



**HAL**  
open science

# Étude de la réactivation d'un rétrovirus endogène dans les ovaires de *Drosophila melanogaster* : expression, invasion et réponse génomique

Marianne Yoth

► **To cite this version:**

Marianne Yoth. Étude de la réactivation d'un rétrovirus endogène dans les ovaires de *Drosophila melanogaster* : expression, invasion et réponse génomique. Génétique. Université Clermont Auvergne, 2022. Français. NNT : 2022UCFAC051 . tel-04049119

**HAL Id: tel-04049119**

**<https://theses.hal.science/tel-04049119v1>**

Submitted on 28 Mar 2023

**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.

Année 2022



ÉCOLE DOCTORALE DES SCIENCES DE LA VIE, SANTE, AGRONOMIE,  
ENVIRONNEMENT

## Thèse

Présentée à l'Université Clermont Auvergne  
Pour l'obtention du grade de DOCTEUR  
(Décret du 5 juillet 1984)

Spécialité : Science de la vie et de la Santé

Marianne YOTH

---

# Étude de la réactivation d'un rétrovirus endogène dans les ovaires de *Drosophila melanogaster* : expression, invasion et réponse génomique

---

Soutenue publiquement le vendredi 30 septembre 2022 devant le jury d'examen :

Président du Jury : TATOUT Christophe  
Rapporteurs : BOIVIN Antoine  
DUHARCOURT Sandra  
Examineur : GILBERT Nicolas  
Directrices de Thèse : BRASSET Émilie  
JENSEN Silke

Équipe « Instabilités génétiques et contrôle par le génome de l'hôte »  
Laboratoire GReD (Génétique, Reproduction et Développement)  
CNRS (UMR6293) – INSERM (U1103)  
Université Clermont Auvergne



Année 2022



ÉCOLE DOCTORALE DES SCIENCES DE LA VIE, SANTE, AGRONOMIE,  
ENVIRONNEMENT

## Thèse

Présentée à l'Université Clermont Auvergne  
Pour l'obtention du grade de DOCTEUR  
(Décret du 5 juillet 1984)

Spécialité : Science de la vie et de la Santé

Marianne YOTH

---

# Étude de la réactivation d'un rétrovirus endogène dans les ovaires de *Drosophila melanogaster* : expression, invasion et réponse génomique

---

Soutenue publiquement le mardi 30 septembre 2022 devant le jury d'examen :

Président du Jury : TATOUT Christophe  
Rapporteurs : BOIVIN Antoine  
DUHARCOURT Sandra  
Examineur : GILBERT Nicolas  
Directrices de Thèse : BRASSET Émilie  
JENSEN Silke

Équipe « Instabilités génétiques et contrôle par le génome de l'hôte »  
Laboratoire GReD (Génétique, Reproduction et Développement)  
CNRS (UMR6293) – INSERM (U1103)  
Université Clermont Auvergne



## Résumé

Pour la survie des espèces, la lignée germinale doit transmettre fidèlement l'information génétique à la descendance. Les éléments transposables (ETs) sont des séquences d'ADN mobiles qui constituent une menace importante pour la stabilité des génomes. Les ETs constituent une fraction importante du génome des eucaryotes supérieurs, ils représentent 20% du génome de la drosophile et plus de 45% du génome humain. Dans les gonades des métazoaires, la mobilisation des ETs est contrôlée par une classe spécifique de petits ARNs, appelés piARNs, qui interagissent avec les protéines de la famille PIWI. Les piARNs sont produits à partir de loci génomiques dédiés, appelés clusters de piARNs, exprimés de manière tissu-spécifique. Les clusters de piARNs sont composés d'une multitude d'ETs, pleine longueur ou bien tronqués, qui représentent le répertoire des ETs que la cellule peut réprimer afin de maintenir la stabilité de son génome. Cependant, il existe au sein des génomes une classe particulière d'ETs, les rétrovirus endogènes, qui lorsqu'ils sont réactivés dans les tissus somatiques adjacents aux cellules germinales, peuvent envahir les cellules germinales, comme le ferait un virus. Bien que la voie des piARNs soit un système immunitaire génomique efficace pour contrôler la transposition des ETs exprimés dans les cellules germinales, les mécanismes développés par la lignée germinale pour se protéger face à une nouvelle invasion d'ET restent mal connus. Pour lever ce verrou scientifique, durant ma thèse j'ai étudié l'invasion des cellules germinales par un rétrovirus endogène présent dans le génome de *Drosophila melanogaster* : ZAM. L'objectif du projet visait à étudier i) les stratégies mises en place par ZAM pour envahir les cellules germinales ii) les conséquences d'une réactivation d'un rétrovirus endogène et iii) les mécanismes de défense mis en place par les cellules germinales pour contrer cette invasion et ainsi protéger le génome germinale d'une transposition massive qui pourrait mettre en danger la survie même des espèces. Mes résultats ont permis de révéler pour la première fois que les piRNAs produits localement, dans les cellules germinales, peuvent agir à l'échelle tissulaire et ainsi protéger la lignée germinale d'une attaque extérieure.



## Abstract

For species survival, the germline must faithfully transmit genetic information to the progeny. Transposable elements (TEs) constitute a significant threat to genome stability due to their capacity to move. In higher eukaryotes, TEs make up a large fraction of the genome, accounting for 20% of the *Drosophila* genome and more than 45% of the *Human* genome. In the metazoan gonads, their expression and so their mobilization is limited by a class of small RNAs called PIWI-interacting RNAs (piRNAs) produced by dedicated genomic loci called piRNA clusters that are expressed in a tissue-specific manner. piRNA clusters are composed of a multitude of TEs either full length or truncated that represent the repertoire of TEs that the cell must repress to maintain the stability of its genome. However, a reactivation of endogenous retroviruses, a particular class of TEs, in somatic tissue adjacent to germ cells, promotes its infection and invasion of the germ line as a virus would do. Although the piRNA pathway is an efficient genomic immunity system, it remains unclear how the germline gains protection against a new transposon invasion. The objective of my thesis was to follow the impact of a TE neo-invasion from an adjacent somatic tissue towards the germ line. For this study, I used the *ZAM* endogenous retrovirus from *Drosophila melanogaster* as a model. I investigated i) the strategies developed by *ZAM* to invade the germ cells ii) the consequences of an endogenous retrovirus reactivation and iii) the defense mechanisms developed by germ cells to counteract *ZAM* invasion and to protect the germline genome from a massive transposition that could endanger the survival of the species. My results revealed for the first time that piRNAs produced locally, in germ cells, act at the tissue scale and thus protect the germ line from an external attack.





## Remerciements

Cette thèse a été un réel travail collaboratif, je tiens à prendre quelques lignes parmi les milliers qui vont suivre, pour remercier l'ensemble des personnes qui ont contribué à rendre cette aventure aussi incroyable. (C'est toujours un peu dommage d'attendre la fin pour dire merci, mais mieux vaut tard que jamais).

Ce travail a été réalisé au sein de l'institut GReD et j'exprime toute ma reconnaissance à Chantal VAURY, ancienne directrice de l'institut, ainsi qu'au nouveau directeur, Krzysztof JAGLA, pour leur accueil au sein de cette structure et pour m'avoir permis d'effectuer cette thèse dans des conditions de travail idéales.

Je remercie les rapporteurs de cette thèse, Sandra DUHARCOURT et Antoine BOIVIN d'avoir accepté de juger mon travail ainsi que Nicolas GILBERT d'avoir accepté le rôle d'examineur. Je vous remercie sincèrement de vous être déplacés jusqu'à Clermont-Ferrand pour évaluer mon travail. Un grand merci également à Christophe TATOUT qui m'a fait l'honneur d'accepter d'être le président de ce jury. En tant que professeur de génétique en Master, il a grandement contribué à faire naître en moi cette passion grandissante pour cette discipline.

J'adresse ensuite un IMMENSE merci à l'ensemble des membres de l'équipe « Instabilité génétique et contrôle par le génome hôte » pour leur accueil et surtout pour m'avoir aidée à grandir aussi bien scientifiquement qu'humainement. J'ai pris un petit bout de savoir (et de savoir être) chez chacun(e) d'entre vous, je ne vous remercierai jamais assez pour votre bienveillance contagieuse. Ça va être dur de partir car c'était un peu ma maison ce laboratoire... et c'est grâce à vous (et aux mouches aussi un peu).

Bien-sûr je remercie profondément ma directrice de thèse, Émilie BRASSET, qui m'a suivie depuis le début de mon arrivée en DUT à Clermont-Ferrand, qui m'a tout d'abord accueillie au sein de son équipe lors de mon stage de DUT puis de Master 2 et enfin pour cette grande aventure qui est la thèse. Ces quelques mots ne pourront jamais résumer tout ce que tu m'as apporté. Merci de m'avoir poussée scientifiquement et humainement à toujours devenir la meilleure version de moi-même, merci d'avoir partagé avec moi ton émulsion scientifique et ta passion communicative. Cette stimulation intellectuelle quotidienne était fantastique et je ne te remercierai jamais assez pour la confiance que tu m'as accordée. J'admire ta force et ta persévérance, ta bienveillance et ton humanité. Tu as été et tu restes un réel exemple pour moi et je suis fière d'avoir pu travailler avec toi ces quelques années.

Merci également à Silke JENSEN qui est devenue récemment ma co-directrice de thèse et ce rôle est plus que mérité. Il est certain que ce travail n'aurait pas pu aboutir sans ton aide quotidienne et sans tes conseils avisés. J'ai toujours été émerveillée par ta curiosité excessive et ta rigueur sans faille qui font de toi une chercheuse hors paire. Merci d'avoir accepté de me transmettre un peu de ton savoir, notamment en bio-informatique, je partais de loin mais tu as toujours su faire preuve de pédagogie et de beaucoup de patience. Ta présence dans l'équipe est une vraie force, tu es toujours prête à aider tout le monde, sans jamais rien attendre en retour. Merci Silke, j'espère repartir du labo avec au moins quelques-unes de tes compétences (avec beaucoup de tes tutos, ça c'est sûr !).

Un énorme merci (et c'est un bien faible mot) à Stéphanie MAUPETIT-MEHOUAS qui m'a épaulée d'arrache-pied durant ces 4 années. Dès mon arrivée, j'ai tout de suite été épatée par tes immenses connaissances théoriques et techniques et par ta capacité de raisonnement. J'ai énormément appris grâce à toi (je n'oublierai jamais les contrôles c'est promis et je ferai toujours mes PCR dans un espace dédié, peu importe où je serai !). Merci de toujours me rappeler que « Fast is slow and slow is fast », merci pour ton soutien quotidien, pour ta bonne humeur et pour toutes ces discussions profondes et sincères. Bien plus qu'une collègue, tu es devenue une amie, alors MERCI de m'avoir aidée dans les moments où tu as su voir que j'en avais vraiment besoin et MERCI pour tous les fous rires et tous les moments de joie. Tu vas énormément me manquer !



Merci ensuite à Nathalie GUEGEN pour son aide quotidienne, je ne compte plus le nombre de fois où tu as accepté de me donner un coup de main quand tu voyais que j'étais (un peu) dépassée par les événements.... Tes conseils techniques/scientifiques/personnels m'ont été d'une grande utilité et c'est une chance d'avoir pu travailler à tes côtés pendant ces quelques années. Merci pour ta disponibilité Nath et merci pour toutes ces discussions qui ont rendu ces longues journées de travail bien plus agréables.

Emilie, Silke, Steph et Nath vous êtes le pilier de cette équipe et si l'ambiance est toujours chaleureuse, conviviale et joyeuse, c'est beaucoup grâce à vous. Mais cette équipe est aussi peuplée d'étudiants et de post-doc qui ont (ou qui ont été) des moteurs pour la faire tourner dans la bonne humeur.

Aline, tu prends la relève et je te souhaite plein de bonnes choses pour cette fin de thèse ! C'était un réel plaisir de partager ces quelques années avec toi, et tu as toujours été incroyablement bienveillante et heureusement que tu étais là pour m'aider au quotidien (sans toi, grand nombre de mes mouches seraient mortes et je ne parle même pas de mes plantes...). Un GRAND merci aussi pour la relecture de cette thèse, tes talents d'écriture et de relecture ont permis d'éliminer environ un millier de fautes... Prends confiance en toi Aline, tu es géniale et tu mérites le meilleur !

Merci également à Abdou qui a rejoint l'équipe récemment en tant que post-doc et qui m'a aidée à garder le moral pendant cette fin de thèse (grâce aux nombreuses viennoiseries et pâtisseries libanaises notamment...). Je te souhaite bon courage pour la suite et j'espère que tu obtiendras rapidement la place que tu mérites au sein de la communauté scientifique.

En parlant de post-doc, je pense bien-évidemment à Benjamin BERTIN, avec qui j'ai eu la chance de partager 2 années riches en échanges scientifiques (et non scientifiques surtout !). T'avoir à mes côtés en tant que post-doc lors de mes débuts en thèse a été un atout majeur, tu m'as toujours conseillée sans me juger (enfin si, un peu, mais toujours avec beaucoup de bienveillance) et tes blagues, ta bonne humeur et nos discussions m'ont terriblement manqué cette dernière année. Je peux t'assurer que j'étais presque au bout du gouffre et que j'ai dû écouter Marcel et les marcassins de nombreuses fois pour éviter de finir rouge opaque à cause du stress. Le p'tit pulpe te remercie pour tout ça.

Une petite pensée pour tous les stagiaires qui sont passés dans l'équipe pendant ces quatre années, un merci tout particulier à Anaëlle qui a pris grand soin de mes précieuses mouches et à Judith que j'ai eu la chance d'encadrer pendant 2 mois.

Merci également à Céline DUC avec qui j'ai eu la chance de commencer ce merveilleux projet.

Merci à Chantal VAURY qui a dérivé la lignée de drosophiles Rev il y a plusieurs années maintenant, ainsi qu'à toutes les personnes qui ont travaillé sur ce projet et l'ont fait grandir avant moi.

Merci à l'ensemble des drosophilistes du 5<sup>ème</sup> étage (et aux pièces rapportées qui viennent prendre le café). Merci à Lili et Marine, les deux petits rayons de soleil de la pièce manip. Merci également à Ronan, Victor et Nico pour les pauses café (et toutes celles que j'ai loupées surtout) et pour les soirées bières un peu trop arrosées (je ne parle pas de moi bien-sûr).

Merci à l'ensemble des membres de l'équipe Mirouse avec qui nous partageons le labo avec toujours beaucoup de plaisir, et notamment Graziella pour ses conseils techniques et à Hervé pour sa bonne humeur quotidienne (et ses compétences extraordinaires en dissections d'ovaires). Merci à tous les membres de l'équipe Waltzer et de l'équipe Jagla qui rendent la pièce de trie de droso toujours agréable et joyeuse !

Merci également à toutes les personnes qui nous facilitent énormément la vie de labo, merci à Fanfan qui est maintenant partie à la retraite, à Nadège et à Eric pour leur aide et leur bonne humeur. Un merci tout particulier à Caro qui est toujours là pour aider tout le monde, tu mérites une médaille de superwoman !

Merci à Oscar de m'avoir guidé dans mes réflexions et actions durant une bonne partie de ma thèse et merci pour tous ces instants précieux.



Et parce qu'il y a quand même une vie à côté de cette thèse, je remercie tous les gens qui rendent mon quotidien riche de sens.

Je ne peux écrire cette thèse sans te remercier Hélène, car tu as été un pilier pendant de nombreuses années. Tous les moments passés ensemble resteront gravés en moi pour toujours, merci pour toutes ces années de complicité.

Merci également à Geoffrey d'avoir été la personne que tu es.

Ma thèse a été riche en émotions et semée de quelques embuches personnelles, merci à tous les gens qui m'ont soutenue et ne m'ont jamais jugée. Merci à Fanny d'être restée à mes côtés pendant toutes ces années, ton soutien est tellement important. Merci à Sabine, Marjo et Elyse pour tous les moments de bonheur partagés.

Je pense également à mon incroyable compagnon de vie. Guillaume, merci d'avoir débarqué dans ma petite vie, et de m'accepter telle que je suis. Merci de me comprendre sans même que j'ai besoin de parler, merci de rendre ma vie un peu plus belle chaque jour, merci pour ton soutien, ta confiance, ton amour. Merci de me rappeler constamment qu'il faut prendre le maximum de chaque opportunité, que la vie est pleine d'étincelles et de me pousser à toujours devenir meilleure. J'ai une chance incroyable de t'avoir trouvé.

Un IMMENSE merci à ma merveilleuse famille, qui ne comprend jamais rien quand je parle des drosophiles, des éléments transposables ou de n'importe quoi en rapport avec la biologie mais qui me soutient toujours. Merci tout simplement d'être là, d'être ce que vous êtes.

Merci à la « Team Gouzou » pour tous les moments de partage et de bonheur. Je pense notamment à Audrey, l'une de mes personnes préférées sur terre, merci pour les discussions sur la vie avec un grand V et d'être une personne aussi incroyable, je suis tellement fière de toi.

Merci à mon frère qui m'a fait le plus grand bonheur de devenir Tata cette année.

Merci à mes deux petites sœurs qui égailent mon quotidien par leur joie de vivre inconditionnelle et à mon papa de m'avoir transmis sa passion pour la vie.

Mais ma plus grande reconnaissance va à ma maman, il me faudrait des pages entières pour exprimer tout mon amour et ma gratitude, mais sache juste que j'ai espoir de devenir un jour une personne aussi humaine, courageuse, forte et généreuse que toi.

J'ai une chance incroyable de vous avoir, tous.

ET enfin.... MERCIIII aux petites mouches aux yeux rouges d'avoir fait vibrer mon âme de scientifique en herbe pendant ces quelques années (et pour toutes les heures de sommeil perdues à tout jamais).

Ce n'est que le début d'une grande aventure je l'espère...

La grande aventure de l'existence c'est de trouver là où l'on est bien.

« Sylvain Tesson »



# **Introduction** **1**

## **Partie 1 - Les éléments transposables** **2**

### **I- Classification, Structure et cycle de transposition.** **3**

I.1- Les transposons à ADN 3

I.2- Les rétrotransposons 4

I.2. a) Les rétrotransposons sans LTR 5

I.2. b) Les rétrotransposons avec LTR 5

I.3- Les rétrovirus endogènes (ERVs) 6

I.3. a) Structure génomique des ERV 6

I.3. b) Cycle de transposition des rétrotransposons 8

I.3. c) Impact de la présence des ERVs dans les génomes. 12

### **II- Relation complexe entre les éléments transposables et leurs hôtes.** **14**

II.1 - Influence des éléments transposables sur l'évolution des génomes 15

II.1. a) Domestication : Comment les génomes s'approprient-ils les ETs ? 15

II.1. b) Effet des ETs sur la régulation des gènes 16

II.2- ETs versus Hôtes : Qui convertira le prochain essai ? 17

Revue : *The Intricate Evolutionary Balance between Transposable Elements and Their Host: Who Will Kick at Goal and Convert the Next Try?*

## **Partie 2 - Régulation des éléments transposables chez la drosophile** **18**

### **I- La drosophile : la star de la génétique.** **18**

I.1. Cycle de vie et ovogenèse chez la Drosophile 19

II.1. a) Cycle de vie de la Drosophile 19

II.1. b) Ovogenèse de la Drosophile 20

I.2. Échanges entre les cellules somatiques folliculaires et les cellules germinales au sein de l'ovaire de la Drosophile 21

### **II- Régulation des éléments transposables dans les ovaires de la drosophile** **22**

II.1- Rôle des siRNA dans la répression des ETs 23

II.2- Découverte du rôle des piARN dans le contrôle des ETs 24





II.2. a) Découverte des piARNs	24
II.2. b) Les piARNs : un rôle dans la régulation des éléments transposables ?	25
II.2 c) Identification de la source de production des piARN	27
II.3- La voie des piARNs : De la biogenèse des piARNs à la répression des ETs	28
II.3 a) Les protéines de la famille PIWI	28
II.3. b) Les clusters de piARNs	29
II.3 c) Maturation des transcrits précurseurs en piARN	34
II.3. d) Rôle des piARNs dans la répression des ETs	36
II.3 e) Répression tissu spécifique des ETs	39
<b>Partie 3 - ZAM : Un modèle d'étude unique</b>	<b>42</b>
<b>I- Découverte des lignées Rev et du retrovirus endogène ZAM</b>	<b>42</b>
I.1- « Z'en ai marre » envahi le génome !	42
I.2- La lignée Rev : une lignée génétiquement instable.	43
<b>II- ZAM : un modèle puissant pour étudier la régulation des ETs.</b>	<b>44</b>
II.1- Relation directe entre la présence de <i>ZAM</i> dans <i>flamenco</i> et sa répression.	44
II.2- <i>ZAM</i> : Un ERV aux capacités d'infection	45
II.2. a) Expression tissu spécifique de <i>ZAM</i> .	45
II.2. b) Mécanismes utilisés pour la transmission intercellulaire de <i>ZAM</i>	46
II.3- Pourquoi utiliser <i>ZAM</i> comme modèle d'étude ?	48
<b>Résultats</b>	<b>50</b>
<b>Partie 1 - L'insertion d'un rétrovirus endogène somatique dans un piRNA cluster à expression germinale permet de protéger les cellules germinales de nouvelle invasion.</b>	<b>51</b>
<i>Article: Trapping a somatic endogenous retrovirus into a germline piRNA cluster immunizes the germline against further invasion</i>	



**Partie 2** - Les piRNA germinaux permettent de contrer les invasions de rétrotransposons provenant des cellules somatiques environnantes dans les ovaires de *Drosophila melanogaster*. **52**

*Article: Germline piRNAs counteract retrotransposon invasion from somatic cells.*

**Partie 3** - Délétion d'un élément transposable d'un cluster de piARN **54**

**I- Fonctionnement de la technique CRISPR-Cas9 et plan d'action** 54

1.1 - Principe général de la technique de "ciseaux génétiques" CRISPR-Cas9 54

1.2 - Lignées de drosophiles utilisées 55

1.3 - Site ciblé pour la mutation 56

**II- Les difficultés rencontrées pour créer la lignée CRISPR-flam $\Delta$ ZAM** 56

**III- Obtention tant attendue de la lignée CRISPR-flam $\Delta$ ZAM** 59

III.1- Vérification de la présence de la délétion 59

III. 2-Différentes lignées CRISPR-flam $\Delta$ ZAM obtenues 60

**IV- Étude de l'effet de la délétion ZAM-flam sur la régulation et l'expression de ZAM** 62

**V- Expériences en cours et à venir sur la lignée CRISPR-flam $\Delta$ ZAM** 62

**Partie 4** – Analyse du cycle rétroviral de ZAM **63**

**I- Rôle du trafic de vitellogénine dans le transit de ZAM** 64

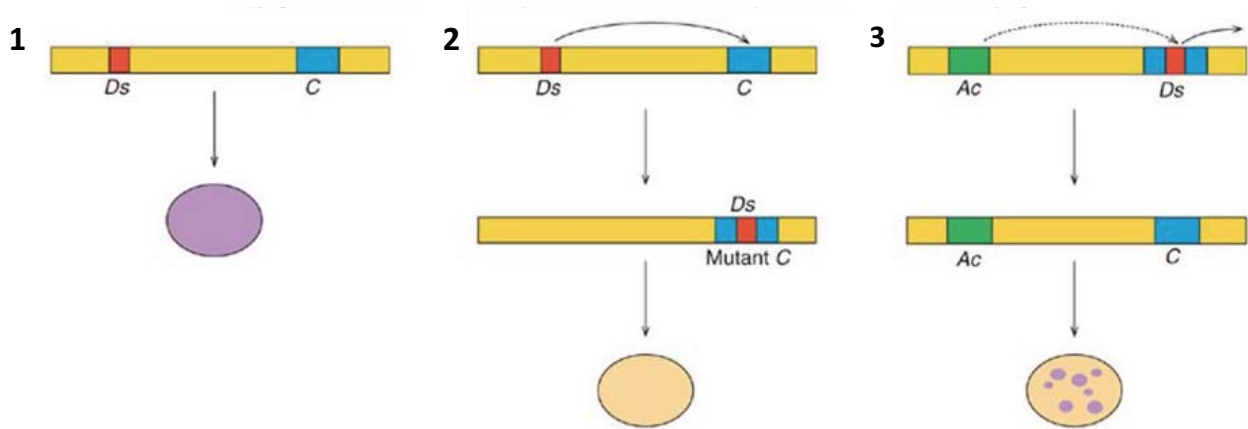
**II- Rôle de la protéine Enveloppe dans le transit de ZAM** 66

**Discussion** **68**

**Conclusion** **84**



# INTRODUCTION



### Figure 1 : Les découvertes de Barbara McClintock

Barbara McClintock étudiait la transmission de la coloration des grains de maïs. En 1944, elle fit une découverte qui bouleversa le domaine de la génétique.

À l'état sauvage, les grains de maïs sont violets (1), or, elle observa que beaucoup de grains de maïs étaient de couleur jaune. Après des analyses cytogénétiques, elle conclut que la coloration jaune des grains de maïs était due à une mutation dans un gène nommé C exprimé dans l'aleurone du maïs. Elle découvrit ensuite que cette mutation était causée par l'insertion d'un élément, nommé Ds, dans ce gène (2). À sa grande surprise, elle observa que des taches violettes apparaissaient parfois sur les grains de maïs jaunes, suggérant que, de façon clonale, cette mutation était réversible. Elle comprit alors que la réapparition de la coloration violette était due à l'excision de l'élément Ds dans certaines cellules de l'aleurone du grain de maïs, permettant de reformer le gène C sauvage (3). L'élément Ds était capable de s'exciser et de s'intégrer ailleurs dans le génome, ce fût une découverte révolutionnaire ! Elle s'aperçut ensuite que pour « bouger » dans le génome, l'élément Ds avait besoin d'un autre élément, nommé Ac. C'est ainsi que le système Ac/Ds et les mécanismes de transposition ont été découverts : c'est le début de l'étude des « gènes sauteurs » maintenant appelés éléments transposables !

Figure extraite du livre « Molecular Biology 4th edition » de Robert F. Weaver (2008)

Je commencerai cette thèse avec une citation de Barbara McClintock :

*« Un organisme n'est pas comme un morceau de plastique, c'est quelque chose qui est constamment affecté par son environnement, qui présente constamment, au cours de sa croissance, des particularités, des faiblesses. »*

Extrait du livre « La passion du Vivant » par Evelyn Fox Keller

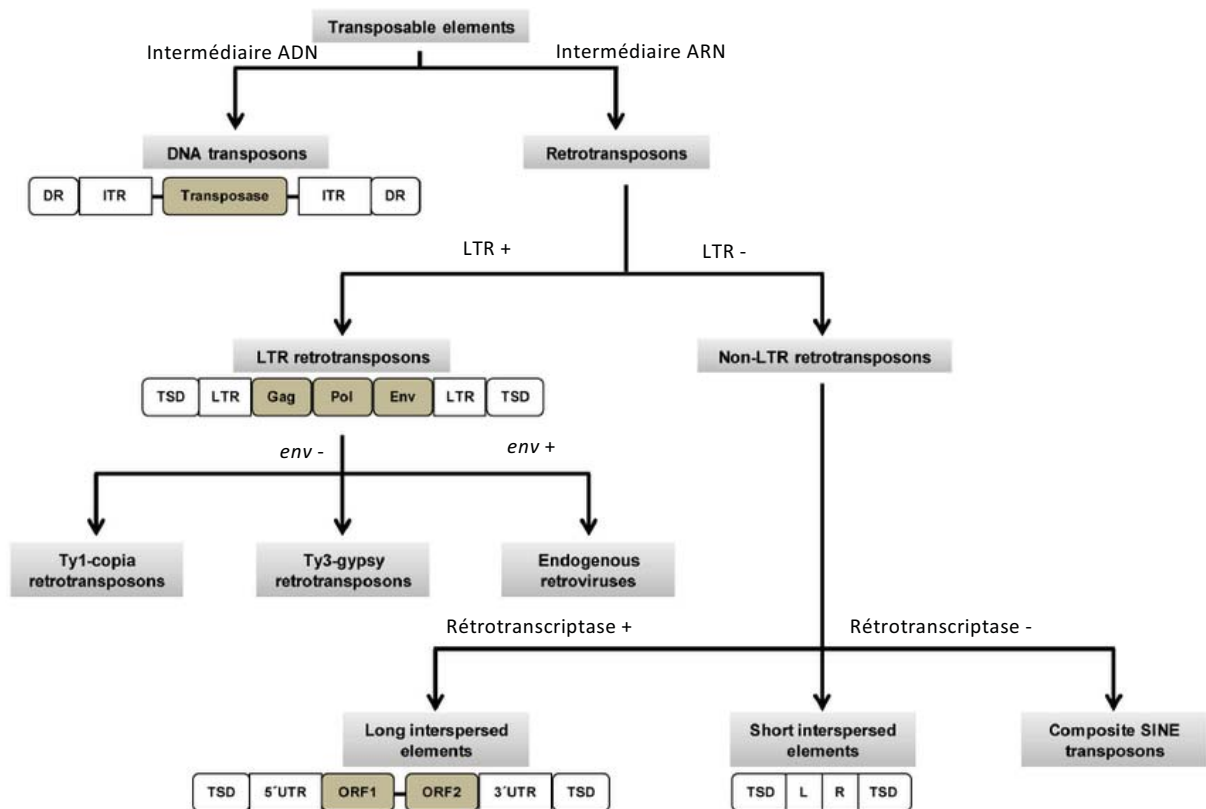
Barbara McClintock était une botaniste, généticienne et cytogénéticienne américaine, prix Nobel de médecine en 1983. Ses découvertes sont à l'origine d'un grand nombre de travaux de recherche, dont le mien.

## **Partie 1 - Les éléments transposables**

Les travaux de recherche de Barbara McClintock publiés en 1950 dans un article intitulé « The origin and behavior of mutable loci in maize » ont déclenché ce que certains appelleront un tremblement de terre génétique ((McClintock, 1950), revue dans : (Jones, 2005)). Ils ont remis en question l'idée même selon laquelle les gènes seraient des structures fixes, peu malléables, transmises de génération en génération. Cette femme à l'intelligence extraordinaire et la curiosité aiguisée, analysa la coloration des grains de maïs pendant de nombreuses années et observa des phénotypes inattendus de changement de couleur. Après des expériences de génétique poussées, elle émit pour la première fois l'hypothèse que les gènes sont beaucoup plus modulables que ce que l'on imaginait. Elle décrit que les gènes peuvent être activés ou désactivés selon les conditions environnementales ou en fonction des stades développementaux. Mais plus impressionnant encore, que certaines séquences d'ADN peuvent changer de position génomique, et même que certaines mutations génétiques peuvent être inversées (McClintock, 1953). C'est ainsi qu'elle découvrit des « gènes sauteurs », aussi appelés « éléments transposables » (ETs) (**Figure 1**).

Dans les années qui ont suivi, cette idée fût très contestée par la communauté scientifique. Cependant, l'engouement pour les ETs s'intensifie de manière extraordinaire lorsqu'on se rend compte dans les années 1970 que ces séquences d'ADN mobiles ne sont pas spécifiques aux plantes mais sont également présentes chez les bactéries, et même chez les animaux (Shapiro et al., 1969; Datta et al., 1971; Kidwell et al., 1977). George Picard a notamment découvert à





**Figure 2 : Classification des éléments transposables**

Il existe deux grandes catégories d'éléments transposables, les DNA transposons et les rétrotransposons. Les rétrotransposons sont sous-divisés en LTR rétrotransposons et non-LTR rétrotransposons qui sont eux-mêmes sous-divisés en plusieurs sous-classes.

Figure adaptée de Ayarpdikannan et Kim 2014 d'après la classification de Finnegan en 1989 et de Anderson et al. en 2008.

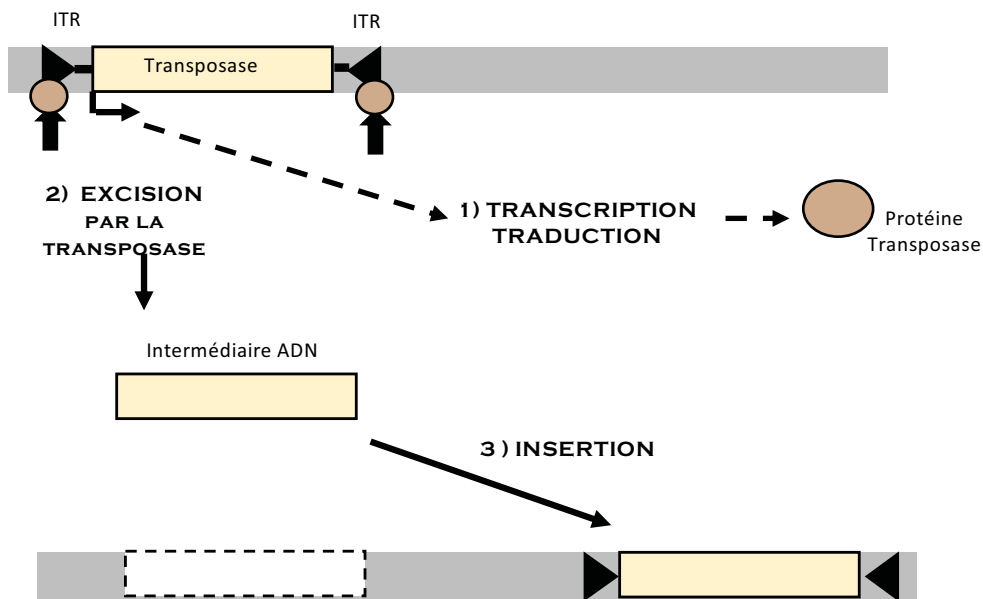
Clermont-Ferrand la présence d'un ET chez la drosophile, nommé le facteur *I*, responsable de phénotypes d'incompatibilités de souches et de stérilité : phénomène appelé la dysgénésie des hybrides (Picard, 1976). Cinquante ans plus tard, l'émergence du séquençage haut débit et des analyses bio-informatiques a permis de révéler la présence d'ETs dans le génome de toutes les espèces analysées jusqu'à présent (Revue dans : (Huang et al., 2012)), à l'exception de la microsporidie *Enterocytozoon bieneusi* (Parisot et al., 2014). De plus, l'annotation de ces ETs a mis en évidence l'existence d'ETs différents présentant une étonnante diversité dans leur structure, leur mécanisme de transposition et leur distribution dans les génomes. Il a donc été nécessaire d'harmoniser et de répertorier ces éléments mobiles qui ont tous pour point commun leur capacité de se déplacer dans les génomes, mais qui sont pourtant si différents lorsqu'on les regarde de plus près.

## I- Classification, structure et cycle de transposition

Les éléments transposables sont définis comme des séquences d'ADN moyennement répétées capables de se déplacer et de se répliquer au sein des génomes. Pour se déplacer d'un locus à un autre, les ETs peuvent utiliser deux mécanismes de transposition distincts : certains transposent *via* un intermédiaire ADN et d'autres utilisent un intermédiaire ARN. C'est sur ce simple critère que David J. Finnegan établit en 1989 la première classification des ETs en séparant les transposons à ADN d'un côté et les rétrotransposons de l'autre (Berg and Howe, 1989; Finnegan, 1989) (**Figure 2**). Depuis, les nouvelles connaissances acquises sur les ETs ont permis d'affiner cette classification, et même de créer de nouvelles classes d'ETs, notamment celle des MITEs (*Miniature Inverted Repeat Transposable Element*) qui transposent selon un mécanisme encore mal connu (Revue dans : (Wicker et al., 2007; Kapitonov and Jurka, 2008)). Cette classification est encore amenée à évoluer avec l'avancée du séquençage génomique qui permet de découvrir et d'annoter constamment de nouveaux ETs dans le génome de multiples espèces. Ainsi, par souci de simplification, dans cette thèse uniquement les deux grandes classes d'ETs seront abordés : les transposons à ADN et les rétrotransposons.

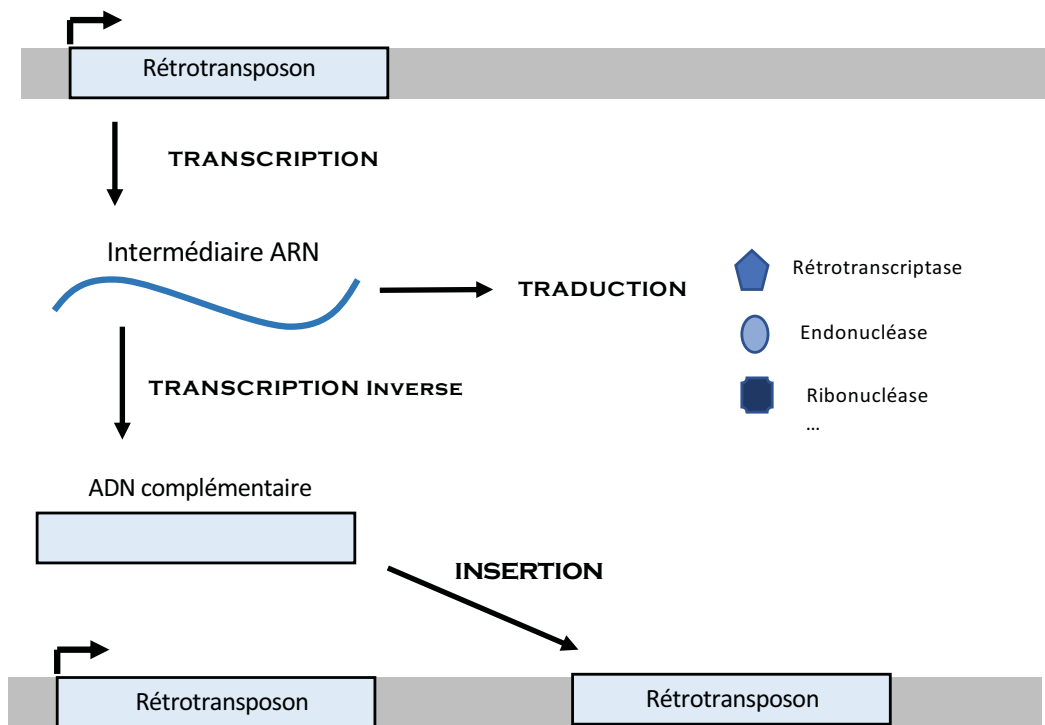
### I.1 - Les transposons à ADN

Les transposons à ADN, aussi appelés les éléments transposables de classe II, se déplacent dans les génomes *via* un mécanisme de « couper-coller ». La transposase, unique enzyme codée par le transposon à ADN, permet l'excision de l'ET de son locus initial via la reconnaissance de



**Figure 3 : Mode de transposition des transposons à ADN (classe II)**

Les transposons à ADN transposent *via* un intermédiaire ADN selon un processus de « couper-coller ». La transposase codée par l'ET excise l'élément de son locus initial *via* la reconnaissance des séquences ITR. Il est ensuite réinséré à un nouveau site génomique.



**Figure 4 : Mode de transposition des rétrotransposons**

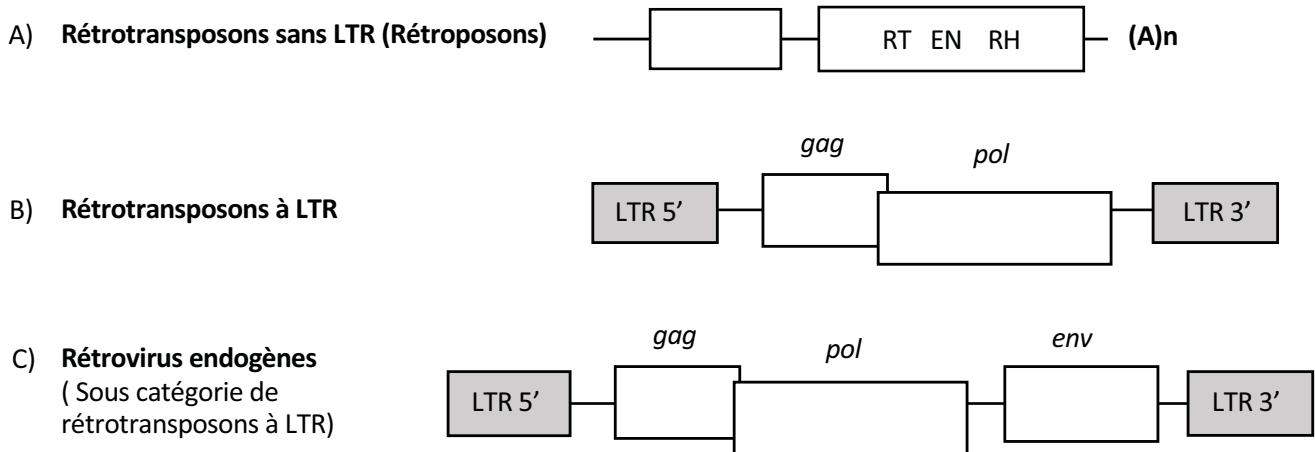
Les rétrotransposons transposent *via* un intermédiaire ARN selon un processus de « copier-coller ». L'élément transposable est d'abord transcrit en un ARN intermédiaire, puis rétrotranscrit en ADN et enfin inséré à un nouveau locus. La copie initiale est conservée à son site d'insertion.

courtes séquences inversées terminales (ITR : *Inverted Terminal Repeat*) présentes à chaque extrémité. L'intermédiaire ADN ainsi généré est ensuite réintégré à un nouveau locus génomique *via* l'utilisation de cette même enzyme (**Figure 3**). Il existe à l'heure actuelle trois grandes familles de transposons à ADN :

- La famille P, composée de l'élément *P* qui a notamment envahi le génome de *Drosophila melanogaster* dans les années 1950 et de quelques transposons à ADN de même type (Preston and Engels, 1989).
- La famille hAT (hobo-Ac-Tam3) qui contient ces trois éléments éponymes présents respectivement dans le génome de *Drosophila melanogaster*, du maïs et du muflier *Antirrhium majus* (Calvi et al., 1991).
- La famille Tc1/mariner, composée de l'élément *Tc1* initialement découvert chez *Caenorhabditis elegans* puis décrit dans de nombreuses espèces telles que les champignons, les nématodes, les drosophiles et l'humain (Emmons et al., 1983; Kachroo et al., 1994; Langin et al., 1995; Smit and Riggs, 1996; Hartl et al., 1997; Ivics et al., 1997).

## I.2 - Les rétrotransposons

Les rétrotransposons, ou éléments de classe I, transposent *via* un mécanisme assimilé à du « copier-coller ». La copie du rétrotransposon est conservée à son site initial et une nouvelle copie est insérée à un nouveau locus. Grâce à ce mode de transposition, les rétrotransposons ont pu envahir et se multiplier dans le génome de nombreuses espèces : ils représentent par exemple près de 50 % du génome humain (contre 3 % pour les transposons à ADN) (Lander et al., 2001). Les rétrotransposons sont subdivisés en plusieurs groupes mais transposent tous grâce à un intermédiaire ARN selon le modèle suivant : la séquence ADN du rétrotransposon est transcrite pour former un ARN intermédiaire qui est ensuite rétrotranscrit en ADN complémentaire (ADNc) qui est alors réintégré dans un nouveau site chromosomique (**Figure 4**). Suivant leur structure, on distingue deux groupes de rétrotransposons : les rétrotransposons à LTR (*Long Terminal Repeat*) et les rétrotransposons sans LTR, aussi appelés rétroposons (**Figure 2**).



**Figure 5 : Structure des rétrotransposons**

A) Les rétrotransposons sans LTR (rétroposons) sont caractérisés par une extrémité 3' riche en adénosine. Ils possèdent un cadre de lecture codant une protéine "RNA-binding" qui s'associe à l'ARN intermédiaire de transposition, et un second cadre de lecture codant les enzymes nécessaires à leur transposition (RT = rétrotranscriptase, EN = endonucléase, RH = ribonucléase H).

B) Les rétrotransposons à LTR sont caractérisés par deux longues répétitions terminales (LTR) qui encadrent deux cadres de lecture : *gag* et *pol*, ayant les mêmes fonctions que chez les rétrovirus.

C) Les rétrovirus endogènes possèdent également des séquences LTR, mais ils ont le cadre de lecture *env*, codant une protéine d'enveloppe comme chez les rétrovirus, en plus des deux autres cadres de lecture *gag* et *pol*.

## I.2. a) Les rétrotransposons sans LTR

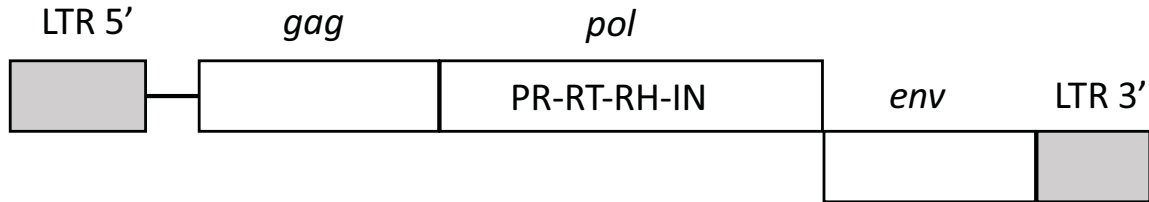
Les rétrotransposons sans LTR sont caractérisés par une extrémité 3' de séquences riches en adénosine (**Figure 5A**). Cette famille de rétrotransposons est fortement représentée au sein du génome des mammifères, elle comprend notamment les rétrotransposons dits « autonomes » (LINEs : *Long Interspersed Nuclear Element*) et les rétrotransposons dits « non-autonomes » (SINEs : *Short interspersed nuclear element*). Ces deux catégories à elles seules représentent plus de 33% du génome humain (Chenais, 2015). Les LINEs possèdent dans leur séquence les gènes codants toutes les enzymes nécessaires à leur transposition (la rétrotranscriptase, l'endonucléase et la RNase H). Ce n'est pas le cas des SINEs qui ne possèdent pas le gène codant ces enzymes (**Figure 2**). Ils dépendent donc d'une complémentation en *trans* par les enzymes des LINEs en *trans* pour pouvoir transposer.

## I.2. b) Les rétrotransposons avec LTR

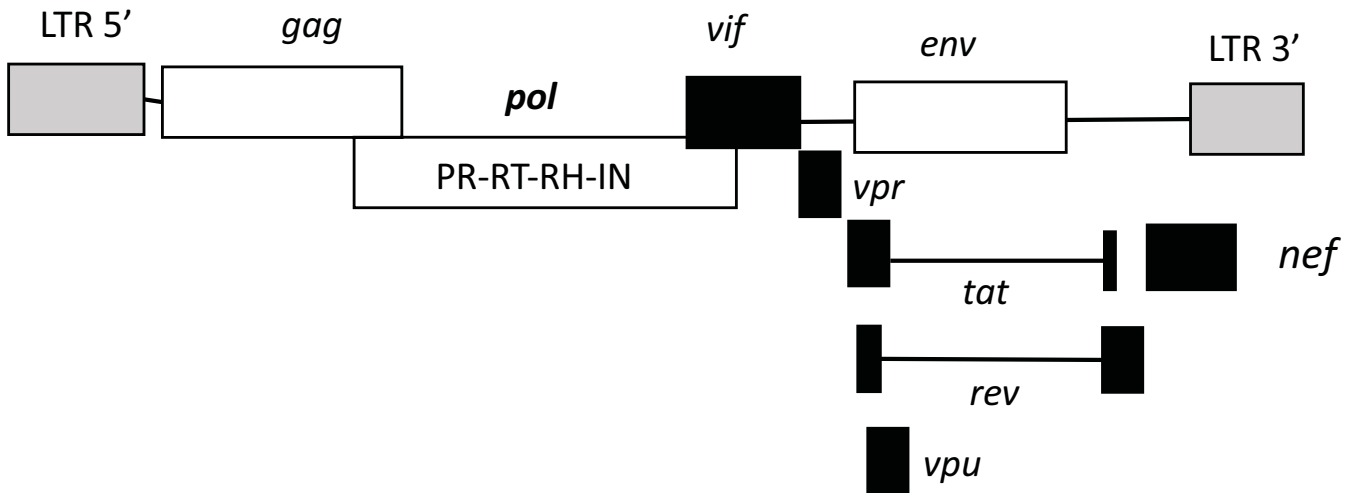
Les rétrotransposons à LTR possèdent en général deux cadres de lecture (ORF = *Open Reading Frame*) : *gag* et *pol*, délimités par de longues répétitions terminales (LTR) (**Figure 5B**). Cette structure est très proche de celle des rétrovirus exogènes sous leur forme provirale (i.e intégré dans le génome). Cependant, la différence majeure réside dans la capacité d'infection inter-individu et intercellulaire des rétrovirus exogènes qui leur est conférée par la présence d'un gène d'*enveloppe* (*env*). Ainsi, l'hypothèse suivante a été émise : les rétrotransposons à LTR pourraient être des ancêtres évolutifs des rétrovirus exogènes, qui seraient alors apparus par un processus complexe comprenant l'acquisition d'un gène *env* (Malik et al., 2000). Cependant, une autre hypothèse a également été envisagée : les rétrotransposons pourraient être le résultat de l'intégration de rétrovirus exogènes au sein des génomes, par un mécanisme d'endogénéisation. Au fil du temps, ces derniers auraient perdu leur gène *env* (Revue dans (Johnson, 2019)). Ces deux hypothèses restent encore débattues et ne sont pas exclusives. Notons qu'il existe au sein des génomes des rétrotransposons à LTR qui possèdent un gène *env*, comme les éléments de type *Gypsy* chez la drosophile, les éléments de type *IAP-E* chez la souris ou encore les éléments *HERV-K* présents dans le génome humain (Péllisson et al., 1994a; Reuss et al., 1996; Boeke and Stoye, 1997; Dewannieux et al., 2005). Les rétrotransposons à LTR avec *env* pourraient alors être issus d'une endogénéisation de rétrovirus exogènes et seraient en transition évolutive entre les stades de rétrovirus exogène et de rétrotransposons à LTR (sans *env*) (Ribet et al., 2008; Johnson, 2019; Keegan et al., 2021). Cette famille de rétrotransposons

# Retroviridae

## Mo-MLV (SOURIS)



## HIV (HOMME)



### Figure 6 : Structure génomique des *Retroviridae*

Le virus Mo-MLV est un rétrovirus infectant les souris. Sa structure est simple : il possède trois cadres de lecture (*gag*, *pol* et *env*) et deux LTR. Le virus HIV infecte l'Homme ; il est considéré comme un rétrovirus complexe car il renferme des cadres de lecture supplémentaires codant des protéines accessoires (*vif*, *vpr*, *tat*, *nef*, *rev*, *vpu*).

Leur structure est très proche des rétrovirus endogènes (Figure 5C), suggérant une origine commune.

PR = protéase, RT = rétrotranscriptase, RH = ribonucléase H, IN = intégrase

à LTR possédant un gène *env* est appelée rétrovirus endogène (ERV) (**Figure 5C**). Les propriétés de transmission/infection/transposition des ERVs restent encore largement méconnues. Je vais essayer de décrire cette famille de façon plus détaillée dans la partie qui suit.

### I.3 - Les rétrovirus endogènes (ERVs)

Les génomes des mammifères hébergent des centaines de milliers de séquences de rétrotransposons, et près de 10 % du génome de l'Homme et de la souris sont occupés par des ERVs, soit plus de 500 000 copies (Huang et al., 2012). La plupart ne sont plus que des fragments d'ERVs, pour la majorité composée uniquement de LTRs solitaires. Cependant, des centaines de copies d'HERV (*Human endogenous retrovirus*) entiers ont été retrouvés dans le génome humain et il a été estimé que 1 % des naissances contiennent une insertion *de novo* d'un rétrovirus endogène (Deininger and Batzer, 2002; Paces et al., 2002; de Parseval et al., 2003). Les ERVs sont considérés comme des reliques d'anciens rétrovirus exogènes ayant infecté la lignée germinale et envahi le génome de leur hôte. Ces rétrovirus endogènes ont conservé de très fortes similitudes de séquence avec les rétrovirus exogènes, notamment avec les *retroviridae* (Lerat and Capy, 1999). Les *retroviridae* sont des virus infectant les vertébrés : on peut par exemple citer le virus Mo-MLV ayant pour hôte la souris et le virus HIV infectant l'Homme. Les informations présentées sur les régions codantes (**Figure 6**) et le cycle de réplication sont principalement extraites du livre « Retrovirus » (Coffin et al., 1997).

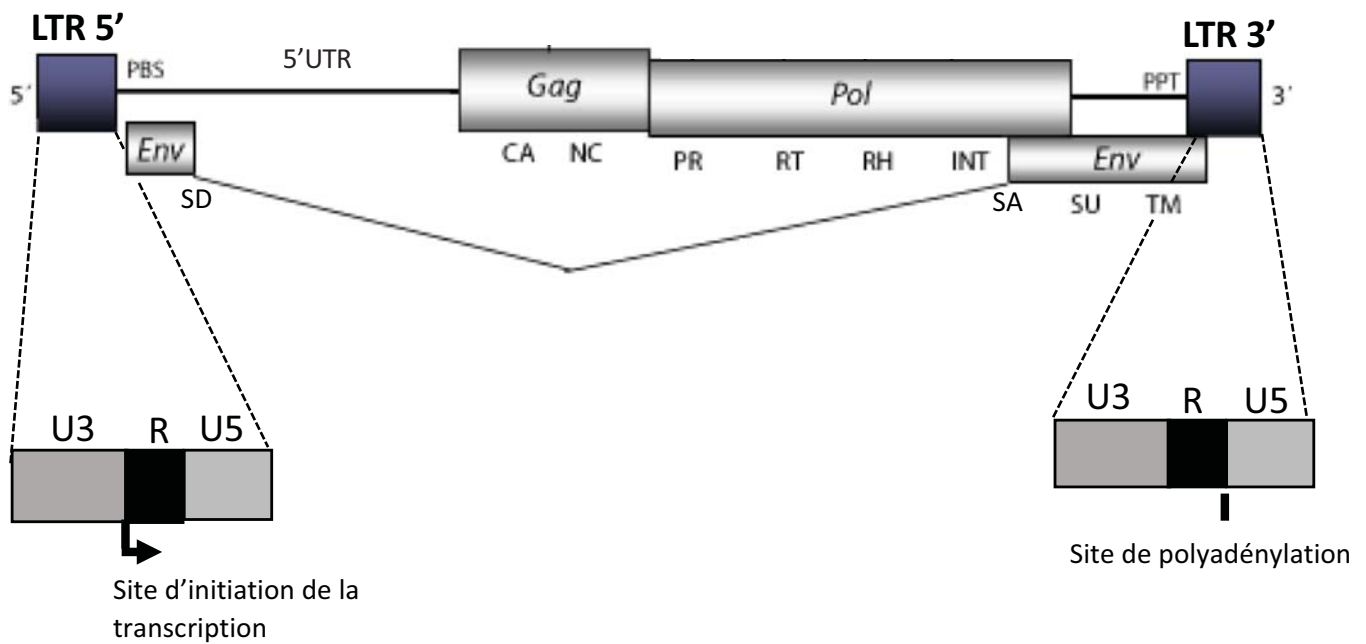
Tels les rétrovirus exogènes, les ERVs ayant conservé le cadre de lecture codant pour une protéine Env peuvent former des pseudo-particules virales enveloppées pouvant se transmettre de cellules à cellules (Chalvet et al., 1999; Brassat et al., 2006; Lee and Bieniasz, 2007; Keegan et al., 2021). Ainsi, à la différence des rétrotransposons à LTR qui ne possèdent que *gag* et *pol*, la présence du gène *env* chez les ERVs leur permet de transposer de manière intercellulaire.

#### I.3. a) Structure génomique des ERVs

Nombreuses des informations sur la structure génomique des ERVs ont été obtenues par comparaison avec la structure des rétrovirus exogènes et ont ainsi été extrapolées. L'ARN génomique des ERVs contient toutes les informations génériques du transposon. En guise d'exemple, l'organisation de la structure génomique du rétrovirus endogène ZAM (Leblanc et al., 1997), présent dans le génome de *Drosophila melanogaster*, est présentée **Figure 7**.



## ZAM : Rétrovirus endogène de *Drosophila melanogaster*



### Figure 7 : Organisation génomique du rétrovirus endogène ZAM

Les LTR sont représentées en bleu de manière détaillée. Chaque LTR est composée de trois régions : U3, R et U5. La partie R est délimitée par le site d'initiation de la transcription et le site de polyadénylation.

CA = Capside, NC = Nucléocapside, PR = protéase, RT = rétrotranscriptase, RH = ribonucléase H, INT = intégrase, SU = sous-unité extracellulaire, TM = domaine transmembranaire, PBS = site de liaison de l'amorce (*Primer Binding Site*), PPT = étendue de polypurines (*Poly Purine Tract*), SD = site donneur d'épissage, SA = site accepteur d'épissage.

Source = <http://gydb.org/index.php/Element:Zam>

⇒ Les gènes gag, pol et env

- **Le gène gag** : ce gène code pour un polypeptide impliqué dans la formation des « *Virus-Like Particle* » (VLP). Les VLP sont des structures qui entourent l'ARN et le protègent lors du cycle de transposition. La polyprotéine Gag est clivée en trois protéines par une protéase (PR), codée par l'ERV lui-même. Les trois protéines de structure ainsi formées sont appelées Matrice (MA), Capside (CA) et Nucléocapside (NC), et s'assemblent pour former la capsid. Les protéines NC possèdent des motifs de liaison aux acides nucléiques (doigts de Zinc) et se lient donc à l'ARN génomique de l'ERV pour former la nucléocapside. Les protéines CA forment une couche de protection à cette nucléocapside. Enfin, les protéines MA forment la couche périphérique de la particule, leur fonction principale est d'orienter la liaison et l'assemblage des VLP à la membrane cellulaire.
- **Le gène pol** : ce gène code également une polyprotéine donnant naissance, après clivage, à des protéines enzymatiques telles qu'une protéase (PR), une rétrotranscriptase (RT), une ribonucléase H (RH) et une intégrase (INT). Ces protéines sont indispensables au cycle de transposition des ERVs.
- **Le gène env** : la polyprotéine Env est synthétisée à partir d'un ARNm épissé. Env est connue pour être indispensable à l'étape d'infection chez les rétrovirus exogènes. Cette polyprotéine est clivée par les protéases cellulaires et non par la protéase codée par l'ERV comme c'est le cas pour Gag et Pol. Ce clivage est médié par les protéases de la famille des furines qui reconnaissent un site de coupure de type RX(K/R)R. Les protéines matures du gène *env* sont ce qu'on appelle, pour les rétrovirus exogènes, les protéines virales Spike. Ce sont par exemple les protéines gp160 du virus HIV, ou les protéines gp71 pour le virus MLV. Ces protéines sont donc constituées de deux sous-unités qui resteront liées par des liaisons non covalentes ou des ponts disulfures. La première sous-unité localisée en N-terminale donne naissance à une protéine de surface (SU) responsable de l'interaction avec le récepteur présent à la surface des cellules cibles. La deuxième sous-unité, dite transmembranaire (TM), possède un domaine hydrophobe responsable de la fusion membranaire entre l'enveloppe du virus et la membrane de la cellule cible lors de l'infection.



⇒ Les régions non codantes

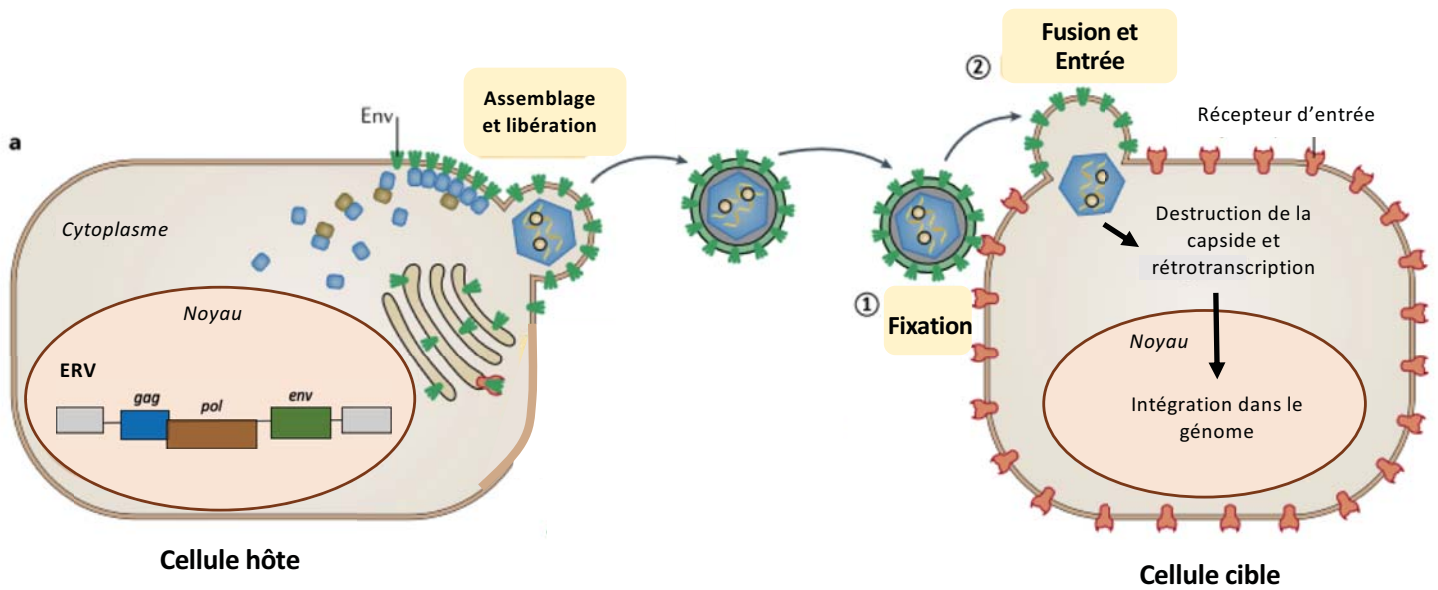
- **Les LTR** : les séquences codantes des ERVs sont encadrées, de part et d'autre, par deux régions LTR identiques. Ces LTR contiennent la plupart des éléments de contrôle transcriptionnel nécessaires à l'expression du rétrotransposon, tel que le promoteur et des séquences régulatrices de type « enhancer ». Chaque LTR est subdivisé en trois domaines fonctionnels : U3, R et U5 (**Figure 7**). La région R est délimitée par le site d'initiation de la transcription (TSS) à gauche et le site de polyadénylation à droite. Le LTR en entier, avec sa structure U3-R-U5, est généré lors de la synthèse de l'ADNc du rétrotransposon par un mécanisme spécifique aux éléments de type rétroviral.
- **La région 5' non traduite (5'UTR : 5' *UnTranslated Region*)** : cette région est localisée entre le LTR 5' et le gène *gag*. De taille très variable, elle fait parfois plus de 300 paires de base (pb).
- **La région PBS (*Primer Binding Site*)** : la région PBS est localisée juste en aval du LTR 5'. Cette région est en partie complémentaire à l'extrémité 3' d'un ARN de transfert (ARNt) exprimé par la cellule. La reconnaissance et l'appariement de la région PBS de l'ARN génomique avec un ARNt cellulaire permet l'amorçage de la transcription inverse.
- **La région PPT (*Poly Purine Tract*)** : la région PPT se localise juste en amont du LTR 3' et permet l'initiation de la synthèse du second brin d'ADN afin de former un ADN double brin pouvant ensuite s'insérer à un nouveau locus génomique.

### I.3. b) Cycle de transposition des rétrotransposons

Outre les homologies de séquences existant entre les rétrovirus exogènes et endogènes, ces deux entités semblent également avoir un cycle de répllication similaire (Echalier, 1989). Ainsi, de nombreuses connaissances concernant le cycle de transposition des ERVs ont été déduites par analyse du cycle de répllication des rétrovirus exogènes.

⇒ Transcription du rétrovirus endogène

L'ERV présent au sein du génome est transcrit par la machinerie transcriptionnelle de la cellule grâce à l'ARN polymérase de type II. L'expression des rétrovirus endogène peut donc être



### Figure 8 : Cycle d'infection d'un rétrovirus endogène

Le rétrovirus endogène (ERV) inséré dans le génome de la cellule hôte est transcrit puis traduit pour produire les protéines Gag, Pol et Env. Les particules se forment sous la face interne de la membrane plasmique grâce aux protéines Gag qui interagissent avec les protéines Env ancrées à la membrane plasmique. Les particules sortent de la cellule par bourgeonnement, incorporant la protéine Env à leur surface. Env permet ensuite la fixation du rétrovirus à la cellule cible via la reconnaissance du récepteur enchâssé dans la membrane de la cellule cible. Cela induit la fusion de la particule avec la membrane de la cellule cible et l'entrée du virus. La capside est ensuite détruite et la transcription inverse est initiée. La rétrotranscription pourrait également être initiée dans la particule virale. L'ADN complémentaire produit peut ensuite s'intégrer dans le génome de la cellule cible.

Figure adaptée de Johnson, 2019

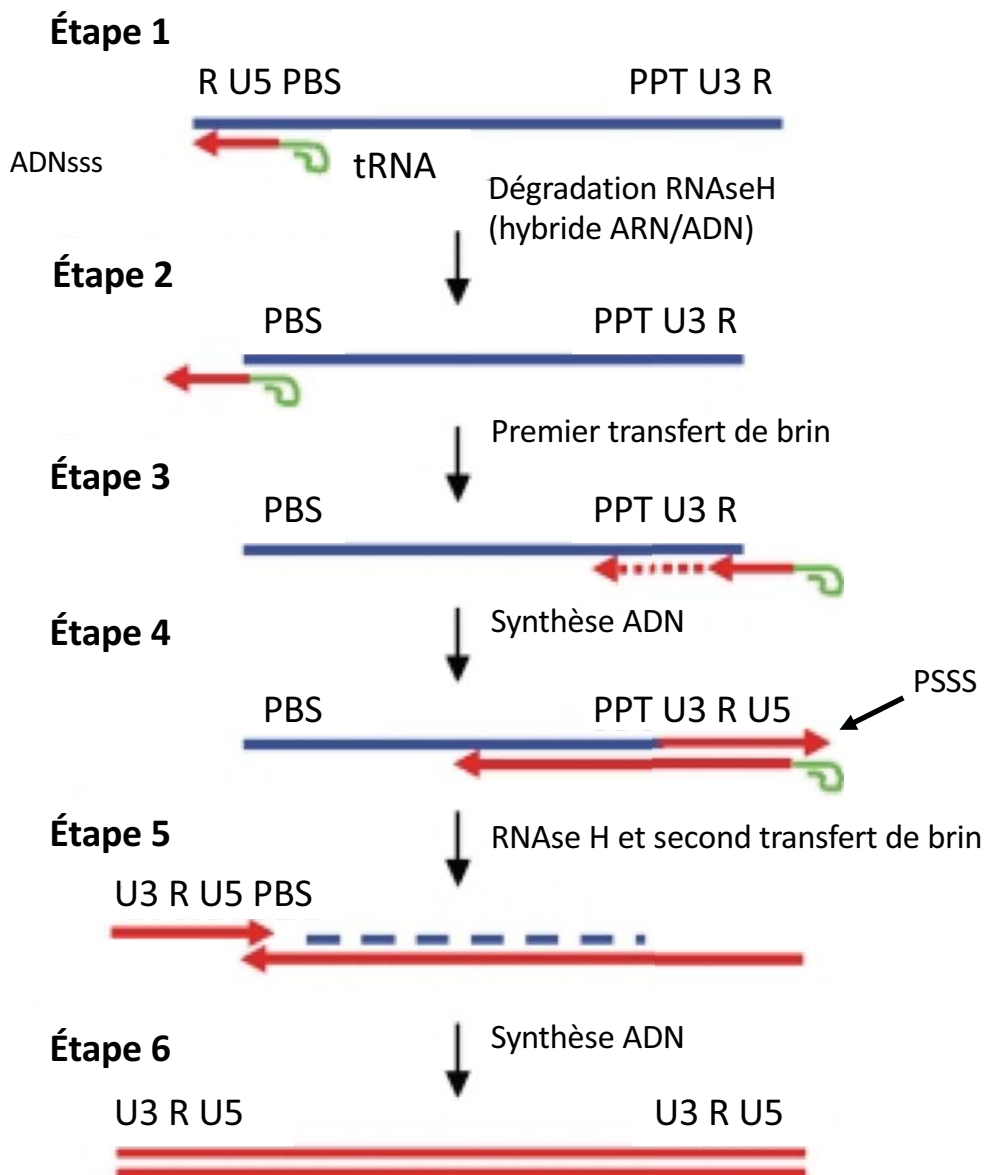
activée par des facteurs de transcription spécifiques, conduisant à une expression tissulaire spécifique. De nombreuses régions régulatrices sont localisées dans l'élément lui-même, mais son expression peut être influencée par les régions régulatrices localisées à proximité du site d'insertion génomique (Cavarec et al., 1997; Meignin et al., 2004).

La transcription est initiée à partir du TSS localisé entre les régions U3 et R du LTR 5', et se termine dans le LTR 3'. L'ARN du rétrovirus ne contient donc pas les LTR en entier mais uniquement les parties R-U5 du LTR en 5' et U3-R du LTR en 3'. Les ARNs des ERVs sont polyadénylés grâce à la présence d'un signal de polyadénylation localisé dans le LTR 3'. Après épissage, deux transcrits sont formés : un transcrit pleine longueur (non épissé), qui code les polyprotéines Gag et Pol, et un transcrit épissé permettant la synthèse de la polyprotéine de l'enveloppe. Les ARNs sont ensuite exportés vers le cytoplasme où ils sont traduits. Dès lors que les protéines Gag et Pol sont synthétisées, la phase d'assemblage des particules peut débuter.

⇒ *Formation et libération des pseudo-particules virales et transmission à la cellule cible*

Comme pour les rétrovirus exogènes, les particules des ERVs se forment au niveau de la membrane plasmique grâce aux protéines Gag qui ont la capacité d'interagir avec la protéine Env localisée à la membrane. L'assemblage permet la formation de particules qui encapsident l'ARN du rétrovirus endogène (supposé pleine longueur). Les particules des rétrovirus exogènes sortent ensuite de la cellule par un processus de bourgeonnement, incorporant la protéine Env à leur surface. Env permet ensuite la liaison de la particule aux récepteurs de la cellule cible, entraînant ainsi l'entrée du virus dans celle-ci (**Figure 8**).

Cependant, le processus permettant la libération des VLP des rétrovirus endogènes est encore mystérieux. En effet, les ERVs codent les mêmes protéines Gag et Env que les rétrovirus exogènes et pourraient donc utiliser le même processus de bourgeonnement pour être libérés et ainsi infecter d'autres cellules. De nombreuses études ont examiné le processus de transmission intercellulaire de *Gypsy*, un ERV exprimé dans les cellules somatiques de l'ovaire de *D. melanogaster* et capable d'envahir des cellules germinales (Marlor et al., 1986; Kim et al., 1994; Péliesson et al., 1994a; Song et al., 1994; Chalvet et al., 1999; Syomin et al., 2001). Une étude récente a notamment démontré que l'infection intercellulaire entre cellules somatiques de *Gypsy* requiert la présence de l'enveloppe et que cette infection ne nécessite pas de contact entre les cellules (Keegan et al., 2021). La transposition intracellulaire semble, elle, être indépendante de la présence de l'enveloppe. Ces données suggèrent que des pseudo-particules virales sont produites par les cellules et sont libérées dans le milieu afin d'infecter d'autres



**Figure 9 : Processus de rétrotranscription du génome rétroviral**

La rétrotranscription du génome viral se déroule en six étapes. La synthèse de l'ADNc commence au niveau du LTR 5', elle est amorcée par la reconnaissance spécifique de la séquence PBS par un ARNt de la cellule hôte (1). S'ensuit une étape de dégradation de l'ARN hybridé à l'ADNc par la ribonucléase H (RNase H) (2), puis un transfert de brin permettant de continuer la rétrotranscription dans le LTR 3' (3). Le PPT permet d'amorcer la rétrotranscription du second brin d'ADNc (4). Une autre étape de dégradation par la RNase H et un second transfert de brin permettent alors d'achever la synthèse des deux brins d'ADNc (5 et 6).

PBS = Primer Binding Site, PPT = Poly Purine Tract, ADN<sub>sss</sub> = ADN strong stop strand-minus DNA, PSSS = *Plus-Strand Strong-Stop DNA*

Figure adaptée de Delviks-Frankenberry et al., 2011

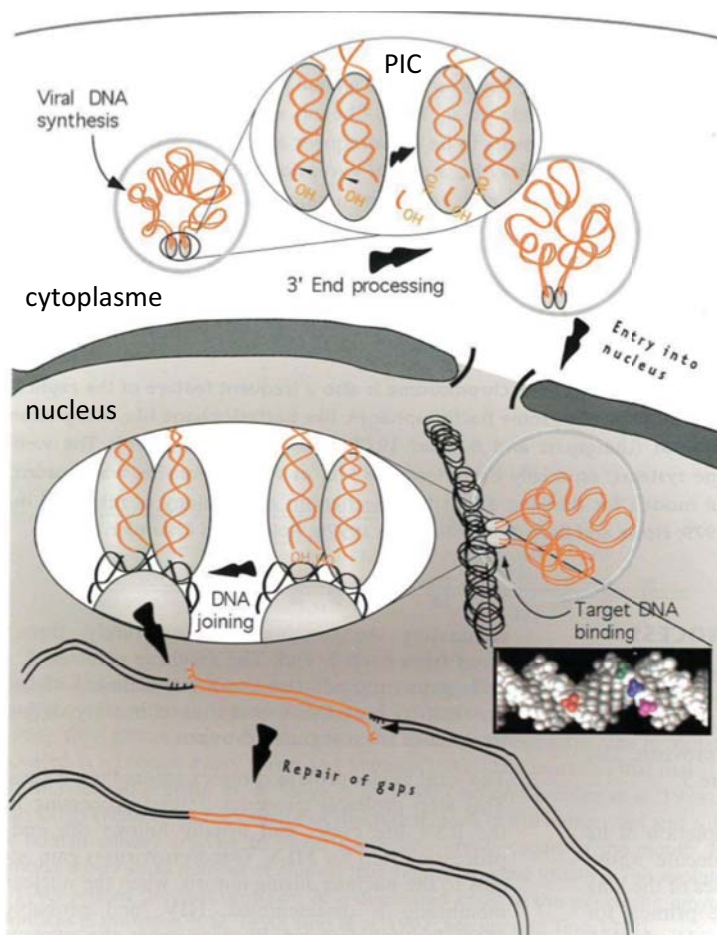
cellules, permettant ainsi à *Gypsy* de transposer de manière intercellulaire. Cependant, à l'heure actuelle, aucune étude n'a permis de révéler la présence de particules bourgeonnantes ou extracellulaires. Il a également été montré que les ERVs peuvent utiliser d'autres processus pour transiter entre les cellules. Ils pourraient détourner des processus cellulaires pour atteindre leurs cellules cibles, c'est par exemple le cas de l'ERV *ZAM* qui semble utiliser le trafic de vitellogénine pour envahir les cellules germinales de l'ovaire (Brasset et al., 2006). Le rôle de l'enveloppe dans le transfert intercellulaire de *ZAM* reste néanmoins encore inconnu. Le transit intercellulaire de *ZAM* sera développé plus loin dans l'introduction et plus d'informations sur les processus d'invasion des rétrotransposons sont disponibles dans la revue présentée dans la partie II de l'introduction.

⇒ Synthèse du nouvel ADN rétroviral et intégration dans l'ADN de la cellule hôte

Lorsque la particule pénètre dans le cytoplasme de la cellule cible, celle-ci subit une décapsidation qui se définit par une déstructuration de la capsid virale. Cette étape est nécessaire pour permettre d'initier la transcription inverse du génome rétroviral en ADN (Dismuke and Aiken, 2006). Il est intéressant de noter que des études récentes ont montré que certains virus exogènes comme le virus HIV-1 peuvent pénétrer encapsidés dans la cellule hôte. La capsid protège ainsi l'ARN génomique viral des machineries antivirales de l'hôte. De plus, des protéines se lient à l'extérieur de la capsid pour faciliter son transit jusqu'au noyau. Des virus encapsidés ont même été retrouvés dans le noyau, suggérant que la capsid peut passer à travers les pores nucléaires (Zila et al., 2021). Dans ce cas-là, la transcription inverse se produit même directement dans la capsid (Yamashita and Engelman, 2017; Bedwell and Engelman, 2021). L'ARN (brin positif) est présent en double copies pleine longueur dans la particule virale, et la transcription inverse va donner naissance à une molécule d'ADN complémentaire double brin encadrée par deux LTR (Revue dans : (Delviks-Frankenberry et al., 2011)). La transcription inverse se déroule en six étapes (**Figure 9**):

- 1- La synthèse de l'ADNc commence au niveau du PBS. En effet, elle est amorcée par la reconnaissance spécifique de la séquence PBS de l'ARN génomique par un ARNt de la cellule hôte. Cela permet à la rétrotranscriptase de synthétiser un petit fragment d'ADNc contenant les régions U5 et R du LTR 5'. Ce court ADNc est appelé « *strong stop strand-minus DNA* » (ADN<sub>ss</sub>).

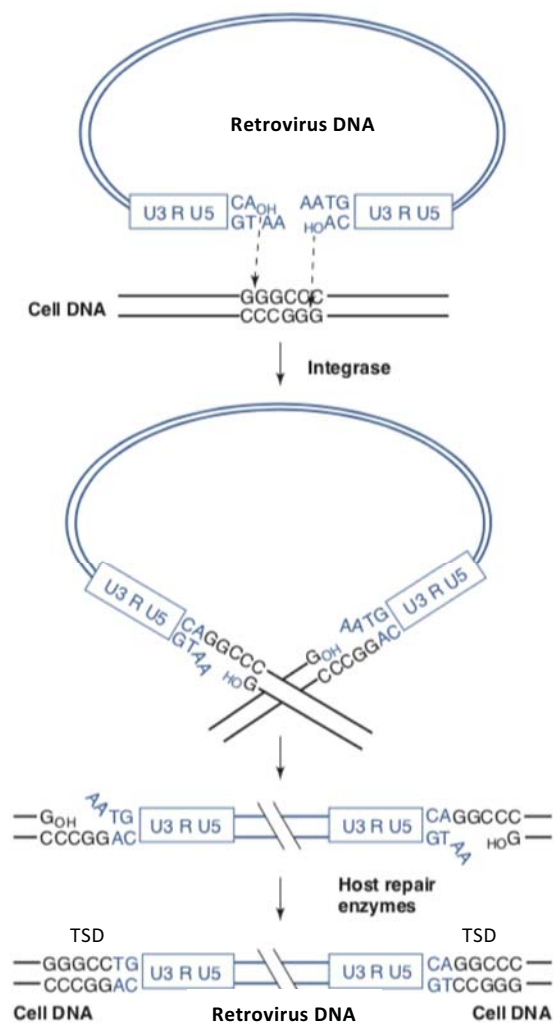




**Figure 10 : Entrée et intégration de l'ADN du rétrovirus**

L'ADN viral synthétisé se lie aux protéines Intégrases pour former le complexe de pré-intégration PIC. L'intégrase induit le clivage de deux nucléotides à chaque extrémité 3' de l'ADN rétroviral, générant des extrémités 3'-OH servant à l'intégration dans l'ADN de la cellule hôte. Le complexe PIC pénètre via les pores nucléaires dans le noyau où l'intégrase clive l'ADN génomique permettant l'intégration du rétrovirus. Après réparation par les protéines cellulaires, le rétrovirus est intégré de manière stable dans le génome.

Figure issue du livre « Retroviruses » (Coffins, 1997)



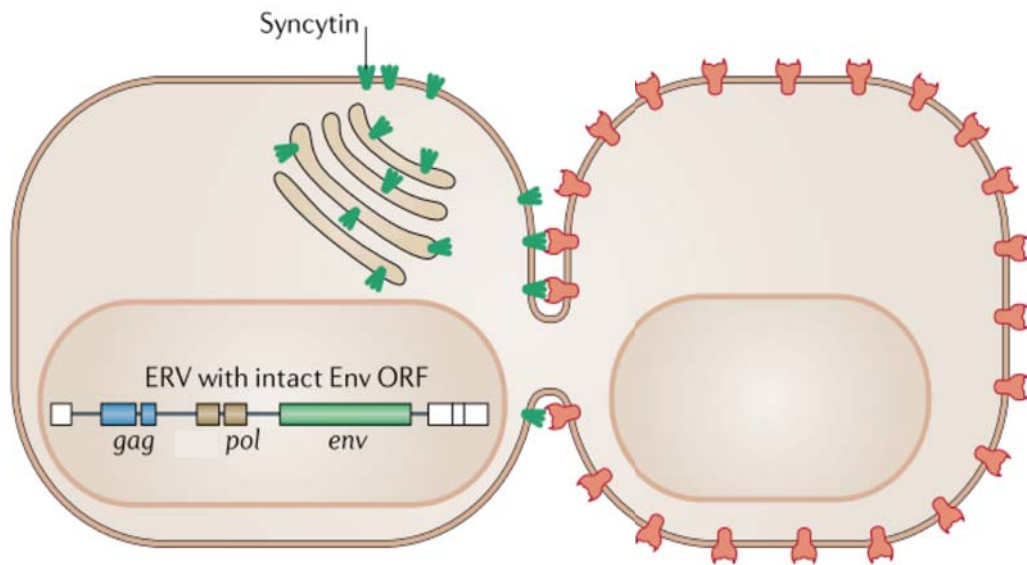
**Figure 11 : Intégration de l'ADN du rétrovirus dans l'ADN de la cellule hôte**

L'intégrase du rétrovirus clive deux nucléotides au niveau de l'ADN rétroviral, libérant des extrémités 3'-OH, puis joint ces extrémités 3'-OH à l'ADN de l'hôte, préalablement coupé par l'intégrase, dans une réaction concertée de clivage-ligation à des sites distants de 4 à 6 nucléotides sur l'ADN de l'hôte (flèches en pointillés). Les enzymes de la cellule hôte réparent ensuite les jonctions. L'insertion génère une duplication au niveau du site d'insertion (TSD). L'ADN du rétrovirus est en bleu ; l'ADN hôte est en noir.

Figure issue du livre « Fundamentals of Molecular Virology » (Acheson, 2011)

- 2- La ribonucléase H (RNase H) dégrade ensuite l'ARN du rétrovirus hybridé avec l'ADNsss.
- 3- L'ADNsss contenant la région R et U5 peut ainsi aller se fixer par complémentarité de séquence à la séquence R du LTR 3' de l'ARN. Cela permet de continuer l'initiation de la transcription inverse à partir du LTR 3' et de rétrotranscrire les séquences internes.
- 4- L'ARN génomique servant de matrice est dégradé en 3', sauf la région PPT qui est résistante à la RNase H et sert alors d'amorce à la synthèse du brin d'ADN positif. La synthèse est stoppée une fois la région U5 et l'ARNt atteints, cette portion du brin nouvellement synthétisé est appelée PSSS (*Plus-Strand Strong-Stop DNA*).
- 5- L'ARNt est alors dégradé, libérant la région PBS du PSSS. Un second transfert de brin permet au PSSS de s'hybrider au niveau de la séquence complémentaire au PBS du brin moins, permettant ainsi d'achever la synthèse des deux brins, plus et moins.
- 6- Une molécule d'ADN double brin encadrée de deux LTR parfaitement homologues est ainsi synthétisée.

La molécule ADNc nouvellement formée est associée à l'Intégrase (IN), protéines codées par le rétrotransposon, pour former le complexe de pré-intégration PIC (*PreIntegration Complex*). Le complexe PIC pénètre ensuite dans le noyau *via* le passage par les pores nucléaires (NPC = *nuclear pore complex*). La protéine IN possède un signal de localisation nucléaire qui permet le recrutement du complexe PIC au niveau des pores nucléaires, ainsi que le recrutement de protéines de transport qui permettront l'entrée du complexe dans le noyau (Matreyek and Engelman, 2013). L'intégrase clive l'ADN du rétrovirus en libérant deux ou trois nucléotides à chaque extrémité 3' des deux brins, ces extrémités 3'OH générées serviront de site d'attachement à l'ADN de la cellule hôte lors de l'intégration. Une fois dans le noyau, l'intégrase clive également l'ADN génomique en générant une coupure décalée variant suivant le rétrovirus (4 à 6 pb). Cette étape permet la génération d'extrémités 5' sortantes. Les jonctions manquantes sont réparées par la synthèse d'ADN, induisant une duplication du site d'insertion (TSD = *Target Site Duplication*). Après une étape de ligation, le rétrovirus est intégré de manière stable dans le génome de la cellule hôte (Coffin et al., 1997; Acheson, 2011) (**Figures 10 et 11**). Il est alors transcrit comme n'importe quel gène cellulaire et un nouveau cycle d'infection/transposition peut débuter.



**Figure 12 : Schéma représentant le rôle des Syncytines, protéines produites à partir du gène *env* d'un rétrovirus endogène domestiqué**

Des rétrovirus endogènes ont été domestiqués pour produire les protéines Syncytines chez de multiples espèces de mammifères, dont l'Homme, la souris et le mouton. Chez l'Homme, les gènes *env* de deux rétrovirus endogènes des familles *HERV-W* et *HERV-FRD* ont été domestiqués pour donner les gènes Syncytin-1 et -2 respectivement. Grâce à leur ancrage membranaire, les Syncytines peuvent se lier aux récepteurs de la cellule cible et induire une fusion membranaire entre les cellules. Ce processus permet la fusion des cellules de la membrane extracellulaire fœtale avec celles du tissu utérin maternel, processus nécessaire à la formation du placenta.

Figure issue de Johnson, 2019.

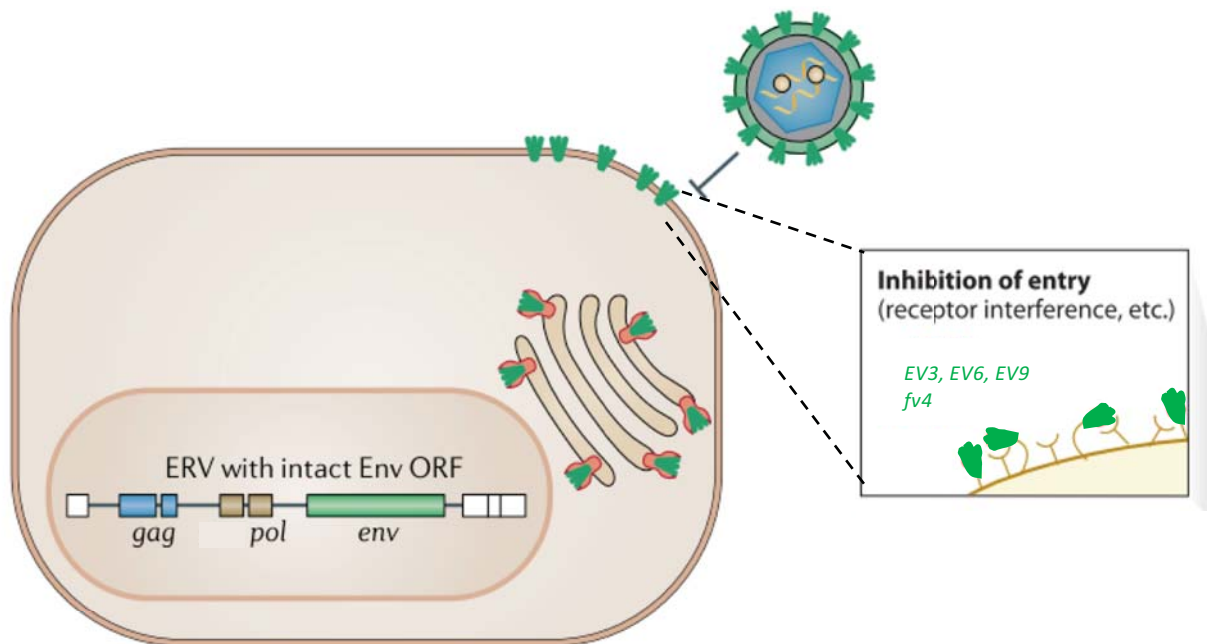
### I.3. c) Impact de la présence des ERVs dans les génomes

Les ERVs sont avant tout des éléments transposables et, depuis leur découverte dans les années 1950, leur impact puissant sur la régulation des gènes n'est plus à démontrer. Mais les ERVs ont également été décrits dans de nombreux processus de domestication. Le mot domestication a été proposé pour la première fois par Weldon J. Miller *et al* pour désigner la conversion d'une séquence codante dérivée de l'élément *P* en un gène intégré de l'hôte (Miller et al., 1992). À l'heure actuelle, on parle de domestication lorsque n'importe quelle fonction d'un ET est cooptée par l'organisme hôte afin de servir de nouvelles fonctions cellulaires et ainsi apporter des nouveautés génétiques adaptatives. La plupart des domestications concernant les ERVs impliquent les gènes *env* et *gag*.

⇒ Les ERV : à l'origine de l'apparition des animaux placentaires

Un des exemples les plus documentés concerne la domestication du gène *env* issu du rétrovirus endogène humain *HERV-W* ayant donné naissance aux protéines Syncytines chez l'Homme (Blond et al., 2000; Mi et al., 2000; Blaise et al., 2003). Ces protéines sont des protéines membranaires notamment responsables de la fusion cellulaire entre les cellules de la membrane extra-embryonnaire fœtale et celles du tissu utérin maternel, processus nécessaire à la formation du placenta. Les glycoprotéines Syncytines, produites à partir de l'ancien gène rétroviral *env*, s'ancrent à la membrane plasmique et peuvent ainsi se lier aux récepteurs de la cellule cible et induire une fusion membranaire entre les cellules (**Figure 12**). Ce mécanisme est comparable à celui permettant l'entrée des rétrovirus dans leur cellule hôte (processus dépendant d'une liaison aux récepteurs pour déclencher la fusion des membranes de la particule et de la cellule). Le plus étonnant est le fait que la domestication d'un ERV pour former les protéines Syncytines s'est produite de manière indépendante dans des ordres de mammifères différents, humain, souris, et mouton entre autres ((Palmarini et al., 2001; Dupressoir et al., 2005; Heidmann et al., 2009), Revue dans : (Dupressoir et al., 2012; Laviaille et al., 2013)). Les ERVs ayant été cooptés ne sont même pas étroitement apparentés, ni intégrés au même site génomique. C'est un exemple remarquable d'évolution convergente !

