexed with DNA guide strand and 19-nt RNA target strand (PDB ID 3HM9), and silkworm PIWI-clade Argonaute Siwi bound to piRNA (PDB ID 5GUH).

Computational models were manually adjusted through the removal of non-predictable N- and C-terminal flexible regions using *COOT* (Emsley *et al.*, 2010) followed by geometry idealization in *PHENIX* (Adams *et al.*, 2011) to adjust the overall geometry. Final model quality was assessed by evaluating average bond lengths, bond angles, clashes, and Ramachandran statistics using Molprobity (Williams *et al.*, 2018) and the *QMEAN* server (Benkert *et*

al., 2009) Structural figures were generated with *PyMol* (Schrödinger, 2015).

4.1.10 Developmental expression profile of *Piwi* genes

Piwi genes expression throughout developmental stages of *Ae. albopictus* were searched for in publicly available RNA-seq data (runs: SRR458468, SRR458471, SRR1663685, SRR1663700, SRR1663754, SRR1663913, SRR1812887, SRR1812889, SRR1845684), which were downloaded, aligned to the current *Ae. albopictus* genome assembly (AaloF1) using Burrows-Wheeler Aligner (BWA-MEM) (Li, 2013) and visualized in Integrative Genomics Viewer (IGV) (Robinson *et al.*, 2011).

Quantitative RT-PCRs (qRT-PCR) were performed on total RNA extracted from embryos, 1st and 4th instar larvae, pupae, and adults using TRIzol (Thermo Fisher Scientific), as previously described. Embryos consisted of two pools of 60 eggs at different time points after oviposition (i.e. 4-8 h, 8-16 h and 16-24 h). Adult samples consisted of males and females kept on a sugar-diet, females 48h after a blood-meal and ovaries of both sugar-fed and blood-fed (48h) females. We used two biological replicates per condition and the RPL34 gene as housekeeping (Reynolds *et al.*, 2013). For each stage, RNA was extracted from pools of 10-15 mosquitoes, except for first instar larvae and embryos when 20 or 60 individuals were used, respectively. Relative quantification of *Piwi* genes was determined using the software qBase+ (Biogazelle). Expression values were normalized with respect to sugar-fed females.

4.1.11 Expression analyses following infection

Fold-change expression values for each *Piwi* gene was assessed for non-

infectious-blood-fed controls, CHIKV-infected and DENV-infected samples after normalization on sugar-fed controls. qRT-PCR experiments were set up for two replicate pools of 15 ovaries and 15 carcasses at days 4, 14 and 4, 21 for CHIKV and DENV, respectively and the corresponding sugar and non-infectious-blood controls. RNA extraction, qRT-PCR and data analyses were performed as described in the previous paragraph (see "Developmental expression profile of Piwi genes"). Fold-change differences significance was assessed using the Analysis of Variance (ANOVA) procedure (Khan and Rayner, 2004; Blanca *et al.*, 2017) as implemented in qBASE+.

4.1.12 Temperature experiments

To assess the differential expression of *Piwi* genes caused by temperature, *Ae. albopictus* mosquitoes were grown at different temperatures from the hatching of eggs until adult emergence. The rest of parameters (i.e. humidity and light cycle) were maintained as standard condition to isolate the effect of temperature on our results. The selected control temperature was 28ºC, near to the 27ºC set as optimum for *Ae. albopictus* growing (Delatte *et al.*, 2009). Eggs were counted and moved to a plastic container inside a thermo-regulable incubator. Equal numbers of eggs, near 370, were incubated for each tested temperature. The higher temperature of 32ºC was selected since it is in the range of the maximum temperatures observed during summer in Pavia. Following the same criteria, the lower selected temperature was 18ºC, which corresponds to the minimum temperatures registered in the same period (Figure 9). Other temperatures were considered and a previous heat shock at

37ºC was tried. However, extreme high mortality at the larval stage was observed, which resulted to be not compatible with the procurement of samples for downstream analysis.

Larvae grew until 4th instar larvae or adult emergence when they were collected and stored in ethanol 100% at -20ºC until RNA extraction. Three replicates were accomplished for each of the six conditions and the RNA extraction was performed over pools of 10 individuals, 10 females in the case of adults.

Quality control was performed over our 18 samples prior to sequence by using TapeStation RNA Screen Tape (Agilent). One of the three replicates for
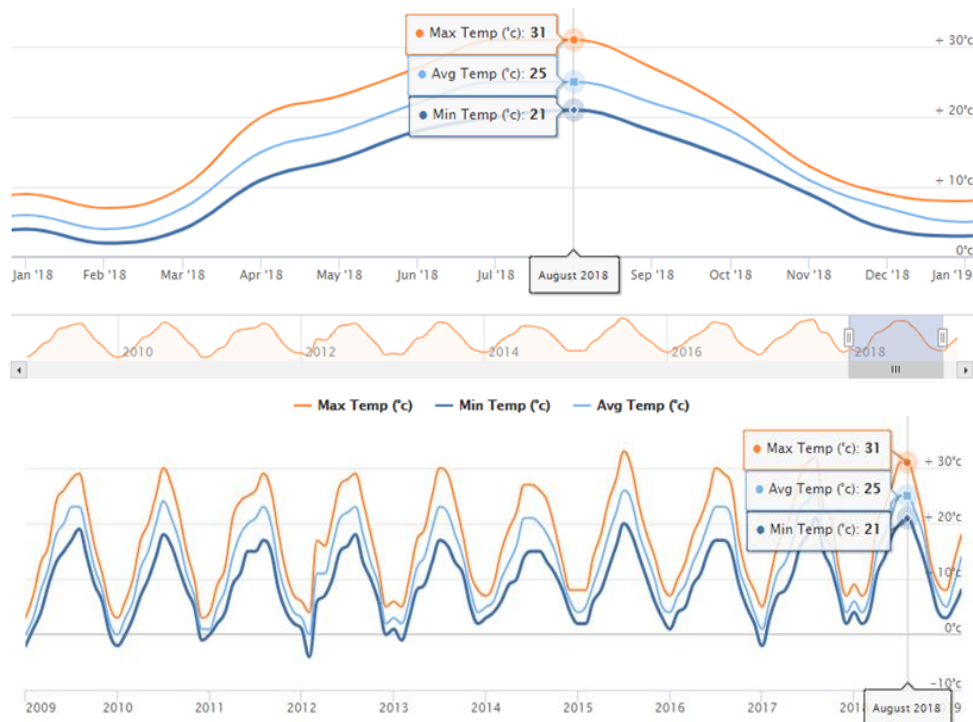
Figure 9. Maximum, minimum and average temperature fluctuations in Pavia during the last decade (World Weather Online).

49

control adults was discarded because it did not pass quality control. Transcriptome sequencing was performed over 17 RNA samples by Macrogen Next-Generation Sequencing services. Total RNA libraries were constructed using Illumina TrueSeq Stranded, rRNA was removed and NovaSeq was run as sequencer. Pair end reads were sequenced to generate high-quality sequence data containing a total of 30 million reads per sample. High-throughput sequencing of messenger RNA generates a large amount of data. Manipulation of this data requires fast, accurate and flexible bioinformatics software which allow us to transform large sets of raw sequencing reads into lists of differentially expressed genes among conditions.

Temperature of growth (18ºC, 28ºC and 32ºC) and developmental stage (adults and larvae) were the variables for which I tested an impact on gene expression. First read quality was analysed for each of the 17 RNA-seq libraries with FASTQC to ensure the validity of ulterior analysis. After this quality check, the new Tuxedo pipeline was applied for differential expression analyses (Pertea *et al.*, 2016). RNA sequencing reads were mapped against the sequences previously obtained for the *Piwi* genes of *Ae. albopictus* using HISAT2 (Hierarchical Indexing for Spliced Alignment of Transcripts) (Kim *et al.*, 2015). Transcripts generated after the alignment were assembled using StringTie (Pertea *et al.*, 2015). Results are generated as count tables which contain both the gene and the transcript level expressions displayed as Fragments per Kilobase of transcript per Million mapped reads (FPKM). Output data generated by StringTie was further analysed using the Ballgown package in R. The overall analyses of

50

transcriptional changes of *Ae. albopictus* following heat challenges are the subject of the Master Thesis of Luis Hernandez, whom I mentored during his internship (Effect of temperature on the transcriptome of the arboviral vector *Aedes albopictus* - Master program in Molecular Biology and Genetics, July 2019). For my PhD program, I focused exclusively on temperature-dependent changes in the expression of *Piwi* genes.

4.2 NIRVS and arboviral infection

4.2.1 Infection experiments

DNA and RNA extractions were performed on the pools of carcasses and ovaries obtained for the late stages of infection and the related controls (see paragraph 4.1.2 for details on the infection experiments procedure). DNA and RNA extractions were performed with NucleoSpin DNA Insect and Nucleospin miRNA by Macherey Nagel, respectively, following manufacturer's instructions.

4.2.2 Small RNA expression analyses from NIRVS and against infecting viruses

DNA and small RNA extraction was performed on CHIKV-infected and DENV-infected ovaries and carcasses at late stages of infection (14 and 21 dpi, respectively), together with the corresponding sugar-fed and blood-fed controls. Whole genome sequencing libraries were generated and sequenced on the Illumina Novaseq 6000 by the Beijing Genomics Institute (BGI) to obtain 150bp paired end reads. Small RNA data was sequenced on the BGI-Seq50 to obtain 40 million reads per sample.

Clean WGS data were aligned to the genome of *Ae. albopictus* (new Canu

assembly, draft version) using BWA (Li and Durbin, 2010), and the presence of previously annotated NIRVSs was tested using GATK (McKenna *et al.*, 2010), setting minimum mapping quality to 20 (0.01 error). New NIRVS formation was investigated using the VyPer pipeline (Forster *et al.*, 2015) and vDNA forms were investigated via PCR, cloning and Sanger sequencing, as previously described (Chapter 4.1.5 Structure of *Piwi* genes).

4.2.3 NIRVS coverage

Small RNA data were aligned using sRNAmapper (Rosenkranz *et al.*, 2015) using -best to the genome of *Ae. albopictus* (new Canu assembly, draft version). BBMap reformat.sh (https://sourceforge.net/projects/bbmap/) was used to filter small RNAs by length. We kept 20-25nt for siRNAs and miRNAs (si-miRNAs), and 26-32nt for piRNAs. For piRNAs, another step was included to filter out reads not starting with T, using custom scripts. This step retains only primary piRNAs (1U bias), which are the piRNAs required for target silencing, whereas secondary piRNAs (10A bias) engage in the ping-pong amplification loop targeting piRNA precursors to increase the amount of primary piRNAs (Wang *et al.*, 2014). Subread featureCounts.sh was used to estimate the amount of si-miRNAs and primary piRNAs mapping against each NIRVS in each condition (Liao *et al.*, 2014). Differential expression analyses were performed base on read counts in edgeR using the weighted trimmed mean of M-values with zero pairing (TMMwzp) normalization method and likelihood ratio test to asses statistical significance (Robinson *et al.*, 2009). Since only two replicates per condition were available, and since a nominal dispersion value can prove to be more realistic than ignoring biological variability, data dispersion for all conditions was set

equal to the dispersion of 14 days ovaries data, considering that data from ovaries are more homogeneous than data from carcasses, which comprise different tissues, and that Chikungunya virus can be readily detected in ovaries even at the first gonotrophic cycle, 5-6 days after an infectious blood-meal (Vazeille *et al.*, 2007; Chompoosri *et al.*, 2016). si-miRNAs and piRNAs analyses were run independently.

4.2.4 Single-molecule counts of small RNAs

High sequence similarity can be found among NIRVS regions, even when they belong to different viral groups of origin. Therefore, single small RNA molecules could be mapped to several NIRVS with equal confidence despite the high stringency implemented in sRNAmapper. To address si-miRNAs and piRNAs expression precisely, reads from each condition that mapped in NIRVS were collapsed based on their sequence identity with fastx-toolkit (Gordon *et al.*, 2014) and each sequence used to count the amount of each small RNA molecule present in every fastq file, using custom scripts. Differential expression analyses were performed in edgeR as previously described, and enrichment of piRNAs on NIRVS and in piRNA clusters were analysed using binomial testing (Rivals *et al.*, 2007). Small RNAs found to be significantly over- or under-expressed between infected and control samples (i.e. mosquitoes offered a not-infectious blood-meal) were blast-searched against the collection of *Ae. albopictus* transcripts allowing no gaps and requiring a seed match of at least 18bp. Thermodynamic analyses on the RNA-RNA pairing were performed using the RNAhybrid software (Krüger and Rehmsmeier, 2006) with a $\Delta G < -20$KJ and 3–10 helix constraints (Krüger and Rehmsmeier, 2006; Kertesz *et al.*, 2007). When I found blast hits

53

of piRNAs to *Ae. albopictus* transcripts, these transcripts were retrieved and analysed using Blast2GO (Götz *et al.*, 2008) to find their Gene Ontology terms. GO analyses were performed merging the results from the Interpro classification and the results from blast on the Culicidae (mosquitoes) database with the subsequent mapping and annotation, as implemented in Blast2GO. Data was visualized with flourish (https://flourish.studio).

4.2.5 Mapping against the infecting viruses

The same small RNA data were also mapped against the genomes of DENV1 and CHIKV in the different conditions. The same tools described above were used to filter small RNAs by length and to count the amount of mapping si-miRNAs and piRNAs. Moreover, strand bias analyses were performed using samtools flag options and PingPongPro (Uhrig and Klein, 2019) was implemented to asses ping-pong signature.

4.3 Life history traits of *Ae. albopictus*

4.3.1 Mosquito strains

Five *Ae. albopictus* strains were used in this study: Foshan (Fo), Rimini (Ri), La Reunion (LaR), Tapachula (Ta) and Crema (Cr). Fo is the genome reference strain: its DNA was used to derive the current genome assembly (Chen *et al.*, 2015). Fo was established in the 1980s from mosquitoes collected in Foshan (China). It was received from Dr. Xiaoguang Chen of the Southern Medical University of Guangzhou (China) in 2013 and has since been maintained in the insectary of the University of Pavia under standard conditions. The Ri strain, also called Fellini, was established in 2004 from 500 eggs collected in an urban environment around the city of Rimini (Dritsou

et al., 2015). The generation 53 of this strain was received from Romano Bellini of the Centro Agricoltura Ambiente "G. Nicoli" in Crevalcore (Italy) in December 2013 and has since been maintained in Pavia. The Cr, Ta and LaR strains were established from eggs collected in the summer-fall of 2016 in Crema (Italy), Tapachula (Mexico) and St. Pierre (La Reunion Island), respectively. The fitness parameters of these three strains were measured within their first eight generations in the laboratory. Mosquitoes of all strains were reared in parallel 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. No specific permits are required to perform laboratory experiments on mosquitoes under the directive 2010/63/EU of the European Parliament and the Council on the protection of animals used for scientific purposes.

4.3.2 Larval-to-pupal developmental time and pupae sex ratio

Ten replicate containers consisting of fifty, one-instar larvae in 150 ml of distilled water were set up for each strain. The evaluated larval density represents an intermediate status between low and high crowding conditions. Larvae were fed every day with the same quantity of food (5 pellets) and water (50 ml) was added every 3rd day to each container. Number of larvae was assessed every day until pupation. Pupae were collected and let emerge to verify their sex.

4.3.3 Wing span

Fifty 3-5-day old mosquitoes per sex were selected randomly for each strain

and their wing lengths measured from the axial incision to the apical margin, not including the fringe of scales using a stereoscope microscope.

4.3.4 Adult lifespan

Fifty 1-day old mosquitoes each were placed in separate cages according their sex and strain. Mosquitoes in each cage were maintained on wet cotton balls with 20% sucrose. Dead mosquitoes were removed and recorded daily until all died.

5. Results

5.1 the piRNA pathway of *Ae. albopictus*

5.1.1 Identification of *Piwi* genes in the *Ae. albopictus* genome

Bioinformatic analyses based on homologous sequences in *Ae. aegypti*, followed by molecular validation and copy number estimations, identified seven *Piwi* genes the genome of *Ae. albopictus*: *Ago3, Piw1/3, Piwi2, Piwi4, Piwi5, Piwi6* and *Piwi7* (Table 1).

Exon-intron boundaries were confirmed for each gene and their expression confirmed by qPCR and Northern Blot. Sequence analysis of the transcripts

Table 1. Copy number estimation of the *Piwi* genes over 4 samples	
Gene	**Copy Number estimation**
Ago3	1.1992 ± 0.3709
Piwi1/3	0.9387 ± 0.1813
Piwi2	1.0832 ± 0.1269
Piwi4	1.0751 ± 0.2819
Piwi5	1.1597 ± 0.1649
Piwi6	1.0280 ± 0.0479
Piwi7	0.9009 ± 0.1407

confirmed the presence of the hallmark domains of Piwis, namely the PAZ, MID and PIWI domains. These domains roughly correspond to the RNA binding site at the amino-terminal, the piRNAs recognition site and the carboxyl-terminus RNase H domain, respectively. The PIWI domain is characterized by a triad of highly conserved residues (DDH or DDE), which were confirmed for each gene.

Phylogenetic analysis with the currently annotated *Piwi* transcripts of Culicinae and Anophelinae mosquitoes confirmed one-to-one orthology between *Ae. albopcitus* and *Ae. aegypti* (Figure 10). It is interesting to notice that *Piwi5, Piwi6* and *Piwi7* grouped close to Aubergine-like transcripts from other species, while *Piwi1/3* and *Piwi2* formed a species-specific clade. In *Ae. aegypti*, *Piwi2* and *Piwi3* are located on chromosome 1 less than 20kb apart, which led us to hypothesize that these two genes in *Ae. albopcitus* may be homogenous due to inter-locus gene conversion caused by nonreciprocal recombination.
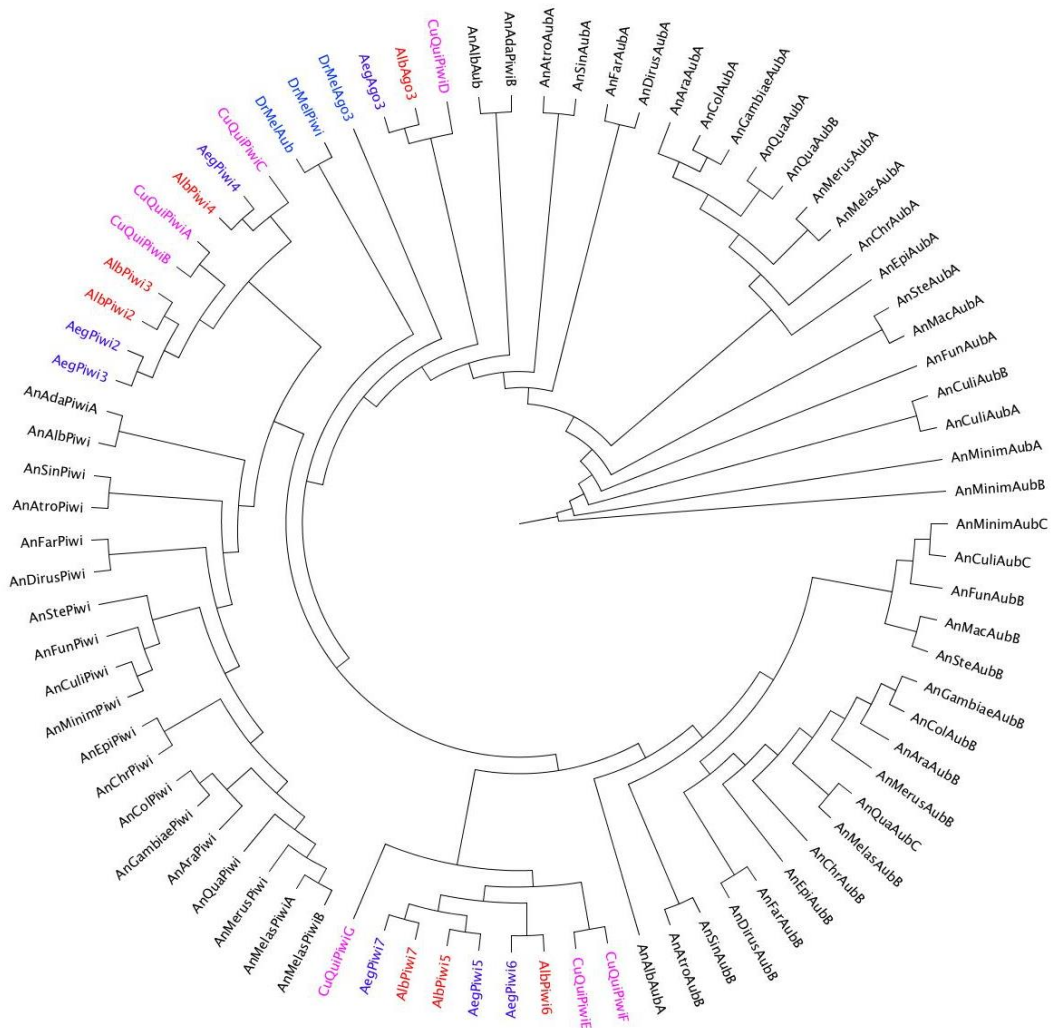
Figure 10. Phylogenetic analysis with the currently annotated *Piwi* transcripts of Culicinae and Anophelinae mosquitoes

5.1.2 *Piwi* gene intraspecies polymorphism

WGS data from two natural populations from La Réunion island and Mexico were combined with data from the Foshan reference strain to evaluate the variability of the *Piwi* genes across populations and assess signs of selective pressure. Briefly, both synonymous and non-synonymous mutations were found for all the genes in all the populations, *Piwi1/3* displaying the lowest level of polymorphism and *Piwi4* and *Piwi5* the highest (Table 2). Between 20 and 80 non-synonymous mutations were found in or in proximity of the PAZ, MID and PIWI domains of *Piwi4*, depending on the population. In-frame deletions were found in the 5' region of the coding sequence of *Piwi5*, two of which (94_99del; 113_118del) were shared across all populations and could be found in homozygosity, suggesting that they are not likely detrimental.

Table 2. Polymorphism of *Aedes albopictus Piwi* genes in mosquitoes from the Foshan strain and wild-caught mosquitoes from La Reunion (Reu) and Mexico (Mex). Number of sequences (*n*), number of sites (*L*), segregating sites (*S*), polymorphism measured as π and θ, and the Tajima's *D* statistic for both synonymous (s) and non-synonymous (a) sites for each gene and population

	n	*L*	*L*$_s$	*L*$_a$	*S*$_s$	*S*$_a$	π_s	π_a	θ_s	θ_a	π_a/π_s	*D*$_s$	*D*$_a$
Ago3													
Pooled	112	2832	680.2	2151.8	316	19	0.0699	0.0005	0.0878	0.0017	0.007	-0.68	-1.95
Foshan	32	2832	680.1	2151.9	124	5	0.0559	0.0004	0.0453	0.0006	0.007	0.89	-0.82
Mex	48	2832	680.2	2151.8	253	14	0.0780	0.0007	0.0838	0.0015	0.009	-0.25	-1.60
Reu	32	2658	643.8	2014.2	189	4	0.0678	0.0002	0.0729	0.0005	0.003	-0.27	-1.50
Piwi1-3													
Pooled	112	2658	644.3	2013.7	136	23	0.0319	0.0010	0.0399	0.0022	0.033	-0.66	-1.51
Foshan	32	2658	644.0	2014.0	10	2	0.0047	0.0003	0.0039	0.0002	0.064	0.68	0.44
Mex	48	2658	644.9	2013.1	117	21	0.0463	0.0017	0.0409	0.0024	0.037	0.48	-0.89
Reu	32	2658	643.8	2014.2	52	4	0.0188	0.0004	0.0201	0.0005	0.021	-0.23	-0.48
Piwi2													
Pooled	112	2625	644.0	1981.0	242	28	0.0760	0.0012	0.0710	0.0027	0.016	0.23	-1.65
Foshan	32	2625	644.0	1981.0	115	10	0.0663	0.0017	0.0443	0.0013	0.026	1.88	1.11
Mex	48	2625	643.9	1981.1	184	15	0.0823	0.0010	0.0644	0.0017	0.012	1.01	-1.28
Reu	32	2625	644.1	1980.9	151	6	0.0712	0.0005	0.0582	0.0008	0.007	0.85	-0.94
Piwi4													
Pooled	112	2592	620.0	1972.1	268	61	0.0729	0.0025	0.0817	0.0058	0.034	-0.36	-1.82
Foshan	32	2592	620.1	1971.9	122	18	0.0610	0.0009	0.0489	0.0023	0.015	0.94	-2.05
Mex	48	2592	619.8	1972.2	181	41	0.0692	0.0035	0.0658	0.0047	0.051	0.19	-0.87
Reu	32	2592	620.1	1971.9	161	45	0.0699	0.0029	0.0645	0.0057	0.041	0.32	-1.79
Piwi5													
Pooled	112	2745	653.1	2091.9	148	23	0.0457	0.0016	0.0428	0.0021	0.035	0.22	-0.66
Foshan	32	2793	664.5	2128.5	58	8	0.0361	0.0018	0.0217	0.0009	0.050	2.47	2.78
Mex	48	2745	652.9	2092.1	137	13	0.0470	0.0017	0.0473	0.0014	0.036	-0.02	0.65
Reu	32	2793	663.4	2129.6	89	6	0.0326	0.0008	0.0333	0.0007	0.025	-0.08	0.40
Piwi6													
Pooled	112	2661	649.0	2012.0	242	8	0.0805	0.0010	0.0705	0.0008	0.013	0.47	0.82
Foshan	32	2661	648.3	2012.8	92	3	0.0632	0.0001	0.0352	0.0004	0.002	2.99	-1.69
Mex	48	2661	649.9	2011.1	213	7	0.0840	0.0001	0.0739	0.0008	0.001	0.50	-2.33
Reu	32	2661	648.5	2012.5	163	4	0.0784	0.0001	0.0624	0.0005	0.001	0.98	-2.01
Piwi7													
Pooled	112	1977	469.8	1507.2	192	33	0.0877	0.0036	0.0772	0.0041	0.041	0.45	-0.42
Foshan	32	1977	469.8	1507.2	118	15	0.0905	0.0034	0.0624	0.0025	0.038	1.71	1.25
Mex	48	1977	469.9	1507.1	150	23	0.0905	0.0034	0.0719	0.0034	0.038	0.93	-0.04
Reu	32	1977	469.6	1507.5	137	17	0.0803	0.0030	0.0724	0.0028	0.037	0.41	0.24

We used the McDonald-Kreitman test to detect signatures of natural selection and found signs of adaptive evolution for *Piwi1/3* and *Piwi6* consistent with divergent positive selection followed by purifying selection (Table 3A). On the contrary, we found a significant deficit of non-synonymous substitutions and/or excess of polymorphic non-synonymous segregating sites for *Piwi4*, which may reflect the effect of intraspecific diversifying selection, as expected in genes involved in immunity. We then tested models for codon evolution to assess whether specific sites along the coding sequences of the *Piwi* genes may be involved in adaptive evolution (Table 3B). We found signs of positive selection at different sites, including sites in the proximity of the PAZ, MID and PIWI domains: for *Piwi1/3* we found one site displaying signs of positive selection in the linker region between PAZ and MID domains (referred to as Linker2) and one site in the MID domain; for *Piwi4* two sites in the PAZ domain; for *Piwi5* two sites in the N-terminus flexible domain (i.e. Flex domain) and for *Piwi6* three sites, two in the Flex and one in the Linker2 domain. Haplotype reconstruction confirmed that these mutations can co-occur on the same gene, with the exception of the Y278D mutation with H287P in *Piwi4* and the mutation A67P with G86S in *Piwi6*.

Table 3. A) McDonald-Kreitman test for each *Piwi* gene using the orthologous sequences of *Ae. aegypti* as outgroup. NI = Neutrality Index; Alpha = proportion of base substitutions fixed by natural selection; *P* estimated using Fisher's exact test. **B)** [1] sites where signs of positive selection (ω > 1) were found; [2] reference amino acid and alternative missense variant; [3] mean omega (ω) value; [4] probability that ω > 1 under the Bayes empirical Bayes (BEB) method (* = P > 0.95; ** = P > 0.99); [5] protein domain based on computational predictions of molecular structures. Domains are as follows: Linker2, linker region between PAZ and MID; PAZ domain; MID domain; and Flex, the Flexible stretch at the N-terminus.

A. McDonald-Kreitman test

	Ago3	*Piwi1/3*	*Piwi2*	*Piwi4*	*Piwi5*	*Piwi6*	*Piwi7*
NI	0.582	0.516	0.9	3.888	0.696	0.154	0.745
alpha	0.418	0.484	0.1	-2.888	0.304	0.846	0.255
P	0.114	0.008	0.785	< 0.001	0.18	< 0.001	0.272

B. Codeml output for sites under positive selection

Gene	Position[1]	Reference>Mutant[2]	ω[3]	P[4]	Domain[5]
AGO3	-	-	-	-	
Piwi1-3	484	E>G	3.026	0.990	Linker2
	485	K>R	2.979	0.965*	Linker2
	548	M>I	3.014	0.984*	MID
Piwi2	-	-	-	-	
Piwi4	278	Y>D	2.522	0.993**	PAZ
	287	H>A,D,P,V	2.532	1.000**	PAZ
Piwi5	89-90	SA>PT	7.813	1.000*	Flex
	139	T>A	7.810	1.000*	Flex
Piwi6	67	A>P	3.560	0.992**	Flex
	86	G>R,S	3.460	0.957*	Flex
	258	V>I	3.581	0.999**	Linker2
Piwi7	-	-	-	-	

Finally, we estimated how variable the *Piwi* genes of *Ae. albopictus* are in comparison to genes that were previously identified as slow-evolving (SGs) and fast-evolving (FGs) (Pischedda *et al.*, 2019). Analyses of the overall level of polymorphism (LoP) of the aforementioned genes for each population showed that *Piwi4*, *Piwi6* and *Piwi7* are comparable to FGs, while *Ago3* and *Piwi5* do not significantly deviate from the LoP values of SGs. *Piwi2* is twice as variable as SGs, while *Piwi1/3* appears to be highly conserved (Figure 11).
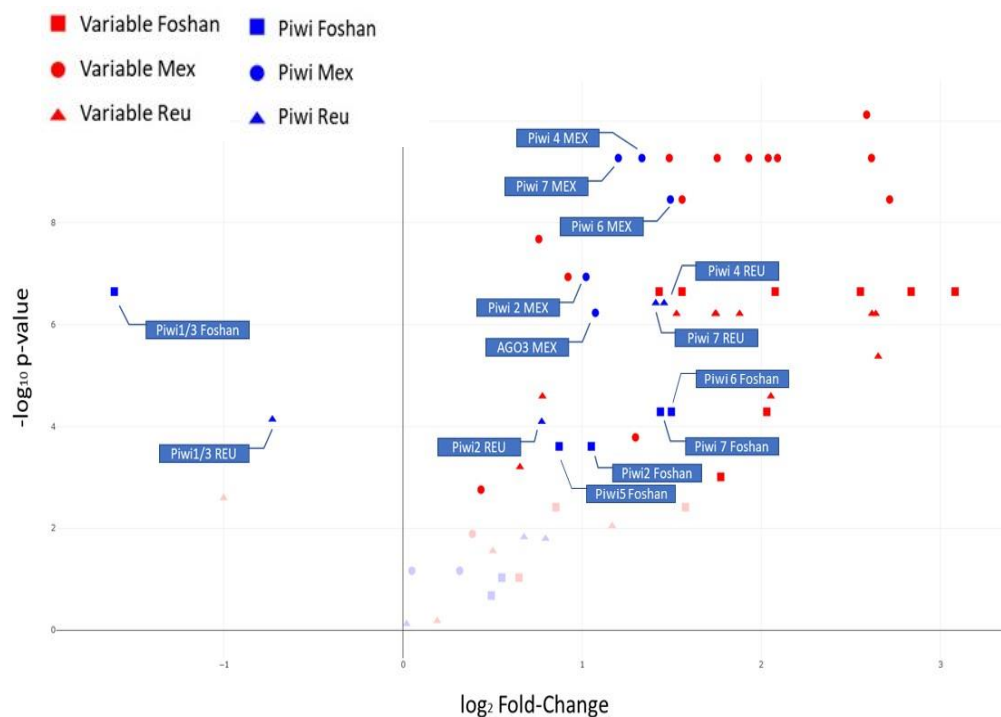
Figure 11. Volcano plot of the Level of polymorphism (LoP) comparison between slow-evolving genes (SGs), fast-evolving genes (FGs) and *Piwi* genes by population. The 0 line is the median value of the LoP of SGs, red shapes represent the values of LoP of FGs and blue shapes represent the LoP of the Piwi genes.

5.1.3 Computational predictions of molecular structures

The most-recent X-ray crystallography structure of Argonaute proteins were used as templates to compute predictions of three-dimensional molecular structures of the Piwi proteins. These models allowed us to gain insight on structural differences among the different Piwi proteins and to map the non-synonymous variants we found to be under selection on the protein 3D structures (Figure 12). Our results revealed that, despite sequence heterogeneity among the different *Piwi* genes, the structure of the resulting proteins is highly conserved. We found high levels of amino acid conservation in the inner pocket corresponding to the RNA binding site, while the protein surface is significantly less conserved. These results are in concordance with what was found in *D. melanogaster*. No polymorphism
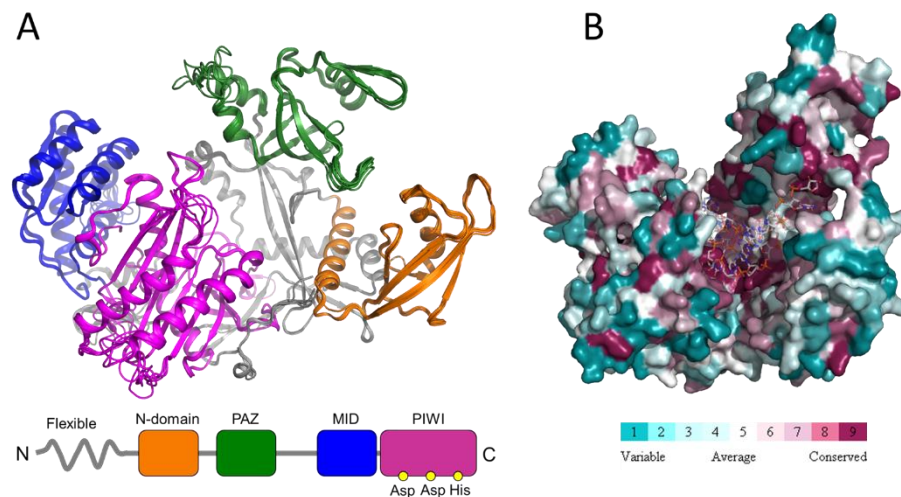
Figure 12. Computational homology models of the *Ae. Albopictus Piwi* proteins. A) Superposition of cartoon representations of Piwi homology models, with highlight of domain organization. B) overview of the ammino acid sequence conservation mapped on three-dimensional homology models

that could lead to a loss of RNA bind and/or slicing activity was found, indicating that all the *Piwis* should have maintained Argonaute-like functions. The variants under selection did not locate in or near the RNA binding site or in the processing sites, indicating that such variants are unlikely to alter protein folding, although they might alter protein stability.

5.1.4 Expression profile of *Piwi* genes throughout mosquito development
We analyzed the expression pattern of the *Piwi* genes during the developmental stages of *Ae. albopcitus*. In particular we looked at early embryogenesis (4-8h after oviposition), late embryogenesis (8-16h and 16-24h), 1st instar larvae, 4th instar larvae, pupae and adult stages. Adults were analyzed as males and females separately; moreover, females were either kept
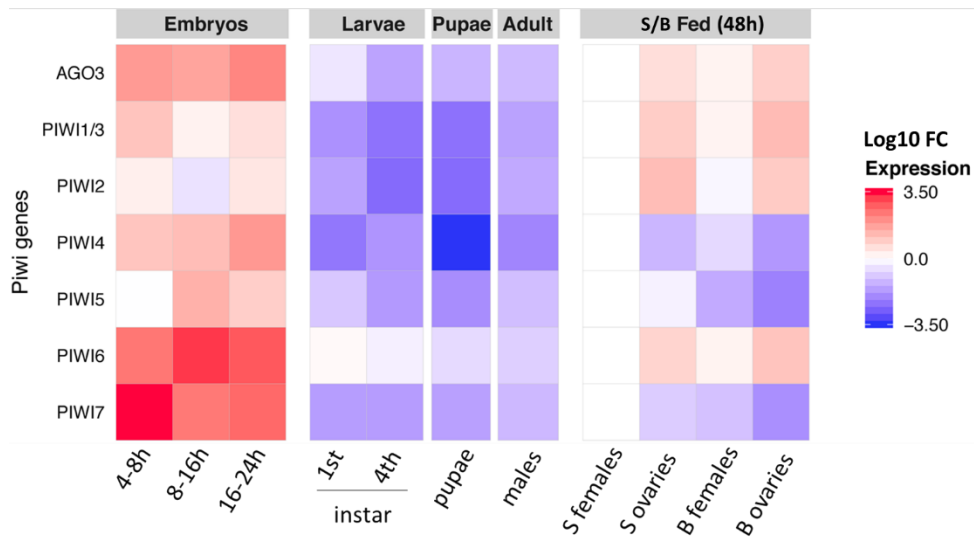
Figure 13. Heatmap representations of log10 transformed fold-change expression values of each *Piwi* gene throughout development

on sugar diet or given a blood meal and then dissected to remove their ovaries, which were analyzed independently (Figure 13).

Our analysis compared the expression levels of sugar fed adult females with all the other conditions. Our results indicate that all the *Piwi* genes are initially expressed during embryogenesis and then their expression drops in juvenile stages (larvae to pupae). *Ago3, Piwi1/3, Piwi2 and Piwi6* showed a second peak of expression in adult females and their ovaries. On the contrary, *Piwi4, Piwi5* and *Piwi7* expression levels increased 100 to 1000 times in sugar-fed females compared to pupae, but not in ovaries nor in blood-fed samples, where the expression levels were not different from the ones of the juvenile stages. Consistently with the data analyzed from publicly available RNA-seq experiments, *Piwi7* is the least expressed gene, making it nearly undetectable after embryonic development.

In conclusion, *Ago3* and all *Piwi* genes were highly expressed in embryos compared to all other developmental stages. In adults, they were all more expressed in females than males. Expression in ovaries was generally higher than in carcasses, in both sugar- and blood-fed females. Differences in carcasses vs. ovaries expression were more pronounced after blood-meal for *Ago3, Piwi1/3* and *Piwi6*, whereas the expression of *Piwi2* was twice higher in sugar- vs. blood-fed ovaries.

5.1.5 Infection experiments

We analysed the transmission efficiency and the number of viral particles in the saliva of 30 individual Foshan mosquitoes for each time point and each virus (i.e. 4 and 14 dpi for CHIKV, 4 and 21 dpi for DENV). The same mosquitoes were then dissected removing their ovaries. Carcasses and ovaries

were pooled into groups of 15 for subsequent DNA and RNA extractions. Early (4 dpi) and late (14 and 21 dpi) infection RNA samples were used for RT-qPCR experiments, while DNA and small RNA extracted from late infection samples were used for the analyses of NIRVS landscape/integration and small RNA production (chapters 5.2.2 and 5.2.3).

Transmission efficiency (i.e. the proportion of mosquitoes with infectious saliva among the total number of mosquitoes tested) analyses revealed that CHIKV could be readily detected at 4 dpi in 40% of the mosquitoes, increasing to nearly 60% at 14 dpi. For DENV, all the saliva retrieved at 4 dpi was found to be non-infectious, while transmission efficiency was found at ~30% at 21 dpi, although with very low number of viral particles (average below 5). These results are in line with previous reports (for instance: Calvez *et al.*, 2017; Amraoui *et al.*, 2019)

5.1.6 Expression of *Piwi* genes following arboviral infection

Piwi genes expression was also evaluated following arboviral infection with DENV 1 and CHIKV 0621 (Figure 14). Carcasses and ovaries were analysed separately during early stages (i.e. 4dpi) and late stages of infection (i.e. 14dpi for CHIKV and 21dpi for DENV). Our results indicate a complex picture where significant differences exist when analysing different tissues, viral species, time point and *Piwi* genes. Briefly, all the *Piwi* genes were upregulated in CHIKV-infected ovaries during the early stages of infection,
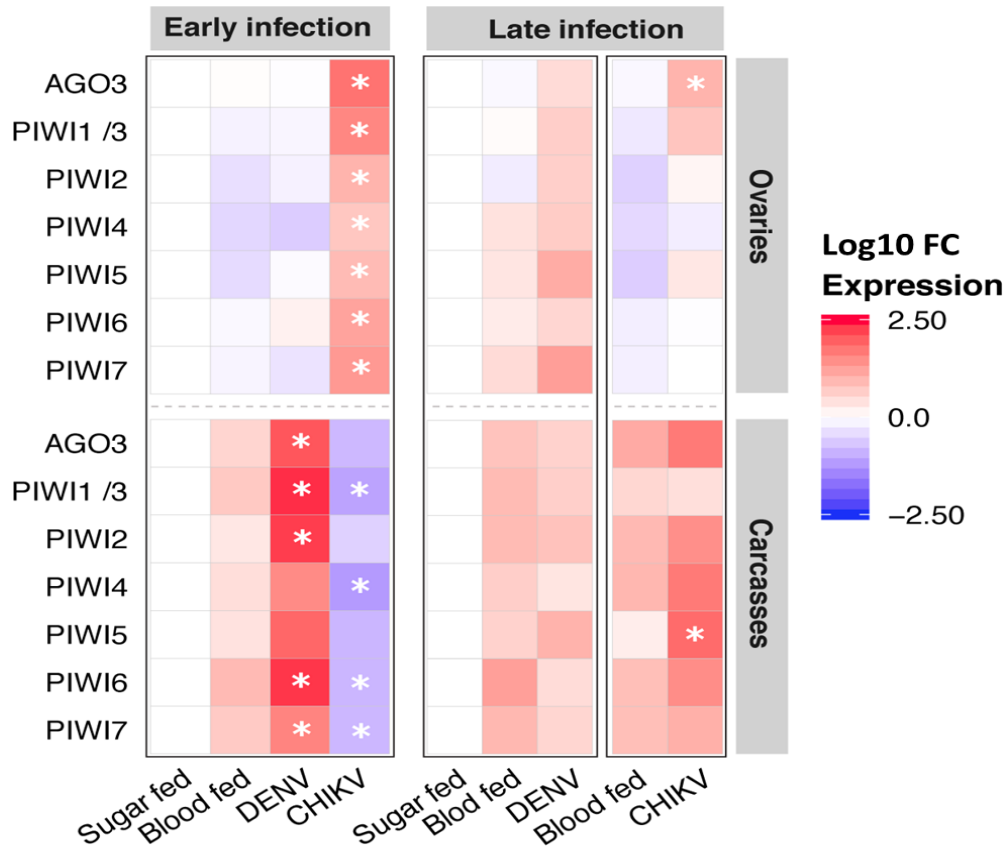
Figure 14. Heatmap representations of log10 transformed fold-change expression pattern of *Piwi* genes following viral infection normalized with respect to sugar-fed samples. Expression was verified in ovaries and carcasses separately, during the early and late stages of infections, that is 4 dpi for both viruses and 14 or 21 dpi for CHIKV and DENV, respectively. Each day post-infection was analysed with respect to sugar and blood-fed controls of the same day. * indicates significant difference (P<0.05) between infected samples and the corresponding blood-fed control.

while no significant increase in expression levels was found in DENV-infected ovaries. This is in line with the different tropisms of the two viruses:

CHIKV is able to spread throughout the mosquito body in the first 2 to 6 dpi, whereas DENV is hardly found in ovaries in the same time window (Liu *et al.*, 2017). Still in ovaries, *Piwi* genes expression levels dropped during the late stages of infection between 3 (*Piwi5*) to 20 (*Piwi7*) times with respect to the values observed at 4 dpi; however, they remained higher than the corresponding expression values in ovaries of both sugar- and blood-fed mosquitoes. On the contrary, 21 dpi DENV-infected ovaries showed a general increment in *Piwi* expression levels, although it was not statistically higher than the ones found for blood-fed mosquitoes.

A different scenario was found in carcasses: at 4dpi, the *Piwi* genes were up-regulated in DENV-infected samples and down-regulated in CHIKV-infected samples. In particular, *Ago3, Piwi1/3, Piwi2, Piwi6* and *Piwi7* were significantly more expressed in DENV-infected carcasses when compared to blood-fed mosquitoes, while *Piwi1/3, Piwi4, Piwi6* and *Piwi7* were significantly under-expressed in CHIKV-infected carcasses. However, the situation changed in late stages of infection, when expression in DENV-infected carcasses decreased to similar levels as in blood-fed control mosquitoes, while in CHIKV-infected carcasses the expression levels rose up to the point that one gene, *Piwi5*, was found to be significantly over-expressed.

