PLOS | NEGLECTED TROPICAL DISEASES

RESEARCH ARTICLE

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

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

1 Department of Biology and Biotechnology, University of Pavia, Pavia, Italy, **2** Arbovirus and Insect Vectors Units, Department of Virology, Institut Pasteur, Paris, France, **3** Verily Life Sciences, South San Francisco, California, United States of America

¤ Current address: Department of Molecular Biology and Biochemistry, University of California, Irvine, California, United States of America
* mariangela.bonizzoni@unipv.it

OPEN ACCESS

Citation: Marconcini M, Hernandez L, Iovino G, Houé V, Valerio F, Palatini U, et al. (2019) Polymorphism analyses and protein modelling inform on functional specialization of *Piwi* clade genes in the arboviral vector *Aedes albopictus*. PLoS Negl Trop Dis 13(12): e0007919. https://doi.org/10.1371/journal.pntd.0007919

Editor: Fabiano Oliveira, National Institutes of Health, UNITED STATES

Received: July 1, 2019

Accepted: November 11, 2019

Published: December 2, 2019

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files

Funding: This research was funded by a European Research Council Consolidator Grant (ERCCoG) under the European Union's Horizon 2020 Programme (Grant Number ERCCoG 682394) to M.B.; by the Italian Ministry of Education, University and Research FARE-MIUR project R1623HZAH5 to M.B.; by the Italian Ministry of

Abstract

Current knowledge of the piRNA pathway is based mainly on studies on *Drosophila melanogaster* where three proteins of the Piwi subclade of the Argonaute family interact with PIWI-interacting RNAs to silence transposable elements in gonadal tissues. In mosquito species that transmit epidemic arboviruses such as dengue and chikungunya viruses, *Piwi* clade genes underwent expansion, are also expressed in the soma and cross-talk with proteins of recognized antiviral function cannot be excluded for some Piwi proteins. These observations underscore the importance of expanding our knowledge of the piRNA pathway beyond the model organism *D. melanogaster*. Here we focus on the emerging arboviral vector *Aedes albopictus* and we couple traditional approaches of expression and adaptive evolution analyses with most current computational predictions of protein structure to study evolutionary divergence among Piwi clade proteins. Superposition of protein homology models indicate possible high structure similarity among all Piwi proteins, with high levels of amino acid conservation in the inner regions devoted to RNA binding. On the contrary, solvent-exposed surfaces showed low conservation, with several sites under positive selection. Analysis of the expression profiles of *Piwi* transcripts during mosquito development and following infection with dengue serotype 1 or chikungunya viruses showed a concerted elicitation of all *Piwi* transcripts during viral dissemination of dengue viruses while maintenance of infection relied on expression of primarily *Piwi5*. Opposite, establishment of persistent infection by chikungunya virus is accompanied by increased expression of all *Piwi* genes, particularly *Piwi4* and, again, *Piwi5*. Overall these results are consistent with functional specialization and a general antiviral role for *Piwi5*. Experimental evidences of sites under positive selection in *Piwi1/3*, *Piwi4* and *Piwi6*, that have complex expression profiles, provide useful knowledge to design tailored functional experiments.

Education, University and Research (MIUR): Dipartimenti Eccellenza Program (2018–2022) Dept. of Biology and Biotechnology "L. Spallanzani", University of Pavia. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Competing interests: The authors have declared that no competing interests exist. Jacob E. Crawford, Bradley White and Teresa Lin are employed by a commercial company, Verily Life Sciences LLC. They have no competing interests.

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 development, 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].

Current knowledge on the piRNA pathway in insects is based mainly on studies on *Drosophila melanogaster* where three proteins of the Piwi subclade, namely Argonaute-3 (AGO3), PIWI and Aubergine (AUB), interact with piRNAs to silence transposable elements (TEs) in gonadal tissues [14]. Interestingly, the piRNA pathway of *D. melanogaster* does not have antiviral activity and no viral integrations have been detected [15]. Additional differences exist between the piRNA pathway of *D. melanogaster* and that of mosquitoes, suggesting that *D. melanogaster* cannot be used as a model to unravel the molecular cross-talk between the siRNA and piRNA pathways leading to antiviral immunity in *Aedes* spp. mosquitoes. For instance, in *Aedes aegypti*, Piwi subclade has undergone expansion with seven proteins (i.e. Ago3, Piwi2, Piwi3, Piwi4, Piwi5, Piwi6 and Piwi7), which are alternatively expressed in somatic and germline cells and interact with both endogenous and vpiRNAs [12,16,17]. Gonadal- or embryonic-specific expression is found for *Piwi1/3* and *Piwi7*, respectively [16]. On the contrary, *Ago3*, *Piwi4*, *Piwi5* and *Piwi6* are highly expressed in *Ae. aegypti* soma and Aag2 cells and all contribute to the production of transposon-derived piRNAs [16,18]. Ago3 and Piwi5 also regulate biogenesis of piRNAs from the replication-dependent histone gene family [19]. Production of vpiRNAs is dependent on Piwi5 and Ago3 during infection of Aag2 cells with the *Alphavirus* CHIKV, Sindbis and Semliki Forest (SF) viruses, but relies also on Piwi6 following infection with the *Flavivirus* DENV2 [18,20–22]. Piwi4 does not bind piRNAs and its knock-down does not alter vpiRNA production upon infection of Aag2 cells with either SFV or DENV2 [18,23]. On the contrary Piwi4 coimmunoprecipate with Ago2, Dcr2, Piwi5, Piwi6 and Ago3, suggesting a bridging role between the siRNA and piRNA pathways [21]. These studies support an antiviral role for Piwi proteins in *Aedes* spp. mosquitoes but given the number of *Piwi* genes in these species, it is a challenge to uncover their distinct physiological roles, if any. In duplicated genes, the presence of sites under positive selection is usually a sign of the acquisition of novel functions [24]. Additionally, under the "arm-race theory", rapid intraspecific evolution is expected for genes with immunity functions because their products should act against fast evolving viruses [25].

Besides being an important basic question, the understanding of functional divergence among Piwi proteins has applied perspectives for the development of novel genetic-based methods of transmission-blocking vector control strategies.

In recent years, the Asian tiger mosquito *Aedes albopictus* has emerged as a novel global arboviral threat. This species is a competent vector for a number of arboviruses, such as chikungunya (CHIKV), dengue (DENV), yellow fever (YFV) and Zika (ZIKV) viruses and is now present in every continent except Antarctica following its quick spread out of its native home range of South East Asia [26]. Establishment of *Ae. albopictus* in temperate regions of the world fostered the re-emergence or the new introduction of arboviruses [27]. For instance, chikungunya outbreaks occurred in Italy in 2007 and 2017 [28,29]; France and Croatia suffered from autochthonous cases of dengue and chikungunya in several occasions since 2010 [30–33] and dengue is remerging in some regions of the United States due to the presence of *Ae. albopictus* [34]. Knowledge on *Ae. albopictus* biology and the molecular mechanisms underlying its competence to arboviruses are still limited in comparison to *Ae. aegypti* despite its increasing public-health relevance.

Here we elucidate the molecular organization, polymorphism and expression of *Piwi* clade genes of *Ae. albopictus* in an evolutionary framework using a combination of molecular, population genomics and computational protein modelling approaches. We show that the genome of *Ae. albopictus* harbours seven *Piwi* genes, namely *Ago3*, *Piwi1/3*, *Piwi2*, *Piwi4*, *Piwi5*, *Piwi6* and *Piwi7*. For the first time in mosquitoes, we show sign of adaptive evolution in *Piwi1/3*,

Piwi4, *Piwi5* and *Piwi6*, including sites in the MID and PAZ domains. Additionally, expression profiles during mosquito development and following infection with the dengue or chikungunya viruses support functional specialization of Piwi proteins, with a prominent and general antiviral role for the transcript of *Piwi5*.

