



HAL
open science

Distribution, evolution, and diversity of retrotransposons at the flamenco locus reflect the regulatory properties of piRNA clusters

Vanessa Zanni, Angéline Eymery, Michael Coiffet, Matthias Zytnicki, Isabelle Luyten, Hadi Quesneville, Chantal Vaury, Silke Jensen

► To cite this version:

Vanessa Zanni, Angéline Eymery, Michael Coiffet, Matthias Zytnicki, Isabelle Luyten, et al.. Distribution, evolution, and diversity of retrotransposons at the flamenco locus reflect the regulatory properties of piRNA clusters. Proceedings of the National Academy of Sciences of the United States of America, 2013, 10.1073/pnas.1313677110 . hal-01189713

HAL Id: hal-01189713

<https://hal.science/hal-01189713>

Submitted on 29 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distribution, evolution, and diversity of retrotransposons at the *flamenco* locus reflect the regulatory properties of piRNA clusters

Vanessa Zanni^{a,b,c,1}, Angéline Eymery^{a,b,c,2}, Michael Coiffet^{a,b,c,3}, Matthias Zytnicki^d, Isabelle Luyten^d, Hadi Quesneville^d, Chantal Vaury^{a,b,c,4}, and Silke Jensen^{a,b,c,4}

^aLaboratoire Génétique, Reproduction, et Développement, Clermont Université, Université d'Auvergne, 63001 Clermont-Ferrand, France; ^bInstitut National de la Santé et de la Recherche Médicale, Unité 1103, 63001 Clermont-Ferrand, France; ^cCentre National de la Recherche Scientifique, Unité Mixte de Recherche 6293, 63001 Clermont-Ferrand, France; and ^dUnité de Recherche en Génomique-Info, Institut National de la Recherche Agronomique, Unité de Recherche 1164, 78026 Versailles Cedex, France

Edited by Jeffrey L. Bennetzen, University of Georgia, Athens, Georgia, and approved October 4, 2013 (received for review July 25, 2013)

Most of our understanding of *Drosophila* heterochromatin structure and evolution has come from the annotation of heterochromatin from the isogenic *y; cn bw sp* strain. However, almost nothing is known about the heterochromatin's structural dynamics and evolution. Here, we focus on a 180-kb heterochromatic locus producing Piwi-interacting RNAs (piRNA cluster), the *flamenco* (*flam*) locus, known to be responsible for the control of at least three transposable elements (TEs). We report its detailed structure in three different *Drosophila* lines chosen according to their capacity to repress or not to repress the expression of two retrotransposons named *ZAM* and *Idefix*, and we show that they display high structural diversity. Numerous rearrangements due to homologous and nonhomologous recombination, deletions and segmental duplications, and loss and gain of TEs are diverse sources of active genomic variation at this locus. Notably, we evidence a correlation between the presence of *ZAM* and *Idefix* in this piRNA cluster and their silencing. They are absent from *flam* in the strain where they are derepressed. We show that, unexpectedly, more than half of the *flam* locus results from recent TE insertions and that most of the elements concerned are prone to horizontal transfer between species of the *melanogaster* subgroup. We build a model showing how such high and constant dynamics of a piRNA master locus open the way to continual emergence of new patterns of piRNA biogenesis leading to changes in the level of transposition control.

RNAi | gene silencing | epigenetics

Over the course of evolution, transposable elements (TEs) have accumulated in the genomes of eukaryotes, where they can account for up to 85% of the DNA (1). Most of these sequences have lost their ability to transpose. They are now stable components of the genomes. Their conservation throughout evolution suggests that they may confer advantageous effects to their hosts. However, transposition of the copies that remain functional could generate deleterious mutations if they were not severely repressed by their host. RNAi, which is a gene-silencing mechanism triggered by small RNAs (reviewed in ref. 2), has been identified as being the main cellular machinery involved in the “taming” of TEs (reviewed in refs. 3–5). RNAi pathways involve small RNAs of diverse families. Among them, Piwi-interacting RNAs (piRNAs) have been shown to be involved in TE silencing in the *Drosophila* ovary. These piRNAs, 23–29 nt long, are bound by the Argonaute proteins Piwi, Argonaute 3, or Aubergine. They are produced by discrete genomic loci named piRNA clusters, which have been described as containing vestiges of TEs (6). One of these loci, the *flamenco* (*flam*) locus, extends over 180 kilobases (kb) on the *Drosophila* X chromosome. It is proximal to the DISCO interacting protein 1 gene (*DIP1*) and close to pericentromeric heterochromatin. Before the identification of piRNAs, this locus had been shown to regulate the *Gypsy* retrotransposon (7, 8). Desset et al. (9) identified a locus called

