



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

## Duplication and diversification of Dipteran Argonaute genes, and the evolutionary divergence of Piwi and Aubergine

**Citation for published version:**

Lewis, SH, Salmela, H & Obbard, D 2016, 'Duplication and diversification of Dipteran Argonaute genes, and the evolutionary divergence of Piwi and Aubergine', *Genome Biology and Evolution*.  
<https://doi.org/10.1093/gbe/evw018>

**Digital Object Identifier (DOI):**

[10.1093/gbe/evw018](https://doi.org/10.1093/gbe/evw018)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Peer reviewed version

**Published In:**

Genome Biology and Evolution

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



Title

Duplication and diversification of Dipteran Argonaute genes, and the evolutionary divergence of Piwi and Aubergine

Running title

Duplication and diversification of Dipteran Argonautes

Authors and affiliations

Samuel H. Lewis<sup>1\*</sup>

Heli Salmela<sup>2</sup>

Darren J. Obbard<sup>3,4</sup>

<sup>1</sup>Department of Genetics, University of Cambridge, Downing Street, Cambridge, CB2 3EH

<sup>2</sup>Department of Biosciences, Centre of Excellence in Biological Interactions, University of Helsinki, Helsinki, Finland

<sup>3</sup>Institute of Evolutionary Biology, University of Edinburgh, Kings Buildings, EH9 3JT, United Kingdom

<sup>4</sup>Centre for Immunity, Infection and Evolution, University of Edinburgh, Kings Buildings, EH9 3JT, United Kingdom

\* Author for correspondence: Samuel H. Lewis, Department of Genetics, University of Cambridge, sam.lewis@gen.cam.ac.uk

Data deposition

All new sequences have been submitted to Genbank as KR012647-KR012696.

## Abstract

Genetic studies of *Drosophila melanogaster* have provided a paradigm for RNAi in arthropods, in which the miRNA and antiviral pathways are each mediated by a single Argonaute (Ago1 and Ago2) and germline suppression of transposable elements is mediated by a trio of Piwi-subfamily Argonaute proteins (Ago3, Aub and Piwi). Without a suitable evolutionary context, deviations from this can be interpreted as derived or idiosyncratic. Here we analyse the evolution of Argonaute genes across the genomes and transcriptomes of 86 Dipteran species, showing that variation in copy number can occur rapidly, and that there is constant flux in some RNAi mechanisms. The lability of the RNAi pathways is illustrated by the divergence of Aub and Piwi (182-156 million years ago), independent origins of multiple Piwi-family genes in *Aedes* mosquitoes (less than 25mya), and the recent duplications of Ago2 and Ago3 in the tsetse fly *Glossina morsitans*. In each case the tissue-specificity of these genes has altered, suggesting functional divergence or innovation, and consistent with the action of dynamic selection pressures across the Argonaute gene family. We find there are large differences in evolutionary rates and gene turnover between pathways, and that paralogues of Ago2, Ago3 and Piwi/Aub show contrasting rates of evolution after duplication. This suggests that Argonautes undergo frequent evolutionary expansions that facilitate functional divergence.

## Key words

RNAi, Argonaute, Piwi, gene duplication, Diptera

## Introduction

Argonaute genes of the Ago and Piwi subfamilies mediate a broad range of processes from development to antiviral immunity, and are found in almost all eukaryotes (Cerutti and Casas-Mollano 2006). They constitute an ancient gene family that was present in the common ancestor of extant prokaryotes and eukaryotes (reviewed in Swarts et al., 2014), and which diverged into Ago and Piwi subfamilies early in eukaryotic evolution (Cerutti and Casas-Mollano 2006; Mukherjee et al. 2013). The Argonautes are effectors in the RNA interference-related (RNAi) pathways, which can be broadly defined as a system of nucleic acid manipulation through complementary base pairing between small RNA (sRNA) guides and long nucleic acid targets. Each sRNA is loaded into an Argonaute protein, which it guides to a target nucleic acid, resulting in cleavage or translational inhibition of the target (reviewed in Sarkies & Miska 2014). Three broad classes of sRNA can be defined based on their sizes and interactors (reviewed in Kim, Han, & Siomi 2009): short interfering RNAs (siRNAs) are ~21-24 nucleotides (nt) long and are produced from viruses, transposable elements (TEs), and some long dsRNA products in the soma; microRNAs (miRNAs) are generally ~22-23nt long and are derived from host-encoded hairpin loops; and Piwi-interacting RNAs (piRNAs) are 24-29nt long, derived largely from intergenic repetitive elements (e.g. TEs) in the germline, and exclusively bind Piwi-subfamily Argonaute proteins.

RNAi is well studied in *Arabidopsis thaliana*, where the Argonaute gene was first identified (Bohmert et al. 1998), and in the nematode *Caenorhabditis elegans*, where the RNAi mechanism was first characterized (Fire et al. 1998). Subsequent studies have reported Argonautes with diverse functions and differences in copy number across different eukaryotic clades (Mukherjee et al. 2013), illustrating that RNAi pathways have a dynamic evolutionary history. For example, in plants RNAi-mediated suppression of TEs is directed by shorter sRNAs than in animals, and is mediated by Agos not Piwis (which they lack completely (Cerutti and Casas-Mollano 2006)) (reviewed in Parent et al. 2012). Differences in Argonaute copy number and function are also found in the animals. In the protostomes, the planarian *Schmidtea mediterranea* has 9 Piwi homologues (Palakodeti et al. 2008), two

of which (*smedwi-2* & *smedwi-3*) play vital roles in regeneration by facilitating the differentiation of pluripotent neoblasts (Reddien et al. 2005; Palakodeti et al. 2008). In contrast, Piwi and their associated piRNAs have been lost independently in several lineages of nematodes, with TE suppression carried out instead by DNA methylation mediated by RNA-dependent RNA polymerase and Dicer (Sarkies et al. 2015). Interestingly, this loss of Piwi has been accompanied by a massive expansion of other Argonaute genes in nematodes, with *C. elegans* encoding 25 Argonautes, 18 of which fall into the divergent worm-specific Ago (WAGO) clade: these associate with a novel class of sRNA (22G-RNAs) and carry out derived functions such as epigenetic memory formation (reviewed in Buck and Blaxter 2013).

Recent genome sequences and experimental data from isolated taxa have also revealed numerous arthropods with duplicates of Argonautes, some of which have novel and divergent functions. For example, the tick *Ixodes scapularis* has 3 Ago2 paralogues, only 2 of which appear to function in antiviral defence (Schnettler et al. 2014). Larger expansions are seen in the aphid *Acyrtosiphon pisum*, which has 2 paralogues of Ago3 and 8 paralogues of Piwi, some of which are expressed in the soma (in contrast to *D. melanogaster*, where they are predominantly germline-specific) (Lu et al. 2011). Additionally, these Piwi paralogues are differentially expressed in aphid reproductive morphs, suggesting that they may have specialized to function in different reproductive strategies (Lu et al. 2011).

Despite this diversity, much of our functional understanding of arthropod Argonautes comes from studies of *D. melanogaster* (Kataoka et al. 2001; Li et al. 2002; Pal-Bhadra et al. 2004; Vagin et al. 2004; Kalmykova et al. 2005; van Rij et al. 2006; Chung et al. 2008; Czech et al. 2008), which has two Ago-subfamily genes. Ago1 binds miRNAs and regulates gene expression by inhibiting translation of host transcripts (reviewed in Eulalio et al. 2008). Ago2 binds siRNAs from two sources. First, virus derived small interfering RNAs (viRNAs), which guide Ago2 to cleave viruses or their transcripts, forming an integral part of the antiviral defence mechanism (Li et al. 2002; van Rij et al. 2006). Second, endogenous (endo)-siRNAs, which are derived from TEs, overlapping UTRs and other repetitive sequences in

the soma (Chung et al. 2008; Czech et al. 2008). *D. melanogaster* also encodes three Piwi-subfamily proteins, which bind piRNAs in the germline and surrounding tissues: Ago3, Aubergine (Aub) and Piwi (reviewed in Iwasaki et al. 2015). The piRNAs are differentiated from miRNAs and siRNAs in *D. melanogaster* by their Dicer-independent production and their amplification through the “Ping-Pong” pathway, a positive feedback loop involving Ago3 and Aub (Li et al. 2009). In *D. melanogaster*, piRNAs guide Piwi to TEs in euchromatin, where it inhibits transposition (Kalmykova et al. 2005) by directing the formation of heterochromatin (Sienski et al. 2012).

