

Research article

Open Access

## Comparative genomics of small RNA regulatory pathway components in vector mosquitoes

Corey L Campbell\*<sup>1</sup>, William C Black IV<sup>1</sup>, Ann M Hess<sup>2</sup> and Brian D Foy<sup>1</sup>

Address: <sup>1</sup>Arthropod-borne Infectious Diseases Laboratory; Microbiology, Immunology, and Pathology Department, Colorado State University, Fort Collins, Colorado, 80523, USA and <sup>2</sup>Department of Statistics and Center for Bioinformatics, Colorado State University, Fort Collins, Colorado, 80523, USA

Email: Corey L Campbell\* - [Corey.campbell@colostate.edu](mailto:Corey.campbell@colostate.edu); William C Black - [wcb4@colostate.edu](mailto:wcb4@colostate.edu); Ann M Hess - [ann.hess@colostate.edu](mailto:ann.hess@colostate.edu); Brian D Foy - [brian.foy@colostate.edu](mailto:brian.foy@colostate.edu)

\* Corresponding author

Published: 18 September 2008

Received: 28 February 2008

BMC Genomics 2008, 9:425 doi:10.1186/1471-2164-9-425

Accepted: 18 September 2008

This article is available from: <http://www.biomedcentral.com/1471-2164/9/425>

© 2008 Campbell et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Small RNA regulatory pathways (SRRPs) control key aspects of development and anti-viral defense in metazoans. Members of the Argonaute family of catalytic enzymes degrade target RNAs in each of these pathways. SRRPs include the microRNA, small interfering RNA (siRNA) and PIWI-type gene silencing pathways. Mosquitoes generate viral siRNAs when infected with RNA arboviruses. However, in some mosquitoes, arboviruses survive antiviral RNA interference (RNAi) and are transmitted via mosquito bite to a subsequent host. Increased knowledge of these pathways and functional components should increase understanding of the limitations of anti-viral defense in vector mosquitoes. To do this, we compared the genomic structure of SRRP components across three mosquito species and three major small RNA pathways.

**Results:** The *Ae. aegypti*, *An. gambiae* and *Cx. pipiens* genomes encode putative orthologs for all major components of the miRNA, siRNA, and piRNA pathways. *Ae. aegypti* and *Cx. pipiens* have undergone expansion of Argonaute and PIWI subfamily genes. Phylogenetic analyses were performed for these protein families. In addition, sequence pattern recognition algorithms MEME, MDScan and Weeder were used to identify upstream regulatory motifs for all SRRP components. Statistical analyses confirmed enrichment of species-specific and pathway-specific cis-elements over the rest of the genome.

**Conclusion:** Analysis of Argonaute and PIWI subfamily genes suggests that the small regulatory RNA pathways of the major arbovirus vectors, *Ae. aegypti* and *Cx. pipiens*, are evolving faster than those of the malaria vector *An. gambiae* and *D. melanogaster*. Further, protein and genomic features suggest functional differences between subclasses of PIWI proteins and provide a basis for future analyses. Common UCR elements among SRRP components indicate that 1) key components from the miRNA, siRNA, and piRNA pathways contain NF-kappaB-related and Broad complex transcription factor binding sites, 2) purifying selection has occurred to maintain common pathway-specific elements across mosquito species and 3) species-specific differences in upstream elements suggest that there may be differences in regulatory control among mosquito species. Implications for arbovirus vector competence in mosquitoes are discussed.

### Background

Small RNA-mediated gene silencing pathways are master regulators of critical cellular processes, from development to germ-line surveillance to anti-viral defense [1-5]. Although small RNA regulatory pathways (SRRPs) operate using distinctly different mechanisms, they share several common features. Small regulatory RNAs of 20 to 30 nucleotides (nts) are used as guide strands for target substrate recognition by an RNase H type nuclease of the Argonaute protein family. Target RNAs are subsequently degraded or otherwise prevented from being translated into protein. The three major pathways are the small interfering RNA (siRNA), microRNA (miRNA) and PIWI small RNA (piRNA) pathways; the general functions of each are summarized in Table 1. Due to the paucity of functional information from mosquitoes, we have relied on protein structure and functional information compiled from *Drosophila* spp., *Caenorhabditis elegans*, and mammals.

The Argonaute protein family contains the Argonaute and PIWI protein subclasses. All proteins in this family contain PAZ (PIWI, Argonaute, Zwillig) and PIWI domains [6]. The PAZ domain is a small RNA binding domain; the PIWI domain is an RNase-H type domain which relies on divalent cation binding to facilitate dsRNA-guided cleavage of ssRNA (reviewed in [7]). Argonaute 1 (Ago1) and Argonaute 2 (Ago2) are in the Argonaute protein subclass and are functionally distinct from PIWIs, in that they rely on Dicer proteins to cleave long dsRNAs into small RNAs, which are then passed to the Agos for further processing and recognition of target RNAs. These proteins physically interact with Dicers at the PIWI domain and use small RNAs formed from double-stranded templates [8]. Ago1 and Ago2 each participate in different gene silencing pathways.

MiRNA pathways of invertebrates have been best characterized in *D. melanogaster* and *C. elegans* (reviewed in [9,10]). In *D. melanogaster*, developmental and housekeeping gene expression is regulated by Ago1 and the small RNA subclass, miRNAs (20–22 nts) [11]. In general, Ago1 acts with Dicer-1, microRNAs (miRNAs) and accessory proteins to control gene expression of housekeeping genes (Figure 1) [12-14]. In *An. gambiae*, the miRNA pathway helps defend against *Plasmodium bergeri* infection [15]. Gene expression can be controlled by a variety of possible

silencing mechanisms, including translational repression, decreased mRNA stability or altered mRNA localization [3,16,17]. These small RNAs are genome-encoded as primary transcript miRNAs (pri-miRNAs) up to 2 kilobases (kB) in length and processed into miRNA precursor (pre-miRNAs) hairpin loops of about 65 nts by Drosha and Pasha (DGCR8) [18,19]. Dicer-1 cleaves pre-miRNAs to form mature miRNAs, which are transferred to Ago1 for silencing of target mRNAs [3,20,21]. In mammals, some miRNAs are encoded in the introns of the genes they target [22].

The siRNA pathway provides defense against RNA viruses in the degradation of double-stranded viral RNAs [1,8,23,24]. SiRNAs, generated from longer dsRNA molecules by the Dicer-2/R2D2 complex, serve as guides for Ago2 to identify and degrade target RNAs. Key proteins in the anti-viral RNA interference (RNAi) pathway, including Ago2 and Dicer-2, are functional in vector mosquitoes and influence the ability of mosquitoes to serve as competent vectors of human virus pathogens [1,25-27]. Further, several recent reports of *Drosophila* have shown that endogenous siRNAs control retrotransposons and mRNAs in somatic cells in an Ago2/Dicer-2 dependent manner [28-30].

In *Drosophila melanogaster*, the PIWI protein subclass includes PIWI, Aubergine (Aub), and Argonaute 3 (Ago3). These proteins contain the signature PAZ and PIWI domains, but have remained enigmatic members of the Argonaute family. They have been implicated in a variety of functions specific to germ-line tissue in *D. melanogaster*, including the suppression of retrotransposon mobilization, maintenance of telomeres, prevention of double-stranded DNA breaks during meiosis, and mRNA silencing in germ-line cells [4,5,31-34]. Ago3, Aub, and PIWI use PIWI-associated RNAs (piRNAs) (24 to 30 nts) in these pathways. The mechanisms for production of this class of non-coding RNAs are not well understood. Models have been proposed, suggesting they are processed from piRNA cluster transcripts and amplified by a "Ping Pong" amplification loop between these transcripts and those derived from active transposons (Figure 1) [35].

Three mosquito species, *Anopheles gambiae*, *Aedes aegypti*, and *Culex pipiens quinquefasciatus* are important vectors of

**Table 1: Argonaute protein family functional groups**

Protein Subfamily	Protein	Small RNA class, length	Putative Function (referenced mosquito studies)
Argonaute	Ago1	miRNA, 20–22 nts	Regulation of endogenous mRNAs during development; control of Plasmodium infection [15,54]
Argonaute	Ago2	siRNA, 20–22 nts	Anti-viral defense [1,25-27]
PIWI	Ago3 PIWI (Ago4) Aub (Ago5)	piRNA, rasiRNA, 24–30 nts	Germline surveillance, telomere maintenance, anti-viral defense [1], and maintenance of retrotransposon silencing



human and animal pathogens [36-41]. The goal of this study is to increase understanding of the three small RNA pathways in vector mosquitoes using comparative genomics, with special emphasis on the key catalytic enzymes of the Argonaute protein family. We performed phylogenetic analyses of Argonaute family proteins and demonstrated expansion of *Ae. aegypti* and *Cx. pipiens* Argonaute and PIWI subfamilies. In addition, we determined whether *cis*-acting regulatory elements upstream of SRRP genes are conserved across mosquito species or among components of specific RNA regulatory pathways, as has been successfully demonstrated in other invertebrates [42]. To this end, previously validated mosquito transcription factor binding sites (TFBSs) and other elements in upstream control regions (UCRs) of all suspected SRRP components were identified. Further, we used motif discovery algorithms to identify novel upstream regulatory sequences. The identification of common *cis*-acting regulatory features lends insight into species-specific differences in regulation of SRRPs and provides a basis for future studies of pathogen defense in vector mosquitoes.

## Results and discussion

The *An. gambiae*, *Ae. aegypti*, and *Cx. pipiens* genomes encode components of the three major RNA regulatory pathways previously identified in *D. melanogaster* (Figure 1 and Additional File 1A) [1,8,11,12,24,26,43-53]. There are gene expansions in some categories; these are described below. Additional File 1A depicts, in gray highlight, the genes for which expression has been confirmed, either experimentally or in EST libraries [1,26]. Importantly, key catalytic residues of the RNase-H domains are conserved in all mosquito Ago1s, Ago2s and PIWI subfamily proteins (Figure 2) [7]. Although a few studies have been done in *Ae. aegypti* and anopheline mosquitoes [1,15,25,54], experimental verification will be required to confirm the roles played by most of these putative orthologs in RNA regulatory pathways.

### Argonaute protein subfamily

*Ago1* is represented at a single genomic locus in *An. gambiae*, *Cx. pipiens* and *D. melanogaster* but seems to be present in two copies in *Ae. aegypti* (Figure 1, Additional File 1A). The flanking sequences of about 1.2 kilobases (kB) upstream and the introns were found to be unique.