⇒ Relation étroite entre rétrovirus endogènes et rétrovirus exogènes : avantage ou inconvénient ?



**Figure 13 : Schéma représentant le rôle potentiel des protéines Env codées par les ERVs dans le blocage de l'infection par un rétrovirus exogène**

Les protéines Env produites par les rétrovirus endogènes peuvent inhiber l'entrée de virus exogènes en bloquant l'accès aux récepteurs cellulaires. Ce phénomène d'interférence a notamment été observé avec les protéines EV3, EV6, EV9 (poulet) et fv4 (souris) issues de gènes dérivés de séquences de rétrovirus endogènes. Ce mécanisme est similaire à celui de « l'interférence à la surinfection » observé chez les virus, qui est la capacité d'une infection virale établie à interférer avec une infection secondaire par le même virus ou par un virus étroitement apparenté.

Figure issue de Johnson, 2019

En plus de leur capacité à induire la fusion membranaire, il semble que certaines protéines Env aient été domestiquées pour leur capacité de compétition. La présence d'*env* de rétrovirus endogènes interfère avec l'infection d'un rétrovirus exogène : les protéines Env peuvent en effet inhiber l'entrée de ce virus exogène en bloquant l'accès aux récepteurs cellulaires (**Figure 13**). De tels phénomènes ont été observés dans un premier temps chez le poulet (les loci *EV3*, *EV6* et *EV9* dérivés de rétrovirus endogènes bloquent l'infection au virus de la leucose aviaire) (Robinson et al., 1981). Ils ont ensuite été retrouvés chez les souris où le locus *fv4*, dérivé du gène *env* du rétrovirus *MLV*, confère une résistance à l'infection du virus de la leucémie murine (Ikeda and Odaka, 1984; Kozak et al., 1984).

Les protéines Env sont normalement ancrées à la membrane cellulaire. Cependant, il a été suggéré que ce phénomène de compétition pour les récepteurs des virus pourrait faire intervenir des protéines Env d'ERV sécrétées (revue dans : (Johnson, 2019)). Dans ce cas, les protéines sécrétées auraient subi des mutations induisant une troncation prématurée, éliminant ainsi le domaine de liaison à la membrane de la protéine Env. Il existe notamment dans le génome de l'Homme un gène, nommé *suppressyn*, codant pour une protéine Env tronquée. Cette protéine se lie aux récepteurs ASCT2 et pourrait bloquer l'entrée de virus utilisant ce même récepteur (Sugimoto et al., 2013). De plus, le génome humain abrite des centaines de loci dérivant de gènes *env* (revue dans : (De Parseval and Heidmann, 2005)); il est probable qu'un certain nombre de ces gènes possèdent des activités antivirales, mais également d'autres activités encore non élucidées. Enfin, quelques études rapportent aussi la domestication des gènes *gag* et *pol* en tant que facteurs de restriction aux infections virales (Marco and Marín, 2009).

Ainsi, la domestication des ERVs pourrait apporter un certain avantage dans la lutte contre les virus exogènes. Néanmoins, il est intéressant de noter que la proximité entre les rétrovirus exogènes et endogènes peut également s'avérer néfaste pour l'hôte : les rétrovirus endogènes pourraient collaborer avec les virus exogènes lors de l'infection par complémentation (Belshaw et al., 2004; Bannert and Kurth, 2006; Halo et al., 2019). Une étude a notamment montré que lorsque des cellules sont infectées par un rétrovirus *RSV* (Virus du Sarcome de Rous) muté pour le gène *env*, le virus peut tout de même infecter et pénétrer dans les cellules lorsque celles-ci possèdent dans leur génome un rétrovirus endogène avec une structure proche et codant une enveloppe fonctionnelle (rétrovirus endogène *RAV-0*) (Tschlis and Coffin, 1980). Cet exemple montre que les rétrovirus exogènes peuvent utiliser des facteurs produits par les ERVs pour optimiser leur propagation.



De plus, les rétrovirus exogènes pourraient effectuer des recombinaisons avec des ERVs résidents dans le génome de la cellule hôte, induisant la création de nouveaux variants (Jenkins and Copeland, 1985). Par exemple, le virus ALV (Virus de la Leucose Aviaire) du sous-groupe J, qui est à l'origine d'épidémie grave chez les volailles, est un recombinant chez lequel un virus ALV exogène a acquis le gène *env* d'un rétrovirus endogène (Benson et al., 1998).

En parlant de désavantages, l'expression d'ERVs a été détectée dans de nombreuses pathologies, notamment dans des cancers, des pathologies auto-immunes et neurologiques (Löwer, 1999; Cherkasova et al., 2016; Zhou et al., 2016; Tam et al., 2019b). Dans ce cas, les protéines exprimées par les ERVs semblent avoir un rôle immunosuppresseur, leur expression est donc désavantageuse pour la survie de l'hôte ((Mangeny and Heidmann, 1998), revue dans : (Gonzalez-Cao et al., 2016)).

Enfin, et nous en parlerons plus en détail dans la partie suivante, l'insertion dans le génome des ERVs peut bien sûr induire des mutations délétères, des recombinaisons et réarrangements génomiques, ou encore altérer l'expression des gènes.

Finalement, rien n'est jamais noir ou blanc. La présence d'ERVs au sein des génomes peut être un réel atout, mais peut aussi se révéler dangereuse. Un contrôle efficace par leur hôte est nécessaire afin de limiter de potentiels effets délétères. Cette vérité est valable pour tous les éléments transposables (et pour tout système en réalité : tout n'est qu'une question d'équilibre). Dans la partie qui suit nous allons rentrer dans le vif du sujet : comment les éléments transposables et leurs hôtes survivent et évoluent ensemble ? Nous essayerons de mieux comprendre cette balance, nécessaire et pourtant si fragile, qui existe entre expression et répression des éléments transposables.

## II- Relation complexe entre les éléments transposables et leurs hôtes

Dès leur découverte dans les années 1950, Barbara McClintock émit l'hypothèse que les éléments transposables avaient la capacité de sculpter les génomes et d'influencer l'expression des gènes. Cette chercheuse avait déjà compris que les éléments transposables étaient sources de mutations et donc potentiellement dangereux, mais elle avait également perçu à quel point ces éléments pouvaient influencer notre génome en bien des façons. Ainsi, elle annonça :

*« It is now known that controlling elements may **modify gene action** in a number of different ways. They may influence the **time of gene action** in the development of a tissue, and also determine the cells in which it will occur. Again, they may influence the **type of action**, with regards to either degree (**quantitative aspects**) or kind (**qualitative aspects**). »*





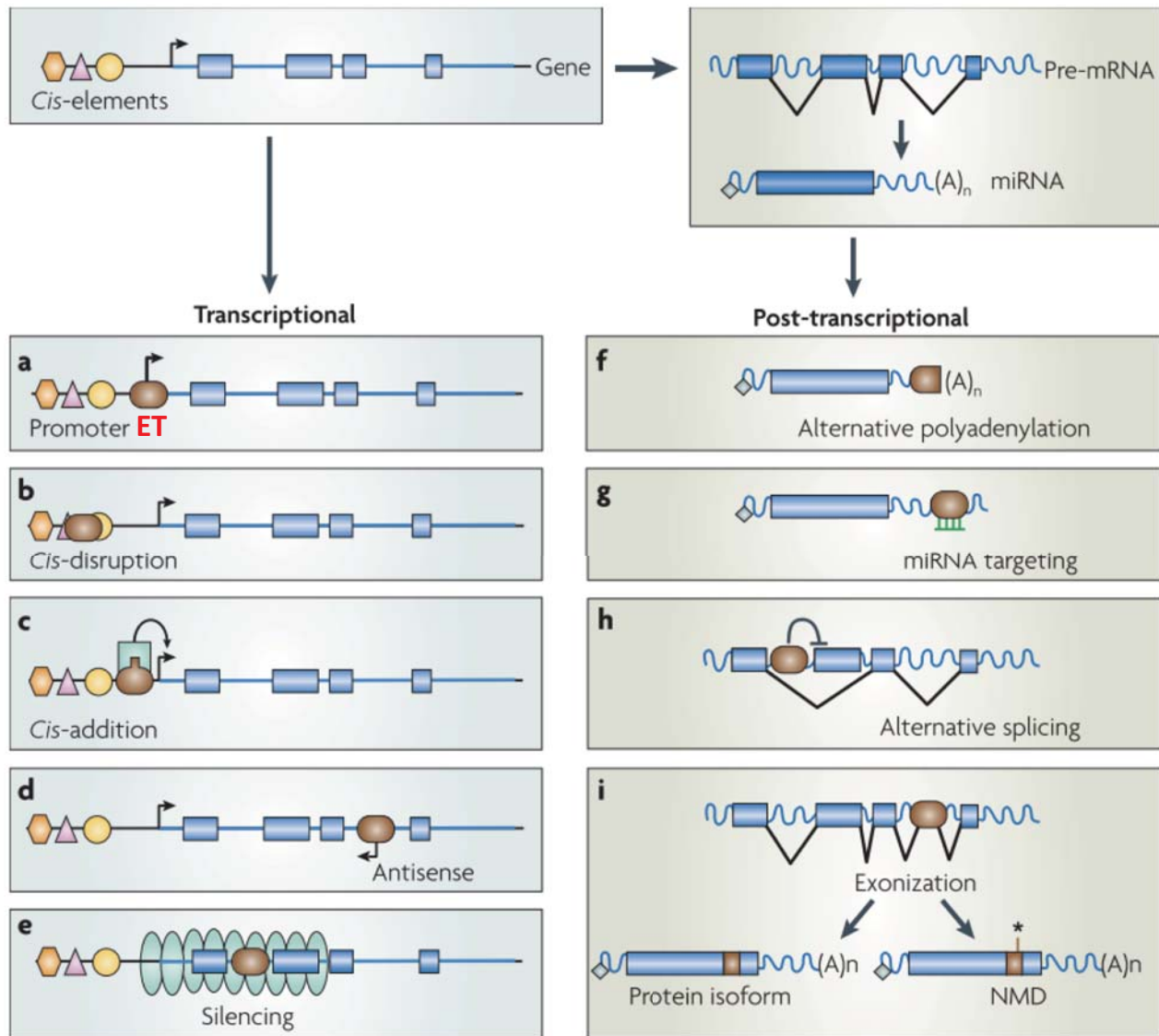
Aujourd'hui, les connaissances accumulées ont confirmé le rôle important des éléments transposables dans la régulation des gènes et, de manière plus générale, sur l'évolution des génomes. Ils présentent des caractéristiques uniques et extrêmement avantageuses pour influencer les génomes : 1) ils sont mobiles et peuvent donc se déplacer dans les génomes, véhiculant ainsi des éléments génétiques, tels que des promoteurs, des enhanceurs et des gènes ; 2) ils sont présents de manière universelle : on a retrouvé ces éléments dans la plupart des espèces, allant des bactéries à l'Homme ; 3) ils semblent très anciens : on soupçonne que les transposons à ADN étaient déjà présents dans l'ancêtre commun des eucaryotes (Feschotte and Pritham, 2007) ; 4) ils peuvent être très abondants dans les génomes : à titre d'exemple 80 % du génome du maïs est composé d'ETs (Schnable et al., 2009) ; 5) les ETs sont des mutagènes puissants !

## **II.1 - Influence des éléments transposables sur l'évolution des génomes**

Les ETs peuvent agir sur les génomes de bien des manières : apport de nouvelles fonctions cellulaires, modification de l'organisation génomique, influence sur l'expression des gènes... Quelques exemples seront donnés dans la partie qui suit, mais ce n'est qu'une faible représentation des nombreuses découvertes faites jusqu'à présent. Et ce n'est probablement que le début, les ETs n'ont pas fini de nous étonner !

### **II.1. a) Domestication : comment les génomes s'approprient-ils les ETs ?**

La domestication, dont nous avons parlé dans le chapitre précédent, est un mécanisme efficace pour acquérir de nouvelles fonctions. Elle n'est pas réservée uniquement aux ERVs, toute fonction potentielle d'un ET peut être cooptée par l'organisme hôte. La transposase, codée par les transposons à ADN, présente notamment des fonctions qui peuvent se révéler très utiles pour son hôte. À titre d'exemple, les protéines RAG1 et RAG2, essentielles pour la recombinaison V(D)J se produisant dans les lymphocytes, sont dérivées de la transposase de l'élément *Transib* ayant conservé sa fonction d'endonucléase (Sakano et al., 1979; Kapitonov and Jurka, 2005).



**Figure 14 : Impact des éléments transposables sur l'expression des gènes**

Au niveau transcriptionnel, un ET (représenté en marron) qui s'est inséré en amont d'un gène peut ajouter des séquences promotrices et introduire un site alternatif d'initiation de transcription (a), perturber le ou les éléments *cis*-régulateurs existants (b), ou encore introduire un nouvel élément *cis* tel qu'un site de fixation à un facteur de transcription (c). Un ET qui s'est inséré dans un intron peut aussi entraîner une transcription antisens et potentiellement interférer avec la transcription sens (d). Enfin, une insertion d'un ET peut induire une hétérochromatinisation du locus (ovales verts), ce qui peut conduire à une inhibition de la transcription d'un ou de plusieurs gènes adjacents (e).

Au niveau post-transcriptionnel, un ET qui s'est inséré dans l'UTR 3' d'un gène peut introduire un site de polyadénylation alternatif (f) ou un site de liaison pour un micro-ARN (g). Un ET qui s'est inséré dans un intron peut interférer avec l'épissage d'un pré-ARNm (h), provoquant la formation de diverses formes d'épissage alternatif (par exemple, rétention d'intron et saut d'exon). Un ET qui s'est inséré dans un intron et qui contient des sites d'épissages peut être incorporé (« exonisé ») comme un exon alternatif (i). Cela peut entraîner la traduction d'une nouvelle isoforme protéique, la déstabilisation ou la dégradation de l'ARNm, en particulier si l'ET exonisé introduit un codon stop prématuré (représenté par un astérisque).

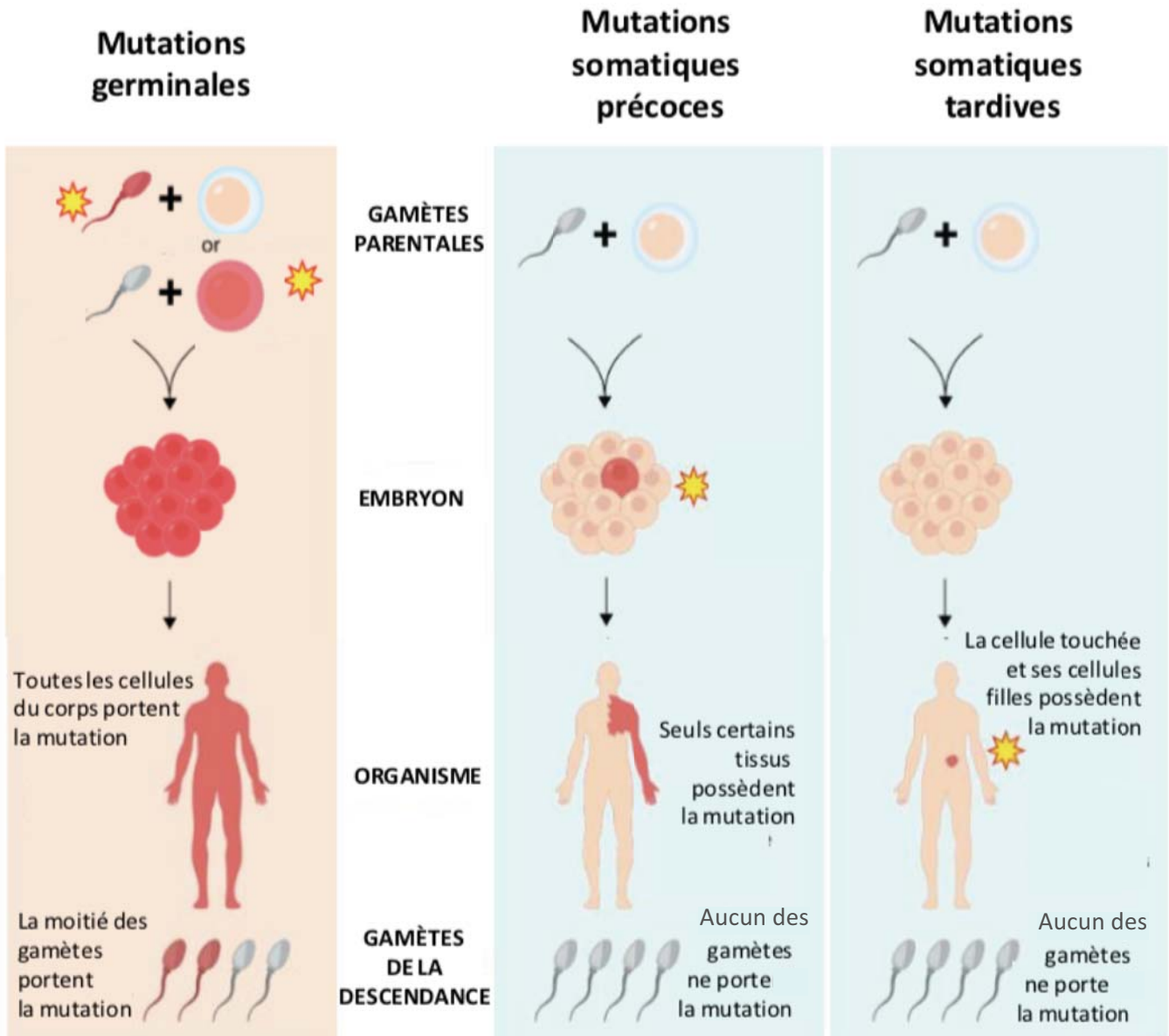
Figure issue de Feschotte, 2008

En ce qui concerne l'impact des ETs sur l'architecture génomique, on peut citer à titre d'exemple la présence de trois rétrotransposons sans LTR - *TART*, *TAHRE* et *Het-A* - au niveau des télomères chez *Drosophila melanogaster* (Biessmann and Mason, 1992). La drosophile ne possède pas de gène codant la télomérase, protéine permettant de conserver l'intégrité des télomères dans les cellules germinales des mammifères. C'est par l'intermédiaire de la transposition successive de ces trois ETs présents au niveau des télomères que les extrémités des chromosomes sont préservées dans les cellules germinales de *D. melanogaster* (Revue dans : (Pardue and DeBaryshe, 2003)).

Autre exemple, les rétrotransposons à LTR sont à l'origine des centromères chez certaines plantes (Cheng and Murata, 2003). Présents en multiples copies dans les génomes, les ETs peuvent recombinaison entre eux et induire des réarrangements génomiques tels que des duplications, des délétions ou encore des inversions. Ces processus contribuent également grandement à la modification globale de l'organisation du génome. La mobilisation des ETs sculpte donc le génome de leur hôte, et les organismes ont parfois réussi à détourner leur fonction première de transposition pour permettre la création de nouvelles fonctions.

## II.1. b) Effet des ETs sur la régulation des gènes

Prenons maintenant un peu de recul. La plupart des insertions n'aboutissent pas à la création de nouvelles fonctions à proprement parler : les cas de domestications des séquences codantes des ETs restent encore rares et exceptionnels. Cependant, de manière générale, en s'insérant aléatoirement dans le génome, les éléments transposables peuvent perturber et modifier des fonctions existant au sein de l'hôte. Leur action peut s'effectuer aussi bien au niveau transcriptionnel que post-transcriptionnel (Feschotte, 2008) (**Figure 14**). Les nouvelles insertions peuvent avoir un impact plus ou moins important en fonction du site génomique. L'insertion d'un ET dans la région codante d'un gène peut bien évidemment perturber son cadre de lecture, et ainsi conduire à la formation d'une protéine tronquée ou à l'apparition précoce d'un codon stop, mais il ne faut pas oublier que les ETs possèdent également leurs propres régions régulatrices : ils ont des promoteurs, des sites de fixation à divers facteurs de transcription, des régions « *enhancer* » (amplificateurs) et « *silencer* » (inactivateurs), ou encore des sites accepteurs et donneurs d'épissage. Ces régions régulatrices sont nécessaires à la biologie de l'élément, mais elles peuvent aussi modifier la régulation de gènes cellulaires voisins. Il a été montré que pas moins de 25 % des régions promotrices des gènes contiennent des séquences dérivées d'ETs, ce qui peut affecter leur expression (Revue dans : (Jordan et al.,



**Figure 15 : Conséquences des mutations germinales *versus* somatiques**

Les mutations dans les cellules germinales sont transmises à la descendance tandis que les mutations somatiques (précoces ou tardives) affectent uniquement l'individu, sans transmission à la descendance.

Figure adaptée de <https://kintalk.org/genetics-101/>

2003)). Les éléments de type SINEs par exemple, comme les séquences *Alu*, sont connus par exemple pour être particulièrement riches en sites de fixation de facteurs de transcription, pouvant ainsi générer de nouveaux enhanceurs (Norris et al., 1995; Babich et al., 1999; Polak and Domany, 2006). On peut ici citer l'exemple du gène *ISL1* qui est régulé depuis plus de 410 millions d'années par l'activité enhancer d'un élément de type SINE (Bejerano et al., 2006). Les séquences *Alu* sont également souvent impliquées dans les modifications d'épissage des gènes dans lesquels ils sont insérés. Un exemple typique est l'insertion de l'élément *Alu* dans un exon du gène *BRAC1* qui provoque une modification d'épissage, évènement corrélé avec l'apparition de cancers du sein (Miki et al., 1996). La mutation de gènes suppresseurs de tumeurs à la suite d'une insertion d'ETs n'est pas un phénomène rare, et de nombreuses études suggèrent l'implication des éléments mobiles dans l'apparition ou la progression des cancers. ((Miki et al., 1992; Zhou et al., 2016), revue dans : (Goodier and Kazazian, 2008; Chénais, 2013; Hancks and Kazazian, 2016; Burns, 2017)).

Il est important de rappeler ici que les mutations, induites par des éléments transposables ou d'autres facteurs, n'ont pas toutes le même impact en fonction du type cellulaire où elles se produisent. Ainsi, les mutations se produisant dans les cellules somatiques pourront affecter l'hôte et induire l'apparition de pathologies chez celui-ci, mais elles ne seront jamais transmises à la descendance. En revanche, si les mutations ont lieu dans la lignée germinale, elles seront transmises à la descendance et retrouvées dans toutes les cellules de la progéniture. Cela peut donc potentiellement induire l'apparition de maladies héréditaires (**Figure 15**).

## **II.2 - ETs *versus* Hôtes : qui convertira le prochain essai ?**

On comprend alors facilement que l'expression des ETs doit être finement régulée. Leur présence dans les génomes semble essentielle pour l'évolution des espèces, mais une transposition massive de ces éléments mobiles peut engendrer des effets néfastes sur son hôte et sur sa descendance. Tout n'est qu'une question de contrôle et d'équilibre. C'est la raison pour laquelle les organismes ont développé des stratégies très efficaces pour protéger leurs génomes face aux ETs, particulièrement dans les cellules germinales. Nous avons analysé, dans la revue intégrée ci-dessous, l'équilibre complexe existant entre les ETs et leurs hôtes (Yoth et al., 2022). Nous avons essayé de mieux comprendre comment les ETs réussissent à contourner les systèmes de répression de leurs hôtes pour se propager dans les génomes et comment, en retour, les organismes s'adaptent pour lutter contre cette invasion massive et contrôler leur propagation. La relation ET-hôte est un magnifique exemple de coadaptation !



Review

# The Intricate Evolutionary Balance between Transposable Elements and Their Host: Who Will Kick at Goal and Convert the Next Try?

Marianne Yoth, Silke Jensen  and Emilie Brasset \* 

iGrED, CNRS, INSERM, Faculté de Médecine, Université Clermont Auvergne, 63000 Clermont-Ferrand, France; marianne.yoth@uca.fr (M.Y.); silke.jensen@uca.fr (S.J.)

\* Correspondence: emilie.brasset@uca.fr

**Simple Summary:** Transposable elements (TEs) are mobile DNA sequences that can jump from one genomic locus to another and that have colonized the genomes of all living organisms. While TE mobilization is an important source of genomic innovations that greatly contribute to the host species evolution, it is also a major threat to genome integrity that can lead to pathologies. In this review, we discuss how TEs successfully bypass the host silencing machineries to propagate in the host's genome and how hosts engage in a fightback against TE invasion and propagation. We highlight how TEs and their hosts have been evolving together to achieve a fine balance between transposition activity and repression.

**Abstract:** Transposable elements (TEs) are mobile DNA sequences that can jump from one genomic locus to another and that have colonized the genomes of all living organisms. TE mobilization and accumulation are an important source of genomic innovations that greatly contribute to the host species evolution. To ensure their maintenance and amplification, TE transposition must occur in the germ cell genome. As TE transposition is also a major threat to genome integrity, the outcome of TE mobility in germ cell genomes could be highly dangerous because such mutations are inheritable. Thus, organisms have developed specialized strategies to protect the genome integrity from TE transposition, particularly in germ cells. Such effective TE silencing, together with ongoing mutations and negative selection, should result in the complete elimination of functional TEs from genomes. However, TEs have developed efficient strategies for their maintenance and spreading in populations, particularly by using horizontal transfer to invade the genome of novel species. Here, we discuss how TEs manage to bypass the host's silencing machineries to propagate in its genome and how hosts engage in a fightback against TE invasion and propagation. This shows how TEs and their hosts have been evolving together to achieve a fine balance between transposition and repression.

**Keywords:** transposable elements; endogenous retroviruses; horizontal transfer; piRNA; piRNA cluster; genomic instability; inheritance



**Citation:** Yoth, M.; Jensen, S.; Brasset, E. The Intricate Evolutionary Balance between Transposable Elements and Their Host: Who Will Kick at Goal and Convert the Next Try? *Biology* **2022**, *11*, 710. <https://doi.org/10.3390/biology11050710>

Academic Editor: Wojciech Makalowski

Received: 7 March 2022

Accepted: 26 April 2022

Published: 6 May 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Transposable elements (TEs) are DNA sequences that can move and multiply within the genome by transposition. They were discovered by Barbara McClintock in the maize genome in 1950 [1]. Since then, TEs have been found in all living organisms in which they have been searched for. The TE community is still debating whether there are organisms without TE sequences in their genome. Their amplification within genomes leads to the formation of families of repeated sequences that can be present, for some of them, in thousands of copies spread in all chromosomes. They are currently classified in two categories [2]: (i) class I elements, also referred to as retrotransposons, that transpose through a copy-and-paste mechanism, and (ii) class II elements, referred to as DNA transposons, that transpose through a cut-and-paste mechanism (excision and re-insertion at



a new locus). TE transposition is a major source of genetic instability, especially through chromosome breakages and insertions that result in mutations, ectopic recombination, and genomic rearrangements. To preserve genome integrity, TE mobilization is strictly controlled by several elaborate defense mechanisms, such as silencing strategies based on KRAB zinc-finger proteins, small RNAs, DNA methylation, and chromatin modifications [3–7]. Moreover, most transposon sequences accumulate mutations that do not allow them to produce the proteins required for their transposition and that may ultimately lead to elimination of all active copies of that mobile element in a population. The combined actions of silencing, mutations, and elimination should result in the complete elimination of TEs from the genome. Yet, TEs represent a large part of the genome in all organisms (prokaryotes, unicellular and multicellular eukaryotes), ranging from 3% in yeast to 85% in maize [8]. Indeed, it is assumed that TE movement and accumulation are an important source of genomic and epigenomic variations that strongly influence the species evolution and adaptation to changing environments [9–12]. However, it remains unclear how TEs persist in the genome, reach such high proportions, and expand in all living species, while transposition is strictly controlled. Here, we review the strategies developed by TEs for effective spreading and to bypass the host silencing machinery. We also describe how organisms fight back to protect against TE invasion and to control their propagation.

## 2. Genome Invasion by Transposable Elements: Strategies for Effective Spreading

### 2.1. Horizontal Transfer: TE Propagation between Species

A major step in understanding how TEs might persist over time was the discovery that some TEs can colonize “naive” genomes through horizontal transfer (HT). HT is the transmission of genetic material between closely or distantly related organisms in the absence of reproduction. These events permit the acquisition of exogenous genetic material and, therefore, are responsible for the appearance of genetic novelties. The first evidence of HT involving TEs (Horizontal Transposon Transfer, HTT) in eukaryotes was the HT and subsequent invasion by the *P-element*, a DNA transposon, between two fruit fly species (i.e., from *Drosophila willistoni* to *Drosophila melanogaster*). *P-elements* rapidly spread through natural populations of *D. melanogaster* between 1950 and 1980, and all flies collected in the wild after 1980 have *P-elements*, unlike laboratory strains derived from flies collected before 1950 [13–16]. The *P-element* in the *D. melanogaster* genome differs by only one nucleotide from that in the *D. willistoni* genome. This demonstrated that the *P-element* found in *D. willistoni* was transferred to *D. melanogaster* some time before 1950. Currently, the number of fully sequenced genomes is sufficiently high to reveal that HTT is a widespread phenomenon in metazoans. For instance, more than 500 putative HTT events have been described between *Drosophila* species [17]. In insects, HTT is not an anecdotic event because up to 24% of all nucleotides of insect genomes might come from HTTs [17]. The same authors showed that DNA transposons transfer horizontally more frequently than retrotransposons. These findings indicate that HTT is a fundamental mechanism implicated in eukaryotic genome evolution. It allows TEs to bypass the host silencing machinery by introduction into naive species that have not yet adapted to silence new TEs.

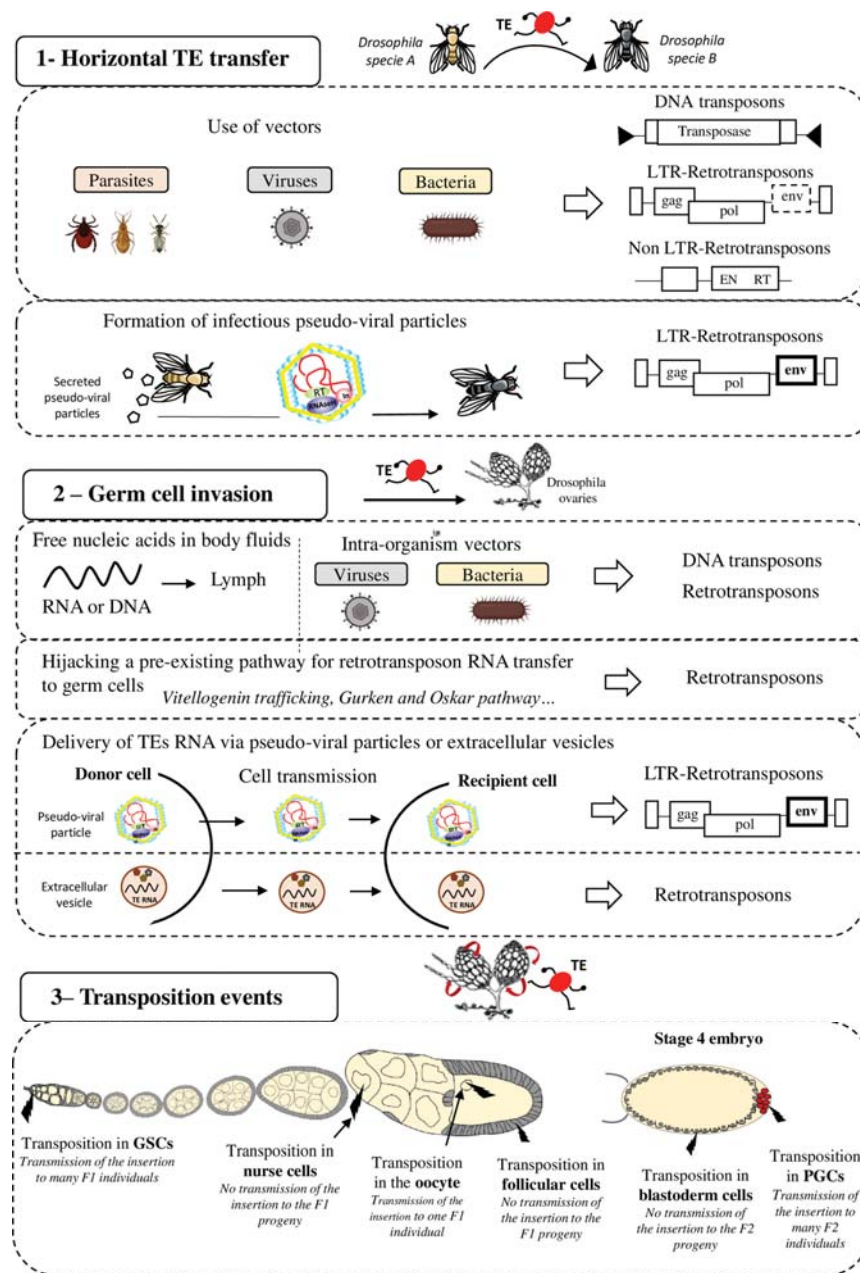
TE mobility and replication characteristics may facilitate the invasion and integration into the host genome. However, the precise mechanisms by which TEs can be shuttled between organisms and the nature of the potential vectors remain speculative (Figure 1 panel 1). It has been suggested that host–parasite interactions favor HTT. For instance, the *P-elements* could have been transmitted from *D. willistoni* to *D. melanogaster* thanks to the mite *Proctolaelaps regalis* [18]. During feeding by piercing and sucking the fly eggs and larvae, this mite might transmit genetic material (e.g., DNA transposons) from one fly to another. Insects, such as wasps or the hemipteran *Rhodnius prolixus*, that feed on the blood of mammals, birds, and reptiles might be involved in HT as vectors (reviewed in [19]). Bacteria and viruses also might be interesting vectors of HTT between species. Indeed, their capacity to transfer DNA and recombine with the host genome might allow them to transport various TE sequences from host to host. Gilbert et al. analyzed

21 genomes of a baculovirus population and demonstrated that a substantial number of TEs from the infected host can transpose into the baculovirus genome [20]. The discovery of a retroposon sequence (Short Interspersed Nucleotide Element, SINE) and its flanking regions coming from the genome of a West African snake (*Echis ocellatus*) in the genome of the taterapox virus (TATV) is another piece of evidence that viruses are frequently used as vectors for HTT [21]. *Wolbachia* is an intracellular parasitic bacterium that infects mainly arthropod species and also some nematodes. This bacterium transfers vertically and horizontally between species and can also transit from cell to cell and infect the host germ cells. Interestingly, many gene transfer events have been detected between *Drosophila* and *Wolbachia*, suggesting that this bacterium is a good candidate vector for HTT between arthropods [22,23]. As a final example, it has been proposed that nematodes may be both great vectors for HTT but also serve as TE reservoirs [24]. Nematodes are ubiquitous organisms and the geographical proximity with many different species increased their chance to participate either as a donor or as a recipient in HTT events. In line with that, many horizontal transfer events involving TEs have been reported to occur between nematodes and unrelated organisms. Future phylogenetic studies will probably reveal many other HTT events involving many different mechanisms and vectors. Combined with geographical and ecological data, these findings will help to unravel the complex and dynamic network of gene transfer and HTT that break down taxon boundaries, and to determine the contribution of ancient HTT events to the evolution of different organisms.

## 2.2. Spread within an Organism and Vertical Transfer to Descendants

### 2.2.1. Germ Cell Invasion: A Life or Death Issue

After invasion of an organism, TEs must then reach the host's germ cells, the only cells whose genetic material will be transmitted to the offspring (vertical transmission) (Figure 1 panel 2). Indeed, if transposition occurs only in somatic cells, the horizontally transferred TEs will die with the host and will never spread within that species. Therefore, successful HTT requires the TE integration into the germ cell genome. Although germline and soma are well distinct cell types in animals, HTT is not rare in complex multicellular eukaryotes, showing that genetic material is transferred from soma to germline. It is still unknown when and in which cells during germ cell development TEs transpose and integrate the germline genome after HTT. Indeed, TE sequences must circulate in the body to reach the germline. They may be transported by their own pseudo-viral particles or by vectors, such as *Wolbachia* bacteria, or even in the form of free RNA or DNA [25]. DNA and RNA molecules can circulate in the extracellular body fluids, such as blood, plasma, lymph, saliva, or milk [26]. It has also been proposed that extracellular vesicles (EVs) could be efficient vectors for transferring TEs between host cells [27]. EVs, such as exosomes or microvesicles, are cell-derived vesicles that deliver biological molecules between different cells and cell types in organisms [28]. LINE1 retrotransposon RNAs have been detected in EVs isolated from cells expressing LINE1 active elements, and these EVs can then deliver LINE1 RNA to recipient cells. Thus, EVs could deliver retrotransposon RNA to neighboring and distant cells, potentially permitting germ cell invasion.



**Figure 1.** Strategies for effective spreading of transposable elements in the *Drosophila* genome. Transposable elements (TE), DNA transposons and LTR and non-LTR retrotransposons, can colonize “naive” genomes through horizontal transfer (**panel 1**). This may occur via vectors (e.g., parasites, viruses, and bacteria) that transfer genetic content from one organism to another. It has also been proposed that LTR retrotransposons (ERV) can form pseudo-viral particles with infectious properties. Then, invading TEs need to reach the host’s germline (**panel 2**) for vertical transmission to the progeny. To this end, TE sequences (excised DNA or RNA for retrotransposons) might circulate in the blood or be transported to germ cells by intra-organism vectors (viruses, bacteria). Retrotransposons can hijack pre-existing cell pathways, such as the vitellogenin trafficking (*ZAM*), Gurken (*I-element*) or Oskar (*TAHRE*) pathways, to transfer their RNA to germ cells. LTR-retrotransposons (ERV), through the formation of infectious pseudo-viral particles, might also be transferred between cells and/or circulate through extracellular fluids to reach the host’s germline. Extracellular vesicles have also been proposed to be efficient vectors for transferring TEs between host cells. The last step for efficient invasion requires transposition in germ cells (**panel 3**) that allows or not (in function of the cell type) the TE transmission and propagation in the specie. GSC = Germinal Stem Cells, PGC = Primordial Germ Cells.

Once the TE reaches the germ cells, transposition into the germ cell genome is the next critical step to ensure its inheritance. It is important to note that gonads are made of different cell types and that transposition can occur at different stages of gametogenesis and in the different cell types (Figure 1 panel 3). The consequences of transposition depend on the type of gonad cells in which transposition occurs. For instance, in *Drosophila* embryos, Primordial Germ Cells (PGCs) give rise to all Germinal Stem Cells (GSCs) present in adult gonads. In adults, GSCs divide asymmetrically to produce one daughter GSC and one cystoblast. The cystoblast begins to differentiate and undergoes four rounds of mitotic divisions to form a cyst of 16 germinal cells. Most of them become nurse cells and only one will differentiate into the future oocyte. The oocyte is the only germ cell that will progress through meiosis and will be fertilized. Nurse cells do not transfer their genetic content to the progeny. Consequently, transposition in nurse cells is not expected to be beneficial to the TE. However, these cells show a high level of polyploidy and produce huge RNA quantities. For TEs, nurse cell invasion could be an intermediate and suitable strategy to reach the oocyte, especially if they transpose through an RNA intermediate, as retrotransposons do. Once inserted in the DNA of nurse cells, TEs might be expressed and produce a huge quantity of RNA that will then be transmitted to the oocyte when nurse cells dump their content into the oocyte. In agreement, the *I-element* (a *Drosophila* retrotransposon) is expressed only in nurse cells and then *I-element* RNA transits to the oocyte to integrate its genome [29]. The RNAs of other TEs (*HMS-Beagle*, *3S18*, *Blood*, *Max*, *TAHRE*, *Burdock*, and *HeT-A*) also can target the oocyte [29–31]. TEs have developed a variety of strategies to optimize their transfer to specific target cells, such as the oocyte, particularly by hijacking pre-existing pathways in the host organism. For instance, *I-elements* target the oocyte nucleus by exploiting the host transport machinery of *gurken* mRNA [32]. *I-element* transcripts contain a small loop of secondary structure that resembles a structure present in *gurken* mRNA. This loop represents a consensus signal for targeting RNAs to the oocyte nucleus by dynein-mediated transport. Similarly, *TAHRE* transcripts migrate to the oocyte germ plasm by mimicking *oskar* RNAs and engaging the Stauf-dependent active transport machinery [30]. Once in the oocyte, TE RNA can be reverse transcribed before integration in the oocyte DNA. However, the highly condensed oocyte genome makes transposition events difficult and unlikely in this cell type. Yet, RNAs present in the ooplasm are transmitted to the embryo. Thus, transposition can also occur in the embryo, particularly in its PGCs. TE integration in the embryo PGCs is particularly advantageous because the new insertions will be inherited by the next generation through the myriad of gametes produced. For example, *P-elements* can transpose to PGCs in embryos and in GSCs in adult ovaries, and this could explain their rapid propagation in *D. melanogaster* populations [33].

#### 2.2.2. Last Step for an Efficient Invasion: Transposition and Fixation in Target Cells

Transposition in the germ cell genome is crucial for TE propagation in a population because it allows the vertical transmission of new insertions. However, the transposition rate is very low. For example,  $10^{-4}$  transposition events per TE copy per generation occur in *Drosophila* natural and laboratory populations [34–36]. Although the transposition rate can be higher in conditions of environmental or genomic stress, this is not sufficient to explain how genome invasion is quickly observed after HTT. According to the model proposed by Le Rouzic and Capy, following invasion by HT, an initial transposition burst occurs [37] that leads to TE accumulation in the genome before the induction of an adaptive response by the host to control transposition. Unfortunately, it is almost impossible to observe transposition bursts in real time because they can be very fast. However, the rapid invasion of *P-elements* in natural *D. melanogaster* populations following HTT strongly suggests that the rate of *P-element* transposition must have been very high at one point. Interestingly, Kofler and al. have demonstrated that *D. simulans* populations that are at their starting point of *P-element* invasion have a high *P-element* transposition rate during the first time of invasion [38]. A similar observation has been made analyzing the transposition dynamics

of the *mariner* DNA transposon after its introduction into *D. melanogaster* populations containing no active *mariner* [39]. The high transposition rate of the introduced *mariner* element leads to the invasion of the population and the colonization of the genome. Finally, by artificially introducing TEs in naive species to mimic HTT, it has been demonstrated that the introduction of DNA transposons (e.g., *Tc1*, *hAT*, and *PiggyBag*) in the genome of species that belong to different kingdoms or domains of life leads to high transposition rates that depend on the TE class and expression pattern [40]. It is important to note that the capacity to transpose may vary among TEs. As an example, it has been proposed that the great success of DNA transposons to transfer horizontally compared to retrotransposons could be explained by their “blurry promoters” [41]. Actually, DNA transposon expression shows very low dependence on host factors; these TEs are more broadly expressed in diverse organisms allowing them to transpose in a large panel of hosts. Another explanation could be that DNA intermediate molecules of DNA transposons are more stable than RNA intermediate molecules used for retrotransposon transposition.

At the end, TEs that can invade genomes are certainly those that can implement an efficient invasion strategy and transposition mechanisms. Moreover, insertions that are neutral or that increase the host fitness have higher chances to be fixed in a population [42]. However, a quantitative population genetics model showed that a TE may persist as long as its deleterious effect on the host is lower than the advantage of transposition explaining that even TEs with negative fitness effects may spread in populations [43,44].

This led to the conclusion that selective pressure is exerted on TEs during the first steps of invasion, depending on their burst capacity and the effect of TE insertions on population fitness. However, there is a “common advantage” for both host and TEs in limiting massive transposition in the whole organism. In fact, transposition in somatic cells can be deleterious by creating detrimental mutations that lead to the host death and concomitantly to the TE disappearance. From this point of view, TEs resemble viruses: they must multiply and spread, but the host also must survive. It has been hypothesized that several TEs, such as the *P-element* and the *I-element*, have developed the ability to be expressed only in germ cells [45–47]. This avoids the deleterious effects of mutations in somatic cells and ensures the transmission of new TE insertions to the progeny. However, transposition in germ cells could also have dangerous outcomes because these mutations are inheritable. Transposition in PGCs (the precursors of all germ cells) is certainly very efficient for TE invasion, but it is highly risky. Indeed, germline transpositions may induce infertility and also deleterious inheritable pathology-causing mutations that potentially endanger the species’ survival. One of the best described deleterious effects of TE transposition in the germline concerns the massive mobilization of TEs that might have contributed to the extinction of Wrangel island mammoths. This small population of mammoths accumulated a large number of detrimental mutations, including deletions and point mutations, and also many TE sequences. This suggests high TE activity in the mammoth germline that led to a very high number of heritable mutations. This high transposition rate in the germline may have contributed, with other factors, to the extinction of this small endangered population [48].

Therefore, TE transposition must be controlled in all cell types to limit its negative effects on the host and its progeny. On the other hand, a too strict control can cause the TE loss from the host genome and deprive the organism of an important source of genetic diversity. For example, *Spermophilus tridecemlineatus* is a rodent in which transposon activity has declined over at least 4 million years. Its genome does not harbor any recent TE, Long Interspersed Nucleotide Element (LINE), SINE, retrotransposon with long terminal repeats (LTR), or DNA transposon activity. Moreover, no functional TE copy is found in the genome of this species because all harbor a huge number of mutations. This is explained by the strong TE silencing, leading to complete inhibition of TE mobilization [49]. To be conserved in a genome, a minimum of transposition is required. Interestingly, in several eukaryotes, temporary relaxation of the TE silencing machinery has been observed in the germline and its associated cells. For instance, during *Drosophila* early oogenesis, there is a short spatiotemporal window when the piRNA pathway seems to be less efficient and

at least some TEs might escape the host control. It has been proposed that this window, termed the ‘Piwi pocket’, allows the insertion of new TEs in the developing germline genome [50–52]. In mammals, transient TE relaxation during germ cell development has been observed mainly during epigenetic reprogramming periods [5,53]. Specifically, during the first wave of global reprogramming that occurs following fertilization, 10% of the transcriptome in 2-cell stage mouse embryos is made of specific TE transcripts, including transcripts from the *MuERV* retroelement [54–56]. The second reprogramming wave occurs in PGCs of the developing mouse embryo. Although no general transcriptional burst has been observed for TEs at this step, some specific TE transcripts (i.e., *LINE1* transcripts) are overrepresented [53], reviewed in [57]. The presence of such spatiotemporal windows during germline development in which TE control is weaker could help to explain the very successful genome invasion by TEs. Once settled in the germline genome, new TE insertions are then vertically transmitted, like any other DNA sequence.

### 2.3. Retrotransposons: A Formidable Capacity of Propagation

#### 2.3.1. Retrotransposons Can Do Intercellular Transposition

Retrotransposons have evolved in a variety of organisms, from protozoa to humans, and display outstanding capacities of rapid invasion and propagation. There are two types of retrotransposons: with and without LTR. Non-LTR retrotransposons lack LTR and have generally two open reading frames of which one encodes a reverse transcriptase and an endonuclease. SINEs do not encode a functional reverse transcriptase and are non-autonomous elements because their transposition relies on enzymes encoded by other non-LTR retrotransposons: the LINEs. Thus, SINEs cannot colonize a naive genome after HT if the genome does not have a corresponding element for trans-complementation.

This part of the review will focus on the other retrotransposon group: LTR retrotransposons. LTR retrotransposons resemble retroviral proviruses. Indeed, they have LTRs at each extremity and open reading frames equivalent to the *gag* and *pol* genes. *Gag* encodes a structural protein involved in the formation of virus-like particles. *Pol* encodes proteins that are necessary for transposition mechanisms: an integrase, a RNase H, a protease, and a reverse transcriptase. Some LTR retrotransposons also harbor the envelope gene (*env*) that encodes a viral surface glycoprotein, and they are called endogenous retroviruses (ERV). ERVs are assumed to be derived from past retroviral infections that have been integrated as permanent residents in host genomes. Like retroviruses, most ERVs can form virus-like particles (VLP). The Env protein interacts with target host cell receptors and allows the fusion of the VLP with the target cell membrane and ERV propagation between cells. ERVs make up approximately 10% of the mouse, rat, and human genomes and they have been extensively studied in *Drosophila* [8]. In the *Drosophila* genome, many ERVs copies are present, such as the very diverse *Gypsy-like* elements including ZAM for instance. These TEs can form VLPs and infect neighboring cells [58,59].

#### 2.3.2. Horizontal Transfer of Retrotransposons: Do They Really Need Vectors?

It has been hypothesized that VLPs produced by ERVs can propagate between organisms, like retrovirus particles, without vectors (Figure 1 panel 1). In this case, these particles could be infectious. VLPs of the *Gypsy* ERV have been found as extracellular particles in the medium in which *D. melanogaster* follicular cells were cultured. This means that *Gypsy* VLPs are secreted by cells [60]. Moreover, these VLPs can infect cultured cells belonging to another *Drosophila* species: *Drosophila hydei*. Interestingly, a recent study showed that *Gypsy* also transmits between cells that are not in contact [61]. Furthermore, experiments in which flies were grown on medium containing crushed pupae that produced *Gypsy* VLPs suggested a possible HTT via food: these flies became infected by *Gypsy* [62]. Once transmitted to a new individual by HTT, retrotransposons could use their retroviral properties to propagate between cells and through body fluids to reach the germline, ensuring their spread in that species.

### 2.3.3. *Drosophila* Germ Cell Invasion by Retrotransposons

Most TEs can insert in the germline by being active directly in these cells. On the other hand, ERVs do not seem to be expressed in germ cells. In *Drosophila* ovaries, when the pathway regulating TEs is abolished in all cell types, ERVs are only expressed in a patch of somatic cells that are called follicular cells and that surround germ cells [63–68]. Indeed, as described for gene transcription, TE transcription requires transcription factors that are present in specific cell subsets [69,70]. For instance, *ZAM* retrotransposon expression requires the presence of Pointed 2, a transcription factor only expressed in a patch of follicular cells [70]. This means that ERV RNAs are not produced directly in germ cells and that transposition into the germ cell genome requires ERV transmission from somatic cells. VLPs formed in the producing somatic cells could infect germ cells via the Env transmembrane protein, but other routes could also be used (Figure 1 panel 2). For instance, the 412 element can infect germ cells, although it does not encode Env [71]. This retrotransposon might use the Env protein encoded by another ERV for germ cell infection. Moreover, the *ZAM* ERV encodes an Env protein, but it reaches germ cells by usurping the endosome/exosome pathway in *Drosophila* ovaries for VLP transfer to the oocyte [59]. This route is normally employed for vitellogenin release and uptake by germ cells. It is not known whether *ZAM* also uses its Env protein to transit to germ cells. The detection of many new *ZAM* insertions in the progeny of flies in which *ZAM* is derepressed in somatic follicular cells indicates that after VLP transfer, *ZAM* can insert into the germ cells genome [72,73]. Therefore, this mechanism of propagation, using transfer from somatic to germ cells, is an efficient way to spread in a population. This suggests that ERVs expressed in somatic cells close to germ cells might create particles that infect the germline or might use a more passive mechanism to reach germ cells for insertion in the genome and vertical transmission.

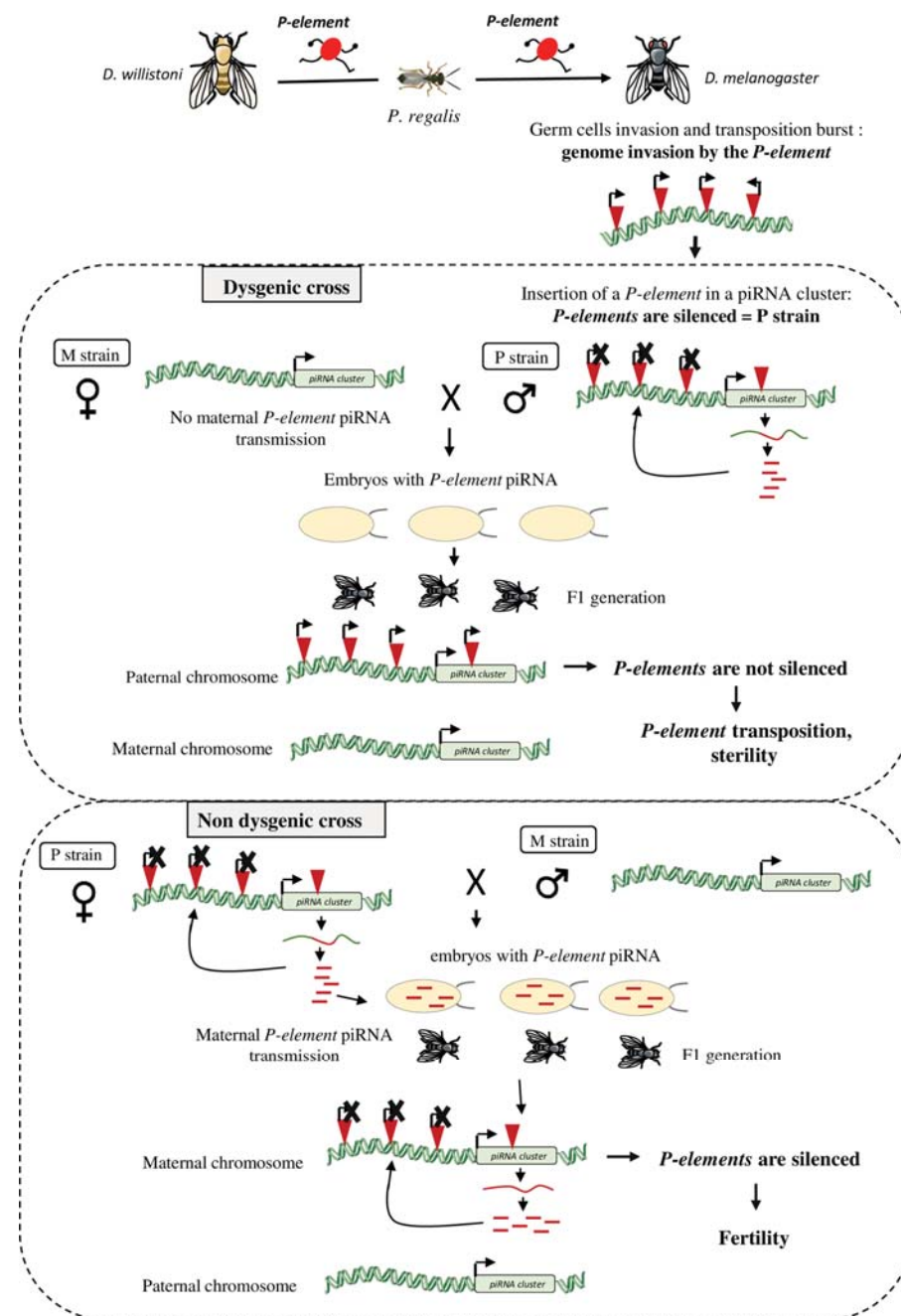
It is clear that even if most ERVs are not expressed in the germline, they have many strategies for efficient propagation. Thanks to their capacity of intercellular and potentially inter-organism transfers, ERVs seem particularly well suited to efficiently propagate in an organism, to its descendants, and also to other species. This could explain why retrotransposons occupy such an important place in eukaryote genomes [8].

## 3. How to Deal with TE Invasion: Host Defense Strategies

### 3.1. Hybrid Dysgenesis and Discovery of the piRNA Pathway

HTT allows TEs to escape the host defenses through infection of a new species without a defense mechanism to block that specific TE. This situation can be compared to what happens when a pathogen enters an organism for the first time. The infected host must develop specific immune defenses for protection against the new invader. The organism response following HTT is largely unknown.

The first data on the host response following HTT were obtained by studying *D. melanogaster* species invasion by the *P-element* (Figure 2). This HTT led to the invasion by *P-elements* of the genome of all wild-caught *D. melanogaster* between 1950 and 1980, demonstrating that at the time of invasion, before 1950, no defense mechanism provided efficient protection against *P-element* transposition. The invaded strains were called P strains and all strains isolated before the HTT (and thus without *P-elements* in their genome) were called M strains. Early studies showed that when females of M strains are crossed with males of P strains, the resulting F1 females are sterile and show rudimentary gonads, a phenomenon called hybrid dysgenesis [74,75]. Hybrid dysgenesis is also observed for other TEs, such as the *I-element* that invaded *D. melanogaster* from *Drosophila simulans* by HTT [76,77], Penelope [78], and *Hobo* [79]. Interestingly, hybrid dysgenesis is usually only observed at certain temperatures. Hybrid dysgenesis was the first evidence that differences in TE genomic content, due to HTT, can induce reproductive incompatibilities and might lead to speciation [80,81].



**Figure 2.** Host response following a TE horizontal transfer: example of the *D. melanogaster* genome invasion by the *P-element*. The *P-element* originally found in the *D. willistoni* genome has been transferred to *D. melanogaster* in the 1950s, possibly via the mite *Proctolaelaps regalis*. After germ cell invasion by the *P-element*, a transposition burst induced the creation of many *P-element* inheritable insertions. The *P-element* invaded the *D. melanogaster* genome and at one point it jumped into a piRNA cluster where piRNAs complementary to this element are produced to silence genomic *P-elements* (P strain). When a male with functional genomic *P-elements* (P strain) is crossed with a female without *P-element* (M strain), their offspring will be sterile (**dysgenic cross**). Indeed, the M female does not transfer any *P-element* piRNA to the embryos. These piRNAs are required to initiate the production of *P-element* piRNAs by the piRNA cluster present on the paternal chromosome of the offspring. Without *P-element* piRNA production in the F1 progeny, genomic *P-elements* are not silenced, leading to sterility. In the reverse cross (P female x M males), the F1 flies are fertile (**non-dysgenic cross**). Here, *P-element* piRNAs produced in the ovaries of the P mother are deposited in the embryos and allow the activation of the maternal *P-element*-containing piRNA cluster for the genomic *P-element* silencing.



On the other hand, the reverse cross (P females × M males) leads to fertile F1 females and the *P-elements* are silenced in the F1 ovaries. This result revealed that some cytoplasmic components are maternally transmitted to the offspring and can trigger a TE-silencing response in the F1 gonads [82,83]. Later, the PIWI-interacting RNA (piRNAs) were discovered as the source of these cytoplasmic components [84,85].

### 3.1.1. Discovery of the piRNA Pathway: A Barrier against Transposition

Many years after the first description of hybrid dysgenesis, several studies simultaneously reported the identification of a small RNA-based immune system, composed of PIWI-interacting RNA (piRNAs). This system, which was called the piRNA pathway [3,86–90], silences TEs genome-wide in male and female germ cells. piRNAs are small RNAs of 23–29 nucleotides in length that bind to PIWI proteins to silence TE activity via homology-dependent mechanisms. The components of this pathway are highly conserved: piRNAs and PIWI proteins are found in many organisms, from protozoans to higher eukaryotes [91,92]. This pathway has been extensively studied in *D. melanogaster* ovaries. These ovaries are formed by two cell types: follicular somatic cells and germ cells, and the piRNA pathway is active in both cell types. This model allowed showing that piRNAs are encoded by dedicated genomic loci called piRNA clusters [3]. piRNA clusters are composed of full length or truncated TEs that define the repertoire of elements recognized and silenced by the piRNA machinery. It is important to note that piRNA clusters have a tissue specific expression. In germ cells, piRNAs are loaded onto the PIWI proteins to trigger transcriptional silencing of TE [93]. Moreover, piRNAs loading onto the Aub and Ago3 proteins allows the amplification of the piRNA pool by a mechanism called the ping-pong cycle and post-transcriptional TE silencing [3,4]. Thus, in germ cells, TEs are silenced at both the transcriptional and post-transcriptional levels. In addition, Aub and Ago3 loaded with piRNAs produced during oogenesis are deposited in the early embryo and are then incorporated into the developing germ cells during embryogenesis [84,94,95]. This maternal deposition is required for efficient TE control in the offspring germline.

### 3.1.2. Role of the piRNA Pathway in Hybrid Dysgenesis

The discovery of the piRNA pathway led to a better understanding of hybrid dysgenesis (Figure 2). Analysis of *P-element* piRNAs showed that they are produced in *D. melanogaster* lines invaded by *P-elements*, but not in lines without *P-elements*. During *P-element* invasion, the TE inserts at many loci, including in a piRNA-producing genomic region often located at the cytological site 1A [96,97]. Trapping *P-elements* into a piRNA cluster allows producing piRNAs complementary to this element and this can induce *P-element* silencing. Very often *P-elements* insert in the same piRNA cluster, specifically in telomere-associated sequences, located at the cytological site 1A [97–100]. *P-element* piRNAs produced in the ovaries of P strain females are then deposited in the embryo and are required to initiate the production of *P-element* piRNAs by a piRNA cluster in the offspring [84,101–104]. *P-element* piRNAs produced in F1 ovaries, together with PIWI proteins, could mediate transcriptional and post-transcriptional silencing of *P-element* euchromatic insertions. However, there is conflicting evidence about the exact effect of piRNA on *P-element* regulation [16]. Some studies have shown that *P-element* piRNAs could directly influence the expression of *P-elements* and thus modulate the level of *P-element* transcripts in germ cells [33,85]. On the other hand, one other study has demonstrated that *P-element* piRNAs act only on the splicing of *P-element* transcripts. *P-element* piRNAs promote the retention of the third intron and therefore inhibits the production of functional P-transposase in germ cells. [105]. Regardless of their mode of action, piRNAs are required to silence active *P-elements* in F1 gonads.

When no *P-element* piRNA is maternally deposited (M strain females), active *P-elements* inherited from the father are not silenced. Following *P-element* transposition, oogenesis is switched off and selective apoptosis is induced in ovarian stem cells by activation of the DNA damage checkpoint, leading to sterility (Figure 2). However, it should be noted

that although males cannot transmit *P-element* piRNAs, the piRNA cluster containing the *P-element* is transmitted. During F1 hybrid dysgenic females culture at 25 °C, de novo production of *P-element* piRNAs is observed, possibly encoded by the paternally inherited *P-element*-containing piRNA cluster. *P-element* expression and transposition decrease and fertility is recovered [33,85]. This phenomenon shows the rapid adaptation to *P-element* expression in a single generation, with these flies inheriting a piRNA cluster for effective protection once activated.

Moreover, it has been shown that *Tirant* probably spread in *D. melanogaster* populations in ~1938, leaving behind old *Tirant* sequences in the heterochromatin [106,107]. These copies accumulated mutations and deletions, leading to highly degenerated *Tirant* sequences. These insertions produce piRNAs, but they are probably too degraded to induce an efficient silencing because a match with less than 10–20% of sequence divergence between piRNA and TE sequence seems to be required [108]. This means that new HTTs or TE reactivation might occur when the regulatory copies created by past invasions are too old and divergent relative to the corresponding active TEs.

Therefore, to be protected against an invading TE, organisms must put in place a new control system that will trigger TE silencing each time the memory copies become too divergent from the active copies.

### 3.2. Response to TE Invasion by Horizontal Transfer

Detection of the first individual of a species concerned by a HTT event is challenging because successful events are probably uncommon and unpredictable. Thus, the initial response of an organism following HTT is difficult to analyze. However, there are many cases where a TE has been transferred to a new species, and propagates in that population. This means that some individuals carry insertions of this new TE and others are devoid of it. The study of these ongoing HTT events allows monitoring the response at the population scale.

#### 3.2.1. Analysis of Ongoing TE Invasion

The *P-element* spread in *D. melanogaster* populations very rapidly because all wild-caught flies collected after 1980 carry *P-element* insertions. Since then, *P-element* propagation in *Drosophila* species continues, and 15 years ago the *P-element* began to invade *D. simulans* populations. Some *D. simulans* flies now have a huge number of *P-elements* in their genome (e.g., *D. simulans* collected in South Africa in 2012), others have only few insertions (*D. simulans* collected in Florida in 2010), and others do not harbor any *P-element* (*D. simulans* collected in the Sub-Saharan region in 2009). Before 1998, *P-elements* were completely absent in all *D. simulans* populations, suggesting an ongoing invasion process.

Kofler and al. studied the response to *P-element* invasion in a *D. simulans* population collected in 2010 and harboring few *P-element* insertions [38]. The copy number increase over generations suggested that *P-elements* are currently active in this population. However, after several generations (22 generations), *P-element* copy numbers in the populations stabilized. This observation was correlated with the new production of piRNAs complementary to *P-elements* and with the ping-pong signature. Based on their findings, the authors proposed a model called “shotgun silencing” composed of three different phases. First, the TE infects a new population where the piRNA pathway cannot silence *P-elements*, and the TE can transpose and multiply rapidly within the population. Second, during this wave of transposition, *P-element* inserted into a piRNA cluster in some individuals of the population resulting in the production of new piRNAs that are complementary to the *P-element*. However, *P-element* mobility is not controlled instantly at the population level, suggesting that these cluster insertions are not fixed in the population. In a third phase, *P-element* piRNA production increases in the population, possibly mediated by an increasing number of *P-element* insertions in piRNA clusters in several individuals, to a level sufficiently high to completely silence *P-elements* in the population. However, it is still unknown if one *P-element* insertion in a piRNA cluster is sufficient to repress this

TE in one individual. Interestingly, it seems that complete silencing of the *P-element* in *D. melanogaster* can be mediated by only one insertion in a specific piRNA cluster at the 1A telomere-associated sequence locus [96,97].

Computer simulations of the dynamics of TE invasion have demonstrated that a single insertion in a single non-recombining cluster, such as the somatic piRNA cluster *flamenco*, is probably sufficient for TE repression [44]. Indeed, the ZAM retrotransposon is only present in one piRNA cluster, *flamenco*, and this single insertion is sufficient to produce enough ZAM-derived piRNAs to repress ZAM expression [109]. However, R. Kofler argues that, concerning piRNA clusters expressed in the germline, several cluster insertions are likely to be required to stop the invasion [44].

This genomic adaptation to invading TEs seems to be rapid and reproducible: most *D. simulans* individuals invaded by *P-elements* acquire a piRNA cluster insertion within a short period of time after invasion. However, the lag time between *P-element* invasion and silencing depends on the TE activity. *P-element* transposition seems to be temperature-dependent, and increases in flies raised at high temperature ( $\geq 23$  °C) compared with flies raised at lower temperature (15 °C) where the P element transposase is less active [33,38,110]. High *P-element* activity results in many de novo insertions and increases the chances of insertion in piRNA clusters. Therefore, *P-element* repression is established faster when flies are raised at higher temperature. This example shows that adaptation to an invading TE can be rapid, but depends on the TE transposition activity.

### 3.2.2. Analysis of the Initial Host Response to HTT

It is thought that *KoRV-A* is a recently introduced exogenous virus in many koala genomes and in the process of becoming an ERV. *KoRV-A* proviral insertions have been detected in almost all koala genomes analyzed, and there are now few naive populations [111]. *KoRV-A* is horizontally transmitted between animals and can infect the germline and integrate in the germ cell genome for vertical transmission [112–114]. Therefore, it is a great model to better understand how an organism and particularly how germ cells respond to retrovirus invasion. Yu et al. [115] suggested that after invasion by *KoRV-A*, an “innate” response is initiated in the gonads. This first phase is based on the recognition of a presumed conserved molecular pattern in *KoRV-A* RNA, leading to its processing into sense piRNA. *KoRV-A* RNAs are processed and are inefficient for active transposition. This finding shows that, during the initial response, insertion in a piRNA cluster is not necessary to trigger sense piRNAs production. This could be an efficient way to suppress virus replication in the first instance but does not appear to be sufficient to stably control *KoRV-A* in the population. The production of antisense piRNA is the second phase of the response and seems to require a TE insertion into a piRNA cluster. This will guide the sequence-specific adaptive immunity and generate a memory of the invader. Thus, the initial TE repression should not be too strong because the TE needs to be still sufficiently expressed to transpose and insert into a piRNA cluster. This study suggests that even when an innate mechanism allows the early repression of the TE, only insertion in a piRNA cluster leads to a robust, durable, and inheritable TE silencing.

Moreover, piRNAs could also be newly produced by euchromatic TE insertions converted into piRNA-producing loci, a phenomenon called paramutation [94,116–118]. Thus, numerous insertions dispersed across the genome may contribute to piRNA production and entire new piRNA generating regions could arise, providing another mechanism for acquiring immunity against new TEs. Indeed, spontaneous formation of novel piRNA clusters from euchromatic TE insertions and transgenic sequences has been observed [94,119,120]. At this time, it is still unknown how new piRNA producing genomic regions arise. It has been suggested that siRNAs could be involved in this process and thus participate in the initial detection of new elements and in the genesis of a stable piRNA response [119,121].

Interestingly, production of endogenous siRNAs (endo-siRNA) targeting TEs in somatic and germ cells had already been observed, suggesting a potential role of these small RNAs in the initial phase of TE control [122–125]. As an example, Rozhkov et al. developed

an artificial system to follow the initial response following TE invasion and have also detected a rapid production of siRNA [126]. The authors introduced a transgene containing an active copy of *Penelope* in the *D. melanogaster* genome to mimic an invasion by a new TE. *Penelope* is a TE found in *Drosophila virilis*, but absent in *D. melanogaster*. After this artificial HT, they observed production of siRNAs against *Penelope* in the transgenic flies. Endo-siRNAs are known to be produced from a dsRNA precursor cleaved by Dicer-2. This precursor does not seem to predominantly derive from master loci, as inferred for TE piRNAs produced by piRNA clusters [123]. The dsRNA might rather be produced by the pairing of sense and antisense transcripts generated by bidirectional transcription of TEs or by the association of transposon sense RNA with antisense RNA transcribed from a piRNA cluster. Moreover, a mutation in the coding gene of Dicer2 or Ago2, two proteins required for siRNA production and targeting, triggers derepression of some TEs in *Drosophila* S2 cells and in *Drosophila* ovaries [122,123,127,128]. However, in the case of the artificial HTT, at the time of the TE introduction in the *D. melanogaster* genome, *Penelope* was not inserted at any loci in the genome. Therefore, this new siRNA production probably comes from the processing of newly produced retrotransposon RNAs. It might be a kind of first “immune defense” initiated by the arrival of an invader in the cell, as observed for viruses. However, siRNA production was not sufficient to silence completely *Penelope* in *D. melanogaster* and it could still transpose occasionally. At later stages of genome invasion, *Penelope*-derived piRNA production was observed and this correlated with the presence of a new *Penelope* insertion in a piRNA cluster [129]. A number of transposition events might be required for insertion in a piRNA cluster and to set up an efficient control of the invading TE via the production of related piRNAs. Moreover, the cooperation between siRNA and piRNA production in the fight against *Penelope* has not been studied yet. It seems that siRNA production precedes or accompanies piRNA production and could then be potentially involved in the activation of piRNA production [119,121]. However, these findings support the idea that the piRNA response is a robust response that must be firmly established to efficiently control the invading TE.

That being said, it is also interesting to note that, while piRNAs seem to be absent in most of the nematode species [130], some evidence suggests that horizontally transferred TEs within these species are controlled very rapidly after their entrance in new genomes. Single-copy horizontally transferred DNA transposons, almost intact copies, have been found in nematode genomes [24]. Thus, it is possible that another efficient mechanism could control the transposition just after the HTT, independently of the production of piRNA.

### 3.3. Germ Cell Response to TE Invasion from Neighboring Somatic Cells

#### 3.3.1. piRNA Clusters Keep the Memory of Past Invasions to Protect Germ Cells

Thanks to piRNAs produced by piRNA clusters, TEs are controlled and this immunity is transmitted to the next generation. Interestingly, a recent study by Gebert et al. has challenged this assumption [116]. The authors have demonstrated that deletion of the three largest piRNA clusters expressed in *Drosophila* germ cells has no effect on TE regulation or fertility. Although much redundancy between multiple germline piRNA clusters is observed for many TEs, this study proposes that, in germ cells, piRNA produced by dispersed elements could be sufficient for TE regulation. However, divergent observations have been made analyzing the *flamenco* piRNA cluster, which is only expressed in the ovary somatic cells [3,67]. A study has been undertaken on a *Drosophila* line called RevI-H2 [73] where the ZAM retrotransposon was deleted from its regulatory piRNA cluster *flamenco*. In wild type flies, ZAM is present in *flamenco*, and 85% of all ZAM-derived piRNAs are produced by the *flamenco* piRNA cluster. *Flamenco* is only expressed in the ovary somatic cells and no ZAM insertion is detected in any other piRNA clusters in the reference genome. This means that *flamenco* is the only regulatory piRNA cluster of ZAM. It is important to note that ZAM deletion from the *flamenco* piRNA cluster results in loss of ZAM control [72,73,109,131], demonstrating the correlation between the presence of a TE in a somatic piRNA cluster and the regulation of this TE. In RevIH2 ovaries, ZAM

is expressed in follicular cells and forms pseudo-viral particles (VLPs) that are assumed to use the endosomal vitellogenin trafficking system to enter the oocyte and invade the germline [59]. *ZAM* deletion from its regulatory piRNA cluster in somatic follicular cells leads to germ cell invasion by *ZAM* VLPs. This condition can be compared to what happens when a TE first invades a new species through horizontal transfer and reaches the germ cells. Therefore, the study of what happens in the germline after *ZAM* invasion can bring new insights into how germ cells control TE propagation. Indeed, the RevI-H2 line adapted to *ZAM* invasion by trapping a new *ZAM* copy in a germline piRNA cluster. This led to the production of *ZAM*-derived piRNAs in the germline [72]. This could result from the specific targeting of *ZAM* to piRNA clusters or from a random transposition event that has been selected through time. Lastly, the time required to trap a TE in a regulatory piRNA cluster is not known, and TE silencing establishment might take several generations.

### 3.3.2. How Long Does It Take to Implement an Efficient Response against an Invading TE?

All these studies revealed that trapping in a piRNA cluster is an efficient process to control transposition of an invading TE and to protect the genome. However, the time required for such an event remains unknown. Depletion in somatic *Drosophila* cells of PIWI, a protein required for TE silencing, leads to derepression of many TEs, including several retrotransposons that may invade germ cells [71]. In this study, new individual TE insertions were detected during more than 73 generations, but did not become fixed in the population. These TE insertions are not particularly enriched at piRNA clusters and another study revealed that the number of TE cluster insertions is insufficient with expectations of the “Trap model” that proposes a specific trapping of TEs in piRNA clusters [38,132]. Moreover, several piRNA cluster insertions per individual seem to be necessary to stop the invasion [44]. All these different arguments might explain why it takes several generations to establish a control. Furthermore, how such control is put in place may vary for each TE and each invasion event. Several parameters should be considered: (i) TE invasions rely on many processes (e.g., cell transfer, pseudo-viral particle formation, vector use) and the response might be different for each mechanism; (ii) TEs have different transposition mechanisms and spatiotemporal expression, resulting in variable transposition rates. A high transposition rate may increase the chances of successful insertion into a piRNA cluster without any specific trapping at these loci; (iii) environmental conditions (e.g., temperature) also influence the transposition rate [33] and modify the response following invasion; (iv) piRNA production varies among all piRNA clusters found in genomes. Moreover, piRNA clusters show a tissue-specific expression and are not all expressed in germ cells where transposition needs to be silenced. This highlights that individual TE insertions into piRNA clusters can have different effects on TE taming.

Thus, although TE trapping in piRNA clusters seems to be an efficient way to sustainably control TE transposition, the response to each TE invasion will be specific. A detailed analysis of many different invasion processes is necessary to better understand the parameters influencing the establishment of the silencing response. Currently, it is still unknown whether TEs insert preferentially into piRNA clusters, or whether they insert randomly in the genome and then insertions in piRNA clusters are retained during evolution due to their beneficial effects. Moreover, TEs may insert, but can also be deleted from a piRNA cluster. Indeed, piRNA clusters in flies and mammals evolve rapidly [85,116,133,134]. Consequently, the piRNA repertoire produced by germ cells evolves constantly and can determine, at a specific moment, TE activity and invasion. Importantly, piRNA clusters keep a memory of past invasions to protect the genome from active and threatening TEs. This strategy can be compared to an adaptive immune system that adjusts to each TE’s characteristics and activity.

## 4. Conclusions

TEs and host control mechanisms evolve constantly to find a balance between transposition and repression. It is now acknowledged that genomes have coevolved with their TEs,

developing strategies to limit transposition. On the other side, TEs propagate in genomes by HTT that allows crossing species boundaries and by targeting germ cells for transposition to ensure their transmission to the next generation. It seems clear that organisms also have evolutionary advantages by allowing some transposition because TEs can be seen as a trigger of a series of events that actively shape the genome architecture and give rise to biological innovations. However, transposition is also a threatening phenomenon because it may create deleterious mutations that could endanger the species' survival. This can explain why organisms develop sophisticated and adaptive strategies to control TE transposition and propagation. Although the initial response following HTT is still largely unknown, it appears that TE insertion in a piRNA cluster is a very efficient mechanism for the long-term control of TE transposition. Studying all TE copies and TE remnants in genomes can provide information about the "history" of each genome facing TE invasion. Each species and each individual has a very specific TE insertion profile, indicating that each organism must deal with specific transposition phenomena and must adapt individually. This is a beautiful example of genome adaptation and demonstrates that a potential source of danger can also be beneficial if well controlled, turning from "junk" into "gold".

**Author Contributions:** M.Y. wrote the manuscript. S.J. and E.B. revised and edited the final version of the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by grants from the Agence Nationale pour la Recherche (ANR-CHApiTRE, ANR-BiopiC, and ANR-EpiTET projects), the Association de Recherche Cancer (contract R17155CC). M.Ys was supported by the Ministère de l'Enseignement Supérieur et de la Recherche (MESR) and the Fondation pour la Recherche Medical (FRM) (FDT202106012950). This research is supported by the French government IDEX-ISITE initiative 16-IDEX-0001 (CAP20-25).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** We thank team members for comments on the manuscript and Stephane Ronseray for discussions. The authors also thank the two reviewers for valuable and critical comments that significantly improved the manuscript.

**Conflicts of Interest:** The authors declare that the research was performed in the absence of any commercial or financial relationship that could be construed as a potential conflict of interest.

