Variable expression levels detected in the Drosophila effectors of piRNA biogenesis

Marie Fablet, Abdou Akkouche, Virginie Braman, Cristina Vieira

Title

Variable expression levels detected in the *Drosophila* effectors of piRNA biogenesis

Authors and Affiliations

Marie Fablet[a], Abdou Akkouche[a,1], Virginie Braman[a], Cristina Vieira[a,2]

[a]Université de Lyon, F-69000, Lyon; Université Lyon 1; CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, F-69622, Villeurbanne, France

[1]Present address : Institut de Génétique Humaine, CNRS, UPR1142; 141 rue de la Cardonille, F-34396, Montpellier, France

[2]Institut Universitaire de France

Corresponding authors:

Cristina Vieira

0033 4 72 44 81 98

cristina.vieira@univ-lyon1.fr

Université de Lyon, F-69000, Lyon; Université Lyon 1; CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, 43 bv 11 novembre 1918, F-69622, Villeurbanne, France

Abstract

piRNAs (piwi-interacting RNAs) are a class of small interfering RNAs that play a major role in the regulation of transposable elements (TEs) in *Drosophila* and are considered of fundamental importance in gonadal development. Genes encoding the effectors of the piRNA machinery are thus often though to be highly constrained. On the contrary, as actors of genetic immunity, these genes have also been shown to evolve rapidly and display a high level of sequence variability. In order to assess the support for these competing models, we analyzed seven genes of the piRNA pathway using a collection of wild-type strains of *Drosophila simulans*, which are known to display significant variability in their TE content between strains. We showed that these genes exhibited wide variation in transcript levels, and we discuss some evolutionary considerations regarding the observed variability in TE copy numbers.

Keywords

transposable element;

retrotransposon;

piwi interacting small RNA;

Drosophila simulans;

wild-type strains;

1. Introduction

In the last decade, major advances were made in our understanding of the epigenetic control of transposable elements (TEs), particularly regarding small RNAs (Saito and Siomi 2010; Senti and Brennecke 2010; Siomi et al. 2011). RNA interference is a widespread phenomenon, and the origin of its effectors dates back to the common ancestor of eukaryotes (Cerutti and Casas-Mollano 2006). Several classes of small interfering RNAs were described, including piRNAs (piwi-interacting RNAs), which are the major regulators of TEs in *Drosophila* (Vagin et al. 2006; Pélisson et al. 2007; Chambeyron et al. 2008). In this study, we will refer to genes involved in the piRNA pathway as GIPPs. Analyses in *Drosophila melanogaster* revealed that mutations in GIPPs led to TE up-regulation (Vagin et al. 2004; Kalmykova et al. 2005; Le Thomas et al. 2013), causing abnormalities in germline development (Schüpbach and Wieschaus 1991; González-Reyes et al. 1997; Cox et al. 1998; Cook et al. 2004; Pane et al. 2007; Li et al. 2009). The piRNA machinery is therefore considered a guardian of genome stability (Senti and Brennecke 2010). In addition, its effectors seem to be involved in many other biological processes, such as splicing and DNA repair (Meister 2013). Because of their biological relevance, GIPPs appear highly constrained and are described as conserved (Meister 2013). We will refer to this as the development-like model.

In contrast, defense against TEs can be viewed as an immunological process, taking place at the genomic scale. Within this framework, evolutionary analyses of GIPPs revealed that they have been recurrently subject to positive selection, as is frequently observed for genes involved in immunity in a broad sense (Obbard et al. 2009a, 2009b; Kolaczkowski et al. 2011). As a consequence, GIPPs belong to the most rapidly evolving known coding sequences (Obbard et al. 2009a, 2009b). We will refer to this as the immunity-like model.

In order to address these apparently contradictory predictions – conservation (based on the development-like model) or rapid evolution (based on the immunity-like model), we used a collection of wild-type strains of *Drosophila simulans*, which are known to display variable TE contents between strains (Vieira et al. 1999; Biémont et al. 2003). We analyzed seven GIPPs: *ago3*, *aub*, and *piwi*, which are the direct effectors of the slicing step; the helicases *spindle E* (*spnE*) and *armitage* (*armi*); and the nucleases *zucchini* (*zuc*) and *squash* (*squ*). Although these genes were extensively studied at the DNA sequence level, a comparative analysis of their expression levels was never performed so far. Based on the development-like model, which suggests that GIPPs are highly constrained, their expression levels are expected to be conserved between strains. On the contrary, genes with the highest rates of sequence polymorphism are known to display the highest variability in expression levels (Nuzhdin et al. 2004; Lemos et al. 2005; Lawniczak et al. 2008). Thus, unlike the above expectation, based on the immunity-like model, GIPPs are predicted to present variable levels of expression between strains.

