Repeated duplication of Argonaute2 is associated with strong selection and testis specialization in Drosophila

Repeated Duplication of Argonaute2 Is Associated with Strong Selection and Testis Specialization in *Drosophila*

Samuel H. Lewis,*,[1] Claire L. Webster,*,[2] Heli Salmela,[†] and Darren J. Obbard*,[‡]
*Institute of Evolutionary Biology and [‡]Centre for Immunity, Infection and Evolution, University of Edinburgh, Ashworth Laboratories, EH9 3FL, United Kingdom and [†]Department of Biosciences, Centre of Excellence in Biological Interactions, University of Helsinki, 00014 Helsinki, Finland

ABSTRACT Argonaute2 (Ago2) is a rapidly evolving nuclease in the *Drosophila melanogaster* RNA interference (RNAi) pathway that targets viruses and transposable elements in somatic tissues. Here we reconstruct the history of Ago2 duplications across the *D. obscura* group and use patterns of gene expression to infer new functional specialization. We show that some duplications are old, shared by the entire species group, and that losses may be common, including previously undetected losses in the lineage leading to *D. pseudoobscura*. We find that while the original (syntenic) gene copy has generally retained the ancestral ubiquitous expression pattern, most of the novel Ago2 paralogs have independently specialized to testis-specific expression. Using population genetic analyses, we show that most testis-specific paralogs have significantly lower genetic diversity than the genome-wide average. This suggests recent positive selection in three different species, and model-based analyses provide strong evidence of recent hard selective sweeps in or near four of the six *D. pseudoobscura* Ago2 paralogs. We speculate that the repeated evolution of testis specificity in *obscura* group Ago2 genes, combined with their dynamic turnover and strong signatures of adaptive evolution, may be associated with highly derived roles in the suppression of transposable elements or meiotic drive. Our study highlights the lability of RNAi pathways, even within well-studied groups such as *Drosophila*, and suggests that strong selection may act quickly after duplication in RNAi pathways, potentially giving rise to new and unknown RNAi functions in nonmodel species.

KEYWORDS Argonaute; RNAi; *Drosophila*; duplication; testis

ARGONAUTE genes are found in almost all eukaryotes, where they play a key role in antiviral immune defense, gene regulation, and genome stability. They perform this diverse range of functions through their role in RNA interference (RNAi) mechanisms, an ancient system of nucleic acid manipulation in which small RNA (sRNA) molecules guide Argonaute proteins to nucleic acid targets through base complementarity (reviewed in Meister 2013). Gene duplication has occurred throughout the evolution of the Argonaute gene family, with ancient duplication events characteristic of some lineages—such as three duplications early in plant evolution (Singh *et al.* 2015) and multiple expansions and losses throughout the evolution of nematodes (reviewed in Buck and Blaxter 2013) and the Diptera (Lewis *et al.* 2016). After duplication, Argonautes have often undergone functional divergence, involving changes in expression patterns and altered sRNA binding partners (Lu *et al.* 2011; Leebonoi *et al.* 2015; Miesen *et al.* 2015). Duplication early in eukaryotic evolution produced two distinct Argonaute subfamilies, Ago and Piwi, which have since been retained in the vast majority of Metazoa (Cerutti and Casas-Mollano 2006). Members of the Ago subfamily are expressed in both somatic and germline tissue, and variously bind sRNAs derived from host transcripts (miRNAs, endo-siRNAs) or transposable elements (TE endo-siRNAs) and viruses (viRNAs). In contrast, in most vertebrates and arthropods, the Piwi subfamily members are expressed primarily in association with the germline (reviewed in Ross *et al.* 2014), and bind sRNAs from TEs and host loci (piRNAs), suggesting that the Piwi subfamily

doi: 10.1534/genetics.116.192336
Manuscript received June 6, 2016; accepted for publication August 12, 2016; published Early Online August 17, 2016.
Available freely online through the author-supported open access option.
Supplemental material is available online at www.genetics.org/lookup/suppl/doi:10.1534/genetics.116.192336/-/DC1.
[1]Corresponding author: Department of Genetics, University of Cambridge, Downing St., Cambridge, CB2 3EH, United Kingdom. E-mail address: sam.lewis@gen.cam.ac.uk
[2]Present address: Life Sciences, University of Sussex, BN1 9QG Sussex, United Kingdom.

specialized to a germline-specific role on the lineages leading to vertebrates and arthropods.

After the early divergence of the Ago and Piwi subfamilies, subsequent duplications gave rise to three Piwi subfamily members [Ago3, Aubergine (Aub), and Piwi] and two Ago subfamily members (Ago1 and Ago2) in *Drosophila melanogaster*. All three Piwi subfamily genes are associated with the germline and bind piRNAs derived from TEs and other repetitive genomic elements: Ago3 and Aub amplify the piRNA signal through the "Ping-Pong" cycle (reviewed in Luteijn and Ketting 2013), and Piwi suppresses transposition by directing heterochromatin formation (Sienski *et al.* 2012). These functional differences are associated with contrasting selective regimes, with Aub evolving under positive selection (Kolaczkowski *et al.* 2011) and more rapidly than Ago3 and Piwi (Obbard *et al.* 2009a). In contrast, Ago1 binds miRNAs and regulates gene expression by inhibiting translation and marking transcripts for degradation (reviewed in Eulalio *et al.* 2008). This function imposes strong selective constraint on Ago1, resulting in slow evolution and very few adaptive substitutions (Obbard *et al.* 2006, 2009a; Kolaczkowski *et al.* 2011). Finally, Ago2 binds siRNAs from viruses (viRNAs) and TEs (endo-siRNAs), and functions in gene regulation (Wen *et al.* 2015), dosage compensation (Menon and Meller 2012), and the ubiquitous suppression of viruses (Li *et al.* 2002; van Rij *et al.* 2006) and TEs (Chung *et al.* 2008; Czech *et al.* 2008). Ago2 also evolves under strong positive selection, with frequent selective sweeps (Obbard *et al.* 2006, 2009a,b, 2011; Kolaczkowski *et al.* 2011), possibly driven by an arms race with virus-encoded suppressors of RNAi (VSRs) (Obbard *et al.* 2006; Marques and Carthew 2007; van Mierlo *et al.* 2014).

In contrast to *D. melanogaster*, from which most functional knowledge of Ago2 in arthropods is derived, an expansion of Ago2 has been reported in *D. pseudoobscura* (Hain *et al.* 2010), providing an opportunity to study how the RNAi pathway evolves after duplication. Given the roles of *D. melanogaster* Ago2 in antiviral defense (Li *et al.* 2002; van Rij *et al.* 2006), TE suppression (Chung *et al.* 2008; Czech *et al.* 2008), dosage compensation (Menon and Meller 2012), and gene regulation (Wen *et al.* 2015), we hypothesized that these *D. pseudoobscura* Ago2 paralogs may have diverged in function. To elucidate the evolution and function of Ago2 paralogs in *D. pseudoobscura* and its relatives, we identified and dated Ago2 duplication events across available *Drosophila* genomes and transcriptomes, tested for divergence in expression patterns between the Ago2 paralogs in *D. subobscura*, *D. obscura*, and *D. pseudoobscura*, and quantified the evolutionary rate and positive selection acting on each of these paralogs. We find that testis specificity of Ago2 paralogs has evolved repeatedly in the *obscura* group, and that the majority of paralogs show evidence of recent positive selection.

Materials and Methods

Identification of Ago2 homologs in the Drosophilidae

We used tBLASTx to identify Ago2 homologs in transcriptomes and genomes of 39 species of the Drosophilidae, using previously characterized Ago2 from the closest possible relative to provide the query for each species. If blast returned partial hits, we aligned all hits from the target species to all Argonautes from the query species and assigned hits to the appropriate Ago lineage based on a neighbor-joining tree. For each query sequence, we then manually curated partial blast hits into complete genes using Geneious v5.6.2 (http://www.geneious.com, Kearse *et al.* 2012) (see Supplemental Material, File S3 for sequence accessions).

Additionally, we used degenerate PCR to identify Ago2 paralogs in *D. azteca* and *D. affinis*, and paralog-specific PCR with a touchdown amplification cycle to validate the Ago2 paralogs identified in *D. subobscura*, *D. obscura*, and *D. pseudoobscura*. For each reaction, unincorporated primers were removed with ExonucleaseI (New England Biolabs) and 5′ phosphates were removed with Antarctic Phosphatase (NEB). The PCR products were sequenced by Edinburgh Genomics using BigDye V3 reagents on a capillary sequencer (Applied Biosystems, Foster City, CA), and Sanger sequence reads were trimmed and assembled using Geneious v.5.6.2 (http://www.geneious.com, Kearse *et al.* 2012). We also used a combination of PCR and blast searches to locate *D. pseudoobscura* Ago2a1 and Ago2a3, which lie on the unplaced "Unknown_contig_265" in release 3.03 of the *D. pseudoobscura* genome (all PCR primers are detailed in Table S4).

Phylogenetic analysis of drosophilid Ago2 paralogs

To characterize the evolutionary relationships between Ago2 homologs in the Drosophilidae, we aligned sequences using translational MAFFT (Katoh *et al.* 2002) with default parameters (File S1). We noted that there is a high degree of codon usage bias (CUB) in *D. pseudoobscura* Ago2e [effective number of codons (ENC) = 34.24] and *D. obscura* Ago2e (ENC = 40.36), and a lesser degree in *D. subobscura* Ago2f (ENC = 45.63) and *D. obscura* Ago2f (ENC = 48.39), and comparison with genome-wide patterns of codon usage bias placed these genes in the lower half of the distribution of ENC (Figure S5). To reduce the impact of CUB, which disproportionately affects synonymous sites, we stripped all third position sites in this analysis (Behura and Severson 2013). We then inferred a gene tree using the Bayesian approach implemented in BEAST v1.8.1 (Drummond *et al.* 2012) under a nucleotide model, assuming a general time reversible (GTR) substitution model, variation between sites modeled by a γ-distribution with four categories, and base frequencies estimated from the data. We used the default priors for all parameters, except tree shape (for which we specified a birth–death speciation model) and the date of the *Drosophila–Sophophora* split. To estimate a time scale for the tree, we specified a normal distribution for the date of this node using values based on mutation rate estimates in Obbard *et al.* 2012, with a mean value of 32 MYA, standard deviation of 7 MYA, and lower and upper bounds of 15 MYA and 50 MYA, respectively. We ran the analysis for 50 million steps, recording samples from the posterior every 1000 steps, and inferred a maximum clade credibility tree with TreeAnnotator v1.8.1 (Drummond *et al.* 2012). Note that

precise date estimates are not a primary focus of this study, but that other calibrations (Russo *et al.* 1995; Tamura 2004) would lead to more ancient estimates of divergence, and thus stronger evidence for selective maintenance.

Domain architecture and structural modeling of Ago2 paralogs in the obscura group

To infer the location of each domain in each paralog identified in *D. subobscura*, *D. obscura*, and *D. pseudoobscura*, we searched the Pfam database (Finn *et al.* 2009). To test for structural differences between the *D. pseudoobscura* paralogs, we built structural models of each paralog, based on the published X-ray crystallographic structure of human Ago2 (Schirle and Macrae 2012). We used the MODELER software in the Discovery Studio 4.0 Modeling Environment [Accelrys Software, San Diego (2013)] to calculate 10 models, selected the most energetically favorable for each protein, and assessed model quality with the 3D-profile option in the software. To assess variation in selective pressure across the structure of each paralog, we mapped variable residues onto each structure (Figure S7) using PyMol v.1.7.4.1 (Schrödinger).

Quantification of virus-induced expression of Ago2 paralogs

We exposed 48- to 96-hr posteclosion virgin males and females of *D. melanogaster*, *D. subobscura*, *D. obscura*, and *D. pseudoobscura* to *Drosophila* C virus (DCV), by puncturing the thorax with a pin contaminated with DCV at a dose of $\sim 4 \times 10^7$ tissue culture infective dose 50 (TCID50) per milliliter. Infection with DCV using this method has previously been shown to lead to a rapid and ultimately fatal increase in DCV titer in *D. melanogaster* and *obscura* group species (Longdon *et al.* 2015). All flies were incubated at 18C under a 12L:12D light cycle, with *D. melanogaster* on Lewis medium and *D. subobscura*, *D. obscura*, and *D. pseudoobscura* on banana medium. We sampled four to seven individuals per species at 0, 8, 16, 24, 48, and 72 hr postinfection. At each time point we extracted RNA using TRIzol reagent (Ambion) and a chloroform/isopropanol extraction, treated twice with TURBO DNase (Ambion), and reverse transcribed using M-MLV reverse transcriptase (Promega, Madison, WI) primed with random hexamers. We then quantified the expression of Ago2 paralogs in these samples by quantitative PCR (qPCR), using Fast Sybr Green (Applied Biosystems) and custom-designed paralog-specific qPCR primer pairs (see Table S5 for primer sequences). Due to their high level of sequence similarity (99.9% identity), no primer pair could distinguish between *D. pseudoobscura* Ago2a1 and Ago2a3, so combined expression of these two genes is presented as "Ago2a." All qPCR reactions for each sample were run in duplicate and scaled to the internal reference gene Ribosomal Protein L32 (RpL32). To capture the widest possible biological variation, the three biological replicates for each species each used a different wild-type genetic background (see Table S3 for backgrounds used).

Quantification of Ago2 paralog expression in different tissues and life stages

For *D. subobscura*, *D. obscura*, and *D. pseudoobscura*, we extracted RNA from the head, testis/ovaries, and carcass of 48- to 96-hr posteclosion virgin adults, with males and females extracted separately. Each sample consisted of 8–15 individuals in *D. subobscura*, 10 individuals in *D. obscura*, and 15 individuals in *D. pseudoobscura*. We then used qPCR to quantify the expression of each Ago2 paralog in each tissue, with two technical replicates per sample (reagents, primers, and cycling conditions as above). We carried out five replicates per species, each using a different wild-type background (see Table S3 for details of backgrounds used). To provide an informal comparison with the expression pattern of Ago2 before duplication (an "ancestral" expression pattern), we used the bases per kilobase of gene model per million mapped bases (BPKM) values for Ago2 calculated from RNA-sequencing (RNA-seq) data from the body (carcass and digestive system), head, ovary, and testis of 4-day-old *D. melanogaster* adults by Brown *et al.* 2014, scaling each BPKM value to the value for RpL32 in each tissue. Due to the design of that experiment, the body data are derived from pooled samples of males and females (Brown *et al.* 2014).

To quantify expression of Ago2 paralogs in *D. pseudoobscura* embryos, we collected eggs within 30 min of laying and used qPCR to measure the expression of each Ago2 paralog (reagents and primers as above) in two separate wild-type genetic backgrounds (MV8 and MV10). As above, we estimated an ancestral expression pattern of Ago2 before duplication from the BPKM values for Ago2 in 0- to 2-hr-old *D. melanogaster* embryos according to Brown *et al.* 2014, scaled to the BPKM value for RpL32 in embryos. To determine any changes in the expression of other *D. pseudoobscura* Argonautes (Ago1, Ago3, Aub, and Piwi) that are associated with Ago2 duplication, we measured their expression in adult tissues and embryos, as detailed above, and compared this with the expression of the Argonautes in *D. melanogaster* as measured by Brown *et al.* 2014.