COM responsible for the control of *ZAM* and *Idefix* retrotransposons. Later, *flam* and *COM* were shown to be one and the same large master locus for regulation of at least these three retrotransposons: *ZAM*, *Idefix*, and *Gypsy* (reviewed in ref. 10). Control of different retrotransposons by the *flam* locus had been shown for diverse *flam* alleles: for *ZAM* and *Idefix* in the Rev line (9) and for *Gypsy* and *ZAM* in *flam* mutant lines (11). Brennecke et al. (6) showed the potential for the *flam* cluster to produce 79% of all ovarian piRNAs that target *ZAM* and 30% and 33% of those matching *Idefix* and *Gypsy*, respectively. It was further reported that a functional pathway links *flam* piRNA to *Gypsy* suppression. Indeed, a substantial reduction of piRNA homologous to *Gypsy* was observed in *flam* mutant lines that are permissive to *Gypsy* transposition (12). Still, the molecular rules allowing this locus to control several retrotransposons are far from being understood.

In previous work, we isolated *Drosophila* lines in which *ZAM* and *Idefix* are either silenced (i.e., “stable”) or derepressed (i.e., “unstable”). Over the course of a *P-element* mediated mutagenesis

Significance

Control of transposable elements (TEs) by RNAi has a large impact on genome evolution in higher eucaryotes. In this paper, we study in detail a Piwi-interacting RNA (piRNA)-producing locus of *Drosophila melanogaster*, *flamenco* (*flam*), known to be responsible for the control of at least three retrotransposons by RNAi. We demonstrate the high structural dynamics of the *flam* locus resulting in loss and gain of TEs and establish a link between such structural variations and its ability to silence retrotransposons. We show that *flam* is a trap for TEs coming in by horizontal transfer from other *Drosophila* species. Overall, our data give unique insights into piRNA cluster regulatory properties, their role in evolution, and expansion and taming of TEs.

Author contributions: H.Q., C.V., and S.J. designed research; V.Z., A.E., M.C., M.Z., I.L., and S.J. performed research; C.V. and S.J. analyzed data; and C.V. and S.J. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. [KF364662](https://doi.org/10.1093/seqs/kf364662)–[KF364665](https://doi.org/10.1093/seqs/kf364665) and [KF410639](https://doi.org/10.1093/seqs/kf410639)–[KF410642](https://doi.org/10.1093/seqs/kf410642)).

¹Present address: Institut Jean-Pierre Bourgin, Institut National de la Recherche Agronomique, Unité Mixte de Recherche 1318, 78026 Versailles Cedex, France.

²Present address: Friedrich Miescher Institute for Biomedical Research, 4058 Basel, Switzerland.

³Present address: Institut de Biochimie et Génétique Cellulaires, Centre National de la Recherche Scientifique, Unité Mixte de Recherche 5095, 33077 Bordeaux Cedex, France.

⁴To whom correspondence may be addressed. E-mail: chantal.vaury@udamail.fr or silke.jensen@udamail.fr.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1313677110/-DCSupplemental.

performed on a stable line called w^{IR6} , we isolated a line, Rev, derived from w^{IR6} , in which *ZAM* and *Idefix* are both derepressed (13–15). Novel mobilizations are regularly observed, and reporter genes containing *ZAM* or *Idefix* fragments are expressed in the Rev line, although silenced in w^{IR6} . We further reported that the observed release in *ZAM* and *Idefix* silencing was due to an unidentified mutation having affected the *flam* locus (9).

Bergman et al. (16) laid out the hypothesis that β -heterochromatic TE nests will act as transposon traps for new TE invasions by horizontal transfer providing an “adaptive immunity” to the host genome. Here, we wondered whether specifically piRNA loci like *flam* might act as a trap for TEs. If so, any trapped TE that inserted into a piRNA cluster should immediately be silenced and further transposition events should not occur. Then, a piRNA locus would be predicted to contain only one copy of a TE per family. To test this model, we first revisit the annotation of the *flam* locus. We show that, as predicted, most of the TEs in the *flam* locus are present only once. More than half of the *flam* locus is made up of recent TEs, and two-thirds of them arose from horizontal transfer, confirming the trap model. We then evaluate the structure of the *flam* region in permissive and repressive strains. Our data indicate that strains permissive for the mobility of *ZAM* and *Idefix* display large deletions that removed *ZAM* and *Idefix* sequences present in *flam*. These results reveal a tight correlation between the presence of these retrotransposons in this locus and their silencing. They also highlight how transposition bursts may occur due to the high molecular dynamics of piRNA clusters like *flam*.