Results

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

Bioinformatic analyses of the current genome assemblies of *Ae. albopictus* (AaloF1) and the C6/36 cell line (canu_80X_arrow2.2), followed by copy number validation, confirmed the presence of seven *Piwi* genes (i.e. *Ago3*, *Piwi1/3*, *Piwi2*, *Piwi4*, *Piwi5*, *Piwi6 and Piwi7*) in *Ae. albopictus* (S1 Table). Genomic DNA sequences were obtained for each exon-intron boundaries, confirming in all *Piwi* genes the presence of the PAZ, MID and PIWI domains, the hallmarks of the Piwi subfamily of Argonaute proteins [35]. For *Ago3*, *Piwi1/3*, *Piwi2*, *Piwi4* and *Piwi6*, single transcript sequences that correspond to predictions based on the identified DNA sequences were retrieved (S1 Dataset). Sequencing results of the transcript from *Piwi5* showed a sequence 27 bp shorter than predicted on the reference genome, due to a 45bp gap followed by a 18b insertion, 110 and 333 bases after the ATG starting codon, respectively. This transcript still includes the PAZ, MID and PIWI domains. The presence of this transcript was further validated by northern-blot (Fig 1). For *Piwi7*, the transcript sequence also appears shorter than predicted (Fig 1). Alignment and phylogenetic analyses, in the context of currently annotated *Piwi* transcripts of Culicinae and Anophelinae mosquitoes, confirmed one-to-one orthologous pairing between *Ae. albopictus Piwi* gene transcripts and those of Ae. aegypti (S2 Table,

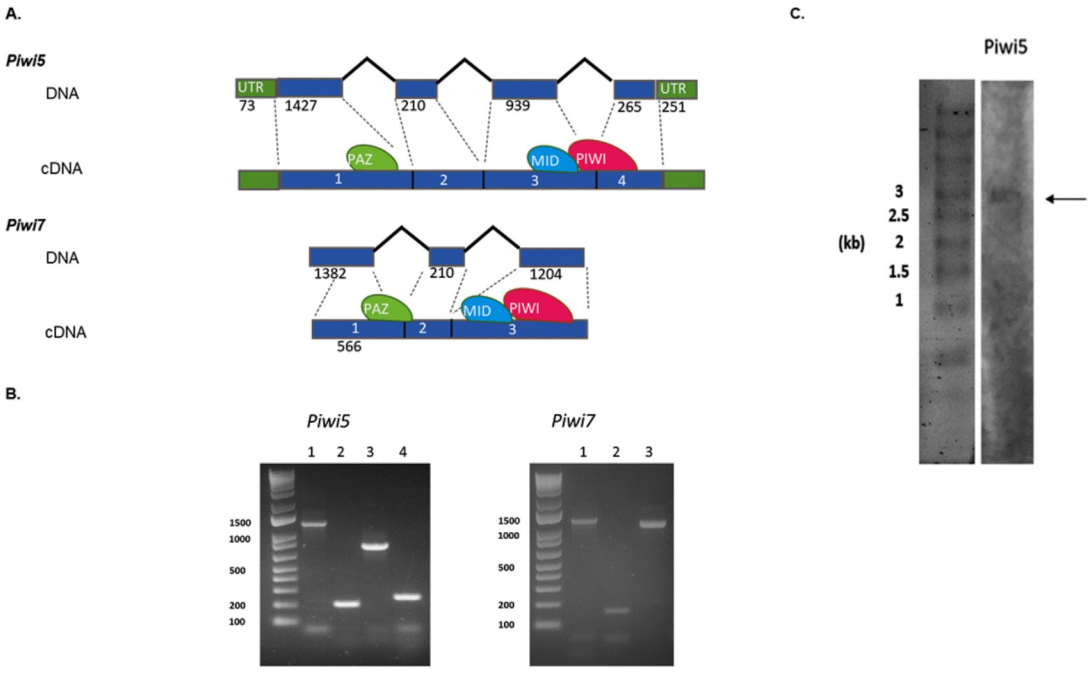

Fig 1. Gene and transcript structure of *Ae. albopictus Piwi5* and *Piwi7*. A) Schematic representation of the DNA structure of *Piwi5* and *Piwi7* genes and their corresponding transcripts as obtained from cDNA amplification of single sugar-fed mosquito samples. Exons and introns are shown by blue boxes and black lines, respectively, with corresponding length in nucleotide below each. The positions of the predicted PAZ, MID and PIWI domains are shown by green, blue and magenta ovals, respectively. Exon numbers correspond to lane numbers. B) Amplification of each exon of Piwi5 and Piwi7 on genomic DNA. Exon numbers correspond to lane numbers. C) Northern-blot results of *Piwi5* indicate the presence of a transcript of 3 kb.

https://doi.org/10.1371/journal.pntd.0007919.g001

S1 Fig). Interestingly, *Piwi5*, *Piwi6* and *Piwi7* transcripts group together and appear more similar to one of the two Aubergine-like transcripts annotated in different Anophelinae species than to *Aedes Piwi2*, *Piwi1/3* and *Piwi4* transcripts. Regarding the latter, *Piwi2* and *Piwi1/3* form a species-specific clade, rather than follow a speciation pattern. Rather, the two genes, which on *Ae. aegypti* map on chromosome 1 and are ~20 kb apart [17], may have originated as a duplication in the ancestor of *Ae. aegypti* and *Ae. albopictus* and be subsequently undergoing interlocus gene conversion. This mechanism causes nonreciprocal recombination, whereby one locus (i.e. part of a gene copy) replaces the homologous sequence of the other copy. The result is concerted evolution of the gene duplicates [36], which in this case eliminates divergence between *Piwi2* and *Piwi1/3* within each species.

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

Across *Drosophila* phylogeny, genes of the piRNA pathway display elevated rates of adaptive evolution [37], with rapidly evolving residues not clustering at the RNA binding site, but being distributed across the proteins [3]. The RNA binding site is found within the PAZ domain, at the amino-terminal part of Piwi proteins [35,38]. On the opposite side, at the carboxyl terminus, the PIWI domain resides. The PIWI domain belongs to the RNase H family of enzymes and the catalytic site is formed by three conserved amino acids (usually aspartate-aspartate-glutamate, DDE or aspartate-aspartate-histidine, DDH) [35,39]. Between the PAZ and PIWI domains the MID domain resides. MID specifies strand- and nucleotide-biases of piRNAs, including their Uridine 5' bias [40,41]. To evaluate the selective pressures acting along these genes, we analysed the polymorphism pattern in *Ae. albopictus* samples from wild-collected populations and from the Foshan reference strain. Synonymous and non-synonymous mutations were found for each gene in all populations (Fig 2), with *Piwi1/3* displaying the lowest polymorphism (Table 1).

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

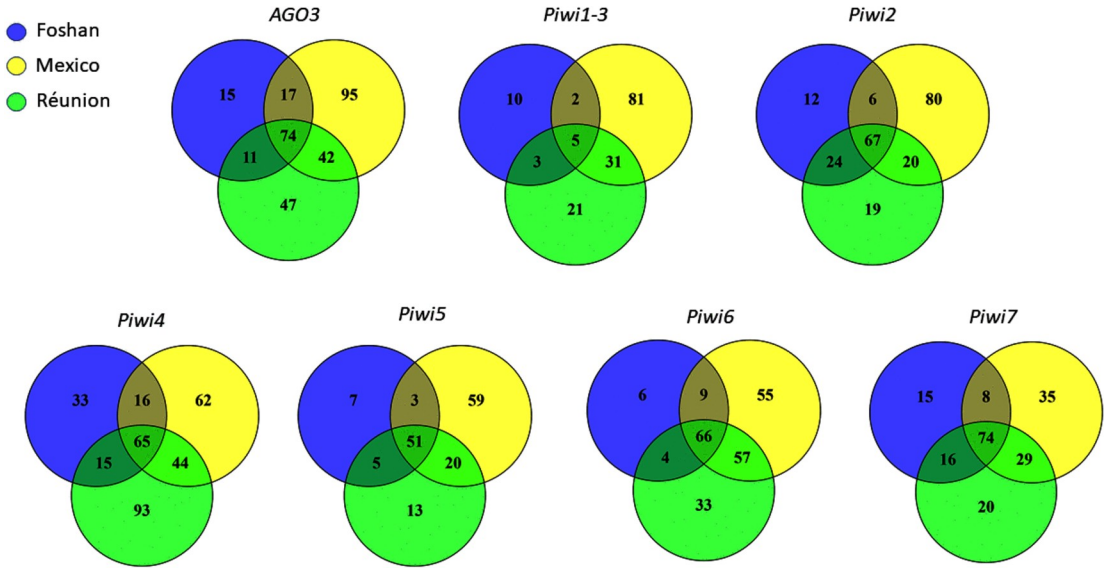

Fig 2. Venn diagrams showing the number of positions harbouring synonymous and non-synonymous mutations in tested samples for each *Piwi* gene.

https://doi.org/10.1371/journal.pntd.0007919.g002

Table 1. Polymorphism of *Aedes albopictus Piwi* genes in mosquitoes from the Foshan strain and wild-caught mosquitoes from La Reunion (Reu) and Mexico (Mex). We report the number of sequences (*n*), as well as the 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 sites (a) for each gene and population (and for the pooled sample).

	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

https://doi.org/10.1371/journal.pntd.0007919.t001

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 domains (S2A Fig), ten of these mutations were shared across all populations (S3 Table). The 5' region of *Piwi5* harboured several indels: two in-frame variants (i.e. 94_99del; 113_118del) were shared across all populations and were present in homozygosity in at least one sample (S2B Fig), suggesting that they are not detrimental. *Ago3* and *Piwi6* have very low non-synonymous nucleotide diversity, suggesting strong constraints at the protein level. The results of the McDonald-Kreitman test [42] further revealed an excess of non-synonymous substitutions compared to the polymorphism pattern in both *Piwi1/3* and *Piwi6*,

Table 2. Insights into Evolutionary divergence of *Piwi* genes in *Ae. albopictus*. 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) Output of Codeml with significant results regarding sites under positive selection.