*Ago1* of the human head louse, *Pediculus humanus* was used as an outgroup for the analysis of Argonaute subfamily proteins. *Ago2* homologs from the silkworm, *Bombyx mori*, and *D. melanogaster* were also included in the analysis. In addition, orthologs from the important pathogen vectors, *Lutzomyia longipalpis*, *Culex tarsalis*, and *Ochlerotatus triseriatus* were included. Based upon the Jones-Taylor-Thornton probability model, the probability of amino acid substitutions among *Ago1* proteins was 0.0247 (coefficient of variation (CV) = 100 \* (standard

deviation/mean) = 100.0%) while among *Ago2* proteins the average probability was 0.9411 (CV = 30.1%) (Figure 3) [55]. Thus, the probability of amino acid substitutions among *Ago2* proteins is about 38-fold greater than among *Ago1* proteins. Because of this high rate of change among *Ago2* proteins, there was only weak bootstrap support (56.6%) for the proteins from mosquitoes. However, among the two genera of the tribe Aedini it is interesting that between *Ochlerotatus triseriatus* and *Ae. aegypti* the substitution probability is only 0.1306, while within and among *Culex* species the probability is 0.3798 (CV = 68.9%). This pattern suggests that anti-viral pathway components are evolving more rapidly in *Culex spp.* than among subgenera of Aedine mosquitoes. An interesting parallel was found among drosophilid species, wherein key anti-viral RNAi component genes, such as *DmeAgo2* and *DmeDicer-2*, were found to be evolving at a higher rate than miRNA pathway components or other immune response genes [56]. Synapomorphies that distinguish *Ago1s* from *Ago2s* can be found in the amino acids surrounding the first catalytic residue (Additional File 2).

*Ago1* protein conservation indicates an evolutionary trend of purifying selection, probably due to conserved mechanisms in developmental pathways. This hypothesis is supported in the conservation of some miRNAs between drosophilids, mosquitoes, and humans <http://microrna.sanger.ac.uk/> [57]. Although many mosquito miRNAs have not been experimentally verified, their presence suggests conservation of developmentally regulated gene silencing pathways. Alignment of *Ago1* proteins further illustrates the degree of conservation with overall amino acid identity of 82 to 89% among all mosquito orthologs and that of *D. melanogaster* (Additional File 1B). This is in contrast to the 30 to 43% amino acid identity among all mosquito *Ago 2* proteins compared to *DmeAgo2* (Additional File 1C).

*Ago2* is a single locus in *Ae. aegypti*, *An. gambiae*, and *D. melanogaster*, but is present in two copies in *Cx. pipiens* (Figure 1). Fragments of two distinct mRNAs have also been isolated from *Cx. tritaeniorhynchus*, suggesting that a gene duplication occurred early in the evolution of *Culex spp.* (Campbell, unpublished). This first evidence of multiple *Ago2* loci in Dipterans suggests that differential regulation of RNAi anti-viral defense could occur in *Culex spp.* vectors. Although anti-viral RNAi activity has not yet been reported for *Culex spp.*, this pathway has been established as an important facet of anti-viral defense in *An. gambiae* and *Ae. aegypti* against alphaviruses and dengue viruses (DENV), respectively [1,25,26].

### PIWI pathway gene expansions

A significant gene expansion has occurred in PIWI subfamily proteins in both *Ae. aegypti* and *Cx. pipiens* (Figure 1) [48,51]. Phylogenetic analyses of PIWIs used *D. melo-*

Ago1	Ago1.Cpi	RPKVFDEPVI...ILYRDGVS...PAPAYY AHL
	Ago1-1.Aae	RPKVFDEPVI...ILYRDGVS...PAPAYY AHL
	Ago1-2.Aae	RPKVFDEPVI...ILYRDGVS...PAPAYY AHL
	Ago1.Aga	RPKVFDEPVI...ILYRDGVS...PAPAYY AHL
	Ago1.Dme	RPKVFNEPVI...ILYRDGVS...PAPAYY AHL
Ago2	Ago2.Otr	MYIGADVTHP...IMYYRDGV...PAPTY Y AHL
	Ago2-1A.Cpi	MFVGGADVTHP...ILYYRDGV...PAPTY Y AHL
	Ago2-1B.Cpi	MFVGGADVTHP...ILYYRDGV...PAPTY Y AHL
	Ago2-2.Cpi	MYVGGADVTHP...IMYYRDGV...PAPTY Y AHL
	Ago2.Aae	MYVGGADVTHP...IMYYRDGV...PAPTY Y AHL
	Ago2.Aga	MYIGADVTHP...ILYYRDGV...PAPTY Y AHL
	Ago2.Dme	MYIGADVTHP...ILYYRDGV...PAPTY LAHL
PIWI	Ago3.Aae	MIAGIDTYHD...IIFRDGV...ACCQY A HK
	Ago3.Aga	MICGIDTYHE...IIFRDGV...ACCQY A HK
	Ago3.Cpi	MIAGIDTYHD...IIFRDGV...ACCQY A HK
	Ago3.Dme	MICGIDSYHD...IIFRDGI...ACCMV SH N
	PIWI1.Cpi	MTVGFVCHD...IFYRDGV...AVCQY A HK
	PIWI2.Cpi	MTIGFVCHD...IFYRDGV...AVCQY A HK
	PIWI4A.Cpi	MCVGFVCRD...IYYRDGV...AVCQY A NK
	PIWI4B.Cpi	MCVGFVCRD...IYYRDGV...AVCQY A NK
	PIWI5A.Cpi	MVIGFVCHD...IYYRDGV...AVCQY A HK
	PIWI6.Cpi	MCLGFVCHD...IYYRDGV...APCQY A HK
	PIWI5B.Cpi	MVIGFVCHD...IYYRDGV...AVCQY A HK
	PIWI3A.Cpi	MTVGFVCHD...IFYRDGV...AVCQY A HK
	PIWI3B.Cpi	MTVGFVCHD...IFYRDGV...AVCQY A HK
	Ago4.Aga	MTIGFVCHD...IFYRDGV...AVCQY A HK
	PIWI1.Aae	MTIGFVCHD...FFYRDGV...AVCQY A HK
	PIWI3.Aae	MTIGFVCHD...FFYRDGV...AVCQY A HK
	PIWI2.Aae	MTVGFVCHD...FFYRDGV...AVCQY A HK
	PIWI6.Aae	MVIGYDVCKD...IVYRDGV...AVCQY A HK
	PIWI7.Aae	MVIGFVCHD...IVYRDGV...AVCQY A HK
	Ago5.Aga	MVVGFDVCHD...LVYRDGV...AVCQY A HK
	PIWI5.Aae	MVIGFVCHD...IVYRDGV...AVCQY A HL
	PIWI4.Aae	MTIGFVCHD...IFYRDGV...AVCQY A HK
	PIWI.Dm	MTIGFDIAKS...VFYRDGV...AVCQY A KK
	Aub.Dm	MTVGFVCHS...LFFRDGV...AVCHY A HK

**Figure 2 Conservation of Argonaute Family Protein Key Catalytic Residues.** Mosquito PIWI and Argonaute orthologs retain key catalytic residues [7]. Stars indicate key residues. Yellow highlight indicates identical amino acids among species. Gene accession numbers: *Oc. triseriatus*, [Genbank: EU182829] or in Additional File 1A.

*nogaster* Aub and PIWI as outgroups (Figure 4). Three well-supported clades are evident. Ago3s form a single clade with *D. melanogaster*. In contrast, none of the mosquito PIWIs formed a clade with the drosophilid proteins, rather, two clades arise based on similarity to *An. gambiae* Ago4 or Ago5.

The first clade has 100% support and contains the Ago3 proteins. Ago3s can be readily distinguished from other Argonaute family proteins by the amino acids surrounding the first conserved catalytic residue (Figure 2 and Additional File 1D). The average rate of change among these is high 0.6872 (CV = 40.7%). The Ago3 proteins among the Culicidae are monophyletic with 100.0% support, and the average rate of change among these is also high 0.4615 (CV = 44.6%).

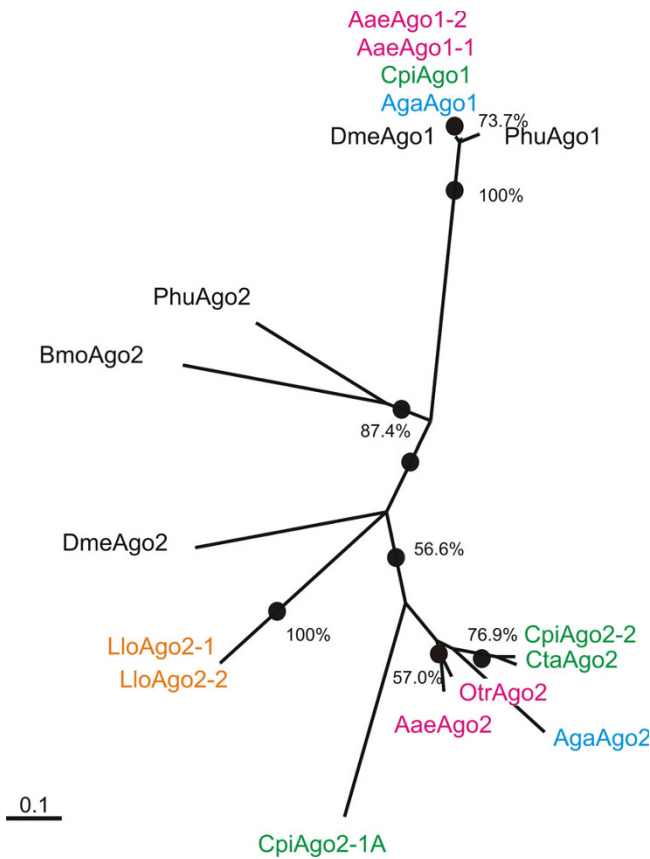
The Ago4-like clade has 100.0% bootstrap support and contains AgaAgo4, CpiPIWI1, CpiPIWI2, CpiPIWI3,

AaePIWI1, AaePIWI2, AaePIWI3 and AaePIWI4. The amino acid substitution probability among these is low relative to the other 2 clades, (0.1943; CV = 39.5%). Members of the Ago4-like group were distinguished by the synapomorphy "ETGIQVLNLLRRAMNGLNLQLV-GRNLY" within the first 260 amino acids (Additional File 2). This synapomorphy is maintained in DmePIWI, but not in DmeAub. Members of the Ago5-like group have a variable sequence in the corresponding region.

