Natural Variation of the Amino-Terminal Glutamine-Rich Domain in Drosophila Argonaute2 Is Not Associated with Developmental Defects

Daniel Hain¹, Brian R. Bettencourt², Katsumoto Okamura³, Tibor Csorba⁴*, Wibke Meyer⁵*, Zhigang Jin³, Jason Biggerstaff⁶, Haruhiko Siomi⁷, Gyorgy Hutvagner⁴, Eric C. Lai³, Michael Welte⁸*, H.-Arno J. Müller¹,*

¹ Division of Cell and Developmental Biology, College of Life Sciences, University of Dundee, Dundee, United Kingdom, 2 Alnylam Pharmaceuticals, Cambridge, Massachusetts, United States of America, 3 Sloan-Kettering Institute, Department of Developmental Biology, New York, New York, United States of America, 4 Wellcome Trust Centre for Gene Regulation and Expression, College of Life Sciences, University of Dundee, Dundee, United Kingdom, 5 Institut für Genetik, Heinrich Heine Universität, Düsseldorf, Germany, 6 Boston University, Boston, Massachusetts, United States of America, 7 Department of Molecular Biology, Keio University School of Medicine, Tokyo, Japan, 8 Department of Biology, University of Rochester, Rochester, New York, United States of America

Introduction

Argonaute proteins play key roles in diverse mechanisms of gene regulation from plants to fungi and animals, including humans. The Argonaute protein family is characterized by a stereotyped domain structure and has been divided into the Argonaute subfamily (related to Arabidopsis thaliana Argonaute) and the PIWI subfamily (related to D. melanogaster PIWI) [1,2]. Organisms may contain a large number of different Argonaute encoding genes, currently with a maximum of 26 family members identified in C. elegans [3]. The presence of multiple, differentially expressed family members in a single species indicates diverse biological functions of Argonaute proteins.

A common feature of all Argonaute proteins is that they bind small RNAs and mediate silencing of target transcripts by translational inhibition or RNA degradation [4]. Argonaute subfamily proteins interact with siRNAs and/or miRNAs and in animals exert diverse functions during development of various
somatic lineages. These functions range from selective degradation of unwanted maternal mRNAs in early development to the regulation of transcripts important for cell fate decisions during neuronal and muscle development as well as degradation of viral RNA in adult organisms [5]. An important function for the RNAi pathway was also demonstrated in innate immune response against viruses in animals and plants [6,7,8,9].

*Drosophila* has proved to be an excellent model system for research on Argonaute protein function [10,11,12]. The *Drosophila melanogaster* genome encodes five Argonaute family members: Pwii subfamily members Aubergine (Aub), Pwii and Ago3 and the Argonaute subfamily members Ago1 and Ago2. Ago1 is essential for miRNA function, while Ago2 predominantly acts in the siRNA pathway [11,13,14]. Ago2 is the carrier of both exogenous and endogenous siRNAs, and functions in virus defense, restriction of transposable elements, and cleavage of mRNA targets [15]. Ago1 and Ago2 also have overlapping functions, e.g. in the control of segment polarity during embryogenesis [16].

*Drosophila* Ago2 is unique among the Argonaute subfamily: Its sequence is highly divergent and clearly distinct from other metazoan Argonautes (including *Drosophila* Ago1) and the four human Argonautes Ago1, Ago2, Ago3 and Ago4 or even Argonautes from fungi or plants [17]. It is a modular protein with the canonical PAZ and PIWI domains in the carboxy-terminal half and an amino-terminal half that is highly enriched in glutamine residues [16]. The carboxy-terminal region (920 amino acids) is broadly conserved in Argonaute family members from plants to fission yeast and vertebrates and has well-characterized biochemical functions, such as binding and cleavage of RNA [1]. In contrast, the amino-terminal portion (in the following referred to as the amino-terminal domain or NTD) is not well characterized. While the corresponding region is typically very short in most other family members (e.g. 24 aa in human Ago2), the Ago2 NTD in *D. melanogaster* is 397 amino acids long and encompasses long stretches of low sequence complexity (40% of all residues are glutamine). The majority of the NTD consists of two types of imperfect glutamine-rich repeats (GRR), called GRR1 (6 amino acid imperfect repeat) and GRR2 (23 amino acid imperfect repeat) ([16]; see below). Previous analysis of the drop out (dop) mutation suggested that altered GRR copy number disrupts Ago2 function and leads to maternal-effect lethality [16].

In the present paper, we demonstrate that *ago2* is a complex locus that produces several distinct transcripts. The protein isoforms encoded by these transcripts have identical C-terminal regions that include the PAZ and Piwi domains. But while two long isoforms contain the usual NTD, a previously uncharacterized short isoform lacks both types of amino-terminal GRRs, reminiscent of mammalian Argonaute proteins. We find that two *ago2* mutant alleles widely employed as null alleles still express the short isoform. Using true null alleles, we show that the dop mutation is not allelic to *ago2* and that GRR variation does not cause striking defects in embryonic development. In fact, we find extensive GRR repeat-number variation among *D. melanogaster* strains isolated from natural and laboratory populations as well as tremendous variability of the Ago2 NTD between insect species. These data suggest that this domain is subject to rapid evolution and has a non-essential, possibly modular function.

**Results and Discussion**

The *ago2* locus generates transcripts encoding putative long and short isoforms

Gene annotation predicts that the *ago2* locus gives rise to the two transcripts AGO2-RB and AGO2-RC (FlyBase, 6_2010). They are largely identical, but differ at their very 5' ends, due to alternative splicing and the use of distinct transcription start sites (Fig. 1A). As a result, the two protein isoforms encoded by these transcripts, AGO2-PB and AGO2-PC, differ only by tiny unique stretches (6 or 9 amino acids) at their amino termini and have a 1208 amino acid region in common.

To functionally characterize Ago2, many studies employ the two putative null alleles *ago2*21B and *ago2*414. These alleles represent two independently derived deletions that eliminate the first two annotated AGO2-RB exons and part of the third intron (Fig. 1A; [11,18]). As they delete the annotated translational start sites for both AGO2-RB and AGO2-RC, these alleles are expected to abolish all *ago2* expression [11,18].

To test this prediction directly, we performed RT-PCR analysis with various primers across the *ago2* locus (Fig. 1B). Experiments that involved a primer specific for exon 1 (primer RB) failed to detect any *ago2* transcripts in these alleles (Fig. 1B), confirming the absence of AGO2-RB and AGO2-RC. However, when we employed primers from exons 3 through 7, we detected amplification products in these supposed null alleles. These products are not due to contamination with genomic DNA templates as their size was the same as in the wild type and sequencing confirmed that they represent correctly spliced transcripts. Furthermore, Northern analysis of total RNA of adult flies confirmed the presence of *ago2* transcript in *ago2*21B (Fig. 1C). These results indicate the presence of previously unannotated transcripts in *ago2*21B and *ago2*414.