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	-	-	-	-	

[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.

https://doi.org/10.1371/journal.pntd.0007919.t002

suggesting that they have been target of adaptive evolution (Table 2A). In contrast, *Piwi4* has a significant deficit of non-synonymous substitutions and/or excess of polymorphic non-synonymous segregating sites (Table 2A). In this gene, Tajima's D is negative but in line with the values of the other *Piwi* genes, and the high non-synonymous polymorphism may reflect selection of intraspecific diversifying selection, as expected in genes involved in immunity. Because positive selection may have acted at the level of very few sites, this not contributing to the gene-level non-synonymous substitution pattern; we explicitly tested models of codon evolution. Signs of positive selection were found at different sites, including one site in the Linker2 and one site in the MID domain of *Piwi1/3*, two sites in the PAZ domain of *Piwi4*, two sites in the Flex domain of *Piwi5* and three sites, two in the Flex and one in the Linker2 domains, of *Piwi6* (Table 2B). Haplotype reconstruction of our samples showed that these mutations can co-occur on the same gene, with the only exception of Y278D+H287P in *Piwi4* and A67P+G86S in *Piwi6*.

Finally, to gain insight on how variable *Piwi* genes are in comparison to slow- and fast -evolving genes of *Ae. albopictus*, we collected variability data of sets of genes previously identified to have slow and high evolutionary rates [43]. For each population, we compared the overall level of polymorphism (LoP) of the *Piwi* genes and of a dataset of fast-evolving genes (FGs) to that measured for a dataset of slow-evolving genes (SGs) as listed in the material and methods section "polymorphisms of *Piwi* genes" [43]. Our results indicate that *Piwi4*, *Piwi6* and

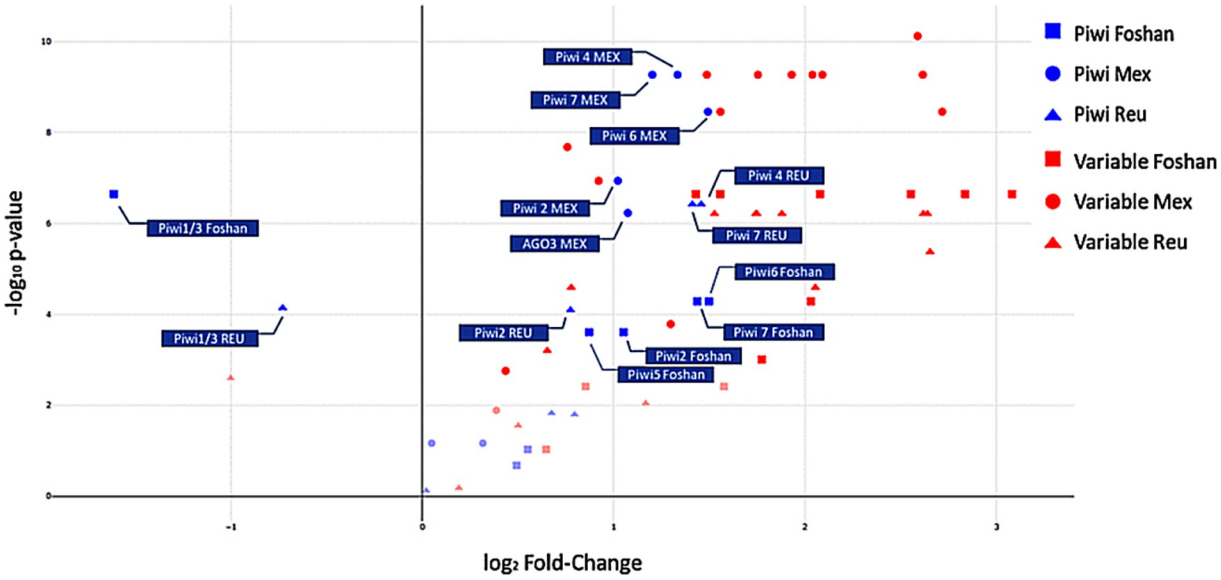

Fig 3. Volcano plot. Level of polymorphism (LoP) comparison between slow-evolving genes (SGs), fast-evolving genes (FGs) and *Piwi* genes by population. Genes on the right side of the panel have LoP values greater than those of SGs, while genes on the left side have smaller LoPs than SGs. The y-axis represents the -log10 p-values of the Kolmogorov-Smirnov test. Faint datapoints are not significant after Bonferroni correction for multiple testing (-log10 0.0024 (0.05/21 genes) = 2.62).

https://doi.org/10.1371/journal.pntd.0007919.g003

Piwi7 have LoP values comparable to those of FGs, while *Ago3* and *Piwi5* do not significantly deviate from the LoP values of SGs. *Piwi1*/3 appears to be conserved (Fig 3).

Computational predictions of molecular structures

The functional significance of the mutations under selection, as well as that of all the shared missense mutations in the PAZ and PIWI domains, was tested by computing predictions of three dimensional molecular structures of the Piwi proteins using the most-recent X-ray crystallography structure of Argonaute proteins as templates [44,45]. Homology modelling revealed high structural conservation among the seven Piwi proteins despite sequence heterogeneity (S2 Fig; Fig 4A).

Similarly, to *D. melanogaster*, the highest levels of amino acid sequence conservation were found in the regions that, based on homology modelling, define the inner pocket of Argonaute molecular assembly where the RNA binds. Significantly lower sequence conservation was found on the proteins surface (Fig 4B). Based on our computational predictions, we could not detect amino acidic polymorphisms that would affect RNA binding or processing, suggesting that all *Ae. albopictus* Piwi proteins may retain the Argonaute-like functions. Mapping of mutations under positive selection (Table 2B) on the homology models and sequence comparisons with known PIWI structural homologs showed that the identified variant amino acids are unlikely to induce severe alterations in protein folding. All mutant variants were found to localize in regions distant from the predicted RNA-binding and/or processing sites, ruling out possible effects associated to alterations in RNA recognition, but raising the intriguing possibility of regulatory roles during interactions with additional binding partners.

Developmental profile of *Ae. albopictus Piwi* genes

To further gain insights on the functional specialization of *Piwi* genes, we assessed their expression profile throughout mosquito development, namely at 4–8 hours (h) after

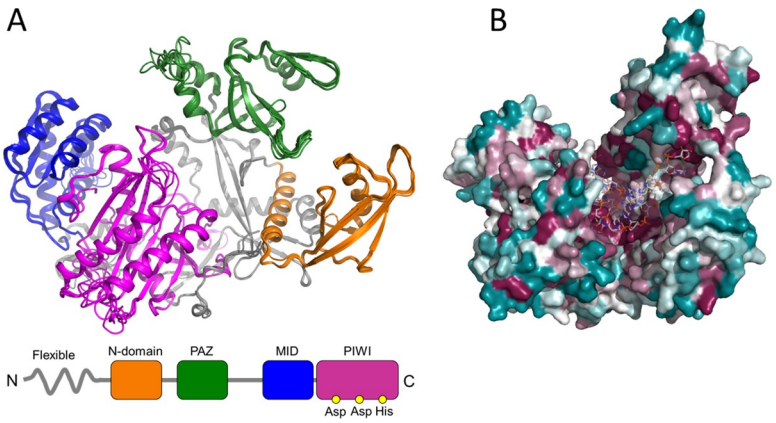

Fig 4. Computational homology models of the *Ae. Albopictus Piwi* proteins. Homology models were generated for the seven *Piwi* genes as described in the methods section. A) Superposition of cartoon representations of Piwi homology models, with highlight of domain organization: the N-terminal domain is shown in orange, the PAZ domain in green, the MID domain in blue and the PIWI domain in magenta. B) *CONSURF* [46]overview of the aminoacid sequence conservation mapped on three-dimensional homology models in a putative RNA-bound arrangement based on the structure of human Argonaute bound to a target RNA (PDB ID 4Z4D), colored from teal (very low conservation) to dark magenta (highly conserved).

https://doi.org/10.1371/journal.pntd.0007919.g004

deposition to capture the maternal-zygotic transition in expression, at late embryogenesis (i.e. 12–16 h and 16–24 h post deposition), at two time points during larval development (i.e. 1st and 4th instar larvae) and at pupal and adult stages (for the latter only we sampled separately males and females). Adult females were dissected to extract ovaries from the carcasses both from females kept on a sugar diet and 48 h after a blood meal, when a peak in *Piwi* gene expression was previously observed [47].

Expression levels of *Ago3*, *Piwi4*, *Piwi5*, *Piwi6* and *Piwi7* are at their peak in the embryonic stages, although at different time points (Fig 5A). Overall, *Ago3*, *Piwi1/3*, *Piwi2* and *Piwi6* have a similar trend during development showing a second peak of expression in adult females and their ovaries, while the expression levels of *Piwi4*, *Piwi5* and *Piwi7* remain constant. In details, *Piwi7* is mostly expressed 4-8h after deposition, *Piwi5* and *Piwi6* are mostly expressed after 8-16h and *Ago3* and *Piwi4* have their pick of expression at 16-24h. On the contrary, *Piwi1/3* and *Piwi2* are mostly expressed in ovaries extracted from blood-fed and sugar-fed females, respectively (Fig 5A, S4A Table). These results are consistent with lack of expression from published RNA-seq data from adult mosquitoes.