To test these opposing predictions, we quantified nucleotide polymorphism and present data on allozyme profiles for seven GIPPs in *D. simulans*. Further, we examined the variation in transcript level for the same GIPPs among 13 strains of *D. simulans*. Our data suggest that there is a high level of variation among strains, which supports the immunity-like model of evolution for GIPPs in *Drosophila*. We propose some evolutionary considerations regarding the associated variable TE contents of the strains.

2. Materials and Methods

2.1. Drosophila stocks

We used wild-type strains of *D. simulans* which originated in Kenya (Makindu), Zimbabwe, Indian Ocean islands (Madagascar, Mayotte, Reunion), Atlantic Ocean islands (Madeira), Portugal (Chicharo), Russia (Moscow), Australia (Canberra, Eden, CannRiver), New Caledonia (Amieu, Noumea) and French Polynesia (Papeete). For the McDonald-Kreitman tests (MK tests), we also used two strains of *D. melanogaster* collected from Senegal and Portugal (Chicharo).

2.2. Sequence library for genes involved in the piRNA pathway

To focus exclusively on coding regions, we amplified sequences from cDNAs, which are devoid of introns. We extracted total RNAs from five adult females from each strain. PCR products were subsequently obtained from cDNAs, isolated using bacterial cloning and sequenced (Sanger sequencing). See supplementary material 1 for GenBank accession numbers.

The obtained sequences were translated and amino-acid sequences were analyzed using BLOSUM62 matrix scores (Henikoff & Henikoff 1992). We considered substitutions with negative scores to belong to distinct allozymes.

2.3. Transcription level measurements

Twenty-five pairs of ovaries from two-four-day-old females were dissected in PBS. Total RNAs were extracted using the RNeasy Kit (Qiagen). cDNAs were produced using the ThermoScript RT-PCR system (Invitrogen) and oligo(dT) primers. The cDNAs were diluted

50-fold and quantified using SYBR Green 1 mix in a LightCycler 480 (Roche Diagnostics) using primers specific to each gene (Supplementary Material 2). Primers were designed in portions of sequences that were conserved between the variants we isolated. The transcript amounts were estimated relative to the amounts of the *rp49* gene, which showed the lowest variation among reference genes. The measurements were performed in three independent experiments.

3. Results and Discussion

3.1. Sequences of GIPPs are variable between wild-type strains

Evolutionary studies of genes involved in immunity, particularly in the defense against viruses and TEs, have revealed that these are the most rapidly evolving genes in the genome and are repeatedly subject to positive selection (Obbard et al. 2009b; Kolaczkowski et al. 2011). Kolaczkowski et al. (2011) reported strong evidence of the adaptive evolution of *spnE* in *D. melanogaster*. They also identified evidence of adaptations in *aub* and *zuc*, but not in *piwi*. Obbard et al. (2009a) found significant deviations from neutrality for *aub* and *armi*. Our intention in this study was not to redo these analyses, however, we know that conclusions of neutrality tests depend on the sample of strains used. Therefore we tested whether data from our sample also led to neutrality rejection. For this purpose, we sequenced exonic portions of the genes in five wild-type strains of *D. simulans* and two wild-type strains of *D. melanogaster* and performed McDonald-Kreitman tests (MK tests). These analyses provided results congruent with the immunity-like model of sequence evolution (MK tests were significant for *armi*, *aub* and *spnE*) (Supplementary Material 3). This is also illustrated by the non-synonymous nucleotide diversity which was significantly larger for GIPPs than for the *alpha-tubulin at 84B* (*α-tub*) reference gene, whereas the synonymous nucleotide diversities were in the same range (fig. 1). Again, this is in agreement with the immunity-like model for GIPPs. However, we cannot exclude the hypothesis of greater tolerance to segregating mildly deleterious mutations for GIPPs. In any case, sequence variability is higher for GIPPs.

The above sequences were translated *in silico*. They corresponded to distinct amino-acid sequences that could be clustered based on chemical profiles. The different clusters are

referred to as allozymes (Supplementary Material 4). We could not find allozymes associated with particularly high (or low) expression level of GIPPs nor could we find allozymes associated with particularly high (or low) TE content. There were no obvious strain-specific associations between the allozymes of the different GIPPs, which led us to conclude that each strain had a unique combination of allelic variants of all seven genes.