Taken together, these observations support the hypothesis of an early activation of the piRNA pathway when mosquitoes were infected with DENV-1, whereas CHIKV infection lead to a later activation. *Piwi5* was elicited both during DENV and CHIKV infection, therefore we can hypothesize a general antiviral role for this gene. However, other *Piwi* genes

seem to be involved in the antiviral response, such as *Piwi6* and *Piwi1/3* or *Piwi4* and *Ago3* during infection with DENV and CHIKV, respectively.

5.1.7 Temperature induced changes in *Piwi* genes expression

Since it started spreading throughout the continents, *Ae. albopictus* has come to inhabit different climatic regions than those of its native range in South East Asia. I therefore investigated the impact that a temperate climate might have on the expression profile of the *Piwi* genes, both at larval and adult stage (Figure 15). Analyses at 18°C and 32°C compared to the 28°C control temperature showed that *Ago3* is significantly up-regulated in larvae at higher temperatures and in adults at lower temperatures. *Piwi2* is also significantly upregulated in larvae at higher temperatures, while it is downregulated in both larvae and adults at lower temperatures. *Piwi5* is down-regulated in larvae at 32°C and *Piwi6* displayed significant differences when comparing 18 to 32°C in larvae. These results suggest a complex scenario where changes in temperature can affect different *Piwi* genes in different manners, with differences also depending on the developmental stage of the mosquito. Unfortunately, much is still unknown on the biological roles of the distinct *Piwi* genes, therefore these results, for the time being, are descriptive and will

require future studies to link such changes to possible phenotypic effects.
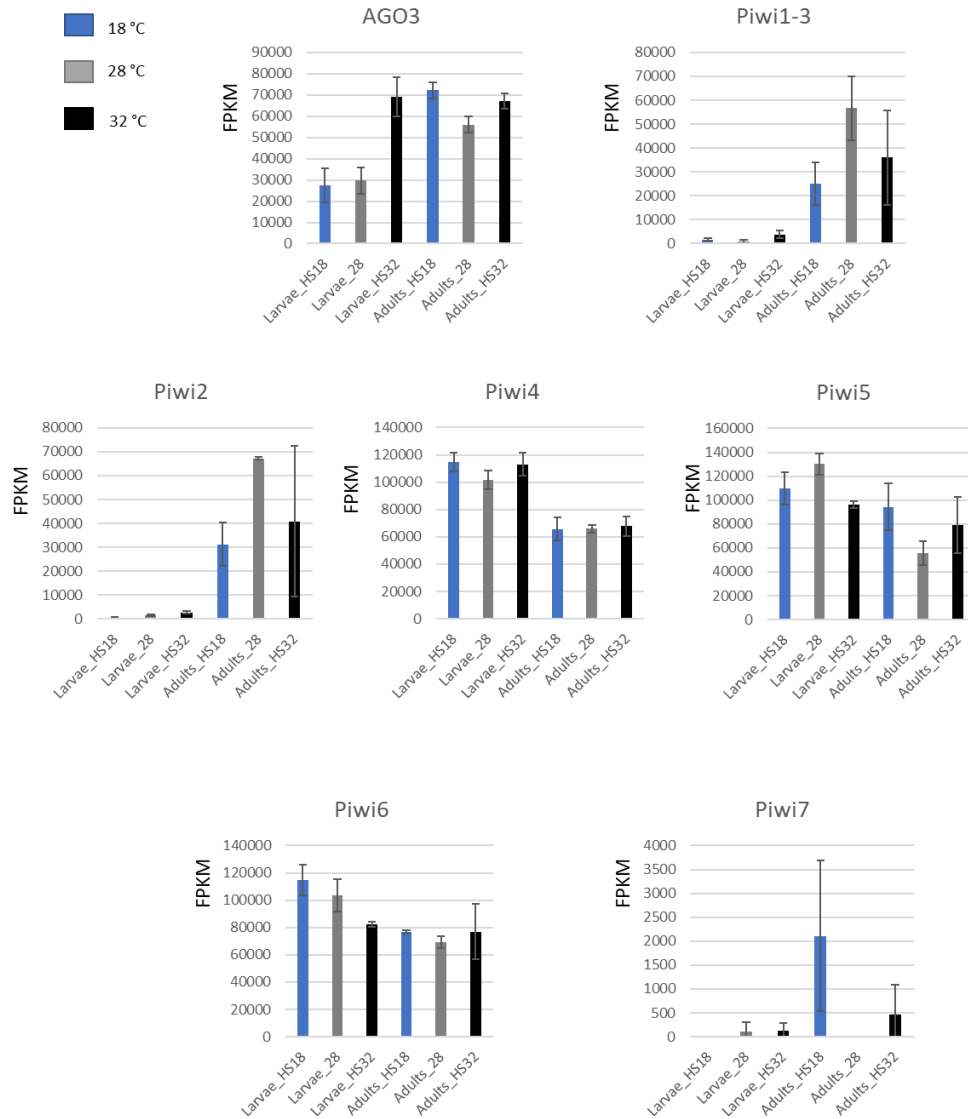
Figure 15. Fragments Per Kilobase Million (FPKM) counts for each *Piwi* gene in larvae and adults reared at low (18°C) standard (28°C) and high (32°C) temperature.

5.2 NIRVS and arboviral infection

I analysed viral DNA forms (vDNA), signs of novel integrations and the small RNA production during the late stages of infection with either CHIKV (14dpi) or DENV (21dpi), in both carcasses and ovaries. In particular, I focused my attention on the different classes of small RNA molecules (i.e. siRNAs, miRNAs and piRNAs), which I identified based on their lengths and known-nucleotide biases (i.e. siRNAs and miRNAs are 20-25nt in length, piRNAs are 26-32nt in length and primary piRNAs are characterized by Uridine at position 1 (Elbashir *et al.*, 2001; Brennecke *et al.*, 2007a; Arensburger *et al.*, 2011; Weick and Miska, 2014; Hu *et al.*, 2015; Maharaj *et al.*, 2015; Liu *et al.*, 2016; Bryant *et al.*, 2019).

5.2.1 vDNAs and novel NIRVS profile during persistent DENV or CHIKV infections

We could not detect any new NIRVS, nor vDNA forms in any of the WGS data analysed. However, through a PCR-based screen using overlapping primers covering the whole viral genome, we could detect a 430bp sequence of the CHIKV Cap gene in CHIKV-infected carcasses and ovaries at both 4 and 14 dpi. In a published experiment, vDNA forms from CHIKV were detected in *Ae. aegypti* mosquitoes of a recently laboratory-adapted strain from Thailand, 9 days post CHIKV infection and using primers that amplify a region upstream of the CHIKV cap gene (Goic *et al.*, 2016).

We therefore hypothesize that either the amount of vDNA forms is too low to be detected by the coverage of our WGS data, different vDNA forms are generated in different experiments depending on the viral genotype and

mosquito strain used or/and vDNA forms are not completely stable and most are lost once infection has become persistent.

The amount of remaining DNA from the infected samples prevents me to design genome walking or other experiments to test whether the amplified fragment is episomal or integrated.

5.2.2 Small RNA profile during persistent DENV or CHIKV infections

si-miRNAs and piRNAs that map against CHIKV- and DENV-1 were identified in all infected samples. Overall si-miRNAs were quantitatively more than piRNAs. Moreover, signals from ovaries were lower than those from carcasses. Strand bias analysis revealed that both si-miRNAs and piRNAs are predominantly in sense orientation with respect to the viral genome, establishing the basis for targeted processing of the negative strand of incoming viral genomes, which is the replicative form of both Flaviviruses and Alphaviruses (Table 4).

Both CHIKV and DENV are positive sense RNA viruses and require a

Table 4. Strand bias analysis for piRNAs and si-miRNAs of ovaries and carcasses at late stages of infection with CHIKV and DENV

		piRNAs		*si-miRNAs*	
	Sample	Forward reads count (RPKM)	Reverse reads count (RPKM)	Forward reads count (RPKM)	Reverse reads count (RPKM)
CHIKV	Ovaries	7.55	1.41	19.98	20.97
	Carcasses	86.51	12.00	417.42	266.92
DENV	Ovaries	0.65	0.10	4.55	3.05
	Carcasses	24.23	1.37	96.86	32.12

negative strand RNA form for the synthesis of their genomic positive RNAs that is packaged into capsids and exit the infected cell. Signs of ping-pong amplification could be found in CHIKV-infected ovaries and carcasses, but not in DENV-infected ovaries and DENV-infected carcasses (Figure 16). Few reads were found for piRNAs in DENV-infected ovaries (211 and 178 in the two replicates, respectively), therefore we cannot exclude that ping-pong signatures may be found at higher depth of coverage. However, count of piRNAs normalized on library sizes indicate that piRNAs production in DENV-infected ovaries is low, with an average RPKM of 0.75. This result implies low piRNA response in DENV-infected ovaries which correlates with a lower expression of *Piwi* genes in ovaries than carcasses of DENV-infected mosquitoes. My results are also in agreement with recently published data showing a lack of ping-pong signature in DENV-2 infected *Ae. albopictus* samples (Wang *et al.*, 2018).

Profile of si-miRNAs showed no hotspots across all samples and for both CHIKV and DENV (Figure 17A). On the contrary, piRNA profiles were not homogenous, with peaks towards the genomic 5'end of both viruses. Specifically, for CHIKV-infected samples, piRNAs coverage increases after position 7500 of the CHIKV genome in both ovary and carcass samples, with distinct peaks in the regions between 7587-7616, 7949-7976, 8645-8673, 10665-10693 and 11005-1106, which include the coding sequences for the structural proteins Cap, gpE3, gpE2 and gpE1. In DENV-infected samples, piRNA coverage was more uniform, but again a concrete increase was noticeable after position 5000 of the DENV reference genome, with distinct peaks between positions 8325-8352, 9991-10017 and 10105-10135, which
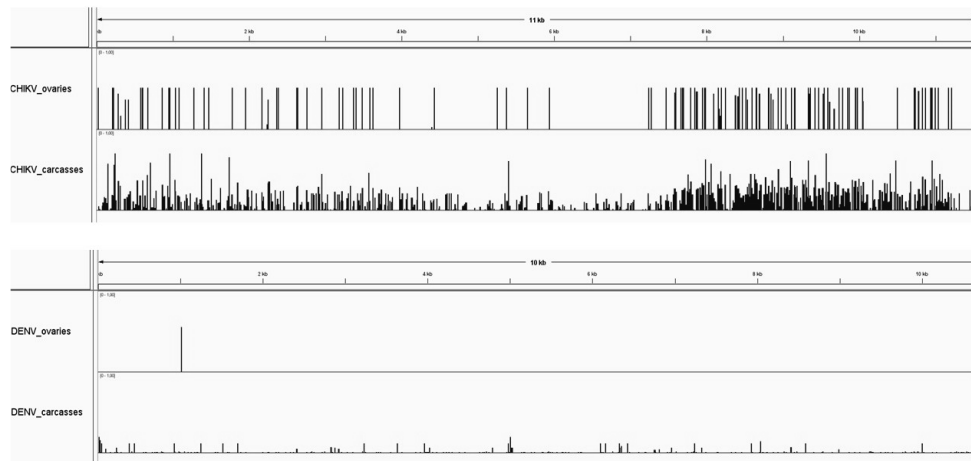
Figure 16. Ping-pong signature in ovaries and carcasses of CHIKV- and DENV-infected samples. Bar height represents the probability that the reads are ping-pong-derived expressed as 1-FDR (False Discovery Rate).

correspond to the coding sequences for the non-structural proteins NS3, NS4A, NS4B and NS5 (Figure 17B).

The identification of a different piRNA profile between DENV- and CHIKV infected samples and the position of the peaks is consistent with what observed with different strains of *Ae. aegypti* and *Ae. albopictus* and also cell-lines infected with CHIKV and DENV, suggesting this is a consistent signal of DENV and CHIKV independent of the combination of *Aedes* spp. strain and viral genotype used (Miesen *et al.*, 2016b). The coverage range in DENV-infected ovaries was the lowest among all tested conditions (peak value = 10 reads), again showing a low piRNAs response in DENV-infected ovaries.
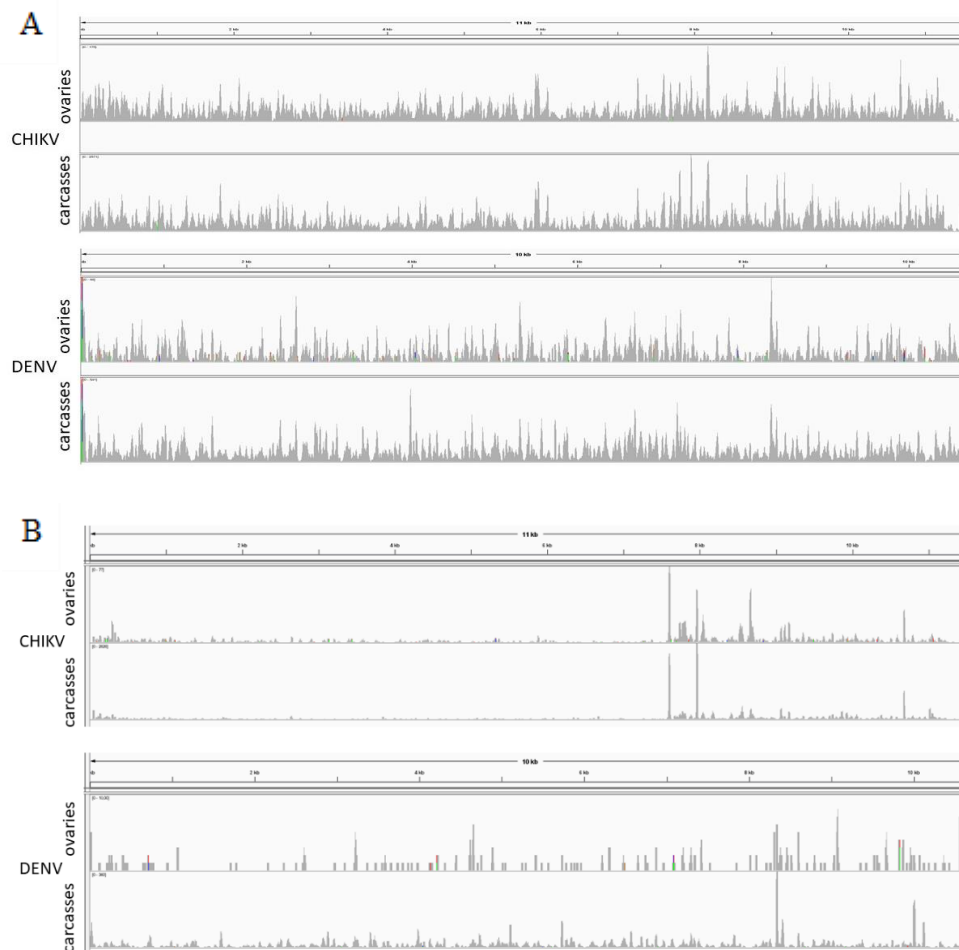
Figure 17. Coverage of si-miRNAs (A) and piRNAs (B) mapping on the viral sequences of CHIKV and DENV in infected ovaries and carcasses. The height of the bars is on relative scales and cannot be compared across conditions.

Small RNAs were also mapped on the NIRVS of the *Ae. albopictus* genome. Because of the sequence identity among some NIRVS, small RNAs may not be unambiguously assigned to single NIRVS. As a consequence, we analysed

77

results at two levels: first we focused on small RNAs and identified those that are significantly differentially expressed (DE) between infected and control (blood-fed) samples; these small RNA molecules may map to different NIRVS (Figure 18).Then we identified the NIRVS with a small RNA profile significantly different between infected and control samples by re-mapping all small RNAs to NIRVSs and assigning reads that mapped to more than one NIRVS to each NIRVS with equal probability; please note that by looking at single NIRVS (level 2) I "normalise" the small RNA profile (Figure 18).
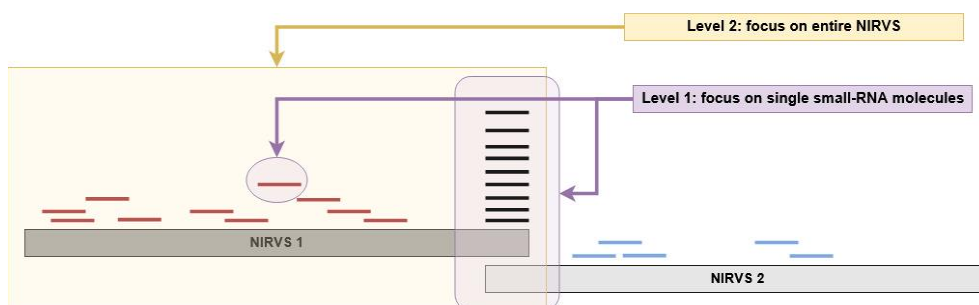
Figure 18. Representation of the strategy used for the small RNA analysis. First, single molecules were counted and referred to all NIRVS they mapped on; this analysis informs on how individual small RNAs are expressed. Secondly, individual NIRVS were considered, with all the small RNA molecules mapping within their boundaries; this analysis informs on the total small RNA profile of entire NIRVS.

Single molecule counts of si-miRNAs and piRNAs revealed that in all conditions a fraction of small RNAs mapping to NIRVS is DE following arboviral infections. Differently than the overall profile observed across the viral genomes, which showed more si-miRNAs than piRNAs, on NIRVS we saw more piRNAs than si-miRNAs, with total piRNA expression being about

8 times higher than that of si-miRNAs. NIRVS also displayed different expression levels during infection, with differences depending on viral species and tissue.

5.2.3 Expression profile of si-miRNAs

Few si-miRNAs were found to be significantly up or down regulated in CHIKV-infected ovaries, CHIKV-infected carcasses and DENV-infected ovaries (1, 2 and 3, respectively), while 154 significantly DE si-miRNAs were found in DENV-infected carcasses. Of the 154 DE si-miRNAs, 131 were over-expressed and 23 under-expressed in DENV-infected with respect to control carcasses (Figure 19). DE si-miRNAs were mapped against NIRVS and binomial test showed that DE si-miRNAs between DENV-infected and control carcasses were enriched in NIRVS located in piRNA clusters (p = 1.50e-6), including Rhabdo-NIRVS (p = 3.49e-15) as well as Flavi-NIRVS (p = 0.003). Flavi-NIRVS with DE si-miRNAs in infected samples included NIRVS corresponding to the NS2, E and NS5 encoding regions of the Flavivirus genome.

Figure 19. Differential expression (DE) of si-miRNAs represented by average log counts-per-million (CPM) on the x-axis and log fold-change on the y-axis. Red dots indicate significantly over-expressed si-miRNAs and blue dots significantly under-expressed si-miRNAs upon infection compared to blood-fed controls.

When I analysed si-miRNAs mapping to each NIRVS rather than looking at individual small RNA molecules, I detected 13 NIRVS whose si-miRNA profile was significantly DE between CHIKV-infected ovaries and CHIKV-control ovaries (6 up-regulated, 7 down-regulated), 51 NIRVS whose si-miRNA profile was significantly DE in CHIKV-infected versus control carcasses (43 up-regulated, 8 down-regulated), 9 NIRVS whose si-miRNA
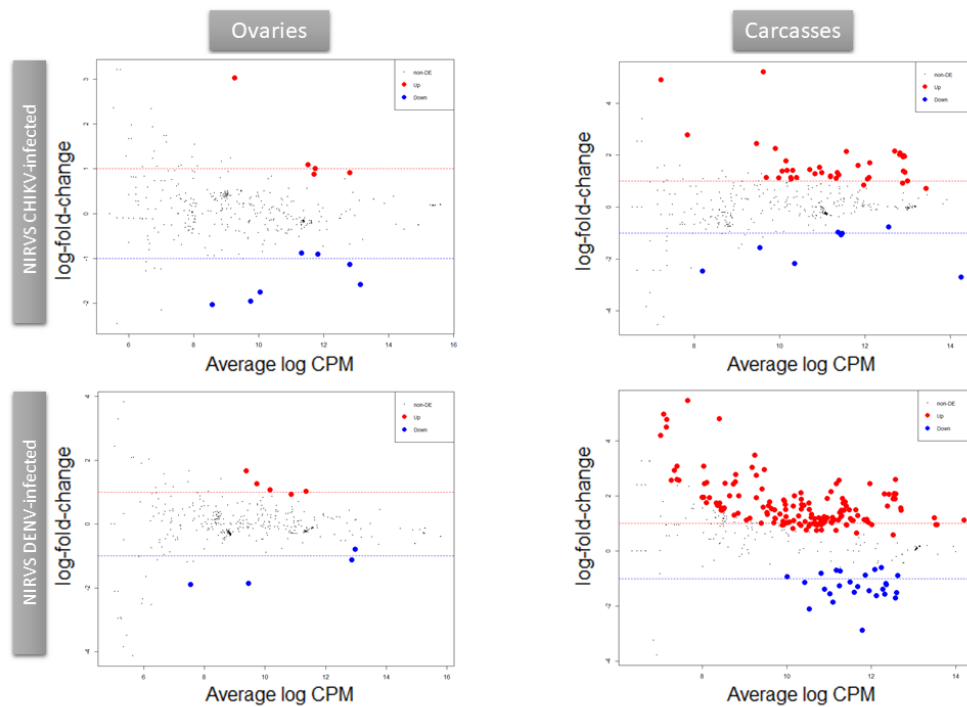
Figure 20. NIRVS profile based on si-miRNAs expression represented by average log counts-per-million (CPM) on the x-axis and log fold-change on the y-axis. Red dots represent NIRVS whose si-miRNA profile is significantly over-expressed and blue dots represent NIRVS whose si-miRNA profile is significantly under-expressed upon infection compared to blood-fed controls.

profile was significantly DE in DENV-infected versus control ovaries (5 up-regulated, 4 down-regulated) and 261 NIRVS whose si-miRNA profile was DE in DENV-infected versus control carcasses (234 up-regulated, 27 down-regulated) (Figure 20). Overall, these results show that a relatively small number of si-miRNAs are significantly DE during persistent infection of CHIKV and DENV whereas the small RNA profile of NIRVS include several

DE NIRVS, with larger changes in DENV-infected carcasses.

5.2.4 Expression profile of piRNAs

For experiments of CHIKV infection, 2 piRNA molecules were found DE in infected vs control ovaries, none in carcasses. The number of piRNA molecules DE between infected vs control samples increased to 118 in ovaries (46 over-expressed, 72 under-expressed) and 111 in carcasses (37 over-expressed, 75 under-expressed) for DENV experiments (Figure 21). The piRNA molecules that are DE in infected vs control samples were mapped on NIRVS and binomial test showed an enrichment in Rhabdo-NIRVS (p = 3.86e-30 and p = 4.84e-12, respectively) for piRNAs that are DE in DENV-infected ovaries and carcasses; in NIRVS mapping in piRNA clusters (p = 0.0004) for piRNAs that are DE in DENV-infected vs control carcasses.

The analyses of the piRNA profile of NIRVS showed 29 and 11 NIRVS with piRNAs that are DE in CHIKV-infected vs control ovaries and carcasses, respectively (i.e. 22 up- and 7 down-regulated in ovaries; 9 up- and 2 down-regulated in carcasses); 22 and 126 NIRVS with piRNAs that are DE in DENV-infected vs control ovaries and carcasses, respectively (5 up-, 17 down-regulated in ovaries; 90 up- and 36 down-regulated in carcasses) (Figure 22). Interestingly, we observed that during viral infection both with CHIKV and DENV down-regulated NIRVS could produce over-expressed small RNAs, in accordance with a small RNA profile not homogeneous, but with hotspots, as observed for the small RNA profile on the viral genome.
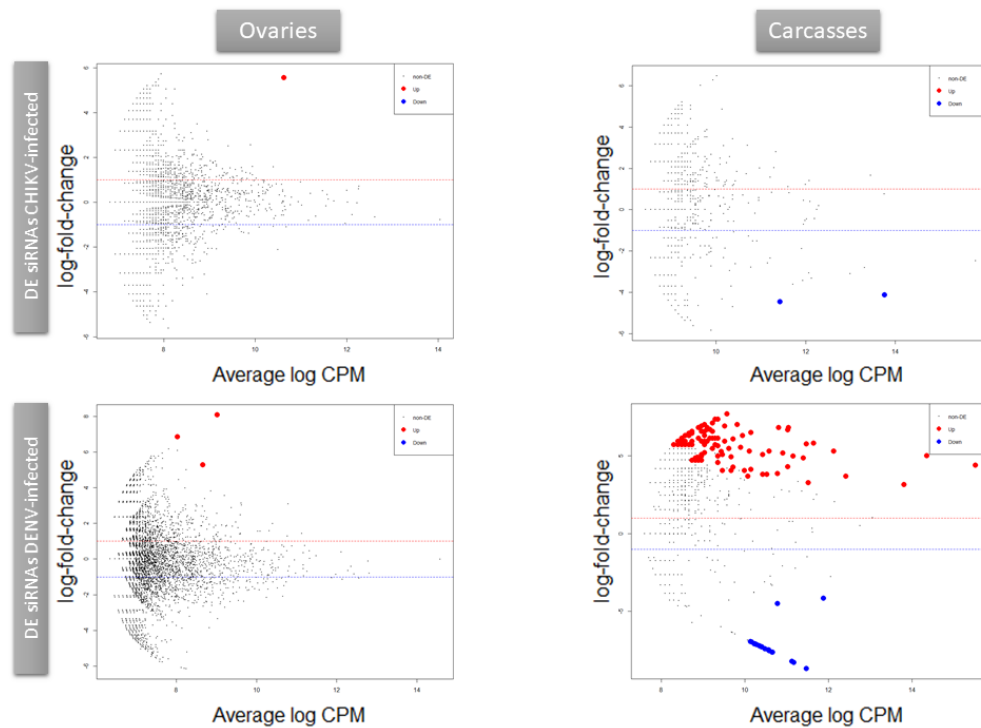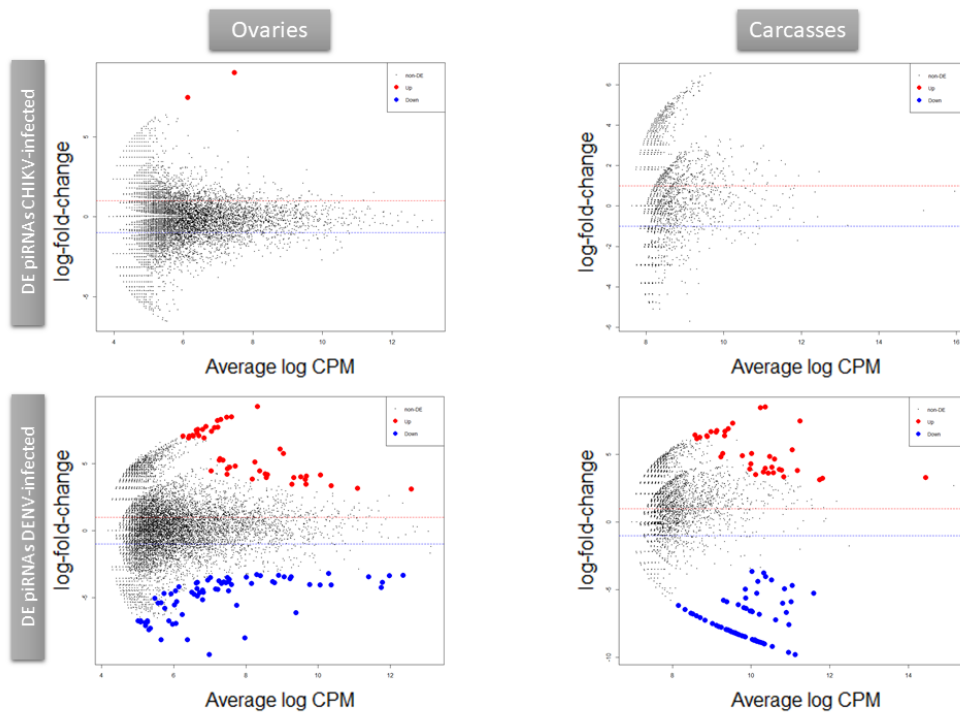
Figure 21. Differential expression (DE) of piRNAs represented by average log counts-per-million (CPM) on the x-axis and log fold-change on the y-axis. Red dots indicate significantly over-expressed piRNAs and blue dots significantly under-expressed piRNAs upon infection compared to blood-fed controls.
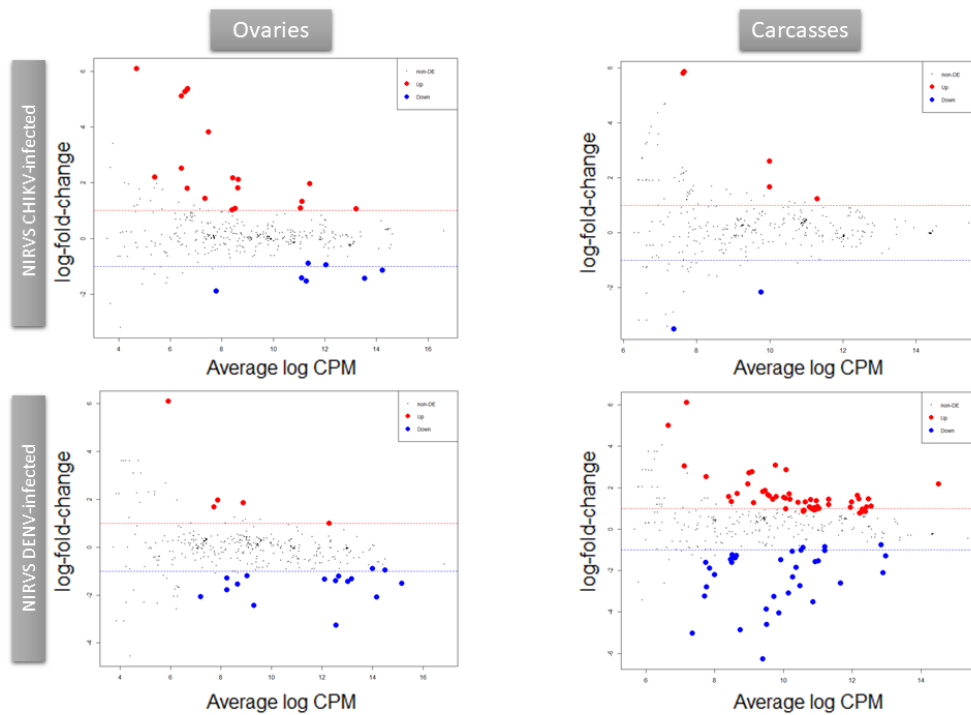
Figure 22. NIRVS profile based on piRNAs expression represented by average log counts-per-million (CPM) on the x-axis and log fold-change on the y-axis. Red dots represent NIRVS whose piRNA profile is significantly over-expressed and blue dots represent NIRVS whose piRNA profile is significantly under-expressed upon infection compared to blood-fed controls.

Figure 23 summarizes the number of NIRVS whose small RNA profile is significantly DE during infection and highlights the number of shared NIRVS among tissues and conditions.
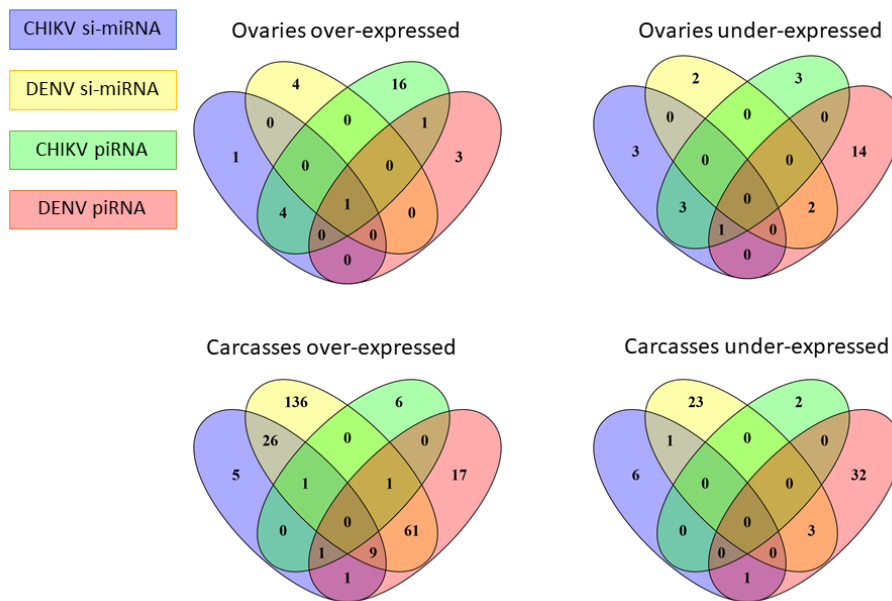
Figure 23. Venn diagram of shared NIRVS presenting DE small RNA profiles during infection compared to blood-fed controls in ovaries and carcasses. NIRVS are grouped according to tissue (ovaries, carcasses) and over- or under-expression of the small RNA profiles.

5.2.5 NIRVS-small RNAs and *Ae. albopictus* transcripts

To check for an effect on the virus, I re-blasted against the genomes of the CHIKV and DENV1, used in infection experiments, the NIRVS-small RNA molecules that had been found to be DE in infected vs control samples. However, I did not obtain any match, a not surprising result based on the low sequence similarity between *Ae. albopictus* NIRVS and either CHIKV or

85
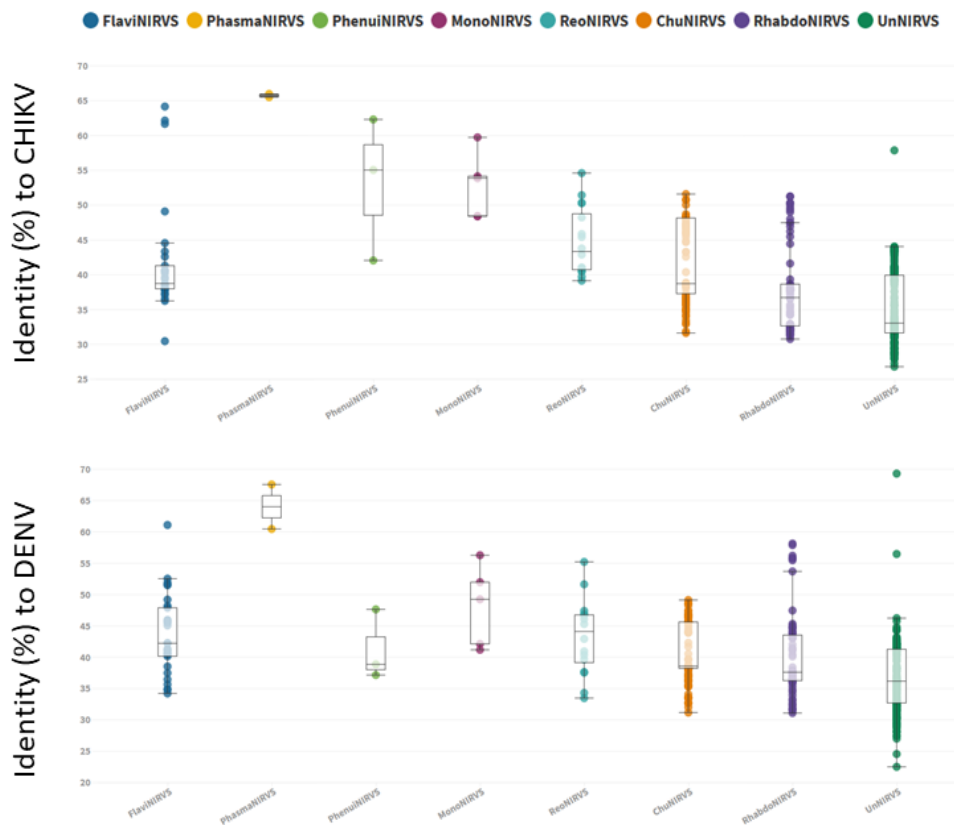
Figure 24. Box plot representing NIRVS identity (divided by viral family) with respect to the genome of the infecting Chikungunya and Dengue viruses.

DENV-1 (Figure 24). Because of this result, I tested whether these small RNAs could also map elsewhere in the *Ae. albopictus* genome. I retrieved the transcript sequences of all blast matches on the transcripts of *Ae. albopictus* and performed Gene ontology analyses using Blast2GO (Götz *et al.*, 2008) to find terms related to the likely targets of regulation (table 5).

Table 5. Detection of transcripts targeted by over- and under-expressed piRNAs and si-miRNAs across the different tested conditions. * indicates that for this condition, no small RNA significantly DE was found in previous analyses, and thus blast analysis was not performed.

Infection	Tissue	piRNAs		si-miRNAs	
		Over-expressed	Under-expressed	Over-expressed	Under-expressed
DENV	Ovaries	17	45	1	0*
	Carcasses	7	54	192	12
CHIKV	Ovaries	2	0*	0	0*
	Carcasses	0*	0*	0*	20

Hereafter, I describe the GO terms, focusing on biological process (P:) and molecular function (F:) of the identified transcripts. For clarity of exposition I describe transcripts identified as hits of DE piRNAs or si-miRNAs, respectively.

GO terms of transcripts that are blast hits for piRNAs. A total of 33 molecular function and 33 biological process GO terms were found associated to transcripts targeted by piRNAs DE between infected and control samples (Table 6A). These terms were variably recurrent. For molecular function, terms included RNA-directed-RNA-polymerase (RDRP), nucleic acid binding (found both in ovaries and carcasses) and helicase activity (found only in carcasses) in DENV-infected samples (Figure 25A,B). For biological process, in the "under-expressed piRNAs" group it is worth mentioning RNA templated transcription (in ovaries and carcasses), DNA integration and transmembrane transport (both found only in carcasses) (Figure 26A). These terms were not found in the "over-expressed piRNAs" group (Figure 26B*).*

A

piRNAs under-expressed in infected vs control

F:catalytic activity 1
F:hydrolase activity 1
F:RNA-directed 5'-3' RNA polymerase activity 5
F:protein binding 8
F:mRNA (guanine-N7-)-methyltransferase activity 4
F:metal ion binding 4
F:zinc ion binding 2
F:binding 2
F:ATP binding 12
F:RNA binding 3
F:nucleic acid binding 5
F:RNA helicase activity 2
F:aspartic-type endopeptidase activity 1
F:motor activity 1
F:protein kinase C activity 1
F:transaminase activity 1

DENV-infected ovaries

DENV-infected carcasses

B

piRNAs over-expressed in infected vs control

F:nucleic acid binding 1
F:RNA binding 1
F:RNA-directed 5'-3' RNA polymerase activity 4
F:ATP binding 3
F:mRNA (guanine-N7-)-methyltransferase activity 3
F:zinc ion binding 2
F:calcium ion binding 2
F:protein binding 2
F:oxidoreductase activity 1
F:chitin binding 1
F:scavenger receptor activity 1
F:serine-type endopeptidase activity 1

DENV-infected ovaries

DENV-infected carcasses

CHIKV-infected ovaries

Figure 25. Alluvial plot representing the connection between molecular function GO terms (left) and the analyzed conditions (right). Numbers reflect the total amount of time a same GO term was found. A) Results for the transcripts targeted by under-expressed piRNAs during infection. B) Results for the transcripts targeted by over-expressed piRNAs during infection.

A

B
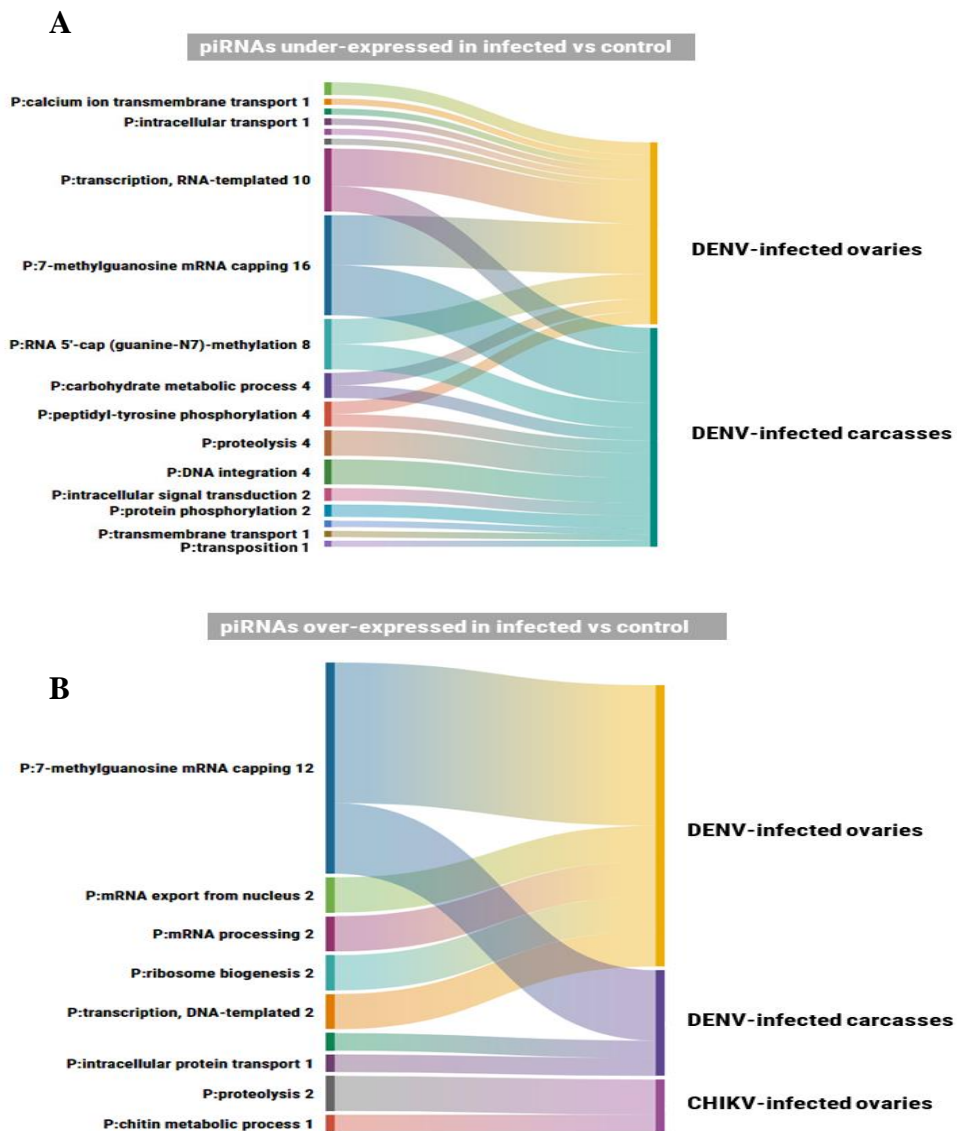
Figure 26. Alluvial plot representing the connection between biological process GO terms (left) and the analyzed conditions (right). Numbers reflect the total amount of time a same GO term was found. A) Results for the transcripts targeted by under-expressed piRNAs during infection. B) Results for the transcripts targeted by over-expressed piRNAs during infection.

GO terms of transcripts that are blast hits for si-miRNAs. For transcripts that are blast hit of NIRVS-si-miRNAs which were identified as DE in infected vs control samples, 23 molecular function and 20 biological process GO terms were retrieved (Table 6B). Mostly, these terms were found in the "over-expressed si-miRNAs" group for DENV-infected carcasses. However, RDRP was again found both in the over- and under-expressed groups and in both ovaries and carcasses (Figure 27A). For biological process, we again found terms related to RNA templated transcription, DNA integration and transposition, these last two only in carcasses (Figure 27B).

It is worth pointing out that terms related to **transcription, RNA- templated** and **RNA-directed 5'-3' RNA polymerase activity** refer to one transcript (AALF020122-RA), which is annotated as rhabdoviral-like polymerase and contains NIRVS_313.