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 5B). 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

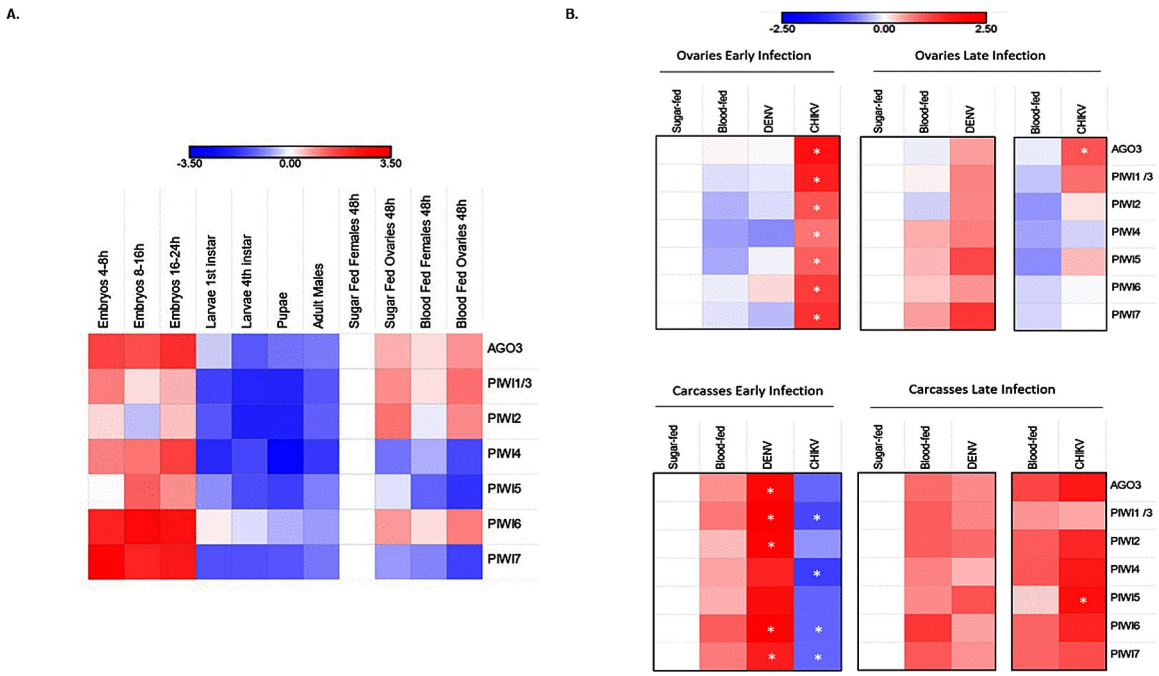

Fig 5. Expression profile of *Piwi* genes. Heatmap representations of log10 transformed fold-change expression values of each *Piwi* gene. A) Developmental expression pattern of the *Piwi* genes normalized on the expression in sugar fed females. B) Expression pattern of *Piwi* genes following viral infection 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.

https://doi.org/10.1371/journal.pntd.0007919.g005

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 folds by 14 dpi and the profile of *Piwi* genes changed. Expression in the ovaries decreased between 3 (*Piwi5*) to 20 (*Piwi7*) times with respect to values observed at 4 dpi, but remained higher than the corresponding expression values in ovaries of both sugar and blood-fed mosquitoes. In carcasses all *Piwi* genes inverted their expression pattern during the infection phase, increasing up to more than 100 times in the case of *Piwi4*, *Piwi5* and *Piwi6*. At 14 dpi, expression of the *Piwi* genes was highest in CHIKV-infected carcasses than in carcasses of sugar- and blood-fed mosquitoes.

For DENV, infection progresses differently than CHIKV. At 4 dpi there is no virus in the salivary gland, where the viral titer was measured at zero. By 21 dpi, DENV has established persistent infection [48]. At 4 dpi expression of *Piwi* genes was lower in DENV- and blood-fed ovaries than in ovaries of sugar-fed mosquitoes. The only exception was *Piwi6*, which was slightly up regulated in ovaries of DENV-infected samples, but slightly down-regulated in ovaried of blood-fed mosquitoes. On the contrary, at the same time point, carcasses of DENV-infected samples showed a drastic increase in the expression of all *Piwi* genes with respect to blood-fed samples; this increase was between 7 to 87 times for *Piwi7* and *Piwi2*, respectively. By 21 dpi, expression in the ovaries increased for all *Piwi* genes, in comparison to what observed both at 4dpi and in blood-fed ovaries, suggesting the increase in expression of *Piwi* genes is related to DENV dissemination. Interestingly, if we compare levels of expression in CHIKV-infected ovaries at 4 dpi and DENV-infected samples at 21 dpi, corresponding to the

time at which both viral species have disseminated throughout the mosquito body, we observe similar levels of fold-change expression of *Piwi4* and *Piwi7*, while *Ago3*, *Piwi1/3* and *Piwi6* show higher fold-change in CHIKV compared to DENV samples. Whether this trend is dependent on the viral species or viral titer requires further investigation. The same type of comparison in carcasses shows a higher fold-change expression level of all *Piwi* genes, particularly *Piwi1/3* and *Piwi5*, in DENV- versus CHIKV-infected samples, even if viral titers are lower for DENV (S4B Table). Overall these results support the hypothesis of a concerted activity of all PIWI proteins during viral dissemination for DENV, and maintenance of infection rely on expression of primarily *Piwi5*. On the contrary, establishment of persistent CHIKV infection was accompanied by elicitation of all Piwi gene expression, particularly *Piwi4* and, again, *Piwi5*.

Discussion

The piRNA pathway does not have antiviral immunity in *D. melanogaster* [15]. In the arboviral vectors *Aedes spp*. mosquitoes, vpiRNAs are found following infections with arboviruses, piRNAs are produced in the soma besides the germline and there has been an expansion on the number of *Piwi* genes, supporting the hypothesis that the piRNA pathway has antiviral immunity[12,49,50]. Besides *Ago3*, the genome of *Ae*. *aegypti* harbours six *Piwi* genes (i.e. *Piwi1/3*, *Piwi2*, *Piwi4*, *Piwi5*, *Piwi6*, *Piwi7)*, some of which show tissue and development-specific expression profile and have been preferentially associated with either TE-derived or viral piRNAs, [16,20,21]. These studies were based on the knowledge of the gene structure of each *Ae*. *aegypti Piwi* gene and the application of *ad hoc* RNAi-based silencing experiments and *in vitro* expression assays, but lack an evolutionary perspective [18–21].

In this work we focused on the emerging arboviral vector *Ae*. *albopictus* and we show how the application of evolutionary and protein modelling techniques helps to unravel functional specialization of Piwi proteins. The genome of *Ae*. *albopictus* harbours one copy of *Ago3* and six *Piwi* genes (i.e. *Piwi1/3*, *Piwi2*, *Piwi4*, *Piwi5*, *Piwi6 and Piwi7*), each a one-to-one orthologue to the *Ae*. *aegypti Piwi* genes. The only exceptions are *Piwi2* and *Piwi1/3*, where the two genes from the same species cluster together. In *Ae*. *aegypti*, these two genes both map on Chromosome 1, separated by ~ 20kb, suggesting they may undergo frequent gene conversion.

All transcripts retain the PAZ and PIWI domains, which are the hallmarks of the Argonaute protein family [35]. By using homology modelling, we obtained predictions of molecular architectures for *Ae*. *albopictus Ago3* and Piwi proteins, onto which we mapped the putative boundaries of each domain. Superpositions and sequence comparisons allowed clear identification of the catalytic DDH triad within the PIWI domain of all modelled proteins. This conservation is consistent with strong sequence matching in the putative RNA binding regions of the PIWI, PAZ and MID domains and suggests the possible maintenance of slicer activity, albeit experimental validation of each isoform is necessary.

The expression of all *Piwi* genes was confirmed throughout the developmental stages and the adult life of the mosquito, both in ovaries and somatic tissues. Interestingly, *Piwi7* transcript expression starkly drops following early embryogenesis, to the point that we could detect it neither in RNA-seq analyses, nor in Northern-blot experiments (S5 Fig). The expression of *Piwi* genes was elicited upon arboviral infection, indirectly confirming the antiviral role of the piRNA pathway. The expression profile of *Piwi* genes showed differences depending on both the species of infecting virus and on when the expression was measured. In CHIKV-infected samples, expression of *Piwi* genes was mostly elicited in ovaries or carcasses at 4 or 14 dpi, respectively. On the contrary, in DENV-infected samples, the highest expression *of Piwi* genes was seen in carcasses 4 dpi. These results are concordant with the timing in piRNAs

accumulation following CHIKV or DENV infection. In *Ae*. *albopictus* mosquitoes infected with CHIKV, secondary piRNAs are not found 3 dpi, but are enriched 9 dpi [9]. In contrast, in *Ae*. *aegypt*i mosquitoes infected with DENV2, piRNAs are the dominant small RNA populations 2 dpi [50].

Overall, these observations and our expression analyses support the hypothesis of an early activation of the piRNA pathway following DENV infection, but a late activation after CHIKV infection. Additionally, our expression analysis is consistent with a generalist antiviral role for *Piwi5*, which is elicited both during DENV and CHIKV infection [20], but suggest a more prominent role for *Piwi6* and *Piwi1/3* or *Piwi4* and *Ago3* during infection with DENV and CHIKV, respectively.