Testing for evolutionary rate changes associated with tissue specificity of Ago2

We used codeml (PAML v4.4, Yang 1997) to fit variants of the M0 model (a single *dN/dS* ratio, ω) to the 65 drosophilid Ago2 homologs shown in Figure 1. All analyses of sequence evolution excluded the highly repetitive N-terminal glutamine-rich repeat regions, as these regions are effectively unalignable and are unlikely to conform to simple models of sequence evolution (Palmer and Obbard 2016). In contrast to the tree topology, which was based on first and second positions only, the alignment for the codeml analysis included all positions (File S2). To compare the evolutionary rates of ubiquitously expressed and testis-specific Ago2 paralogs, we fitted a model specifying one ω for the Ago2 paralogs that were shown to be testis-specific by qPCR (7 homologs), and another ω for the rest of the tree (58 homologs). We also fitted two models to account for rate variation between the *obscura* group Ago2 subclades. The first model specified a separate ω for the Ago2a subclade

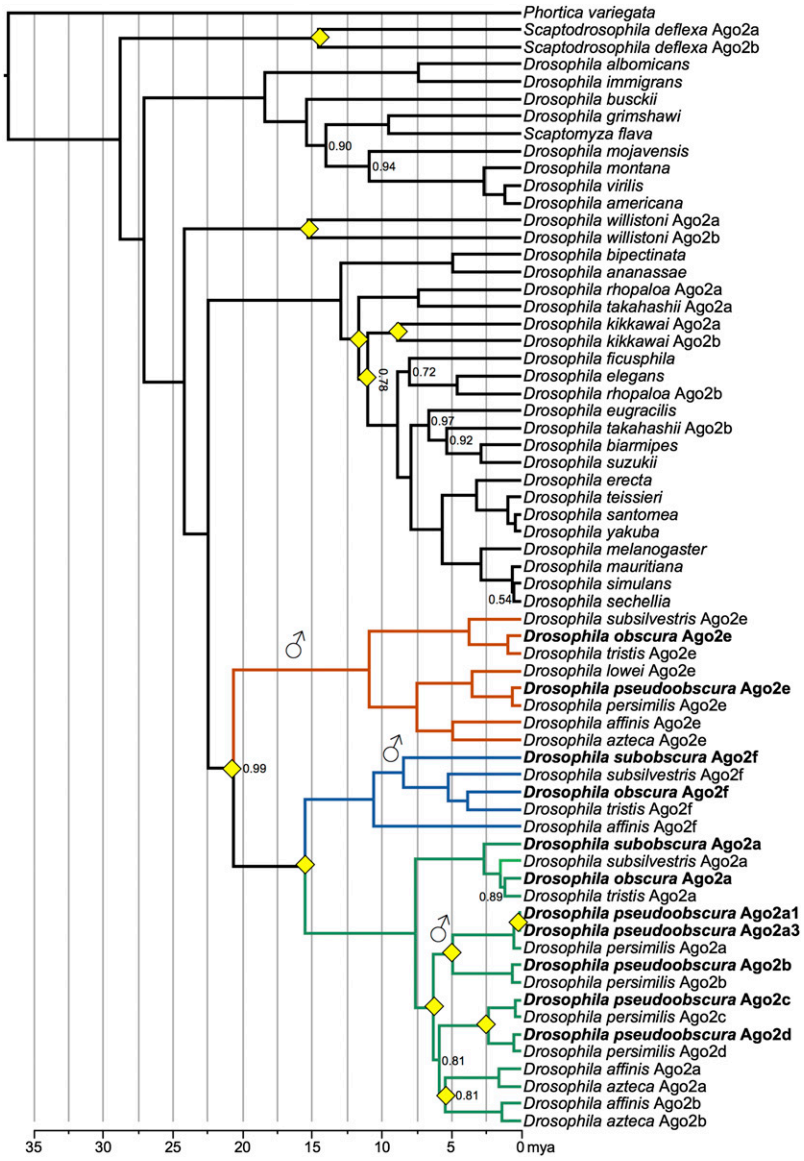


Figure 1 An approximately time-scaled Bayesian gene tree of Ago2 in the Drosophilidae. Duplication events are marked by yellow diamonds, Bayesian posterior support is shown for nodes for which it is <100% and the genes and species that are the focus of the present study are marked in boldface type. Ago2 has duplicated at least 12 times in the Drosophilidae: 7 times in the *obscura* group, twice early in the melanogaster group, and once each in the lineages leading to *D. willistoni*, *S. deflexa* and *D. kikkawai*. There has also been a potentially recent duplication of Ago2a on the *D. affinis*/*D. azteca* lineage (~5 MYA), although the low support for this node may suggest that these paralogs could also nest within the *D. pseudoobscura*/*D. persimilis* expansion, with one paralog sister to the Ago2a–Ago2b subclade and the other sister to the Ago2c–Ago2d subclade. After duplication, Ago2 paralogs in the *obscura* group have specialized to the testis 3 times independently (marked with ♂) and have been retained for an extended period of time (>10 MY in the case of Ago2e), suggesting an adaptive basis for testis specificity. The labeling a–e of paralogous clades corresponds to Hain *et al.* (2010) and is retained for consistency with subsequent publications, which also use these labels, while clade f is newly reported here. All genes were identified by BLAST, apart from the following, which were found by PCR: *D. teissieri* Ago2; *D. santomea* Ago2; *D. azteca* Ago2a, Ago2b, and Ago2e; and *D. pseudoobscura* Ago2a1 and Ago2a3.

(17 homologs), the Ago2e subclade (8 homologs), the Ago2f subclade (5 homologs), and the rest of the tree (35 homologs). The second model additionally incorporated an extra ω specified for the *D. pseudoobscura*–*D. persimilis* Ago2a–Ago2b subclade (3 homologs, all of which are testis-specific, in contrast with the rest of the *obscura* group Ago2a subclade). We used Akaike weights to assess which model provided the best fit to the data, given the number of parameters. As mentioned above, the high CUB seen in some Ago2 paralogues may affect PAML analyses by decreasing synonymous site divergence (*dS*) in those lineages, thereby inflating the *dN*/*dS* ratio (ω). However, we find no link between levels of CUB and the value of ω, suggesting that CUB is not impacting our PAML analyses.

Sequencing of Ago2 paralog haplotypes from D. subobscura, D. obscura, and D. pseudoobscura

To obtain genotype data for the Ago2 paralogs in *D. subobscura*, *D. obscura*, and *D. pseudoobscura*, we sequenced the Ago2 paralogs from six males and six females of each species, each from a different wild-collected line (detailed in Table S3, sequence polymorphism data in File S4). We extracted genomic DNA from each individual using the DNeasy Blood and Tissue kit (QIAGEN, Valencia, CA) and amplified and Sanger sequenced each Ago2 paralog from each individual (reagents and PCR primers as above, sequencing primers detailed in Table S5). We trimmed and assembled Sanger sequence reads using Geneious v.5.6.2 (http://www.geneious.com, Kearse *et al.* 2012), and identified polymorphic sites by eye. After sequencing Ago2a (annotated as a single gene in the *D. pseudoobscura* genome), we discovered two very recent Ago2a paralogs (which we denote Ago2a1 and Ago2a3), which had been cross-amplified. For each *D. pseudoobscura* individual, we therefore resequenced Ago2a3 using one primer targeted to its neighboring locus GA22965, and used this sequence to resolve polymorphic sites in the Ago2a1/Ago2a3 composite sequence, thereby gaining both genotypes for each individual.

For each Ago2 paralog, we inferred haplotypes from these sequence data using PHASE (Stephens *et al.* 2001), apart from the X-linked paralogs (Ago2a1, Ago2a3, and Ago2d) in *D. pseudoobscura* males, for which phase was obtained directly from the sequence data. The hemizygous haploid X-linked sequences were used in phase inference and should substantially improve the inferred phasing of female genotypes.

To quantify differences between paralogs in their population genetic characteristics, we aligned haplotypes using translational MAFFT (Katoh *et al.* 2002), and used DnaSP v.5.10.01 (Librado and Rozas 2009) to calculate the following summary statistics for each Ago2 paralogue: π (pairwise diversity, with Jukes–Cantor correction as described in Lynch and Crease 1990) at nonsynonymous (π_a) and synonymous (π_s) sites, Tajima's *D* (Tajima 1989) and ENC (Wright 1990). To compare the ENC for each gene with the genome as a whole, we used codonW v1.4.2 (Peden 1995) to calculate the ENC for the longest ORF from each gene or transcript in the genomes or transcriptomes of *D. subobscura*, *D. obscura*, and *D. pseudoobscura* (ORF sets detailed below). In each species, we then compared the ENC values of each Ago2 paralog with this genome-wide ENC distribution.

Testing for positive selection on Ago2 paralogs in the obscura group

We used McDonald–Kreitman (MK) tests (McDonald and Kreitman 1991) to test for positive selection on each Ago2 paralog. For each paralog, we chose an outgroup with divergence at synonymous sites (K_S) in the range 0.1–0.2 where possible. However, the prevalence of duplications and losses of Ago2 paralogs in the *obscura* group meant that for some tests no suitably divergent extant outgroup existed. In these cases, we reconstructed hypothetical ancestral sequences using the M0 model provided by codeml from PAML (Yang 1997). To assess the effect of these outgroup choices on our results, we repeated each test with another outgroup and found no effect of outgroup choice on the significance of any tests, and only marginal differences in estimates of α and ω_α (results of tests using primary and alternative outgroups are detailed in Table S1 and Table S2).

A complementary approach to identifying positive selection is to test for reduced diversity at a locus compared with the genome as a whole. To compare the diversity of each *D. pseudoobscura* Ago2 paralog with the genome-wide distribution of synonymous site diversity, we used genomic data for 12 lines generated by McGaugh *et al.* 2012. We mapped short reads to the longest ORF for each gene in the R3.2 gene set using Bowtie2 v2.1.0 (Langmead *et al.* 2009) and estimated synonymous site diversity (θ_W based on fourfold synonymous sites) at each ORF using PoPoolation (Kofler *et al.* 2011). We then plotted the distribution of synonymous site diversity, limited to genes in the size range of 0.75–3 kb for comparability with the Ago2 paralogs, and compared the fourfold synonymous site diversity levels of each *D. pseudoobscura* Ago2 paralog with this distribution. Some *D. pseudoobscura* paralogs are located on autosomes (Ago2b, Ago2c, and Ago2e) and some on the X chromosome (Ago2a1, Ago2a3, and Ago2d). Therefore, because of the different population genetic expectations for autosomal and X-linked genes (Vicoso and Charlesworth 2006), we examined separate distributions for autosomal and X-linked genes. To provide an additional test for reduced diversity at *D. pseudoobscura* Ago2 paralogs, we performed maximum-likelihood Hudson–Kreitman–Aguadé (HKA) tests (Wright and Charlesworth 2004), using divergence from *D. affinis* and intraspecific polymorphism data for 84 *D. pseudoobscura* loci generated by Haddrill *et al.* (2010). We performed 63 tests to encompass all one-, two-, three-, four-, five-, and six-way combinations of the paralogs and calculated Akaike weights from the resulting likelihood estimates to provide an estimate of the level of support for each combination.

To infer a genome-wide distribution of synonymous site diversity for *D. obscura* and *D. subobscura*, for which genomic data are unavailable, we used pooled transcriptome data from wild-collected adult male flies that had previously been generated for surveys of RNA viruses (van Mierlo *et al.* 2014; Webster *et al.* 2016). To generate a *de novo* transcriptome for each species, we assembled short reads with Trinity r20140717 (Grabherr *et al.* 2011). For each species, we mapped short reads from the pooled sample to the longest ORF for each transcript, estimated synonymous site diversity at each locus using PoPoolation (Kofler *et al.* 2011), and plotted the distribution of diversity (as described above for *D. pseudoobscura*). The presence of heterozygous sites in males (identified by Sanger sequencing) confirmed that all Ago2 paralogs in *D. subobscura* and *D. obscura* are autosomal: we therefore compared the synonymous site diversity for these paralogs with the autosomal distribution and do not show the distributions for putatively X-linked genes. Our use of transcriptome data for *D. obscura* and *D. subobscura* will bias the resulting diversity distributions in three ways. First, variation in expression level will cause individuals displaying high levels of expression to be overrepresented among reads, downwardly biasing diversity. Second, highly expressed genes are easier to assemble, and highly expressed genes tend to display lower genetic diversity (Pal *et al.* 2001; Lemos *et al.* 2005). Third, high-diversity genes are harder to assemble, *per se*. However, as all three biases will tend to artifactually reduce diversity in the genome-wide data set relative to Ago2, this makes our finding that Ago2 paralogs display unusually low diversity conservative.

Identifying selective sweeps in Ago2 paralogs of D. pseudoobscura

To test whether the unusually low diversity seen in the *D. pseudoobscura* Ago2 paralogs is due to recent selection or generally reduced diversity in that region of the genome, we compared diversity at each paralog to diversity in their neighboring regions. We obtained sequence data for the 50 kb either side of each of these paralogs from the 11 whole genomes detailed in McGaugh *et al.* 2012 (SRA044960.1, SRA044955.2, and SRA044956.1). Note that the very high

similarity of these Ago2 paralogues means that they cannot be accurately assembled from short read data, and are not present in the data from McGaugh *et al.* 2012. For each genome, we therefore replaced the poorly assembled region corresponding to the paralog with one of our own Sanger-sequenced haplotypes, making a set of 11 ~102-kb sequences for each paralog. We aligned these sequences using PRANK (Löytynoja and Goldman 2005) with default settings and calculated Watterson's θ at all sites in a sliding window across each alignment, with a window size of 5 kb and a step of 1 kb. For Ago2a1 and Ago2a3, which are located in tandem, we analyzed the same genomic region. Since our Ago2 haplotypes were sampled from a different North American population of *D. pseudoobscura* than those of McGaugh *et al.* 2012, an apparent reduction in local diversity might result from differences in diversity between the two populations. We therefore also repeated these analyses on a data set in which our Sanger sequenced haplotypes were removed, leaving missing data.

To test explicitly for selective sweeps at each region, we used Sweepfinder (Nielsen *et al.* 2005b) to calculate the likelihood and location of a sweep in or near each Ago2 paralog. We specified a grid size of 20,000, a folded frequency spectrum for all sites, and included invariant sites. To infer the significance of any observed peaks in the composite likelihood ratio, we used ms (Hudson 2002) to generate 1000 samples of 11 sequences under a neutral coalescent model. We generated separate samples for each region surrounding an Ago2 paralog, conditioning on the number of polymorphic sites observed in that region, the sequence length equal to the alignment length, and an effective population size of 10^6 (based on a previous estimate for *D. melanogaster* by Li and Stephan 2006). We specified the recombination rate at 5 cM/Mb, a conservative value based on previous estimates for *D. pseudoobscura* (McGaugh *et al.* 2012), which will lead to larger segregating linkage groups and therefore a more stringent significance threshold.

Data availability

All new sequences produced in this study have been submitted to GenBank as KX016642–KX016771.

Results

Ago2 has undergone numerous ancient and recent duplications in the obscura group

Ago2 duplications had previously been noted in *D. pseudoobscura* (Hain *et al.* 2010), but their age and distribution in other species was unknown. We used BLAST (Altschul *et al.* 1997) and PCR to identify 65 Ago2 homologs in 39 species sampled across the Drosophilidae, including 30 homologs in 9 *obscura* group species. Using PCR and Sanger sequencing, we verified that the paralogs in *D. subobscura*, *D. obscura*, and *D. pseudoobscura* are genuine distinct loci, and not artifacts of erroneous assembly. Additionally, we verified that all paralogs possess introns and so are most likely to be the product of segmental duplication rather than retrotransposition. This is perhaps unsurprising, given that segmental duplicates are generally retained at a higher rate than retrotransposed duplicates, despite the rate of retrotransposition being higher than segmental duplication (Hahn 2009).

To characterize the relationships between Ago2 homologs in the *obscura* group and the other Drosophilidae, and estimate the date of the duplication events that produced them, we carried out a strict clock Bayesian phylogenetic analysis (Figure 1). This showed that there are early diverging Ago2 clades in the *obscura* group: the Ago2e subclade that diverged from other Ago2 paralogs ~21 MYA (±10 MY) and the Ago2a and Ago2f subclades that were produced by a gene duplication event ~16 MYA (±7 MY). Subsequently there have been a series of more recent duplications in the *D. pseudoobscura* subgroup Ago2a–d lineage. Using published genomes, transcriptomes, and PCR, we were unable to identify Ago2e in *D. subobscura*, Ago2e or Ago2f in *D. lowei*, or Ago2f in *D. pseudoobscura*, *D. persimilis*, and *D. azteca*. While apparent losses may reflect a lack of genomic data (*D. subobscura*, *D. lowei*, and *D. azteca*), incomplete genome assemblies (*D. pseudoobscura* and *D. persimilis*), or unexpressed genes in transcriptome surveys, we attempted to validate the losses observable in *D. pseudoobscura* and *D. subobscura* by extensive PCR and were again unable to recover these genes from those two species.

In release 3.03 of the *D. pseudoobscura* genome, the paralogs Ago2b–Ago2e have confirmed locations, but Ago2a1 and Ago2a3 (the very recent paralogs newly identified here) lie in tandem on an unplaced contig with a third incomplete copy (Ago2a2) between them. We used PCR to confirm the existence, orientation, and relative positioning of these genes and to identify the location of this contig, which lies in reverse orientation on chromosome XL-group1a (predicted coordinates 3,463,701–3,489,689). We then combined this information with our phylogenetic analysis to reconstruct the positional evolution of *D. pseudoobscura* Ago2 paralogs (Figure S1). We found that *D. pseudoobscura* Ago2d is syntenic with *D. melanogaster* Ago2, indicating that Ago2d is the ancestral paralog in this species. We also found that Ago2 paralogs have translocated throughout the *D. pseudoobscura* genome (Figure S1) and are situated on autosomes (Ago2b, Ago2c, and Ago2e) and both arms of the X chromosome (Ago2a1, Ago2a3, and Ago2d). It should be noted that a lack of genomic data precludes similar synteny analysis for any other *obscura* group species; our naming of the Ago2 paralogs in these species as Ago2a (or Ago2a and Ago2b in the case of *D. affinis* and *D. azteca*) reflects their position within the Ago2a subclade, rather than a syntenic relationship or otherwise with *D. pseudoobscura* Ago2a1 and Ago2a3.