Figure 27. Alluvial plot for molecular function (A) and biological process (B) GO terms of transcripts targeted by over or under-expressed si-miRNAs during DENV infection

A

| | DENV-infected | | | | CHIKV-infected |
| | Under-expressed piRNAs | | Over-expressed piRNAs | | |
Process	Ovaries	Carcasses	Ovaries	Carcasses	Ovaries
P:7-methylguanosine mRNA capping	2	2	2	1	0
P:endoplasmic reticulum to Golgi vesicle-mediated transport	0	0	0	1	0
P:intracellular protein transport	0	0	0	1	0
P:7-methylguanosine mRNA capping	2	2	2	1	0
P:mRNA export from nucleus	0	0	2	0	0
P:mRNA processing	0	0	2	0	0
P:ribosome biogenesis	0	0	2	0	0
P:transcription, DNA-templated	1	0	1	0	0
P:chitin metabolic process	0	0	0	0	1
P:proteolysis	0	2	0	0	1
P:transcription, RNA-templated	3	2	0	0	0
P:7-methylguanosine mRNA capping	2	2	2	1	0
P:RNA 5'-cap (guanine-N7)-methylation	2	2	0	0	0
P:calcium ion transmembrane transport	1	0	0	0	0
P:carbohydrate metabolic process	1	1	0	0	0
P:GPI anchor biosynthetic process	1	0	0	0	0
P:intracellular transport	1	0	0	0	0
P:peptidyl-tyrosine phosphorylation	1	1	0	0	0
P:proton transmembrane transport	1	0	0	0	0
P:regulation of transcription, DNA-templated	1	0	0	0	0
P:transcription, DNA-templated	1	0	1	0	0
P:7-methylguanosine mRNA capping	2	2	2	1	0
P:biosynthetic process	0	1	0	0	0
P:carbohydrate metabolic process	1	1	0	0	0
P:DNA integration	0	4	0	0	0
P:intracellular signal transduction	0	2	0	0	0
P:peptidyl-tyrosine phosphorylation	1	1	0	0	0
P:protein phosphorylation	0	2	0	0	0
P:proteolysis	0	2	0	0	1
P:RNA 5'-cap (guanine-N7)-methylation	2	2	0	0	0
P:transcription, RNA-templated	3	2	0	0	0
P:transmembrane transport	0	1	0	0	0
P:transposition	0	1	0	0	0

| | DENV-infected | | | | CHIKV-infected |
| | Under-expressed piRNAs | | Over-expressed piRNAs | | |
Function	Ovaries	Carcasses	Ovaries	Carcasses	Ovaries
F:ATP binding	4	8	2	1	0
F:protein binding	4	4	1	0	1
F:RNA-directed 5'-3' RNA polymerase activity	3	2	3	1	0
F:mRNA (guanine-N7-)-methyltransferase activity	2	2	2	1	0
F:nucleic acid binding	1	4	1	0	0
F:metal ion binding	2	2	0	0	0
F:RNA binding	1	2	1	0	0
F:zinc ion binding	1	1	1	1	0
F:calcium ion binding	1	0	1	1	0
F:beta-N-acetylhexosaminidase activity	1	1	0	0	0
F:binding	1	1	0	0	0
F:non-membrane spanning protein tyrosine kinase activity	1	1	0	0	0
F:RNA helicase activity	0	2	0	0	0
F:actin binding	0	1	0	0	0
F:aspartic-type endopeptidase activity	0	1	0	0	0
F:calcium transmembrane transporter activity, phosphorylative mechanism	1	0	0	0	0
F:catalytic activity	1	0	0	0	0
F:chitin binding	0	0	0	0	1
F:DNA-binding transcription factor activity	1	0	0	0	0
F:hydrolase activity	1	0	0	0	0
F:metalloendopeptidase activity	0	1	0	0	0
F:motor activity	0	1	0	0	0
F:oxidoreductase activity	0	0	0	1	0
F:protein dimerization activity	1	0	0	0	0
F:protein kinase activity	0	1	0	0	0
F:protein kinase C activity	0	1	0	0	0
F:proton-transporting ATP synthase activity, rotational mechanism	1	0	0	0	0
F:proton-transporting ATPase activity, rotational mechanism	1	0	0	0	0
F:pyridoxal phosphate binding	0	1	0	0	0
F:scavenger receptor activity	0	0	0	0	1
F:serine-type endopeptidase activity	0	0	0	0	1
F:transaminase activity	0	1	0	0	0
F:transmembrane transporter activity	0	1	0	0	0

92

B

Function	Over-expressed siRNAs		Under-expressed siRNAs
	Ovaries	Carcasses	
F:ARF guanyl-nucleotide exchange factor activity	0	1	0
F:aspartic-type endopeptidase activity	0	1	0
F:ATP binding	1	7	1
F:calcium transmembrane transporter activity, phosphorylative mechanism	0	1	0
F:catalytic activity	0	2	0
F:cystathionine beta-synthase activity	0	1	0
F:GTP binding	0	1	0
F:lipid transporter activity	0	1	0
F:mRNA (guanine-N7-)-methyltransferase activity	1	2	1
F:nuclear export signal receptor activity	0	1	0
F:nucleic acid binding	0	3	0
F:protein binding	0	2	0
F:protein kinase activity	0	1	0
F:proton-transporting ATP synthase activity, rotational mechanism	0	1	0
F:proton-transporting ATPase activity, rotational mechanism	0	1	0
F:Ran GTPase binding	0	1	0
F:RNA binding	0	2	0
F:RNA helicase activity	0	1	0
F:RNA-directed 5'-3' RNA polymerase activity	1	2	1
F:serine-type endopeptidase activity	0	2	0
F:serine-type endopeptidase inhibitor activity	0	1	0
F:translation elongation factor activity	0	1	0
F:zinc ion binding	0	2	0

Process	Over-expressed siRNAs		Under-expressed siRNAs
	Ovaries	Carcasses	
P:7-methylguanosine mRNA capping	1	2	1
P:calcium ion transmembrane transport	0	1	0
P:cysteine biosynthetic process from serine	0	1	0
P:cysteine biosynthetic process via cystathionine	0	1	0
P:DNA integration	0	1	0
P:intracellular protein transport	0	2	0
P:lipid transport	0	1	0
P:negative regulation of endopeptidase activity	0	1	0
P:nuclear export	0	1	0
P:protein phosphorylation	0	1	0
P:protein refolding	0	1	0
P:proteolysis	0	3	0
P:proton transmembrane transport	0	1	0
P:regulation of alternative mRNA splicing, via spliceosome	0	1	0
P:regulation of ARF protein signal transduction	0	1	0
P:RNA 5'-cap (guanine-N7)-methylation	1	2	1
P:transcription, RNA-templated	1	2	1
P:translational elongation	0	1	0
P:transposition	0	1	0
P:vesicle-mediated transport	0	1	0

Table 6. Gene ontology terms retrieved for transcripts targeted by DE piRNAs (A) and si-miRNAs (B) during infection. GO terms related to molecular function and biological process are reported.

5.3 Life history traits of *Ae. albopictus*

Quantifying mosquito fitness traits is key to predict epidemics of mosquito-borne diseases and to organize effective control programs. Due to the recent expansion of *Ae. albopictus* on global scale and the different climatic regions where this mosquito species now dwells, different life history traits are likely to have arisen. To gain insight on such differences, we analyzed and compared the life history traits of three field-collected populations (Crema, Cr; Thailand, Ta and La Reunion island, LaR) and two laboratory strains (Foshan, Fo and Rimini, Ri).

Overall, average larval-to-pupal developmental time was of 6.738 days ± 0.793. Development of Fo and Ri mosquitoes was generally delayed one day than that of mosquitoes from the three recently-derived strains (Figure 28); differences were statistically significant between Fo and Cr mosquitoes (H=4.48, p=0.0034) and between Ri and both Ta (H=6.48, p=0.011) and Cr mosquitoes (H=7.406, p=0.0065). Larval viability was > 0.9 in all strains and emerging adults showed a significantly higher percentage of males or females in the Fo and Ta strains, respectively (Table 7).
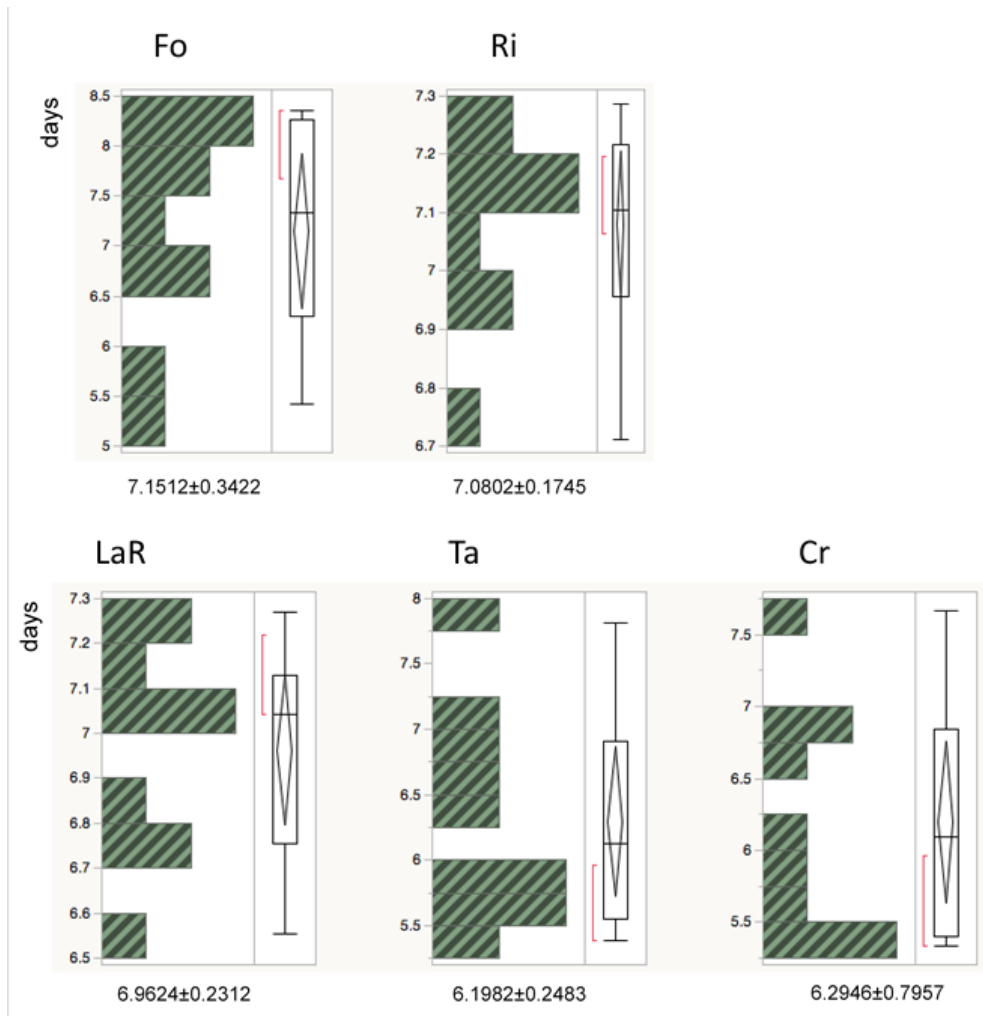
Figure 28. Distribution of the larval-to-pupal developmental time across ten replicates of 50 larvae each for each strain.

Table 7. Distribution of the larval-to-pupal developmental time across ten replicates of 50 larvae each for each strain.

strain	Larval viability[1]	Males
Fo	0.976±0.020	25.5±3.47 (51%)
Ri	0.928±0.037	23.4±3.20 (46.8%)
LaR	0.928±0.042	23.9±2.91 (47.8%)
Ta	0.958±0.037	22.7±2.31 (45.2%)
Cr	0.948±0.010	23.2±1.72 (46.4%)

strain	Pupae sex ratio[2]	
	Females	Mann-Whitney test
Fo	21±2.58 (42%);	Z=-2.65; p=0.008
Ri	23±2.58 (46%);	Z=-0.26; p=0.795
LaR	24±2.41 (48%);	Z=0; p=1
Ta	26.1±2.18 (52.2%);	Z=2.68; p=0.007
Cr	24.2±1.66 (48.4%);	Z=0.98; p=0.327

[1]Larval viability is the proportion of larvae becoming pupae
[2]Sex was scored when adult emerged, starting from 50 larvae in each of the 10 pans/strain. Numbers of females and males do not add to 50 because of larval and pupal mortality.

Average wingspan was 1.924±0.274 mm for males and 2.441±0.254 for females. Fo and Ri mosquitoes showed the smallest and largest wingspan across all analyzed strains, respectively. In all strains, females were significantly larger than males (Table 8A). Differences in wingspan were statistically significant also across strains, with the exception of mosquitoes from the LaR and Cr strains (Table 8B and 8C).

Survival curves for male and female mosquitoes of each strain are shown in Figure 29. In all strains, females had longer average life-span (50.0996 days ± 1.1356) than males (45.0197 days ± 1.2095) (Table 9). Survival curves of

> **Table 8.** Results of T test-based comparisons of wing length between males and females within strains (A), between males across strains (B) and between females across strains (C).

A-Within strains, males and females

Strains	Males (mean ± SD)	Females (mean ± SD)	T-test (t-Ratio, prob.)
Fo	1.547±0.074	2.074±0.066	-32.548, <0.001
Ri	2.264±0.063	2.784±0.084	-37.949, <0.001
LaR	2.066±0.069	2.542±0.070	-47.009, <0.001
Ta	1.680±0.073	2.294±0.109	-30.089, <0.001
Cr	2.054±0.054	2.512±0.075	-41.266, <0.001

B-Across strains, males (t-Ratio, probability)

	Fo	Ri	LaR	Ta
Fo				
Ri	57.349, <0.001			
LaR	30.188, <0.001	-18.021, <0.001		
Ta	11.608, <0.001	-22.867 <0.001	-27.571, <0.001	
Cr	29.562, <0.001	-16.562 <0.001	-1.137, 0.261	31.017, <0.001

C-Across strains, females (t-Ratio, probability)

	Fo	Ri	LaR	Ta
Fo				
Ri	48.502, <0.001			
LaR	37.228, <0.001	-15.697, <0.001		
Ta	10.0748, <0.001	-45.299 <0.001	-12.497, <0.001	
Cr	45.246, <0.001	-17.697 <0.001	-2.132, 0.038	10.934, <0.001

females showed the majority of mosquitoes starting to dye after 40 days, whereas survival curves of males showed a more gradual mortality over time. Significant differences between sexes were observed in the two long-adapted strains, Foshan and Rimini, and in the La Reunion strain (Table 9A). Survival curves of females of the Rimini strain were statistically-significantly different than those of all recently-derived strains, with Rimini females have a longer average survival time.
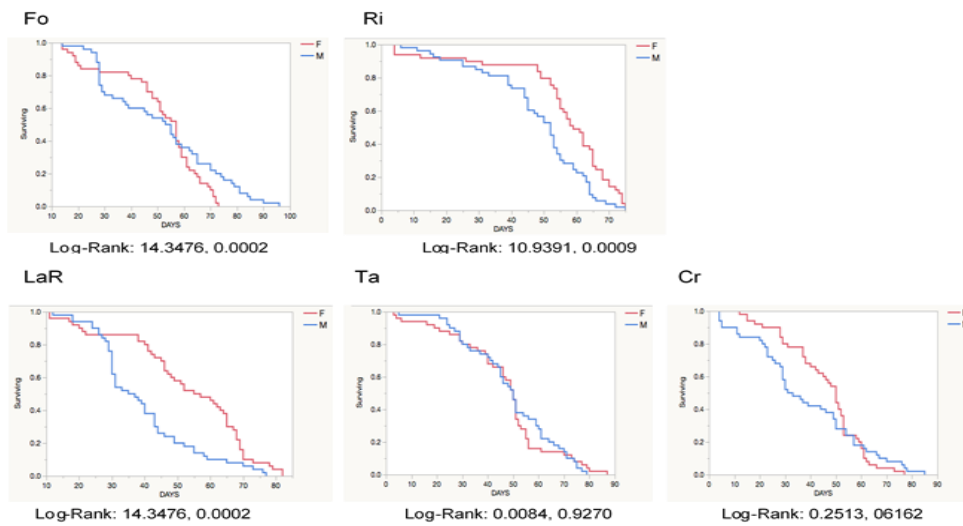
Figure 29. Survival curves of males (M) and females (F) of each strain. The Log-Rank values (ChiSquare, probability) for the comparison of the survival curves between males and females are shown below each figure.

98

5.4 Side projects

Based on my expertise in population genetics and analyses of small RNAs, I was involved in a number of projects, besides my main phD project. I briefly summarize here the two main side projects I contributed to.

5.4.1 Pyrethroid resistance in *Ae. albopictus* mosquitoes

I worked with Dr Alessandra Tancredi in her project on tracing the temporal and geographical distribution of resistance to pyrethroids in *Ae. albopictus* mosquitoes. Application of pyrethroids is among the most widely-used interventions for vector control, especially in the presence of an arboviral outbreak. Studies are emerging that reveal phenotypic resistance in *Ae. albopictus* populations and monitor mutations at the target site, the *para* sodium channel gene, primarily on a local scale (Kamgang *et al.*, 2011; Vontas *et al.*, 2012; Karunaratne *et al.*, 2013; Ngoagouni *et al.*, 2016; Smith *et al.*, 2016; Moyes *et al.*, 2017; Estep *et al.*, 2018; Pichler *et al.*, 2018). Dr. Tancredi genotyped 512 *Ae. albopictus* mosquitoes sampled between 2011 and 2018 from twelve geographic sites, including those from the native home range and invaded areas, at the all currently known five codons of the *para* sodium channel gene with mutations predictive of resistance phenotype to pyrethroids in *Aedes* spp. mosquitoes (hereafter called *kdr* mutations) (Du *et al.*, 2015; Haddi *et al.*, 2017; Moyes *et al.*, 2017; Saavedra-Rodriguez *et al.*, 2018). Besides analyzing the geographic widespread of *kdr* mutations in *Ae. albopictus* populations, we also wanted to trace the origin of *kdr* mutations to test whether *kdr* mutations arose once and then spread, or they evolved independently in different populations. Among our samples, we had several

mosquitoes from Mexico, which showed unique *kdr* mutations (i.e. mutations at position 410 and 989). We focused on this sample and tested whether the origin of Mexican mosquitoes as derived by microsatellite data correlates with their *kdr* haplotypes. To do this I determined the genetic make-up of Mexican mosquitoes using microsatellite markers and analyzed the genetic connectivity of southern Mexico mosquitoes with mosquitoes from home range, the Reunion Island, America and Europe.

Levels of genetic variability in mosquitoes from Mexico were comparable to what detected in previously-analyzed worldwide mosquitoes (Manni *et al.*, 2017). For instance, the number of effective alleles/locus ranged between 1.687 (Hawaii) and 3.841 (Thailand), with a value of 2.307 for Chiapas; the observed heterozygosity ranged between 0.283 (Hawaii) and 0.520 (La Reunion Island), with a value of 0.472 for Mexico. Despite mosquitoes from Mexico had a similar number of effective alleles as other tested populations, they also showed private alleles at different loci (i.e. Aealbmic2; Aealbmic3; Aealbmic17; Aealbmic9), a mark of genetic distinctiveness. Fst values of mosquitoes from Chiapas were lower in comparison to those of mosquitoes from other invasive populations such as northern Italy, Virginia and Greece, suggesting invasion into southern Mexico is a secondary invasion event, possibly from different regions (Figure 30A) (Pech-May *et al.*, 2016). I further examined genetic structuring of individuals using the bayesian method implemented in STRUCTURE (Evanno *et al.*, 2005). The natural logarithm of the likelihood of the data, following the Evanno method (ΔK) (Evanno *et al.*, 2005), showed two picks at K = 2 and K = 6, suggesting two scenarios of genetic clustering for the tested populations (Figure 30B). Individual
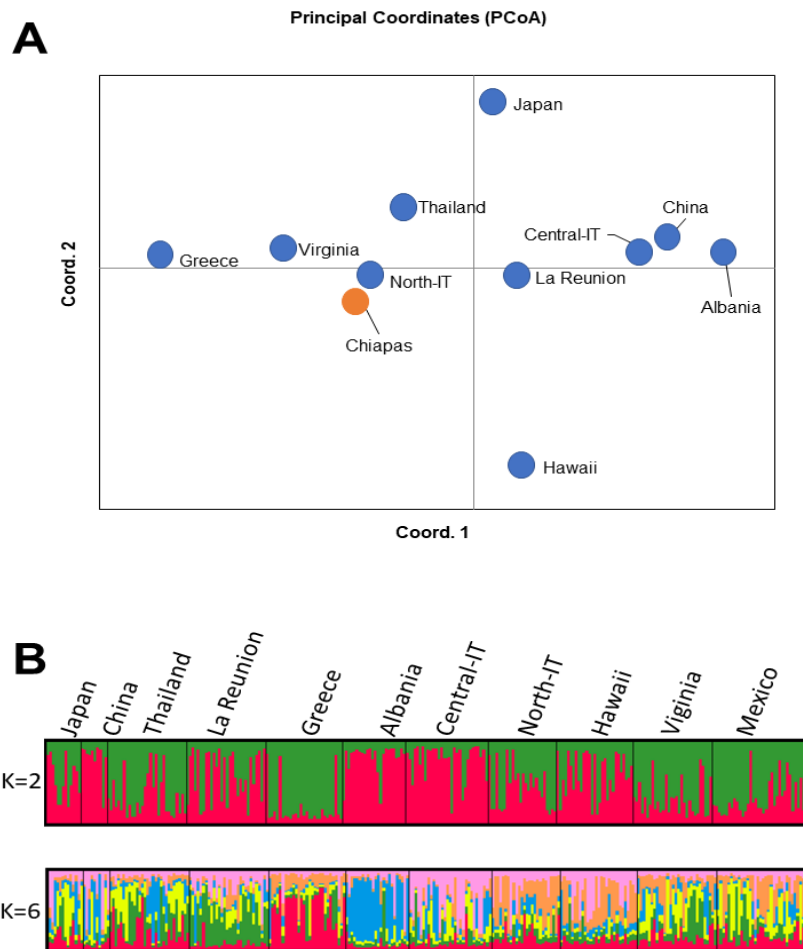
Figure 30. Genetic structure of *Ae. albopictus* populations as revealed by microsatellite markers. A) Principal coordinate analysis (PCOA) based on the calculated Fst values. Mosquitoes from Chiapas as shown by an orange dot. B) Graphical representation of the Bayesian cluster analysis using STRUCTURE with K=2 and K=6.

mosquitoes from the examined populations were assigned to each cluster with a certain probability value for the two scenarios of 2 and 6 clusters. In both

scenarios, extensive genetic admixture among *Ae. albopictus* geographic samples was evident as previously observed (Kotsakiozi *et al.*, 2017; Manni *et al.*, 2017) (Figure 30B). Regarding mosquitoes from Mexico, in the two-cluster model, they clustered with Virginia, Greece and north Italy. In the six-cluster model, the Mexican mosquito population appeared divided into subgroups showing similarities with mosquitoes from either Virginia, Greece or north Italy, supporting the hypothesis of secondary invasions from multiple regions.

Mosquitoes from Mexico showed a pattern of PY-resistance mutations different than that of mosquitoes from Virginia, Greece and northern Italy to which they were most genetically-close, in support of the hypothesis of independent and local origin of *kdr* mutations.

5.4.2 Probing NIRVS adaptation and resistance to infections with cognate viruses

I worked with Dr. Cristina Crava, a postdoc in the laboratory in her project on viral integrations in *Ae. aegypti* mosquitoes. We previously demonstrated that NIRVS are more abundant in *Aedes* vs *Anopheles* mosquitoes and they are not distributed homogeneously across the genome, but they are embedded within TE fragments in piRNA clusters and produce piRNAs (Palatini *et al.*, 2017). Additionally, we previously showed that NIRVS-piRNAs are mostly antisense to the incoming viral genomes, establishing the basis for targeted processing of cognate viruses (Palatini *et al.*, 2017).

The hypothesis of the involvement of NIRVS in immunity in mosquitoes through the piRNA pathway entail that the landscape of NIRVS across mosquito populations should be different depending on viral exposure. Alternatively, NIRVS landscape could be conserved, with few sequences providing adaptive advantage across a wide range of viruses.

Dr. Crava analyzed the patter of distribution of NIRVS in wild-collected *Ae. aegypti* mosquitoes and identified novel viral integrations. I contributed to her project by doing a detailed analysis of the piRNA profile across *Ae. aegypti* NIRVS to test whether the presence of different NIRVSs corresponding to the same region of the viral genome would favor amplification of the piRNA signal.

Briefly, I mapped deep-sequencing data from published resources (SRR5961506 and SRR5961505), including samples from carcasses and ovaries, on NIRVS sequences using bowtie with a minimum seed match of 18nt. Aligned reads were filtered by length using BBMap reformat.sh (https://sourceforge.net/projects/bbmap/), keeping only piRNA sized reads (23-32nt). I performed another filtering step using custom scripts to select all reads starting with T to retain only primary piRNAs (1U bias) (Wang *et al.*, 2014). BEDTools Intersect (Quinlan and Hall, 2010) removed reads mapping outside annotated NIRVS. Finally, all reads with 100% identity were collapsed with fastx-toolkit (Gordon *et al.*, 2014) and used for quantification. Due to sequence similarity and overlap among NIRVS, it is impossible to quantify reads mapping uniquely to most NIRVS. To avoid any bias, I used custom scripts to quantify each piRNA size sequence in each original fastq file and then identified all the NIRVSs to which each piRNA sequence could

be mapped to. Counts for each experiment were normalized based on the library size by Quantile-to-Quantile Normalization as implemented in edgeR (Robinson *et al.*, 2009). Normalized counts of Flavi-derived and Rhabdo-derived piRNA sized sequences were mapped to the Xishuangbanna virus and the Ohlsdorf virus, respectively, transposing the position of each small RNA from the first-hit NIRVSs to the relative position on the corresponding virus. The data was visualized as bed graphs in Integrative Genomics Viewer (IGV) (Robinson *et al.*, 2011).

Through these analyses, I observed that piRNAs are not evenly distributed, but spike in distinct NIRVSs. In the case of FlaviNIRVS, the highest pick of piRNAs was seen in the regions where more NIRVS overlap, such as the regions around the NS1 coding sequence and the regions that code for S7 and NS3, which contains the NTPase/helicase motif (Figure 31). Additionally, the piRNA hotspots are mostly shared between carcasses and ovaries, with signals from carcasses being quantitatively higher. On the contrary, piRNAs from RhabdoNIRVS were distributed across the nucleoprotein, glycoprotein and polymerase encoding sequences and signals in ovaries were generally quantitatively higher than that in the carcasses. A discontinuous profile with picks is similar to that observed for viral siRNAs (vsiRNAs) produced after arboviral infection of mosquitoes (Miesen *et al.*, 2016a), even if hotspots of vpiRNAs and NIRVS-derived piRNAS are different. To check for NIRVS-piRNA overlapping across different viral species, NIRVS-piRNAs were re-blasted against our viral database. Interestingly, different piRNA sequences deriving from FlaviNIRVS have their best viral hit against arboviruses and

piRNAs from two NIRVS blast against various regions of the viral genome supporting the hypothesis that NIRVSs are a redundant system.
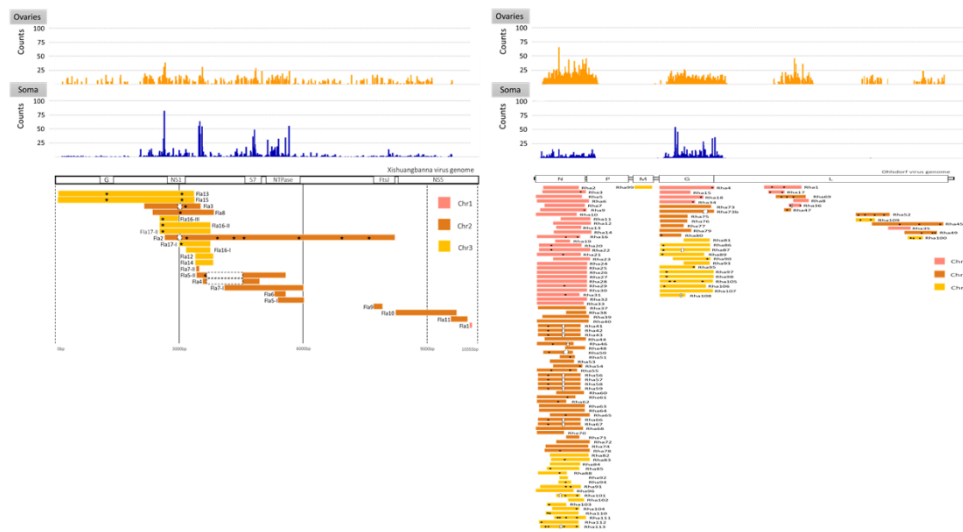
Figure 31. Representation of NIRVS position along an ideal flavivirus (left) and arbovirus (right). Barplots on top indicate the normalized counts of piRNAs produced from the corresponding NIRVS.

6. Discussion

6.1 *Piwi* genes

Recent studies indicate that the piRNA pathway may be involved in antiviral immunity in *Aedes* spp. mosquitoes (Miesen *et al.*, 2016b; Petit *et al.*, 2016). This new role of the piRNA pathway is likely to depend on the expansion and subsequent functional specialization of its core components (Campbell *et al.*, 2008b; Hess *et al.*, 2011; Miesen *et al.*, 2016b). Studies in *Ae. aegypti* identified one *Ago3* and seven *Piwi* genes (i.e. *Piwi1, Piwi2, Piwi3, Piwi4, Piwi5, Piwi6, Piwi7*), which display differences in their tissue and development expression profile and associate with either TE-derived or viral piRNAs (Akbari *et al.*, 2013; Miesen *et al.*, 2015; Varjak *et al.*, 2017). These studies were based on the knowledge of the gene structure of each *Ae. aegypti Piwi* genes and implemented RNAi-based silencing experiments and *in vitro* expression assays, but lacked an evolutionary perspective (Miesen *et al.*, 2015; Miesen *et al.*, 2016a; Girardi *et al.*, 2017; Varjak *et al.*, 2017). A possibility that links the PIWI pathway to antiviral defense resides in the recent discovery of NIRVS in *Aedes spp.* mosquitoes (Palatini *et al.*, 2017), in particular in reference to NIRVS located in piRNA clusters, involved in the production of precursor piRNA transcripts.

In my PhD thesis work I focused on the emerging arboviral vector *Ae. albopictus* to unravel the structure and expression profile of the *Piwi* genes in this mosquito. I found that the genome of *Ae. albopictus* harbors one copy of

Ago3 and six *Piwi* genes (i.e. *Piwi1/3, Piwi2, Piwi4, Piwi5, Piwi6 and Piwi7*), These genes are one-to-one orthologues to the *Piwi* genes of *Ae. aegypti*, with the only exceptions of *Piwi2* and *Piwi1/3*, where the two genes from the same species cluster together. In *Ae. aegypti*, *Piwi2* and *Piwi3* are located on Chromosome 1, separated by ~ 20kb, suggesting they may undergo frequent gene conversion.

In collaboration with Prof. Forneris, I used homology modelling to build predictions of molecular architectures for the *Piwi* genes of *Ae. albopictus* and define the boundaries of the PAZ, MID and PIWI domains, which are the hallmarks of the Argonaute protein family (Joshua-Tor, 2006). Superpositions and sequence comparisons indicate that all the *Piwi* genes have similar 3d architecture and retain high levels of conservation in the RNA binding pocket, including the characteristic DDH triad in the catalytic site of the PIWI domain. These results suggest that all the proteins likely exert slicing activity, although tailored experiments and confirmations are required. The expression profile of the *Piwi* genes indicates that these genes are particularly important during embryogenesis. Four of them (i.e. *Ago3, Piwi1/3, Piwi2 and Piwi6*) are highly expressed also in adult females, in particular in ovaries and following blood-feeding. The expression of *Piwi* genes was also elicited upon arboviral infection, indirectly confirming the antiviral role of the piRNA pathway. Differences in expression were found depending on tissues, species of the infecting virus and day post infection. In CHIKV-infected samples, expression of *Piwi* genes was mostly elicited after 4 days post infection in the ovaries, while in the carcasses the expression values at the same time point were even lower than in controls. After 14 dpi,

in CHIKV-infected ovaries only *Ago3* remained significantly up-regulated, while carcasses displayed a general increase in *Piwi* genes expression, particularly for *Piwi5*. Differently from CHIKV infection, the highest values of *Piwi* genes expression in DENV-infected samples was seen in carcasses 4 dpi. These results are concordant with the timing in piRNAs accumulation previously described for CHIKV or DENV infection in *Aedes* spp. mosquitoes: infection experiments with CHIKV in *Ae. albopictus* showed that secondary piRNAs are not found 3 dpi, but are enriched 9 dpi (Goic *et al.*, 2016). In contrast, experiments with DENV2 in *Ae. aegypti* mosquitoes indicate that piRNAs are the dominant small RNA populations 2 dpi (Hess *et al.*, 2011). Overall, these observations and our expression analyses support the hypothesis of an early activation of the piRNA pathway following DENV infection limited to carcasses, while CHIKV infection leads to an early activation in ovaries and a late activation in carcasses. *Piwi5* was elicited both during DENV and CHIKV infection, in line with studies in *Ae. aegypti* that ascribed a general antiviral activity to the same protein (Miesen *et al.*, 2015). However, our results indicate that other *Piwi* genes such as *Ago3*, *Piwi1/3* and *Piwi6* can be elicited during infection and might have a more prominent role, which will need further investigation.

6.2 small RNAs and NIRVS

Recent studies have shown evidence that eukaryotic genomes, including mosquitoes, harbor integrated viral sequences from non-retroviral RNA viruses now called NIRVS (Blair *et al.*, 2019). The biological role of these

sequences is currently under investigation and an intriguing hypothesis suggests that NIRVS are remnants of past infections involved in shaping antiviral immunity and, with it, vector competence (Flegel, 2007; Flegel and Sritunyalucksana, 2011; Park *et al.*, 2013; Goic *et al.*, 2016; Miesen *et al.*, 2016b; Olson and Bonizzoni, 2017; Palatini *et al.*, 2017).

In my PhD program I studied small RNAs production during persistent infection with the Dengue-1 and Chikungunya viruses and I observed that both 26-32nt sized piRNAs and 20-25nt sized small RNAs (si-miRNAs) are produced to target the infecting virus, and in particular its negative strand (i.e. the replicative form of both CHIKV and DENV). Small RNA molecules are also identified mapping to NIRVS and some of them appear to be up or down regulated in infected vs control samples. In total, 78% of the DE NIRVS-si-miRNAs starts with U, a typical sign of Dicer processing (Starega-Roslan *et al.*, 2015). The profile of small RNAs on the viral genome and that on NIRVS is similarly organised in hotspots, but while si-miRNAs are quantitatively more than piRNAs on the viral genome, the opposite is true for NIRVS. NIRVS-piRNAs and NIRVS-si-miRNAs do not target the infecting viral sequence, rather they are found to map on endogenous transcripts related to a plethora of Gene ontology terms. These results suggest a regulatory role of NIRVS in *Ae. albopictus* by their effect, via RNAi, on endogenous transcripts in response to arboviral infection. These results also connect NIRVS to recent studies in which profiling of the transcriptional changes induced in *Aedes spp.* mosquitoes by viral infection has shown that a large number of elicited genes act outside the core immunity pathways in ways yet to be explored (Behura *et al.*, 2011; Colpitts *et al.*, 2011; Bonizzoni *et al.*, 2012; Zhao *et al.*, 2019)

and may represent candidate antiviral target genes (Feng *et al.*, 2017; Terradas and McGraw, 2019; Zhao *et al.*, 2019). Recent studies highlighted the importance of several gene types/families involved in different aspects of the antiviral response in mosquitoes, including virus-cytoskeleton interactions, cellular trafficking and chitin metabolism (Sigle and McGraw, 2019). Indeed, it was shown that proteins such as actin, tubulin, motor proteins and myosin can have altered expression profile during infection with DENV, CHIKV and WNV (Chee and AbuBakar, 2004; Colpitts *et al.*, 2011; Bonizzoni *et al.*, 2012; Shrinet *et al.*, 2017), and chitin genes were found to be differentially expressed upon infection with DENV, YFV and WNV (Colpitts *et al.*, 2011). Regulation of miRNAs has also been demonstrated in *Aedes* spp mosquitoes during infection with DENV, CHIKV and ZIKV and bioinformatic predictions of miRNAs target and gene ontology analyses provided a wide variety of candidate genes, which include Toll receptors, serine proteases, scavenger receptors, endopeptidases, dead-box helicases, V-ATP synthase, ubiquitin, G proteins and many more (Liu *et al.*, 2015; Maharaj *et al.*, 2015; Dubey *et al.*, 2017; Saldaña *et al.*, 2017). In accordance with this data, I found that NIRVS-encoded small RNAs may target genes related to actin and chitin binding, motor activity, ATP synthase activity, scavenger receptor activity, intracellular transport, helicase activity, GTP binding and serine endopeptidase activity, suggesting that NIRVS may be regulators of several aspects of mosquito immunity.

My GO analysis also revealed terms related to viral replication, such as RDRP and RNA templated transcription, other GO terms relate to molecular

mechanisms that may lead to NIRVS formation during infection, such as transposition and DNA integration (Goic *et al.*, 2016; Poirier *et al.*, 2018). Several studies have shown that viruses can exploit host miRNAs for their benefit, and miRNAs-target binding, although generally considered to lead to negative regulation, can also lead to positive regulation (Asgari, 2014). Therefore, a major goal of future studies will be to assess the functional outcome of the regulation of endogenous genes upon infection and shed light on the possible interactions between antiviral defense and viral exploitation of the RNAi mechanisms.

7. Conclusions and perspectives

The role of the piRNA pathway in mosquitoes goes beyond the preservation of genomic integrity as shown by works in *D. melanogaster* and may represent a key component of vector competence (Akbari *et al.*, 2013; Miesen *et al.*, 2016b; Olson and Bonizzoni, 2017; Matthews *et al.*, 2018). Studies in *Ae. aegypti* have shed light on the antiviral role of the different components of the Piwi clade proteins and explored the possibility of interaction with the siRNA pathway (Schnettler *et al.*, 2013; Miesen *et al.*, 2016a; Varjak *et al.*, 2017). In this work, I expand the knowledge on the *Piwi* genes of *Ae. albopcitus*, a highly invasive species and competent vector for arboviral pathogens. My results confirm elicitation of the piRNA pathway upon arboviral infection, reveal the presence of polymorphic sites and an overall conserved organization of domains, so that further detailed studies are required to clarify the role of the different Piwi proteins in response to different viral species and in different tissues.

In the second part of my work, I show that the piRNA pathway is connected to integrated viral sequences named NIRVS. The transfer of genetic material between separate evolutionary lineages (horizontal transfer) has had an important role in the evolution of genomes. Despite being overshadowed by the higher prevalence and more understood gene transfer among prokaryotes, the exchange of genetic material from DNA viruses and retroviruses to eukaryotic cells is a well-recognised phenomenon.

Recent discoveries expand the range of viral types that can transfer genetic material to their hosts to non-retroviral RNA viruses. The frequency of this

phenomenon and its biological relevance are not fully understood. Studies in different animals have shown that integrated viral sequences can influence immunity responses against similar viruses, such as in bees, squirrels and shrimps (Flegel, 2007; Maori *et al.*, 2007; Fujino *et al.*, 2014), or even act against bacteria (Di Lelio *et al.*, 2019). My work shows that NIRVS play a role during infection in *Ae. albopictus,* producing piRNAs and siRNAs/miRNAs, and helps identify possible target genes for a more focused analysis. Future studies will aim both at a further classification of the different types of NIRVS-derived small RNAs that I identified (i.e. distinction of the signal between siRNAs and miRNAs) and the assessment of the role of identified target genes in vector competence. To this end, RNAseq data from the infected samples that were analysed for their small RNA profile will be produced and analysed to confirm the predictions of the effect of small RNAs on the identified candidate genes. Additionally, genetic modification of NIRVS is in progress to address their potential role during infection. My work allowed to identify the ideal NIRVS candidate to target, such as NIRVS with DE small RNA profile that are shared across conditions, NIRVS that target transcripts with GO terms related to antiviral activity, and others, like NIRVS_313, a Rhabdo-NIRVS that targets an endogenous gene encoding for a RDRP.

References

Abascal, F., Zardoya, R. and Telford, M. J. (2010) 'TranslatorX: Multiple alignment of nucleotide sequences guided by amino acid translations', *Nucleic Acids Research*. doi: 10.1093/nar/gkq291.

Adams, P. D. *et al.* (2011) 'The Phenix software for automated determination of macromolecular structures', *Methods*. doi: 10.1016/j.ymeth.2011.07.005.

Adelman, Z. N., Anderson, M. A. E., Wiley, M. R., Murreddu, M. G., Samuel, G. H., Morazzani, E. M. and Myles, K. M. (2013) 'Cooler Temperatures Destabilize RNA Interference and Increase Susceptibility of Disease Vector Mosquitoes to Viral Infection', *PLoS Neglected Tropical Diseases*. doi: 10.1371/journal.pntd.0002239.

Agarwal, A., Parida, M. and Dash, P. K. (2017) 'Impact of transmission cycles and vector competence on global expansion and emergence of arboviruses', *Reviews in Medical Virology*. doi: 10.1002/rmv.1941.

Akbari, O. S., Antoshechkin, I., Amrhein, H., Williams, B., Diloreto, R., Sandler, J. and Hay, B. A. (2013) 'The Developmental Transcriptome of the Mosquito *Aedes aegypti* , an Invasive Species and Major Arbovirus Vector', *G3&#58; Genes|Genomes|Genetics*. doi: 10.1534/g3.113.006742.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) 'Basic local alignment search tool', *Journal of Molecular Biology*. doi: 10.1016/S0022-2836(05)80360-2.

Amraoui, F., Ayed, W. Ben, Madec, Y., Faraj, C., Himmi, O., Btissam, A., Sarih, M. and Failloux, A. B. (2019) 'Potential of aedes albopictus to cause the emergence of arboviruses in Morocco', *PLoS Neglected Tropical Diseases*. doi: 10.1371/journal.pntd.0006997.

Anderson, J. R. and Rico-Hesse, R. (2006) 'Aedes aegypti vectorial capacity is determined by the infecting genotype of dengue virus', *American Journal of Tropical Medicine and Hygiene*.

Angleró-Rodríguez, Y. I., MacLeod, H. J., Kang, S., Carlson, J. S., Jupatanakul, N. and Dimopoulos, G. (2017) 'Aedes aegypti molecular responses to Zika Virus: Modulation of infection by the toll and Jak/Stat immune pathways and virus host factors', *Frontiers in Microbiology*. doi: 10.3389/fmicb.2017.02050.

Arensburger, P., Hice, R. H., Wright, J. A., Craig, N. L. and Atkinson, P. W. (2011) '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*, 12. doi: 10.1186/1471-2164-12-606.

Asgari, S. (2014) 'Role of microRNAs in vrbovirus/vector interactions', *Viruses*. doi: 10.3390/v6093514.

Bartel, D. P. (2004) 'MicroRNAs: Genomics, Biogenesis, Mechanism, and Function', *Cell*. doi: 10.1016/S0092-8674(04)00045-5.

Bartel, D. P. (2009) 'MicroRNAs: Target Recognition and Regulatory Functions', *Cell*. doi: 10.1016/j.cell.2009.01.002.

Bartholomay, L. C., Fuchs, J. F., Cheng, L. L., Beck, E. T., Vizioli, J., Lowenberger, C. and Christensen, B. M. (2004) 'Reassessing the role of defensin in the innate immune response of the mosquito, Aedes aegypti', *Insect Molecular Biology*. doi: 10.1111/j.0962-1075.2004.00467.x.

Baruffi, L., Damiani, G., Guglielmino, C. R., Bandii, C., Malacrida, A. R. and Gasperi, G. (1995) 'Polymorphism within and between populations of ceratitis: Comparison between RAPD and multilocus enzyme electrophoresis data', *Heredity*, 74(4), pp. 425–437. doi: 10.1038/hdy.1995.60.

Beckham, J. D., Pastula, D. M., Massey, A. and Tyler, K. L. (2016) 'Zika Virus as an Emerging Global Pathogen', *JAMA Neurology*. doi: 10.1001/jamaneurol.2016.0800.

Behura, S. K., Gomez-Machorro, C., Harker, B. W., deBruyn, B., Lovin, D. D., Hemme, R. R., Mori, A., Romero-Severson, J. and Severson, D. W. (2011) 'Global cross-talk of genes of the mosquito Aedes aegypti in response to dengue virus infection', *PLoS Neglected Tropical Diseases*. doi: 10.1371/journal.pntd.0001385.

Benedict, M. Q., Levine, R. S., Hawley, W. A. and Lounibos, L. P. (2007) 'Spread of The Tiger: Global Risk of Invasion by The Mosquito *Aedes albopictus*', *Vector-Borne and Zoonotic Diseases*. doi: 10.1089/vbz.2006.0562.

Benkert, P., Künzli, M. and Schwede, T. (2009) 'QMEAN server for protein model quality estimation', *Nucleic Acids Research*. doi: 10.1093/nar/gkp322.

Bernhardt, S. A., Simmons, M. P., Olson, K. E., Beaty, B. J., Blair, C. D. and Black, W. C. (2012) 'Rapid Intraspecific Evolution of miRNA and siRNA Genes in the Mosquito Aedes aegypti', *PLoS ONE*, 7(9). doi: 10.1371/journal.pone.0044198.