AgaAgo5 is basal to the Ago5-like clade, which has 93.0% bootstrap support and contains CpiPIWI4, CpiPIWI5, CpiPIWI6, AaePIWI5, AaePIWI6 and AaePIWI7. The amino acid substitution probability is high (0.3422; CV = 37.8%). These analyses indicate that the expanded PIWI protein subfamily comprises several different gene families with some resulting from recent gene duplication events (Figure 4, Additional File 1E). Expansion of Argonaute family proteins has occurred in other organisms, as well. One example is in *C. elegans*, in which a gene expansion has occurred, evidently to aid in sequential amplification of siRNA signal. A key difference between the *C. elegans* and mosquito Argonaute family gene expansions is that the *C. elegans* secondary Agos (SAGOs) lack the key metal binding residues required for RNase H catalytic activity and so are thought to play a secondary role in RNAi [58], whereas, these key catalytic residues are present in mosquito PIWIs (Figure 2).

In drosophilids, PIWI pathway components have been implicated in the control of both retrotransposons and the long terminal repeats of telomeres [31,33]. Our analysis suggests the PIWI subfamily gene expansion initially occurred in ancestral Culicinae mosquitoes before the divergence of *Ae. aegypti* and *Cx. pipiens* ancestors [59,60]. Expansions in PIWI pathway components may have been adaptive for controlling the increased burden of retrotransposons in *Ae. aegypti* and *Cx. pipiens* relative to *An. gambiae*. Although retrotransposons are present in the *An. gambiae* genome, the percentage of the genome bearing TEs is much less than that of *Ae. aegypti* [46,61-63]. About 47% of the *Ae. aegypti* genome harbors both class I and class II transposable elements (TEs), whereas the *An. gambiae* genome contains about 16% TEs [46,51]. TEs are also present in *Culex spp.* genomes, however the percentage of the genome occupied has not yet been reported [63,64]. Retrotransposons, also known as Class I TEs, account for at least half of the TE load in both anopheline and aedine genomes. Therefore, the aedine genome carries more than twice the retrotransposon load than anophelines carry.

In a related context, DmePIWI has also been implicated in maintenance of heterochromatin. DmePIWI associates directly with chromatin and heterochromatin protein 1a (HP1a) [65]. DmePIWI-HP1a interactions occur through conserved PIWI protein motifs, PxVxV or PxVxM. Of the



**Figure 3**  
**Argonaute Subfamily Tree.** Argonaute protein subfamily maximum likelihood tree with bootstrap values. *Oc. triseriatus*, [Genbank: [EU182829](#)]; *Cx tarsalis*, [Genbank: [EU182828](#)]; *Lutzomyia longipalpis*, [Genbank: [AM094709.1](#), [AM094708.1](#)]; Phum, *Pedicularis humanus*, [Vectorbase: [phum002582](#)]. All other accession numbers are listed in Additional File 1A. Bar equals 0.1000 amino acid substitution probability.

mosquito PIWIs, only CpiPIWI4A, CpiPIWI4B, AaePIWI5 and AaePIWI6 contain conserved PxVxV sites, and no proteins bear PxVxM motifs (data not shown). Importantly, all four of these proteins are in the Ago5-like protein class. In contrast, neither AgaAgo4 nor AgaAgo5 carry these motifs.

**Accessory Proteins**

The RNA helicases of *Drosophila* and mosquitoes are clearly a large and complex family. Other than a genetic connection to anti-viral defense and retrotransposon maintenance in drosophilids, little is known about Rm62 in Dipterans. Using a high stringency search cut-off E value of  $E = 10^{-80}$ , multiple mosquito genes were identified that are homologous to *DmeRm62* RNA helicase (Figure 1 and Additional Files 1A, F). Although reciprocal genome searches identified 7 homologous RNA helicase genes in *Drosophila*, only *DmeRm62* has been linked to

RNA interference [24,66]. Phylogenetic analysis demonstrated that there are independent orthologous groups of Rm62-like proteins (Additional File 3).

Rm62 and Armitage-like proteins carry predicted DEAD box ATP-dependent helicase domains. *DmeArmitage* is a transcriptional repressor of specific mRNAs during male germ-line development and has been shown in a genetic screen to be involved in anti-viral defense [24,67]. Interestingly, the Armitage gene duplication in *Aedes* and *Culex* has resulted in two different types of multi-domain helicases (Figure 1). Although both aedine Armitage-like paralogs carry a type III DNA restriction enzyme domain, as does *DmeArmitage*, only one of the two *Culex* paralogs carries this domain (CPIJ001247). Instead, the second paralog is missing the restriction enzyme domain and carries a predicted DNA helicase domain (CPIJ001245) (data not shown). In contrast, AgaArmitage has two RNA helicase domains, in addition to the DNA restriction enzyme domain. This interesting diversity in protein domains among mosquito species suggests that, if Armitage participates in anti-viral defense as in *D. melanogaster*, it may be pivotal to species-specific differences.

**Upstream Control Regions**

We took a conservative approach to determine whether cis-acting elements are conserved among SRRP component upstream regions. Our goals were two-fold: 1) to identify previously validated mosquito TFBSs and 2) novel upstream regulatory control elements in regions corresponding to -1000 nts to +100 nts. This region corresponds to the upstream genomic sequence, the 5' untranslated region of the transcript, the translation start site and some coding nucleotides in the first exon. Consensus sequences for the upstream motifs are listed in Table 2. We determined whether elements in each of these categories are conserved within each pathway or across mosquito species.

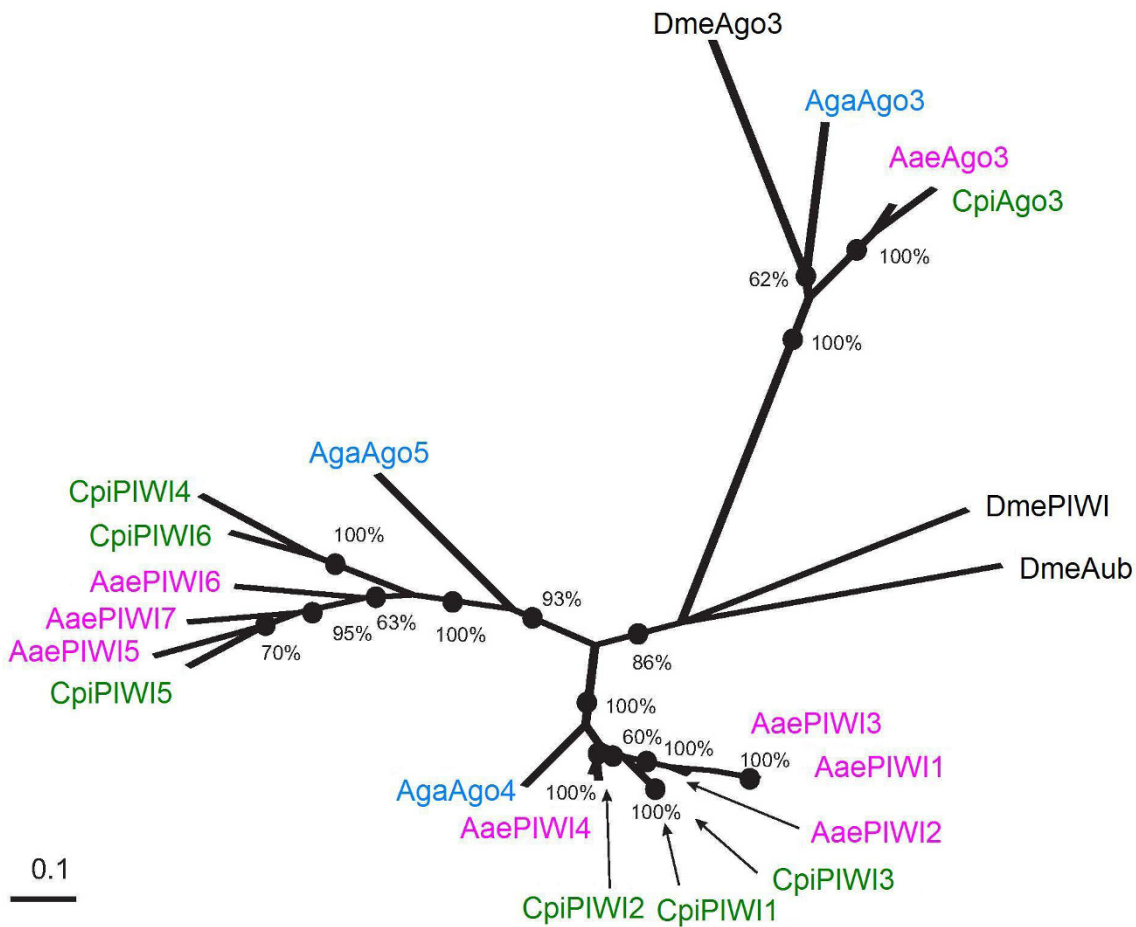
Within the last 150 million years, *An. gambiae* and *Ae. Aegypti* became distinct species, and more recently, *Cx. pipiens* diverged. One might expect that among-species comparisons would yield fewer common elements than within-species elements. However, non-coding DNA is subject to the same stabilizing selection as protein coding regions, and, in drosophilids, non-coding DNA is actually less polymorphic than protein coding regions at synonymous sites [68]. If purifying selection has occurred in UCRs of mosquito SRRP genes, common elements could be conserved across species, thus adding support to the functional significance of upstream regulatory elements, regardless of whether they were unique across the entire genome. Furthermore, selective pressures on non-coding DNA can be significant and therefore exacerbate the search for TFBSs [68]. Therefore, to maintain a conserva-

tive line of inquiry, we chose to examine only those TFBSs that have been experimentally confirmed in mosquitoes. The search for novel motifs was expected to yield conserved cis-acting regulatory elements.