## References

1. McClintock, B. Induction of Instability at Selected Loci in Maize. *Genetics* **1953**, *38*, 579–599. [[CrossRef](#)] [[PubMed](#)]
2. Wicker, T.; Sabot, F.; Hua-Van, A.; Bennetzen, J.L.; Capy, P.; Chalhoub, B.; Flavell, A.; Leroy, P.; Morgante, M.; Panaud, O.; et al. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **2007**, *8*, 973–982. [[CrossRef](#)] [[PubMed](#)]
3. Brennecke, J.; Aravin, A.A.; Stark, A.; Dus, M.; Kellis, M.; Sachidanandam, R.; Hannon, G.J. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* **2007**, *128*, 1089–1103. [[CrossRef](#)] [[PubMed](#)]
4. Gunawardane, L.S.; Saito, K.; Nishida, K.M.; Miyoshi, K.; Kawamura, Y.; Nagami, T.; Siomi, H.; Siomi, M.C. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* **2007**, *315*, 1587–1590. [[CrossRef](#)] [[PubMed](#)]
5. Slotkin, R.K.; Vaughn, M.; Borges, F.; Tanurdžić, M.; Becker, J.D.; Feijó, J.A.; Martienssen, R.A. Epigenetic Reprogramming and Small RNA Silencing of Transposable Elements in Pollen. *Cell* **2009**, *136*, 461–472. [[CrossRef](#)] [[PubMed](#)]
6. Wolf, D.; Goff, S.P. Embryonic stem cells use ZFP809 to silence retroviral DNAs. *Nature* **2009**, *458*, 1201–1204. [[CrossRef](#)]
7. Wolf, G.; Yang, P.; Fuchtbauer, A.C.; Fuchtbauer, E.-M.; Silva, A.M.; Park, C.; Wu, W.; Nielsen, A.L.; Pedersen, F.S.; Macfarlan, T.S. The KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses. *Genes Dev.* **2015**, *29*, 538–554. [[CrossRef](#)]
8. Huang, C.R.L.; Burns, K.H.; Boeke, J.D. Active Transposition in Genomes. *Annu. Rev. Genet.* **2012**, *46*, 651–675. [[CrossRef](#)]
9. Zeh, D.W.; Zeh, J.A.; Ishida, Y. Transposable elements and an epigenetic basis for punctuated equilibria. *BioEssays* **2009**, *31*, 715–726. [[CrossRef](#)]
10. Cordaux, R.; Batzer, M.A. The impact of retrotransposons on human genome evolution. *Nat. Rev. Genet.* **2009**, *10*, 691–703. [[CrossRef](#)]
11. Feschotte, C.; Pritham, E.J. DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* **2007**, *41*, 331–368. [[CrossRef](#)] [[PubMed](#)]

12. Oliver, K.R.; Greene, W.K. Transposable elements: Powerful facilitators of evolution. *BioEssays* **2009**, *31*, 703–714. [[CrossRef](#)] [[PubMed](#)]
13. Daniels, S.B.; Peterson, K.R.; Strausbaugh, L.D.; Kidwell, M.G.; Chovnik, A. Evidence for horizontal transmission of the P transposable element between *Drosophila* species. *Genetics* **1990**, *124*, 339–355. [[CrossRef](#)] [[PubMed](#)]
14. Anxolabéhère, D.; Nouaud, D.; Périquet, G.; Tchen, P. P-element distribution in Eurasian populations of *Drosophila melanogaster*: A genetic and molecular analysis. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 5418–5422. [[CrossRef](#)]
15. Anxolabéhère, D.; Kidwell, M.G.; Periquet, G. Molecular characteristics of diverse populations are consistent with the hypothesis of a recent invasion of *Drosophila melanogaster* by mobile P elements. *Mol. Biol. Evol.* **1988**, *5*, 252–269. [[CrossRef](#)]
16. Ghanim, G.E.; Rio, D.C.; Teixeira, F.K. Mechanism and regulation of P element transposition. *Open Biol.* **2020**, *10*, 200244. [[CrossRef](#)]
17. Peccoud, J.; Loiseau, V.; Cordaux, R.; Gilbert, C. Massive horizontal transfer of transposable elements in insects. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 4721–4726. [[CrossRef](#)]
18. Houck, M.A.; Clark, J.B.; Peterson, K.R.; Kidwell, M.G. Possible horizontal transfer of *Drosophila* genes by the mite *Proctolaelaps regalis*. *Science* **1991**, *253*, 1125–1128. [[CrossRef](#)]
19. Loreto, E.L.S.; Carareto, C.M.A.; Capy, P. Revisiting horizontal transfer of transposable elements in *Drosophila*. *Heredity* **2008**, *100*, 545–554. [[CrossRef](#)]
20. Gilbert, C.; Peccoud, J.; Chateigner, A.; Moumen, B.; Cordaux, R.; Herniou, E.A. Continuous Influx of Genetic Material from Host to Virus Populations. *PLoS Genet.* **2016**, *12*, e1005838. [[CrossRef](#)]
21. Piskurek, O.; Okada, N. Poxviruses as possible vectors for horizontal transfer of retroposons from reptiles to mammals. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 12046–12051. [[CrossRef](#)] [[PubMed](#)]
22. Dunning Hotopp, J.C.; Clark, M.E.; Oliveira, D.C.S.G.; Foster, J.M.; Fischer, P.; Muñoz Torres, M.C.; Giebel, J.D.; Kumar, N.; Ishmael, N.; Wang, S.; et al. Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science* **2007**, *317*, 1753–1756. [[CrossRef](#)] [[PubMed](#)]
23. Raychoudhury, R.; Baldo, L.; Oliveira, D.C.S.G.; Werren, J.H. Modes of acquisition of *Wolbachia*: Horizontal transfer, hybrid introgression, and codivergence in the *Nasonia* species complex. *Evol. Int. J. Org. Evol.* **2009**, *63*, 165–183. [[CrossRef](#)] [[PubMed](#)]
24. Palazzo, A.; Escuder, E.; D’Addabbo, P.; Lovero, D.; Marsano, R.M. A genomic survey of Tc1-mariner transposons in nematodes suggests extensive horizontal transposon transfer events. *Mol. Phylogenet. Evol.* **2021**, *158*, 107090. [[CrossRef](#)]
25. White, P.M.; Pietri, J.E.; Debec, A.; Russell, S.; Patel, B.; Sullivan, W. Mechanisms of Horizontal Cell-to-Cell Transfer of *Wolbachia* spp. in *Drosophila melanogaster*. *Appl. Environ. Microbiol.* **2017**, *83*, e03425-16. [[CrossRef](#)]
26. Stroun, M.; Lyautey, J.; Lederrey, C.; Mulcahy, H.E.; Anker, P. Alu repeat sequences are present in increased proportions compared to a unique gene in plasma/serum DNA: Evidence for a preferential release from viable cells? *Ann. N. Y. Acad. Sci.* **2001**, *945*, 258–264. [[CrossRef](#)]
27. Kawamura, Y.; Sanchez Calle, A.; Yamamoto, Y.; Sato, T.A.; Ochiya, T. Extracellular vesicles mediate the horizontal transfer of an active LINE-1 retrotransposon. *J. Extracell. Vesicles* **2019**, *8*, 1643214. [[CrossRef](#)]
28. Tkach, M.; Théry, C. Communication by Extracellular Vesicles: Where We Are and Where We Need to Go. *Cell* **2016**, *164*, 1226–1232. [[CrossRef](#)]
29. Wang, L.; Dou, K.; Moon, S.; Tan, F.J.; Zhang, Z.Z. Hijacking Oogenesis Enables Massive Propagation of LINE and Retroviral Transposons. *Cell* **2018**, *174*, 1082–1094.e12. [[CrossRef](#)]
30. Tiwari, B.; Kurtz, P.; Jones, A.E.; Wylie, A.; Amatruda, J.F.; Boggupalli, D.P.; Gonsalvez, G.B.; Abrams, J.M. Retrotransposons Mimic Germ Plasm Determinants to Promote Transgenerational Inheritance. *Curr. Biol.* **2017**, *27*, 3010–3016.e3. [[CrossRef](#)]
31. Shpiz, S.; Kwon, D.; Uneva, A.; Kim, M.; Klenov, M.; Rozovsky, Y.; Georgiev, P.; Savitsky, M.; Kalmykova, A. Characterization of *Drosophila* telomeric retroelement TAHRE: Transcription, transpositions, and RNAi-based regulation of expression. *Mol. Biol. Evol.* **2007**, *24*, 2535–2545. [[CrossRef](#)] [[PubMed](#)]
32. Van De Bor, V.; Hartwood, E.; Jones, C.; Finnegan, D.; Davis, I. gurken and the I factor retrotransposon RNAs share common localization signals and machinery. *Dev. Cell* **2005**, *9*, 51–62. [[CrossRef](#)] [[PubMed](#)]
33. Moon, S.; Cassani, M.; Lin, Y.A.; Wang, L.; Dou, K.; Zhang, Z.Z. A Robust Transposon-Endogenizing Response from Germline Stem Cells. *Dev. Cell* **2018**, *47*, 660–671.e3. [[CrossRef](#)] [[PubMed](#)]
34. Suh, D.S.; Choi, E.H.; Yamazaki, T.; Harada, K. Studies on the transposition rates of mobile genetic elements in a natural population of *Drosophila melanogaster*. *Mol. Biol. Evol.* **1995**, *12*, 748–758. [[CrossRef](#)]
35. Nuzhdin, S.V.; Mackay, T.F. The genomic rate of transposable element movement in *Drosophila melanogaster*. *Mol. Biol. Evol.* **1995**, *12*, 180–181. [[CrossRef](#)]
36. Maside, X.; Assimacopoulos, S.; Charlesworth, B. Rates of movement of transposable elements on the second chromosome of *Drosophila melanogaster*. *Genet. Res.* **2000**, *75*, 275–284. [[CrossRef](#)]
37. Le Rouzic, A.; Capy, P. The first steps of transposable elements invasion: Parasitic strategy vs. genetic drift. *Genetics* **2005**, *169*, 1033–1043. [[CrossRef](#)]
38. Kofler, R.; Senti, K.A.; Nolte, V.; Tobler, R.; Schlötterer, C. Molecular dissection of a natural transposable element invasion. *Genome Res.* **2018**, *28*, 824–835. [[CrossRef](#)]
39. Robillard, É.; Le Rouzic, A.; Zhang, Z.; Capy, P.; Hua-Van, A. Experimental evolution reveals hyperparasitic interactions among transposable elements. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 14763–14768. [[CrossRef](#)]

40. Schaack, S.; Gilbert, C.; Feschotte, C. Promiscuous DNA: Horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends Ecol. Evol.* **2010**, *25*, 537–546. [[CrossRef](#)]
41. Zhang, H.H.; Peccoud, J.; Xu, M.R.X.; Zhang, X.G.; Gilbert, C. Horizontal transfer and evolution of transposable elements in vertebrates. *Nat. Commun.* **2020**, *11*, 1362. [[CrossRef](#)] [[PubMed](#)]
42. Capy, P. Taming, Domestication and Exaptation: Trajectories of Transposable Elements in Genomes. *Cells* **2021**, *10*, 3590. [[CrossRef](#)] [[PubMed](#)]
43. Hickey, D.A. Selfish DNA: A sexually transmitted nuclear parasite. *Genetics* **1982**, *101*, 519–531. [[CrossRef](#)] [[PubMed](#)]
44. Kofler, R. Dynamics of transposable element invasions with piRNA clusters. *Mol. Biol. Evol.* **2019**, *36*, 1457–1472. [[CrossRef](#)]
45. Laski, F.A.; Rio, D.C.; Rubin, G.M. Tissue specificity of Drosophila P element transposition is regulated at the level of mRNA splicing. *Cell* **1986**, *44*, 7–19. [[CrossRef](#)]
46. Chaboissier, M.C.; Busseau, I.; Prosser, J.; Finnegan, D.J.; Bucheton, A. Identification of a potential RNA intermediate for transposition of the LINE-like element I factor in Drosophila melanogaster. *EMBO J.* **1990**, *9*, 3557–3563. [[CrossRef](#)]
47. del Carmen Seleme, M.; Disson, O.; Robin, S.; Brun, C.; Teninges, D.; Bucheton, A. In vivo RNA localization of I factor, a non-LTR retrotransposon, requires a cis-acting signal in ORF2 and ORF1 protein. *Nucleic Acids Res.* **2005**, *33*, 776–785. [[CrossRef](#)]
48. Rogers, R.L.; Slatkin, M. Excess of genomic defects in a woolly mammoth on Wrangel island. *PLoS Genet.* **2017**, *13*, e1006601. [[CrossRef](#)]
49. Platt, R.N.; Ray, D.A. A non-LTR retroelement extinction in Spermophilus tridecemlineatus. *Gene* **2012**, *500*, 47–53. [[CrossRef](#)]
50. Dufourt, J.; Vaury, C. During a short window of Drosophila oogenesis, piRNA biogenesis may be boosted and mobilization of transposable elements allowed. *Front. Genet.* **2014**, *5*, 8–11. [[CrossRef](#)]
51. Théron, E.; Maupetit-Mehouas, S.; Pouchin, P.; Baudet, L.; Brasset, E.; Vaury, C. The interplay between the argonaute proteins piwi and aub within drosophila gemarium is critical for oogenesis, piRNA biogenesis and TE silencing. *Nucleic Acids Res.* **2018**, *46*, 10052–10065. [[CrossRef](#)] [[PubMed](#)]
52. Dufourt, J.; Dennis, C.; Boivin, A.; Gueguen, N.; Théron, E.; Goriaux, C.; Pouchin, P.; Ronsseray, S.; Brasset, E.; Vaury, C. Spatio-temporal requirements for transposable element piRNA-mediated silencing during Drosophila oogenesis. *Nucleic Acids Res.* **2014**, *42*, 2512–2524. [[CrossRef](#)]
53. Seisenberger, S.; Andrews, S.; Krueger, F.; Arand, J.; Walter, J.; Santos, F.; Popp, C.; Thienpont, B.; Dean, W.; Reik, W. The Dynamics of Genome-wide DNA Methylation Reprogramming in Mouse Primordial Germ Cells. *Mol. Cell* **2012**, *48*, 849–862. [[CrossRef](#)] [[PubMed](#)]
54. Evsikov, A.V.; De Vries, W.N.; Peaston, A.E.; Radford, E.E.; Fancher, K.S.; Chen, F.H.; Blake, J.A.; Bult, C.J.; Latham, K.E.; Solter, D.; et al. Systems biology of the 2-cell mouse embryo. *Cytogenet. Genome Res.* **2004**, *105*, 240–250. [[CrossRef](#)] [[PubMed](#)]
55. Fadloun, A.; Le Gras, S.; Jost, B.; Ziegler-Birling, C.; Takahashi, H.; Gorab, E.; Carninci, P.; Torres-Padilla, M.E. Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. *Nat. Struct. Mol. Biol.* **2013**, *20*, 332–338. [[CrossRef](#)]
56. Kigami, D.; Minami, N.; Takayama, H.; Imai, H. MuERV-L is one of the earliest transcribed genes in mouse one-cell embryos. *Biol. Reprod.* **2003**, *68*, 651–654. [[CrossRef](#)]
57. Maupetit-Mehouas, S.; Vaury, C. Transposon Reactivation in the Germline May Be Useful for Both Transposons and Their Host Genomes. *Cells* **2020**, *9*, 1172. [[CrossRef](#)]
58. Chalvet, F.; Teyssset, L.; Terzian, C.; Prud'homme, N.; Santamaria, P.; Bucheton, A.; Pélisson, A. Proviral amplification of the Gypsy endogenous retrovirus of Drosophila melanogaster involves env-independent invasion of the female germline. *EMBO J.* **1999**, *18*, 2659–2669. [[CrossRef](#)]
59. Brasset, E.; Taddei, A.R.; Arnaud, F.; Faye, B.; Fausto, A.M.; Mazzini, M.; Giorgi, F.; Vaury, C. Viral particles of the endogenous retrovirus ZAM from Drosophila melanogaster use a pre-existing endosome/exosome pathway for transfer to the oocyte. *Retrovirology* **2006**, *3*, 25. [[CrossRef](#)]
60. Syomin, B.V.; Fedorova, L.I.; Surkov, S.A.; Ilyin, Y.V. The endogenous Drosophila melanogaster retrovirus gypsy can propagate in Drosophila hydei cells. *Mol. Gen. Genet.* **2001**, *264*, 588–594. [[CrossRef](#)]
61. Keegan, R.M.; Talbot, L.R.; Chang, Y.H.; Metzger, M.J.; Dubnau, J. Intercellular viral spread and intracellular transposition of Drosophila gypsy. *PLoS Genet.* **2021**, *17*, e1009535. [[CrossRef](#)] [[PubMed](#)]
62. Kim, A.; Terzian, C.; Santamaria, P.; Pélisson, A.; Prud'homme, N.; Bucheton, A. Retroviruses in invertebrates: The gypsy retrotransposon is apparently an infectious retrovirus of Drosophila melanogaster. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 1285–1289. [[CrossRef](#)] [[PubMed](#)]
63. Leblanc, P.; Desset, S.; Giorgi, F.; Taddei, A.R.; Fausto, A.M.; Mazzini, M.; Dastugue, B.; Vaury, C. Life Cycle of an Endogenous Retrovirus, ZAM, in Drosophila melanogaster. *J. Virol.* **2002**, *74*, 10658–10669. [[CrossRef](#)]
64. Tcheressiz, S.; Calco, V.; Arnaud, F.; Arthaud, L.; Dastugue, B.; Vaury, C. Expression of the Idefix retrotransposon in early follicle cells in the gemarium of Drosophila melanogaster is determined by its LTR sequences and a specific genomic context. *Mol. Genet. Genomics* **2002**, *267*, 133–141. [[CrossRef](#)] [[PubMed](#)]
65. Malone, C.D.; Brennecke, J.; Dus, M.; Stark, A.; McCombie, W.R.; Sachidanandam, R.; Hannon, G.J. Specialized piRNA Pathways Act in Germline and Somatic Tissues of the Drosophila Ovary. *Cell* **2009**, *137*, 522–535. [[CrossRef](#)]
66. Olivieri, D.; Sykora, M.M.; Sachidanandam, R.; Mechtler, K.; Brennecke, J. An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in Drosophila. *EMBO J.* **2010**, *29*, 3301–3317. [[CrossRef](#)] [[PubMed](#)]

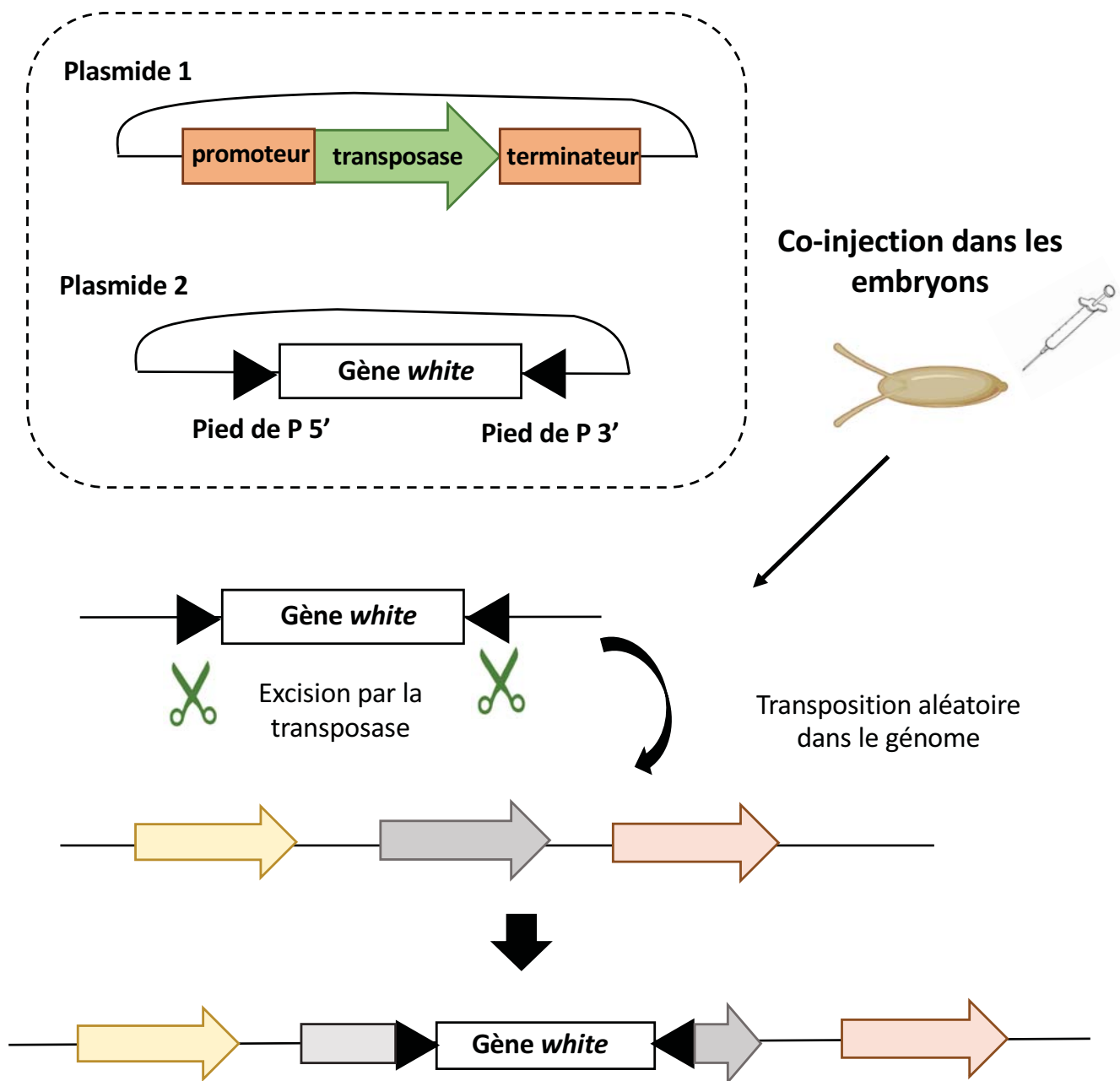


67. Pélişson, A.; Song, S.U.; Prud'homme, N.; Smith, P.A.; Bucheton, A.; Corces, V.G. Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the *Drosophila flamenco* gene. *EMBO J.* **1994**, *13*, 4401–4411. [[CrossRef](#)]
68. Sokolova, O.A.; Mikhaleva, E.A.; Kharitonov, S.L.; Abramov, Y.A.; Gvozdev, V.A.; Klenov, M.S. Special vulnerability of somatic niche cells to transposable element activation in *Drosophila* larval ovaries. *Sci. Rep.* **2020**, *10*, 1076. [[CrossRef](#)]
69. Cavarec, L.; Jensen, S.; Casella, J.F.; Cristescu, S.A.; Heidmann, T. Molecular cloning and characterization of a transcription factor for the copia retrotransposon with homology to the BTB-containing *lola* neurogenic factor. *Mol. Cell. Biol.* **1997**, *17*, 482–494. [[CrossRef](#)]
70. Meignin, C.; Dastugue, B.; Vaury, C. Intercellular communication between germ line and somatic line is utilized to control the transcription of ZAM, an endogenous retrovirus from *Drosophila melanogaster*. *Nucleic Acids Res.* **2004**, *32*, 3799–3806. [[CrossRef](#)]
71. Barckmann, B.; El-Barouk, M.; Pélişson, A.; Mugat, B.; Li, B.; Franckhauser, C.; Fiston Lavier, A.-S.; Mirouze, M.; Fablet, M.; Chambeyron, S. The somatic piRNA pathway controls germline transposition over generations. *Nucleic Acids Res.* **2018**, *46*, 9524–9536. [[CrossRef](#)] [[PubMed](#)]
72. Duc, C.; Yoth, M.; Jensen, S.; Mouni e, N.; Bergman, C.M.; Vaury, C.; Brasset, E. Trapping a somatic endogenous retrovirus into a germline piRNA cluster immunizes the germline against further invasion. *Genome Biol.* **2019**, *20*, 127. [[CrossRef](#)] [[PubMed](#)]
73. Dessel, S.; Conte, C.; Dimitri, P.; Calco, V.; Dastugue, B.; Vaury, C. Mobilization of two retroelements, ZAM and Idefix, in a novel unstable line of *Drosophila melanogaster*. *Mol. Biol. Evol.* **1999**, *16*, 54–66. [[CrossRef](#)] [[PubMed](#)]
74. Kidwell, M.G.; Kidwell, J.F.; Sved, J.A. Hybrid dysgenesis in *Drosophila melanogaster*: A syndrome of aberrant traits including mutation, sterility and male recombination. *Genetics* **1977**, *86*, 813–833. [[CrossRef](#)]
75. Rubin, G.M.; Kidwell, M.G.; Bingham, P.M. The molecular basis of P-M hybrid dysgenesis: The nature of induced mutations. *Cell* **1982**, *29*, 987–994. [[CrossRef](#)]
76. Picard, G. Non mendelian female sterility in *Drosophila melanogaster*: Hereditary transmission of I factor. *Genetics* **1976**, *83*, 107–123. [[CrossRef](#)]
77. Busseau, I.; Chaboissier, M.C.; Pélişson, A.; Bucheton, A. I factors in *Drosophila melanogaster*: Transposition under control. *Genetica* **1994**, *93*, 101–116. [[CrossRef](#)]
78. Evgen'ev, M.B.; Zelentsova, H.; Shostak, N.; Kozitsina, M.; Barskyi, V.; Lankenau, D.H.; Corces, V.G. Penelope, a new family of transposable elements and its possible role in hybrid dysgenesis in *Drosophila virilis*. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 196–201. [[CrossRef](#)]
79. Blackman, R.K.; Grimaila, R.; Macy, M.; Koehler, D.; Gelbart, W.M. Mobilization of hobo elements residing within the decapentaplegic gene complex: Suggestion of a new hybrid dysgenesis system in *Drosophila melanogaster*. *Cell* **1987**, *49*, 497–505. [[CrossRef](#)]
80. Serrato-Capuchina, A.; Matute, D.R. The role of transposable elements in speciation. *Genes* **2018**, *9*, 254. [[CrossRef](#)]
81. Belyayev, A. Bursts of transposable elements as an evolutionary driving force. *J. Evol. Biol.* **2014**, *27*, 2573–2584. [[CrossRef](#)] [[PubMed](#)]
82. Engels, W.R. Hybrid dysgenesis in *Drosophila melanogaster*: Rules of inheritance of female sterility. *Genet. Res.* **2007**, *89*, 407–424. [[CrossRef](#)] [[PubMed](#)]
83. Kidwell, M.G. Hybrid dysgenesis in *Drosophila melanogaster*: Factors affecting chromosomal contamination in the P-M system. *Genetics* **1983**, *104*, 317–341. [[CrossRef](#)] [[PubMed](#)]
84. Brennecke, J.; Malone, C.D.; Aravin, A.A.; Sachidanandam, R.; Stark, A.; Hannon, G.J. An Epigenetic Role for Maternally Inherited piRNAs in Transposon Silencing. *Science* **2008**, *1387*, 1387–1392. [[CrossRef](#)]
85. Khurana, J.S.; Wang, J.; Xu, J.; Koppetsch, B.S.; Thomson, T.C.; Nowosielska, A.; Li, C.; Zamore, P.D.; Weng, Z.; Theurkauf, W.E. Adaptation to P element transposon invasion in *Drosophila melanogaster*. *Cell* **2011**, *147*, 1551–1563. [[CrossRef](#)]
86. Aravin, A.; Gaidatzis, D.; Iovino, N.; Morris, P.; Brownstein, M.J.; Kuramochi-miyagawa, S.; Nakano, T.; Chien, M.; Russo, J.J.; Ju, J.; et al. A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* **2006**, *442*, 203–207. [[CrossRef](#)]
87. Girard, A.; Sachidanandam, R.; Hannon, G.J.; Carmell, M.A. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* **2006**, *442*, 199–202. [[CrossRef](#)]
88. Grivna, S.T.; Beyret, E.; Wang, Z.; Lin, H. A novel class of small RNAs in mouse spermatogenic cells. *Genes Dev.* **2006**, *20*, 1709–1714. [[CrossRef](#)]
89. Lau, N.C.; Seto, A.G.; Kim, J.; Kuramochi-Miyagawa, S.; Nakano, T.; Bartel, D.P.; Kingston, R.E. Characterization of the piRNA complex from rat testes. *Science* **2006**, *313*, 363–367. [[CrossRef](#)]
90. Watanabe, T.; Takeda, A.; Tsukiyama, T.; Mise, K.; Okuno, T.; Sasaki, H.; Minami, N.; Imai, H. Identification and characterization of two novel classes of small RNAs in the mouse germline: Retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes Dev.* **2006**, *20*, 1732–1743. [[CrossRef](#)]
91. Gainetdinov, I.; Colpan, C.; Arif, A.; Cecchini, K.; Zamore, P.D. A Single Mechanism of Biogenesis, Initiated and Directed by PIWI Proteins, Explains piRNA Production in Most Animals. *Mol. Cell* **2018**, *71*, 775–790.e5. [[CrossRef](#)] [[PubMed](#)]
92. Bhattacharya, S.; Bakre, A.; Bhattacharya, A. Mobile genetic elements in protozoan parasites. *J. Genet.* **2002**, *81*, 73–86. [[CrossRef](#)] [[PubMed](#)]
93. Sienski, G.; D nertas, D.; Brennecke, J. Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell* **2012**, *151*, 964–980. [[CrossRef](#)] [[PubMed](#)]

94. De Vanssay, A.; Bougé, A.L.; Boivin, A.; Hermant, C.; Teyssset, L.; Delmarre, V.; Antoniewski, C.; Ronsseray, S. Paramutation in *Drosophila* linked to emergence of a piRNA-producing locus. *Nature* **2012**, *490*, 112–115. [[CrossRef](#)]
95. Le Thomas, A.; Stuwe, E.; Li, S.; Du, J.; Marinov, G.; Rozhkov, N.; Chen, Y.C.A.; Luo, Y.; Sachidanandam, R.; Toth, K.F.; et al. Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. *Genes Dev.* **2014**, *28*, 1667–1680. [[CrossRef](#)]
96. Marin, L.; Lehmann, M.; Nouaud, D.; Izaabel, H.; Anxolabéhère, D.; Ronsseray, S. P-element repression in *Drosophila melanogaster* by a naturally occurring defective telomeric P copy. *Genetics* **2000**, *155*, 1841–1854. [[CrossRef](#)]
97. Ronsseray, S.; Lehmann, M.; Anxolabehere, D. The maternally inherited regulation of P elements in *Drosophila melanogaster* can be elicited by two P copies at cytological site 1A on the X chromosome. *Genetics* **1991**, *129*, 501–512. [[CrossRef](#)]
98. Ronsseray, S.; Josse, T.; Boivin, A.; Anxolabéhère, D. Telomeric transgenes and trans-silencing in *Drosophila*. *Genetica* **2003**, *117*, 327–335. [[CrossRef](#)]
99. Casier, K.; Delmarre, V.; Gueguen, N.; Hermant, C.; Viodé, E.; Vaury, C.; Ronsseray, S.; Brasset, E.; Teyssset, L.; Boivin, A. Environmentally-induced epigenetic conversion of a piRNA cluster. *Elife* **2019**, *8*, e39842. [[CrossRef](#)]
100. Marie, P.P.; Ronsseray, S.; Boivin, A. From embryo to adult: PiRNA-mediated silencing throughout germline development in *Drosophila*. *G3 Genes Genomes Genet.* **2017**, *7*, 505–516. [[CrossRef](#)]
101. Hermant, C.; Boivin, A.; Teyssset, L.; Delmarre, V.; Asif-Laidin, A.; Van Den Beek, M.; Antoniewski, C.; Ronsseray, S. Paramutation in *Drosophila* requires both nuclear and cytoplasmic actors of the piRNA pathway and induces cis-spreading of piRNA production. *Genetics* **2015**, *201*, 1381–1396. [[CrossRef](#)] [[PubMed](#)]
102. de Vanssay, A.; Bougé, A.L.; Boivin, A.; Hermant, C.; Teyssset, L.; Delmarre, V.; Antoniewski, C.; Ronsseray, S. piRNAs and epigenetic conversion in *Drosophila*. *Fly* **2013**, *7*, 237–241. [[CrossRef](#)] [[PubMed](#)]
103. Todeschini, A.L.; Teyssset, L.; Delmarre, V.; Ronsseray, S. The epigenetic Trans-silencing effect in *Drosophila* involves maternally-transmitted small RNAs whose production depends on the piRNA pathway and HP1. *PLoS ONE* **2010**, *5*, e11032. [[CrossRef](#)] [[PubMed](#)]
104. Josse, T.; Teyssset, L.; Todeschini, A.L.; Sidor, C.M.; Anxolabéhère, D.; Ronsseray, S. Telomeric trans-silencing: An epigenetic repression combining RNA silencing and heterochromatin formation. *PLoS Genet.* **2007**, *3*, e158. [[CrossRef](#)] [[PubMed](#)]
105. Teixeira, F.K.; Okuniewska, M.; Malone, C.D.; Coux, R.X.; Rio, D.C.; Lehmann, R. PiRNA-mediated regulation of transposon alternative splicing in the soma and germ line. *Nature* **2017**, *552*, 268–272. [[CrossRef](#)]
106. Mugnier, N.; Gueguen, L.; Vieira, C.; Biémont, C. The heterochromatic copies of the LTR retrotransposons as a record of the genomic events that have shaped the *Drosophila melanogaster* genome. *Gene* **2008**, *411*, 87–93. [[CrossRef](#)]
107. Moltó, M.D.; Paricio, N.; López-Preciado, M.A.; Semeshin, V.F.; Martínez-Sebastián, M.J. Tirant: A new retrotransposon-like element in *Drosophila melanogaster*. *J. Mol. Evol.* **1996**, *42*, 369–375. [[CrossRef](#)]
108. Schwarz, F.; Wierzbicki, F.; Senti, K.-A.; Kofler, R. Tirant stealthily invaded natural *Drosophila melanogaster* populations during the last century. *Mol. Biol. Evol.* **2021**, *38*, 1482–1497. [[CrossRef](#)]
109. Zanni, V.; Eymery, A.; Coiffet, M.; Zytnicki, M.; Luyten, I.; Quesneville, H.; Vaury, C.; Jensen, S. Distribution, evolution, and diversity of retrotransposons at the flamenco locus reflect the regulatory properties of piRNA clusters. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 19842–19847. [[CrossRef](#)]
110. Schaefer, R.E.; Kidwell, Z.M.G.; Fausto-sterling, A. Hybrid dysgenesis in *Drosophila melanogaster*: Morphological and cytological studies of ovarian dysgenesis. *Public Health* **1979**, 1141–1152. [[CrossRef](#)]
111. Tarlinton, R.E.; Meers, J.; Young, P.R. Retroviral invasion of the koala genome. *Nature* **2006**, *442*, 79–81. [[CrossRef](#)] [[PubMed](#)]
112. Chappell, K.J.; Brealey, J.C.; Amarilla, A.A.; Watterson, D.; Hulse, L.; Palmieri, C.; Johnston, S.D.; Holmes, E.C.; Meers, J.; Young, P.R. Phylogenetic Diversity of Koala Retrovirus within a Wild Koala Population. *J. Virol.* **2017**, *91*, e01820-16. [[CrossRef](#)] [[PubMed](#)]
113. Denner, J.; Young, P.R. Koala retroviruses: Characterization and impact on the life of koalas. *Retrovirology* **2013**, *10*, 108. [[CrossRef](#)] [[PubMed](#)]
114. Ishida, Y.; Zhao, K.; Greenwood, A.D.; Roca, A.L. Proliferation of endogenous retroviruses in the early stages of a host germ line invasion. *Mol. Biol. Evol.* **2015**, *32*, 109–120. [[CrossRef](#)]
115. Yu, T.; Koppetsch, B.; Chappell, K.; Pagliarani, S.; Johnston, S.; Silverstein, N.J.; Luban, J.; Weng, Z.; Thauerkauf, W.E. The PiRNA Response to Retroviral Invasion of the Koala Genome. *Cell* **2019**, *179*, 632–643. [[CrossRef](#)]
116. Gebert, D.; Neubert, L.K.; Lloyd, C.; Gui, J.; Lehmann, R.; Teixeira, F.K. Large *Drosophila* germline piRNA clusters are evolutionarily labile and dispensable for transposon regulation. *Mol. Cell* **2021**, *81*, 3965–3978. [[CrossRef](#)]
117. Andersen, P.R.; Tirian, L.; Vunjak, M.; Brennecke, J. A heterochromatin-dependent transcription machinery drives piRNA expression. *Nature* **2017**, *549*, 54–59. [[CrossRef](#)]
118. Mohn, F.; Sienski, G.; Handler, D.; Brennecke, J. The Rhino-Deadlock-Cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in *Drosophila*. *Cell* **2014**, *157*, 1364–1379. [[CrossRef](#)]
119. Olovnikov, I.A.; Kalmykova, A.I. piRNA clusters as a main source of small RNAs in the animal germline. *Biochemistry* **2013**, *78*, 572–584. [[CrossRef](#)]
120. Shpiz, S.; Ryazansky, S.; Olovnikov, I.; Abramov, Y.; Kalmykova, A. Euchromatic Transposon Insertions Trigger Production of Novel Pi- and Endo-siRNAs at the Target Sites in the *Drosophila* Germline. *PLoS Genet.* **2014**, *10*, e1004138. [[CrossRef](#)]

121. Luo, Y.; He, P.; Kanrar, N.; Toth, K.F.; Aravin, A.A. Maternally inherited siRNAs initiate piRNA cluster formation. *bioRxiv* **2022**, preprint. [[CrossRef](#)]
122. Czech, B.; Malone, C.D.; Zhou, R.; Stark, A.; Schlingeheyde, C.; Dus, M.; Perrimon, N.; Kellis, M.; Wohlschlegel, J.A.; Sachidanandam, R.; et al. An endogenous small interfering RNA pathway in *Drosophila*. *Nature* **2008**, *453*, 798–802. [[CrossRef](#)] [[PubMed](#)]
123. Chung, W.J.; Okamura, K.; Martin, R.; Lai, E.C. Endogenous RNA Interference Provides a Somatic Defense against *Drosophila* Transposons. *Curr. Biol.* **2008**, *18*, 795–802. [[CrossRef](#)] [[PubMed](#)]
124. Ghildiyal, M.; Zamore, P.D. Small silencing RNAs: An expanding universe. *Nat. Rev. Genet.* **2009**, *10*, 94–108. [[CrossRef](#)]
125. Okamura, K.; Chung, W.J.; Ruby, J.G.; Guo, H.; Bartel, D.P.; Lai, E.C. The *Drosophila* hairpin RNA pathway generates endogenous short interfering RNAs. *Nature* **2008**, *453*, 803–806. [[CrossRef](#)]
126. Rozhkov, N.V.; Aravin, A.A.; Zelentsova, E.S.; Schostak, N.G.; Sachidanandam, R.; McCombie, W.R.; Hannon, G.J.; Evgen'ev, M.B. Small RNA-based silencing strategies for transposons in the process of invading *Drosophila* species. *RNA* **2010**, *16*, 1634–1645. [[CrossRef](#)]
127. Rehwinkel, J.; Natalin, P.; Stark, A.; Brennecke, J.; Cohen, S.M.; Izaurralde, E. Genome-Wide Analysis of mRNAs Regulated by Droscha and Argonaute Proteins in *Drosophila melanogaster*. *Mol. Cell. Biol.* **2006**, *26*, 2965–2975. [[CrossRef](#)]
128. Kawamura, Y.; Saito, K.; Kin, T.; Ono, Y.; Asai, K.; Sunohara, T.; Okada, T.N.; Siomi, M.C.; Siomi, H. *Drosophila* endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* **2008**, *453*, 793–797. [[CrossRef](#)]
129. Rozhkov, N.V.; Schostak, N.C.; Zelentsova, E.S.; Yushenova, I.A.; Zatssepina, O.G.; Evgen'ev, M.B. Evolution and dynamics of small RNA response to a retroelement invasion in *Drosophila*. *Mol. Biol. Evol.* **2013**, *30*, 397–408. [[CrossRef](#)]
130. Sarkies, P.; Selkirk, M.E.; Jones, J.T.; Blok, V.; Boothby, T.; Goldstein, B.; Hanelt, B.; Ardila-Garcia, A.; Fast, N.M.; Schiffer, P.M.; et al. Ancient and novel small RNA pathways compensate for the loss of piRNAs in multiple independent nematode lineages. *PLoS Biol.* **2015**, *13*, e1002061. [[CrossRef](#)]
131. Desset, S.; Meignin, C.; Dastugue, B.; Vaury, C. COM, a heterochromatic locus governing the control of independent endogenous retroviruses from *Drosophila melanogaster*. *Genetics* **2003**, *164*, 501–509. [[CrossRef](#)] [[PubMed](#)]
132. Mohamed, M.; Dang, N.T.M.; Ogyama, Y.; Burlet, N.; Mugat, B.; Boulesteix, M.; Mérel, V.; Veber, P.; Salces-Ortiz, J.; Severac, D.; et al. A Transposon Story: From TE Content to TE Dynamic Invasion of *Drosophila* Genomes Using the Single-Molecule Sequencing Technology from Oxford Nanopore. *Cells* **2020**, *9*, 1776. [[CrossRef](#)] [[PubMed](#)]
133. George, P.; Jensen, S.; Pogorelnik, R.; Lee, J.; Xing, Y.; Brasslet, E.; Vaury, C.; Sharakhov, I.V. Increased production of piRNAs from euchromatic clusters and genes in *Anopheles gambiae* compared with *Drosophila melanogaster*. *Epigenet. Chromatin* **2015**, *8*, 50. [[CrossRef](#)] [[PubMed](#)]
134. Wierzbicki, F.; Kofler, R.; Signor, S. Evolutionary dynamics of piRNA clusters in *Drosophila*. *Mol. Ecol.* **2021**, 1–17. [[CrossRef](#)]





**Figure 16 : Principe de la mutagenèse P chez *Drosophila melanogaster***

Deux plasmides différents sont co-injectés dans l'embryon (mutant pour le gène *white*). Le premier code pour la transposase de l'élément P, sous le contrôle d'un promoteur fort pour permettre son expression très précocement dans les cellules germinales (Transgène 1). Le deuxième transgène contient un élément P modifié où les séquences internes ont été remplacées par des séquences d'intérêt associées à un gène *white*, permettant d'analyser la présence du transgène par la coloration des yeux : rouges dans ce cas, blancs en absence du transgène. Le gène *white* est borné de séquences ITR de P (aussi appelées « pieds de P ») qui permettent la reconnaissance et l'excision du transposon modifié par la transposase (Transgène 2). Après injection dans l'embryon, la transposase exprimée dans les cellules à l'origine des futures cellules germinales catalyse la coupure au niveau des séquences ITR présentes sur le transgène 2. Après excision, la transposase permet l'intégration du transposon modifié de manière aléatoire dans le génome. Après intégration, le gène *white* est utilisé comme indicateur visuel pour suivre l'insertion avec succès du transposon. Il est ensuite facile d'analyser le site d'insertion du transposon par des méthodes de biologie moléculaire.

## Partie 2 - Régulation des éléments transposables chez la drosophile

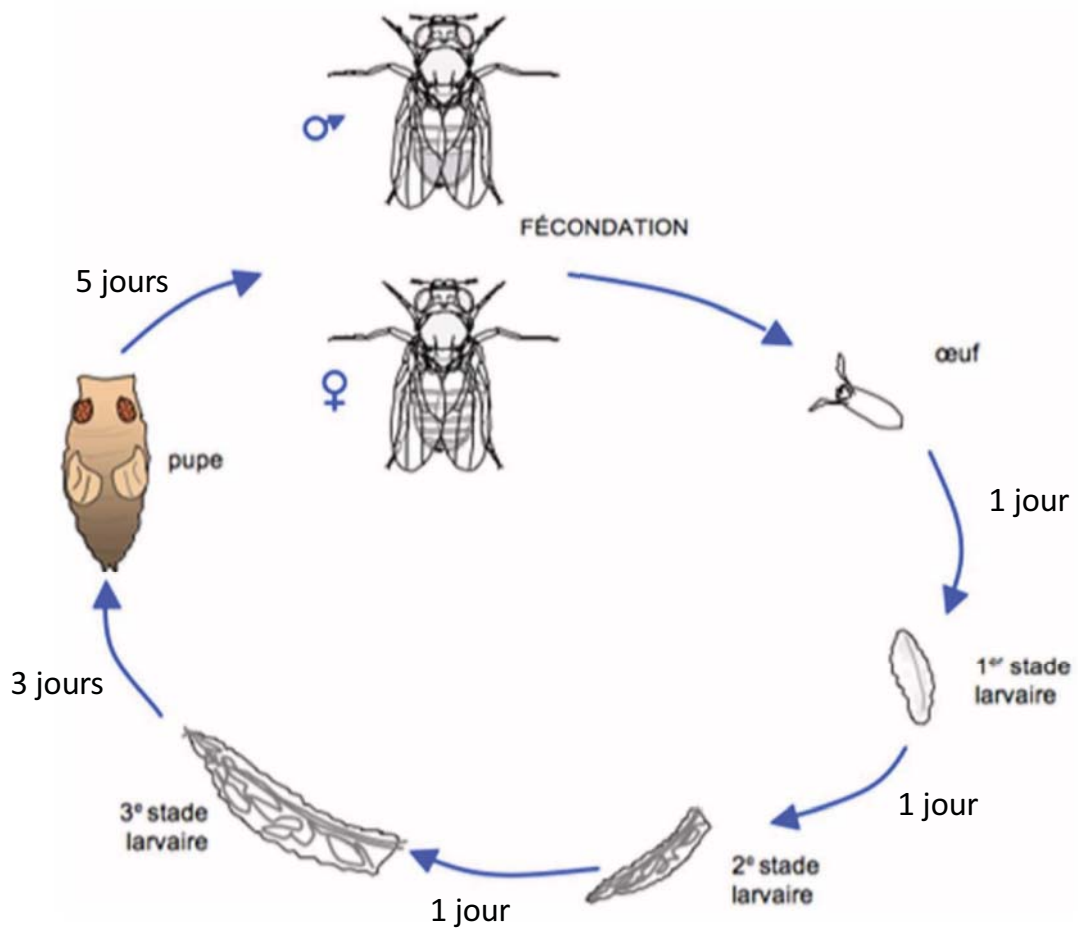
### I- La drosophile : la star de la génétique

Depuis le début des années 1900, la drosophile est utilisée pour sonder les mystères de l'hérédité.

Thomas Hunt Morgan était un précurseur dans le domaine de la génétique, il élevait une population de petites mouches de 2-3 mm communément appelées « mouches du vinaigre ». En 1910, il découvrit une petite variation chez un mâle : cette mouche avait les yeux blancs contrairement aux autres qui les avaient rouge brique. Il suivit alors la descendance de cet individu et étudia la transmission de ce caractère. Il découvrit que les mâles aux yeux blancs avaient une mutation dans un gène localisé sur le chromosome X, mutation récessive visible uniquement à l'état homozygote. Le gène en question sera nommé « *white* » pour le phénotype « yeux blanc » associé à sa mutation, inaugurant une tradition des généticiens de la drosophile à nommer les gènes d'après le phénotype de leurs allèles mutants. Morgan et son équipe ont ainsi décrit des dizaines de mutations, et il reçut en 1933 le prix Nobel de médecine en proposant les chromosomes comme support de l'hérédité.

En moins d'un siècle, plus de 200 000 publications concernant la drosophile seront accumulées par des milliers de chercheurs, faisant de la drosophile l'organisme modèle animal le mieux connu. Mais pourquoi un tel engouement pour cet insecte ? La première raison est d'ordre pratique : son temps de génération est très court (10 jours à 25 °C), son élevage est facile et peu coûteux, et son génome est composé uniquement de quatre petits chromosomes (environ 180 Mégabases au total). Une femelle peut pondre à elle seule jusqu'à quarante œufs par jour, et il peut y avoir environ 25 générations par an ! Avec un taux de reproduction aussi important, il est facile d'obtenir un grand nombre de descendants, permettant un suivi des populations et l'étude de leur évolution en accéléré (en comparaison de l'Homme qui nécessite des milliers d'années pour faire de même...).

De plus, l'étude des chromosomes polytènes, qui sont des chromosomes géants présents dans les glandes salivaires, a permis de cartographier relativement aisément le génome de la drosophile. Et en 1991, le projet de séquençage du génome de la souche l'espèce *Drosophila melanogaster* débute pour se terminer neuf ans plus tard (Adams et al., 2000). Malgré la distance évolutive importante avec l'Homme, on découvre que de nombreux gènes sont



**Figure 17 : Le cycle de vie de la drosophile**

L'embryon pondu sur le milieu éclot 24 heures après en larve. Trois stades larvaires vont ensuite se succéder durant 5 jours (à 25°C) puis, après le 3<sup>e</sup> stade larvaire, la larve s'encapsule dans une pupa. Le stade pupal dure également 5 jours. Durant cette période, la larve va se métamorphoser en mouche adulte qui émerge à l'issue du stade pupal. Au bout de quelques heures, les femelles vont s'accoupler avec les mâles et stockeront le sperme de ces derniers dans la spermathèque afin de féconder leurs œufs durant plusieurs jours, avant de s'accoupler à nouveau.

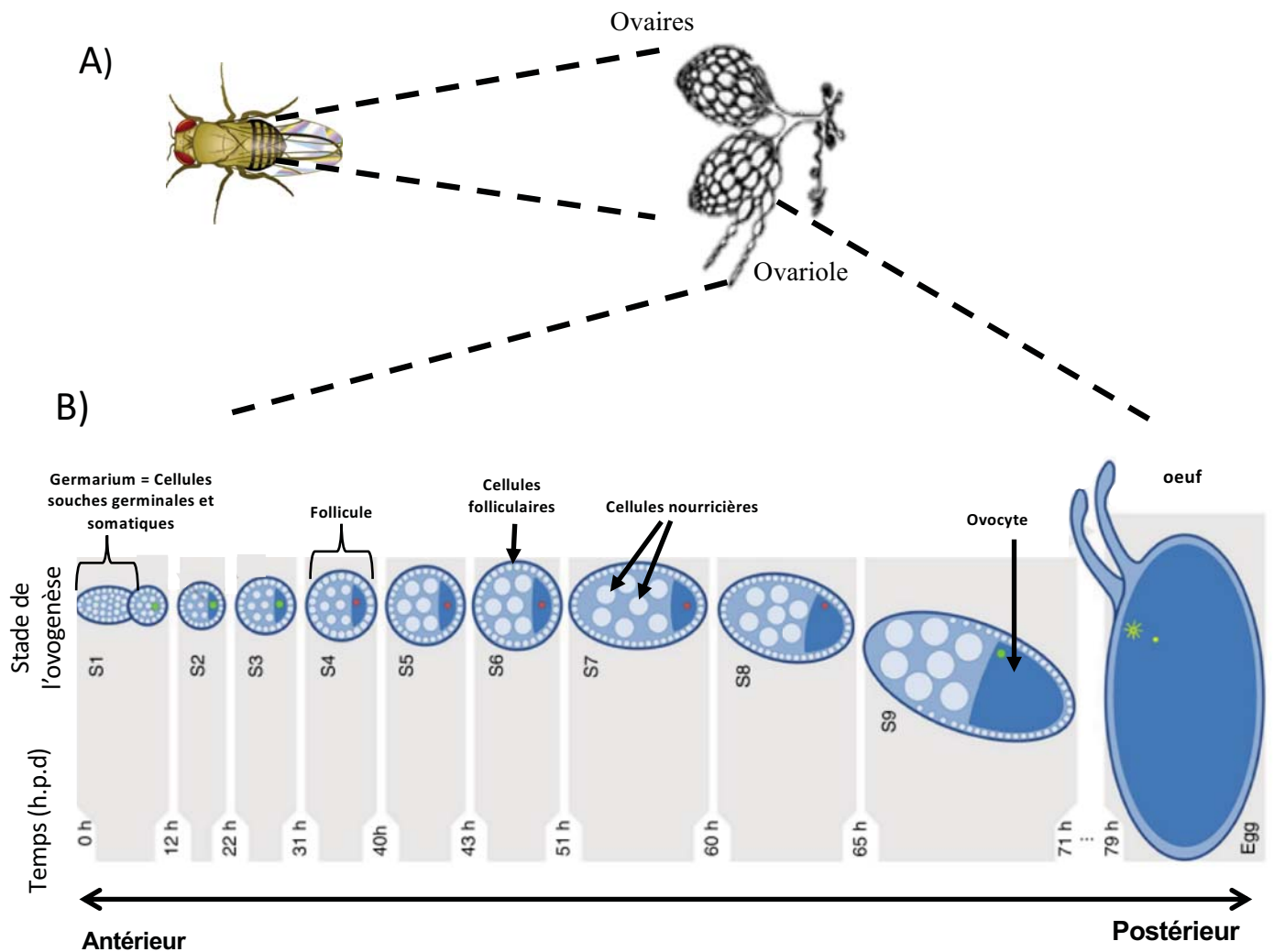
conservés : près de 60 % des gènes de la drosophile sont conservés entre les deux espèces, et plus de 75 % des gènes impliqués dans des pathologies humaines ont leur équivalent chez la drosophile (Revue dans : (Bier, 2005)). En outre, dès le début de son utilisation, de nombreuses méthodes ont été développées pour induire des mutations : rayon X, irradiation aux UV, mais aussi utilisation des éléments transposables ! Les connaissances acquises sur les éléments transposables ont donc aussi permis de faire des avancées en matière de génétique, participant à l'identification du rôle de nombreux gènes. En effet, la « mutagenèse P », utilisant les propriétés de la transposase de l'élément *P*, permet l'insertion aléatoire de cet élément dans le génome, ce qui a largement été utilisée pour générer des mutants dans les années 90 (Spradling and Rubin, 1982; Spradling et al., 1995) (**Figure 16**). Depuis, de nombreux autres outils de manipulation génétique ont été développés, faisant de la drosophile un modèle d'étude puissant. On peut par exemple citer la transgénèse ciblée qui permet notamment d'insérer un fragment d'ADN d'intérêt à un locus spécifique, ou bien l'utilisation du système CRISPR-Cas9 pour créer des mutations/délétions/insertions ciblées de manière tissu spécifique ou même de manière inductible. Le système d'ARN interférence est également un outil largement utilisé pour étudier la fonction d'un gène. Ainsi, à l'heure actuelle, cette petite mouche est impliquée au total dans six prix Nobel de physiologie ou médecine dont ceux de la découverte du système immunitaire et du cycle circadien chez la *Drosophile*.

## **I.1. Cycle de vie et ovogenèse chez la *Drosophile***

### **I.1. a) Cycle de vie de la *Drosophile***

Le cycle de vie de la drosophile se divise en quatre phases : embryonnaire, larvaire, pupal et adulte (**Figure 17**). La durée du cycle est variable selon la température (pour le plus grand bonheur des drosophilistes qui peuvent ainsi accélérer ou ralentir le développement des mouches à leur guise). Le cycle, qui normalement dure environ 10 jours à 25 °C, peut en effet durer seulement 8 jours à 30 °C et jusqu'à 50 jours à 12 °C. Tout commence par la ponte : les femelles pondent des œufs fécondés qui ont une longueur d'environ 0,5 mm. 24 h après la ponte, la larve sort de l'œuf lors de l'éclosion et croît pendant plusieurs jours (cinq jours à 25 °C) avant de s'encapsuler dans une pupa. Le stade pupal dure également 5 jours et, durant cette période, la larve se métamorphose en adulte. Après éclosion, les mâles et femelles doivent attendre quelques heures avant de pouvoir s'accoupler (6 à 8 h). Une fois accouplées, les femelles ont la particularité de stocker le sperme des mâles dans leur spermathèque et peuvent





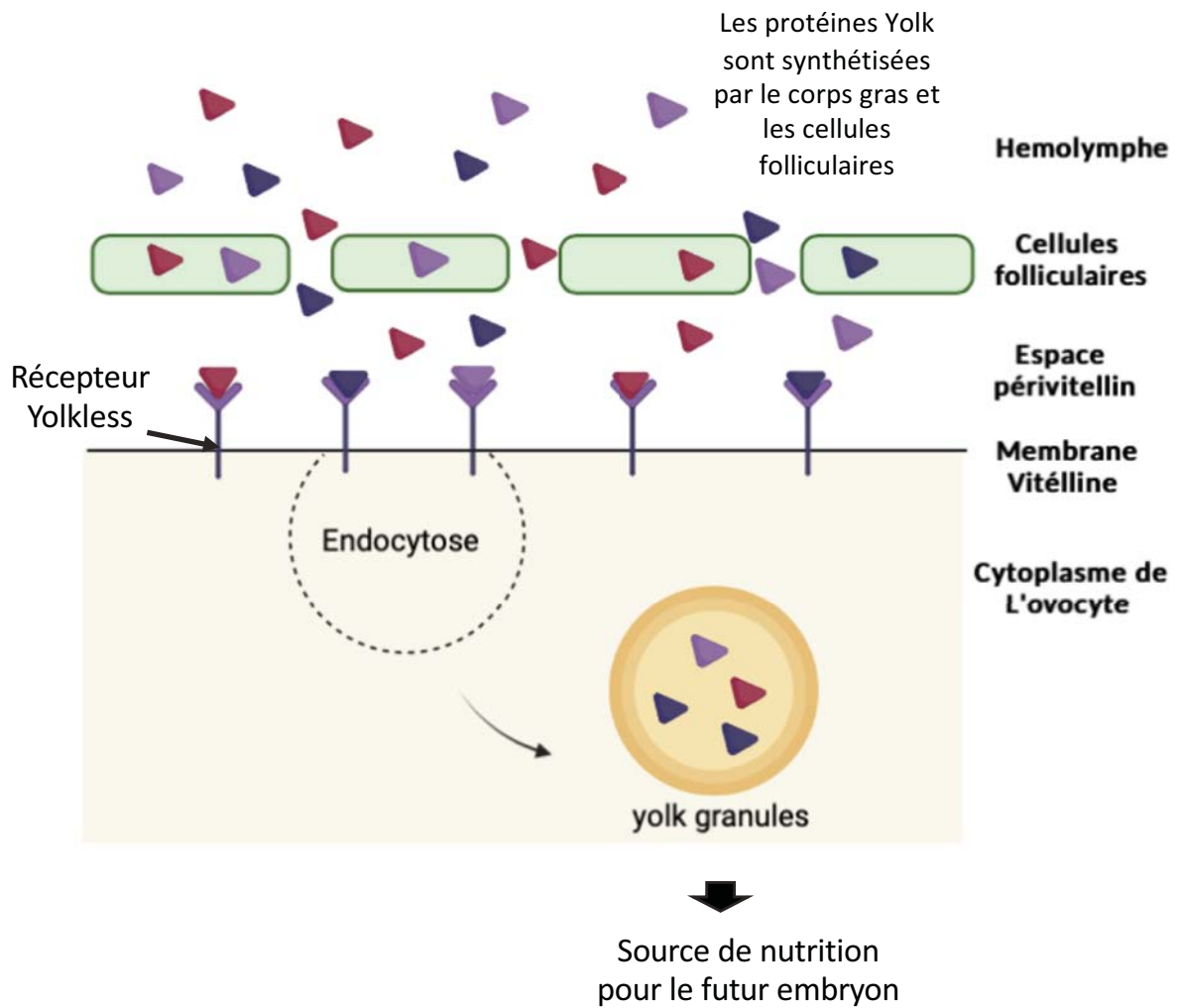
### Figure 18 : Tissus reproducteurs de la drosophile femelle

La drosophile femelle possède dans son abdomen une paire d'ovaire. Chaque ovaire est constitué d'une quinzaine d'ovarioles (A). Chaque ovariole est composée, selon un axe antéro-postérieur, d'une succession de follicules à des stades progressifs de maturation, du germarium qui contient les cellules souches germinales et folliculaires jusqu'à l'œuf mature. Chaque follicule est composé de seize cellules germinales entourées d'une monocouche de cellules somatiques folliculaires. Parmi les seize cellules germinales présentes au sein d'un follicule, quinze sont des cellules nourricières et une seule se différencie en ovocyte. Le temps de développement par rapport au début de l'ovogenèse est exprimé en heures après la division des cellules souches germinale (h.p.d. = heures post-germ line stem cell division).  
Figure adaptée de Navarro-Costa et al, 2016

ainsi utiliser celui-ci pour féconder leurs ovocytes durant toute leur vie. C'est la raison pour laquelle les généticiens doivent trier les femelles dites « vierges », c'est-à-dire âgées de moins de 6 h, pour débiter tout nouveau croisement avec des mâles afin de contrôler le génotype de la progéniture. Ainsi, les drosophilistes passent un temps assez incroyable à « trier des femelles vierges » et à « repiquer les mouches » (ce qui signifie mettre les adultes sur un nouveau milieu nutritif pour entretenir la lignée). Mais ces petits inconvénients ne sont rien en comparaison des avantages multiples apportés par l'utilisation de ces petites mouches !

### I.1. b) Ovogenèse de la *Drosophile*

Les éléments transposables peuvent s'exprimer dans de nombreux tissus. Néanmoins, dans le cadre de cette étude, nous étudions leur expression et leur répression dans le tissu reproducteur de la femelle. Celui-ci est composé de deux ovaires, reliés entre eux par un oviducte suivi d'un utérus et de ses annexes. Chaque drosophile femelle possède une paire d'ovaires renfermant chacun environ seize ovarioles (**Figure 18A**). Chaque ovariole est organisée en succession de follicules à des stades progressifs de maturation, selon l'axe antéro-postérieur, aboutissant à la formation d'un œuf. Le germarium, localisé dans la partie antérieure, contient les cellules souches germinales et les cellules souches somatiques folliculaires. Les deux cellules souches germinales (GCS) se divisent de façon asymétrique, aboutissant à la formation d'une cellule souche germinale fille qui restera dans la niche et d'un cystoblaste qui subira quatre divisions mitotiques successives pour former un cyste de seize cellules interconnectées. À la suite de ces divisions, une cellule parmi les seize formées sera sélectionnée pour former l'ovocyte alors que les quinze autres deviendront des cellules nourricières (**Figure 18B**). L'ovocyte est alors bloqué en prophase 1 de méiose. Les cellules nourricières, quant à elles, synthétisent un grand nombre d'ARNs messagers et de protéines qui seront déversés dans l'ovocyte durant les stades tardifs de l'ovogenèse. Le cyste contenant donc un ovocyte et quinze cellules nourricières. Il est encapsulé par une monocouche de cellules somatiques, appelées cellules folliculaires, pour former ce que l'on nomme un follicule. Le follicule se développe au travers de quatorze stades bien définis selon des critères morphologiques (Bastock and St Johnston, 2008). L'ovogenèse dure un peu plus de 3 jours et génère un ovocyte mature dans la partie la plus postérieure de l'ovaire où il est relargué dans l'oviducte et fécondé lors de son passage par le sperme stocké dans la spermathèque de la mouche.



**Figure 19 : Schéma de la vitellogenèse chez la drosophile**

Les protéines Yolk sont des protéines synthétisées par le corps gras et par les cellules somatiques folliculaires de l'ovaire. Après avoir été relarguées dans l'hémolymphe, les protéines Yolk se fixent sur le récepteur Yolkless présent sur la membrane vitelline qui sépare les cellules folliculaires de l'ovocyte. Cela induit l'endocytose des protéines Yolk qui s'associent à des lipides pour former des vésicules appelées « *yolk granule* ». Ces granules sont transmises à l'embryon et seront une source d'énergie pour son développement.

## **I.2. Échanges entre les cellules somatiques folliculaires et les cellules germinales au sein de l'ovaire de la drosophile**

Les cellules germinales sont entourées d'une monocouche de cellules folliculaires. Ces deux types cellulaires communiquent *via* de nombreux échanges moléculaires afin de maintenir l'homéostasie et la fonction de l'ovaire. Ces échanges permettent le transfert de molécules nutritives et de signaux développementaux notamment impliqués dans la détermination des axes antéro-postérieur et dorso-ventral ((Vachias et al., 2014) revue dans :(Antel and Inaba, 2020)). Par exemple, le morphogène Gurken est produit par l'ovocyte dans les stades précoces de l'ovogenèse, puis migre vers la partie postérieure de l'ovocyte avant de s'accumuler à l'interface entre l'ovocyte et les cellules folliculaires (Saunders and Cohen, 1999). En se fixant sur un récepteur présent à la membrane des cellules folliculaires, Gurken active la voie EGFR dans les cellules folliculaires (González-Reyes et al., 1995; Tian et al., 2013). Cela aboutit à une spécialisation des cellules folliculaires qui, en retour, envoient un signal pour réorganiser le cytosquelette de l'ovocyte afin qu'il définisse au final les axes embryonnaires (Revue dans : (Nilson and Schüpbach, 1999)).

Par ailleurs, le noyau de l'ovocyte devient rapidement quiescent et la transcription de son génome ne se réactive qu'après fécondation. De ce fait, la croissance de l'ovocyte dépend de la traduction d'ARNm qui ont produits et stockés dans les cellules nourricières et transférés vers l'ovocyte, ainsi que du matériel provenant des cellules voisines. Par exemple, les protéines *Yolk* sont des protéines qui s'accumulent dans l'ovocyte et serviront de réserve de nutriments pour le futur embryon. Les protéines *Yolk* sont produites en partie par les cellules folliculaires adjacentes et par le corps gras. Elles sont transférées par endocytose à l'ovocyte à partir du stade 8 de l'ovogenèse, ce processus est appelé la vitellogenèse (**Figure 19**).

Il est également intéressant de noter que les cellules somatiques et germinales sont connectées par des jonctions gap qui permettent l'échange de molécules telles que des ions, des sucres, des acides aminés... (Bohrmann and Zimmermann, 2008). Certaines études suggèrent même un rôle de ces jonctions communicantes dans le transfert de petits ARNs (Revue dans : (Dinger et al., 2008; Brink et al., 2012; Zong et al., 2016)). Cependant, cette hypothèse reste encore inexplorée au sein des ovaires de drosophiles. De nombreuses vésicules extracellulaires, telles que les exosomes, permettent également l'échange de composés entre ces deux types cellulaires. Les exosomes contiennent des molécules de taille plus importante comme des cytokines, des facteurs de croissance, des lipides, et même des acides nucléiques tels que des ARNs (Valadi et al., 2007; Raposo and Stoorvogel, 2013; Tkach and Théry, 2016). Des ARNm



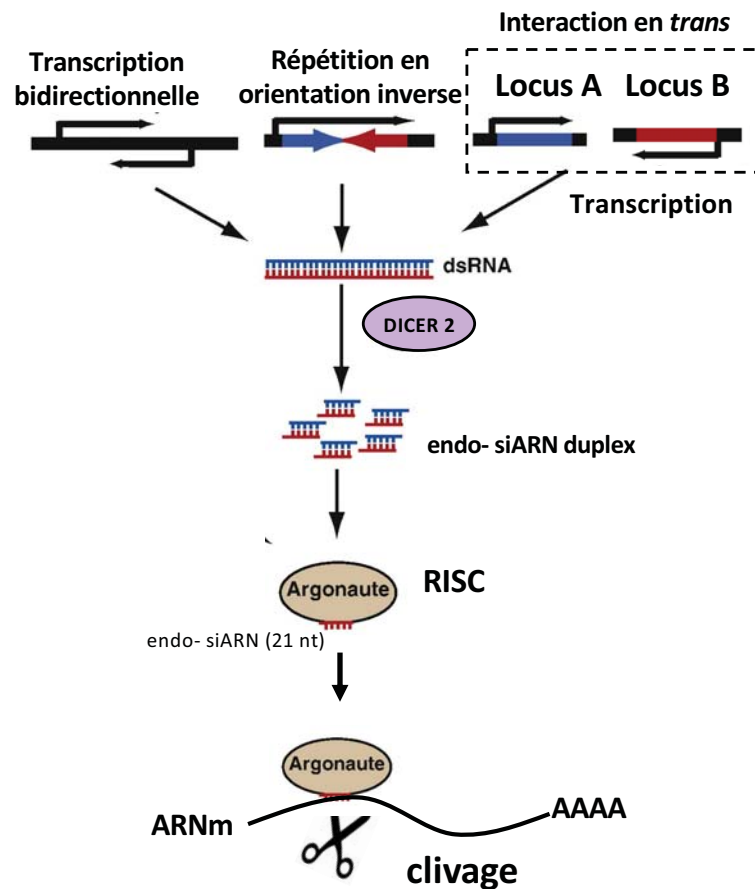
d'éléments transposables ont même été détectés dans des vésicules comme les exosomes, suggérant un rôle de ces vésicules dans le transfert intercellulaire des ETs (Balaj et al., 2011; Kawamura et al., 2019). Par exemple, chez l'humain, des ARNs du rétrotransposon *LINE1* ont été détectés dans des exosomes isolés à partir de cellules possédant des copies actives de *LINE1* dans leur génome (Kawamura et al., 2019). Les exosomes peuvent ainsi délivrer l'ARN de *LINE1* à des cellules cibles. Cette étude révèle le rôle potentiel des exosomes dans le transfert d'ARN de rétrotransposon entre des cellules proches ou éloignées.

Comme décrit dans la revue ci-dessus (Yoth et al., 2022), de nombreux éléments transposables sont exprimés dans les ovaires de la drosophile mais présentent un profil d'expression cellule-spécifique, voire parfois même stade-spécifique. En effet, l'expression des ETs est contrôlée, comme celle des gènes, par la présence de facteurs de transcription spécifiques (Cavarec et al., 1997; Meignin et al., 2004). La diversité de profil d'expression des ETs suggère que ces derniers sont probablement régulés par un panel large et varié de facteurs de transcription. Cependant, les ETs peuvent utiliser les différents mécanismes d'échange intercellulaire existant au sein de l'ovaire pour potentiellement transiter et transposer dans une cellule où ils ne sont pas exprimés (cf revue ci-dessus : (Yoth et al., 2022)). C'est notamment le cas de certains rétrovirus endogènes, exprimés dans les cellules folliculaires de l'ovaire et transmis aux cellules germinales, pouvant ainsi transposer dans le génome de ces dernières (Kim et al., 1994; Brasset et al., 2006). Ainsi, un contrôle efficace des éléments transposables dans l'ensemble des cellules de l'ovaire est nécessaire pour éviter toute transposition dans les cellules germinales, particulièrement dans l'ovocyte.

## **II- Régulation des éléments transposables dans les ovaires de la drosophile**

Face au potentiel mutagène des ETs, les organismes ont développé des mécanismes sophistiqués pour contrôler leur expression et leur transposition. Ce contrôle est d'autant plus important dans les cellules germinales, seules cellules de l'organisme à transférer le matériel génétique et épigénétique à la descendance. Au sein des génomes eucaryotes, les ETs peuvent être contrôlés essentiellement à deux niveaux : au niveau transcriptionnel et au niveau post-transcriptionnel. La répression transcriptionnelle consiste à mettre en place une structure hétérochromatique au niveau des promoteurs des ETs pour inhiber leur expression. La répression post-transcriptionnelle permet quant à elle de dégrader les transcrits d'ETs qui auraient échappé au contrôle transcriptionnel. Ces deux mécanismes utilisent des mécanismes

## Voie des endo-siARNs



**Figure 20 : Voie des siARNs chez *Drosophila melanogaster***

La production des siARNs s’amorce par la formation d’un ARN double brin reconnu dans le cytoplasme par la protéine DICER2 afin de cliver l’ARN double brins en duplex de 21 nts. Ces duplex sont ensuite chargés par une protéine de la famille Argonaute, ici la protéine AGO2. Après sélection d’un des deux brins, le complexe RISC reconnaît par complémentarité de séquence les transcrits à cliver.

Figure adaptée de Cecere and Grishok, 2014

d'ARN interférence faisant intervenir des petits ARNs non codants et des protéines de la famille ARGONAUTE. Il existe trois catégories principales de petits ARNs chez les animaux : i) les miARNs (microARN), d'une longueur de 22 nucléotides (nts) et qui sont pris en charge par la protéine AGO1 ; ii) les siARNs (small interfering RNAs), 21 nts de long et interagissant avec la protéine AGO2 ; iii) les piARNs (PIWI-interacting RNAs), dont la taille varie entre 23 et 29 nts, qui sont chargés par les protéines appartenant à la sous-famille PIWI des protéines ARGONAUTES. Les siARNs et les piARNs jouent un rôle bien établi dans le contrôle des ETs. Les miARNs n'ont pas de rôle clairement établi dans la répression des ETs, même s'ils pourraient être impliqués dans la production et/ou la stabilité des piARNs (Mugat et al., 2015).

## II.1 - Rôle des siARNs dans la répression des ETs

La voie des siARNs a initialement été décrite comme étant une voie de défense contre l'invasion de matériels génétiques extérieurs, notamment lors des infections virales (revue dans : (Kobayashi and Tomari, 2016)). Les siARNs sont souvent produits à partir de séquences étrangères, dégradées en petits ARNs, puis chargés sur la protéine AGO2. En 2008, le séquençage des petits ARNs de différents organes de *Drosophila melanogaster* a révélé que les siARNs pouvaient également être codés par le génome : on les a alors appelés les endo-siARNs (car provenant de régions endogènes) (Czech et al., 2008; Ghildiyal et al., 2008; Okamura and Lai, 2008). Les endo-siARNs sont produits à partir d'ARNs précurseurs double brins issus de la transcription génomique. Les précurseurs sont ensuite clivés par la protéine DICER2 pour aboutir à la formation de petits fragments ARN double brins de 21 nts. Ces petits ARNs double brin sont ensuite pris en charge par la protéine AGO2 avec un seul des deux brins qui reste lié à la protéine AGO2 et l'autre brin qui est dégradé. La liaison petit ARN-AGO2 forme ce que l'on appelle un complexe RISC (*RNA-Induced Silencing Complex*). Ce complexe cible alors des ARNs cytoplasmiques par complémentarité de séquence et induit leur dégradation (revue dans : (Ghildiyal and Zamore, 2009; Ding, 2010; Cecere and Grishok, 2014; Kobayashi and Tomari, 2016)) (**Figure 20**).

Il a également été montré que des endo-siARNs dérivés de séquences d'ETs sont produits chez *Drosophila melanogaster* (Chung et al., 2008; Czech et al., 2008; Okamura et al., 2008). Ils ont également été retrouvés chez la souris et le cochon notamment (Okamura and Lai, 2008; Kong et al., 2019). L'hypothèse que la fonction naturelle de l'ARN interférence soit la répression des ETs a d'abord été émise en 1999 (Tabara et al., 1999). Le rôle des endo-siARNs sur la répression des ETs a ensuite été confirmé. En effet, chez *Drosophila melanogaster*, une mutation de la





protéine DICER2 ou AGO2 engendre une augmentation de l'expression des ETs dans les cellules somatiques de la drosophile (Chung et al., 2008; Ghildiyal et al., 2008; Kawamura et al., 2008; Okamura and Lai, 2008). En ce qui concerne leur rôle dans les tissus reproducteurs, les siARNs semblent réprimer les ETs dans les ovocytes de souris pour éviter une dérégulation massive lors de la reprogrammation du génome de l'ovocyte (Tam et al., 2008; Watanabe et al., 2008; Ohnishi et al., 2010). La mutation de la voie des siARNs induit également une stérilité des souris. Chez la drosophile, l'altération de cette voie au sein de la lignée germinale entraîne une légère dérégulation d'un petit nombre d'ETs, sans induction de stérilité (Rehwinkel et al., 2006; Kawamura et al., 2008). La voie des siARNs ne semble donc pas être la voie majoritaire permettant la répression des ETs dans les cellules germinales de la drosophile. Cependant, les siARNs pourraient coopérer avec une autre voie essentielle à la répression des ETs dans les gonades : la voie des piARNs (Olovnikov and Kalmykova, 2013; Shpiz et al., 2014; Luo et al., 2022).

## II.2 - Découverte du rôle des piARNs dans le contrôle des ETs

### II.2. a) Découverte des piARNs

La découverte des piARNs est assez récente. En 2001, des petits ARNs double brins de 25 à 27 nts dérivés des deux brins d'ADN du locus « *Suppresseur de Stellate* », localisé sur le chromosome Y de *Drosophila melanogaster*, sont découverts (Aravin et al., 2001). Ces petits ARNs permettent d'inhiber l'expression du gène *Stellate* chez le mâle: la délétion du locus *Suppresseur de Stellate* induit une perte des petits ARNs double brins produits à partir de ce locus et une forte expression de la protéine *Stellate*. L'expression de cette protéine au sein des testicules de drosophile engendre alors la formation d'agrégats qui induisent une stérilité des mâles. Ces petits ARNs sont dans un premier temps appelés rasiARNs (*Repeat-associated small interfering RNAs*) car ils permettent de réprimer des séquences répétées du génome telles que *Stellate*. Quelques années plus tard, les premières immunoprécipitations de protéines associées à des petits ARNs et l'évolution du séquençage ont permis de faire une découverte qui va accélérer de manière extraordinaire les connaissances sur les voies d'interférences ARN. Des protéines de la sous-famille PIWI, exprimées dans les testicules de souris (MILI et MIWI), semblaient lier spécifiquement des petits ARNs de 23 à 29 nts (Aravin et al., 2006; Girard et al., 2006; Grivna et al., 2006; Lau et al., 2006; Vagin et al., 2006; Watanabe et al., 2006). Ces petits ARNs avaient une taille clairement distincte de celle des miARNs et des siARNs,



suggérant que ces protéines lient un type différent de petits ARNs. Leur taille est en revanche similaire à celle observée pour les petits ARNs dérivés du locus « *Suppresseur de Stellate* », supposant qu'ils appartiendraient à la même catégorie. Ces petits ARNs, initialement baptisés rasiARNs, sont alors renommés piARNs pour « *PIWI-interacting RNA* ». Les avancées en matière de NGS (*Next Generation Sequencing*) ont permis de révéler que les piARNs sont présents chez de nombreux animaux, mais semblent exprimés uniquement (ou principalement) dans les gonades. Les piARNs ont même été retrouvés dans des espèces très éloignées phylogénétiquement comme les mollusques (Jehn et al., 2018), les planaires (vers plats) (Lakshmanan et al., 2021), les arthropodes (Lewis et al., 2018), les reptiles (Sun et al., 2020) ou encore chez les mammifères (Roovers et al., 2015). De manière surprenante, la taille et la composition/séquence de ces piARNs s'avèrent très variables selon les espèces. Néanmoins, ils présentent des caractéristiques communes : ils possèdent à leur extrémité 3' un groupement 2'-O-méthyle et commencent pour la plupart par une Uridine monophosphatée à l'extrémité 5'.

## II.2. b) Les piARNs : un rôle dans la régulation des éléments transposables ?

Il existe donc des petits ARNs de 23 à 29 nts liés à des protéines de la famille PIWI et exprimés essentiellement uniquement dans les gonades des animaux. Jusqu'en 2007, leur fonction restait hypothétique. Toutefois, plusieurs indices obtenus dans les années précédentes laissent à penser que ces piARNs pourraient avoir un rôle dans la répression des éléments transposables :

- Le séquençage des premiers piARNs (autrefois appelés rasiARNs) chez la drosophile et le Poisson-zèbre révèle la présence de piARNs ayant des séquences enrichies en éléments transposables (Aravin et al., 2001, 2003; Chen et al., 2005).
- Les mutations des gènes codants les protéines de la famille PIWI, connues pour interagir avec les piARNs, entraînent une dérégulation des éléments transposables dans les ovaires de *Drosophila melanogaster*. Par exemple, la mutation du gène codant la protéine Piwi entraîne une dérégulation de *Gypsy* (Sarot et al., 2004), et la mutation du gène codant Aubergine entraîne une dérégulation des éléments transposables *TART* (Savitsky et al., 2006) et de l'élément *P* (Reiss et al., 2004). Toutes ces études suggéraient déjà un rôle probable de l'interférence ARN dans le contrôle des éléments transposables. De plus, l'expression spécifique des protéines de la famille Piwi dans les

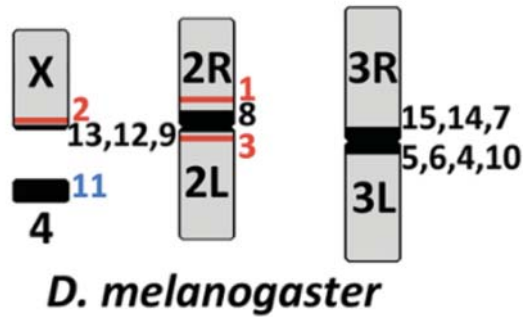


cellules germinales est un phénomène conservé parmi des espèces même très éloignées phylogénétiquement, suggérant une fonction essentielle de ces protéines dans les tissus reproducteurs (Cox et al., 1998; Kuramochi-Miyagawa et al., 2001; Reddien et al., 2005).

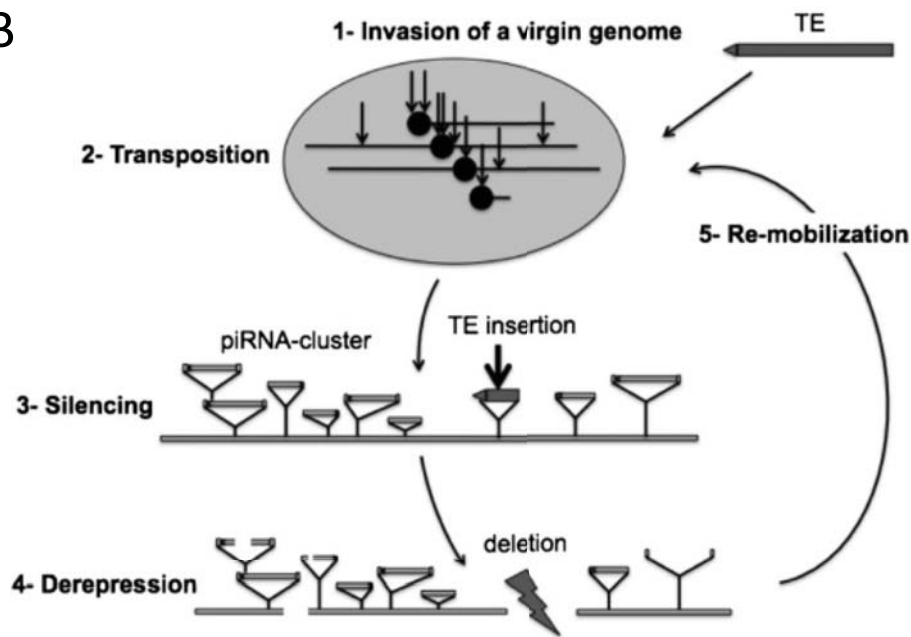
- Le contrôle des ETs avait également été corrélé à la présence de loci génomiques particuliers. Alain Bucheton et Alain Péliesson furent les premiers à montrer, dès les années 1990, que le contrôle du rétrotransposon *Gypsy* dans les ovaires de *D. melanogaster* était médié par une région péricentromérique localisée sur le chromosome X, nommée *flamenco* (Péliesson et al., 1994a; Bucheton, 1995). Ce locus a ensuite été identifié comme étant une région de contrôle de deux autres ETs : *ZAM* et *Idefix* (Desset et al., 2003). De la même façon, l'insertion de l'élément *P* à un locus particulier serait suffisante pour permettre son contrôle (Ronsseray et al., 1989, 1991; Biémont et al., 1990). Cette corrélation directe entre la présence d'un « locus régulateur » et l'expression d'un élément transposable est fortement ressemblante à la situation observée avec le *Suppresseur de Stellate* qui réprime l'expression de *Stellate* dans les testicules de la drosophile grâce à la production de rasiARNs.
- Enfin, les études faites sur la dysgénésie hybride (voir revue ci-dessus (Yoth et al., 2022)) ont révélé que l'immunité contre un ET pouvait être transmise à la génération suivante par les cellules germinales de la femelle (Kidwell et al., 1977; Bregliano et al., 1980; Rubin et al., 1982). Ceci suggère que la protection contre l'expression des ETs comporte à la fois des composants cytoplasmiques et génétiques. Les petits ARNs pourraient être ces composants transmis par la femelle à sa progéniture.

Au final, toutes ces informations suggèrent que : i) les ETs seraient régulés par des régions spécifiques du génome, et que ii) la voie d'interférence ARN faisant intervenir les protéines de la famille PIWI pourrait être impliquée dans ce contrôle. Mais ces deux mécanismes sont-ils liés ? Quel est le lien entre « région de contrôle » et « voie d'interférence ARN » ? Comment ces piARNs sont-ils produits et sont-ils réellement impliqués dans le contrôle de tous les ETs au sein des tissus reproducteurs ? Tant de questions mystérieuses qui trouveront réponse à la suite de trois études parues en 2007 qui vont bousculer le domaine de recherche des ETs (Aravin et al., 2007; Brennecke et al., 2007; Gunawardane et al., 2007).

A



B



**Figure 21 : Rôle des clusters de piARNs dans la répression des ETs**

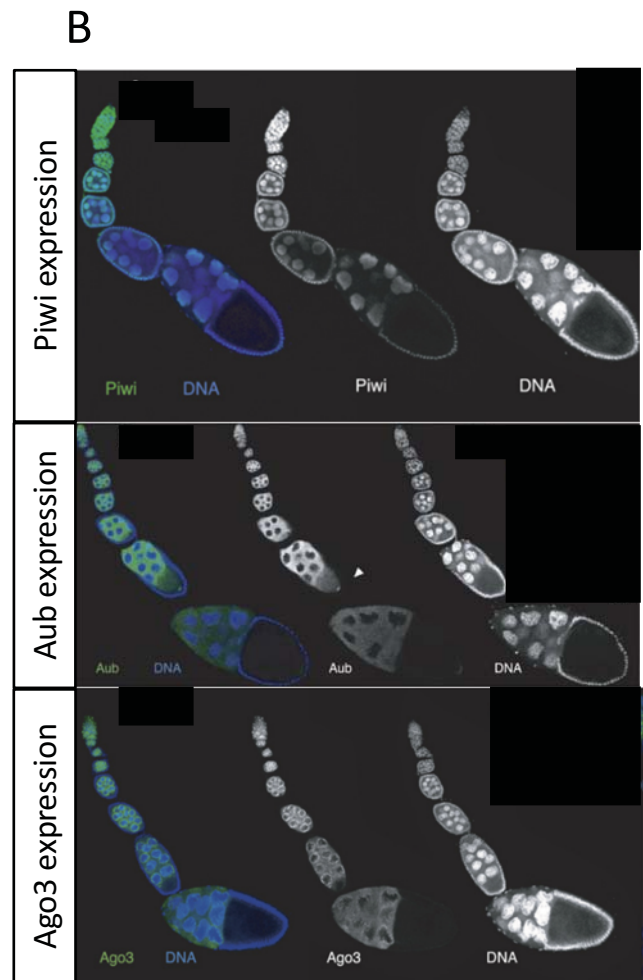
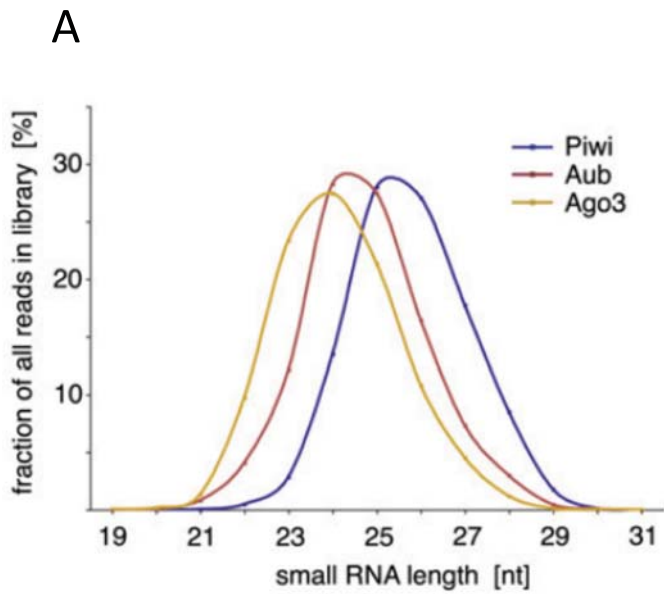
- A) Le génome de la drosophile est constitué de 13 000 gènes répartis sur quatre paires de chromosomes : la paire de chromosomes sexuels X/Y et trois paires de chromosomes autosomes (2, 3 et 4). Le chromosome 4 est très petit et porte très peu de gènes. Les clusters de piARNs présents dans le génome de *D.melanogaster* sont numérotés à droite des bras chromosomiques, les couleurs (gris euchromatique, noir hétérochromatine péricentrique, rouge hétérochromatine intercalaire) indiquent le type de chromatine. Figure issue de George et al, 2015.
- B) Lors d'une invasion génique par un nouvel ET se produit (1), L'ET transpose à de multiples loci (2). L'insertion dans un cluster de piARN induit la biogénèse des piARNs correspondants et conduit à la répression de l'ET (3). Plusieurs événements mutationnels suppriment des portions du cluster de piARN et, finalement, l'ET lui-même (4). L'élimination de l'ET du cluster abolit la production des piARNs correspondants et conduit à la remobilisation de l'ET (5). Figure issue de Zanni et al, 2013.

## II.2 c) Identification de la source de production des piARNs

C'est ainsi que les premières études identifiant l'origine génomique des piARNs sont nées. Deux scientifiques ont élaboré une stratégie astucieuse pour identifier de manière certaine les loci génomiques producteurs de piARNs : aligner sur le génome uniquement les piARNs qui ne s'alignent qu'à un seul endroit sur le génome (appelés piARNs uniques). En effet, lorsque tous les piARNs sont alignés sur le génome, et que certains proviennent de séquences répétées, ces derniers peuvent s'aligner à de nombreux loci, rendant impossible la caractérisation du locus initialement producteur. C'est donc par l'alignement de ces piARNs uniques sur le génome de *Drosophila melanogaster* et de la souris que les régions productrices de piARNs sont identifiées, elles sont alors appelées clusters de piARNs (Aravin et al., 2007; Brennecke et al., 2007). La taille de ces clusters peut varier de quelques kilobases à plusieurs centaines de kilobases et ils sont répartis sur les différents chromosomes. Chez la drosophile, 142 clusters producteurs de piARNs, principalement localisés dans l'hétérochromatine péri-centromérique et télomérique, ont été identifiés dans un premier temps (**Figure 21A**) et, depuis, les progrès de séquençage et d'annotations génomiques ont révélé qu'il en existe probablement bien plus (George et al., 2015). Ces régions génomiques présentent des caractéristiques très spécifiques : elles sont dépourvues de gènes et sont au contraire fortement enrichies en séquences répétées et en éléments transposables ! Une grande partie des ETs insérés dans les clusters de piARNs sont des copies tronquées ou fortement mutées qui ne sont pour la plupart plus capables de transposer de manière autonome. Parmi ces régions productrices de piARNs, on retrouve notamment le locus *flamenco*, qui avait préalablement été identifié comme région de contrôle d'au moins trois ETs : *Gypsy*, *ZAM* et *Idefix* (Péligsson et al., 1994a; Bucheton, 1995; Desset et al., 2003). Toutes ces nouvelles informations, cumulées avec les données obtenues lors des 30 années passées, permettent de dresser un tableau « complet » sur les clusters de piARNs (même si rien n'est jamais vraiment complet en science...) !

L'insertion des ETs dans les clusters de piARNs permet de produire des petits ARNs complémentaires qui peuvent réprimer les copies actives d'ETs insérés ailleurs dans le génome par un mécanisme d'interférence ARN. Ainsi, le mécanisme de répression par les clusters de piARNs est très semblable aux mécanismes de réponse immunitaire adaptative : lorsqu'un nouvel élément transposable arrive dans un génome, son intégration dans un cluster de piARNs permet une répression en *trans* des autres copies insérées dans le génome. De cette manière, le génome garde une mémoire de tous les éléments qui doivent être réprimés dans la cellule, et les





**Figure 22 : Caractéristiques des protéines de la famille PIWI chez *Drosophila melanogaster***

A) Taille des piARNs en nucléotides (nts) liés par les différentes protéines de la famille PIWI

B) La protéine Piwi est présente dans le noyau des cellules germinales et somatiques de l'ovaire. Les protéines Aub et Ago3 sont retrouvées uniquement dans le cytoplasme des cellules germinales avec une accumulation périnucléaire (appelé « nuage »).

Figure issue de Brennecke et al, 2007

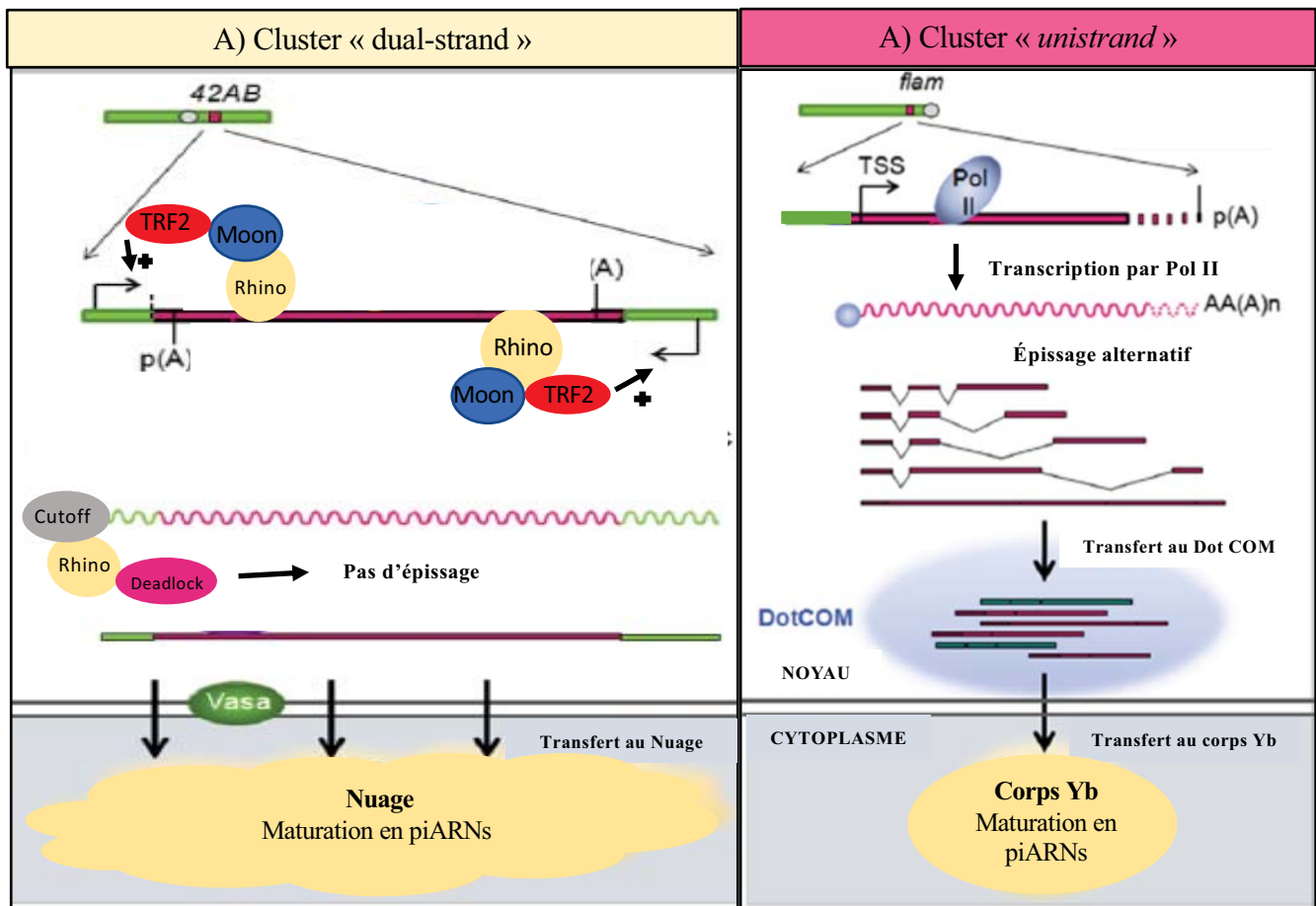
nouvelles insertions d'ETs dans les clusters de piARNs permettent de faire évoluer constamment cette mémoire génomique (**Figure 21B**). Les éléments transposables, en s'insérant à un locus génomique particulier, participent donc à leur propre régulation : un système évolutif et, nous allons le voir dans la partie suivante, très efficace !

### II.3 - La voie des piARNs : de la biogenèse des piARNs à la répression des ETs

La voie des piARNs a par la suite été étudiée en détail, et beaucoup de connaissances ont été acquises grâce aux études réalisées sur les ovaires de *D. melanogaster*. L'utilisation de ce modèle a permis d'effectuer des cribles génétiques de grande envergure, conduisant à la découverte de nombreux acteurs impliqués dans la régulation des ETs et dans la production des piARNs. Je résumerai donc dans la partie suivante les données obtenues sur la voie des piARNs dans les tissus reproducteurs de la drosophile.

#### II.3 a) Les protéines de la famille PIWI

Le génome de la drosophile code trois protéines de la sous-famille PIWI sont codées par le génome de la drosophile : Piwi, Argonaute 3 (Ago3) et Aubergine (Aub). Elles sont caractérisées par la présence de quatre domaines : le domaine N-terminal, le domaine PAZ, le domaine Mid et le domaine PIWI. Les domaines PAZ et Mid leur permettent de fixer la partie 3' des piARNs. Le domaine PIWI possède quant à lui une activité endonucléasique qui induit le clivage de l'ARN complémentaire cible (Lingel et al., 2003; Yan et al., 2003; Song et al., 2004). Ainsi, lorsque les piARNs sont pris en charge par ces protéines, le complexe va reconnaître par complémentarité de séquence un transcrite et induire son clivage ; on parle alors de régulation post-transcriptionnelle. Les analyses d'immunoprécipitation et de séquençage des petits ARNs associés aux protéines de la famille PIWI ont révélé que les populations de petits ARNs liés par chacune des protéines sont différentes. Tout d'abord, ces piARNs présentent une taille différente : les piARNs liés par Piwi font un pic autour de 26 nts, ceux liés par Aub autour de 25 nts et ceux liés par Ago3 sont davantage des ARNs autour de 24 nts (**Figure 22A**). Les piARNs liés par Piwi et Aub possèdent principalement une Uridine à la position 5' alors que cet enrichissement préférentiel est quasiment absent pour les piARNs liés à Ago3. Cela suggère un rôle potentiellement différent de chacune de ces protéines dans la répression des éléments transposables.



**Figure 23 : Description de la transcription des clusters de piARNs dans les cellules germinales et folliculaires de l’ovaire**

A) Dans les cellules germinales, la transcription des clusters de piARNs « dual-strand » nécessite l’intervention des protéines Rhino, Moonshiner (Moon) et TRF2. L’épissage des transcrits est inhibé par l’association de ceux-ci avec les protéines du complexe RDC (Rhino, Deadlock, Cutoff). Les transcrits précurseurs sont pris en charge par la protéine Vasa et sont directement adressés au cytoplasme pour être maturés dans le nuage.

B) Dans les cellules folliculaires, les clusters de piARNs sont transcrits par l’ARN polymérase II. Les transcrits sont polyadénylés, coiffés puis épissés. Ils s’accumulent ensuite à la périphérie du noyau au niveau d’une structure nommée Dot COM avant d’être transférés vers le cytoplasme pour être maturés en piARNs au sein d’un foyer multi-protéique appelé le corps Yb.