Results and Discussion

Most of the TEs in *flam* Are Present Only Once in This Locus. We predicted that if a TE is silenced as soon as it inserts into the *flam* locus, it should be present only once in *flam* because further transposition events should simply not occur. However, the current annotation [ref. 17 and Flybase (<http://flybase.org>)] appeared to contradict this prediction. Actually, there were several copies annotated for the same TEs (i.e., for *ZAM*, *Idefix*, *Stalker*, *Stalker2*, *Stalker4*, *gypsy6*, *springer*, *mdg1*, *F-element*). Several hypotheses could be made: Either multiple insertions of these TEs occurred, contradicting our prediction, or single insertions occurred, followed by duplication, or single insertions were misrepresented as multiple insertions because of erroneous annotation.

We conducted a bioinformatics analysis of the *flam* locus to check these hypotheses. To generate reannotation the 180-kb sequence of the *flam* locus of the sequenced Iso1A strain was analyzed by CENSOR using Repbase (www.girinst.org) (18, 19). The whole sequence was also aligned to itself, in both orientations, to analyze sequence redundancy. Among the 52 different TEs that are present at this locus and present in Repbase, we found 49 being

present as a unique copy in *flam*, supporting the single-insertion prediction. Only the *F-element*, a long interspersed nucleotidic element (LINE), and the LTR retrotransposons *mdg1* and *Stalker4* are multicopy. Two *F-elements* are present in *flam*, one full-length copy inserted in the Plus orientation (distal to proximal) and a second truncated copy inserted in the Minus orientation (proximal to distal) (Fig. 1 and Table S1). Six distinct *mdg1* copies inserted in both orientations have 97–99.4% sequence identity with the referenced element. These *mdg1* copies are mainly clustered in one site, as is observed in another locus from chromosome 2 (2R:6509912..6524619, *Drosophila melanogaster* genome release 5.1).

We identified segmental duplication events and found that several *mdg1* and *Stalker4* fragments show the same breakpoints as other *mdg1* and *Stalker4* fragments in the locus [Table S1; *mdg1(flam)*4, 5, and 6; *Stalker4(flam)*2, 3, 7, and 8]. Segmental duplications in the *flam* region had also been shown by Bergman et al. (16), who proposed that they play an important role in the genesis of TE-rich regions of the genome.

Another case of sequence redundancy concerns solo-LTRs. Actually, we observed TEs for which, in addition to one full-length copy, a solo-LTR is present elsewhere in the *flam* locus. This is the case for *mdg1* and an element that we later identified as being *Pifo* (see below). The presence of the *Pifo* solo-LTR in addition to a full-length *Pifo* indicates that *Pifo* inserted twice in the *flam* locus. This might indicate that the solo-LTR does not generate enough piRNA to repress *Pifo* transposition.

The presence of several copies in *flam* suggests that the *F-element*, *mdg1*, and *Stalker4* might not be silenced by their presence in *flam*. Actually, even if some of the *mdg1* and *Stalker4* copies appear to originate from segmental duplication, others seem to arise from reiterated insertions. For the remaining 49 TEs present in *flam*, our data confirm the single-insertion prediction supporting that these TEs are candidates to be silenced by their presence in this particular piRNA cluster.

In Silico Analysis of the *flam* Locus Reveals a Large Number of New TEs and Gives a Basis for New Annotation of the Locus. Although the vast majority of the *flam* TEs are single copy in this locus, as shown by our reannotation above, some of them had been previously annotated as if they were present several times in *flam*. This was also the case for *ZAM* and *Idefix* (*flam* annotation is provided in Flybase and in ref. 17). Because we aimed to analyze *ZAM*- and *Idefix*-related elements in *flam* in silenced and derepressed *Drosophila* strains, we first had to elucidate the quality of the existing *flam* annotation. Moreover, annotations offered a glimpse of the *flam* locus as a genomic region highly fragmented and mainly composed of degenerated vestiges of TEs. To verify that this was indeed the case, we first used CENSOR, utilizing Repbase to extract all *flam* sequences that display less than 95% sequence identity with known TEs from *D. melanogaster*, corresponding to

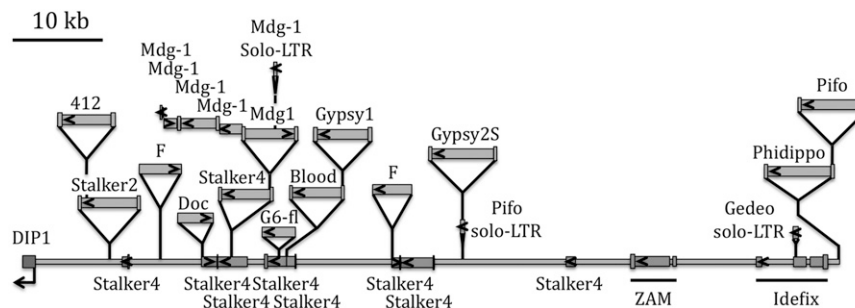


Fig. 1. Structure of the *flam* piRNA cluster in *D. melanogaster*. Recent TEs are presented in the upper part, and ancient elements are shown as the base line of the locus. Only recent elements, *ZAM*, *Idefix*, and the repeated *Stalker4* copies, are shown individually (a detailed analysis of all TEs in *flam* is presented in Table S1). The centromere of the X chromosome is on the right-hand (proximal) side. The sense of transcription for the DIP1 gene and orientation of the TEs are indicated by arrows.