These transcripts could also be detected with primers upstream of the third exon splice site, a region annotated as intronic for AGO2-RB and AGO2-RC (Fig. 1B, primer TS). Apparently, the alternate transcripts originate 5’ to exon 3. This property makes it possible to design isoform-specific primers to examine the expression of the alternative transcript and of AGO2-RB/AGO2-RC separately. Using these primers, we found that wild-type flies express both types of transcripts. Quantitative RT-PCR analysis on samples from adult flies indicates that the shorter transcript is expressed approximately 30% less than AGO2-RB/AGO2-RC (Fig. 1D).

These alternative transcripts are predicted to encode a shorter protein isoform of Ago2, in the following referred to as Ago2*short*. Since we have not been able to verify protein production from this transcript in vivo, we tested whether a cDNA encoding Ago2*short* can in principle produce a protein of the expected size. The Berkeley *Drosophila* genome project (BDGP) had isolated an EST clone, GM07030, which is predicted to encode Ago2*short*. We generated a transgene, in which HA-tagged GM07030 is driven under the control of UAS elements. Uniform expression of this transgene in the embryo results in a protein product of approximately 85 kDa, consistent with the predicted molecular weight of Ago2*short* (Fig. 1E).

A truncated Ago2 protein lacking GRRs was recently shown to support full biochemical reconstitution of RISC indicating that the N-terminal region does not contain domains that are essential for core RNAi [19]. In the lack of an NTD, Ago2*short* is reminiscent of the mammalian Argonautes; indeed, BLAST comparisons reveal significant similarity between Ago2*short* and human Ago2 throughout essentially the entire length of the proteins (not shown). We conclude that the *ago2* locus has the potential to encode at least three protein isoforms, only two of which encompass the unusual NTD that sets *D. melanogaster* Ago2 apart from other family members.

**Identification of true null alleles of *ago2***

The available putative null alleles of *ago2* remove 5’ exons with limited coding potential [11,18]. The existence of an alternative
Figure 1. The ago2\textsuperscript{218} and ago2\textsuperscript{414} alleles produce an alternative transcript encoding a putative short form of the protein. (A) ago2 genomic organization indicating 8 exons of the two alternatively spliced gene products Ago2-RB and Ago2-RC (UTRs are indicated in orange; CDS is indicated in blue). Also shown is the extent of the chromosomal deletions generated by imprecise excision of the EP(3)3417 P-element insertion, located 12 bp downstream of the transcriptional start site (TSS) for Ago2-RB. The ago2\textsuperscript{414} allele creates a deletion of 2336 bp and ago2\textsuperscript{321} creates a deletion of 2514 bp. (B) Cartoon of genomic organization of Ago2-RB and the positions of the PCR primers (RB, TS, GM (forward and reverse, respectively), E5, and PZ). RT-PCR experiments using primer pairs as indicated; all PCR products were subjected to DNA sequencing. The RB/PZ primer pair amplifies a product of 3488 bp indicative of a transcript encoding for the long isoform; this product is only detected in w\textsuperscript{1118} and dop\textsuperscript{1}, but not in ago2\textsuperscript{218} and ago2\textsuperscript{414}. GM/PZ amplifies a 2200 bp product, which is detected in all genotypes. The TS primer was designed to bind within the second intron. The TS/GM or TS/E5 primer pairs revealed products of the predicted sizes and sequences of an alternative transcript encoding for the short isoform (w\textsuperscript{1118} samples). All template polyA\textsuperscript{+} RNA was prepared from ovaries. (C) 10 μg total RNA from adult flies was subjected to Northern blotting. The ago2\textsuperscript{414} allele produces detectable ago2 transcripts whereas no transcript was detected in ago2\textsuperscript{218} homozygous flies. Note that the probe was designed against the downstream of ago2\textsuperscript{64} deletion and a putative truncated mRNA was detected in ago2\textsuperscript{64} homozygous flies. (D) Quantitative RT
isofom implies that ago2

and ago2

may retain some Ago2

function. Indeed, although Northern analysis of embryos had suggested that ago2

is transcript null [11], we observed substantial amounts of nearly full length transcript in adult flies (Fig. 1C).

We therefore examined additional excision events derived from EP(3)3417 to identify bona fide Ago2 null alleles. This screen identified two alleles that were transcript null in the adult, ago2

and ago2

(Fig. 1C). These alleles contain the same 5’ breakpoint located 34 bp downstream of the transcriptional start site of AGO2-RB, but have different 3’ breakpoints (Fig. 1A). The deletion in ago2

removes 2514 bp and is very similar to ago2

, but retains a section of the original EP element; this insertion may contribute to why it is transcript null. ago2

carries a deletion of 5750 bp and accumulates a truncated transcript consistent with this deletion. This deletion removes the PAZ domain and much of the PIWI domain, and is therefore unquestionably a null allele.

Homozygotes of ago2

and ago2

stocks were viable, but unlike ago2

and ago2

, they exhibited low fertility in both males and females. 55% of embryos obtained from ago2

homozygous mothers did not hatch (n = 221), while ago2

homozygous females did not lay any eggs. 50% ovaries from ago2

and 90% ovaries from ago2

homozygotes are rudimentary. Surprisingly, none of the phenotypes seen in ago2

or ago2

homozygotes except male sterility were seen in ago2

/ ago2

trans-heterozygotes. Moreover, ago2

/Df or ago2

/Df hemizygotes (using the molecularly defined deficiencies Df(3L)ED218 and Df(3L)BSC5580) displayed none of the above-mentioned phenotypes. Therefore, the fertility phenotypes observed in ago2

and ago2

homozygotes are likely attributable to background mutations. While a detailed analysis of these novel ago2 null alleles is beyond the scope of this paper, an exciting prospect for the future is to determine whether these alleles corroborate or extend other phenotypes that were previously reported for ago2

or ago2

alleles. Such phenotypes include defective heterochromatin formation [20], defective neuromuscular junction and egg chamber development [21], and defects in embryonic syncytial divisions and pole cell formation [22]. The partial penetrance of these defects in ago2

and ago2

homozygotes might result from residual Ago2 function due to Ago2

expression.

Ago2 is the carrier of endogenous small interfering RNAs (endo-siRNAs) [15]. Although a cohort of endo-siRNA substrates have been identified, along with a number of targeted transcripts, the biological function of endogenous RNAs remains poorly defined. The major phenotypes are molecular in nature, and include upregulation of transcripts derived from transposable elements and certain direct targets of hairpin RNA-derived siRNAs [23,24,25,26,27,28]. A detailed analysis of the true ago2 null alleles ago2

and ago2

might therefore provide the most stringent test of the overall requirements of the endo-siRNA pathway for various aspects of development.

GRR variation does not noticeably disrupt development

Our expression analysis indicates that D. melanogaster is able to produce Ago2 variants with and without the glutamine-rich NTD. This observation suggests that the NTD serves some important function as its expression is retained even though NTD-less forms can apparently be produced.