3.2. Transcription levels of GIPPs are variable between wild-type strains

We quantified transcript levels in ovaries using RT-qPCR. We enlarged our sample to a panel of 13 wild-type strains of *D. simulans* and observed significant variation in the transcript levels of all GIPPs (fig. 2). We performed the same experiment on four housekeeping genes that are often used as reference genes in expression experiments : *18S*, *adh*, *α-tub* and *CG13919*, and found lower variation. The coefficients of variation (square root of the variance divided by the mean) were significantly different between both categories of genes (Wilcoxon test, p-value = 0.012) (fig. 2). We also performed ANOVA1 analyses to compute η^2 coefficients, which account for the amount of variability explained by the strains (η^2 equals sum of squares between groups divided by total sum of squares). ANOVAs were significant for all GIPPs, with high η^2 values (fig. 2). These results indicate that GIPPs are significantly more variable in their transcript levels compared to reference genes. The transcription data from GIPPs were used to compute distance matrices among the strains. No significant correlations were established between these matrices and the matrices of geographic distances (Mantel test, p-value = 0.55), suggesting the absence of geographical patterns such as isolation by distance.

3.3. Evolutionary considerations regarding TE copy numbers

In *D. melanogaster* laboratory strains mutated for GIPPs, a global up-regulation of TEs is observed, leading to a high rate of mutation and aberrant phenotypes (Klattenhoff and Theurkauf 2008). However, the extent of TE up-regulation depends on the nature of the mutant allele. For instance, Lu et Clark (2010) present data on two distinct mutant alleles of *aub*, aub^{QC42} and aub^{HN}, which exhibit wide differences in fold changes of TE mRNAs for 32 families (bias in TE copy numbers was excluded by the authors). Therefore we propose to test the hypothesis that the wild-type variability in GIPPs revealed by this study is associated with variability in TE activity and thus in TE content.

Based on *in situ* hybridization studies on polytene chromosomes, we know that 25 families of TEs exhibit high variability in euchromatic copy numbers between wild-type strains of *D. simulans* (Vieira et al. 1999; Biémont et al. 2003). We tested whether the transcriptional activity of GIPPs was correlated to the total number of TE copies per genome, and found no cases in which it was significant at the 0.05 level (Spearman correlation tests) (Supplementary Material 5). We then performed the test for each TE family separately, and found a few significant cases (Spearman correlation tests, see Supplementary Material 5). However, they probably correspond to noisy false positives (Supplementary Material 5).

In addition, the total number of TE insertions appeared to increase with the distance from East Africa, the cradle of the species (Lachaise et al. 1988) (Spearman correlation test, p-value = 0.013). Therefore, we looked for a relationship between transcription levels of GIPPs and distance from Africa. We could not find significant correlations between the distance from East Africa and the transcriptional activity of any of the tested GIPPs (Spearman correlation tests) (Supplementary Material 5).

In conclusion, the variability in the piRNA pathway is higher than predicted by the

development-like model, and appears to be independent of TE copy number. Also, it appears to be independent of the colonization history of the species, since no association was detected with the geographic distances. Further, Dowell et al. (2010) demonstrated in yeast that genes that appeared to be essential in a well-studied genetic background were no longer so in another. Similarly, variability in genetic architecture may be the reason why expression levels of GIPPs are allowed to vary in wild-type strains of *D. simulans* without observable fitness decrease. In addition, some interplay and compensation may exist with other pathways, such as siRNAs and endo-siRNAs (reviewed in Malone and Hannon 2009).

A last consideration is that differences in time scale may explain the observed independence between GIPP transcription levels and TE copy numbers. Indeed, *in situ* hybridization on polytene chromosomes detects copies that reflect relatively ancient activity which persist in the host genome, while the piRNA pathway allows for the short-term regulation of active TEs. When a new active TE copy enters a naive genome, regulation *via* piRNAs starts as soon as the TE inserts into a piRNA cluster (Khurana et al. 2011). As time goes by, TE sequences start to degenerate, so even if the piRNA control terminates, no transposition can occur and the considered TE family is no longer harmful for the host genome. Therefore, variability in the efficiency of the piRNA machinery is expected to lead to variability only in the first steps of TE invasion but may not interfere with older TE insertions. Indeed, Kelleher and Barbash (2013) recently demonstrated in *D. melanogaster* that piRNA-mediated silencing is particularly robust for recently active TE families.