However, comprehensive analysis of Argonaute evolution at a eukaryotic, or even metazoan, scale is hindered by limited taxon sampling, wide variation in evolutionary rate, and the presence of ancient and recent duplications and losses (discussed by Philippe et al. 2011). The Diptera provide an opportunity to study Argonaute evolution in an order that is densely sampled and less divergent, but still shows variation in Argonaute copy number and function. Previous reports of Argonaute duplication in the Diptera have been limited to isolated taxa, such as the house fly *Musca domestica* (Scott et al. 2014), *Drosophila pseudoobscura* (Hain et al. 2010), and three mosquito species (Campbell et al. 2008). These mosquito duplicates appear to have evolved derived functions: several Piwi paralogues in *Aedes aegypti* (Vodovar et al. 2012; Schnettler et al. 2013) and *Aedes albopictus* (Morazzani et al. 2012) are expressed in the soma, and at least one of the somatically-expressed Piwi duplicates in *Aedes albopictus* appears to have functionally diverged to a novel antiviral function (Schnettler et al. 2013).

Gene duplications, such as those that gave rise to the diversity of eukaryotic RNAi pathways, are often associated with changes in evolutionary rate (reviewed in Hahn 2009), and *Drosophila* duplicates that evolve a new function often evolve more rapidly (Assis and Bachtrog 2013). However, the subsequent duration of this rate change can vary considerably, either changing only briefly following duplication (e.g. Nielsen et al. 2010), or persisting in all branches subtending the duplication event (e.g. Morandin et al. 2014). Additionally, rate change after duplication can be symmetrical or asymmetrical between the

resulting paralogues: if both paralogues specialize to different pre-existing functions (subfunctionalization) they are expected to have roughly symmetrical evolutionary rates; whereas if one paralogue undergoes neofunctionalization it is expected to evolve more rapidly than the other paralogue, resulting in asymmetrical rates (Hittinger and Carroll 2007). Such a difference is seen after duplication of desaturase genes in *Drosophila*, which play key roles in evolutionary divergence and speciation through their contribution to the formation of cuticular hydrocarbons (Keays et al. 2011). These characteristic patterns of selection following duplication therefore enable us to use analyses of evolutionary rate to gain an insight into functional evolution.

Here we take advantage of the diversity available in the sequenced genomes and transcriptomes of Diptera to analyse patterns of Argonaute duplication and sequence evolution across 86 species. Contrasting rates of duplication and evolution are commonly associated with differences in function and selection pressure. We find a higher rate of protein evolution in Ago2 and Ago3, a higher rate of gene turnover in Ago2 and Piwi/Aub, and we estimate the date of the duplication that led to the separate Piwi and Aub subclades. We also find that paralogues of Ago2, Ago3 and Piwi/Aub evolve more rapidly after duplication, indicating potential divergence into novel and strongly selected functions.

## Materials and Methods

### Identification of Argonaute homologues

We used tblastx and tblastn (Altschul et al. 1997) to identify Argonaute homologues in the genomes and transcriptomes of 86 Dipteran species found in Genbank, Flybase, Vectorbase, Diptex, the NCBI Transcriptome Shotgun Assembly or other unpublished transcriptomes (see Supplementary Materials for a detailed list of sources; novel sequences have been submitted to Genbank as KR012647-KR012696). For each species, we used Argonautes from the closest well-annotated relative as queries, or *D. melanogaster* if no homologue from a close relative was available. Where blast returned multiple partial hits, we assigned hits to the correct query sequence by aligning all hits from the target species to all



Argonautes from the query species, and inferring a neighbour-joining tree. For each query sequence, partial blast hits were then manually curated into complete genes using Geneious v5.6.2 (<http://www.geneious.com>, Kearse et al. 2012). For some species of *Drosophila*, PCR and Sanger sequencing was used as no transcriptomic or genomic data were available (see Supplementary Materials for details of genes).

#### Phylogenetic analysis of Dipteran Argonautes

We initially assigned homologues into subclades (Ago1, Ago2, Ago3 and Piwi/Aub) based on a Bayesian gene tree rooted between the Ago and Piwi subfamilies, with ambiguous alignment positions removed using Gblocks (Castresana 2000) and with the wasp *Nasonia vitripennis* as the outgroup for each subclade. We repeated this analysis with three other arthropod species as outgroup (*Bemisia tabaci*, *Bombyx mori* and *Tribolium castaneum*), and find that in each case the same Dipteran genes are classified into the Ago1, Ago2, Ago3 and Piwi/Aub subclades. To minimise the loss of information when removing ambiguous positions, we re-inferred separate Bayesian gene trees for each subclade with no outgroup, using new alignments with ambiguous positions identified by eye and removed (see Supplementary Materials for alignments). Sequences were aligned using translational MAFFT (Kato et al. 2002) with default parameters. All phylogenies were inferred using the Bayesian approach implemented in MrBayes v3.2.1 (Ronquist and Huelsenbeck 2003) under a nucleotide model, assuming a GTR substitution model with 3 unlinked codon-position classes, gamma-distributed rate variation between sites with no invariant sites, and inferred base frequencies. We ran each analysis for a minimum of 50 million steps, or as long as necessary for the tree topologies to reach stationarity (standard deviation of split frequencies between duplicate independent runs <0.01; PSRF~1 and ESS>1000 for all parameters). Samples from the posterior were recorded every 10,000 steps, and a maximum clade credibility tree was inferred from two duplicate runs using TreeAnnotator (Drummond et al. 2012).

#### Gene turnover rates



To quantify the rate of gene duplication and loss during Argonaute evolution, we estimated the rate of gene turnover ( $\lambda$ , the number of gains or losses per million years) for each Argonaute subclade using CAFE v3.1 (Han et al. 2013). We also tested whether subclades differed significantly in their rates of gene turnover by using 1000 replicates of CAFE's Monte Carlo resampling procedure. This generates an expected distribution of gene family sizes under a birth-death model, conditioned on the species topology and a set  $\lambda$  value (which we fixed at the value estimated for each subclade), thus providing an estimate of the  $p$ -value for each of the other subclades. To mitigate potential bias introduced by incomplete genome assemblies, turnover analyses only included species that had at least one gene in each subclade (66 of total 86 species). To assess the potential impact of searching transcriptomes, which will only detect expressed genes (and may therefore lead to erroneous inference of gene loss and falsely inflate the rate of gene turnover), we repeated these analyses with rates of gene gain and loss estimated separately. We find similar results when comparing rates of gene gain and gene turnover, suggesting that missing data has a negligible effect on our estimates of gene turnover rate.

To provide the independent species-level tree topology for all 66 species that is required for this analysis, we manually combined the high-confidence multigene phylogenies presented in Wiegmann et al. 2011 and Misof et al. 2014. Where these reference trees lacked the relevant taxa (e.g. relationships below the level of family), we either referred to other published multi-gene phylogenies (van der Linde et al. 2010, Zhang et al. 2010 and Dyer et al. 2008 for Drosophilidae, *Bactrocera* and *Glossina*, respectively), or inferred a Bayesian phylogeny using the arginine kinase gene (Culicidae, MrBayes parameters as above). Conditional on this species topology, we estimated relative branch lengths using BEAST v1.7 (Drummond et al. 2012) and a translational MAFFT alignment of the 1:1:1 orthologue Ago1, constraining the dates of key nodes to previously inferred dates derived from fossil evidence (as used by Wiegmann et al. 2011: root=245mya, Brachycera=200mya, Cyclorrhapha=150mya, Schizophora=70mya). As our primary concern is the difference in

relative rates of gene gain and loss for the different subclades, inaccuracies of the absolute timescale will have minimal impact on our conclusions.

#### Evolutionary rate and positively selected residues

To infer the relative rates of synonymous and non-synonymous substitution ( $dN/dS=\omega$ ) averaged across all sites, we used codeml (PAML, Yang 1997) to fit model M0 (single  $\omega$ ) separately for each subclade (Ago1, Ago2, Ago3 & Piwi/Aub), conditional on the alignment and tree topology. To test for significant differences between these subclade-specific rates, we fixed  $\omega$  for each subclade at the value estimated for each of the other subclades, and used Akaike weights to compare the likelihood of these fixed  $\omega$  values to the likelihood of the  $\omega$  value estimated from the data for that subclade.