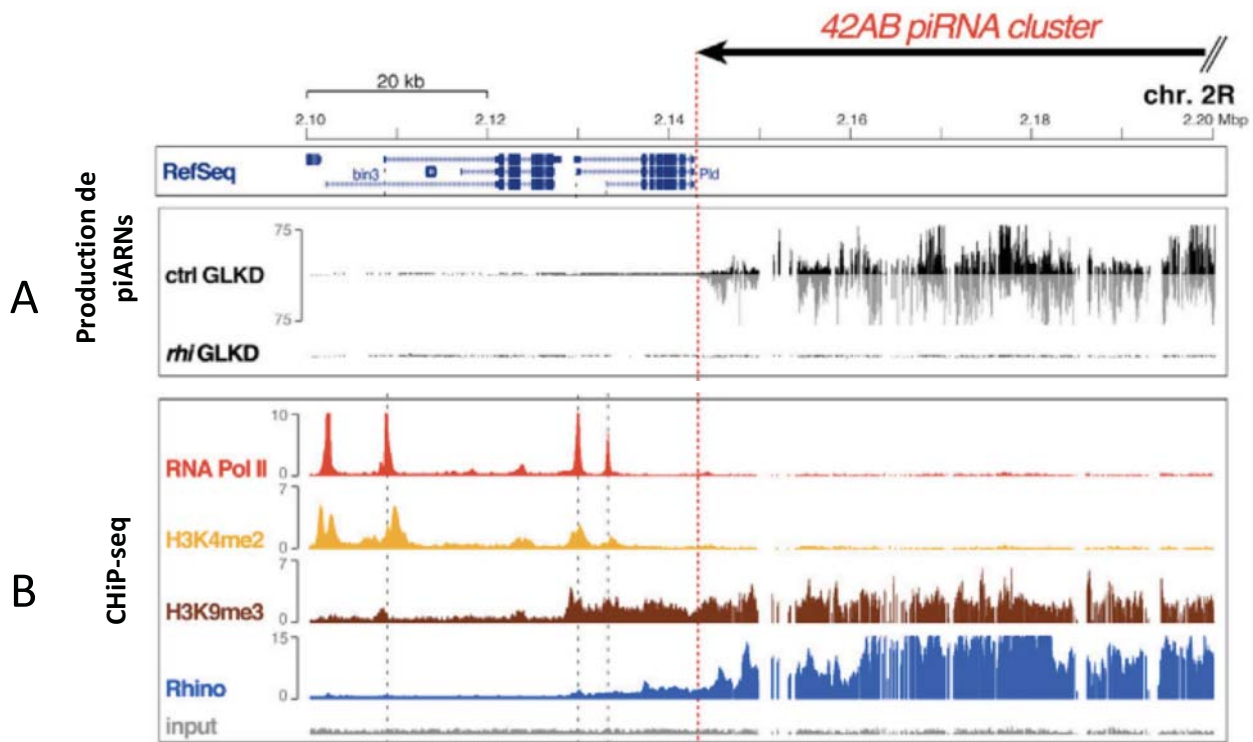
Figure adaptée de Théron et al., 2014

Des immunomarquages ont été réalisés afin d'analyser l'expression de ces protéines dans les ovaires de drosophile. Rappelons que ces ovaires sont composés de deux types cellulaires : les cellules germinales, comprenant l'ovocyte, et les cellules folliculaires - d'origine somatique - entourant les cellules germinales. Les trois protéines de la famille PIWI sont exprimées dans les cellules germinales. Cependant, Piwi se localise exclusivement dans le noyau, contrairement aux protéines Aub et Ago3 qui sont présentes uniquement dans le cytoplasme des cellules germinales et s'accumulent au niveau périnucléaire dans une structure appelée le nuage (Brennecke et al., 2007; Gunawardane et al., 2007). Dans les cellules folliculaires de l'ovaire, la protéine Piwi est aussi présente dans le noyau. Cependant, Aub et Ago3 n'y sont pas exprimées (**Figure 22B**). La présence d'une seule des trois protéines de la famille PIWI dans les cellules folliculaires suggère que la voie des piARNs est active mais que celle-ci est probablement différente de la voie des piARNs présente dans les cellules germinales.

### II.3. b) Les clusters de piARNs

Les études de Julius Brennecke et Alexei A. Aravin publiées en 2007 ont par conséquent révélé que les piARNs sont issus de la transcription des clusters de piARNs (Aravin et al., 2007; Brennecke et al., 2007). Chez *Drosophila melanogaster*, ces clusters représentent 3,5 % du génome mais produisent 92 % des piARNs totaux (Brennecke et al., 2007). L'alignement des piARNs sur ces clusters a montré que certains clusters sont transcrits à partir d'un seul des deux brins de l'ADN, on parle de clusters « *unistrand* » ; pour d'autres, les deux brins sont transcrits de façon convergente, ce sont les clusters de piARNs « *dual-strand* ». (Brennecke et al., 2007; Malone et al., 2009; Senti and Brennecke, 2010). Les clusters de piARNs « *unistrand* » sont majoritairement exprimés dans les cellules folliculaires de l'ovaire, même s'il existe des clusters « *unistrand* » tel que le cluster *20A* également exprimé dans les cellules germinales. Les clusters « *dual-strand* » sont quant à eux exprimés exclusivement dans les cellules germinales. La transcription de ces deux catégories de clusters, ainsi que les protéines impliquées dans leur transcription respective, sont différentes.

De manière générale, la transcription des clusters de piARNs soulève une question majeure : comment ces clusters, localisés dans des régions hétérochromatiques hypercondensées, peuvent être transcrits ? Comment échappent-ils à la répression globale présente dans ces régions ? Et comment l'expression tissu spécifique de ces clusters est-elle médiée ?



**Figure 24 : Caractéristiques transcriptionnelles du cluster de piARNs «dual-strand» 42AB exprimé dans les cellules germinales des ovaires de *D. melanogaster***

A) L'alignement des piARNs uniques sur la séquence du 42AB montre que les piARNs sont produits à partir des deux brins de l'ADN (noir : plus, gris : moins) en condition contrôle (ctrl GLKD). L'absence de Rhino dans les cellules germinales (Rhino GLKD) induit la perte de production des piARNs par le cluster 42AB.

B) Des expériences de CHIP-seq montrent l'enrichissement de la polymérase II (Pol II) sur le cluster 42AB (en rouge). L'ARN polymérase II n'est pas spécifiquement enrichie sur un promoteur bien défini au niveau du cluster 42AB. Les CHIP-seq réalisés avec des anticorps ciblant les marques histones H3K4me2 et H3K9me3 (en jaune et en marron respectivement) montrent une forte accumulation de la marque répressive H3K9me3 sur le cluster 42AB. La présence de H3K9me3 est corrélée avec la présence de la protéine Rhino (en bleu).

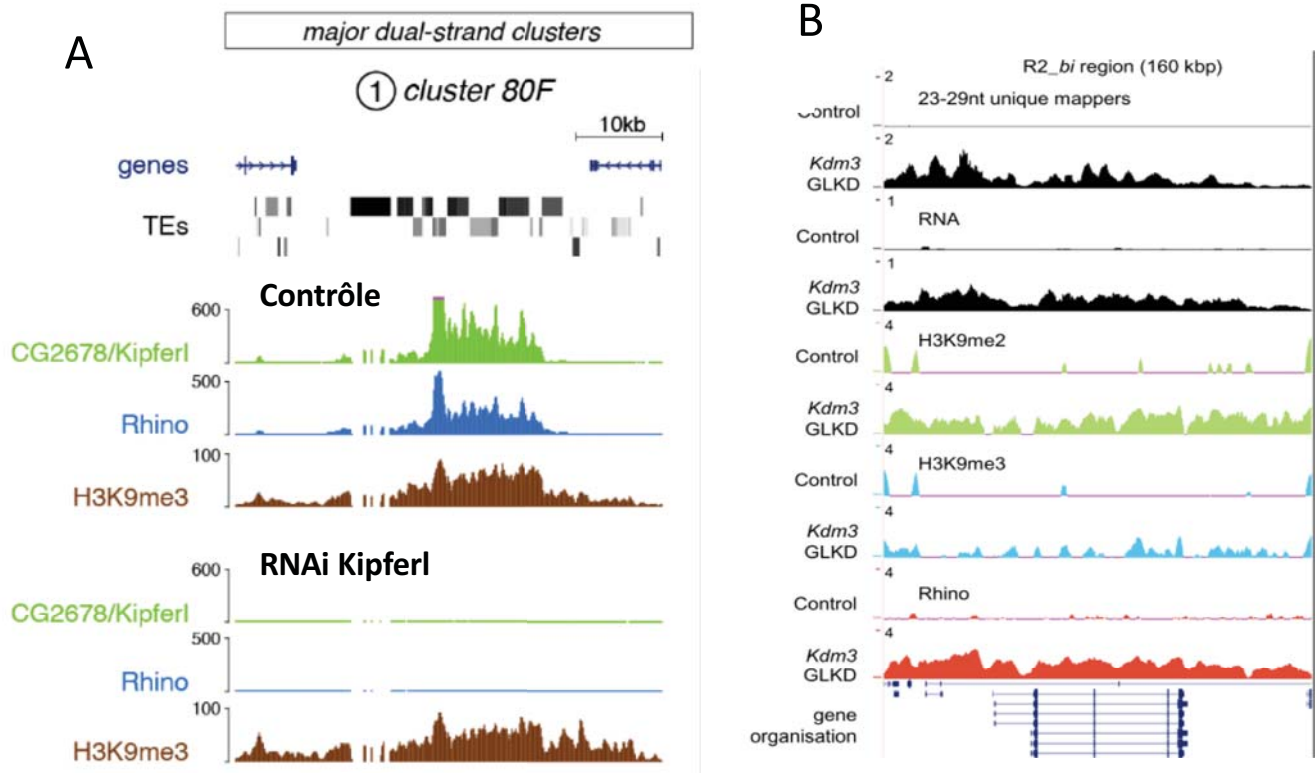
Figure adaptée Mohn et al., 2014

- **Les clusters « dual-strand »**

Les clusters « dual-strand » sont exprimés exclusivement dans les cellules germinales des ovaires de drosophiles. Les producteurs majeurs sont les clusters de piARNs *42AB*, *38C* et *80F* (Brennecke et al., 2007). Les clusters « dual-strand » sont composés d'une multitude d'insertions d'ETs orientées de manière aléatoire. La transcription de ces clusters s'effectue à partir des deux brins d'ADN, permettant de produire des piARNs sens et antisens par rapport aux séquences d'ETs présente dans les clusters (**Figure 23A**). La transcription de ces clusters a été principalement décrite en étudiant le cluster de piARNs *42AB*, producteur majoritaire de piARNs dans les cellules germinales (le cluster *42AB* produit effectivement 30 % des piARNs totaux retrouvés dans les ovaires).

La transcription des clusters « dual-strand » n'est pas canonique : on parle de transcription « non canonique » car non dépendante de promoteurs standards comme c'est le cas pour les gènes (Klattenhoff et al., 2009; Zhang et al., 2012; Le Thomas et al., 2014; Mohn et al., 2014). En effet, des sites d'initiation de transcription ont été observés tout le long de ces clusters. La transcription des clusters de piARNs « dual-strand » est permise grâce à la fixation d'une protéine clef tout le long de ces clusters : la protéine Rhino, exprimée spécifiquement dans les cellules germinales de l'ovaire. Rhino, paralogue de la protéine HP1 (*Heterochromatin Protein 1*), se lie au niveau des clusters de piARNs grâce à son chromodomaine qui reconnaît les marques chromatinienne Histone H3-Lysine-9-tri-méthyl (H3K9me3). Alors que le recrutement de HP1 induit la répression des gènes et la compaction de la chromatine chez les mammifères, le recrutement de Rhino chez la drosophile a un effet opposé. Rhino recrute plusieurs protéines nécessaires à l'expression des clusters de piARNs et à l'export des transcrits précurseurs (Zhang et al., 2012, 2014, 2018; Mohn et al., 2014; Hur et al., 2016; Andersen et al., 2017; ElMaghraby et al., 2019; Kneuss et al., 2019). En absence de Rhino, les clusters de piARNs « dual-strand » ne sont plus transcrits et sont convertis en hétérochromatine canonique, induisant une perte de la production des piARNs (**Figure 24A**). Ainsi, l'expression des clusters de piARNs « dual-strand » requiert préalablement un dépôt des marques H3K9me3 sur ces clusters, puis une reconnaissance spécifique de ces marques par Rhino (**Figure 24B**). Les marques H3K9me3 sont déposées sur les clusters « dual-strand » durant l'embryogenèse *via* les complexes piARN/Piwi hérités maternellement (Akkouche et al., 2017). Ce dépôt est médié par la protéine Eggless (Rangan et al., 2011).

Cependant, les marques H3K9me3 ne sont pas uniquement et spécifiquement présentes sur les clusters de piARNs, et une large partie de l'hétérochromatine enrichie en H3K9me3 n'est pas



**Figure 25 : Rôle des protéines Kipferl et Kdm3 dans la détermination de l'identité cluster de piARNs.**

A) Les expériences de CHIP-Seq montrent que la protéine doigt de Zinc Kipferl est enrichie sur le cluster de piARNs « dual-strand » *80F* (vert). Kipferl est enrichie sur les mêmes régions que Rhino (bleu) et la marque chromatinienne H3K9me3 (marron).

En l'absence de Kipferl dans les cellules germinales (RNAi Kipferl : panel du bas), Rhino n'est plus recrutée sur le cluster *80F* alors que les marques H3K9me3 sont encore présentes.

Figure adaptée de Baumgartner et al., 2022

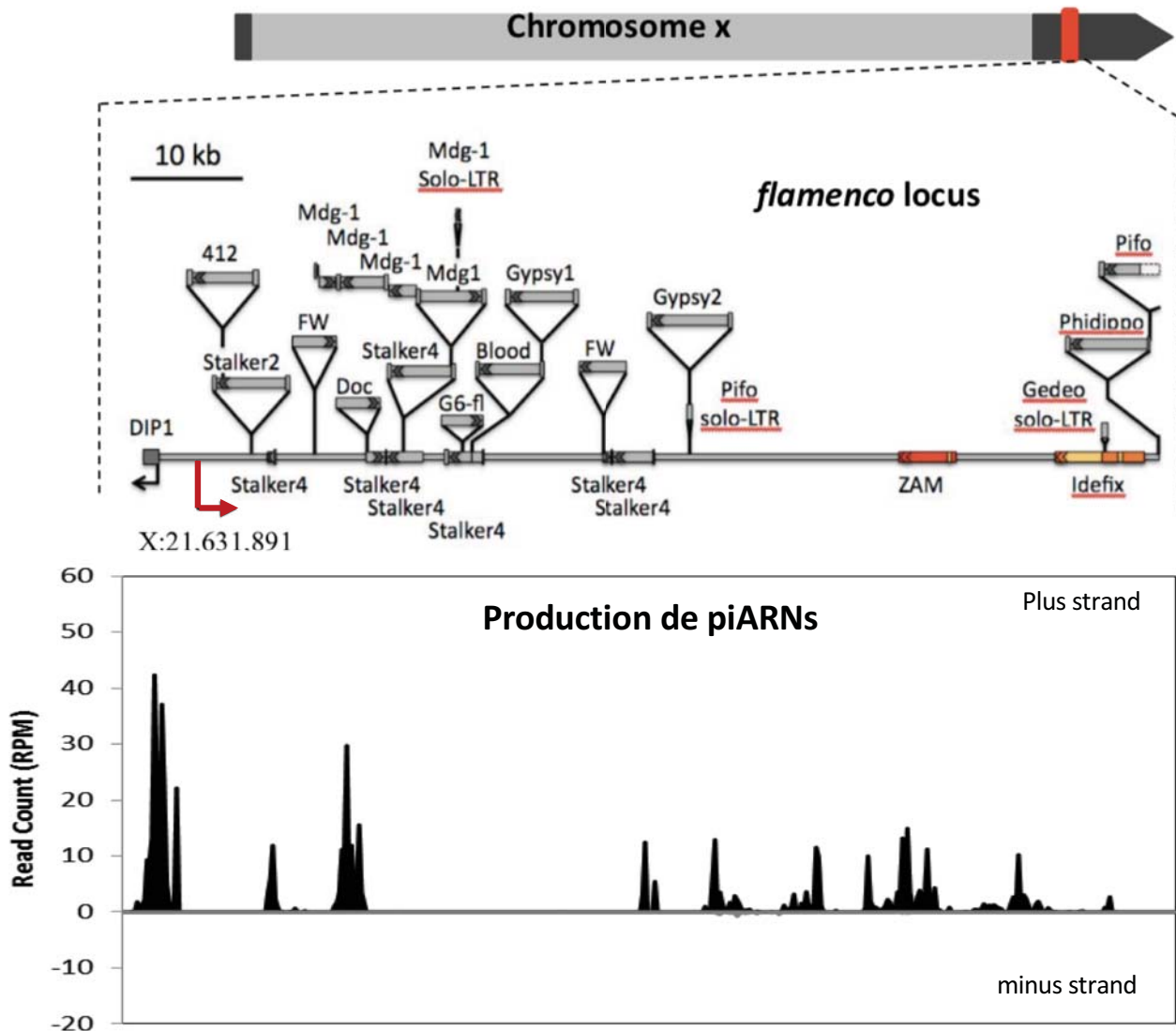
B) Le séquençage de petits ARNs en condition Kdm3-GLKD montre que la région génique *R2\_bi* commence à produire des piARNs (23-29nt unique mappers). Le Kdm3-GLKD induit également une augmentation de l'expression du locus *R2\_bi* (RNA-seq) ainsi qu'une accumulation des marques H3K3me3 et H3K9me2 (ChIP-seq). La protéine Rhino est maintenant recrutée sur le locus *R2\_bi*. Figure issue de Casier et al., 2022

ou peu liée par Rhino. La question suivante se pose alors : comment Rhino est-elle recrutée de manière spécifique sur les clusters de piARNs ? Une étude récente a montré que le recrutement spécifique de Rhino sur les clusters de piARNs est permis en partie grâce à la protéine à Zinc Finger nommée Kipferl (Baumgartner et al., 2022). Kipferl reconnaît des motifs ADN, riches en guanosine de type GRGGN, retrouvés sur les clusters de piARNs. Ce facteur, associé à un enrichissement de méthylation de marques H3K9, permettrait le recrutement de Rhino spécifiquement sur les clusters de piARNs. En absence de Kipferl, Rhino n'est plus recruté sur certains clusters de piARNs « dual-strand », bien que les marques H3K9me3 soient toujours présentes (**Figure 25A**). Cela induit une perte d'expression de ces clusters. Néanmoins, certaines régions productrices de piARNs restent enrichies en Rhino même en l'absence de Kipferl, notamment le cluster de piARNs *42AB*, suggérant qu'un autre mécanisme de recrutement de Rhino pourrait exister.

Le recrutement de Rhino sur certaines régions génomiques enrichies en H3K9me3, marque déposée par Eggless, est indispensable à l'acquisition de l'identité en tant que clusters de piARNs de certaines régions génomiques. Il était donc admis qu'une région génomique non enrichie en H3K9me3 ne pouvait, par défaut, devenir une région productrice de piARNs. Cependant, une étude récente a montré que la production de piARNs par certaines régions géniques est activement inhibée par des protéines qui évitent l'émergence de cluster de piARNs (Casier et al., 2022). En effet, l'histone déméthylase KDM3 permet la déméthylation active de certains loci géniques afin d'éviter le recrutement de la protéine Rhino sur ces régions. La perte de KDM3 induit un enrichissement en marques H3K9me2 et H3K9me3 sur certaines régions. La présence de ces deux marques hétérochromatiques induit le recrutement de la protéine Rhino et la production de piARNs par ces loci géniques (**Figure 25B**). Ces piARNs, produits à partir des deux brins d'ADN, fonctionnent comme des « piARN auto-immuns ». En effet ces piARNs nouvellement produits peuvent cibler par complémentarité de séquence les ARNm de gènes importants pour le développement embryonnaire conduisant à des défauts phénotypiques sévères. Il est donc nécessaire que certaines régions soient activement déméthylées afin d'éviter leur transformation en cluster de piARNs et donc la production de piARNs par ces régions géniques non enrichies en ETs.

Ainsi, l'environnement épigénétique préfigure le recrutement de Rhino et donc la détermination des clusters de piARNs. Certains facteurs tels que KDM3 permettent d'éliminer les marques H3K9me3 sur des régions qui ne doivent absolument pas être converties en loci producteurs de piARNs et d'autres facteurs, tels que Kipferl, permettent le recrutement de Rhino sur des régions enrichies en H3K9me3 très spécifiques.





**Figure 26 : Structure du cluster de piARNs *flamenco* chez *Drosophila melanogaster***

Le cluster *flamenco* est un cluster « uni-strand » exprimé uniquement dans les cellules somatiques folliculaires des ovaires. Il est localisé dans la région péricentromérique du chromosome X. *flamenco* est composé, comme tous les clusters de piARNs, d'un enchevêtrement d'éléments transposables. Ces ETs sont orientés en antisens par rapport au sens de transcription du cluster (flèche rouge). Seuls les ETs récents ont été représentés au-dessus de la ligne. Cinquante-deux ETs ont été identifiés au total dans *flamenco*.

L'alignement des piARNs uniques sur la séquence de *flamenco* montre que les piARNs sont produits uniquement à partir d'un seul des deux brins de l'ADN. Cela permet de générer des piARNs antisens par rapport aux séquences d'ETs qui le composent.

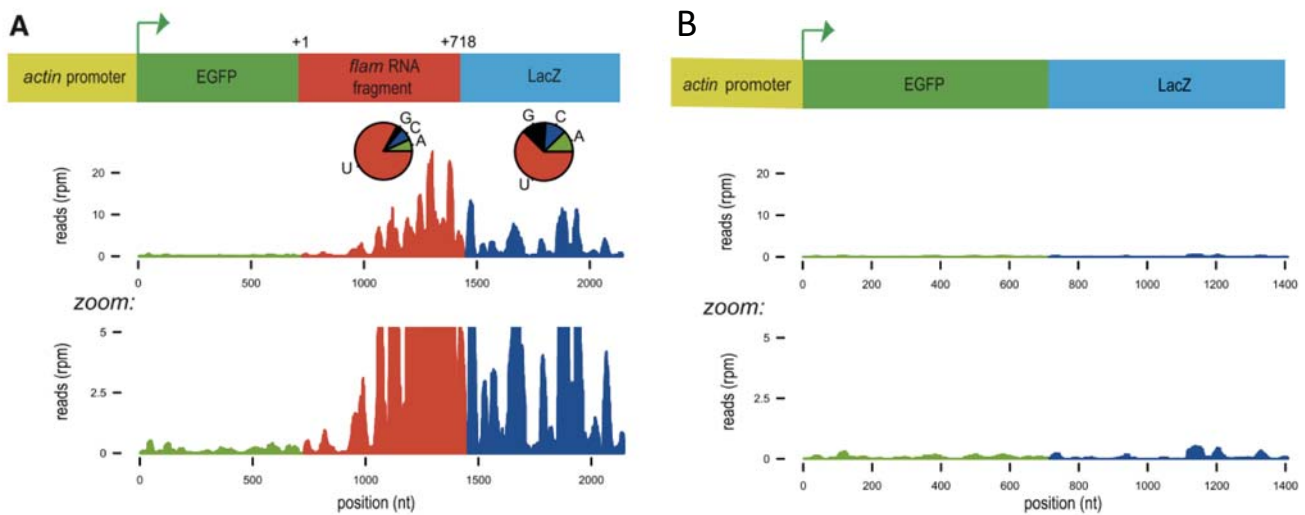
Figure adaptée de Zanni et al., 2013.

Une fois mobilisée sur les clusters de piARNs « dual-strand », Rhino recrute la protéine Moonshiner, paralogue de TFIIA-L, qui, en association avec le facteur TRF2, permet le recrutement de l'ARN polymérase II et l'initiation de la transcription (Andersen et al., 2017). Rhino s'associe également avec le dimère protéique Deadlock-Cutoff afin de former le complexe RDC (Rhino-Deadlock-Cutoff) dont le rôle est d'inhiber d'une part l'épissage du transcrit précurseur et d'autre part la mise en place d'une coiffe en 5' (Pane et al., 2011; Mohn et al., 2014; Zhang et al., 2014). Le transcrit généré sera dans ce cas directement exporté vers une structure périnucléaire spécifique des cellules germinales appelée le nuage pour être maturé en piARNs (Lim and Kai, 2007; Théron et al., 2014) (**Figure 23A**).

Le système de transcription des clusters de piARNs « dual-strand » est donc particulièrement astucieux : la présence de marques hétérochromatiques H3K9me3 empêche l'expression des éléments transposables présents dans les clusters de piARNs, et ces mêmes marques sont également reconnues par la machinerie de transcription des clusters de piARNs afin de permettre la création de transcrits maturés en piARNs. Ainsi, l'inaccessibilité de l'ADN hétérochromatique médié par H3K9me3 est transformé en empreinte spécifique permettant l'activation de la transcription non canonique des clusters de piARNs. Pour renforcer cette action, une étude a démontré que le recrutement de la protéine Maelstrom sur les clusters de piARNs « dual-strand » permettrait de bloquer la transcription canonique des ETs présents dans les clusters afin d'éviter : i) que des copies potentiellement encore actives d'ETs ne s'expriment et ne transposent, et ii) que la transcription des ETs composant les clusters de piARNs ne rentre en compétition avec la transcription du cluster lui-même (Chang et al., 2019).

- **Les clusters « *uni-strand* »**

Les clusters de piARNs « uni-strand » sont les seuls clusters exprimés au sein des cellules folliculaires. L'analyse des piARNs provenant de cellules OSS (cellules dérivées des cellules souches somatiques ovariennes) a permis de mettre en évidence que 85 % des piARNs dans ces cellules sont produits à partir d'un seul et unique locus localisé dans la région péricentromérique du chromosome X (Brennecke et al., 2007; Duc et al., 2019). Cette grande région de plus de 180 kb est le cluster de piARNs *flamenco* (*flam*), identifié préalablement comme « centre de contrôle » des ETs *Gypsy*, *ZAM* et *Idefix* (Péllisson et al., 1994b; Bucheton, 1995; Desset et al., 2003). On sait maintenant que ces ETs sont insérés dans le cluster de piARNs *flam*. Mais ce cluster est aussi composé d'une multitude d'autres ETs, souvent tronqués et/ou mutés, tous



**Figure 27 : La région +1 ; +718 du cluster de piARNs *flamenco* permet la maturation de la séquence placée en aval en piARNs.**

Un transgène contenant le fragment de *flamenco* +1 ; +718 (en rouge) encadré par deux gènes rapporteurs, EGFP en vert et LacZ en bleu, a été transfecté en cellules OSS (A). Ce même transgène sans le fragment de *flamenco* a également été transfecté en contrôle (B). Les transgènes sont exprimés de façon ubiquitaire grâce au promoteur actine. La production de piARNs par les transgènes a été analysée par séquençage des piARNs liés à la protéine Piwi. L'alignement des piARNs sur les constructions est représenté ci-dessus. Des piARNs sont produits à partir de la séquence LacZ (en bleu) présente en aval du fragment de *flamenco* +1 ; +718 (en rouge) mais aucun piARN provenant de la séquence eGFP placée en amont (en vert) n'est observé. Lorsqu'aucun fragment de *flamenco* n'est présent dans le transgène, aucune production de piARN n'est observée.

Figure adaptée de Homolka et al., 2015

orientés du centromère vers le télomère - et donc en antisens par rapport au sens de transcription du cluster. Cette organisation permet de générer des piARNs antisens par rapport à l'ARNm des ETs qui le composent (**Figure 26**). *flam* est le producteur majoritaire de piARNs au sein des cellules somatiques folliculaires de l'ovaire, et donc le régulateur principal des ETs dans ces cellules.

Les premières études analysant la transcription de *flamenco* ont été réalisées au sein de notre équipe. Coline Goriaux et ses collaborateurs identifient en 2014 le site majeur d'initiation de la transcription de *flam*. Ce dernier est localisé à la position X : 21 631 891 (Release 6), 1733 pb en amont du gène *DIP1* (Goriaux et al., 2014a). Le promoteur de *flam* est un promoteur de type INR DPE dépendant de l'ARN polymérase II, et l'activation de sa transcription est sous le contrôle entre autres du facteur de transcription Cubitus Interruptus (Ci). De plus, comme les transcrits issus de séquences géniques, les transcrits des *flam* sont coiffés à leur extrémité 5' et épissés (Mohn et al., 2014). Les ARNs produits s'accumulent ensuite dans le noyau, en périphérie, au niveau d'une structure appelée Dot COM (Dennis et al., 2013). Enfin, la maturation en piARNs se produit dans le cytoplasme, au sein d'un corps multiprotéique appelé le corps Yb (**Figure 23B**). La transcription de *flam* est donc similaire à celle des gènes : on parle de transcription « canonique ». Cela soulève alors une question majeure : comment les transcrits de *flam*, qui ressemblent fortement à des transcrits de gènes, sont-ils reconnus pour être maturés en piARNs ? Comment la cellule distingue-t-elle les transcrits émanant de clusters de piARNs « unistrand », devant être maturés en piARNs, des transcrits provenant de gènes devant être traduits ?

Des études préliminaires menées au sein de l'équipe, en collaboration avec l'équipe de Ramesh Pillai, suggèrent qu'une séquence ou structure présente dans l'ARN de *flam* serait reconnue par l'une des protéines de la voie des piARNs, permettant ainsi la maturation du transcrit en piARNs. En effet, des études réalisées en cellule OSS ont démontré que la présence d'une séquence de 718 nts de *flam* (+1 à +718 nts par rapport à son site d'initiation de transcription) insérée en amont d'un gène rapporteur est suffisante pour induire la production de piARNs provenant de l'ARN du gène rapporteur (**Figure 27**) (Homolka et al., 2015b). Cependant, ces résultats restent encore à démontrer *in vivo*. La particularité de cette région et les facteurs impliqués dans la reconnaissance des transcrits de clusters de piARNs « unistrand » restent également énigmatiques. Des études récentes ont néanmoins montré que le recrutement artificiel de la protéine Yb sur un transcrit induit sa maturation en piARNs. La fixation de Yb, en association avec la protéine Armi, pourrait ainsi permettre d'identifier les transcrits comme



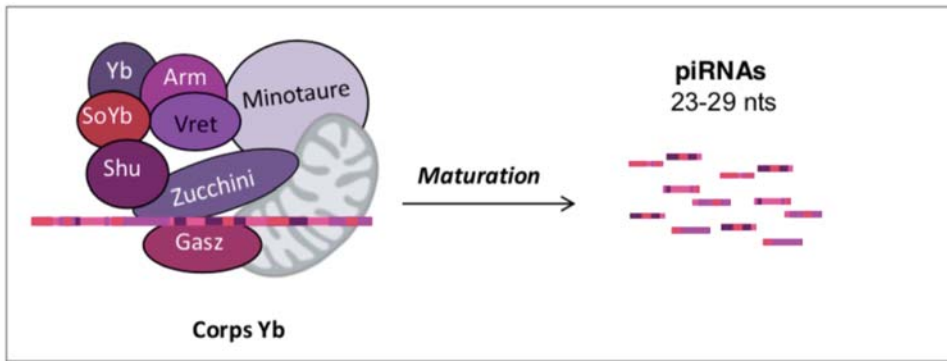
transcrits de clusters de piARNs et induire leur relocalisation vers le centre de maturation (Pandey et al., 2017).

À la différence des clusters de piARNs « dual-strand », le cluster de piARNs *flamenco* uni-strand est donc transcrit à partir d'un seul promoteur bien défini. *flamenco* est faiblement enrichi en marques H3K9me3 et Rhino n'est donc pas recrutée sur ce cluster. En effet, la présence de H3K9me3 sur les clusters de piARNs « dual-strand » est médiée par le ciblage du cluster par des piARNs complémentaires, liés à Piwi, déposés maternellement. Or, aucun piARN produit par *flam* n'est transmis maternellement. Aucun piARN ne peut donc cibler *flam*. Ce cluster est transcrit de façon canonique, permettant de générer un long transcrypt précurseur qui sera ensuite mûré en piARNs. Cependant, cette observation soulève une autre question : comment les ETs insérés dans *flam* sont-ils réprimés en absence de marques hétérochromatiques sur ce cluster ? Maelstrom, protéine impliquée dans la répression des ETs présents dans les clusters de piARNs « dual-strand », semble faiblement recruté sur *flam*. L'absence de ce facteur ne semble pas induire une dérégulation des ETs présents dans *flam* (Chang et al., 2019). Et même si un grand nombre d'ETs présentes dans *flam* sont dégénérées, de nouvelles insertions se produisent fréquemment, conduisant à la présence de copies entières et fonctionnelles pour la transposition. Il doit donc exister un mécanisme permettant d'éviter le recrutement de facteurs de transcription au niveau des promoteurs des ETs localisés dans *flam* et l'expression de ces copies potentiellement actives. Ce mécanisme reste à l'heure actuelle, à ma connaissance, inconnu.

### II.3 c) Maturation des transcrits précurseurs en piARNs

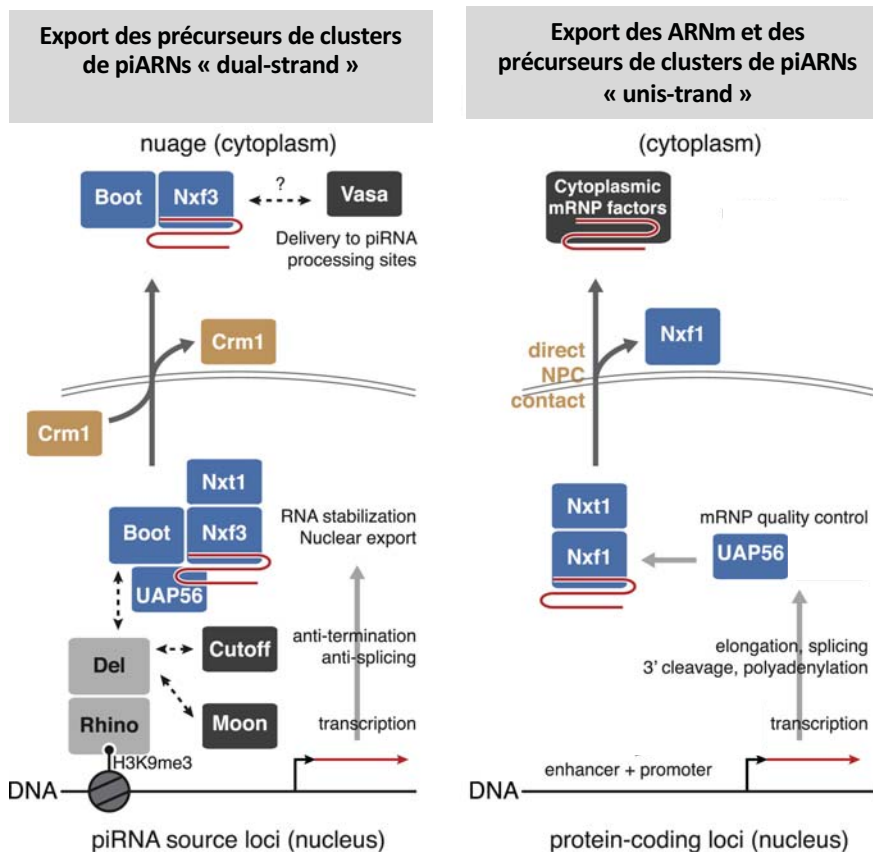
- **Maturation des transcrits précurseurs en piARNs dans les cellules folliculaires**

Les transcrits du cluster de piARNs uni-strand *flamenco*, produits dans les cellules folliculaires, s'accumulent tout d'abord dans un focus nucléaire, appelé le « Dot COM », avant d'être exportés vers le cytoplasme (Dennis et al., 2013). Ensuite, l'exportation des transcrits précurseurs vers le cytoplasme est ensuite dépendante des exportines Nxt1 et Nxf1, et de la protéine UAP56 ; ce processus est habituellement utilisé pour le transport des ARNs messagers vers le cytoplasme (Dennis et al., 2016). Puis, les précurseurs sont ensuite internalisés dans le



**Figure 28 : Représentation schématique du corps Yb**

Le corps Yb est une structure multiprotéique responsable de la maturation des transcrits de clusters de piARNs en piARNs dans les cellules folliculaires de l’ovaire. Le corps Yb est notamment composé des protéines Yb, Armitage (Arm), Sister of Yb (SoYB), Vreteno (Vret), Shutdown (Shu), Zucchini, Minotaure et de GasZ.



**Figure 29 : Comparaison des voies d’exportation des précurseurs issus des clusters de piARNs « dual-strand » et « uni-strand »**

Les précurseurs de clusters de piARNs « uni-strand », utilisent, comme les ARNm, un mécanisme d’exportation dépendant des exportines Nxt1 et Nxf1 grâce à la reconnaissance de signatures spécifiques sur les ARNs telles que la polyadénylation ou les marques d’épissage. De leur côté, les précurseurs des clusters de piARNs « dual-strand », qui ne possèdent pas ces signatures, utilisent un système d’export spécifique dépendant des protéines Bootlegger (Boot), Nxf1, Nxf3 et Cmr1 pour être exportés vers le nuage.

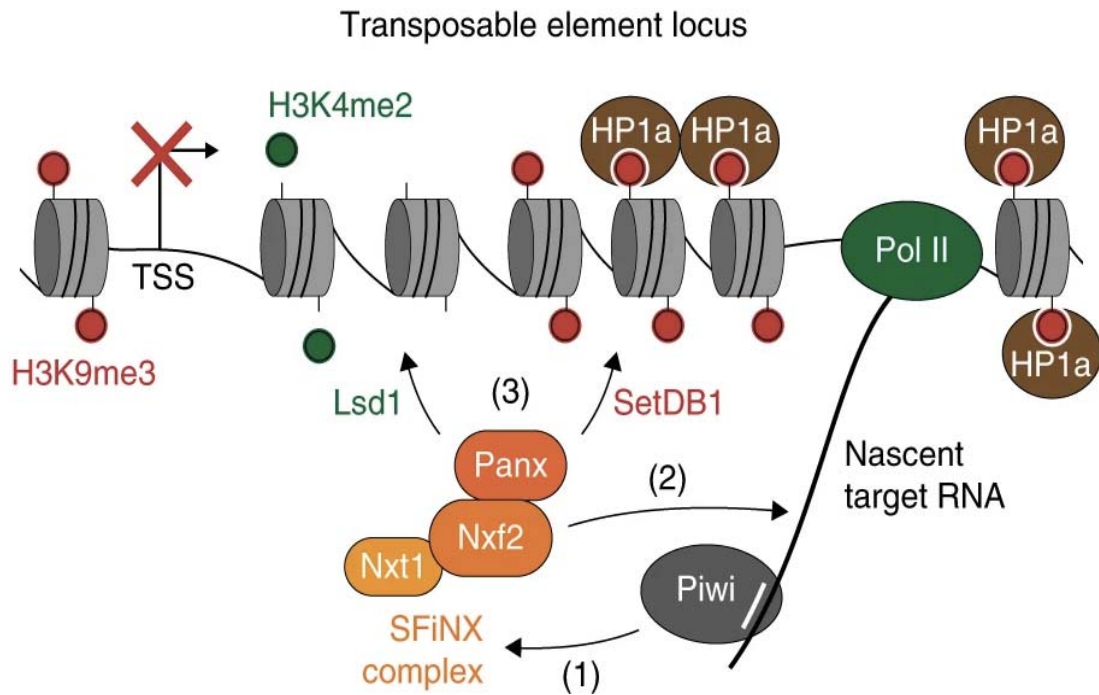
Figure adaptée de ElMaghraby et al., 2019

corps Yb, structure périnucléaire multiprotéiques composée d'une panoplie de protéines nécessaires à la biogenèse des piARNs. On y retrouve entre autres les protéines Yb, Armitage (Armi), Zucchini (Zuc), Vreteno (Vret), Shutdown (Shu) et Sister of Yb (SoYb) (Gunawardane et al., 2007; Saito et al., 2010; Handler et al., 2013). La protéine Zuc est l'endonucléase localisée à la membrane des mitochondries, responsable de la coupure des extrémités 5' des piARNs. Zuc possède deux cofacteurs nécessaires à son activité : GasZ et Minotaure (Vagin et al., 2013) (**Figure 28**). Les fragments générés sont de tailles variables et sont en antisens par rapport aux séquences des ARNm des ETs. Ces pré-piARNs sont ensuite raccourcis à leur extrémité 3' par l'enzyme Nibbler, ayant une activité exonucléase, afin de générer des fragments de 23 à 29 nts (Hayashi et al., 2016). Enfin, la méthyltransférase Hen1 méthyle l'extrémité 3' (2'-O-méthylation), conduisant à la formation d'un piARN mature pouvant être chargé sur la protéine Piwi (Horwich et al., 2007).

- **Maturation des transcrits précurseurs en piARNs dans les cellules germinales**

Dans les cellules germinales, l'export des transcrits des clusters de piARNs dual-strand s'effectue de manière singulière. En effet, les transcrits précurseurs des clusters de piARNs dual-strand ne sont pas maturés, c'est-à-dire ni épissés ni polyadénylés, contrairement aux ARNm devant être exportés vers le cytoplasme. Or, des mécanismes de surveillance présents dans le noyau garantissent que les transcrits « non maturés » ne soient pas exportés vers le cytoplasme et soient à la place dégradés. Pour contourner ce mécanisme de dégradation, les précurseurs de clusters de piARNs dual-strand utilisent un mécanisme d'exportation bien spécifique *via* l'utilisation du facteur d'export Nxf3 (ElMaghraby et al., 2019). À la suite de la transcription, les transcrits précurseurs des clusters de piARNs dual-strand sont reconnus par le complexe RDC qui permet notamment d'éviter leur épissage et la polyadénylation. Ce complexe RDC, notamment la protéine Cutoff, permet également le recrutement des protéines UAP56 et Bootlegger qui, à leur tour, recrutent les facteurs Nxf3 et Nxf1. Nxf3 exporte ensuite l'ARN précurseur dans le cytoplasme par un processus dépendant de l'exportine Cmr1 (**Figure 29**). Nxf3 est codée par un gène issu d'une duplication puis d'une néofonctionnalisation du gène *nxf1* impliqué dans l'exportation des ARN messagers. La création de ce nouveau gène permet de contourner les systèmes de surveillance nucléaires des ARNs et permet un export des précurseurs des clusters de piARNs dual-strand jusqu'au





**Figure 30 : Répression transcriptionnelle des ETs par le complexe SFiNX**

La protéine Piwi chargée par un piARN antisens reconnaît le transcrit naissant d'un ET par complémentarité de séquence. Lors de la fixation de la cible, Piwi interagit avec le complexe multiprotéique SFiNX (1). Le recrutement du complexe SFiNX est médié grâce à la reconnaissance du transcrit naissant de l'ET par la protéine Nxf2 (2). Ceci permet à Panoramax de recruter des modificateurs de la chromatine : Lsd1 pour déméthyliser H3K4me2, et SetDB1 pour déposer les marques H3K9me3. L'hétérochromatinisation du promoteur engendrée permet de bloquer l'accès à la machinerie transcriptionnelle. Les marques H3K9me3 sont ensuite reconnues par la protéine HP1a qui maintient la répression.

Figure issue de Batki et al., 2019

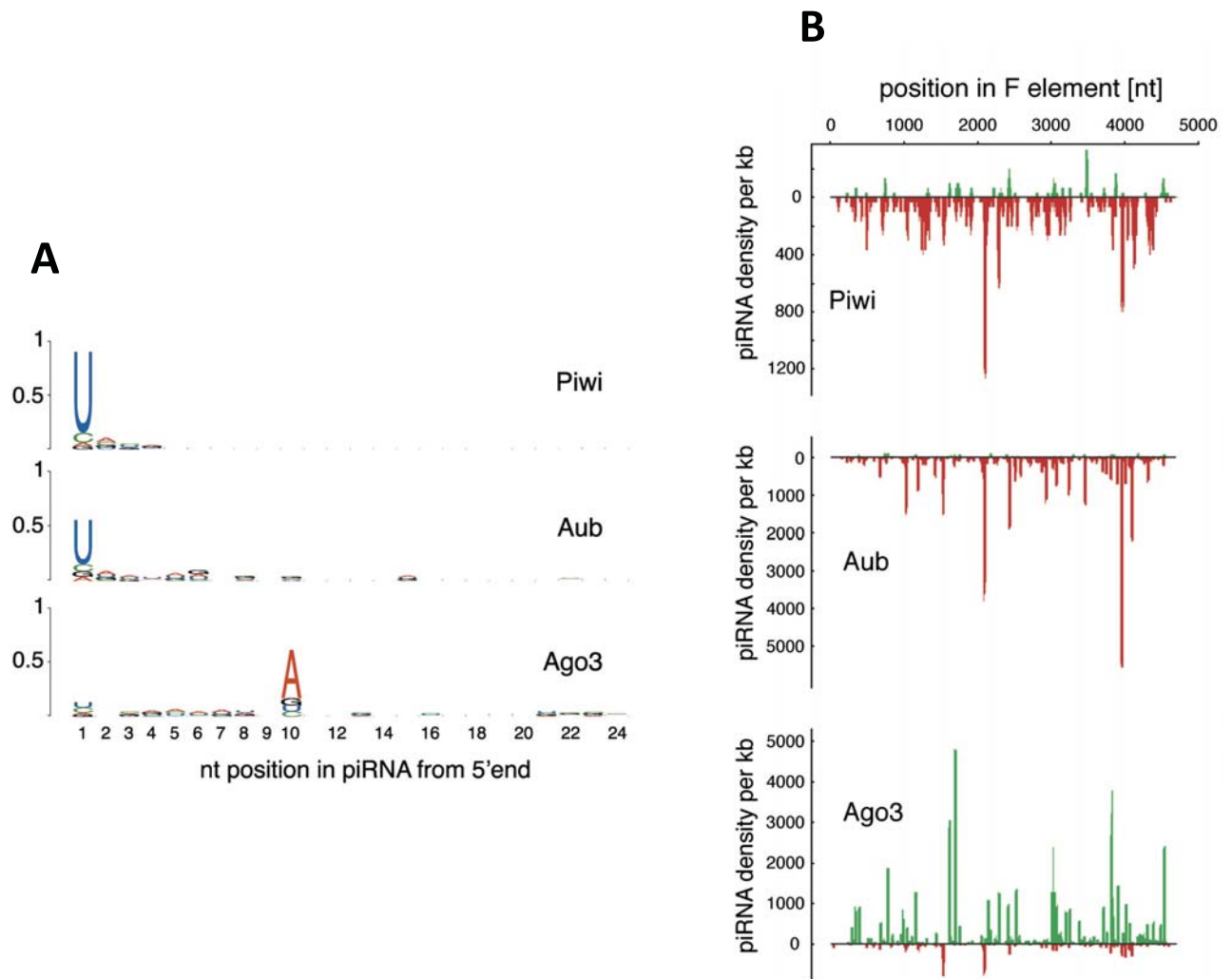
cytoplasme. Une fois dans le cytoplasme, les précurseurs sont pris en charge par la protéine Vasa afin d'être acheminés vers leur lieu de maturation : le nuage (Zhang et al., 2012). Le nuage est une structure électriquement dense qui entoure le noyau des cellules germinales composé des protéines nécessaires à la biogenèse des piARNs. On y retrouve les mêmes protéines que dans le corps Yb, présent dans les cellules folliculaires, à l'exception de la protéine Yb elle-même (Malone et al., 2009). Ainsi, dans les cellules germinales, la maturation du précurseur s'effectue dans le nuage, puis les piARNs peuvent être pris en charge par la protéine Piwi, mais peuvent également être liés par une autre protéine de la famille PIWI : la protéine Aubergine (Aub).

Dans les cellules germinales, la maturation des transcrits issus des clusters de piARN peut également être initiée par un autre processus : les transcrits de cluster de piARNs peuvent être reconnus par les complexes piARN/PIWI hérités maternellement, induisant le clivage du précurseur pour générer l'extrémité 5' du piARN. Les piARNs permettant d'initier la coupure en 5' sont alors appelés « *initiator* piARNs », anciennement appelés « piARNs primaires ». Les piARNs générés après maturation des transcrits sont quant à eux appelés « *responder* piARNs », anciennement « piARNs secondaires ». (Ozata et al., 2019).

### II.3. d) Rôle des piARNs dans la répression des ETs

- **Répression transcriptionnelle des éléments transposables**

Les piARNs matures formés dans les cellules folliculaires et germinales sont chargés par la protéine Piwi, formant le complexe piRISC (*piRNA Induced Silencing Complex*). Le complexe piRISC est importé dans le noyau grâce au signal d'adressage nucléaire (NLS) présent à l'extrémité N-terminale de la protéine Piwi (Sienski et al., 2012). Le cofacteur Asterix permet la stabilisation du complexe piRISC (Dönertas et al., 2013). Le piRISC cible ensuite par complémentarité de séquence les ARNm d'éléments transposables en cours de transcription. Lors de sa fixation au transcrit cible, Piwi interagit avec un complexe répresseur appelé le complexe SFiNX (*Silencing Factor interacting Nuclear Export variant*) (Batki et al., 2019). Ce complexe multiprotéique est composé des protéines Panoramix (Panx), Nxf2 et Nxt1. Le recrutement du complexe SFiNX est médié grâce à la reconnaissance du transcrit naissant par la protéine Nxf2 (**Figure 30**). Panoramix recrute ensuite des modificateurs de la chromatine tels que la méthyltransférase H3K9 Eggless (aussi appelée SetDB1) et la déméthyltransférase



**Figure 31 : Caractérisation des piARNs liés aux protéines Piwi, Aub et Ago3**

A) Weblogo représentant l'enrichissement de chaque nucléotide à chaque position des piARNs liés par les protéines Piwi, Aub et Ago3. Les piARNs liés par les protéines Piwi et Aub présentent préférentiellement une uridine (U) à la position 1 alors que les piARNs liés par Ago3 présentent une adénine (A) à la position 10.

B) Alignement des piARNs obtenus par immunoprécipitation des protéines Piwi, Aub et Ago3 sur la séquence de l'élément transposable *F*. Les protéines Piwi et Aub lient majoritairement des piARNs antisens par rapport au sens de transcription de l'ET (en rouge). À l'inverse, les piARNs sens par rapport au sens de transcription de l'ET (vert) sont préférentiellement pris en charge par la protéine Ago3.

Figure issue de Brennecke et al., 2007

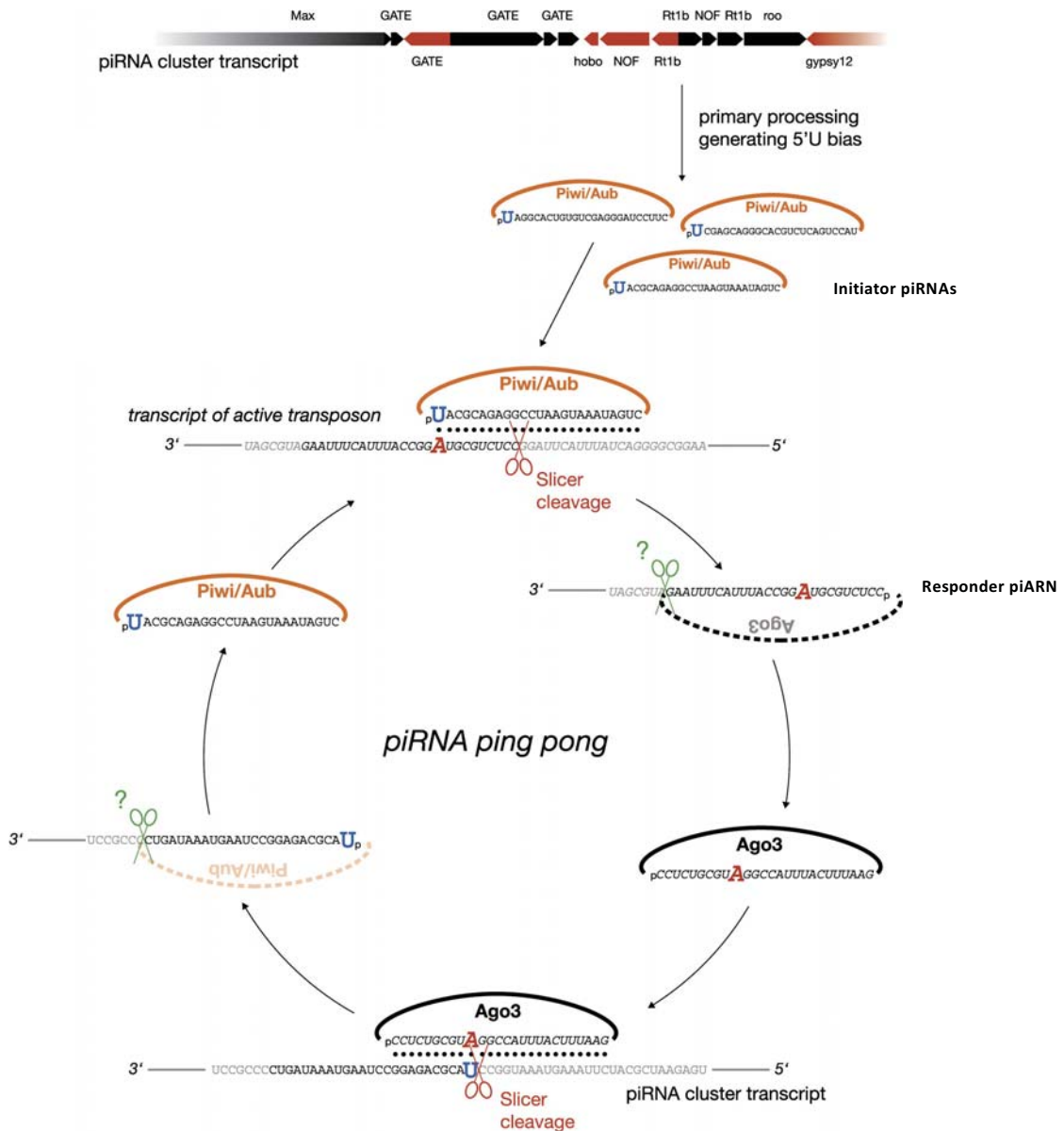
des marques H3K4-methyl Lsd1. Ces protéines induisent une hétérochromatinisation des promoteurs des ETs et une répression de leur expression. Les marques H3K9me3 déposées sont reconnues par la protéine HP1a qui maintient et verrouille la répression transcriptionnelle des ETs. L'implication de la protéine Nxf2 dans la répression transcriptionnelle des ETs est quelque peu inattendue. En effet, Nxf2 appartient à la famille des NXF (Facteurs d'Export Nucléaire des ARNs), tout comme Nxf1 et Nxf3 dont nous avons parlé dans la partie « exportation des précurseurs des clusters de piARNs ». Nxf2 a donc conservé son habileté à lier des ARNs mais ne semble plus impliquée dans l'export de ces ARNs. De ce fait, il semblerait que la famille des protéines NXF soit fortement exposée aux innovations génétiques par duplication puis néofonctionnalisation de leurs gènes. Les variants Nxf2 et Nxf3, découverts récemment chez *Drosophila melanogaster*, semblent dorénavant remplir des fonctions clés dans la défense du génome (Batki et al., 2019; ElMaghraby et al., 2019).

Enfin, des études réalisées en cellules OSS (Ovarian Somatic Sheet Cells) ont également montré le rôle-clef d'un autre acteur dans la répression transcriptionnelle des ETs : il s'agit du facteur Maelstrom (Mael) dont nous avons parlé plus haut puisqu'il intervient dans la répression des ETs localisés dans les clusters de piARNs « dual-strand ». Mael est recrutée par la protéine Piwi et son absence induit une perte de la répression transcriptionnelle des ETs sans affecter la production des piARNs (Sienski et al., 2012). Son mode d'action reste encore à ce jour inconnu.

L'action de tous ces facteurs a donc au final pour but d'induire une hétérochromatinisation des promoteurs des ETs afin de les rendre inaccessibles à la machinerie de transcription. Cette répression transcriptionnelle se produit dans les cellules germinales et folliculaires de l'ovaire de drosophile afin d'empêcher l'expression des ETs.

- **Répression post-transcriptionnelle des éléments transposables et amplification du pool de piARNs dans les cellules germinales**

La répression post-transcriptionnelle des ETs est médiée par deux protéines de la famille PIWI : Aubergine (Aub) et Argonaute 3 (Ago3). Ces deux protéines sont exprimées exclusivement dans les cellules germinales (**Figure 22B**). Par conséquent, dans les cellules folliculaires, seule la répression transcriptionnelle des ETs, médiée par Piwi, est effectuée. La répression post-transcriptionnelle des ETs est ainsi un mécanisme spécifique des cellules germinales. Les protéines Aub et Ago3 sont localisées au niveau du nuage dans les cellules germinales et lient



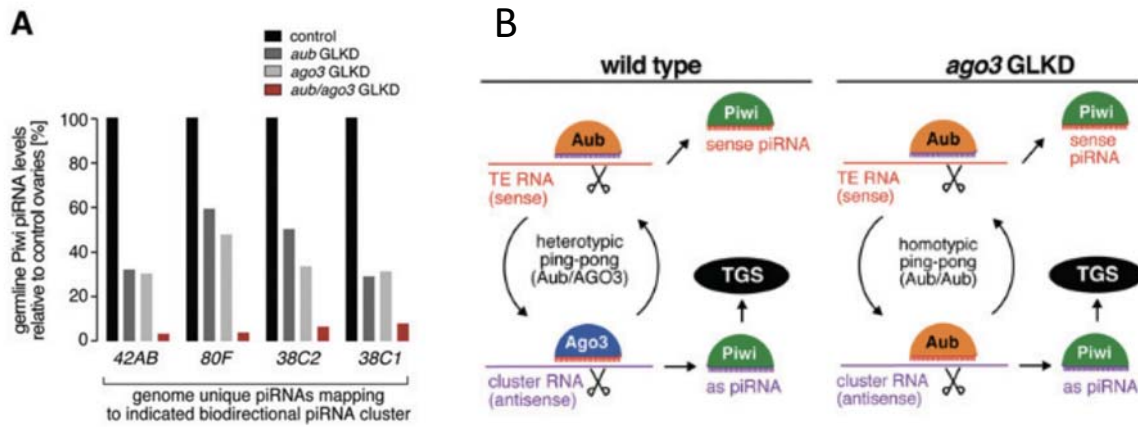
**Figure 32 : Représentation schématique du cycle ping-pong**

Le complexe piARN/Aub reconnaît, par complémentarité de séquence, le transcrite d'un ET actif. La protéine Aub, via son activité endonucléasique, génère une coupure entre le dixième et le onzième nucléotide de la région ciblée par « l'initiator piARN » (antisens). Le « responder piARN » généré (sens) possède une adénosine (A) en position 10, il est alors pris en charge par la protéine Ago3. Le complexe Ago3/piARN va à son tour reconnaître, par complémentarité de séquence, le transcrite provenant d'un cluster de piARNs. Grâce à son activité endonucléasique, Ago3 coupe également entre le dixième et le onzième nucléotide de la région ciblée par le piARN, générant ainsi l'extrémité 5' du piARN pris en charge par la protéine Aub. Ce cycle permet une amplification du pool de piARNs dans les cellules germinales et une répression post-transcriptionnelle des ETs.

Figure issue de Brennecke et al., 2007

des piARNs pour former des complexes piRISC. Le séquençage des petits ARNs liés à ces deux protéines a révélé que Aub lie majoritairement des piARNs avec une Uridine en 5' et antisens par rapport à la séquence de l'ARNm des ETs. La protéine Ago3 lie quant à elle des piARNs différents puisqu'ils ne possèdent pas cette Uridine préférentielle en 5' mais ont la particularité de présenter préférentiellement une Adénine en position 10. De plus, les piARNs chargés par Ago3 sont des piARNs sens par rapport à la séquence de l'ARNm des ETs (Brennecke et al., 2007) (**Figure 31**). À l'inverse de la protéine Piwi, Aub et Ago3 ne possèdent pas de signal NLS, leur action s'effectue donc dans le cytoplasme des cellules germinales.

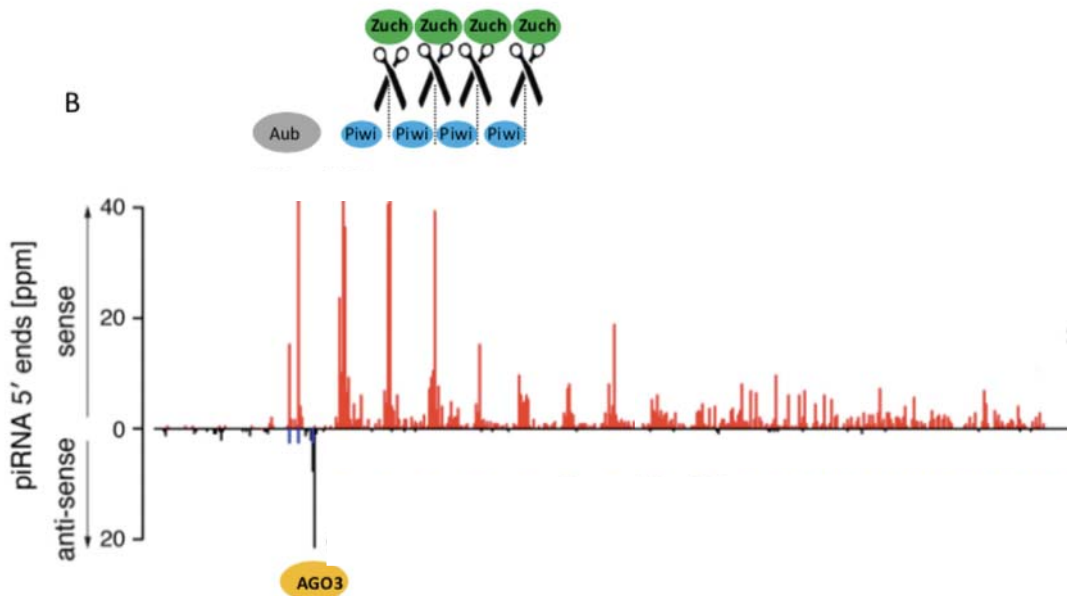
La répression post-transcriptionnelle des ETs est basée sur un mécanisme appelé le cycle ping-pong. Les piARNs antisens issus de la maturation des ARNs précurseurs des clusters de piARNs sont pris en charge par la protéine Aub ou par la protéine Piwi. Le complexe piARN/Aub reconnaît ensuite, par complémentarité de séquence, l'ARNm cytoplasmique d'un ET qui aurait échappé au contrôle transcriptionnel. La protéine Aub possède une activité endonucléase qui lui permet de générer une coupure dans l'ARNm de l'ET. Les piARNs liés par Aub sont des « *initiator* piARNs » puisqu'ils initient la coupure du transcrite pour former l'extrémité 5' du piARN. Cette coupure s'effectue à une position bien particulière : entre le dixième et le onzième nucléotide de la région ciblée par l'« *initiator* piARN » (Gunawardane et al., 2007). Le clivage par Aub permet donc : i) de cliver l'ARNm de l'ET, et par conséquent d'éviter sa traduction et sa transposition, ii) de générer des nouveaux piARNs alors appelés « *responder* piARNs » (Ozata et al., 2019). Ces piARNs sont orientés sens par rapport à la séquence de l'ARNm de l'ET puisqu'ils sont produits directement à partir de la maturation de l'ARNm de l'ET. De plus, le premier nucléotide – une Uridine - des « *initiator* piARN » (antisens) correspond au dixième nucléotide – une Adénine - du « *responder* piARN » (sens) chargé par Ago3 : les piARNs sens et antisens partagent ainsi une complémentarité de séquence sur dix nucléotides à partir de leur extrémité 5'. Les piARNs chargés par la protéine Ago3 subissent une maturation à leur extrémité 3' ainsi qu'une 2'O-méthylation par la méthyltransférase Hen1, et peuvent alors à leur tour jouer le rôle d'« *initiator* piARNs ». Le complexe piARN/Ago3 cible ensuite les transcrits précurseurs issus des clusters de piARNs et induisent un clivage de ce dernier en piARNs antisens (**Figure 32**). Ces piARNs antisens générés pourront après être pris en charge par Aub ou par Piwi. Ainsi, en plus d'induire une répression post-transcriptionnelle des ETs, ce mécanisme permet d'amplifier le pool de piARNs présents dans les cellules germinales : on parle de boucle d'amplification ou amplification ping-pong (Brennecke et al., 2007; Gunawardane et al., 2007). Les piARNs issus de ce cycle ping-pong sont identifiables grâce à la présence d'une complémentarité de séquence caractéristique



**Figure 33 : Implication des protéines Aub et Ago3 dans la biogenèse des piARNs des clusters de piARNs « dual-strand »**

- A) La perte de Aub ou de Ago dans les cellules germinales induit une diminution de la production des piARNs produits par les clusters de piARNs « dual-strand » tels que le *42AB*, *80F*, *38C1* et *38C2*. En absence des deux protéines Ago3 et Aub, la production des piARNs par ces clusters est quasiment abolie.
- B) En condition sauvage (wild type), le ping-pong est préférentiellement réalisé via une interaction hétérotypique entre Aub et Ago3. En absence de Ago3, un ping-pong homotypique Aub/Aub peut se produire, permettant un maintien du ping-pong dans les cellules germinales. Les piARNs générés par le cycle ping-pong participent à la formation des piARNs antisens pris en charge par Piwi pour effectuer la répression transcriptionnelle des ETs.

Figure issue de Senti et al., 2015



**Figure 34 : Mécanisme de Phasing**

Suite au cycle ping-pong réalisé par les protéines Aub (en gris) et Ago3 (en orange), le reste du transcrite de l'ET est maturé de façon phasique par la protéine Zucchini (en vert avec des ciseaux) et la protéine Piwi (en bleu) afin de générer des « trail piARNs ». Figure adaptée de Mohn et al., 2015

de leur extrémité 5' sur dix nucléotides entre les piARNs sens et antisens ; c'est ce qu'on appelle la signature ping-pong.

La maturation des transcrits des clusters de piARNs « dual-strand » est donc dépendante des piARNs produits par le cycle ping-pong. La perte de Aub ou de Ago3 induit une forte diminution (sans une perte totale) de la production des piARNs par les clusters de piARNs « dual-strand ». Seule la perte des deux protéines induit une absence totale de production de piARNs (**Figure 33A**). En l'absence de Ago3, un ping-pong dit homotypique peut s'effectuer, faisant uniquement intervenir la protéine Aub (Senti et al., 2015) (**Figure 33B**). En revanche, le ping-pong est indépendant de la présence des autres protéines de la voie des piARNs tel que les protéines Piwi, Armi ou Vret (Malone et al., 2009).

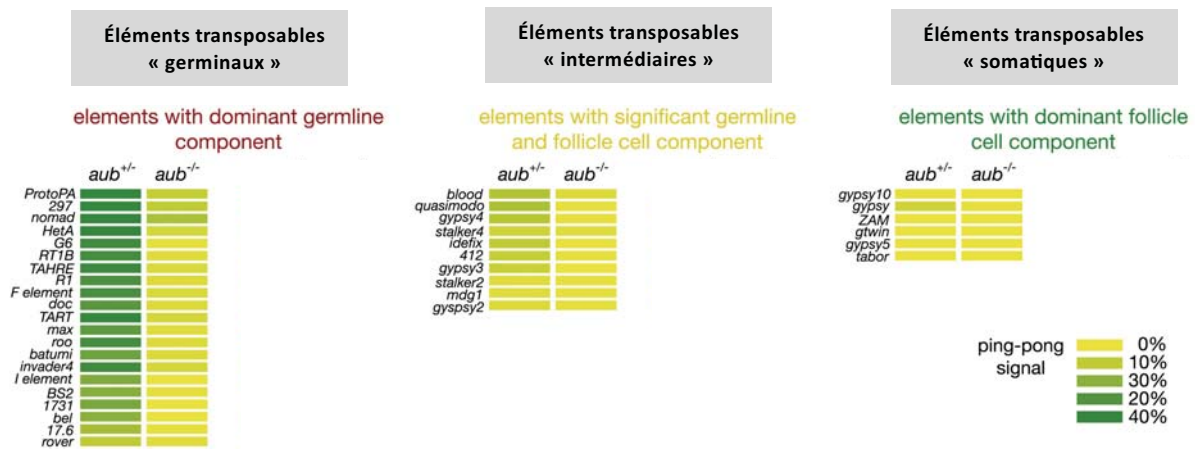
Enfin, le ping-pong permet aussi la production des piARNs appelés « *trail* piARNs » (Han et al., 2015; Mohn et al., 2015). Ils sont générés à partir d'un transcrit préalablement coupé en 5' par Aub ou Ago3 lors du cycle ping-pong. La coupure endonucléasique du transcrit ciblé par les protéines Aub ou Ago3 induit la maturation du reste du transcrit de manière phasique par les protéines Piwi et Zucchini (Zuc). Pour cela, les protéines Piwi se fixent sur le reste du transcrit (**Figure 34**). La protéine Zucchini coupe de part et d'autre des protéines Piwi, produisant ainsi, de façon cyclique, des piARNs présentant une extrémité 5'U. Cette nouvelle source de production de piARNs est appelée le "*Phasing*" ou « *Inchworming* » (Homolka et al., 2015a)

Les piARNs produits par le ping-pong permettent au final de réprimer efficacement les ETs dans les cellules germinales. Ces piARNs seront également transmis maternellement à l'embryon (Brennecke et al., 2008). Il a été montré que les piARNs liés par Aub et Piwi hérités maternellement sont nécessaires à la réactivation des clusters de piARNs dont ils sont issus dans la descendance (Akkouche et al., 2017). Seuls les piARNs produits à partir de clusters de piARNs exprimés dans les cellules germinales sont transmis maternellement (Malone et al., 2009). Par exemple, les piARNs du cluster de piARN *flamenco* ne sont pas transmis par la mère à la descendance. Les mécanismes permettant la réactivation des clusters de piARNs « unistrand » dans la descendance restent encore énigmatiques à l'heure actuelle.

### II.3 e) Répression tissu spécifique des ETs

Ainsi, même si la voie des piARNs est présente dans les deux types de cellules qui composent l'ovaire de la drosophile, il existe des différences notables. Les protéines Aub et Ago3 n'étant pas exprimées dans les cellules folliculaires, uniquement la biogenèse primaire des piARNs se

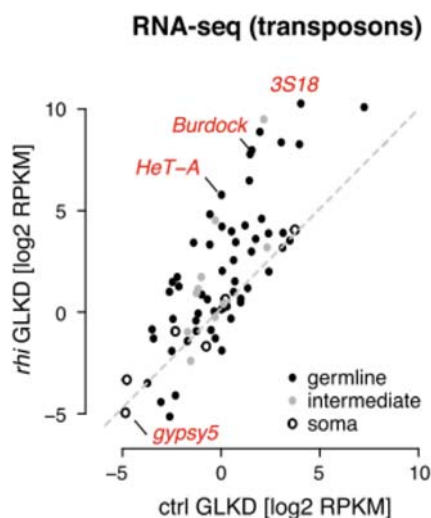




**Figure 35 : Exemples d'éléments transposables « germinaux », « intermédiaires » et « somatiques »**

L'analyse de la signature ping-pong permet de déterminer l'origine des piARNs. Les piARNs possédant une forte signature ping-pong sont produits par les cellules germinales de l'ovaire. Les ETs sont dits « germinaux » s'ils sont ciblés par des piARNs avec une signature ping-pong supérieur à 20 %. Cette signature est perdue en condition mutante pour Aub, démontrant que c'est la voie de piARNs dans les cellules germinales qui permet la production de ces piARNs. Les ETs « intermédiaires » sont quant à eux ciblés par des piARNs présentant une faible signature ping-pong. Enfin, les ETs « somatiques » sont ciblés par les piARNs qui ne présentent pas de signature ping-pong car ils sont uniquement produits par les cellules folliculaires de l'ovaire où Aub et Ago3 sont absents.

Figure issue de Malone et al., 2009



**Figure 36 : Analyse de l'expression des éléments transposables « germinaux », « intermédiaires » et « somatiques » en absence de Rhino dans les cellules germinales**

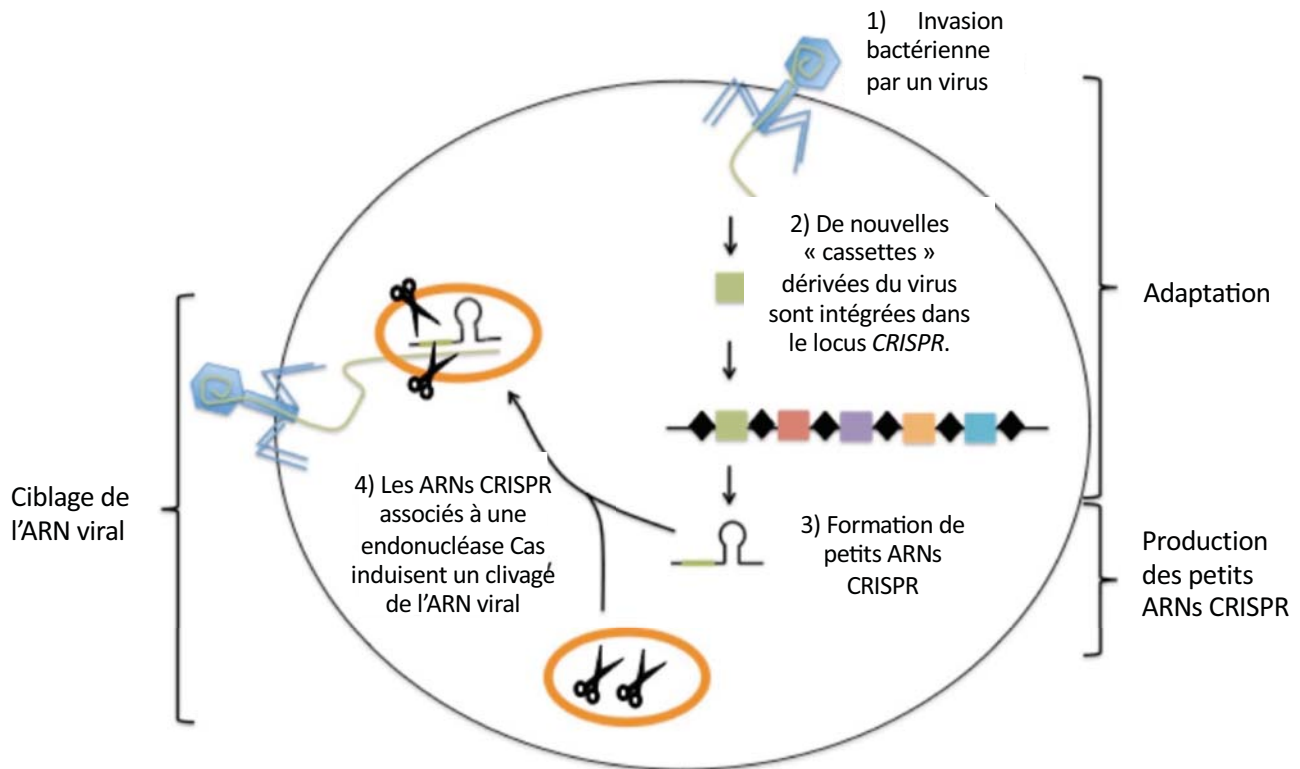
Diagramme de dispersion illustrant les niveaux d'expression des ETs des classes indiquées dans les ovaires témoins par rapport aux ovaires Rhino GLKD ("Germ-Line Knock-Down", induction d'un RNAi contre Rhino dans les cellules germinales). Le RNAi contre Rhino induit une dérégulation des ETs « germinaux » (points noirs) tels que *Burdock* ou *HeT-A*, et dans une moindre mesure des ETs « intermédiaires » (points gris). L'expression des ETs « somatiques », tels que *Gypsy5*, n'est pas affectée par la mutation de Rhino (points blancs).

Figure issue de Mohn et al., 2014

produit dans ces cellules. De plus, comme évoqué précédemment, il existe une expression tissu spécifique des clusters de piARNs. Par exemple, le cluster *42AB* n'est exprimé que dans les cellules germinales alors que *flamenco* n'est exprimé que dans les cellules folliculaires. Or, chaque cluster de piARNs possède un répertoire d'ETs spécifique, et produit donc un pool de piARNs spécifique. Cela signifie que les piARNs produits dans les cellules germinales sont différents des piARNs produits dans les cellules folliculaires et ciblent des ETs différents. Des analyses des piARNs produits dans les ovaires de drosophiles mutantes pour différents acteurs de la voie des piARNs ont permis de révéler qu'il existe trois catégories d'ETs (Malone et al., 2009) (**Figure 35**) :

- les ETs dits « somatiques », pour lesquels des piARNs les ciblant ne sont produits que dans les cellules folliculaires. Ces piARNs ne présentent donc pas de signature ping-pong. On pourra par exemple citer *gypsy10*, *ZAM*, *gtwin*, *tabor* et *mdg1*.
- les ETs dits « germinaux », pour lesquels des piARNs les ciblant ne sont produits que dans les cellules germinales. Ces piARNs présentent une forte signature ping-pong et sont retrouvés dans les embryons précoces. On retrouve par exemple *HeTA*, *Burdock*, *l'élément I*, *rover* et *roo*.
- les ETs dits « intermédiaires », pour lesquels des piARNs les ciblant sont produits dans les cellules folliculaires et dans les cellules germinales. Ces piARNs présentent une faible signature ping-pong et sont retrouvés dans les embryons précoces. C'est par exemple le cas des ETs *Idefix*, *Gypsy4*, *412*, *blood* et *quasimodo*.

Les ETs « somatiques » tels que *ZAM* ou *gtwin* ne sont insérés que dans le cluster de piARNs *flamenco*, uniquement exprimé dans les cellules folliculaires de l'ovaire (Zanni et al., 2013). Dans la plupart des lignées de drosophile trouvées en laboratoire, aucune insertion de ces ETs n'est présente dans des clusters de piARNs exprimés dans les cellules germinales (Malone et al., 2009). L'expression tissu spécifique des clusters de piARNs conduit donc à une régulation tissu spécifique des ETs. On pourrait ainsi imaginer que cette spécificité de régulation soit dangereuse car cela signifie que les cellules germinales ne produisent pas de piARNs contre tous les ETs présents dans le génome et ne peuvent donc pas réguler certains de ces ETs. Le système est en réalité bien plus efficace que ça. En effet, les ETs « somatiques », pour lesquels les piARNs les ciblant ne sont produits que dans les cellules folliculaires, sont exprimés exclusivement dans les cellules folliculaires de l'ovaire. Ainsi, une mutation des acteurs de la voie des piARNs dans les cellules germinales, comme *Rhino*, n'induit aucune dérégulation des



**Figure 37 : Le système CRISPR-Cas9, le système immunitaire antiviral des procaryotes**  
 le système CRISPR-Cas9 protège les bactéries contre des infections par des phages (virus) et plasmides. Lors d'une infection, des fragments du génome viral appelés « cassettes » sont intégrés à un locus bactérien spécifique, le locus *CRISPR*. Le locus *CRISPR* est transcrit en un long précurseur qui est mûré pour générer des petits ARNs non codants (ARNs guides). En cas de nouvelles infections, les petits ARNs guides, associés à une endonucléase Cas9, ciblent par complémentarité de séquence le génome viral et le dégradent. Ce système permet donc de garder une mémoire des infections passées pour mieux protéger les bactéries face aux futures invasions.

ETs « somatiques » (Li et al., 2009; Mohn et al., 2014; Batki et al., 2019) (**Figure 36**). Il en est de même pour les ETs « germinaux » qui ne sont exprimés que dans les cellules germinales. Les ETs somatiques sont majoritairement des rétrotransposons, et certains possèdent un gène *env* faisant d'eux des rétrovirus endogènes potentiellement infectieux (Péligon et al., 1994b; Leblanc et al., 2000; Tcheressiz et al., 2002; Malone et al., 2009; Olivieri et al., 2010; Sokolova et al., 2020). De ce fait, les rétrotransposons exprimés dans les cellules folliculaires de l'ovaire, par le biais de la formation de particules infectieuses ou par un mécanisme plus passif, peuvent infecter les cellules germinales (Chalvet et al., 1999; Brassat et al., 2006). Après infection, l'ET peut transposer dans le génome des cellules germinales et ces nouvelles insertions seront transmises à la descendance. Ainsi, l'expression des ETs dans les cellules folliculaires peut également être dangereuse pour les cellules germinales. La présence de la voie des piARNs et de *flamenco* dans les cellules folliculaires pourrait donc être une stratégie de protection qui a co-évolué dans ces cellules afin de contrôler efficacement les ETs dans ces cellules somatiques en contact direct avec les cellules germinales.

La voie des piARNs semble donc s'adapter à chaque ETs et évolue constamment afin de protéger efficacement la lignée germinale contre l'activité des ETs. Les clusters de piARNs représentent une « mémoire » de tous les ETs qui doivent être réprimés dans la cellule. Cette voie de répression a souvent été comparée à une sorte de « vaccination génétique » et au système CRISPR-Cas9 présent chez les bactéries. L'existence de ce système immunitaire antiviral chez les procaryotes a été découverte en 2007, même année que celle de la découverte de la voie des piARNs (Barrangou et al., 2007). Lors d'une infection par un virus (phage), des fragments du génome viral sont intégrés sous forme de « cassettes » à un locus spécifique du génome bactérien, le locus *CRISPR* (Clustered Regularly Interspaced Short Palindromic Repeats). Ce locus crée des petits ARNs guides qui sont alors complémentaires au génome viral. Ainsi, en cas de nouvelles infections par ce même virus, les ARNs guides pourront cibler le génome du virus et, *via* leur association avec une endonucléase Cas, induire sa dégradation pour empêcher la réplication virale (**Figure 37**). La voie des piARNs dans les cellules germinales serait alors une sorte de « système immunitaire adaptatif » qui protège ces cellules face aux dangers de la transposition. Ce système s'adapte constamment aux ETs qu'il rencontre et garde une mémoire des invasions passées afin de maintenir la stabilité et l'intégrité du génome des cellules les plus importantes de l'organisme.



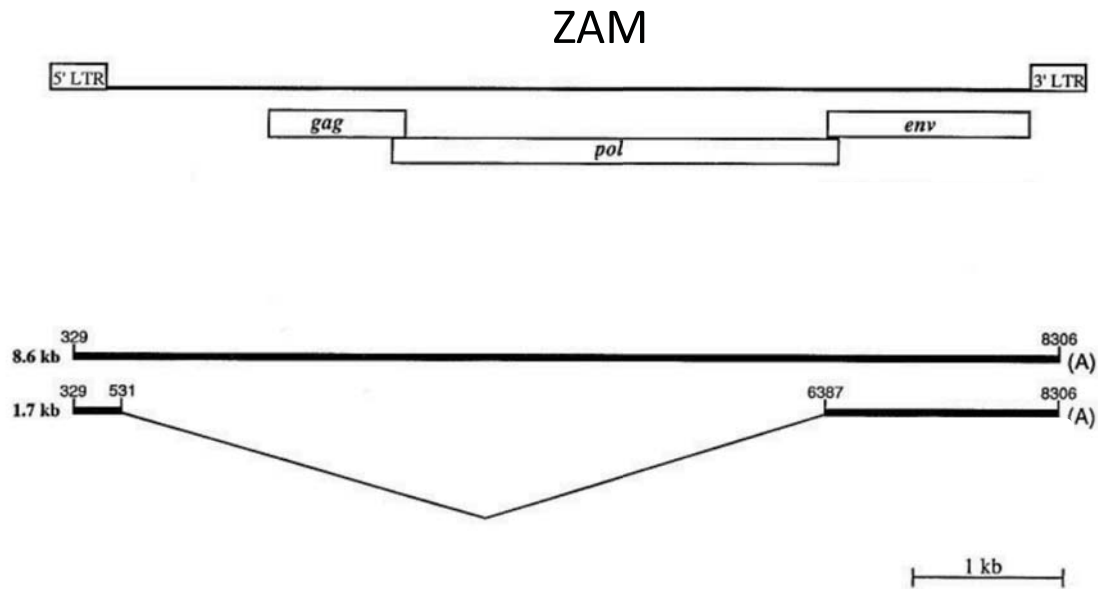
## Partie 3 - ZAM : Un modèle d'étude unique

### I- Découverte des lignées Rev et du rétrovirus endogène ZAM

#### I.1 - Quand « Z'en ai Marre » envahi le génome !

L'histoire de ZAM et de la lignée Rev commence en 1997 au sein de notre équipe qui était anciennement dirigée par Chantal Vaury. L'équipe étudiait l'effet mutagène d'un élément transposable, le *facteur I* (ou « *élément I* »). Une lignée appelée  $w^{IR6}$  avait été établie car celle-ci présentait une coloration des yeux atypique ; cette modification phénotypique était due à l'insertion du *facteur I* dans le premier intron du gène responsable de la coloration des yeux, le gène *white* (Lajoinie et al., 1995). L'insertion du *facteur I* génère l'apparition d'un nouveau site d'épissage et induit la création de messagers aberrants. Ainsi, dans la lignée  $w^{IR6}$ , les messagers épissés permettant la formation d'une protéine White fonctionnelle sont produits en moindre quantité, cela induit une modification de la couleur des yeux : ceux de la lignée  $w^{IR6}$  sont brun-orangé au lieu de la coloration rouge brique des mouches sauvages. Chantal Vaury souhaitait alors générer des révertants de ce phénotype, comme l'avait précédemment fait Barbara McClintock avec la coloration des grains de maïs. Elle réalisa une mutagenèse P (**figure 16**) sur la lignée  $w^{IR6}$  afin d'identifier de potentiel « modificateur de ce phénotype ». Elle a de ce fait établi plusieurs lignées, et c'est au cours de l'entretien de l'une d'entre elles qu'elle observa un mâle révertant aux yeux rouge brique comme en condition sauvage ! Cette lignée fût alors appelée la lignée  $w^{IR6Rev}$  (puis Rev pour simplifier). Les analyses génétiques et moléculaires ont rapidement révélé que la nouvelle mutation responsable de la réversion phénotypique affectait le gène *white* : une nouvelle insertion de 8,6 kb s'était produite 3 kb en amont du site d'initiation de la transcription (TSS) du gène *white*. Pascal Leblanc a ensuite cloné et séquencé des fragments de cette insertion, et l'alignement sur les banques de données a révélé des homologies de séquence avec les gènes *env* d'éléments mobiles de type rétrovirus endogènes. Malgré de nombreuses tentatives, Pascal Leblanc n'a jamais réussi à cloner l'insertion entière (ni personne encore aujourd'hui !) ; il a ainsi baptisé ce nouvel élément « ZAM » pour « Z'en Ai Marre ».

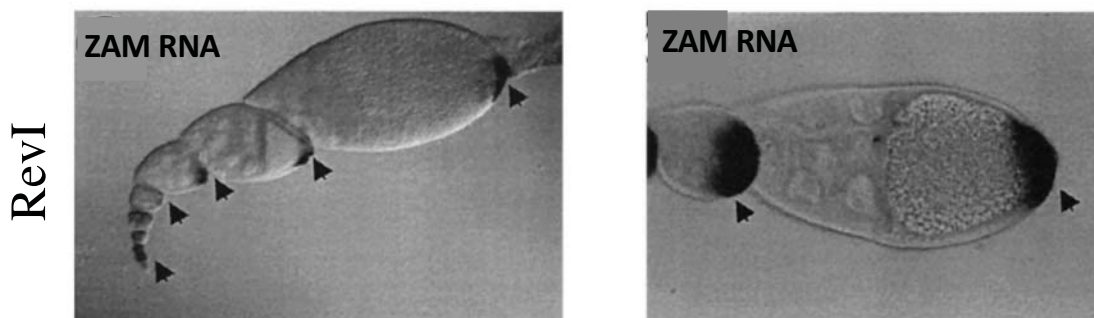
L'utilisation de différentes techniques de clonage a néanmoins permis de séquencer petit à petit complètement ZAM, et ainsi de révéler sa structure (Leblanc et al., 1997). ZAM est donc un



**Figure 38 : Structure génomique et transcrits de ZAM**

ZAM est un rétrovirus endogène de 8,6 kb composé de trois cadres de lecture : *gag*, *pol* et *env*, encadrés par deux séquences LTR. Un transcrit pleine longueur est formé, et un second transcrit subgénomique, résultant d'un épissage, permet l'expression du gène *env*. Les chiffres indiquent les nucléotides d'initiation de la transcription, le site donneur, le site accepteur et la position du site de polyadénylation.

Figure issue de Leblanc et al.,1997



**Figure 39 : Hybridation *in situ* des ARNm de ZAM dans la lignée RevI**

Les ARNm de ZAM sont détectés par hybridation *in situ* dans les ovaires de la lignée RevI. ZAM est exprimé dans les cellules folliculaires des ovaires, son expression débute dès les stades précoces (germarium) et se localise ensuite en postérieur des follicules.