Blair, C. D., Olson, K. E. and Bonizzoni, M. (2019) 'The Widespread Occurrence and Potential Biological Roles of Endogenous Viral Elements in

Insect Genomes', *Current Issues in Molecular Biology*. doi: 10.21775/cimb.034.013.

Blanca, M. J., Alarcón, R., Arnau, J. and Bendayan, R. (2017) 'Non-normal data: Is ANOVA still a valid option?', *Psicothema*. doi: 10.7334/psicothema2016.383.

Bolling, B. G., Weaver, S. C., Tesh, R. B. and Vasilakis, N. (2015) 'Insect-specific virus discovery: Significance for the arbovirus community', *Viruses*. doi: 10.3390/v7092851.

Bonizzoni, M., Dunn, W. A., Campbell, C. L., Olson, K. E., Marinotti, O. and James, A. A. (2012) 'Complex Modulation of the Aedes aegypti Transcriptome in Response to Dengue Virus Infection', *PLoS ONE*. doi: 10.1371/journal.pone.0050512.

Bonizzoni, M., Gasperi, G., Chen, X. and James, A. A. (2013) 'The invasive mosquito species Aedes albopictus: Current knowledge and future perspectives', *Trends in Parasitology*. doi: 10.1016/j.pt.2013.07.003.

Bono, J. M., Matzkin, L. M., Castrezana, S. and Markow, T. A. (2008) 'Molecular evolution and population genetics of two Drosophila mettleri cytochrome P450 genes involved in host plant utilization', *Molecular Ecology*. doi: 10.1111/j.1365-294X.2008.03823.x.

Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R. and Hannon, G. J. (2007a) 'Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila', *Cell*. doi: 10.1016/j.cell.2007.01.043.

Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R. and Hannon, G. J. (2007b) 'Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila', *Cell*, 128(6), pp. 1089–1103. doi: 10.1016/j.cell.2007.01.043.

Bryant, W. B., Mills, M. K., Olson, B. J. S. C. and Michel, K. (2019) 'Small RNA-Seq analysis reveals miRNA expression dynamics across tissues in the malaria vector, Anopheles gambiae', *G3: Genes, Genomes, Genetics*. doi: 10.1534/g3.119.400104.

Buhagiar, J. A. (2009) 'A second record of Aedes (Stegomyia) albopictus (Diptera: Culicidae) in Malta', *European Mosquito Bulletin*, 27(November), pp. 65–67.

Calvez, E., Guillaumot, L., Girault, D., Richard, V., O'Connor, O., Paoaafaite, T., Teurlai, M., Pocquet, N., Cao-Lormeau, V. M. and Dupont-Rouzeyrol, M. (2017) 'Dengue-1 virus and vector competence of

Aedes aegypti (Diptera: Culicidae) populations from New Caledonia', *Parasites and Vectors*. doi: 10.1186/s13071-017-2319-x.

Caminade, C., Medlock, J. M., Ducheyne, E., McIntyre, K. M., Leach, S., Baylis, M. and Morse, A. P. (2012) 'Suitability of European climate for the Asian tiger mosquito Aedes albopictus: Recent trends and future scenarios', *Journal of the Royal Society Interface*, 9(75), pp. 2708–2717. doi: 10.1098/rsif.2012.0138.

Campbell, C. L., Black IV, W. C., Hess, A. M. and Foy, B. D. (2008a) 'Comparative genomics of small RNA regulatory pathway components in vector mosquitoes', *BMC Genomics*, 9. doi: 10.1186/1471-2164-9-425.

Campbell, C. L., Keene, K. M., Brackney, D. E., Olson, K. E., Blair, C. D., Wilusz, J. and Foy, B. D. (2008b) 'Aedes aegypti uses RNA interference in defense against Sindbis virus infection', *BMC Microbiology*, 8, pp. 1–12. doi: 10.1186/1471-2180-8-47.

Carrington, L. B., Armijos, M. V., Lambrechts, L. and Scott, T. W. (2013) 'Fluctuations at a Low Mean Temperature Accelerate Dengue Virus Transmission by Aedes aegypti', *PLoS Neglected Tropical Diseases*. doi: 10.1371/journal.pntd.0002190.

Chee, H. Y. and AbuBakar, S. (2004) 'Identification of a 48 kDa tubulin or tubulin-like C6/36 mosquito cells protein that binds dengue virus 2 using mass spectrometry', *Biochemical and Biophysical Research Communications*, 320(1), pp. 11–17. doi: 10.1016/j.bbrc.2004.05.124.

Chen, X.-G. *et al.* (2015) 'Genome sequence of the Asian Tiger mosquito, *Aedes albopictus* , reveals insights into its biology, genetics, and evolution', *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1516410112.

Chompoosri, J., Thavara, U., Tawatsin, A., Boonserm, R., Phumee, A., Sangkitporn, S. and Siriyasatien, P. (2016) 'Vertical transmission of Indian Ocean Lineage of chikungunya virus in Aedes aegypti and Aedes albopictus mosquitoes', *Parasites and Vectors*. doi: 10.1186/s13071-016-1505-6.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X. and Ruden, D. M. (2012) 'A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3', *Fly*. doi: 10.4161/fly.19695.

Collet, J. M., Fuentes, S., Hesketh, J., Hill, M. S., Innocenti, P., Morrow, E. H., Fowler, K. and Reuter, M. (2016) 'Rapid evolution of the intersexual

117

genetic correlation for fitness in Drosophila melanogaster', *Evolution*. doi: 10.1111/evo.12892.

Colpitts, T. M., Cox, J., Vanlandingham, D. L., Feitosa, F. M., Cheng, G., Kurscheid, S., Wang, P., Krishnan, M. N., Higgs, S. and Fikrig, E. (2011) 'Alterations in the aedes aegypti transcriptome during infection with west nile, dengue and yellow fever viruses', *PLoS Pathogens*. doi: 10.1371/journal.ppat.1002189.

Cox, J., Mota, J., Sukupolvi-Petty, S., Diamond, M. S. and Rico-Hesse, R. (2012) 'Mosquito Bite Delivery of Dengue Virus Enhances Immunogenicity and Pathogenesis in Humanized Mice', *Journal of Virology*. doi: 10.1128/jvi.00534-12.

Crans, W. J., Sprenger, D. A. and Mahmood, F. (1996) 'The blood-feeding habits of Aedes sollicitans (Walker) in relation to Eastern Equine Encephalitis virus in coastal areas of New Jersey. II. Results of experiments with caged mosquitoes and the effects of temperature and physiological age on host selection', *Journal of Vector Ecology*, 21 (1)(1), pp. 1–5.

Day, J. F., Ramsey, A. M. and Zhang, J.-T. (2015) 'Environmentally Mediated Seasonal Variation in Mosquito Body Size', *Environmental Entomology*. doi: 10.1093/ee/19.3.469.

Delatte, H., Gimonneau, G., Triboire, A. and Fontenille, D. (2009) ' Influence of Temperature on Immature Development, Survival, Longevity, Fecundity, and Gonotrophic Cycles of Aedes albopictus , Vector of Chikungunya and Dengue in the Indian Ocean ', *Journal of Medical Entomology*. doi: 10.1603/033.046.0105.

Denli, A. M., Tops, B. B. J., Plasterk, R. H. A., Ketting, R. F. and Hannon, G. J. (2004) 'Processing of primary microRNAs by the Microprocessor complex', *Nature*. doi: 10.1038/nature03049.

Dritsou, V. *et al.* (2015) ' A draft genome sequence of an invasive mosquito: an Italian Aedes albopictus ', *Pathogens and Global Health*. doi: 10.1179/2047773215y.0000000031.

Du, Y., Nomura, Y., Zhorov, B. S. and Dong, K. (2015) 'Rotational Symmetry of Two Pyrethroid Receptor Sites in the Mosquito Sodium Channel.', *Molecular pharmacology*. American Society for Pharmacology and Experimental Therapeutics, 88(2), pp. 273–80. doi: 10.1124/mol.115.098707.

Dubey, S. K., Shrinet, J., Jain, J., Ali, S. and Sunil, S. (2017) 'Aedes aegypti microRNA miR-2b regulates ubiquitin-related modifier to control

chikungunya virus replication', *Scientific Reports*. doi: 10.1038/s41598-017-18043-0.

Dubrulle, M., Mousson, L., Moutailier, S., Vazeille, M. and Failloux, A. B. (2009) 'Chikungunya virus and Aedes mosquitoes: Saliva is infectious as soon as two days after oral infection', *PLoS ONE*. doi: 10.1371/journal.pone.0005895.

Edgar, R. C. (2004) 'MUSCLE: Multiple sequence alignment with high accuracy and high throughput', *Nucleic Acids Research*. doi: 10.1093/nar/gkh340.

Elbashir, S. M., Lendeckel, W. and Tuschl, T. (2001) 'RNA interference is mediated by 21- and 22-nucleotide RNAs', *Genes and Development*. doi: 10.1101/gad.862301.

Emsley, P., Lohkamp, B., Scott, W. G. and Cowtan, K. (2010) ' Features and development of Coot ', *Acta Crystallographica Section D Biological Crystallography*. doi: 10.1107/s0907444910007493.

Estep, A. S., Sanscrainte, N. D., Waits, C. M., Bernard, S. J., Lloyd, A. M., Lucas, K. J., Buckner, E. A., Vaidyanathan, R., Morreale, R., Conti, L. A. and Becnel, J. J. (2018) 'Quantification of permethrin resistance and *kdr* alleles in Florida strains of *Aedes aegypti* (L.) and *Aedes albopictus* (Skuse)', *PLOS Neglected Tropical Diseases*. Edited by P. J. McCall, 12(10), p. e0006544. doi: 10.1371/journal.pntd.0006544.

Evanno, G., Regnaut, S. and Goudet, J. (2005) 'Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study', *Molecular Ecology*. John Wiley & Sons, Ltd (10.1111), 14(8), pp. 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x.

Feng, T., Sun, T., Li, G., Pan, W., Wang, K. and Dai, J. (2017) 'DEAD-box helicase DDX25 is a negative regulator of type I interferon pathway and facilitates RNA virus infection', *Frontiers in Cellular and Infection Microbiology*. doi: 10.3389/fcimb.2017.00356.

Flegel, T. W. (2007) 'Update on viral accommodation, a model for host-viral interaction in shrimp and other arthropods', *Developmental and Comparative Immunology*. doi: 10.1016/j.dci.2006.06.009.

Flegel, T. W. and Sritunyalucksana, K. (2011) 'Shrimp Molecular Responses to Viral Pathogens', *Marine Biotechnology*. doi: 10.1007/s10126-010-9287-x.

Flot, J. F. (2010) 'Seqphase: A web tool for interconverting phase input/output files and fasta sequence alignments', *Molecular Ecology*

Resources. doi: 10.1111/j.1755-0998.2009.02732.x.

Fonseca, N. A., Morales-Hojas, R., Reis, M., Rocha, H., Vieira, C. P., Nolte, V., Schlötterer, C. and Vieira, J. (2013) 'Drosophila americana as a model species for comparative studies on the molecular basis of phenotypic variation', *Genome Biology and Evolution*. doi: 10.1093/gbe/evt037.

Förstemann, K., Horwich, M. D., Wee, L., Tomari, Y. and Zamore, P. D. (2007) 'Drosophila microRNAs Are Sorted into Functionally Distinct Argonaute Complexes after Production by Dicer-1', *Cell*. doi: 10.1016/j.cell.2007.05.056.

Forster, M., Szymczak, S., Ellinghaus, D., Hemmrich, G., Rühlemann, M., Kraemer, L., Mucha, S., Wienbrandt, L., Stanulla, M. and Franke, A. (2015) 'Vy-PER: Eliminating false positive detection of virus integration events in next generation sequencing data', *Scientific Reports*. doi: 10.1038/srep11534.

Fort, P., Albertini, A., Van-Hua, A., Berthomieu, A., Roche, S., Delsuc, F., Pasteur, N., Capy, P., Gaudin, Y. and Weill, M. (2012) 'Fossil rhabdoviral sequences integrated into arthropod genomes: Ontogeny, evolution, and potential functionality', *Molecular Biology and Evolution*. doi: 10.1093/molbev/msr226.

Fragkoudis, R., Chi, Y., Siu, R. W. C., Barry, G., Attarzadeh-Yazdi, G., Merits, A., Nash, A. A., Fazakerley, J. K. and Kohl, A. (2008) 'Semliki Forest virus strongly reduces mosquito host defence signaling', *Insect Molecular Biology*. doi: 10.1111/j.1365-2583.2008.00834.x.

Fros, J. J., Liu, W. J., Prow, N. A., Geertsema, C., Ligtenberg, M., Vanlandingham, D. L., Schnettler, E., Vlak, J. M., Suhrbier, A., Khromykh, A. A. and Pijlman, G. P. (2010) 'Chikungunya Virus Nonstructural Protein 2 Inhibits Type I/II Interferon-Stimulated JAK-STAT Signaling', *Journal of Virology*, 84(20), pp. 10877–10887. doi: 10.1128/JVI.00949-10.

Fujino, K., Horie, M., Honda, T., Merriman, D. K. and Tomonaga, K. (2014) 'Inhibition of Borna disease virus replication by an endogenous bornavirus-like element in the ground squirrel genome', *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1407046111.

Gangaraju, V. K. and Lin, H. (2009) 'MicroRNAs: Key regulators of stem cells', *Nature Reviews Molecular Cell Biology*. doi: 10.1038/nrm2621.

Garrison, E. and Marth, G. (2012) 'Haplotype-based variant detection from short-read sequencing -- Free bayes -- Variant Calling -- Longranger', *arXiv*

preprint arXiv:1207.3907. doi: arXiv:1207.3907 [q-bio.GN].

Gatt, P., Deeming, J. C. and Schaffner, F. (2009) 'First record of *Aedes (Stegomyia) albopictus* (Skuse) (Diptera: Culicidae) in Malta', *European Mosquito Bulletin.*

Georgel, P., Naitza, S., Kappler, C., Ferrandon, D., Zachary, D., Swimmer, C., Kopczynski, C., Duyk, G., Reichhart, J. M. and Hoffmann, J. A. (2001) 'Drosophila Immune Deficiency (IMD) Is a Death Domain Protein that Activates Antibacterial Defense and Can Promote Apoptosis', *Developmental Cell.* doi: 10.1016/S1534-5807(01)00059-4.

Geuking, M. B., Weber, J., Dewannieux, M., Gorelik, E., Heidmann, T., Hengartner, H., Zinkernagel, R. M. and Hangartner, L. (2009) 'Recombination of retrotransposon and exogenous RNA virus results in nonretroviral cDNA integration', *Science.* doi: 10.1126/science.1167375.

Ghildiyal, M., Seitz, H., Horwich, M. D., Li, C., Du, T., Lee, S., Xu, J., Kittler, E. L. W., Zapp, M. L., Weng, Z. and Zamore, P. D. (2008) 'Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells', *Science.* doi: 10.1126/science.1157396.

Giangaspero, A., Marangi, M., Latrofa, M. S., Martinelli, D., Traversa, D., Otranto, D. and Genchi, C. (2013) 'Evidences of increasing risk of dirofilarioses in southern Italy', *Parasitology Research*, 112(3), pp. 1357–1361. doi: 10.1007/s00436-012-3206-1.

Girardi, E., Miesen, P., Pennings, B., Frangeul, L., Saleh, M. C. and Van Rij, R. P. (2017) 'Histone-derived piRNA biogenesis depends on the ping-pong partners Piwi5 and Ago3 in Aedes aegypti', *Nucleic Acids Research.* doi: 10.1093/nar/gkw1368.

Gjenero-Margan, I. *et al.* (2011) 'Autochthonous dengue fever in Croatia, August- September 2010', *Eurosurveillance*, 16(9), pp. 1–4. doi: 19805 [pii].

Goic, B., Stapleford, K. A., Frangeul, L., Doucet, A. J., Gausson, V., Blanc, H., Schemmel-Jofre, N., Cristofari, G., Lambrechts, L., Vignuzzi, M. and Saleh, M. C. (2016) 'Virus-derived DNA drives mosquito vector tolerance to arboviral infection', *Nature Communications.* doi: 10.1038/ncomms12410.

Gordon, A., Hannon, G. J. and Gordon (2014) 'FASTX-Toolkit', *[Online] http://hannonlab. cshl. edu/fastx_ toolkit http://hannonlab. cshl. edu/fastx_ toolkit.*

Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., Robles, M., Talón, M., Dopazo, J. and Conesa, A. (2008)

'High-throughput functional annotation and data mining with the Blast2GO suite', *Nucleic Acids Research*. doi: 10.1093/nar/gkn176.

Gould, E. A. and Higgs, S. (2009) 'Impact of climate change and other factors on emerging arbovirus diseases', *Transactions of the Royal Society of Tropical Medicine and Hygiene*. doi: 10.1016/j.trstmh.2008.07.025.

Gratz, N. G. (2004) 'Critical review of the vector status of Aedes albopictus', *Medical and Veterinary Entomology*, 18(3), pp. 215–227. doi: 10.1111/j.0269-283X.2004.00513.x.

Guida, V., Cernilogar, F. M., Filograna, A., De Gregorio, R., Ishizu, H., Siomi, M. C., Schotta, G., Bellenchi, G. C. and Andrenacci, D. (2016) 'Production of small noncoding RNAs from the flamenco locus is regulated by the gypsy retrotransposon of Drosophila melanogaster', *Genetics*. doi: 10.1534/genetics.116.187922.

Haddi, K., Tomé, H. V. V., Du, Y., Valbon, W. R., Nomura, Y., Martins, G. F., Dong, K. and Oliveira, E. E. (2017) 'Detection of a new pyrethroid resistance mutation (V410L) in the sodium channel of *Aedes aegypti*: A potential challenge for mosquito control', *Scientific Reports*. doi: 10.1038/srep46549.

Hahn, M. W. (2009) 'Distinguishing among evolutionary models for the maintenance of gene duplicates', *Journal of Heredity*. doi: 10.1093/jhered/esp047.

Hawley, W. A. (1988) 'The biology of Aedes albopictus.', *Journal of the American Mosquito Control Association. Supplement*.

Hess, A. M., Prasad, A. N., Ptitsyn, A., Ebel, G. D., Olson, K. E., Barbacioru, C., Monighetti, C. and Campbell, C. L. (2011) 'Small RNA profiling of Dengue virus-mosquito interactions implicates the PIWI RNA pathway in anti-viral defense', *BMC Microbiology*. doi: 10.1186/1471-2180-11-45.

Hollidge, B. S., González-Scarano, F. and Soldan, S. S. (2010) 'Arboviral encephalitides: Transmission, emergence, and pathogenesis', *Journal of Neuroimmune Pharmacology*. doi: 10.1007/s11481-010-9234-7.

Horwich, M. D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P. and Zamore, P. D. (2007) 'The Drosophila RNA Methyltransferase, DmHen1, Modifies Germline piRNAs and Single-Stranded siRNAs in RISC', *Current Biology*. doi: 10.1016/j.cub.2007.06.030.

Hu, W., Criscione, F., Liang, S. and Tu, Z. (2015) 'MicroRNAs of two medically important mosquito species: Aedes aegypti and Anopheles

stephensi', *Insect Molecular Biology*. doi: 10.1111/imb.12152.

Huang, Z., Kingsolver, M. B., Avadhanula, V. and Hardy, R. W. (2013) 'An Antiviral Role for Antimicrobial Peptides during the Arthropod Response to Alphavirus Replication', *Journal of Virology*. doi: 10.1128/jvi.03360-12.

Hurlbut, H. S. (1973) 'The effect of environmental temperature upon the transmission of St. Louis encephalitis virus by Culex pipiens quinquefasciatus', *Journal of Medical Entomology*. doi: 10.1093/jmedent/10.1.1.

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

Joshua-Tor, L. (2006) 'The argonautes', in *Cold Spring Harbor Symposia on Quantitative Biology*. doi: 10.1101/sqb.2006.71.048.

Kamgang, B., Marcombe, S., Chandre, F., Nchoutpouen, E., Nwane, P., Etang, J., Corbel, V. and Paupy, C. (2011) 'Insecticide susceptibility of *Aedes aegypti* and *Aedes albopictus* in Central Africa.', *Parasites and Vectors*, 4, p. 79. doi: 10.1186/1756-3305-4-79.

Karunaratne, S. H. P. P., Weeraratne, T. C., Perera, M. D. B. and Surendran, S. N. (2013) 'Insecticide resistance and, efficacy of space spraying and larviciding in the control of dengue vectors *Aedes aegypti* and *Aedes albopictus* in Sri Lanka', *Pesticide Biochemistry and Physiology*, 107(1), pp. 98–105. doi: 10.1016/j.pestbp.2013.05.011.

Kawaoka, S., Izumi, N., Katsuma, S. and Tomari, Y. (2011) '3' End Formation of PIWI-Interacting RNAs In Vitro', *Molecular Cell*, 43(6), pp. 1015–1022. doi: 10.1016/j.molcel.2011.07.029.

Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U. and Segal, E. (2007) 'The role of site accessibility in microRNA target recognition', *Nature Genetics*. doi: 10.1038/ng2135.

Khan, A. and Rayner, G. D. (2004) 'Robustness to non-normality of common tests for the many-sample location problem', *Journal of Applied Mathematics and Decision Sciences*. doi: 10.1155/S1173912603000178.

Kim, D., Langmead, B. and Salzberg, S. L. (2015) 'HISAT: a fast spliced aligner with low memory requirements', *Nature Methods*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 12, p. 357.

Kingsolver, M. B., Huang, Z. and Hardy, R. W. (2013) 'Insect antiviral

innate immunity: Pathways, effectors, and connections', *Journal of Molecular Biology*. doi: 10.1016/j.jmb.2013.10.006.

Klepsatel, P., Gáliková, M., De Maio, N., Ricci, S., Schlötterer, C. and Flatt, T. (2013) 'Reproductive and post-reproductive life history of wild-caught Drosophila melanogaster under laboratory conditions', *Journal of Evolutionary Biology*. doi: 10.1111/jeb.12155.

Koonin, E. V. (2017) 'Evolution of RNA- and DNA-guided antivirus defense systems in prokaryotes and eukaryotes: Common ancestry vs convergence', *Biology Direct*. doi: 10.1186/s13062-017-0177-2.

Kotsakiozi, P., Richardson, J. B., Pichler, V., Favia, G., Martins, A. J., Urbanelli, S., Armbruster, P. A. and Caccone, A. (2017) 'Population genomics of the Asian tiger mosquito, *Aedes albopictus* : insights into the recent worldwide invasion', *Ecology and Evolution*. John Wiley & Sons, Ltd, 7(23), pp. 10143–10157. doi: 10.1002/ece3.3514.

Kraemer, M. U. G. *et al.* (2015) 'The global distribution of the arbovirus vectors Aedes aegypti and Ae. Albopictus', *eLife*. doi: 10.7554/eLife.08347.

Krüger, J. and Rehmsmeier, M. (2006) 'RNAhybrid: MicroRNA target prediction easy, fast and flexible', *Nucleic Acids Research*. doi: 10.1093/nar/gkl243.

Lai, Z., Markovets, A., Ahdesmaki, M., Chapman, B., Hofmann, O., Mcewen, R., Johnson, J., Dougherty, B., Barrett, J. C. and Dry, J. R. (2016) 'VarDict: A novel and versatile variant caller for next-generation sequencing in cancer research', *Nucleic Acids Research*. doi: 10.1093/nar/gkw227.

Lambrechts, L., Quillery, E., Noël, V., Richardson, J. H., Jarman, R. G., Scott, T. W. and Chevillon, C. (2013) 'Specificity of resistance to dengue virus isolates is associated with genotypes of the mosquito antiviral gene Dicer-2', *Proceedings of the Royal Society B: Biological Sciences*, 280(1751). doi: 10.1098/rspb.2012.2437.

Lee, W. S., Webster, J. A., Madzokere, E. T., Stephenson, E. B. and Herrero, L. J. (2019) 'Mosquito antiviral defense mechanisms: A delicate balance between innate immunity and persistent viral infection', *Parasites and Vectors*. doi: 10.1186/s13071-019-3433-8.

Legros, M., Otero, M., Aznar, V. R., Solari, H., Gould, F. and Lloyd, A. L. (2016) 'Comparison of two detailed models of Aedes aegypti population dynamics', *Ecosphere*. doi: 10.1002/ecs2.1515.

Di Lelio, I., Illiano, A., Astarita, F., Gianfranceschi, L., Horner, D.,

Varricchio, P., Amoresano, A., Pucci, P., Pennacchio, F. and Caccia, S. (2019) 'Evolution of an insect immune barrier through horizontal gene transfer mediated by a parasitic wasp', *PLoS Genetics*. doi: 10.1371/journal.pgen.1007998.

Lequime, S. and Lambrechts, L. (2014) 'Vertical transmission of arboviruses in mosquitoes: A historical perspective', *Infection, Genetics and Evolution*. doi: 10.1016/j.meegid.2014.07.025.

Lequime, S. and Lambrechts, L. (2017) ' Discovery of flavivirus-derived endogenous viral elements in Anopheles mosquito genomes supports the existence of Anopheles -associated insect-specific flaviviruses ', *Virus Evolution*. doi: 10.1093/ve/vew035.

Li, H. (2013) 'Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM'. Available at: https://arxiv.org/abs/1303.3997 (Accessed: 13 March 2019).

Li, H. and Durbin, R. (2010) 'Fast and accurate long-read alignment with Burrows-Wheeler transform', *Bioinformatics*. doi: 10.1093/bioinformatics/btp698.

Liao, Y., Smyth, G. K. and Shi, W. (2014) 'FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features', *Bioinformatics*. doi: 10.1093/bioinformatics/btt656.

Librado, P. and Rozas, J. (2009) 'DnaSP v5: A software for comprehensive analysis of DNA polymorphism data', *Bioinformatics*. doi: 10.1093/bioinformatics/btp187.

Liu-Helmersson, J., Stenlund, H., Wilder-Smith, A. and Rocklöv, J. (2014) 'Vectorial capacity of Aedes aegypti: Effects of temperature and implications for global dengue epidemic potential', *PLoS ONE*. doi: 10.1371/journal.pone.0089783.

Liu, P., Dong, Y., Gu, J., Puthiyakunnon, S., Wu, Y. and Chen, X. G. (2016) 'Developmental piRNA profiles of the invasive vector mosquito Aedes albopictus', *Parasites and Vectors*. doi: 10.1186/s13071-016-1815-8.

Liu, Y., Zhou, Y., Wu, J., Zheng, P., Li, Y., Zheng, X., Puthiyakunnon, S., Tu, Z. and Chen, X. G. (2015) 'The expression profile of Aedes albopictus miRNAs is altered by dengue virus serotype-2 infection', *Cell and Bioscience*. doi: 10.1186/s13578-015-0009-y.

Liu, Z., Zhang, Z., Lai, Z., Zhou, T., Jia, Z., Gu, J., Wu, K. and Chen, X. G. (2017) 'Temperature increase enhances Aedes albopictus competence to transmit dengue virus', *Frontiers in Microbiology*. doi:

10.3389/fmicb.2017.02337.

Luplertlop, N., Surasombatpattana, P., Patramool, D., Dumas, E., Wasinpiyamongkol, L., Saune, L., Hamel, R., Bernard, E., Sereno, D., Thomas, F. R., Piquemal, D., Yssel, H., Briant, L. and Missé, D. (2011) 'Induction of a peptide with activity against a broad spectrum of pathogens in the Aedes aegypti salivary gland, following infection with Dengue Virus', *PLoS Pathogens*. doi: 10.1371/journal.ppat.1001252.

MACDONALD, G. (1956) 'Theory of the eradication of malaria.', *Bulletin of the World Health Organization*.

Maharaj, P. D., Widen, S. G., Huang, J., Wood, T. G. and Thangamani, S. (2015) 'Discovery of Mosquito Saliva MicroRNAs during CHIKV Infection', *PLoS Neglected Tropical Diseases*. doi: 10.1371/journal.pntd.0003386.

Manni, M., Guglielmino, C. R., Scolari, F., Vega-Rúa, A., Failloux, A.-B., Somboon, P., Lisa, A., Savini, G., Bonizzoni, M., Gomulski, L. M., Malacrida, A. R. and Gasperi, G. (2017) 'Genetic evidence for a worldwide chaotic dispersion pattern of the arbovirus vector, *Aedes albopictus*', *PLOS Neglected Tropical Diseases*. Edited by C. Apperson. Public Library of Science, 11(1), p. e0005332. doi: 10.1371/journal.pntd.0005332.

Maori, E., Tanne, E. and Sela, I. (2007) 'Reciprocal sequence exchange between non-retro viruses and hosts leading to the appearance of new host phenotypes', *Virology*. doi: 10.1016/j.virol.2006.11.038.

Marklewitz, M., Zirkel, F., Kurth, A., Drosten, C. and Junglena, S. (2015) 'Evolutionary and phenotypic analysis of live virus isolates suggests arthropod origin of a pathogenic RNA virus family', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.1502036112.

Matthews, B. J. *et al.* (2018) 'Improved reference genome of Aedes aegypti informs arbovirus vector control', *Nature*. doi: 10.1038/s41586-018-0692-z.

McDonald, J. H. and Kreitman, M. (1991) 'Adaptive protein evolution at the Adh locus in Drosophila', *Nature*. doi: 10.1038/351652a0.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. and DePristo, M. A. (2010) 'The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.', *Genome research*. doi: 10.1101/gr.107524.110.

Medlock, J. M., Avenell, D., Barrass, I. and Leach, S. (2006) 'Analysis of

the potential for survival and seasonal activity of Aedes albopictus (Diptera: Culicidae) in the United Kingdom.', *Journal of vector ecology : journal of the Society for Vector Ecology*.

Miesen, P., Girardi, E. and Van Rij, R. P. (2015) 'Distinct sets of PIWI proteins produce arbovirus and transposon-derived piRNAs in Aedes aegypti mosquito cells', *Nucleic Acids Research*, 43(13), pp. 6545–6556. doi: 10.1093/nar/gkv590.

Miesen, P., Ivens, A., Buck, A. H. and van Rij, R. P. (2016a) 'Small RNA Profiling in Dengue Virus 2-Infected Aedes Mosquito Cells Reveals Viral piRNAs and Novel Host miRNAs', *PLoS Neglected Tropical Diseases*. doi: 10.1371/journal.pntd.0004452.

Miesen, P., Joosten, J. and van Rij, R. P. (2016b) 'PIWIs Go Viral: Arbovirus-Derived piRNAs in Vector Mosquitoes', *PLoS Pathogens*, 12(12), pp. 1–17. doi: 10.1371/journal.ppat.1006017.

Monsanto-Hearne, V. and Johnson, K. N. (2018) 'MiRNAs in insects infected by animal and plant viruses', *Viruses*. doi: 10.3390/v10070354.

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

Moser, L. A., Lim, P.-Y., Styer, L. M., Kramer, L. D. and Bernard, K. A. (2016) 'Parameters of Mosquito-Enhanced West Nile Virus Infection', *Journal of Virology*. doi: 10.1128/jvi.02280-15.

Moureau, G., Cook, S., Lemey, P., Nougairede, A., Forrester, N. L., Khasnatinov, M., Charrel, R. N., Firth, A. E., Gould, E. A. and De Lamballerie, X. (2015) 'New insights into flavivirus evolution, taxonomy and biogeographic history, extended by analysis of canonical and alternative coding sequences e0117849', *PLoS ONE*. doi: 10.1371/journal.pone.0117849.

Moyes, C. L., Vontas, J., Martins, A. J., Ng, L. C., Koou, S. Y., Dusfour, I., Raghavendra, K., Pinto, J., Corbel, V., David, J. P. and Weetman, D. (2017) 'Contemporary status of insecticide resistance in the major *Aedes* vectors of arboviruses infecting humans', *PLoS Neglected Tropical Diseases*, 11(7), pp. 1–20. doi: 10.1371/journal.pntd.0005625.

Nag, D. K., Brecher, M. and Kramer, L. D. (2016) 'DNA forms of arboviral RNA genomes are generated following infection in mosquito cell cultures', *Virology*. doi: 10.1016/j.virol.2016.08.022.

127

Ngoagouni, C., Kamgang, B., Brengues, C., Yahouedo, G., Paupy, C., Nakouné, E., Kazanji, M. and Chandre, F. (2016) 'Susceptibility profile and metabolic mechanisms involved in *Aedes aegypti* and *Aedes albopictus* resistant to DDT and deltamethrin in the Central African Republic', *Parasites and Vectors*, 9(1), pp. 1–13. doi: 10.1186/s13071-016-1887-5.

Obbard, D. J., Jiggins, F. M., Halligan, D. L. and Little, T. J. (2006) 'Natural selection drives extremely rapid evolution in antiviral RNAi genes', *Current Biology*, 16(6), pp. 580–585. doi: 10.1016/j.cub.2006.01.065.

Oliveros, J. C. (2016) 'Venny. An interactive tool for comparing lists with Venn's diagrams. https://bioinfogp.cnb.csic.es/tools/venny/index.html'.

Olivieri, D., Sykora, M. M., Sachidanandam, R., Mechtler, K. and Brennecke, J. (2010) 'An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in Drosophila', *EMBO Journal*. Nature Publishing Group, 29(19), pp. 3301–3317. doi: 10.1038/emboj.2010.212.

Olson, K. E. and Bonizzoni, M. (2017) 'Nonretroviral integrated RNA viruses in arthropod vectors: an occasional event or something more?', *Current Opinion in Insect Science*. Elsevier Inc, 22(1), pp. 45–53. doi: 10.1016/j.cois.2017.05.010.

Palatini, U., Miesen, P., Carballar-Lejarazu, R., Ometto, L., Rizzo, E., Tu, Z., van Rij, R. P. and 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*. doi: 10.1186/s12864-017-3903-3.

Palmer, W. H., Varghese, F. S. and Van Rij, R. P. (2018) 'Natural variation in resistance to virus infection in dipteran insects', *Viruses*. doi: 10.3390/v10030118.

Paradkar, P. N., Trinidad, L., Voysey, R., Duchemin, J. B. and Walker, P. J. (2012) 'Secreted Vago restricts West Nile virus infection in Culex mosquito cells by activating the Jak-STAT pathway', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.1205231109.

Park, S. Y., Choi, E. and Jeong, Y. S. (2013) 'Integrative effect of defective interfering RNA accumulation and helper virus attenuation is responsible for the persistent infection of Japanese encephalitis virus in BHK-21 cells', *Journal of Medical Virology*. doi: 10.1002/jmv.23665.

Parker, B. M. (1986) 'Hatchability of Eggs of Aedes taeniorhynchus

(Diptera: Culicidae): Effects of Different Temperatures and Photoperiods during Embryogenesis', *Annals of the Entomological Society of America*, 79(6), pp. 925–930. doi: 10.1093/aesa/79.6.925.

Patel, K. J., Rueda, L. M., Axtell, R. C. and Stinner, R. E. (1990) 'Temperature-Dependent Development and Survival Rates of Culex quinquefasciatus and Aedes aegypti (Diptera: Culicidae)', *Journal of Medical Entomology*, 27(5), pp. 892–898. doi: 10.1093/jmedent/27.5.892.

Paupy, C., Delatte, H., Bagny, L., Corbel, V. and Fontenille, D. (2009) 'Aedes albopictus, an arbovirus vector: From the darkness to the light', *Microbes and Infection*. Elsevier Masson SAS, 11(14–15), pp. 1177–1185. doi: 10.1016/j.micinf.2009.05.005.

Pech-May, A., Moo-Llanes, D. A., Puerto-Avila, M. B., Casas, M., Danis-Lozano, R., Ponce, G., Tun-Ku, E., Pinto-Castillo, J. F., Villegas, A., Ibáñez-Piñon, C. R., González, C. and Ramsey, J. M. (2016) 'Population genetics and ecological niche of invasive Aedes albopictus in Mexico', *Acta Tropica*. Elsevier, 157, pp. 30–41. doi: 10.1016/J.ACTATROPICA.2016.01.021.

Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. and Salzberg, S. L. (2016) 'Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown', *Nature Protocols*. doi: 10.1038/nprot.2016.095.

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T. and Salzberg, S. L. (2015) 'StringTie enables improved reconstruction of a transcriptome from RNA-seq reads', *Nature Biotechnology*. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved., 33, p. 290.

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

Pichler, V. *et al.* (2018) 'First evidence of resistance to pyrethroid insecticides in Italian Aedes albopictus populations 26 years after invasion', *Pest Management Science*. doi: 10.1002/ps.4840.

Pingen, M., Bryden, S. R., Pondeville, E., Schnettler, E., Kohl, A., Merits, A., Fazakerley, J. K., Graham, G. J. and McKimmie, C. S. (2016) 'Host Inflammatory Response to Mosquito Bites Enhances the Severity of Arbovirus Infection', *Immunity*. doi: 10.1016/j.immuni.2016.06.002.

Pischedda, E., Scolari, F., Valerio, F., Carballar-Lejarazú, R., Catapano,

P. L., Waterhouse, R. M. and Bonizzoni, M. (2019) 'Insights Into an Unexplored Component of the Mosquito Repeatome: Distribution and Variability of Viral Sequences Integrated Into the Genome of the Arboviral Vector Aedes albopictus', *Frontiers in Genetics*. doi: 10.3389/fgene.2019.00093.

Poirier, E. Z. *et al.* (2018) 'Dicer-2-Dependent Generation of Viral DNA from Defective Genomes of RNA Viruses Modulates Antiviral Immunity in Insects', *Cell Host and Microbe*. doi: 10.1016/j.chom.2018.02.001.

Pool, J. E. and Aquadro, C. F. (2007) 'The genetic basis of adaptive pigmentation variation in Drosophila melanogaster', *Molecular Ecology*. doi: 10.1111/j.1365-294X.2007.03324.x.

Qiu, P., Pan, P. C. and Govind, S. (1998) 'A role for the Drosophila Toll/Cactus pathway in larval hematopoiesis', *Development*.

Quinlan, A. R. and Hall, I. M. (2010) 'BEDTools: A flexible suite of utilities for comparing genomic features', *Bioinformatics*. doi: 10.1093/bioinformatics/btq033.

Reiner, R. C. *et al.* (2013) 'A systematic review of mathematical models of mosquito-borne pathogen transmission: 1970-2010.', *Journal of the Royal Society, Interface*. doi: 10.1098/rsif.2012.0921.

Reynolds, J. A., Poelchau, M. F., Rahman, Z., Armbruster, P. A. and Denlinger, D. L. (2013) 'NIH Public Access', 58(7), pp. 966–973. doi: 10.1016/j.jinsphys.2012.04.013.Transcript.

Rezza, G. *et al.* (2007) 'Infection with chikungunya virus in Italy: an outbreak in a temperate region', *Lancet*, 370(9602), pp. 1840–1846. doi: 10.1016/S0140-6736(07)61779-6.

Rimmer, A., Phan, H., Mathieson, I., Iqbal, Z., Twigg, S. R. F., Wilkie, A. O. M., Mcvean, G. and Lunter, G. (2014) 'Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications', *Nature Genetics*. doi: 10.1038/ng.3036.

Rivals, I., Personnaz, L., Taing, L. and Potier, M. C. (2007) 'Enrichment or depletion of a GO category within a class of genes: Which test?', *Bioinformatics*. doi: 10.1093/bioinformatics/btl633.

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G. and Mesirov, J. P. (2011) 'Integrative genomics viewer', *Nature Biotechnology*. doi: 10.1038/nbt.1754.

Robinson, M. D., McCarthy, D. J. and Smyth, G. K. (2009) 'edgeR: A Bioconductor package for differential expression analysis of digital gene

expression data', *Bioinformatics*. doi: 10.1093/bioinformatics/btp616.

Rosenkranz, D., Han, C. T., Roovers, E. F., Zischler, H. and Ketting, R. F. (2015) 'Piwi proteins and piRNAs in mammalian oocytes and early embryos: From sample to sequence', *Genomics Data*. doi: 10.1016/j.gdata.2015.06.026.

Ross, R. (1911) *The prevention of Malaria*. New York, NY: Dutton.

Rstudio Team (2016) 'RStudio: Integrated Development for R', *[Online] RStudio, Inc., Boston, MA*. doi: 10.1007/978-81-322-2340-5.

La Ruche, G. *et al.* (2010) 'First two autochthonous dengue virus infections in metropolitan France, september 2010', *Eurosurveillance*. doi: 19676 [pii].

Saavedra-Rodriguez, K., Maloof, F. V., Campbell, C. L., Garcia-Rejon, J., Lenhart, A., Penilla, P., Rodriguez, A., Sandoval, A. A., Flores, A. E., Ponce, G., Lozano, S. and Black, W. C. (2018) 'Parallel evolution of *vgsc* mutations at domains IS6, IIS6 and IIIS6 in pyrethroid resistant *Aedes aegypti* from Mexico', *Scientific Reports*. doi: 10.1038/s41598-018-25222-0.

Saldaña, M. A., Etebari, K., Hart, C. E., Widen, S. G., Wood, T. G., Thangamani, S., Asgari, S. and Hughes, G. L. (2017) 'Zika virus alters the microRNA expression profile and elicits an RNAi response in Aedes aegypti mosquitoes', *PLoS Neglected Tropical Diseases*. doi: 10.1371/journal.pntd.0005760.

Sánchez-Vargas, I., Scott, J. C., Poole-Smith, B. K., Franz, A. W. E., Barbosa-Solomieu, V., Wilusz, J., Olson, K. E. and Blair, C. D. (2009) 'Dengue virus type 2 infections of Aedes aegypti are modulated by the mosquito's RNA interference pathway', *PLoS Pathogens*, 5(2). doi: 10.1371/journal.ppat.1000299.

Sanders, H. R., Foy, B. D., Evans, A. M., Ross, L. S., Beaty, B. J., Olson, K. E. and Gill, S. S. (2005) 'Sindbis virus induces transport processes and alters expression of innate immunity pathway genes in the midgut of the disease vector, Aedes aegypti', *Insect Biochemistry and Molecular Biology*. doi: 10.1016/j.ibmb.2005.07.006.

Sansone, C. L., Cohen, J., Yasunaga, A., Xu, J., Osborn, G., Subramanian, H., Gold, B., Buchon, N. and Cherry, S. (2015) 'Microbiota-dependent priming of antiviral intestinal immunity in Drosophila', *Cell Host and Microbe*. doi: 10.1016/j.chom.2015.10.010.

Schirle, N. T., Sheu-Gruttadauria, J., Chandradoss, S. D., Joo, C. and MacRae, I. J. (2015) 'Water-mediated recognition of t1-adenosine anchors Argonaute2 to microRNA targets', *eLife*. doi: 10.7554/elife.07646.

Schnettler, E., Donald, C. L., Human, S., Watson, M., Siu, R. W. C., McFarlane, M., Fazakerley, J. K., Kohl, A. and Fragkoudis, R. (2013) 'Knockdown of piRNA pathway proteins results in enhanced semliki forest virus production in mosquito cells', *Journal of General Virology*, 94(PART7), pp. 1680–1689. doi: 10.1099/vir.0.053850-0.

Schrödinger, L. (2015) 'The PyMOL molecular graphics system, version 1.8', *https://www.pymol.org/citing*.

Schroeder, M. P. and Lopez-Bigas, N. (2015) 'muts-needle-plot: Mutations Needle Plot v0.8.0'. doi: 10.5281/ZENODO.14561.

Schuffenecker, I. *et al.* (2006) 'Genome microevolution of chikungunya viruses causing the Indian Ocean outbreak', *PLoS Medicine*. doi: 10.1371/journal.pmed.0030263.

Shapiro, L. L. M., Whitehead, S. A. and Thomas, M. B. (2017) 'Quantifying the effects of temperature on mosquito and parasite traits that determine the transmission potential of human malaria', *PLoS Biology*. doi: 10.1371/journal.pbio.2003489.

Shrinet, J., Srivastava, P. and Sunil, S. (2017) 'Transcriptome analysis of Aedes aegypti in response to mono-infections and co-infections of dengue virus-2 and chikungunya virus', *Biochemical and Biophysical Research Communications*. doi: 10.1016/j.bbrc.2017.01.162.

Sigle, L. T. and McGraw, E. A. (2019) 'Expanding the canon: Non-classical mosquito genes at the interface of arboviral infection', *Insect Biochemistry and Molecular Biology*. doi: 10.1016/j.ibmb.2019.04.004.

Sim, S., Jupatanakul, N. and Dimopoulos, G. (2014) 'Mosquito immunity against arboviruses', *Viruses*. doi: 10.3390/v6114479.