Ago2 paralogs in D. subobscura, D. obscura, and D. pseudoobscura are probably functional

Our phylogenetic analysis (Figure 1) revealed that the Ago2 paralogs in the *obscura* group have retained coding sequences for millions of generations, showing that they have

remained functional for this period. They have also retained PAZ and PIWI domains and a bilobal structure (characteristic of Argonaute proteins), suggesting that they are part of a functional RNAi pathway. In *D. melanogaster* Ago2 plays a key role in antiviral immunity, but is ubiquitously and highly expressed in both males and females and is not strongly induced by viral challenge (Figure 2a, Aliyari *et al.* 2008). To test whether this expression pattern has been conserved after Ago2 duplication, or whether any Ago2 paralogs have become inducible by viral challenge, we measured the expression of each Ago2 paralog in female and male *D. subobscura*, *D. obscura*, and *D. pseudoobscura* after infection with DCV. These species are separated by ~10 MY of evolution and represent the three major clades within the *obscura* group. Members of the *obscura* group are highly susceptible to DCV, supporting high viral titres and displaying rapid mortality (Longdon *et al.* 2015). We found that only one paralog is expressed in both sexes at a high level in *D. subobscura* (Ago2a), *D. obscura* (Ago2a), and *D. pseudoobscura* (Ago2c). These paralogs show a similar pattern of expression to *D. melanogaster* Ago2, being expressed constitutively throughout the time course rather than induced by viral infection (Figure 2). Unexpectedly, and with only one exception, the other Ago2 paralogs in all species were expressed exclusively in males (Figure 2, b–d), raising the possibility that these duplicates have specialized to a sex-specific role. The one exception was *D. pseudoobscura* Ago2d, which is the ancestral paralog in this species (inferred by synteny), and for which we could not detect any expression.

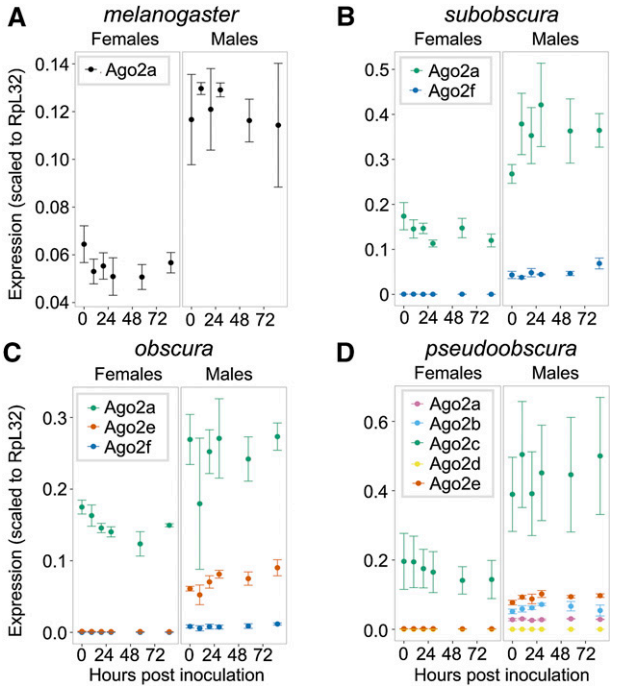


Figure 2 Expression patterns of Ago2 paralogs under challenge with DCV. In each *obscura* group species, only one Ago2 paralog has retained the ancestral pattern of ubiquitous stable expression in each sex (illustrated by *D. melanogaster*). In contrast, all other paralogs are expressed in males only (in *D. pseudoobscura* females, Ago2a, Ago2b, Ago2d, and Ago2e are all unexpressed throughout the time course). The only exception to this is *D. pseudoobscura* Ago2d, which is unexpressed in either sex. The high degree of sequence similarity between Ago2a1 and Ago2a3 prevented us from amplifying these genes separately in qPCR, and here they are combined as "Ago2a." Error bars indicate one standard error estimated from two technical replicates in each of three different genetic backgrounds. Apparent differences in expression between sexes and species should be interpreted with caution, as these may be driven by differences in expression levels of the reference gene (RpL32).

Ago2 paralogs have repeatedly specialized to the testis

To determine whether the strongly male-biased expression pattern is associated with a testis-specific role, we quantified the tissue-specific expression patterns of Ago2 paralogs in *D. subobscura*, *D. obscura*, and *D. pseudoobscura*. In *D. melanogaster*, the single copy of Ago2 was expressed in all adult tissues (Figure 3D), and transcripts were present in the embryo (Figure S2). In *D. subobscura*, *D. obscura*, and *D. pseudoobscura*, we found that the Ago2 paralogs exhibited striking differences in their tissue-specific patterns of expression (Figure 3, A–C). In each species, one paralog has retained the ancestral ubiquitous expression pattern in adult tissues. In contrast, every other paralog was expressed only in the testis, except for the nonexpressed *D. pseudoobscura* Ago2d. None of the testis-specific paralogs in *D. pseudoobscura* was detectable in embryos (Figure S2).

Interestingly, the ubiquitously expressed paralog in *D. subobscura* and *D. obscura* is the ancestral gene (Ago2a in both cases, as inferred by synteny with *D. melanogaster*), but in *D. pseudoobscura* another paralog (Ago2c) has evolved the ubiquitous expression pattern, and the ancestral gene (Ago2d) was not expressed at a detectable level in any tissue. When interpreted in the context of the phylogenetic relationships between these paralogs, the most parsimonious explanation is that testis specificity evolved at least three times: first at the base of the Ago2e clade, second at the base of the Ago2f clade, and third at the base of the *D. pseudoobscura–D. persimilis* Ago2a–Ago2b subclade (Figure 1).

Testis specificity is associated with faster protein evolution

To test for differences in evolutionary rate between testis-specific and ubiquitously expressed Ago2 paralogs, we fitted sequence evolution models to the set of drosophilid Ago2 sequences depicted in Figure 1 using codeml (PAML, Yang 1997). These tests estimate separate *dN/dS* ratios (ω) for different subclades in the gene tree, providing a test for differential rates of protein evolution. We found that most support (Akaike weight = 0.99) falls behind a model specifying a different ω for each obscura group Ago2 subclade, and another separate ω for the *D. pseudoobscura–D. persimilis* Ago2a–Ago2b subclade. Under this model, the testis-specific *D. pseudoobscura–D. persimilis* Ago2a–Ago2b subclade has the highest rate of protein evolution (ω = 0.32 ± 0.047 SE), followed by the testis-specific Ago2f subclade (ω = 0.21 ± 0.014), the ubiquitous Ago2a subclade (ω = 0.19 ± 0.012), the testis-specific Ago2e subclade (ω = 0.16 ± 0.010), and

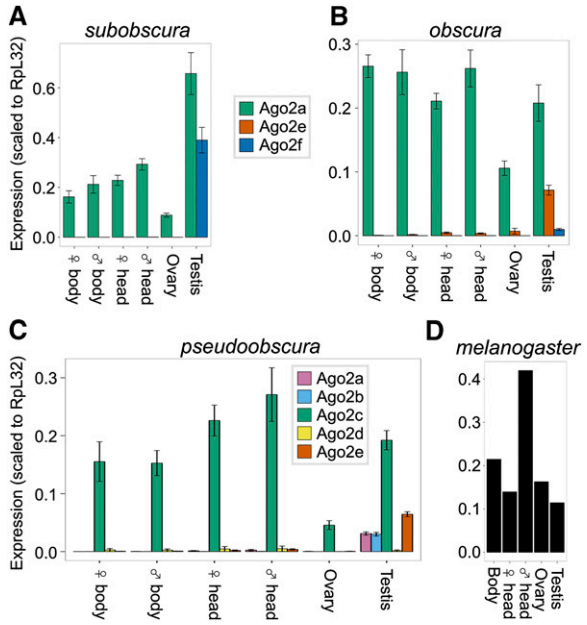


Figure 3 Tissue-specific expression patterns of Ago2 paralogs. In each of the three *obscura* group species tested, one paralog has retained the ancestral ubiquitous expression pattern, while the others have specialized to the testis (with the exception of *D. pseudoobscura* Ago2d). The high degree of sequence similarity between Ago2a1 and Ago2a3 prevented us from amplifying these genes separately in qPCR, and here they are combined as "Ago2a." Error bars indicate one standard error estimated from two technical replicates in each of five different genetic backgrounds. *D. melanogaster* expression levels were taken from a single RNA-seq experiment (Brown *et al.* 2014).

finally the other Drosophilid Ago2 sequences ($\omega = 0.12 \pm 0.002$). This shows that the evolution of testis specificity was accompanied by an increase in the rate of protein evolution following two of the three duplications. We also used the Bayes empirical Bayes sites test in codeml to identify codons evolving under positive selection across the entire gene tree, and the branch-sites test to identify codons under positive selection in the *obscura* group Ago2 subclade. While we found no positively selected codons with the sites test, we identified three codons under positive selection (297, 338, and 360) in the *obscura* group Ago2 subclade with the branch-sites test (likelihood ratio test M8 *vs.* M8a, $P < 0.005$).

MK tests identify strong positive selection on D. pseudoobscura Ago2e

Changes in evolutionary rate after the evolution of testis specificity may occur as a result of changes in positive selection or changes in selective constraint. However, unless there are multiple substitutions within single codons, this will be hard to detect using methods such as codeml. Therefore, as a second test for positive selection on Ago2 paralogs in *D. subobscura*, *D. obscura*, and *D. pseudoobscura*, we gathered intraspecies polymorphism data for each Ago2 paralog in these species (File S4) and performed MK tests (Table S1). The MK test uses a comparison of the numbers of fixed differences between species at nonsynonymous (D_n) and synonymous (D_s) sites, and polymorphisms within a species at nonsynonymous (P_n) and synonymous (P_s) sites to infer the action of positive selection. If all mutations are either neutral or strongly deleterious, the D_n/D_s ratio should be approximately equal to the P_n/P_s ratio; however, if there is positive selection, an excess of nonsynonymous differences is expected (McDonald and Kreitman 1991). The majority of MK tests were nonsignificant (Fisher's exact test, $P > 0.1$), despite often displaying relatively high K_A/K_S ratios *e.g.*, *D. pseudoobscura* Ago2a1 ($K_A/K_S = 0.34$), Ago2b ($K_A/K_S = 0.43$), and Ago2d ($K_A/K_S = 0.36$). However, the low diversity at these loci (<10 polymorphic sites in most cases; see below) means that the MK approach has little power, and that estimates of the proportion of substitutions that are adaptive (α) are likely to be poor. In contrast to the other loci, our MK analysis identified strong positive selection acting on *D. pseudoobscura* Ago2e—which has relatively high genetic diversity—with α at 100% ($\alpha = 1.00$; Fisher's exact test, $P = 0.0004$). This result is driven by the extreme dearth of nonsynonymous-to-synonymous polymorphisms (0 P_n to 17 P_s), despite substantial numbers of fixed differences (77 D_n to 120 D_s), and its statistical significance is robust to the choice of outgroup (Table S2).

The majority of Ago2 paralogs have extremely low levels of sequence diversity

When strong selection acts to reduce genetic diversity at a locus, it can also reduce diversity at linked loci before recombination can break up linkage (Maynard Smith and Haigh 1974). Recent positive selection can therefore be inferred from a reduction in synonymous site diversity compared with other genes. Because MK tests can detect only multiple long-term substitutions, and are hampered by low diversity, diversity-based approaches offer a complementary way to detect very recent strong selection. We therefore compared the synonymous site diversity at each Ago2 paralog in *D. pseudoobscura* with the distribution of genome-wide synonymous site diversity. We found that all *D. pseudoobscura* paralogs have unusually low diversity, relative to other loci: Ago2a1, Ago2b, and Ago2c fall into the lowest percentile, Ago2a3 and Ago2d into the second lowest percentile, and Ago2e into the eighth lowest percentile (Figure S4). A multilocus extension of the HKA test (ML-HKA, Wright and Charlesworth 2004) confirmed that the diversity of Ago2a1–Ago2e is significantly lower than the *D. pseudoobscura* genome as a whole (Akaike weight = 0.98).

Unfortunately, population-genomic data are not available for *D. subobscura* and *D. obscura*, preventing a similar analysis. However, we found similar results for Ago2a and Ago2e when comparing the diversity of *D. subobscura* and *D. obscura* Ago2 paralogs to levels of diversity inferred from transcriptome data (data from Webster *et al.* 2016), suggesting that this effect is not limited to *D. pseudoobscura*, and these genes may therefore have been recent targets of selection in multiple species. In *D. obscura*, Ago2a and Ago2e fall into the 2nd and 4th lowest diversity percentile, respectively, whereas Ago2f falls into the 19th percentile (Figure S4). In *D. subobscura*,

Ago2a falls into the 7th percentile, whereas Ago2f falls into the 16th percentile (Figure S4). The prevalence of low intraspecific diversity for testis-specific paralogs is consistent with recent selective sweeps, suggesting that positive selection, not merely relaxation of constraint, has contributed to the increased evolutionary rate seen after specialization to the testis.

Four of six D. pseudoobscura Ago2 duplicates show a strong signature of recent hard selective sweeps

The impact of selection on linked diversity (a selective sweep) is expected to leave a characteristic footprint in local genetic diversity around the site of selection, and this forms the basis of explicit model-based approaches to detect the recent action of positive selection (Nielsen *et al.* 2005a). For *D. pseudoobscura*, population genomic data for 11 haplotypes is available from McGaugh *et al.* (2012), permitting an explicit model-based test for recent hard selective sweeps near to Ago2 paralogs. We therefore combined our Ago2 data with 111-kb-long haplotypes from McGaugh *et al.* (2012) to analyze the neighboring region around each paralog. Ago2a1 and Ago2a3 form a tandem repeat and were therefore analyzed together as a single potential sweep. We found strong evidence for recent selective sweeps at or very close to Ago2a1/3, Ago2b, and Ago2c, which display sharp troughs in their diversity levels and large peaks in the composite likelihood of a sweep, which far exceed a significance threshold derived from coalescent simulation ($P < 0.01$; Figure 4). These localized reductions in diversity remain when our own Ago2 haplotype data are removed, showing the results are robust to the fact that our Ago2 sequence data are derived from a different population to the genome-wide data of McGaugh *et al.* (2012) [Figure S6; note that sequence data for Ago2 paralogs cannot be derived from the data of McGaugh *et al.* (2012), because of their extreme similarity]. In addition, there is ambiguous evidence for a sweep at Ago2d, in the form of one significant ($P < 0.01$) likelihood peak just upstream of the paralog, but two other peaks ~1 kb and ~3 kb further upstream. There is no evidence for a hard sweep at Ago2e, which has no diversity trough or likelihood peak.

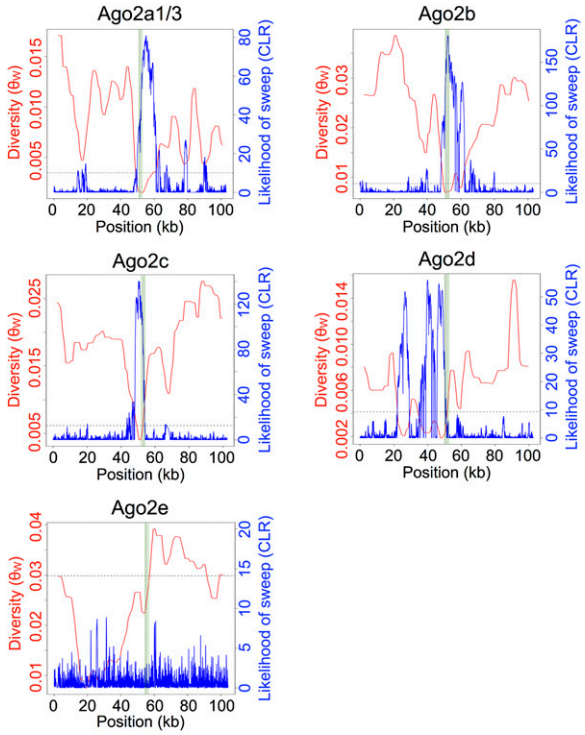


Figure 4 Selective sweeps at *D. pseudoobscura* Ago2 paralogs. For each paralog, diversity at all sites (Watterson's θ) is displayed in red, and the likelihood of a sweep centered at that site (composite likelihood ratio, CLR) is displayed in blue. The gene region containing the paralog is represented by the shaded vertical bar, and the significance threshold for the CLR is displayed by the horizontal dotted line ($P < 0.01$, derived from the 10th-highest CLR out of 1000 coalescent simulations, assuming constant recombination rate and N_e). There is strong evidence for sweeps at Ago2a, Ago2b, and Ago2c, indicated by troughs in their diversity levels and peaks in the likelihood of a sweep.