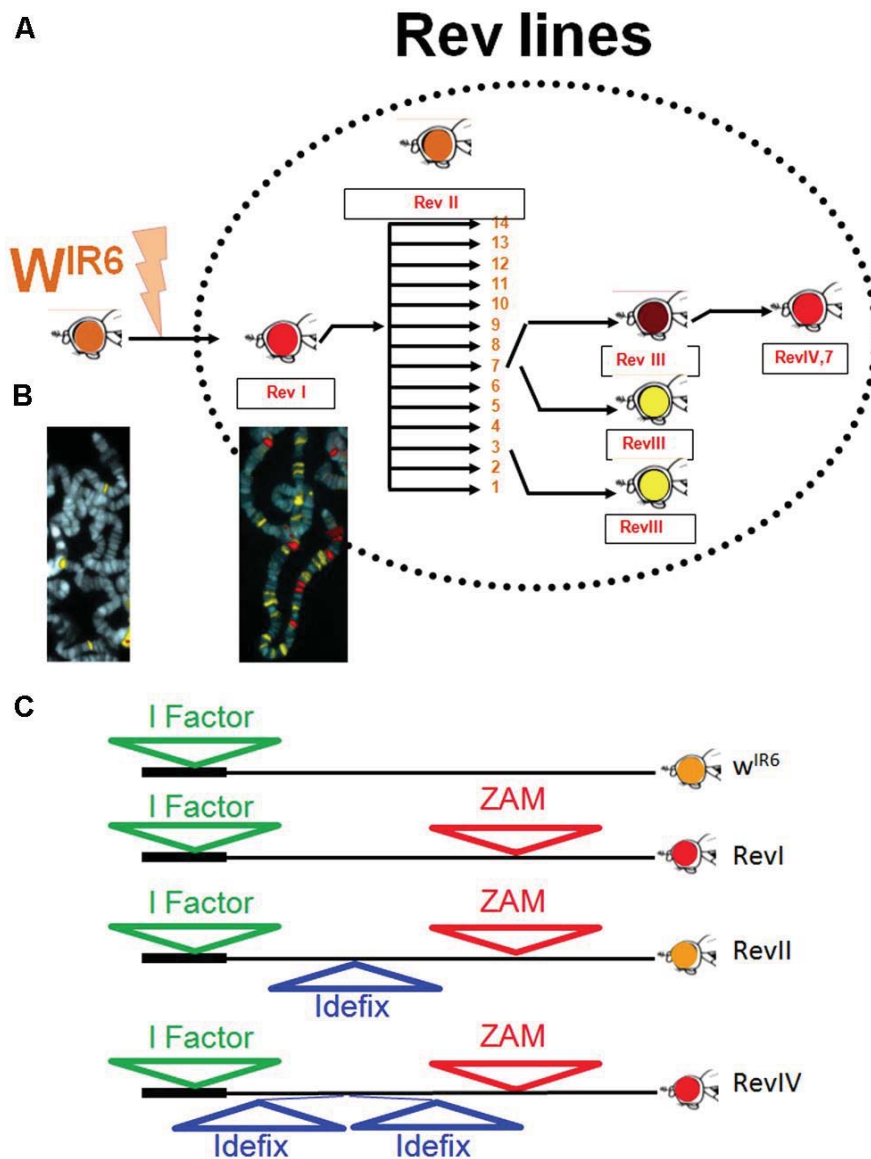
Figure issue de Leblanc et al., 2000

rétrovirus endogène de type *gypsy*. Il est composé de trois cadres de lecture – *gag*, *pol* et *env* - encadrés par deux séquences LTR. *ZAM* possède également un site donneur et un site accepteur d'épissage, et la présence de transcrits épissés subgénomiques de *ZAM* ont été observés. Ces transcrits épissés permettent l'expression du gène *env* (**Figure 38**). De plus, une séquence enhancer a été identifiée dans la région 5' UTR de *ZAM*. Ainsi, dans la lignée Rev, l'insertion de *ZAM* en amont du gène *white* permet d'augmenter la transcription du gène et donc d'induire une forte production de la protéine White responsable de la coloration rouge brique des yeux (Conte et al., 2002). Sophie Desset avait également réalisé des expériences d'hybridation *in situ* sur chromosomes polytènes et de Southern blot sur la lignée Rev, révélant que *ZAM* ne s'était pas seulement inséré dans le gène *white*, puisqu'une vingtaine d'autres copies de *ZAM* ont aussi été trouvées insérées ailleurs, essentiellement sur les bras chromosomiques, alors que la lignée parentale  $w^{IR6}$  ne possède que quelques copies de *ZAM* (Desset et al., 1999) (**Figure 40B**). L'augmentation du nombre de copies de *ZAM* montre que cet élément a été fortement mobilisé dans la lignée Rev. Les analyses de Northern blot ont également révélé que l'augmentation du nombre de copies était corrélée à une augmentation des transcrits de *ZAM* dans la lignée Rev. La présence d'ARNs de *ZAM* a aussi été détectée dans les ovaires de la lignée Rev par hybridation *in situ*. *ZAM* est exprimé dans les cellules folliculaires somatiques de l'ovaire uniquement : son expression débute très précocement dans les cellules somatiques du germarium, puis les transcrits de *ZAM* se localisent au niveau postérieur du follicule à partir des stades 6-7 de l'ovogenèse (Leblanc et al., 2000) (**Figure 39**).

## I.2 - La lignée Rev : une lignée génétiquement instable

La lignée Rev s'est avérée être une lignée génétiquement instable. Au cours de son entretien, de nouveaux mutants pour la couleur des yeux apparaissaient sporadiquement. Les mutants secondaires dérivant de la souche Rev avaient tous la même coloration des yeux brun-orangé retrouvée chez la souche parentale  $w^{IR6}$ . Ces mutants secondaires ont donné naissance aux lignées RevII, et la ligne mère Rev a été renommée la lignée RevI. Onze lignées RevII ont été établies, et toutes possédaient la même mutation : une nouvelle insertion d'un rétrotransposon, nommé *Idefix*, 1,7 kb en amont du TSS du gène *white*. Il a été montré que l'élément *Idefix* porte une séquence insulateur, localisée dans la région U3 du LTR, qui est capable de bloquer le dialogue entre l'enhancer de *ZAM* et le promoteur du gène *white* (Conte et al., 2002; Brassset et al., 2007; Fablet et al., 2009). Les lignées RevII ont également donné naissance à de nouvelles





**Figure 40 : Description des lignées Rev**

- (A) Historique de la lignée Rev obtenue après une mutagenèse P réalisée sur la lignée  $w^{IR6}$ . La lignée  $w^{IR6}$  avait les yeux orange. Dans les lignées Rev, des mutations récurrentes affectant la couleur des yeux ont donné naissance à des lignées dérivées successivement appelées RevI, RevII, RevIII, et RevIV.
- (B) Cartographie FISH sur chromosomes polytènes de copies *ZAM* (rouge) et *Idefix* (jaune) dans les lignées  $w^{IR6}$  (gauche) et Rev (droite). Les ETs *ZAM* et *Idefix* ont envahi le génome de la lignée Rev.
- (C) Structure moléculaire de différents allèles du gène *white* présents dans les lignées Rev. Dans la lignée  $w^{IR6}$ , le *facteur I* est intégré dans le premier intron du gène *white* (en orientation minus). Cette insertion altère la production de la protéine White et induit une coloration orange des yeux. Dans la lignée RevI, un élément *ZAM* s'est inséré 3 kb en amont du site d'initiation de la transcription du gène *white*. *ZAM* possède une activité enhancer qui augmente l'expression du gène *white*. Cela permet de retrouver une coloration rouge brique des yeux. Dans la lignée RevII, une insertion d'*Idefix* entre *ZAM* et le gène *white* s'est produite : l'activité enhancer de *ZAM* est bloquée par une séquence insulateur présente dans l'ET *Idefix*. Les yeux sont donc à nouveau orange. Dans la lignée RevIV, deux insertions d'*Idefix* sont présentes en tête-bêche entre *ZAM* et le gène *white*, bloquant l'action insulatrice d'*Idefix* : *ZAM* peut à nouveau favoriser la transcription du gène *white*, les yeux des drosophiles sont alors rouge brique.

Figure issue de Goriaux et al., 2014b

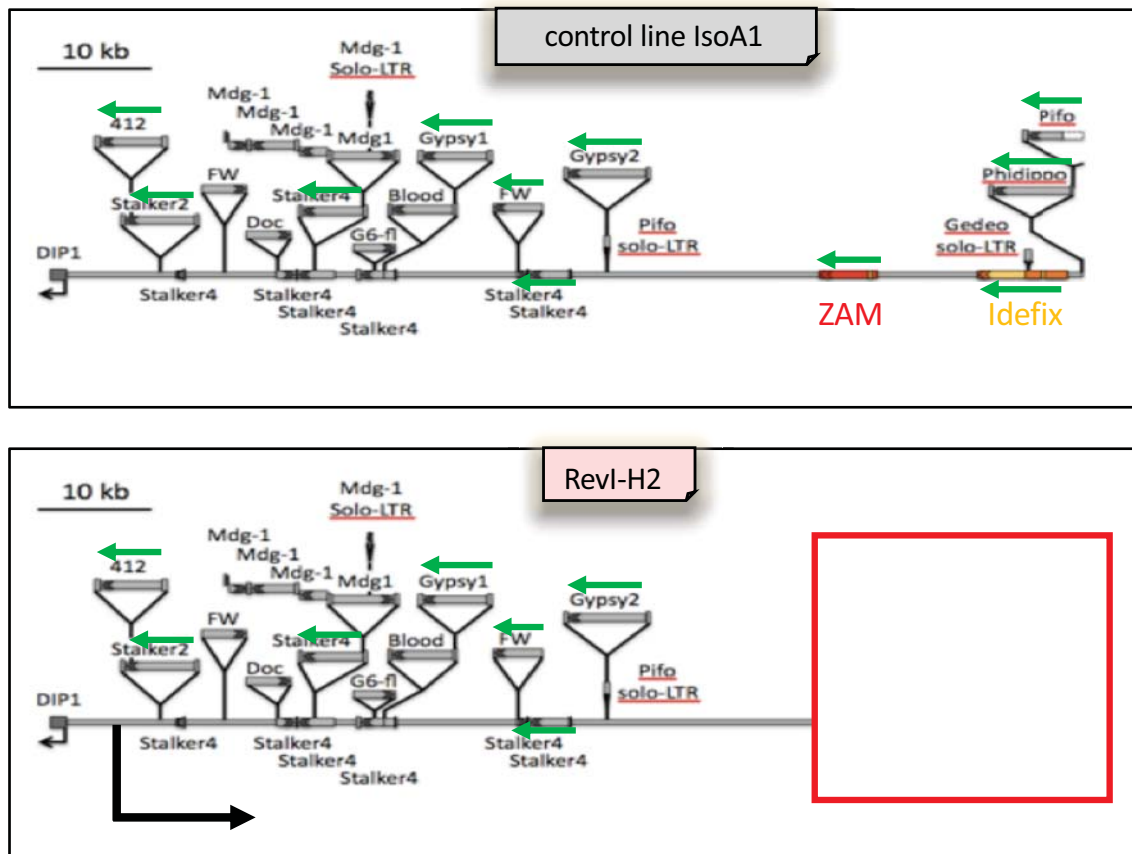
lignées mutantes qui ont tout naturellement été appelées lignées RevIII. Un seul mutant RevIV a été isolé à partir de la lignée RevIII (Revue dans : (Goriaux et al., 2014b)) (**Figure 40**). Les analyses génétiques ont par la suite révélé que la lignée Rev a également subi une invasion massive de l'élément *Idefix* (Desset et al., 1999). Les lignées Rev, aussi appelées « souches instables Rev », ont ensuite été utilisées afin d'étudier l'activité de ces deux éléments transposables : *ZAM* et *Idefix*. Ces lignées constituent un formidable modèle génétique qui a permis de faire des avancées considérables dans la compréhension de la régulation des éléments transposables et, aujourd'hui encore, nous utilisons une de ces lignées au sein de l'équipe afin de répondre à de multiples questions. La lignée RevI-H2 (lignée fille de la lignée Rev) est à la base de mon projet de thèse. Je l'ai observée, analysée et utilisée quasiment quotidiennement durant 4 ans (et, comme beaucoup d'autres avant moi, je l'ai adorée, parfois un peu détestée, mais elle m'a toujours fascinée). Je n'ai jamais vu apparaître dans mon stock de drosophiles RevI-H2 de nouveaux phénotypes mutants... et c'est bien là tout le sujet de ma thèse.

## **II- ZAM : un modèle puissant pour étudier la régulation des ETs**

### **II.1 - Relation directe entre la présence de ZAM dans *flamenco* et sa répression**

Les études génétiques ont ensuite montré que la dérégulation de *ZAM* dans la lignée Rev était causée par une mutation du locus *COM* localisé sur le chromosome X, qui s'avérait être le locus *flamenco* (Desset et al., 2003). Des allèles mutants de *flam* avaient déjà été générés, appelés *flam<sup>kg</sup>* et *flam<sup>bg</sup>* pour lesquels une dérégulation d'un autre rétrovirus endogène, *Gypsy*, avait été observée (Robert et al., 2001). Cette étude a aussi montré un rôle de *flam* dans la morphogénèse de l'épithélium folliculaire puisque ces deux mutants présentent des défauts sévères de développement des follicules. Avant même que la voie des piARNs et les clusters de piARNs ne soient découverts, l'équipe de Chantal VAURY avait déjà émis l'hypothèse que l'insertion de *ZAM* dans un locus particulier (*COM/flam*) permettait sa répression, et que cette répression était potentiellement médiée par l'interférence ARN.

En effet, en 2003, 4 ans avant la découverte de la voie des piARNs, Sophie Desset et ses collaborateurs écrivent en discussion de leur article (Desset et al., 2003) quant au locus « 20A2-3 » qui correspond en fait au locus *COM/flam*:



**Figure 41 : La lignée RevI-H2 présente une large délétion dans la partie proximale du cluster de piARNs *flamenco*.**

Le séquençage génomique de la lignée RevI-H2 a permis de révéler une large délétion dans la partie proximale du cluster de piARNs *flamenco* (la plus proche du centromère). Cette délétion comprend les ETs *ZAM*, *Idefix*, *Gedeo*, *Phidippo* et *Pifo*. Délétion de X:21638367..21684449 et au-delà (génomme de *D. melanogaster* release 5.1)

« *Since the 20A2-3 locus (COM/ flam) is composed mainly of vestiges of transposable elements, aberrant transcripts initiated from these vestiges and homologous to ZAM, Idefix, and possibly other elements can potentially be synthesized and then become a basis for their repression by a mechanism involving cosuppression... Although RNA interference is an attractive hypothesis, it remains to be demonstrated that this cellular defense system is indeed the mechanism accounting for ZAM and Idefix silencing. Regardless of the mechanism involved, however, our results clearly indicate that a control system for retroelements is present in this particular region of the Drosophila genome and prevents high rates of transposition. If altered, this system could easily trigger the reactivation of dormant retrotransposons and lead to sudden bursts of mobilization that affect an individual and thereby, potentially, a species.*»

En 2007, la découverte de la voie des piARNs a permis de confirmer que la répression de ZAM médiée par le locus *flamenco* s'effectue en effet grâce à une voie d'interférence faisant intervenir des piARNs (Brennecke et al., 2007; Desset et al., 2008). Par la suite, des expériences d'amplification par PCR, le séquençage et l'annotation du génome de la drosophile et une analyse détaillée du locus *flamenco*, ont permis de révéler que, dans la lignée Rev, le locus *flamenco* présente une large délétion (X:21638367...21684449 et au-delà, génome de *D. melanogaster* Release 5) (Zanni et al., 2013). Cette délétion s'est produite dans la partie proximale de *flam* et a induit la délétion des ETs ZAM et Idefix (de même que d'autres ETs comme *Phidippo* et *Pifo*) (Zanni et al., 2013) (**Figure 41**). Ainsi, dans la lignée RevI, où ZAM et Idefix ont été massivement mobilisés, ces ETs avaient été éliminés du cluster de piARNs *flamenco*. Cette étude présente pour la première fois la corrélation directe entre la présence d'un ET dans un cluster de piARNs et sa régulation.

## II.2 - ZAM : Un ERV aux capacités d'infection

### II.2. a) Expression tissu spécifique de ZAM

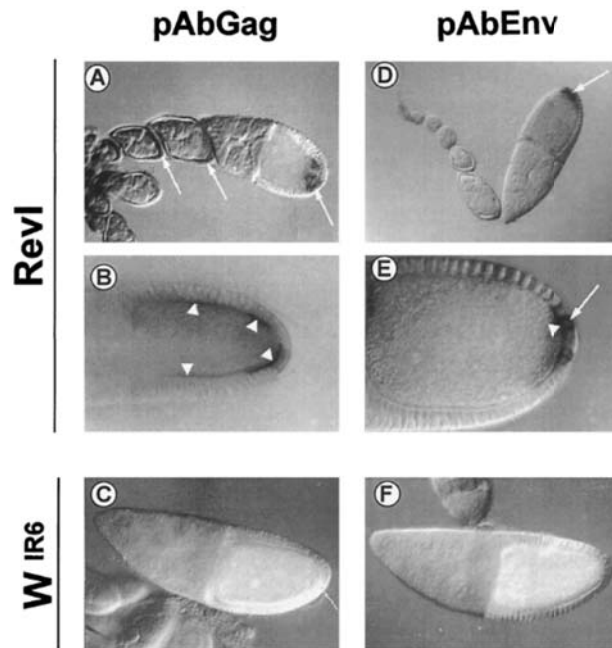
Les expériences d'hybridation *in situ* et l'utilisation de transgènes d'expression ont montré que l'expression de ZAM est restreinte uniquement aux cellules folliculaires de l'ovaire. Aucune expression de ZAM n'est observée dans les cellules germinales (**Figure 39**). En effet, l'expression de ZAM est contrôlée par un facteur de transcription, le facteur Pointed 2 (Meignin



et al., 2004). *pointed 2* est un gène cible de la voie EGFR. Cette voie de signalisation est spécifiquement activée dans les cellules folliculaires de l'ovaire, en partie postérieure des follicules, grâce à la fixation du morphogène Gurken sur son récepteur EGF (« *Epidermal Growth Factor* »). *ZAM* possède un site de fixation de *Pointed 2* localisé dans la partie U3 du LTR de *ZAM* (à la position 79) et sa fixation permet d'activer l'expression de l'ET. En absence de *Pointed 2*, ou d'une voie de signalisation EGFR fonctionnelle, l'expression de *ZAM* est abolie. Cependant, *Pointed 2* est également impliqué dans différents processus de différenciation au cours du développement de l'œil ou des tubules rénaux, son expression n'est donc pas restreinte uniquement aux cellules folliculaires ovariennes (O'Neill et al., 1994; Sudarsan et al., 2002). Or, aucune expression de *ZAM* n'a été détecté dans ces tissus. Cela suggère que *Pointed 2* est nécessaire à l'expression de *ZAM* mais n'est pas suffisante. D'autres cofacteurs sont probablement requis pour permettre l'expression de *ZAM*. Ainsi, l'expression des éléments transposables est contrôlée à deux niveaux différents : 1) un contrôle exercé par les mécanismes de répression cellulaire, tels que la voie des piARNs, et 2) un contrôle exercé par leur dépendance à l'égard des facteurs de transcription et d'autres cofacteurs. Même en absence de répression, les ETs ne peuvent donc s'exprimer que dans les cellules où les facteurs de transcription nécessaires à leur activation transcriptionnelle sont présents. Lorsque les mécanismes de contrôle des ETs sont altérés, *ZAM* est uniquement exprimé dans les cellules folliculaires de l'ovaire. Cette observation est retrouvée pour la plupart des rétrotransposons à LTR. Or, la présence de multiples insertions de *ZAM* dans le génome de la lignée RevI révèle que ces insertions ont été maintenues et transmises à travers les générations : ces insertions se sont donc produites dans les cellules germinales de la lignée RevI. Ces données suggèrent que *ZAM* est transmis des cellules folliculaires aux cellules germinales où il pourra alors être mobilisé.

## II.2. b) Mécanismes utilisés pour la transmission intercellulaire de *ZAM*

Des études se sont donc intéressées aux mécanismes permettant aux rétrotransposons à LTR, exprimés dans les cellules folliculaires, de transposer dans le génome des cellules germinales (voir revue (Yoth et al., 2022) - Partie 2.2). Rappelons que *ZAM* est un rétrovirus endogène et possède par conséquent un cadre de lecture codant pour la protéine Enveloppe, connue pour intervenir dans l'infection intercellulaire chez les virus. Les premières études sur la lignée RevI ont montré que les protéines Gag et Env de *ZAM* étaient produites dans les cellules folliculaires des ovaires. Gag semble exprimée dans toutes les cellules somatiques où l'ARN de *ZAM* est

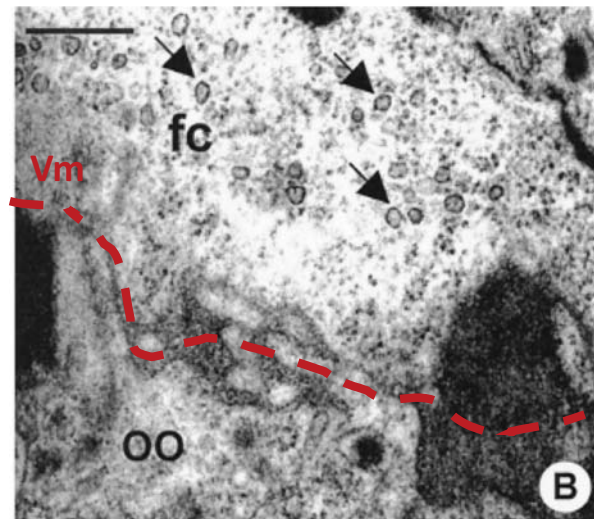


**Figure 42 : Expression des protéines Gag et Env dans les ovaires des lignées RevI et  $w^{IR6}$**

L'immunomarquage a été réalisé avec des anticorps polyclonaux pAbGag (A, B et C) et pAbEnv (D, E et F) sur des ovaires de la lignée RevI et la lignée contrôle  $w^{IR6}$ .

A) Un marquage de la protéine Gag est détecté dans chaque follicule dans quelques cellules folliculaires somatiques qui entourent l'ovocyte. (B) Grossissement de la partie postérieure d'un follicule de stade 10 précoce. Le marquage est indiqué par des flèches. (C) Follicule de stade 10 d'une femelle  $w^{IR6}$ . Aucune protéine Gag n'est détectée par l'anticorps pAbGag. (D) Un marquage intense de Env est observé dans quelques cellules folliculaires localisées en postérieure des follicules de stade 10 (flèche) de la lignée RevI. (E) Grossissement de la région postérieure du follicule de stade 10 de la lignée RevI. (F) Follicule de stade 10 d'une femelle  $w^{IR6}$ . Aucune protéine Env n'est détectée par l'anticorps pAbEnv.

Figure issue de Leblanc et al., 2000



**Figure 43 : Image de microscopie confocale révélant la présence de particules virales de ZAM dans les follicules ovariens de la lignée RevI**

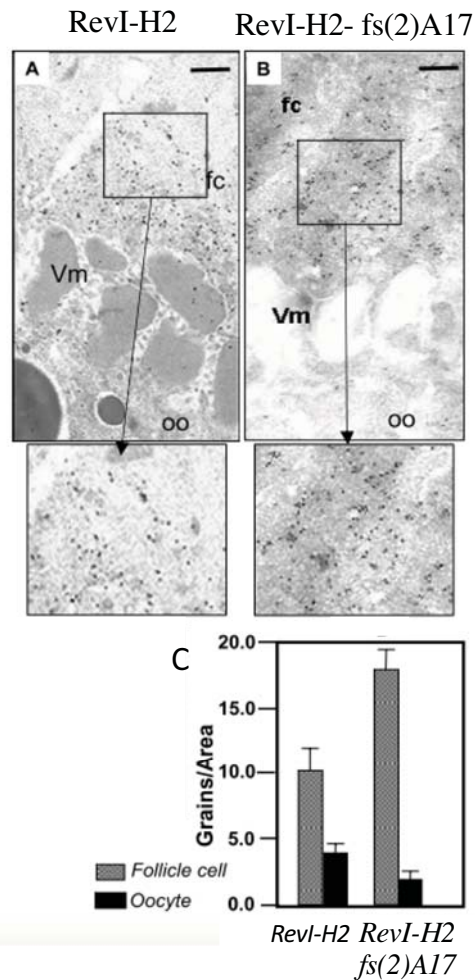
Des acquisitions d'ultrastructures de follicules de stade 9 ont révélé la présence de nombreuses particules virales d'environ 45 nm dans les cellules folliculaires de la lignée RevI (montrées par des flèches). Ces particules s'accumulent en bordure de la membrane vitelline (pointillés rouges).

Fc = cellules folliculaires, oo = ovocyte et Mc = Membrane vitelline séparant les cellules folliculaires (fc) de l'ovocyte (oo). Échelle = 0,3  $\mu$ m

Figure issue de Leblanc et al., 2000

déecté tandis que la protéine Env est uniquement détectée dans un petit nombre de cellules folliculaires localisées en postérieur des follicules à partir des stades 9/10 de l'ovogenèse. Aucune expression de ces deux protéines n'est détectée dans la lignée  $w^{IR6}$  (Leblanc et al., 2000) (**Figure 42**). Tous les composants nécessaires pour la formation de pseudo-particules virales sont donc exprimés dans les cellules folliculaires postérieures de la lignée RevI. Pascal Leblanc et ses collaborateurs ont alors recherché la présence de particules virales dans les ovaires grâce à des expériences de microscopie électronique. Des particules ellipsoïdales ou annulaires d'un diamètre approximatif de 45 nm ont été observés dans les cellules folliculaires au niveau postérieur des follicules de stades 8 à 10 (**Figure 43**). Aucune particule n'a été observée dans la lignée contrôle  $w^{IR6}$ . Cela suggère que, dans les cellules folliculaires, *ZAM* pourrait former des pseudo-particules virales qui auraient la capacité d'infecter les cellules germinales. Cependant, aucune particule bourgeonnante à partir des cellules folliculaires n'a été observée. En revanche, des particules de *ZAM* ont été observées en bordure des cellules folliculaires en association avec les précurseurs de la membrane vitelline. Des particules ont également été observées dans le cytoplasme de l'ovocyte, à l'intérieur des Yolk granules qui sont formées à la suite de l'endocytose des protéines Yolk synthétisées par les cellules folliculaires et le corps gras (schéma de la vitellogenèse **Figure 19**). Il est intéressant de noter que l'incorporation de la vitellogénine se produit à partir du stade 8-9, stade où une très forte expression de l'ARN de *ZAM*, mais aussi des protéines Gag et Env, a été observée dans la lignée RevI. Ainsi, *ZAM* pourrait utiliser le trafic de vitellogénine pour atteindre l'ovocyte. Emilie BRASSET a donc ensuite analysé le transit des particules de *ZAM* dans un contexte mutant pour un des composants des Yolk granules nécessaires à l'incorporation de vitellogénine (Brasset et al., 2006). Des expériences de microscopie électroniques en utilisant des particules d'or associées à un anticorps anti-Gag de *ZAM* ont montré que, lorsque le trafic de vitellogénine est bloqué, les particules de *ZAM* s'accumulent fortement à la membrane des cellules folliculaires (**Figure 44**). Ce résultat suggère qu'en absence d'endocytose de la vitellogénine, les particules de *ZAM* sont moins transférées à l'ovocyte et s'accumulent donc dans les cellules folliculaires. *ZAM* pourrait alors utiliser une voie d'endocytose/exocytose préexistante au sein des ovaires pour transiter des cellules folliculaires vers l'ovocyte. Mais alors, quel est le rôle de l'enveloppe dans la transmission intercellulaire de *ZAM*? *ZAM* peut-il infecter les cellules germinales comme le ferait un rétrovirus ou utilise-t-il seulement une voie de transmission passive comme le trafic de vitellogénine? Toutes ces questions restent à l'heure actuelle non résolues. Il n'est pas exclu que plusieurs mécanismes soient utilisés pour le transit de *ZAM*. De





**Figure 44 : Détection de particules d'or couplées à un anticorps anti-Gag de ZAM en microscopie électronique dans un contexte RevI et RevI mutant pour l'endocytose de vitellogénine**

Dans un contexte RevI, de nombreuses particules de ZAM, révélées par un anticorps anti-Gag, ont été détectées proche de la membrane vitelline dans les cellules folliculaires, dans la membrane vitelline et, en moindre mesure, dans le cytoplasme de l'ovocyte. B) Chez les drosophiles RevI mutantes pour l'endocytose des protéines Yolk (*fs(2)A17*), les particules virales s'accumulent dans l'épithélium folliculaire et ne sont plus retrouvées dans la membrane vitelline, ni dans l'ovocyte. C) L'histogramme montre la répartition des particules de ZAM détectées dans les lignées RevI-H2 et RevI-H2 *fs(2)A17* mutante pour la voie d'endocytose de la vitellogénine. Les particules ont été comptées dans les cellules folliculaires et, dans l'ovocyte, les particules de ZAM s'accumulent dans les cellules folliculaires en condition mutante pour le trafic de vitellogénine.

Figure adaptée de Brasslet et al., 2006

plus, le transit de *ZAM* via les mécanismes d'endocytose de la vitellogénine pourrait nécessiter que les particules de *ZAM* soient enveloppées. Des expériences supplémentaires sont nécessaires pour comprendre pleinement les mécanismes utilisés par *ZAM* pour transiter vers les cellules germinales. Enfin, le stade du développement où l'intégration dans le génome des cellules germinales se produit reste également énigmatique. Cette intégration pourrait avoir lieu dans l'ovocyte ou dans l'embryon précoce.

### **II.3 - Pourquoi utiliser *ZAM* comme modèle d'étude ?**

*ZAM* sera utilisé comme modèle d'étude pour mon projet de thèse, mais les données obtenues seront transposables aux ETs de façon plus générale. Finalement, pourquoi avoir choisi *ZAM* spécifiquement pour cette étude ? Voici un petit récapitulatif expliquant ce qui fait de *ZAM* un modèle d'étude unique.

#### **1) Le mécanisme de contrôle de *ZAM* est « simple » et bien connu.**

Les données obtenues dans l'équipe ces dernières années ont permis d'avoir une vision assez précise des mécanismes contrôlant l'expression de *ZAM* dans les ovaires de *Drosophila melanogaster*. Son profil d'expression et les facteurs impliqués dans son activation transcriptionnelle sont en grande partie connus. L'avantage majeur de *ZAM* est que cet ET n'est contrôlé que par un seul cluster de piARNs : *flamenco*. Aucune autre insertion de *ZAM* n'a été retrouvée dans aucun autre cluster de piARNs présent dans le génome de *Drosophila melanogaster*. Ce phénomène est rare puisque la plupart des ETs sont insérés dans plusieurs clusters de piARNs, créant une redondance. Ainsi, lorsque *flamenco* est affecté, la répression de *ZAM* est directement affectée.

#### **2) Nous possédons dans l'équipe des outils puissants pour l'étude de *ZAM*.**

La lignée RevI-H2 est un outil génétique puissant puisque cette lignée présente une délétion dans le cluster de piARNs *flamenco* induisant une dérégulation de *ZAM* dans les ovaires. De plus, nous possédons des outils moléculaires très pratiques pour suivre l'expression et la régulation de *ZAM* comme des transgènes senseurs, des sondes permettant de détecter la présence d'ARN ou d'ADN de *ZAM*, ou bien encore des anticorps dirigés contre les protéines Gag et Env de *ZAM*.



### 3) *ZAM* est un rétrovirus endogène aux propriétés d'invasion.

*ZAM* est exprimé dans les cellules folliculaires de l'ovaire et il a la capacité d'envahir les cellules germinales. Les mécanismes utilisés par *ZAM* pour envahir les cellules germinales restent à définir clairement. À l'heure actuelle, les mécanismes précis permettant aux ERVs de transiter entre les cellules, ainsi que les réponses cellulaires mises en place à la suite de cette invasion, restent à explorer. De manière plus globale, les études sur *ZAM* pourraient donc permettre de mieux comprendre la relation complexe existant entre les ERVs et leurs hôtes.

## OBJECTIFS DE THÈSE

L'objectif général de ma thèse était d'étudier la réactivation d'un rétrovirus endogène dans les ovaires de *Drosophila melanogaster*. Nous avons tenté de répondre à plusieurs questions :

- Quel est le profil d'expression d'un rétrovirus endogène tel que *ZAM* ? Cet ET est-il réellement capable d'envahir les cellules germinales ?
- Comment *ZAM* transite-t-il des cellules folliculaires vers les cellules germinales ?
- Comment les cellules germinales se protègent-elles face à l'invasion d'un rétrovirus endogène ? Existe-t-il des systèmes de protection mis en place dans les ovaires permettant de protéger le génome des cellules germinales contre une mobilisation massive d'un rétrovirus endogène venant de cellules voisines ?



# RÉSULTATS



## **Partie 1 : L'insertion d'un rétrovirus endogène somatique dans un cluster de piARNs à expression germinale permet de protéger les cellules germinales de nouvelles invasions.**

Cette première étude a pour but d'étudier comment les cellules germinales se protègent face à l'invasion d'un rétrovirus endogène. Pour répondre à cette question, nous avons utilisé la lignée RevI-H2 qui présente une large délétion dans la partie 3' du cluster de piARNs *flamenco*. Cette délétion induit une dérégulation de *ZAM* dans les cellules somatiques folliculaires des ovaires et ce rétrovirus endogène semble capable d'envahir les cellules germinales et de s'intégrer dans le génome de ces cellules. Avant mon arrivée, Émilie Brassset avait réalisé un séquençage des petits ARNs présents dans les ovaires de la lignée RevI-H2 et avait ainsi découvert que des piARNs ciblant *ZAM* étaient produits dans ces ovaires alors que *ZAM* était absent de son cluster régulateur *flamenco*. Cette observation inattendue fût à l'origine du projet, et c'est pourquoi nous avons cherché à comprendre l'origine et le rôle de ces piARNs.

En collaboration avec Céline Duc, qui était Post-Doctorante dans l'équipe d'Émilie Brassset, nous avons montré que les piARNs ciblant *ZAM* produits dans les ovaires de la lignée RevI-H2 sont sens et antisens, et ils présentent une signature ping-pong. Cette signature démontre que ces piARNs dirigés contre *ZAM* sont produits dans les cellules germinales, une production germinale spécifique de la lignée RevI-H2 qui n'avait jamais été observée dans aucune autre lignée de drosophile auparavant. Les piARNs ciblant *ZAM* produits dans les cellules germinales de la lignée RevI-H2 permettent la répression d'un transgène senseur contenant un fragment de *ZAM* fusionné à la GFP dans les cellules germinales. Cette production de piARNs est corrélée avec la présence de *ZAM* dans un cluster de piARNs « dual-strand », insertion spécifique de la lignée RevI-H2. Nous avons alors émis l'hypothèse que la lignée germinale peut se protéger de l'invasion d'un rétrovirus endogène en piégeant cet ET dans un cluster de piARNs exprimé dans les cellules germinales. Cette insertion permet la production de piARNs ciblant *ZAM* dans les cellules germinales.

ARTICLE 1






RESEARCH

Open Access



# Trapping a somatic endogenous retrovirus into a germline piRNA cluster immunizes the germline against further invasion

Céline Duc<sup>1,2†</sup>, Marianne Yoth<sup>1†</sup>, Silke Jensen<sup>1</sup>, Nolwenn Mouniée<sup>1</sup>, Casey M. Bergman<sup>3</sup>, Chantal Vaury<sup>1\*</sup> and Emilie Brassat<sup>1\*</sup> 

## Abstract

**Background:** For species survival, the germline must faithfully transmit genetic information to the progeny. Transposable elements (TEs) constitute a significant threat to genome stability due to their mobility. In the metazoan germline, their mobilization is limited by a class of small RNAs called PIWI-interacting RNAs (piRNAs) produced by dedicated genomic loci called piRNA clusters. Although the piRNA pathway is an adaptive genomic immunity system, it remains unclear how the germline gains protection from a new transposon invasion.

**Results:** To address this question, we analyze *Drosophila melanogaster* lines harboring a deletion within flamenco, a major piRNA cluster specifically expressed in somatic follicular cells. This deletion leads to derepression of the retrotransposon *ZAM* in the somatic follicular cells and subsequent germline genome invasion. In this mutant line, we identify *de novo* production of sense and antisense *ZAM*-derived piRNAs that display a germinal molecular signature. These piRNAs originated from a new *ZAM* insertion into a germline dual-strand piRNA cluster and silence *ZAM* expression specifically in germ cells. Finally, we find that *ZAM* trapping in a germinal piRNA cluster is a frequent event that occurs early during the isolation of the mutant line.

**Conclusions:** Transposons can hijack the host developmental process to propagate whenever their silencing is lost. Here, we show that the germline can protect itself by trapping invading somatic-specific TEs into germline piRNA clusters. This is the first demonstration of “auto-immunization” of a germline endangered by mobilization of a surrounding somatic TE.

**Keywords:** Transposable elements, piRNAs, piRNA cluster, Genome stability, Inheritance, *Drosophila*, Germline

## Background

Germ cells are the only cell type within multicellular organisms that can transfer genetic and epigenetic material to the offspring. Due to their capacity to move, transposable elements (TEs), a major component of eukaryotic genomes, constitute a significant threat to the germline genome integrity [1–3]. Indeed, their mobilization could lead to gene disruption or chromosomal rearrangements. To limit TE mobilization in the germline, a class of small RNAs of 23 to 29 nucleotides (nt) in length, called PIWI-interacting RNA (piRNAs), are expressed in the reproductive tissue and

silence TE activity via homology-dependent mechanisms [4–7].

The piRNA pathway has been extensively studied in the *Drosophila melanogaster* ovary that comprises about 16 ovarioles, each of which contains a succession of follicles composed of germline and somatic follicular cells [8]. In *D. melanogaster*, piRNAs are encoded by dedicated genomic loci that are called piRNA clusters [9]. These clusters are composed of full length or truncated TEs that define the repertoire of elements that are recognized and silenced by the piRNA machinery. Two classes of piRNA clusters have been defined on the basis of their transcriptional properties: (i) unidirectional or uni-strand and (ii) bidirectional or dual-strand piRNA clusters [9]. Unidirectional clusters are expressed predominantly in somatic follicular cells of ovaries, while

\* Correspondence: [chantal.vaury@uca.fr](mailto:chantal.vaury@uca.fr); [emilie.brassat@uca.fr](mailto:emilie.brassat@uca.fr)

† Céline Duc and Marianne Yoth contributed equally to this work.

<sup>1</sup>GREd, Université Clermont Auvergne, CNRS, INSERM, Faculté de Médecine, 63000 Clermont-Ferrand, France

Full list of author information is available at the end of the article



bidirectional clusters are transcribed in germline cells. Therefore, TEs are silenced in both cell types by piRNAs via different mechanisms [10, 11]. Transcription of piRNA clusters produces long piRNA precursors that are diced into piRNAs. In germline cells, these piRNAs are loaded on the Piwi protein to form a complex that triggers TE transcriptional silencing [12]. In addition to Piwi, two other PIWI-family proteins, Aub and Ago3, participate in the post-transcriptional control of TEs. They act to amplify the piRNA pool by a mechanism called the ping-pong cycle [9, 13]. Moreover, Aub- and Ago3-bound piRNAs are deposited in the embryo to ensure the re-initiation of piRNA clusters and efficient TE control in the offspring germline [14–16]. In somatic follicular cells, whose genome does not contribute to the next generation but which could be the origin of transposon invasion, a simplified version of the piRNA pathway is active because only the Piwi protein is expressed [17, 18]. The tissue-specific expression of piRNA clusters, which contain different TE sequences, suggests a tissue-specific regulation of certain classes of elements. For instance, *flamenco* is the best characterized piRNA cluster predominantly expressed in somatic follicular cells. The *flamenco* locus is a uni-strand cluster that extends over more than 180 kb and is located in the pericentromeric heterochromatin of *D. melanogaster* X chromosome [19–21]. Most TEs inserted in *flamenco* belong to the long terminal repeat (LTR) group of retrotransposons and are oriented opposite to the cluster transcription direction. Across the entire spectrum of transposons described in *flamenco*, maternally deposited piRNAs targeting some TEs, such as *ZAM* or *gypsy*, are underrepresented in the embryonic piRNA pool [17]. This suggests that piRNAs matching these TEs are not produced by any germline piRNA cluster and that they originate from the main somatic piRNA cluster, *flamenco*. Thus, these TEs should be exclusively silenced in somatic follicular cells. In the absence of efficient silencing of these TEs in somatic follicular cells, the oocyte genome is exposed to internal threats. Indeed, when the silencing of *ZAM* or *gypsy* is released in somatic follicular cells, these (and potentially other) retrovirus-like TEs can infect germline cells [22, 23]. Therefore, the stability of the germline genome requires efficient silencing of TEs also in somatic follicular cells.

The piRNA pathway has often been compared to an adaptive immune system, because it conveys the memory of previous transposon invasions by storing TE sequence information within piRNA clusters [17]. This model leads to several major questions. Particularly, it is not known whether some TE classes are regulated only in specific tissues and whether and how germ cells can counteract TE invasion from the surrounding somatic follicular cells. To gain insights into these issues, we

used *D. melanogaster* lines in which *ZAM* expression is either silenced (i.e., “stable,” w<sup>IR6</sup> line) or derepressed (i.e., “unstable,” RevI-H2 also named RevI in [21]). The RevI-H2 line was derived from the w<sup>IR6</sup> line after P-mediated mutagenesis [24, 25] and displays a large deletion of the proximal (i.e., the region closest to the centromere) part of *flamenco* corresponding to the region containing its only *ZAM* insertion [26]. This suggests a tight correlation between the presence of *ZAM* in the *flamenco* locus and the repression of all functional genomic copies of *ZAM* in the somatic follicular cells [26].

Here, we found that in the w<sup>IR6</sup> ovaries, *ZAM* was silenced only in follicular cells with an absence of a germline-specific silencing mechanism. Conversely, in the RevI-H2 line, *ZAM* was derepressed in somatic follicular cells and silenced in the germline following its rapid trapping into a germline piRNA cluster. This represents an efficient mechanism of protection against TE invasion from the surrounding somatic tissues.

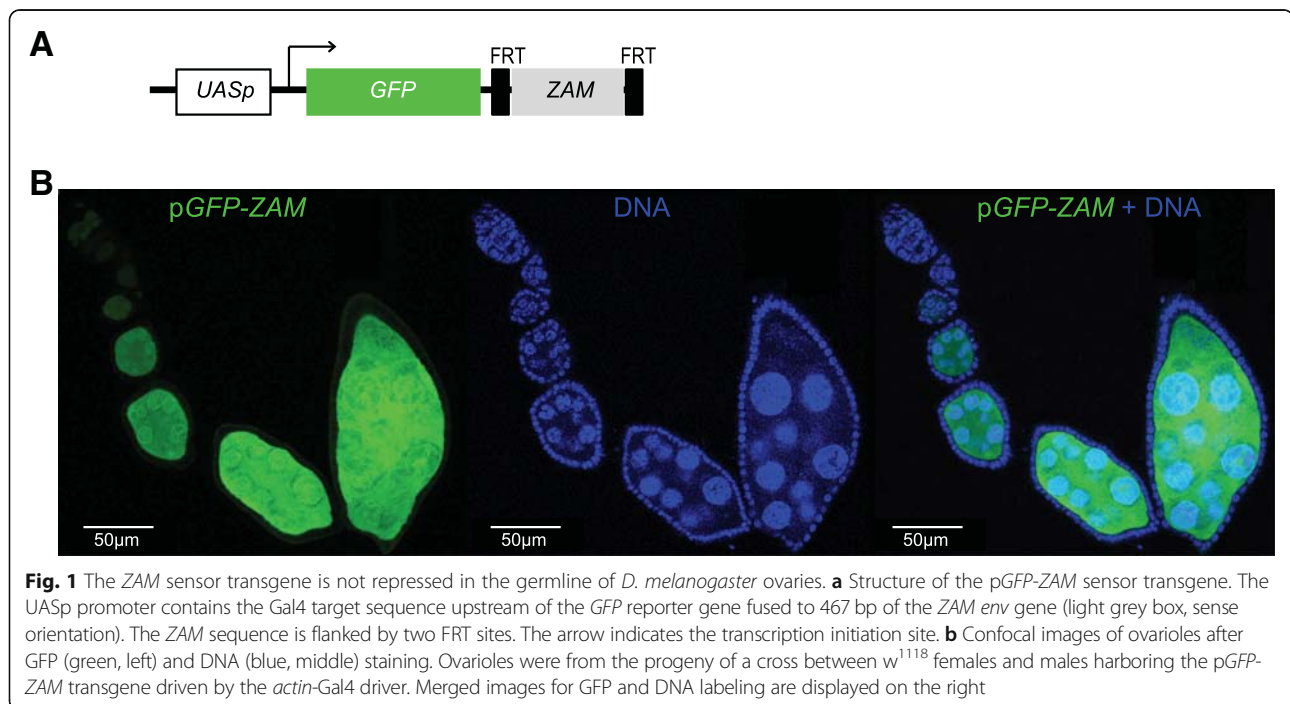
## Results

### *ZAM* is silenced in a tissue-specific manner

Previous studies have reported that distinct tissue-specific piRNA populations are expressed in the germline and in somatic follicular cells [17]. This suggests a tissue-specific repression of TEs. Here, we used *ZAM* to monitor the germline capacity to repress TEs for which no germline piRNA is produced. *ZAM* is a prototypic somatic TE [17, 27] and *ZAM*-derived piRNAs are highly depleted in the early embryonic piRNA population that mirrors the germline piRNA population [17]. To monitor *ZAM* repression, we generated a sensor transgene that expresses the *GFP* reporter gene under the control of an inducible Upstream Activation Sequence promoter (UASp) and harbors a *ZAM* fragment in its 3'UTR (p*GFP-ZAM*) (Fig. 1a). Transgene expression analysis in both somatic and germline cells using the *actin*-Gal4 driver showed that p*GFP-ZAM* was completely silenced in somatic cells (Fig. 1b). This indicated that *ZAM*-derived piRNAs, which are produced by *flamenco* in these cells, targeted the transgene and efficiently guided its silencing. Conversely, in germline cells, its expression was not inhibited, as shown by the strong GFP signal (Fig. 1b). This showed that *ZAM* silencing is specific to somatic follicular cells suggesting that it is mediated by the somatic *flamenco* cluster, as shown by genetic evidence [21, 28] and that there are no *ZAM*-derived piRNAs from any germline piRNA cluster.

### *ZAM*-derived piRNAs are produced in the germline in response to follicular cell instability

*ZAM* silencing release in somatic follicular cells could expose the oocyte genome to internal threats arising from the surrounding follicular cells. To analyze how

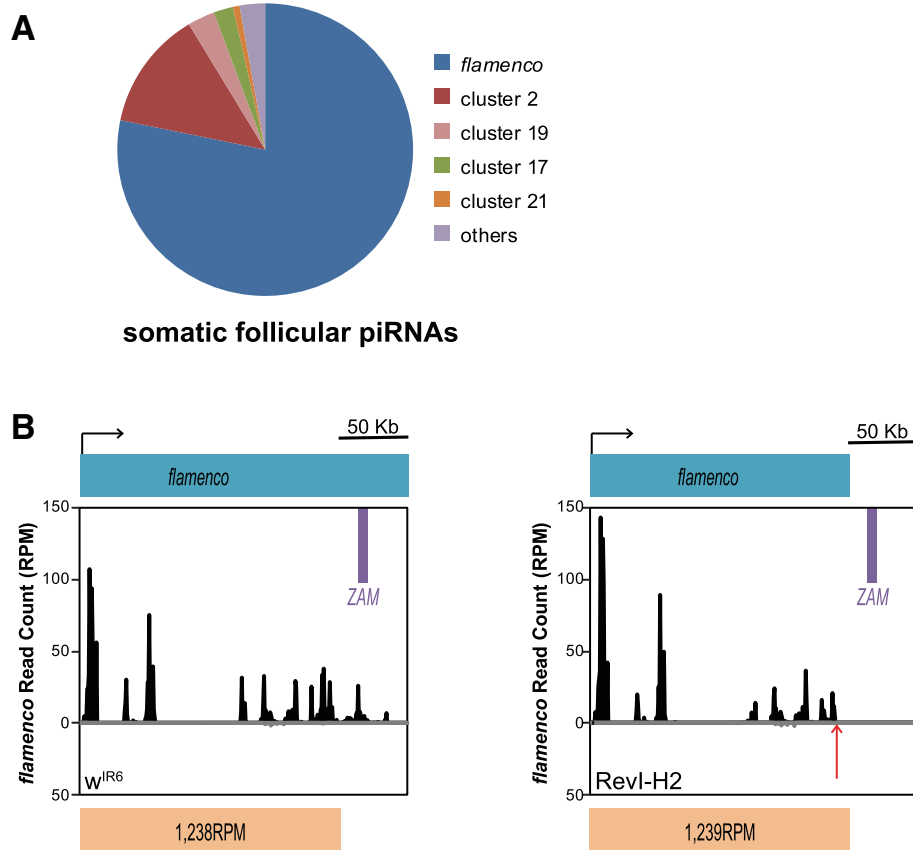


the germline may protect itself against TE mobilization from the surrounding follicular cells, we used RevI-H2 flies harboring a deletion in the proximal part of *flamenco* [21] that eliminates the region in which ZAM is inserted [26] (Additional file 1: Figure S1A), but does not affect germline development. In contrast, as the *flamenco* piRNA cluster is the main source of piRNAs (78%) produced in somatic follicular cells (Fig. 2a), other mutations affecting *flamenco* expression, such as *flamKG* and *flamBG*, lead to disruption of piRNA production, but also to impairment of ovarian germline stem cell differentiation and division, thus preventing further analysis of how the germline might respond to any TE mobilization initiated in the surrounding follicular cells [28]. In addition, the close relationship between the parental  $w^{IR6}$  and derived RevI-H2 allowed us to closely control for the genetic background.

To determine whether the *flamenco* deletion in RevI-H2 was associated with changes in piRNA production at this locus, we sequenced and compared ovarian small RNAs from the RevI-H2 line and the parental  $w^{IR6}$  line. This highlighted the complete loss of piRNAs produced at the deleted locus in RevI-H2 samples compared with the  $w^{IR6}$  control line (Fig. 2b). Conversely, the global production of piRNAs uniquely mapping to the *flamenco* locus upstream of the deletion was not affected by the deletion (1,238 and 1,239 Reads Per Million for the RevI-H2 and  $w^{IR6}$  samples, respectively) (Fig. 2b and Additional file 1: Figure S1B).

As expected from earlier studies, in the  $w^{IR6}$  control line, 88% of ZAM-derived piRNAs mapped to piRNA

clusters [9] (without mismatch) and 86% of them mapped the *flamenco* locus (Fig. 3a). Detailed analysis showed that piRNAs were predominantly antisense to the ZAM sequence (Fig. 3b), in agreement with ZAM insertion in the antisense orientation relative to *flamenco* transcription orientation (Additional file 1: Figure S1A) [26]. Moreover, 90% of ZAM-derived piRNAs displayed a uridine bias at the 5' end, a feature of mature primary piRNAs (Fig. 3c). As ZAM is absent from the RevI-H2 *flamenco* locus and is derepressed in somatic follicular cells of RevI-H2 ovaries [21], we hypothesized that production of ZAM-derived piRNAs would be abolished in RevI-H2 ovaries. However, sequencing of ovarian small RNAs revealed that antisense ZAM-derived piRNAs were considerably increased (three times) in RevI-H2 ovaries compared with  $w^{IR6}$  ovaries (Fig. 3d). Moreover, many more ZAM-derived sense piRNAs were produced in RevI-H2 than in  $w^{IR6}$  ovaries (Fig. 3e). To identify the cellular origin of these ZAM-derived piRNAs, we performed a nucleotide profile analysis. We identified a bias for uracil at the first position (1U) and for adenine at the tenth position (10A) (Fig. 3f). This is a typical feature of piRNAs generated by the ping-pong amplification mechanism that occurs exclusively in germline cells. We then checked the ping-pong signature (i.e., a 10-nucleotide overlap between sense and antisense pairs of ZAM-derived piRNAs) [9, 13] and found significant enrichment for this signature in the RevI-H2 line, but not in the parental  $w^{IR6}$  line (Fig. 3g). Moreover, in RevI-H2 samples, 34% of the ZAM-derived piRNAs possessed ping-pong partners (PPP), i.e., piRNAs which



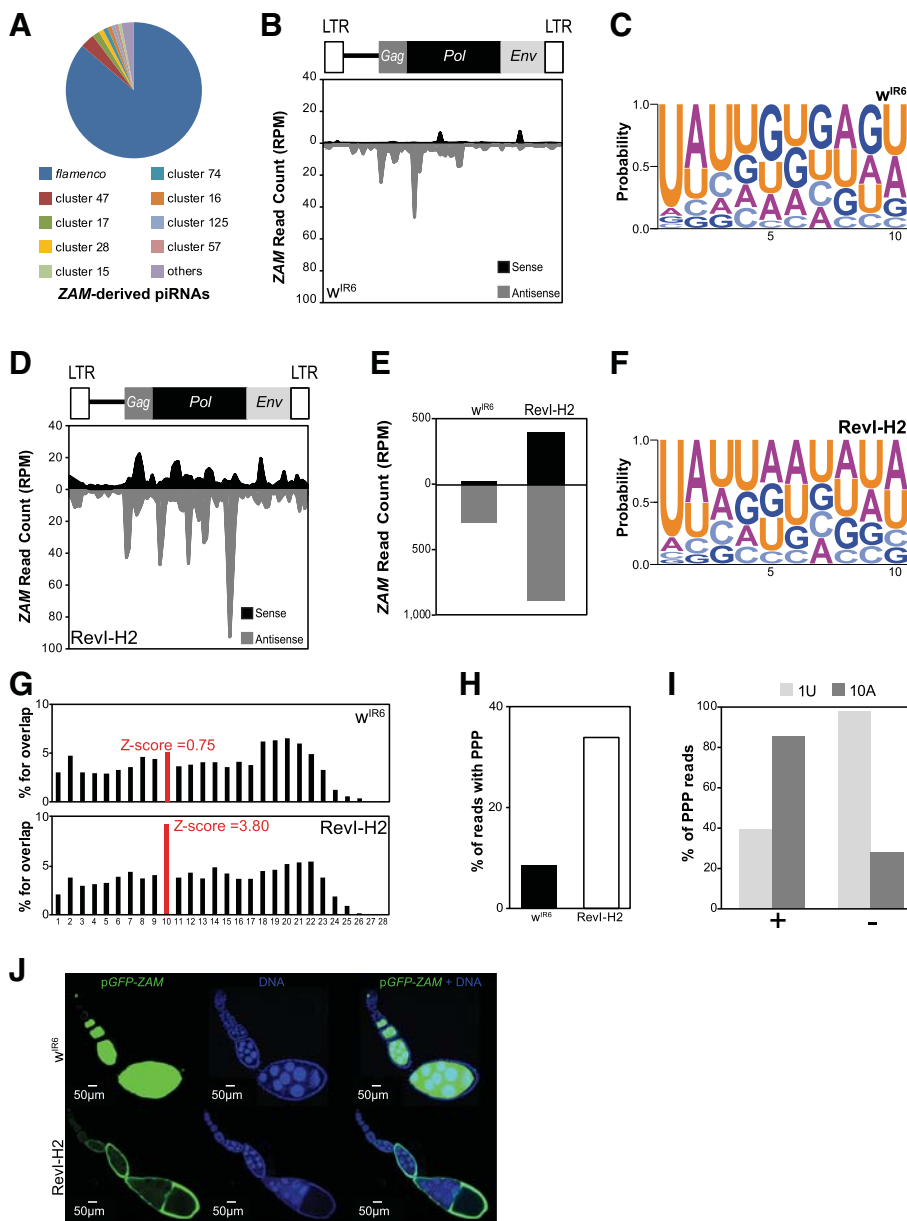
**Fig. 2** Deletion of some TE fragments in *flamenco* does not impair the global piRNA production from this piRNA cluster. **a** Pie chart showing the proportion of unique piRNAs that map to each of the 142 piRNA clusters in ovarian somatic sheath cells (no mismatch allowed, piRNA clusters defined as in [9]). **b** Density profile of unique piRNAs from the  $w^{IR6}$  (left) and RevI-H2 (right) lines that map to the *flamenco* piRNA cluster. Sense and antisense reads are presented in black and grey, respectively. Almost no antisense reads map to the *flamenco* piRNA cluster. *ZAM* location in *flamenco* is displayed by a purple box. The *flamenco* deletion distal break-point in RevI-H2 [26] (Additional file 1: Figure S1B) is indicated by a red arrow and the sense of transcription by a black arrow. The count of piRNA reads per million (RPM) mapping the non-deleted region of *flamenco*, indicated below, does not differ between  $w^{IR6}$  and RevI-H2

present a 10-nt 5'-overlap between sense and antisense *ZAM*-derived piRNAs (Fig. 3 h). In addition, they harbored the typical 10A and 1 U bias (Fig. 3i, Additional file 1: Figure S2A-B). This abundant production of sense and antisense *ZAM*-derived piRNAs and the ping-pong signature enrichment were similar to the results obtained for piRNAs derived from *Burdock*, a typical target of the germline piRNA pathway (Additional file 1: Figure S2C-H). Altogether, these findings strongly suggested a germinal origin of the *ZAM*-derived piRNAs produced in the RevI-H2 line.

Aub and Ago3, the two main proteins involved in piRNA production through the ping-pong mechanism, were expressed only in the germline in both  $w^{IR6}$  and RevI-H2 ovaries (Additional file 1: Figure S3A-B). This excluded a ping-pong-mediated ectopic production of *ZAM*-derived piRNAs in somatic cells of RevI-H2 ovaries. Moreover, we found that these new *ZAM*-derived piRNAs in RevI-H2 were maternally deposited in early

embryos (Additional file 1: Figure S3C-D) and possessed the same characteristics as those produced in adult ovaries (Additional file 1: Figure S3E-G). Taken together, our data strongly suggested that these *ZAM*-derived piRNAs were produced in the germline of RevI-H2 ovaries. This is intriguing because *ZAM* has been classified as a somatic TE, only expressed in somatic cells [17, 21].

To monitor the silencing potential of *ZAM*-derived piRNAs produced in the germline of the RevI-H2 ovaries, we followed the GFP expression of the pGFP-*ZAM* sensor transgene in the presence of the *actin-Gal4* driver. In  $w^{IR6}$  control ovaries, the transgene was completely silenced in somatic cells and strongly expressed in germline cells (Fig. 3j) as observed for  $w^{1118}$  (Fig. 1b). Conversely, in RevI-H2 ovaries, the transgene was silenced in the germline and strongly expressed in somatic cells. When the *ZAM* sequence was excised upon recombination between the flanking FRTs giving rise to a pGFP transgene lacking targets for *ZAM* piRNAs, GFP



**Fig. 3** De novo production of functional ZAM-derived piRNAs in the germline of the Revl-H2 line. **a** Pie chart summarizing the proportion of ZAM-derived piRNAs (allowing up to 3 mismatches) that map to the 142 piRNA clusters in  $w^{IR6}$  (no mismatch allowed, piRNA clusters defined as in [9]). **b** Density profile of ZAM-derived piRNAs along the 8.4 kb ZAM sequence in  $w^{IR6}$  ovaries (allowing up to 3 mismatches). Sense and antisense reads are represented in black and grey, respectively. ZAM organization is displayed above the profile. LTR, long terminal repeats. **c** Logo of nucleotide bias for the first ten positions in ZAM-derived piRNAs produced in  $w^{IR6}$  ovaries. The nucleotide height represents its relative frequency at that position. **d** Density profile of ZAM-derived piRNAs along the ZAM sequence produced in Revl-H2 ovaries (allowing up to 3 mismatches). Sense and antisense reads are represented in black and grey, respectively. **e** Bar diagram showing the total amount of ZAM-derived piRNAs produced in  $w^{IR6}$  and Revl-H2 ovaries, quantified from the profiles shown in **b** and **c**, respectively. **f** Logo of nucleotide bias for the first ten positions of ZAM-derived piRNAs produced in Revl-H2 ovaries. **g** Histogram showing the percentage of 5'-overlap between sense and antisense ZAM-derived piRNAs (23–29 nt) in  $w^{IR6}$  (top) and Revl-H2 (bottom) ovaries. The proportion of 10-nt overlapping pairs is in red, and the Z-score is indicated. **h** Bar diagram indicating the percentage of ZAM-derived piRNAs with ping-pong partners (PPP) in the  $w^{IR6}$  and Revl-H2 lines. **i** Analysis of nucleotide bias for sense (+) and antisense (–) ZAM-derived piRNAs with PPP in Revl-H2 ovaries. The percentage of PPP with a uridine at position 1 (1U) and with an adenosine at position 10 (10A) is shown. **j** Confocal images of ovarioles after GFP (green, left panels) and DNA (blue, middle panels) staining. Ovarioles were from the progeny of a cross between  $w^{IR6}$  or Revl-H2 females and males carrying the pGFP-ZAM sensor transgene driven by *actin-Gal4*. Right panels, merged images of GFP and DNA labeling

is strongly expressed in RevI-H2 and  $w^{IR6}$  somatic and germline cells indicating that the *ZAM* fragment in the fusion transcript is responsible for *GFP* repression (Additional file 1: Figure S3H). To confirm that the *pGFP-ZAM* transgene silencing is piRNA mediated, we knocked down the germline expression of *Aub* or *AGO3* and monitored *GFP* expression. We showed that the *pGFP-ZAM* transgene is strongly expressed in the germline of both *Aub-KD* and *AGO3-KD* (Additional file 1: Figure S3I-J), confirming that the transgene is silenced by a piRNA-mediated mechanism. These results indicated that RevI-H2 germline cells produce *ZAM*-derived piRNAs that efficiently guide sensor silencing. Conversely, *GFP* is strongly expressed in RevI-H2 somatic follicular cells that do not produce *ZAM*-derived piRNAs due to the deletion of the proximal part of *flamenco*.

Taken together, we concluded that in RevI-H2 ovaries, functional *ZAM*-derived piRNAs are produced in the germline from a new *ZAM* insertion somewhere outside the deleted region of the *flamenco* cluster.

#### **ZAM transposed into a pre-existing germline piRNA cluster**

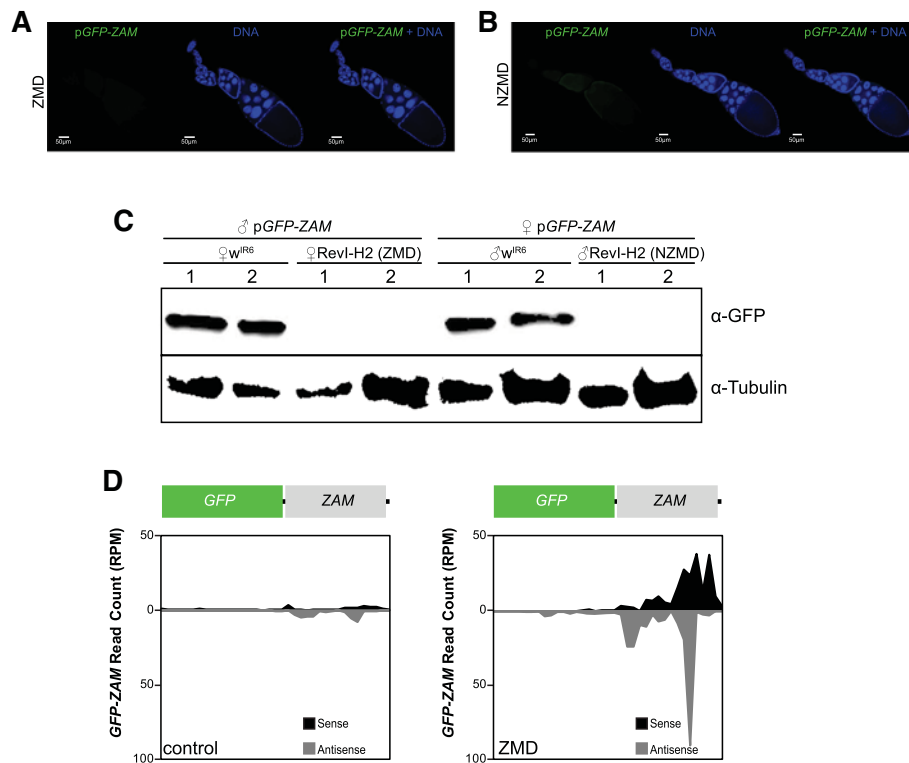
*ZAM*-derived piRNA production in the RevI-H2 line could be explained by insertion of a new copy of *ZAM* into a pre-existing germline piRNA cluster or by the *de novo* creation of a piRNA cluster in the germline induced by a new *ZAM* insertion. To discriminate between these hypotheses, we studied the activity of this putative piRNA cluster in the progeny obtained by crossing  $w^{IR6}$  and RevI-H2 flies. Since germline piRNAs are maternally deposited in the embryo and this transgenerational piRNA inheritance triggers piRNA biogenesis in the progeny [15, 16], we predicted that if *ZAM*-derived piRNAs in RevI-H2 arose from a *de novo*-formed piRNA cluster, repression operated by this cluster should only be observed when the locus is inherited from the mother. Conversely, if the germline repression of *ZAM* is due to an insertion in a pre-existing piRNA cluster, then piRNA produced by this cluster should induce repression when inherited from either parent (Additional file 1: Figure S4A).

We named ZMD (for maternal deposition of *ZAM*-derived piRNAs) the progeny obtained by crossing a RevI-H2 female and a control male and NZMD (no maternal deposition of *ZAM*-derived piRNAs) the progeny of a RevI-H2 male and a control female. In both crosses, the control line was the line harboring the *pGFP-ZAM* transgene, the expression of which is driven in germline cells by the *nanos-Gal4* driver in the  $w^{1118}$  background. In both ZMD and NZMD progenies, the sensor transgene was completely silenced in germline cells, as shown by immunofluorescence and western blot analysis

(Fig. 4a–c). This finding suggested that the unknown piRNA cluster that can silence the sensor transgene in the germline does not need maternal deposition of *ZAM*-derived piRNAs to be active. Indeed, the maternal deposition of the general piRNA population, required to activate piRNA clusters in the progeny, was sufficient for the production of *ZAM*-derived piRNAs in the progeny. Therefore, we concluded that the *ZAM*-derived piRNAs produced in the RevI-H2 germline arose from a *ZAM* sequence inserted into a pre-existing germline cluster. To further analyze the sensor silencing and to rule out the possibility that the transgene has become a piRNA cluster by itself, we sequenced and compared ovarian small RNAs from the ZMD progeny (Fig. 4d, right panel) and from a control line in which the *pGFP-ZAM* transgene is expressed in the germline (in the  $w^{IR6}$  genetic background: Fig. 4d, left panel, and Additional file 1: Figure S4B–C). The results indicated that the sensor transgene was not a *de novo* piRNA cluster because the upstream *GFP* sequence produced very few piRNAs, while a significant amount of piRNAs mapped to the *ZAM* fragment in the ZMD progeny (Fig. 4d, right panel). These data suggested the presence of a new *ZAM* insertion in a pre-existing germline piRNA cluster.

To genetically map this germline piRNA cluster that produces *ZAM*-derived piRNAs in the germline, we isolated each chromosome of the RevI-H2 line and established three lines harboring (i) the X chromosome from RevI-H2 ( $X^{RevI-H2}$ ; II; III and referred as  $X^{RevI-H2}$ ); (ii) the autosomal chromosome II from RevI-H2 ( $X$ ; II  $^{RevI-H2}$ ; III and referred as  $II^{RevI-H2}$ ); or (iii) the autosomal chromosome III from RevI-H2 ( $X$ ; II; III  $^{RevI-H2}$  and referred as  $III^{RevI-H2}$ ). It should be noted that the  $II^{RevI-H2}$  and  $III^{RevI-H2}$  lines carry a wild-type *flamenco* locus, while the  $X^{RevI-H2}$  line harbors the *flamenco* deletion present in RevI-H2. To identify which chromosome was required for germline production of *ZAM*-derived piRNAs, we assessed the *GFP* expression of the *pGFP-ZAM* sensor transgene driven by *nanos-Gal4* in each line. We found that the transgene was silenced in the germline of the  $X^{RevI-H2}$  line, like in RevI-H2 (Additional file 1: Figure S4D–E). Conversely, it was expressed in the  $II^{RevI-H2}$  and  $III^{RevI-H2}$  germlines (Additional file 1: Figure S4D–E). This indicates that in RevI-H2 ovaries, *ZAM*-derived piRNAs are produced from a germline piRNA cluster localized on the X chromosome.

We sought to identify the precise genomic location of this new *ZAM* insertion in a germline piRNA cluster localized on the X chromosome by performing whole genome sequencing of RevI-H2. We first searched for new *ZAM* insertions in euchromatin within the RevI-H2 genome using the McClintock pipeline [29], which identified seven *ZAM* insertions on the X-chromosome, including the known *ZAM* insertion in the *white* locus at X: 23,474,



**Fig. 4** ZAM-derived piRNAs are produced from a pre-existing germline piRNA cluster in RevI-H2 ovaries. **a–b** Confocal images of ovarioles after GFP (green, left panels) and DNA (blue, middle panels) staining. Merged images of the GFP and DNA signals are displayed on the right. Ovarioles were from the progeny of a cross between RevI-H2 females and control males (ZAM maternal deposition, ZMD) in **a** and from a cross between RevI-H2 males and control females (No ZAM maternal deposition, NZMD) in **b**. In both crosses, the pGFP-ZAM line in which ZAM expression is driven in germline cells by a *nanos*-Gal4 driver was the control line. **c** Western blotting of proteins extracted from ovaries of progenies of crosses between w<sup>IR6</sup> or RevI-H2 and the same control line as in **a** and **b**. The lines used for the crosses are indicated above. Proteins were from two biological replicates (1 and 2) prepared from 5 pairs of ovaries; α-tubulin was used as loading control. **d** Density profile of piRNAs mapping along the GFP-ZAM transgene sequence (allowing up to 3 mismatches). Sense and antisense reads are in black and grey, respectively. The profiles are for crosses between w<sup>IR6</sup> (left, control) or RevI-H2 (right, ZMD) females and control males harboring the pGFP-ZAM transgene

449..23,513,109. None of these X-chromosome insertions are found in known piRNA clusters. Since the component methods in McClintock do not efficiently detect new TE insertions within repetitive regions, TE nests or piRNA clusters, we used a complementary approach to identify chimeric reads containing both ZAM sequence and genomic sequence, which were not uniquely mappable on the reference genome. Using this approach, we identified a new ZAM insertion that mapped to a R1 element sequence found at multiple locations within cluster 9 [30], a dual-strand piRNA cluster located next to the X chromosome centromere (Additional file 1: Figure S4F). Analysis of paired-end reads followed by PCR and sequencing in RevI-H2 (Additional file 1: Figure S4G) localized the ZAM insertion to one of the three possible sites within this piRNA cluster spanning the interval X: 23,474,449..23,513,109. A previous inverse PCR study in RevI-H2 also detected this ZAM insertion but could not map it to a specific genomic location [24]. PCR analysis confirmed that this ZAM

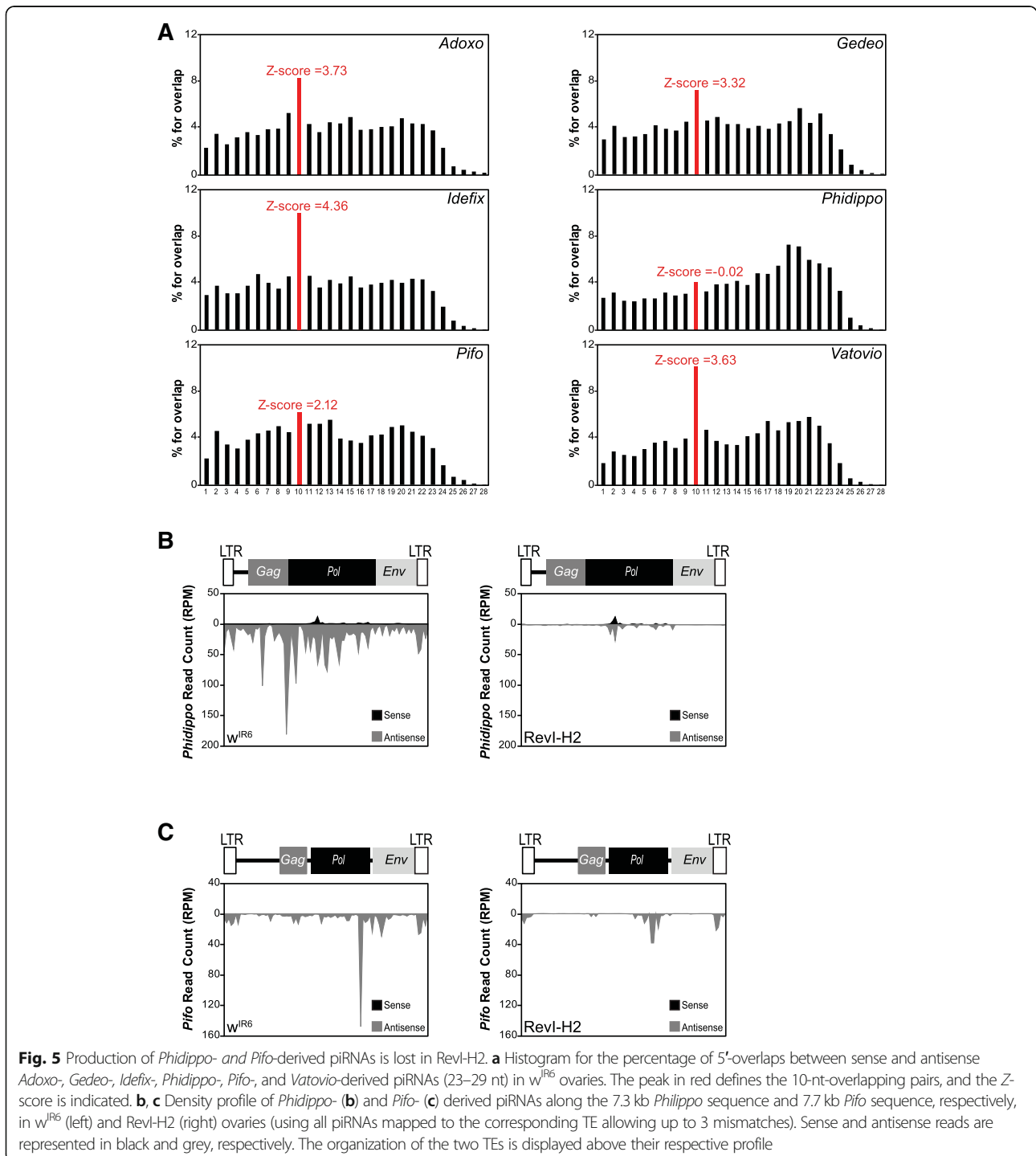
insertion is absent from the stable lines w<sup>IR6</sup>, ISO1A, and w<sup>1118</sup> (Additional file 1: Figure S4H).

These analyses demonstrated that the RevI-H2 line possesses a ZAM insertion in a pre-existing germline piRNA cluster located on the X chromosome.

#### Analysis of TEs lost with the *flamenco* deletion in RevI-H2 reveals various patterns of piRNA production

Besides ZAM, several other transposons are contained within the *flamenco* deletion in RevI-H2: *Adoxo*, *Gedeo*, *Idefix*, *Phidippo*, *Pifo*, *Uxumo*, and *Vatovio* (Additional file 1: Figure S1A). To verify whether the genomic deletion also affected the epigenetic regulation of other transposons, we analyzed the piRNA population produced by RevI-H2 ovaries against these different elements. We focused our analysis on *Phidippo* and *Pifo* because they appeared to be mainly silenced by *flamenco*. Indeed, in the control line w<sup>IR6</sup>, *Phidippo*- and *Pifo*-derived piRNAs did not harbor a ping-pong signature (Fig. 5a) and were mainly antisense (Fig. 5b, c).





Conversely, *Adoxo*-, *Gedeo*-, *Idefix*-, and *Vatovio*-derived piRNAs displayed a ping-pong signature (Fig. 5a, Additional file 1: Figure S5A). Moreover, 37% of *Phidippo*- and 54% of *Pifo*-derived piRNAs that mapped to piRNA clusters [9] mapped *flamenco*, and 21% of *Phidippo*-derived piRNAs mapped cluster 17 (Additional file 1: Figure S5B-C). Notably, cluster 17 has been proposed to be part of the *flamenco* cluster [26], raising the

percentage of *Phidippo*-mapping piRNAs that map to the extended *flamenco* to 58%.

In the RevI-H2 line, production of *Phidippo*- and *Pifo*-derived piRNAs was almost abolished (Fig. 5b, c), differently from what observed for *ZAM*-derived piRNAs (Fig. 3d). In contrast to *ZAM*, which must have an active copy outside the *flamenco* region that gave rise to the new *ZAM* insertions in RevI-H2 (such as the reference

genome copies at 2R:1,808,663..1,817,084 and 3L:24,168,844..24,176,114), no additional active copy of *Phidippo* or *Pifo* has been identified in the reference genome, besides the one in the *flamenco* locus. This indicated that the *Pifo*- and *Phidippo*-derived piRNAs are produced almost exclusively by *flamenco* and that in the absence of additional functional copies, these TEs could not invade the genome, differently from *ZAM*.

#### Transposition of *ZAM* in a germline piRNA cluster is an early event

The Rev line was first identified two decades ago [24] based on a phenotypic reversion of the mutated eye phenotype of  $w^{IR6}$  flies due to a de novo *ZAM* insertion upstream of the *white* gene. A series of homozygous RevI lines (RevI-H1, RevI-H2, and RevI-H3) were then derived from the initial Rev line. Several secondary mutations affecting eye color were recovered from the initial RevI-H2 line, and new lines were successively isolated and called RevII ([25]; see [31] for further description). To further trace when the germline acquired the potential to silence *ZAM*, we sought to determine when the *ZAM* insertion into a germline piRNA cluster occurred. We sequenced ovarian small RNAs from RevII-7 (which was derived 20 years ago from RevI-H2). Detailed analysis of *ZAM*-derived piRNAs in RevII-7 samples showed that *ZAM*-derived sense and antisense piRNAs were produced to an extent similar to what observed in the RevI-H2 line (Fig. 6a). These piRNAs displayed the typical ping-pong signature: a bias for 1U and 10A (Fig. 6b) and the enrichment of 10-nt 5'-overlaps (Fig. 6c). Moreover, 25% of the *ZAM*-derived piRNAs had a PPP with the typical 10A and 1U bias (for the sense and antisense PPPs, respectively) (Additional file 1: Figure S6A-F). We concluded that the *ZAM* insertion event into a germline piRNA cluster occurred before the RevII lines were derived from the RevI-H2 line.

Thus, the *ZAM* insertion event may have occurred very early when the three RevI lines (RevI-H1, RevI-H2, and RevI-H3) were established from the initial Rev line. Sequencing of small RNAs from RevI-H3 ovaries and analysis of *ZAM*-derived piRNAs showed again the production of sense and antisense piRNAs, but with a high bias for sense piRNAs (Fig. 6d), differently from what observed in the RevII-7 and RevI-H2 lines (Figs. 6a and 3d). The bias for 1U and 10A (Fig. 6e) and the enrichment of the 10-nt 5'-overlap were also present in the RevI-H3 line (Fig. 6f), but to a smaller extent than in the RevI-H2 and RevII-7 lines. In RevI-H3 samples, 20% of the *ZAM*-derived piRNAs possessed a PPP with the typical 10A and 1U bias (Additional file 1: Figure S6B-F). These results suggested that the RevI-H3 line, which was independently established at the same time as RevI-H2, also carries a *ZAM* insertion in a germline piRNA

cluster. However, the differences observed for *ZAM*-derived piRNAs produced in RevI-H2 and RevI-H3 suggested that there may be secondary changes to the piRNA cluster 9 or that *ZAM* inserted into another piRNA cluster in RevI-H3 different from the one identified in RevI-H2.

In addition to providing context about the timing of the germline invasion, the RevII-7 and RevI-H3 allowed us to determine the conservation of *ZAM* repression over time in independent stocks. To monitor the efficiency of the various *ZAM*-derived piRNAs produced in the germline of the RevII-7 and RevI-H3 lines, we followed the GFP expression of the pGFP-*ZAM* sensor transgene. Like for the RevI-H2 line, the transgene was completely silenced in germline cells and strongly expressed in somatic cells in both RevII-7 and RevI-H3 (Additional file 1: Figure S6G).

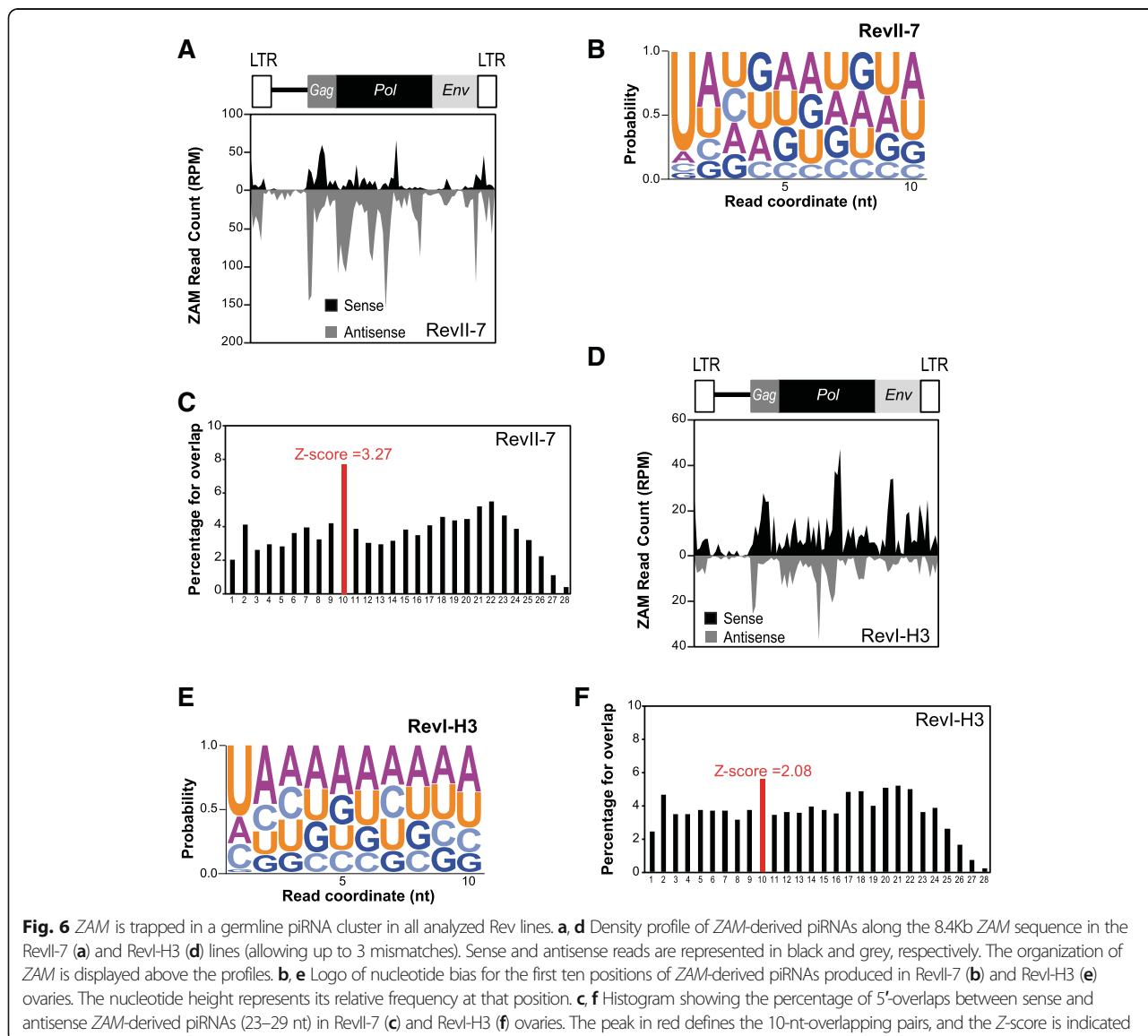
To conclude, analysis of the various Rev mutant lines suggested that *ZAM* transposition into a germline piRNA cluster (leading to de novo *ZAM*-derived piRNAs production) is an early and frequent event essential for germline protection against invasion by mobile elements from the surrounding somatic tissue.

#### Discussion

TEs have colonized the genome of all living organisms. To ensure their vertical transmission and amplification in multicellular organisms, mobile element transposition has to take place in germ cells. In turn, germ cells have developed specialized strategies to protect the integrity of their genome and thus the species continuity. Using the prototypic somatic element *ZAM* from *D. melanogaster*, we discovered that the germline can rapidly evolve to control the activity of TEs after invasion from the surrounding somatic tissues by trapping copies of the invading element into germline piRNA clusters. This ensures the production of piRNAs against the invading TE and germline genome protection.

#### The germline can adapt to the threat of active transposon invasion from surrounding somatic tissues

The *flamenco* locus is a master piRNA cluster, expressed only in somatic follicular cells that do not transfer any genetic information to the progeny. It produces somatic piRNAs characterized by the absence of the ping-pong signature. The very efficient TE silencing in somatic tissue by *flamenco* protects the germline genome against invasion by somatic TEs. The expression pattern of TE-derived piRNAs suggests that several TEs (*gtwin*, *gypsy*, *Tabor*, *gypsy5*, *gypsy10*, and *ZAM*) are almost exclusively controlled by *flamenco*-derived piRNAs [17]. In this study, we demonstrated that in control ovaries, *ZAM* is repressed exclusively in somatic follicular cells and no *ZAM*-derived piRNAs are produced in the germline,



leaving the germline genome vulnerable to ZAM invasion when its control is lost in somatic follicular cells. In agreement, the pGFP-ZAM sensor transgene was not silenced in the germline. This observation also reveals that antisense ZAM-derived piRNAs produced in somatic follicular cells are cell autonomous and do not transit to the germline to ensure ZAM silencing in this compartment.

In fly ovaries, in addition to the piRNA pathway, the short interfering RNA (siRNA) pathway also is active and involved in TE silencing [32, 33]. In addition, it has been reported that, during artificial horizontal transfers of the TE *Penelope* from *D. virilis* to *D. melanogaster*, only 21-nt siRNAs are detected in the ovary. However, they cannot completely silence *Penelope* which remained capable of occasional transposition [34]. In the case of

ZAM, the strong expression of the sensor transgene in the germline cells suggests that neither the siRNA pathway nor any other silencing pathway can silence this TE in the germline.

We previously showed that in the unstable RevI-H2 line in which ZAM silencing is released in somatic follicular cells due to a deletion in *flamenco*, ZAM particles produced within follicular cells use the endosomal vitellogenin trafficking system, which is active during late oogenesis, to enter the closely apposed oocyte and invade the germline [23]. At the time of the invasion, no ZAM-derived piRNAs were produced in the germline. Therefore, this condition could be compared to what happens when a TE first enters a new species through horizontal transfer [35–38]. For instance, the P element was introduced from *D. willistoni* to *D. melanogaster* by

horizontal transfer and a copy of P inserted at the subtelomeric heterochromatin 1A site, which corresponds to a region that gives rise to multiple small RNAs [10, 38]. This insertion is sufficient to elicit a strong P repression in *D. melanogaster* P strains [39–41]. Studies on P-M dysgenic hybrid system showed that in the F1 hybrid adult females, the invading paternally inherited P element escapes silencing and mobilizes due to the absence of maternally deposited P-derived piRNAs. With age, fertility is restored and the P element is silenced suggesting also an adaptation to P element transposon invasion. However, in contrast to what we observed for ZAM, P-derived piRNAs are produced from paternally inherited clusters [42, 43]. Our detailed analysis of piRNAs produced by the RevI-H2 ovaries revealed that this line adapted to ZAM invasion by trapping a new ZAM copy in a germline piRNA cluster, leading to the production of ZAM-derived piRNAs in the germline. Hence, the RevI-H2 line is the first example in which the germline, which does not have initially the genetic capacity to produce ZAM-derived piRNAs, needs to protect itself from invasion caused by the sudden loss of control of an endogenous somatic cell-specific TE, the expression of which is normally repressed and should not have been a risk for the progeny.

#### Retention of invading somatic TEs in germline piRNA clusters protects the germline from further invasion

In this study, we observed the de novo production of sense and antisense ZAM-derived piRNAs in RevI-H2 ovaries. Analysis of the ZMD and NZMD progenies showed that the piRNA cluster that trapped a ZAM copy was activated by maternal deposition of piRNAs other than ZAM-derived piRNAs. This finding strongly suggests that the ZAM insertion occurred in an existing germline piRNA cluster. The specific features of these ZAM-derived piRNAs (10-nt overlap and 1 U and 10A bias) indicate that they are produced through the germline-specific ping-pong cycle. Moreover, they successfully silenced the pGFP-ZAM sensor transgene in germline cells of RevI-H2 ovaries. As ZAM is not normally expressed in the germline, the sense transcripts, which are engaged in the ping-pong cycle and produce piRNAs, could arise (i) from a ZAM copy in a germline piRNA cluster, (ii) from dispersed ZAM copies inserted in the vicinity of germline promoters, or (iii) from invading ZAM mRNAs produced from somatic cells.

Among the 142 piRNA clusters identified in the *D. melanogaster* genome, most of them are significantly enriched in pericentromeric and telomeric heterochromatin [9], regions that concentrate most TEs [44]. We previously proposed a model in which piRNA clusters play the role of TE traps [26]. This model relies on the capacity of TEs to transpose into piRNA clusters that

passively acquire new TE content. Thus, TEs that “jump” into piRNA clusters can produce the corresponding piRNAs and silence homologous elements. This mechanism should constitute an adaptive advantage that can then be fixed by evolutionary selection. How piRNA clusters are formed and then produce piRNAs to repress a novel invasive TE is not well understood yet. Our findings indicate that de novo piRNAs can be produced by germline cells after ZAM invasion from another cellular lineage (i.e., somatic follicular cells) and successfully counteract the invasion. This suggests that invasive TEs can be trapped by piRNA clusters. ZAM trapping into a pre-existing piRNA cluster could result from a random transposition event. However, we found that in all the Rev lines analyzed, a germline piRNA cluster trapped a ZAM copy. Therefore, TE trapping by piRNA clusters seems to be a frequent event, and/or there is selective pressure to maintain a newly inserted ZAM copy in a germline piRNA cluster. The chromatin structure or some physical constraints, such as the nuclear organization of piRNA clusters in the genome, may play a role in transposon trapping. It has been suggested that in *Arabidopsis thaliana*, a nuclear structure, termed KNOT, in which TE-enriched regions of all five chromosomes are entangled, is a preferential insertion site for TEs [45]. In addition, the low recombination rate of these heterochromatic regions might facilitate TE accumulation for further development into piRNA clusters [46].

#### Conclusion

In our model system, ZAM internal invasion of the germline from another cell type mimics a TE horizontal transfer. This constitutes a unique opportunity to investigate the germline behavior after TE invasion in a system that experimentally imitates evolution. However, we cannot exclude that ZAM silencing is progressive, thus requiring several generations for complete repression. Finally, it is thought that piRNA clusters allow germ cells to record the TEs to which they have been exposed to over time, resulting in their silencing by the piRNA pathway. For this reason, the content of all piRNA clusters could be considered as the genetic vaccination record of that fly line or population.