Sinha, N. K., Iwasa, J., Shen, P. S. and Bass, B. L. (2018) 'Dicer uses distinct modules for recognizing dsRNA termini', *Science*. doi: 10.1126/science.aaq0921.

Smith, D. L., Battle, K. E., Hay, S. I., Barker, C. M., Scott, T. W. and McKenzie, F. E. (2012) 'Ross, Macdonald, and a theory for the dynamics and control of mosquito-transmitted pathogens', *PLoS Pathogens*. doi: 10.1371/journal.ppat.1002588.

Smith, L. B., Kasai, S. and Scott, J. G. (2016) 'Pyrethroid resistance in *Aedes aegypti* and *Aedes albopictus*: Important mosquito vectors of human diseases', *Pesticide Biochemistry and Physiology*. Elsevier B.V., 133, pp. 1–12. doi: 10.1016/j.pestbp.2016.03.005.

Souza-Neto, J. A., Powell, J. R. and Bonizzoni, M. (2019) 'Aedes aegypti

vector competence studies: A review', *Infection, Genetics and Evolution*. doi: 10.1016/j.meegid.2018.11.009.

Souza-Neto, J. A., Sim, S. and Dimopoulos, G. (2009) 'An evolutionary conserved function of the JAK-STAT pathway in anti-dengue defense', *Proceedings of the National Academy of Sciences*, 106(42), pp. 17841–17846. doi: 10.1073/pnas.0905006106.

Starega-Roslan, J., Galka-Marciniak, P. and Krzyzosiak, W. J. (2015) 'Nucleotide sequence of miRNA precursor contributes to cleavage site selection by Dicer', *Nucleic Acids Research*. doi: 10.1093/nar/gkv968.

Stein, C. B., Genzor, P., Mitra, S., Elchert, A. R., Ipsaro, J. J., Benner, L., Sobti, S., Su, Y., Hammell, M., Joshua-Tor, L. and Haase, A. D. (2019) 'Decoding the 5′ nucleotide bias of PIWI-interacting RNAs', *Nature Communications*. Springer US, 10(1), p. 828. doi: 10.1038/s41467-019-08803-z.

Stephens, M. and Scheet, P. (2005) 'Accounting for Decay of Linkage Disequilibrium in Haplotype Inference and Missing-Data Imputation', *The American Journal of Human Genetics*. doi: 10.1086/428594.

Stephens, M., Smith, N. J. and Donnelly, P. (2001) 'A new statistical method for haplotype reconstruction from population data.', *American journal of human genetics*. doi: 10.1086/319501.

Stern, D. L. and Orgogozo, V. (2008) 'The loci of evolution: How predictable is genetic evolution?', *Evolution*. doi: 10.1111/j.1558-5646.2008.00450.x.

Tajima, F. (1983) 'Evolutionary relationship of DNA sequences in finite populations.', *Genetics*.

Tajima, F. (1989) 'Statistical method for testing the neutral mutation hypothesis by DNA polymorphism', *Genetics*.

Terradas, G. and McGraw, E. A. (2019) 'Using genetic variation in Aedes aegypti to identify candidate anti-dengue virus genes', *BMC Infectious Diseases*. BMC Infectious Diseases, 19(1), pp. 1–14. doi: 10.1186/s12879-019-4212-z.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) 'CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice', *Nucleic Acids Research*. doi: 10.1093/nar/22.22.4673.

Tomari, Y., Matranga, C., Haley, B., Martinez, N. and Zamore, P. D. (2004) 'A protein sensor for siRNA asymmetry', *Science*. doi:

10.1126/science.1102755.

Tomasello, D. and Schlagenhauf, P. (2013) 'Chikungunya and dengue autochthonous cases in Europe, 2007-2012', *Travel Medicine and Infectious Disease*. doi: 10.1016/j.tmaid.2013.07.006.

Uhrig, S. and Klein, H. (2019) 'PingPongPro: A tool for the detection of piRNA-mediated transposon-silencing in small RNA-Seq data', *Bioinformatics*. doi: 10.1093/bioinformatics/bty578.

Varjak, M., Kean, J., Vazeille, M., Failloux, A. and Kohl, A. (2017) 'Aedes aegypti Piwi4 Is a Noncanonical', *mSphere*, 2(3), pp. e00144-17.

Varjak, M., Leggewie, M. and Schnettler, E. (2018) 'The antiviral piRNA response in mosquitoes?', pp. 1–12. doi: 10.1099/jgv.0.001157.

Vazeille, M., Moutailler, S., Coudrier, D., Rousseaux, C., Khun, H., Huerre, M., Thiria, J., Dehecq, J. S., Fontenille, D., Schuffenecker, I., Despres, P. and Failloux, A. B. (2007) 'Two Chikungunya isolates from the outbreak of La Reunion (Indian Ocean) exhibit different patterns of infection in the mosquito, Aedes albopictus', *PLoS ONE*. doi: 10.1371/journal.pone.0001168.

Venturi, G. *et al.* (2017) 'Detection of a chikungunya outbreak in Central Italy Detection of a chikungunya outbreak in Central', *Euro Surveill*, 22(39), pp. 1–4. doi: 10.2807/1560.

Vodovar, N., Bronkhorst, A. W., van Cleef, K. W. R., Miesen, P., Blanc, H., van Rij, R. P. and Saleh, M. C. (2012) 'Arbovirus-derived piRNAs exhibit a ping-pong signature in mosquito cells', *PLoS ONE*, 7(1). doi: 10.1371/journal.pone.0030861.

Vontas, J., Kioulos, E., Pavlidi, N., Morou, E., della Torre, A. and Ranson, H. (2012) 'Insecticide resistance in the major dengue vectors *Aedes albopictus* and *Aedes aegypti*', *Pesticide Biochemistry and Physiology*, 104(2), pp. 126–131. doi: 10.1016/j.pestbp.2012.05.008.

Wang, W., Yoshikawa, M., Han, B. W., Izumi, N., Tomari, Y., Weng, Z. and Zamore, P. D. (2014) 'The initial uridine of primary piRNAs does not create the tenth adenine that is the hallmark of secondary piRNAs', *Molecular Cell*. doi: 10.1016/j.molcel.2014.10.016.

Wang, X. H., Aliyari, R., Li, W. X., Li, H. W., Kim, K., Carthew, R., Atkinson, P. and Ding, S. W. (2006) 'RNA interference directs innate immunity against viruses in adult Drosophila', *Science*. doi: 10.1126/science.1125694.

Wang, Y., Jin, B., Liu, P., Li, J., Chen, X. and Gu, J. (2018) 'PiRNA

profiling of dengue virus type 2-infected Asian tiger mosquito and midgut tissues', *Viruses*, 10(4), pp. 1–20. doi: 10.3390/v10040213.

Watterson, G. A. (1975) 'On the number of segregating sites in genetical models without recombination', *Theoretical Population Biology*. doi: 10.1016/0040-5809(75)90020-9.

Weaver, S. C. and Reisen, W. K. (2010) 'Present and future arboviral threats', *Antiviral Research*. doi: 10.1016/j.antiviral.2009.10.008.

Webb, B. and Sali, A. (2017) 'Protein structure modeling with MODELLER', in *Methods in Molecular Biology*. doi: 10.1007/978-1-4939-7231-9_4.

Weick, E. M. and Miska, E. A. (2014) 'piRNAs: From biogenesis to function', *Development (Cambridge)*. doi: 10.1242/dev.094037.

Whitfield, Z. J., Dolan, P. T., Kunitomi, M., Tassetto, M., Seetin, M. G., Oh, S., Heiner, C., Paxinos, E. and Andino, R. (2017) 'The Diversity, Structure, and Function of Heritable Adaptive Immunity Sequences in the Aedes aegypti Genome', *Current Biology*. doi: 10.1016/j.cub.2017.09.067.

Williams, C. J. *et al.* (2018) 'MolProbity: More and better reference data for improved all-atom structure validation', *Protein Science*. doi: 10.1002/pro.3330.

Xi, Z., Ramirez, J. L. and Dimopoulos, G. (2008) 'The Aedes aegypti toll pathway controls dengue virus infection', *PLoS Pathogens*. doi: 10.1371/journal.ppat.1000098.

Xiao, X., Zhang, R., Pang, X., Liang, G., Wang, P. and Cheng, G. (2015) 'A Neuron-Specific Antiviral Mechanism Prevents Lethal Flaviviral Infection of Mosquitoes', *PLoS Pathogens*, 11(4), pp. 1–33. doi: 10.1371/journal.ppat.1004848.

Yáñez-Mó, M. *et al.* (2015) 'Biological properties of extracellular vesicles and their physiological functions', *Journal of Extracellular Vesicles*. doi: 10.3402/jev.v4.27066.

Yang, Z. (2007) 'PAML 4: Phylogenetic analysis by maximum likelihood', *Molecular Biology and Evolution*. doi: 10.1093/molbev/msm088.

Yuan, J. S., Burris, J., Stewart, N. R., Mentewab, A. and Neal, C. N. (2007) 'Statistical tools for transgene copy number estimation based on real-time PCR', *BMC Bioinformatics*, 8(SUPPL. 7), pp. 1–12. doi: 10.1186/1471-2105-8-S7-S6.

Zanni, V., Eymery, A., Coiffet, M., Zytnicki, M., Luyten, I., Quesneville, H., Vaury, C. and Jensen, S. (2013) 'Distribution, evolution, and diversity

of retrotransposons at the flamenco locus reflect the regulatory properties of piRNA clusters', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.1313677110.

Zhang, M., Zheng, X., Wu, Y., Gan, M., He, A., Li, Z., Liu, J. and Zhan, X. (2010) 'Quantitative analysis of replication and tropisms of dengue virus type 2 in Aedes albopictus', *American Journal of Tropical Medicine and Hygiene*. doi: 10.4269/ajtmh.2010.10-0193.

Zhao, L., Alto, B. W. and Shin, D. (2019) 'Transcriptional profile of Aedes aegypti leucine-rich repeat proteins in response to zika and chikungunya viruses', *International Journal of Molecular Sciences*. doi: 10.3390/ijms20030615.

Zimmermann, L., Stephens, A., Nam, S. Z., Rau, D., Kübler, J., Lozajic, M., Gabler, F., Söding, J., Lupas, A. N. and Alva, V. (2018) 'A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core', *Journal of Molecular Biology*. doi: 10.1016/j.jmb.2017.12.007.

List of original manuscripts

1) **Houé V, Gabiane G, Dauga C, Suez M, Madec Y, Mousson L, <u>Marconcini M</u>, Yen PS, de Lamballerie X, Bonizzoni M, Failloux AB.** Evolution and biological significance of flaviviral elements in the genome of the arboviral vector *Aedes albopictus*. *Emerg Microbes Infect*. 2019;8(1):1265-1279. doi: 10.1080/22221751.2019.1657785. PMID: 31469046; PMCID: PMC6735342.

2) **<u>Marconcini M</u>, Hernandez L, Iovino G, Houé V, Valerio F, Palatini U, Pischedda E, Crawford JE, White BJ, Lin T, Carballar-Lejarazu R, Ometto L, Forneris F, Failloux AB, Bonizzoni M. (in press, PLOS NTDs)** Polymorphism analyses and protein modelling inform on functional specialization of Piwi clade genes in the arboviral vector *Aedes albopictus.* Preprint available on: bioRxiv 677062; doi: https://doi.org/10.1101/677062.

EMi · Taylor & Francis
Taylor & Francis Group

Evolution and biological significance of flaviviral elements in the genome of the arboviral vector *Aedes albopictus*

Vincent Houé[a,b], Gaelle Gabiane[a], Catherine Dauga[c], Marie Suez[d], Yoann Madec[e], Laurence Mousson[a], Michele Marconcini [f], Pei-Shi Yen[a], Xavier de Lamballerie[g,h], Mariangela Bonizzoni [f] and Anna-Bella Failloux [a]

[a]Department of Virology, Arboviruses and Insect Vectors, Institut Pasteur, Paris, France; [b]Sorbonne Université, Collège Doctoral, Paris, France; [c]Institut Pasteur, Center for Bioinformatics, BioStatistics and Integrative Biology (C3BI), Paris, France; [d]Institut de Biologie Paris-Seine, Paris, France; [e]Department of Infection and Epidemiology, Institut Pasteur, Epidemiology of Emerging Diseases, Paris, France; [f]Department of Biology and Biotechnology, University of Pavia, Pavia, Italy; [g]Aix Marseille Université, IRD French Institute of Research for Development, EHESP French School of Public Health, EPV UMR_D 190 'Emergence des Pathologies Virales', Marseille, France; [h]IHU Méditerranée Infection, APHM Public Hospitals of Marseille, Marseille, France

ABSTRACT
Since its genome details are publically available, the mosquito *Aedes albopictus* has become the central stage of attention for deciphering multiple biological and evolutionary aspects at the root of its success as an invasive species. Its genome of 1,967 Mb harbours an unusual high number of non-retroviral integrated RNA virus sequences (NIRVS). NIRVS are enriched in piRNA clusters and produce piRNAs, suggesting an antiviral effect. Here, we investigated the evolutionary history of NIRVS in geographically distant *Ae. albopictus* populations by comparing genetic variation as derived by neutral microsatellite loci and seven selected NIRVS. We found that the evolution of NIRVS was far to be neutral with variations both in their distribution and sequence polymorphism among *Ae. albopictus* populations. The Flaviviral elements AlbFlavi2 and AlbFlavi36 were more deeply investigated in their association with dissemination rates of dengue virus (DENV) and chikungunya virus (CHIKV) in *Ae. albopictus* at both population and individual levels. Our results show a complex association between NIRVS and DENV/CHIKV opening a new avenue for investigating the functional role of NIRVS as antiviral elements shaping vector competence of mosquitoes to arboviruses.

Introduction

In less than four decades, the Asian tiger mosquito *Aedes albopictus* (Skuse, 1894) has become a public health concern owing to its ability to transmit several human pathogenic viruses such as Chikungunya virus (CHIKV), Dengue viruses (DENV) and Zika virus (ZIKV) [1]. This vector is responsible of several recent arboviral outbreaks (reviewed by [2]) and astonished by the speed at which it has conquered all continents except Antarctica [3,4], thus becoming one of the world's 100-most invasive species according to the Global Invasive Species Database. Originally native to tropical forests of South-East Asia, *Ae. albopictus* was confined for centuries to few regions in Asia. Starting in the eighteenth or nineteenth centuries, *Ae. albopictus* was introduced in the islands of the Indian Ocean by Asian immigrants. From the late 1970s, this species took advantage of the increase in global trade to invade most tropical and sub-tropical regions of the world [5]. Additionally, the capacity of its eggs to undergo

photoperiodic diapause in winter, favoured *Ae. albopictus* colonization into temperate regions [6]. *Aedes albopictus* was first reported in Europe in 1979 [7], in North America in 1985 [8], in South America in 1986 [9] and in Africa in 1990 [10].

Migratory routes used by *Ae. albopictus* to expand from its Asian cradle can be defined using population genetic approaches [11–13]. Mosquito populations form groups of interbreeding individuals, which coexist in space and time. These genetic units are interconnected through gene flow. The current worldwide distribution of *Ae. albopictus* is a direct consequence of increasing human activities with rapid irradiation of populations [14–18]. Populations established locally presented a high genetic variability evidencing a mixture of individuals of distinct origins with disparate susceptibilities to arboviruses, which are reflected by different vector competences [17]. Vector competence refers to the ability of a mosquito population to become infected after an infectious blood meal, to support viral

replication, dissemination and transmission to a new host in a subsequent blood-meal [19]. The level of vector competence depends on the tripartite interactions among the mosquito genotype, the virus genotype and environmental factors under GxGxE interactions [20].

In the early 2000s, non-retroviral integrated RNA virus sequences (NIRVS) were discovered in different metazoans, including mosquitoes [21,22]. *Aedes* mosquitoes can host NIRVS originated from different viruses related to arboviruses: mainly insect-specific viruses (ISVs) including insect-specific flaviviruses (ISFs) and other viruses belonging to the *Mononegavirales* order (such as rhabdoviruses) [22–25]. Additionally, most NIRVS correspond to fragmented viral open-reading frames, are flanked by transposable elements (TEs), enriched in PIWI-interacting RNA (piRNA) clusters and produce piRNAs [29]. piRNA clusters are genomic regions composed of fragmented sequences of TEs which are expressed as long primary single-stranded RNAs and processed into fragments of 24–30 nucleotides called piRNAs. In the model organism *Drosophila melanogaster*, piRNAs are primarily produced in germline cells and target TE transcripts based on sequence complementarity to protect from heritable lesions [26–29]. The landscape of TE fragments within piRNA clusters defines the regulatory properties of *D. melanogaster* strains to TE invasion [30]. The analogy between TE fragments and viral sequences in piRNA supports the hypothesis that viral sequences may contribute to mosquito susceptibility to subsequent viral infections. If viral integrations contribute in controlling virus replication, with consequences on vector competence, positive selection should be expected [31]. On the contrary, if NIRVS stand for fossil records, these sequences should reach fixation and evolve at a neutral rate [32,33].

In this study, we selected 19 *Ae. albopictus* populations that cover the geographical distribution of the species, where CHIKV and DENV were circulating, to study the evolutionary dynamics of seven selected NIRVS [24]. The occurrence of NIRVS in populations was compared to processes driving population genetic differentiation observed on neutral loci (i.e. microsatellites). We showed that (i) based on microsatellite marker polymorphism, populations of *Ae. albopictus* are distributed into two different genetic clusters, one of them divided into four subclusters, without any correlation between genetic and geographical distances, (ii) the distribution of NIRVS in geographic populations and their polymorphism are not related to genetic divergences within and between populations as depicted by microsatellite markers, suggesting that NIRVS are not evolving neutrally, and (iii) all NIRVS studied, except AlbFlavi1 may have influence on vector competence to DENV and CHIKV, at the population level.

Materials and methods

Ethic statements

Mice were housed in the Institut Pasteur animal facilities accredited by the French Ministry of Agriculture for performing experiments on live rodents. Work on animals was performed in compliance with French and European regulations on care and protection of laboratory animals (EC Directive 2010/63, French Law 2013-118, February 6th, 2013). All experiments were approved by the Ethics Committee #89 and registered under the reference APAFIS#6573-201606l412077987 v2.

Mosquitoes

To assess genetic variation within and among *Ae. albopictus* populations, 19 samples from different geographic locations were studied (Table 1). These populations ranging from 10 to 30 mosquitoes were selected in the native range of *Ae. albopictus* where arboviral outbreaks and epidemics occurred [17,34,35]. Only the 13 populations of 19–20 mosquitoes and the Foshan colony used as control were selected to characterize NIRVS polymorphism [36].

Mosquitoes were collected in the field as immature stages (larvae, pupae, eggs). Frozen mosquitoes (<13th generation) were used for population genetic and NIRVS analyses. For the pilot experiments for DENV and CHIKV infections, F1 mosquitoes from Tibati population (Cameroon) and F18 mosquitoes from Foshan laboratory strain were used. Larvae obtained after immersion in dechlorinated tap water of field-collected eggs were distributed in pans of 200 individuals. Immature stages were fed every two days with a yeast tablet dissolved in 1 L of dechlorinated tap water and incubated at $26 \pm 1°C$. Emerging adults were placed in cages and maintained at $28 \pm 1°C$ with a light/dark cycle of 12/12 h at 80% relative humidity and supplied with a 10% sucrose solution. Females were exposed three times a week to anesthetized mice (OF1 mice; Charles River Laboratories, MA, USA) as a source of blood for producing eggs.

Microsatellite genotyping

Genomic DNA was extracted from individual mosquito using the Nucleospin Tissue kit (Macherey-Nagel, Hoerdt, France) according to manufacturer's instructions. Briefly, mosquitoes were individually homogenized in 180 μL lysis buffer supplemented with 25 μL of Proteinase K. To bind total nucleic acids, homogenates were passed through columns. Silica membranes were further desalted and DNA was collected in 100 μL of elution buffer. Quality and quantity of DNA were then assessed using the

Table 1. Details on *Aedes albopictus* populations analyzed.

	Population name	Continent	Country	City	Generation	Year of collection	Mosquitoes females	Males
1	Alessandria	Europe	Italy	Alessandria	F2	2012	10	10
2	Ulcinj	Europe	Montenegro	Ulcinj	F1	2013	10	10
3	Cagnes-sur-mer	Europe	France	Cagnes-sur-mer	F13	2000	10	0
4	Montsecret	Europe	France	Montsecret	F4	2002	0	10
5	Bar-sur-Loup	Europe	France	Bar-sur-Loup	F1	2011	10	0
6	Tirana	Europe	Albania	Tirana	F6	2016	15	15
7	Franceville	Africa	Gabon	Franceville	F2	2015	10	10
8	Mfilou	Africa	Congo	Mfilou (Brazzaville)	F3	2012	10	10
9	Bertoua	Africa	Cameroon	Bertoua	F5	2008	10	4
10	Saint-Denis	Africa	La Réunion	Saint-Denis	F2	1998/2006	10	10
11	Rabat	Africa	Morocco	Rabat	F1	2017	15	15
12	Vero Beach	America	USA	VeroBeach	F5	2016	10	10
13	Rio	America	Brazil	Rio de Janeiro	F1	2001	10	9
14	Jurujuba	America	Brazil	Jurujuba	F1	2014	10	0
15	Manaus	America	Brazil	Manaus	F1	2015	15	15
16	PMNI	America	Brazil	PMNI (Nova Iguaçu)	F1	2015	15	15
17	Binh Duong	Asia	Vietnam	Binh Duong (Ben Cat) Phu Hoa	F9	2014	10	10
18	Sarba	Asia	Lebanon	Sarba	F0	2011	10	10
19	Foshan	Asia	China	Foshan	Lab colony	–	10	10
20	Oahu	America	Hawaii	Oahu	Lab colony	1999	10	10

Note: Except the colony Foshan, all 19 populations were genetically characterized using 10 microsatellite markers, and 13 populations (in bold) were selected for studying NIRVS diversity.

Nanodrop 2000 Spectrophotometer (*Thermo* Scientific™, MA, USA) and a PCR was performed using *histone h3* reference gene (NCBI: XM_019696438.1) as control. Eleven microsatellite loci were amplified using PCR specific primers flanking the repeated region [16]. PCR reaction mixtures in a final volume of 15 µL contained 50 ng genomic DNA, 1X PCR buffer, 1.5 mM MgCl2, 0.27 mM dNTPs (Invitrogen™, CA, USA), 1U Taq polymerase (Invitrogen™), and 10 µM of each primer (one was 5′ labelled with a fluorescent dye). PCR cycling conditions consisted in a step at 94°C for 5 min, followed by 29 cycles at 94°C for 30 s, 58°C for 30 s, and 72°C for 30 s with a final step at 72°C for 10 min. Aliquots of PCR products were visualized in 2% agarose gels stained with ethidium bromide under UV light. Each PCR product was then diluted 1:10 in ddH2O water and 2 µL of this dilution was added to 10 µL of a mixture of deionized formamide and GeneScan-500 ROX size standard (Applied Biosystems, CA, USA). Genotyping was processed in an ABI3730XL sequence analyser (Applied Biosystems) and data analyzed using GeneScan and Genemapper softwares.

Genetic diversity of populations

For statistical analyses, mosquitoes collected from each sampling site were assumed to represent local populations.

Amounts of heterozygosity at various levels of population structure were explored by using F_{IS} and F_{ST} values. The F_{IS} inbreeding coefficient indicates the level of heterozygosity within each population. The F_{ST} fixation index measures the reduction in heterozygosis due to random genetic drift between populations. The Wright's F-statistics were computed in Genetix v.4.05.2. software [37] and tested using 10^4 iterations

according to Weir and Cockerham (1984). The number of alleles (N_A), allelic richness (A_R), expected (H_E), observed (H_O) heterozygosis, F_{IS} by locus and F_{ST} by populations were obtained. The significance of F_{IS} and F_{ST} were analyzed using FSTAT v.2.9.3.2 [38].

Departures from Hardy-Weinberg equilibrium, linkage disequilibrium between loci and molecular variance (AMOVA) over all populations were analyzed with Arlequin v3.5.2.2 to estimate intra- and inter-population variation [39]. The mean frequency of null alleles (a mutation in microsatellite flanking regions leading to an absence of amplification products) per population was also very low (i.e. 0.06) and ranged from 0.000 to 0.118, meaning that the selected microsatellite loci were successfully amplified and appropriate for population genetic analysis. Relationships between geographical and genetic F_{ST} distances were tested between populations using Mantel test implemented in GenAlEx v 6.5 [40].

Graphic representation of relatedness among populations

Genetic relationships between populations were estimated by using PHYLIP 3.69, as previously described [41]. Cavalli-Sforza & Edwards's (CSE) chord distance for each pair of populations was calculated (GENDIST module). The resulting distance matrix was used to create a phylogram based on Neighbour-Joining (NEIGHBOUR module). Node confidence was inferred via 100 bootstrap replicates (modules SEQ- BOOT, GENDIST, NEIGHBOUR and CONSENSE).

Test of isolation by distance

To test the hypothesis whether the geographical pattern of genetic differentiation is caused by isolation

by distance (IBD), we ran Mantel tests for pairwise matrices between geographical distances (kilometres) and genetic differentiation (measured as $F_{ST}/(1-F_{ST})$).

Genetic structure of populations

Genetic population structure was assessed with individual-based Bayesian clustering method implemented in the program STRUCTURE v.2.3.4 [42]. The likelihood of each possible number of genetic populations (K), ranging from 1 to 20, was calculated after 10 independent runs for each value of K, using a burn-in of 500,000 replications, 500,000 Markov Chain Monte Carlo steps, assuming an admixture model, with frequencies correlated between populations and without the use of sampling location as a prior. The most likely number of populations (K) was estimated by the ΔK method described by Evanno et al. [43] with Structure Harvester software (http://taylor0.biology.ucla.edu/structureHarvester/). The results were then graphically displayed with DISTRUCT 1.1 [44].

NIRVS in natural populations

Six NIRVS (AlbFlavi1, AlbFlavi2, AlbFlavi4, AlbFlavi10, AlbFlavi36, AlbFlavi41), plus CSA [21], were studied. They were chosen based on their unique occurrence in different regions of the Foshan genome, except for AlbFlavi41, known to be duplicated. Because of its length [21], CSA was characterized using two sets of primers (CSA-NS3 and CSA-JJL), and considered in analysis as two separated datasets (Supplementary Table 1). NIRVS were searched on the same mosquitoes as those used for microsatellite genotyping.

PCR primers flanking each NIRVS (Supplementary Table 1) were designed using PRIMER3 [45]. PCR reactions were performed in a final volume of 25 μL using 5 ng of DNA, PCR buffer 1X, 2.9 mM MgCl₂, 0.25 mM dNTP mix, 0.25 μM of each primer and 1.25 unit of *Taq* DNA polymerase (Invitrogen™). Amplifications were done using *T100™ Thermal Cycler* (Bio-Rad) according to the following cycle conditions: 95°C for 5 min, 35 cycles at 95°C for 15 s, 59–64°C for 90 s, 72°C for 90 s, and a final elongation step at 72°C for 10 min. PCR products were electrophoresed on 1.5% of agarose gels stained with ethidium bromide and visualized under UV light. All negative samples were confirmed with a second-round PCR.

The number of each NIRVS per population (frequency) was represented using Graph Pad Prism version 6 (Graph Pad Software, CA, USA).

NIRVS between populations

NIRVS composition of each population was expressed in terms of relative abundance corresponding to the percentage of each NIRVS relative to the total number of tested mosquitoes. To assess variations of NIRVS abundance between populations, we calculated Bray Curtis dissimilarities, a metrics, widely used, including data standardization [46]. A dendrogram was generated by using this dissimilarity matrix and neighbour-joining method, to visualize NIRVS compositional differences across populations [46].

Mantel tests between NIRVS and microsatellites

Mantel tests implemented in GenAlEx v 6.5 [40], were used to evaluate the statistical significance of the correlation between two or more distance matrices, using permutation tests. The significance of associations between matrixes of φ′st distance among populations as derived by microsatellite markers or NIRVS distribution was tested using the Mantel test with 999 permutations.

NIRVS association with vector competence of Ae. albopictus at the population level

Vector competence is assessed using several parameters [47]. The dissemination efficiency (DE) refers to the proportion of mosquitoes able to disseminate the virus beyond the midgut barrier after ingestion of infectious blood meal and active viral replication in midgut epithelial cells. Published data on DEs to DENV and CHIKV were retrieved for mosquito populations with close or identical geographical proximity with the populations analyzed (Supplementary Table 2). DEs were described using median and inter-quartile range (IQR). Logistic linear regression models were used to test association between the presence of NIRVS and DEs at the *Ae. albopictus* population level. *P*-values <0.05 were considered significant.

Sequence polymorphism and evolutionary dynamics of AlbFlavi2 and AlbFlavi36

Sequence polymorphism of AlbFlavi2 and AlbFlavi36

Amplification products for AlbFlavi2 and AlbFlavi36 were purified by using NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel) according to the manufacturer's instructions and sequenced by Eurofins Genomics (Cochin hospital platform, Paris, France). If the quality of chromatogram profiles assessed with Geneious® 10.1.3 did not meet the standard required, amplicons were subsequently cloned into pCR™II-TOPO® vector using *TOPO® TA Cloning®Kit (Thermo Fisher Scientific)* and transformed into One Shot® TOP10 Chemically Competent *E. coli* (Invitrogen™). Sequences were aligned in Geneious® 10.1.3 using MUSCLE algorithm [48]. p-distance values, representing proportion of nucleotide sites at which two

sequences being compared are different, were calculated after alignments using MEGA 10.0.5.

Phylogeny based on AlbFlavi2 and AlbFlavi36

DNA sequences from AlbFlavi2 and AlbFlavi36 were aligned with MUSCLE algorithm [48]. Exogenous virus sequences were used as outgroup to determine the direction of character transformations. Phylogenetic trees were obtained by parsimony analysis implemented in the PAUP* software package (version 4.0), by using gap as 5th state or not (data not shown) and nearest neighbour interchange (NNI) for tree rearrangements [49].

Pilot experiment assessing the association between AlbFlavi2/Albflavi36 and vector competence of Ae. albopictus at the individual level

Logistic linear regression models were used to test association between the presence of AlbFlavi2 and AlbFlavi36 and viral dissemination in single mosquitoes from the Foshan colony and a field-collected African population (Tibati collected in Tibati, Cameroon in 2018). Foshan and Tibati mosquitoes were experimentally infected with DENV (DENV-2, Accession number: MK268692 [50]) and CHIKV (06.21, Accession number: AM258992 [51]) provided in blood meals as described in Amraoui et al. 2019 [52]. All surviving mosquitoes were examined individually at 14 days post-infection to define (i) the infectious status; presence of infectious particles in heads was estimated by focus fluorescent assay on C6/36 Ae. albopictus cells [52], and (ii) the presence of AlbFlavi2 and/or AlbFlavi36 by PCR on DNA extracted from bodies and thorax. Statistical analyses were conducted using the Stata software (StataCorp LP, Texas, and USA). P-values <0.05 were considered significant.

Results

Characterization of populations using microsatellite data

A total of 363 individuals (Table 1) were genotyped at 11 microsatellite markers. Ten of the 11 microsatellite loci, by showing a frequency upper to 0.95 for the most common allele, were considered polymorphic and as such, were included in the analysis. Only the A17 locus was monomorphic and then excluded from the analyses. No significant linkage disequilibrium between any pairs of loci was observed indicating that the 10 loci were statistically independent from each other. Each locus was tested for within-population deviations from HWE implementing Bonferroni correction for multiple testing: 20 deviations out of 190

Table 2. Genetic diversity at each microsatellite locus for all mosquito populations.

	N_A	N	Allelic richness	PIC	H_o	H_E	F_{IS}
A1	4	288		0.44	0.16	0.47	**0.59**
A2	10	360	2.43	0.55	0.63	0.62	−0.15
A3	18	341	3.14	0.74	0.62	0.78	0.08
A5	19	334	2.90	0.75	0.52	0.78	0.13
A6	11	330	2.59	0.68	0.54	0.68	−0.04
A9	9	326	2.73	0.65	0.59	0.70	0.01
A11	9	323	2.81	0.71	0.53	0.75	0.15
A14	9	351	1.71	0.41	0.17	0.40	**0.38**
A15	6	347	1.25	0.13	0.04	0.15	**0.57**
A16	14	279		0.83	0.66	0.84	0.04
Mean	10.9	327.9	2.44	0.59	0.45	0.62	**0.17**

Notes: N_a, number of alleles; N, number of individuals examined at a locus; PIC, polymorphism information content; Ho, observed heterozygosity; He, expected heterozygosity; F_{IS}, the inbreeding coefficient. Values in bold are significant at the 0.1% after Bonferroni correction.

combinations were found but did not cluster at any population or locus.

The loci displayed a mean PIC of 0.59 suggesting that they were sufficiently informative for assessing the degree of variability and structuring of mosquito populations (Table 2). The number of alleles, averaged over all loci, ranged from 4 to 19, with a mean value of 10.9 per locus, and a mean allelic richness at 2.44 (Table 2). Eight of the 10 loci presented a significant difference between observed and expected heterozygosity, with a mean F_{IS} of 0.17 (p-value <0.001; Table 2), supporting an excess of homozygotes in Ae. albopictus populations.

Population diversity

The overall mean number of alleles per population was 3.7 varying between 2.0 and 5.0. Private alleles were rarely found, their mean frequency per population ranged from 0.00 to 0.079 (Table 3). Mean F_{IS} value per population was −0.09, with values ranging from −0.0285 to 0.338 (Table 3). Estimation of the molecular variance within and among populations (AMOVA) revealed that most of the variation (89.6%) was detected within individuals whereas only 10.4% occurred among populations.

Population genetic structure

The overall differentiation across all 19 populations was high ($F_{ST} = 0.239$) with F_{ST} being highly significant (p-value <0.001), suggesting a significant genetic structure. To identify genetic clusters among tested individuals, we implemented 25 independent simulations in the software Structure according to the method of Evanno; the uppermost level of structuring in the model was observed at K = 2 (ΔK = 2293.4; Supplementary Figure 1(A)). The best assignment of individuals made by Structure led to two clusters: (i) cluster 1, which includes populations from Brazil (i.e. PMNI and Manaus), Northern Africa (i.e. Rabat) and a population from Albania, and (ii) cluster 2, which includes the remaining populations (Figure 1 and

Table 3. Analysis of genetic variability of different geographical populations of *Aedes albopictus*.

Populations	Country	Continent	N	n_a	n_a/n	n_p	n_p/n	A_p	Ho	H_E	A_n	F_{IS}
Vero Beach	USA	America	20	4.4	0.22	2	0.10	0.07	00.48	0.53	0.08	0.12***
Oahu	Hawaii		20	2.6	0.13	0	0.00	0.00	0.25	0.34	0.08	0.30***
Manaus	Brazil		30	3.4	0.11	3	0.10	0.07	0.32	0.47	0.12	0.34***
PMNI			30	3.6	0.12	1	0.03	0.03	0.39	0.51	0.11	0.25***
Rio			19	3.4	0.18	0	0.00	0.00	0.42	0.50	0.08	0.18***
Jurujuba			10	3.3	0.33	0	0.00	0.00	0.47	0.42	0.00	−0.06
Montsecret	France	Europe	10	2.0	0.20	0	0.00	0.00	0.50	0.37	0.01	−0.29
Cagnes-sur-Mer			10	2.8	0.28	1	0.10	0.06	0.55	0.46	0.01	−0.14
Bar-sur-Loup			10	3.8	0.38	2	0.20	0.08	0.50	0.52	0.05	0.13*
Alessandria	Italy		20	4.1	0.21	2	0.10	0.05	0.43	0.50	0.06	0.18***
Ulcinj	Montenegro		20	3.1	0.16	0	0.00	0.00	0.40	0.39	0.03	0.02
Tirana	Albania		30	3.8	0.13	3	0.10	0.07	0.35	0.43	0.08	0.21***
Rabat	Morocco		30	3.2	0.11	0	0.00	0.00	0.42	0.47	0.06	0.11***
Bertoua	Cameroon		14	4.3	0.31	1	0.07	0.07	0.46	0.52	0.08	0.16***
Franceville	Gabon	Africa	20	4.8	0.24	1	0.05	0.03	0.44	0.56	0.08	0.25***
Mfilou	Congo		20	4.0	0.20	0	0.00	0.00	0.60	0.55	0.02	−0.06
Saint-Denis	La Réunion		20	5.1	0.26	6	0.30	0.05	0.52	0.55	0.07	0.08
Sarba	Lebanon	Asia	10	3.5	0.35	4	0.40	0.08	0.55	0.44	0.00	−0.18
Binh Duong	Vietnam		20	5.0	0.25	3	0.15	0.08	0.2	0.41	0.05	0.11**
Mean			19.1	3.7	0022	1.5	0.09	0.04	0.45	0,47	0.06	−0.09

Notes: N, population size; n_a, mean number of alleles; n_a/N, mean number of alleles/individual; n_p, number of private alleles; n_p/N, mean number of private alleles/individual; A_p, mean frequency of private alleles; H_O, mean observed heterozygosity; H_E, mean expected heterozygosity; A_n, mean frequency of null alleles; F_{IS}, the inbreeding coefficient.
**p-value $\leq$ 0.01; ***p-value $\leq$ 0.001.

Supplementary Figure 1(B)). Although Structure identified only two clusters, their genetic differentiation was significantly high ($F_{ST} = 0.12$, p-value ≤ 0.001). Additionally, these two genetic clusters were found to deviate significantly (p-value ≤ 0.01) from Hardy-Weinberg proportions, with F_{IS} values of 0.23 and 0.36, respectively after implementing 10^4 permutations on allele frequencies in the software GENETIX. This result suggested a Wahlund effect which indicated a genetic substructure. As a consequence, we re-analyzed the two clusters separately and detected substructures in Cluster 2 (Figure 1 and Supplementary Figure 1 (C)). Sub-structuring of Cluster 2 emphasized differentiation among populations from South America (Jurujuba and Rio), Europe (i.e. all French and Italian populations), Africa (Bertoua and Mfilou) and Asia. On the contrary, populations from Florida (Vero Beach), Middle East (Sarba), La Reunion Island (Saint Denis) and Gabon (Franceville) appeared genetically mixed suggesting that these sites may represent cross-roads in *Ae. albopictus* colonization out of Asia (Figure 1).

Spatial and genetic data

When considering the genetic differentiation according to geographical distances using Mantel test, only 21% of the genetic variability was explained by the geographical distance (p-value = 0.03).

NIRVS analysis

We further chose to study NIRVS distribution in 13 *Ae. albopictus* populations selected previously for assessing genetic diversity based on epidemiological data of arboviral outbreaks and *Ae. albopictus* widespread

out if its native range [17,34,35]; we analyzed 10 females and 10 males per population (Table 1). NIRVS distribution was not homogenous across populations (Figure 2, Supplementary Figure 2). AlbFlavi1 (Figure 2(A)) and CSA-JJL (Figure 2(H)) were the most widespread NIRVS, being detected in 85–100% of tested mosquitoes, respectively. On the opposite, AlbFlavi10 was the rarest NIRVS, being found in 17% of the tested mosquitoes. Despite displaying the highest presence of AlbFlavi2 and AlbFlavi4 (present in most populations), Brazilian populations showed the lowest number of NIRVS, with the absence of Alb-Flavi36 and AlbFlavi10 and the lower frequency of Alb-Flavi41 compared to the other *Ae. albopictus* populations (Figure 2).

Interestingly, when targeting the CSA locus, no amplification was obtained for Brazilian mosquitoes using the NS3-CSA primer set (Supplementary Table 1; Figure 2(G)), which was not consistent with the results obtained with the JJL-CSA primer set that amplified the same NIRVS but targeting a different region (Figure 2(H)). This suggests then that recombination events occurred at this genomic position for Manaus, Rio and PMNI mosquitoes.

Comparison between microsatellite data and NIRVS abundance profiles

To define whether NIRVS evolve under selective pressures or randomly, we compared (i) the relatedness between populations based on microsatellite polymorphism (Figure 3(A)) with (ii) the similarity between populations based on NIRVS abundance profiles (Figure 3(B)). At the population level, the Neighbour-Joining clustering analyses applied to genetic frequencies (Figures 1 and 3(A), Supplementary
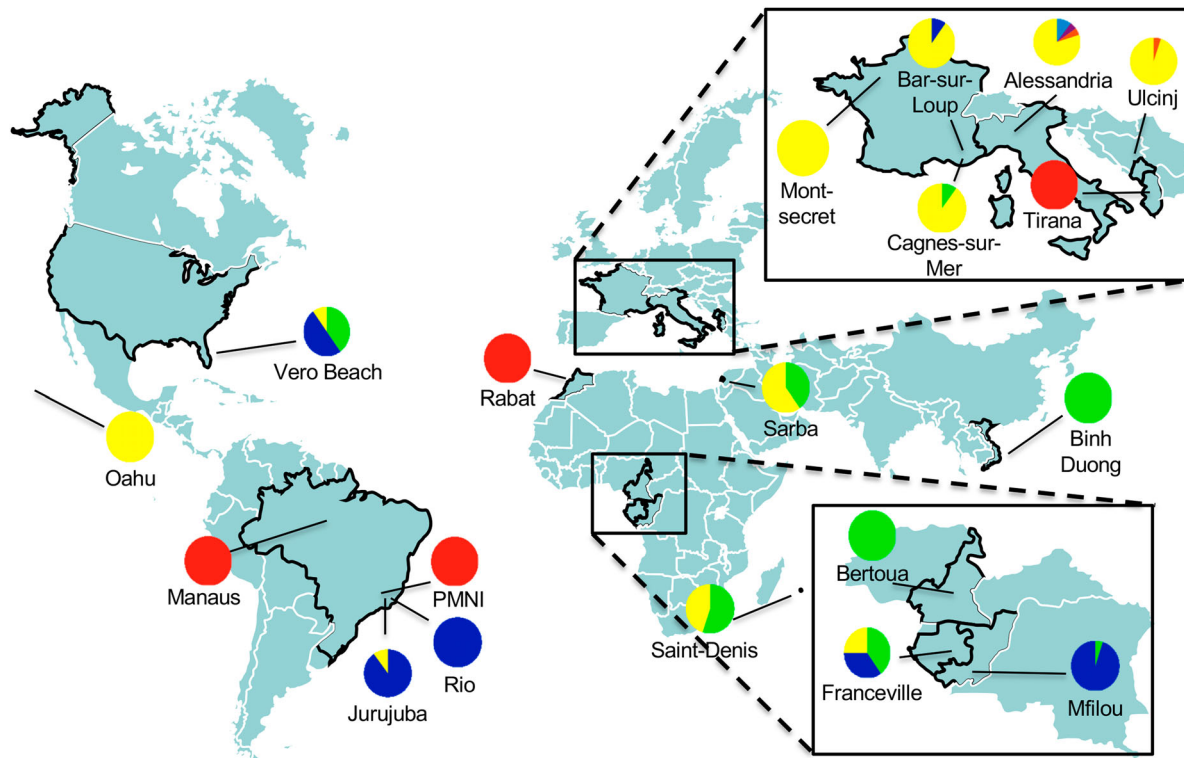
Figure 1. Estimated population structure of 363 individuals (19 populations) using 10 microsatellite markers. Map with sampling sites of populations with colour pie charts showing genotype frequencies, according to Cluster 1 (red) and Cluster 2, which the latest subdivided into 4 subclusters (blue, green, yellow and orange), deduced from the ΔK curve obtained (Supplementary Figure 1(A–C)).

Figure 1(B)) mainly revealed two clusters of populations: one with Tirana, Rabat, Manaus and PMNI populations, and the second with the remaining populations (i.e. Binh Duong, Oahu, Vero Beach, Rio, Franceville, Alessandria, Ulcinj and Mfilou). Regarding NIRVS contents, two major clusters distantly related from each other were also obtained (Figure 3 (B)), one including only Brazilian populations (i.e. Manaus, PMNI and Rio) sharing a low abundance of NIRVS, and the other including geographically distant populations. Therefore, populations from Vero Beach and Rio shared closely related genetic relationships and harboured clearly different abundance profiles of NIRVS (Figure 3(A,B)). Conversely, genetic distinct populations from Vero Beach, Rabat and Alessandria shared the same contents of NIRVS. In short, closely related populations can have different NIRVS contents and geographic and/or genetic distant populations can contain similar NIRVS abundance profiles. Thus, we observed significantly different relatedness between populations according to the marker used, that may indicate that random genetic drift was not the main force shaping these NIRVS distribution. In addition, no correlation was identified between the two matrices ($R^2 = 0.0006$), showing that NIRVS composition of populations was not related to the genetic structure caused by random genetic drift (Supplementary Figure 2).