Materials and methods

Mosquitoes

Aedes albopictus mosquitoes of the Foshan strain were used in this study. This strain was established in 1981 in the Center for Disease Control and Prevention of Guangdong Province in China. It has been at the University of Pavia since 2013 [10,51]. Mosquitoes are reared under constant conditions, at 28˚C and 70–80% relative humidity with a 12/12h light/dark cycle. Larvae are reared in plastic containers, at a controlled density to avoid competition for food. Food is provided daily in the form of fish food (Tetra Goldfish Gold Colour). Adults are kept in 30 cm^3 cages and fed with cotton soaked in 0.2 g/ml sucrose as a carbohydrate source. Adult females are fed with defibrinated mutton blood (Biolife Italiana) using a Hemotek blood feeding apparatus. Mosquitoes from Mexico and La Reunion island were collected in 2017 as adults and maintained in ethanol 70% before shipment to Italy. All samples were processed at the University of Pavia.

Mosquito infections

Foshan mosquitoes were infected with DENV serotype 1, genotype 1806 or CHIKV 06.21. DENV-1 (1806) was isolated from an autochthonous case from Nice, France in 2010 [52]. CHIKV 06–21 was isolated from a patient on La Reunion Island in 2005 [53]. Both strains were kindly provided by the French National Reference Center for Arboviruses at the Institut Pasteur. CHIKV 06–21 and DENV-1 1806 were passaged twice on cells to constitute the viral stocks for experimental infections of mosquitoes, on C6/36 cells for CHIKV 06–21 and on African green monkey kidney Vero cells for DENV-1 1806. Viral titers of stocks were estimated by serial dilutions and expressed in focus-forming units (FFU)/mL.

Four boxes containing 60 one-week-old females were exposed to an infectious blood-meal composed by 2 mL of washed rabbit red blood cells, 1 mL of viral suspension and 5 mM of ATP. The titer of the blood-meal was 10^7 PFU/mL for CHIKV and $10^{6.8}$ PFU/mL for DENV. Fully engorged females were placed in cardboard boxes and fed with a 10% sucrose solution. Mosquitoes were incubated at 28˚C until analysis.

In parallel, mosquitoes were fed with uninfected blood-meal or kept on a sugar-diet and grown in the same conditions. Thirty mosquitoes were killed to be analyzed at days 4 and 14 post-infection (pi) for CHIKV, and at days 4 and 21 pi for DENV. To estimate transmission, saliva was collected from individual mosquitoes as described in [54]. After removing wings and legs from each mosquito, the proboscis was inserted into a 20 μL tip containing 5 μL of Fetal Bovine Serum (FBS) (Gibco, MA, USA). After 30 min, FBS containing saliva was expelled in 45 μL of Leibovitz L15 medium (Invitrogen, CA, USA) for titration. Transmission efficiency refers to the proportion of mosquitoes with infectious saliva among tested mosquitoes (which correspond to engorged mosquitoes at day 0 pi having survived until the day of examination).

The number of infectious particles in saliva was estimated by focus fluorescent assay on C6/36 *Ae*. *albopictus* cells. Samples were serially diluted and inoculated into C6/36 cells in 96-well plates. After incubation at 28˚C for 3 days (CHIKV) or 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.

At the same time points mosquitoes that had been fed a non-infectious blood or kept on a sugar diet were sampled and dissected as above.

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

The sequences of the *Ae*. *aegypti Piwi* genes [55] 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. Deduced coding sequences (CDS) where analysed in Prosite (Prosite.expasy.org/prosite.html) to screen for the typical PAZ and PIWI domains of Argonaute proteins [56].

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, on genomic DNA from four mosquitoes and using gene-specific primers, after having verified their efficiency (S4 Table). DNA was extracted using DNA Isolation DNeasy Blood & Tissue Kit (Qiagen). Estimates of gene copy number were performed based on the $2^{-\Delta CT}$ method using *Piwi6* and the para sodium channel genes (AALF000723) as references [57].

Structure of *Piwi* genes

DNA extracted from whole mosquitoes and dissected ovaries [58] was used as template in PCR amplifications to confirm the presence and the genome structure of each bioinformatically-identified *Piwi* gene. Primers were designed to amplify each exon, with particular attention to detect differences between paralogous *Piwi* genes (S1 Table). 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 visualized under UV light after gel electrophoresis using 1–1.5% agarose gels stained with ethidium bromide and a 100 bp or 1 kb molecular marker. PCR products were either directly sequenced or 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.

Piwi gene transcript sequences and phylogeny

RNA was extracted using a standard TRIzol protocol from pools of 5 adult female mosquitoes to verify the transcript sequence of each *Piwi* gene. Sets of primers were designed for each gene to amplify its entire transcript sequence (S4 Table). PCR reactions were performed using a High Fidelity taq-polymerase (Platinum SuperFi DNA Polymerase, Invitrogen) following manufacturer's instructions. PCR products were cloned using the TOPO TA Cloning Kit (Invitrogen) and plasmid DNA, purified using the QIAprep Spin Miniprep Kit, was sequenced. Rapid amplification of cDNA ends (RACE) PCRs were performed using First-Choice RLM-RACE Kit (Thermo Fisher Scientific) to analyse 5' and 3' ends of the transcript

sequences following manufacturer's instructions. Amplification products were cloned and sequenced as previously indicated.

Sequences of the identified *Ae. albopictus Piwi* gene transcripts were aligned to sequences of *Culicidae* and *D. melanogaster Piwi* transcripts, as downloaded from VectorBase (www.vectorbase.org), using MUSCLE [59]. Maximum-likelihood based phylogenetic inference was based on RAxML after 1000 bootstrap resampling of the original dataset. Phylogeny reconstruction was done through the CIPRESS portal (http://www.phylo.org/index.php/). Resulting tree was 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) [60] with custom parameters. Polymorphisms was tested by Freebayes [61]. Annotation of the detected mutations, as well counts of synonymous and non-synonymous variants, were performed in snpEff [62]. Frameshifts and non-synonymous variants were plotted using muts needle-plot [63]. Venn diagrams of positions with mutations in the three tested samples were built using Venny 2.1 [64]. Haplotype reconstruction was performed using seqPHASE [65] and PHASE [66,67]. The inferred haplotypes were analysed with DnaSP [68], which estimated the number of segregating sites and the level of nucleotide diversity π [69] in both synonymous and non-synonymous sites. We manually calculated, for synonymous and non-synonymous positions separately, the nucleotide diversity estimator theta [70] and Tajima's D statistic [71], which are a function of the total number of sites and the number of segregating sites (both estimated by DnaSP), and of sample size (see references for detailed formulas). We also tested for signatures of adaptive evolution using the McDonald-Kreitman test [42] (as implemented in DnaSP), which compares the rate of polymorphism and substitutions in synonymous and non-synonymous sites. For this analysis we used alignments that included the orthologous sequences from *Ae. aegypti*.

Haplotype sequences for each gene from each individual were also aligned in TranslatorX [72] using Clustalw [73] and used for Maximum-likelihood based phylogenetic inference

based on RAxML after 1000 bootstrap under the GTRGAMMA model. The relative rates of synonymous and nonsynonymous mutations (dN/dS = ω) averaged across sites was calculated using Codeml in PAML version 4.9 [74], as implemented in PAMLX [75]. Signs of selective pressure for each detected mutation were investigated comparing the M1a (nearly-neutral) *versus* the M2a (positive selection) site models by inferring ω estimations and posterior probabilities under the Bayes empirical Bayes (BEB) approach as implemented in Codeml [74]. This analysis was performed with the following default parameters: runmode = 0, clock = 0, Mgene = 0, CodonFreq = 0, estFreq = 0, fix_blength = 0, optimization method = 0, icode = 0, Seqtype = 1, fix α = 0, ncatG = 5, Small_Diff = 5e-7, n = 1, aaDist = 0.

The 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 in [43]. Briefly, SNPs and INDELs were inferred using four Variant callers (i.e. Freebayes [61], Platypus [76], Vardict [77] and GATK UnifiedGenotyper [78]) and the data merged and filtered with custom scripts. Filters include: minimum phred mapping quality = 20 (corresponding to 0.01 error rate), minimum phred base quality = 20, minimum allele frequency = 0.2, minimum allele observation = 2, minimum coverage = 8, maximum depth = 5000. The LoP for each individual was calculated as the number of variants averaged over the region length and the median value for each population was used for subsequent analyses. Statistical analyses were performed in R studio [79]. Fold-change differences were computed 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 via the Kolmogorov-Smirnov test and the p-value threshold was adjusted with the Bonferroni correction.

Homology modelling

Computational structural investigations were carried out initially through the identification of the closest homologs based on sequence similarity (using *NCBI Blast* [80]) and secondary structure matching (using *HHPRED* [81]), using the whole PDB as source collection. Homology model were then generated using *MODELLER* [82] by selecting only homologous RNA-bound structures as template models: *Kluyveromyces polysporus* Argonaute with a guide RNA (PDB ID 4F1N), Human Argonaute2 Bound to t1-G Target RNA (PDB ID 4z4d [83], *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* [84] followed by geometry idealization in *PHENIX* [85] to adjust the overall geometry. Final model quality was assessed by evaluating average bond lengths, bond angles, clashes, and Ramachandran statistics using Molprobity [86] and the *QMEAN* server [87]. The sequence alignment was generated using EBI muscle [88] and depicted using ESPRIPT3 [89]. Structural figures were generated with *PyMol* [90].