4. Conclusion

Each one of our wild-type strains captured a subset of the variation that can be found in nature. We may envision that, as illustrated by certain of our wild-type strains, individuals

in nature bearing a less efficient combination of GIPPs (either due to less efficient protein sequences or to too low expression levels) will suffer from TE up-regulation in the germline, which can end up in TE copy number increase. We can expect that if the increase in TE copy number is too high, it will be selected against. If TE up-regulation and transposition are too strong, selection will favor more efficient GIPP alleles or expression levels. Past TE copy number increases are still visible in the genome until deletion removes TE insertions. On the contrary, past versions of the piRNA pathway leave no trace in our strains : according to rapid evolution, GIPP combinations that have allowed moderate TE transposition are replaced with more efficient alleles. This may be the reason why we cannot detect correlations between TE abundance and GIPP status : we lack the data on GIPP status at the very precise time of TE copy number increase.

Acknowledgements

We thank Sarah Schaack, Matthieu Boulesteix, Rita Rebollo and Christian Biémont for their helpful comments and careful reading of the manuscript, and we thank Séverine Chambeyron and the members of the Vieira group for useful discussions. We thank the DTAMB (FR41, University Lyon 1) for the qPCR facility.

This work was supported by the Centre National de la Recherche Scientifique (Agence Nationale de la Recherche grant GENEMOBILE), the Institut Universitaire de France, and the Region Rhone Alpes CIBLE 2008.

Figure 1. Box plot of the nucleotide diversity (π) of GIPPs in *D. simulans*.

The non-synonymous (N) and synonymous (S) nucleotide diversity (π) values calculated for seven GIPPs in wild-type strains of *D. simulans* are illustrated. The values obtained for *α-tub* are plotted in black circles.

Figure 2. Transcript levels in the ovaries of the *D. simulans* wild-type strains.

Vertical axes represent the enrichment relative to *rp49* transcripts. (A and C) Coefficients of variation (CV, square root of the variance divided by the mean) and η^2 coefficients from ANOVA analysis of strain effect are indicated for each gene (η^2 equals sum of squares between groups divided by total sum of squares). NS: non significant strain effect. Error bars correspond to standard deviations calculated on three independent biological replicates. (A) GIPPs: *ago3*, *armi*, *aub*, *piwi*, *spnE*, *squ* and *zuc*; (C) reference genes: *18S*, *adh*, *α-tub* and *CG13919*. (B) Distributions of expression level of GIPPs among strains.

References

Biémont, C. et al., 2003. Worldwide distribution of transposable element copy number in natural populations of Drosophila simulans. *Evolution; International Journal of Organic Evolution*, 57(1), p.159-167.

Cerutti, H. & Casas-Mollano, J.A., 2006. On the origin and functions of RNA-mediated silencing: from protists to man. *Current Genetics*, 50(2), p.81-99.

Chambeyron, S. et al., 2008. piRNA-mediated nuclear accumulation of retrotransposon transcripts in the Drosophila female germline. *Proceedings of the National Academy of Sciences of the United States of America*, 105(39), p.14964-14969.

Cook, H.A. et al., 2004. The Drosophila SDE3 homolog armitage is required for oskar mRNA silencing and embryonic axis specification. *Cell*, 116(6), p.817-829.

Cox, D.N. et al., 1998. A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. *Genes & development*, 12(23), p.3715-3727.

Dowell, R.D. et al., 2010. Genotype to phenotype: a complex problem. *Science (New York, N.Y.)*, 328(5977), p.469.

González-Reyes, A., Elliott, H. & St Johnston, D., 1997. Oocyte determination and the origin of polarity in Drosophila: the role of the spindle genes. *Development (Cambridge, England)*, 124(24), p.4927-4937.

Henikoff, S. & Henikoff, J.G., 1992. Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences of the United States of America*, 89(22), p.10915-10919.

Kalmykova, A.I., Klenov, M.S. & Gvozdev, V.A., 2005. Argonaute protein PIWI controls mobilization of retrotransposons in the Drosophila male germline. *Nucleic acids research*, 33(6), p.2052-2059.

Kelleher, E.S. & Barbash, D.A., 2013. Analysis of piRNA-mediated silencing of active TEs in Drosophila melanogaster suggests limits on the evolution of host genome defense. *Molecular biology and evolution*, 30(8), p.1816-1829.

Khurana, J.S. et al., 2011. Adaptation to P element transposon invasion in Drosophila melanogaster. *Cell*, 147(7), p.1551-1563.

Klattenhoff, C. & Theurkauf, W., 2008. Biogenesis and germline functions of piRNAs. *Development (Cambridge, England)*, 135(1), p.3-9.