Our previous work had suggested a biological role for the NTD. We discovered variants in the GRR pattern and found that they were associated with severe disruption of embryonic development [16]. In particular, we had found that embryos from mothers homozygous for dop mutations show delays in cellularization, display abnormal transport of lipid droplets, and fail to hatch. Sequencing of genomic DNA of the dop

allele revealed that in this strain ago2 has 4 copies of GRR1 and 10 copies of GRR2; in contrast, the precursor chromosome (red e) exhibited the expected 4xGRR1/11xGRR2 pattern of the published wild-type sequence. For an independently derived dop allele (dop

), ago2 displayed 3xGRR1 and 11xGRR2. A functional connection between dop and ago2 was suggested by the observation that a chromosomal deletion (Df(3L)XG9) uncovering ago2 does not complement dop alleles and that the maternal lethality of dop mutations is partially rescued by expression of an ago2 cDNA [16]. Finally, dop alleles genetically interact with other genes involved in RNA silencing. These results suggested that deviation from the wild-type number of GRRs severely impairs Ago2 functions and results in developmental defects [16].

The identification of true null alleles of ago2 (above) enabled us to perform a stringent genetic test of this idea. dop homozygous as well as dop/Df(3L)XG9 mothers are completely sterile. However, mothers carrying one dop allele (either dop

or dop

) and one ago2 null allele (ago2

or ago2

) were fertile. For example, the embryos obtained from dop

/ago2

heterozygotes hatched at rates similar to wild type (Fig. 2A).

The embryos from transheterozygous (ago2

/dop) mothers also did not exhibit the developmental defects characteristic of dop mutant embryos. Embryos obtained from dop

or dop

homozygous mothers show normal syncytial development, but plasma membrane growth during cellularization is strongly delayed (Fig. 2B; [16]). Embryos obtained from transheterozygous mothers do not display such delay in cellularization (Fig. 2B). Furthermore, lipid droplets are severely mislocalized in dop mutant embryos during germ-band extension (Fig. 2C). In contrast, lipid-droplet distribution in embryos from ago2/dop mothers is similar to the wild type (Fig. 2C). These results demonstrate that dop is not allelic to ago2. In addition, there is no evidence that ago2 with non-standard GRR pattern results in developmental defects.

GRR variation does not affect RISC assembly

Depletion of Ago2 from cell lines or flies results in a complete block of the RNAi response [10]. Furthermore, biochemical data indicate that Ago2 is an essential and sufficient component of RISC, the central machinery that directs mRNA cleavage by siRNAs [10,12,29,30]. Ago2

is essential for this process, because ago2

and ago2

homoyzogous mothers lack most, if not all, of the RNAi response [11,16]. Since the dop

chromosome (characterized by the 4/10 GRR ago2 pattern) exhibits a mild suppression of experimental RNAs in vivo [16], variation in Ago2

might modulate the formation of RISC.

To test this possibility directly, we assayed RISC-formation in vitro in embryo extracts obtained from wild-type and dop

embryos, using gel shift experiments with siRNA [31]. RISC formation as observed by assembly of complexes over time was
unimpaired in dop1 mutant embryos when compared to wild type (Fig. 3). We therefore conclude that the variation of GRR pattern in dop mutants does not affect formation of RISC. This finding is consistent with reports that recombinant Ago2 lacking the first 277 aa of the NTD (this truncation removes all GRR1 and seven GRR2 repeats) supports RISC assembly [19].

**GRR copy number varies extensively between different fly strains**

The complementation tests above demonstrate that dop and ago2 represent distinct genes. Yet, in two independently derived lines, dop mutations are associated with changes in the GRR pattern of ago2 [16]. This fortuitous association might be explained if changes in the GRR pattern of ago2 occur at high frequency.

We therefore examined GRR1 and GRR2 copy number in twenty-two laboratory strains of diverse origins. In these strains, GRR1 copy number varied between 2 and 4, and GRR2 copy number between 10 and 16 (Table 1, Fig. 4, Fig. S1, and data not shown). For example, copies of the wild-type strain Oregon R obtained from two different laboratories had 4x GRR1 and 13x GRR2 or 3x GRR1 and 16x GRR2, respectively. Various combinations of GRR1 and GRR2 copy numbers can occur. In the following, we refer to these combinations as GRR haplotypes.

**Figure 2. Null alleles of ago2 complement dop1 mutants.** (A) Percentage of larvae hatching from indicated genotypes (n>300 per genotype). (B) Bright-field images of post-gastrula embryos (stage 10) show that the periphery of dop1 mutant embryos is abnormally transparent compared to wild type (wt) [16]. These transparency changes are characteristic of altered lipid-droplet distribution [57]. Embryos obtained from mothers transheterozygous for dop1 and either ago2221 or ago2454 exhibit a normal pattern lipid-droplet distribution. (C) Immunolabeling of the membrane protein Neurotactin (red) and DNA (DAPI, blue) in embryos during and after cellularization. Embryos from dop1 homozygous females exhibit a severe delay in membrane ingression marked by the absence of Neurotactin staining at the interface between adjacent nuclei [16]. Embryos from dop1/ago2221 transheterozygous mothers show normal membrane ingression during cellularization.

doi:10.1371/journal.pone.0015264.g002

**Figure 3. Comparison of RISC formation in wild type and dop1 mutant embryos.** In-vitro silencing complexes formation was analyzed by incubating dilutions of cell free extracts (conc: µg/µl total protein) of wild type or dop1 Drosophila embryo extracts with pre-labelled siRNA. The formed silencing complexes were separated on a native acrylamide gel. The free-siRNAs and the R1, R2 and holo-RISC complexes are shown (The star marks an unknown band).

doi:10.1371/journal.pone.0015264.g003
To assess if variation in GRR copy number is a consequence of extended laboratory culture, we examined ten strains recently isolated from the wild. These strains are isogenic for third chromosomes that had been extracted from a single wild North American population [32]; we analyzed their GRR haplotypes within five years after they had been established. These strains exhibited GRR haplotype diversity equal to or greater than that found in the long-established laboratory strains (Table 1).

Table 1. GRR pattern in Ago2 in different D. melanogaster strains.