Developmental expression profile of *Piwi* genes

Publicly available RNA-seq data (runs: SRR458468, SRR458471, SRR1663685, SRR1663700, SRR1663754, SRR1663913, SRR1812887, SRR1812889, SRR1845684) were downloaded and aligned using Burrows-Wheeler Aligner (BWA-MEM) [60] to the current *Ae. albopictus* genome assembly (AaloF1). Aligned reads were visualized in Integrative Genomics Viewer

(IGV) [91]. Total RNA was extracted from embryos, 1st and 4th instar larvae, pupae, and adults using Trizol (Thermo Fisher Scientific). Embryos consisted of two pools of 60 eggs at different time points (i.e. 4-8h, 8-16h and 16-24h). Adult samples consisted of males and females kept on a sugar-diet and females fed an uninfected blood-meal. Blood-fed females were dissected to separate ovaries from the carcasses 48 h after blood-meal. These parameters were based on the results of previous studies on *Anopheles stephensi* and *Ae*. *aegypti* that showed high *Piwi* gene expression during early embryogenesis or 48-72h post blood meal [47]. For each stage, RNA was extracted from pools of 10 mosquitoes, except for first instar larvae and embryos when 20 and 60 individuals were used respectively.

RNA was DNaseI-treated (Sigma-Aldrich) and reverse-transcribed in a 20 μl reaction using the qScript cDNA SuperMix (Quantabio) following the manufacturer's instructions. Quantitative RT-PCRs (qRT-PCR) were performed as previously described using two biological replicates per condition and the RPL34 gene as housekeeping [92]. Relative quantification of *Piwi* genes was determined using the delta-delta-Ct method implemented in the software qBase+ (Biogazelle). Expression values were normalized with respect to those obtained from sugar-fed females.

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 (S4 Table) 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 [93,94] as implemented in qBASE+.

Supporting information

S1 Table. List of the core components of the piRNA pathway in *Ae. aegypti* and their orthologous in *Ae. albopictus*.
(PDF)

S2 Table. List of Transcript IDs and abbreviations of the Culicidae and Drosophilidae species included in the phylogenetic analyses.
(PDF)

S3 Table. Number of non-synonymous mutations found in mosquitoes of the Foshan strain (Foshan) and wild-caught samples from Mexico (Mex) and the island of La Reunion (Reu) divided by type (i.e. missense [M], frameshift [F], indel [I] and nonsense [N]) and number of sites in which mutations were found in all tested samples.
(PDF)

S4 Table. Relative expression values (log10 fold-change) of Piwi genes during development (A) and following viral infection (B) normalized with respect to sugar-fed samples. Samples (2 pools per condition, 15 individuals each) were analysed at 4 days post infection (early infection) and at 14 and 21 days post infection for CHIKV and DENV, respectively (late infection). Each condition was normalized to the corresponding sugar-fed control and compared to the corresponding Blood-fed control. Ovaries and carcasses were analysed independently. * indicates statistically significant difference between infected and non-infected blood-fed samples (ANOVA framework). Relative expression values may mask differences in levels of expression.

For instance, the Ct values of Piwi6, Piwi7 and Piwi1/3 in ovaries 4 days post infection with CHIKV were 30, 33.39 and 25.20, respectively. Ovaries of blood-fed samples at the same time point showed Ct values of 30.30, 33.93 and 26.55 for Piwi6, Piwi7 and Piwi1/3. When relative expression was calculated with respect to Ct values of RPL34, fold-changes in gene expression were comparable among the three genes in both conditions, but Ct values clearly indicate that Piwi7 is less expressed than both Piwi1/3 and Piwi6. These considerations were taken into account when describing results.
(PDF)

S5 Table. List of primers used for CDS analyses, copy number estimation, qPCR experiments and Northern Blot probe design.
(PDF)

S1 Dataset. CDS of the seven Piwi genes of *Ae. albopictus*. The sequence of the PAZ, MID and PIWI domains is in bold, underline and bold-italics, respectively.
(PDF)

S1 Fig. Maximum likelihood cladogram generated from the alignment 862 of transcript sequences of annotated Piwi genes in Culicidae. Transcript IDs and species abbreviations are as listed in S2 Table. AlbPiwi3 is the same as Piwi1/3 in the text. Piwi gene transcripts from *Ae. albopictus* are in red, from *Ae. aegypti* in purple, from *Culex quinquefasciatus* in pink. Transcripts from *D. melanogaster* Ago3, Piwi and Aubergine genes are included for reference and shown in blue. All nodes were supported by bootstrap values higher than 50% with the exception of the three nodes with a black dot.
(PDF)

S2 Fig. Polymorphism of Piwi4 and Piwi5. Lollipop plots representing position, amount and type of mutation along the coding sequences of Piwi4 and Piwi5 in mosquitoes of the Foshan strain, from la Reunion Island (Reu) and Mexico (Mex) as inferred by Freebayes and SnpEFF analyses. Only missense (blue), nonsense (red) and indels (orange) and frameshift (yellow) are shown. The PAZ, MID and PIWI domains are shown in green, blue and magenta, respectively. DDH residues positions are highlighted in the PIWI domain.
(PDF)

S3 Fig. Sequence alignment of *Aedes albopictus* Piwi proteins. Domain boundaries inferred from structural predictions are highlighted by coloured lines using the same colour coding as in Fig 4 (Orange: N-terminus; Green: PAZ; Blue: MID; Magenta: PIWI). Conserved DDH residues found in PIWI are indicated by a black triangle (▲). The "acc" line indicates relative solvent accessibility, ranging from blue (accessible) to white (buried). The sequence alignment was generated using EBI muscle [88] and depicted using ESPRIPT3 [89].
(PDF)

S4 Fig. Per-site read depth of the sequenced libraries of mosquitoes from Tapachula and Tampon.
(PDF)

S5 Fig. Northern blot results for Piwi5 and Piwi7.
(PDF)

Acknowledgments

We thank Monica Ruth Waghacore for insectary work. We thank Verily for having generated libraries and sequenced the mosquitoes from Tapachula and Tampon.

Author Contributions

Conceptualization: Rebeca Carballar-Lejarazu, Federico Forneris, Anna-Bella Failloux, Mariangela Bonizzoni.

Data curation: Michele Marconcini, Mariangela Bonizzoni.

Formal analysis: Michele Marconcini.

Funding acquisition: Mariangela Bonizzoni.

Investigation: Michele Marconcini, Luis Hernandez, Giuseppe Iovino, Vincent Houé, Federica Valerio, Umberto Palatini, Elisa Pischedda, Rebeca Carballar-Lejarazu, Lino Ometto, Federico Forneris.

Methodology: Michele Marconcini, Luis Hernandez, Giuseppe Iovino, Vincent Houé, Federica Valerio, Umberto Palatini, Bradley J. White, Teresa Lin, Rebeca Carballar-Lejarazu, Lino Ometto, Federico Forneris.

Project administration: Michele Marconcini, Anna-Bella Failloux, Mariangela Bonizzoni.

Resources: Jacob E. Crawford, Bradley J. White, Teresa Lin, Anna-Bella Failloux, Mariangela Bonizzoni.

Software: Umberto Palatini, Elisa Pischedda.

Supervision: Lino Ometto, Federico Forneris, Anna-Bella Failloux, Mariangela Bonizzoni.

Validation: Federico Forneris.

Writing – original draft: Michele Marconcini, Federico Forneris, Mariangela Bonizzoni.

Writing – review & editing: Jacob E. Crawford, Lino Ometto, Federico Forneris, Anna-Bella Failloux, Mariangela Bonizzoni.