**Transcription factor binding sites: NF-kappa B related sites**

There is empirical evidence that RNA viruses activate innate immunity pathways in Dipterans. The Toll and JAK-STAT (janus kinase/signal transducers and activators of transcription) pathways are activated in drosophilids upon infection with *Drosophila C* virus or *Drosophila X* virus [69,70]. A search for STAT binding motifs (Table 2) among all pathway component UCRs yielded no matches, thus suggesting that the JAK-STAT pathway does not activate transcription of effectors of mosquito miRNA, siRNA, or PIWI pathways. Mosquito NF kappa B-related proteins, REL1 and REL2, participate in transcriptional induction of immune effectors during bacterial and fungal infections (summarized in [71]). REL1 stimulates downstream effec-

tors via the Toll pathway, while REL2 stimulates downstream effectors via the immunodeficiency (Imd) pathway [72]. During Sindbis (SINV) infection of *Ae. aegypti*, REL1 transcripts are enriched early in infection [73]. Identification of REL TFBSs in UCRs of SRRP genes would suggest a link between small regulatory RNA pathways and the innate immune response cascade. Consensus NF kappa B-related binding motifs in *Ae. aegypti* and *An. gambiae* are "GGKGATYYAC" (NFkB1) and "KGGGAWHMMM" (NFkB2), respectively. A cross-species consensus is "KGKGAWHHMM" (NFkB4) [72]. These REL TFBS consensus motifs could be bound by either AaeREL1 or AaeREL2. Search of all mosquito SRRP components for NFkB1 yielded hits on only two Rm62-like UCRs, and NFkB2 revealed few hits across all species (Additional File 4). The cross-species 10 nt consensus, NFkB4, revealed more hits, some of which are redundant to those identified by NFkB2 (Figures 5, 6, 7). To identify putative motifs under more relaxed conditions, a shortened version of the



**Figure 4**

**PIWI Subfamily Tree.** PIWI protein subfamily maximum likelihood tree with bootstrap values. All accession numbers are listed in Additional File 1A. Bar equals 0.1000 amino acid substitution probability.

NFkB2 consensus, "GGGAWHM" (NFkB7), was used. In this case, the relaxed search was followed by a more rigorous requirement of  $\geq 2$  motifs per UCR. Proportions of UCRs containing these motifs are listed in Table 3.

Schematics showing distribution of NFkB4 motifs are shown in Figures 5, 6 and 7. Hits were detected in key component UCRs of all three mosquito species and SRRPs. For example, *AaeDrosha*, *AgaDrosha*, *AaeTSN*, *CpiTSN* and *AgaTSN* UCRs share NFkB4 sites, and all three species have NFkB4 sites upstream of PIWI pathway component UCRs. Although there were no significant pathway-specific or species-specific differences in the presence of NF kappa B-related binding sites, the presence of these motifs in key UCRs suggests a possible link to the innate immune response or development.

The drosophilid ortholog of Tudor staphylococcal nuclease (TSN) is a component of the RISC and has multiple functional attributes, such as transcriptional co-activation activity, non-specific ssRNA cleavage, cleavage of hyper-edited dsRNA substrates, and an un-defined role in anti-viral defense [43,74]. Microarray analysis of *Ae. aegypti* transcripts during a SINV infection, showed early enrichment of *AaeREL1* transcripts at 1 day post-infection (dpi), followed by stimulation of *AaeTSN* at 4 dpi [73]. Analysis of anti-viral RNAi component transcript levels in *Ae. aegypti* infected with alphaviruses or flaviviruses showed periodic enrichment of *TSN* transcripts in a virus-dependent manner [25](Campbell, unpublished). Importantly, *TSN* UCRs from all three species contained NFkB4 motifs (Figure 7, Additional File 4). This evidence further suggests that a link exists between the anti-viral pathway and the innate immune response pathway of mosquitoes.

#### **Broad Complex motifs**

Broad complex (BRC) transcription factors act during ecdysone-dependent regulation of gene expression in invertebrates [75]. Ecdysone, a steroid hormone, is of key importance in embryogenesis and development in insects. Of the four types of BRC TFBSs, BRC\_Z1 is the most abundant in miRNA and piRNA pathway component UCRs, however, the enrichment is not statistically significant (Table 3). In mosquitoes, BRC\_Z1 is thought to be a transcriptional repressor that ensures proper temporal gene expression control for the yolk protein precursor gene, vitellogenin [76]. The presence of a BRC\_Z1 motif upstream of nearly all of the miRNA pathway components and many PIWI pathway components suggests the requirement for temporal control of these pathways, as well. BRC\_Z2 is required for 20-hydroxyecdysone-mediated transcriptional activation [76]. BRC\_Z2 motifs were identified on key genes of all three pathways, sometimes in tandem with BRC\_Z1 motifs (Figures 5, 6, 7 and Additional File 4).

#### **GATA factors**

GATA factors control tissue- and temporally-specific gene expression. GATA binding sites are upstream of a variety of genes in mosquitoes, including lysosomal protease, gut trypsin, and vitellogenin genes [77-79]. Expression of gut trypsin and vitellogenin genes are tightly linked to blood-feeding. GATA factors or GATA repressors bind these sites to either activate or repress transcription. Multiple GATA sites were identified in all UCRs, indicating no obvious species-specific or pathway-specific variation (Additional File 4). Therefore, this element was not analyzed further.

#### **Novel Elements**

Novel UCR elements could serve as genomic structural elements, mRNA stability elements or TFBSs. We identified features in upstream non-coding regions that may have been selectively maintained over the course of evolution. To accomplish this, we compared UCRs across mosquito species and across small RNA regulatory pathways for all three mosquito species. This focused approach allowed us to use a subset of each mosquito genome as background. The presence of *cis* upstream elements across all three mosquito species for a single pathway would suggest that they may be important regulatory features of that pathway. Although some components, such as those of the RISC, may act in multiple pathways; for our analysis, we categorized the groups according to the outline in Figure 1. The sequence motif search programs MEME (Multiple EM for motif elicitation), MDScan, and Weeder were used [80-83]. These programs identify over-represented sequence patterns or motifs in a given dataset compared to all other datasets. The top two elements discovered for each species and pathway were then identified among all UCRs, and significant differences among pathways or species were noted. Selected motifs are listed in Table 3, along with the proportion of UCRs with at least one species-specific or pathway-specific hit, and full descriptions are in Additional File 4. Fisher's Exact test was used to determine whether the UCR elements or TFBS motifs were enriched in a pathway-specific or species-specific manner. The Benjamini-Hochberg (BH) multiple testing adjustment was used to correct for false positives, thus, for any BH score over 0.05, the Fisher's Exact test score was considered insignificant. In addition, we performed full genome searches for all novel elements and removed those that did not show enrichment in SRRP UCRs over the rest of the genome (Additional File 4). Together, these methods allowed identification of *cis*-elements that could be important regulatory features of SRRP pathways.

Identification of species-specific *cis* elements could provide an important foundation for future characterization of species-specific differences in regulation of pathogen defense pathways. By convention, strong regulatory sites are those represented in tandem repeats (Figures 5, 6, 7,



**Table 2: UCR elements used in this Study**

GATA	See Additional File 4	[78,79]
NfKb1	GGKGATYYAC	Aedine consensus [72]
NfKb2	KGGGAWHMMM	Anopheline consensus [72]
NfKb4	KGKGAWHHMM	Cross-species consensus
NfKb7	GGGAWHM	Subset of NfKb-related TFBSs, analyzed with higher stringency
JAK-STAT	TTCTAGGAA, TTTCTAAGAAA	[69,95]
BR-C Z1	TAAWWRACAARW	[96]
BR-C Z2	TTWWCTATTT	[96]
BR-C Z3	WAAACWWRW	[96]
BR-C Z4	RKAAASA	[96]
M1	MSMYCCACCCMCTCC	Aedes-specific
M2	SMCWCCCACCYMICY	Aedes-specific
M7	RCMRCARCARCACA	Anopheles-specific
M8	SCRMRCAGCARCAGCARCM	Anopheles-specific
M9	KTGTGTGTGTGTYG	Anopheles-specific
M12	AAGAATTTAAGAATT	Culex-specific
M13	TTYTTAAATTYTYAA	Culex-specific
M17	AACTTTTT	Culex-specific
M19	CCCMCCCCCCCCYY	miRNA pathway-specific
M21	SRBBSGCSGKRSSGG	miRNA pathway-specific
M24	CCRCMRCARCMGCWRCMRCM	siRNA pathway-specific
M26	TYTGAG	siRNA pathway-specific
M30	CGAGCKRCTSSRASYWGGTT	PIWI pathway-specific

IUPAC consensus codes used, wherein "K" is T or G; "Y" is T or C; "W" is T or A; "R" is A or G; "S" is C or G; "M" is C or A; "H" is A, T, or C; "V" is A, C or G; "B" is C, T, or G; "D" is A, T, or G.

Additional File 4) [84]. Due to the experimental design, elements that were identified in the species-specific search might also be enriched in a pathway-specific manner. For example, the *An. gambiae* elements M7 and M8 are enriched over the remainder of the genome in an siRNA pathway-specific manner for both *An. gambiae* and *Ae. aegypti* (Additional File 4). In addition, elements M12 and M13 are enriched in *Cx. pipiens* in a miRNA pathway-specific manner over the remainder of the genome, even though Fisher's Exact test did not indicate the *Cx. pipiens* has significantly more of these elements than *Ae. aegypti* or *An. gambiae* (Table 3). In contrast, the M1 and M2 elements of *Aedes aegypti* are strongly enriched in all SRRPs over the remainder of the aedine genome (Additional File 4). M9 (*An. gambiae*) and M17 (*Cx. pipiens*) were enriched in a species-specific manner over all other SRRP UCRs according to Fisher's Exact test (Table 3), however, they were not enriched in SRRP upstream regions compared to the remainder of the genome. Therefore, they are not likely to be important to regulation of small RNA metabolism.

Pathway-specific upstream elements were also identified. Of the five elements identified, M24 was the most interesting, because, it was enriched among the siRNA pathway UCRs across all three species. Conservation of within-pathway features across species supports the hypothesis of purifying selection in UCRs of small RNA regulatory components.

**Implications for Vector Competence**

Of the Argonaute protein subfamily members, Ago2, the anti-viral RNase H-type nuclease, is significantly more diverse among mosquito species than the miRNA pathway nuclease, Ago1. This finding suggests that anti-viral defense effectors are evolving at a faster rate than those involved in housekeeping functions. Anti-viral defense systems likely evolved to protect against entomopathogenic viruses. In turn, some entomopathogenic viruses probably evolved into arboviruses. When considering these pathways, it is important to keep in mind that evolutionary pressure exists on both the arbovirus and the vector mosquito to modulate the immune response.