Discussion

Testis specificity may indicate a loss of antiviral function

We have found that Ago2 paralogs in the *obscura* group have repeatedly evolved divergent expression patterns after duplication, with the majority of paralogs specializing to the testis. This is the first report of testis specificity for any arthropod Ago2, which is ubiquitously expressed in *D. melanogaster* (Celniker *et al.* 2009), and provides a strong indication that these paralogs have diverged in function. This testis specificity (Figure 3) suggests that these Argonautes are likely to have lost their ancestral ubiquitous antiviral role. Additionally, the constant level of expression of testis-specific paralogs under DCV infection (Figure 2) suggests that they have not evolved an inducible response to viral infection, either restricted to the testis or in other tissues. In contrast, one paralog in each species has retained the ubiquitous expression pattern seen in *D. melanogaster* (*D. subobscura* Ago2a, *D. obscura* Ago2a, and *D. pseudoobscura* Ago2c; Figure 3), suggesting that these paralogs have retained roles in antiviral defense (Li *et al.* 2002; van Rij *et al.* 2006), dosage compensation (Menon and Meller 2012), and/or somatic TE suppression (Chung *et al.* 2008; Czech *et al.* 2008).

Both ubiquitous and testis-specific Ago2 paralogs show evidence of recent positive selection

We identified selective sweeps at the ubiquitously expressed Ago2 paralog in *D. pseudoobscura* Ago2c, and very low diversity in the ubiquitously expressed Ago2 paralogs of *D. subobscura* and *D. obscura* (Ago2a), suggesting that all of these genes may have recently experienced strong positive selection. Four randomly chosen genes that are testis-specific in *D. melanogaster* (Mikhaylova *et al.* 2008) do not fall into the low-diversity tails of the genome-wide diversity distributions of *D. obscura* [β-tubulin (85D), Hsp60C, sungrazer and roughex] and *D. subobscura* [β-tubulin (85D), Hsp60C, sungrazer, and Calcutta cup], suggesting that the low diversity of testis-specific Ago2 paralogs in these species is not a general

consequence of testis-specific expression. This is consistent with previous findings of strong selection and rapid evolution of Ago2 in *D. melanogaster* (Obbard *et al.* 2006, 2009b, 2011), which has also experienced recent sweeps in *D. melanogaster*, *D. simulans*, and *D. yakuba* (Obbard *et al.* 2011), and across *Drosophila* more broadly (Kolaczkowski *et al.* 2011). It has previously been suggested that this is driven by arms-race coevolution with viruses (Obbard *et al.* 2009a; Kolaczkowski *et al.* 2011), some of which encode viral suppressors of RNAi (VSRs) that block Ago2 function (Bronkhorst and van Rij 2014). The presence of VSR-encoding viruses, such as Nora virus, in natural *obscura* group populations (Webster *et al.* 2016), combined with the host specificity that can be displayed by VSRs (van Mierlo *et al.* 2014), suggest that arms-race dynamics may also be driving the rapid evolution of ubiquitously expressed Ago2 paralogs in the *obscura* group.

Potential testis-specific functions

In contrast to their ancestral ubiquitous expression pattern, the dominant fate for Ago2 paralogs in the *obscura* group appears to have been specialization to the testis. Paralogs often undergo a brief period of testis specificity soon after duplication (Assis and Bachtrog 2013, 2015), and this has given rise to the "out-of-the-testis" hypothesis, in which new paralogs are initially testis-specific before evolving functions in other tissues (Kaessmann 2010). However, two lines of evidence suggest an adaptive basis for the testis specificity observed for the *obscura* group Ago2 paralogs. First, testis specificity has been retained for >10 MY in Ago2e and Ago2f, in contrast to the broadening of expression over time expected under the out-of-the-testis hypothesis (Kaessmann 2010; Assis and Bachtrog 2013). Second, all testis-specific Ago2 paralogs in *D. pseudoobscura* show evidence either of long-term positive selection (MK test for the high-diversity Ago2e) or of recent selective sweeps (in low-diversity Ago2a1/3 and Ago2b), and the testis-specific *D. obscura* Ago2e displays a reduction in diversity, potentially driven by selection.

Under a subfunctionalization model for Ago2 testis specialization, five candidate selective pressures seem likely: testis-specific dosage compensation, antiviral defense, gene regulation, TE suppression, and/or the suppression of meiotic drive. Of these, testis-specific dosage compensation seems the least likely to drive testis specificity because the MSL complex, which Ago2 directs to X-linked genes to carry out dosage compensation in the soma of *D. melanogaster*, is absent from testis (Conrad and Akhtar 2012). Testis-specific antiviral defense seems similarly unlikely, as the only known paternally transmitted *Drosophila* viruses (Sigmaviruses; Rhabdoviridae) pass through both the male and female gametes (Longdon and Jiggins 2012), and so the potential benefits of testis specificity seem unclear. Alternatively, testis-specific Ago2 duplicates could be coevolving with other testis-specific genes through the hairpin RNA pathway, in which siRNAs generated from endogenous hairpin-forming RNAs (hpRNAs) bind Ago2 and regulate the expression of host genes (Okamura *et al.* 2008). In *D. melanogaster*, hpRNA-derived siRNAs target testis-specific genes involved in male fertility and coevolve with these targets to maintain base complementarity (Wen *et al.* 2015). If a similar pathway operates in the *obscura* group, Ago2 paralogs could have specialized to the hpRNA pathway in order to regulate testis-specific genes more effectively.

Finally, the suppression of TEs or meiotic drive seem promising candidate selective forces. First, numerous TEs transpose preferentially in the testis, such as *Penelope* in *D. virilis* (Rozhkov *et al.* 2010) and *copia* in *D. melanogaster* (Pasyukova *et al.* 1997; Morozova *et al.* 2009), which could impose a selection pressure on Ago2 paralogs to provide a testis-specific TE suppression mechanism. It should be noted that all members of the canonical anti-TE Piwi subfamily (Ago3, Aub, and Piwi) are also expressed in *obscura* group testis (Figure S3), suggesting that if Ago2 paralogs have specialized to suppress TEs, they are doing so alongside the existing TE suppression mechanism. Second, testis specificity could have evolved to suppress meiotic drive, which is prevalent (in the form of sex-ratio distortion) in the *obscura* group (Gershenson 1928; Sturtevant and Dobzhansky 1936; Wu and Beckenbach 1983; Jaenike 2001; Unckless *et al.* 2015), and which is suppressed by RNAi-based mechanisms in other species (Tao *et al.* 2007; Kotelnikov *et al.* 2009; Gell and Reenan 2013). A high level of meiotic drive in the *obscura* group could therefore impose selection for the evolution of novel suppression mechanisms, leading to the repeated specialization of Ago2 paralogs to the testis.

Prospects for novel functions during the evolution of RNAi

The functional specialization that we observe for *obscura* group Ago2 paralogs raises the prospect of undiscovered derived functions following Argonaute expansions in other lineages. Ago2 has duplicated frequently across the arthropods, with expansions present in insects (*D. willistoni* (Figure 1) and *Musca domestica*, Scott *et al.* 2014), crustaceans (*Penaeus monodon*, Leebonoi *et al.* 2015), and chelicerates (*Tetranychus urticae*, *Ixodes scapularis*, *Mesobuthus martensii*, and *Parasteatoda tepidariorum*, Palmer and Jiggins 2015). The prevalence of testis specificity in *obscura* group Ago2 paralogs raises the possibility that specialization to the germline may be more widespread following Argonaute duplication. The expression of Ago2 paralogs has previously been characterized in *P. monodon* and shows that one paralog has indeed specialized to the germline of both males and females, but not the testis alone (Leebonoi *et al.* 2015). Publicly available RNA-seq data from the head, gonad, and carcass of male and female *M. domestica* (GSE67065, Meisel *et al.* 2015) suggest that neither *M. domestica* Ago2 paralog has specialized to the testis (Figure S8). However, public data from the head, thorax, and abdomen of male and female *D. willistoni* (GSE31723, Meisel *et al.* 2012) show that one *D. willistoni* Ago2 paralog (FBgn0212615) is expressed ubiquitously, while the other (FBgn0226485) is expressed only in the male abdomen

(Figure S8), consistent with the evolution of testis specificity after duplication. This raises the possibility that a testis-specific selection pressure may be driving the retention and specialization of Ago2 paralogs across the arthropods.

In conclusion, we have identified rapid and repeated evolution of testis specificity after the duplication of Ago2 in the *obscura* group, associated with low genetic diversity and signatures of strong selection. Ago2 and other RNAi genes have undergone frequent expansions in different eukaryotic lineages (Mukherjee *et al.* 2013; Lewis *et al.* 2016) and have been shown to switch between ubiquitous and germline- or ovary-specific functions in isolated species. This study provides evidence for the evolution of a new testis-specific RNAi function and suggests that positive selection may act on young paralogs to drive the rapid evolution of novel RNAi mechanisms across the eukaryotes.

Acknowledgments

We thank Ben Longdon and Brian Charlesworth for providing us with strains of *D. obscura* and *D. pseudoobscura*, respectively, and Francis Jiggins for providing us with DCV. This work was supported by a Natural Environment Research Council doctoral training grant (NERC DG NE/J500021/1 to S.H.L.), the Academy of Finland (265971 to H.S.), a University of Edinburgh chancellor's fellowship, a Wellcome Trust research career development fellowship (WT085064 to D.J.O.), and a Wellcome Trust strategic award to the Centre for Immunity, Infection and Evolution (WT095831).

Literature Cited

Aliyari, R., Q. Wu, H.-W. Li, X.-H. Wang, F. Li *et al.*, 2008 Mechanism of induction and suppression of antiviral immunity directed by virus-derived small RNAs in Drosophila. Cell Host Microbe 4: 387–397.

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang *et al.*, 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25: 3389–3402.

Assis, R., and D. Bachtrog, 2013 Neofunctionalization of young duplicate genes in Drosophila. Proc. Natl. Acad. Sci. USA 110: 17409–17414.

Assis, R., and D. Bachtrog, 2015 Rapid divergence and diversification of mammalian duplicate gene functions. BMC Evol. Biol. 15: 138.

Behura, S. K., and D. W. Severson, 2013 Codon usage bias: causative factors, quantification methods and genome-wide patterns: with emphasis on insect genomes. Biol. Rev. Camb. Philos. Soc. 88: 49–61.

Bronkhorst, A. W., and R. P. van Rij, 2014 The long and short of antiviral defense: small RNA-based immunity in insects. Curr. Opin. Virol. 7C: 19–28.

Brown, J. B., N. Boley, R. Eisman, G. E. May, M. H. Stoiber *et al.*, 2014 Diversity and dynamics of the Drosophila transcriptome. Nature 512: 393–399.

Buck, A. H., and M. Blaxter, 2013 Functional diversification of Argonautes in nematodes: an expanding universe. Biochem. Soc. Trans. 41: 881–886.

Celniker, S. E., L. A. L. Dillon, M. B. Gerstein, K. C. Gunsalus, S. Henikoff *et al.*, 2009 Unlocking the secrets of the genome. Nature 459: 927–930.

Cerutti, H., and J. A. Casas-Mollano, 2006 On the origin and functions of RNA-mediated silencing: from protists to man. Curr. Genet. 50: 81–99.

Chung, W.-J., K. Okamura, R. Martin, and E. C. Lai, 2008 Endogenous RNA interference provides a somatic defense against Drosophila transposons. Curr. Biol. 18: 795–802.

Conrad, T., and A. Akhtar, 2012 Dosage compensation in Drosophila melanogaster: epigenetic fine-tuning of chromosome-wide transcription. Nat. Rev. Genet. 13: 123–134.

Czech, B., C. D. Malone, R. Zhou, A. Stark, C. Schlingeheyde *et al.*, 2008 An endogenous small interfering RNA pathway in Drosophila. Nature 453: 798–802.

Drummond, A. J., M. A. Suchard, D. Xie, and A. Rambaut, 2012 Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29: 1969–1973.

Eulalio, A., E. Huntzinger, and E. Izaurralde, 2008 Getting to the root of miRNA-mediated gene silencing. Cell 132: 9–14.

Finn, R. D., J. Mistry, and J. Tate, P Coggill, A. Heger *et al.*, 2009 The Pfam protein families database. Nucleic Acids Res. 38: D211–D222.

Gell, S. L., and R. A. Reenan, 2013 Mutations to the piRNA pathway component aubergine enhance meiotic drive of segregation distorter in Drosophila melanogaster. Genetics 193: 771–784.

Gershenson, S., 1928 A new sex-ratio abnormality in Drosophila obscura. Genetics 13: 488–507.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson *et al.*, 2011 Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29: 644–652.

Haddrill, P. R., L. Loewe, and B. Charlesworth, 2010 Estimating the parameters of selection on nonsynonymous mutations in *Drosophila pseudoobscura* and *D. miranda*. Genetics 185: 1381–1396.

Hahn, M., 2009 Distinguishing among evolutionary models for the maintenance of gene duplicates. Heredity 100: 605–617.

Hain, D., B. R. Bettencourt, K. Okamura, T. Csorba, W. Meyer *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.

Hudson, R. R., 2002 Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics 18: 337–338.

Jaenike, J., 2001 Sex chromosome meiotic drive. Annu. Rev. Ecol. Syst. 32: 25–49.

Kaessmann, H., 2010 Origins, evolution, and phenotypic impact of new genes. Genome Res. 20: 1313–1326.

Katoh, K., K. Misawa, K. Kuma, and T. Miyata, 2002 MAFFT : a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30: 3059–3066.

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung *et al.*, 2012 Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28: 1647–1649.

Kofler, R. P., N. Orozco-terWengel, R. V. De Maio, V. Pandey, Nolte *et al.*, 2011 PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. PLoS One 6: e15925.

Kolaczkowski, B., D. N. Hupalo, and A. D. Kern, 2011 Recurrent adaptation in RNA interference genes across the Drosophila phylogeny. Mol. Biol. Evol. 28: 1033–1042.

Kotelnikov, R. N., M. S. Klenov, Y. M. Rozovsky, L. V. Olenina, and M. V. Kibanov, *et al.*, 2009 Peculiarities of piRNA-mediated post-transcriptional silencing of Stellate repeats in testes of Drosophila melanogaster. Nucleic Acids Res. 37: 3254–3263.

Langmead, B., C. Trapnell, M. Pop, and S. L. Salzberg, 2009 Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10: R25.

Leebonoi, W., S. Sukthaworn, S. Panyim, and A. Udomkit, 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.

Lemos, B., B. R. Bettencourt, C. D. Meiklejohn, and D. L. Hartl, 2005 Evolution of proteins and gene expression levels are coupled in Drosophila and are independently associated with mRNA abundance, protein length, and number of protein-protein interactions. Mol. Biol. Evol. 22: 1345–1354.

Lewis, S. H., H. Salmela, and D. J. Obbard, 2016 Duplication and diversification of Dipteran Argonaute genes, and the evolutionary divergence of Piwi and Aubergine. Genome Biol. Evol. 8: 507–518.

Li, H., and W. Stephan, 2006 Inferring the demographic history and rate of adaptive substitution in Drosophila. PLoS Genet. 2: 1580–1589.

Li, H., W. X. Li, and S. W. Ding, 2002 Induction and suppression of RNA silencing by an animal virus. Science 296: 1319–1321.

Librado, P., and J. Rozas, 2009 DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451–1452.

Longdon, B., and F. M. Jiggins, 2012 Vertically transmitted viral endosymbionts of insects: Do sigma viruses walk alone? Proc. Biol. Sci. 279: 3889–3898.

Longdon, B., J. D. Hadfield, J. P. Day, S. C. L. Smith, J. E. McGonigle *et al.*, 2015 The causes and consequences of changes in virulence following pathogen host shifts. PLoS Pathog. 11: e1004728.

Löytynoja, A., and N. Goldman, 2005 An algorithm for progressive multiple alignment of sequences with insertions. Proc. Natl. Acad. Sci. USA 102: 10557–10562.

Lu, H.-L., S. Tanguy, C. Rispe, J.-P. Gauthier, T. Walsh *et al.*, 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, M. J., and R. F. Ketting, 2013 PIWI-interacting RNAs: from generation to transgenerational epigenetics. Nat. Rev. Genet. 14: 523–534.

Lynch, M., and T. J. Crease, 1990 The analysis of population survey data on DNA sequence variation. Mol. Biol. Evol. 7: 377–394.

Marques, J. T., and R. W. Carthew, 2007 A call to arms: coevolution of animal viruses and host innate immune responses. Trends Genet. 23: 359–364.

Maynard Smith, J., and J. Haigh, 1974 The hitch-hiking effect of a favourable gene. Genet. Res. 23: 23–35.

McDonald, J. H., and M. Kreitman, 1991 Adaptive protein evolution at the Adh locus in Drosophila. Nature 351: 652–654.