<table>
<thead>
<tr>
<th>Line</th>
<th>GRR1</th>
<th>GRR2</th>
<th>Length of NTD</th>
<th>Glu in NTD</th>
<th>% Glu in NTD</th>
</tr>
</thead>
<tbody>
<tr>
<td>JCPA120</td>
<td>2</td>
<td>13</td>
<td>431</td>
<td>168</td>
<td>39%</td>
</tr>
<tr>
<td>JCPA6</td>
<td>2</td>
<td></td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JCPA103</td>
<td>3</td>
<td>10</td>
<td>368</td>
<td>146</td>
<td>40%</td>
</tr>
<tr>
<td>JCPA54</td>
<td>3</td>
<td></td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JCPA35</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JCPA129</td>
<td>3 n.d.</td>
<td>11</td>
<td>397</td>
<td>159</td>
<td>40%</td>
</tr>
<tr>
<td>JCPA122</td>
<td>4</td>
<td>11</td>
<td>397</td>
<td>159</td>
<td>40%</td>
</tr>
<tr>
<td>JCPA31</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JCPA113</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JCPA2</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canton-S</td>
<td>n.d.</td>
<td>11</td>
<td>397</td>
<td>159</td>
<td>40%</td>
</tr>
<tr>
<td>w-14 melbourne</td>
<td>3</td>
<td>11</td>
<td>391</td>
<td>155</td>
<td>40%</td>
</tr>
<tr>
<td>Tai255.1</td>
<td>4</td>
<td>13</td>
<td>443</td>
<td>177</td>
<td>40%</td>
</tr>
<tr>
<td>Or R*</td>
<td>3</td>
<td>16</td>
<td>506</td>
<td>199</td>
<td>39%</td>
</tr>
<tr>
<td>Or R***</td>
<td>3</td>
<td>11</td>
<td>391</td>
<td>155</td>
<td>40%</td>
</tr>
<tr>
<td>dop[46]</td>
<td>4</td>
<td>10</td>
<td>374</td>
<td>150</td>
<td>40%</td>
</tr>
<tr>
<td>red e</td>
<td>4</td>
<td>11</td>
<td>397</td>
<td>159</td>
<td>40%</td>
</tr>
<tr>
<td>In(1)AB</td>
<td>4</td>
<td>11</td>
<td>397</td>
<td>159</td>
<td>40%</td>
</tr>
<tr>
<td>TM6B</td>
<td>4</td>
<td>12</td>
<td>420</td>
<td>168</td>
<td>40%</td>
</tr>
<tr>
<td>TM8</td>
<td>n.d.</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TM1</td>
<td>n.d.</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LVM</td>
<td>n.d.</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dp(3;3)Cam36</td>
<td>n.d.</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Genomic DNA encompassing the GRR1 or GRR2 repeats was amplified using PCR. The number of repeats was estimated by gel electrophoresis as in Fig. S1. Top: The JCPA strains were derived from a single wild North American population in 2001 [32]. Middle: Wild-type laboratory strains; note that two strains characterized as Or R (obtained from different laboratories) had distinct GRR repeat patterns. Bottom: A selection of other laboratory strains; numbers in brackets or determined by sequencing the PCR fragments (Fig. 3A).

To assess if variation in GRR copy number is a consequence of extended laboratory culture, we examined ten strains recently isolated from the wild. These strains are isogenic for third chromosomes that had been extracted from a single wild North American population [32]; we analyzed their GRR haplotypes within five years after they had been established. These strains exhibited GRR haplotype diversity equal to or greater than that found in the long-established laboratory strains (Table 1).

Our results indicate that in D. melanogaster the NTD of Ago2 can undergo drastic changes in length and sequence composition. As we have not attempted a comprehensive survey, we suspect that GRR variation is not restricted to the haplotypes we identified. Considering just the instances listed in Table 1, NTD length can vary by over 35% (from 368 to 506 amino acids). These changes in the NTD likely arise by repeat expansion and contraction mediated by unequal crossing over [33]. Interestingly, the glutamine fraction of the NTD remains fairly constant (~40%).

Failure to detect phenotypic consequences of NTD variability

The NTD variability we uncovered provides an independent test whether GRR copy number and the dop phenotype are linked. Many of the strains in Table 1 are homozygous viable and at least grossly normal, suggesting that the exact composition of the amino-terminal domain of Ago2 is nonessential for viability and that many GRR haplotypes are compatible with normal development. In particular, the Tai255.1 strain, a wild-type D. melanogaster strain isolated 1983 at Ivory Coast [34] had the same 3/11 GRR haplotype found in the dop allele. However, unlike dop or dop embryos, Tai255.1 embryos develop apparently normally (Fig. 5). They cellularized normally and displayed normal lipid-droplet transport (Fig. 5A,B). We therefore conclude that GRR variation in ago2 does not disrupt embryogenesis and in particular is not responsible for the developmental defects observed in dop mutants.

We next examined if GRR variation influences expression of long and short Ago2 isoforms. Eight wild type third-chromosome extracted 3CPA strains all express both long and short ago2 transcripts (Fig. 6). We conducted quantitative real-time PCR using ago2 long- and short-specific primers and cDNA prepared from 3-day old females [35]. ago2long expression varied significantly among the strains (ANOVA; p<0.05), and was consistently higher than ago2short expression (Fig. 6; average ago2long expression is 11% higher than ago2short). In contrast, ago2short expression did not vary significantly among the strains, and we detected no correlation between ago2long and ago2short expression levels. Last, there was no obvious relationship between GRR haplotypes and ago2long or ago2short expression in these strains. These results suggest independent regulation of transcripts for ago2long and ago2short.

In summary, we find no evidence that GRR variation is associated with dramatic changes in embryonic development or survival, in RISC formation or with changes in isoform expression. It is conceivable that GRR variation does not have any functional consequences and simply arises as “noise” due to the propensity of repeat regions to undergo unequal crossing over. Alternatively, GRR variation may be involved in more subtle functions of Ago2 than the ones tested here. While RISC formation per se can occur in the absence of most of the GRRs [19], regulation of Ago2 activity or efficient recognition of certain targets might be modulated by the NTD. For example, the long Ago2 isoform, which includes the NTD, is implicated in anti-viral responses as part of the RNAi pathway [36]. It was proposed that this reflects adaptive evolutionary in response to changes in viral pathogens, as part of a virus-host arms race [37]. The observed changes in these RNAi pathway genes indeed bear the signature of positive selection [36].

Extensive remodeling of the NTD between closely related species

Previous studies have uncovered unusually high variability in non-NTD regions of Ago2 [36,37]. Comparison of the C-terminal 839 amino acids of Ago2 (i.e. largely excluding the NTD) in pairs of recently diverged Drosophila species revealed rapid evolutionary changes. These observations placed ago2 among the fastest evolving genes in Drosophilids, along with other components of the RNAi pathway [36]. It was proposed that this reflects adaptive evolutionary in response to changes in viral pathogens, as part of a virus-host arms race [37]. The observed changes in these RNAi pathway genes indeed bear the signature of positive selection [36].
The previously described evolutionary changes in Ago2 are concentrated in regions of unknown function [36]. In contrast, regions of Ago2 associated with core functions are conserved across all Drosophila species. The function of the NTD is also unknown. We therefore examined the NTDs in closely related species to determine whether their sequence evolution matches the former or latter pattern.

For this analysis, we focused on the three species most closely related to Drosophila melanogaster, namely Drosophila simulans, Drosophila mauritiana, and Drosophila sechellia. These species diverged <1 Mya (sechellia versus simulans) or ~2.3 Mya (D. melanogaster versus D. sechellia/simulans/mauritiana lineages) ago from each other, and the sequences of most genes are highly conserved in all four species [38,39].

We first examined the DNA sequences that correspond to exon 3 of ago2 in D. melanogaster. Here, this exon encodes 330 amino acids of the NTD (including all GRR2 repeats) plus the following 92 amino acids, whose sequence is conserved in Argonautes from fungi, plants and animals. The latter, non-NTD stretch displayed very high amino acid sequence conservation between the four Drosophila species (Fig. S2, S3, and data not shown). And in each of the four species, the sequence immediately 5' to this conserved region has the potential to encode glutamine-rich, repetitive stretches, but the exact sequence is highly divergent ([16], and Fig. S2). For example, the 23 aa GRR2 repeats characteristic of D. melanogaster are not apparent in any of the other three species. This observation suggests that NTDs can change drastically between even closely related species.