Differences in UCR element motif patterns suggest that there are likely to be species-specific differences in transcriptional regulation of siRNA pathway components that could affect arbovirus vector competence. Loosely categorized, *An. gambiae* primarily transmits malaria parasites, *Cx. pipiens* transmits encephalitic arboviruses, and *Ae. aegypti* transmits important hemorrhagic arboviruses. For a given arthropod, low vector competence could arise from an effective immune response that clears the pathogen and prevents pathogen escape to the salivary glands. With this model in mind, we hypothesize that *An. gambiae* has a more effective antiviral immune response than either *Cx. pipiens* or *Ae. aegypti*. Characterization of these putative regulatory differences awaits further exploration in each species. The demonstrated ability of some mosquito species to serve as competent arbovirus vectors in

**Table 3: Selected UCR Motifs**

Motif	FET p value	BH	Enriched over genome?	Proportion of UCRs with Motif			Median Motifs per UCR		
				miRNA	siRNA	PIWI	miRNA	siRNA	PIWI
NFkB2	not sig. <sup>^</sup>	not sig.	No	0.188	0.250	0.100	1.0	1.0	1.0
NFkB4	not sig.	not sig.	No	0.313	0.438	0.460	1.0	1.0	1.0
NFkB7*	not sig.	not sig.	No	0.188	0.375	0.280	1.0	1.0	1.0
BRC_Z1	not sig.	not sig.	No	0.813	0.313	0.540	1.0	1.0	1.0
BRC_Z2	not sig.	not sig.	No	0.563	0.688	0.620	1.0	1.0	1.0
Pathway-specific							miRNA	siRNA	PIWI
M19 (miRNA)	not sig.	not sig.	Yes	0.188	0.100	0.033	4.5	1	1
M21 (miRNA)	not sig.	not sig.	Yes	0.219	0.100	0.067	2	2	1
M24 (siRNA)	not sig.	not sig.	Yes	0.063	0.400	0.200	1	14	1
M26 (siRNA)	not sig.	not sig.	Yes	0.031	0.250	0.033	2	3	2
M30 (PIWI)	not sig.	not sig.	Yes	0.000	0.000	0.060	NA	NA	5
Species-Specific				Aae	Aga	Cpi	Aae	Aga	Cpi
M1 (Aedes)	not sig.	not sig.	Yes	0.188	0.100	0.033	1.5	1.5	1
M2 (Aedes)	not sig.	not sig.	Yes	0.219	0.100	0.067	2	2	1
M3 (Aedes)	not sig.	not sig.	Yes	0.094	0.000	0.033	9	NA	1
M7 (Anopheles)#	not sig.	not sig.	Yes	0.063	0.400	0.200	10.5	1.5	1
M8 (Anopheles) #	not sig.	not sig.	Yes	0.031	0.250	0.033	10	1	1
M9 (Anopheles)	< 0.001	0.002	No	0.000	0.400	0.300	NA	6	1
M12 (Culex)+	not sig.	not sig.	Yes	0.031	0.000	0.067	1	NA	21
M13 (Culex)+	not sig.	not sig.	Yes	0.031	0.000	0.067	1	NA	21
M17 (Culex)	< 0.001	0.002	No	0.094	0.100	0.533	1	1	1

"^", see Additional File 4 for details. "\*", for NFkB7 only, significance scores were based on the presence of 2 or more hits per UCR. "#", enriched in an siRNA pathway specific manner. "+", enriched in an miRNA pathway specific manner. "BH", Benjamini-Hochberg test, cut-off = 0.05. "FET", Fisher's Exact test, cutoff = 0.01. "NA", not applicable. "Aae", *Ae. aegypti*, "Aga", *An. gambiae*, "Cpi", *Cx. pipiens*.

spite of the presence of anti-viral RNAi components, begs the question of what mechanisms are used by arboviruses to evade anti-viral RNAi defense.

This report brings to bear both the importance of PIWI family gene expansion in vector mosquitoes and the need for research focus in this area. A few exploratory studies have found evidence that PIWI proteins could be important in anti-viral defense. Transient silencing of *Ago3* by dsRNA injection increased O'nyong-nyong infection of *An. gambiae* [1]. However, characterization of PIWI pathway activity in mosquito anti-viral defense is clearly required before additional conclusions can be drawn. In *Drosophila*, an apparent paradox is found, wherein PIWI pathway proteins are reportedly expressed only in ovaries and are restricted to germline maintenance, however, PIWI pathway components DmeAub, DmeArmitage and DmeRm62 have been implicated in anti-viral processes [5,24,67,85]. Recent reports describing the presence of endogenous siRNAs (endo-siRNAs) are beginning to shed light in this area [29,30]. Small RNAs were found to control retrotransposons in somatic tissue in *D. melanogaster* in a Dicer2/Ago2 dependent manner.

**Conclusion**

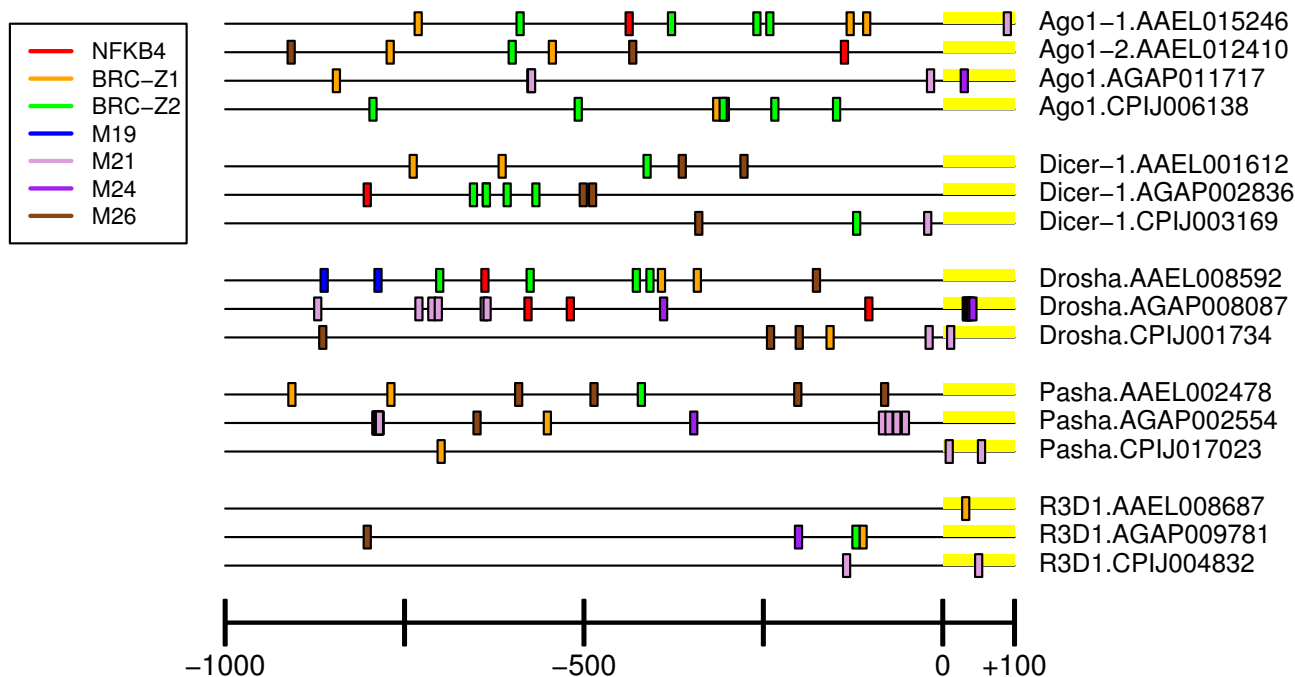
This genomics study provides important contextual information for both vector biologists and the RNAi community and highlights important differences between vector

mosquitoes and model organisms, such as *Drosophila*. Together, these data suggest a need to further investigate the relationship between vector competence and specific small RNA pathways across mosquito species, as well as potential arbovirus strategies for evasion of these defense mechanisms. However, it is still unclear what aspects of mosquito biology are driving evolution and gene expansion of Argonaute family genes.

**Methods**

**Putative ortholog identification**

Mosquito Argonaute/PIWI family genes were identified by tblastx search of the *Ae. aegypti* and *Cx. pipiens quinquefasciatus* genomes at Vectorbase.org, using the *An. gambiae* RNAi orthologs [86]. Putative orthologs of other SRRP components were identified by tblastx search of the databases using drosophilid RNAi genes. Argonaute and Rm62 family hits with E values of 0.0 -10<sup>-80</sup> were included in the analyses. Many other putative orthologs did not allow cut-offs with this high stringency, therefore, for all other groups, top hits with E values of < 10<sup>-40</sup> were used. Gene expression in each species was corroborated by the presence of a mosquito EST in the public database, as determined by blastn search, E value < 10<sup>-40</sup>. Putative isoforms from *Cx. pipiens*, based on alternate gene predictions at the same locus, were designated with a letter following the allele number (Additional File 1A).



**Figure 5**  
**miRNA pathway UCR Schematic.** Graphical representation of key cis UCR elements of miRNA pathway components. Scale is 1000 bases upstream to about 100 bases downstream of the beginning of the transcript.

The conservative naming conventions used for the PIWI genes could not be followed for Ago1 and Ago2. The gene names "Argonaute 1" and "Argonaute 2" are understood in the broad scientific community to be associated with the miRNA and siRNA pathways, respectively. To use a different numbering scheme for the expanded "Argonaute 1" and "Argonaute 2" gene families would have further added to confusion, therefore, we chose to assign the newly identified genes as paralogs rather than orthologs, even though the genes were not located on the same super-contig. The high level of sequence similarity provides additional support for this decision. For example, the open reading frames of *AaeAGO1-1* and *AaeAGO1-2* have 99.1% nucleotide identity but different upstream flanking sequences and introns. Similar searches were done for *CpiAGO2-1A* and *CpiAGO2-2*. Importantly, *CpiAGO2-1A* and *CpiAGO2-2* share only 48% nt identity but both are classified as 'Ago2-type' proteins according to amino acid sequence similarities and synapomorphies.