#### Methods

##### Fly stocks, transgenic lines, and crosses

All experiments were performed at 20 °C. The strains *nanos*-Gal4, *actin*-Gal4,  $w^{1118}$ ,  $w^{IR6}$ , and the various Rev lines [25, 47] were from the GRD collection. The FM7c (#2177) strain, the RNAi lines against *white* (#35573), *Aub* (#35201), and *AGO3* (#35232) were from the Bloomington Drosophila Stock Center. The pGFP-ZAM sensor transgene (located on chromosome 2) was

generated by inserting part of the *ZAM env* region into the UASp-*GFP* vector containing FLP1 Recombination Target (FRT) sequences [48] after *NotI/BamHI* digestion. The *ZAM env* region was amplified by Taq polymerase using the primers 5'-GAAGCGCCGCCGGGACT-CACGACTGATGTG-3' and 5'-GAAGGATCCCGGAG GAATTGGTGGAGCGA-3'. The FRT-ZAM-FRT construct is in sense orientation relative to the *GFP* gene. Gal4-driven p*GFP-ZAM* sensor lines were established by crossing the p*GFP-ZAM* line with the *actin-Gal4* or *nanos-Gal4* driver lines.

### Immunofluorescence

Ovaries from 3- to 5-day-old flies were dissected in Schneider's *Drosophila* Medium, fixed in 4% formaldehyde/PBT for 15 min, rinsed three times with PBT ( $\times$  1 PBS, 0.1% Triton, 1% BSA), and incubated in PBT for at least 1 h and then with goat anti-GFP (ab5450, Abcam; 1/1000), mouse anti-Ago3 (1/500) [9], or rabbit anti-Aub (1/500) [9] antibodies overnight. After 3 washes in PBT, ovaries were incubated with the corresponding secondary antibodies (1/1000), coupled to Alexa-488, Cy3, or Alexa-488, respectively, for 90 min. After two washes, DNA was stained with the TOPRO-3 stain (1/1000). Three-dimensional images were acquired on Leica SP5 and Leica SP8 confocal microscopes using a  $\times$  20 objective and analyzed using the Fiji software [49]. Images of the progeny of  $w^{IR6}$  and Rev crosses were processed with the same parameters.

### Protein extraction and western blotting

At least 5 pairs of ovaries from 3- to 5-day-old flies were dissected in 200  $\mu$ l of lysis buffer (17.5 mM HEPES, 1.3 mM MgCl<sub>2</sub>, 0.38 M NaCl, 0.18 mM EDTA, 22% glycerol, 0.2% Tween-20, and protease inhibitor cocktail from Roche). After sonication, supernatants were recovered and 400  $\mu$ g of proteins were loaded on precast 4–15% acrylamide gels. Western blots were probed using anti-GFP (Ozyme; #JL-8; 1/1000) and anti-tubulin (to confirm equal loading) (Sigma, #DM1A, 1/5000) antibodies, followed by an anti-mouse (Abliance; 1/1000) secondary antibody and then the Clarity Western ECL reagent (BioRad). Densitometric analysis was performed on non-saturated signals using the Image Lab™ software (BioRad).

### Small RNA sequencing and bioinformatics analysis of piRNAs

Total RNA was isolated from 80–100 pairs of ovaries from 3- to 5-day-old flies or from ovarian somatic sheath (OSS) cell culture (for analysis of piRNA production by somatic follicular cells) with TRIzol Reagent (Ambion). After 2S RNA depletion, deep sequencing of 18–30-nt small RNAs was performed by Fasteris S.A.

(Geneva/CH) on an Illumina Hi-Seq 4000. Illumina small RNA-Seq reads were loaded into the small RNA pipeline sRNAPipe [50] for mapping to the various genomic sequence categories of the *D. melanogaster* genome (release 6.03). All libraries were normalized to the total number of genome-mapped reads (no mismatch). For the analysis, 23–29-nt RNAs were selected as piRNAs. All the analyses were performed using piRNAs mapped to TEs (0 to 3 mismatches) or genome-unique piRNAs mapped to piRNA clusters, as defined by [9] (no mismatch allowed), the strand relative to the transposon or the genome being determined [9]. The window size was of 428 nt for *flamenco*, 91 nt for *ZAM*, 80 nt for *Burdock*, 87 nt for *Pifo*, and 85 nt for *Phidippo* to establish the density profile of piRNAs and dependent of the TE size. The ping-pong signature was assessed by counting the proportion of sense piRNAs with an overlap of 10 nt with antisense piRNAs, based on piRNAs mapping to the analyzed TE (0 to 3 mismatches). The proportions of 1- to 28-nt-long overlaps were determined, and the percentage of 10-nt overlaps defined as ping-pong signature. The Z-score was determined on the proportions of 1- to 23-nt-long overlaps and considered significant for values  $> 1.96$ . The nucleotide frequency for each position within the 10-nt overlap was determined for the piRNAs mapping to the analyzed TE (0 to 3 mismatches) with ping-pong partners. Logos were generated with the WebLogo web server [51].

### Genome sequencing and analyses of new ZAM insertions in RevI-H2

Genomic DNA from RevI-H2 was extracted from a sample of mixed sex adult flies using standard protocol. Input DNA was tagged using the Illumina Nextera DNA sample preparation kit (Cat. No. FC-121-1030). Following a cleanup using the Zymo-Spin kit (Cat. No. D4023), the purified, tagged DNA was then amplified via limited-cycle PCR that also added the indices (i7 and i5) and sequencing primers. AMPure XP beads (Cat. No. A63881) were then used to purify and size select the library DNAs. The libraries were then normalized to 2 nM and pooled prior to cluster generation using a cBot instrument. The loaded flow cell was then paired-end sequenced (2  $\times$  101 bp) on an Illumina HiSeq2500 instrument.

To identify new ZAM insertions in the RevI-H2 genome, we used two complementary approaches. First, we used the McClintock system which aims to identify non-reference TE insertions using multiple component TE detection systems (commit: 9f53a5b4e1fc977b22a77-babfb24461face407d3, options -m "popoolationte retroseq temp ngs\_te\_mapper te-locate"). Because McClintock component methods do not efficiently detect new TE insertions within repetitive regions, we developed a second

approach to identify candidate TE insertions in piRNA clusters. Chimeric reads containing genomic sequence and *ZAM* 5'- or 3'-sequence were isolated from the unmapable reads. *ZAM* sequences were then stripped off from these chimeric reads, and the resulting flanking sequences mapped to the *D. melanogaster* Release 6.03 genome. This approach identified a novel *ZAM* insertion in piRNA cluster 9 as defined by [30]. We validated the presence of the insertion by PCR on DNA extracted from RevI-H2 flies. The following primers were used for Additional file 1: Figure S4G: primer F 5'-CTCACCATTTCCTCCTTGAC-3' and primer R 5'-CTCCAATCATCTCCTCAA-3'. Sequencing of the amplicon was done by GATC Biotech.

## Additional files

**Additional file 1: Figure S1.** The RevI-H2 line carries a deletion that removes *ZAM* from the *flamenco* piRNA cluster. **Figure S2.** In RevI-H2 ovaries, piRNAs derived from *Burdock*, the prototypic germinal TE, present similar features as those derived from *ZAM*. **Figure S3.** *ZAM*-derived piRNAs are de novo produced by the germline of RevI-H2 ovaries. **Figure S4.** *ZAM*-derived piRNAs originate from a germline piRNA cluster localized on the X chromosome. **Figure S5.** *Phidippo*- and *Pifo*-derived piRNAs are mainly produced by the *flamenco* cluster. **Figure S6.** *ZAM*-derived piRNAs produced in the different Rev lines display similar features. (PDF 3179 kb)

**Additional file 2:** Review history. (DOCX 515 kb)

## Acknowledgements

We thank Nathalie Guéguen, Françoise Pellissier, and Nadège Anglaret for the technical assistance. We thank Arpita Sarkar who constructed the plasmid p*GFP-ZAM*, Marion Delattre for the assistance sequencing the *w*<sup>IR6</sup> and RevI-H2 small RNA libraries, and J. Brennecke for Aub and AGO3 antibodies.

We thank all members of the team for discussion and critical comments.

## Review history

The review history is available at Additional file 2.

## Authors' contributions

EB and CD conceived the study. EB, CD, MY, and NM designed and performed the experiments. EB and CD gathered and analyzed the small RNA-Seq data. CD, EB, and SJ performed the bioinformatic analyses. CB and SJ gathered and analyzed the DNA-Seq data. EB and CD analyzed the data. SJ and CV participated in the discussions about the project and critically read the manuscript. CD and EB wrote the article. All authors read and approved the final manuscript.

## Funding

This work was supported by grants from the Agence Nationale pour la Recherche (ANR-PlasTiSiPi and ANR-EpiTET projects). N.M. and M.Y. were supported by the Ministère de l'Enseignement Supérieur et de la Recherche (MESR) and the Ligue contre le Cancer. E.B. received a grant from the Region Auvergne. This research is supported by the University of Georgia Research Foundation and the French government IDEX-ISITE initiative 16-IDEX-0001 (CAP20-25).

## Availability of data and materials

RNA-Seq (SRR7651646, OSS cell culture; SRR7630522, ovaries from 3- to 5-day-old *w*<sup>IR6</sup> flies; SRR7630528, ovaries from 3- to 5-day-old RevI-H2 flies; SRR7630524, ovaries from 3- to 5-day-old RevI-7 flies; SRR7630523, ovaries from 3- to 5-day-old RevI-H3 flies; SRR7630525, early embryos from 3- to 5-day-old RevI-H2 flies; SRR7630526, ovaries from 3- to 5-day-old F1 flies from a cross between a *w*<sup>IR6</sup> female and a control line harboring the p*GFP-ZAM* transgene whose expression is driven by the *nanos*-Gal4 driver; SRR7630527, ovaries from 3- to 5-day-old F1 flies from a cross between a RevI-H2 female

and a control line harboring the p*GFP-ZAM* transgene whose expression is driven by the *nanos*-Gal4 driver) and DNA-Seq from RevI-H2 (SRR8975035) datasets supporting the conclusions of this article are available in the SRA repository under accession number SRP155919 [52].

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup>GREd, Université Clermont Auvergne, CNRS, INSERM, Faculté de Médecine, 63000 Clermont-Ferrand, France. <sup>2</sup>Present address: UFIP UMR-CNRS 6286, Epigénétique: prolifération et différenciation, Faculté des Sciences et des Techniques, 2 rue de la Houssinière, 44322 Nantes, France. <sup>3</sup>Department of Genetics and Institute of Bioinformatics, University of Georgia, 120 E. Green St, Athens, GA 30602, USA.

Received: 2 January 2019 Accepted: 11 June 2019

Published online: 21 June 2019

## References

- Hancks DC, Kazazian HH. Roles for retrotransposon insertions in human disease. *Mobile DNA*. 2016;7:9.
- Levin HL, Moran JV. Dynamic interactions between transposable elements and their hosts. *Nat Rev Genet*. 2011;12:615–27.
- Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet*. 2007;8:272–85.
- Aravin A, Gaidatzis D, Pfeffer S, Lagos-Quintana M, Landgraf P, Iovino N, et al. A novel class of small RNAs bind to MILI protein in mouse testes. *Nat Cell Biol*. 2006;442(7099):203–7.
- Girard A, Sachidanandam R, Hannon GJ, Carmell MA. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature*. 2006;442:199–202.
- Grivna ST, Pyhtila B, Lin H. MIWI associates with translational machinery and PIWI-interacting RNAs (piRNAs) in regulating spermatogenesis. *Proc Natl Acad Sci U S A*. 2006;103:13415–20.
- Lau NC, Seto AG, Kim J, Kuramochi-Miyagawa S, Nakano T, Bartel DP, et al. Characterization of the piRNA complex from rat testes. *Science*. 2006;313:363–7.
- King RC, Aggarwal SK, Aggarwal U. The development of the female *Drosophila* reproductive system. *J Morphol*. 1968;124:143–66.
- Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, et al. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell*. 2007;128:1089–103.
- Senti K-A, Brennecke J. The piRNA pathway: a fly's perspective on the guardian of the genome. *Trends Genet*. 2010;26:499–509.
- Siomi MC, Sato K, Pezic D, Aravin AA. PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Biol*. 2011;12:246–58.
- Siensi G, Dönertas D, Brennecke J. Transcriptional silencing of transposons by Piwi and Maelstrom and its impact on chromatin state and gene expression. *Cell*. 2012;151:964–80.
- Gunawardane LS1, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, Siomi H, Siomi MC. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science*. 2007; 315(5818):1587–1590.
- Brennecke J, Malone CD, Aravin AA, Sachidanandam R, Stark A, Hannon GJ. An epigenetic role for maternally inherited piRNAs in transposon silencing supplementary data. *Science*. 2008;322:1387–92.
- Le Thomas A, Stuwe E, Li S, Du J, Marinov G, Rozhkov N, et al. Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. *Genes & Development*. 2014;28:1667–80.
- de Vanssay A, Bougé A-L, Boivin A, Hermant C, Teyssset L, Delmarre V, et al. Paramutation in *Drosophila* linked to emergence of a piRNA-producing locus. *Nature*. 2012;490:112–5.

17. Malone CD, Brennecke J, Dus M, Stark A, McCombie WR, Sachidanandam R, et al. Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell*. 2009;137:522–35.
18. Li C, Vagin W, Lee S, Xu J, Ma S, Xi H, et al. Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell*. 2009;137:509–21.
19. Pelisson A, Song SU, Prud'homme N, Smith PA, Bucheton A, Corces VG. Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the *Drosophila* flamenco gene. *EMBO J*. 1994;13:4401–11.
20. Prud'homme N, Gans M, Masson M, Terzian C, Bucheton A. Flamenco, a gene controlling the gypsy retrovirus of *Drosophila melanogaster*. *Genetics*. 1995;139(2):697–711.
21. Desset S, Meignin C, Dastugue B, Vaury C. COM, a heterochromatic locus governing the control of independent endogenous retroviruses from *Drosophila melanogaster*. *Genetics*. 2003;164:501–9.
22. Chalvet F, Teyssset L, Terzian C, Prud'homme N, Santamaria P, Bucheton A, et al. Proviral amplification of the Gypsy endogenous retrovirus of *Drosophila melanogaster* involves env-independent invasion of the female germline. *EMBO J*. 1999;18:2659–69.
23. Brasset E, Taddei AR, Arnaud F, Faye B, Fausto AM, Mazzini M, et al. Viral particles of the endogenous retrovirus ZAM from *Drosophila melanogaster* use a pre-existing endosome/exosome pathway for transfer to the oocyte. *Retrovirology*. 2006;3:25.
24. Leblanc P, Desset S, Dastugue B, Vaury C. Invertebrate retroviruses: ZAM a new candidate in *D.melanogaster*. *EMBO J*. 1997;16:7521–31.
25. Desset S, Conte C, Dimitri P, Calco V, Dastugue B, Vaury C. Mobilization of two retroelements, ZAM and Idefix, in a novel unstable line of *Drosophila melanogaster*. *Mol Biol Evol*. 1999;16:54–66.
26. Zanni V, Eymery A, Coiffet M, Zytznicki M, Luyten I, Quesneville H, et al. Distribution, evolution, and diversity of retrotransposons at the flamenco locus reflect the regulatory properties of piRNA clusters. *Proc Natl Acad Sci U S A*. 2013;110:19842–7.
27. Desset S, Buchon N, Meignin C, Coiffet M, Vaury C. In *Drosophila melanogaster* the COM Locus Directs the Somatic Silencing of Two Retrotransposons through both Piwi-Dependent and -Independent Pathways. *PLoS One*. 2008;3(2):e1526.
28. Mevel-Ninio M, Pelisson A, Kinder J, Campos AR, Bucheton A. The flamenco locus controls the gypsy and ZAM retroviruses and is required for *Drosophila* oogenesis. *Genetics*. 2007;175:1615–24.
29. Nelson MG, Linheiro RS, Bergman CM. McClintock: an integrated pipeline for detecting transposable element insertions in whole-genome shotgun sequencing data. *G3 (Bethesda)*. 2017;7:2763–78.
30. George P, Jensen S, Pogorelnik R, Lee J, Xing Y, Brasset E, et al. Increased production of piRNAs from euchromatic clusters and genes in *Anopheles gambiae* compared with *Drosophila melanogaster*. *Epigenetics & Chromatin*. 2015;8:50.
31. Goriaux C, Théron E, Brasset E, Vaury C. History of the discovery of a master locus producing piRNAs: the flamenco/COM locus in *Drosophila melanogaster*. *Front. Gene*. 2014;5:257.
32. Czech B, Malone CD, Zhou R, Stark A, Schlingehayde C, Dus M, et al. An endogenous small interfering RNA pathway in *Drosophila*. *Nature*. 2008;453:798–802.
33. Kawamura Y, Saito K, Kin T, Ono Y, Asai K, Sunohara T, et al. *Drosophila* endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature*. 2008;453:793–7.
34. Rozhkov NV, Aravin AA, Zelentsova ES, Schostak NG, Sachidanandam R, McCombie WR, et al. Small RNA-based silencing strategies for transposons in the process of invading *Drosophila* species. *RNA*. 2010;16:1634–45.
35. Panaud O. Horizontal transfers of transposable elements in eukaryotes: The flying genes. *C. R. Biol*. 2016;339:296–9.
36. Baidouri El M, Carpentier M-C, Cooke R, Gao D, Lasserre E, Llauro C, et al. Widespread and frequent horizontal transfers of transposable elements in plants. *Genome Res*. 2014;24:831–8.
37. Barckmann B, El-Barouk M, Pelisson A, Mugat B, Li B, Franckhauser C, et al. The somatic piRNA pathway controls germline transposition over generations. *Nucleic Acids Res*. 2018;46:9524–36.
38. Daniels SB, Peterson KR, Strausbaugh LD, Kidwell MG, Chovnick A. Evidence for horizontal transmission of the P transposable element between *Drosophila* species. *Genetics*. 1990;124:339–55.
39. Ronsseray S, Lehmann M, Anxolabéhère D. Copy number and distribution of P and I mobile elements in *Drosophila melanogaster* populations. *Chromosoma*. 1989;98:207–14.
40. Marin L, Lehmann M, Nouaud D, Izaabel H, Anxolabéhère D, Ronsseray S. P-Element repression in *Drosophila melanogaster* by a naturally occurring defective telomeric P copy. *Genetics*. 2000;155:1841–54.
41. Brennecke J, Malone CD, Aravin AA, Sachidanandam R, Stark A, Hannon GJ. An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science*. 2008;322:1387–92.
42. Khurana JS, Wang J, Xu J, Koppetsch BS, Thomson TC, Nowosielska A, et al. Adaptation to P element transposon invasion in *Drosophila melanogaster*. *Cell*. 2011;147:1551–63.
43. Moon S, Cassani M, Lin YA, Wang L, Dou K, Zhang ZZ. A robust transposon-endogenizing response from germline stem cells. *Developmental Cell*. 2018;47:660–71.
44. Bergman CM, Quesneville H, Anxolabéhère D, Ashburner M. Recurrent insertion and duplication generate networks of transposable element sequences in the *Drosophila melanogaster* genome. *Genome Biol*. 2006;7(11):R112.
45. Grob S, Schmid MW, Grossniklaus U. Hi-C analysis in *Arabidopsis* identifies the KNOT, a structure with similarities to the flamenco locus of *Drosophila*. *Mol Cell*. 2014;55:678–93.
46. Lu J, Clark AG. Population dynamics of PIWI-interacting RNAs (piRNAs) and their targets in *Drosophila*. *Genome Res*. 2010;20:212–27.
47. Leblanc P, Desset S, Dastugue B, Vaury C. Invertebrate retroviruses: ZAM a new candidate in *D. melanogaster*. *EMBO J*. 1997;16:7521–31.
48. Dufourt J, Dennis C, Boivin A, Gueguen N, Theron E, Goriaux C, et al. Spatio-temporal requirements for transposable element piRNA-mediated silencing during *Drosophila* oogenesis. *Nucleic Acids Res*. 2014;42:2512–24.
49. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. *Nat Meth*. 2012;9:676–82.
50. Pogorelnik R, Vaury C, Pouchin P, Jensen S, Brasset E. sRNAPipe: a Galaxy-based pipeline for bioinformatic in-depth exploration of small RNAseq data. *Mobile DNA*. 2018;9:25.
51. Crooks GE, Hon G, Chandonia J-M, Brenner SE. WebLogo: a sequence logo generator. *Genome Res*. 2004;14:1188–90.
52. Duc C, Yoth M, Jensen S, Mounié N, Bergman CM, Vaury C, Brasset E. Trapping a somatic endogenous retrovirus into a germline piRNA cluster immunizes the germline against further invasion. *Data sets*. SRA. 2019. <https://www.ncbi.nlm.nih.gov/sra/SRP155919>

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

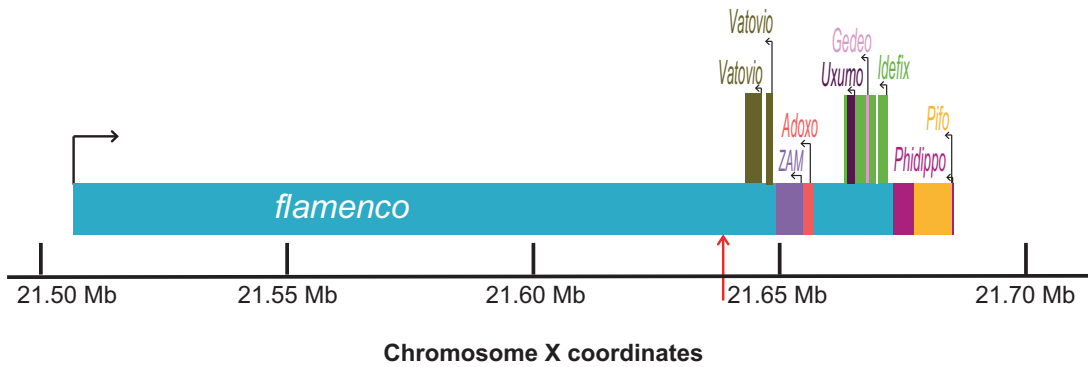
- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

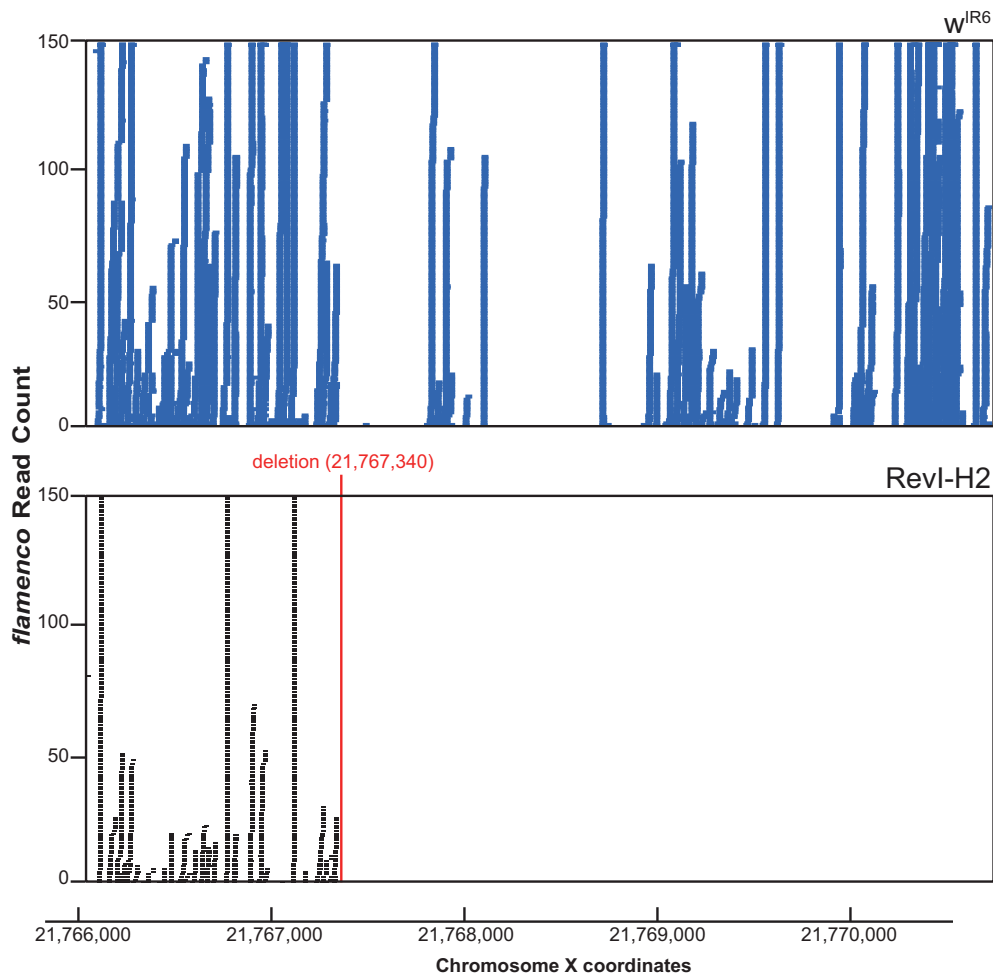
Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)



A

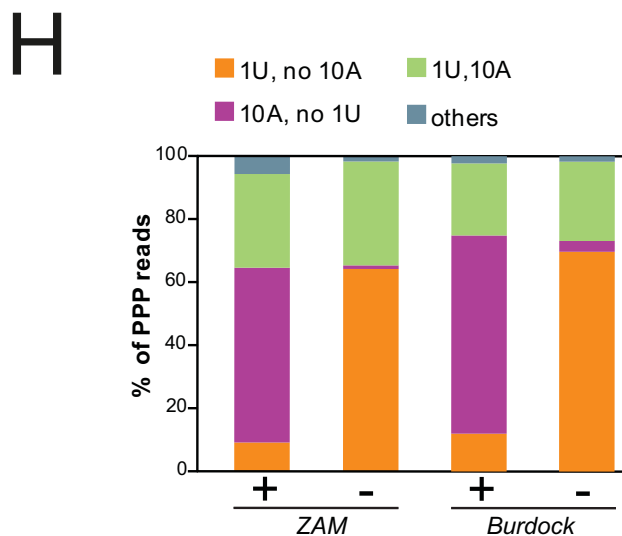
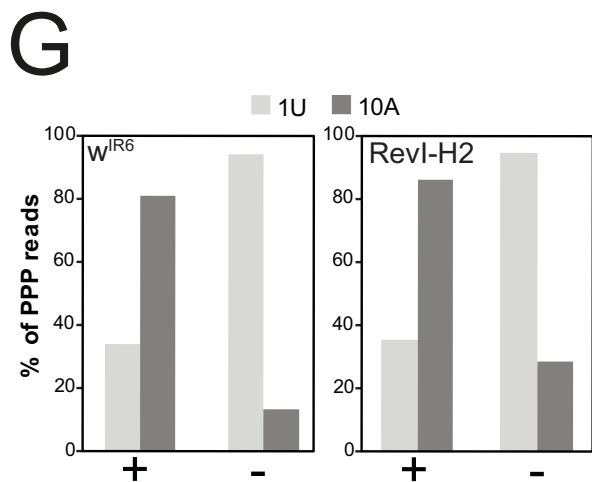
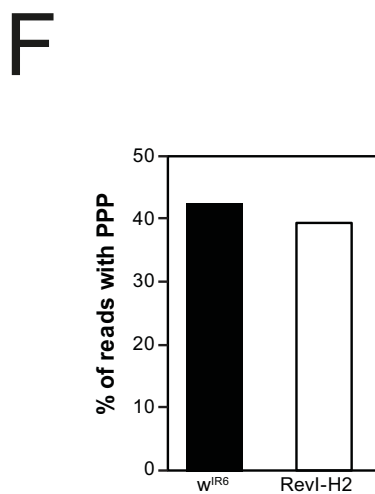
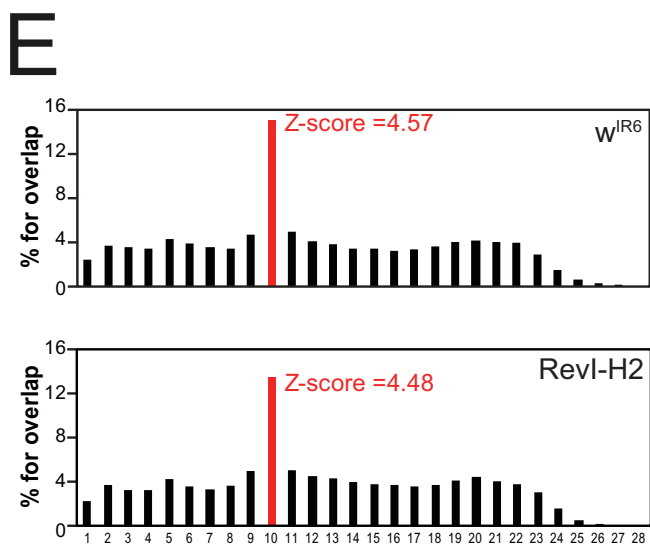
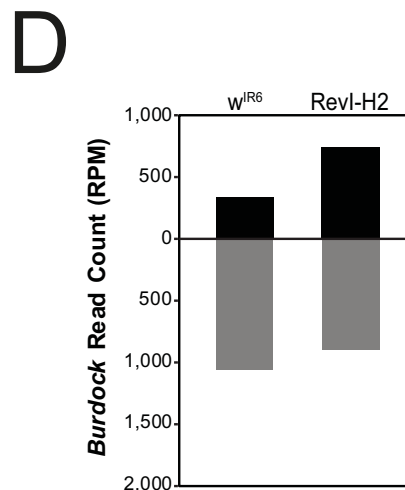
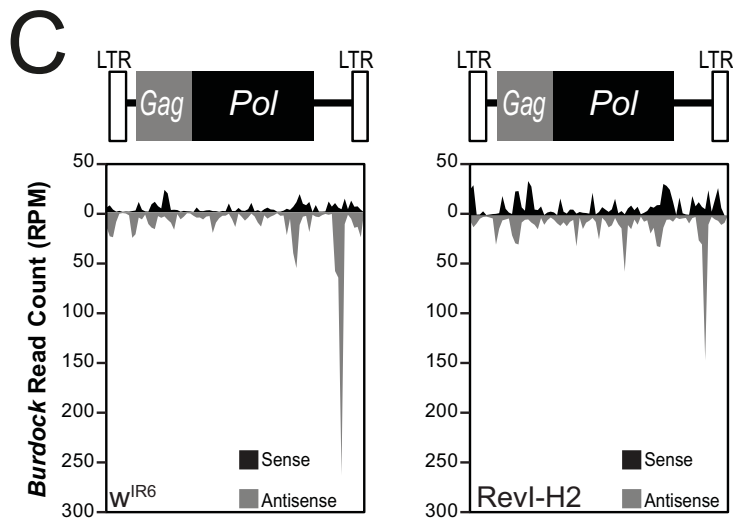


B



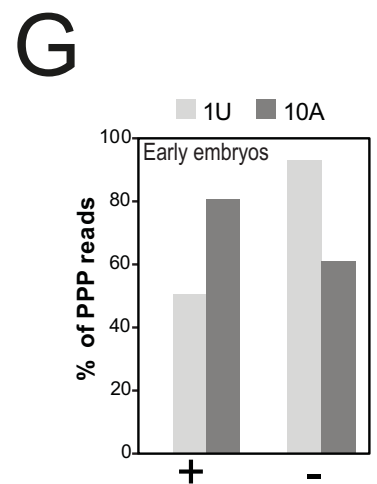
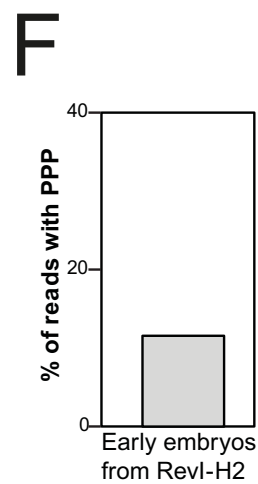
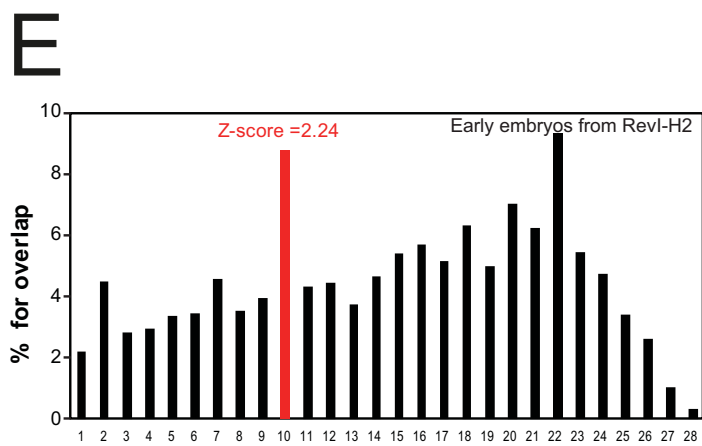
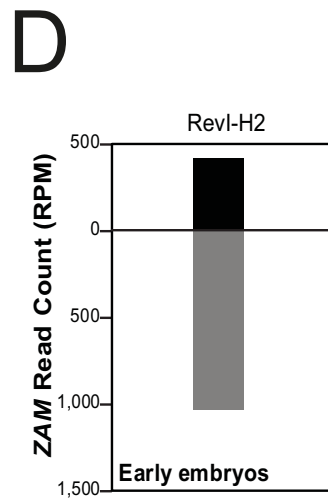
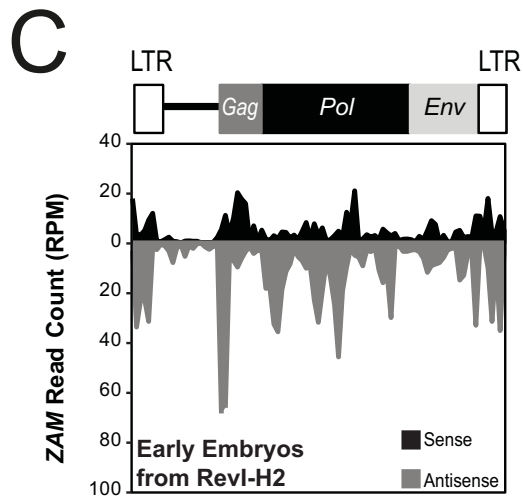
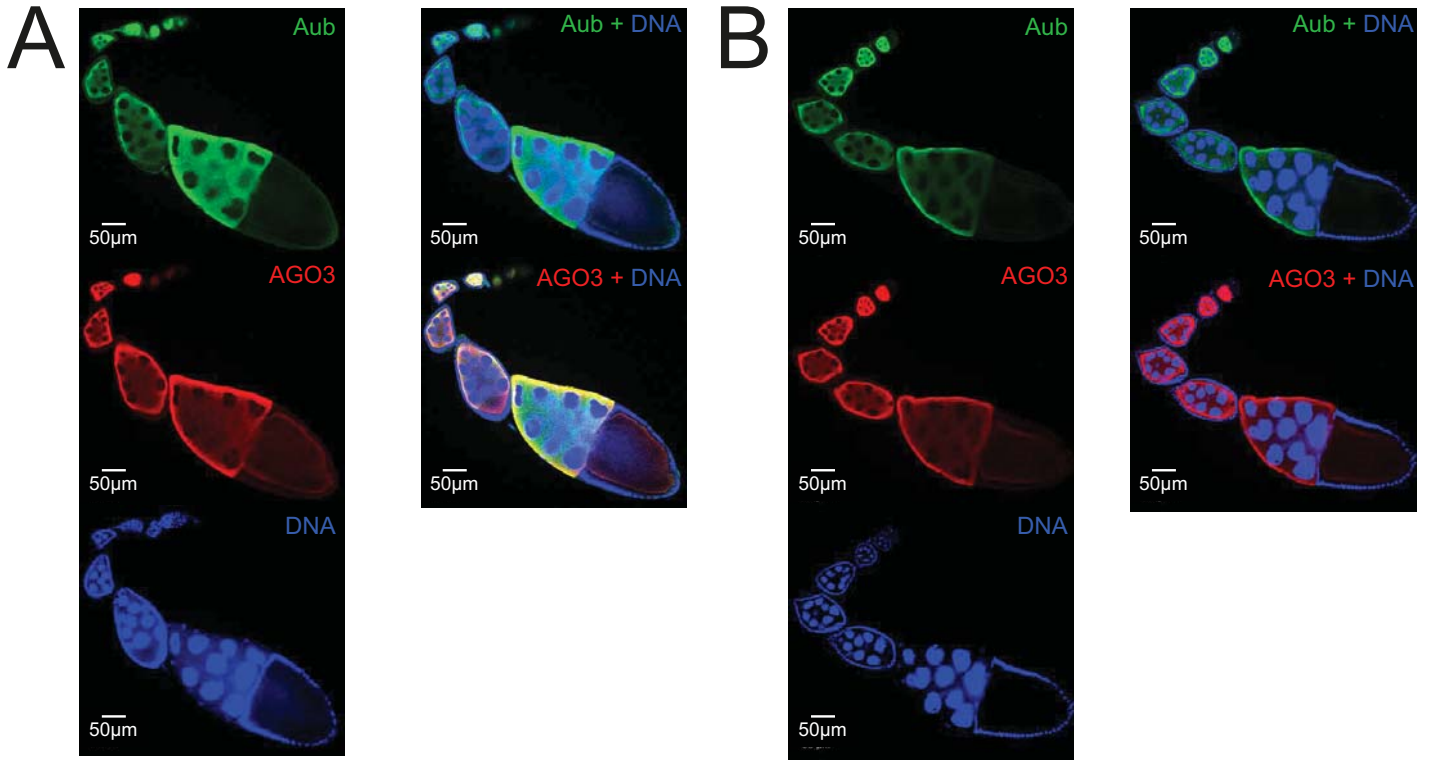
**Figure S1. The Revl-H2 line carries a deletion that removes *ZAM* from the *flamenco* piRNA cluster.**  
**a** Structure of the *flamenco* piRNA cluster in *D. melanogaster*. TEs located in the region of *flamenco* deleted in Revl-H2 are presented individually above *flamenco* according to [26]. The centromere of the X chromosome is on the right-hand (proximal) side. Sense-strand transcription for *flamenco* and TE orientation are indicated by black arrows. The *flamenco* deletion distal break-point in Revl-H2 [26] is indicated by a red arrow. The chromosome coordinates are according to release 5 of the *D. melanogaster* genome. **b** Genome browser panel showing *flamenco* piRNA levels in *w<sup>IR6</sup>* and Revl-H2 line. The refined Release 6.03 coordinate for the break-point of the Revl-H2 deletion in *flamenco* reported in [26] is displayed by a red line.

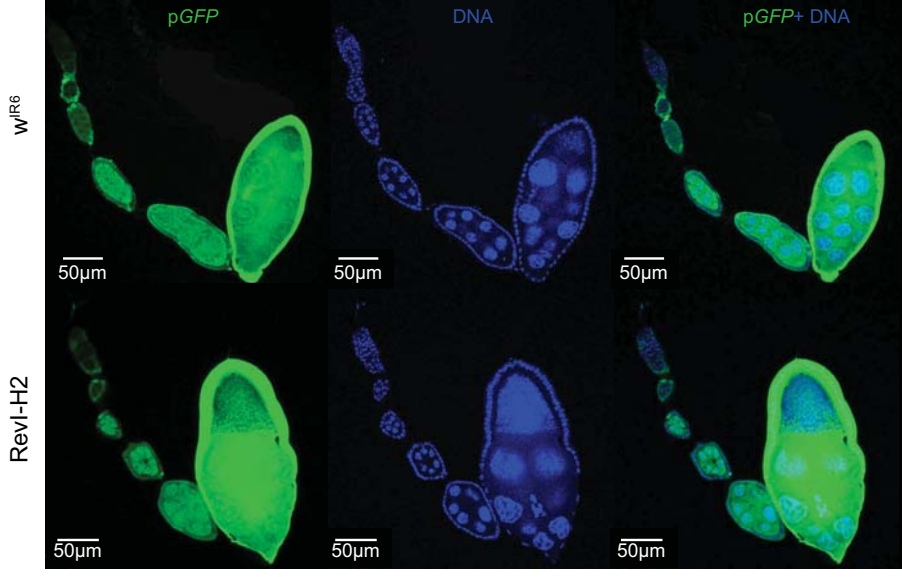
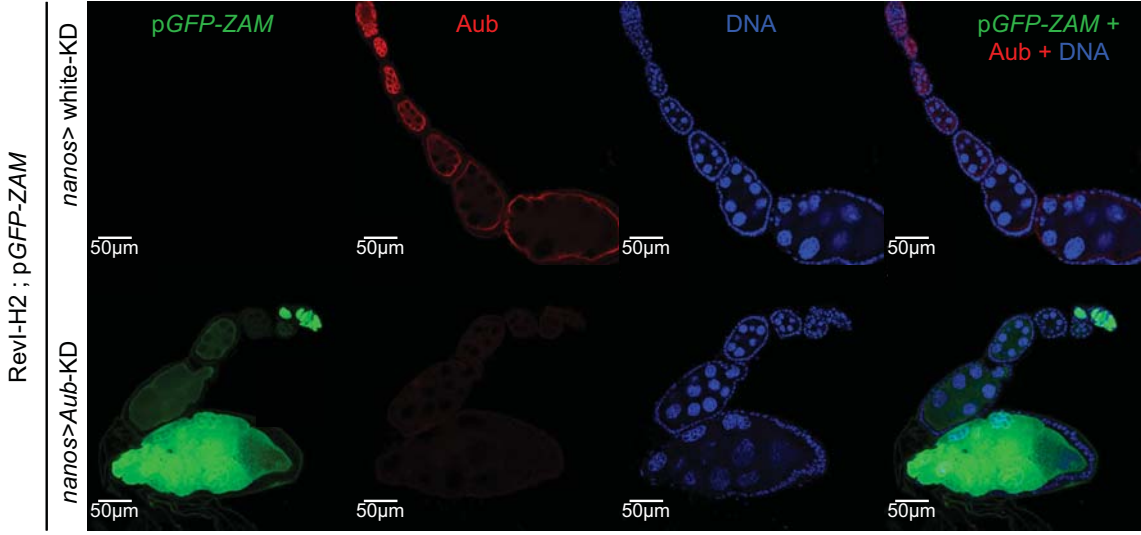
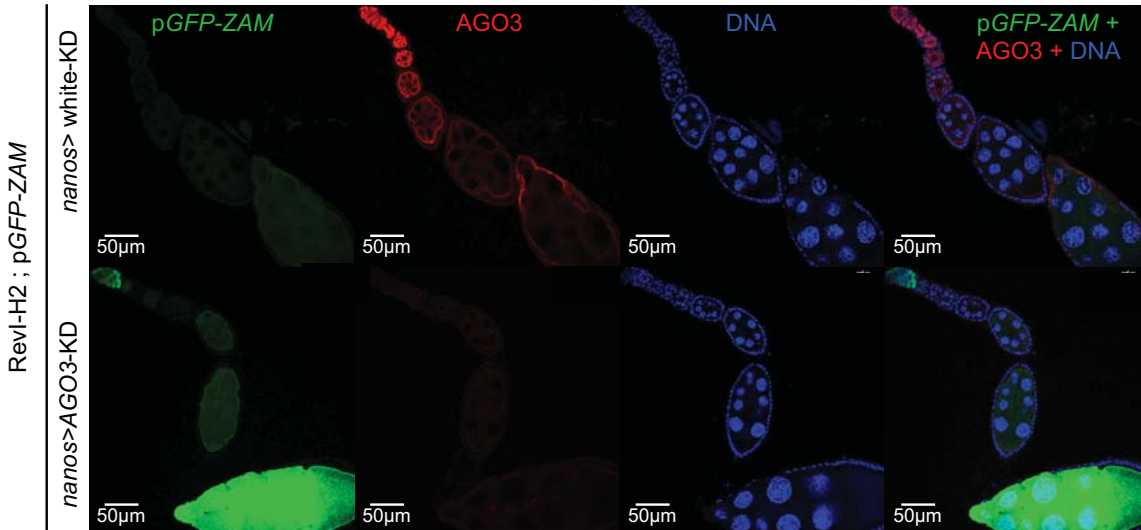




**Figure S2. In Revl-H2 ovaries, piRNAs derived from *Burdock*, the prototypic germinal TE, present similar features as those derived from *ZAM*.**

**a-b** Logo of nucleotide bias for the first ten positions of sense (*a*) and antisense (*b*) *ZAM*-derived piRNAs with ping-pong partner (PPP) produced in Revl-H2 ovaries. The nucleotide height represents its relative frequency at that position. **c** Density profile of *Burdock*-derived piRNAs along the 6.4kb *Burdock* sequence in  $w^{IR6}$  (left) and Revl-H2 (right) ovaries (all mapping piRNAs allowing up to 3 mismatches). Sense and antisense reads are in black and grey, respectively. The organization of *Burdock* is displayed above the profiles. **d** The total amount of *Burdock*-derived piRNAs produced in  $w^{IR6}$  and Revl-H2 ovaries was quantified from the profiles in *c*. **e** Histogram showing the percentage of 5'-overlaps between sense and antisense *Burdock*-derived piRNAs (23-29nt) in  $w^{IR6}$  (top) and Revl-H2 (bottom) ovaries. The peak in red defines the proportion of 10nt-overlapping pairs and the Z-score is indicated. **f** Bar diagram indicating the percentage of *Burdock*-derived piRNAs with ping-pong partner (PPP) in the  $w^{IR6}$  and Revl-H2 lines. **g** Analysis of the nucleotide bias for sense (+) and antisense (-) *Burdock*-derived piRNAs with PPP in  $w^{IR6}$  (left) and Revl-H2 (right). The percentages of PPPs with a 1U and 10A are displayed. **h** Analysis of nucleotide bias for sense (+) and antisense (-) *ZAM*- and *Burdock*-derived piRNAs with PPPs in Revl-H2 ovaries.



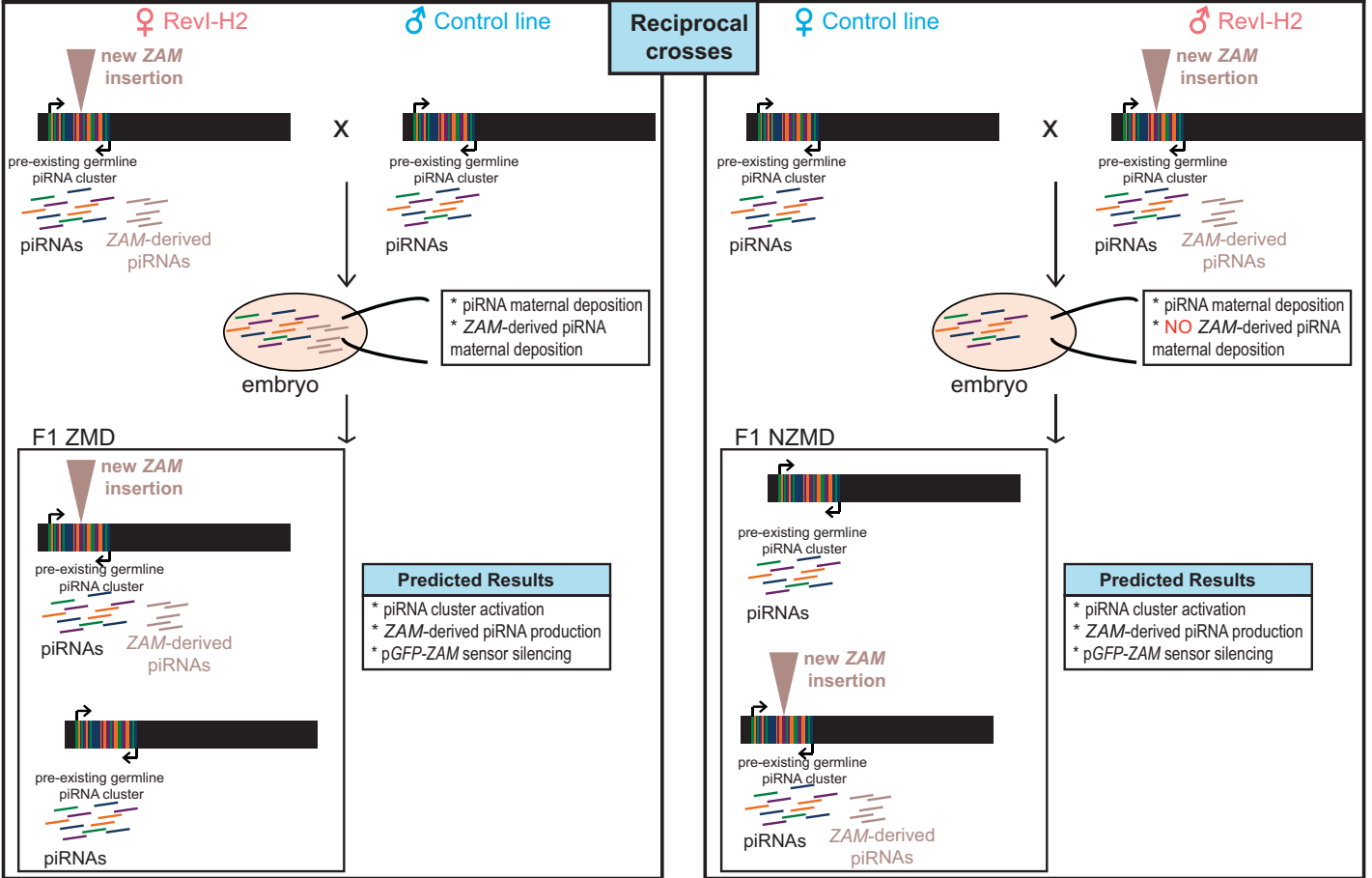
**H****I****J**

**Figure S3. ZAM-derived piRNAs are *de novo* produced by the germline of Revl-H2 ovaries.**

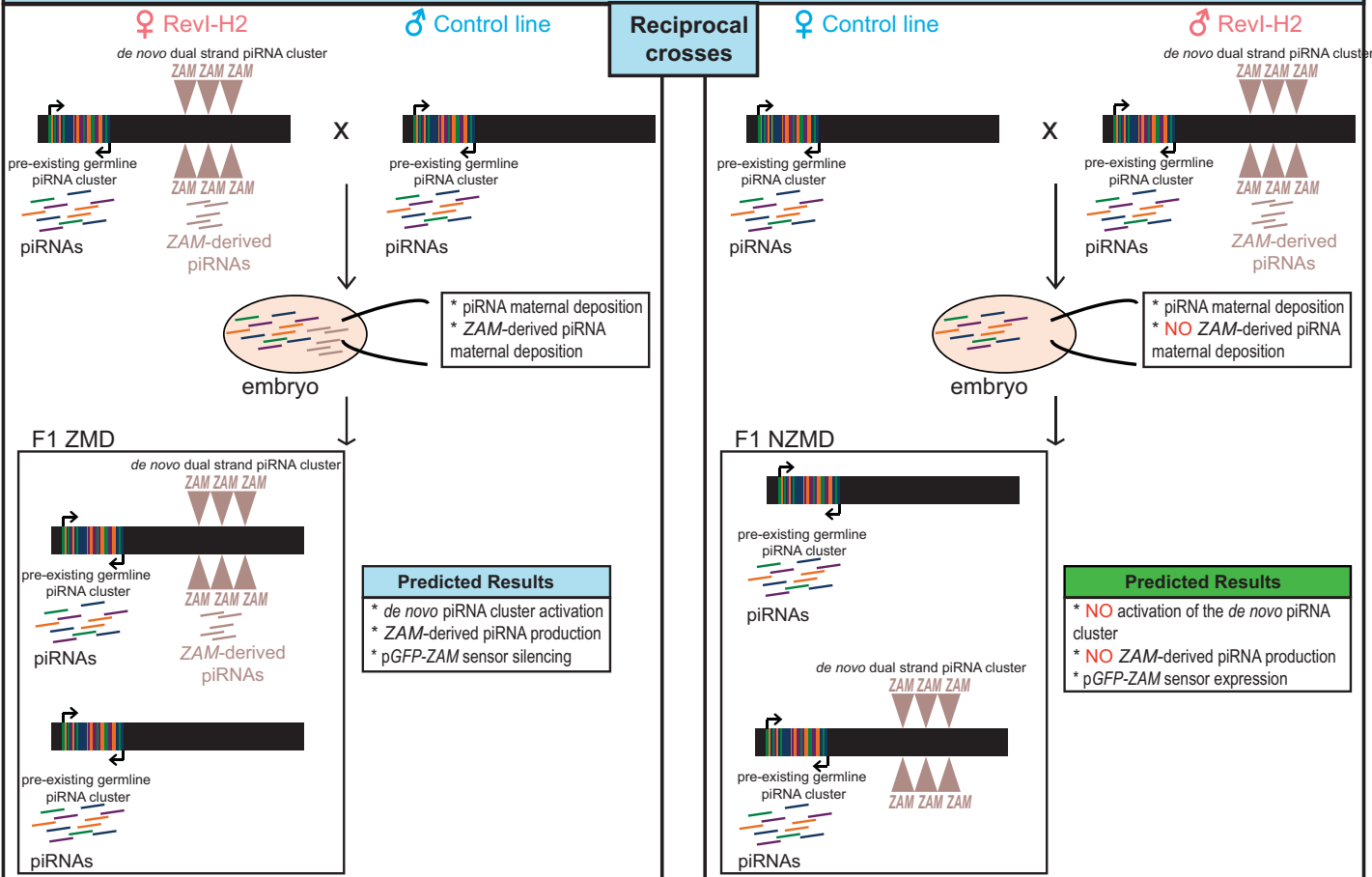
**a-b** Confocal images of ovarioles from *w<sup>IR6</sup>* (*a*) and Revl-H2 (*b*) ovaries after labeling with anti-Aub (green, top) and -Ago3 (red, middle) antibodies and DNA (blue, bottom) staining. Merged images of the Aub or Ago3 signal and DNA staining are displayed on the right panels. **c** Density profile of ZAM-derived piRNAs along the ZAM sequence produced in early embryos from Revl-H2 (allowing up to 3 mismatches). Sense and antisense reads are represented in black and grey, respectively. ZAM organization is displayed above the profile. **d** The total amount of ZAM-derived piRNAs produced in early embryos from Revl-H2 was quantified from the profile in Fig. S3C. Sense and antisense reads are represented in black and grey, respectively. **e** Histogram showing the percentage of 5'-overlaps between sense and antisense ZAM-derived piRNAs (23-29nt) in early embryos from Revl-H2. The peak in red defines the proportion of 10nt-overlapping pairs and the Z-score is indicated. **f** Bar diagram indicating the percentage of ZAM-derived piRNAs with ping-pong partner (PPP) in early Revl-H2 embryos. **g** Analysis of the nucleotide bias for sense (+) and antisense (-) ZAM-derived piRNAs with PPPs in early Revl-H2 embryos. The percentages of PPPs with a 1U and those with a 10A are displayed. **h** Confocal images of ovarioles after GFP (green, left panels) and DNA (blue, middle panels) staining. Ovarioles were from the progeny of a cross between *w<sup>IR6</sup>* or Revl-H2 females and males carrying the pGFP-ZAM sensor transgene – obtained after excision of the ZAM sequence upon recombination between the flanking FRTs – driven by *actin-Gal4*. Right panels, merged images of GFP and DNA labeling. **i-j** Confocal images of ovarioles after GFP (green), Aub (red, *i*) or AGO3 (red, *j*) and DNA (blue) staining. Ovarioles were from the progeny of a cross between Revl-H2 females carrying the *nanos-Gal4* driver and males carrying the pGFP sensor transgene with either of the RNAi *Aub*- or *AGO3-KD* (Knock-Down). The white-KD was used as a control. Right panels, merged images of GFP, Aub (*i*) or AGO3 (*j*) and DNA labeling.

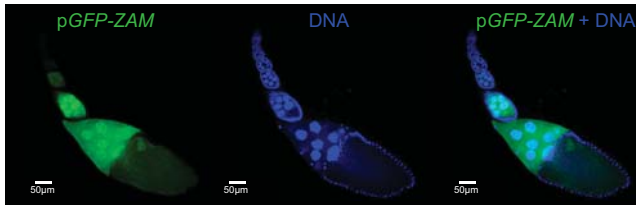
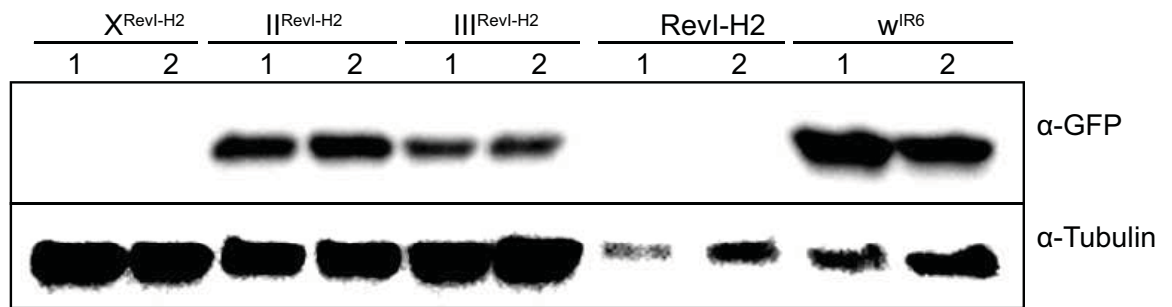
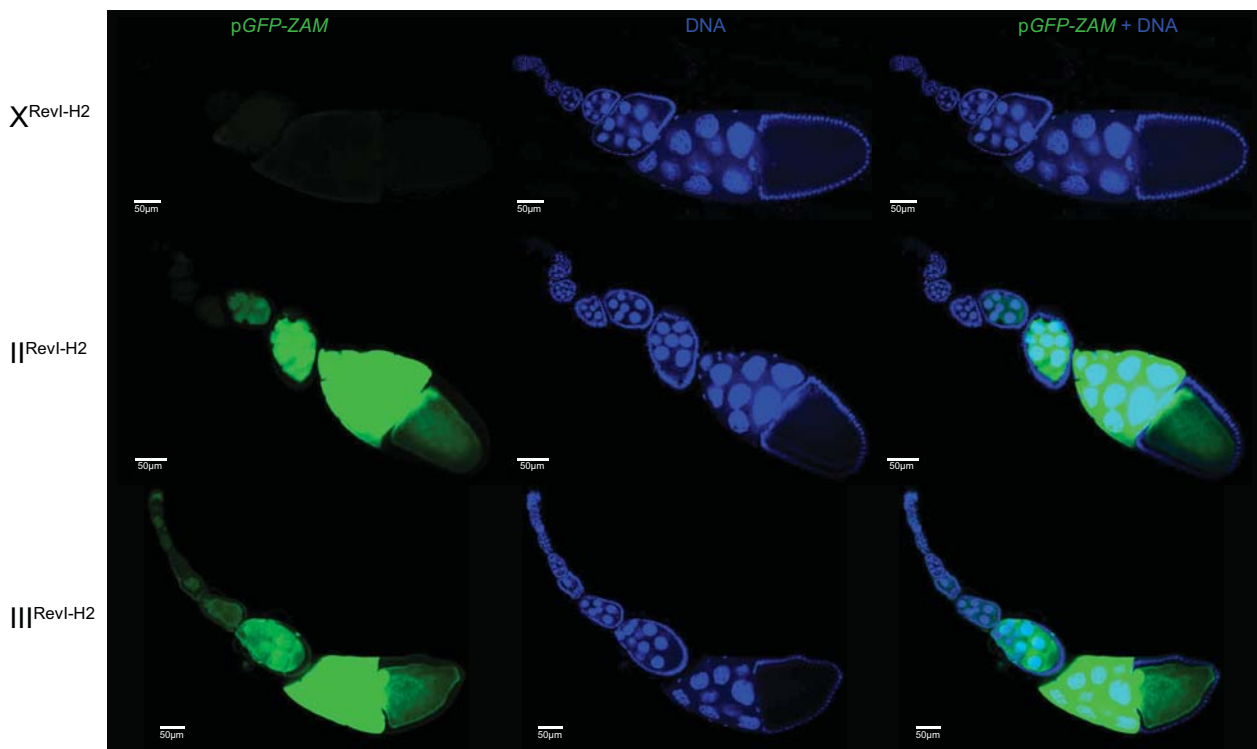
A

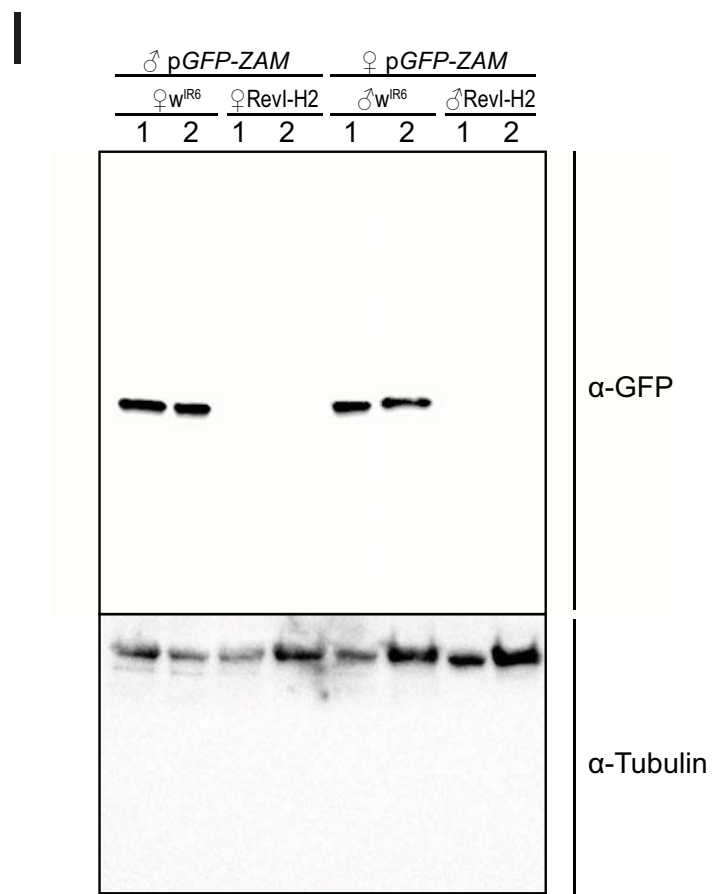
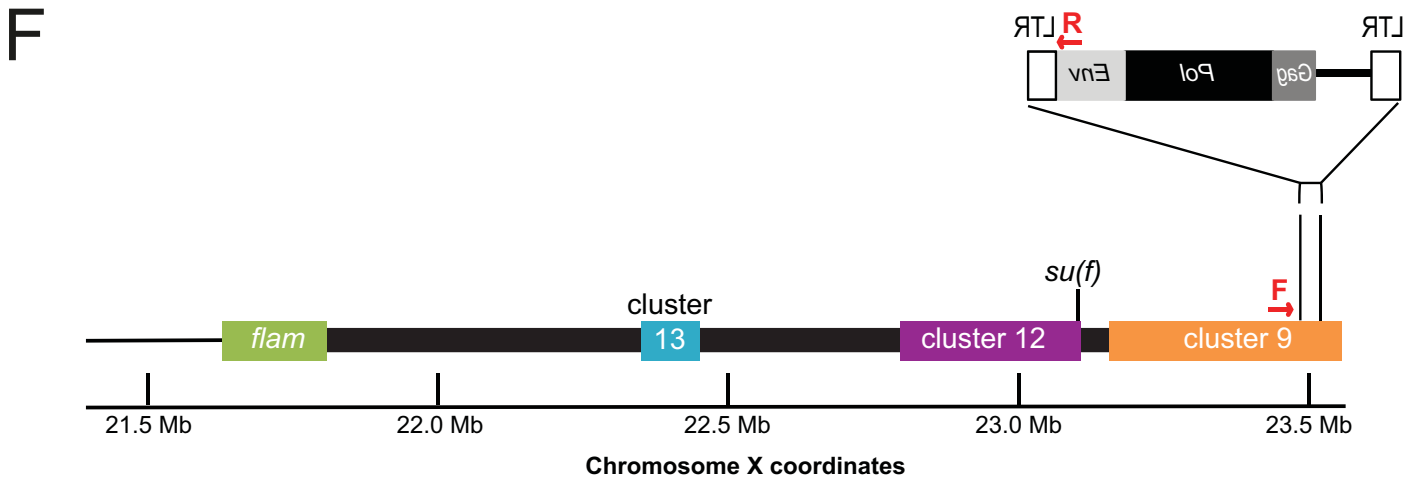
**Hypothesis n°1: ZAM new insertion occurred in a pre-existing germline piRNA cluster**



**Hypothesis n°2: ZAM new insertions formed a *de novo* germline piRNA cluster**



**B**F1 progeny (cross female  $w^{IR6}$  with male  $pGFP-ZAM$ )**C**F1 progeny (cross male  $w^{IR6}$  with female  $pGFP-ZAM$ )**D****E**



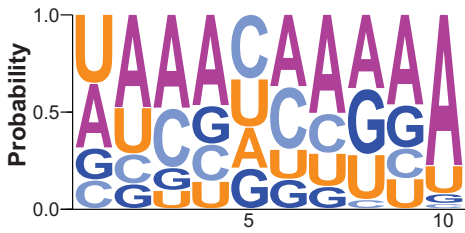


**Figure S4. ZAM-derived piRNAs originate from a germline piRNA cluster localized on the X chromosome.**

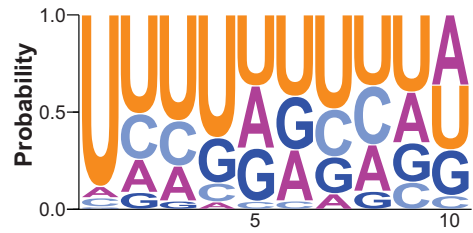
**a** Schema representing crosses performed to generate the ZMD and NZMD F1 progeny analyzed in Figure 4. The schema details both hypothesis for the new *ZAM* insertion in the RevI-H2 line. Only the X chromosome is displayed on the schema. The control line is the p*GFP-ZAM* line in which expression is driven in germline cells by a *nanos-Gal4* driver. **b-c** Confocal images of ovarioles after staining for GFP (green, left panels) and DNA (blue, middle panels). Merged images of the GFP and DNA signals are on the right. Ovarioles were from the progeny of a cross between  $w^{IR6}$  females and control males (*b*) and from the reciprocal cross between  $w^{IR6}$  males and control females (*c*). In both crosses, the p*GFP-ZAM* line in which expression is driven in germline cells by a *nanos-Gal4* driver was the control line. **d** Western blot analysis of proteins extracted from ovaries of progenies of crosses between  $X^{Rev}$ ,  $II^{Rev}$  or  $III^{Rev}$  females with control males. The control line was the same as in *b* and *c*. Proteins were from two biological replicates (1&2) prepared from 5 pairs of ovaries and  $\alpha$ -tubulin was the loading control. **e** Confocal images of ovarioles after staining for GFP (green, left panels) and DNA (blue, middle panels). Merged images of the GFP and DNA signals are on the right. Ovarioles were from the progeny of a cross between  $X^{Rev}$ ,  $II^{Rev}$  or  $III^{Rev}$  females with control males. The control line was the same as in *b* and *c*. **f** Structure of the region where the new *ZAM* insertion was identified in RevI-H2. The insertion is located in the dual-strand piRNA cluster 9 (according to ranking of piRNA clusters identified in *D. melanogaster* by [30]). *ZAM* is in genomic minus strand orientation. The chromosome coordinates are according to release 6 of the *D. melanogaster* genome. The *su(f)* genetic marker is indicated. The pericentromeric part of the X chromosome is displayed by a black box, euchromatin as a black line, *flam* and the other piRNA clusters by colored boxes. Primers (F and R) used to confirm the presence of the *ZAM* insertion by PCR (Fig. S4H) are depicted by red arrows. The *ZAM* insertion was localized in region X: 23,474,449..23,513,109 by genome sequencing of RevI-H2 (3 identical possible insertion sites). **g** The sequenced junctions between *ZAM* and the genomic flanks. The *ZAM* sequence is displayed in black capital letter and the genomic flanking sequence matching to cluster 9 in orange lower case letters. Target site duplications (TSD) are boxed. **h** PCR analysis of the new *ZAM* insertion identified in RevI-H2. Used primers are displayed in Fig. S4F. Only RevI-H2 presents an amplicon showing that the  $w^{IR6}$ , Iso1A and  $w^{1118}$  lines are devoid of this insertion and confirming the presence of this *ZAM* insertion located in cluster 9 in RevI-H2. **i** Western blotting of proteins extracted from ovaries of progenies of crosses between  $w^{IR6}$  or RevI-H2 and the same control line as in Fig. 4a. The lines used for the crosses are indicated above. Proteins were from two biological replicates (1&2) prepared from 5 pairs of ovaries;  $\alpha$ -tubulin was used as loading control. Images displayed in this panel are the unprocessed versions of those displayed in Fig. 4c.

# A

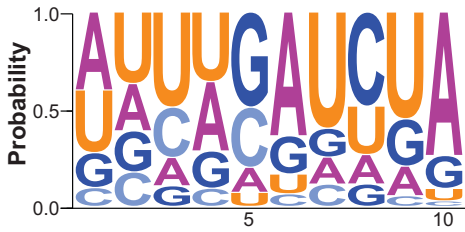
sense *Adoxo*-derived reads with PPP in  $w^{IR6}$



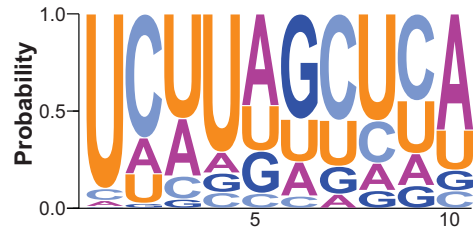
antisense *Adoxo*-derived reads with PPP in  $w^{IR6}$



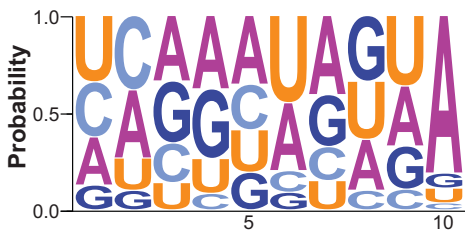
sense *Gedeo*-derived reads with PPP in  $w^{IR6}$



antisense *Gedeo*-derived reads with PPP in  $w^{IR6}$



sense *Idefix*-derived reads with PPP in  $w^{IR6}$



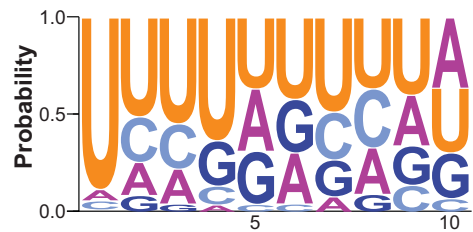
antisense *Idefix*-derived reads with PPP in  $w^{IR6}$



sense *Vatovio*-derived reads with PPP in  $w^{IR6}$

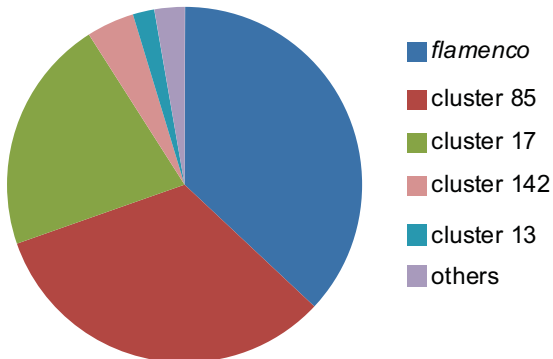


antisense *Vatovio*-derived reads with PPP in  $w^{IR6}$



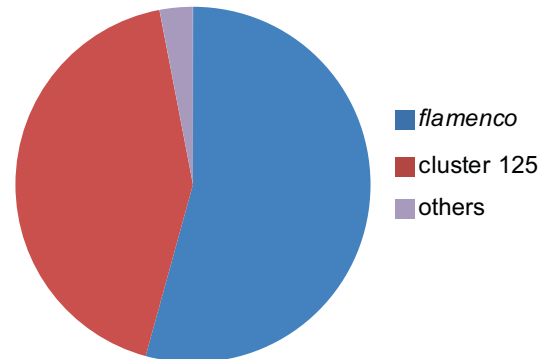
# B

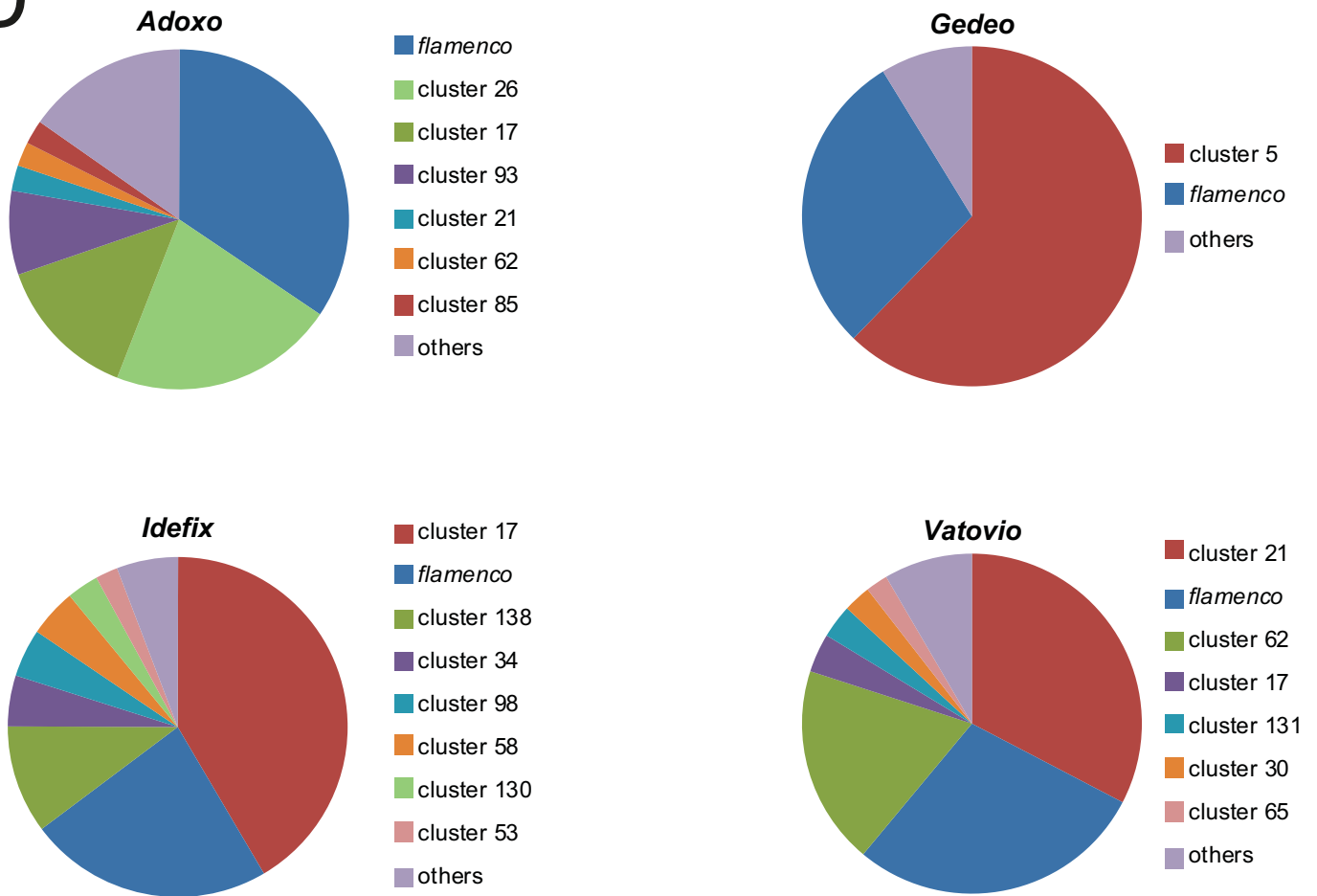
*Phidippo*



# C

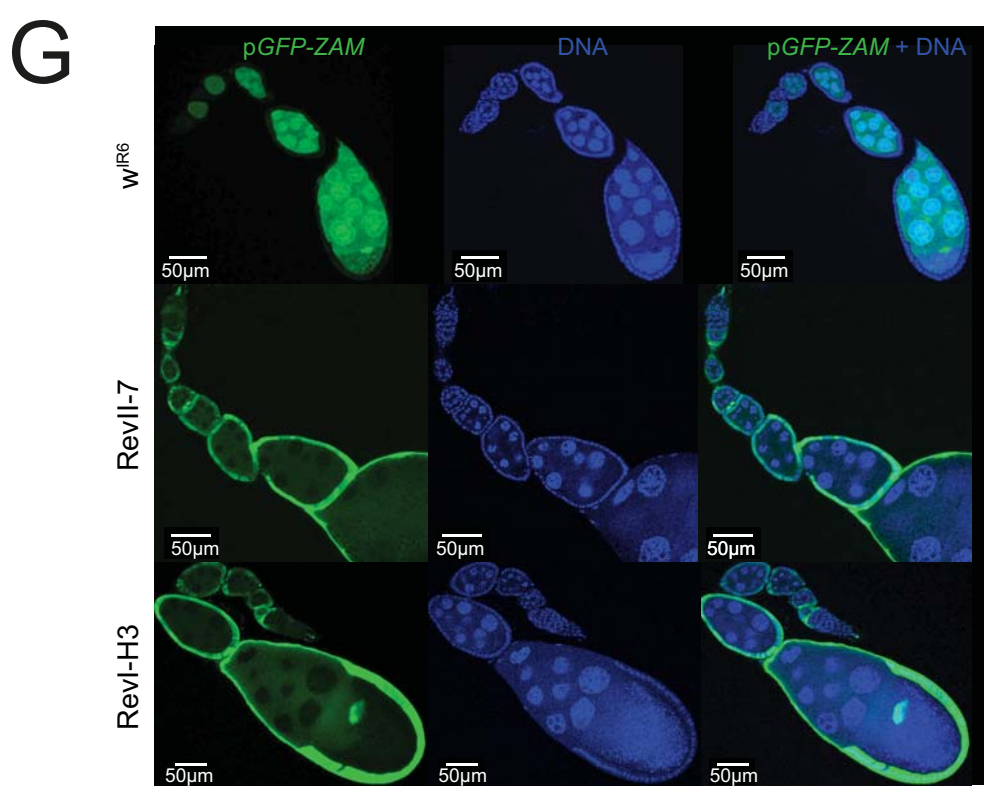
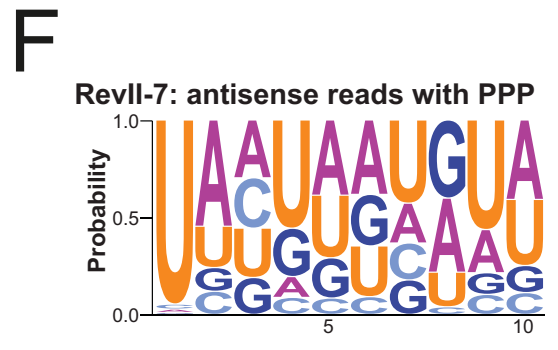
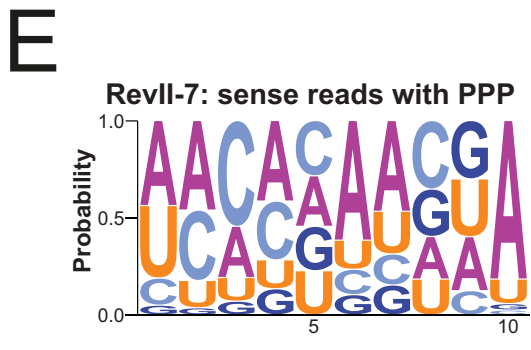
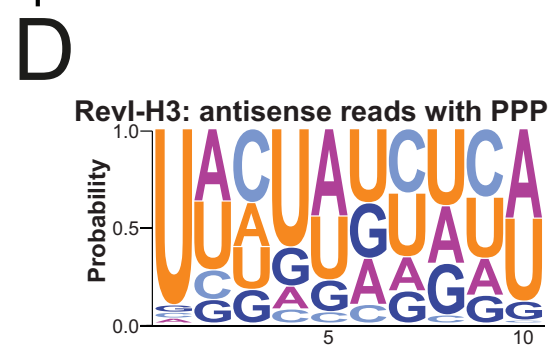
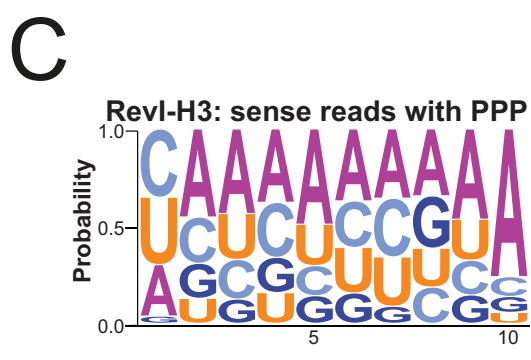
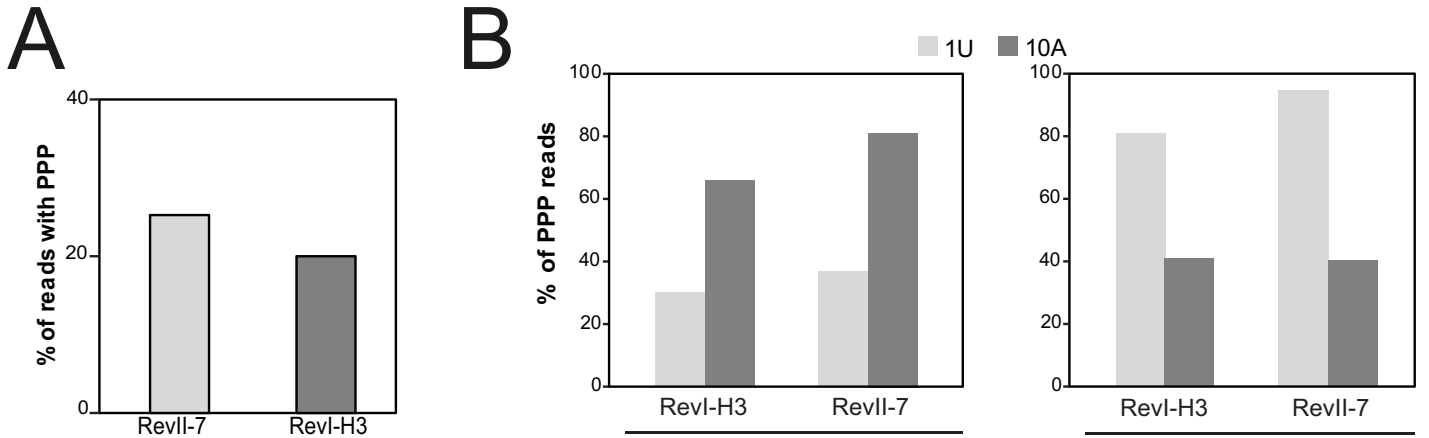
*Pifo*

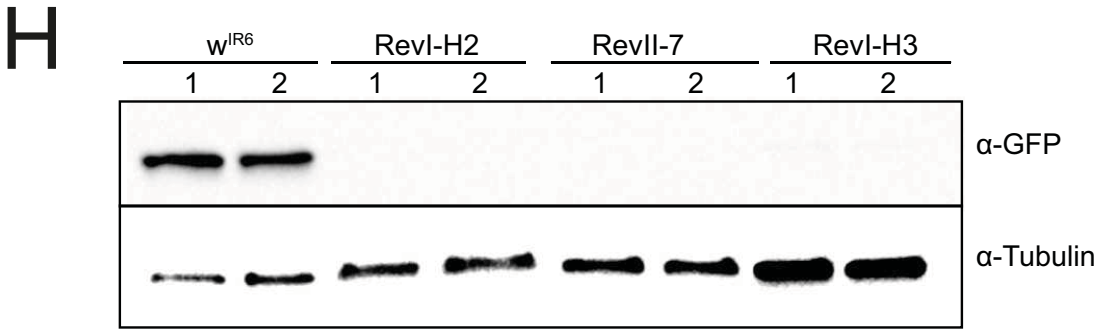


**D**

**Figure S5. *Phidippo*- and *Pifo*-derived piRNAs are mainly produced by the *flamenco* cluster.**

**a** Logo of nucleotide bias for the first ten positions of *Adoxo*-, *Gedeo*-, *Idefix*- and *Vatovio*-derived piRNAs with ping-pong partner (PPP) produced in  $w^{IR6}$  ovaries. The nucleotide height represents its relative frequency at that position. **b-c** Pie charts showing the proportion of *Phidippo*- (*a*) and *Pifo*-derived piRNAs (*b*) mapped (allowing up to 3 mismatches) to the 142 piRNA clusters (allowing no mismatch, piRNA clusters as in [9]) in the  $w^{IR6}$  line. **d** Pie charts showing the proportion of *Adoxo*-, *Gedeo*-, *Idefix*- and *Vatovio*-derived piRNAs mapped (allowing up to 3 mismatches) to the 142 piRNA clusters (allowing no mismatch, piRNA clusters as in [9]) in the  $w^{IR6}$  line.





**Figure S6. ZAM-derived piRNAs produced in the different Rev lines display similar features.**

**a** Percentage of ZAM-derived piRNAs with ping-pong partner (PPP) in the RevII-7 and RevI-H3 lines. **b** Analysis of the nucleotide bias for sense (+) and antisense (-) ZAM-derived piRNAs with PPPs in the RevII-7 and RevI-H3 lines. The percentages of PPPs with a 1U and a 10A are shown. Both lines had a 10A bias for sense piRNAs and a 1U bias for antisense piRNAs. **c-f** Logo of nucleotide bias for the first ten positions of sense (*c, e*) and antisense (*d, f*) ZAM-derived piRNAs with ping-pong partner (PPP) produced in RevI-H3 (*c-d*) and RevII-7 (*e-f*) ovaries. The nucleotide height represents its relative frequency at that position. **g** Confocal images of ovarioles after staining for GFP (green, left panels) and DNA (blue, middle panels). Merged images of the GFP and DNA signals are on the right. Ovarioles were from the progeny of crosses between *w<sup>IR6</sup>* (top), RevII-7 (middle) or RevI-H3 (bottom) females with males that harbor the pGFP-ZAM sensor transgene driven by the *actin-Gal4* driver. **h** Western blot analysis of proteins extracted from ovaries of progenies of crosses between *w<sup>IR6</sup>*, RevI-H2, RevII-7 or RevI-H3 females with a control male. The pGFP-ZAM line in which ZAM expression is driven in germline cells by a *nanos-Gal4* driver was the control line. Proteins were from two biological replicates (1&2) prepared from 5 pairs of ovaries and  $\alpha$ -tubulin was the loading control.





## **Partie 2 : Les piARNs germinaux permettent de contrer les invasions de rétrovirus endogènes provenant des cellules somatiques environnantes dans les ovaires de *Drosophila melanogaster*.**

Nous avons donc démontré dans l'étude précédente que des piARNs ciblant *ZAM* sont produits dans les cellules germinales de la lignée RevI-H2, probablement grâce à l'insertion de *ZAM* dans un cluster de piARNs « dual-strand » exprimé dans les cellules germinales. Nous avons alors émis l'hypothèse que piéger un ET dans un cluster de piARNs à la suite d'une invasion permettrait de protéger le génome de la lignée germinale contre de nouvelles invasions et transpositions. Dans cette deuxième étude, nous avons analysé plus en détail le rôle de ces piARNs ciblant *ZAM* produits dans les cellules germinales de la lignée RevI-H2.

Une étude génomique a été réalisée sur une lignée isogénique RevI-H2 dérivée récemment, nommée RevI-H2i2, qui a permis de révéler que le génome de cette lignée isogénique contient au moins dix-sept insertions de *ZAM* alors que seulement deux insertions sont présentes dans le génome de lignée de référence (Iso1A). La majorité de ces insertions étaient déjà présentes dans le génome de la lignée initiale RevI-H2, séquencée 10 ans auparavant. Ces résultats suggèrent que la lignée RevI-H2 a subi une invasion génomique par l'ET *ZAM* dans le passé mais qu'aucune nouvelle insertion n'a eu lieu depuis 10 ans : la lignée RevI-H2 semble aujourd'hui être stable. Cependant, des expériences de *single molecular* RNA FISH (smFISH) ont révélé que *ZAM* est toujours exprimé dans les cellules folliculaires de la lignée RevI-H2i2. Ainsi, nous avons analysé si les piARNs produits dans les cellules germinales de la lignée RevI-H2i2 peuvent contrer l'invasion de *ZAM* et ainsi protéger les cellules germinales de nouvelles insertions génomiques. Pour cela, nous avons altéré la production des piARNs dans les cellules germinales de la lignée RevI-H2 et analyser la localisation cellulaire des ARN de *ZAM*. L'abolition de la voie des piARNs dans les cellules germinales de la lignée RevI-H2i2 induit une forte invasion de l'ovocyte. De nombreux ARNs de *ZAM* sont retrouvés dans le cytoplasme de l'ovocyte alors qu'aucun ARNm de *ZAM* n'est détecté dans l'ovocyte d'une lignée RevI-H2i2 où la voie des piARNs est fonctionnelle. Ces résultats suggèrent que la production de piARNs ciblant *ZAM* dans les cellules germinales de la lignée RevI-H2i2 permet de dégrader post-transcriptionnellement les ARNs de *ZAM* provenant des cellules folliculaires.