McGaugh, S. E., C. S. S. Heil, B. Manzano-Winkler, L. Loewe, S. Goldstein *et al.*, 2012 Recombination modulates how selection affects linked sites in Drosophila. PLoS Biol. 10: e1001422.

Meisel, R. P., J. H. Malone, and A. G. Clark, 2012 Disentangling the relationship between sex-biased gene expression and X-linkage. Genome Res. 22: 1255–1265.

Meisel, R. P., J. G. Scott, and A. G. Clark, 2015 Transcriptome differences between alternative sex determining genotypes in the house fly, Musca domestica. Genome Biol. Evol. 7: 2051–2061.

Meister, G., 2013 Argonaute proteins: functional insights and emerging roles. Nat. Rev. Genet. 14: 447–459.

Menon, D. U., and V. H. Meller, 2012 A role for siRNA in X-chromosome dosage compensation in *Drosophila melanogaster*. Genetics 191: 1023–1028.

Miesen, P., E. Girardi, and R. P. van Rij, 2015 Distinct sets of PIWI proteins produce arbovirus and transposon-derived piRNAs in Aedes aegypti mosquito cells. Nucleic Acids Res. 43: 6545–6556.

Mikhaylova, L. M., K. Nguyen, and D. I. Nurminsky, 2008 Analysis of the *Drosophila melanogaster* testes transcriptome reveals coordinate regulation of paralogous genes. Genetics 179: 305–315.

Morozova, T. V., E. A. Tsybulko, and E. G. Pasyukova, 2009 Regularory elements of the copia retrotransposon determine different levels of expression in different organs of males and females of Drosophila melanogaster. Genetika 45: 169–177.

Mukherjee, K., H. Campos, and B. Kolaczkowski, 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, R., C. Bustamante, A. G. Clark, S. Glanowski, T. B. Sackton *et al.*, 2005a A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol. 3: e170.

Nielsen, R., S. Williamson, Y. Kim, M. J. Hubisz, A. G. Clark *et al.*, 2005b Genomic scans for selective sweeps using SNP data. Genome Res. 15: 1566–1575.

Obbard, D. J., F. M. Jiggins, D. L. Halligan, and T. J. Little, 2006 Natural selection drives extremely rapid evolution in antiviral RNAi genes. Curr. Biol. 16: 580–585.

Obbard, D. J., K. H. J. Gordon, A. H. Buck, and F. M. Jiggins, 2009a The evolution of RNAi as a defence against viruses and transposable elements. Philos. Trans. R. Soc. Lond. B Biol. Sci. 364: 99–115.

Obbard, D. J., J. J. Welch, K.-W. Kim, and F. M. Jiggins, 2009b Quantifying adaptive evolution in the Drosophila immune system. PLoS Genet. 5: e1000698.

Obbard, D. J., F. M. Jiggins, N. J. Bradshaw, and T. J. Little, 2011 Recent and recurrent selective sweeps of the antiviral RNAi gene Argonaute-2 in three species of Drosophila. Mol. Biol. Evol. 28: 1043–1056.

Obbard, D. J., J. MacLennan, K. W. Kim, A. Rambaut, P. M. O'Grady *et al.*, 2012 Estimating divergence dates and substitution rates in the Drosophila phylogeny. Mol. Biol. Evol. 29: 3459–3473.

Okamura, K., W.-J. Chung, J. G. Ruby, H. Guo, D. Bartel *et al.*, 2008 The Drosophila hairpin RNA pathway generates endogenous short interfering RNAs. Nature 453: 803–807.

Pal, C., B. Papp, and L. D. Hurst, 2001 Highly expressed genes in yeast evolve slowly. Genetics 158: 927–931.

Palmer W. H., and D. J. Obbard, 2016 Variation and evolution of the glutamine-rich repeat region of Drosophila Argonaute-2. G3 (Bethesda) 6: 2563–2572.

Palmer, W. J., and F. M. Jiggins, 2015 Comparative genomics reveals the origins and diversity of arthropod immune systems. Mol. Biol. Evol. 32: 2111–2129.

Pasyukova, E., S. Nuzhdin, W. Li, and A. J. Flavell, 1997 Germ line transposition of the copia retrotransposon in Drosophila melanogaster is restricted to males by tissue-specific control of copia RNA levels. Mol. Gen. Genet. 255: 115–124.

Peden, J., 1995 Analysis of codon usage bias. Ph.D. Thesis, University of Nottingham, Nottingham, United Kingdom.

Ross, R. J., M. M. Weiner, and H. Lin, 2014 PIWI proteins and PIWI-interacting RNAs in the soma. Nature 505: 353–359.

Rozhkov, N. V., A. A. Aravin, E. S. Zelentsova, N. G. Schostak, R. Sachidanandam *et al.*, 2010 Small RNA-based silencing strategies for transposons in the process of invading Drosophila species. RNA 16: 1634–1645.

Russo, C. A., N. Takezaki, and M. Nei, 1995 Molecular phylogeny and divergence times of Drosophilid species. Mol. Biol. Evol. 12: 391–404.

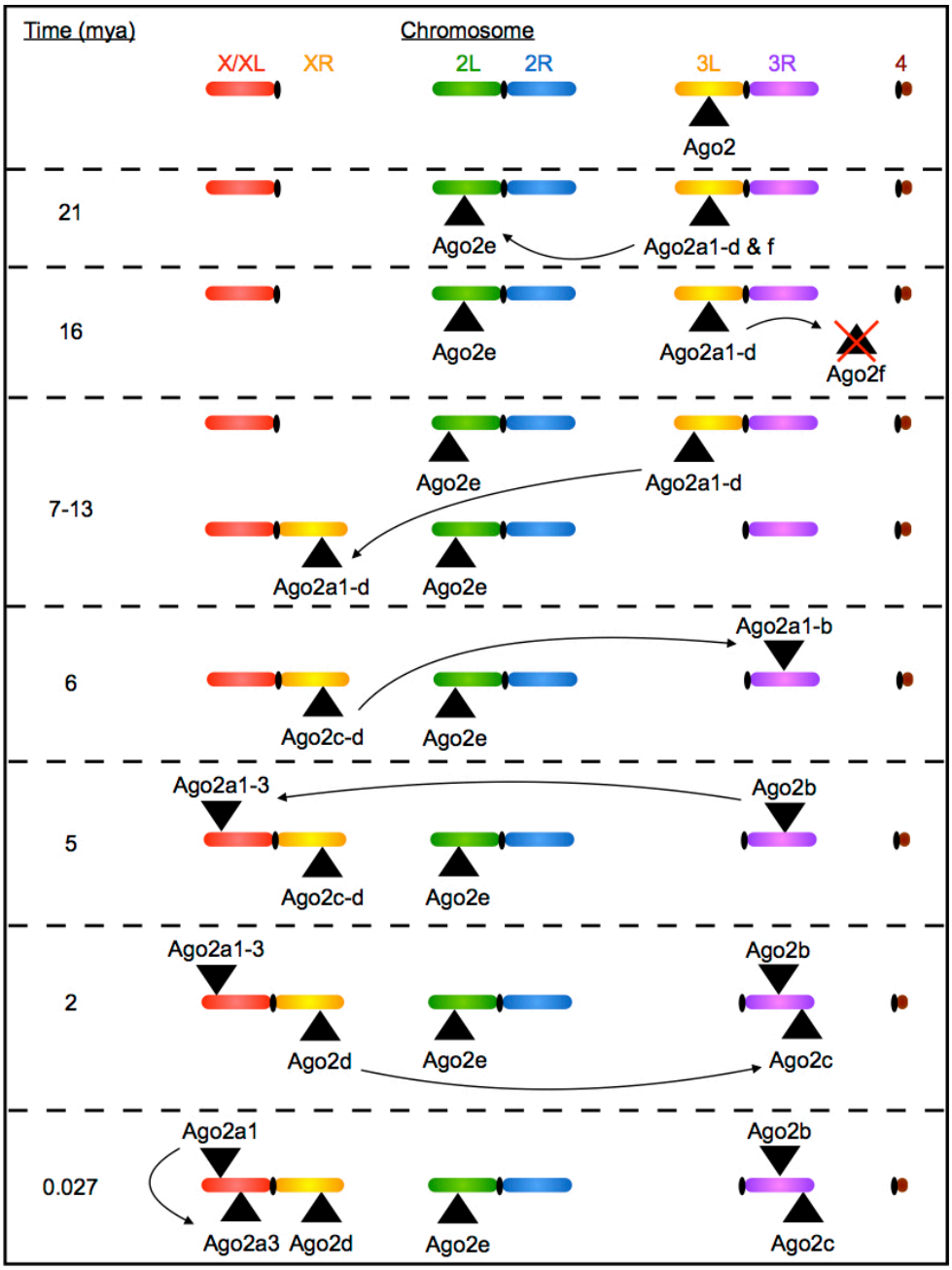
Schirle, N. T., and I. J. Macrae, 2012 The crystal structure of human Argonaute2. Science 336: 1037–1040.

Scott, J. G., W. C. Warren, L. W. Beukeboom, D. Bopp, A. G. Clark *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.

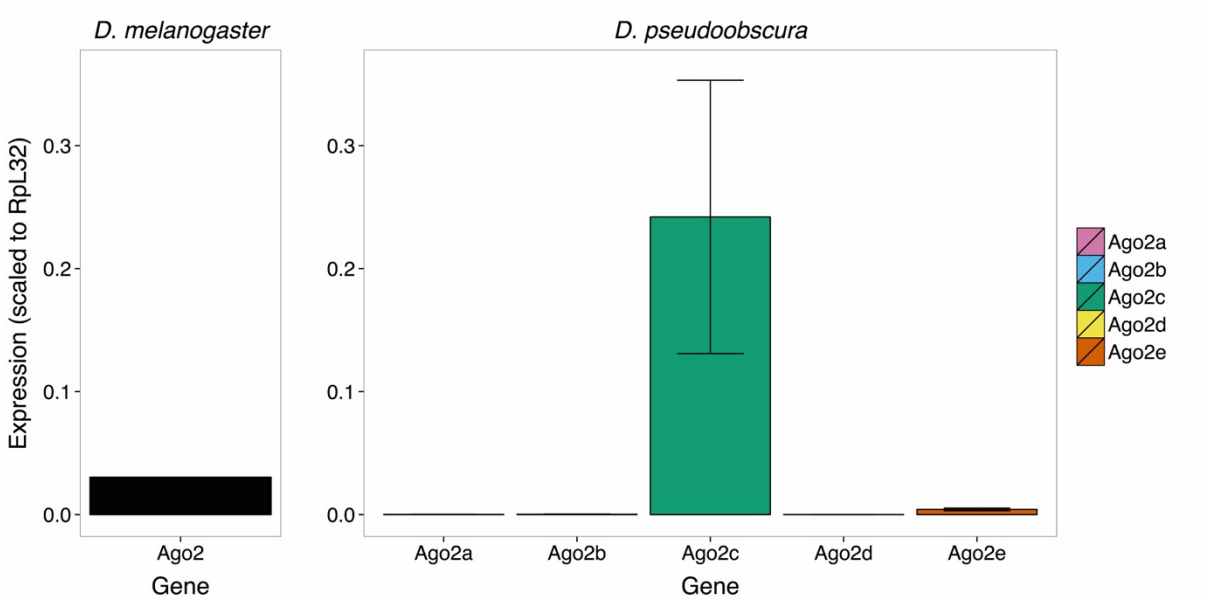
Sienski, G., D. Dönertas, and J. Brennecke, 2012 Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. Cell 151: 964–980.

Singh, R. K., K. Gase, I. T. Baldwin, and S. P. Pandey, 2015 Molecular evolution and diversification of the Argonaute family of proteins in plants. BMC Plant Biol. 15: 1–16.
Stephens, M., N. J. Smith, and P. Donnelly, 2001 A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. 68: 978–989.
Sturtevant, A. H., and T. Dobzhansky, 1936 Geographical distribution and cytology of "Sex Ratio" in Drosophila pseudoobscura and related species. Genetics 21: 473–490.
Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585–595.
Tamura, K., 2004 Temporal patterns of fruit fly (Drosophila) evolution revealed by mutation clocks. Mol. Biol. Evol. 21: 36–44.
Tao, Y., L. Araripe, S. B. Kingan, Y. Ke, H. Xiao *et al.*, 2007 A sex-ratio meiotic drive system in Drosophila simulans. II: An X-linked distorter. PLoS Biol. 5: 2576–2588.
Unckless, R. L., A. M. Larracuente, and A. G. Clark, 2015 Sex-ratio meiotic drive and Y-linked resistance in Drosophila affinis. Genetics 199: 831–840.
van Mierlo, J. T., G. J. Overheul, B. Obadia, K. W. R. van Cleef, C. L. Webster *et al.*, 2014 Novel Drosophila viruses encode host-specific suppressors of RNAi. PLoS Pathog. 10: e1004256.
van Rij, R. P., M.-C. Saleh, B. Berry, C. Foo, A. Houk *et al.*, 2006 The RNA silencing endonuclease Argonaute 2 mediates specific antiviral immunity in Drosophila melanogaster. Genes Dev. 20: 2985–2995.
Vicoso, B., and B. Charlesworth, 2006 Evolution on the X chromosome: unusual patterns and processes. Nat. Rev. Genet. 7: 645–653.
Webster, C. L., B. Longdon, S. H. Lewis, and D. J. Obbard, 2016 Twenty five new viruses associated with the Drosophilidae (Diptera). Evol. Bioinform. Online. 12(Suppl 2): 13–25.
Wen, J., H. Duan, F. Bejarano, K. Okamura, L. Fabian *et al.*, 2015 Adaptive regulation of testis gene expression and control of male fertility by the Drosophila Harpin RNA pathway. Mol. Cell 57: 165–178.
Wright, F., 1990 The "effective number of codons" used in a gene. Gene 87: 23–29.
Wright, S. I., and B. Charlesworth, 2004 The HKA test revisited: a maximum-likelihood-ratio test of the standard neutral model. Genetics 168: 1071–1076.
Wu, C. I., and A. T. Beckenbach, 1983 Evidence for extensive genetic differentiation between the sex-ratio and the standard arrangement of Drosophila pseudobscura and D. persimilis and identification of hybrid sterility factors. Genetics 105: 71–86.
Yang, Z., 1997 PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. 13: 555–556.