Relationships between NIRVS landscape and vector competence (from published data)

Our results showed that NIRVS were not neutrally distributed across populations supporting the hypothesis of a biological function of NIRVS, such as antiviral functions, as it has been previously suggested. To analyze whether NIRVS contribute to the control of arboviral replication, we used logistic regression models to evaluate potential association between NIRVS distribution frequencies among populations and dissemination efficiencies (DEs) of DENV and CHIKV in corresponding or geographically close populations selected from the literature (Supplementary Table 2). On this basis, we classified our tested populations depending on their frequencies of each NIRVS using the median. Several associations between NIRVS and DENV DEs were found. Whereas high AlbFlavi2 and CSA-JJL frequencies were significantly associated with high DEs, high AlbFlavi10, AlbFlavi36, AlbFlavi41 and CSA-NS3 frequencies were however correlated to low DENV DEs (Table 4). Opposite associations to CHIKV DEs were also observed but with fewer NIRVS. Indeed, high frequencies of AlbFlavi4 were associated with high CHIKV DEs while high distribution of AlbFlavi36 and CSA-NS3 among populations was correlated to low CHIKV DEs. Together, these results indicated
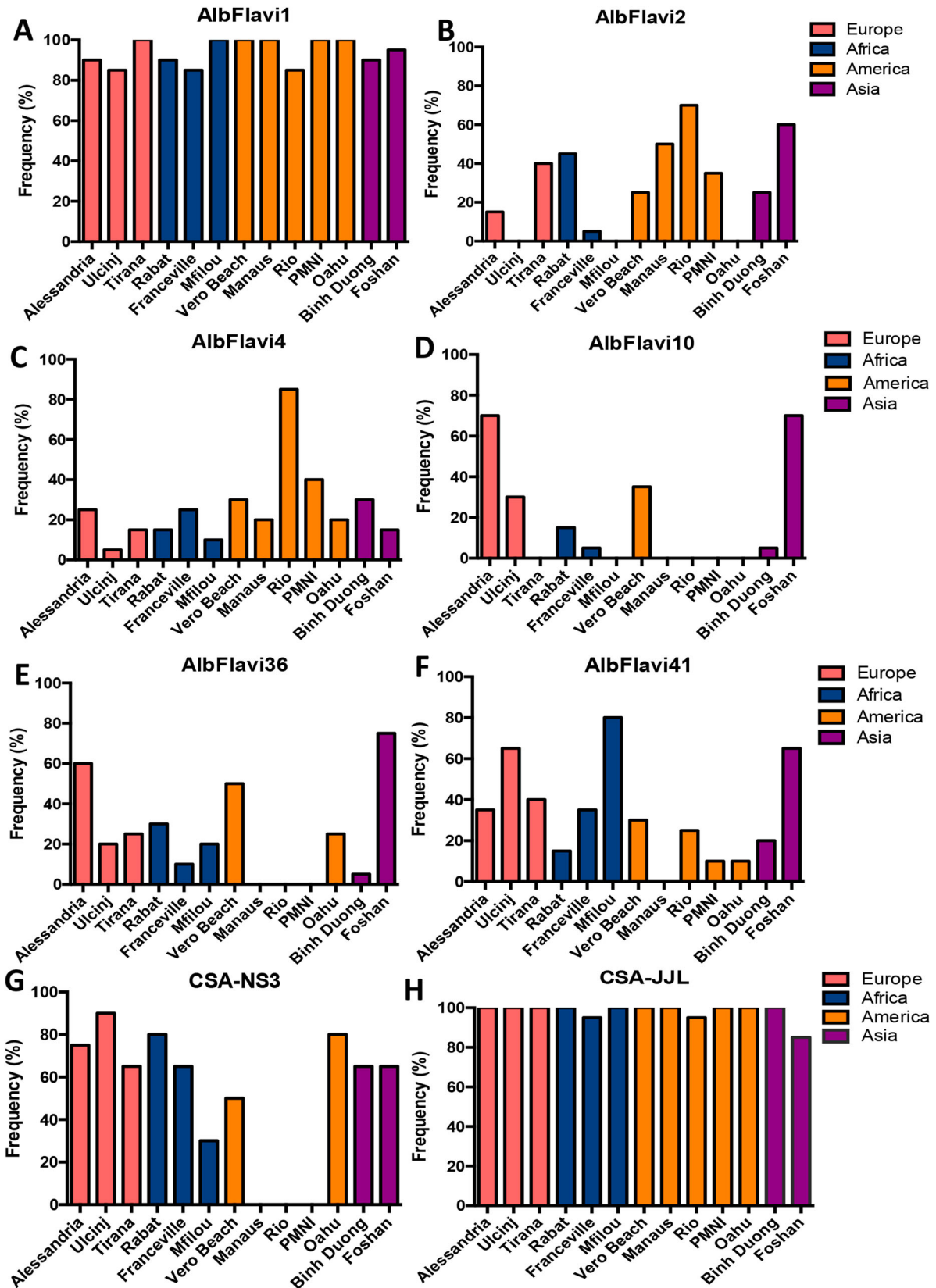
Figure 2. NIRVS variability among *Aedes albopictus* populations. The frequency of AlbFlavi1 (A), AlbFlavi2 (B), AlbFlavi4 (C), Alb-Flavi10 (D), AlbFlavi36 (E), AlbFlavi41 (F) and CSA (G and H) was assessed for 20 individuals in each *Ae. albopictus* population (except the Rio population with 19 individuals). Populations were clustered according to their continent of origin. Oahu and Foshan correspond to laboratory colonies. The variability of CSA was assessed using two sets of primers: CSA-NS3 (G) and CSA-JJL (H).

that the presence of NIRVS was associated with changes in DEs of arboviruses in *Ae. albopictus* at the population level.

Focus on AlbFlavi2 and AlbFlavi36

Besides its association to high DENV DEs, we further focused on AlbFlavi2 located in a piRNA cluster and

Figure 3. *Aedes albopictus* population clustering based on microsatellite and NIRVS loci. (A) Dendrogram of *Ae. albopictus* populations based on the analysis of 8 microsatellite loci of 12 *Aedes albopictus* populations using Cavalli-Sforza & Edwards's genetic distance and Neighbour-Joining method. Bootstrap values were indicated when >50%. (B) Dendrogram of *Ae. albopictus* populations based on Bray Curtis distance representing dissimilarities between NIRVS composition and abundances.
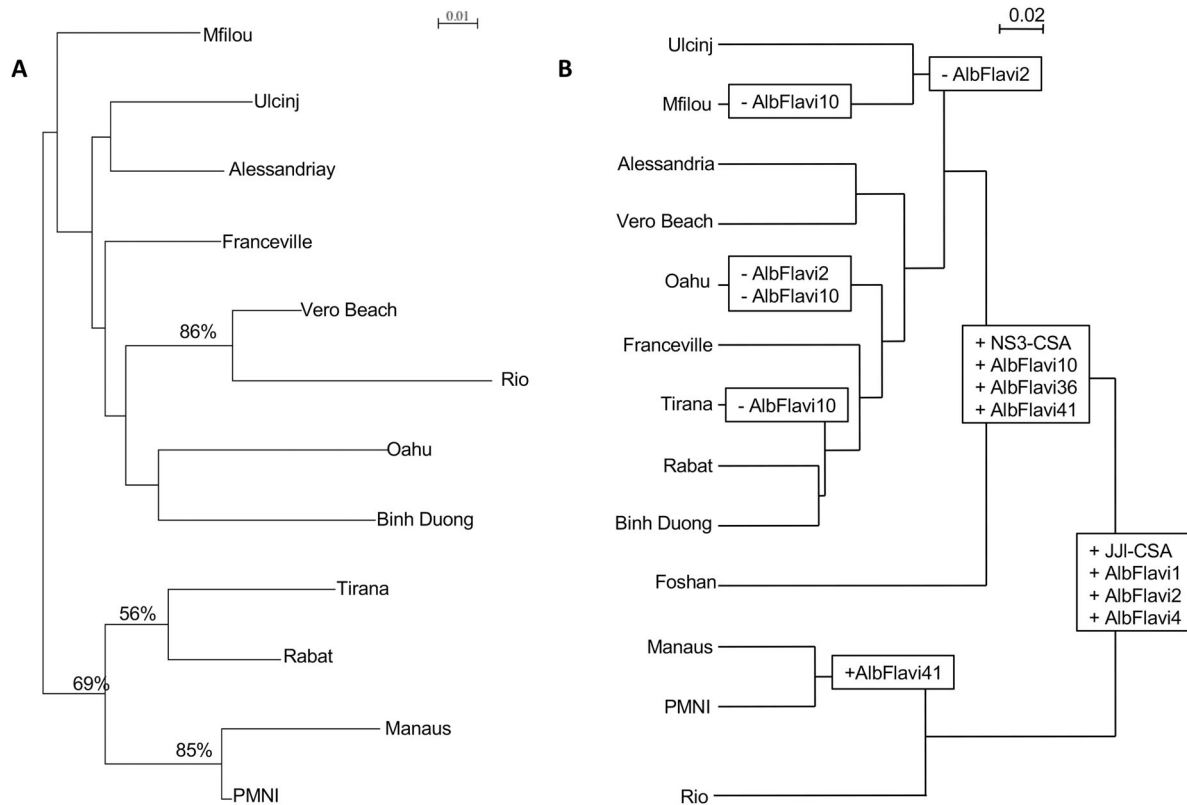
Table 4. Association between NIRVS and arboviral dissemination efficiencies in *Aedes albopictus* populations (logistic regression models).

NIRVS	N	OR (95% CI)	*p*-value
DENV			
AlbFlavi1	707	0.82 (0.61–1.11)	NS
AlbFlavi2		3.07 (2.3–4.2)	***
AlbFlavi4		1.46 (1.0–2.1)	NS
AlbFlavi10		0.46 (0.3–0.6)	***
AlbFlavi36		0.46 (0.3–0.6)	***
AlbFlavi41		0.43 (0.29–0.62)	***
CSA-NS3		0.68 (0.5–0.9)	**
CSA-JJL		3.87 (1.0–14)	*
CHIKV			
AlbFlavi1	360	1.53 (0.9–2.6)	NS
AlbFlavi2		1.3 (0.7–2.1)	NS
AlbFlavi4		2.14 (1.2–3.6)	**
AlbFlavi10		0.60 (0.3–1.0)	NS
AlbFlavi36		0.36 (0.2–0.6)	***
AlbFlavi41		0.61 (0.4–1.0)	NS
CSA-NS3		0.56 (0.3–1.0)	*
CSA-JJL		0.68 (0.3–1.7)	NS

Notes: Dissemination efficiencies (DE) data were assessed from the same or geographically close *Ae. albopictus* populations (see Supplementary Table 2). Populations were characterized as high or low frequencies for each NIRVS by using the median and were analyzed with DE data by logistic regression models to find any association. Odds ratio (OR) >1 and <1 indicated positive and negative association respectively. 95% CI: 95% confidence intervals; N: sample size.

presenting a large geographical distribution encompassing all continents. Moreover, AlbFlavi36 located in an intergenic region was not found in populations from South America and presented an opposite pattern of association to DENV DEs; it was used for

comparison to AlbFlavi2. Therefore, sequence polymorphism, evolution and potential association to vector competence to arboviruses were assessed in *Ae. albopictus* at the individual level.

Sequence polymorphism and evolution of AlbFlavi2 and AlbFlavi36

AlbFlavi2 and AlbFlavi36 of individual mosquitoes were successfully amplified in 10 out of the 13 *Ae. albopictus* populations studied (Figure 2). Sequences of AlbFlavi2 and AlbFlavi36 amplified fragments were aligned with the closely related exogenous viral sequences: the Kamiti River virus (KRV) for AlbFlavi2 and the Aedes Flavivirus (AeFV) for AlFlavi36.

The p-distance calculated between AlbFlavi2 sequences revealed a moderate diversity, with values ranging from 0% to 3.9% (data not shown). All AlbFlavi2 sequences underwent deletion events that were sometimes shared by several individuals from different populations or observed in a specific *Ae. albopictus* population (Figure 4; Supplementary AlbFlavi36).

The p-distance values calculated between AlbFlavi36 sequences were low and varied only from 0% to 0.9% (data not shown). Therefore, AlbFlavi36 sequences mapping to an intergenic region [24] appeared more similar between individuals than AlbFlavi2 sequences (which are integrated in piRNA cluster 27 [53]).
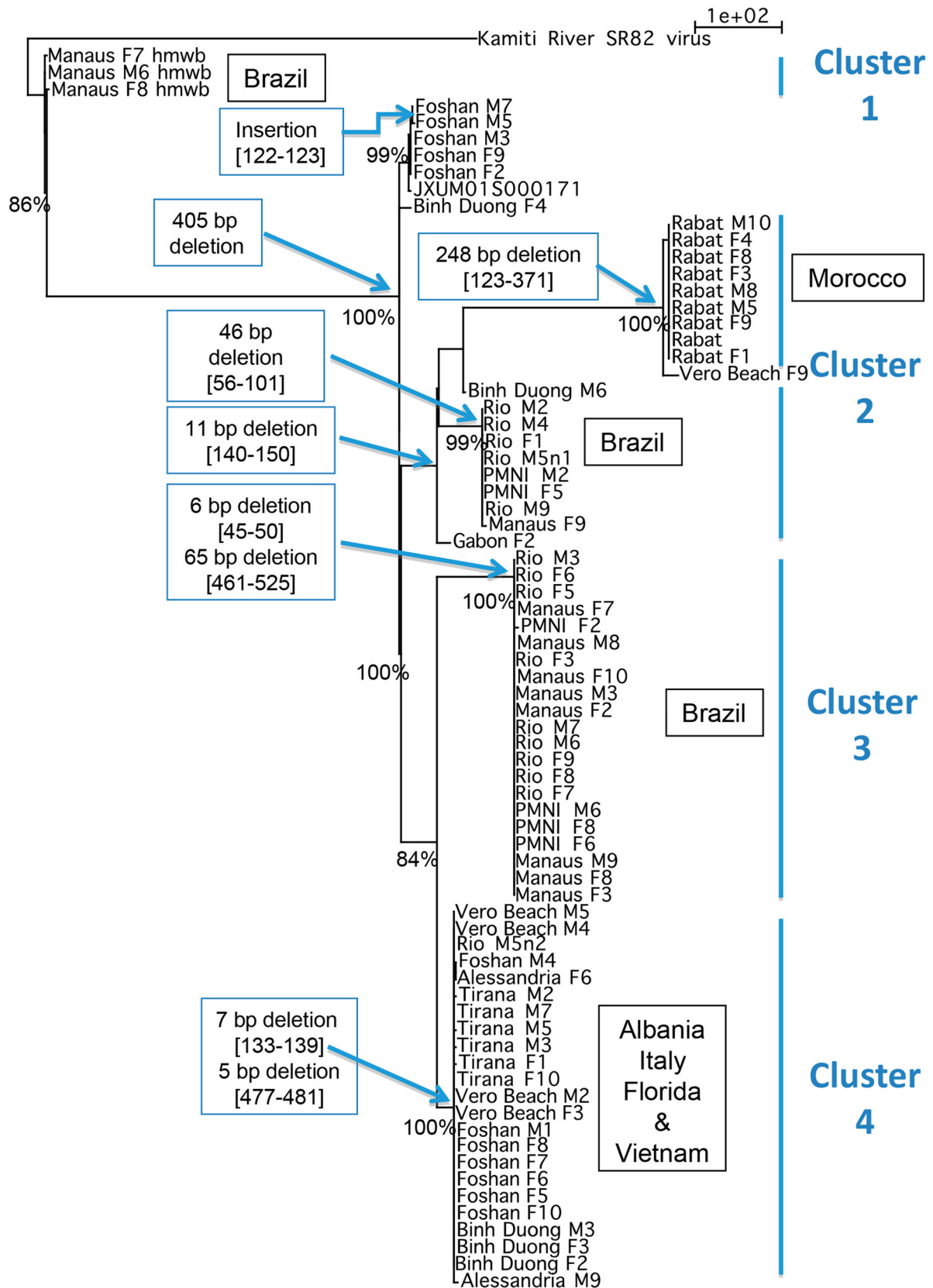
Figure 4. Divergence of AlbFlavi2 among *Aedes albopictus* individuals. Phylogram of AlbFlavi2 sequences based on parsimony with gaps considered as 5th nucleotides. Each node was found in 98–100% of the trees obtained through NNI rearrangements. Significant bootstrap values were indicated at nodes. hmwb: high molecular weight band Values; in brackets: alignment coordinates of deletion. The same result was obtained by parsimony without gap as 5th nucleotide, except for the sequence cluster of mosquitoes from Morocco.

We further performed phylogenetic analysis to describe the evolutionary history of AlbFlavi2 and Alb-Flavi36 (Figure 4; Supplementary Figure 4). As

expected, AlbFlavi2 displayed higher divergence than AlbFlavi36. The resulted trees based on AlbFlavi2 supported four major clusters sharing sequences with the

same deletion events (Figure 4). One of them, defined by a 11 bp deletion was subdivided into two subclusters of sequences sharing respectively a 248 bp and a 46 bp deletion. Overall, these clusters did not show any strict relationship with geographical origin of populations.

AlbFlavi2 from Brazilian populations (PMNI, Manaus and Rio) appeared clearly polyphyletic. Three clusters represented only sequences from Brazil: an older cluster of sequences sharing a 405 bp insertion and two other clusters of sequences sharing respectively a 46 bp and a 6–65 bp deletions. Only AlbFlavi2 from Rabat population formed a monophyletic group, a unique cluster of sequences characterized by a 248 bp deletion.

One cluster contained most of the sequences from Tirana (Albania), Vero Beach (USA), Alessandria (Italy) and Binh Duong (Vietnam). However, some sequences from Binh Duong, Vero Beach and Alessandria populations, were also isolated or found in other clusters. Moreover, half of the AlbFlavi2 sequences from Foshan formed a unique cluster whereas the other half shared the heterogeneous cluster of sequences described above.

The polymorphism of AlbFlavi36 was low in *Ae. albopictus* populations and all the sequences showed close relationships with each other, without any significant bootstrap values (Supplementary Figure 4). However, sequences from African mosquitoes (Congo, Gabon and Morocco) in branching deeply in the tree may be more ancient than sequences of Asian (Binh Duong, Foshan) and European mosquitoes.

Collectively, phylogenetic studies revealed different histories of NIRVS in mosquito populations and particularly, complex history of AlbFlavi2 evolving by both mutation and deletion events.
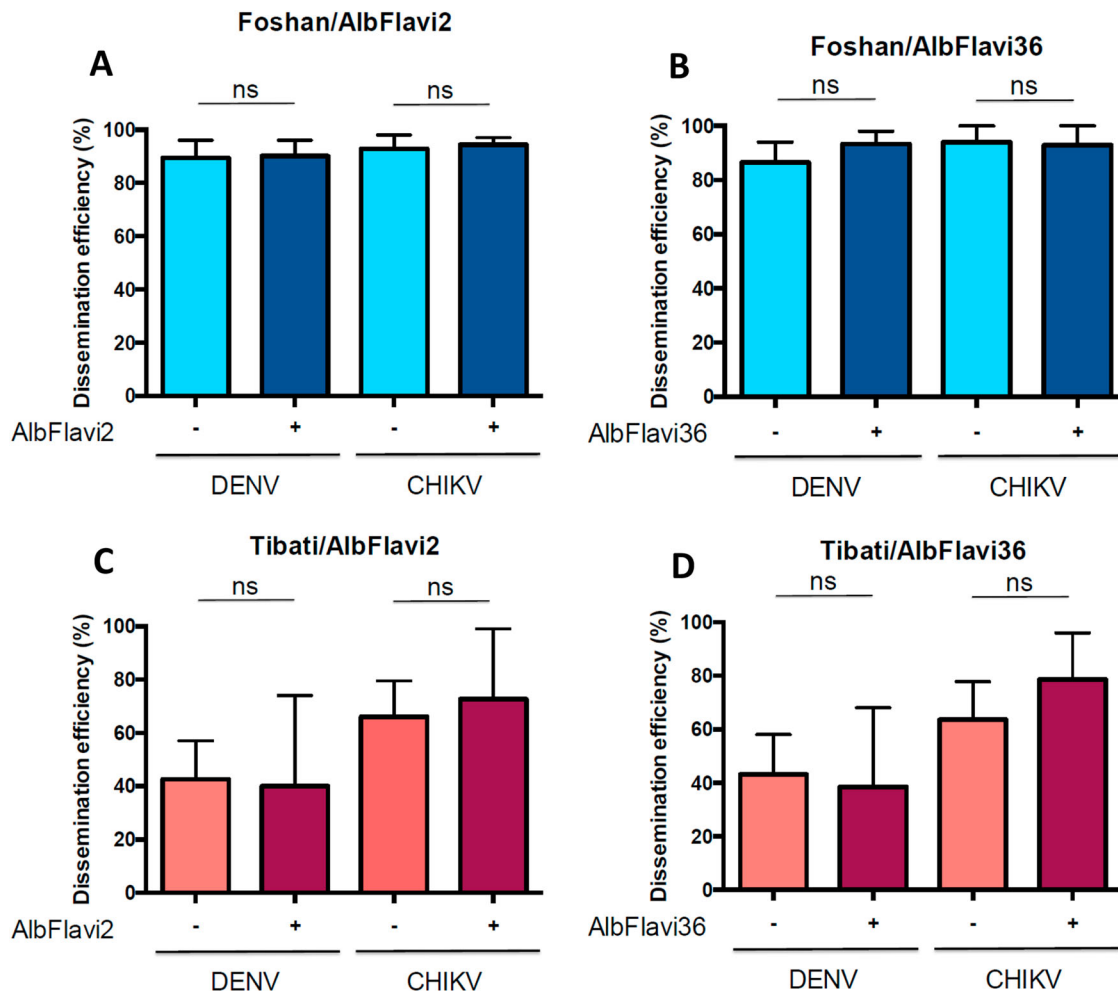
Figure 5. Pilot analysis showing the association between frequencies of AlbFlavi2/AlbFlavi36 and DENV/CHIKV dissemination efficiencies (DE) in *Aedes albopictus* populations. The Foshan colony and the Tibati population (Cameroon, generation F1) were used for the analysis. (A) Presence/absence of AlbFlavi2 and DEs to DENV and CHIKV obtained for the Foshan colony. (B) Presence/absence of AlbFlavi36 and DEs to DENV and CHIKV obtained for the Foshan colony. (C) Presence/absence of AlbFlavi2 and DEs to DENV and CHIKV obtained for the Tibati population. (D) Presence/absence of AlbFlavi36 and DEs to DENV and CHIKV obtained for the Tibati population. DEs were obtained for both viruses at 14 days post-infection. In total, 191 and 122 individuals were examined for presence of AlbFlavi2/AlbFlavi36 after infection DENV and CHIKV, respectively. Interactions of populations and frequencies of AlbFlavi2/AlbFlavi36 with DEs were tested using logistic regression models.

Association between AlbFlavi2/Albflavi36 and vector competence in *Ae. albopictus* at the individual level

To further investigate the biological role of AlbFlavi2 and AlbFlavi36, we extended our analyses of association between the presence of these two NIRVS and viral dissemination from a population level to an individual level. To do so, mosquitoes of the Foshan colony strain and the field-collected Tibati population (Cameroon) were infected with DENV-2 and CHIKV. At 14 days post-infection, mosquito dissemination status and the presence of both AlbFlavi2 and AlbFlavi36 were determined for 313 mosquitoes (Figure 5). When comparing DENV dissemination, Foshan better disseminated than Tibati (respectively 89.8% and 42.2%; *p*-value <10^{-4}; data not shown). However, the presence of AlbFlavi2 or AlbFlavi36 in mosquitoes from Foshan or Tibati was not significantly associated to DENV dissemination (*p*-values > 0.209; Figure 5(A–D)). Moreover, the same results were obtained regarding the association between Alb-Flavi2/AlbFlavi36 presence and CHIKV dissemination (*p*-values > 0.3; Figure 5(A–D)). In all, using both laboratory colony and field-collected mosquitoes, no association was found between any of the two NIRVS (AlbFlavi2 and AlbFlavi36) and DENV/CHIKV dissemination.

Discussion

The presence of sequences with similarities to Flaviviruses in the genome of *Aedes* spp. mosquitoes and their enrichment in piRNA clusters support the hypothesis that viral integrations, at least some, are not simply viral fossils, but could have a biological role [24,25,54]. Here we demonstrated that the NIRVS studied are not neutrally distributed among *Ae. albopictus* populations and all of them (except Alb-Flavi1) were significantly associated to vector competence to DENV and CHIKV.

NIRVS are not neutral markers

NIRVS are endogenous viral sequences located in protein-coding gene exons, intergenic regions, and PIWI-interacting RNA (piRNA) clusters [55]. Contrary to the other categories, piRNA genes are evolving rapidly under positive selection to generate a high diversity of piRNAs [56–63]. Approximately 30 sequences of flaviviral ORFs (primarily NS1 and NS5) were detected in the Foshan colony [24] including the early-detected Flavivirus-like sequences [21,64]. It has been suggested that the variable number and frequency of NIRVS across *Ae. albopictus* populations contribute to variations of genome size among mosquitoes [36]. Our study targeted seven NIRVS (AlbFlavi1, AlbFlavi2, AlbFlavi4, Alb-Flavi10, AlbFlavi36, AlbFlavi41, CSA), for which their

evolution under selection pressures or simply by genetic drift remained unknown.

Contrary to NIRVS deriving from flaviviruses in *Ae. aegypti* (unpublished data), the presence of the tested NIRVS was highly variable among worldwide populations, with the lowest number being detected in mosquitoes from Brazil. This suggests that these NIRVS have not reached fixation in any of the *Ae. albopictus* populations tested (with potential exception for Alb-Flavi1 and a fragment of CSA). Interestingly, Brazilian populations (Manaus, Rio, and PMNI) displayed the lowest number of NIRVS studied with a total absence of AlbFlavi10, AlbFlavi36 and CSA-NS3. Considering the ancient origin of NIRVS, estimated between 6.5 thousands to 2.5 million years ago [65], we hypothesized that they might have been less exposed to ISFs in the past, leading to fewer NIRVS in their genomes or that populations may have acquired NIRVS in the past which have been lost over time.

We showed that the microsatellite polymorphism of populations did not match with the NIRVS abundance profiles. Therefore, because microsatellites are sequences that neutrally evolve in the genome, we speculate that the evolution of NIRVS is far from neutral and NIRVS could provide benefits to the host. NIRVS may have been produced for specific purposes in the host rather than being the consequences of random endogenization of exogenous viral fragments.

The focus on NIRVS from piRNA cluster (AlbFlavi2) and intergenic region (AlbFlavi36) revealed different sequence polymorphism and evolutionary histories despite their relatively wide distribution among *Ae. albopictus* populations. Whereas AlbFlavi36 appeared monophyletic with highly conserved sequences among populations, the phylogenetic analyses showed a particularly complex history for Albflavi2 evolving by both mutation and deletion events. We postulate that owing to its high variability, AlbFlavi2 may act similarly to TE fragments and piRNA genes, and evolve according to exposures to related exogenous virus burden.

Association between NIRVS and vector competence to arboviruses

Similar to some endogenous retroviruses (ERVs) that have an effect on viruses from different subfamilies and genera [66], we questioned whether NIRVS deriving from ISF sequences affect the dissemination of different arboviruses and contribute to the regulation of vector competence of *Ae. albopictus* mosquitoes. Because NIRVS located in piRNA clusters, such as Alb-Flavi2, produce piRNAs, are differentially distributed among populations and appeared older integrations than NIRVS located in codons [65], they were proposed to function as novel mosquito antiviral immune factors. However, contradictory results were obtained using both population- and individual-level analysis.

At the population level, many NIRVS were present in *Ae. albopictus* populations. The frequency of some of them appears correlated (positively or negatively) to DENV dissemination. Whereas some NIRVS were associated with high viral dissemination (i.e. AlbFlavi2 and CSA-JJL), others were significantly related to low viral dissemination (AlbFlavi10, AlbFlavi36, AlbFlavi41 and CSA-NS3), suggesting different functions of NIRVS deriving from ISFs. Moreover, fewer NIRVS were positively (i.e. AlbFlavi4) and negatively (i.e. AlbFlavi36 and CSA-NS3) correlated to CHIKV dissemination. These results are consistent with the contradictory results obtained in studies assessing the impact of exogenous ISFs on arbovirus fitness in mosquitoes. Indeed, whereas some ISFs have shown to repress arboviral replication [67], others have been proved to facilitate infection [68]. Further experiments should be performed to confirm these results, as only 20 mosquitoes per population were tested for NIRVS distribution.

At the level of individuals, our pilot experiment based on mosquitoes from both laboratory colony and field did not show any significant association between dissemination of DENV and CHIKV, and NIRVS, AlbFlavi2 or AlbFlavi36. while this does not seem surprising for the alphavirus CHIKV considering that NIRVS examined are homologous to flavivirus sequences, it is still questionable for DENV. Since it appears that they do not evolve neutrally and were preserved from purifying selection, NIRVS such as CSA has demonstrated to produce a transcript of 4671 nt long [21,69]; its functional role should be investigated. Lastly, other components should be considered such as the virome [70] in addition to the anatomical barriers in the mosquito such as the salivary glands for viral transmission [19].

To conclude, our results clearly show that NIRVS in *Ae. albopictus* follow processes different from that of neutral genes such as microsatellites and most NIRVS are far from reaching fixation. Flaviviral integrations are differentially distributed among *Ae. albopictus* populations and are here suggested to be associated with the vector competence to arboviruses by mechanisms that remain to be elucidated. Finally, this study opens the way to new perspectives on evolution and biological functions of NIRVS, in part on vector competence.

Acknowledgements

Author contributions

ABF and XDL designed the experiments; VH performed the research; GG did the microsatellite genotyping; CD did the phylogenetic reconstruction; MS and MM did the analysis of genetic structure; YD did the statistical analysis; LM and PSY provided a technical help; MB edited the paper; VH, CD, ABF wrote the paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

ORCID

Michele Marconcini http://orcid.org/0000-0002-6785-9106
Mariangela Bonizzoni http://orcid.org/0000-0003-0568-8564
Anna-Bella Failloux http://orcid.org/0000-0001-6890-0820

References

[1] Paupy C, Delatte H, Bagny L, et al. *Aedes albopictus*, an arbovirus vector: from the darkness to the light. Microbes Infect. 2009;11:1177–1185. doi:10.1016/j.micinf.2009.05.005.

[2] Bonizzoni M, Gasperi G, Chen X, et al. The invasive mosquito species *Aedes albopictus*: current knowledge and future perspectives. Trends Parasitol. 2013;29:460–468. doi:10.1016/j.pt.2013.07.003.

[3] Caminade C, Medlock JM, Ducheyne E, et al. Suitability of European climate for the Asian tiger mosquito *Aedes albopictus*: recent trends and future scenarios. J R Soc Interface. 2012;9:2708–2717. doi:10.1098/rsif.2012.0138.

[4] Benedict MQ, Levine RS, Hawley WA, et al. Spread of the tiger: global risk of invasion by the mosquito *Aedes albopictus*. Vector Borne Zoonotic Dis. 2007;7:76–85. doi:10.1089/vbz.2006.0562.

[5] Lounibos LP. Invasions by insect vectors of human disease. Annu Rev Entomol. 2002;47:233–266. doi:10.1146/annurev.ento.47.091201.145206.

[6] Hawley WA. The biology of *Aedes albopictus*. J Am Mosq Control Assoc Suppl. 1988;1:1–39.

[7] Adhami J, Reiter P. Introduction and establishment of Aedes (Stegomyia) albopictus skuse (Diptera: Culicidae) in Albania. J Am Mosq Control Assoc. 1998;14:340–343.

[8] Sprenger D, Wuithiranyagool T. The discovery and distribution of *Aedes albopictus* in Harris County, Texas. J Am Mosq Control Assoc. 1986;2:217–219.

[9] Consoli RAGB OR. Principais mosquitos de importância sanitária no Brasil. Rio de Janeiro. *Editora FIOCRUZ*. 1994;228.

[10] Cornel AJ, Hunt RH. *Aedes albopictus* in Africa? First records of live specimens in imported tires in Cape Town. J Am Mosq Control Assoc. 1991;7:107–108.

[11] Goubert C, Minard G, Vieira C, et al. Population genetics of the Asian tiger mosquito *Aedes albopictus*, an invasive vector of human diseases. Heredity (Edinb). 2016;117:125–134. doi:10.1038/hdy.2016.35.

[12] Jackson H, Strubbe D, Tollington S, et al. Ancestral origins and invasion pathways in a globally invasive bird correlate with climate and influences from bird trade. Mol Ecol. 2015;24:4269–4285. doi:10.1111/mec.13307.

[13] Powell JR, Tabachnick WJ. History of domestication and spread of Aedes aegypti–a review. Mem Inst Oswaldo Cruz. 2013;108(Suppl 1):11–17. doi:10.1590/0074-0276130395.

[14] Maynard AJ, Ambrose L, Cooper RD, et al. Tiger on the prowl: invasion history and spatio-temporal genetic structure of the Asian tiger mosquito *Aedes albopictus* (Skuse 1894) in the Indo-Pacific. PLoS Negl Trop Dis. 2017;11:e0005546. doi:10.1371/journal.pntd.0005546.

[15] Kotsakiozi P, Richardson JB., Pichler V, et al. Population genomics of the Asian tiger mosquito, *Aedes albopictus*: insights into the recent worldwide invasion. Ecol Evol. 2017;7:10143–10157. doi:10.1002/ece3.3514.

[16] Manni M, Gomulski LM, Aketarawong N, et al. Molecular markers for analyses of intraspecific genetic diversity in the Asian tiger mosquito, *Aedes albopictus*. Parasit Vectors. 2015;8:188. doi:10.1186/s13071-015-0794-5.

[17] Manni M, Guglielmino CR, Scolari F, et al. Genetic evidence for a worldwide chaotic dispersion pattern of the arbovirus vector, *Aedes albopictus*. PLoS Negl Trop Dis. 2017;11:e0005332. doi:10.1371/journal.pntd.0005332.

[18] Mousson L, Dauga C, Garrigues T, et al. Phylogeography of *Aedes* (Stegomyia) *aegypti* (L.) and *Aedes* (Stegomyia) *albopictus* (Skuse) (Diptera: Culicidae) based on mitochondrial DNA variations. Genet Res. 2005;86:1–11. doi:10.1017/S0016672305007627.

[19] Kramer LD, Ebel GD. Advances in virus research. Adv Virus Res. 2003;60:187–232.

[20] Zouache K, Fontaine A, Vega-Rua A, et al. Three-way interactions between mosquito population, viral strain and temperature underlying chikungunya virus transmission potential. Proc Biol Sci. 2014;281. doi:10.1098/rspb.2014.1078.

[21] Crochu S, et al. Sequences of flavivirus-related RNA viruses persist in DNA form integrated in the genome of *Aedes* spp. mosquitoes. J Gen Virol. 2004;85:1971–1980. doi:10.1099/vir.0.79850-0.

[22] Katzourakis A, Gifford RJ. Endogenous viral elements in animal genomes. PLoS Genet. 2010;6:e1001191. doi:10.1371/journal.pgen.1001191.

[23] Fort P, Albertini A, Van-Hua A, et al. Fossil rhabdoviral sequences integrated into arthropod genomes: ontogeny, evolution, and potential functionality. Mol Biol Evol. 2012;29:381–390. doi:10.1093/molbev/msr226.

[24] Palatini U, Miesen P, Carballar-Lejarazu R, et al. Comparative genomics shows that viral integrations are abundant and express piRNAs in the arboviral vectors *Aedes aegypti* and *Aedes albopictus*. BMC Genomics. 2017;18:512. doi:10.1186/s12864-017-3903-3.

[25] Whitfield ZJ, Dolan PT, Kunitomi M, et al. The diversity, structure, and function of heritable adaptive immunity sequences in the *Aedes aegypti* genome. Curr Biol. 2017;27:3511–3519. doi:10.1016/j.cub.2017.09.067. e3517.

[26] Brennecke J, Aravin AA, Stark A, et al. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. Cell. 2007;128:1089–1103. doi:10.1016/j.cell.2007.01.043.

[27] Pelisson A, Sarot E, Payen-Groschene G, et al. A novel repeat-associated small interfering RNA-mediated silencing pathway downregulates complementary sense gypsy transcripts in somatic cells of the *Drosophila* ovary. J Virol. 2007;81:1951–1960. doi:10.1128/JVI.01980-06.

[28] Saito K, Nishida KM, Mori T, et al. Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. Genes Dev. 2006;20:2214–2222. doi:10.1101/gad.1454806.

[29] Vagin VV, Sigova A, Li C, et al. A distinct small RNA pathway silences selfish genetic elements in the germline. Science. 2006;313:320–324. doi:10.1126/science.1129333.

[30] Zanni V, Eymery A, Coiffet M, 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;110:19842–19847. doi:10.1073/pnas.1313677110.

[31] Frank JA, Feschotte C. Co-option of endogenous viral sequences for host cell function. Curr Opin Virol. 2017;25:81–89. doi:10.1016/j.coviro.2017.07.021.

[32] Aswad A, Katzourakis A. Paleovirology and virally derived immunity. Trends Ecol Evol. 2012;27:627–636. doi:10.1016/j.tree.2012.07.007.

[33] Katzourakis A. Paleovirology: inferring viral evolution from host genome sequence data. Philos Trans R Soc Lond B Biol Sci. 2013;368:20120493. doi:10.1098/rstb.2012.0493.

[34] Musso D, Rodriguez-Morales AJ, Levi JE, et al. Unexpected outbreaks of arbovirus infections: lessons learned from the Pacific and tropical America. Lancet Infect Dis. 2018;18:e355–e361. doi:10.1016/S1473-3099(18)30269-X.

[35] Fritzell C, Rousset D, Adde A, et al. Current challenges and implications for dengue, chikungunya and Zika seroprevalence studies worldwide: A scoping review. PLoS Negl Trop Dis. 2018;12:e0006533. doi:10.1371/journal.pntd.0006533.

[36] Chen XG, Jiang X, Gu J, 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:E5907–E5915. doi:10.1073/pnas.1516410112.

[37] GENETIX 4.05 Logiciel Sous Windows TM Pour la Génétique des Populations. 1996–2004.

[38] FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). 2001.

[39] Excoffier L, Lischer HE. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour. 2010;10:564–567. doi:10.1111/j.1755-0998.2010.02847.x.

[40] Peakall R, Smouse PE. Genalex 6.5: genetic analysis in Excel. Population genetic software for teaching and research–an update. Bioinformatics. 2012;28:2537–2539. doi:10.1093/bioinformatics/bts460.

[41] Vazeille M, Zouache K, Vega-Rúa A, et al. Importance of mosquito "quasispecies" in selecting an epidemic arthropod-borne virus. Sci Rep. 2016;6:29564. doi:10.1038/srep29564.

[42] Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155:945–959.

[43] Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. 2005;14:2611–2620. doi:10.1111/j.1365-294X.2005.02553.x.

[44] Rosenberg N. DISTRUCT: a program for the graphical display of population structure. Mol Ecol Notes. 2004;4:137–138. doi:10.1046/j.1471-8286.2003.00566.x.

[45] Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol. 2000;132:365–386.

[46] Ricotta C, Podaní J. On some properties of the Bray-Curtis dissimilarity and their ecological meaning. Ecol Complex. 2017;31:201–205.

[47] Vega-Rua A, Zouache K, Girod R, et al. High level of vector competence of Aedes aegypti and Aedes albopictus from ten American countries as a crucial factor in the spread of chikungunya virus. J Virol. 2014;88:6294–6306. doi:10.1128/JVI.00370-14.

[48] Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 2004;5:113. doi:10.1186/1471-2105-5-113.

[49] PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sunderland (MA), 1998.

[50] Vazeille-Falcoz M, Mousson L, Rodhain F, et al. Variation in oral susceptibility to dengue type 2 virus of populations of Aedes aegypti from the islands of Tahiti and Moorea, French Polynesia. Am J Trop Med Hyg. 1999;60:292–299.

[51] Schuffenecker I, Iteman I, Michault A, et al. Genome microevolution of chikungunya viruses causing the Indian Ocean outbreak. PLoS Med. 2006;3:e263. doi:10.1371/journal.pmed.0030263.

[52] Amraoui F, Ben Ayed W, Madec Y, et al. Potential of Aedes albopictus to cause the emergence of arboviruses in Morocco. PLoS Negl Trop Dis. 2019;13:e0006997. doi:10.1371/journal.pntd.0006997.

[53] Liu P, Dong Y, Gu J, et al. Developmental piRNA profiles of the invasive vector mosquito Aedes albopictus. Parasit Vectors. 2016;9:524. doi:10.1186/s13071-016-1815-8.

[54] Olson KE, Bonizzoni M. Nonretroviral integrated RNA viruses in arthropod vectors: an occasional event or something more? Curr Opin Insect Sci. 2017;22:45–53. doi:10.1016/j.cois.2017.05.010.

[55] Houe V, Bonizzoni M, Failloux AB. Endogenous non-retroviral elements in genomes of Aedes mosquitoes and vector competence. Emerg Microbes Infect. 2019;8:542–555. doi:10.1080/22221751.2019.1599302.

[56] Lee YC, Langley CH. Long-term and short-term evolutionary impacts of transposable elements on Drosophila. Genetics. 2012;192:1411–1432. doi:10.1534/genetics.112.145714.

[57] Vermaak D, Henikoff S, Malik HS. Positive selection drives the evolution of rhino, a member of the heterochromatin protein 1 family in Drosophila. PLoS Genet. 2005;1:96–108. doi:10.1371/journal.pgen.0010009.

[58] Lewis SH, Quarles KA, Yang Y, et al. Pan-arthropod analysis reveals somatic piRNAs as an ancestral defence against transposable elements. Nat Ecol Evol. 2018;2:174–181. doi:10.1038/s41559-017-0403-4.

[59] Heger A, Ponting CP. Evolutionary rate analyses of orthologs and paralogs from 12 Drosophila genomes. Genome Res. 2007;17:1837–1849. doi:10.1101/gr.6249707.

[60] Obbard DJ, Gordon KH, Buck AH, et al. The evolution of RNAi as a defence against viruses and transposable elements. Philos Trans R Soc Lond B Biol Sci. 2009;364:99–115. doi:10.1098/rstb.2008.0168.

[61] Kolaczkowski B, Hupalo DN, Kern AD. Recurrent adaptation in RNA interference genes across the Drosophila phylogeny. Mol Biol Evol. 2011;28:1033–1042. doi:10.1093/molbev/msq284.

[62] Yi M, Chen F, Luo M, et al. Rapid evolution of piRNA pathway in the teleost fish: implication for an adaptation to transposon diversity. Genome Biol Evol. 2014;6:1393–1407. doi:10.1093/gbe/evu105.

[63] Simkin A, Wong A, Poh YP, et al. Recurrent and recent selective sweeps in the piRNA pathway. Evolution. 2013;67:1081–1090. doi:10.1111/evo.12011.

[64] Roiz D, Vazquez A, Seco MP, et al. Detection of novel insect flavivirus sequences integrated in Aedes albopictus (Diptera: Culicidae) in Northern Italy. Virol J. 2009;6:93. doi:10.1186/1743-422X-6-93.

[65] Pischedda E, Scolari F, Valerio F, et al. Insights into an unexplored component of the mosquito repeatome: distribution and variability of viral sequences integrated into the genome of the arboviral vector Aedes albopictus. Front Genet. 2019;10:93. doi:10.3389/fgene.2019.00093.

[66] Yap MW, Colbeck E, Ellis SA, et al. Evolution of the retroviral restriction gene Fv1: inhibition of non-MLV retroviruses. PLoS Pathog. 2014;10:e1003968. doi:10.1371/journal.ppat.1003968.

[67] Hobson-Peters J, Yam AWY, Lu JWF, et al. A new insect-specific flavivirus from northern Australia suppresses replication of West Nile virus and Murray Valley encephalitis virus in co-infected mosquito cells. PLoS One. 2013;8:e56534. doi:10.1371/journal.pone.0056534.

[68] Kuwata R, Isawa H, Hoshino K, et al. Analysis of mosquito-borne flavivirus superinfection in Culex tritaeniorhynchus (Diptera: Culicidae) cells persistently infected with Culex Flavivirus (Flaviviridae). J Med Entomol. 2015;52:222–229. doi:10.1093/jme/tju059.

[69] Suzuki Y, Frangeul L, Dickson LB, et al. Uncovering the repertoire of endogenous flaviviral elements in Aedes mosquito genomes. J Virol. 2017;91. doi:10.1128/JVI.00571-17.

[70] Bolling BG, Weaver SC, Tesh RB, et al. Insect-specific virus discovery: significance for the arbovirus community. Viruses. 2015;7:4911–4928. doi:10.3390/v7092851.