Enfin, en utilisant la technique d'édition du génome CRISPR-Cas9, j'ai éliminé *ZAM* du cluster de piARNs *flamenco* (lignée CRISPR *flam* $\Delta$ *ZAM*) afin d'analyser l'impact direct sur la régulation et l'expression de *ZAM*. Cette lignée ne présente aucune autre insertion de *ZAM* dans





aucun autre cluster de piARNs. La délétion *flam* $\Delta$ *ZAM* induit une perte quasi-totale des piARNs ciblant *ZAM*. La perte des piARNs régulateurs de *ZAM* induit une forte dérégulation de *ZAM* dans les cellules folliculaires, ainsi qu'une invasion massive de l'ovocyte par *ZAM*. Ce phénomène est associé à une atrophie sévère des ovaires et à une subfertilité des femelles. Ces observations démontrent qu'en absence de production de piARNs contre cet élément dans les cellules germinales, *ZAM* envahit l'ovocyte puisque ses ARNs ne peuvent pas être dégradés post-transcriptionnellement. De plus, aucune réponse primaire faisant intervenir des petits ARNs, tels que des piARNs ou des siARNs, n'a été détectée à la suite de la réactivation de *ZAM*. En conclusion, cette étude a permis de montrer que la production de piARNs par les cellules germinales semble être une réponse adaptative efficace pour contrer l'invasion par un rétrovirus endogène provenant des cellules somatiques adjacentes. Notre étude révèle pour la première fois que les piARNs produits dans les cellules germinales, en plus de permettre la répression des ETs exprimés dans ces cellules, sont impliqués dans un mécanisme de protection plus large à l'échelle tissulaire.

## ARTICLE 2



# Germline piRNAs counteract retrotransposon invasion from somatic cells.

Marianne Yoth<sup>1</sup>, Nathalie Gueguen<sup>1</sup>, Silke Jensen<sup>1</sup>, Emilie Brasset<sup>1</sup>

<sup>1</sup> GReD, Université Clermont Auvergne, CNRS, INSERM, Faculté de Médecine, 63000 Clermont-Ferrand, France.

## Abstract

In the metazoan gonads, TE mobilization is limited by small RNAs called PIWI-interacting RNAs (piRNAs). piRNAs arise from heterochromatic source loci termed piRNA clusters that are expressed in a tissue-specific manner. The relationship between piRNA clusters and the control of TEs is far from being completely understood. In the present study, we show how piRNAs produced locally, in the germ cells, counteract endogenous retrovirus invasion arriving from adjacent tissue. While the piRNA pathway is fully functional, we succeed to reactivate an endogeneous retrovirus, *ZAM*, by deleting, using Crispr CAS9, the unique *ZAM* copy inserted in a piRNA cluster. As long as no *ZAM*-derived-piRNA is produced in the germline, reactivation of *ZAM* in the somatic cells promotes oocyte invasion resulting in fertility defects. Surprisingly we found that no initial small RNA response takes place soon after the reactivation. Our data reveal how cooperation between somatic and germline piRNAs can lead to an efficient control of endogenous retrovirus at a tissue scale.



## Introduction

The gonad is a major organ in animals because of its role in producing germ cells, the only cells in an organism that transfer genetic information to the next generation. However, germ cells require a set of supporting somatic cells, called gonadal somatic cells. Studies in model organisms such as *Caenorhabditis elegans*, mice and drosophila have revealed the essential role of somatic cells in regulating germline development<sup>1-5</sup>.

Transposable elements (TEs), discovered by Barbara McClintock in the 1940's<sup>6</sup>, are mobile genetic entities, which are present in nearly all organisms. The integrity of the eukaryotic genome relies on transcriptional and post-transcriptional repression of TE activity. In animals, genome defense against TEs in the gonads relies essentially on the piRNA pathway. piRNAs, the central players in this regulatory pathway, are small RNAs of 23 to 29 nucleotides interacting with proteins of the PIWI family<sup>7-9</sup>. They originate from transposon-rich genomic loci called piRNA clusters<sup>10-12</sup>. In *Drosophila*, there are germline and soma-expressed piRNA clusters. Although most piRNA clusters are expressed in germ cells in a bidirectional manner and produce piRNAs from both genomic strands, other clusters are specifically expressed in somatic cells surrounding the ovary and in a unidirectional manner producing mostly antisense transposon piRNAs<sup>10</sup>. Transposons were then classified according to their targeting by germline, somatic, or both piRNAs. For example, HeT-A was classified as a germline, accord2 as an intermediate, and *ZAM* as a somatic transposon<sup>13</sup>.

Surprisingly, while a depletion of piRNA pathway components in the germline or somatic lineage leads to TE derepression and sterility, a deletion of the three largest germline clusters, has no impact neither on transposon regulation nor on sterility<sup>14-16</sup>. However, *flamenco* (*flam*), the major somatic piRNA cluster, is reported to be directly involved in the regulation of at least 3 TEs: *ZAM*, *Idefix* et *Gypsy*<sup>17-19</sup>. Indeed, control of different retrotransposons by the *flam* locus had been shown for diverse *flam* alleles: for *ZAM* and *Idefix* in the RevI-H2 line and for *Gypsy* and *ZAM* in *flam1*, *flam-KG*, *flam-BG* mutant lines<sup>20-22</sup>. The significant differences of TE control in germline and somatic cells highlight that the relationship between piRNA clusters and the control of TE is far from being completely understood.

Using the RevI-H2 line, displaying a large deletion of the proximal (i.e., the region closest to the centromere) part of *flamenco* corresponding to the region containing its only *ZAM* insertions, we previously showed that, in this line, *ZAM* somatic piRNA production is abolished while germline sense and antisense *ZAM*-derived piRNAs are newly produced<sup>23,24</sup>. However,

whether and how germline piRNAs contribute to protect the germline against invasion of TEs coming from surrounding somatic cells is not understood.

Here we report, by using *ZAM* as a model, that when a retrotransposon is derepressed in the somatic cells surrounding germ cells, it invades the germline. Furthermore, no initial small RNA response takes place after invasion of the germ cells. Many years after *flam* mutation appeared in the RevI-H2 line, *ZAM* is still expressed in somatic cells surrounding germ cells and no new *ZAM* insertion in the *flamenco* piRNA cluster has been detected. However, once the germline produces piRNAs against this new invader, either from new inserted dispersed copies or because it has jumped into a piRNA cluster expressed in the germline, we revealed an unexpected arrest of the invasion while the retrotransposon is still activated in the surrounding tissue. Our findings have revealed the critical role of the germline piRNA pathway as an effective defense mechanism to counteract transposon invasion arriving from adjacent tissue. Our work provides a unique opportunity to better understand the role of each type of piRNA cluster in TE regulation at the cellular and tissue scales.

## **Material and Methods**

### **Fly stocks, transgenic lines, and crosses**

Flies were maintained at 25°C under light/dark cycles and 60% humidity. Flies between 3 and 6 days after hatching were used for experiments. The RNAi lines against actors of the piRNA pathway used in this study are listed in **Supplementary Table S1**. Isogenic lines from the RevIH2 stock were generated by crossing during 5 generations a single female with a single male.

### **RNA fluorescent in situ hybridization (smRNA FISH) on ovaries and embryos.**

*ZAM* smFISH was performed using 48 unique probes targeting *ZAM* transcripts on a region that is absent from the *ZAM* inserted in the *flamenco* piRNA cluster to detect only transcripts of active *ZAM* copies (oligo sequences in **Supplementary Table S2**). Ovaries from 3 to 6-day-old flies were dissected in Scheider's *Drosophila* Medium and fixed in Fixing Buffer (4% formaldehyde, 0.3% Triton X-100, 1x PBS) for 20min at room temperature, rinsed three times in 0,3% Triton X-100, 1x PBS, and permeabilized in 70% ethanol at 4°C overnight. Permeabilized ovaries were rehydrated in RNA-FISH wash buffer [10% formamide in 2x SSC] for 10 min. Ovaries were resuspended in 50µL hybridization buffer [10% dextran sulfate, 10% formamide in 2x SSC] supplemented with 1µL of RNA-FISH probes. Hybridization was performed at 37°C overnight with rotation. Ovaries were then washed twice with RNA-FISH wash buffer for 30min at 37°C and two times with 2xSSC solution and DNA were stained with DAPI (1/500 dilution in 2x SSC) for 20min at room temperature. Ovaries were mounted in 30µL vectashield mounting medium and imaged on a Zeiss LSM- 980 or Zeiss LSM-800 confocal-microscope. The resulting images were processed using FIJI/ImageJ. RNaseA

treatment was performed to confirm the RNA signal specificity by adding 1mg/mL of RNaseA in 2x SSC for 2 hours before the hybridization step.

For embryo staining, flies were caged and fed yeast paste. Embryos (0-2h) were collected, dechorionated in 50% bleach for 4min and rinsed in water. Eggs were fixed in 4% paraformaldehyde/heptane for 20min, devitellinized by vigorous shaking in 100% Methanol and stored in methanol at -20°C. Embryos are rehydrated with 1:1 Methanol: PBT (1xPBS, 0,1% Tween-20) for 5 min and 2 times in PBT. Embryos were resuspended in RNA-FISH wash buffer (10% formamide in 2x SSC) for 10 min and were processed for RNA FISH as described for ovaries. Immunostaining combined with RNA FISH was performed by adding primary antibody to the RNA-FISH probes in the hybridation buffer and incubate overnight at 37°C. Embryos were washed twice with RNA-FISH wash buffer and incubated 90min with the secondary antibody (1/1000 diluted in 2x SSC). After 2 final washes in 2x SSC, embryos were mounted in 30 $\mu$ L vectashield mounting medium.

### **Immunofluorescence on ovaries and embryos.**

Ovaries from 3- to 5-day-old flies were dissected in supplemented Schneider medium, and ovarioles were separated and the muscle sheath was removed before fixation to obtain undistorted follicles. Ovaries are then fixed in 4% formaldehyde/PBS 1X, 2% Tween-20 for 15min, rinsed three times with PBS 1X Tween 2% and incubate 2 hours in PBT (1x PBS, 0.1% Triton X-100, 1% BSA) for blocking. Ovaries were incubated in PBT containing primary antibodies at 4°C overnight (antibodies are described in **Supplementary Table S3**). After 3 washes in PBT, ovaries were incubated with the corresponding secondary antibodies (1/1000 diluted in PBT), coupled to Alexa-488 or Cy3 for 90 min. After two washes in 1x PBS, DNA was stained with DAPI (1/500 dilution in 1x PBS) for 20min. Ovaries were mounted in 30 $\mu$ L vectashield mounting medium and imaged on a Zeiss LSM- 980 or Zeiss LSM-800 confocal-microscope. The resulting images were processed using FIJI/ImageJ.

### **RT-qPCR analysis of transposon expression**

10-20 pairs of dissected ovaries were homogenized in TRIzol reagent (Ambion). Following DNase I treatment, cDNA was prepared from 1 $\mu$ g RNA by random priming of total RNA using Superscript IV Reverse Transcriptase (ThermoFisher Scientific). Quantitative PCR was performed with Roche FastStart SYBR Green Master and the Lightcycler® 480 Instrument. cDNA was used as template for RT-qPCR quantification of transposon mRNA levels (primer sequences in **Supplementary Table S4**). All experiments were conducted with three biological replicates and with technical triplicates. Steady-state RNA levels were calculated from the threshold cycle for amplification by the  $2^{-\Delta\Delta C_t}$  method. *rp49* was used for normalization (**Supplementary Table S1**).

### **Immunoprecipitation of Piwi RNP Complexes**

For each genotype, 50 pairs of ovaries from 3-5 days-old flies were dissected, lysed in 1 ml lysis buffer (20 mM HEPES-NaOH at pH 7.0, 150 mM NaCl, 2.5 mM MgCl<sub>2</sub>, 250 mM Sucrose, 0.05% NP40, 0.5% Triton X-100, 1x Roche-Complete). Samples were cleared by centrifugation at 10000  $\times$  g for 10 min at 4°C. Extracts were incubated for 4h at 4°C with rotation with rabbit polyclonal anti-Piwi antibodies (4 $\mu$ g per sample) and the whole was incubated overnight at 4°C with rotation with Dynabeads Protein A (50  $\mu$ l, invitrogen, 10002D). The beads have been equilibrated with NT2 buffer (NT2: 50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1 mM MgCl<sub>2</sub>, 0.05% NP40) prior to the incubation. The beads were washed twice with ice-cold NT2 and twice with NT2 with the concentration of NaCl adjusted to 300 mM. Nucleic



acids that co-immunoprecipitated with Piwi were isolated by treatment of the beads with 0.7mg/ml proteinase K in 0.3ml proteinase K buffer (0,5%SDS, 10mM TrisHcL pH=7,4, 50mM NaCl, 5mM EDTA), followed by phenol (at neutral pH)/chloroform extraction and ethanol precipitation.

### **Small RNA sequencing**

Regulatory small RNA extraction was performed as described in <sup>25</sup>. In brief, 50 pairs of ovaries from 3-5 days-old flies were lysed and Argonaute-sRNA complexes were isolated using TraPR ion exchange spin columns. sRNAs were subsequently purified using TRIzol (Ambion). Total RNA extraction has also been performed for some samples (80- 100 ovaries for each sample) using classical TRIzol extraction, 2S depletion and size selection have been realized before the sequencing. All sequencing have been performed by Fasteris S.A (Geneva/CH) on an Illumina NextSeq550 instrument (13-15 millions of reads per sample).

Illumina small RNA-Seq reads were loaded into the small RNA pipeline sRNAPipe (Pogorelcnik et al., 2018) for mapping to the various genomic sequence categories of the *D. melanogaster* genome (release 6.03). For the analysis, 23-29nt genome mappers were selected as piRNAs. 21nt genome mappers were selected as siRNAs. Unique-piRNAs are designed as 23-29 reads that mapped uniquely across the reference genome. piRNAs or unique-piRNAs were mapped to TEs (0-3 mismatches) or to piRNAs clusters, as defined by (Brennecke et al, 2007) (no mismatch allowed). The window size was of 91nt for *ZAM* and 140nt for the *flamenco* region (X:21,769,891...21,789,891) to establish the density profile of piRNAs. To allow for comparison between samples, all read counts were normalized by the number of reads used as input (piRNA genome or unique piRNA genome). The ping-pong signature was assessed by counting the proportion of sense piRNAs with an overlap of 10 nt with antisense piRNAs, based on piRNAs mapping to *ZAM* (0 to 3 mismatches). The proportions of 1- to 28-nt-long overlaps were determined, and the percentage of 10-nt overlaps defined as ping-pong signature. The Z-score was determined on the proportions of 1- to 23-nt-long overlaps and considered significant for values > 1.96. Scatter plot were done with RStudio with antisense piRNA or siRNA mapping to individual internal sequence of TEs (0-3 mismatches for piRNAs and 1-3 mismatches for siRNAs). All small RNA sequencing realized for this study are available in **Supplementary Table S5**.

### **Sterility test**

For the fertility tests, RevIH2i2 homozygous females were compared to w<sup>1118</sup> females and homozygous CRISPR *flamΔZAM* females carrying two functional *ZAM* copies obtained from the RevI-H2i2 X chromosome were compared to CRISPR *flamΔZAM* females without functional *ZAM* copie in their genome. Thirty 3-4 days old virgin females of each genotype were individually mated with three w<sup>1118</sup> males and eggs were collected during 24 hours at 25°C. The number of eggs laid by each individual female was determined and eggs were then kept at 25°C for another 24 hours before being counted to determine the egg hatching rate. The experiment has been made twice for a total of approximatively 60 replicates for each condition.

### **Generation of the CRISPR *flamΔZAM* line by CRISPR/Cas9**

Deletion of the *ZAM* copy inserted in the *flamenco* piRNA cluster was generated using the pCFD6 plasmid expressing two single-guide RNA under of the Gal4/UAS system. Oligonucleotides were design targeting unique regions located upstream and downstream of *ZAM-flam*: sgRNA\_ *flamΔZAM\_1*: TTGTAGCGCTCTTCTTCTCT, sgRNA *flamΔZAM\_2*: AGCGCAACCACGTACAGCGA. The plasmid has been injected by Bestgene company in embryos from the stock #9736 (integration genomic of the plasmid on the site 53B2). Obtained

flies were then crossed with *nanos-gal4; UAS-Cas9* flies (BL#54593). After establishment of 100 stocks, homozygote females were screened by PCR using primers specific of the *ZAM-flam* (no amplification = positive flies). Positive lines were selected for PCR amplification of the target locus and PCR amplicons were subjected to sequencing.

## Results

### **1- In an ancient *flamenco* mutant line, *ZAM* is still expressed and no new *ZAM* copy is inserted in the *flamenco* locus.**

In the RevI-H2 line, a deletion of the proximal part of *flamenco* that eliminates the region in which *ZAM* was inserted, a complete loss of *ZAM* piRNAs produced by *flamenco* in the somatic cells has been described<sup>23</sup>. The, RevI-H2i2, RevI-H2i3 and RevI-H2i4 are 3 isogenic lines derived from the RevI-H2 line that has more than 25 years laboratory history<sup>26</sup>. The *ZAM* transposon actively transposed in the initial RevI-H2 line but it was not known whether *ZAM* is still transcriptionally active in the actual RevI-H2-derived lines. We show that the 3 isogenic Rev-H2 lines are still permissive for *ZAM* expression and have a very similar *ZAM* expression pattern (**Fig 1a, Supplementary Fig.1a**). Indeed, smFISH experiments (Single Molecule Fluorescence In Situ Hybridization) revealed that *ZAM* RNA is produced very early during oogenesis in the somatic cells of the germarium in some ovarioles. Then, after stage 8, *ZAM* transcripts accumulate in a patch of follicle cells located at the posterior side of the follicles (**Fig. 1a**). No *ZAM* transcripts were detected in the germ cells, neither in the nurse cells nor in the oocyte. RNase A treatment led to a complete loss of *ZAM* staining and no *ZAM* RNAs are detected in the *w<sup>1118</sup>*, *w<sup>118</sup>* or Iso1A control ovaries (**Supplementary Fig.1b,c**). We then focused our studies on the RevI-H2i2 line and validated these smFISH results by RT-qPCR confirming that *ZAM* is derepressed in the RevI-H2i2 ovaries while other soma-specific TEs which are not included in the deletion, such as *Gypsy*, or germline-specific TEs, such as *Burdock*, are not upregulated (**Fig. 1b**). *ZAM* is an endogenous retrovirus encoding *Gag*, *Pol* and *Env* proteins. We visualized *ZAM*-encoded *Gag* and *Env* proteins by immunostaining and observed that both proteins accumulate in the posterior follicle cells in late follicles (**Fig. 1c**). All these results showed that *ZAM* is still expressed in the RevI-H2i2 line. The long reads obtained by Oxford Nanopore Technology (ONT) sequencing of the RevI-H2i2 genome allowed us to identify *ZAM* insertions, even when they are inserted in repeated sequences, but no *ZAM* insertion was detected in the *flamenco* piRNA cluster. Thus no *ZAM* piRNAs can be produced from *flamenco*, the major somatic piRNA source locus, explaining why *ZAM* is still expressed in the follicle cells in this line. Although *ZAM* is expressed in the somatic follicle cells, the fertility of this RevIH2i2 line is not affected. All together these results suggest that the expression of *ZAM* in the somatic follicle cells has no deleterious effect on the RevI-H2i2 line.

## **2-The RevI-H2i2 is a stable line although ZAM is actively expressed in the follicle cells.**

We took advantage of the ONT long-read sequencing of the RevI-H2i2 genome and identified in total 18 *ZAM* copies in the RevI-H2i2 genome (**Fig. 2a**). Only one of these *ZAM* copies is also present in the reference genome (<http://flybase.org>) and 17 are new insertions that are very similar to the *ZAM* reference element, which had been identified in the initial RevI-H2 line<sup>26</sup>. Of these, 9 are full-length *ZAM* elements, one of these has a deleted 5'-UTR, 3 have large internal deletions in the coding regions, and 5 are what we called "*ZAM*-variants", they are full-length but have one same small deletion at the C-terminal end of the *pol* gene. This small deletion corresponds to positions 5494-6120 in *ZAM* internal sequence (Repbase *ZAM\_I*, <https://www.girinst.org/repbase/>,<sup>26</sup>). This deleted part of the *pol* gene doesn't correspond to any known protein domain and the fact that there are 5 *ZAM*-variants at different locations in the RevI-H2i2 line indicates that this *ZAM* copy is competent for transposition.

The full-length copies and *ZAM*-variants all potentially encode *Gag*, *Pol* and *Env* proteins. 14 of the 17 new *ZAM* insertions could be clearly localised (**Fig.2a**). Interestingly, in the initial RevI-H2 line, a *ZAM* insertion had been found in a germline piRNA cluster, piRNA cluster 9 close to the X chromosome centromere<sup>12,23</sup>. Here, thanks to ONT sequencing, we identified in all three *ZAM* insertions in this piRNA cluster 9, one full-length and two *ZAM*-variants, and another full-length *ZAM* insertion that is either in piRNA cluster 13 or in piRNA cluster 56 (piRNA clusters as in<sup>12</sup>).

Unexpectedly, when we reassessed the Illumina sequencing results performed ten years ago on the initial RevI-H2 line in light of the latest ONT results, we found that at least 14 of all present new *ZAM* insertions in RevI-H2i2, including all *ZAM* insertions in piRNA clusters, were already present ten years ago. This result indicates that the genomic *ZAM* profile is the same as ten years ago and suggests that no new *ZAM* insertions occurred since then, that the RevI-H2 line is stable.

## **3- Germline ZAM piRNAs counteract ZAM invasion**

We aimed to find out how the RevI-H2i2 line can be stable despite *ZAM* expression in the somatic follicle cells. As previously reported<sup>27</sup>, the *ZAM* transposon actively transposed in the initial RevI-H2 line and four *ZAM* insertions were now found in germline piRNA clusters. The germline cells of RevI-H2 produce high amounts of *ZAM*-derived piRNAs with a ping-pong signature (i.e. an enrichment of 10-nucleotide overlap between sense and antisense *ZAM*-derived piRNAs). These piRNAs are efficient to silence a *ZAM* sensor transgene in the germline<sup>23</sup>. The sensor expresses the GFP reporter gene under the control of an inducible Upstream Activation Sequence promoter (UASp) and harbors a *ZAM* fragment in its 3'UTR (pGFP-*ZAM*).

We first checked that the RevI-H2i2 line harbor the same characteristics. The RevI-H2i2 line disclosed sense and antisense *ZAM*-derived-piRNA production with a ping-pong signature and silenced the *ZAM*

sensor transgene in the germ cells (**Fig. 2b, c, d**). We obtained the same results with the RevI-H2i3 and RevI-H2i4 (**Supplementary Fig. 2a, b, c**). Furthermore, we validated that all three RevI-H2 isogenic lines harbor the deletion of *ZAM* copy from *flamenco* and insertions of *ZAM* in the germline piRNA cluster 9 (**Supplementary Fig. 2d**).

To test whether piRNAs produced in germ cells can thwart *ZAM* invasion from somatic cells to the germline, we abolished the piRNA pathway in the germ cells, by germline-specific knock-down (GLKD), in RevI-H2i2 genetic background, and followed the subcellular localization of *ZAM* RNA in ovaries. We first verified that we indeed lost *ZAM* piRNAs. We observed a drop in the level of antisense piRNAs mapping *ZAM* and germ-specific TEs in ovaries in which the piRNA pathway genes *argonaute3 (ago3)*, *zuccini (zuc)*, or *piwi* were knocked down in a RevI-H2i2 genetic background (**Fig. 3a**). Furthermore, in Ago3-GLKD condition, the ping-pong signature for *ZAM* and for many germline specific TEs was abolished, while ping-pong is maintained in control white-GLKD and in *zuc*- and *piwi*-GLKD ovaries (**Fig. 3b, c**).

We then analyzed *ZAM* RNA subcellular localization by smFISH in the RevI-H2i2 control line (RevI-H2i2 white-GLKD) in which *ZAM*-derived piRNAs were absent from follicle cells but produced in the germline. *ZAM* RNA expression was restricted to follicle cells and almost no *ZAM* RNA signal was detected in the oocyte (**Fig. 3d**). In contrast, all lines with RevI-H2i2 background and germline-specific knockdown of piRNA pathway actors (*vreteno (vret)*, *zuc*, *ago3*, *armitage (armi)*) clearly showed *ZAM* expression in the posterior follicle cells and *ZAM* RNA invasion in the oocyte. Indeed, *ZAM* RNA spread throughout the ooplasm but was clearly enriched at the posterior pole adjacent to *ZAM*-expressing follicle cells. Actually, 30 to 70% of stage 10 follicles showed a strong signal of *ZAM* RNA in the oocyte (**Fig. 3d, e**). The strongest invasion phenotype was observed for the RevI-H2i2 Ago3-GLKD condition where we have reported a drastic decrease in the production of *ZAM*-derived piRNA (**Fig. 3a**). In this condition, the accumulation of *ZAM* RNA in the ooplasm is correlated with a strong increase of *ZAM* RNA present in total ovaries assayed by RT-qPCR (**Fig. 3f**). These results showed that *ZAM* RNAs transit from the somatic follicle cells where they are produced to the oocyte when *ZAM* derived piRNAs are absent or lowered in the germline. Taken together our data strongly suggest that *ZAM* piRNAs produced by the germline piRNA pathway trigger post-transcriptional silencing of *ZAM* RNA arriving from the somatic follicular cells. In line with this conclusion no *ZAM* RNA was detected in RevI-H2i2 stage 10 ooplasm and *ZAM* expression was not affected when Piwi, which acts through transcriptional gene silencing, was knocked down in the germline (**Fig. 3e, Supplementary Fig. 3a, b**).

We confirmed this finding by using a line where no *ZAM*-derived piRNAs are produced in the germline (*w<sup>1118</sup>* genetic background) and in which we knocked down the piRNA pathway in the follicle cells by somatic knock-down (sKD) of *armi*, *vret*, *piwi* or *yb*. sKD of a piRNA pathway actor such as *vret* or *yb* led to a strong decrease in *ZAM* piRNA production (**Fig. 4a, Supplementary Fig. 4a**). Antisense piRNA production in the somatic follicle cells was globally affected, the amount of Piwi-bound piRNAs that

map soma-specific TEs was strongly decreased upon the loss of a piRNA pathway actor (**Fig. 4b**, **Supplementary Fig. 4b**). *ZAM*, and other soma-specific TEs such as *Gypsy*, were derepressed as assessed by RT-qPCR (**Fig. 4c**). smFISH results revealed that *ZAM* RNAs were not restricted to the posterior follicle cells. Indeed, as in ovaries harboring germline-specific knockdown of the piRNA pathway in RevI-H2i2 background, *ZAM* RNA was detected in the oocyte. Actually, *ZAM* RNA was detected in the ooplasm in 90% of the observed stage 10 follicles (**Fig. 4d, e**, **Supplementary Fig. 4c**). This result confirmed that when no *ZAM* piRNAs are produced in the germline, *ZAM* RNA expressed in the somatic follicle cells transits to the oocyte.

#### **4- *ZAM* RNA transcribed in the follicle cells transits to the oocyte and is deposited in early embryos**

Even if *ZAM* ought to be transcribed only in follicle cells<sup>28,29</sup>, we wanted to rule out the possibility that *ZAM* RNAs originated from the germinal nurse cells. *ZAM* RNAs strongly accumulate at the posterior pole of the oocyte suggesting that these RNAs arrive from the adjacent follicle cells. Moreover, the *ZAM* RNA staining observed in the ooplasm in the RevI-H2i2 Ago3-GLKD condition was only detected late in oogenesis at stage 10 and no staining was observed at any stage in the cytoplasm of nurse cells (**Fig. 5a**). This pattern was very different from what is observed for Burdock transcripts, known to be expressed in the germline: Burdock RNAs were detected in the cytoplasm of the nurse cells and accumulated in the anterior part of the oocyte, close to the nurse cells, and this even in earlier stages (**Fig. 5a**). These data indicate that, while Burdock RNAs originate from the germinal nurse cells, the *ZAM* RNAs detected in the oocyte originate from the somatic follicle cells.

Throughout the animal kingdom, the first stages of embryonic development are controlled by transcripts and proteins deposited by the mother during oogenesis<sup>30</sup>. In *Drosophila*, most of the mRNAs transmitted to the embryo are dumped from nurse cells into the oocyte during oogenesis. Although *ZAM* transcripts are originating from somatic follicle cells, when *ZAM* RNA transmission to the oocyte was observed, they were also found in the early embryos. Indeed, smFISH experiments performed on early embryos originating from a RevI-H2i2 mother in which the piRNA pathway was knocked down in the germline, or from a mother without germinal *ZAM* piRNAs in which the piRNA pathway was knocked down in the somatic follicle cells revealed a strong accumulation of *ZAM* RNAs in early embryo (before the zygotic transition - 2 hours after laying) (**Fig. 5b**, **Supplementary Fig. 5a**). Furthermore, *ZAM* RNAs accumulated at the posterior pole of early embryos where future germ cells will cellularize (**Fig. 5c – Supplementary Fig. 5b**). *ZAM* transcripts are then detected until the cellularization of the blastoderm around stage 8-9 of embryogenesis (**Fig. 5d**). In the progeny of the RevI-H2i2 line, no *ZAM* RNAs were detected, although *ZAM* RNAs are produced in the somatic follicle cells of the RevI-H2i2 ovaries confirming that *ZAM* invasion of the germline is not allowed in this line (**Fig. 5b**). Taken together our data revealed an efficient post-transcriptional silencing of *ZAM* RNA arriving from somatic follicle cells

into the oocyte by piRNAs produced in the germline and dumped into the oocyte thus likely limiting the transposition of the *ZAM* somatic retrotransposon into the germline genome.

### 5- *de novo* *ZAM* reactivation leads to subfertility

Loss of the piRNA pathway in the germline or in the somatic follicle cells leads to derepression of many TEs and results in DNA damage and developmental aberrations<sup>31-34</sup>. We sought to exclude the possibility that these defects could underlie the *ZAM* invasion into the oocyte observed when the piRNA pathway is knocked down. To this end, using CRISPR-Cas9, we *de novo* deleted the *ZAM* copy in the *flamenco* piRNA cluster in a line carrying the X chromosome of the IsoA1 reference line and in which the piRNA pathway is fully functional. In this line, the *flamenco* *ZAM* copy is localized at the genomic position X:21,778,810...21,783,994 and we used two guides to create a deletion spanning over X:21,777,135...21,784,062 (6926 pb). We have named the resulting line the *flam* $\Delta$ *ZAM* line. Mapping of unique piRNAs to the *flamenco* locus highlighted the complete loss of piRNA production at the deleted region compared to a control line (**Fig. 6a**). We noted that only piRNAs produced at the position where *ZAM* was inserted are lost, the global production of unique piRNAs mapping upstream and downstream of *ZAM* is not affected. We confirmed that the deletion induced a strong decrease of all piRNAs mapping to internal regions of *ZAM* (allowing 0-3 mismatches). However antisense piRNAs targeting the *ZAM* LTR were still produced (**Fig. 6b**). In line with these results, we observed by PCR amplification and DNA sequencing of the CRISPR-Cas9 target locus that only the internal regions of *ZAM* had been deleted from the *flamenco* cluster resulting in the retention of a solo-LTR at the initial *ZAM* insertion site.

In addition, only the production of *ZAM*-derived antisense piRNAs was strongly altered, the production of antisense piRNAs mapping other TEs was not affected (**Fig. 6c**). This result shows that the deletion of *ZAM* from *flamenco* impairs *ZAM*-derived piRNA production and does not affect the global production of piRNAs in ovaries. Thus, we ended up with a fly line in which the piRNA pathway was functional and no *ZAM* piRNAs were produced. We then analysed the expression of *ZAM* RNA by smFISH on *flam* $\Delta$ *ZAM* ovaries. We expected *ZAM* derepression but surprisingly no *ZAM* expression was detected in the ovaries of the *flam* $\Delta$ *ZAM* line (**Fig. 6d- panel 1**). We hypothesized that there was no functional *ZAM* copy in the *flam* $\Delta$ *ZAM* line. To test this, we exchanged either the chromosome II or III by the chromosome II or III of the RevI-H2i2 line, which contain respectively 4 or 3 potentially functional *ZAM* copies. Indeed, considering that these copies transposed recently in the RevI-H2i2 line it is likely that they are still functional. *ZAM* smFISH revealed that *ZAM* RNA is strongly expressed when these RevI-H2i2 *ZAM* copies are added to the genome of the *flam* $\Delta$ *ZAM* line (**Fig. 6d – panels 2 and 3**). We also observed a strong invasion of the oocyte, approximately 70% of the stage 10 follicles presented *ZAM* RNA in the ooplasm. We observed a similar expression pattern when we added the 2 *ZAM* copies located close to the telomere of the X chromosome of the RevI-H2i2 to the *flam* $\Delta$ *ZAM*

genome (**Fig. 6d – panel 4**). Overall, our findings show that in absence of *ZAM*-derived piRNAs neither in the germline nor in the somatic follicle cells, transcripts from functional *ZAM* copies massively invade the oocyte. As mentioned previously, a solo-LTR of *ZAM* is still present in the *flamenco* piRNA cluster of the *flamΔZAM* line. Our results demonstrate that the presence of this ancestral *ZAM* LTR of 454bp is not sufficient to silence *ZAM* expression in follicle cells (**Supplementary Fig. 6**)

Moreover, the *flamΔZAM* flies containing active copies of *ZAM* displayed atrophied ovaries associated with reduced fertility compared to the *flamΔZAM* line lacking functional *ZAM* copies (**Fig 6. e,f,g**). 90% of *flamΔZAM* females did not lay eggs and the hatching rate of the few eggs laid was about 50%. Taken together, our data establish that the reactivation of one single retrotransposon in a patch of somatic follicle cells is sufficient to induce severe fertility defects. On the other hand, the RevI-H2 and RevI-H2-derived lines are totally fertile. These results argue that when *ZAM* piRNAs are produced in the germline, they contribute to ensure genome integrity and fertility by protecting the oocyte against *ZAM* invasion from adjacent cells.

## **6- No initial small RNA response takes place after *ZAM* retrotransposon reactivation**

The reactivation of an endogenous retrovirus such as *ZAM* in a genetic context where the piRNA pathway is functional provides a unique opportunity to directly examine the genomic immune response to endogenous retroviral reactivation and germline invasion. piRNAs and siRNAs are the two main classes of small RNAs produced to respectively control the expression of transposable elements and counteract viral infection. To characterize the small RNA response upon endogenous retrovirus reactivation we sequenced small RNAs that are complexed with Argonaute proteins (named here "regulatory" piRNAs) from the *flamΔZAM* ovaries containing 4 or 3 functional *ZAM* copies (introduced with RevI-H2i2 chromosome II or III respectively). Regardless of the *flamΔZAM* line analyzed, *de novo* *ZAM* piRNAs of 23 to 29 nt were not produced upon *ZAM* reactivation and this despite germline invasion. Indeed, only few sense and antisense *ZAM* piRNAs were produced in the ovaries of the *flamΔZAM* line containing functional *ZAM* copies (**Fig. 7a – Supplementary Fig. 7a**). We also note that *ZAM* was the only one in this case and that the level of antisense regulatory piRNAs targeting other TEs was not reduced (**Supplementary Fig. 7b**).

sKD of the piRNA pathway on the contrary affected the piRNA levels for all somatic TEs including *ZAM* (*vret*- and *yb*-sKD) (**Fig. 4b, Supplementary Fig. 4b**). This drop in antisense regulatory *ZAM* piRNA allowed *ZAM* expression in the follicle cells and a strong invasion of the oocyte, as observed in the *flamΔZAM* line. In the same *vret*-sKD line, this event did not induce production of *ZAM*-derived piRNAs, only very few regulatory *ZAM*-derived piRNA were detected (**Fig. 4a, Supplementary Fig. 4a**). Thus, under these different conditions, where *ZAM* is reactivated and invades the oocyte, we did not detect any new production of regulatory piRNAs that could prevent *ZAM* activity.

Surprisingly we noted that when we sequenced total small RNA, not only those bound to Argonaute proteins, sense *ZAM*-mapping small RNAs of various size, including 23-29nt small RNAs, were observed upon *vret*- or *yb*-sKD, while these were not found in the *white*-sKD control ovaries (**Fig. 7b – Supplementary Fig. 7c**). However, sense 23-29nt *ZAM*-derived small RNAs detected in *yb*- or *vret*-sKD ovaries do not display uridine bias at the 5' end, a feature of mature primary piRNAs (**Fig. 7c**). Furthermore, these sense *ZAM*-mapping small RNAs are not immunoprecipitated with the Piwi protein. Almost no sense or antisense *ZAM*-mapping Piwi bound piRNAs were detected after *vret*- or *yb*-sKD (**Fig. 7d, Supplementary Fig. 7d**). These results suggest that these small RNAs mapping *ZAM* sequence are probably not piRNAs.

Since *ZAM* is an endogenous retrovirus, we investigated whether a siRNA response was triggered upon *ZAM* reactivation. For this, we analyzed the production of 21-nt small RNA complexed with Argonaute proteins ("regulatory" siRNAs). We observed that a similar level of *ZAM* 21-nt small RNA was produced in *flam* $\Delta$ *ZAM* lines with functional *ZAM* copies and was only slightly increased in *vret*-sKD condition, compared to a control line where no *ZAM* expression was observed (*white*-sKD) (**Fig. 7e**). Globally, the production of 21-nt small RNA was the same for all TEs upon *ZAM* reactivation (**Supplementary Fig. 7e**). Overall, our results suggest that no, or very few, siRNAs that could protect ovaries from *ZAM* activity are produced upon activation of this endogenous retrovirus in follicle cells.

Therefore, we propose that no innate small RNA response is triggered in the lines where *ZAM* is newly reactivated, suggesting that *ZAM* will not be brought under control until specific adaptive immunity and immune memory are established.

## Discussion

The discovery that piRNAs produced in the germ cells, besides restricting TE expression in the germline itself, counteract the invasion of endogenous retroviruses from adjacent cells reveals a novel role of piRNAs produced locally, in the germ cells, as an effective defense mechanism at the tissue scale. When piRNAs against *ZAM* are produced in the germline, the transcripts of this endogenous retrovirus are limited to a patch of somatic follicle cells and the flies are perfectly fertile. However, in absence of *ZAM*-derived piRNAs in the germline, transcripts from functional *ZAM* copies massively invade the oocyte and are even transferred to the embryos. At the molecular level, our study revealed that piRNAs produced in the nurse cells and dumped into the oocyte can target RNAs produced by the follicular cells that are delivered into the oocyte. It is tempting to speculate that the *ZAM* RNA arriving from the somatic cells is targeted by complementary antisense piRNAs and degraded through the ping-pong cycle. This



would result in the production of *ZAM* sense piRNA, thus participating in the amplification of the pool of piRNAs against *ZAM* and strengthen the defense mechanism against the invasion.

Transposition in the germ cell genome is crucial for TE propagation in a population because it allows the vertical transmission of new insertions. At the time of invasion (e.g. in the *flam* $\Delta$ *ZAM* line), no *ZAM*-derived piRNAs are produced in either somatic or germ cells. This condition could therefore be compared to what happens when a TE first enters a new species by horizontal transfer. According to several studies, an initial transposition burst occurs that leads to TE accumulation in the genome before the induction of an adaptive response by the host to control transposition<sup>35-38</sup>. We found at least 17 new *ZAM* insertions in the RevI-H2i2 genome testifying that *ZAM* actively transposed at a given time. Surprisingly, we identified 3 *ZAM* insertions within the same germline piRNA cluster. It has been proposed that two different mechanisms can explain the *de novo* piRNA production required to specifically silence a novel invading TE. Indeed, besides being produced from a new insertion of the TE into a pre-existing piRNA cluster, piRNAs could also be newly produced by euchromatic TE insertions converted into piRNA-producing loci<sup>14,39-41</sup>. Interestingly, we observed that *ZAM* insertions located on chromosomes II and III of the RevI-H2i2 line are not sufficient to produce regulatory *ZAM*-derived piRNA. However, *ZAM* is supposed to be transcriptionally inactive in the germline as *ZAM* expression is regulated by the somatic transcription factor Pointed2<sup>28</sup>. Thus, although the contribution of individual copies needs to be investigated, our data strongly argue that germline *ZAM* piRNAs emanate at least in part from piRNA clusters. In line with our findings, computer simulations of the dynamics of TE invasion have suggested that, concerning TE silencing in the germline, several insertions in piRNA clusters are likely to be required to stop the invasion<sup>42</sup>. However, concerning TE silencing in the soma, we demonstrate here, as has previously been suggested<sup>24</sup>, that a single insertion in a somatic piRNA cluster such as *flamenco* is sufficient for TE silencing in the somatic follicle cells.

In this study, we now show indeed that in the somatic follicle cells, the *flamenco* piRNA cluster acts alone as the principal regulator of transposon activity. A precise deletion of the internal part of the endogenous retrovirus *ZAM*, inserted within *flamenco*, leads to a complete derepression of functional *ZAM* copies. Surprisingly, the reactivation of this endogenous retrovirus in only a patch of somatic cells leads to subfertility, while the germline can deal without its three most highly expressed germline piRNA loci. Indeed, the deletion of the 42AB, 38C, and 20A affected neither fertility nor TE expression<sup>14</sup>. Here we show that the somatic cells can't do without piRNA clusters. Redundancy in piRNA production between piRNA clusters as in the germline, with probably multiple piRNA clusters including stand-alone transposon insertions converted into piRNA-producing loci involved in the silencing of a given TE, is not observed in the somatic cells. Moreover, many TE copies have acquired mutations and are nonfunctional either for transcription or transposition. In the *flam* $\Delta$ *ZAM* line, although no *ZAM* piRNAs were produced, no *ZAM* expression was observed. However, when we brought functional copies of *ZAM*, *ZAM* was expressed demonstrating that depending of the genetic background functional copies can be absent. In general, the contents of TEs vary considerably between populations<sup>40,43</sup>.

Interestingly we have identified *ZAM* copies having a deletion affecting the *pol* gene. There are almost the same number of this *ZAM*-variant that have transposed in the RevI-H2 line (5 copies) as there are classical *ZAM* elements that have transposed (8 copies) suggesting that this *ZAM*-variant is very efficient for transposition. Analysis made on different organisms have shown that TEs are rapidly diversifying, like the mariner DNA transposon, which is present in 68 different versions in grass genomes <sup>44</sup>. Such rapid diversification certainly promotes speciation of new families of active TEs.

One interesting question is how *ZAM* RNA transits into the oocyte and at which stage during *Drosophila* oogenesis *ZAM* transposes in germ cells. We noticed that *ZAM* RNA, although transcribed in the somatic follicle cells, is deposited and accumulated at the posterior pole of early embryos, where future germ cells will cellularize. To amplify, *ZAM* might have evolved to preferentially mobilize in the dividing primordial germ cells of the offspring instead of the developing oocyte, which is not in a phase favorable to transposition since it is in a prolonged arrest of prophase I associated with highly compacted chromatin <sup>45</sup>.

Previous experiments have demonstrated that *ZAM* RNAs accumulate in the follicle cells when vitellogenesis is defective, suggesting that *ZAM* transmission to the oocyte requires functional vitellogenin trafficking <sup>46</sup>. Moreover, endogenous retroviruses such as *ZAM* and *gypsy* are able to form pseudo-viral particles in the follicle cells when they are expressed <sup>29,46-48</sup>. *ZAM* Gag and Env proteins are expressed in the RevI-H2i2 follicle cells. The *ZAM env* gene is similar to the retroviral *env* gene, which is responsible for the infectious properties <sup>26</sup>. Indeed, the *ZAM env* is composed of a glycosylated hydrophilic polypeptide (SU), a transmembrane domain (TM) and an Arg-X-Lys/Arg-Arg conserved domain, which is considered to be a Furin consensus proteolytic cleavage site <sup>26,49,50</sup>. A recent study has demonstrated that intracellular retrotransposition of the Gypsy endogenous retrovirus can occur in absence of the Env protein while intercellular transmission requires a functional Env <sup>48</sup>. At this point, it is still unknown if *ZAM* can transpose intracellularly, in the follicle cells, and which exact mechanism(s), depending or not on the Env protein, are used for *ZAM* transmission to the oocyte.

Once a virus enters into the target cells, three models for retrovirus uncoating have been proposed: The capsid (1) is disassembled soon after virus entry, (2) remains intact until it comes into contact with the nuclear pore complex, (3) is partially disassembled during the cytoplasmic traffic to the nucleus (for review <sup>51</sup>). The way of capsid disassembly has not been investigated for endogenous retroviruses. In our study, we observe the loss of *ZAM* RNA in the oocyte when germline *ZAM*-derived piRNAs are produced. Either the capsid of *ZAM* is dissociated soon after its entry in the oocyte and thus piRNAs can easily target *ZAM* RNA, or *ZAM* RNA is still encapsidated in the oocyte but the capsid does not protect the *ZAM* RNA from piRNAs and their associated endonuclease. Previous studies on virus-like particles of the yeast retrotransposon Ty suggest that these particles form an open structure leaving the RNA at the interior of the capsid accessible to RNase <sup>52,53</sup>.

It has been proposed that in different species an “innate” response either at the piRNA or at the siRNA level could be initiated in the gonads upon retrovirus invasion (reviewed in <sup>54, 55-60</sup>). We have shown that no clear innate small RNA response is triggered upon *ZAM* reactivation in *Drosophila* ovaries. Total small-RNA sequencing under condition of *ZAM* reactivation revealed the presence of sense *ZAM*-mapping small RNAs but these lack characteristics of siRNA or piRNAs. We hypothesize that these sense small RNAs are degradation products that are produced when there is an excessive amount of *ZAM* RNAs. Taken together, these results show that no initial small RNA response is mounted upon *ZAM* reactivation, leaving the follicle cells, where *ZAM* is expressed, but also the oocyte unprotected against *ZAM* transposition. Our findings suggest that the piRNA response is a robust response that, in the case of *ZAM*, must be established either in the follicle cells or in the germ cells to efficiently control the invading TE. However, the time required for developing a robust piRNA response to sustainably control TE transposition remains unknown. Lastly, it had been concluded that the piRNA pathway had no antiviral role in *Drosophila* <sup>61</sup>. But what if, during infection, parts of the viral genome are integrated into a germline piRNA cluster?

## REFERENCES

1. Byrd, D. T. & Kimble, J. Scratching the niche that controls *Caenorhabditis elegans* germline stem cells. *Semin. Cell Dev. Biol.* **20**, 1107–1113 (2009).
2. Kidder, G. M. & Vanderhyden, B. C. Bidirectional communication between oocytes and follicle cells: ensuring oocyte developmental competence. *Can. J. Physiol. Pharmacol.* **88**, 399–413 (2010).
3. Saunders, P. T. K. Germ cell-somatic cell interactions during spermatogenesis. *Reprod. (Cambridge, England). Suppl.* **61**, 91–101 (2003).
4. Jenc, J. C. Somatic gonadal cells: the supporting cast for the germline. *Genesis* **49**, 753–775 (2011).
5. Hall, D. H. *et al.* Ultrastructural features of the adult hermaphrodite gonad of *Caenorhabditis elegans*: relations between the germ line and soma. *Dev. Biol.* **212**, 101–123 (1999).
6. McClintock, B. The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. U. S. A.* **36**, 344–355 (1950).
7. Aravin, A. *et al.* A novel class of small RNAs bind to MILI protein in mouse testes. **442**, 203–207 (2006).
8. Grivna, S. T., Beyret, E., Wang, Z. & Lin, H. A novel class of small RNAs in mouse spermatogenic cells. *Genes Dev.* **20**, 1709–1714 (2006).
9. Lau, N. C. *et al.* Characterization of the piRNA complex from rat testes. *Science (80-. )*. **313**, 363–367 (2006).
10. Brennecke, J. *et al.* Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in *Drosophila*. *Cell* **128**, 1089–1103 (2007).
11. Aravin, A. a, Hannon, G. J. & Brennecke, J. The Piwi-piRNA Pathway Provides an Adaptive Defense in the Transposon Arms Race. *Science (80-. )*. **318**, 761–764 (2007).
12. George, P. *et al.* Increased production of piRNAs from euchromatic clusters and genes in *Anopheles gambiae* compared with *Drosophila melanogaster*. *Epigenetics and Chromatin* **8**, (2015).
13. Malone, C. D. *et al.* Specialized piRNA Pathways Act in Germline and Somatic Tissues of the *Drosophila* Ovary. *Cell* **137**, 522–535 (2009).
14. Gebert, D. *et al.* Large *Drosophila* germline piRNA clusters are evolutionarily labile and dispensable for transposon regulation. *Mol. Cell* 1–14 (2021) doi:10.1016/j.molcel.2021.07.011.
15. Handler, D. *et al.* The genetic makeup of the *drosophila* piRNA pathway. *Mol. Cell* **50**, 762–777 (2013).
16. Olivieri, D., Sykora, M. M., Sachidanandam, R., Mechtler, K. & Brennecke, J. An in vivo RNAi assay

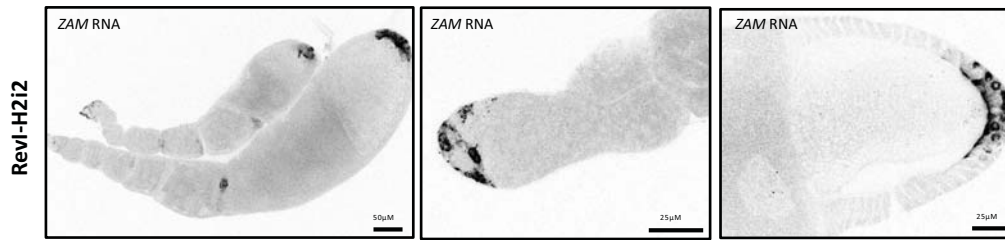
- identifies major genetic and cellular requirements for primary piRNA biogenesis in *Drosophila*. *EMBO J.* **29**, 3301–3317 (2010).
17. Desset, S., Meignin, C., Dastugue, B. & Vaury, C. COM, a heterochromatic locus governing the control of independent endogenous retroviruses from *Drosophila melanogaster*. *Genetics* **164**, 501–509 (2003).
  18. Pélisson, A. *et al.* 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**, 4401–4411 (1994).
  19. Bucheton, A. The relationship between the flamenco gene and gypsy in *Drosophila* : how to tame a retrovirus. **11**, 349–353 (1995).
  20. Robert, V., Prud'homme, N., Kim, A., Bucheton, A. & Pélisson, A. Characterization of the flamenco region of the *Drosophila melanogaster* genome. *Genetics* **158**, 701–713 (2001).
  21. Mével-Ninio, M., Pelisson, A., Kinder, J., Campos, A. R. & Bucheton, A. The flamenco Locus Controls the gypsy and ZAM Retroviruses and Is Required for *Drosophila* Oogenesis. **1624**, 1615–1624 (2007).
  22. Lécher, P., Bucheton, A. & Pélisson, A. Expression of the *Drosophila* retrovirus gypsy as ultrastructurally detectable particles in the ovaries of flies carrying a permissive flamenco allele. *J. Gen. Virol.* **78**, 2379–2388 (1997).
  23. Duc, C. *et al.* Trapping a somatic endogenous retrovirus into a germline piRNA cluster immunizes the germline against further invasion. *Genome Biol.* **20**, 1–14 (2019).
  24. Zanni, V. *et al.* Distribution, evolution, and diversity of retrotransposons at the flamenco locus reflect the regulatory properties of piRNA clusters. *Proc. Natl. Acad. Sci.* **110**, 19842–19847 (2013).
  25. Grentzinger, T. *et al.* A universal method for the rapid isolation of all known classes of functional silencing small RNAs. *Nucleic Acids Res.* **48**, (2020).
  26. Leblanc, P., Desset, S., Dastugue, B. & Vaury, C. Invertebrate retroviruses: ZAM a new candidate in *D. melanogaster*. *EMBO J.* **16**, 7521–7531 (1997).
  27. Desset, S. *et al.* Mobilization of two retroelements, ZAM and Idefix, in a novel unstable line of *Drosophila melanogaster*. *Mol. Biol. Evol.* **16**, 54–66 (1999).
  28. Meignin, C., Dastugue, B. & Vaury, C. Intercellular communication between germ line and somatic line is utilized to control the transcription of ZAM, an endogenous retrovirus from *Drosophila melanogaster*. *Nucleic Acids Res.* **32**, 3799–3806 (2004).
  29. Leblanc, P. *et al.* Life Cycle of an Endogenous Retrovirus, ZAM, in *Drosophila melanogaster*. *J. Virol.* **74**, 10658–10669 (2000).
  30. Atallah, J. & Lott, S. E. Evolution of maternal and zygotic mRNA complements in the early *Drosophila* embryo. *PLoS Genet.* **14**, e1007838 (2018).
  31. Aravin, A. A. & Hannon, G. J. Small RNA silencing pathways in germ and stem cells. *Cold Spring Harb. Symp. Quant. Biol.* **73**, 283–290 (2008).
  32. Khurana, J. S. & Theurkauf, W. piRNAs, transposon silencing, and *Drosophila* germline development. *J. Cell Biol.* **191**, 905–913 (2010).
  33. Saito, K. & Siomi, M. C. Small RNA-mediated quiescence of transposable elements in animals. *Dev. Cell* **19**, 687–697 (2010).
  34. Klattenhoff, C. *et al.* *Drosophila* rasiRNA pathway mutations disrupt embryonic axis specification through activation of an ATR/Chk2 DNA damage response. *Dev. Cell* **12**, 45–55 (2007).
  35. Le Rouzic, A. & Capy, P. The first steps of transposable elements invasion: Parasitic strategy vs. genetic drift. *Genetics* **169**, 1033–1043 (2005).
  36. Kofler, R., Senti, K. A., Nolte, V., Tobler, R. & Schlötterer, C. Molecular dissection of a natural transposable element invasion. *Genome Res.* **28**, 824–835 (2018).
  37. Robillard, É., Le Rouzic, A., Zhang, Z., Capy, P. & Hua-Van, A. Experimental evolution reveals hyperparasitic interactions among transposable elements. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 14763–14768 (2016).
  38. Schaack, S., Gilbert, C. & Feschotte, C. Promiscuous DNA: Horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends Ecol. Evol.* **25**, 537–546 (2010).
  39. Mohn, F., Sienski, G., Handler, D. & Brennecke, J. The Rhino-Deadlock-Cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in *Drosophila*. *Cell* **157**, 1364–1379 (2014).
  40. Luo, S. *et al.* The evolutionary arms race between transposable elements and piRNAs in *Drosophila*

- melanogaster. *BMC Evol. Biol.* **20**, 1–18 (2020).
41. Shpiz, S., Ryazansky, S., Olovnikov, I., Abramov, Y. & Kalmykova, A. Euchromatic Transposon Insertions Trigger Production of Novel Pi- and Endo-siRNAs at the Target Sites in the *Drosophila* Germline. *PLoS Genet.* **10**, (2014).
  42. Kofler, R. Dynamics of transposable element invasions with piRNA clusters. *Mol. Biol. Evol.* **36**, 1457–1472 (2019).
  43. Lerat, E. *et al.* Population-specific dynamics and selection patterns of transposable element insertions in European natural populations. *Mol. Ecol.* **28**, 1506–1522 (2019).
  44. Feschotte, C. & Wessler, S. R. Mariner-like transposases are widespread and diverse in flowering plants. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 280–285 (2002).
  45. Navarro-Costa, P. *et al.* Early programming of the oocyte epigenome temporally controls late prophase I transcription and chromatin remodelling. *Nat. Commun.* **7**, 12331 (2016).
  46. Brassat, E. *et al.* Viral particles of the endogenous retrovirus ZAM from *Drosophila melanogaster* use a pre-existing endosome/exosome pathway for transfer to the oocyte. *Retrovirology* **3**, 1–9 (2006).
  47. Chalvet, F. *et al.* Proviral amplification of the Gypsy endogenous retrovirus of *Drosophila melanogaster* involves env-independent invasion of the female germline. *EMBO J.* **18**, 2659–2669 (1999).
  48. Keegan, R. M., Talbot, L. R., Chang, Y. H., Metzger, M. J. & Dubnau, J. Intercellular viral spread and intracellular transposition of *Drosophila* gypsy. *PLoS Genet.* **17**, 1–22 (2021).
  49. Klenk, H. D. & Garten, W. Host cell proteases controlling virus pathogenicity. *Trends Microbiol.* **2**, 39–43 (1994).
  50. Hosaka, M. *et al.* Arg-X-Lys/Arg-Arg motif as a signal for precursor cleavage catalyzed by furin within the constitutive secretory pathway. *J. Biol. Chem.* **266**, 12127–12130 (1991).
  51. Yamashita, M. & Engelman, A. N. Capsid-Dependent Host Factors in HIV-1 Infection. *Trends Microbiol.* **25**, 741–755 (2017).
  52. Palmer, K. J. *et al.* Cryo-electron microscopy structure of yeast Ty retrotransposon virus-like particles. *J. Virol.* **71**, 6863–6868 (1997).
  53. Burns, N. R. *et al.* Symmetry, flexibility and permeability in the structure of yeast retrotransposon virus-like particles. *EMBO J.* **11**, 1155–1164 (1992).
  54. Kolliopoulou, A. *et al.* PIWI pathway against viruses in insects. *Wiley Interdiscip. Rev. RNA* **10**, e1555 (2019).
  55. Rozhkov, N. V. *et al.* Small RNA-based silencing strategies for transposons in the process of invading *Drosophila* species. *Rna* **16**, 1634–1645 (2010).
  56. Rozhkov, N. V. *et al.* Evolution and dynamics of small RNA response to a retroelement invasion in *Drosophila*. *Mol. Biol. Evol.* **30**, 397–408 (2013).
  57. Yu, T. *et al.* The PiRNA Response to Retroviral Invasion of the Koala Genome. *SSRN Electron. J.* 1–12 (2019) doi:10.2139/ssrn.3391942.
  58. Kawamura, Y. *et al.* *Drosophila* endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* **453**, 793–797 (2008).
  59. Rehwinkel, J. *et al.* Genome-Wide Analysis of mRNAs Regulated by Drosha and Argonaute Proteins in *Drosophila melanogaster*. *Mol. Cell. Biol.* **26**, 2965–2975 (2006).
  60. Katsuma, S. *et al.* Transcriptome profiling reveals infection strategy of an insect maculavirus. *DNA Res. an Int. J. rapid Publ. reports genes genomes* **25**, 277–286 (2018).
  61. Petit, M. *et al.* piRNA pathway is not required for antiviral defense in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci.* **113**, E4218–E4227 (2016).

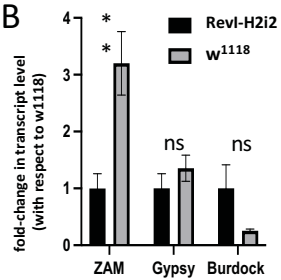




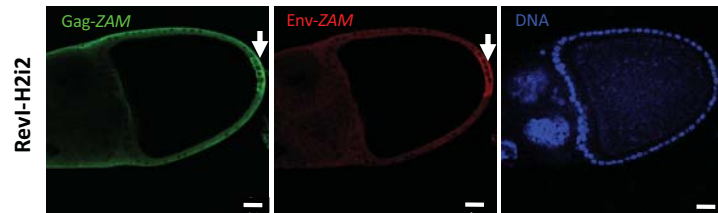
A



B



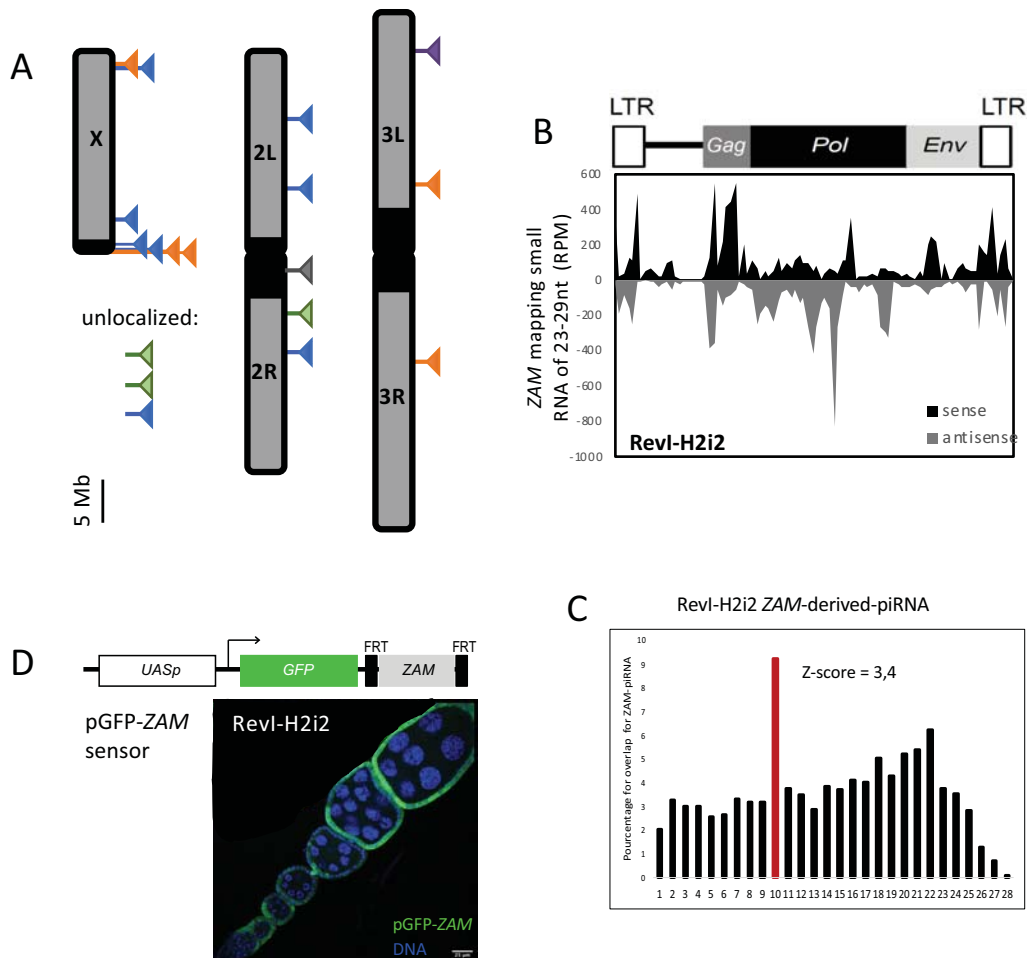
C



**Figure 1 : ZAM is expressed in the follicular cells of the RevIH2i2 isogenic line**

(A) Color-inverted projection of confocal images of ovarioles, germarium and stage 10 follicles depict ZAM RNA-FISH (in black) in ovaries of the RevI-H2i2 isogenic line (Z projection on 20 slides). (B) Fold change in steady state RNA level of TEs in RevI-H2i2 ovaries based on RT-qPCR. Tested TEs are indicated above the graph (the sequences of primers are indicated in Supplementary Table x). RNA levels are normalized to w1118. At least 3 biological replicates and 2 technical replicates were used. Statistical tests were made using prisme Mann Whitney test. \* corresponds to a  $P$  value  $<0.05$ , \*\* to a  $P$  value  $<0.01$  and \*\*\* to a  $P$  value  $<0.001$ . Error bars indicate SD. (C) Confocal sections of stage 10 egg chambers from RevI-H2i2 showing ZAM-encoded Gag (green) and Env (red) proteins. Dapi staining marks DNA in blue.



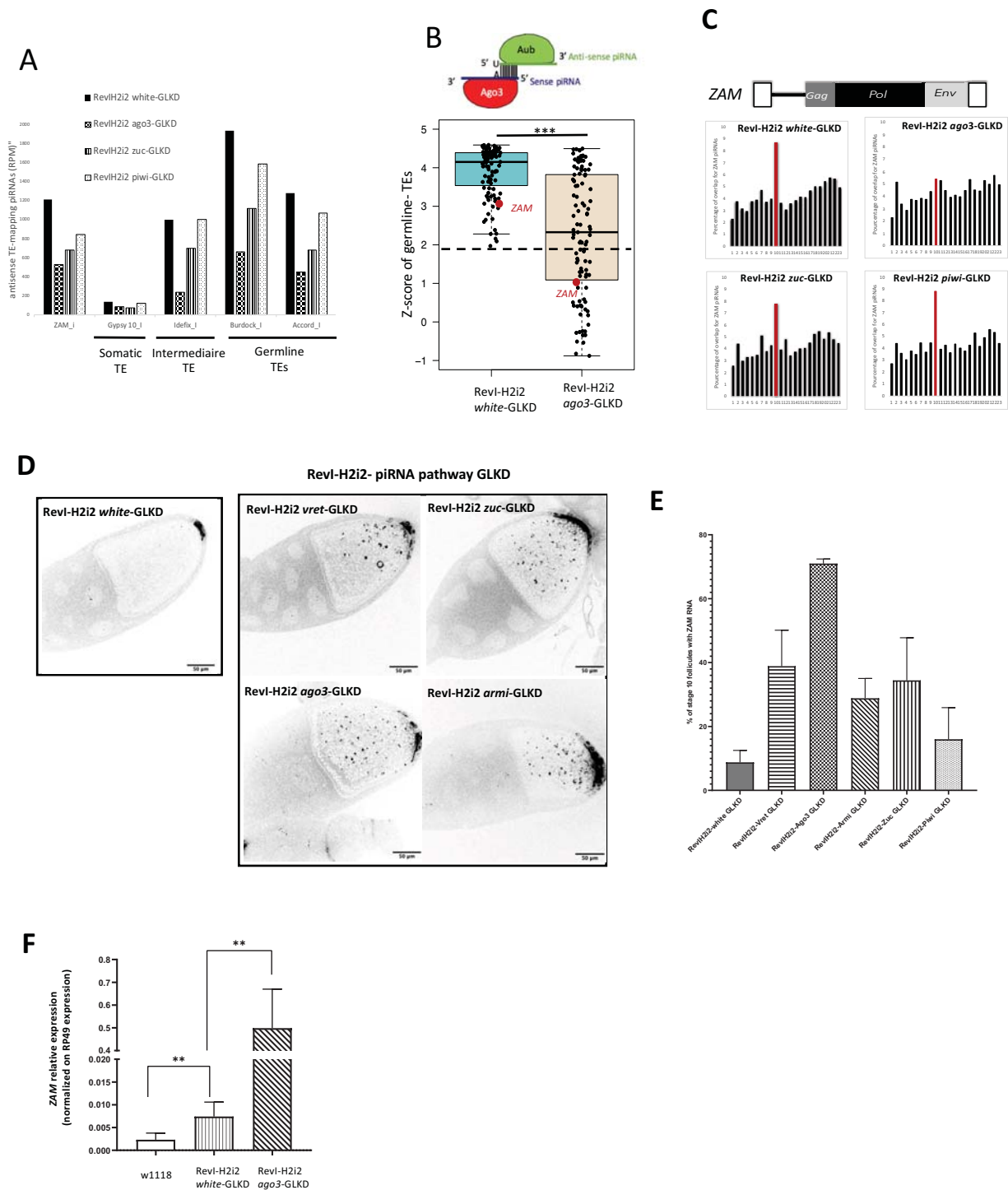


**Figure 2. ZAM elements in the RevI-H2i2 genome.**

(A) Each triangle represents a ZAM insertion. In blue full-length ZAM elements, in orange ZAM-variants, in which positions 5505-6117 of ZAM internal sequences (ZAM\_I in RepBase) are deleted, in violet: ZAM full-length, 5'-UTR deleted, in green ZAM with large internal deletions other than the ZAM-variant, in grey the ZAM copy that is also present in the reference Release 6 genome.

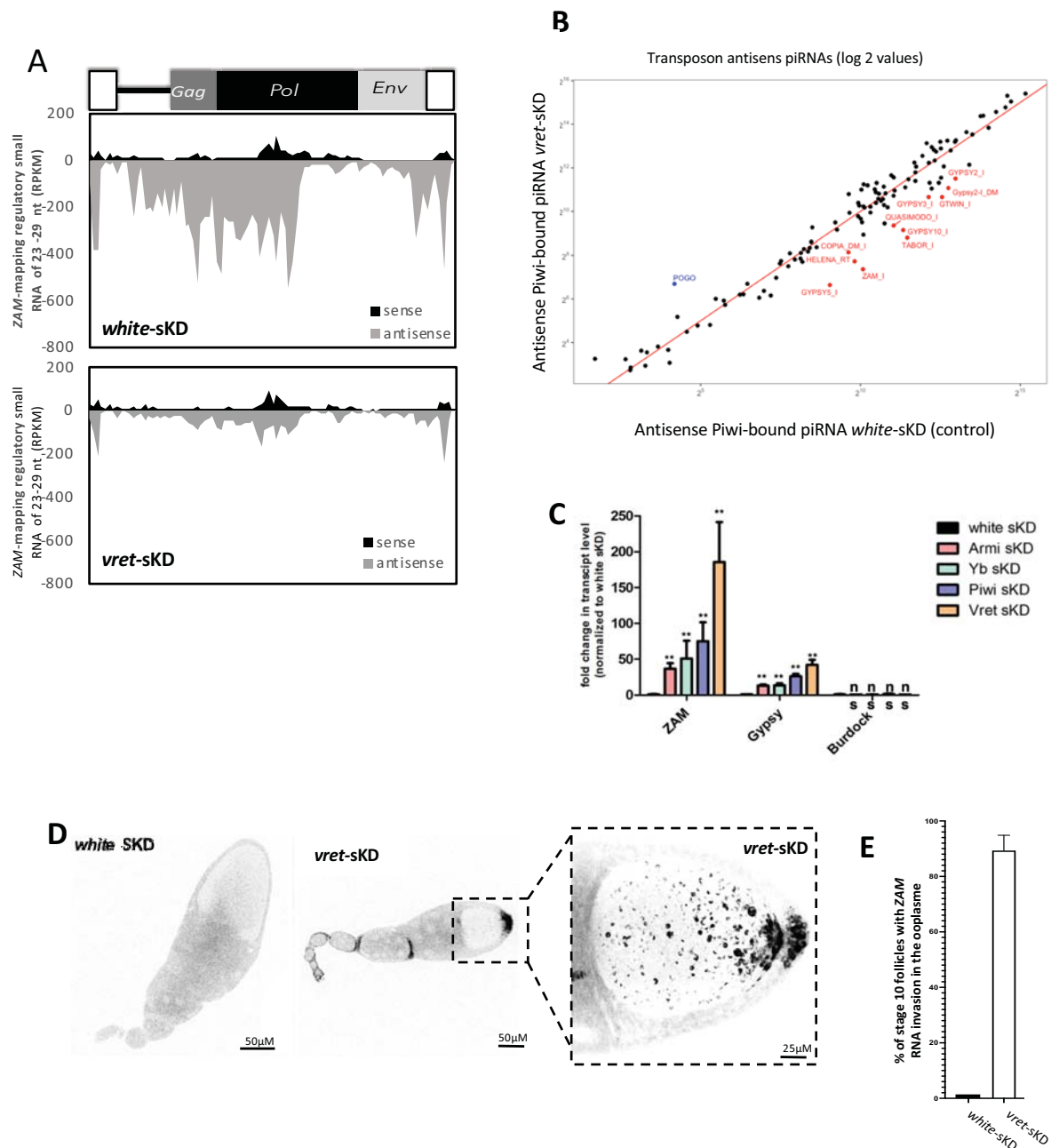
In the RevI-H2i2 genome, 18 recent ZAM insertions were detected. Only one of these ZAM insertions is also present in the reference genome (<http://flybase.org>, genomic position R6 2R:1808663-1817084), and 17 are new insertions. Seven of the latter are located on the X chromosome, four on chromosome 2 and three on chromosome 3. Four of the X chromosome ZAM copies are in X pericentromeric heterochromatin, all others are euchromatic. Three ZAM insertions could not be clearly localized since they are in highly repeated regions. Two of these are in ribosomal DNA (thus on X or Y chromosomes) while the third is in a non-assembled part of the reference genome (NW\_001845412.1). No ZAM insertions were found on chromosome 4 (1.35 Mb, not shown).

One of the full-length ZAM elements is closely related to the ZAM copy inserted in the Snap25 gene in the reference genome: It carries the same 1159 bp deletion within the 5'-UTR (ZAM\_I positions 60-1218) but is inserted at a different location. The Snap25 ZAM copy itself is absent in the RevI-H2i2 genome. From the 3 ZAM copies with large internal deletions, 2 have one same deletion ranging from ZAM\_I position 2332 to 5754, deleting major parts of *gag* and *pol* genes. One has an additional deletion of positions 6116-7265, deleting C-terminal part of *pol* and more than half of the *env* gene. The deletion in the third deleted ZAM copy spans over ZAM\_I positions 2929-5615. (B) Density profile of ZAM-mapping regulatory piRNAs along the ZAM sequence produced in RevI-H2i2 isogenic line ovaries (allowing up to 3 mismatches). Sense and antisense reads are represented in black and grey, respectively. (C) Histogram showing the percentage of 5'-overlap between sense and antisense ZAM-derived piRNAs (23-29 nt) in RevI-H2i2 isogenic line ovaries. The proportion of 10-nt 5' overlap pairs is in red and the Z-score is indicated. (D) Confocal images of ovarioles of after GFP (green) and DNA (blue) staining. Ovarioles were from the progeny of a cross between RevI-H2i2 females and males carrying the pGFP-ZAM sensor transgene driven by actin-Gal4.



**Figure 3 : Germline ZAM piRNAs produced in RevI-H2i2 germ cells counteract ZAM invasion**

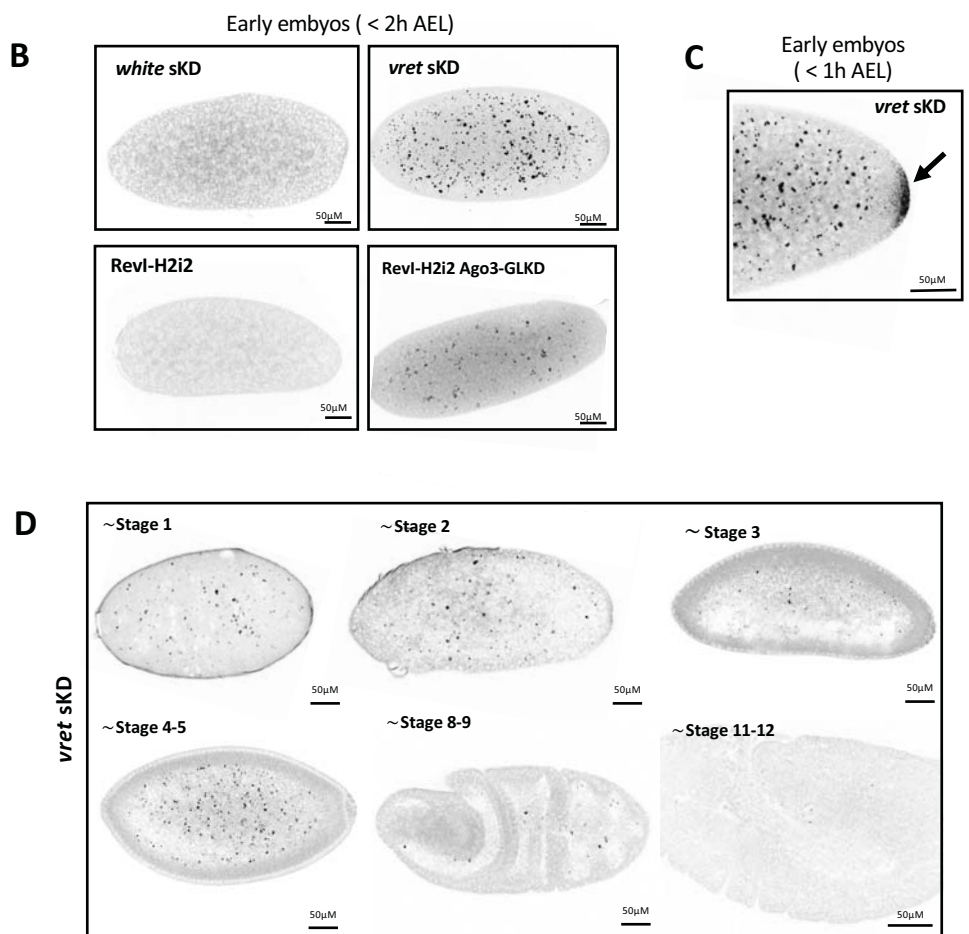
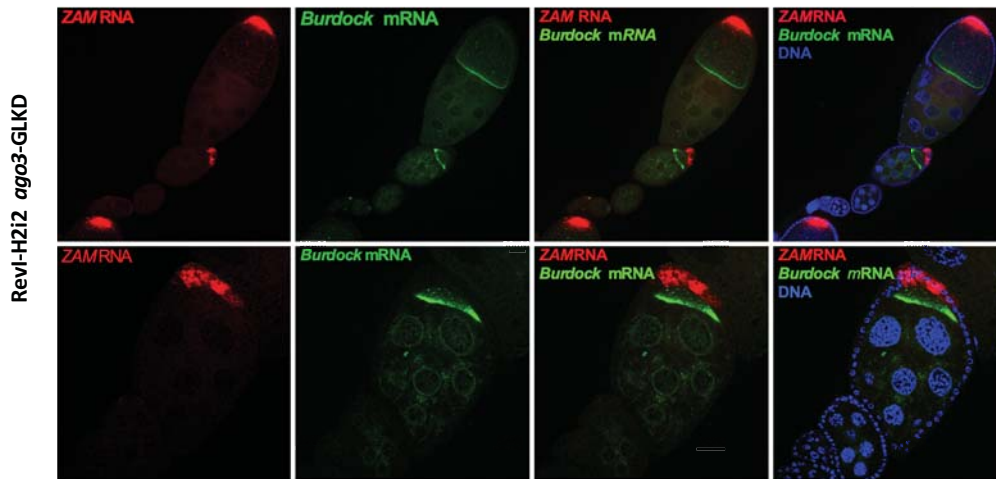
(A) Antisense piRNA mapping TEs sequences with 0-3mismatch in RevI-H2i2 line harboring *white*, *ago3*, *zuc* or *piwi* GLKD. Normalised per millions of piRNA genome. (B) Box plot displaying Z-score distribution of The proportion of 10-nt 5' overlap for TEs piRNAs in RevI-H2i2 *white*-GLKD and RevI-H2i2 *ago3*-GLKD conditions. Each dot represent a TE that show a Z-score >1.96 (dotted line) in a control condition (RevI-H2i2 *white*-GLKD). ZAM has been highlighted in red. Midline indicates the median value, box ranges from the first to the third quartile. Error bars indicate SD. Statistical tests were made using prisme Mann Whitney test. \*\*\* correspond to a  $P$  value <0.001.(C) Histogram showing the percentage of 5'-overlap between sense and antisense ZAM-derived piRNAs (23–29 nt) in RevI-H2i2 *white*, *ago3*, *zuc* or *piwi* GLKD ovaries. The proportion of 10-nt overlapping pairs is in red. (D) Color-inverted confocal images of stage 10 egg chambers depict ZAM RNA-FISH (in black) in ovaries of the indicated genotype. (E) Bar plot showing the pourcentage of stage 10 follicles with ZAM RNA in the ooplasm. The pourcentage has been assessed by ZAM smFISH on RevI-H2i2 ovaries harboring the indicated GLKD. This experiment has been done 3 times and approximately 50 follicles per conditions have been observed for each experiment. Error bars indicate SD. (F) Bar plot showing fold changes in steady-state RNA levels of ZAM in total ovarian RNAs from the indicated genotypes (normalized to rp49). At least 3 biological replicates and 2 technical replicates were used. Statistical tests were made using prisme Mann Whitney test. \*\* corresponds to a  $P$  value <0.01. Error bars indicate SD.



**Figure 4 : ZAM invades the oocyte when no ZAM-derived piRNAs are produced in the germ cells.**

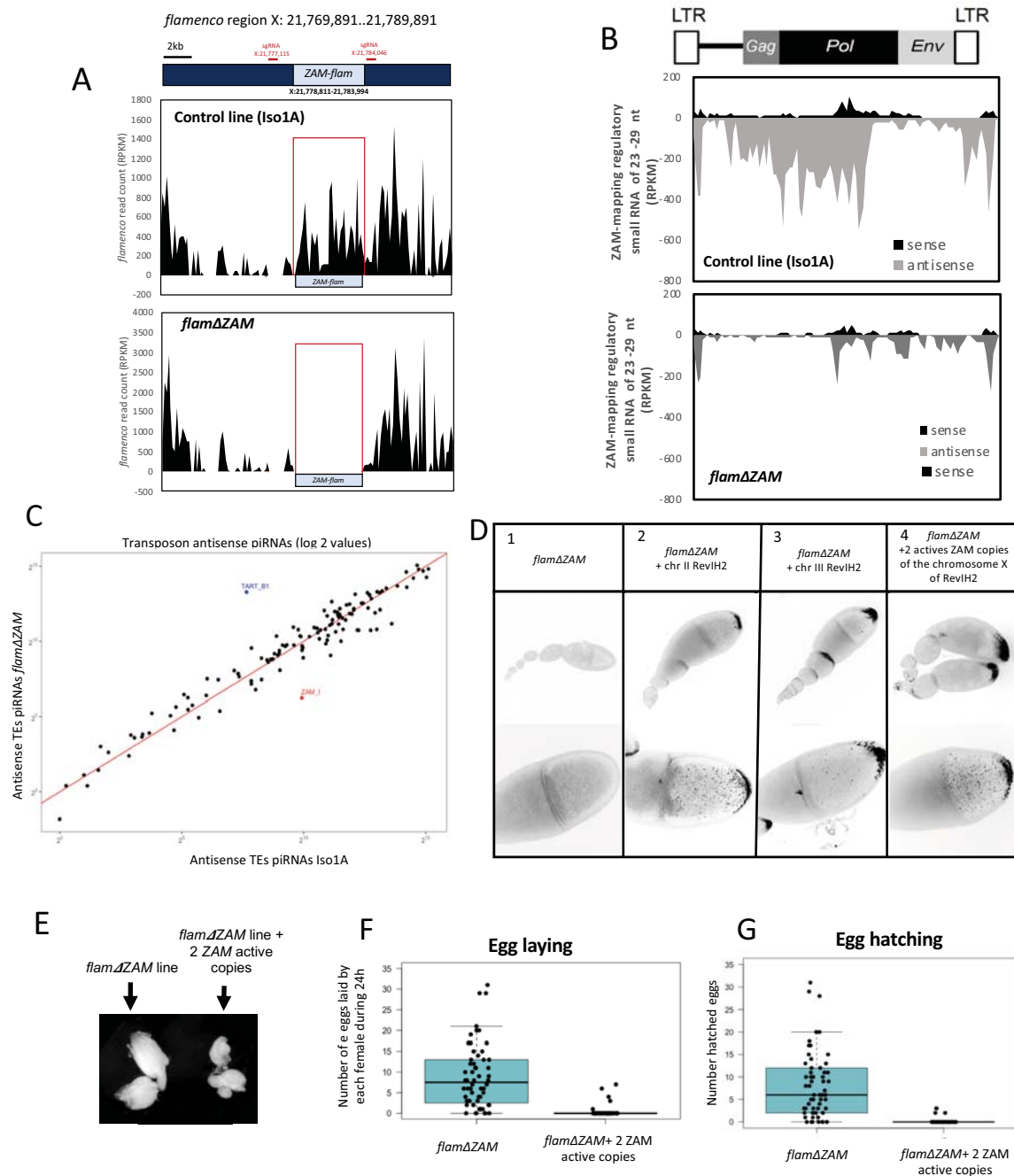
(A) Density profile of ZAM mapping regulatory piRNAs along the ZAM sequence produced in *white* and *vret-sKD* ovaries (allowing up to 3 mismatches). Sense and antisense reads are represented in black and grey, respectively. (B) Scatter plot showing log<sub>2</sub> of antisens Piwi-bound piRNA (obtained by Piwi immunoprecipitation) mapping individuals TEs in control ovaries (*white-sKD*) versus *vreteno-sKD* ovaries. 0-3 mismatches are allowed for the mapping of antisens piRNA on TEs and normalised per millions of genome piRNAs. \*TEs in blue >3 ratio *vret-sKD/white-sKD*, TEs in red <0,3 ratio *vret-sKD/white-sKD*. (C) Fold change in steady state RNA level of TEs in the indicated sKD based on RT-qPCR. Tested TEs are indicated above the graph (the sequences of primers are indicated in Supplementary Table x). RNA levels are normalized to *w<sup>1118</sup>*. At least 3 biological replicates and 2 technical replicates was used. Statistical tests were made using prisme Mann Whitney test. \*\* corresponds to a *P* value <0.01. Error bars indicate SD. (D) Color-inverted confocal images of ovarioles or stage 10 egg chambers depict ZAM RNA-FISH (in black) in ovaries of *white-sKD* or *vret-sKD* flies. ZAM RNA is observed in follicular cells and invasion in the ooplasm in a *vret-sKD* condition. (E) Bar plot showing the pourcentage of stage 10 follicles with ZAM RNA invasion in the ooplasm. The pourcentage has been assessed by ZAM smFISH on *white-sKD* (no ZAM expression detected) or on *vret-sKD* ovaries. This experiment has been done 2 times and approximately 50 follicles per conditions has been observed for each experiment. Error bars indicate SD.

A

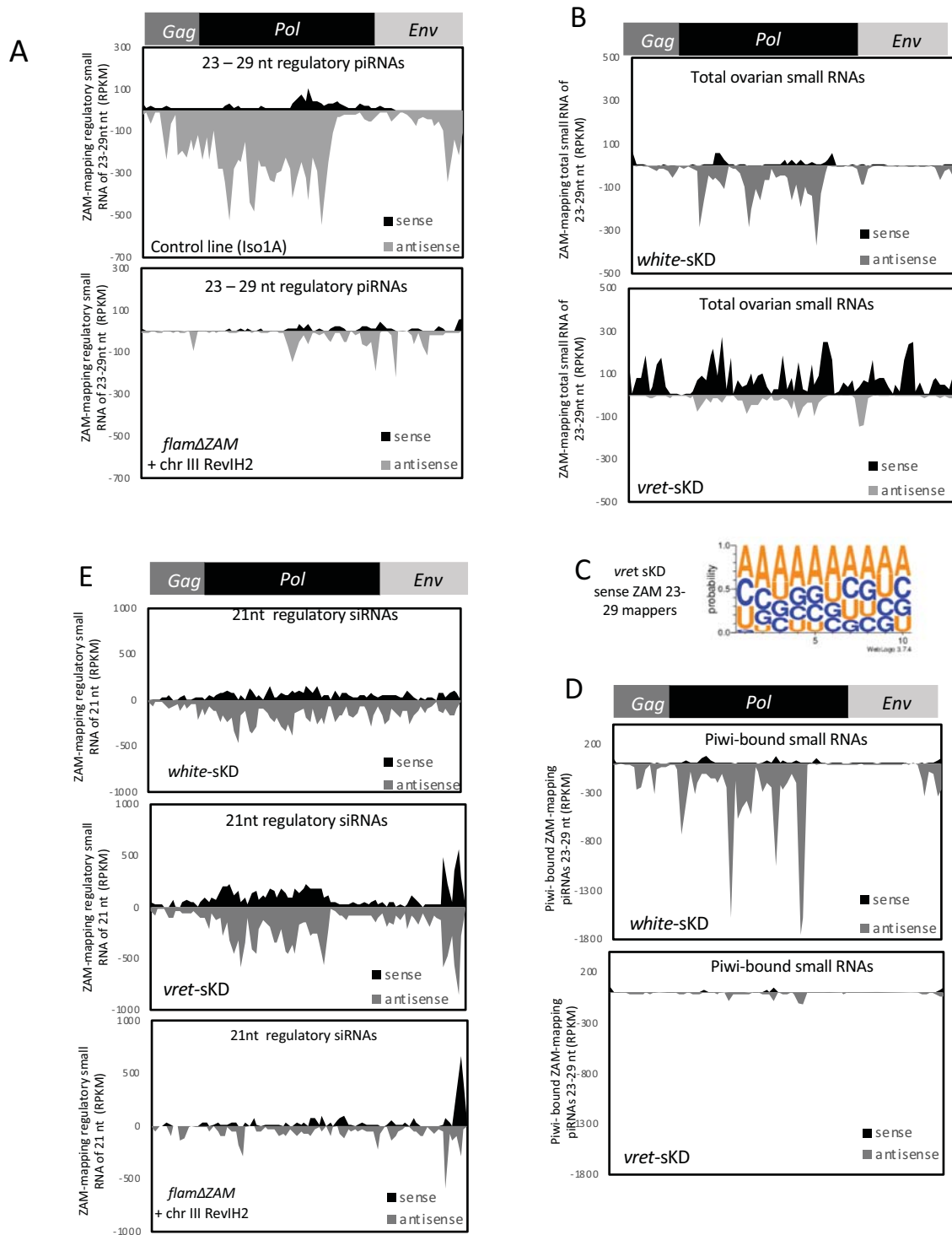


**Figure 5: ZAM RNA transcribed in the follicle cells transits to the oocyte and is deposited in early embryos .**

(A) Confocal sections of ovarioles (upper panels) or an isolate stage 8 egg chambers (lower panel) from a Revl-H2i2 *ago3-GLKD* line. smFISH as been perform to detect *Burdock* RNA (green) and ZAM mRNA (red). Dapi staining marks DNA in blue. (B) Color-inverted confocal sections of early embryos collected before 2 hours after egg laying (AEL). Mother are from the indicated genotype. ZAM RNA (in black) is detected by smFISH. (C) Zoom on the posterior pole of early embryo collected less that 1h AEL showing ZAM RNA accumulated at the posterior pole of the *vret-sKD* embryo. (D) Color-inverted confocal sections of embryos at indicated stage of development laid by *vret-sKD* females showing ZAM RNA. ZAM RNA is observed in high quantity until stage 5 and few ZAM RNA persist until stage 8-9 of embryogenesis.