To estimate the change in evolutionary rate after duplication, and to test whether duplicates experienced a transient or sustained change in evolutionary rate, we fitted two variants of the M0 model, each with two separate  $\omega$  parameters estimated for different branches of the gene tree (illustrated in Figure 1). To test for a transient change in evolutionary rate directly after duplication, we fitted a model (which we term “Immediate”) that specified one  $\omega$  for branches immediately after a duplication event, and another  $\omega$  for all other branches. To test for a sustained change in evolutionary rate following duplication, we fitted a second model (which we term “All descendants”) that specified one  $\omega$  for all branches arising from a duplication event, and another  $\omega$  for all other branches. For each subclade, Akaike weights were used to estimate the relative support for the M0, “Immediate” and “All descendants” models.

To test for asymmetrical evolutionary rates after a particular duplication event, we fitted two variants of the M0 model. The first model (which we term “Asymmetrical rates”) estimated three separate  $\omega$  parameters for different branches of the gene tree: one  $\omega$  for the branches in one lineage produced by the duplication event; a second  $\omega$  for the branches in the other lineage; and a third  $\omega$  for the rest of the tree. The second model (which we term “Symmetrical rates”) estimated two  $\omega$  ratios: one  $\omega$  for both lineages arising from the duplication event, and a second  $\omega$  for the rest of the tree. The large number of nested

duplication events means that an exhaustive test of all cases in which some duplication events result in asymmetrical rates is intractable; we therefore focused on ten key duplication events in the Ago2 and Piwi/Aub subclades (illustrated in Figure S1). For each duplication event, we used a likelihood ratio test to compare the fit of the Asymmetrical and Symmetrical rates models. Following the method outlined above, we fitted “Immediate” and “All descendants” versions of each of these models.

To test for positively selected residues in each subclade, we used likelihood ratio tests (LRTs) to compare the fit of two models, each with two site classes. In both models,  $\omega$  of the first ‘constrained’ site class was a discretised beta-distribution with 8 classes. The models differ in that in the first model (the null model “M8a” in codeml)  $\omega$  of the second ‘positively selected’ site class is fixed at 1 (neutrality), while in the second model (the “M8” model)  $\omega$  of the second site class is constrained to exceed 1. If the LRT indicated a significantly better fit for M8 than M8a given the parameters in the model, individual residues were classed as positively selected if they had a Bayes Empirical Bayes (BEB) posterior probability of >95% that  $\omega > 1$ .

To assess the potential impact of false positives introduced by misalignments (see Jordan and Goldman 2012), we ran M0 and M8 codeml analyses on two alignments for each subclade, the first with no trimming of ambiguous alignment positions (which may represent genuinely rapidly evolving sites), and the second with ambiguous alignment positions identified by eye and removed. All estimates and statistical comparisons of evolutionary rates outlined above were very similar with and without alignment screening: we therefore report results estimated from the untrimmed alignments. Although we could not rule out gene conversion between paralogues (which can lead to erroneous support for positive selection (Casola and Hahn 2009)), we found very few positively selected sites, so this effect is likely to have little or no effect on our analyses.

#### Domain mapping and structural modelling

To investigate the distribution of rapidly evolving sites across the domain architecture of each Argonaute gene, we inferred the location of each domain in each Argonaute gene by

searching the Pfam database (Finn et al. 2009), and then mapped the mean estimate of  $\omega$  for each residue across the gene (derived from the BEB posterior distribution under the M8 model in PAML (Yang 1997)). To describe evolutionary rate heterogeneity in the protein structures of each gene, we built structural models based on published X-ray crystallography structures: the *D. melanogaster* Ago1 structure was based on human Ago1 (Faehnle et al. 2013), and the structures of *D. melanogaster* Ago2, Ago3 and Piwi were based on human Ago2 (Schirle and Macrae 2012). We used the MODELER software in the Discovery Studio 4.0 Modeling Environment (Accelrys Software Inc., San Diego: Accelrys Software Inc., 2013) to calculate ten models, and selected the most energetically favourable for each protein. The model optimization level was set to High, and loop refinement was included. The model quality was assessed with the 3D-profile option in the software, which compares the compatibility of the 3D structure and the sequence. For *D. melanogaster* Ago2, we replaced the inferred PAZ domain structure with the *D. melanogaster* Ago2 PAZ domain structure that has previously been resolved using X-ray crystallography (Song et al. 2003). We then mapped  $\omega$  onto each residue of the structure using PyMol v.1.7.4.1 (Schrödinger, LLC). For both analyses, we used estimates of  $\omega$  from trimmed alignments to provide a conservative estimate of residue-specific evolutionary rate. Sites that were trimmed out of the alignment were excluded when mapping  $\omega$  across domains, and were set as  $\omega=0$  when mapping  $\omega$  across structures.

## Results

### Duplications of Ago2, Ago3 and Piwi/Aub occur in different Dipteran lineages

To explore the evolutionary dynamics of Argonautes in the Diptera, we quantified the rate of duplication and evolution of Argonautes from 86 Dipteran species. We find numerous expansions of Ago2 and Piwi/Aub (including the origin of canonical Piwi and Aub themselves from their Piwi-subfamily ancestor; illustrated in Figures 2 & 3). This is in sharp contrast to Ago1, which is present as a single copy orthologue in all Diptera (Figures 2 & S2), and Ago3, which has duplicated only rarely (Figures 2 & S3).

We also find that the expansions of Ago2 and Piwi/Aub have occurred in different taxa and at different times (Figures 2 & 3). Most duplications of Ago2 have occurred in the Brachycera, with numerous duplications within the *Glossina* (<84mya), and the *Drosophila obscura* group (<50mya) (Hain et al. 2010). Perhaps surprisingly, even in the melanogaster group there appear to have been at least three duplications: a duplicate shared between *D. rhopaloa* and *D. takahashii* (DRHO009538 & DTAK011769 respectively) implying multiple losses, and a duplication within the lineage leading to *D. kikkawai*. Although incomplete genomes and some uncertainty in the gene tree topology mean that the losses are uncertain, implications for our study are minimal, as our analysis of gene turnover uses only gene counts, and losses do not factor into our comparison of evolutionary rates before and after duplication. Single duplications of Ago2 have occurred in the Brachycerans *Drosophila willistoni*, *Scaptodrosophila deflexa*, *Musca domestica* and *Megaselia abdita*, and in the Nematocerans *Belgica antarctica*, *Culex pipiens* and *Sitodiplosis mosellana* (illustrated in Figure 3).

In contrast, most duplications of Piwi/Aub have occurred in the Nematocera. Numerous duplications have occurred in the mosquitoes (*Aedes spp.*, *Anopheles spp.* and *Culex quinquefasciatus*) <65mya, and multiple copies are seen in *Lutzomyia longipalpis*, *Sitodiplosis mosellana*, *Chironomus riparius*, *Belgica antarctica* and *Corethrella appendiculata* (shown in Figure 3). A duplication at the base of the Brachycera between 182 and 156mya gave rise to the separate Aub and Piwi subclades (as they occur in *D. melanogaster*, labelled in Figures 2 & 3). Within these subclades duplications have occurred rarely, only being observed in Piwi of the drosophilids *Phortica variegata* and *Scaptodrosophila deflexa*, and in Aub of *Teleopsis* species.

#### Ago2 and Piwi/Aub have significantly higher duplication rates than Ago1 and Ago3

To quantify these contrasting patterns of duplication, we used CAFE (Han et al. 2013) to estimate the rate of gene turnover ( $\lambda$ , the number of gains or losses per million years) in each Argonaute subclade. We find that gene turnover rate varies considerably among the subclades, with Ago2 ( $\lambda=0.0022$ ), Ago3 ( $\lambda=0.0003$ ) and Piwi/Aub ( $\lambda=0.0012$ ) having

significantly higher gene turnover rates than Ago1 ( $\lambda=1.1516 \times 10^{-10}$ ) ( $p < 0.001$  based on the expected distribution of gene family sizes under a birth-death model, with  $\lambda$  fixed at the value estimated for Ago1). We also find that Ago2 and Piwi/Aub have significantly ( $p < 0.001$ ) higher gene turnover rates than Ago3, but do not differ significantly from each other ( $p = 0.198$ ).