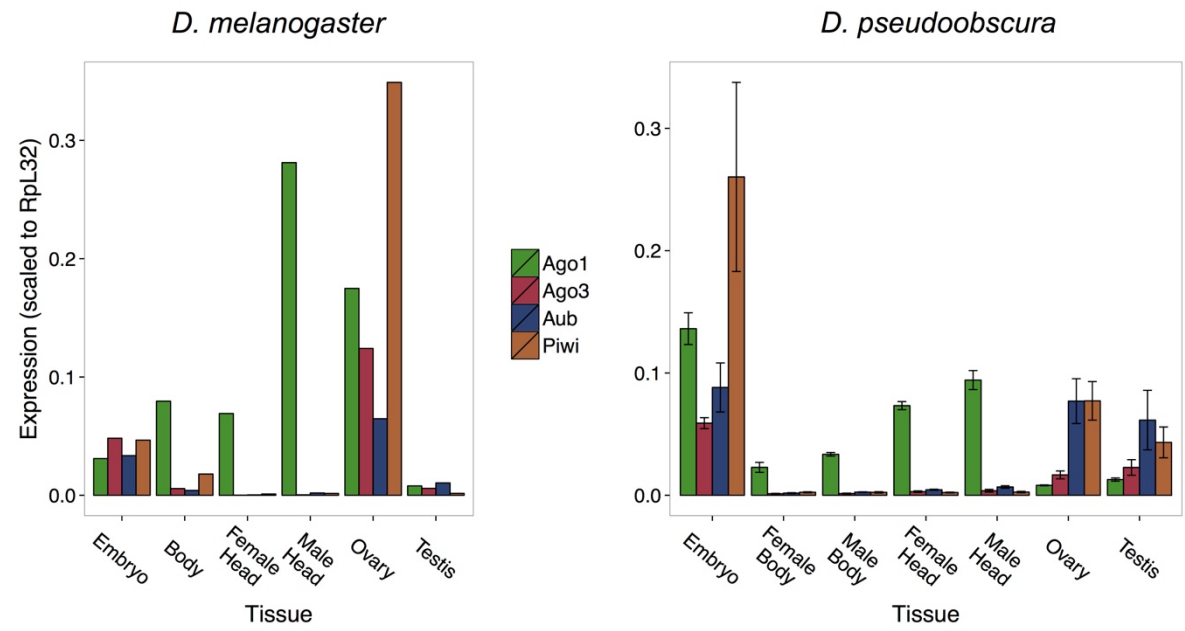
Communicating editor: D. A. Barbash



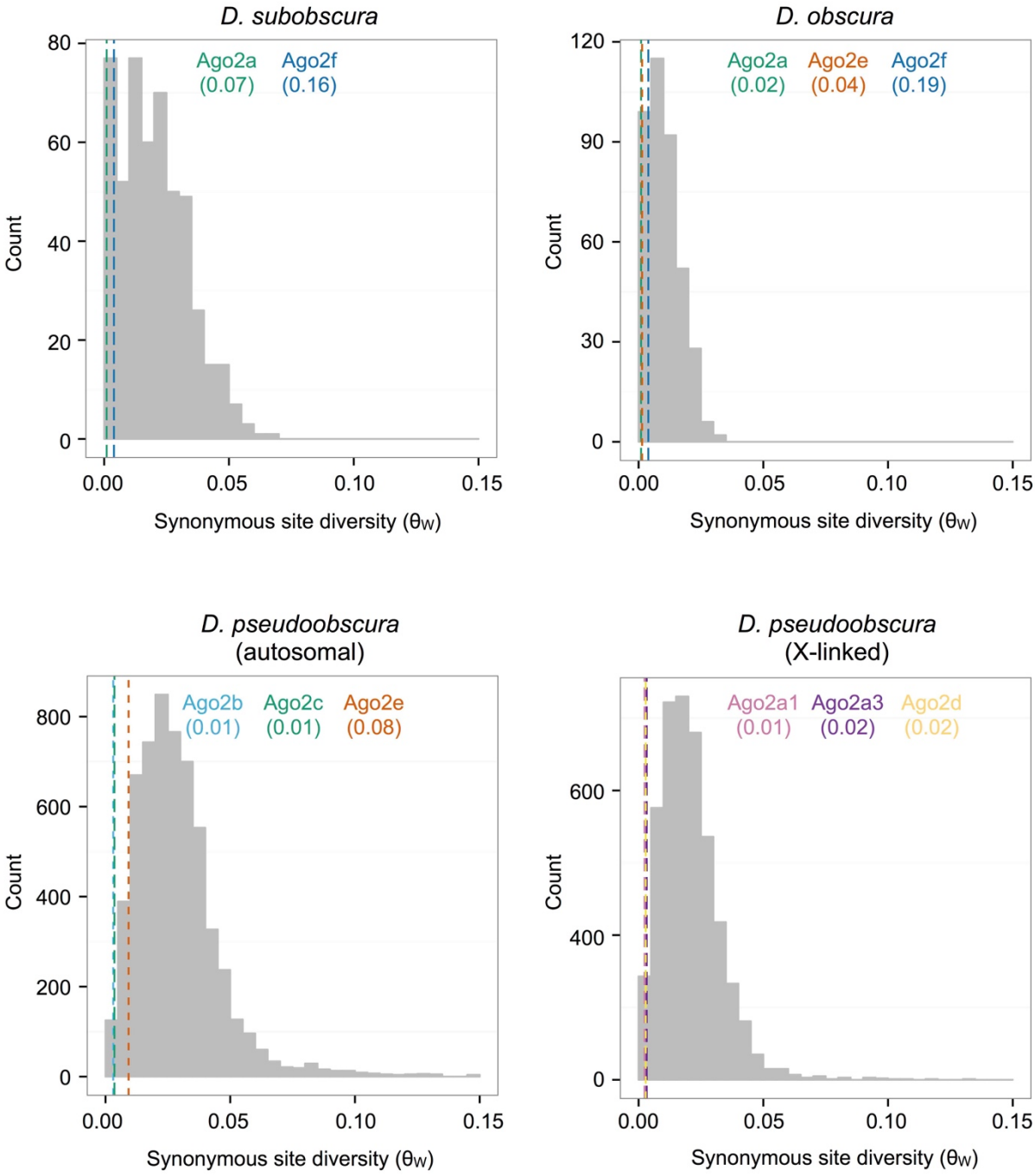
Supplementary Figure S1: The course of duplications and translocations of Ago2 paralogues in *D. pseudoobscura*.
A complex series of duplications and translocations has produced six Ago2 paralogues in *D. pseudoobscura*, located on four different chromosome arms. Chromosome arms correspond to Muller Elements A (X/XL), B (2L), C (2R), D (3L), E (3R) & F (4) (see Schaeffer et al. 2008, Fig. 1). Firstly, the Ago2a1-f ancestor duplicated ~21mya to form Ago2a1-d&f and Ago2e, the latter of which moved onto chromosome 2L. Next, the Ago2a1-d&f ancestor duplicated ~16mya, forming Ago2a1-d and Ago2f, with Ago2f subsequently being lost in the *D. pseudoobscura* lineage. Then, the 3L arm fused with the X chromosome, moving Ago2a1-d onto the X: this happened 7-15mya, after the divergence of the obscura group into Palearctic (e.g. *D. subobscura*) and Nearctic (e.g. *D. pseudoobscura*) clades (see Segarra and Aguadé 1992). Ago2a1-d then duplicated ~6mya, forming Ago2c-d and Ago2a1-b, the latter of which moved onto chromosome 2. After this, Ago2a1-b duplicated ~5mya, producing Ago2b and Ago2a1-3, the latter of which moved onto the left arm of the X chromosome. This was followed by a duplication of Ago2c-d ~2mya, forming Ago2d and Ago2c, the latter of which moved onto chromosome 2. Finally, Ago2a1-3 duplicated ~27kya, producing Ago2a1 and Ago2a3 in tandem. Note that due to differences in evolutionary rate between branches, the timings of these events should be treated with caution.



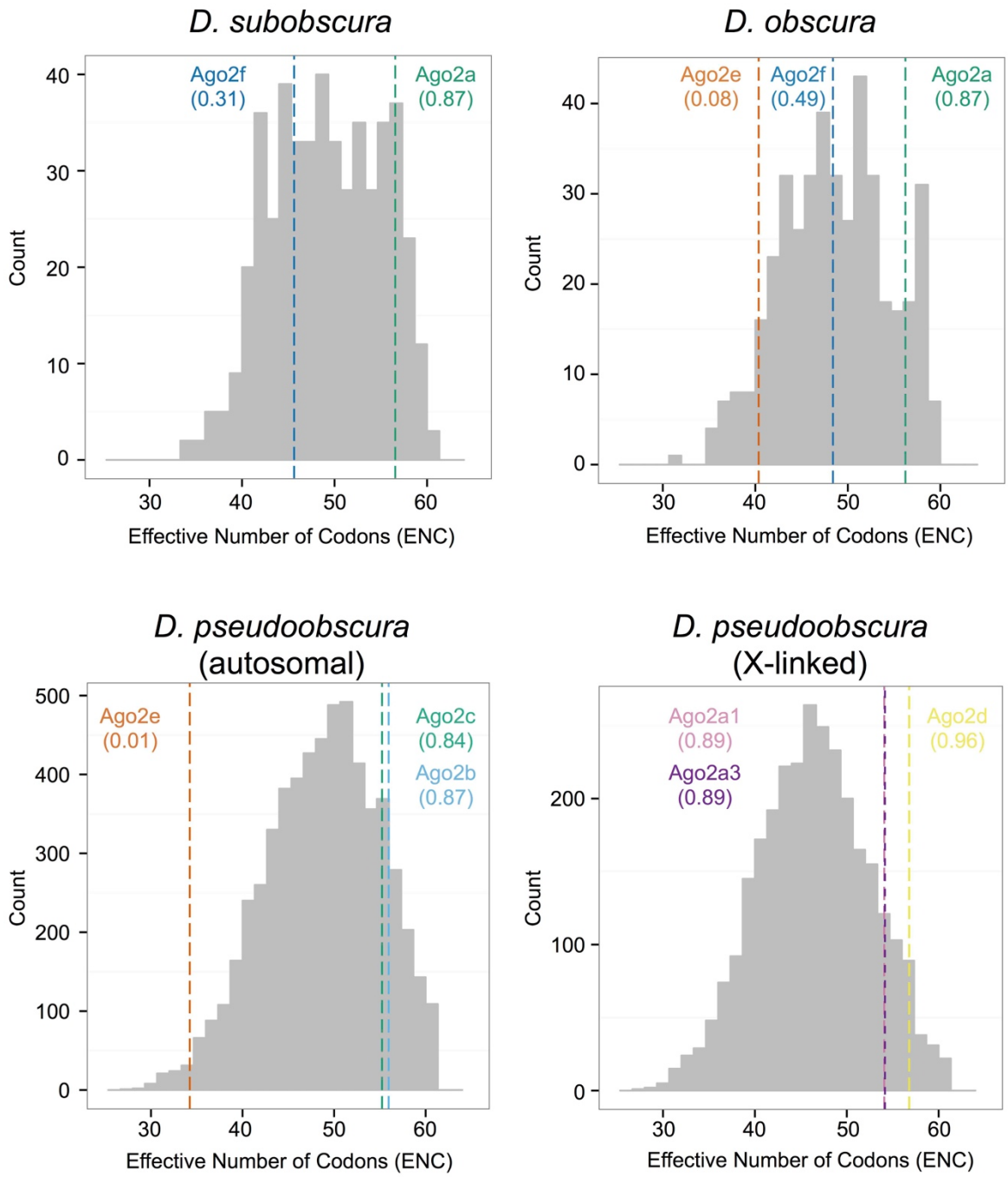
Supplementary Figure S2: The expression of *D. pseudoobscura* Ago2 paralogues in embryos. Error bars indicate 1 standard error estimated from 2 technical replicates in each of two different genetic backgrounds. *D. melanogaster* expression levels were taken from a single publicly-available RNA-seq experiment (Brown et al. 2014). Ago2c is highly expressed in embryos, but none of the testis-specific Ago2 paralogues (Ago2a, Ago2b & Ago2e) are expressed.



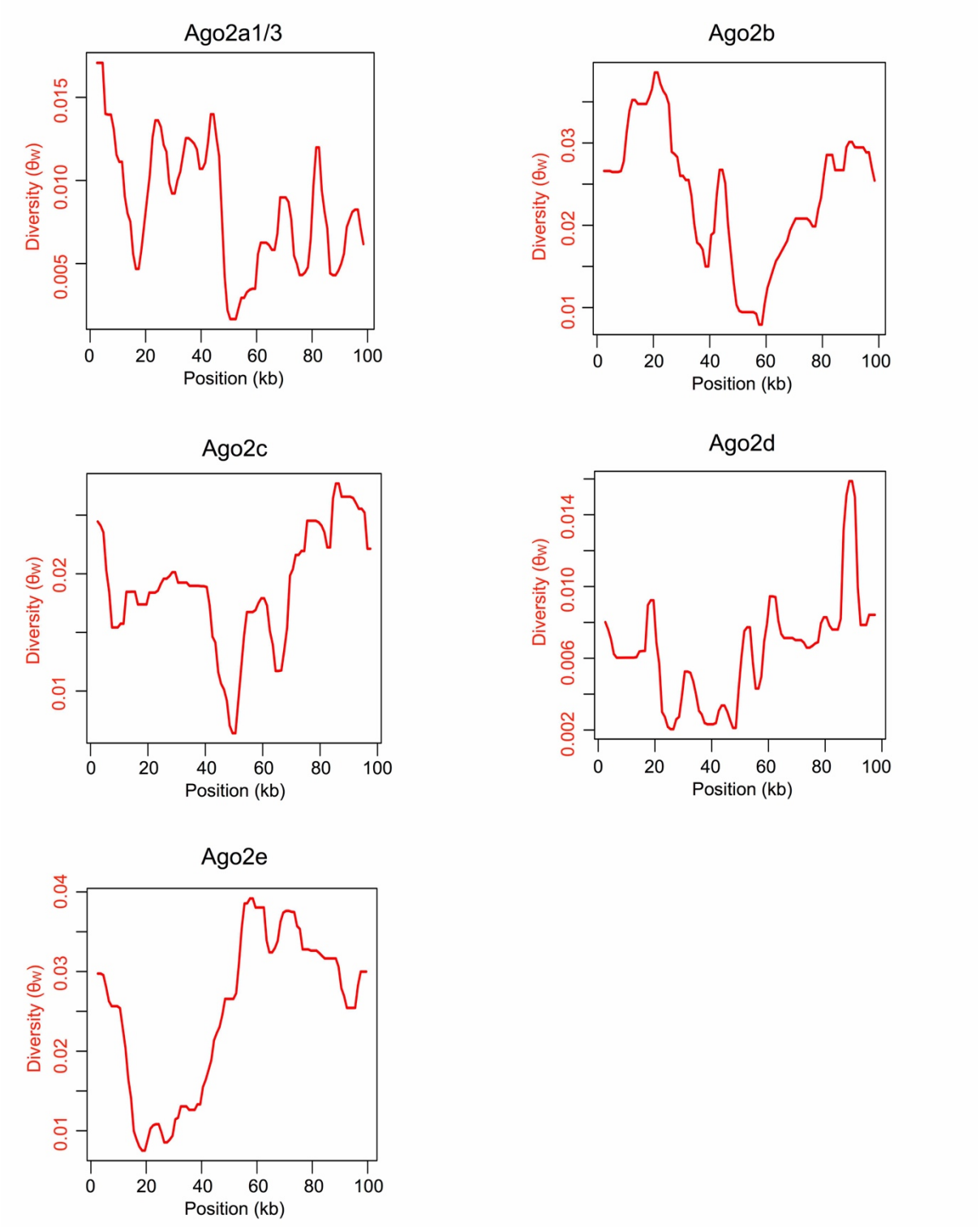
Supplementary Figure S3: The tissue-specific expression patterns of other members of the Argonaute gene family (Ago1, Ago3, Aub & Piwi) in *D. melanogaster* and *D. pseudoobscura*. For *D. pseudoobscura* embryo, error bars indicate 1 standard error estimated from 2 technical replicates in each of two different genetic backgrounds. For all other *D. pseudoobscura* tissues, error bars indicate 1 standard error estimated from 2 technical replicates in each of five different genetic backgrounds. *D. melanogaster* expression levels were taken from a single RNA-seq experiment (Brown et al. 2014). In *D. pseudoobscura*, Ago1 is expressed in all tissues, but the other genes are only expressed in the embryo and germline.



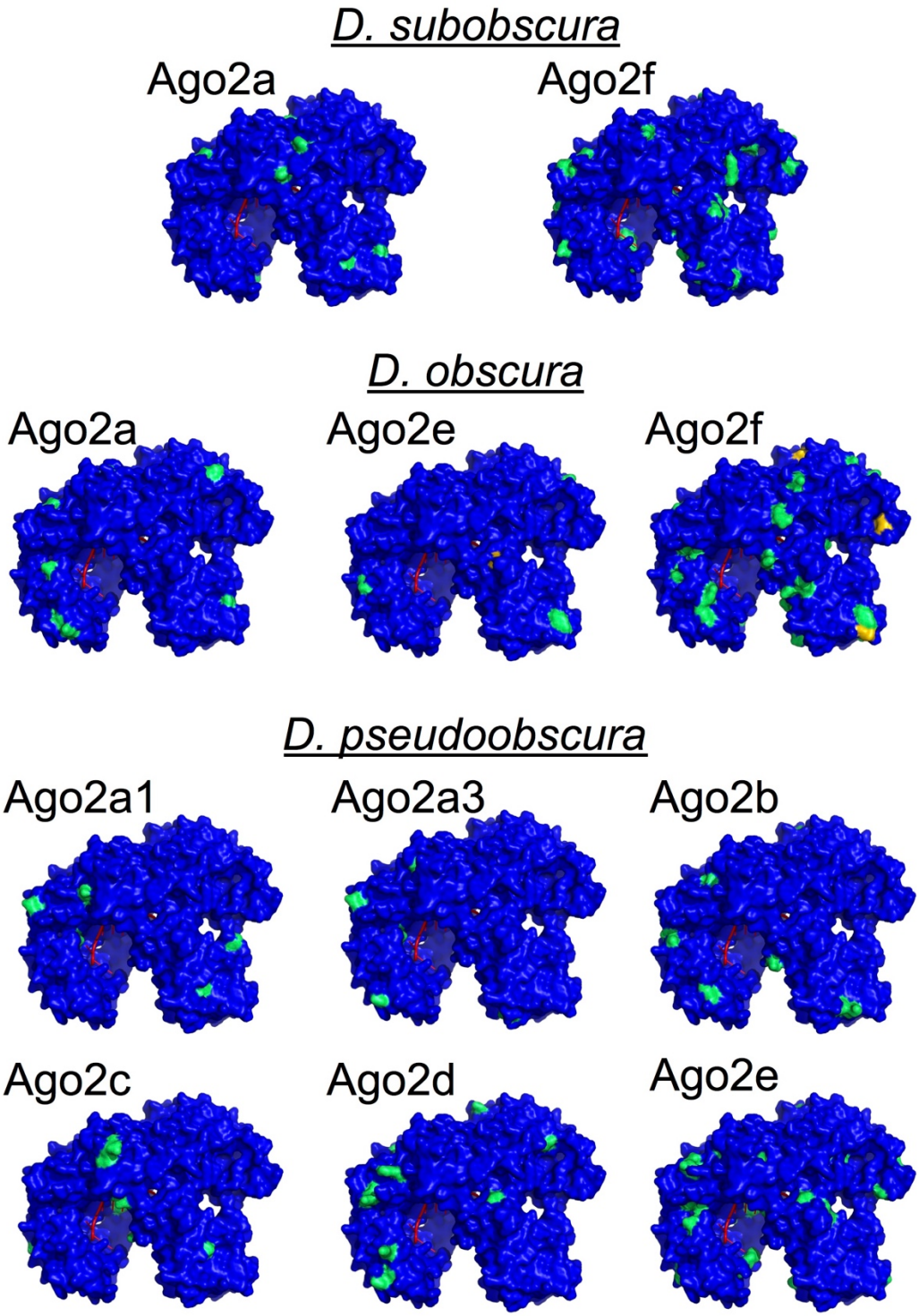
Supplementary Figure S4: The distribution of synonymous site diversity across genes, derived from genome (*D. pseudoobscura*) or transcriptome (*D. subobscura* & *D. obscura*) data. The percentile of the distribution into which each paralogue falls is indicated in brackets under the paralogue name. In each species, members of the Ago2a and Ago2e subclades have very low diversity compared with the genome as a whole.