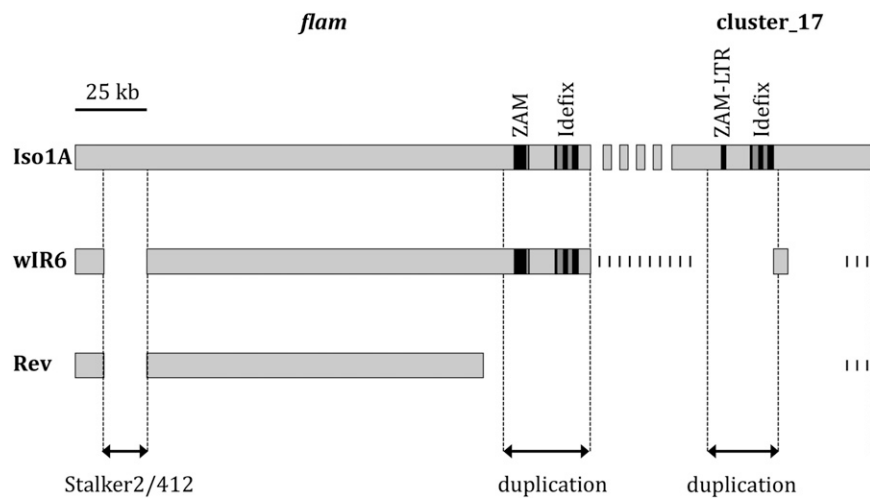


Fig. 2. Proposed structure of *flam* and *cluster_17* in Iso1A, w^{IR6} , and Rev strains. Light gray boxes represent sections of *flam* (Left) and *cluster_17* (Right) that are present in Iso1A, w^{IR6} , or Rev. ZAM and *Idefix* sequences are shown in black, deleted sections are shown as missing zones (white), zones that might be present are shown as short vertical bars, and the segmental duplication of *flam* in *cluster_17* is indicated.

more than one-third of the *flam* locus. With these sequences, we searched for homologous sequences in the genomes of other *Drosophila* species (BLAST at <http://flybase.org>).

By this approach, we identified nine retrotransposons so far undescribed and six for which similar but not identical elements had been described (*Gypsy2S*, *Pifo*, *GYPSY6S*, *mdg1_DSe*, *Gypsy20S_DYa*, and *G6-fl*). These TEs are reported in Table S2 (their sequences are provided in Dataset S1). All have significant coding capacities. They can be found within the *melanogaster* subgroup in species such as *D. melanogaster*, *Drosophila simulans*, *Drosophila sechellia*, and *Drosophila erecta*. However, species outside of this subgroup, such as *Drosophila ananassae*, *Drosophila pseudoobscura*, and *Drosophila persimilis*, are devoid of these TEs. The general map of the *flam* locus is presented in Fig. 1, and three interesting results may be pointed out. First, several regions that were annotated as being fragments of different TEs, or not annotated at all, correspond to one and the same TE in our revised annotation (Tables S1 and S2). As an example, a region that presented some homology with different fragments of *Gypsy10* (92% sequence identity) was revealed to display 99.2% sequence identity over the entire length with an element that we identified in *D. sechellia* (Table S2). We named this element *Phidippo*. Second, a region that was not annotated in Flybase showed more than 99% sequence identity with *Gypsy2* but seemed to be interrupted by an *HMSBEAGLE* fragment (Table S1). This element was revealed to be a full-length, entirely coding *Gypsy2* element that we named *Gypsy2S* (Table S2). Third, sequences that are not annotated in Flybase but had some limited sequence identity with ZAM and *Tirant* fragments correspond in fact to a newly identified element that we named *Agoriino* and a *Pifo* element. *Pifo* was previously identified in *Drosophila yakuba* (20) and a 4,414-bp degenerate *Pifo* in *D. melanogaster* (21). Consequently, only one vestige of ZAM is present at position X:21649838..21655021. The same is true for *Idefix*; a single copy is found in *flam* at position X:21663930..21672608 (Fig. 1 and Table S1).

These results show that most of the sequences composing *flam* are long TE copies and confirm that some elements annotated as being present several times are present only once using our reannotation based on improved query sequences. Moreover, interspecies in silico analysis using *flam* sequence as a starting point allowed the identification of previously undescribed TEs within the *melanogaster* subgroup.