#### Argonautes show contrasting rates of protein evolution before and after duplication

To quantify the rate of protein evolution in each Argonaute subclade, and to identify any sites evolving under positive selection, we fitted models using codeml (PAML; Yang 1997). These analyses revealed that Ago2 has the highest non-synonymous to synonymous substitution ratio ( $\omega = 0.14 \pm 0.0015$ ), followed by Ago3 ( $\omega = 0.12 \pm 0.0015$ ), Piwi/Aub ( $\omega = 0.09 \pm 0.0009$ ), and lastly Ago1 ( $\omega = 0.01 \pm 0.0002$ ). All rates were significantly different from each other (Akaike weight = 1.000 to 3 d.p. in all cases). Scans for positively-selected sites identified five candidate sites in Ago3 and one in Piwi/Aub; however, in both cases the M8 model was not significantly more likely than the null M8a model (for  $\omega$  estimates and likelihoods under all models see Supplementary Tables 1-3).

To test whether the relative rate of protein evolution changes following duplication, we calculated the likelihood of the data for Ago2, Ago3 and Piwi/Aub under two models: the first with a separate evolutionary rate for branches immediately after a duplication event (the “Immediate” model); and the second with a separate rate for all branches subtending a duplication event (the “All descendants” model) (see Figure 1). For Ago2 and Ago3, the “All descendants” model had all support (Akaike weight = 1.000 to 3 d.p. for each). For Piwi/Aub, however, the “Immediate” model had all support (Akaike weight = 1.000 to 3 d.p.). In each case the evolutionary rate increased after duplication, with Ago2 having the highest rate and Piwi the lowest (Figure 4).

To test for asymmetry between the evolutionary rates of paralogues after duplication, we calculated the likelihood of the data for 10 key duplication events in the Ago2 and Piwi/Aub subclades under two models: the “Asymmetrical rates” model specified one evolutionary rate for one lineage produced by the duplication event, a second rate for the other lineage, and a third for the rest of the tree; whereas the “Symmetrical rates” model



specified one evolutionary rate for both lineages produced by the duplication event, and a second rate for the rest of the tree. For Ago2, we find that the Asymmetric rates model does not provide a significantly better fit for the *Glossina sp.*, *Belgica antarctica* or *Culex pipiens* duplication events (LRT,  $p > 0.1$  in all cases). However, the Asymmetrical rates model fits significantly better for the two branches immediately after the *obscura* group Ago2e-Ago2a/f event (LRT,  $p < 0.005$ ), and for all branches subtending this event (LRT,  $p < 0.005$ ). Under the Asymmetrical rates (All descendants) model, the Ago2e clade ( $\omega = 0.17 \pm 0.011$ ) and Ago2a/Ago2f clade ( $\omega = 0.22 \pm 0.009$ ) are evolving considerably faster than the rest of the tree ( $\omega = 0.13 \pm 0.001$ ). For the *obscura* group Ago2a-Ago2f event, the Asymmetrical rates (All descendants) model does not provide a significantly better fit (LRT,  $p > 0.1$ ), but the Immediate version of this model does give a significantly better fit (LRT,  $p < 0.005$ ). Under this model, the branch at the base of the Ago2a clade has a much lower evolutionary rate ( $\omega = 0.06 \pm 0.029$ ) than the rest of the tree ( $\omega = 0.14 \pm 0.001$ ), and the branch at the base of the Ago2f clade has a much higher evolutionary rate ( $\omega = 0.42 \pm 0.142$ ).

For Piwi/Aub, we find that the Asymmetric rates (All descendants) model provides a significantly better fit for the Piwi-Aub divergence (LRT,  $p < 0.01$ ) and the duplication event early in mosquito evolution (LRT,  $p < 0.005$ ). The Aub lineage ( $\omega = 0.09 \pm 0.002$ ) has a higher evolutionary rate than the Piwi lineage ( $\omega = 0.08 \pm 0.002$ ); however, the similarity between the rate of Aub and the rest of the tree ( $\omega = 0.09 \pm 0.002$ ) suggests that this difference may be caused by constraint on Piwi rather than positive selection on Aub. After the mosquito Piwi/Aub duplication event, we see a large difference in evolutionary rates, with one clade evolving much more rapidly ( $\omega = 0.14 \pm 0.004$ ) than the other ( $\omega = 0.04 \pm 0.003$ ), and the rest of the tree ( $\omega = 0.08 \pm 0.001$ ).

#### Ago2 displays hotspots of evolution at the RNA binding pocket entrance

To investigate the distribution of rapidly evolving residues, we mapped  $\omega$  estimates onto the domains and structures of each Argonaute. We found that rapidly evolving residues are spread across all domains of Ago2, Ago3 and Piwi/Aub (Figure S4). We also found that Ago2 appears to have clusters of more rapidly evolving residues at the entrance to the RNA



binding pocket (Figure 5), which are not found in Ago3 (Figure S5) or Piwi/Aub (Figure 5). In contrast, the residues that directly contact the sRNA guide are conserved in all Argonautes (Figures 5 & S5).

### Discussion

Our results reveal contrasting patterns of selection and duplication during Dipteran Argonaute evolution. The low evolutionary rate and lack of gene turnover in Ago1 are in agreement with previous studies in *D. melanogaster* (Obbard et al. 2006; Obbard et al. 2009), and are consistent with the idea that Ago1 is carrying out a conserved gene regulatory role in the Diptera as a whole. In contrast, the better fit of the “All descendants” model to duplications in Ago2 and Ago3 (Figure 4) indicates that paralogues in these subclades have experienced a sustained increase in evolutionary rate, possibly driven by the acquisition of new functions.

This result is particularly noteworthy in Ago2, which is already among the top 3% of the fastest evolving proteins in *D. melanogaster* (Obbard et al. 2006). Our structural modelling suggests that one possible hotspot of adaptive evolution for these paralogues may be the entrance of the RNA binding pocket (Figure 5). Relaxation of selection pressures on these residues is unexpected as they form alpha-helices, rigid secondary structures that are needed for the stability of the tertiary structure of the protein (Panchenko et al. 2005; Liu et al. 2008); instead, their rapid evolution may be caused by undetected positive selection. The pocket is formed by the PAZ and MID domains, which bind the sRNA guide and form the channel in which the target RNA sits during cleavage (Schirle and Macrae 2012). While the molecular interactions between the sRNA guide and the inside of the binding pocket have been characterised (reviewed in Swarts et al. 2014), less is known about the function of the residues at the entrance to the pocket. However, given the location of these rapidly-evolving residues at the mouth of the binding pocket away from the sRNA guide (Figure 5), such positive selection could be driving differences in target RNA binding and cleavage. Alternatively, selection could be imposed by viral suppressors of RNAi, which are encoded

by numerous viruses to inhibit the antiviral RNAi response, and several of which prevent target cleavage by Ago2 (Wang et al. 2006; van Mierlo et al. 2012; van Mierlo et al. 2014). While we do not find evidence of positive selection in our sites analysis across the Diptera as a whole, signatures of selection are evident when we apply branch-sites analyses to Ago2 in the Drosophilidae, as has been reported previously (Kolaczkowski et al. 2011). Such selection may be acting on Ago2 in the Diptera as a whole, but its signature may be masked by saturation of synonymous sites (Anisimova et al. 2001; Clark et al. 2007).

Functional differences between most Dipteran Argonaute paralogues have not been characterized experimentally. However, transcriptome data are available for some *G. morsitans* tissues, including ‘lactating’ and non-lactating females (Benoit et al. 2014) and salivary glands from parasitized and unparasitized individuals (Telleria et al. 2014) (SRA accessions SRX287393, SRX287395, SRX342351 and SRX342350 respectively). Using these data we explored the possibility of functional divergence in *G. morsitans* Ago2 and Ago3 paralogues, and found differential expression between both sets of paralogues, as well as high expression of Ago3b in the salivary glands, which increased upon infection with *Trypanosoma brucei* (Figure S6). Although this observation awaits replication, the canonical germline-specific role of Ago3 in *D. melanogaster* (Li et al. 2009) makes any expression of *G. morsitans* Ago3b in the salivary glands unexpected, and suggests that this paralogue has undergone rapid functional divergence to a role beyond TE suppression. Strikingly, this reflects the general patterns noted for somatically-expressed Piwis across the eukaryotes, which have evolved diverse roles in epigenetic regulation, genome rearrangement and somatic development (reviewed in Ross et al. 2014).