As the Ago2-encoding genomic sequences of D. simulans and D. mauritiana were nearly identical but differed from both D. sechellia and D. melanogaster, we performed a detailed comparison of the entire NTD between D. melanogaster, D. simulans and D. sechellia. The analysis above does not reveal the full extent of the NTDs or if these putative NTDs are indeed expressed. We therefore predicted possible upstream exons and designed primers in these exons and in exon 5 for RT-PCR analysis. For both D. simulans and D. sechellia, this strategy allowed us to amplify cDNA fragments spanning these exons; sequencing revealed that these fragments indeed encompass the putative exon 3, including the highly variable glutamine-rich stretches. Thus, like D. melanogaster, these species express unusual Ago2 NTDs.

The NTDs of the three species differ extensively from each other in length (D. sechellia: 211 aa; D. simulans: 277 aa; D. melanogaster: 368 to 506 aa (Table 1)); their glutamine content, however, was fairly similar (36–40%) (Table 2; Fig. S3). These length differences do not arise from variation in copy numbers of similar repeats, as between...
Figure 5. Altered GRR pattern is not associated with developmental defects characteristic of dop alleles. (A) Kymographs from time-lapse microscopy of w[1118], Tai255.1 and dop[1] embryos. Upper panels show primary data and lower panels indicate the embryo surface membrane (red) and the ingressing plasma membrane (yellow) over time (t). In the wild type (wt), membrane ingression is initially slow at 0.29 μm/min) and then increased in the second phase to 0.72 μm/min. Similar values were observed for Tai255.1. dop[1] mutant embryos exhibit a strong reduction in membrane ingression during slow phase (0.12 μm/min); fast phase is normal. (B) Bright-field images of post-gastrula stage embryos indicating distribution of lipid droplets in wild-type (wt), Tai255.1 and dop[1] mutant embryos. In wild-type and Tai255.1 embryos, the periphery of the embryo is opaque due to the presence of lipid droplets within the ectoderm and mesoderm cell layer. dop[1] mutant embryos exhibit a characteristic lipid-droplet transport defect indicated by a concentration of lipid droplets in the center of the embryo and a transparent embryonic periphery. doi:10.1371/journal.pone.0015264.g005

Figure 6. Expression of ago2short and ago2long in wild-type D. melanogaster strains. Levels of Ago2long (blue) and Ago2short (red) in the 3CPA strains. Each symbol represents mean expression level relative to Rpl32, indicated as ratio of qPCR critical thresholds ("CT"). Bars represent one SEM. X axis denotes strain. doi:10.1371/journal.pone.0015264.g006
Table 2. NTD of Ago2 orthologs in *Drosophila* species.

<table>
<thead>
<tr>
<th>species</th>
<th>name</th>
<th>Length of NTD</th>
<th>Glu in NTD</th>
<th>% Glu</th>
</tr>
</thead>
<tbody>
<tr>
<td>melanogaster</td>
<td>Dmel\textunderscore Ago2-PC</td>
<td>397</td>
<td>159</td>
<td>40%</td>
</tr>
<tr>
<td>melanogaster</td>
<td>Dmel\textunderscore Ago2short</td>
<td>0</td>
<td>0</td>
<td>N.A.</td>
</tr>
<tr>
<td>simulans</td>
<td>Dsim\textunderscore Ago2</td>
<td>277</td>
<td>101</td>
<td>36%</td>
</tr>
<tr>
<td>sechellia</td>
<td>Dsec\textunderscore Ago2</td>
<td>211</td>
<td>83</td>
<td>39%</td>
</tr>
<tr>
<td>yakuba</td>
<td>Dyak\textunderscore Ago2</td>
<td>357</td>
<td>153</td>
<td>43%</td>
</tr>
<tr>
<td>erecta</td>
<td>Dere\textunderscore Ago2</td>
<td>298</td>
<td>124</td>
<td>42%</td>
</tr>
<tr>
<td>ananassae</td>
<td>Dana\textunderscore Ago2</td>
<td>214</td>
<td>78</td>
<td>36%</td>
</tr>
<tr>
<td>pseudoobscura</td>
<td>Dps\textunderscore Ago2a</td>
<td>226</td>
<td>43</td>
<td>19%</td>
</tr>
<tr>
<td>pseudoobscura</td>
<td>Dps\textunderscore Ago2b</td>
<td>182</td>
<td>36</td>
<td>20%</td>
</tr>
<tr>
<td>pseudoobscura</td>
<td>Dps\textunderscore Ago2c</td>
<td>157</td>
<td>55</td>
<td>35%</td>
</tr>
<tr>
<td>pseudoobscura</td>
<td>Dps\textunderscore Ago2d</td>
<td>149</td>
<td>34</td>
<td>23%</td>
</tr>
<tr>
<td>pseudoobscura</td>
<td>Dps\textunderscore Ago2e</td>
<td>46</td>
<td>3</td>
<td>7%</td>
</tr>
<tr>
<td>persimilis</td>
<td>Dper\textunderscore Ago2a</td>
<td>5</td>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>persimilis</td>
<td>Dper\textunderscore Ago2b</td>
<td>200</td>
<td>48</td>
<td>24%</td>
</tr>
<tr>
<td>persimilis</td>
<td>Dper\textunderscore Ago2c</td>
<td>142</td>
<td>57</td>
<td>40%</td>
</tr>
<tr>
<td>persimilis</td>
<td>Dper\textunderscore Ago2d</td>
<td>142</td>
<td>34</td>
<td>24%</td>
</tr>
<tr>
<td>persimilis</td>
<td>Dper\textunderscore Ago2e</td>
<td>46</td>
<td>3</td>
<td>7%</td>
</tr>
<tr>
<td>willistoni</td>
<td>Dwil\textunderscore Ago2a</td>
<td>202</td>
<td>56</td>
<td>28%</td>
</tr>
<tr>
<td>willistoni</td>
<td>Dwil\textunderscore Ago2b</td>
<td>190</td>
<td>51</td>
<td>18%</td>
</tr>
<tr>
<td>mojavensis</td>
<td>Dmoj\textunderscore Ago2</td>
<td>247</td>
<td>57</td>
<td>23%</td>
</tr>
<tr>
<td>virilis</td>
<td>Dviri\textunderscore Ago2</td>
<td>335</td>
<td>96</td>
<td>29%</td>
</tr>
<tr>
<td>grimshawi</td>
<td>Dgr\textunderscore Ago2</td>
<td>316</td>
<td>99</td>
<td>31%</td>
</tr>
</tbody>
</table>

Predicted Ago2 proteins in the 12 sequenced *Drosophila* species were examined for length and glutamine content of the amino-terminal domain. The NTD was defined by comparison to *Homo sapiens* Ago2, as described in Experimental Procedures. For details of how these proteins were predicted and for FlyBase or GenBank names, see Fig. S4. For *D. melanogaster*, the canonical 3xGRR1, 11xGRR2 long isoform and the short isoform are listed. doi:10.1371/journal.pone.0015264.t002

distinct *D. melanogaster* strains (Table 1). Rather, NTD organization is quite different between species. The primary sequences of the *simulans* and *sechellia* NTDs are clearly related; both species contain a highly similar sequence of ~118 aa. But *D. simulans* Ago2 has imperfect, tandem copies of this sequence while *D. sechellia* Ago2 has a single copy (Fig. S2B). Instead, in *D. sechellia* Ago2, a 15 aa stretch internal to the 118 aa region is repeated three times. Neither the 118 aa nor the 15 aa repeat displays striking similarity to the 6 aa (GRR1) or 23 aa (GRR2) repeats in *D. melanogaster*. In contrast, the remainder of the Ago2 sequence in all three species is highly similar and can easily be aligned (Fig. S3). Thus, the primary sequence of the NTD is extremely variable, much more so than the rest of the protein (which already is among the most rapidly evolving proteins in *Drosophila* [36]).