Structural Modifications Affecting *flam* May Explain Differential Silencing of TEs in Permissive and Nonpermissive Strains. We aimed to determine what makes the difference between *Drosophila* strains that are permissive or nonpermissive to the expression of particular retrotransposons, ZAM and *Idefix* in this case. It was already known that ZAM and *Idefix* expression depends on the *flam* locus, but nothing was established on how this works at the molecular level. Three lines were analyzed: one permissive strain, Rev, and two nonpermissive strains, w^{IR6} and the sequenced isogenic *yc; cn bw sp* strain (here referenced as the Iso1A strain).

The genomic structure of *flam* was first examined through PCR amplifications. Primer design was inspired by the work of de La Roche Saint André and Bregliano (22), which showed how to amplify specifically one of many repeated sequences based on one nucleotide difference. Although most of the region of *flam* corresponds to repeated sequences of the *Drosophila* genome, we were able to build a collection of more than 180 primers distributed along the 180-kb locus (primer sequences are provided in Dataset S2). Twenty-five PCR DNA fragments all along the locus were analyzed in the first set of amplifications. When differences were observed, the regions concerned were analyzed in detail by other PCR amplifications and sequenced.

The first difference between the Iso1A strain and both the w^{IR6} and Rev lines was revealed in region X:21511046..21526566. This region encompasses a 412 element inserted within a *Stalker2* element (Fig. 1). These two elements are absent both from w^{IR6} and Rev (Fig. 2). PCR and sequencing of this region revealed that the target site of *Stalker2* and 412 insertion in Iso1A is clearly an empty site in w^{IR6} and Rev [i.e., without any relics of 412 or *Stalker2* sequences (compare with GenBank accession nos. KF364662 and KF364663, respectively)]. This result indicates that *Stalker2* and 412 recently jumped into this position in the Iso1A strain.

Second, two short deletions found around the same region, one in w^{IR6} and one in Rev: positions X:21527735..21527905 were found deleted in w^{IR6} , and positions X:21526698..21526797 are absent in Rev (compare with GenBank accession nos. KF364664 and KF364663, respectively). The fact that each of these deletions exists in one strain but not in the other indicates that they occurred after the isolation of Rev from w^{IR6} some 20 years ago.

Most importantly, the third difference observed in the *flam* structure between the Iso1A, w^{IR6} , and Rev lines pointed out the potential relationship between its structure and its TE-silencing properties. This difference was found within the region

X:21638001..21684449 containing the only *ZAM* and *Idefix* copies present in *flam* as reported above. Both are Minus-oriented. The *ZAM* copy shows significant homology to the active copy, with 94.1% sequence identity extending over 4,786 bp with an internal deletion from positions 4,298 and 6,968 (Rebase *ZAM_I* sequence). Its sequence is interrupted by a *Gypsy4*-like solo-LTR. The *Idefix* copy has 92% sequence identity with canonical *Idefix*. This copy is fragmented by deletion of *Idefix* sequences and foreign insertions (Table S1). All these *Idefix* fragments add up to 3,150 bp distributed over 8,679 bp. PCR amplifications in the region containing *ZAM* and *Idefix* copies were successful in Iso1A and w^{IR6} but did not give any amplified fragments in Rev (Fig. S1). By inverse PCR, we amplified the DNA fragment adjacent to the supposed breakpoint close to position X:21637392. Sequencing of the amplified fragments showed a complex rearrangement in the Rev line, resulting in a large deletion (X:21638367..21684449) containing *ZAM*, *Idefix*, *Phidippo*, and *Pifo* (sequence in GenBank accession nos. KF410639–KF410642). It must be noted that this deletion could be longer in its proximal part, but it was not possible to analyze it further due to the available sequence of *flam* ending in the middle of *Pifo* at position X:21684449. Nevertheless, we were able to extend the analysis to the entire *Pifo* element by amplifying the whole element, and part of the *Phidippo*-LTR in which it is inserted, by PCR in Iso1A and to sequence it (sequence in GenBank accession no. KF364665). Because the large deletion in Rev encompasses the *ZAM* and *Idefix* elements completely, these results reveal a strict correlation between the presence of *ZAM* and *Idefix* within the *flam* locus and their silencing in Iso1A and w^{IR6} strains and, inversely, between their absence from the locus and their derepression in Rev.