![UNIVERSITÀ DI PAVIA]

MICHELE MARCONCINI <michele.marconcini01@universitadipavia.it>

Notification of Formal Acceptance of "Polymorphism analyses and protein modelling inform on functional specialization of Piwi clade genes in the arboviral vector Aedes albopictus" (PNTD-D-19-01060R2) - [EMID:d16303ed06d1f59b]

PLOS Neglected Tropical Diseases <em@editorialmanager.com>　　　22 novembre 2019 14:18
Rispondi a: PLOS Neglected Tropical Diseases <plosntds@plos.org>
A: Michele Marconcini <michele.marconcini01@ateneopv.it>

You are being carbon copied ("cc:'d") on an e-mail "To" "Mariangela Bonizzoni" mariangela.bonizzoni@unipv.it; m.bonizzoni@unipv.it
CC: "Giuseppe Iovino" iovinogiuseppe91@gmail.com, "Luis Hernandez" luishpbachiller@gmail.com, "Vincent Houé" vincent.houe@pasteur.fr, "Federica Valerio" federica.valerio02@universitadipavia.it, "Umberto Palatini" umberto.palatini01@universitadipavia.it, "Elisa Pischedda" elisa.pischedda01@universitadipavia.it, "Jacob Crawford" jacobcrawford@verily.com, "Bradley J. White" bradwhite@verily.com, "Teresa Lin" linteresa@verily.com, "Rebeca Carballar-Lejarazu" rebecacarballar@gmail.com, "Anna-Bella Failloux" anna-bella.failloux@pasteur.fr, "Federico Forneris" federico.forneris@unipv.it, "Lino Ometto" lino.ometto@unipv.it, "Michele Marconcini" michele.marconcini01@ateneopv.it

Dear Dr. Bonizzoni,

We are delighted to inform you that your manuscript, "Polymorphism analyses and protein modelling inform on functional specialization of Piwi clade genes in the arboviral vector Aedes albopictus," has been formally accepted for publication in PLOS Neglected Tropical Diseases.

We have now passed your article onto the PLOS Production Department who will complete the rest of the publication process. All authors will receive a confirmation email upon publication.

The corresponding author will soon be receiving a typeset proof for review, to ensure errors have not been introduced during production. Please review the PDF proof of your manuscript carefully, as this is the last chance to correct any scientific or type-setting errors. Please note that major changes, or those which affect the scientific understanding of the work, will likely cause delays to the publication date of your manuscript. Note: Proofs for Front Matter articles (Editorial, Viewpoint, Symposium, Review, etc...) are generated on a different schedule and may not be made available as quickly.

Soon after your final files are uploaded, the early version of your manuscript will be published online unless you opted out of this process. The date of the early version will be your article's publication date. The final article will be published to the same URL, and all versions of the paper will be accessible to readers.

Thank you again for supporting open-access publishing; we are looking forward to publishing your work in PLOS Neglected Tropical Diseases.

Best regards,

Serap Aksoy
Editor-in-Chief
PLOS Neglected Tropical Diseases

Shaden Kamhawi
Editor-in-Chief
PLOS Neglected Tropical Diseases

_____
In compliance with data protection regulations, you may request that we remove your personal registration details at any time. (Use the following URL: https://www.editorialmanager.com/pntd/login.asp?a=r). Please contact the publication office if you have any questions.

1 **Polymorphism analyses and protein modelling inform on functional specialization of *Piwi***
2 **clade genes in the arboviral vector *Aedes albopictus***
3

4 Short Title: Piwi clade genes of *Aedes albopictus*
5

6 Michele Marconcini[1], Luis Hernandez[1], Giuseppe Iovino[1], Vincent Houé[2], Federica Valerio[1],
7 Umberto Palatini[1], Elisa Pischedda[1], Jacob E. Crawford[3], Bradley J. White[3], Teresa Lin[3], Rebeca
8 Carballar-Lejarazu[1]*, Lino Ometto[1], Federico Forneris[1], Anna-Bella Failloux[2], Mariangela
9 Bonizzoni[1&]
10

11 [1]Department of Biology and Biotechnology, University of Pavia, Italy
12 [2]Arbovirus and Insect Vectors Units, Department of Virology, Institut Pasteur, Paris, France
13 [3] Verily Life Sciences LLC, San Francisco, USA
14 *Current Address: University of California, Irvine, Department of Molecular Biology and
15 Biochemistry, California, USA
16 [&]Corresponding author
17

18 **Keywords**

19 *Aedes albopictus*, piRNA pathway, arbovirus, development
20

ABSTRACT

21

22   Current knowledge of the piRNA pathway is based mainly on studies on *Drosophila*

23 *melanogaster* where three proteins of the Piwi subclade of the Argonaute family interact with PIWI-

24 interacting RNAs to silence transposable elements in gonadal tissues. In mosquito species that

25 transmit epidemic arboviruses such as dengue and chikungunya viruses, *Piwi* clade genes

26 underwent expansion, are also expressed in the soma and cross-talk with proteins of recognized

27 antiviral function cannot be excluded for some Piwi proteins. These observations underscore the

28 importance of expanding our knowledge of the piRNA pathway beyond the model organism *D.*

29 *melanogaster*.

30   Here we focus on the emerging arboviral vector *Aedes albopictus* and we couple traditional

31 approaches of expression and adaptive evolution analyses with most current computational

32 predictions of protein structure to study evolutionary divergence among Piwi clade proteins.

33 Superposition of protein homology models indicate possible high structure similarity among all Piwi

34 proteins, with high levels of amino acid conservation in the inner regions devoted to RNA binding.

35 On the contrary, solvent-exposed surfaces showed low conservation, with several sites under

36 positive selection. Analysis of the expression profiles of *Piwi* transcripts during mosquito

37 development and following infection with Dengue serotype 1 or Chikungunya viruses showed a

38 concerted elicitation of all *Piwi* transcripts during viral dissemination of dengue viruses while

39 maintenance of infection relied on expression of primarily *Piwi5*. Opposite, establishment of

40 persistent infection by Chikungunya virus is accompanied by increased expression of all *Piwi*

41 genes, particularly *Piwi4* and, again, *Piwi5*. Overall these results are consistent with functional

42 specialization and a general antiviral role for *Piwi5*. Experimental evidences of sites under positive

43 selection in *Piwi1-3*, *Piwi4* and *Piwi6*, that have complex expression profiles, provide useful

44 knowledge to design tailored functional experiments.

45

2

AUTHOR SUMMARY

Argonautes are ancient proteins involved in many cellular processes, including innate

immunity. Early in eukaryote evolution, Argonautes separated into Ago and Piwi clades, which

maintain a dynamic evolutionary history with frequent duplications and losses. The use *of Drosophila*

*melanogaste*r as a model organism proved fundamental to understand the function of Argonautes.

However, recent studies showed that the patterns and observations made in *D. melanogaster*,

including the number of Argonautes, their expression profile and their function, are a rarity among

Dipterans.

In vectors of epidemic arboviruses such as Dengue and Chikungunya viruses, *Piwi* genes

underwent expansion, are expressed in the soma, and some of them appear to have antiviral

functions. Besides being an important basic question, the identification of which (and how) *Piwi*

genes have antiviral functions may be used for the development of novel genetic-based strategies

of vector control. Here we coupled population genetics models with computational predictions of

protein structure and expression analyses to investigate the evolution and function of *Piwi* genes of

the emerging vector *Aedes albopictus*. Our data support a general antiviral role for *Piwi*5. Instead,

the detection of complex expression profiles with the presence of sites under positive selection in

*Piwi*1/3, *Piwi*4 and *Piwi*6 requires tailored functional experiments to clarify their antiviral role.

INTRODUCTION

First discovered for their role in plant developmental, proteins of the Argonaute family were found in all domains of life, where they are essential for a wide variety of cellular processes, including innate immunity [1,2].

Recent studies provided evidence of evolutionary expansion and functional divergence of Argonautes in Dipterans, including examples in both the Ago and Piwi subclades [3]. Differences in function and copy number have also been found in other taxa such as nematodes [4], oomycetes [5] and higher plants [6], showing that this protein family is subject to a dynamic evolutionary history. In eukaryotes, Argonautes are key components of RNA interference (RNAi) mechanisms, which can be distinguished in three main pathways: the small interfering RNA (siRNA), microRNA (miRNA) and the PIWI-interacting RNA (piRNA) pathways.

The siRNA pathway is the cornerstone of antiviral defense in insects. The canonical activity of this pathway is the Argonaute 2 (Ago2)-dependent cleavage of viral target sequences. Ago2 is guided to its target through an RNA-induced silencing complex (RISC) loaded with 21-nucleotide (nt)-long siRNAs. siRNAs are produced from viral double-strand RNAs intermediates by the RNAase-III endonuclease activity of Dicer-2 (Dcr2) and define the target based on sequence complementarity [7]. Dcr2 also possesses a DExD/H helicase domain that mediates the synthesis of viral DNA (vDNA) fragments [8]. vDNAs appear to further modulate antiviral immunity [8]. vDNA fragments are synthesized in both circular and linear forms, in complex arrangements with sequences from retrotransposons, but details of their mode of action have not been elucidated yet [8,9]. We and others recently showed that the genomes of *Aedes spp.* mosquitoes harbor fragmented viral sequences, which are integrated next to transposon sequences, are enriched in piRNA clusters and produced PIWI-interacting RNAs (piRNAs) [10,11]. The similar organization between vDNAs and viral integrations, along with the production of piRNAs of viral origin (vpiRNAs) following arboviral infection of *Aedes* spp. mosquitoes, led to the hypothesis that the piRNA pathway function cooperatively with the siRNA pathway in the acquisition of tolerance to infection [10,12,13].

4

91   Current knowledge on the piRNA pathway in insects is based mainly on studies on

92   *Drosophila melanogaster* where three proteins of the Piwi subclade, namely Argonaute-3 (AGO3),

93   PIWI and Aubergine (AUB), interact with piRNAs to silence transposable elements (TEs) in

94   gonadal tissues [14]. Interestingly, the piRNA pathway of *D. melanogaster* does not have antiviral

95   activity and no viral integrations have been detected [15]. Additional differences exist between the

96   piRNA pathway of *D. melanogaster* and that of mosquitoes, suggesting that *D. melanogaster*

97   cannot be used as a model to unravel the molecular cross-talk between the siRNA and piRNA

98   pathways leading to antiviral immunity in *Aedes* spp. mosquitoes. For instance, in *Aedes aegypti,*

99   Piwi subclade has undergone expansion with seven proteins (i.e. Ago3, Piwi2, Piwi3, Piwi4, Piwi5,

100   Piwi6 and Piwi7), which are alternatively expressed in somatic and germline cells and interact with

101   both endogenous and vpiRNAs [12,16,17]. Gonadal- or embryonic-specific expression is found for

102   *Piwi1-3* and *Piwi7*, respectively [16]. On the contrary, *Ago3*, *Piwi4*, *Piwi5* and *Piwi6* are highly

103   expressed in *Ae. aegypti* soma and Aag2 cells and all contribute to the production of transposon-

104   derived piRNAs [16,18]. Ago3 and Piwi5 also regulate biogenesis of piRNAs from the replication-

105   dependent histone gene family [19]. Production of vpiRNAs is dependent on Piwi5 and Ago3

106   during infection of Aag2 cells with the *Alphavirus* CHIKV, Sindbis and Semliki Forest (SF) viruses,

107   but relies also on Piwi6 following infection with the *Flavivirus* DENV2 [18,20–22]. Piwi4 does not

108   bind piRNAs and its knock-down does not alter vpiRNA production upon infection of Aag2 cells

109   with either SFV or DENV2 [18,23]. On the contrary Piwi4 coimmunoprecipate with Ago2, Dcr2,

110   Piwi5, Piwi6 and Ago3, suggesting a bridging role between the siRNA and piRNA pathways [21].

111   These studies support an antiviral role for Piwi proteins in *Aedes* spp. mosquitoes but given the

112   number of *Piwi* genes in these species*,* it is a challenge to uncover their distinct physiological roles,

113   if any. In duplicated genes, the presence of sites under positive selection is usually a sign of the

114   acquisition of novel functions [24]. Additionally, under the "arm-race theory", rapid intraspecific

115   evolution is expected for genes with immunity functions because their products should act against

116   fast evolving viruses [25].

117   Besides being an important basic question, the understanding of functional divergence

118   among Piwi proteins has applied perspectives for the development of novel genetic-based

119   methods of transmission-blocking vector control strategies.

120   In recent years, the Asian tiger mosquito *Aedes albopictus* has emerged as a novel global

121   arboviral threat. This species is a competent vector for a number of arboviruses, such as

122   chikungunya (CHIKV), dengue (DENV), yellow fever (YFV) and Zika (ZIKV) viruses and is now

123   present in every continent except Antarctica following its quick spread out of its native home range

124   of South East Asia [26]. Establishment of *Ae. albopictus* in temperate regions of the world fostered

125   the re-emergence or the new introduction of arboviruses [27]. For instance, Chikungunya

126   outbreaks occurred in Italy in 2007 and 2017 [28,29]; France and Croatia suffered from

127   autochthonous cases of Dengue and Chikungunya in several occasions since 2010 [30–33] and

128   dengue is remerging in some regions of the United States due to the presence of *Ae. albopictus*

129   [34]. Knowledge on *Ae. albopictus* biology and the molecular mechanisms underlying its

130   competence to arboviruses are still limited in comparison to *Ae. aegypti* despite its increasing

131   public-health relevance.

132   Here we elucidate the molecular organization, polymorphism and expression of *Piwi* clade

133   genes of *Ae. albopictus* in an evolutionary framework using a combination of molecular, population

134   genomics and computational protein modelling approaches. We show that the genome of *Ae.*

135   *albopictus* harbours seven *Piwi* genes, namely *Ago3*, *Piwi1-3*, *Piwi2*, *Piwi4*, *Piwi5*, *Piwi6* and *Piwi7*.

136   For the first time in mosquitoes, we show sign of adaptive evolution in *Piwi1-3*, *Piwi4*, *Piwi5* and

137   *Piwi6*, including sites in the MID and PAZ domains. Additionally, expression profiles during

138   mosquito development and following infection with the Dengue or Chikungunya viruses support

139   functional specialization of Piwi proteins, with a prominent and general antiviral role for the

140   transcript of *Piwi5*.

141

142   **RESULTS**

143   **Seven *Piwi* genes are present in the genome of *Ae. albopictus***

144   Bioinformatic analyses of the current genome assemblies of *Ae. albopictus* (AaloF1) and the

6

145   C6/36 cell line (canu_80X_arrow2.2), followed by copy number validation, confirmed the presence

146   of seven *Piwi* genes (i.e. *Ago3, Piwi1/3, Piwi2, Piwi4, Piwi5, Piwi6 and Piwi7*) in *Ae. albopictus* (S1

147   Table). Genomic DNA sequences were obtained for each exon-intron boundaries, confirming in all

148   *Piwi* genes the presence of the PAZ, MID and PIWI domains, the hallmarks of the Piwi subfamily of

149   Argonaute proteins [35]. For *Ago3, Piwi1/3, Piwi2, Piwi4* and *Piwi6*, single transcript sequences

150   that correspond to predictions based on the identified DNA sequences were retrieved (S1 Dataset).

151   Sequencing results of the transcript from *Piwi5* showed a sequence 27 bp shorter than predicted

152   on the reference genome, due to a 45bp gap followed by a 18b insertion, 110 and 333 bases after

153   the ATG starting codon, respectively. This transcript still includes the PAZ, MID and PIWI domains.

154   The presence of this transcript was further validated by northern-blot (Fig 1). For *Piwi7*, the

155   transcript sequence also appears shorter than predicted (Fig 1). Alignment and phylogenetic

156   analyses, in the context of currently annotated *Piwi* transcripts of Culicinae and Anophelinae

157   mosquitoes, confirmed one-to-one orthologous pairing between *Ae. albopictus Piwi* gene

158   transcripts and those of Ae. aegypti (S2 Table, S1 Fig). Interestingly, *Piwi5, Piwi6* and *Piwi7*

159   transcripts group together and appear more similar to one of the two Aubergine-like transcripts

160   annotated in different Anophelinae species than to *Aedes Piwi2, Piwi1/3* and *Piwi4* transcripts.

161   Regarding the latter*, Piwi2* and *Piwi1/3* form a species-specific clade, rather than follow a

162   speciation pattern. Rather, the two genes, which on *Ae. aegypti* map on chromosome 1 and are

163   ~20 kb apart [17], may have originated as a duplication in the ancestor of *Ae. aegypti* and *Ae.*

164   *albopictus* and be subsequently undergoing interlocus gene conversion. This mechanism causes

165   nonreciprocal recombination, whereby one locus (i.e. part of a gene copy) replaces the

166   homologous sequence of the other copy. The result is concerted evolution of the gene duplicates

167   [36], which in this case eliminates divergence between *Piwi2* and *Piwi1/3* within each species.

168

169   **Figure 1. Gene and transcript structure of *Ae. albopictus Piwi5* and *Piwi7***. A) Schematic

170   representation of the DNA structure of *Piwi5* and *Piwi7* genes and their corresponding transcripts

171   as obtained from cDNA amplification of single sugar-fed mosquito samples. Exons and introns are

172   shown by blue boxes and black lines, respectively, with corresponding length in nucleotide below

173 each. The positions of the predicted PAZ, MID and PIWI domains are shown by green, blue and

174 magenta ovals, respectively. Exon numbers correspond to lane numbers. B) Amplification of each

175 exon of Piwi5 and Piwi7 on genomic DNA. Exon numbers correspond to lane numbers. C)

176 Northern-blot results of *Piwi5* indicate the presence of a transcript of 3 kb.

177

Piwi genes display high levels of polymorphism across populations and show signs of adaptive evolution

180 Across *Drosophila* phylogeny, genes of the piRNA pathway display elevated rates of adaptive

181 evolution [37], with rapidly evolving residues not clustering at the RNA binding site, but being

182 distributed across the proteins [3]. The RNA binding site is found within the PAZ domain, at the

183 amino-terminal part of Piwi proteins [35,38]. On the opposite side, at the carboxyl terminus, the

184 PIWI domain resides. The PIWI domain belongs to the RNase H family of enzymes and the

185 catalytic site is formed by three conserved amino acids (usually aspartate-aspartate-glutamate,

186 DDE or aspartate-aspartate-histidine, DDH) [35,39]. Between the PAZ and PIWI domains the MID

187 domain resides. MID specifies strand- and nucleotide-biases of piRNAs, including their Uridine 5'

188 bias [40,41]. To evaluate the selective pressures acting along these genes, we analysed the

189 polymorphism pattern in *Ae. albopictus* samples from wild-collected populations and from the

190 Foshan reference strain. Synonymous and non-synonymous mutations were found for each gene

191 in all populations (Fig 2), with *Piwi1/3* displaying the lowest polymorphism (Table 1).

192 **Table 1. Polymorphism of *Aedes albopictus Piwi* genes in mosquitoes from the Foshan strain**
193 **and wild-caught mosquitoes from La Reunion (Reu) and Mexico (Mex)**. We report the number
194 of sequences (*n*), as well as the number of sites (*L*), segregating sites (*S*), polymorphism measured
195 as π and θ, and the Tajima's *D* statistic for both synonymous (s) and non-synonymous sites (a) for
196 each gene and population (and for the pooled sample).
197

	n	*L*	L_s	L_a	S_s	S_a	π_s	π_a	θ_s	θ_a	π_a/π_s	D_s	D_a
Ago3													
Pooled	112	2832	680.2	2151.8	316	19	0.0699	0.0005	0.0878	0.0017	0.007	-0.68	-1.95
Foshan	32	2832	680.1	2151.9	124	5	0.0559	0.0004	0.0453	0.0006	0.007	0.89	-0.82
Mex	48	2832	680.2	2151.8	253	14	0.0780	0.0007	0.0838	0.0015	0.009	-0.25	-1.60
Reu	32	2658	643.8	2014.2	189	4	0.0678	0.0002	0.0729	0.0005	0.003	-0.27	-1.50
Piwi1-3													
Pooled	112	2658	644.3	2013.7	136	23	0.0319	0.0010	0.0399	0.0022	0.033	-0.66	-1.51

Foshan	32	2658	644.0	2014.0	10	2	0.0047	0.0003	0.0039	0.0002	0.064	0.68	0.44
Mex	48	2658	644.9	2013.1	117	21	0.0463	0.0017	0.0409	0.0024	0.037	0.48	-0.89
Reu	32	2658	643.8	2014.2	52	4	0.0188	0.0004	0.0201	0.0005	0.021	-0.23	-0.48
Piwi2													
Pooled	112	2625	644.0	1981.0	242	28	0.0760	0.0012	0.0710	0.0027	0.016	0.23	-1.65
Foshan	32	2625	644.0	1981.0	115	10	0.0663	0.0017	0.0443	0.0013	0.026	1.88	1.11
Mex	48	2625	643.9	1981.1	184	15	0.0823	0.0010	0.0644	0.0017	0.012	1.01	-1.28
Reu	32	2625	644.1	1980.9	151	6	0.0712	0.0005	0.0582	0.0008	0.007	0.85	-0.94
Piwi4													
Pooled	112	2592	620.0	1972.1	268	61	0.0729	0.0025	0.0817	0.0058	0.034	-0.36	-1.82
Foshan	32	2592	620.1	1971.9	122	18	0.0610	0.0009	0.0489	0.0023	0.015	0.94	-2.05
Mex	48	2592	619.8	1972.2	181	41	0.0692	0.0035	0.0658	0.0047	0.051	0.19	-0.87
Reu	32	2592	620.1	1971.9	161	45	0.0699	0.0029	0.0645	0.0057	0.041	0.32	-1.79
Piwi5													
Pooled	112	2745	653.1	2091.9	148	23	0.0457	0.0016	0.0428	0.0021	0.035	0.22	-0.66
Foshan	32	2793	664.5	2128.5	58	8	0.0361	0.0018	0.0217	0.0009	0.050	2.47	2.78
Mex	48	2745	652.9	2092.1	137	13	0.0470	0.0017	0.0473	0.0014	0.036	-0.02	0.65
Reu	32	2793	663.4	2129.6	89	6	0.0326	0.0008	0.0333	0.0007	0.025	-0.08	0.40
Piwi6													
Pooled	112	2661	649.0	2012.0	242	8	0.0805	0.0010	0.0705	0.0008	0.013	0.47	0.82
Foshan	32	2661	648.3	2012.8	92	3	0.0632	0.0001	0.0352	0.0004	0.002	2.99	-1.69
Mex	48	2661	649.9	2011.1	213	7	0.0840	0.0001	0.0739	0.0008	0.001	0.50	-2.33
Reu	32	2661	648.5	2012.5	163	4	0.0784	0.0001	0.0624	0.0005	0.001	0.98	-2.01
Piwi7													
Pooled	112	1977	469.8	1507.2	192	33	0.0877	0.0036	0.0772	0.0041	0.041	0.45	-0.42
Foshan	32	1977	469.8	1507.2	118	15	0.0905	0.0034	0.0624	0.0025	0.038	1.71	1.25
Mex	48	1977	469.9	1507.1	150	23	0.0905	0.0034	0.0719	0.0034	0.038	0.93	-0.04
Reu	32	1977	469.6	1507.5	137	17	0.0803	0.0030	0.0724	0.0028	0.037	0.41	0.24

Figure 2. Venn diagrams showing the number of positions harbouring synonymous and non-synonymous mutations in tested samples for each *Piwi* gene.

As expected, the laboratory strain Foshan showed the lowest levels of variability and Tajima's D values that contrast (in sign) from those of the other populations and from the pooled sample, consistent with a strong bottleneck associated to the strain establishment. In *Piwi4*, between 20 and 80 non-synonymous variants could be found inside and in proximity of the PAZ, MID and PIWI

9

207   domains (S2 Fig.A), ten of these mutations were shared across all populations (S3 Table). The 5'

208   region of *Piwi5* harboured several indels: two in-frame variants (i.e. 94_99del; 113_118del) were

209   shared across all populations and were present in homozygosity in at least one sample (S2 Fig.B),

210   suggesting that they are not detrimental. *Ago3* and *Piwi6* have very low non-synonymous

211   nucleotide diversity, suggesting strong constraints at the protein level. The results of the

212   McDonald-Kreitman test [42] further revealed an excess of non-synonymous substitutions

213   compared to the polymorphism pattern in both *Piwi1/3* and *Piwi6*, suggesting that they have been

214   target of adaptive evolution (Table 2.A). In contrast, *Piwi4* has a significant deficit of non-

215   synonymous substitutions and/or excess of polymorphic non-synonymous segregating sites (Table

216   2.A). In this gene, Tajima's D is negative but in line with the values of the other *Piwi* genes, and the

217   high non-synonymous polymorphism may reflect selection of intraspecific diversifying selection, as

218   expected in genes involved in immunity. Because positive selection may have acted at the level of

219   very few sites, this not contributing to the gene-level non-synonymous substitution pattern; we

220   explicitly tested models of codon evolution. Signs of positive selection were found at different sites,

221   including one site in the Linker2 and one site in the MID domain of *Piwi1/3,* two sites in the PAZ

222   domain of *Piwi4*, two sites in the Flex domain of *Piwi5* and three sites, two in the Flex and one in

223   the Linker2 domains, of *Piwi6* (Table 2.B). Haplotype reconstruction of our samples showed that

224   these mutations can co-occur on the same gene, with the only exception of Y278D+H287P in

225   *Piwi4* and A67P+G86S in *Piwi6*.

226   **Table 2**. **Insights into Evolutionary divergence of *Piwi* genes in *Ae. albopictus.*** A)

227   McDonald-Kreitman test for each *Piwi* gene using the orthologous sequences of *Ae. aegypti*

228   as outgroup. NI = Neutrality Index; Alpha = proportion of base substitutions fixed by natural

229   selection; *P* estimated using Fisher's exact test. B) Output of Codeml with significant results

230   regarding sites under positive selection.

231

A. McDonald-Kreitman test	Ago3	Piwi1/3	Piwi2	Piwi4	Piwi5	Piwi6	Piwi7
NI	0.582	0.516	0.9	3.888	0.696	0.154	0.745
alpha	0.418	0.484	0.1	-2.888	0.304	0.846	0.255
P	0.114	0.008	0.785	< 0.001	0.18	< 0.001	0.272

B. Codeml output for sites under positive selection

Gene	Position[1]	Reference>Mutant[2]	ω[3]	P[4]	Domain[5]
AGO3	-	-	-	-	
Piwi1-3	484	E>G	3.026	0.990*	Linker2
	485	K>R	2.979	0.965*	Linker2
	548	M>I	3.014	0.984*	MID
Piwi2	-	-	-	-	
Piwi4	278	Y>D	2.522	0.993**	PAZ
	287	H>A,D,P,V	2.532	1.000**	PAZ
Piwi5	89-90	SA>PT	7.813	1.000**	Flex
	139	T>A	7.810	1.000**	Flex
Piwi6	67	A>P	3.560	0.992**	Flex
	86	G>R,S	3.460	0.957*	Flex
	258	V>I	3.581	0.999**	Linker2
Piwi7	-	-	-	-	

232 Table 2 legend. [1] sites where signs of positive selection ($\omega > 1$) were found; [2] reference amino acid
233 and alternative missense variant; [3] mean omega (ω) value; [4] probability that $\omega > 1$ under the Bayes
234 empirical Bayes (BEB) method (* = P > 0.95; ** = P > 0.99); [5] protein domain based on
235 computational predictions of molecular structures. Domains are as follows: Linker2, linker region
236 between PAZ and MID; PAZ domain; MID domain; and Flex, the Flexible stretch at the N-
237 terminus.

238

239

240 Finally, to gain insight on how variable *Piwi* genes are in comparison to slow- and fast

241 -evolving genes of *Ae. albopictus*, we collected variability data of sets of genes previously identified

242 to have slow and high evolutionary rates [43]. For each population, we compared the overall level

243 of polymorphism (LoP) of the *Piwi* genes and of a dataset of fast-evolving genes (FGs) to that

244 measured for a dataset of slow-evolving genes (SGs) as listed in the material and methods section

245   "polymorphisms of *Piwi* genes" [43]. Our results indicate that *Piwi4*, *Piwi6* and *Piwi7* have LoP

246   values comparable to those of FGs, while *Ago3* and *Piwi5* do not significantly deviate from the LoP

247   values of SGs. *Piwi1*/3 appears to be conserved (Fig 3).

248

249   **Figure 3. Volcano plot.** Level of polymorphism (LoP) comparison between slow-evolving genes

250   (SGs), fast-evolving genes (FGs) and *Piwi* genes by population. Genes on the right side of the

251   panel have LoP values greater than those of SGs, while genes on the left side have smaller LoPs

252   than SGs. The y-axis represents the -log10 p-values of the Kolmogorov-Smirnov test. Faint

253   datapoints are not significant after Bonferroni correction for multiple testing (-log10 0.0024 (0.05/21

254   genes) = 2.62).

255

256   **Computational predictions of molecular structures**

257   The functional significance of the mutations under selection, as well as that of all the shared

258   missense mutations in the PAZ and PIWI domains, was tested by computing predictions of three

259   dimensional molecular structures of the Piwi proteins using the most-recent X-ray crystallography

260   structure of Argonaute proteins as templates [44,45]. Homology modelling revealed high structural

261   conservation among the seven Piwi proteins despite sequence heterogeneity (S2 Fig.; Fig 4.A).

262   Similarly, to *D. melanogaster*, the highest levels of amino acid sequence conservation were

263   found in the regions that, based on homology modelling, define the inner pocket of Argonaute

264   molecular assembly where the RNA binds. Significantly lower sequence conservation was found

265   on the proteins surface (Fig 4.B). Based on our computational predictions, we could not detect

266   amino acidic polymorphisms that would affect RNA binding or processing, suggesting that all *Ae.*

267   *albopictus* Piwi proteins may retain the Argonaute-like functions. Mapping of mutations under

268   positive selection (Table 2.B) on the homology models and sequence comparisons with known

269   PIWI structural homologs showed that the identified variant amino acids are unlikely to induce

270   severe alterations in protein folding. All mutant variants were found to localize in regions distant

271   from the predicted RNA-binding and/or processing sites, ruling out possible effects associated to

272   alterations in RNA recognition, but raising the intriguing possibility of regulatory roles during

273   interactions with additional binding partners.

274

275   **Figure 4. Computational homology models of the *Ae. Albopictus Piwi* proteins.** Homology

276   models were generated for the seven *Piwi* genes as described in the methods section. A)

277   Superposition of cartoon representations of Piwi homology models, with highlight of domain

278   organization: the N-terminal domain is shown in orange, the PAZ domain in green, the MID domain

279   in blue and the PIWI domain in magenta. B) *CONSURF* [92]overview of the aminoacid sequence

280   conservation mapped on three-dimensional homology models in a putative RNA-bound

281   arrangement based on the structure of human Argonaute bound to a target RNA (PDB ID 4Z4D),

282   colored from teal (very low conservation) to dark magenta (highly conserved).

283

Developmental profile of *Ae. albopictus Piwi* genes

285   To further gain insights on the functional specialization of *Piwi* genes, we assessed their

286   expression profile throughout mosquito development, namely at 4-8 hours (h) after deposition to

287   capture the maternal-zygotic transition in expression, at late embryogenesis (i.e. 12-16 h and 16-

288   24 h post deposition), at two time points during larval development (i.e. 1st and 4th instar larvae)

289   and at pupal and adult stages (for the latter only we sampled separately males and females). Adult

290   females were dissected to extract ovaries from the carcasses both from females kept on a sugar

291   diet and 48 h after a blood meal, when a peak in *Piwi* gene expression was previously observed

292   [46].

293   Expression levels of *Ago3, Piwi4, Piwi5, Piwi6* and *Piwi7* are at their peak in the embryonic stages,

294   although at different time points (Fig 5). Overall, *Ago3, Piwi1/3, Piwi2* and *Piwi6* have a similar

295   trend during development showing a second peak of expression in adult females and their ovaries,

296   while the expression levels of *Piwi4*, *Piwi5* and *Piwi7* remain constant. In details, *Piwi7* is mostly

297   expressed 4-8h after deposition, *Piwi5* and *Piwi6* are mostly expressed after 8-16h and *Ago3* and

298   *Piwi4* have their pick of expression at 16-24h. On the contrary, *Piwi1/3* and *Piwi2* are mostly

299   expressed in ovaries extracted from blood-fed and sugar-fed females, respectively (Fig 5.A, S4.A

13

table). These results are consistent with lack of expression from published RNA-seq data from adult mosquitoes.

Figure 5. Expression profile of *Piwi* genes. Heatmap representations of log10 transformed fold-change expression values of each *Piwi* gene. A) Developmental expression pattern of the *Piwi* genes normalized on the expression in sugar fed females. B) Expression pattern of *Piwi* genes following viral infection normalized with respect to sugar-fed samples. Expression was verified in ovaries and carcasses separately, during the early and late stages of infections, that is 4 dpi for both viruses and 14 or 21 dpi for CHIKV and DENV, respectively. Each day post-infection was analysed with respect to sugar and blood-fed controls of the same day. * indicates significant difference (P<0.05) between infected samples and the corresponding blood-fed control.

Overall, at the adult stages, *Ago3* and all *Piwi* genes were more expressed in females than males. Expression in ovaries was higher than in the corresponding carcasses, in both sugar- and blood-fed females. Differences in carcasses vs. ovaries expression were more pronounced after blood-meal for *Ago3, Piwi1/3* and *Piwi6*, while expression of *Piwi2* was doubled in sugar-fed vs. blood-fed ovaries.

Piwi genes expression following viral infection

Finally, we assessed whether the expression pattern of *Piwi* genes was altered upon DENV and CHIKV infection (Fig 5.B). The expression profile of the *Piwi* genes was different following CHIKV- and DENV-infection and also when comparing samples of carcasses and ovaries. In ovaries, during CHIKV infection all *Piwi* genes were significantly up-regulated compared to both sugar- and blood-fed mosquitoes. Four days post infection (dpi), the expression of *Ago3, Piwi1/3, Piwi6* and *Piwi7* was between 4 to 10 folds higher than that of *Piwi2, Piwi4* and *Piwi5*, which nevertheless were upregulated with respect to ovaries of sugar- and blood-fed mosquitoes. An opposite profile was seen in the carcasses, where all *Piwi* genes, particularly *Piwi1/3* and *Piwi4*, were down-regulated. At 4 dpi, CHIKV has already disseminated throughout the mosquito body, has reached the salivary glands and is able to be transmitted. CHIKV viral titer was reduced ten

328　folds by 14 dpi and the profile of *Piwi* genes changed. Expression in the ovaries decreased

329　between 3 (*Piwi5*) to 20 (*Piwi7*) times with respect to values observed at 4 dpi, but remained higher

330　than the corresponding expression values in ovaries of both sugar and blood-fed mosquitoes. In

331　carcasses all *Piwi* genes inverted their expression pattern during the infection phase, increasing up

332　to more than 100 times in the case of *Piwi4, Piwi5* and *Piwi6*. At 14 dpi, expression of the *Piwi*

333　genes was highest in CHIKV infected carcasses than in carcasses of sugar- and blood-fed

334　mosquitoes.

335　For DENV, infection progresses differently than CHIKV. At 4 dpi there is no virus in the salivary

336　gland, where the viral titer was measured at zero. By 21 dpi, DENV has established persistent

337　infection [47]. At 4 dpi expression of *Piwi* genes was lower in DENV- and blood-fed ovaries than in

338　ovaries of sugar-fed mosquitoes. The only exception was *Piwi6*, which was slightly up regulated in

339　ovaries of DENV-infected samples, but slightly down-regulated in ovaried of blood-fed mosquitoes.

340　On the contrary, at the same time point, carcasses of DENV-infected samples showed a drastic

341　increase in the expression of all *Piwi* genes with respect to blood-fed samples; this increase was

342　between 7 to 87 times for *Piwi7* and *Piwi2*, respectively. By 21 dpi, expression in the ovaries

343　increased for all *Piwi* genes, in comparison to what observed both at 4dpi and in blood-fed ovaries,

344　suggesting the increase in expression of *Piwi* genes is related to DENV dissemination.

345　Interestingly, if we compare levels of expression in CHIKV-infected ovaries at 4 dpi and DENV-

346　infected samples at 21 dpi, corresponding to the time at which both viral species have

347　disseminated throughout the mosquito body, we observe similar levels of fold-change expression

348　of *Piwi4* and *Piwi7*, while *Ago3, Piwi1/3* and *Piwi6* show higher fold-change in CHIKV compared to

349　DENV samples. Whether this trend is dependent on the viral species or viral titer requires further

350　investigation. The same type of comparison in carcasses shows a higher fold-change expression

351　level of all *Piwi* genes, particularly *Piwi1/3* and *Piwi5*, in DENV- versus CHIKV-infected samples,

352　even if viral titer 300 are lower for DENV (S4.B Table). Overall these results support the hypothesis

353　of a concerted activity of all PIWI proteins during viral dissemination for DENV, and maintenance of

354　infection rely on expression of primarily *Piwi5*. On the contrary, establishment of persistent CHIKV

15

355 infection was accompanied by elicitation of all Piwi gene expression, particularly *Piwi4* and, again,

356 *Piwi5*.

357 **DISCUSSION**

358   The piRNA pathway does not have antiviral immunity in *D. melanogaster* [15]. In the

359 arboviral vectors *Aedes spp.* mosquitoes, vpiRNAs are found following infections with arboviruses,

360 piRNAs are produced in the soma besides the germline and there has been an expansion on the

361 number of *Piwi* genes, supporting the hypothesis that the piRNA pathway has antiviral

362 immunity[12,48,49]. Besides *Ago3*, the genome of *Ae. aegypti* harbours six *Piwi* genes (i.e.

363 *Piwi1/3, Piwi2, Piwi4, Piwi5, Piwi6, Piwi7),* some of which show tissue and development-specific

364 expression profile and have been preferentially associated with either TE-derived or viral piRNAs,

365 [16,20,21]. These studies were based on the knowledge of the gene structure of each *Ae. aegypti*

366 *Piwi* gene and the application of *ad hoc* RNAi-based silencing experiments and *in vitro* expression

367 assays, but lack an evolutionary perspective [18–21].

368 In this work we focused on the emerging arboviral vector *Ae. albopictus* and we show how

369 the application of evolutionary and protein modelling techniques helps to unravel functional

370 specialization of Piwi proteins. The genome of *Ae. albopictus* harbours one copy of *Ago3* and six

371 *Piwi* genes (i.e. *Piwi1/3, Piwi2, Piwi4, Piwi5, Piwi6 and Piwi7*), each a one-to-one orthologue to the

372 *Ae. aegypti Piwi* genes. The only exceptions are *Piwi2* and *Piwi1/3*, where the two genes from the

373 same species cluster together. In *Ae. aegypti*, these two genes both map on Chromosome 1,

374 separated by ~ 20kb, suggesting they may undergo frequent gene conversion.

375 All transcripts retain the PAZ and PIWI domains, which are the hallmarks of the Argonaute protein

376 family [35].

377 By using homology modelling, we obtained predictions of molecular architectures

378 for *Ae. albopictus Ago3* and Piwi proteins, onto which we mapped the putative boundaries of each

379 domain. Superpositions and sequence comparisons allowed clear identification of the catalytic

380 DDH triad within the PIWI domain of all modelled proteins. This conservation is consistent with

381 strong sequence matching in the putative RNA binding regions of the PIWI, PAZ and MID domains

16

382 and suggests the possible maintenance of slicer activity, albeit experimental validation of each

383 isoform is necessary.

384 The expression of all *Piwi* genes was confirmed throughout the developmental stages and the

385 adult life of the mosquito, both in ovaries and somatic tissues. Interestingly, *Piwi7* transcript

386 expression starkly drops following early embryogenesis, to the point that we could detect it neither

387 in RNA-seq analyses, nor in Northern-blot experiments (S5 fig.).

388 The expression of *Piwi* genes was elicited upon arboviral infection, indirectly confirming the

389 antiviral role of the piRNA pathway. The expression profile of *Piwi* genes showed differences

390 depending on both the species of infecting virus and on when the expression was measured. In

391 CHIKV-infected samples, expression of *Piwi* genes was mostly elicited in ovaries or carcasses at 4

392 or 14 dpi, respectively. On the contrary, in DENV-infected samples, the highest expression *of Piwi*

393 genes was seen in carcasses 4 dpi. These results are concordant with the timing in piRNAs

394 accumulation following CHIKV or DENV infection. In *Ae. albopictus* mosquitoes infected with

395 CHIKV, secondary piRNAs are not found 3 dpi, but are enriched 9 dpi [9]. In contrast, in *Ae.*

396 *aegypti* mosquitoes infected with DENV2, piRNAs are the dominant small RNA populations 2 dpi

397 [49].

398 Overall, these observations and our expression analyses support the hypothesis of an early

399 activation of the piRNA pathway following DENV infection, but a late activation after CHIKV

400 infection.

401 Additionally, our expression analysis is consistent with a generalist antiviral role for *Piwi5*, which is

402 elicited both during DENV and CHIKV infection [20], but suggest a more prominent role for *Piwi6*

403 and *Piwi1/3* or *Piwi4* and *Ago3* during infection with DENV and CHIKV, respectively.

MATERIAL AND METHODS

Mosquitoes

406 *Aedes albopictus* mosquitoes of the Foshan strain were used in this study. This strain was

407 established in 1981 in the Center for Disease Control and Prevention of Guangdong Province in

408 China. It has been at the University of Pavia since 2013 [10,50]. Mosquitoes are reared under

409   constant conditions, at 28°C and 70-80% relative humidity with a 12/12h light/dark cycle. Larvae

410   are reared in plastic containers, at a controlled density to avoid competition for food. Food is

411   provided daily in the form of fish food (Tetra Goldfish Gold Colour). Adults are kept in 30 cm^3 cages

412   and fed with cotton soaked in 0.2 g/ml sucrose as a carbohydrate source. Adult females are fed

413   with defibrinated mutton blood (Biolife Italiana) using a Hemotek blood feeding apparatus.

414   Mosquitoes from Mexico and La Reunion island were collected in 2017 as adults and maintained in

415   ethanol 70% before shipment to Italy. All samples were processed at the University of Pavia.

Mosquito infections

417   Foshan mosquitoes were infected with DENV serotype 1, genotype 1806 or CHIKV 06.21.

418   DENV-1 (1806) was isolated from an autochthonous case from Nice, France in 2010 [51]. CHIKV

419   06-21 was isolated from a patient on La Reunion Island in 2005 [52]. Both strains were kindly

420   provided by the French National Reference Center for Arboviruses at the Institut Pasteur. CHIKV

421   06-21 and DENV-1 1806 were passaged twice on cells to constitute the viral stocks for

422   experimental infections of mosquitoes, on C6/36 cells for CHIKV 06-21 and on African green

423   monkey kidney Vero cells for DENV-1 1806. Viral titers of stocks were estimated by serial dilutions

424   and expressed in focus-forming units (FFU)/mL.

425   Four boxes containing 60 one-week-old females were exposed to an infectious blood-meal

426   composed by 2 mL of washed rabbit red blood cells, 1 mL of viral suspension and 5 mM of ATP.

427   The titer of the blood-meal was 10^7 PFU/mL for CHIKV and $10^{6.8}$ PFU/mL for DENV. Fully

428   engorged females were placed in cardboard boxes and fed with a 10% sucrose solution.

429   Mosquitoes were incubated at 28°C until analysis.

430   In parallel, mosquitoes were fed with uninfected blood-meal or kept on a sugar-diet and

431   grown in the same conditions. Thirty mosquitoes were killed to be analyzed at days 4 and 14 post-

432   infection (pi) for CHIKV, and at days 4 and 21 pi for DENV. To estimate transmission, saliva was

433   collected from individual mosquitoes as described in [53]. After removing wings and legs from each

434   mosquito, the proboscis was inserted into a 20 µL tip containing 5 µL of Fetal Bovine Serum (FBS)

435   (Gibco, MA, USA). After 30 min, FBS containing saliva was expelled in 45 µL of Leibovitz L15

436   medium (Invitrogen, CA, USA) for titration. Transmission efficiency refers to the proportion of

437 mosquitoes with infectious saliva among tested mosquitoes (which correspond to engorged

438 mosquitoes at day 0 pi having survived until the day of examination). The number of infectious

439 particles in saliva was estimated by focus fluorescent assay on C6/36 *Ae. albopictus* cells.

440 Samples were serially diluted and inoculated into C6/36 cells in 96-well plates. After incubation at

441 28°C for 3 days (CHIKV) or 5 days (DENV), plates were stained using hyperimmune ascetic fluid

442 specific to CHIKV or DENV-1 as primary antibody. A Fluorescein-conjugated goat anti-mouse was

443 used as the second antibody (Biorad). Viral titers were 16,266±50,446 FFU and 155±125 FFU for

444 CHIKV at 14 dpi and DENV at 21 dpi, respectively.

445   At the same time points mosquitoes that had been fed a non-infectious blood or kept on a

446 sugar diet were sampled and dissected as above.