In *D. melanogaster*, the NTD displays drastic intra-specific variation (Table 1). To test if similar variability might occur in *D. simulans*, we amplified by PCR the genomic region corresponding to exon 3 from nine different strains and sequenced the PCR fragments. In total, we recovered five haplotypes that vary from each other by small deletions and insertions (Fig. S2A). We conclude that the NTD evolves rapidly both within and between species.

The NTD of Ago2 is highly variable throughout insects

In the genus *Drosophila*, nearly full genome sequences are available for nine additional species [40]. This resource provides an opportunity to examine the pattern of Ago2 evolution across longer evolutionary times, on the order of 50 Mya. We first identified all Argonaute family members in the predicted proteomes of all twelve species, by reciprocal blastp best-hit analysis and comparison of FlyBase annotations in all 12 species. We conducted multiple alignment of the protein sequences, neighbor-joining tree building and 1000 bootstrap trials via CLUSTALX (data not shown). Examination of the phylogram revealed a subset of these proteins as most closely related to Ago2 of *D. melanogaster*, clearly separate from Ago1, Piwi, Aubergine, and Ago3 orthologs. Interestingly, *D. willistoni* had two Ago2 paralogs and *D. pseudoobscura* and *D. persimilis* each had five Ago2 paralogs (see also Fig. 7). Multiple Ago2s in a single genome have also been reported for the mosquito *Culex quinquefasciatus* [41] and the flour beetle *Tribolium castaneum* [42] (Table 3). Presumably, multiple ago2 genes in a single organism allow those variants to be adapted for specific tasks. We speculate that in *D. melanogaster* the multiple isoforms generated from a single ago2 gene allow similar specialization.

The protein annotation for the 11 non-*melanogaster* genomes employs a consensus between eight distinct gene prediction algorithms [40]. For most of the Ago2 candidates, either this consensus or at least one of the individual algorithms indicated the presence of a glutamine-rich NTD, suggesting that such amino-terminal domains are common throughout the genus. In total, 18 out of the 21 Ago2 sequences in Table 2 have an NTD of more than 240 amino acids and an NTD glutamine content above 12%. Although the presence of these NTDs will have to be independently confirmed in the future, for several of these candidates, cDNA or microarray evidence support that the NTD region is indeed expressed (Table S1).

Although many of these predicted Ago2 proteins have glutamine-rich NTDs, the primary sequence, length and glutamine content vary greatly between family members (Table 2, Fig. S4): This is even true for Ago2 paralogs in the same species, such as for the Ago2s of *D. willistoni* that have NTDs of 244 and 356 amino acids, respectively. And some of the *D. persimilis* and *D. pseudoobscura* copies even seem to lack a glutamine-rich NTD entirely. In addition, there are radical changes in repeat organization between species [16]. For example, in various Ago2s, we identified repeats of 11, 12, 13, 16, 22 and 31 amino acids in length. Even in the sister species *D. persimilis* and *D. pseudoobscura* copies which diverged ~0.5 Mya ago [43], the closely related Ago2 pair Dps\textunderscore Ago2c and Dper\textunderscore Ago2c are identical across much of the NTD and the rest of the coding sequence, but are distinguished from each other by unique repeats (Fig. S5).

To determine if Ago2 NTDs are found outside the genus *Drosophila*, we searched various databases for Ago2 candidates and identified them in a wide variety of insect species, including flies, beetles, bees, wasps and lice, and even the crustacean *Daphnia pulex* (Fig. S4). A comparison with Argonaute family members from fungi, plants and animals (i.e., the dataset analyzed in [17]) indicated that insect plus *Daphnia* Ago2s are a monophyletic assemblage, well supported by bootstrap trials (data not shown). As previously noticed for *D. melanogaster* Ago2 [17], this entire clade is highly divergent and distinct from all other Argonautes. And like the *Drosophila* Ago2s, most members of this clade are predicted to carry glutamine-rich amino-terminal domains ([16], Fig. 7, Fig. S4, Table 3). In several instances, cDNA data support that these NTDs are expressed (Table S1). These domains are often, but not always, organized as multiple copies of imperfect repeats. As in the genus *Drosophila*, the lengths and sequences of these repeats vary substantially ([16], Fig. S4), and the NTDs overall do not display extensive similarity in primary sequence. Thus, highly variable
Argonaute2 NTDs are apparently ancient and predate the origin of insects. At the moment, it is impossible to determine whether drift or selection is the primary determinant of this variability. It is conceivable that the dramatic changes in NTD primary sequence are simply a consequence of the high instability of repetitive DNA sequences. If the NTD has no function, it may be free to vary with little constraints. In this scenario, selective sweeps due to positive selection on the rest of the Ago2 sequence, e.g., frequency-dependent adaptation of the anti-viral response [36], may cause certain NTD haplotypes that happened to be present in those ago2 variants to rise in frequency via hitchhiking. Indeed, the number and heterozygosity of NTD alleles in the natural PA population is striking, especially when compared with other protein-coding repeats which may experience similar selective pressures. Considering the unique GRR1/GRR2 haplotypes listed in Table 1 (excluding missing data), there are 7 NTD alleles segregating among the PA strains. Heterozygosity, measured as one minus the sum of the allele frequencies, is 0.815. A recent survey of 58 protein-coding repeats in the same strains found lower diversity overall [44] the mean and maximum number of alleles were 2.83 ± 0.2 and 8, respectively. Mean and maximum heterozygosity in the same survey were 0.457 ± 0.03 and 0.83. This survey considered primarily homotypic amino-acid repeat loci, not mixed or “minisatellite” repeats as are found in the Ago2 NTD. Analyses of mixed protein-coding repeats in natural D. melanogaster.
been proposed as protein interaction domains [50] and changes in the number of glutamine-rich repeats can modulate interaction strength [51,52], RGG or NTD variation might modulate the affinity of the Argonaute protein to particular targets. Indeed, mutant versions of TbAgo1 show progressively stronger RNAi responses, the more copies of the RGG motif are present [40].

### Conclusion

The evolution of the Argonaute protein family resulted in members with specific tasks in small RNA functions within a single organism. Here we uncover additional complexity: Many insects express multiple Ago2 variants, either from independent genes or by generating multiple isoforms from a single gene. In addition, there is variation between different individuals of the same species, due to extensive variation in NTD organization. The functional significance of this variation remains unknown; we find that it does not grossly alter embryonic development or RISC assembly. But given the known role of Ago2 in protection against viruses, we speculate that this variability might allow fine-tuning of anti-viral responses. Such a modulatory role of the NTD would also be consistent with the rapid evolution of this domain, as observed in comparisons across insects. A potential association between NTD haplotype and Ago2 activity might provide novel insights into the plasticity of RNAi pathways in general and the innate viral response mechanisms in particular.