During the in silico analysis of the *flam* locus, we found that a 30-kb region in the proximal part is very similar to a region in piRNA cluster_17 (6). Cluster_17 is in an unmapped scaffold (positions U:964336..1041768, with “U” meaning “unmapped”). The similar regions are X:21643521..21673908 in *flam* and U:1003370..1028994 in cluster_17. Our analysis suggests that part of cluster_17 is a segmental duplication of the *flam* region. The main differences between the two regions are the following. In cluster_17, the *ZAM* element is reduced to a solo-LTR, the *Pifo* and *Phidippo* elements are both absent, and the *mdg1*-like solo-LTR in which the *Phidippo* element inserted in *flam* is complete. These differences should be due to recombination between LTRs for *ZAM* and to *Pifo* and *Phidippo* insertions into *flam* that occurred after duplication of the region. The sequence identity of the two regions ranges from 96–98%.

Like *flam*, cluster_17 produces piRNAs in follicle cells and in their derived ovary somatic sheet cells (12, 23), suggesting that it could be involved in *ZAM* and *Idefix* silencing by production of corresponding piRNAs. Thus, it was important to search for any structural modifications of cluster_17 in Rev, where this silencing is lost. PCR amplifications with primer pairs within region U:1010113..1029656, including all *ZAM* and *Idefix* sequences, and in region U:988011..988850 failed to amplify fragments in w^{IR6} and Rev. These results indicate that the *ZAM* and *Idefix* sequences are absent from cluster_17 in w^{IR6} and Rev. PCR amplifications further revealed that this deletion is larger in Rev than in w^{IR6} because a region proximal to *Idefix* (U:1003864..1004568) is present in w^{IR6} (Fig. 2; PCR results are provided in Fig. S2).

In the Rev strain, we observed a deletion of the proximal *flam* region and also a deletion within cluster_17. Thus, we hypothesized that cluster_17 might be close to the *flam* cluster. To test this hypothesis, we used the Df(1)DCB1-35b line (24) containing a large deficiency of the *flam* locus and surrounding regions (covering regions 19F1-h26) to perform a genetic complementation test as in the study by Desset et al. (9). The Df(1)DCB1-35b line was crossed with the w^{IR6} or the Rev line, and with Iso1A for control, to obtain females that are heterozygous for

the 19F1-h26 deficiency and the *flam* allele from respective strains. These females present the same phenotype for *ZAM* and *Idefix* silencing as the parent Iso1A, w^{IR6} , or Rev line (9). Genomic DNA from these females was extracted and tested for the presence of the different regions of cluster_17 as before. These PCR experiments gave the same results as for the Iso1A, w^{IR6} , and Rev strains above (Fig. S2). Actually, all regions of cluster_17 that gave negative PCR results in the w^{IR6} or Rev line also gave negative results in the hybrid progeny from crosses with the Df(1)DCB1-35b line, indicating that the deficiency cannot complement the w^{IR6} or Rev line for cluster_17. These results suggest that cluster_17 is contained in the 19F1-h26 deficiency, thus close to *flam*, and might be localized in the 75-kb non-sequenced region that is proximal to *flam* (Fig. 2).

***flam* Locus Contains More Than 50% of Recent TEs and Is a Trap for TEs Coming in by Horizontal Transfer from Other *Drosophila* Species.**

CENSOR analysis using Rebase shows that 52.7% of the *flam* sequences (in length) have more than 98% sequence identity with reference TEs, including the newly identified *Phidippo* element. This low sequence divergence suggests that these elements inserted recently at the locus. To verify this point, we examined their coding capacity, the identity of LTRs, and target site duplications. As expected for recently transposed TEs, we found that 12 elements in *flam* have long coding capacities (Fig. 1 and Table S3). They are either full length or have a single internal deletion. Three of these elements are LINEs, and nine are LTR-retrotransposons. At least seven of the latter have two identical LTRs (unknown for *Phidippo* because its 5'-end in the very proximal part of *flam* has not been sequenced yet). All of them have target site duplications. Nevertheless, four of them are interrupted by insertion of another full-length TE (*Stalker2*, *Stalker4*, *Blood*, and *Phidippo*). Among all recent insertions, three display a solo-LTR indicating that their insertion has been followed by homologous recombination between the LTRs: *Gedeo*, *mdg1*, and *Pifo*. Taking into account all newly identified TEs, the percentage of recent TEs raises to 55.2% of the total length of the *flam* piRNA cluster (corresponding to 98,668 bp; Table S4). These data illustrate the constant and high dynamics of TE insertions in the locus.

Interestingly, several results indicated horizontal transfer. First, an element having some limited homology with *ZAM* revealed 99.6% sequence identity with an element named *Pifo* from *D. yakuba* for which data in a study by Bartolomé et al. (20) suggested horizontal transfer between this species and *D. melanogaster*. Because only two other *Pifo* elements can be detected in the genome of Iso1A (*Pifo*_Dm B and C in Table S2), in addition to the *Pifo* full-length element and *Pifo* solo-LTR in *flam*, it is likely that *Pifo* has invaded the Iso1A strain recently and was rapidly silenced, possibly after insertion within the *flam* locus.