The better fit of the “Immediate” model to duplications of Piwi/Aub (Figure 4) suggests that the evolutionary rate of paralogues in these subclades has been constrained soon after duplication, which may indicate a burst of adaptation to specialize to existing (but distinct) roles. For many duplicates, the branches immediately after duplication are also terminal branches, which clouds the difference between the “Immediate” and “All descendants” models. In contrast, the divergence of separate Aub and Piwi (*sensu stricto*) lineages

resulted from a much older duplication in the Piwi-subfamily Piwi/Aub subclade. Our asymmetry analysis suggests that this divergence was accompanied by a reduction in evolutionary rate, particularly in the Piwi lineage, indicating that these lineages are evolving under tight constraint. We estimate that this divergence, which happened at the base of the Brachycera, occurred between 182 and 156mya (Figure 2). However, the ambiguous identities of the two Piwi/Aub paralogues in *Hermetia illucens* and *Tabanus bromius* (Figure 3) mean that this duplication could have occurred slightly earlier (~200mya).

Under either scenario, Piwi/Aub paralogues in the vast majority of Nematoceran taxa (including all mosquitoes) are equally homologous to Aub and Piwi, which in *D. melanogaster* have specialized to distinct roles in the Ping-Pong piRNA amplification cycle and TE silencing respectively, suggesting that the ancestral Piwi/Aub gene may have had multiple conflicting functions (reviewed in Luteijn and Ketting 2013). It may be that the increased duplication rate of Piwi/Aub in the Nematocera is a result of multiple independent resolutions of this conflict, causing piRNA biogenesis and TE silencing to rely on different suites of Argonaute genes in the Nematocera. This is supported by our asymmetry analysis, which finds that the Piwi/Aub expansion in mosquitoes resulted in asymmetrical evolutionary rates in the resulting lineages, with the rapid evolution of one lineage consistent with the evolution of a novel function. Notably, this rapidly evolving lineage includes a Piwi/Aub paralogue in *A. aegypti* (Piwi5) that has recently been shown to have a highly derived function in the production of virus-derived piRNAs (Miesen et al. 2015). *A. aegypti* is a major vector of several arboviruses including yellow fever virus and chikungunya virus, and also has an exceptionally high TE load (Arensburger et al. 2011); although little is known about the total viral load of *A. aegypti*, it is possible that the combined viral and TE loads impose contrasting selection pressures, thereby driving the expansion of Piwi/Aub. Moreover, the numerous instances of expansion followed by functional divergence demonstrate that the Piwi/Aub subclade is not constrained to a germline-specific anti-TE role, but can evolve novel and highly derived functions.

## Conclusion

We show that Dipteran Argonautes differ widely in their rates of gene turnover and protein evolution, with duplication driving an increase in evolutionary rate that suggests frequent functional divergence. Our results provide an insight into the selection pressures driving the evolution of RNAi mechanisms across the eukaryotes, which are integral to a range of cellular and genomic processes. Our finding that Argonautes undergo frequent expansions and contractions indicates that expansions in other taxa, such as the WAGO clade of nematodes (Buck and Blaxter 2013) and the Piwi clade of the aphid *A. pisum* (Lu et al. 2011), are not isolated cases; instead, these are further examples of a general pattern of rapid gene turnover in some Argonaute clades. Additionally, our finding that duplication drives rapid evolution suggests that Argonautes evolve new functions frequently and rapidly, as exemplified by Ago4 in the shrimp *Penaeus monodon* (Leebonoi et al. 2015) and the smedwi clade of planarians (Reddien et al. 2005). This combination of rapid gene turnover and frequent functional divergence illustrates a high degree of evolutionary lability in Argonaute function across a wide range of taxa, and may drive the functional overlap frequently observed between different Argonaute subclades across the eukaryotes. Our work also highlights the selection pressures exerted by parasites, shown by the higher rate of gene turnover for Piwi/Aub and higher evolutionary rate of Ago2, which play roles in defence against TEs and viruses respectively. This provides further evidence of the importance of parasites in evolution (Dawkins and Krebs 1979), and demonstrates how host-parasite interactions can drive genome evolution and generate phenotypic novelty.

## Acknowledgements

This work was supported by a Natural Environment Research Council Doctoral Training Grant (NERC DG NE/J500021/1 to SHL), the Academy of Finland (265971 to HS), a University of Edinburgh Chancellor's Fellowship (to DJO) and a Wellcome Trust strategic award to the Centre for Immunity, Infection and Evolution (WT095831 to the CIIE). We thank all authors who gave us access to unreleased genomes and transcriptomes. We also thank

Esther Schnettler for reading an earlier version of the manuscript, and two anonymous reviewers for valuable feedback.

## References

- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389–3402.
- Anisimova M, Bielawski JP, Yang Z. 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol Biol Evol.* 18:1585–1592.
- Arensburger P, Hice RH, Wright JA, Craig NL, Atkinson PW. 2011. The mosquito *Aedes aegypti* has a large genome size and high transposable element load but contains a low proportion of transposon-specific piRNAs. *BMC Genomics* 12:606.
- Assis R, Bachtrog D. 2013. Neofunctionalization of young duplicate genes in *Drosophila*. *Proc. Natl. Acad. Sci.* 110:17409–17414.
- Benoit JB, Attardo GM, Michalkova V, Krause TB, Bohova J, Zhang Q, Baumann AA, Mireji PO, Takáč P, Denlinger DL, et al. 2014. A Novel Highly Divergent Protein Family Identified from a Viviparous Insect by RNA-seq Analysis: A Potential Target for Tsetse Fly-Specific Abortifacients. *PLoS Genet.* 10:6–10.
- Bohmert K, Camus I, Bellini C, Bouchez D, Caboche M, Benning C. 1998. AGO1 defines a novel locus of *Arabidopsis* controlling leaf development. *EMBO J.* 1:170–180.
- Buck AH, Blaxter M. 2013. Functional diversification of Argonautes in nematodes: an expanding universe. *Biochem. Soc. Trans.* 41:881–886.
- Campbell CL, Black WC, Hess AM, Foy BD. 2008. Comparative genomics of small RNA regulatory pathway components in vector mosquitoes. *BMC Genomics* 9:425.
- Casola C, Hahn MW. 2009. Gene conversion among paralogs results in moderate false detection of positive selection using likelihood methods. *J. Mol. Evol.* 68:679–687.
- Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17:540–552.
- Cerutti H, Casas-Mollano JA. 2006. On the origin and functions of RNA-mediated silencing: from protists to man. *Curr. Genet.* 50:81–99.
- Chung W-J, Okamura K, Martin R, Lai EC. 2008. Endogenous RNA interference provides a somatic defense against *Drosophila* transposons. *Curr. Biol.* 18:795–802.
- Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, Kellis M, Gelbart W, Iyer VN, et al. 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450:203–218.
- Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, Dus M, Perrimon N, Kellis M, Wohlschlegel JA, Sachidanandam R, et al. 2008. An endogenous small interfering RNA pathway in *Drosophila*. *Nature* 453:798–802.
- Dawkins R, Krebs JR. 1979. Arms races between and within species. *Proc. R. Soc. B* 205:489–511.



- Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29:1969–1973.
- Dyer NA, Lawton SP, Ravel S, Choi KS, Lehane MJ, Robinson AS, Okedi LM, Hall MJR, Solano P, Donnelly MJ. 2008. Molecular phylogenetics of tsetse flies (Diptera: Glossinidae) based on mitochondrial (COI, 16S, ND2) and nuclear ribosomal DNA sequences, with an emphasis on the palpalis group. *Mol. Phylogenet. Evol.* 49:227–239.
- Eulalio A, Huntzinger E, Izaurralde E. 2008. Getting to the Root of miRNA-Mediated Gene Silencing. *Cell* 132:9–14.
- Faehnle CR, Elkayam E, Haase AD, Hannon GJ, Joshua-Tor L. 2013. The Making of a Slicer: Activation of Human Argonaute-1. *Cell Rep.* 3:1901–1909.
- Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, et al. 2009. The Pfam protein families database. *Nucleic Acids Res.* 38:D211–D222.
- Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391:806–811.
- Hahn MW. 2009. Distinguishing among evolutionary models for the maintenance of gene duplicates. *J. Hered.* 100:605–617.
- Hain D, Bettencourt BR, Okamura K, Csorba T, Meyer W, Jin Z, Biggerstaff J, Siomi H, Hutvagner G, Lai EC, et al. 2010. Natural variation of the amino-terminal glutamine-rich domain in *Drosophila* argonaute2 is not associated with developmental defects. *PLoS One* 5:e15264.
- Han MV, Thomas GWC, Lugo-Martinez J, Hahn MW. 2013. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol. Biol. Evol.* 30:1987–1997.
- Hittinger CT, Carroll SB. 2007. Gene duplication and the adaptive evolution of a classic genetic switch. *Nature* 449:677–681.
- Iwasaki YW, Siomi MC, Siomi H. 2015. PIWI-Interacting RNA: Its Biogenesis and Functions. *Annu. Rev. Biochem.* 84:405–433.
- Jordan G, Goldman N. 2012. The effects of alignment error and alignment filtering on the sitewise detection of positive selection. *Mol. Biol. Evol.* 29:1125–1139.
- Kalmykova AI, Klenov MS, Gvozdev VA. 2005. Argonaute protein PIWI controls mobilization of retrotransposons in the *Drosophila* male germline. *Nucleic Acids Res.* 33:2052–2059.
- Kataoka Y, Takeichi M, Uemura T. 2001. Developmental roles and molecular characterization of a *Drosophila* homologue of Arabidopsis Argonaute1, the founder of a novel gene superfamily. *Genes to Cells* 6:313–325.



- Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT : a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30:3059–3066.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647–1649.
- Keays MC, Barker D, Wicker-Thomas C, Ritchie MG. 2011. Signatures of selection and sex-specific expression variation of a novel duplicate during the evolution of the *Drosophila* desaturase gene family. *Mol. Ecol.* 20:3617–3630.
- Kim VN, Han J, Siomi MC. 2009. Biogenesis of small RNAs in animals. *Nat. Rev. Mol. Cell Biol.* 10:126–139.
- Kolaczkowski B, Hupalo DN, Kern AD. 2011. Recurrent adaptation in RNA interference genes across the *Drosophila* phylogeny. *Mol. Biol. Evol.* 28:1033–1042.
- Leebonoi W, Sukthaworn S, Panyim S, Udomkit A. 2015. A novel gonad-specific Argonaute 4 serves as a defense against transposons in the black tiger shrimp *Penaeus monodon*. *Fish Shellfish Immunol.* 42:280–288.
- Li C, Vagin V V, Lee S, Xu J, Ma S, Xi H, Seitz H, Horwich MD, Syrzycka M, Honda BM, et al. 2009. Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell* 137:509–521.
- Li H, Li WX, Ding SW. 2002. Induction and suppression of RNA silencing by an animal virus. *Science.* 296:1319–1321.
- van der Linde K, Houle D, Spicer GS, Stepan SJ. 2010. A supermatrix-based molecular phylogeny of the family Drosophilidae. *Genet. Res. (Camb).* 92:25–38.
- Liu J, Zhang Y, Lei X, Zhang Z. 2008. Natural selection of protein structural and functional properties: a single nucleotide polymorphism perspective. *Genome Biol.* 9:R69.
- Lu H-L, Tanguy S, Rispe C, Gauthier J-P, Walsh T, Gordon K, Edwards O, Tagu D, Chang C, Jaubert-Possamai S. 2011. Expansion of genes encoding piRNA-associated argonaute proteins in the pea aphid: diversification of expression profiles in different plastic morphs. *PLoS One* 6:e28051.
- Luteijn MJ, Ketting RF. 2013. PIWI-interacting RNAs: from generation to transgenerational epigenetics. *Nat. Rev. Genet.* 14:523–534.
- van Mierlo JT, Bronkhorst AW, Overheul GJ, Sadanandan SA, Ekström J-O, Heestermans M, Hultmark D, Antoniewski C, van Rij RP. 2012. Convergent evolution of argonaute-2 slicer antagonism in two distinct insect RNA viruses. *PLoS Pathog.* 8:e1002872.
- van Mierlo JT, Overheul GJ, Obadia B, van Cleef KWR, Webster CL, Saleh M-C, Obbard DJ, van Rij RP. 2014. Novel *Drosophila* Viruses Encode Host-Specific Suppressors of RNAi.

- PLoS Pathog. 10:e1004256.
- Miesen P, Girardi E, van Rij RP. 2015. Distinct sets of PIWI proteins produce arbovirus and transposon-derived piRNAs in *Aedes aegypti* mosquito cells. *Nucleic Acids Res.*
- Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, Frandsen PB, Ware J, Flouri T, Beutel RG, et al. 2014. Phylogenomics resolves the timing and pattern of insect evolution. *Science* 346:763–767.
- Morandin C, Havukainen H, Kulmuni J, Dhaygude K, Trontti K, Helanterä H. 2014. Not only for egg yolk - functional and evolutionary insights from expression, selection, and structural analyses of *Formica* ant vitellogenins. *Mol. Biol. Evol.* 31:2181–2193.
- Morazzani EM, Wiley MR, Murreddu MG, Adelman ZN, Myles KM. 2012. Production of virus-derived ping-pong-dependent piRNA-like small RNAs in the mosquito soma. *PLoS Pathog.* 8:e1002470.
- Mukherjee K, Campos H, Kolaczowski B. 2013. Evolution of animal and plant dicers: early parallel duplications and recurrent adaptation of antiviral RNA binding in plants. *Mol. Biol. Evol.* 30:627–641.
- Nielsen MG, Gadagkar SR, Gutzwiller L. 2010. Tubulin evolution in insects: gene duplication and subfunctionalization provide specialized isoforms in a functionally constrained gene family. *BMC Evol. Biol.* 10:113.
- Obbard DJ, Gordon KHJ, Buck AH, Jiggins FM. 2009. The evolution of RNAi as a defence against viruses and transposable elements. *Philos. Trans. R. Soc. London Biol. Sci.* 364:99–115.
- Obbard DJ, Jiggins FM, Halligan DL, Little TJ. 2006. Natural selection drives extremely rapid evolution in antiviral RNAi genes. *Curr. Biol.* 16:580–585.
- Pal-Bhadra M, Leibovitch BA, Gandhi SG, Rao M, Bhadra U, Birchler JA, Elgin SCR. 2004. Heterochromatic silencing and HP1 localization in *Drosophila* are dependent on the RNAi machinery. *Science* 303:669–672.
- Palakodeti D, Smielewska M, Lu Y-C, Yeo GW, Graveley BR. 2008. The PIWI proteins SMEDWI-2 and SMEDWI-3 are required for stem cell function and piRNA expression in planarians. *RNA* 14:1174–1186.
- Panchenko AR, Wolf YI, Panchenko LA, Madej T. 2005. Evolutionary plasticity of protein families: Coupling between sequence and structure variation. *Proteins Struct. Funct. Genet.* 61:535–544.
- Parent J-S, Martínez de Alba AE, Vaucheret H. 2012. The origin and effect of small RNA signaling in plants. *Front. Plant Sci.* 3:1–9.
- Philippe H, Brinkmann H, Lavrov D V., Littlewood DTJ, Manuel M, Wörheide G, Baurain D. 2011. Resolving difficult phylogenetic questions: Why more sequences are not enough. *PLoS Biol.* 9.

- Reddien PW, Oviedo NJ, Jennings JR, Jenkin JC, Sánchez Alvarado A. 2005. SMEDWI-2 is a PIWI-like protein that regulates planarian stem cells. *Science* 310:1327–1330.
- van Rij RP, Saleh M-C, Berry B, Foo C, Houk A, Antoniewski C, Andino R. 2006. The RNA silencing endonuclease Argonaute 2 mediates specific antiviral immunity in *Drosophila melanogaster*. *Genes Dev.* 20:2985–2995.
- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.
- Ross RJ, Weiner MM, Lin H. 2014. PIWI proteins and PIWI-interacting RNAs in the soma. *Nature* 505:353–359.
- Sarkies P, Miska EA. 2014. Small RNAs break out: the molecular cell biology of mobile small RNAs. *Nat. Rev. Mol. Cell Biol.* 15:525–535.
- Sarkies P, Selkirk ME, Jones JT, Blok V, Boothby T, Goldstein B, Hanelt B, Ardila-Garcia A, Fast NM, Schiffer PM, et al. 2015. Ancient and Novel Small RNA Pathways Compensate for the Loss of piRNAs in Multiple Independent Nematode Lineages. *PLoS Biol.* 13:e1002061.
- Schirle NT, Macrae IJ. 2012. The Crystal Structure of Human Argonaute2. *Science* 336:1037–1040.
- Schnettler E, Donald CL, Human S, Watson M, Siu RWC, McFarlane M, Fazakerley JK, Kohl A, Fragkoudis R. 2013. Knockdown of piRNA pathway proteins results in enhanced Semliki Forest virus production in mosquito cells. *J. Gen. Virol.* 94:1680–1689.
- Schnettler E, Tykalová H, Watson M, Sharma M, Sterken MG, Obbard DJ, Lewis SH, McFarlane M, Bell-Sakyi L, Barry G, et al. 2014. Induction and suppression of tick cell antiviral RNAi responses by tick-borne flaviviruses. *Nucleic Acids Res.* 42:1–11.
- Scott JG, Warren WC, Beukeboom LW, Bopp D, Clark AG, Giers SD, Hediger M, Jones AK, Kasai S, Leichter CA, et al. 2014. Genome of the house fly, *Musca domestica* L., a global vector of diseases with adaptations to a septic environment. *Genome Biol.* 15:466–482.
- Sienski G, Dönertas D, Brennecke J. 2012. Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell* 151:964–980.
- Song J-J, Liu J, Tolia NH, Schneiderman J, Smith SK, Martienssen RA, Hannon GJ, Joshua-Tor L. 2003. The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nat. Struct. Biol.* 10:1026–1032.
- Swarts DC, Makarova K, Wang Y, Nakanishi K, Ketting RF, Koonin E V, Patel DJ, van der Oost J. 2014. The evolutionary journey of Argonaute proteins. *Nat. Struct. Mol. Biol.* 21:743–753.
- Telleria EL, Benoit JB, Zhao X, Savage AF, Regmi S, e Silva TLA, O'Neill M, Aksoy S. 2014.