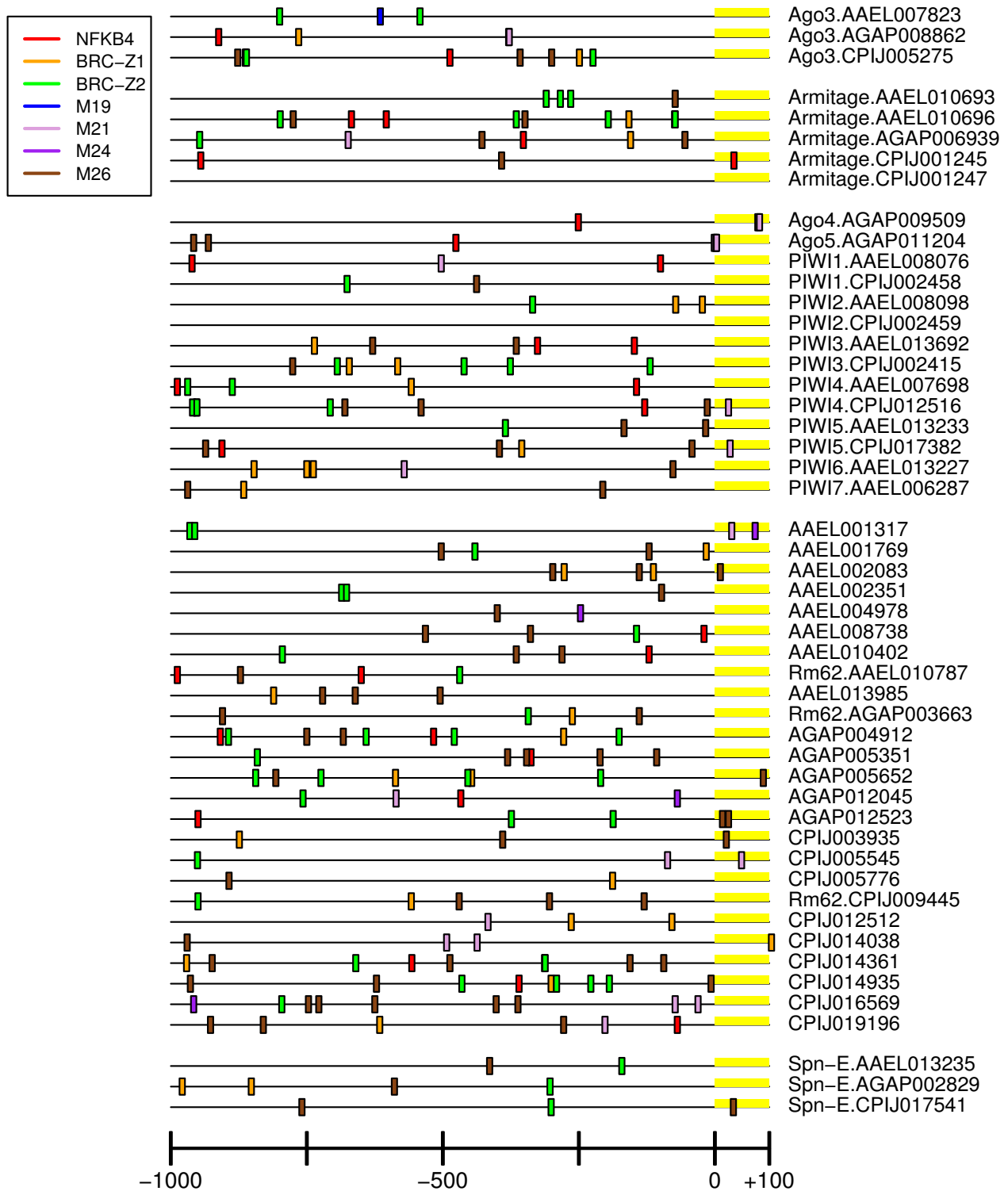
**Phylogenetic analyses**

Full-length PIWI subfamily proteins were compared. The C-terminal 250 amino acids of Argonaute subfamily proteins were compared. Partial Ago sequences were used, because full-length sequences were not available for some species. A Gonnet matrix was used in a CLUSTALW alignment; this weighted matrix is based on the supposition

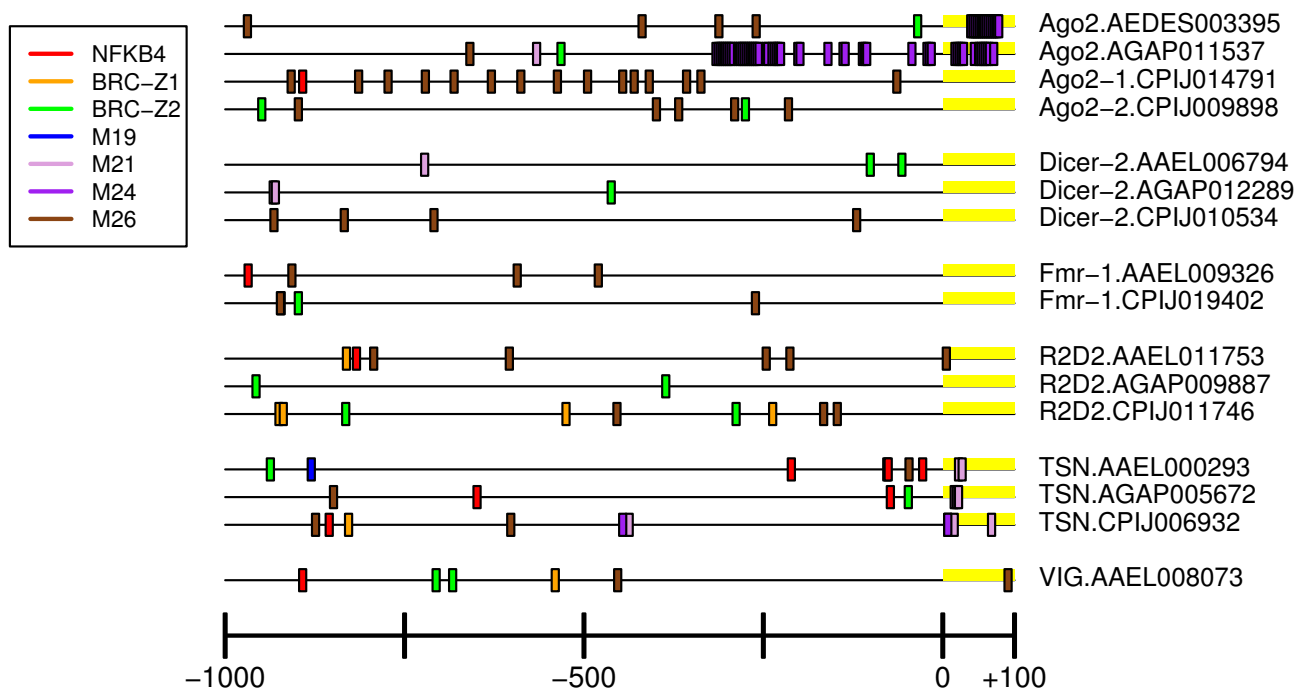
that any given amino acid substitution is influenced by neighboring amino acids, and thus provides a basis for characterizing protein families [87]. The alignment was used to generate a maximum likelihood phylogenetic tree using PROML (PHYLIP [88]). Phylogenetic relationships among proteins were estimated using a maximum likelihood analysis of amino acid sequences with the Jones-Taylor-Thornton probability model of amino acid changes [55]. This method assumes that taxa evolve independently and that substitutions at each amino acid site occur with a probability specified in the PAM 250 matrix. In addition, all amino acids are included in the analysis rather than just the phylogenetically informative sites. A bootstrap analysis was done with 1,000 pseudoreplicates.

**Motif Searches and Analyses**

Genomic sequences corresponding to approximately -1000 to +100 was extracted from genomic, 5' untranslated region, and partial coding region of each gene as listed at Vectorbase.org [86]. Motif searches were performed using MEME, Weeder, and MDScan [81-83,89-91]. Searches of the upstream control regions were performed separately for each of the following groups: (1) *Ae. aegypti*, (2) *An. gambiae*, (3) *Cx. pipiens*, (4) miRNA genes across species, (5) siRNA genes across species and (6) PIWI genes across



**Figure 6**  
**PIWI pathway UCR Schematic.** Graphical representation of key cis UCR elements of PIWI pathway components. Rm62-like genes list the accession number without any additional annotation. Scale is 1000 bases upstream to about 100 bases downstream of the beginning of the transcript.



**Figure 7**  
**siRNA pathway UCR Schematic.** Graphical representation of key cis UCR elements of siRNA pathway components. Scale is 1000 bases upstream to about 100 bases downstream of the beginning of the transcript.

species. The top two consensus sequences from each of the searches were included in subsequent analyses.

String analysis (R Biostrings, version 2.4.8) was used to locate and count the number of matches for each of the novel motifs and the previously described GATA, BRC Z1–Z4 and NFkB motifs [92]. The consensus sequences and the number of mismatches allowed for each motif are shown in Additional File 4. Motifs present multiple times in all UCRs were removed from further analysis. In addition, NFkB1 was removed from further consideration because it was not contained in any UCRs.

Full genome searches were performed to ensure that each element is enriched in the pathway or species of interest over the rest of the genome. This validation method provided support for the hypothesis that the particular motif was enriched in the pathway-specific or species-specific manner in which it was originally discovered. Each species genome was searched for the given sequence using the same number of mismatches denoted in Additional File 4. The output number was adjusted to the number of occurrences per supercontig length. The 95<sup>th</sup> percentile of this "mod count" was used as a cut-off to determine which motifs were enriched in each category. Motifs reported as

species-specific were enriched at or above the 95% percentile for the species in which it was identified.

Motifs reported as pathway-specific were enriched at or above the 95% percentile for the pathway in which it was identified for at least two of the three species. Experimentally identified motifs are interesting for biological reasons, therefore, significance by Fisher's Exact test or genome search was not required.

Fisher's Exact test adds further support and highlights motifs that are over-represented in a species or pathway-specific manner among all UCRs tested. The median number of hits was calculated using only those UCRs which contained at least one hit. Fisher's exact test was used to test for a relationship between each pathway (or species) versus the absence or presence for each motif, using a lower cut-off of one hit per motif. The Benjamini-Hochberg (BH) multiple testing adjustment method, which controls the False Discovery rate, was also performed, using a cut-off of 0.05 [93]. This cut-off ensures that 5% or fewer should be false discoveries or false positives and was used to identify motifs with significantly different proportions among pathways or species. The BH adjusted p-values as well as the gene proportions for

which each motif was observed at least once are shown in Table 3.

### Authors' contributions

CLC designed the study, performed bioinformatics analyses and wrote the manuscript. WCB4 performed the phylogenetic analyses and contributed to the manuscript. AH performed the UCR searches, summarized the results and carried out the statistical analyses. BDF edited the manuscript. All authors read and approved the manuscript.

### Additional material

#### Additional File 1

*SRRP components in vector mosquitoes compared to drosophilids. SRRP homologs, Argonaute Family and Rm62-like Protein alignments.*  
Click here for file  
[http://www.biomedcentral.com/content/supplementary/1471-2164-9-425-S1.pdf]

#### Additional File 2

*Argonaute Protein Family synapomorphies. Signature protein features of protein classes.*  
Click here for file  
[http://www.biomedcentral.com/content/supplementary/1471-2164-9-425-S2.pdf]

#### Additional File 3

*Rm62-like Protein Tree. Maximum Likelihood tree of Mosquito Rm62-like proteins compared to Drosophila Rm62-like helicases.*  
Click here for file  
[http://www.biomedcentral.com/content/supplementary/1471-2164-9-425-S3.jpeg]

#### Additional File 4

*UCR motif statistical analyses. Full genome search data; locations of elements in each UCR.*  
Click here for file  
[http://www.biomedcentral.com/content/supplementary/1471-2164-9-425-S4.xls]

### Acknowledgements

We are thankful for the availability of sequence data through Vectorbase and the mosquito genome sequencing projects. We also thank P. Arensburg for help with annotation of *Culex* genes. This work was funded by grant AI060960 from the National Institutes of Health.