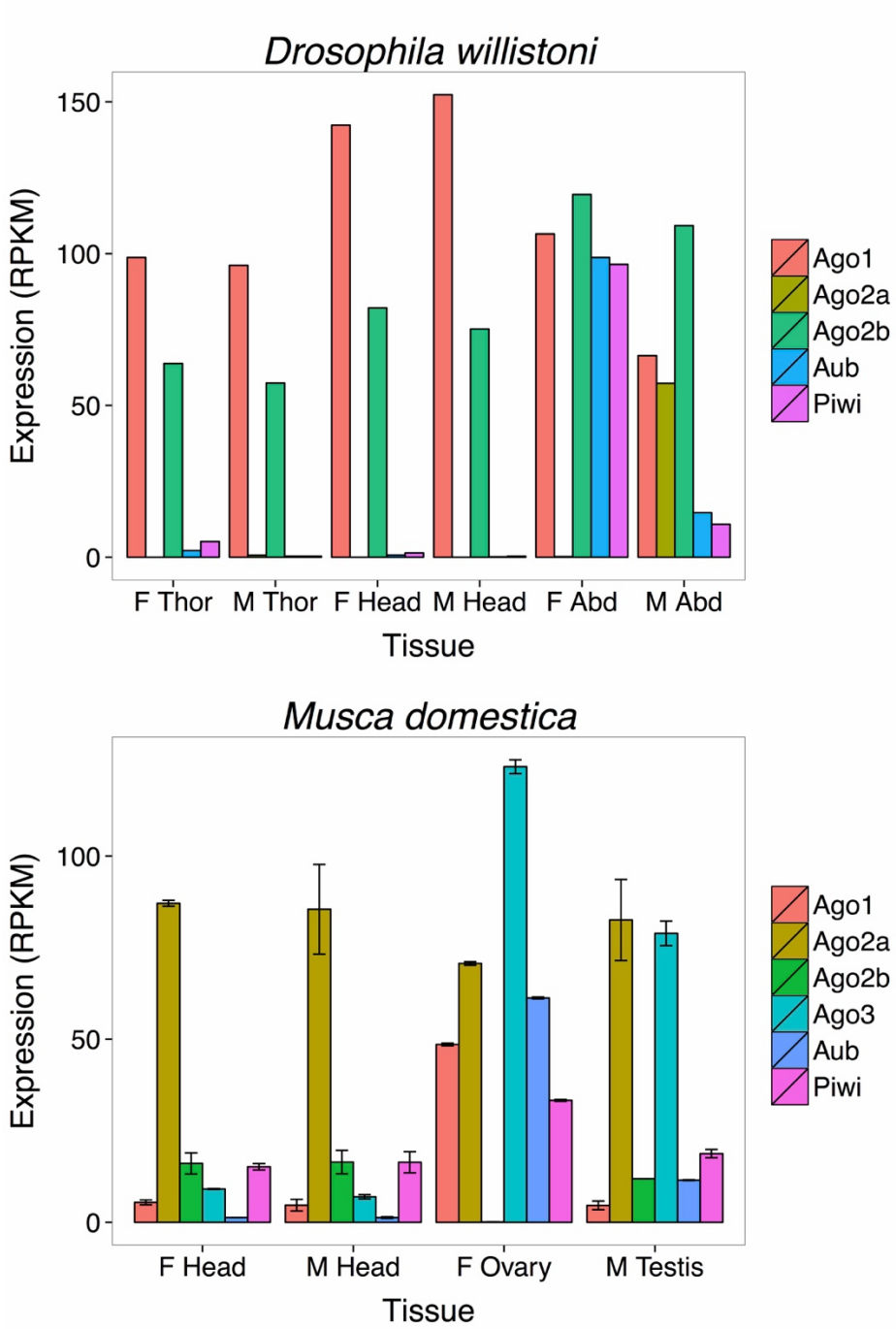
Supplementary Figure S5: The distribution of codon usage bias, derived from genome (*D. pseudoobscura*) or transcriptome (*D. subobscura* & *D. obscura*) data. The percentile of the distribution into which each paralogue falls is indicated in brackets under the paralogue name. Ago2e has a very low effective number of codons (ENC) compared with the genome as a whole, indicating a high degree of codon usage bias.



Supplementary Figure S6: Genetic diversity in the regions surrounding each *D. pseudoobscura* Ago2 paralogue, with Ago2 paralogue haplotype sequences removed. After specifying Ago2 paralogue sequence data as missing information, sharp troughs in diversity remain at Ago2a, Ago2b and Ago2c, indicating a selective sweep.



Supplementary Figure S7: The distribution of variable sites across the protein structures of Ago2 paralogues in *D. subobscura*, *D. obscura* and *D. pseudoobscura*. For each paralogue, residues with one variable site are coloured green, residues with two variable sites are coloured yellow, and a small RNA guide molecule is depicted in red. Variable sites are distributed across the protein structure of each paralogue, suggesting that they are not evolving through divergence of one part of the structure; however, the small number of variable sites for most paralogues mean that these conclusions are necessarily tentative.



Supplementary Figure S8: The tissue-specific expression patterns of the Argonaute gene family in *D. willistoni* and *M. domestica*. Transcriptome data for *D. willistoni* were taken from (Meisel et al. 2012), and transcriptome data for *M. domestica* were taken from (Meisel et al. 2015). For both species, we mapped reads to coding sequences using Bowtie 2.1 (Langmead et al. 2009), counted reads mapping to each coding sequence using HTSeq (Anders et al. 2015), and converted counts to reads per kilobase per million reads (RPKM (Mortazavi et al. 2008)) to account for coding sequence length and sequencing depth. For *M. domestica*, error bars indicate two biological replicates, each in a different genetic background.

Supplementary Table S1: McDonald-Kreitman test results. Pn & Ps are the number of within-species polymorphisms after singletons have been removed. All values are displayed to 2dp, except ω_α which is displayed to 4dp.

Paralogue	Pn	Ps	Outgroup	Dn	Ds	α	ω(α)	p value
D. subobscura Ago2a	2	9	*D. subobscura* / *D. obscura* Ago2a ancestor	26	72	0.39	0.0016	0.73
D. subobscura Ago2f	5	13	*D. subobscura* / *D. obscura* Ago2f ancestor	129	150	0.55	0.0011	0.15
D. obscura Ago2a	3	5	*D. subsilvestris* Ago2a	39	65	0.00	0.00	1.00
D. obscura Ago2f	15	12	*D. subsilvestris* Ago2f	92	106	-0.44	-0.0013	0.42
D. obscura Ago2e	1	5	*D. subsilvestris* Ago2e	87	145	0.67	0.0015	0.42
D. pseudoobscura Ago2a1	5	1	*D. pseudoobscura* Ago2a1-d ancestor	74	66	-3.46	-0.0157	0.22
D. pseudoobscura Ago2a3	5	4	*D. pseudoobscura* Ago2a1-d ancestor	72	71	-0.23	-0.0010	1.00
D. pseudoobscura Ago2b	5	1	*D. pseudoobscura* Ago2a1-d ancestor	78	54	-2.46	-0.0136	0.40
D. pseudoobscura Ago2c	8	8	*D. pseudoobscura* Ago2a1-d ancestor	47	78	-0.66	-0.0025	0.42
D. pseudoobscura Ago2d	5	3	*D. pseudoobscura* Ago2a1-d ancestor	85	72	-0.41	-0.0017	0.73
D. pseudoobscura Ago2e	0	17	*D. pseudoobscura* Ago2e – *D. affinis* Ago2e ancestor	77	120	1.00	0.0027	0.00

Supplementary Table S2: McDonald-Kreitman test results with alternative outgroups. Pn & Ps are the number of within-species polymorphisms after singletons have been removed. All values are displayed to 2dp, except ω_α which is displayed to 4dp.

Paralogue	Pn	Ps	Outgroup	Dn	Ds	α	ω(α)	p value
D. subobscura Ago2a	2	9	*D. subsilvestris* Ago2a	58	120	0.54	0.0014	0.51
D. subobscura Ago2f	5	11	*D. subsilvestris* Ago2f	165	190	0.48	0.0008	0.31
D. obscura Ago2a	3	5	*D. tristis* Ago2a	20	33	0.01	0.00	1.00
D. obscura Ago2f	20	15	*D. tristis* Ago2f	19	23	-0.22	-0.0008	0.71
D. obscura Ago2e	1	5	*D. tristis* Ago2e	90	82	0.76	0.0105	0.38
D. pseudoobscura Ago2a1	5	1	*D. pseudoobscura* Ago2b	113	80	-2.54	-0.0095	0.40
D. pseudoobscura Ago2a3	5	4	*D. pseudoobscura* Ago2b	110	82	0.07	-0.0002	1.00
D. pseudoobscura Ago2b	3	0	*D. pseudoobscura* Ago2a1	100	73	N/A	N/A	N/A
D. pseudoobscura Ago2c	8	8	*D. pseudoobscura* Ago2d	60	45	0.25	0.0016	0.60
D. pseudoobscura Ago2d	5	3	*D. pseudoobscura* Ago2c	62	50	-0.34	-0.0021	0.73
D. pseudoobscura Ago2e	0	17	*D. lowei* Ago2e	66	133	1.00	0.0024	0.00

Supplementary Table S3: Genetic backgrounds used in each experiment. Line refers to an individual isofemale line, and Origin refers to the geographic location where the female who founded that line was caught.

			Use		
Species	**Line**	**Origin**	**DCV challenge**	**Tissue expression**	**Sequencing**
D. subobscura	DPG1	Blackford Hill, Scotland	Y	Y	Male
D. subobscura	DPG2	Blackford Hill, Scotland	Y	Y	Male
D. subobscura	DPG3	Blackford Hill, Scotland			Male
D. subobscura	DPG4	Blackford Hill, Scotland		Y	Male
D. subobscura	DPG5	Blackford Hill, Scotland			Male
D. subobscura	DPG6	Blackford Hill, Scotland		Y	Male
D. subobscura	DPG7	Blackford Hill, Scotland			Female
D. subobscura	DPG8	Blackford Hill, Scotland			Female
D. subobscura	DPG9	Blackford Hill, Scotland			Female
D. subobscura	DPG10	Blackford Hill, Scotland			Female
D. subobscura	DPG11	Blackford Hill, Scotland			Female
D. subobscura	DPG12	Blackford Hill, Scotland	Y	Y	Female
D. obscura	DPG1	Blackford Hill, Scotland	Y	Y	Male
D. obscura	DPG2	Blackford Hill, Scotland		Y	Male
D. obscura	DPG3	Blackford Hill, Scotland	Y	Y	Male
D. obscura	DPG4	Blackford Hill, Scotland			Male
D. obscura	DPG5	Blackford Hill, Scotland			Male
D. obscura	DPG6	Blackford Hill, Scotland			Male
D. obscura	DPG7	Blackford Hill, Scotland			Female
D. obscura	DPG8	Blackford Hill, Scotland			Female
D. obscura	DPG9	Blackford Hill, Scotland			Female
D. obscura	DPG10	Blackford Hill, Scotland			Female
D. obscura	DPG11	Blackford Hill, Scotland			Female
D. obscura	DPG12	Blackford Hill, Scotland			Female
D. obscura	DA45	Moor Lane, England	Y	Y	
D. obscura	DA46	Moor Lane, England		Y	
D. pseudoobscura	MV2	Mesa Verde, CO, USA	Y	Y	Male
D. pseudoobscura	MV8	Mesa Verde, CO, USA	Y	Y (embryo)	Male
D. pseudoobscura	MV10	Mesa Verde, CO, USA	Y	Y (embryo)	Male
D. pseudoobscura	MV11	Mesa Verde, CO, USA		Y	Male
D. pseudoobscura	MV13	Mesa Verde, CO, USA			Male
D. pseudoobscura	MV19	Mesa Verde, CO, USA			Male
D. pseudoobscura	MV15	Mesa Verde, CO, USA		Y	Female
D. pseudoobscura	MV21	Mesa Verde, CO, USA			Female
D. pseudoobscura	MV25	Mesa Verde, CO, USA		Y	Female
D. pseudoobscura	MV26	Mesa Verde, CO, USA			Female
D. pseudoobscura	MV28	Mesa Verde, CO, USA			Female
D. pseudoobscura	MV32	Mesa Verde, CO, USA		Y	Female
D. melanogaster	FR32	Montpellier, France	Y		
D. melanogaster	FR35	Montpellier, France	Y		
D. melanogaster	FR39	Montpellier, France	Y		

Supplementary Table S4: Primers used for PCR and qPCR amplification of Ago2 paralogues.

D. subobscura Ago2 paralogue PCR primers

Paralogue	Name	Sequence
Ago2a	Dsubob_Ago2_1229_5_F	CCAAGAAGTGAAAGTAACAGATCG
Ago2a	Dsubob_Ago2_1229_M_F	CCACTGAATCGCAAGGATTCTGC
Ago2a	Dsubob_Ago2_1229_M_R	GGCGAATACCAAAGCGACTGATGG
Ago2a	Dsubob_Ago2_1229_3_R	CTTTGGGGAACGGAACTTGGTGAC
Ago2f	Dsubob_Ago2_21203_5_F	CACGCCTTTGAGGTGTACAGAAAGC
Ago2f	Dsubob_Ago2_21203_3_R	CACCAAATGTGCCAGATAGACCG

D. subobscura Ago2 paralogue qPCR primers

Name	Sequence
Dsubob_Ago2a_q_F_2_1	CCAACGAGAGGAAGGCCAAGATTATAC
Dsubob_Ago2a_q_R_2	CCAGGCGAATACCAAAGCGACT
Dsubob_Ago2f_q_F_3	GATTTCAAGCGGCTCCAATGTG
Dsubob_Ago2f_q_R_3_1	GTTTGCGCTGCACCGTAAACAG
Obs_group_RpL32_q_F	CTTAGTTGTCGCACAAATGG
Obs_group_RpL32_q_R	TGCGCTTGTTGGAACCGTAAC

D. obscura Ago2 paralogue PCR primers

Paralogue	Name	Sequence
Ago2a	Dobs_Ago2_2809_5_F	GGACAAGTATCTGTCAATTATCTCGACG
Ago2a	Dobs_Ago2_2809_18680_3_R	CTTGGGGAGAACGGAACTTGG
Ago2f	Dobs_Ago2_18680_5_F	CCTTTGAGCTGTTCAGAGTGGAAC
Ago2f	Dobs_Ago2_18680_M_F_2	GTAAATTGAGCCCCCAGTGTGTGTTGA
Ago2f	Dobs_Ago2_18680_M_R	CCAGCTGAGTGCGCGGGTTATC
Ago2f	Dobs_Ago2_all_3_R	TGGCGCCAGTCAGATAGACACG
Ago2e	Dobs_Ago2_24803_5_F	CGMGGTACACTGGGCAGAATCG
Ago2e	Dobs_947_Ago2_24803_F	GGTGAATCGCAAGGACTCCACGCT
Ago2e	Dobs_Ago2_24803_M_R	CAGTTCGGCTTTCTGTTTCAGTTC
Ago2e	Dobs_2269_Ago2_24803_R	GGCTCCACGTTGTTGTATTTGTTGTG