### Materials and Methods

#### Drosophila strains and culture

*Drosophila* strains and embryos were cultured under standard laboratory conditions. The following strains were used in this study: Oregon R, Canton-S (gift from L. Griffith), ago2*116 (gift of F.-B. Gao), ago2*114, ago2*154, ago2*151. Screening of deletion mutants was carried out as previously described [11]. The EP transposon in the EP(3)3417 line was mobilized by crossing EP(3)3417 virgins with males possessing A2-3 transposase. Deletions were determined as described previously [11]. The D. melanogaster 3CPA strains were donated by B.P. Lazzaro and had been established in 2001 via extracting 3rd chromosomes segregating in a single Pennsylvania (USA) population over balancers and crossing into a common background [32]. D. melanogaster strains Tai255.1, w-14 melbourne and Oregon R** as well as D. simulans strains Oxnard, Tsimbazaza, vermilion, maz1, maz6 and C167.4 were a gift from H. Hollocher. These strains are described in Sainz et al. [53]. Additional strains were obtained from the Tucson Drosophila Stock Center; three D. simulans strains (sim2 = #14021-0251.047, sim3 = #14021-0251.004), one D. mauritiana strain (#14021-0241.46) and two D. sechellia strains (#14021-0248.08, #14021-0248.25). All other strains in Table 1 were obtained from the Bloomington Drosophila Stock Center.

#### Quantitative real-time PCR

Groups of 15 female flies aged 3-5 days were anesthetized under light CO2 and flash-frozen in liquid N2. Total RNA was prepared from each group, and cDNA was synthesized using oligo-dT primers according to standard protocols (see BETTENCOURT et al. 2008). Three replicate preparations were obtained for each of the 3CPA strains in Figure 6. qRT-PCR was conducted using primers specific for Ago2*154 and Ago2*151; sequences are available upon request. The ribosomal protein gene RpL32 was also examined as a control according to BETTENCOURT et al. (2008). A minimum of two replicate PCR reactions per strain/gene were conducted. Reactions were conducted and analyzed on a BIO-RAD myIQ

---

**Table 3. NTD of Ago2 orthologs in other species.**

<table>
<thead>
<tr>
<th>Species</th>
<th>Name</th>
<th>Length of NTD</th>
<th>% Glu</th>
</tr>
</thead>
<tbody>
<tr>
<td>Culex quinquefasciatus</td>
<td>CqU-Ago2-1</td>
<td>87</td>
<td>20</td>
</tr>
<tr>
<td>Culex quinquefasciatus</td>
<td>CqU-Ago2-2</td>
<td>208</td>
<td>41</td>
</tr>
<tr>
<td>Aedes aegypti</td>
<td>Aae-Ago2</td>
<td>173</td>
<td>77</td>
</tr>
<tr>
<td>Apis mellifera</td>
<td>Amel-Ago2</td>
<td>366</td>
<td>148</td>
</tr>
<tr>
<td>Nasonia vitripennis</td>
<td>Nvit-Ago2</td>
<td>222</td>
<td>50</td>
</tr>
<tr>
<td>Bombyx mori</td>
<td>Bmor-Ago2</td>
<td>232</td>
<td>11</td>
</tr>
<tr>
<td>Tribolium castaneum</td>
<td>TcAs-Ago2</td>
<td>49</td>
<td>5</td>
</tr>
<tr>
<td>Tribolium castaneum</td>
<td>TcAs-Ago2b</td>
<td>69</td>
<td>18</td>
</tr>
<tr>
<td>Acrythosiphon pisum</td>
<td>Apis-Ago2</td>
<td>146</td>
<td>26</td>
</tr>
<tr>
<td>Pediculus humanus</td>
<td>Phum-Ago2</td>
<td>429</td>
<td>92</td>
</tr>
<tr>
<td>Daphnia pulex</td>
<td>Dpul-Ago2</td>
<td>272</td>
<td>67</td>
</tr>
<tr>
<td>Homo sapiens</td>
<td>HsAp-Ago2</td>
<td>24</td>
<td>1</td>
</tr>
</tbody>
</table>

Predicted Ago2 proteins in various insects and the crustacean *Daphnia pulex* were examined for length and glutamine content of the amino-terminal domain, using the same approach as for Table 2. Human Ago2 is included for comparison.

DOI:10.1371/journal.pone.0015264.t003
thermocycler using a standard 40-cycle protocol, and amplification products were verified via electrophoresis and melt-curve.

**In vivo observations and immunohistology**

For in vivo observation, embryos were collected and staged on yeasted apple juice plates and mounted in halocarbon oil 27 (Sigma-Aldrich) on microscope slides. Video-microscopy was performed on a Zeiss (Germany) microscope equipped with Nomarski optics, and time-lapse movies were taken using the OpenLab software (Improvision, UK). For immunohistology, embryos were either heat-fixed or fixed using modified Stefanini’s fixative and stained with antibodies essentially as described elsewhere [54]. The following antibodies were used: mouse anti Neurotactin (DSHB, Iowa, USA); mouse anti Arm [53]. All secondary fluorochrome conjugated antibodies were from DiNovo (Germany) or Molecular Probes (Eugene, OR). Imaging was performed on a Leica SP2 Confocal microscope, and image processing was performed with Adobe Photoshop.

**Molecular Biology**

Molecular cloning was performed following standard protocols. RT-PCR was performed using ‘one step RT PCR kit’ (Qiagen, Germany) with embryonic poly A+ RNA as templates. Primer sequences are available upon request. Products from RT-PCR were confirmed by sequencing (Seqlab, Göttingen, Germany). GenBank accession numbers of full-length cDNA for *D. melanogaster ago2* are BT003546 (EST RE04347, RB isoform) and BT099682 (EST RE06670, RC isoform). In addition, the sequence of the EST GM07030 (GenBank accession number AV094751) is consistent with expression of the *ago2*short isoform. PCR to analyze NTDs in *D. melanogaster, simulans, mauritiana* and *sechellia* was performed with primers located in introns surrounding exons 2 and 3 of *ago2*, respectively. Protein extraction and immunoblotting was performed as described before [16].

**Northern blotting**

RNA samples were prepared from 2-4 day old male flies using Trizol. 10 µg total RNA was loaded in each lane on a denaturing agarose gel and transferred to Genescreen plus membrane. The template for the probe was prepared by PCR using an *ago2* cDNA fragment cloned in pGEMT-easy vector. The primers used for PCR amplification are available upon request. Probes were labeled using Random Primed DNA Labeling Kit (Roche) and 32P-dCTP, and hybridized at 65°C in 500 mM Church phosphate buffer containing 7% SDS and 1 mM EDTA.