### References

- Keene KM, Foy BD, Sanchez-Vargas I, Beaty BJ, Blair CD, Olson KE: **RNA interference acts as a natural antiviral response to O'nyong-nyong virus (Alphavirus; Togaviridae) infection of Anopheles gambiae.** *Proc Natl Acad Sci USA* 2004, **101(49)**:17240-17245.
- Nakahara K, Kim K, Sciulli C, Dowd SR, Minden JS, Carthew RW: **Targets of microRNA regulation in the Drosophila oocyte proteome.** *Proc Natl Acad Sci USA* 2005, **102(34)**:12023-12028.
- Olsen PH, Ambros V: **The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation.** *Dev Biol* 1999, **216(2)**:671-680.
- Saito K, Nishida KM, Mori T, Kawamura Y, Miyoshi K, Nagami T, Siomi H, Siomi MC: **Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the Drosophila genome.** *Genes Dev* 2006, **20(16)**:2214-2222.
- Vagin VV, Sigova A, Li C, Seitz H, Gvozdev V, Zamore PD: **A distinct small RNA pathway silences selfish genetic elements in the germline.** *Science* 2006, **313(5785)**:320-324.
- Cerutti L, Mian N, Bateman A: **Domains in gene silencing and cell differentiation proteins: the novel PAZ domain and redefinition of the Piwi domain.** *Trends Biochem Sci* 2000, **25(10)**:481-482.
- Joshua-Tor L: **The Argonautes.** *Cold Spring Harb Symp Quant Biol* 2006, **71**:67-72.
- Hammond SM, Boettcher S, Caudy AA, Kobayashi R, Hannon GJ: **Argonaute2, a Link Between Genetic and Biochemical Analyses of RNAi.** *Science* 2001, **293(5532)**:1146-1150.
- Behura SK: **Insect microRNAs: Structure, function and evolution.** *Insect Biochem Mol Biol* 2007, **37(1)**:3-9.
- Mello CC, Conte D Jr: **Revealing the world of RNA interference.** *Nature* 2004, **431(7006)**:338-342.
- Williams RV, Rubin GM: **ARGONAUTE1 is required for efficient RNA interference in Drosophila embryos.** *Proc Natl Acad Sci USA* 2002, **99(10)**:6889-6894.
- Hutvagner G, Zamore PD: **A microRNA in a multiple-turnover RNAi enzyme complex.** *Science* 2002, **297(5589)**:2056-2060.
- Miyoshi K, Tsukumo H, Nagami T, Siomi H, Siomi MC: **Slicer function of Drosophila Argonautes and its involvement in RISC formation.** *Genes Dev* 2005, **19(23)**:2837-2848.
- Okamura K, Ishizuka A, Siomi H, Siomi MC: **Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways.** *Genes Dev* 2004, **18(14)**:1655-1666.
- Winter F, Edaye S, Huttenhofer A, Brunel C: **Anopheles gambiae miRNAs as actors of defence reaction against Plasmodium invasion.** *Nucleic Acids Res* 2007, **35(20)**:6953-6962.
- Liu J, Valencia-Sanchez MA, Hannon GJ, Parker R: **MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies.** *Nat Cell Biol* 2005, **7(7)**:719-723.
- Souret FF, Kastenmayer JP, Green PJ: **AtXRN4 degrades mRNA in Arabidopsis and its substrates include selected miRNA targets.** *Mol Cell* 2004, **15(2)**:173-183.
- Han J, Lee Y, Yeom KH, Nam JW, Heo I, Rhee JK, Sohn SY, Cho Y, Zhang BT, Kim VN: **Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex.** *Cell* 2006, **125(5)**:887-901.
- Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S, et al.: **The nuclear RNase III Drosha initiates microRNA processing.** *Nature* 2003, **425(6956)**:415-419.
- Behm-Ansmant I, Rehwinkel J, Doerks T, Stark A, Bork P, Izaurralde E: **mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCPI1:DCP2 decapping complexes.** *Genes Dev* 2006, **20(14)**:1885-1898.
- Ketting RF, Fischer SE, Bernstein E, Sijen T, Hannon GJ, Plasterk RH: **Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans.** *Genes Dev* 2001, **15(20)**:2654-2659.
- Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A: **Identification of mammalian microRNA host genes and transcription units.** *Genome Res* 2004, **14(10A)**:1902-1910.
- van Rij RP, Saleh MC, Berry B, Foo C, Houk A, Antoniewski C, Andino R: **The RNA silencing endonuclease Argonaute 2 mediates specific antiviral immunity in Drosophila melanogaster.** *Genes Dev* 2006, **20(21)**:2985-2995.
- Zamboni RA, Vakharia VN, Wu LP: **RNAi is an antiviral immune response against a dsRNA virus in Drosophila melanogaster.** *Cellular Microbiology* 2006, **8(5)**:880-889.
- Campbell CL, Keene KM, Brackney DE, Olson KE, Blair CD, Wilusz J, Foy BD: **Aedes aegypti uses RNA interference in defense against Sindbis virus infection.** *BMC Microbiol* 2008, **8**:47.
- Franz AVW, Sanchez-Vargas I, Adelman ZN, Blair CD, Beaty BJ, James AA, Olson KE: **Engineering RNA interference-based resistance to dengue virus type 2 in genetically modified Aedes aegypti.** *Proc Natl Acad Sci USA* 2006, **103(11)**:4198-4203.
- Sanchez-Vargas I, Travanty EA, Keene KM, Franz AVW, Beaty BJ, Blair CD, Olson KE: **RNA interference, arthropod-borne viruses, and mosquitoes.** *Virus Res* 2004, **102(1)**:65-74.