447 **Bioinformatic identification of *Piwi* genes in the *Ae. albopictus* genome**

448   The sequences of the *Ae. aegypti Piwi* genes [54] were used as query to find orthologs in

449 the reference genome of the *Ae. albopictus* Foshan strain (AaloF1 assembly) and in the genome of

450 the *Ae. albopictus* C6/36 cell line (canu_80X_arrow2.2 assembly) using the BLAST tool in

451 Vectorbase. Deduced coding sequences (CDS) where analysed in Prosite

452 (Prosite.expasy.org/prosite.html) to screen for the typical PAZ and PIWI domains of Argonaute

453 proteins [55].

454 **Copy number of *Piwi* genes**

455   qPCR reactions were performed using the QuantiNova SYBR Green PCR Kit (Qiagen)

456 following the manufacturer's instructions on an Eppendorf Mastercycler RealPlex4, on genomic

457 DNA from four mosquitoes and using gene-specific primers, after having verified their efficiency

458 (S4 Table). DNA was extracted using DNA Isolation DNeasy Blood & Tissue Kit (Qiagen).

459 Estimates of gene copy number were performed based on the $2^{-\Delta CT}$ method using *Piwi6* and the

460 para sodium channel genes (AALF000723) as references [56].

461 **Structure of *Piwi* genes**

462   DNA extracted from whole mosquitoes and dissected ovaries [57] was used as template in

463 PCR amplifications to confirm the presence and the genome structure of each bioinformatically-

464   identified *Piwi* gene. Primers were designed to amplify each exon, with particular attention to detect

465   differences between paralogous *Piwi* genes (S1 Table). The DreamTaq Green PCR Master Mix

466   (Thermo Scientific) was used for PCR reactions with the following parameter: 94°C for 3 minutes,

467   40 cycles at 94°C for 30 sec, 55°C-62°C for 40 sec, 72°C for 1-2 minutes and final extension step

468   of 72°C for 10 minutes. PCR products were visualized under UV light after gel electrophoresis

469   using 1-1.5% agarose gels stained with ethidium bromide and a 100 bp or 1 kb molecular marker.

470   PCR products were either directly sequenced or cloned using the TOPO® TA Cloning® Kit

471   strategy (Invitrogen) following the manufacturer's instructions. DNA plasmids were purified using

472   the QIAprep Spin Miniprep Kit and sequenced.

473   ***Piwi* gene transcript sequences and phylogeny**

474   RNA was extracted using a standard TRIzol protocol from pools of 5 adult female

475   mosquitoes to verify the transcript sequence of each *Piwi* gene. Sets of primers were designed for

476   each gene to amplify its entire transcript sequence (S4 Table). PCR reactions were performed

477   using a High Fidelity taq-polymerase (Platinum SuperFi DNA Polymerase, Invitrogen) following

478   manufacturer's instructions. PCR products were cloned using the TOPO® TA Cloning® Kit

479   (Invitrogen) and plasmid DNA, purified using the QIAprep Spin Miniprep Kit, was sequenced. Rapid

480   amplification of cDNA ends (RACE) PCRs were performed using FirstChoice RLM-RACE Kit

481   (Thermo Fisher Scientific) to analyse 5' and 3' ends of the transcript sequences following

482   manufacturer's instructions. Amplification products were cloned and sequenced as previously

483   indicated.

484   Sequences of the identified *Ae. albopictus Piwi* gene transcripts were aligned to sequences

485   of *Culicidae* and *D. melanogaster Piwi* transcripts, as downloaded from VectorBase

486   (www.vectorbase.org), using MUSCLE [58]. Maximum-likelihood based phylogenetic inference was

487   based on RAxML after 1000 bootstrap resampling of the original dataset. Phylogeny reconstruction

488   was done through the CIPRESS portal (http://www.phylo.org/index.php/). Resulting tree was

489   visualised using FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

Northern Blot analysis

10µg of total RNA from a pool of 10 sugar-fed females was run in a 1% x 2% agarose/formaldehyde gel (1 g agarose, 10 ml 10x MOPS buffer, 5.4 ml 37% formaldehyde, 84.6 ml DEPC water). Gels were washed twice in 20x SSC for 15 minutes prior to blotting. RNA was transferred to a Amersham Hybond-N+ nylon membrane (GE healthcare) using 20x SSC and cross-linked using UV light exposure for 1 minute. Probes were labelled with biotin using Biotin-High Prime (Roche). Hybridization and detection of biotinylated probes was performed using the North2South™ Chemiluminescent Hybridization and Detection Kit (Thermo Fisher Scientific) following manufacturer instructions.

Polymorphisms of *Piwi* genes

We investigated *Piwi* gene polymorphism by looking at the distribution of single nucleotide polymorphism in whole genome sequence data from a total of 56 mosquitoes, of which 24 from Mexico, 16 from the island of La Reunion island and 16 from the reference Foshan strain. Whole genome sequencing libraries were generated and sequenced on the Illumina HiSeqX platform at the Genomics Laboratory of Verily in South San Francisco, California to generate 150 basepair paired end reads. Whole Genome Sequencing data alignments have been deposited to the SRA archive (BioProjects PRJNA484104 and PRJNA562979). Libraries from Tampon and Tapachula had an average of 225216071 reads, meaning an average coverage of 15X based on the AaloF1 assembly (S4 Fig.).

Illumina reads were mapped to *Piwi* gene transcript sequences using Burrows-Wheeler Aligner (BWA-MEM) [59] with custom parameters. Polymorphisms was tested by Freebayes [60]. Annotation of the detected mutations, as well counts of synonymous and non-synonymous variants, were performed in snpEff [61]. Frameshifts and non-synonymous variants were plotted using muts needle-plot [62]. Venn diagrams of positions with mutations in the three tested samples were built using Venny 2.1 [63]. Haplotype reconstruction was performed using seqPHASE [64] and PHASE [65,66]. The inferred haplotypes were analysed with DnaSP [67], which estimated the number of segregating sites and the level of nucleotide diversity π [68] in both synonymous and non-synonymous sites. We manually calculated, for synonymous and non-synonymous positions

21

518   separately, the nucleotide diversity estimator theta [69] and Tajima's D statistic [70], which are a

519   function of the total number of sites and the number of segregating sites (both estimated by

520   DnaSP), and of sample size (see references for detailed formulas). We also tested for signatures

521   of adaptive evolution using the McDonald-Kreitman test [42] (as implemented in DnaSP), which

522   compares the rate of polymorphism and substitutions in synonymous and non-synonymous sites.

523   For this analysis we used alignments that included the orthologous sequences from *Ae. aegypti*.

524   Haplotype sequences for each gene from each individual were also aligned in TranslatorX

525   [71] using Clustalw [72] and used for Maximum-likelihood based phylogenetic inference based on

526   RAxML after 1000 bootstrap under the GTRGAMMA model. The relative rates of synonymous and

527   nonsynonymous mutations (dN/dS = ω) averaged across sites was calculated using Codeml in

528   PAML version 4.9 [73], as implemented in PAMLX [74]. Signs of selective pressure for each

529   detected mutation were investigated comparing the M1a (nearly-neutral) *versus* the M2a (positive

530   selection) site models by inferring ω estimations and posterior probabilities under the Bayes

531   empirical Bayes (BEB) approach as implemented in Codeml [73]. This analysis was performed with

532   the following default parameters: runmode = 0, clock= 0, Mgene = 0, CodonFreq = 0, estFreq = 0,

533   fix_blength = 0, optimization method= 0, icode= 0, Seqtype = 1, fix α = 0, ncatG = 5, Small_Diff =

534   5e-7, n = 1, aaDist = 0.

535   The level of polymorphism (LoP) for slow-evolving genes (SGs) (AALF008224,

536   AALF005886, AALF020750, AALF026109, AALF014156 ,AALF018476 ,AALF014287,

537   AALF004102, AALF003606, AALF019476, AALF028431, AALF018378, AALF027761,

538   AALF014448), fast-evolving genes (FGs) (AALF010748, AALF022019, AALF024551,

539   AALF017064, AALF004733, AALF018679, AALF028390, AALF026991, AALF014993,

540   AALF009493, AALF010877, AALF012271, AALF009839, AALF019413) and the *Piwi* genes was

541   calculated for each population following the pipeline as in [43]. Briefly, SNPs and INDELs were

542   inferred using four Variant callers (i.e. Freebayes [60], Platypus [75], Vardict [76] and GATK

543   UnifiedGenotyper [77]) and the data merged and filtered with custom scripts. Filters include:

544   minimum phred mapping quality = 20 (corresponding to 0.01 error rate), minimum phred base

545   quality = 20, minimum allele frequency = 0.2, minimum allele observation = 2, minimum coverage =

546   8, maximum depth = 5000. The LoP for each individual was calculated as the number of variants

547   averaged over the region length and the median value for each population was used for

548   subsequent analyses. Statistical analyses were performed in R studio [78]. Fold-change

549   differences were computed as the ratio of the median LoP for each *Piwi* gene and each FG gene

550   over the median LoP of the SG genes. Statistical differences in LoP distribution was assessed via

551   the Kolmogorov-Smirnov test and the p-value threshold was adjusted with the Bonferroni

552   correction.

553   **Homology modelling**

554   Computational structural investigations were carried out initially through the identification of

555   the closest homologs based on sequence similarity (using *NCBI Blast* [79]) and secondary

556   structure matching (using *HHPRED* [80]), using the whole PDB as source collection. Homology

557   model were then generated using *MODELLER* [81] by selecting only homologous RNA-bound

558   structures as template models: *Kluyveromyces polysporus* Argonaute with a guide RNA (PDB ID

559   4F1N), Human Argonaute2 Bound to t1-G Target RNA (PDB ID 4z4d [82], *T. thermophilus*

560   Argonaute complexed with DNA guide strand and 19-nt RNA target strand (PDB ID 3HM9), and

561   silkworm PIWI-clade Argonaute Siwi bound to piRNA (PDB ID 5GUH).

562   Computational models were manually adjusted through the removal of non-predictable N-

563   and C-terminal flexible regions using *COOT* [83] followed by geometry idealization in *PHENIX* [84]

564   to adjust the overall geometry. Final model quality was assessed by evaluating average bond

565   lengths, bond angles, clashes, and Ramachandran statistics using Molprobity [85] and the *QMEAN*

566   server [86] Structural figures were generated with *PyMol* [87].

567   **Developmental expression profile of *Piwi* genes**

568   Publicly available RNA-seq data (runs: SRR458468, SRR458471, SRR1663685,

569   SRR1663700, SRR1663754, SRR1663913, SRR1812887, SRR1812889, SRR1845684) were

570   downloaded and aligned using Burrows-Wheeler Aligner (BWA-MEM) [59] to the current *Ae.*

571   *albopictus* genome assembly (AaloF1). Aligned reads were visualized in Integrative Genomics

572   Viewer (IGV) [88]. Total RNA was extracted from embryos, 1st and 4th instar larvae, pupae, and

573 adults using Trizol (Thermo Fisher Scientific). Embryos consisted of two pools of 60 eggs at

574 different time points (i.e. 4-8h, 8-16h and 16-24h). Adult samples consisted of males and females

575 kept on a sugar-diet and females fed an uninfected blood-meal. Blood-fed females were dissected

576 to separate ovaries from the carcasses 48 h after blood-meal. These parameters were based on

577 the results of previous studies on *Anopheles stephensi* and *Ae. aegypti* that showed high *Piwi*

578 gene expression during early embryogenesis or 48-72h post blood meal [46]. For each stage, RNA

579 was extracted from pools of 10 mosquitoes, except for first instar larvae and embryos when 20 and

580 60 individuals were used respectively.

581 RNA was DNaseI-treated (Sigma-Aldrich) and reverse-transcribed in a 20 µl reaction using

582 the qScript cDNA SuperMix (Quantabio) following the manufacturer's instructions. Quantitative RT-

583 PCRs (qRT-PCR) were performed as previously described using two biological replicates per

584 condition and the RPL34 gene as housekeeping [89]. Relative quantification of *Piwi* genes was

585 determined using the delta-delta-Ct method implemented in the software qBase+ (Biogazelle).

586 Expression values were normalized with respect to those obtained from sugar-fed females.

Expression analyses following infection

588 Fold-change expression values for each *Piwi* gene was assessed for non-infectious-blood-fed

589 controls, CHIKV-infected and DENV-infected samples after normalization on sugar-fed controls.

590 qRT-PCR experiments (S4 Table) were set up for two replicate pools of 15 ovaries and 15

591 carcasses at days 4, 14 and 4, 21 for CHIKV and DENV, respectively and the corresponding sugar

592 and non-infectious-blood controls. RNA extraction, qRT-PCR and data analyses were performed

593 as described in the previous paragraph (see "Developmental expression profile of *Piwi* genes").

594 Fold-change differences significance was assessed using the Analysis of Variance (ANOVA)

595 procedure [90,91] as implemented in qBASE+.

596

REFERENCES

1. Bohmert K, Camus I, Bellini C, Bouchez D, Caboche M, Banning C. AGO1 defines a novel locus of Arabidopsis controlling leaf development. EMBO J. 1998; doi:10.1093/emboj/17.1.170

2. Swarts DC, Makarova K, Wang Y, Nakanishi K, Ketting RF, Koonin E V., et al. The evolutionary journey of Argonaute proteins. Nature Structural and Molecular Biology. 2014. doi:10.1038/nsmb.2879

3. 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: 507–518. doi:10.1093/gbe/evw018

4. Buck AH, Blaxter M. Functional diversification of Argonautes in nematodes: an expanding universe: Figure 1. Biochem Soc Trans. 2013; doi:10.1042/BST20130086

5. Bollmann SR, Press CM, Tyler BM, Grünwald NJ. Expansion and divergence of argonaute genes in the oomycete genus *phytophthora*. Front Microbiol. 2018; doi:10.3389/fmicb.2018.02841

6. Singh RK, Gase K, Baldwin IT, Pandey SP. Molecular evolution and diversification of the Argonaute family of proteins in plants. BMC Plant Biol. 2015; doi:10.1186/s12870-014-0364-6

7. Bronkhorst AW, Van Rij RP. The long and short of antiviral defense: Small RNA-based immunity in insects. Current Opinion in Virology. 2014. doi:10.1016/j.coviro.2014.03.010

8. Poirier EZ, Goic B, Tomé-Poderti L, Frangeul L, Boussier J, Gausson V, et al. Dicer-2-Dependent Generation of Viral DNA from Defective Genomes of RNA Viruses Modulates Antiviral Immunity in Insects. Cell Host Microbe. 2018; doi:10.1016/j.chom.2018.02.001

9. 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; doi:10.1038/ncomms12410

10. Palatini U, Miesen P, Carballar-Lejarazu R, Ometto L, Rizzo E, Tu Z, et al. Comparative genomics shows that viral integrations are abundant and express piRNAs in the arboviral vectors *Aedes aegypti* and *Aedes albopictus*. BMC Genomics. 2017; doi:10.1186/s12864-017-3903-3

11. Whitfield ZJ, Dolan PT, Kunitomi M, Tassetto M, Seetin MG, Oh S, et al. The Diversity, Structure, and Function of Heritable Adaptive Immunity Sequences in the *Aedes aegypti* Genome. Curr Biol. 2017; doi:10.1016/j.cub.2017.09.067

12. Miesen P, Joosten J, van Rij RP. PIWIs Go Viral: Arbovirus-Derived piRNAs in Vector Mosquitoes. PLoS Pathog. 2016;12: 1–17. doi:10.1371/journal.ppat.1006017

13. Olson KE, Bonizzoni M. Nonretroviral integrated RNA viruses in arthropod vectors: an occasional event or something more? Curr Opin Insect Sci. Elsevier Inc; 2017;22: 45–53.

637   doi:10.1016/j.cois.2017.05.010

638   14.   Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, et al. Discrete Small

639   RNA-Generating Loci as Master Regulators of Transposon Activity in *Drosophila*. Cell.

640   2007;128: 1089–1103. doi:10.1016/j.cell.2007.01.043

641   15.   Petit M, Mongelli V, Frangeul L, Blanc H, Jiggins F, Saleh M-C. piRNA pathway is not

642   required for antiviral defense in *Drosophila melanogaster*. Proc Natl Acad Sci. 2016;

643   doi:10.1073/pnas.1607952113

644   16.   Akbari OS, Antoshechkin I, Amrhein H, Williams B, Diloreto R, Sandler J, et al. The

645   Developmental Transcriptome of the Mosquito *Aedes aegypti* , an Invasive Species and

646   Major Arbovirus Vector. G3&#58; Genes|Genomes|Genetics. 2013;

647   doi:10.1534/g3.113.006742

648   17.   Matthews BJ, Dudchenko O, Kingan SB, Koren S, Antoshechkin I, Crawford JE, et al.

649   Improved reference genome of *Aedes aegypti* informs arbovirus vector control. Nature.

650   2018; doi:10.1038/s41586-018-0692-z

651   18.   Miesen P, Ivens A, Buck AH, van Rij RP. Small RNA Profiling in Dengue Virus 2-Infected

652   *Aedes* Mosquito Cells Reveals Viral piRNAs and Novel Host miRNAs. PLoS Negl Trop Dis.

653   2016; doi:10.1371/journal.pntd.0004452

654   19.   Girardi E, Miesen P, Pennings B, Frangeul L, Saleh MC, Van Rij RP. Histone-derived piRNA

655   biogenesis depends on the ping-pong partners Piwi5 and Ago3 in *Aedes aegypti*. Nucleic

656   Acids Res. 2017; doi:10.1093/nar/gkw1368

657   20.   Miesen P, Girardi E, Van Rij RP. Distinct sets of PIWI proteins produce arbovirus and

658   transposon-derived piRNAs in *Aedes aegypti* mosquito cells. Nucleic Acids Res. 2015;43:

659   6545–6556. doi:10.1093/nar/gkv590

660   21.   Varjak M, Kean J, Vazeille M, Failloux A, Kohl A. *Aedes aegypti* Piwi4 Is a Noncanonical.

661   mSphere. 2017;2: e00144-17.

662   22.   Varjak M, Leggewie M, Schnettler E. The antiviral piRNA response in mosquitoes? 2018; 1–

663   12. doi:10.1099/jgv.0.001157

664   23.   Schnettler E, Donald CL, Human S, Watson M, Siu RWC, McFarlane M, et al. Knockdown of

665   piRNA pathway proteins results in enhanced semliki forest virus production in mosquito

666   cells. J Gen Virol. 2013;94: 1680–1689. doi:10.1099/vir.0.053850-0

667   24.   Hahn MW. Distinguishing among evolutionary models for the maintenance of gene

668   duplicates. Journal of Heredity. 2009. doi:10.1093/jhered/esp047

669   25.   Obbard DJ, Jiggins FM, Halligan DL, Little TJ. Natural selection drives extremely rapid

670   evolution in antiviral RNAi genes. Curr Biol. 2006;16: 580–585.

671   doi:10.1016/j.cub.2006.01.065

672   26.   Bonizzoni M, Gasperi G, Chen X, James AA. The invasive mosquito species *Aedes*

673   *albopictus*: Current knowledge and future perspectives. Trends in Parasitology. 2013.

674   doi:10.1016/j.pt.2013.07.003

675   27.   Rezza G. Dengue and chikungunya: long-distance spread and outbreaks in naïve areas.

676   Pathog Glob Health. 2014; doi:10.1179/2047773214Y.0000000163

677   28.   Bonilauri P, Bellini R, Calzolari M, Angelini R, Venturi L, Fallacara F, et al. Chikungunya

678   virus in *Aedes albopictus*, Italy. Emerging Infectious Diseases. 2008.

679   doi:10.3201/eid1405.071144

680   29.   Venturi G, Di Luca M, Fortuna C, Elena Remoli M, Riccardo F, Severini F, et al. Detection of

681   a chikungunya outbreak in Central Italy Detection of a chikungunya outbreak in Central.

682   Euro Surveill. 2017;22: 1–4. doi:10.2807/1560

683   30.   Gjenero-Margan I, Aleraj B, Krajcar D, Lesnikar V, Klobučar A, Pem-Novosel I, et al.

684   Autochthonous dengue fever in Croatia, August- September 2010. Eurosurveillance.

685   2011;16: 1–4. doi:19805 [pii]

686   31.   Marchand E, Prat C, Jeannin C, Lafont E, Bergmann T, Flusin O, et al. Autochthonous case

687   of dengue in France, October 2013. Eurosurveillance. 2013; doi:10.2807/1560-

688   7917.ES2013.18.50.20661

689   32.   Delisle E, Rousseau C, Broche B, Leparc-Goffart I, L'Ambert G, Cochet A, et al.

690   Chikungunya outbreak in Montpellier, France, September to October 2014. Euro Surveill.

691   2015; doi:10.2807/1560-7917.ES2015.20.17.21108

692   33.   Calba C, Guerbois-Galla M, Franke F, Jeannin C, Auzet-Caillaud M, Grard G, et al.

693   Preliminary report of an autochthonous chikungunya outbreak in France, July to September

694   2017. Eurosurveillance. 2017; doi:10.2807/1560-7917.ES.2017.22.39.17-00647

695   34.   Bouri N, Sell TK, Franco C, Adalja AA, Henderson DA, Hynes NA. Return of epidemic

696   dengue in the United States: Implications for the public health practitioner. Public Health

697   Rep. 2012; doi:10.1177/003335491212700305

698   35.   Joshua-Tor L. The argonautes. Cold Spring Harbor Symposia on Quantitative Biology. 2006.

699   doi:10.1101/sqb.2006.71.048

700   36.   Innan H, Kondrashov F. The evolution of gene duplications: Classifying and distinguishing

701   between models. Nature Reviews Genetics. 2010. doi:10.1038/nrg2689

702   37.   Stone SS, Haldar JP, Tsao SC, Hwu W -m. W, Sutton BP, Liang Z-P. Accelerating

703   advanced MRI reconstructions on GPUs. J Parallel Distrib Comput. 2008;68: 1307–1318.

704   doi:10.1016/j.jpdc.2008.05.013

705   38.   Yan KS, Yan S, Farooq A, Han A, Zeng L, Zhou M-M. Structure and conserved RNA binding

706   of the PAZ domain. Nature. 2003; doi:10.1038/nature02129

707   39.   Song JJ, Smith SK, Hannon GJ, Joshua-Tor L. Crystal structure of argonaute and its

708   implications for RISC slicer activity. Science (80-). 2004; doi:10.1126/science.1102514

709   40.   Cora E, Pandey RR, Xiol J, Taylor J, Sachidanandam R, McCarthy AA, et al. The MID-PIWI

710   module of Piwi proteins specifies nucleotide- and strand-biases of piRNAs. RNA. 2014;

27

711   doi:10.1261/rna.044701.114

712   41.   Stein CB, Genzor P, Mitra S, Elchert AR, Ipsaro JJ, Benner L, et al. Decoding the 5′
713   nucleotide bias of PIWI-interacting RNAs. Nat Commun. Springer US; 2019;10: 828.
714   doi:10.1038/s41467-019-08803-z

715   42.   McDonald JH, Kreitman M. Adaptive protein evolution at the Adh locus in *Drosophila*.
716   Nature. 1991; doi:10.1038/351652a0

717   43.   Pischedda E, Scolari F, Valerio F, Carballar-Lejarazú R, Catapano PL, Waterhouse RM, et
718   al. Insights Into an Unexplored Component of the Mosquito Repeatome: Distribution and
719   Variability of Viral Sequences Integrated Into the Genome of the Arboviral Vector *Aedes*
720   *albopictus*. Front Genet. 2019; doi:10.3389/fgene.2019.00093

721   44.   Schirle NT, MacRae IJ. The crystal structure of human argonaute2. Science (80-). 2012;
722   doi:10.1126/science.1221551

723   45.   Matsumoto N, Nishimasu H, Sakakibara K, Nishida KM, Hirano T, Ishitani R, et al. Crystal
724   Structure of Silkworm PIWI-Clade Argonaute Siwi Bound to piRNA. Cell. 2016;
725   doi:10.1016/j.cell.2016.09.002

726   46.   Macias V, Coleman J, Bonizzoni M, James AA. piRNA pathway gene expression in the
727   malaria vector mosquito *Anopheles stephensi*. Insect Mol Biol. 2014;23: 579–586.
728   doi:10.1111/imb.12106

729   47.   Bonizzoni M, Dunn WA, Campbell CL, Olson KE, Marinotti O, James AA. Complex
730   Modulation of the Aedes aegypti Transcriptome in Response to Dengue Virus Infection.
731   PLoS One. 2012; doi:10.1371/journal.pone.0050512

732   48.   Campbell CL, Keene KM, Brackney DE, Olson KE, Blair CD, Wilusz J, et al. *Aedes aegypti*
733   uses RNA interference in defense against Sindbis virus infection. BMC Microbiol. 2008;8: 1–
734   12. doi:10.1186/1471-2180-8-47

735   49.   Hess AM, Prasad AN, Ptitsyn A, Ebel GD, Olson KE, Barbacioru C, et al. Small RNA
736   profiling of Dengue virus-mosquito interactions implicates the PIWI RNA pathway in anti-
737   viral defense. BMC Microbiol. 2011; doi:10.1186/1471-2180-11-45

738   50.   Chen X-G, Jiang X, Gu J, Xu M, Wu Y, Deng Y, et al. Genome sequence of the Asian Tiger
739   mosquito, *Aedes albopictus* , reveals insights into its biology, genetics, and evolution. Proc
740   Natl Acad Sci. 2015; doi:10.1073/pnas.1516410112

741   51.   La Ruche G, Souarès Y, Armengaud A, Peloux-Petiot F, Delaunay P, Desprès P, et al. First
742   two autochthonous dengue virus infections in metropolitan France, september 2010.
743   Eurosurveillance. 2010; doi:19676 [pii]

744   52.   Schuffenecker I, Iteman I, Michault A, Murri S, Frangeul L, Vaney MC, et al. Genome
745   microevolution of chikungunya viruses causing the Indian Ocean outbreak. PLoS Med.
746   2006; doi:10.1371/journal.pmed.0030263

747   53.   Dubrulle M, Mousson L, Moutailier S, Vazeille M, Failloux AB. Chikungunya virus and *Aedes*

748   mosquitoes: Saliva is infectious as soon as two days after oral infection. PLoS One. 2009;

749   doi:10.1371/journal.pone.0005895

750 54.   Campbell CL, Black IV WC, Hess AM, Foy BD. Comparative genomics of small RNA

751   regulatory pathway components in vector mosquitoes. BMC Genomics. 2008;9.

752   doi:10.1186/1471-2164-9-425

753 55.   Arensburger P, Hice RH, Wright JA, Craig NL, Atkinson PW. The mosquito *Aedes aegypti*

754   has a large genome size and high transposable element load but contains a low proportion

755   of transposon-specific piRNAs. BMC Genomics. 2011;12. doi:10.1186/1471-2164-12-606

756 56.   Yuan JS, Burris J, Stewart NR, Mentewab A, Neal CN. Statistical tools for transgene copy

757   number estimation based on real-time PCR. BMC Bioinformatics. 2007;8: 1–12.

758   doi:10.1186/1471-2105-8-S7-S6

759 57.   Baruffi L, Damiani G, Guglielmino CR, Bandii C, Malacrida AR, Gasperi G. Polymorphism

760   within and between populations of ceratitis: Comparison between RAPD and multilocus

761   enzyme electrophoresis data. Heredity (Edinb). 1995;74: 425–437. doi:10.1038/hdy.1995.60

762 58.   Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput.

763   Nucleic Acids Res. 2004; doi:10.1093/nar/gkh340

764 59.   Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.

765   2013; Available: https://arxiv.org/abs/1303.3997

766 60.   Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing -- Free

767   bayes -- Variant Calling -- Longranger. arXiv Prepr arXiv12073907. 2012;

768   doi:arXiv:1207.3907 [q-bio.GN]

769 61.   Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating

770   and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome

771   of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012;

772   doi:10.4161/fly.19695

773 62.   Schroeder MP, Lopez-Bigas N. muts-needle-plot: Mutations Needle Plot v0.8.0. 2015;

774   doi:10.5281/ZENODO.14561

775 63.   Oliveros JC. Venny. An interactive tool for comparing lists with Venn's diagrams. 2007--

776   2015. h ttp. bioinfogp cnb csic es/tools/venny/index html Accessed. 2016;

777 64.   Flot JF. Seqphase: A web tool for interconverting phase input/output files and fasta

778   sequence alignments. Mol Ecol Resour. 2010; doi:10.1111/j.1755-0998.2009.02732.x

779 65.   Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction

780   from population data. Am J Hum Genet. 2001; doi:10.1086/319501

781 66.   Stephens M, Scheet P. Accounting for Decay of Linkage Disequilibrium in Haplotype

782   Inference and Missing-Data Imputation. Am J Hum Genet. 2005; doi:10.1086/428594

783 67.   Librado P, Rozas J. DnaSP v5: A software for comprehensive analysis of DNA

784   polymorphism data. Bioinformatics. 2009; doi:10.1093/bioinformatics/btp187

785 68.   Tajima F. Evolutionary relationship of DNA sequences in finite populations. Genetics. 1983;

786 69.   Watterson GA. On the number of segregating sites in genetical models without

787   recombination. Theor Popul Biol. 1975; doi:10.1016/0040-5809(75)90020-9

788 70.   Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA

789   polymorphism. Genetics. 1989;

790 71.   Abascal F, Zardoya R, Telford MJ. TranslatorX: Multiple alignment of nucleotide sequences

791   guided by amino acid translations. Nucleic Acids Res. 2010; doi:10.1093/nar/gkq291

792 72.   Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: Improving the sensitivity of

793   progressive multiple sequence alignment through sequence weighting, position-specific gap

794   penalties and weight matrix choice. Nucleic Acids Res. 1994; doi:10.1093/nar/22.22.4673

795 73.   Yang Z. PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;

796   doi:10.1093/molbev/msm088

797 74.   Xu B, Yang Z. PamlX: A graphical user interface for PAML. Mol Biol Evol. 2013;

798   doi:10.1093/molbev/mst179

799 75.   Rimmer A, Phan H, Mathieson I, Iqbal Z, Twigg SRF, Wilkie AOM, et al. Integrating

800   mapping-, assembly- and haplotype-based approaches for calling variants in clinical

801   sequencing applications. Nat Genet. 2014; doi:10.1038/ng.3036

802 76.   Lai Z, Markovets A, Ahdesmaki M, Chapman B, Hofmann O, Mcewen R, et al. VarDict: A

803   novel and versatile variant caller for next-generation sequencing in cancer research. Nucleic

804   Acids Res. 2016; doi:10.1093/nar/gkw227

805 77.   McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The

806   Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA

807   sequencing data. Genome Res. 2010; doi:10.1101/gr.107524.110

808 78.   Rstudio Team. RStudio: Integrated Development for R. [Online] RStudio, Inc., Boston, MA.

809   2016. doi:10.1007/978-81-322-2340-5

810 79.   Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J

811   Mol Biol. 1990; doi:10.1016/S0022-2836(05)80360-2

812 80.   Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, et al. A Completely

813   Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. J Mol

814   Biol. 2018; doi:10.1016/j.jmb.2017.12.007

815 81.   Webb B, Sali A. Protein structure modeling with MODELLER. Methods in Molecular Biology.

816   2017. doi:10.1007/978-1-4939-7231-9_4

817 82.   Schirle NT, Sheu-Gruttadauria J, Chandradoss SD, Joo C, MacRae IJ. Water-mediated

818   recognition of t1-adenosine anchors Argonaute2 to microRNA targets. Elife. 2015;

819   doi:10.7554/elife.07646

820 83.   Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot . Acta

821   Crystallogr Sect D Biol Crystallogr. 2010; doi:10.1107/s0907444910007493

822   84.   Adams PD, Afonine P V., Bunkóczi G, Chen VB, Echols N, Headd JJ, et al. The Phenix
823   software for automated determination of macromolecular structures. Methods. 2011.
824   doi:10.1016/j.ymeth.2011.07.005
825   85.   Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, et al. MolProbity:
826   More and better reference data for improved all-atom structure validation. Protein Sci. 2018;
827   doi:10.1002/pro.3330
828   86.   Benkert P, Künzli M, Schwede T. QMEAN server for protein model quality estimation.
829   Nucleic Acids Res. 2009; doi:10.1093/nar/gkp322
830   87.   Schrödinger L. The PyMOL molecular graphics system, version 1.8.
831   https://www.pymol.org/citing. 2015;
832   88.   Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al.
833   Integrative genomics viewer. Nature Biotechnology. 2011. doi:10.1038/nbt.1754
834   89.   A. Reynolds J, Poelchau MF, Rahman Z, Armbruster PA, Denlinger DL. Transcript profiling
835   reveals mechanisms for lipid conservation during diapause in the mosquito, *Aedes*
836   *albopictus.* J Insect Physiol. 2012; doi:10.1016/j.jinsphys.2012.04.013
837   90.   Khan A, Rayner GD. Robustness to non-normality of common tests for the many-sample
838   location problem. J Appl Math Decis Sci. 2004; doi:10.1155/S1173912603000178
839   91.   Blanca MJ, Alarcón R, Arnau J, Bendayan R. Non-normal data: Is ANOVA still a valid
840   option? Psicothema. 2017; doi:10.7334/psicothema2016.383
841   92.   Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. ConSurf 2010: Calculating evolutionary
842   conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res.
843   2010; doi:10.1093/nar/gkq399
844   93.   Madeira F, Park, Y M, Lee J, Buso N, Gur T, Madhusoodanan N, et al. The EMBL-EBI
845   search and sequence analysis tools APIs in 2019. Nucleic Acids Res. 2019;
846   94.   Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript
847   server. Nucleic Acids Res. 2014; doi:10.1093/nar/gku316
848
849

SUPPORTING INFORMATION

851 **S1 Table.** List of the core components of the piRNA pathway in *Ae. aegypti* and their orthologous

852 in *Ae. albopictus*.

853 **S2 Table.** List of Transcript IDs and abbreviations of the Culicidae and Drosophilidae species

854 included in the phylogenetic analyses.

855 **S3 Table.** Number of non-synonymous mutations found in mosquitoes of the Foshan strain

856 (Foshan) and wild-caught samples from Mexico (Mex) and the island of La Reunion (Reu) divided

857 by type (i.e. missense [M] , frameshift [F], indel [I] and nonsense [N]) and number of sites in which

858 **S4 Table.** Relative expression values (log10 fold-change) of Piwi genes during development (A)

859 and following viral infection (B) normalized with respect to sugar-fed samples. Samples (2 pools

860 per condition, 15 individuals each) were analysed at 4 days post infection (early infection) and at

861 14 and 21 days post infection for CHIKV and DENV, respectively (late infection). Each condition

862 was normalized to the corresponding sugar-fed control and compared to the corresponding Blood-

863 fed control. Ovaries and carcasses were analysed independently. * indicates statistically significant

864 difference between infected and non-infected blood-fed samples (ANOVA framework). Relative

865 expression values may mask differences in levels of expression. For instance, the Ct values of

866 Piwi6, Piwi7 and Piwi1/3 in ovaries 4 days post infection with CHIKV were 30, 33.39 and 25.20,

867 respectively. Ovaries of blood-fed samples at the same time point showed Ct values of 30.30,

868 33.93 and 26.55 for Piwi6, Piwi7 and Piwi1/3. When relative expression was calculated with

869 respect to Ct values of RPL34, fold-changes in gene expression were comparable among the three

870 genes in both conditions, but Ct values clearly indicate that Piwi7 is less expressed than both

871 Piwi1/3 and Piwi6. These considerations were taken into account when describing results.

872 **S5 Table.** List of primers used for CDS analyses, copy number estimation, qPCR experiments and

873 Northern Blot probe design.

874 **S1 Dataset.** CDS of the seven Piwi genes of *Ae. albopictus*. The sequence of the PAZ, MID and

875 PIWI domains is in bold, underline and bold-italics, respectively.

876 **S1 Fig.** Maximum likelihood cladogram generated from the alignment 862 of transcript sequences

877 of annotated Piwi genes in Culicidae. Transcript IDs and species abbreviations are as listed in S2

878   Table. AlbPiwi3 is the same as Piwi1/3 in the text. Piwi gene transcripts from *Ae. albopictus* are in

879   red, from *Ae. aegypti* in purple, from *Culex quinquefasciatus* in pink. Transcripts from *D.*

880   *melanogaster* Ago3, Piwi and Aubergine genes are included for reference and shown in blue. All

881   nodes were supported by bootstrap values higher than 50% with the exception of the three nodes

882   with a black dot.

883   **S2 Fig.** Polymorphism of Piwi4 and Piwi5. Lollipop plots representing position, amount and type

884   of mutation along the coding sequences of Piwi4 and Piwi5 in mosquitoes of the Foshan strain,

885   from la Reunion Island (Reu) and Mexico (Mex) as inferred by Freebayes and SnpEFF analyses.

886   Only missense (blue), nonsense (red) and indels (orange) and frameshift (yellow) are shown. The

887   PAZ, MID and PIWI domains are shown in green, blue and magenta, respectively. DDH residues

888   positions are highlighted in the PIWI domain.

889   **S3 Fig.** Sequence alignment of *Aedes albopictus* Piwi proteins. Domain boundaries inferred from

890   structural predictions are highlighted by coloured lines using the same colour coding as in figure 4

891   (Orange: N-terminus; Green: PAZ; Blue: MID; Magenta: PIWI). Conserved DDH residues found in

892   PIWI are indicated by (▲). The "acc" line indicates relative solvent accessibility, ranging from blue

893   (accessible) to white (buried). The sequence alignment was generated using EBI muscle [93] and

894   depicted using ESPRIPT3 [94] .

895   **S4 Fig**. Per-site read depth of the sequenced libraries of mosquitoes from Tapachula and Tampon.

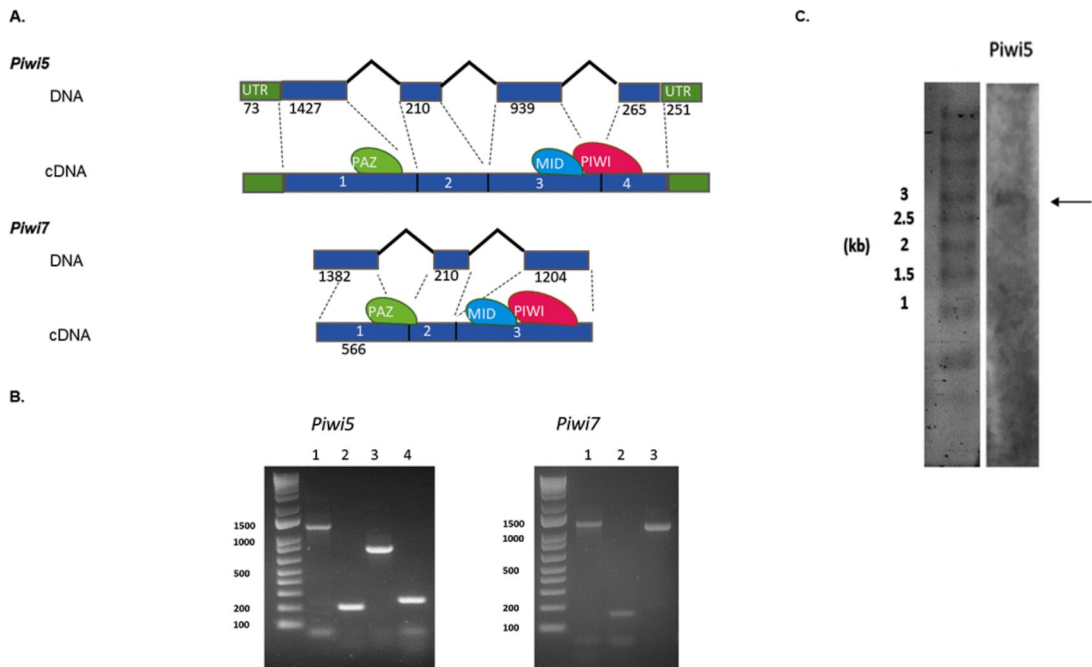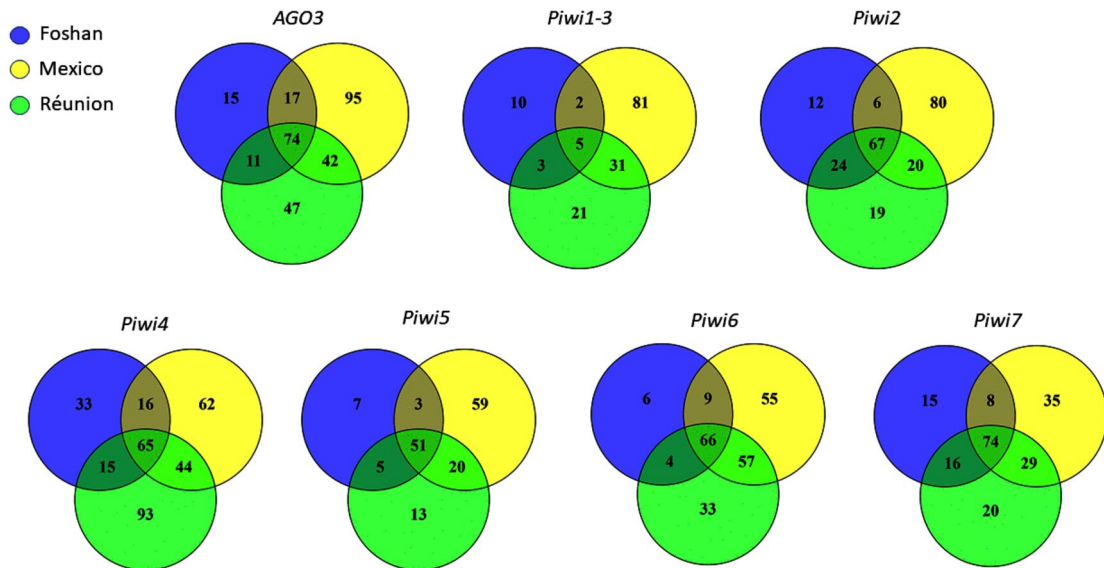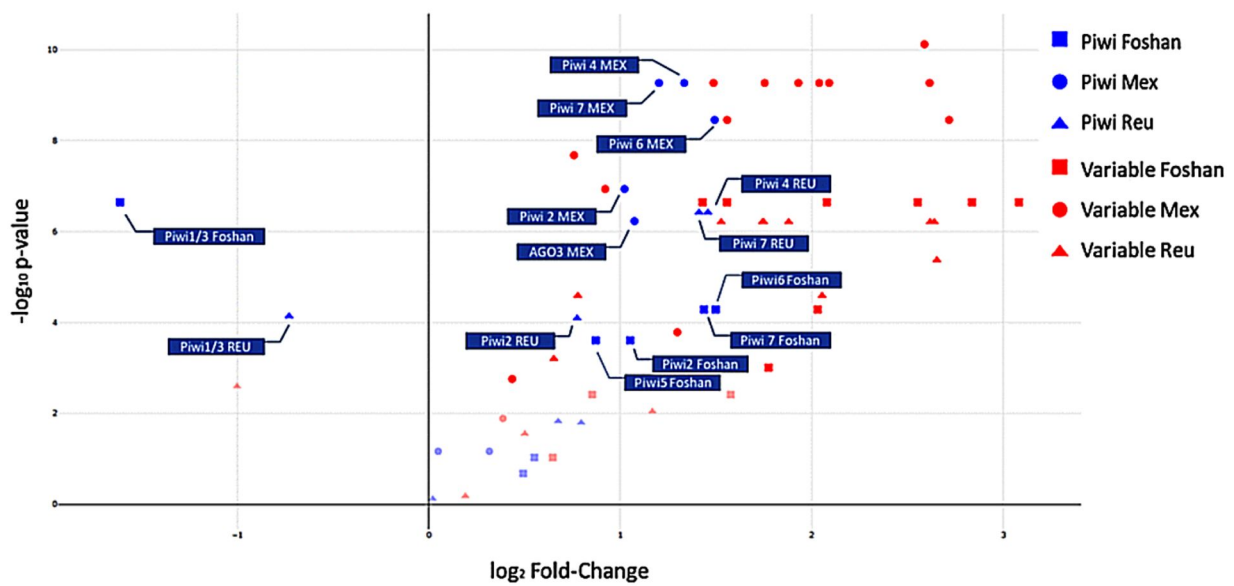896   **S5 Fig**. Northern blot results for Piwi5 and Piwi7.

33

Figure 1

A.

B.

C.

34

Figure 2

Figure 3

Figure 4

Figure 5