**RISC formation assay**

*Drosophila* embryo extract preparation and synthetic siRNA labelling and annealing were described previously [56]. Reactions used different dilutions of cell free extract of wild-type or *dpp* mutant embryos, 5 nM of P32-end-labelled siRNA and 1x lysis buffer containing 10% v/v of glycerol in a total volume of 10 µL. Reactions were incubated for 30 min to allow silencing complex formation. Native gel electrophoresis for separation of silencing complexes was essentially as described previously [31]. The in vitro reaction mixtures were diluted with 10 µL of loading buffer (1x lysis buffer, 6% Ficoll 400), and part of the sample was analyzed on a 3.9% (39:1 acrylamide-bisacrylamide) native acrylamide gel in 1xTBE buffer. The gels were dried and exposed to a storage phosphor screen.

**Ago2 sequence comparisons**

For various species, Ago2 family members were identified by searching FlyBase annotations or GenBank entries. For Ago2 proteins from *Drosophila* species, most candidates showed a glutamine-rich NTD. For those candidates that did not, we examined the eight distinct gene prediction algorithms from which the FlyBase consensus prediction is derived [40]. In several additional cases, at least one of these algorithms predicted a glutamine-rich NTD (details provided in Fig. S4). DUF1785, PAZ and PIWI domains were estimated using the Conserved Domain Search at NCBI. To estimate the extent of the NTD, the candidates were aligned with human Ago2 using the “Align two sequences by using BLAST” function at NCBI; the NTD was defined as those sequences amino-terminal to where significant alignment was detected. Similar results are obtained when candidates are aligned to any of the other human Argonaute family members (not shown). For human Ago2, the NTD was defined by aligning to *D. melanogaster* Ago2. For the phylogram in Fig. 7, we conducted multiple alignment of the protein sequences, neighbor-joining tree building and 1000 bootstrap trials via CLUSTALX. In similar phylogenoms based on all *Drosophila* Argonaute family members (not shown), the branch lengths for Ago2s are much longer than for Ago1s, consistent with the observation that Ago2 evolves much more rapidly than Ago1 [36]. These increased branch lengths hold true even if the NTD sequences of Ago2 are omitted from the analysis.

**Supporting Information**

*Figure S1* Amino acid sequence of GR2R repeats in various *D. melanogaster* strains. GR2R repeats display slight sequence variations, indicated by different colors. Fly strains differ both in number and type of repeat. The 3CP122 pattern is identical to the one in the current FlyBase annotation for Ago2, and was also found in strains In(1)AB, w-14 melbourne, red e, and dop46. Note that the same overall repeat copy number can be achieved via distinct primary sequences (compare left and right columns below). This variability in sequence likely results in even greater number of distinct haplotypes, and the ten haplotypes in Table 1, which are based solely on repeat copy number, probably underestimate the true variability among the 32 strains surveyed. (PDF)

*Figure S2* Variability in the Ago2 NTD for *Drosophila simulans, mauritiana, and sechellia* strains. The genomic DNA corresponding to exon 3 of *ago2* was sequenced in a number of strains of the *D. simulans* species complex. The amino acid sequences for the corresponding NTD stretches were compared. A) Nine *D. simulans* and one *D. mauritiana* strain displayed highly similar sequences that could easily be aligned. Details on these strains are given in materials and methods. A total of six different haplotypes differ from each other by small insertions and deletions as well as single amino-acid changes. Strains sim2 and sim3 had the same sequence as strain Oxnard; strains maz1 and C167.4 had the same sequence as maz6. B) Comparison of *D. simulans* (strain maz6 is shown) and *D. sechellia* (two strains analyzed) sequences shows that they are composed of highly related stretches (indicated in red, green, and yellow) whose overall arrangement differs. In *D. simulans*, there is an imperfect, tandem repeat of the red + green + yellow stretches (118 or 121 aa long). In *D. sechellia*, a single copy of the red stretch is followed by three repeats (15 aa long) of the green stretch, and then one copy of the yellow stretch. (PDF)

*Figure S3* Ago2 from *Drosophila melanogaster*, *simulans* and *sechellia* are dramatically variable in the NTD, but highly similar in the rest of the protein. Ago2 from *D.
Figure S4 Sequences of Ago2 family members in insects and in Daphnia pulex
For various species, Ago2 family members were identified by searching FlyBase annotations or GenBank entries. In a few cases, predicted proteins were corrected based on cDNA or after resequencing problematic genomic regions; if alternative predictions for the protein included a glutamine-rich NTD, those predictions are listed below. In these sequences, four domains were identified: NTD (light blue), DUF1785 (pink), PAZ (green) and Piwi (yellow). The latter three domains were estimated using the Conserved Domain Search at NCBI. The extent of the NTD was based on comparison to human Ago2 (see Materials and Methods). Glutamine residues are highlighted in red and bold. For D. pseudoobscura, D. persimilis, and D. willistoni, FlyBase annotations suggest the existence of multiple Ago2 paralogs. Sequences of these putative paralogs were compared to ensure that they did not represent alternative isoforms of the same gene. For annotations that showed only minor differences from each other, only one such annotation was included in the list below. The phylogram in Fig. 7 suggests that the thus chosen candidates represent bona fide paralogs. Dsime\Ag2 Drosophila simulans NTD sequence based on our cDNA sequencing; rest of sequence based on Genbank entry EDX10559.1. The FlyBase annotation for this gene (GD14553) is largely identical across the NTD, DUF and PAZ domains, but predicts a deletion of much of the Piwi domain. Dsec\Ag2 Drosophila sechellia NTD sequence based on our cDNA sequencing; rest of sequence based on combining FlyBase annotations GM25537 and GM25538 and resequencing the genomic DNA. Dwil\Ag2 Drosophila willistoni is largely identical across the NTD, DUF and PAZ domains, but major differences from each other, only one such annotation was included in the list below. The phylogram in Fig. 7 suggests that the thus chosen candidates represent bona fide paralogs. Dspe\Ag2 Drosophila pseudobscura Dvi\Ag2 Drosophila virilis Based on alternative NSCAN prediction for FlyBase annotation XP_001607164.1. Bmor\Ag2 Bombyx mori (domestic silkworm) GenBank entry NP_001107842 (Argonaute-2a). Tcas\Ag2b Tribolium castaneum (red flour beetle) GenBank entry NP_001107828 (Argonaute-2b). Apis\Ag2 A. mellifera (honey bee) GenBank entry XP_001945621. Plm\Ag2 P. americana (human louse) GenBank entry NP_001041714. Hsap\Ag2 Homo sapiens GenBank entry NP_036290.

Table S1 Evidence for expression of Ago2 NTDs in various insect species. For various insect Ago2s, the genomic regions corresponding to the NTD are indeed transcribed. Evidence includes cDNA analysis of Ago2 specifically, EST data from high-throughput sequencing efforts, and RNAseq data. If the transcribed region encompasses the entire NTD, this is indicated as “full” in the table.

Acknowledgments
We thank the Drosophila stock centers in Bloomington (Indiana, USA), Tuscon (Arizona, USA) and the Developmental Studies Hybridoma Bank (Iowa, USA) and H. Holocher, B. Lazzaro, and L. Griffith for fly stocks and antibodies, respectively. We thank T. Volkmann, R. Webster, N. Levin and N. Rizzo for expert technical assistance, and Y. Guo and D.-H. Kim for help with cDNA cloning.

Author Contributions

References