28. Chung WJ, Okamura K, Martin R, Lai EC: **Endogenous RNA Interference Provides a Somatic Defense against Drosophila Transposons.** *Curr Biol* 2008, **18(11)**:795-802.
29. Ghildiyal M, Seitz H, Horwich MD, Li C, Du T, Lee S, Xu J, Kittler EL, Zapp ML, Weng Z, et al.: **Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells.** *Science* 2008, **320(5879)**:1077-1081.
30. Kawamura Y, Saito K, Kin T, Ono Y, Asai K, Sunohara T, Okada TN, Siomi MC, Siomi H: **Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells.** *Nature* 2008, **453(7196)**:793-797.
31. Kalmykova AI, Klenov MS, Gvozdev VA: **Argonaute protein PIWI controls mobilization of retrotransposons in the Drosophila male germline.** *Nucleic Acids Res* 2005, **33(6)**:2052-2059.
32. Sarot E, Payen-Groschene G, Bucheton A, Pelisson A: **Evidence for a piwi-dependent RNA silencing of the gypsy endogenous retrovirus by the Drosophila melanogaster flamenco gene.** *Genetics* 2004, **166(3)**:1313-1321.
33. Savitsky M, Kwon D, Georgiev P, Kalmykova A, Gvozdev V: **Telomere elongation is under the control of the RNAi-based mechanism in the Drosophila germline.** *Genes Dev* 2006, **20(3)**:345-354.
34. Klattenhoff C, Bratu DP, McGinnis-Schultz N, Koppetsch BS, Cook HA, Theurkauf WE: **Drosophila rasiRNA pathway mutations disrupt embryonic axis specification through activation of an ATR/Chk2 DNA damage response.** *Dev Cell* 2007, **12(1)**:45-55.
35. Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ: **Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila.** *Cell* 2007, **128(6)**:1089-1103.
36. Black WCt, Bennett KE, Gorrochotegui-Escalante N, Barillas-Mury CV, Fernandez-Salas I, de Lourdes Munoz M, Farfan-Ale JA, Olson KE, Beaty BJ: **Flavivirus susceptibility in Aedes aegypti.** *Arch Med Res* 2002, **33(4)**:379-388.
37. Gubler DJ: **The global emergence/resurgence of arboviral diseases as public health problems.** *Arch Med Res* 2002, **33(4)**:330-342.
38. Guerra CA, Snow RW, Hay SI: **Mapping the global extent of malaria in 2005.** *Trends Parasitol* 2006, **22(8)**:353-358.
39. Kamath S, Das AK, Parikh FS: **Chikungunya.** *J Assoc Physicians India* 2006, **54**:725-726.
40. Kramer LD, Styer LM, Ebel GD: **A Global Perspective on the Epidemiology of West Nile Virus.** *Annu Rev Entomol* 2008, **53**:61-81.
41. Kuniholm MH, Wolfe ND, Huang CJ, Mpoudi-Ngole E, Tamoufe U, LeBreton M, Burke DS, Gubler DJ: **Seroprevalence and distribution of Flaviviridae, Togaviridae, and Bunyaviridae arboviral infections in rural Cameroonian adults.** *Am J Trop Med Hyg* 2006, **74(6)**:1078-1083.
42. Stark A, Lin MF, Kheradpour P, Pedersen JS, Parts L, Carlson JW, Crosby MA, Rasmussen MD, Roy S, Deoras AN, et al.: **Discovery of functional elements in 12 Drosophila genomes using evolutionary signatures.** *Nature* 2007, **450(7167)**:219-232.
43. Caudy AA, Ketting RF, Hammond SM, Denli AM, Bathoorn AM, Tops BB, Silva JM, Myers MM, Hannon GJ, Plasterk RH: **A micrococcal nuclease homologue in RNAi effector complexes.** *Nature* 2003, **425(6956)**:411-414.
44. Caudy AA, Myers M, Hannon GJ, Hammond SM: **Fragile X-related protein and VIG associate with the RNA interference machinery.** *Genes Dev* 2002, **16(19)**:2491-2496.
45. Cook HA, Koppetsch BS, Wu J, Theurkauf WE: **The Drosophila SDE3 homolog armitage is required for oskar mRNA silencing and embryonic axis specification.** *Cell* 2004, **116(6)**:817-829.
46. Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wincker P, Clark AG, Ribeiro JM, Wides R, et al.: **The genome sequence of the malaria mosquito Anopheles gambiae.** *Science* 2002, **298(5591)**:129-149.
47. Ishizuka A, Siomi MC, Siomi H: **A Drosophila fragile X protein interacts with components of RNAi and ribosomal proteins.** *Genes Dev* 2002, **16(19)**:2497-2508.
48. Lawson D, Arensburger P, Atkinson P, Besansky NJ, Bruggner RV, Butler R, Campbell KS, Christophides GK, Christley S, Dialynas E, et al.: **VectorBase: a home for invertebrate vectors of human pathogens.** *Nucleic Acids Res* 2007:D503-505.
49. Lee YS, Nakahara K, Pham JW, Kim K, He Z, Sontheimer EJ, Carthew RW: **Distinct roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA silencing pathways.** *Cell* 2004, **117(1)**:69-81.
50. Liu Q, Rand TA, Kalidas S, Du F, Kim HE, Smith DP, Wang X: **R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway.** *Science* 2003, **301(5641)**:1921-1925.
51. Nene V, Wortman JR, Lawson D, Haas B, Kodira C, Tu ZJ, Loftus B, Xi Z, Megy K, Grabherr M, et al.: **Genome Sequence of Aedes aegypti, a Major Arbovirus Vector.** *Science* 2007, **316(5832)**:1718-1723.
52. Pham JW, Pellino JL, Lee YS, Carthew RW, Sontheimer EJ: **A Dicer-2-dependent 80s complex cleaves targeted mRNAs during RNAi in Drosophila.** *Cell* 2004, **117(1)**:83-94.
53. Sharakhova MV, Hammond MP, Lobo NF, Krzywinski J, Unger MF, Hillenmeyer ME, Bruggner RV, Birney E, Collins FH: **Update of the Anopheles gambiae PEST genome assembly.** *Genome Biol* 2007, **8(1)**:R5.
54. Mead EA, Tu Z: **Cloning, characterization, and expression of microRNAs from the Asian malaria mosquito, Anopheles stephensi.** *BMC Genomics* 2008, **9**.
55. Jones DT, Taylor WR, Thornton JM: **The rapid generation of mutation data matrices from protein sequences.** *Comput Appl Biosci* 1992, **8(3)**:275-282.
56. Obbard DJ, Jiggins FM, Halligan DL, Little TJ: **Natural selection drives extremely rapid evolution in antiviral RNAi genes.** *Curr Biol* 2006, **16(6)**:580-585.
57. Griffiths-Jones S: **The microRNA Registry.** *Nucleic Acids Res* 2004:D109-111.
58. Yigit E, Batista PJ, Bei Y, Pang KM, Chen CC, Tolia NH, Joshua-Tor L, Mitani S, Simard MJ, Mello CC: **Analysis of the C. elegans Argonaute family reveals that distinct Argonautes act sequentially during RNAi.** *Cell* 2006, **127(4)**:747-757.
59. Foley DH, Bryan JH, Yeates D, Saul A: **Evolution and systematics of Anopheles: insights from a molecular phylogeny of Australasian mosquitoes.** *Mol Phylogenet Evol* 1998, **9(2)**:262-275.
60. Krzywinski J, Grushko OG, Besansky NJ: **Analysis of the complete mitochondrial DNA from Anopheles funestus: an improved dipteran mitochondrial genome annotation and a temporal dimension of mosquito evolution.** *Mol Phylogenet Evol* 2006, **39(2)**:417-423.
61. Besansky NJ: **A retrotransposable element from the mosquito Anopheles gambiae.** *Mol Cell Biol* 1990, **10(3)**:863-871.
62. Biedler J, Tu Z: **Non-LTR retrotransposons in the African malaria mosquito, Anopheles gambiae: unprecedented diversity and evidence of recent activity.** *Mol Biol Evol* 2003, **20(11)**:1811-1825.
63. Biedler JK, Tu Z: **The Juan non-LTR retrotransposon in mosquitoes: genomic impact, vertical transmission and indications of recent and widespread activity.** *BMC Evol Biol* 2007, **7**:112.
64. Coy MR, Tu Z: **Genomic and evolutionary analyses of Tango transposons in Aedes aegypti, Anopheles gambiae and other mosquito species.** *Insect Mol Biol* 2007, **16(4)**:411-421.
65. Brower-Toland B, Findley SD, Jiang L, Liu L, Yin H, Dus M, Zhou P, Elgin SC, Lin H: **Drosophila PIWI associates with chromatin and interacts directly with HPIa.** *Genes Dev* 2007, **21(18)**:2300-2311.
66. Lei EP, Corces VG: **RNA interference machinery influences the nuclear organization of a chromatin insulator.** *Nat Genet* 2006, **38(8)**:936-941.
67. Aravin AA, Klenov MS, Vagin VV, Bantignies F, Cavalli G, Gvozdev VA: **Dissection of a natural RNA silencing process in the Drosophila melanogaster germ line.** *Mol Cell Biol* 2004, **24(15)**:6742-6750.
68. Andolfatto P: **Adaptive evolution of non-coding DNA in Drosophila.** *Nature* 2005, **437(7062)**:1149-1152.
69. Dostert C, Jouanguy E, Irving P, Troxler L, Galiana-Arnoux D, Hetru C, Hoffmann JA, Imler JL: **The Jak-STAT signaling pathway is required but not sufficient for the antiviral response of drosophila.** *Nat Immunol* 2005, **6(9)**:946-953.
70. Zambon RA, Nandakumar M, Vakharia VN, Wu LP: **The Toll pathway is important for an antiviral response in Drosophila.** *Proc Natl Acad Sci USA* 2005, **102(20)**:7257-7262.
71. Waterhouse RM, Kriventseva EV, Meister S, Xi Z, Alvarez KS, Bartholomay LC, Barillas-Mury C, Bian G, Blandin S, Christensen BM, et al.: **Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes.** *Science* 2007, **316(5832)**:1738-1743.

72. Shin SW, Kokoza V, Bian G, Cheon HM, Kim YJ, Raikhel AS: **RELI, a homologue of Drosophila dorsal, regulates toll antifungal immune pathway in the female mosquito Aedes aegypti.** *J Biol Chem* 2005, **280(16)**:16499-16507.
73. Sanders HR, Foy BD, Evans AM, Ross LS, Beaty BJ, Olson KE, Gill SS: **Sindbis virus induces transport processes and alters expression of innate immunity pathway genes in the midgut of the disease vector, Aedes aegypti.** *Insect Biochem Mol Biol* 2005, **35(11)**:1293-1307.
74. Scadden AD: **The RISC subunit Tudor-SN binds to hyper-edited double-stranded RNA and promotes its cleavage.** *Nat Struct Mol Biol* 2005, **12(6)**:489-496.
75. von Kalm L, Crossgrove K, Von Seggern D, Guild GM, Beckendorf SK: **The Broad-Complex directly controls a tissue-specific response to the steroid hormone ecdysone at the onset of Drosophila metamorphosis.** *Embo J* 1994, **13(15)**:3505-3516.
76. Zhu J, Chen L, Raikhel AS: **Distinct roles of Broad isoforms in regulation of the 20-hydroxyecdysone effector gene, Vitellogenin, in the mosquito Aedes aegypti.** *Mol Cell Endocrinol* 2007, **267(1-2)**:97-105.
77. Dittmer NT, Raikhel AS: **Analysis of the mosquito lysosomal aspartic protease gene: an insect housekeeping gene with fat body-enhanced expression.** *Insect Biochem Mol Biol* 1997, **27(4)**:323-335.
78. Giannoni F, Muller HM, Vizioli J, Catteruccia F, Kafatos FC, Crisanti A: **Nuclear factors bind to a conserved DNA element that modulates transcription of Anopheles gambiae trypsin genes.** *J Biol Chem* 2001, **276(1)**:700-707.
79. Martin D, Piulachs MD, Raikhel AS: **A novel GATA factor transcriptionally represses yolk protein precursor genes in the mosquito Aedes aegypti via interaction with the CtBP corepressor.** *Mol Cell Biol* 2001, **21(1)**:164-174.
80. Bailey TL, Elkan C: **The value of prior knowledge in discovering motifs with MEME.** *Proc Int Conf Intell Syst Mol Biol* 1995, **3**:21-29.
81. Bailey TL, Williams N, Misleh C, Li WW: **MEME: discovering and analyzing DNA and protein sequence motifs.** *Nucleic Acids Res* 2006, **W369-373**.
82. Liu XS, Brutlag DL, Liu JS: **An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments.** *Nat Biotechnol* 2002, **20(8)**:835-839.
83. Pavesi G, Mauri G, Pesole G: **An algorithm for finding signals of unknown length in DNA sequences.** *Bioinformatics* 2001, **17(Suppl 1)**:S207-214.
84. Davidson EH: **Genomic Regulatory Systems: Development and Evolution.** San Diego: Academic Press; 2001.
85. Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, Siomi H, Siomi MC: **A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila.** *Science* 2007, **315(5818)**:1587-1590.
86. **VectorBase: a home for invertebrate vectors of human pathogens** [<http://www.vectorbase.org/>]. [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=17145709](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17145709)
87. Gonnet GH, Cohen MA, Benner SA: **Exhaustive matching of the entire protein sequence database.** *Science* 1992, **256(5062)**:1443-1445.
88. Felsenstein J: **PHYLIP-Phylogeny Inference Package (Version 3.2).** *Cladistics* 1989, **5**:164-166.
89. **An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments** [<http://ai.stanford.edu/~xslu/MDscan/>]. [http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list\\_uids=12101404](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12101404)
90. **An algorithm for finding signals of unknown length in DNA sequences** [[http://www.pesolelab.it/index.php?mod=04\\_Tools](http://www.pesolelab.it/index.php?mod=04_Tools)].
91. Pages H, Gentleman R, DebRoy S: **Biostrings: String objects representing biological sequences, and matching algorithms. R package version 2.4.8.**
92. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *J R Statist Soc Bulletin* 1995, **57**:289-300.
93. Ambros V, Chen X: **The regulation of genes and genomes by small RNAs.** *Development* 2007, **134(9)**:1635-1641.
94. Lin CC, Chou CM, Hsu YL, Lien JC, Wang YM, Chen ST, Tsai SC, Hsiao PW, Huang CJ: **Characterization of two mosquito STATs, AaSTAT and CtSTAT. Differential regulation of tyrosine phosphorylation and DNA binding activity by lipopolysaccharide treatment and by Japanese encephalitis virus infection.** *J Biol Chem* 2004, **279(5)**:3308-3317.
95. Knuppel R, Dietze P, Lehnberg W, Frech K, Wingender E: **TRANSFAC retrieval program: a network model database of eukaryotic transcription regulating sequences and proteins.** *J Comput Biol* 1994, **1(3)**:191-198.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