- Insights into the Trypanosome-Host Interactions Revealed through Transcriptomic Analysis of Parasitized Tsetse Fly Salivary Glands. *PLoS Negl. Trop. Dis.* 8.
- Vagin V V, Klenov MS, Stolyarenko AD, Kotelnikov RN, Gvozdev VA. 2004. The RNA interference proteins and Vasa locus are involved in the silencing of retrotransposons in the female germline of *Drosophila melanogaster*. *RNA Biol.* 1:54–58.
- Vodovar N, Bronkhorst AW, van Cleef KWR, Miesen P, Blanc H, van Rij RP, Saleh M-C. 2012. Arbovirus-derived piRNAs exhibit a ping-pong signature in mosquito cells. *PLoS One* 7:e30861.
- Wang X-H, Aliyari R, Li W-X, Li H-W, Kim K, Carthew R, Atkinson P, Ding S-W. 2006. RNA interference directs innate immunity against viruses in adult *Drosophila*. *Science* 312:452–454.
- Wiegmann BM, Trautwein MD, Winkler IS, Barr NB, Kim J-W, Blagoderov V, Caravas J, Narayanan S, Schmidt-Ott U, Kampmeier GE, et al. 2011. Episodic radiations in the fly tree of life. *Proc. Natl. Acad. Sci.* 108:5690–5695.
- Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* 13:555–556.
- Zhang B, Liu YH, Wu WX, Wang Z Le. 2010. Molecular Phylogeny of *Bactrocera* Species (Diptera: Tephritidae: Dacini) Inferred from Mitochondrial Sequences of 16S rDNA and COI Sequences. *BioOne* 93:369–377.

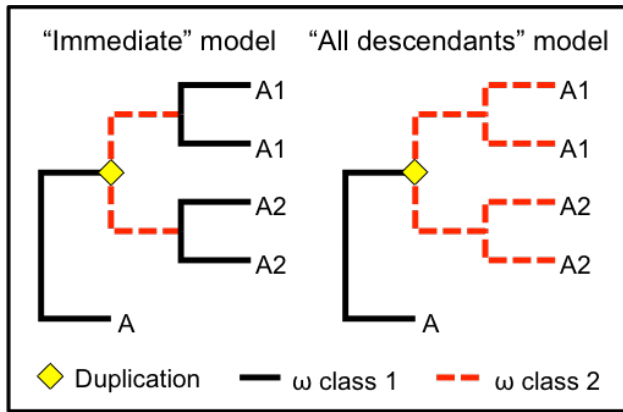


Figure 1: The two models fitted to branches after duplication events. “Immediate” models the expectation if selection pressures change only briefly after duplication, whereas “All descendants” models the expectation if paralogues evolve at a consistently different rate.

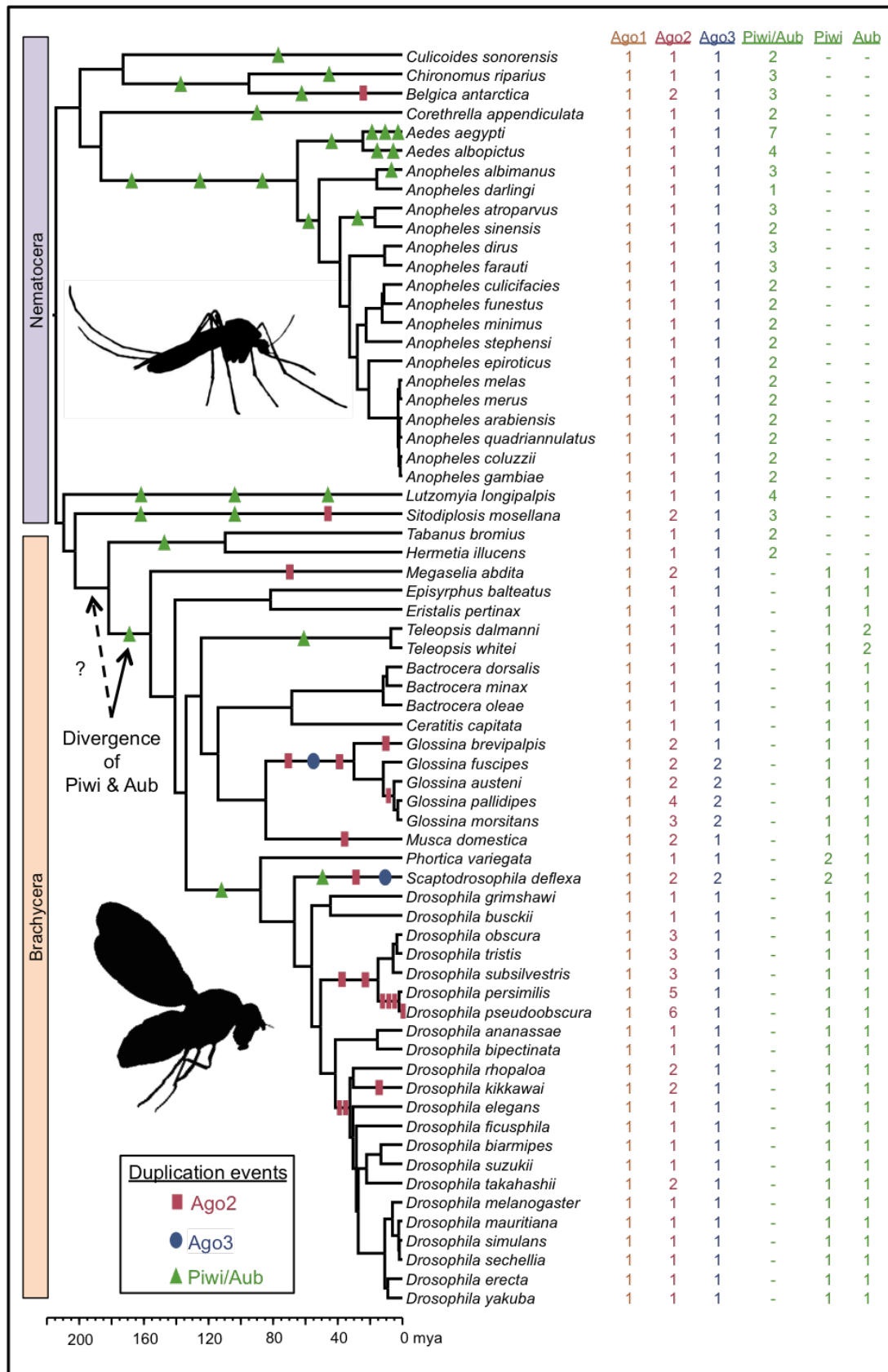


Figure 2: Counts of each Argonaute subclade. Shown are counts for a subsample of 66 Dipteran species with at least one gene in each subclade (out of a total of 86 species). Gene duplication events were inferred by parsimony, and are illustrative only (gene loss is not depicted due to space constraints, thus for some taxa gene counts do not correspond to the number of gene duplications). The rate of gene turnover differs between different Argonautes and lineages, and the divergence of Piwi & Aub occurred 182-156mya. Silhouettes by Warren Photographic and Ramiro Morales-Hojas.