To evaluate whether horizontal transfer might concern other recent TEs present in *flam*, we determined levels of sequence identity between the *flam* TEs and related elements in other species of the *melanogaster* subgroup. High sequence similarity suggests horizontal transfer (reviewed in ref. 25). Sequence similarity was compared with the one found for *R1*. *R1* is a non-LTR retrotransposon found in many species. It has been shown to evolve at a rate similar to that of nuclear genes in the *melanogaster* subgroup (26). By comparing the sequence identity values of a given TE and *R1*, we can test horizontal vs. vertical transfer. If the sequence identity of the TE is significantly higher than for *R1* in the same species, a recent origin by horizontal transfer is very likely. For eight of the 12 recent full-length TEs identified in the *flam* locus, we found closely related elements in *D. simulans*, *D. sechellia*, *D. yakuba* and *D. erecta* showing more than 98% sequence identity (Table S5). This is significantly higher than the sequence identity expected by vertical transmission. These

Rounds of high transposition rate can trigger genetic instabilities and disease-associated mutations, but there is no doubt that they also play an essential role in the evolution of species. Actually, the current *Drosophila* genome witnesses multiple transposition bursts over time for most of the TE families, resulting in ancient and recent copies being present in the genome (examples from this study are *Blood*, *Stalker2*, *Stalker4*, *Gypsy1*, and *Phidippo*). The case of *Pifo* depicted here is different and certainly represents a case of a new invasion of *D. melanogaster*, because no ancient *Pifo* elements can be found in the genome. Such high dynamicity of piRNA clusters should also remodel heterochromatic regions in other *Drosophila* species. In *D. erecta* and *D. yakuba*, *flam* loci have been shown to contain a large amount of TEs that are completely different from the *D. melanogaster flam* elements (12). These data illustrate the dynamics of piRNA clusters and their coevolution with the rest of the genome regarding TE content. They also highlight the essential role that piRNA clusters might play in speciation by remodeling via TE control of large genomic regions.

Materials and Methods

Bioinformatics. Sequences and genomic positions were from Flybase (<http://flybase.org>) *Drosophila melanogaster* genome release 5.1. The TE content of the *flam* locus was first analyzed by CENSOR using Repbase (www.girinst.org) (18, 19). Elements that had more than 98% sequence identity with the library (or with newly identified TEs) were defined as recent TEs. The sequences having less than 98% sequence identity with the library were considered as ancient TEs. TE annotations were from Flybase (*Drosophila melanogaster* genome release 5.1).

Search for TEs in Different *Drosophila* Species. Sequences homologous to TEs present in the *flam* locus were extracted from the *Drosophila* genome database by standard nucleotide–nucleotide BLAST search without a low-complexity filter at the Flybase web page (<http://flybase.org>).

Sequence Alignments. Pairwise sequence alignments were done using bl2seq without a filter at Mobyle@Pasteur (<http://mobyle.pasteur.fr>).

***Drosophila* Strains.** Flies were reared on standard medium at 22 °C ± 1°. The w^{IR6}, Rev, and Iso1A lines are from the collection of the Institut National de la Santé et de la Recherche Médicale U1103. The Df(1)DCB1-35b line (24) was provided by the Bloomington *Drosophila* Stock Center at Indiana University.

PCR Amplifications. PCR amplifications were carried out under standard conditions in 50 μL of either Recombinant Taq DNA polymerase (Invitrogen) for short fragments to amplify or the Expand Long Template PCR System (Roche) for fragments longer than 3 kb, according to the manufacturers' recommendations. Amplification was allowed for 40 cycles. Oligonucleotides (compare with [Dataset S2](#)) were essentially designed as in the work of de La Roche Saint André and Bregliano (22). They were purchased from Eurogentec. Sequencing of some amplified DNA fragments was done by GATC Biotech.

ACKNOWLEDGMENTS. We thank Françoise Pellissier and Agostinha De Sousa for technical assistance. We thank Phillip George for critical reading of the manuscript. This work was supported by grants from the Région Auvergne and the European Union, the Association pour la Recherche contre le Cancer, and the Ministère de l'Enseignement Supérieur et de la Recherche.