D. obscura Ago2 paralogue qPCR primers

Name	**Sequence**
Dobs_Ago2a_q_F_2	GCTTTCCAAGGTTTCAGCAAGCTC
Dobs_Ago2a_q_R_2_1	CCAACATGCAAGCATAGAAGGT
Dobs_Ago2f_q_F_3_1	GCACTCCGTCCACGTACG
Dobs_Ago2f_q_R_3	CTCATTCGGATGGACAATGATCCT
Dobs_Ago2e_q_F_4	CAACTACAACAAGATGCGGGACCTTG
Dobs_Ago2e_q_R_4	GAAGTGCGGATCCAGGCTCT
Obs_group_RpL32_q_F	CTTAGTTGTCGCACAAATGG
Obs_group_RpL32_q_R	TGCGCTTGTTGGAACCGTAAC

D. pseudoobscura Ago2 paralogue PCR primers

Paralogue	**Name**	**Sequence**
Ago2a1&3	Dper.mir.A_F	TGGAGGTTGTGTTGGCAGTA
Ago2a1&3	DpseAgo2A_R	CTANACGAARTACATAGGRTTCGTCTTC
Ago2a3	Dpse_GA22965_3_out_F_2	CCAAGAGGACGAAAACACTGATTGG
Ago2a3	DpseAgo2A_R	CTANACGAARTACATAGGRTTCGTCTTC
Ago2b	Dper.mir.D_F	CAGTACGATGTGAAGATCACGTCAGTAT
Ago2b	Dpse_Ago2_UnivR	GCCAGTRAGRTAGACACGTCC
Ago2c	Dpse_Ago2c_5_F	GGCCGTACCCTGACTTACACTGTGGAAC
Ago2c	Dpse_Ago2c_M_F	CTTGGAAAACGACTTCATTGTGGTGC
Ago2c	Dpse_Ago2c_M_R	CTGTAATTCCGTATATCGGCCTTCTG
Ago2c	Dpse_Ago2c_3_R	CACGAAATACATGGGGTTCGTTTTCAT
Ago2d	DpseAgo2B_MF	ATGCCAGCTGTGGCCTACCA
Ago2d	Dpse_Ago2d_M_F	CTGGATGGGAAGCAAACGACGG
Ago2d	qrtD_R	GAAGTCAGTGCCCAGGCGT
Ago2d	Dpse_Ago2d_3_R	GGAACTCTGGAACAATCAACCGCTTTT
Ago2e	pse_Ago2E_5_F	CGAGGTGGCTGTGAACTACCTGCAG
Ago2e	pse_Ago2E_3_R	CATGGGGTTCCTGCTGGACAGG

D. pseudoobscura qPCR primers

Gene	Name	Sequence
Ago1	Obsgroup_Ago1_q_F_2	GTGAAGTTCACCAAGGAGATCAAGG
Ago1	Obsgroup_Ago1_q_R_2	GGTTACATTGCAGACACGATACTTGC
Ago2a	Dpse_Ago2a_q_F_3	ATGGTTATTCAGAAGAGTCGCAAAG
Ago2a	Dpse_Ago2a_q_R_3	CTAGTTCACGTTCATCCTTGTAGTACAG
Ago2b	Dpse_Ago2b_q_F	GGGAAAGGAAAATAAATATAAACCGAA
Ago2b	Dpse_Ago2b_q_R	CGCACCTGTAGCTTTTAGTTGA
Ago2c	Dpse_Ago2c_q_F	AAGGAGGCGGACAACAGAG
Ago2c	Dpse_Ago2c_q_R	TGTGCTTGCTGACCCTGAG
Ago2d	Dpse_Ago2d_q_F	TCAGATTGAGTACAAAAACAAGTTG
Ago2d	Dpse_Ago2d_q_R	CCCTGAAAATCGACCACTCTTA
Ago2e	qrtE_F_3	GAACTACAACAAGATGCGGGACTTCG
Ago2e	qrtE_R_2	GCTTGGGTCCAGGCTCTTGG
Ago3	Dpse_Ago3_q_F_1	CGAAAGCAGTTCGATCCTTCATGTCC
Ago3	Dpse_Ago3_q_R_1	CGTCACAGCAGAGCATTAATCCTCC
Aub	Dpse_Aub_q_F_1	GCATTCAACAAGCGCTTGCAATC
Aub	Dpse_Aub_q_R_1	ACGCGAGCTGGGATGTCCAC
Piwi	Dpse_Piwi_q_F_1	GCGTATGGGCATATTGTCAAATCACG
Piwi	Dpse_Piwi_q_R_1	GGCCACACAGCACAATTGAATC
RpL32	Fly rp49 qPCR F-a	TGCCAAGTTGTCGCACAAATGG
RpL32	Fly rp49 qPCR R-i	TACGCTTGTTGGAGCCGTAAC

Supplementary Table S5: Primers used for Sanger sequencing of Ago2 paralogues.

D. subobscura Ago2a sequencing primers

Name	Sequence
Dsubob_Ago2_1229_5_F	CCAAGAAGTGAAAGTAACAGATCG
Dsubob_s_800_Ago2a_F	CTAGACTTCAGGCGTAACGATATCG
Dsubob_Ago2_1229_M_F	CCACTGAATCGCAAGGATTCTGC
Dsubob_s_1635_Ago2a_F	GCAATATGGCATTCTCACACAATG
Dsubob_s_2085_Ago2a_F	CTGCTGCAAGATGCACATTAAGC
Dsubob_Ago2_1229_3_R	CTTTGGGGAACGGAACTTGGTGAC
Dsubob_s_1965_Ago2a_R	GTGCTCAAGGGTTATTGACTCCATGTC
Dsubob_s_1420_Ago2a_R	GTACTGCATCTTACCGTACAGTATGG
Dsubob_Ago2_1229_M_R	GGCGAATACCAAAGCGACTGATGG
Dsubob_s_710_Ago2a_R	CCACATTGACAAATGGACGATCACC

D. subobscura Ago2f sequencing primers

Name	Sequence
Dsubob_Ago2_21203_5_F	CACGCCTTTGAGGTGTACAGAAAGC
Dsubob_s_440_Ago2f_F	GCTGGTCGCTCCTTCTTCAAGC
Dsubob_s_1480_Ago2f_F	GCTGCAGCACGGCATACTGAC
Dsubob_s_1935_Ago2f_F	CGTGTTGCAAGAAGCACATTCG
Dsubob_Ago2_21203_3_R	CACCAAATGTGCCAGATAGACCG
Dsubob_s_1820_Ago2f_R	GCAAGTGCTCCAGCGTGATGG
Dsubob_s_1500_Ago2f_R	GTCAGTATGCCGTGCTGCAG
Dsubob_s_624_Ago2f_R	CGTGATATCGCTCAATGTACTCCAC

D. obscura Ago2a sequencing primers

Name	Sequence
Dobs_Ago2_2809_5_F	GGACAAGTATCTGTCAATTATCTCGACG
Dobs_s_720_Ago2_2809_18680_F	AATCTTGGCGACGGCTACGAAGCTC
Dobs_s_1215_Ago2_2809_18680_F	CCATGATTAGGTATGCTGCCACATC
Dobs_s_1715_Ago2_2809_18680_F	GGCTGAGCTGCAGTATGGCATTCT
Dobs_s_1984_Ago2_2809_18680_F	GCAATATCGCTTGCAACGCTCTG
Dobs_s_2586_Ago2_2809_F	GGTCAGCCATCAGTCCATTCAGG
Dobs_Ago2_2809_18680_3_R	CTTGGGGAGAACGGAACTTGG
Dobs_s_1922_Ago2_2809_18680_R	CGCTGATCGGGCGATGGATG
Dobs_s_1435_Ago2_2809_18680_R	CCATGCGCCACGAACCGTT
Dobs_s_798_Ago2_2809_18680_R	TCCACATTGACAAACGGACG

D. obscura Ago2f sequencing primers

Name	**Sequence**
Dobs_Ago2_18680_5_F	CCTTTGAGCTGTTCAGAGTGGAAC
Dobs_s_720_Ago2_2809_18680_F	AATCTTGGCGACGGCTACGAAGCTC
Dobs_s_1215_Ago2_2809_18680_F	CCATGATTAGGTATGCTGCCACATC
Dobs_s_1715_Ago2_2809_18680_F	GGCTGAGCTGCAGTATGGCATTCT
Dobs_Ago2_18680_M_F_2	GTAAATTGAGCCCCCAGTGTGTGTTGA
Dobs_s_1984_Ago2_2809_18680_F	GCAATATCGCTTGCAACGCTCTG
Dobs_Ago2_all_3_R	TGGCGCCAGTCAGATAGACACG
Dobs_s_1922_Ago2_2809_18680_R	CGCTGATCGGGCGATGGATG
Dobs_Ago2_18680_M_R	CCAGCTGAGTGCGCGGGTTATC
Dobs_s_1435_Ago2_2809_18680_R	CCATGCGCCACGAACCGTT
Dobs_s_802_Ago2_18680_R	CGCGTATAGCAGTGCTATGAC
Dobs_s_798_Ago2_2809_18680_R	TCCACATTGACAAACGGACG

D. obscura Ago2e sequencing primers

Name	**Sequence**
Dobs_Ago2_24803_5_F	CGMGGTACACTGGGCAGAATCG
Dobs_s_553_Ago2_24803_F	GTGAATGTGGACATCACACACAAGTG
Dobs_s_1044_Ago2_24803_F	GCAGTACTTCAGCCACAACACGG
Dobs_947_Ago2_24803_F	GGTGAATCGCAAGGACTCCACGCT
Dobs_s_1377_Ago2_24803_F	GAGCCTGGATCCGCACTTCAAGG
Dobs_s_1881_Ago2_24803_F	GGTCAGCGATGGGCAGTTCC
Dobs_2269_Ago2_24803_R	GGCTCCACGTTGTTGTATTTGTTGTG
Dobs_s_1812_Ago2_24803_R	GGCTGTTATCGACTCCATGTCCTC
Dobs_Ago2_24803_M_R	CAGTTCGGCTTTCTGTTTCAGTTC
Dobs_s_1064_Ago2_24803_R	GTGTTGTGGCTGAAGTACTGCAG
Dobs_s_578_Ago2_24803_R	CACTTGTGTGTGATGTCCACATTCAC

D. pseudoobscura Ago2a1/Ago2a3 sequencing primers

Name	Sequence
Dper.mir.A_F	TGGAGGTTGTGTTGGCAGTA
Dpse_s_1561_Ago2ab_F	CTGAARCACATTTAYTTGCCTATCG
Dpse_s_2011_Ago2ab_F	GTCAATCTGTGCCTGGATRCCAARG
Dpse_s_2487_Ago2ab_F	GTCGATAACCCTKGAGCACTTGCGTG
DpseAgo2A_R	CTANACGAARTACATAGGRTTCGTCTTC
Dpse_s_2808_Ago2ab_R	GTTGTATTGCGATGGARCTCCGYTCG
Dpse_s_2220_Ago2ab_R	CTCGACTGTGATCTGCTTGAKGC
Dpse_s_1611_Ago2ab_R	CTGYCCATCSTCAATGCGACATAG

D. pseudoobscura Ago2a3 sequencing primers

Name	Sequence
Dper.mir.A_F	TGGAGGTTGTGTTGGCAGTA
Dpse_s_1561_Ago2ab_F	CTGAARCACATTTAYTTGCCTATCG
Dpse_s_2011_Ago2ab_F	GTCAATCTGTGCCTGGATRCCAARG
Dpse_s_2487_Ago2ab_F	GTCGATAACCCTKGAGCACTTGCGTG
DpseAgo2A_R	CTANACGAARTACATAGGRTTCGTCTTC
Dpse_s_2808_Ago2ab_R	GTTGTATTGCGATGGARCTCCGYTCG
Dpse_s_2220_Ago2ab_R	CTCGACTGTGATCTGCTTGAKGC
Dpse_s_1611_Ago2ab_R	CTGYCCATCSTCAATGCGACATAG
Dpse_GA22965_3_out_F_2	CCAAGAGGACGAAAACACTGATTGG

D. pseudoobscura Ago2b sequencing primers

Name	Sequence
Dper.mir.D_F	CAGTACGATGTGAAGATCACGTCAGTAT
Dpse_s_912_Ago2b_F	CAGGAAGACGCAGGAATCGGAAG
Dpse_s_1561_Ago2ab_F	CTGAARCACATTTAYTTGCCTATCG
Dpse_s_2011_Ago2ab_F	GTCAATCTGTGCCTGGATRCCAARG
Dpse_s_2487_Ago2ab_F	GTCGATAACCCTKGAGCACTTGCGTG
Dpse_Ago2_UnivR	GCCAGTRAGRTAGACACGTCC
Dpse_s_2808_Ago2ab_R	GTTGTATTGCGATGGARCTCCGYTCG
Dpse_s_2220_Ago2ab_R	CTCGACTGTGATCTGCTTGAKGC
Dpse_s_1611_Ago2ab_R	CTGYCCATCSTCAATGCGACATAG
Dpse_s_1087_Ago2b_R	CAACCTCCAGACACTGCAAGGCTC

D. pseudoobscura Ago2c sequencing primers

Name	Sequence
Dpse_Ago2c_5_F	GGCCGTACCCTGACTTACACTGTGGAAC
Dpse_s_1248_Ago2_c_F	GACTACAGGCGTTACGATATCGAATC
Dpse_s_1683_Ago2_c_F	CCAAGATCATCCACTTGCTCG
Dpse_Ago2c_M_R	CTGTAATTCCGTATATCGGCCTTCTG
Dpse_s_1476_Ago2_c_R	CAACATGCAAGCAGAGCAGGTTC
Dpse_Ago2c_M_F	CTTGGAAAACGACTTCATTGTGGTGC
Dpse_s_2508_Ago2_c_F	CATGGAGTCGATAACTCTTGAGCACTTTC
Dpse_Ago2c_3_R	CACGAAATACATGGGGTTCGTTTTCAT
Dpse_s_2533_Ago2_c_R	GTGCTCAAGAGTTATCGACTCCATGTC
Dpse_s_1943_Ago2_c_R	CGACCACACTCATGTAGTTAATCTTACC

D. pseudoobscura Ago2d sequencing primers

Name	Sequence
DpseAgo2B_MF	ATGCCAGCTGTGGCCTACCA
Dpse_s_612_Ago2d_F	GTCAGAGCCCGGTAAAGCCTTTG
qrtD_R	GAAGTCAGTGCCCAGGCGT
Dpse_s_754_Ago2d_R	CAATGATCGTCATGGCCTTCGGAAAG
Dpse_Ago2d_M_F	CTGGATGGGAAGCAAACGACGG
Dpse_s_1443_Ago2_d_F	GTCAATGTGTGCCTGAATGACAACG
Dpse_s_1933_Ago2_d_F	GAGCACTTGCGTGTCTATCATCAGTACC
Dpse_Ago2d_3_R	GGAACTCTGGAACAATCAACCGCTTTT
Dpse_s_1911_Ago2_d_R	CCTGTATCTCCTCCAAGTCAGAGC
Dpse_s_1348_Ago2_d_R	CGACCACTCTTATGTAGTCAATCTTACC

D. pseudoobscura Ago2e sequencing primers

Name	Sequence
pse_Ago2E_5_F	CGAGGTGGCTGTGAACTACCTGCAG
Dpse_s_926_Ago2e_F	CAACTGTGATGGCACGAAGGTGAC
pse_Ago2E_M_F	GGCTTGTGGCACATCGACAGGTC
Dpse_s_1921_Ago2e_F	GCGTCCTACAACATGCAGTACCG
pse_Ago2E_3_R	CATGGGGTTCCTGCTGGACAGG
Dpse_s_2117_Ago2e_R	CATCCCTCGCAGCTCCTCGTTCC
Dpse_s_1575_Ago2e_R	CTTGGGTCCAGGCTCTTGGCGTC
Dpse_s_1055_Ago2e_R	GCACAGCTCAATGGGCAGATAGACGG

File S1. Alignment of drosophilid Ago2 homologues with third positions stripped, used to infer Figure 1. (.fasta, 103 KB)

www.genetics.org/lookup/suppl/doi:10.1534/genetics.116.192336/-/DC1/FileS1.fasta

File S2. Alignment of drosophilid Ago2 homologues, used for PAML analyses. (.fasta, 159 KB)

www.genetics.org/lookup/suppl/doi:10.1534/genetics.116.192336/-/DC1/FileS2.fasta

File S3. Sequence metadata for drosophilid Ago2 homologues. (.xlsx, 14 KB)

www.genetics.org/lookup/suppl/doi:10.1534/genetics.116.192336/-/DC1/FileS3.xlsx

File S4. Sequence data for haplotypes of Ago2 paralogues in *D. subobscura*, *D. obscura* and *D. pseudoobscura*. (.fasta, 276 KB)

www.genetics.org/lookup/suppl/doi:10.1534/genetics.116.192336/-/DC1/FileS4.fasta