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; https://doi.org/10.1093/emboj/17.1.170 PMID: 9427751
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. https://doi.org/10.1038/nsmb.2879 PMID: 25192263
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. https://doi.org/10.1093/gbe/evw018 PMID: 26868596
4. Buck AH, Blaxter M. Functional diversification of Argonautes in nematodes: an expanding universe: Figure 1. Biochem Soc Trans. 2013; https://doi.org/10.1042/BST20130086 PMID: 23863149
5. Bollmann SR, Press CM, Tyler BM, Grünwald NJ. Expansion and divergence of argonaute genes in the oomycete genus phytophthora. Front Microbiol. 2018; https://doi.org/10.3389/fmicb.2018.02841 PMID: 30555430
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; https://doi.org/10.1186/s12870-014-0364-6 PMID: 25626325
7. Bronkhorst AW, Van Rij RP. The long and short of antiviral defense: Small RNA-based immunity in insects. Current Opinion in Virology. 2014. https://doi.org/10.1016/j.coviro.2014.03.010 PMID: 24732439
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; https://doi.org/10.1016/j.chom.2018.02.001 PMID: 29503180

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; https://doi.org/10.1038/ncomms12410 PMID: 27580708

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; https://doi.org/10.1186/s12864-017-3903-3 PMID: 28676109

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; https://doi.org/10.1016/j.cub.2017.09.067 PMID: 29129531

12. Miesen P, Joosten J, van Rij RP. PIWIs Go Viral: Arbovirus-Derived piRNAs in Vector Mosquitoes. PLoS Pathog. 2016; 12: 1–17. https://doi.org/10.1371/journal.ppat.1006017 PMID: 28033427

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. https://doi.org/10.1016/j.cois.2017.05.010 PMID: 28805638

14. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, et al. Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in Drosophila. Cell. 2007; 128: 1089–1103. https://doi.org/10.1016/j.cell.2007.01.043 PMID: 17346786

15. Petit M, Mongelli V, Frangeul L, Blanc H, Jiggins F, Saleh M-C. piRNA pathway is not required for antiviral defense in *Drosophila melanogaster*. Proc Natl Acad Sci. 2016; https://doi.org/10.1073/pnas.1607952113 PMID: 27357659

16. Akbari OS, Antoshechkin I, Amrhein H, Williams B, Diloreto R, Sandler J, et al. The Developmental Transcriptome of the Mosquito *Aedes aegypti*, an Invasive Species and Major Arbovirus Vector. G3: Genes|Genomes|Genetics. 2013; https://doi.org/doi:10.1534/g3.113.006742

17. Matthews BJ, Dudchenko O, Kingan SB, Koren S, Antoshechkin I, Crawford JE, et al. Improved reference genome of *Aedes aegypti* informs arbovirus vector control. Nature. 2018; https://doi.org/10.1038/s41586-018-0692-z PMID: 30429615

18. Miesen P, Ivens A, Buck AH, van Rij RP. Small RNA Profiling in Dengue Virus 2-Infected Aedes Mosquito Cells Reveals Viral piRNAs and Novel Host miRNAs. PLoS Negl Trop Dis. 2016; https://doi.org/10.1371/journal.pntd.0004452 PMID: 26914027

19. Girardi E, Miesen P, Pennings B, Frangeul L, Saleh MC, Van Rij RP. Histone-derived piRNA biogenesis depends on the ping-pong partners Piwi5 and Ago3 in *Aedes aegypti*. Nucleic Acids Res. 2017; https://doi.org/10.1093/nar/gkw1368 PMID: 28115625

20. Miesen P, Girardi E, Van Rij RP. Distinct sets of PIWI proteins produce arbovirus and transposon-derived piRNAs in *Aedes aegypti* mosquito cells. Nucleic Acids Res. 2015; 43: 6545–6556. https://doi.org/10.1093/nar/gkv590 PMID: 26068474

21. Varjak M, Kean J, Vazeille M, Failloux A, Kohl A. *Aedes aegypti* Piwi4 Is a Noncanonical PIWI Protein Involved in Antiviral Responses. mSphere. 2017; 2: e00144–17. https://doi.org/10.1128/mSphere.00144-17 PMID: 28497119

22. Varjak M, Leggewie M, Schnettler E. The antiviral piRNA response in mosquitoes? 2018; 1–12. https://doi.org/10.1099/jgv.0.001157 PMID: 30372405

23. Schnettler E, Donald CL, Human S, Watson M, Siu RWC, McFarlane M, et al. Knockdown of piRNA pathway proteins results in enhanced semliki forest virus production in mosquito cells. J Gen Virol. 2013; 94: 1680–1689. https://doi.org/10.1099/vir.0.053850-0 PMID: 23559478

24. Hahn MW. Distinguishing among evolutionary models for the maintenance of gene duplicates. Journal of Heredity. 2009. https://doi.org/10.1093/jhered/esp047 PMID: 19596713

25. Obbard DJ, Jiggins FM, Halligan DL, Little TJ. Natural selection drives extremely rapid evolution in antiviral RNAi genes. Curr Biol. 2006; 16: 580–585. https://doi.org/10.1016/j.cub.2006.01.065 PMID: 16546082

26. Bonizzoni M, Gasperi G, Chen X, James AA. The invasive mosquito species *Aedes albopictus*: Current knowledge and future perspectives. Trends in Parasitology. 2013. https://doi.org/10.1016/j.pt.2013.07.003 PMID: 23916878

27. Rezza G. Dengue and chikungunya: long-distance spread and outbreaks in naïve areas. Pathog Glob Health. 2014; https://doi.org/10.1179/2047773214Y.0000000163 PMID: 25491436

28. Bonilauri P, Bellini R, Calzolari M, Angelini R, Venturi L, Fallacara F, et al. Chikungunya virus in *Aedes albopictus*, Italy. Emerging Infectious Diseases. 2008. https://doi.org/10.3201/eid1405.071144 PMID: 18439383

29. Venturi G, Di Luca M, Fortuna C, Elena Remoli M, Riccardo F, Severini F, et al. Detection of a chikungunya outbreak in Central Italy Detection of a chikungunya outbreak in Central. Euro Surveill. 2017; 22: 1–4. https://doi.org/10.2807/1560

30. Gjenero-Margan I, Aleraj B, Krajcar D, Lesnikar V, Klobučar A, Pem-Novosel I, et al. Autochthonous dengue fever in Croatia, August- September 2010. Eurosurveillance. 2011; 16: 1–4. 19805 [pii]

31. Marchand E, Prat C, Jeannin C, Lafont E, Bergmann T, Flusin O, et al. Autochthonous case of dengue in France, October 2013. Eurosurveillance. 2013; https://doi.org/10.2807/1560-7917.ES2013.18.50.20661 PMID: 24342514

32. Delisle E, Rousseau C, Broche B, Leparc-Goffart I, L'Ambert G, Cochet A, et al. Chikungunya outbreak in Montpellier, France, September to October 2014. Euro Surveill. 2015; https://doi.org/10.2807/1560-7917.ES2015.20.17.21108 PMID: 25955774

33. Calba C, Guerbois-Galla M, Franke F, Jeannin C, Auzet-Caillaud M, Grard G, et al. Preliminary report of an autochthonous chikungunya outbreak in France, July to September 2017. Eurosurveillance. 2017; https://doi.org/10.2807/1560-7917.ES.2017.22.39.17–00647

34. Bouri N, Sell TK, Franco C, Adalja AA, Henderson DA, Hynes NA. Return of epidemic dengue in the United States: Implications for the public health practitioner. Public Health Rep. 2012; https://doi.org/10.1177/003335491212700305 PMID: 22547856

35. Joshua-Tor L. The argonautes. Cold Spring Harbor Symposia on Quantitative Biology. 2006. https://doi.org/10.1101/sqb.2006.71.048 PMID: 17381282

36. Innan H, Kondrashov F. The evolution of gene duplications: Classifying and distinguishing between models. Nature Reviews Genetics. 2010. https://doi.org/10.1038/nrg2689 PMID: 20051986

37. Stone SS, Haldar JP, Tsao SC, Hwu W -m. W, Sutton BP, Liang Z-P. Accelerating advanced MRI reconstructions on GPUs. J Parallel Distrib Comput. 2008; 68: 1307–1318. https://doi.org/10.1016/j.jpdc.2008.05.013 PMID: 21796230

38. Yan KS, Yan S, Farooq A, Han A, Zeng L, Zhou M-M. Structure and conserved RNA binding of the PAZ domain. Nature. 2003; https://doi.org/10.1038/nature02129 PMID: 14615802

39. Song JJ, Smith SK, Hannon GJ, Joshua-Tor L. Crystal structure of argonaute and its implications for RISC slicer activity. Science (80-). 2004; https://doi.org/10.1126/science.1102514 PMID: 15284453

40. Cora E, Pandey RR, Xiol J, Taylor J, Sachidanandam R, McCarthy AA, et al. The MID-PIWI module of Piwi proteins specifies nucleotide- and strand-biases of piRNAs. RNA. 2014; https://doi.org/10.1261/rna.044701.114 PMID: 24757166

41. Stein CB, Genzor P, Mitra S, Elchert AR, Ipsaro JJ, Benner L, et al. Decoding the 5′ nucleotide bias of PIWI-interacting RNAs. Nat Commun. Springer US; 2019; 10: 828. https://doi.org/10.1038/s41467-019-08803-z PMID: 30783109

42. McDonald JH, Kreitman M. Adaptive protein evolution at the Adh locus in Drosophila. Nature. 1991; https://doi.org/10.1038/351652a0 PMID: 1904993

43. Pischedda E, Scolari F, Valerio F, Carballar-Lejarazú R, Catapano PL, Waterhouse RM, 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; https://doi.org/10.3389/fgene.2019.00093 PMID: 30809249

44. Schirle NT, MacRae IJ. The crystal structure of human argonaute2. Science (80-). 2012; https://doi.org/10.1126/science.1221551 PMID: 22539551

45. Matsumoto N, Nishimasu H, Sakakibara K, Nishida KM, Hirano T, Ishitani R, et al. Crystal Structure of Silkworm PIWI-Clade Argonaute Siwi Bound to piRNA. Cell. 2016; https://doi.org/10.1016/j.cell.2016.09.002 PMID: 27693359

46. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. ConSurf 2010: Calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res. 2010; https://doi.org/10.1093/nar/gkq399 PMID: 20478830