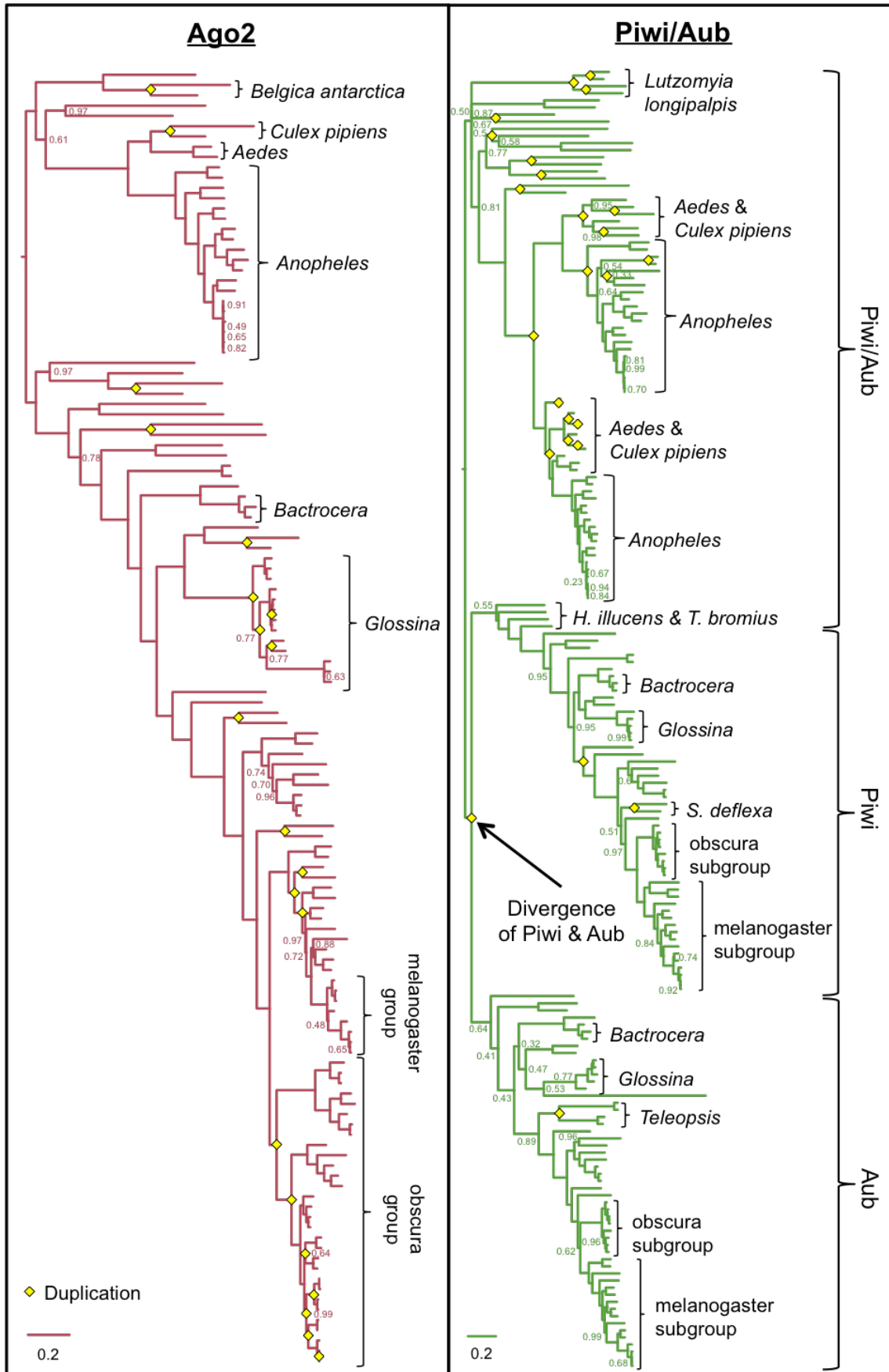


Figure 3: Bayesian gene trees of Ago2 and Piwi/Aub. Ago2 has expanded rapidly in *Glossina* and the obscura group of *Drosophila*, whereas Piwi/Aub has undergone numerous duplications in *Aedes*, *Anopheles* and many other Nematoceran taxa.



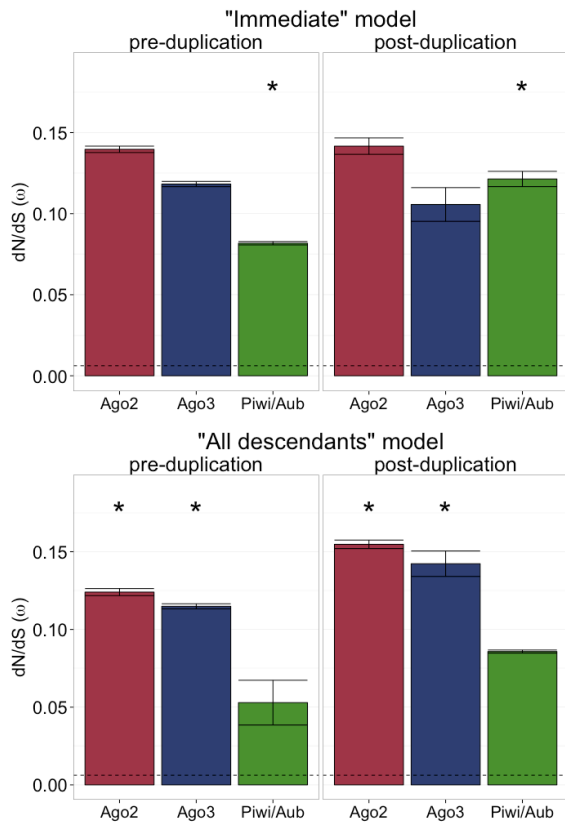


Figure 4: Evolutionary rate estimates before and after duplication, under the "Immediate" and "All descendants" models. Asterisks indicate the most highly supported model, and the dashed line indicates the  $\omega$  value for Ago1 under the M0 model. Duplicates of Piwi/Aub evolve more quickly immediately after duplication, whereas Ago2 and Ago3 paralogues experience a sustained increase in evolutionary rate.

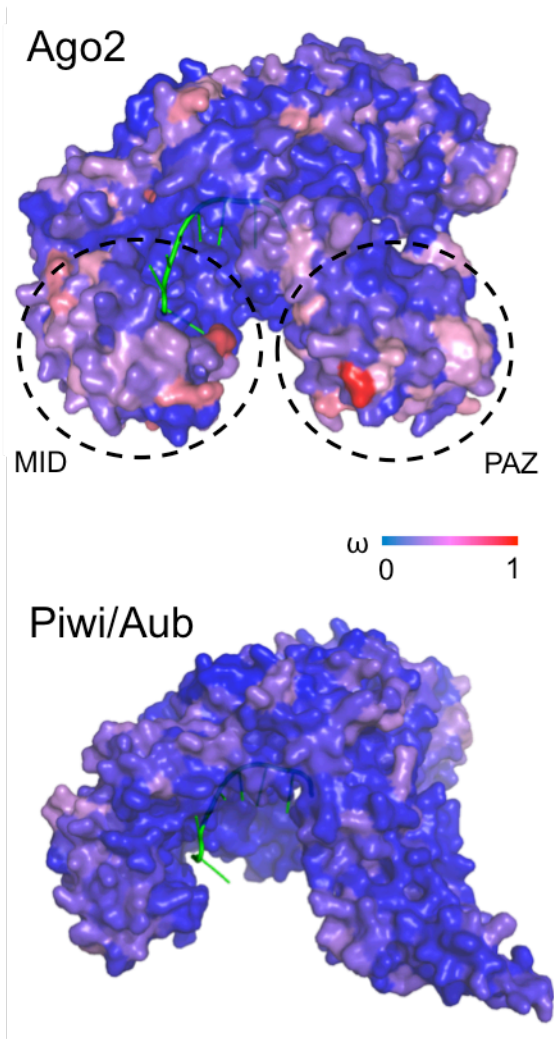


Figure 5: Evolutionary rates mapped onto 3D structures of Ago2 and Piwi/Aub, each binding a sRNA guide.

In Ago2, hotspots of evolution are seen at the entrance of the RNA binding pocket; in contrast, evolutionary rate ( $\omega$ ) across the structure of Piwi/Aub is uniformly low. The MID and PAZ protein domains are indicated for Ago2.