- Schnable PS, et al. (2009) The B73 maize genome: Complexity, diversity, and dynamics. *Science* 326(5956):1112–1115.
- Ghildiyal M, Zamore PD (2009) Small silencing RNAs: An expanding universe. *Nat Rev Genet* 10(2):94–108.
- Malone CD, Hannon GJ (2009) Small RNAs as guardians of the genome. *Cell* 136(4):656–668.
- Siomi MC, Sato K, Pezic D, Aravin AA (2011) PIWI-interacting small RNAs: The vanguard of genome defence. *Nat Rev Mol Cell Biol* 12(4):246–258.
- Siomi MC, Saito K, Siomi H (2008) How selfish retrotransposons are silenced in *Drosophila* germline and somatic cells. *FEBS Lett* 582(17):2473–2478.
- Brennecke J, et al. (2007) Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 128(6):1089–1103.
- Pélisson A, et al. (1994) Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the *Drosophila flamenco* gene. *EMBO J* 13(18):4401–4411.
- Prud'homme N, Gans M, Masson M, Terzian C, Bucheton A (1995) Flamenco, a gene controlling the gypsy retrovirus of *Drosophila melanogaster*. *Genetics* 139(2):697–711.
- Desset S, Meignin C, Dastugue B, Vauray C (2003) COM, a heterochromatic locus governing the control of independent endogenous retroviruses from *Drosophila melanogaster*. *Genetics* 164(2):501–509.
- Finnegan DJ (2012) Retrotransposons. *Curr Biol* 22(11):R432–R437.
- Mével-Ninio M, Pelisson A, Kinder J, Campos AR, Bucheton A (2007) The flamenco locus controls the gypsy and ZAM retroviruses and is required for *Drosophila* oogenesis. *Genetics* 175(4):1615–1624.
- Malone CD, et al. (2009) Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell* 137(3):522–535.
- Leblanc P, Desset S, Dastugue B, Vauray C (1997) Invertebrate retroviruses: ZAM a new candidate in *D. melanogaster*. *EMBO J* 16(24):7521–7531.
- Desset S, et al. (1999) Mobilization of two retroelements, ZAM and Idefix, in a novel unstable line of *Drosophila melanogaster*. *Mol Biol Evol* 16(1):54–66.
- Tcheressiz S, et al. (2002) Expression of the Idefix retrotransposon in early follicle cells in the germline of *Drosophila melanogaster* is determined by its LTR sequences and a specific genomic context. *Mol Genet Genomics* 267(2):133–141.
- Bergman CM, Quesneville H, Anxolabéhère D, Ashburner M (2006) Recurrent insertion and duplication generate networks of transposable element sequences in the *Drosophila melanogaster* genome. *Genome Biol* 7(11):R112.
- Quesneville H, et al. (2005) Combined evidence annotation of transposable elements in genome sequences. *PLoS Comput Biol* 1(2):166–175.
- Jurka J, Klonowski P, Dagman V, Pelton P (1996) CENSOR—A program for identification and elimination of repetitive elements from DNA sequences. *Comput Chem* 20(1):119–121.
- Jurka J, et al. (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* 110(1-4):462–467.
- Bartolomé C, Bello X, Maside X (2009) Widespread evidence for horizontal transfer of transposable elements across *Drosophila* genomes. *Genome Biol* 10(2):R22.
- Yu B, et al. (2011) Structural analysis of a 4414-bp element in *Drosophila melanogaster*. *Genet Mol Res* 10(2):717–730.
- de La Roche Saint André C, Bregliano JC (1998) Evidence for a multistep control in transposition of I factor in *Drosophila melanogaster*. *Genetics* 148(4):1875–1884.
- Lau NC, et al. (2009) Abundant primary piRNAs, endo-siRNAs, and microRNAs in a *Drosophila* ovary cell line. *Genome Res* 19(10):1776–1785.
- Schalet A, Lefevre G, Jr. (1973) The localization of "ordinary" sex-linked genes in section 20 of the polytene X chromosome of *Drosophila melanogaster*. *Chromosoma* 44(2):183–202.
- Loreto EL, Carareto CM, Capy P (2008) Revisiting horizontal transfer of transposable elements in *Drosophila*. *Heredity (Edinb)* 100(6):545–554.
- Eickbush DG, Lathe WC, 3rd, Francino MP, Eickbush TH (1995) R1 and R2 retrotransposable elements of *Drosophila* evolve at rates similar to those of nuclear genes. *Genetics* 139(2):685–695.
- Vermaak D, Henikoff S, Malik HS (2005) Positive selection drives the evolution of rhino, a member of the heterochromatin protein 1 family in *Drosophila*. *PLoS Genet* 1(1):96–108.
- Vermaak D, Malik HS (2009) Multiple roles for heterochromatin protein 1 genes in *Drosophila*. *Annu Rev Genet* 43:467–492.
- Werren JH (2011) Selfish genetic elements, genetic conflict, and evolutionary innovation. *Proc Natl Acad Sci USA* 108(Suppl 2):10863–10870.
- Le Rouzic A, Capy P (2005) The first steps of transposable elements invasion: Parasitic strategy vs. genetic drift. *Genetics* 169(2):1033–1043.