47. Macias V, Coleman J, Bonizzoni M, James AA. piRNA pathway gene expression in the malaria vector mosquito *Anopheles stephensi*. Insect Mol Biol. 2014; 23: 579–586. https://doi.org/10.1111/imb.12106 PMID: 24947897

48. Bonizzoni M, Dunn WA, Campbell CL, Olson KE, Marinotti O, James AA. Complex Modulation of the *Aedes aegypti* Transcriptome in Response to Dengue Virus Infection. PLoS One. 2012; https://doi.org/10.1371/journal.pone.0050512 PMID: 23209765

49. Campbell CL, Keene KM, Brackney DE, Olson KE, Blair CD, Wilusz J, et al. *Aedes aegypti* uses RNA interference in defense against Sindbis virus infection. BMC Microbiol. 2008; 8: 1–12. https://doi.org/10.1186/1471-2180-8-1

50. Hess AM, Prasad AN, Ptitsyn A, Ebel GD, Olson KE, Barbacioru C, et al. Small RNA profiling of Dengue virus-mosquito interactions implicates the PIWI RNA pathway in anti-viral defense. BMC Microbiol. 2011; https://doi.org/10.1186/1471-2180-11-45 PMID: 21356105

51. Chen X-G, Jiang X, Gu J, Xu M, Wu Y, Deng Y, et al. Genome sequence of the Asian Tiger mosquito, *Aedes albopictus*, reveals insights into its biology, genetics, and evolution. Proc Natl Acad Sci. 2015; https://doi.org/10.1073/pnas.1516410112 PMID: 26483478

52. La Ruche G, Souarès Y, Armengaud A, Peloux-Petiot F, Delaunay P, Desprès P, et al. First two autochthonous dengue virus infections in metropolitan France, september 2010. Eurosurveillance. 2010; 19676 [pii]

53. Schuffenecker I, Iteman I, Michault A, Murri S, Frangeul L, Vaney MC, et al. Genome microevolution of chikungunya viruses causing the Indian Ocean outbreak. PLoS Med. 2006; https://doi.org/10.1371/journal.pmed.0030263 PMID: 16700631

54. Dubrulle M, Mousson L, Moutailier S, Vazeille M, Failloux AB. Chikungunya virus and Aedes mosquitoes: Saliva is infectious as soon as two days after oral infection. PLoS One. 2009; https://doi.org/10.1371/journal.pone.0005895 PMID: 19521520

55. Campbell CL, Black IV WC, Hess AM, Foy BD. Comparative genomics of small RNA regulatory pathway components in vector mosquitoes. BMC Genomics. 2008; 9. https://doi.org/10.1186/1471-2164-9-425 PMID: 18801182

56. Arensburger P, Hice RH, Wright JA, Craig NL, Atkinson PW. The mosquito *Aedes aegypti* has a large genome size and high transposable element load but contains a low proportion of transposon-specific piRNAs. BMC Genomics. 2011; 12. https://doi.org/10.1186/1471-2164-12-606 PMID: 22171608

57. Yuan JS, Burris J, Stewart NR, Mentewab A, Neal CN. Statistical tools for transgene copy number estimation based on real-time PCR. BMC Bioinformatics. 2007; 8: 1–12. https://doi.org/10.1186/1471-2105-8-1

58. Baruffi L, Damiani G, Guglielmino CR, Bandii C, Malacrida AR, Gasperi G. Polymorphism within and between populations of ceratitis: Comparison between RAPD and multilocus enzyme electrophoresis data. Heredity (Edinb). 1995; 74: 425–437. https://doi.org/10.1038/hdy.1995.60 PMID: 7759289

59. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004; https://doi.org/10.1093/nar/gkh340 PMID: 15034147

60. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013; Available: https://arxiv.org/abs/1303.3997

61. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing—Free bayes—Variant Calling—Longranger. arXiv Prepr arXiv12073907. 2012; arXiv:1207.3907 [q-bio.GN]

62. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. 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 (Austin). 2012; https://doi.org/10.4161/fly.19695 PMID: 22728672

63. Schroeder MP, Lopez-Bigas N. muts-needle-plot: Mutations Needle Plot v0.8.0. 2015; https://doi.org/10.5281/ZENODO.14561

64. Oliveros JC. Venny. An interactive tool for comparing lists with Venn's diagrams. https://bioinfogp.cnb.csic.es/tools/venny/index.html. 2016;

65. Flot JF. Seqphase: A web tool for interconverting phase input/output files and fasta sequence alignments. Mol Ecol Resour. 2010; https://doi.org/10.1111/j.1755-0998.2009.02732.x PMID: 21565002

66. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet. 2001; https://doi.org/10.1086/319501 PMID: 11254454

67. Stephens M, Scheet P. Accounting for Decay of Linkage Disequilibrium in Haplotype Inference and Missing-Data Imputation. Am J Hum Genet. 2005; https://doi.org/10.1086/428594 PMID: 15700229

68. Librado P, Rozas J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009; https://doi.org/10.1093/bioinformatics/btp187 PMID: 19346325

69. Tajima F. Evolutionary relationship of DNA sequences in finite populations. Genetics. 1983;

70. Watterson GA. On the number of segregating sites in genetical models without recombination. Theor Popul Biol. 1975; https://doi.org/10.1016/0040-5809(75)90020-9

71. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989;

72. Abascal F, Zardoya R, Telford MJ. TranslatorX: Multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. 2010; https://doi.org/10.1093/nar/gkq291 PMID: 20435676

73. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994; https://doi.org/10.1093/nar/22.22.4673 PMID: 7984417

74. Yang Z. PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007; https://doi.org/10.1093/molbev/msm088 PMID: 17483113

75. Xu B, Yang Z. PamlX: A graphical user interface for PAML. Mol Biol Evol. 2013; https://doi.org/10.1093/molbev/mst179 PMID: 24105918

76. Rimmer A, Phan H, Mathieson I, Iqbal Z, Twigg SRF, Wilkie AOM, et al. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. Nat Genet. 2014; https://doi.org/10.1038/ng.3036 PMID: 25017105

77. Lai Z, Markovets A, Ahdesmaki M, Chapman B, Hofmann O, Mcewen R, et al. VarDict: A novel and versatile variant caller for next-generation sequencing in cancer research. Nucleic Acids Res. 2016; https://doi.org/10.1093/nar/gkw227 PMID: 27060149

78. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; https://doi.org/10.1101/gr.107524.110 PMID: 20644199

79. Rstudio Team. RStudio: Integrated Development for R. [Online] RStudio, Inc., Boston, MA. 2016. https://doi.org/10.1007/978-81-322-2340-5

80. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990; https://doi.org/10.1016/S0022-2836(05)80360-2

81. Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, et al. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. J Mol Biol. 2018; https://doi.org/10.1016/j.jmb.2017.12.007 PMID: 29258817

82. Webb B, Sali A. Protein structure modeling with MODELLER. Methods in Molecular Biology. 2017. https://doi.org/10.1007/978-1-4939-7231-9_4 PMID: 28986782

83. Schirle NT, Sheu-Gruttadauria J, Chandradoss SD, Joo C, MacRae IJ. Water-mediated recognition of t1-adenosine anchors Argonaute2 to microRNA targets. Elife. 2015; https://doi.org/10.7554/elife.07646 PMID: 26359634

84. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr Sect D Biol Crystallogr. 2010; https://doi.org/10.1107/s0907444910007493 PMID: 20383002

85. Adams PD, Afonine P V., Bunkóczi G, Chen VB, Echols N, Headd JJ, et al. The Phenix software for automated determination of macromolecular structures. Methods. 2011. https://doi.org/10.1016/j.ymeth.2011.07.005 PMID: 21821126

86. Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, et al. MolProbity: More and better reference data for improved all-atom structure validation. Protein Sci. 2018; https://doi.org/10.1002/pro.3330 PMID: 29067766

87. Benkert P, Künzli M, Schwede T. QMEAN server for protein model quality estimation. Nucleic Acids Res. 2009; https://doi.org/10.1093/nar/gkp322 PMID: 19429685

88. Madeira F, Park Y M, Lee J, Buso N, Gur T, Madhusoodanan N, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. 2019;

89. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. 2014; https://doi.org/10.1093/nar/gku316 PMID: 24753421

90. Schrödinger L. The PyMOL molecular graphics system, version 1.8. https://www.pymol.org/citing. 2015;

91. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nature Biotechnology. 2011. https://doi.org/10.1038/nbt.1754 PMID: 21221095

92. A. Reynolds J, Poelchau MF, Rahman Z, Armbruster PA, Denlinger DL. Transcript profiling reveals mechanisms for lipid conservation during diapause in the mosquito, *Aedes albopictus*. J Insect Physiol. 2012; https://doi.org/10.1016/j.jinsphys.2012.04.013 PMID: 22579567

93. Khan A, Rayner GD. Robustness to non-normality of common tests for the many-sample location problem. J Appl Math Decis Sci. 2004; https://doi.org/10.1155/S1173912603000178

94. Blanca MJ, Alarcón R, Arnau J, Bendayan R. Non-normal data: Is ANOVA still a valid option? Psicothema. 2017; https://doi.org/10.7334/psicothema2016.383 PMID: 29048317