Kolaczkowski, B., Hupalo, D.N. & Kern, A.D., 2011. Recurrent adaptation in RNA interference genes across the Drosophila phylogeny. *Molecular Biology and Evolution*, 28(2), p.1033-1042.

Lachaise, D. et al., 1988. Historical biogeography of the Drosophila melanogaster species subgroup. , 22, p.159-225.

Lawniczak, M.K.N. et al., 2008. Genomic analysis of the relationship between gene expression variation and DNA polymorphism in Drosophila simulans. *Genome biology*, 9(8), p.R125.

Lemos, B. et al., 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. *Molecular biology and evolution*, 22(5), p.1345-1354.

Li, C. et al., 2009. Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell*, 137(3), p.509-521.

Lu, J. & Clark, A.G., 2010. Population dynamics of PIWI-interacting RNAs (piRNAs) and their targets in Drosophila. *Genome research*, 20(2), p.212-227.

Malone, C.D. & Hannon, G.J., 2009. Small RNAs as guardians of the genome. *Cell*, 136(4), p.656-668.

Meister, G., 2013. Argonaute proteins: functional insights and emerging roles. *Nature reviews. Genetics*, 14(7), p.447-459.

Nuzhdin, S.V. et al., 2004. Common pattern of evolution of gene expression level and protein sequence in Drosophila. *Molecular biology and evolution*, 21(7), p.1308-1317.

Obbard, D.J., Welch, J.J., et al., 2009. Quantifying adaptive evolution in the Drosophila immune system. *PLoS Genetics*, 5(10), p.e1000698.

Obbard, D.J., Gordon, K.H.J., et al., 2009. The evolution of RNAi as a defence against viruses and transposable elements. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 364(1513), p.99-115.

Olivieri, D. et al., 2010. An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in Drosophila. *The EMBO journal*, 29(19), p.3301-3317.

Pane, A., Wehr, K. & Schüpbach, T., 2007. zucchini and squash encode two putative nucleases required for rasiRNA production in the Drosophila germline. *Developmental cell*, 12(6), p.851-862.

Pélisson, A. et al., 2007. A novel repeat-associated small interfering RNA-mediated silencing pathway downregulates complementary sense gypsy transcripts in somatic cells of the Drosophila ovary. *Journal of virology*, 81(4), p.1951-1960.

Saito, K. & Siomi, M.C., 2010. Small RNA-mediated quiescence of transposable elements in animals. *Developmental cell*, 19(5), p.687-697.

Schüpbach, T. & Wieschaus, E., 1991. Female sterile mutations on the second chromosome of

Drosophila melanogaster. II. Mutations blocking oogenesis or altering egg morphology. *Genetics*, 129(4), p.1119-1136.

Senti, K.-A. & Brennecke, J., 2010. The piRNA pathway: a fly's perspective on the guardian of the genome. *Trends in genetics: TIG*, 26(12), p.499-509.

Siomi, M.C. et al., 2011. PIWI-interacting small RNAs: the vanguard of genome defence. *Nature reviews. Molecular cell biology*, 12(4), p.246-258.

Le Thomas, A. et al., 2013. Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes & development*, 27(4), p.390-399.

Vagin, V.V. et al., 2006. A distinct small RNA pathway silences selfish genetic elements in the germline. *Science (New York, N.Y.)*, 313(5785), p.320-324.

Vagin, V.V. et al., 2004. The RNA interference proteins and vasa locus are involved in the silencing of retrotransposons in the female germline of Drosophila melanogaster. *RNA biology*, 1(1), p.54-58.

Vieira, C. et al., 1999. Wake up of transposable elements following Drosophila simulans worldwide colonization. *Molecular Biology and Evolution*, 16(9), p.1251-1255.

Figure 1

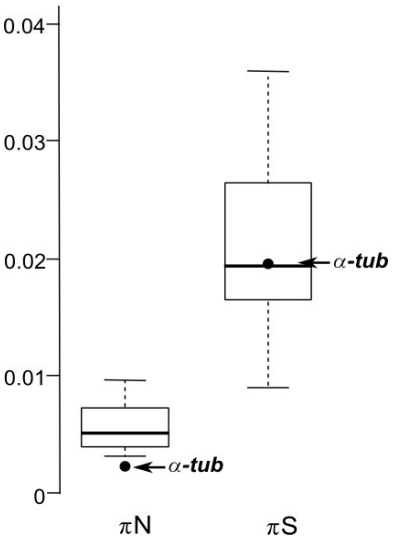

Figure 2