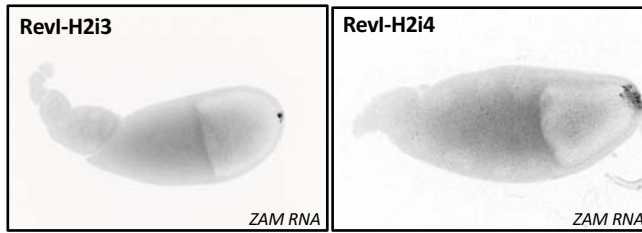
**Figure 6 : Deletion of ZAM from the *flamenco* piRNA cluster leads to ZAM reactivation and induces subfertility.** (A) 10kb of the *flamenco* piRNA cluster where the ZAM insertion is localized has been represented (Position in the reference genome release 6.05 : X: 21,769,891..21,789,891). The position of ZAM insertion in *flam* (ZAM-*flam*) is indicated. Position of sgRNAs used for the deletion of the ZAM-*flam* by CRISPR/Cas9 are shown in red. The density plot represents the mapping of unique piRNAs (with 0 mismatches) from a control line (Iso1A) and from the *flam*ΔZAM line on the 10kb region of *flamenco* represented above. A complete loss of unique piRNAs mapping ZAM-*flam* is observed in the *flam*ΔZAM line. (B) Density profile of ZAM-mapping regulatory piRNAs along the ZAM sequence produced in control and *flam*ΔZAM ovaries (allowing up to 3 mismatches). Sense and antisense reads are represented in black and grey, respectively. (C) Scatter plot showing log2 of antisense regulatory piRNAs mapping individuals TEs in control ovaries (Iso1A) versus *flam*ΔZAM ovaries. 0-3mismatches are allowed for the mapping of antisense piRNAs on TEs and normalised per millions of genome piRNA. \*TE in blue >3 ratio *flam*ΔZAM/Iso1A, TEs in red <0,3 ratio *flam*ΔZAM/Iso1A. (D) Color-inverted confocal images of ovarioles (upper panels) or stage 10 egg chambers (lower panels) from indicated genotype showing ZAM mRNA (black). (E) Ovaries of a *flam*ΔZAM line and of the same *flam*ΔZAM line where 2 active copies of ZAM have been introduced in the genome by genetic cross with the RevI-H2i2 line (the 2 copies located on the X chromosome). (F) Number of eggs laid per fly per day in *flam*ΔZAM flies and *flam*ΔZAM with 2 active copies of ZAM flies. (G) Number of hatched eggs in each individual *flam*ΔZAM or *flam*ΔZAM with 2 active copies of ZAM.



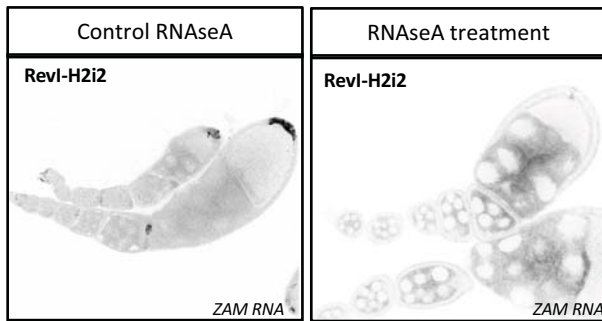
**Figure 7 : No piRNA and siRNA production after ZAM retrotransposon reactivation**

(A) Density profile of ZAM-mapping regulatory piRNAs (23-29nt small-RNAs complexed with Argonaute proteins) along the ZAM internal sequence produced in control line (Iso1A), and *flamΔZAM*+ chr III RevIH2 ovaries (allowing up to 3 mismatches). Sense and antisense reads are represented in black and grey, respectively. (B) Density profile of ZAM-mapping total small RNA of 23 to 29nt along the ZAM internal sequence produced in *white-sKD* and *vret-sKD* ovaries (allowing up to 3 mismatches). Sense and antisense reads are represented in black and grey, respectively. (C) Logo of nucleotide bias for the first ten positions in ZAM mapping sense small RNA of 23 to 29nt (from total ovarian small RNA) produced in *vret-sKD* ovaries. The nucleotide height represents its relative frequency at that position. (D) Density profile of ZAM-mapping Piwi-bound piRNAs along the ZAM internal sequence produced in *white-sKD* and *vret-sKD* ovaries (allowing up to 3 mismatches). Sense and antisense reads are represented in black and grey, respectively. (E) Density profile of ZAM-mapping regulatory siRNAs (21nt small-RNAs complexed with Argonaute proteins) along the ZAM internal sequence produced in *white-sKD*, *vret-sKD* and *flamΔZAM*+ chr III RevIH2 ovaries (allowing 1 mismatch). Sense and antisense reads are represented in black and grey, respectively.

A



B

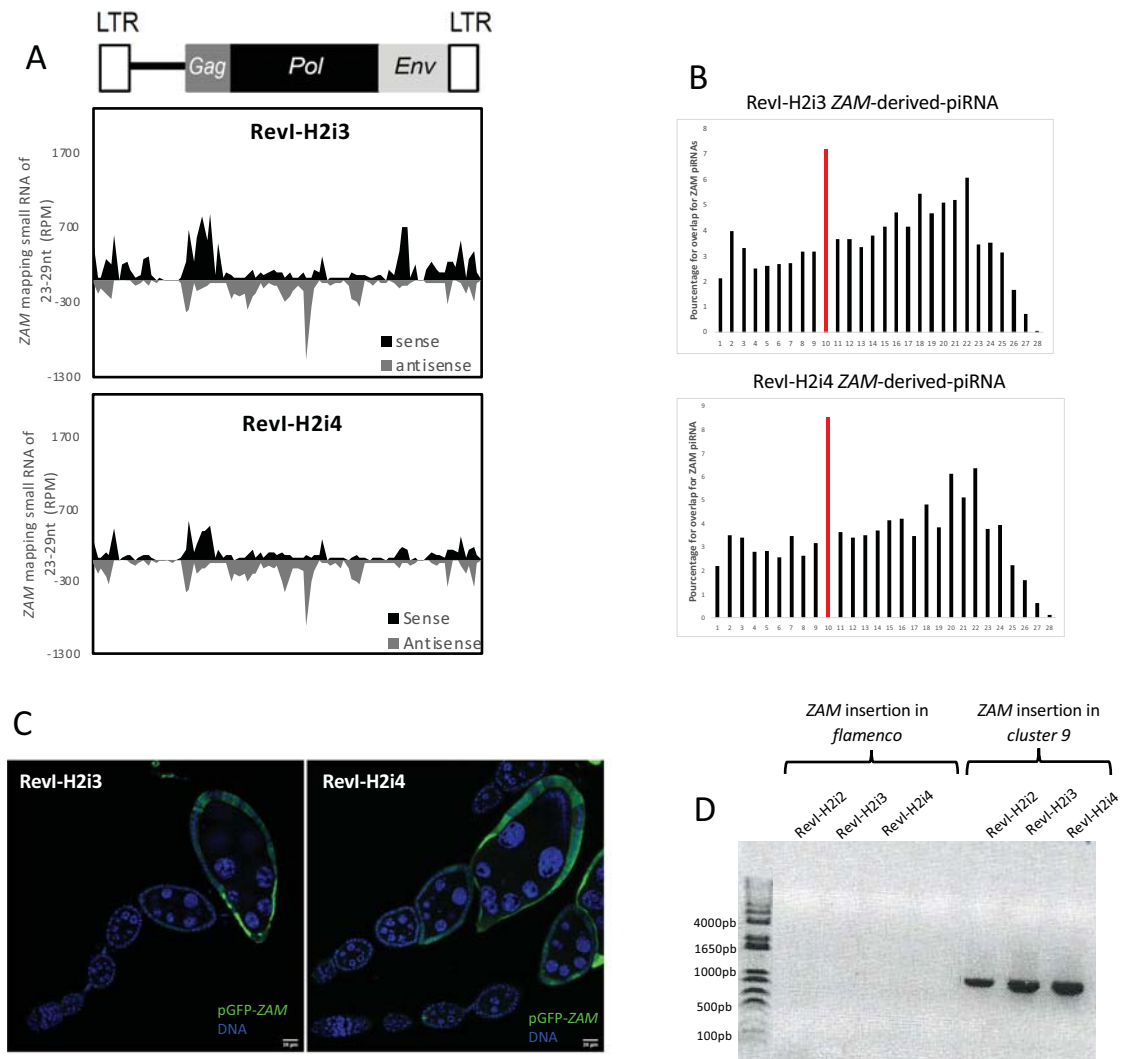


C



**Fig S1 (related to Figure 1)**

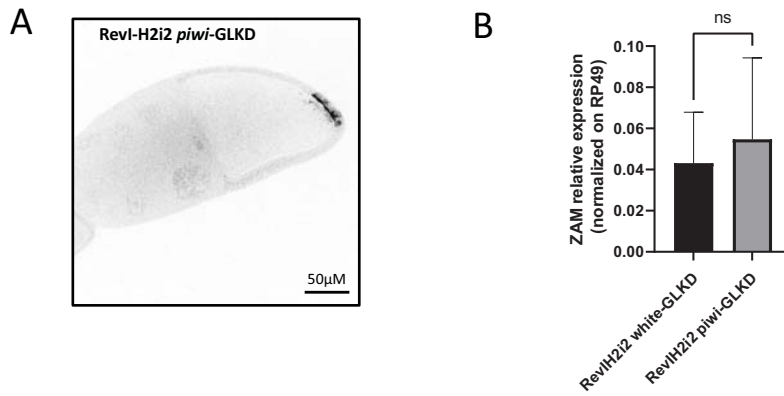
(A) Color-inverted confocal images of ovarioles depict ZAM RNA (in black) in ovaries of RevI-H2 isogenic lines (B) Color-inverted confocal images of ovarioles of RevI-H2i2 ovaries showing ZAM RNA (in black) without (left panel) and with (right panel) RNase A treatment. (C) Color-inverted confocal images of ovarioles showing that ZAM RNA is totally absent in Iso1A, w1118 and wIR6 ovaries



**Fig S2 (related to Figure 2)**

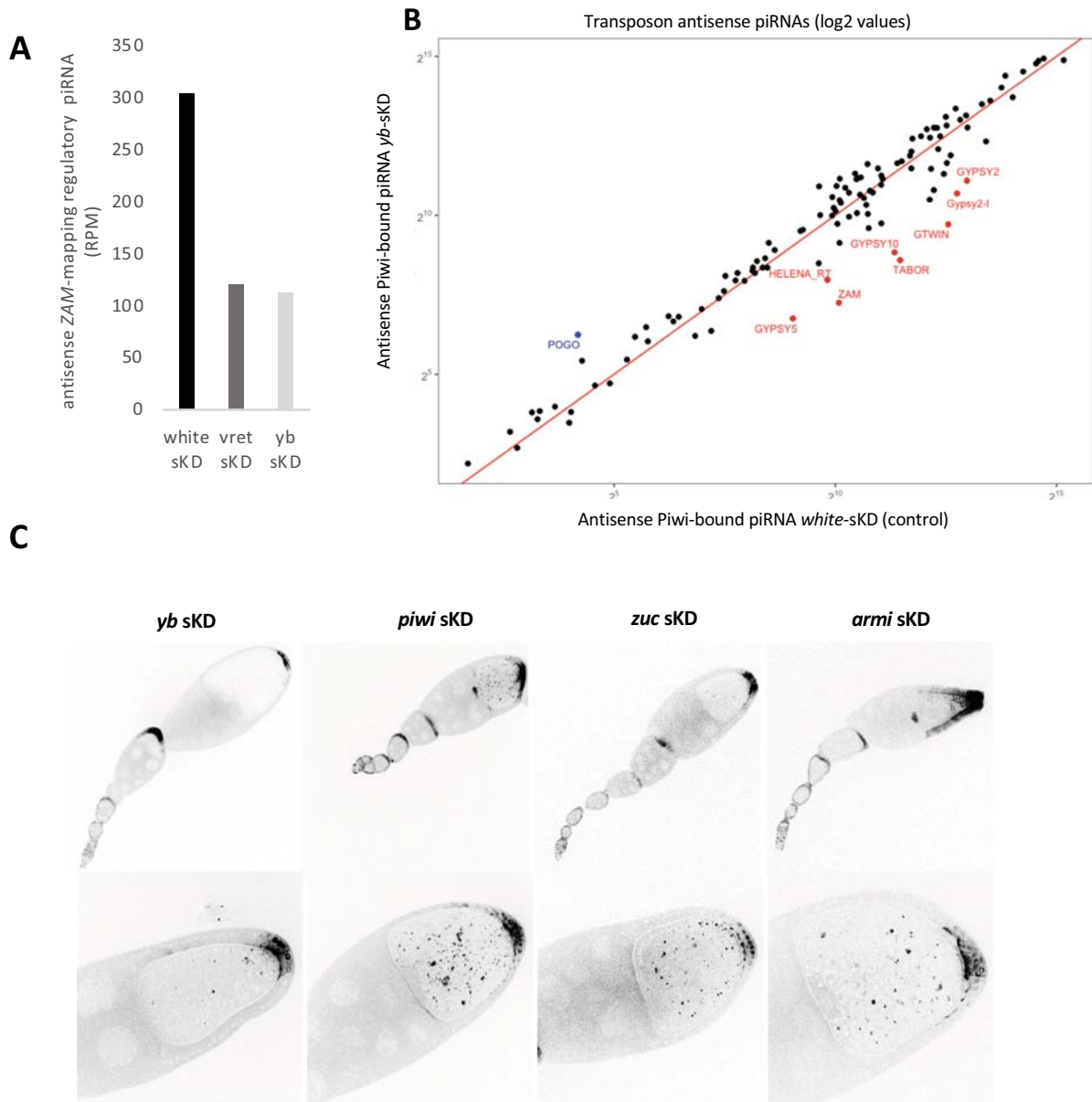
(A) Genomic DNA from RevI-H2 isogenic lines (RevI-H2i2, RevI-H2i3 and RevI-H2i4) was amplified by PCR using primers that can detect either the presence of *ZAM* insertion in *flamenco* or *ZAM* insertion in the *cluster 9* that is specific of the RevI-H2 line. For both conditions, one primer target a region upstream of the *ZAM* insertion and one primer is located in the *ZAM* sequence. Amplification can only occur if *ZAM* is inserted at the targetted locus (sequence of the primer are available table x). For all RevI-H2 isogenic lines tested no amplification are observed for *ZAM* insertion in *flamenco* while a PCR product of ~ 1000pb is observed for the *ZAM* insertion in the *cluster 9*. (B) Density profile of *ZAM*-mapping regulatory piRNAs along the *ZAM* sequence produced in RevI-H2 isogenic lines ovaries (allowing up to 3 mismatches). Sense and antisense reads are represented in black and grey, respectively. (C) Histogram showing the percentage of 5'-overlap between sense and antisense *ZAM*-derived piRNAs (23–29 nt) in RevI-H2 isogenic lines ovaries. The proportion of 10-nt 5' overlap pairs is in red. (D) Confocal images of ovarioles with the sensor transgene (green) and DNA (blue) staining. Ovarioles were from the progeny of a cross between RevI-H2i3 or RevI-H2i4 females and males carrying the pGFP-*ZAM* sensor transgene driven by actin-Gal4.





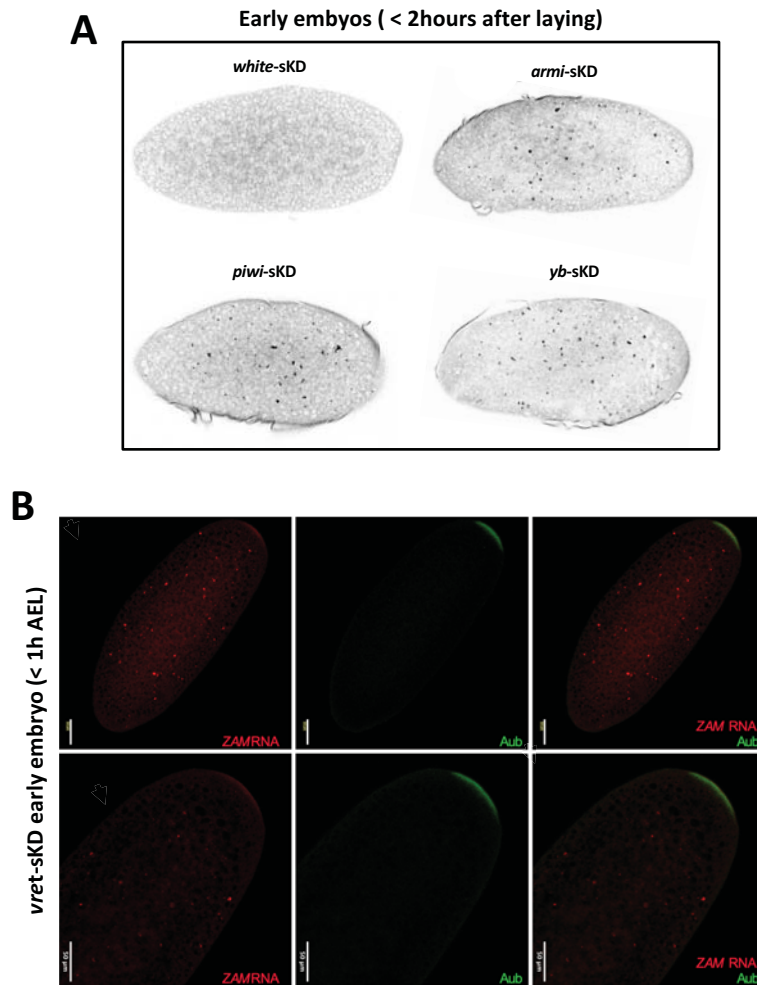
**Fig S3 (related to Figure 3)**

(A) Color-inverted confocal images of stage 10 egg chambers from RevI-H2i2 *piwi*-GLKD ovaries showing ZAM mRNA. (B) Bar plot showing fold changes in steady-state RNA levels of ZAM in total ovarian RNA from the indicated genotypes (normalized to rp49). At least 3 biological replicates and 2 technical replicates were used. Statistical tests were made using prisme Mann Whitney test. ns corresponds to a  $P$  value  $>0.05$ . Error bars indicate SD.



**FigS4 (related to Figure 4):**

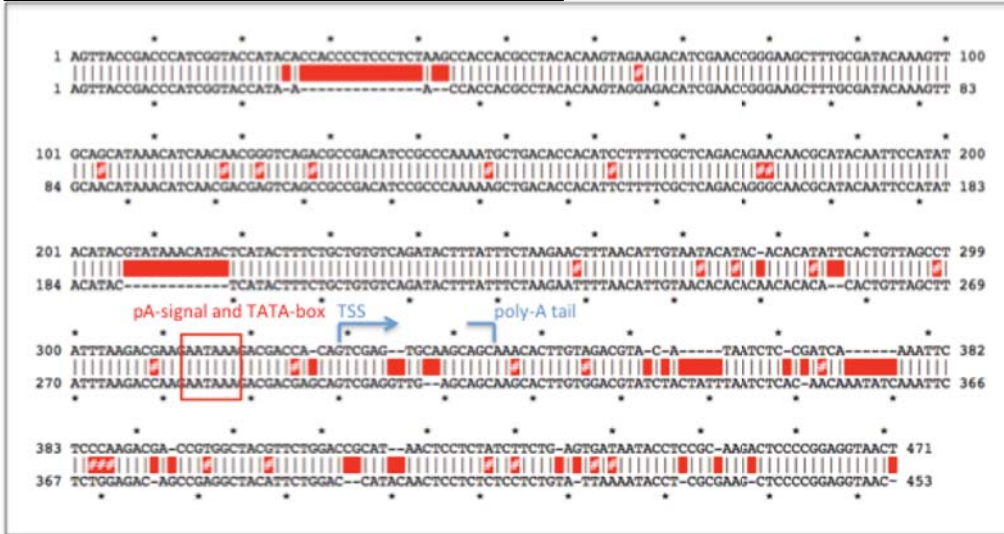
(A) Antisense piRNAs mapping *ZAM* sequences with 0-3 mismatches in *white*, *vret* or *yb*-sKD (normalised per millions of genome piRNAs). (B) Scatter plot showing log<sub>2</sub> of antisense piwi-bound piRNA (obtained by Piwi immunoprecipitation) mapping individuals TEs in control ovaries (*white*-sKD) versus *yb*-sKD ovaries. 0-3 mismatches are allowed for the mapping of antisense piRNAs on TEs and normalised per millions of piRNA genome. \*TEs in blue >3 ratio *yb*-sKD/*white*-sKD, TEs in red < 0,3 ratio *yb*-sKD/*white*-sKD (C) Color-inverted confocal images of ovarioles (upper panels) or stage 10 egg chambers (lower panels) from indicated sKD showing *ZAM* RNA (black). *ZAM* RNA is observed in follicular cells and in the ooplasm in all conditions.



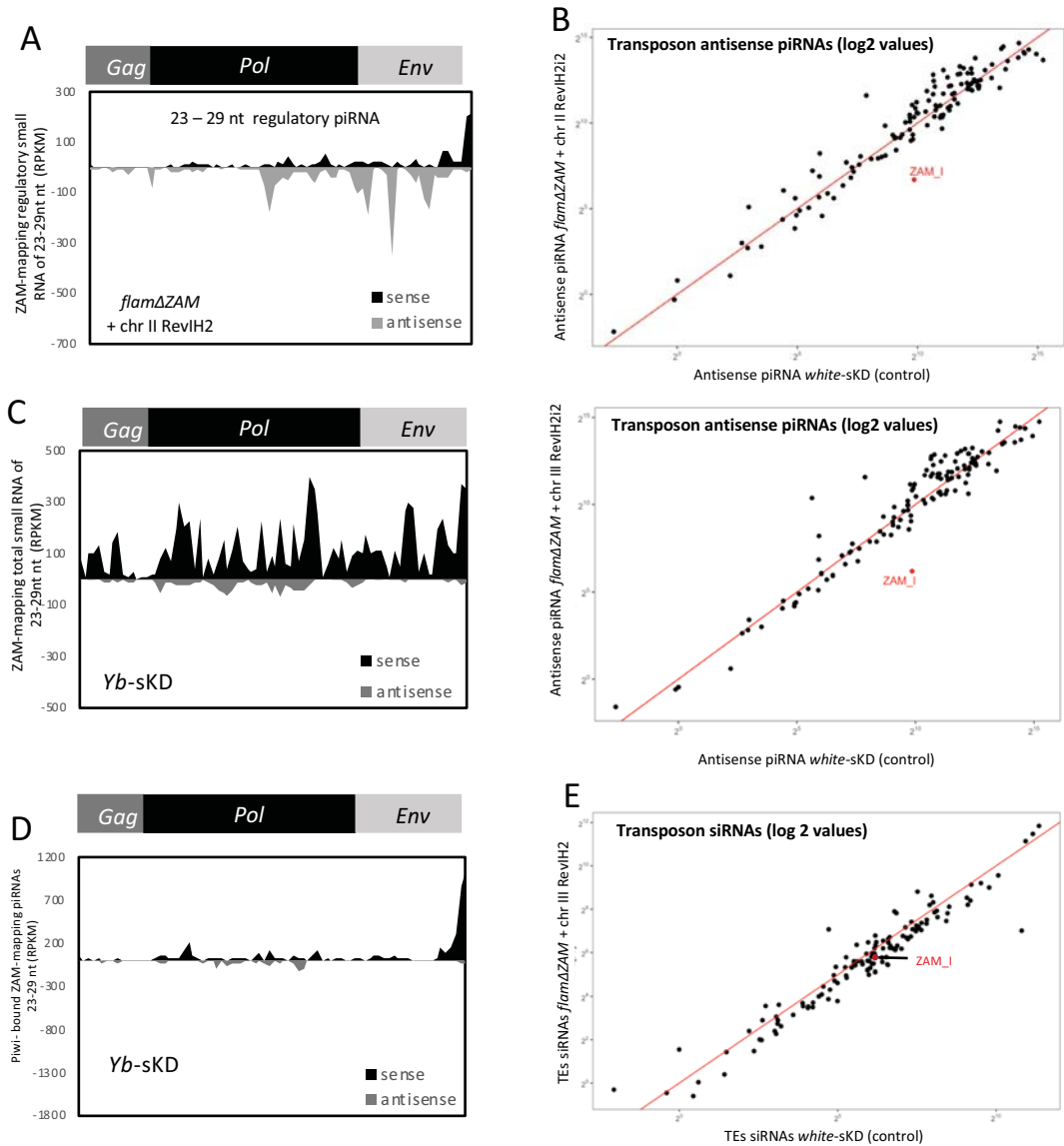
**Fig S5 (related to Figure 5):**

(A) Color-inverted confocal sections of early embryos collected before 2 hours after egg laying (AEL). Mothers harbor sKD of the piRNA pathway as indicated. *ZAM* RNA (in black) is detected by smFISH. (B) Projection of confocal images of embryos at stage 1 of development (< 1h After Egg Laying) laid by *vret-sKD* females showing *ZAM* mRNA detected by smFISH and Aub protein detected by immunostaining. Both accumulated at the posterior pole of the embryo where future germ cells will cellularize (Z-Projection on 25 stacks).

TATA-box and pA-signal: pos. 312 GAATAAAG  
 TSS: pos. 329 GTCGAGTGC  
 poly-A tail at pos. 8307 (345) CAAGCAGC-(A)<sub>n</sub>  
 (Leblanc et al., 1997)



**Fig. S6 (related to Figure 6):** Sequence alignment of the RevI-H2 consensus ZAM LTR and the ZAM solo-LTR, which remains after CRISPR/Cas9 deletion of the internal ZAM sequence in *flamenco* of the *flam*ΔZAM line. Shown are the RevI-H2 consensus ZAM LTR (upper sequence) and the ZAM solo-LTR (lower sequence) with their TATA-box and polyadenylation signal (pA-signal) at pos. 312 of the ZAM consensus (upper sequence), the transcription start site (TSS, at pos. 329), the start of the poly-A tail at pos. 345(CAAGCAGC-(A)<sub>n</sub>), according to Leblanc et al., 1997.



**Fig S7 (Related to Figure 7)**

(A) Density profile of ZAM-mapping regulatory piRNAs (23-29nt small-RNAs complexed with Argonaute proteins) along the ZAM internal sequence produced in *flamΔZAM*+ chr II RevIH2 ovaries (allowing up to 3 mismatches). Sense and antisense reads are represented in black and grey, respectively. (B) Scatter plot showing log<sub>2</sub> of antisense regulatory piRNAs mapping individuals TEs in control ovaries (*white sKD*) versus *flamΔZAM* + chr II RevIH2 (upper panel) or chr III RevIH2 (lower panel) ovaries. 0-3 mismatches are allowed for the mapping of antisense piRNAs on TEs and normalised per millions of piRNA genome. (C) Density profile of ZAM-mapping total small RNA of 23 to 29nt along the ZAM sequence produced in *yb sKD* ovaries (allowing up to 3 mismatches). Sense and antisense reads are represented in black and grey, respectively. (D) Density profile of ZAM-mapping Piwi-bound piRNAs along the ZAM internal sequence produced in *yb-sKD* ovaries (allowing up to 3 mismatches). Sense and antisense reads are represented in black and grey, respectively. (E) Scatter plot showing log<sub>2</sub> of antisense regulatory siRNAs mapping individuals TEs in control ovaries (*white sKD*) versus *flamΔZAM* + chr III RevIH2. 1-3 mismatches are allowed for the mapping of antisense siRNAs on TEs and normalised per millions of siRNA genome.

## Supplementary Table S1 – Drosophila lines

Category	Gene name	CG Number	Stock number	Genotype	Source	Notes
RNAi line	white	CG2759	BDSC_35573	$\{1\} \{sc\} \{v1\} \{sev21\}; Pj\{t7.7\} \{v11.8\}; TRIP.G100094\{attP2\}$	Bloomington	
RNAi line	armitage	CG11513	BDSC_34789	$\{1\} \{sc\} \{v1\} \{sev21\}; Pj\{t7.7\} \{v11.8\}; TRIP.HMS2009\{attP2\}$	Bloomington	
RNAi line	yb	CG6272	VDRC_110056	$\{sc\} \{v1\} \{sev21\}; Pj\{t7.7\} \{v11.8\}; TRIP.HMS2009\{attP2\}$	VDRC	
RNAi line	Pw1	CG6122	BDSC_37481	$\{1\} \{sc\} \{v1\} \{sev21\}; Pj\{t7.7\} \{v11.8\}; TRIP.G10026\{attP40\}$	Bloomington	
RNAi line	histone	CG6999	VDRC_24927	$\{sc\} \{v1\} \{sev21\}; Pj\{t7.7\} \{v11.8\}; TRIP.HMS2009\{attP2\}$	VDRC	
RNAi line	Jucchini	CG13314	BDSC_35227	$\{1\} \{sc\} \{v1\} \{sev21\}; Pj\{t7.7\} \{v11.8\}; TRIP.G10011\{attP2\}$	Bloomington	
RNAi line	Argonaute 3	CG49300	BDSC_35232	$\{1\} \{sc\} \{v1\} \{sev21\}; Pj\{t7.7\} \{v11.8\}; TRIP.G10017\{attP2\}$	Bloomington	
RNAi line	Aubergine	CG6137	BDSC_35201	$\{1\} \{sc\} \{v1\} \{sev21\}; Pj\{t7.7\} \{v11.8\}; TRIP.G10079\{attP2\}$	Bloomington	
Wildtype	white (w1118)		BDSC_3605	w1118::	Bloomington	
nos-Gal4 driver			BDSC_4937	w1118; UASnos-Ch-GAL4-VP16-nos.UTR;CG6325[MVD1]	Bloomington	
TraffJam gal4 driver				w1118; UASJam4	Unknown - Cultivated in the lab for many years	
Rev1-H211					Produced in the lab	Isogenic line derived from the Rev1-H2 stock
Rev1-H212					Produced in the lab	Isogenic line derived from the Rev1-H2 stock
Rev1-H213					Produced in the lab	Isogenic line derived from the Rev1-H2 stock
Rev1-H214					Produced in the lab	Isogenic line derived from the Rev1-H2 stock

## Supplementary Table S2 – smFISH probes

Oligo number	ZAM (design for this paper)	Burdock ( from Senti and al.,2015)
1	GAAAAAGAAATGTTGGCAAT	aggttatcactgtcactcat
2	AATATGGGCAGTAAAATACT	gcttgacactagggttctca
3	AGACCATATATTTATGGCGT	tgggtgcgctgaccatttatt
4	TACAGATCACAAGCCACTAA	tttttctgactgttttgc
5	TGAATTTCAAAGAACCACAAC	ccaccaatattctgatgact
6	AAAAATAATTCGTTGGAGAC	cgctagttgtctttgtttt
7	TCCAACCTCATGGAATACAAT	aactgagcttcaaaattccat
8	TCACAAGAAAGTTCACAAA	aacgacactctctacgatcg
9	TGTAATTGCAGACGCCTTAA	tgattgttttgactgcga
10	GTAGAGCGGACCCAAATTTA	tcattcatctctgcaattg
11	AACTACAACGAAACACTGAC	gggttctcactgtaatttcc
12	TGTTAAGCCTTGCCCCACAT	ctacataagtttccacctct
13	CCGAAAAAAGCTATTAACGAA	tatatctagaggctgctac
14	AACACGCAACTCATACTAGA	gccatcaaatctggcagag
15	TCTTACCAACTACAACACC	cttccctcgtaattttgaa
16	AGGAAAAAATATTACAGCC	ccaagagcttggtaaaatgt
17	AAAACTTGAAAGGAAACCTT	aagccaatactgtattcct
18	AAAACCTAACAGGATTGCG	gctttgaaatgaacggagt
19	CATTCTGGCGGACGATAAT	tgtgaaatcaagcggctga
20	TCCGCATTAATCGAAAAAGC	tagatagatggcctttgt
21	TTCTCCAGAAATCGTAGGA	gtcgcgaagttagatgctct
22	ACCCCGAACAAAACAAATTC	tcagttaagatcatgtctcc
23	GGAATATCACACTAACAGCA	tcttgttgtaagtagggtc
24	GAGGCATAGACGAAACATTC	gccaaggcactcaaaatgt
25	TTCACTCAAACGAGAAACC	gtcctgtactttcattcag
26	ACTTCCCAATATGAAAAAC	ctcgggttacaatacacgt
27	AATTAGGAATTGCGAAACCT	atgtcgtcaacgatttctt
28	GTCTAAACTCAAATACGAC	atctttggacttttgcaa
29	AGGCAACCACAAAATATAGT	tggctgactgcacctctg
30	ATTTGAAACCCAGAAACCC	aaggcgaattgtaacgct
31	GAAACCCCTGCACATAATAC	ccctgtctcaaattttta
32	ACCTGACAATCATAGACAAA	tgccctttgctgattttctg
33	TCTCAAAATTCGAGCTGTC	tgagtttctatccttctgc
34	TACCCCATCCCAATAGAAA	tttttagttgcatgggat
35	CGGCATCAATTGCATCAAAG	cactgacgctgtctaaaat
36	TTCAGTCAATTCGGACTACC	ggattgttaactggactgg
37	CAAAAACTAATACAGGACC	ttctctgatgcatggat
38	AGGAGTAGAATTTGCAACG	ggagtttgcaaaaatttcc
39	CCATGTCAATCTCCAGC	aatattccctgagtcggaa
40	AGTAATCTCCAGTAGAACG	ggttcgatgcgaattgatt
41	ACACTCCTCTTGACAGAGA	ctgactgtgtttgtctgtt
42	AATACTAGACACCGGAAAA	ttgcatcaccaaaatgtc
43	ACACAAATACCTACAGACC	cactcgttgctttttgat
44	ACGAAGAAATAATGTCAGAA	ttttggccctgtaaacctg
45	ATAACAACGCAATCCACTCC	ggtaagttagttgatcctctg
46	CCACCAACACACCCCTTTT	ctcactgacctttctcat
47	TGGTAGGACCCATTTATTCG	ccctttgataaccgtctgta
48	ATACCAATAATGAGCATGA	ataatcatcaacctcgctt

## Supplementary Table S3 - Antibody

Antibody	Source	Dilution	Fluorochrome	Species	Reference
<b>Anticorps primaires</b>					
Goat anti-GFP	abcam	1/1000			ab5450
Rabbit anti-Aubergine	Homemade (Eurogentec) - polyclonal antibody	1/1000			
Rabbit anti-Gag-ZAM	Homemade	1/100			
Rabbit anti Env-ZAM	Homemade	1/1000			
<b>Anticorps secondaires</b>					
Alexa Fluor 488 donkey anti-goat IgG (H+L)	Invitrogen (Fisher)	1/1000	Alexa 488	goat	A11055
Donkey anti-rabbit Alexa 488 IgG (H+L) 2mg/ml	life tech	1/1000	Alexa 488	rabbit	A21206
Goat anti-rat Cy3	Jackson ImmunoResearch	1/1000	Cy3	rat	112-165-143
<b>Marqueur fluo</b>					
DAPI	Sigma	1/500			D8417-1MG

## Supplementary Table S4 - Primers

	Name	Sequence	Used for
Rp49	323-rp49r	AAACGCGGTTCTGCATGAG	RT-qPCR
	324-rp49F	GACGCTCAAGGGACAGTATCTG	
Burdock	135_primer_sens_Burdock1	GGACAAACTCCATTCCGACTC	RT-qPCR
	136_primer_antisens_Burdock_1	TCCCTGAGCCTGACTGTGT	
Gypsy	1864-Gypsy-s	CGCTCGAAGGTTACCAGGTAGGTTCC	RT-qPCR
	1865-Gypsy-as	CTTACGTTCTGCGAGCGGTCT	
ZAM	1181-Zam_env_sens_qpcr	CTCCCAATCATCTCCTCCAA	RT-qPCR
	1182-zam_env_antisens_qpcr	AGAAGTCGGGGTGCTCTTCT	
ZAM-flam	353_ZAM-FLAM_vasaCas9_F4	TCGCTTAGGAGGGTAGCTTG	CRISPR SCREEN
	355_ZAM-FLAM_vasaCas9_R4	CTTAACACCTACTGTCACTGTC	
ZAM insertion in flamenco	COM21648188	ctagtgtctttggtaccatgg	Screen
	COM21649620AS	tatcaactatcagactatcact	
ZAM insertion in cluster 9	308_ZAM insertion Site 3-4-5_F	CTCACCATTCTCCTTGAC	Screen
	Q5 : 1181-Zam_env_sens_qpcr	CTCCCAATCATCTCCTCCAA	

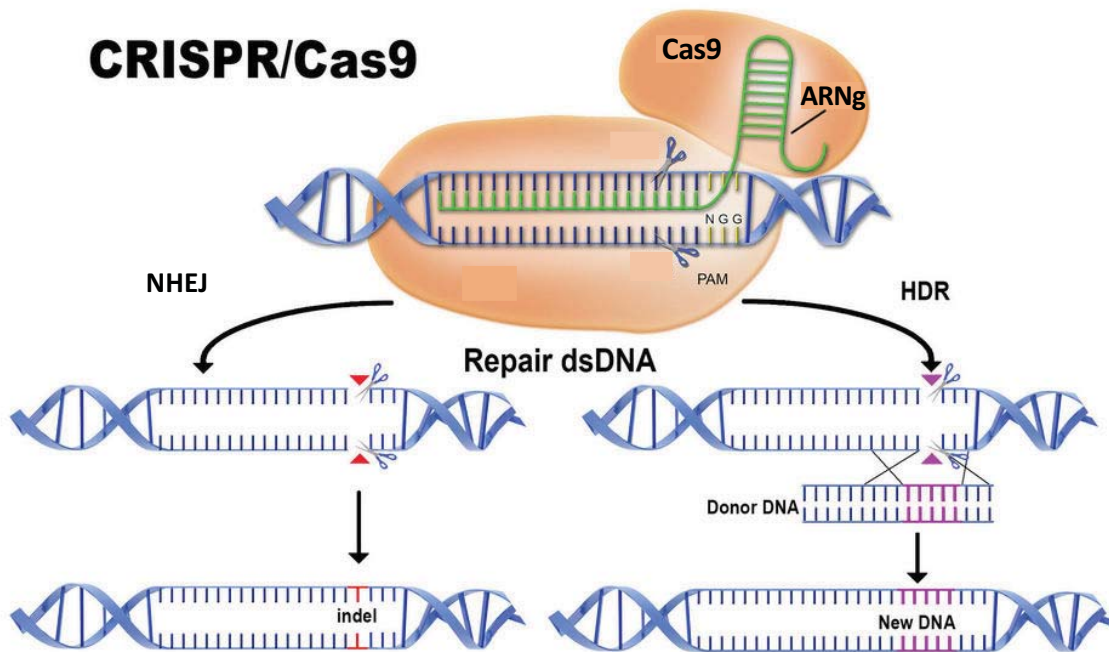
## Supplementary Table S5 – small-RNA seq

### Small RNA seq

- sRNA from Rev1-H2i1
- sRNA from Rev1-H2i2
- sRNA from Rev1-H2i3
- sRNA from Rev1-H2i4
- sRNA- from Rev1-H2i2- nos-Gal4- expressed dsRNA-white
- sRNA- from Rev1-H2i2- nos-Gal4- expressed dsRNA Ago3
- sRNA- from Rev1-H2i2- nos-Gal4- expressed dsRNA Piwi
- sRNA- from Rev1-H2i2- nos-Gal4- expressed dsRNA Zuc
- sRNA from tj-Gal4-expressed dsRNA-Vret
- sRNA from tj-Gal4-expressed dsRNA-white
- sRNA from CRISPR flamΔZAM
- Total RNA from tj-Gal4-expressed dsRNA-Vret
- Total RNA from tj-Gal4-expressed dsRNA-white
- Total RNA from tj-Gal4-expressed dsRNA-yb







**Figure 1 : Principe du CRISPR/Cas9 et des différentes voies de réparation**

Les ARNg dirigent la protéine Cas9 vers une séquence cible de 20 nts permettant de générer une cassure double brin au site cible. La cassure est réparée par l'une des deux voies de réparation cellulaire : la jonction des extrémités non homologues (NHEJ) ou la réparation dirigée par homologie (HDR). La réparation NHEJ induit une jonction des deux extrémités de la cassure et génère des variations de séquences (insertions ou délétions : *indel*) alors que la recombinaison homologue répare la cassure en utilisant comme matrice une séquence homologue (*donor DNA*).

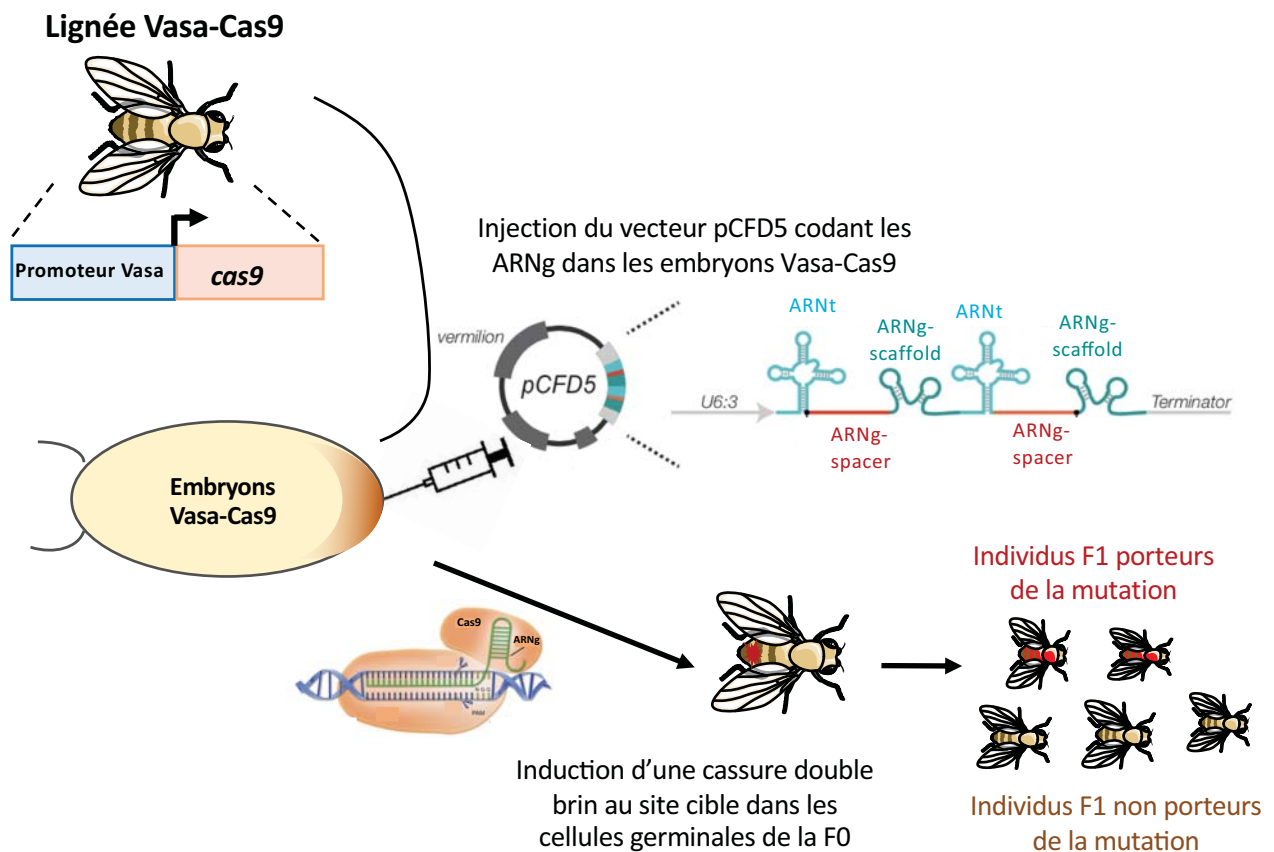
### Partie 3 : Délétion d'un élément transposable dans un cluster de piARNs

Dans cette partie, j'ai souhaité revenir sur la génération de la lignée CRISPR-*flam* $\Delta$ ZAM. Nous souhaitons éliminer la copie de ZAM insérée dans le cluster de piARNs *flamenco* afin de suivre en temps réel l'impact sur l'expression et la régulation de ZAM. Dans une étude publiée récemment, Daniel Gebert et ses collaborateurs ont éliminé des clusters de piARNs complets en utilisant la recombinaison homologe (Gebert et al., 2021). Cependant, réaliser une délétion ciblée par CRISPR/Cas9 dans un cluster de piARNs, région très répétée du génome, n'a encore jamais été accompli. Éliminer précisément un seul et unique élément transposable d'un cluster de piARNs permet de suivre de manière précise l'effet d'une telle délétion sur la régulation de cet ET, ainsi que les conséquences de la dérégulation de cet ET spécifiquement. Créer cette lignée a été un vrai défi qui a nécessité de nombreux essais (quatre pour être exact, et plus de deux années d'acharnement). En effet, nous étions face à plusieurs difficultés que je vais détailler dans la partie suivante. Cependant, la lignée finalement obtenue permettra d'aborder de nombreuses questions, jusqu'alors non résolues, et sera un outil robuste pour mener différents projets.

#### I- Fonctionnement de la technique CRISPR-Cas9 et plan d'action

##### I. 1) Principe général de la technique de "ciseaux génétiques" CRISPR-Cas9

La découverte du système CRISPR-Cas9 a permis de faire progresser de manière significative la capacité à effectuer des modifications ciblées du génome. Chez la drosophile, le système CRISPR-Cas9 est dorénavant largement utilisé pour supprimer, remplacer, marquer et modifier les gènes et autres régions du génome. C'est un outil puissant et efficace, et les récentes innovations techniques ont permis de faciliter son utilisation. Le système de *Streptococcus pyogenes* est le plus utilisé, il est composé uniquement de deux éléments : une endonucléase Cas9 et un petit ARN spécifique appelé ARN guide (ARNg). Les ARNg interagissent avec la protéine Cas9 et guident l'endonucléase vers des séquences d'ADN cibles spécifiques complémentaires à l'ARNg sur 20 nts. Après fixation sur la région cible, la protéine Cas9 utilise ses deux domaines nucléase pour générer une cassure double brin (**Figure 1**). La seule exigence majeure pour le potentiel site de clivage est la présence d'un motif adjacent "*protospacer*" (PAM) de 3 pb de type NGG qui doit être localisé immédiatement en 3' de la séquence cible de



**Figure 2 : Stratégie d'édition du génome par la technique CRISPR-Cas9 - Utilisation de la lignée Vasa-Cas9 et injection du vecteur pCFD5 exprimant les ARNg**

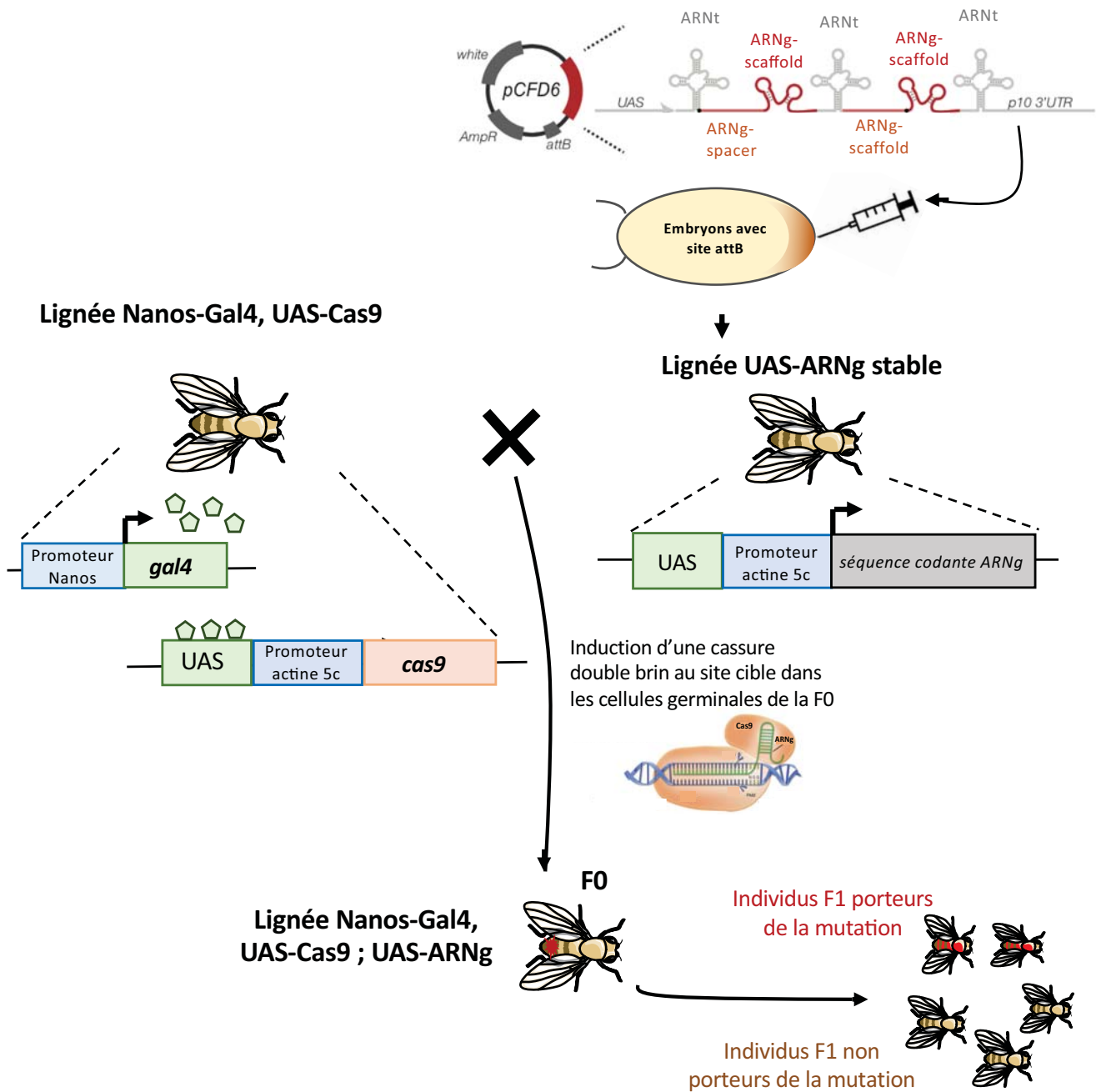
Dans la lignée transgénique Vasa-Cas9, l'expression de l'endonucléase Cas9 est sous la dépendance du promoteur germlinal du gène *vasa*. Le vecteur permettant l'expression de l'ARNg (pCFD5) est injecté dans les embryons Vasa-Cas9 au pôle postérieur, là où se formeront les futures cellules germinales. Le vecteur pCFD5 permet l'expression d'un ou de plusieurs ARNg sous le contrôle d'un seul promoteur puissant et ubiquitaire dépendant de la polymérase III : le promoteur U6:3. L'ARN guide (ARNg) est composé de deux parties : la séquence « *scaffold* » d'une part est nécessaire pour le recrutement de la Cas9, la partie « *spacer* », d'autre part, est une séquence de 20 nts qui définit le site génomique ciblé à modifier. Les séquences codant chaque ARNg sont séparées par une séquence codant un ARN de transfert (ARNt). L'ARN polycistronique généré est ensuite pris en charge par le système endogène de maturation des ARNt de la cellule qui clive les ARNt, libérant ainsi les ARNg. L'expression de la protéine Cas9 et des ARNg permet de générer des cassures double brin au niveau des sites génomiques cibles dans les cellules germinales. Au final, les cellules germinales ainsi mutées de manière ciblée donneront naissance à des individus F1 porteurs de la mutation alors que les autres cellules germinales donneront naissance à des individus F1 non porteurs de la mutation.

20 nts. Après induction d'une cassure double brin, celle-ci est réparée par l'une des deux voies de réparation cellulaire : la jonction des extrémités non homologues (*non-homologous end joining*, NHEJ) ou la réparation dirigée par homologie (*homology-directed repair*, HDR). Afin de générer des mutations transmises aux descendants et ainsi établir une lignée stable possédant la mutation, celle-ci doit être générée dans les cellules germinales.

### *I. 2) Lignées de drosophiles utilisées*

La première nécessité pour effectuer une modification génétique avec le système CRISPR-Cas9 est la production de l'endonucléase Cas9. Il existe plusieurs possibilités qui permettent d'exprimer la protéine Cas9 dans la lignée germinale. Le moyen le plus simple est d'utiliser une souche de drosophile transgénique où le gène codant la protéine Cas9 a été introduit dans le génome et se trouve sous le contrôle d'un promoteur exprimé dans les cellules germinales tel que les promoteurs des gènes *vasa* ou *nanos*. Dans notre cas, nous avons utilisé la lignée Vasa-Cas9 (**Figure 2**). L'ARNg est ensuite apporté par injection d'un vecteur de type pCFD5 où l'expression de l'ARNg est contrôlée par un promoteur ubiquitaire fort (U6 :3). Ce vecteur est injecté directement dans les embryons Vasa-Cas9, au niveau postérieur où se formeront les cellules polaires, futures cellules germinales. L'expression de l'ARNg et de la protéine Cas9 dans les cellules germinales de l'embryon permet d'induire une cassure au niveau du site cible. Nous avons utilisé cette technique pour les deux premiers essais mais, après les échecs rencontrés, nous avons ensuite essayé une autre technique, connu pour être plus efficace.

En effet, il est aussi possible d'utiliser une lignée où l'expression de la Cas9 est inductible grâce au système UAS-Gal4. L'expression de la protéine Cas9 est alors sous le contrôle d'un promoteur dépendant d'une région activatrice UAS (*Upstream Activation Sequence*). La fixation de la protéine Gal4 sur la région UAS est nécessaire pour activer l'expression du gène codant la Cas9. Le gène permettant l'expression de la protéine Gal4 est quant à lui placé sous le contrôle d'un promoteur choisi pour son expression spécifique dans les cellules d'intérêts, ici les cellules germinales (pilote Nanos-Gal4). Deux transgènes sont alors présents dans le génome de cette lignée : le transgène Nanos-Gal4 et le transgène UAS-Cas9. Une autre lignée transgénique est également générée pour permettre l'expression de l'ARNg. Dans ce cas, la séquence codant l'ARNg est insérée en aval d'un promoteur de type UAS dans un vecteur de type pCFD6. Ce vecteur est injecté dans une lignée de drosophile qui possède un site de recombinaison de type attB permettant l'intégration du vecteur codant les ARNg à un locus génomique spécifique. Il suffit ensuite de croiser la lignée Nanos-Gal4 ; UAS-Cas9 avec la lignée portant le transgène UAS-ARNg afin d'obtenir une lignée où tous les transgènes sont



**Figure 3 : Stratégie d'édition du génome par la technique CRISPR- Cas9 - Utilisation des lignées Nanos-Gal4 ; UAS-Cas9 et UAS-ARNg**

Dans la lignée *Nanos-Gal4;UAS-Cas9*, l'expression de la protéine Cas9 est inducible grâce à la fixation de la protéine Gal4 sur la séquence activatrice UAS. La protéine Gal4 est elle-même sous le contrôle du promoteur du gène *nanos* exprimé spécifiquement dans les cellules germinales. Une lignée *UAS-ARNg* est ensuite générée grâce à l'injection du vecteur *pCFD6* codant les ARNg dans des embryons possédant un site *attP* dans leur génome et exprimant l'intégrase *phiC31*. Le vecteur possède également un site *attB* permettant son insertion dans le génome de l'embryon au site *attP* par recombinaison, créant ainsi le transgène *UAS-ARNg*. Ce transgène permet l'expression d'un ou de plusieurs ARNg sous le contrôle du système *UAS-Gal4*. En effet, la fixation de la protéine Gal4 sur la séquence UAS est nécessaire pour l'expression des ARNg à partir du transgène *UAS-ARNg*. La lignée transgénique *UAS-ARNg* est ensuite croisée avec la lignée *Nanos-Gal4 ; UAS-Cas9* afin de permettre une expression des ARNg et de la Cas9 dans les cellules germinales grâce au pilote *Nanos-Gal4*. Cela induit une cassure d'ADN double brin au site cible dans les cellules germinales (au stade embryonnaire, larvaire ou adulte). Les cellules germinales mutées donneront naissance à des individus F1 porteurs de la mutation alors que les autres cellules germinales donneront naissance à des individus F1 non porteurs de la mutation.

présents et où l'expression de la protéine Cas9 et de l'ARNg sont sous la dépendance du promoteur germinale Nanos *via* le pilote Nanos-Gal4 (**Figure 3**). L'avantage de cette technique est que les transgènes sont tous intégrés de manière permanente dans le génome, ce qui permet l'expression de la protéine Cas9 et de l'ARNg dans les cellules germinales de la drosophile à tous les stades de développement grâce au pilote Nanos-Gal4. Ainsi, la coupure par la Cas9 pourra se produire dans les cellules germinales à partir de l'embryogenèse jusqu'au stade adulte. Cette stratégie s'est révélée efficace et nous a permis d'obtenir les lignées mutantes souhaitées.

### *I. 3) Site ciblé pour la mutation*

Nous avons pour but d'éliminer la copie de *ZAM* présente dans le cluster de piARNs *flamenco* (copie appelée *ZAM-flam*). Dans le génome de référence de *Drosophila melanogaster* (lignée Iso1A), *ZAM-flam* est localisée aux positions génomiques X:21,778,811...21,783,994. La copie *ZAM-flam* n'est pas une copie pleine longueur puisqu'une grosse partie du gène *pol* et plus de deux tiers du gène *env* de *ZAM* ne sont pas présents. La séquence consensus de *ZAM* fait normalement 8433 pb alors que le *ZAM-flam* fait seulement 5185 pb. Le but de la mutagenèse CRISPR-Cas9 était de générer une délétion dans le cluster afin d'éliminer la copie *ZAM-flam* entière. Pour effectuer une délétion, il est nécessaire d'utiliser deux ARNs guides qui ciblent des séquences de part et d'autre de la région à éliminer et génèrent ainsi des cassures double brins de chaque côté. Le fragment localisé entre les deux cassures sera éliminé lorsque l'ADN sera réparé par NHEJ. Les deux ARNg sont clonés dans le même vecteur d'expression, sous la dépendance du même promoteur.

## **II - Les difficultés rencontrées pour créer la lignée CRISPR-*flam* $\Delta$ *ZAM***

La plus grosse difficulté de cette expérience concerne la nature même de la région ciblée pour la délétion. *flamenco* est une région riche en éléments transposables, ces derniers ne sont pas présents uniquement dans *flam* mais sont également présents à d'autres endroits dans le génome. Ainsi, les séquences flanquantes de *ZAM-flam* sont hautement répétées. Par conséquent, il est particulièrement difficile de trouver des guides spécifiques qui ne cibleront que les régions entourant la copie *ZAM-flam* sans viser d'autres régions génomiques. Il faut également savoir que l'assemblage génomique de telles régions répétées est complexe, induisant fréquemment des erreurs dans la séquence fournie par les bases de données. De plus,



les clusters de piARNs tels que *flamenco* évoluent différemment et rapidement dans chaque lignée de drosophile, en fonction de l'activité des éléments transposables et des nouvelles invasions subies (Zanni et al., 2013; Gebert et al., 2021; Wierzbicki et al., 2021). Ainsi, les clusters de piARNs peuvent être très différents en fonction de la lignée de drosophile utilisée. Il est donc possible que les ARNg créés ciblent le *flamenco* d'une lignée de drosophile mais soient incapables de cibler cette région dans une autre lignée car celle-ci peut présenter des polymorphismes, et même être totalement différente, voire absente.

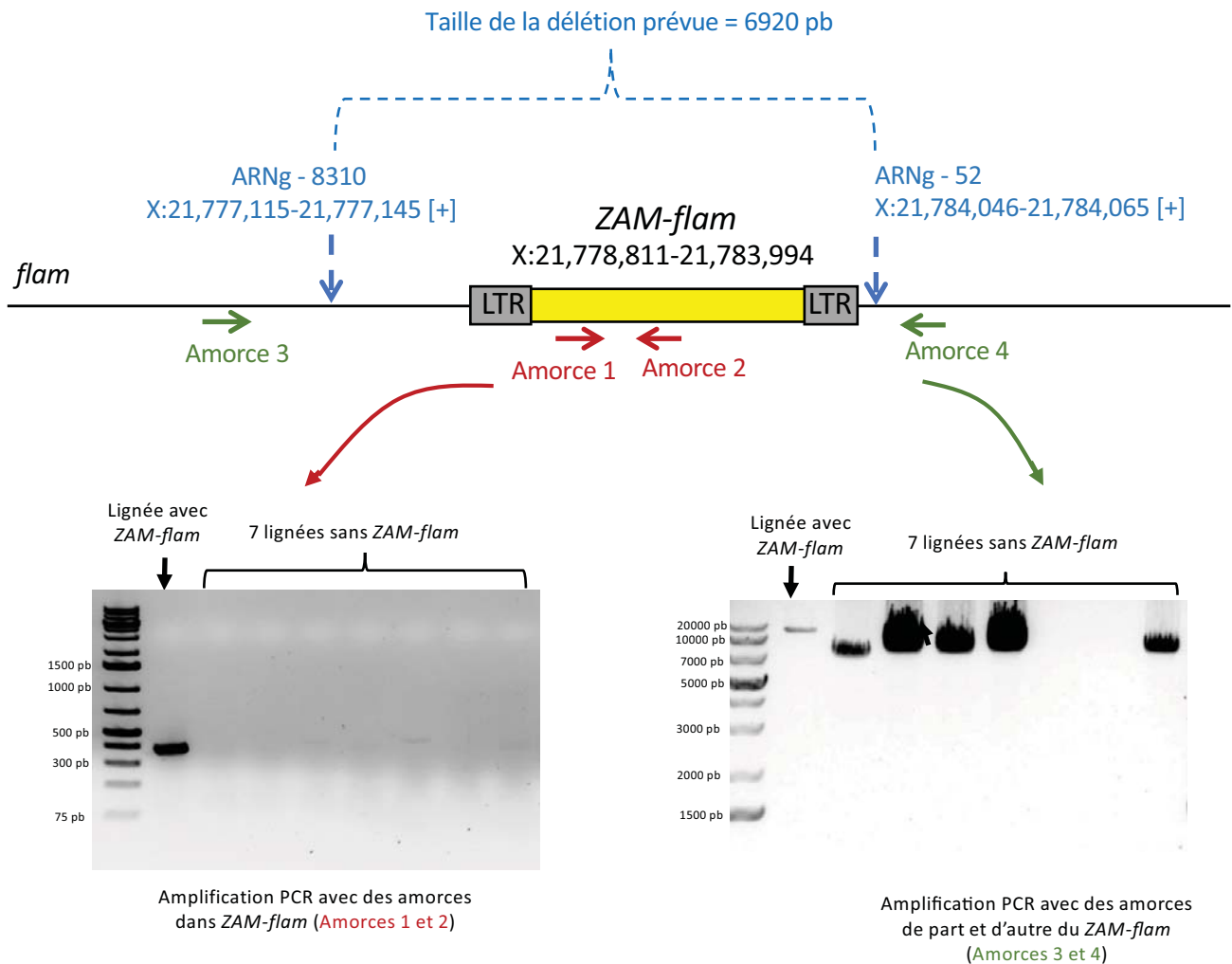
Face à ces difficultés, qui se sont révélées à nous petit à petit, les premiers essais se sont conclus par un échec. Je me suis heurtée aux problèmes de polymorphismes entre les génomes et aux erreurs d'assemblage génomique. J'ai dans un premier temps utilisé un générateur d'ARNg en ligne pour la technique du CRISPR-Cas9 (<http://targetfinder.flycrispr.neuro.brown.edu>). Le problème majeur fût que les ARNg générés étaient incapables de se fixer sur la région souhaitée dans la lignée Vasa-Cas9 utilisée dans les premières expériences. En effet, le génome de référence utilisé pour le design des ARNg sur ce site web est celui de la lignée de drosophile Iso1A, or le cluster *flam* se révèle être relativement différent entre les lignées Vasa-Cas9 et Iso1A.

J'ai donc, dans un second temps, créé de nouveaux ARNg en vérifiant que les séquences reconnues par les ARNg étaient bien présentes dans le génome de la lignée Vasa-Cas9 en utilisant comme référence le génome Vasa-cas9\_dmel disponible sur les bases de données. Malgré ces vérifications, ce second essai s'est aussi conclu en échec.

Nous avons alors émis l'hypothèse que la position génomique de *flamenco* limitait son accessibilité aux ARNg et à la Cas9. En effet, *flam* est localisé dans l'hétérochromatine péricentromérique du chromosome X qui est fortement condensée. J'ai par conséquent essayé d'utiliser plusieurs ARNg en amont et en aval de *ZAM-flam* (quatre ARNg : deux de chaque côté de *ZAM-flam*) afin de maximiser les chances de coupures. Mais, là encore, ce troisième essai a échoué...

Pour faire face à ces problèmes, j'ai reçu l'aide précieuse de Silke Jensen, experte en analyse de séquence, particulièrement dans l'étude des séquences répétées. Après des travaux poussés, nous avons constaté que le génome de la lignée Vasa-Cas9 utilisée pour le design des ARNg était en réalité mal assemblé, notamment au niveau des régions répétées. Ces régions ont été remplacées par les régions localisées au même site génomique dans la lignée de référence Iso1A dans l'assemblage final. Cette mauvaise annotation a induit des erreurs dans la séquence fournie par rapport à la séquence réellement présente dans le génome de cette lignée. Malheureusement, la plupart des ARNg que j'avais créés se trouvaient dans des régions mal assemblées, expliquant





#### Figure 4 : Délétion du *ZAM-flam* par la technique CRISPR-Cas9

Deux ARNg (ARNg-8310 et ARNg-52) sont utilisés pour induire la délétion de la copie *ZAM-flam* localisée dans le cluster de piARNs *flamenco*. Les ARNg ciblent des régions localisées de part et d'autres de *ZAM-flam* indiquée par des flèches en pointillés bleus. Cela devrait induire une délétion de 6920 pb comprenant la séquence de *ZAM-flam*.

Les amorces 1 et 2 (flèches rouges) ont été utilisées pour le crible par amplification PCR chez les individus F1. Aucune amplification n'est visible dans les lignées où la délétion s'est produite (migration sur gel présentée à gauche) alors qu'une bande caractéristique de 364 pb est présente lorsque *ZAM-flam* est toujours présente. Les amorces 3 et 4 (flèches vertes) ont été utilisées pour amplifier la région entière ciblée par la délétion (migration sur gel présentée à droite). Une amplification d'une taille supérieure à 10 kb est observée pour la lignée contrôle et une amplification de taille inférieure (6 à 7 kb) est visible pour 5 lignées où la copie *ZAM-flam* est absente.

Séquences des amorces :

Amorce 1 : GTGAGTGTATCCAGATGGAGT

Amorce 2 : ACACCGAAGTCAACTATTCAAT

Amorce 3 : TCTAACGGCTGACCTACTGT

Amorce 4 : GTAAACCCGTCGGAACAAAGT

pourquoi les ARNg ne pouvaient en réalité pas reconnaître la séquence cible dans le génome de la lignée Vasa-Cas9.

Nous avons alors décidé d'aborder une stratégie de design plus poussée, et Silke a généré la séquence des ARNg sans l'utilisation d'un outil dédié, en se basant directement sur les prérequis nécessaires pour la génération de ARNg CRISPR, à savoir :

- la présence d'un motif PAM (5'-NGG-3') localisé immédiatement en 3' de la séquence de 20 nts ciblée par l'ARNg,
- la région ciblée par l'ARNg qui doit être unique dans le génome : aucune autre région génomique ne doit présenter plus de 15 nts de complémentarité avec l'ARNg,
- le pourcentage de CG présent dans la séquence cible qui doit être supérieur à 35 %.

Afin d'être sûres que les guides cibleront la région souhaitée, nous avons décidé de réaliser la délétion sur la lignée dont le génome a été séquencé avec la technologie Nanopore permettant la génération de *reads* de grande taille, et donc un meilleur assemblage du génome (lignée Iso1A). Pour éviter toute erreur d'assemblage, Silke a généré les ARNg en utilisant directement les *reads* obtenus lors du séquençage génomique. Cela a donc permis de vérifier que la séquence ciblée est réellement présente dans le génome de la lignée utilisée et se trouve localisée au bon endroit. J'ai également vérifié, par amplification PCR (*Polymerase Chain Reaction*) et séquençage du produit d'amplification, la séquence de la région ciblée par les ARNg. Notre travail a ainsi permis de générer plusieurs ARNg localisés de part et d'autre de la copie *ZAM-flam* pour lesquels nous étions sûres qu'ils étaient capables de cibler les régions souhaitées dans la lignée Iso1A. Nous avons tout d'abord sélectionné deux ARNg : un localisé en amont de *ZAM-flam* et l'autre localisé en aval. La séquence localisée en amont de *ZAM-flam* est fortement répétée et nous a obligé à utiliser un ARNg localisé 1,7 kb en amont de *ZAM-flam*, il était impossible de trouver une séquence cible plus proche. Le deuxième guide était quant à lui localisé 52 pb en aval de *ZAM-flam*. La délétion totale prévue était de 6920 pb (**Figure 4**).

Nous avons également décidé d'utiliser les lignées où l'expression de la Cas9 et des ARNg sont inductibles (Nanos-Gal4; UAS-Cas9 et UAS-ARNg) afin d'augmenter nos chances de réussite. Le problème majeur était que nous ne connaissions pas le fond génétique de ces lignées. Nous n'étions donc pas sûres que les ARNg, générés en utilisant le génome de la lignée Iso1A, puissent reconnaître la séquence de *flamenco* présente dans ces lignées-là. Pour



détourner ce problème, Émilie Brasset nous a proposé une stratégie, certes chronophage mais très efficace, qui fût le secret de la réussite. *flamenco* étant localisé sur le chromosome X et les transgènes Nanos-Gal4, UAS-Cas9 et UAS-ARNg sur les chromosomes 2 et 3, nous avons remplacé les chromosomes X dans les deux lignées Nanos-Gal4 ; UAS-Cas9 et UAS-ARNg par le chromosome X de la lignée Iso1A. Nous étions ainsi sûres que les ARNg créés pourraient se fixer sur le *flamenco* présent dans ces lignées puisque nous avons utilisé cette séquence pour le design initial des ARNs guides.

### III – Obtention tant attendue de la lignée CRISPR-*flam* $\Delta$ ZAM

#### III. 1) Vérification de la présence de la délétion

Après croisement des lignées Nanos-Gal4 ; UAS-Cas9 et UAS-ARNg, des individus F0 contenant les éléments génomiques suivants ont été obtenus :

- le chromosome X de la lignée Iso1A,
- le transgène UAS-ARNg, permettant de produire les deux ARNg, inséré sur le chromosome 2,
- le transgène UAS-Cas9, permettant de produire la protéine Cas9 ainsi que le pilote Nanos-Gal4 induisant l'expression de la protéine Gal4 dans les cellules germinales, insérés sur le chromosome 3.

La délétion de *ZAM-flam* devait alors se produire dans les cellules germinales de ces individus F0 grâce à l'expression tissu-spécifique de la Cas9 et des deux ARNg. Cependant, la mutation peut ne pas toucher toutes les cellules germinales d'un individu F0. Uniquement les cellules germinales possédant la mutation pourront alors donner naissance, après reproduction, à des descendants F1 où, cette fois-ci, toutes les cellules de l'organisme porteront la mutation. Certains descendants pourront ne pas porter la délétion car ils seront issus de cellules germinales non mutées. Il est donc primordial d'analyser chaque individu F1 individuellement car ils sont chacun issu de cellules germinales différentes et, par conséquent, des événements différents ont pu se produire dans chacun d'entre eux. J'ai ainsi établi 500 lignées à partir d'individus F1 différents et analysé par crible PCR si ces lignées possédaient la délétion *flam* $\Delta$ ZAM. Pour cela, j'ai utilisé des amorces qui s'hybrident spécifiquement et uniquement au niveau de séquences internes de *ZAM-flam* : si la copie est éliminée comme attendu, aucune

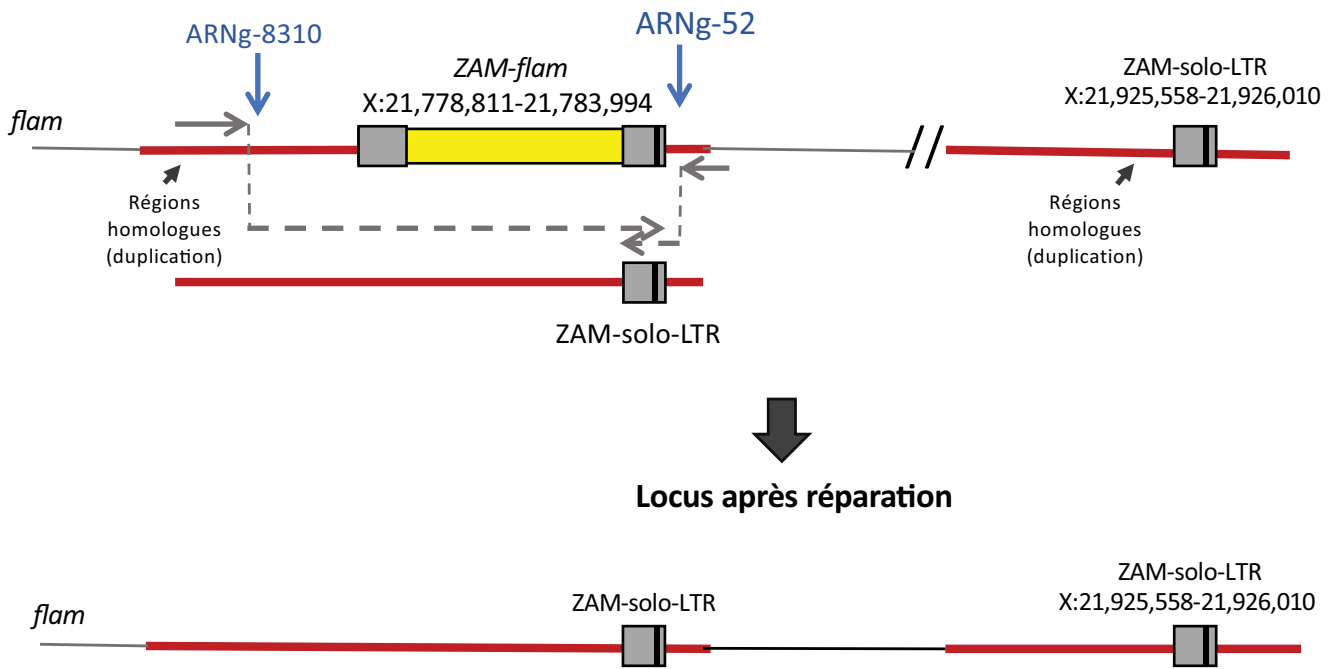


amplification ne sera possible. J'ai dans un premier temps analysé 94 lignées parmi lesquelles 7 ne présentaient pas d'amplification en utilisant des amorces localisées dans le *ZAM-flam* (**Figure 4** – migration sur gel présentée à gauche).

### III. 2) Différentes lignées CRISPR-*flam*Δ*ZAM* obtenues

Afin de déterminer la séquence exacte de la région contenant initialement le *ZAM-flam* dans les sept lignées potentiellement mutantes, j'ai utilisé des amorces localisées de part et d'autre des régions ciblées par les ARNg pour amplifier cette région par PCR. Encore une fois, le caractère hautement répété des régions localisées en amont et en aval de la région *ZAM-flam* ciblée nous a obligé à utiliser des amorces localisées assez loin des sites de coupure ciblés par les ARNg. Le produit d'amplification attendu fait 10079 pb lorsque l'ADN d'une lignée contrôle est utilisé. Un produit de taille compatible est effectivement observé pour la lignée contrôle (**Figure 4** – migration sur gel présentée à droite). Si *ZAM-flam* est absente de son locus initial, une amplification plus courte de 3159 pb ( $10079 - 6920 = 3159$ ) devrait être observée. De manière très surprenante, un produit d'amplification de taille de 6 à 7 kb a été observé dans cinq des lignées potentiellement mutantes (nous n'avons pas réussi à amplifier la région par PCR dans les deux autres lignées). Le produit d'amplification est tout de même plus petit que dans une lignée contrôle où *ZAM-flam* est encore présent, témoignant qu'une délétion semble bien s'être produite dans cette région (**Figure 4** – à droite). Ces premiers résultats suggèrent que la délétion effectuée n'est pas la délétion initialement prévue et qu'elle est potentiellement plus petite. Nous avons donc fait séquencer le produit d'amplification afin de déterminer la séquence exacte de la région ciblée pour la délétion dans les lignées potentiellement mutantes. Les résultats obtenus ont confirmé ceux observés précédemment : une partie de la région qui aurait normalement dû être éliminée est toujours présente dans les cinq lignées analysées. Après alignement sur le génome de référence de la lignée Iso1A, nous avons constaté qu'uniquement les séquences internes de *ZAM* ont été délétées et un solo-LTR de *ZAM* est maintenant présent au site où le *ZAM-flam* était initialement inséré. Les régions génomiques localisées en amont et en aval du *ZAM-flam* sont également toujours présentes, bien qu'un peu différentes. Alors comment expliquer ces résultats ?

Revenons sur le principe de la technique CRISPR-Cas9. Après coupure par la Cas9, les cassures d'ADN double brin générées doivent être réparées par un des deux systèmes de réparation cellulaires : la réparation NHEJ induit une ligation des deux extrémités des cassures,



**Figure 5 : Réparation par recombinaison homologue dans la lignée CRISPR *flam*□*ZAM***

Un solo-LTR de *ZAM* est présent dans *flamenco* environ 140 kb en aval de *ZAM-flam* (position X:21,925,558-21,926,010). Les régions localisées de part et d'autres de *ZAM-flam* et du solo-LTR (représentées en rouge) sont homologues car ces deux régions sont issues d'une duplication. Suite à la génération des cassures double brin par la Cas9 de part et d'autres de la copie *ZAM-flam*, la région homologue contenant le solo-LTR a servi de matrice pour la réparation dirigée par homologie (HDR). Cela aboutit à la présence d'un solo-LTR à la place de *ZAM-flam* initialement insérée à cette position, les séquences internes de *ZAM-flam* ont quant à elles été totalement éliminées.

là où la HDR, réparation dirigée par homologie, utilise une séquence homologue pour réparer la cassure. Nous espérons que les cassures soient réparées par NHEJ, induisant alors une délétion complète de la région localisée entre les deux coupures. Cependant, la délétion n'étant ici pas totale, il semblerait que les cassures aient plutôt été réparées par HDR. La HDR utilise une séquence homologue à la région de la cassure pour permettre la réparation. Le chromosome homologue ou la chromatide-sœur sont majoritairement utilisés pour permettre une réparation fidèle de la séquence endommagée. Cependant, dans notre cas, la réparation n'a pas été fidèle car (1) les parties internes de *ZAM-flam* ont été délétées dans les sept lignées analysées, et (2) la région encore présente entre les sites ciblés par les ARNg possède de nombreux polymorphismes par rapport à la séquence initiale. L'hypothèse la plus probable est que la séquence utilisée pour la réparation soit une séquence différente située ailleurs dans le génome. L'analyse de la séquence de *flamenco* montre que 30 kb de la région proximale de ce cluster a été dupliquée et se retrouve donc également présente environ 140 kb en aval. Cette région comprend un reliquat de *ZAM-flam* (Zytnicki et al., 2013). En effet, le *ZAM* présent dans cette région dupliquée est réduit, pour les cinq lignées mutantes identifiées, à un solo-LTR (localisé aux positions X :21,925,558-21,926,010). Cela suggère que, dans cette région dupliquée, une recombinaison homologue entre les LTRs de *ZAM* s'est produite après la duplication, éliminant les séquences internes de ce *ZAM*. De plus, certains ETs sont présents uniquement dans une des deux régions, témoignant que la duplication est probablement assez ancienne et que chacune des deux régions a ensuite évolué différemment. Cependant, les régions proches des séquences *ZAM* présentent encore près de 98 % d'homologie de séquence. Nous émettons donc l'hypothèse que la duplication, localisée à environ 140 kb en aval du *ZAM* ciblée par les ARNs guides, ait été utilisée pour la réparation après la création des cassures par la Cas9 (**Figure 5**). L'utilisation de cette région dupliquée comme matrice pour la réparation aurait ainsi abouti à la présence d'un solo-LTR à la place de la copie *ZAM-flam* normalement présente. Cette hypothèse est supportée par le fait que le fragment LTR, qui est maintenant présent dans nos lignées CRISPR à la place du *ZAM-flam*, est au moins sur une partie de sa longueur 100 % homologue au solo-LTR localisé dans la duplication. Nous avons également constaté pour certaines lignées mutantes que plusieurs matrices ont été utilisées pour la réparation (phénomène de conversion génique), conduisant à la création d'une région composée de séquences provenant de plusieurs loci génomiques différents.

Pour résumer, les ARNg ont reconnu les séquences présentes en amont et en aval de la copie *ZAM-flam* et ont permis de recruter l'endonucléase Cas9 qui a induit des cassures double brin de chaque côté. Les cassures ont ensuite été réparées par HDR en utilisant comme matrice une





région homologue, le solo-LTR de *ZAM* présent dans la duplication, combinée avec d'autres régions homologues dans le génome. Cela a induit la perte des séquences internes de *ZAM* qui, dans les cinq lignées mutées caractérisées par séquençage, sont maintenant totalement absentes de *flamenco*. Parmi ces cinq mutantes obtenues, une seule a été utilisée pour les analyses, cette lignée sera appelée : *flam* $\Delta$ *ZAM*. Les résultats obtenus avec cette lignée sont présentés dans la partie 5 de l'article 2.

#### **IV - Étude de l'effet de la délétion *ZAM-flam* sur la régulation et l'expression de *ZAM***

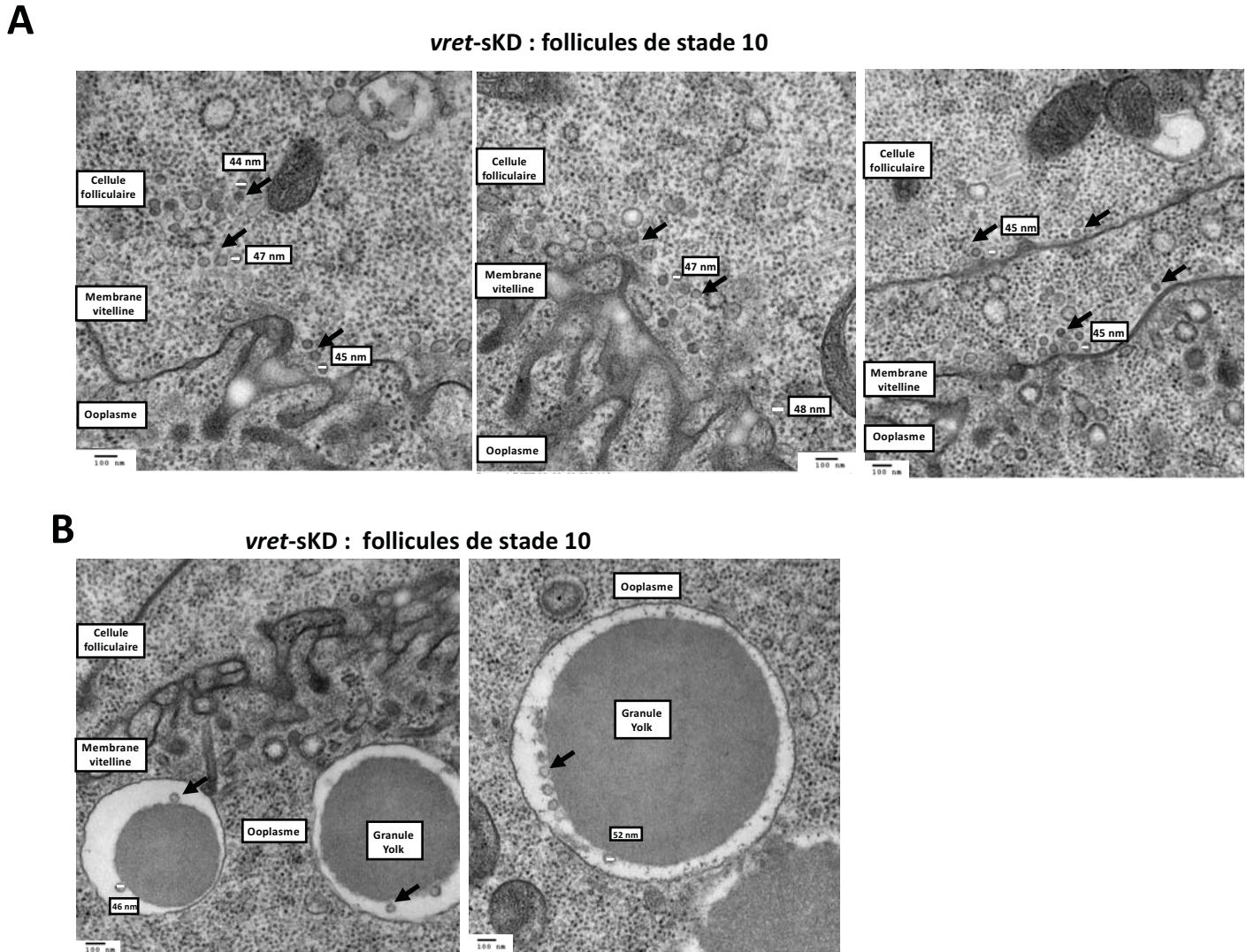
Les résultats obtenus avec la lignée *flam* $\Delta$ *ZAM* sont présentés dans la partie 5 de l'article 2. Pour résumer rapidement, nous avons pu montrer que l'élimination de *ZAM* de son cluster de piARNs régulateur, *flamenco*, induit une perte de production des piARNs ciblant cet ET. En absence de piARNs régulateurs, lorsque des copies actives de *ZAM* sont insérées dans le génome de la lignée *flam* $\Delta$ *ZAM*, une forte expression de *ZAM* est détectée dans les cellules folliculaires. De plus, *ZAM* envahi massivement l'ovocyte dans cette condition. Ces résultats confirment qu'en absence de production de piARNs dans les cellules germinales, la réactivation de *ZAM* conduit à une invasion de l'ovocyte : les ARNs de *ZAM* provenant des cellules folliculaires ne peuvent pas être dégradés post-transcriptionnellement. L'étude de la réponse primaire suite à la réactivation de *ZAM* a révélé qu'aucune production de piARNs ou de siARNs ne se met en place immédiatement pour contrôler l'invasion et potentiellement éviter de nouvelles insertions génomiques de *ZAM*. Nous avons également montré que la réactivation de *ZAM* dans les cellules somatiques folliculaires induit des défauts sévères de fertilité associés à une atrophie des ovaires.

#### **V- Expériences en cours et à venir sur la lignée CRISPR-*flam* $\Delta$ *ZAM***

Nous souhaitons poursuivre l'étude de la réponse cellulaire mise en place suite à une réactivation d'un ERV. Pour cela, nous avons réalisé dernièrement un séquençage des ARNs présents dans les ovaires de la lignée CRISPR-*flam* $\Delta$ *ZAM* possédant deux copies actives de *ZAM* (RNA-seq polyA) et les résultats devraient être disponibles prochainement. Nous souhaitons analyser si la dérégulation de *ZAM* affecte l'expression des gènes, notamment ceux codants des protéines impliquées dans la voie des piARNs. Un projet futur est également d'analyser si une réponse immunitaire est déclenchée à la suite de la réactivation d'un ERV. Pour



d'analyser si une réponse immunitaire est déclenchée à la suite de la réactivation d'un ERV. Pour cela, nous analyserons l'expression de gènes connus pour être impliqués dans la réponse antivirale. Les résultats obtenus avec les tests de fertilité montrent que la dérégulation de *ZAM* dans les cellules germinales induit une quasi-stérilité des femelles. Nous allons tout de même tenter d'établir une lignée homozygote CRISPR-*flam* $\Delta$ *ZAM* avec soit deux copies fonctionnellement actives de *ZAM* (localisées sur le chromosome X de la lignée RevI-H2), soit trois copies fonctionnelles de *ZAM* (localisées sur le chromosome III du génome RevI-H2), soit quatre copies fonctionnelles de *ZAM* (localisées sur le chromosome II de la lignée RevI-H2). Par ces expériences, nous souhaitons réaliser une étude transgénérationnelle afin d'analyser la dynamique de transposition de *ZAM* à travers les générations, ainsi que la mise en place d'une potentielle réponse adaptative. Cela nous permettrait d'analyser la présence de nouvelles insertions génomiques de *ZAM* au cours des générations grâce à des techniques sur chromosomes polytènes ou de séquençage génomique via la technologie de séquençage Oxford Nanopore. La production de petits ARNs (piARNs, siARNs) sera également analysée au cours des générations afin de détecter la mise en place d'une potentielle réponse qui permettrait de réprimer l'expression et la transposition de *ZAM*.



**Figure 1 : Détection de pseudo-particules virales de ZAM par microscopie électronique.**

(A) Des pseudo-particules d'environ 45 nm (montrées par les flèches noires), qui pourraient être des particules de ZAM, s'accumulent dans les cellules folliculaires à proximité de la membrane vitelline, qui sépare les cellules folliculaires de l'ovocyte, dans les follicules de stade 10 des drosophiles *vret-sKD*. (B) Des particules sont également détectées dans les granules Yolk localisés dans le cytoplasme de l'ovocyte des follicules de stade 10 des drosophiles *vret-sKD*.

## Partie 4 : Analyse du cycle rétroviral de *ZAM*

Les études passées analysant le cycle rétroviral de *ZAM* ont été réalisées en utilisant la lignée RevIH2 comme modèle, où une dérégulation de *ZAM* a été constatée dans les cellules folliculaires des ovaires (Leblanc et al., 2000; Brassat et al., 2006). L'équipe de Chantal Vaury avait détecté la présence de particules de *ZAM* dans les ovaires de la lignée RevIH2 par microscopie électronique (Leblanc et al., 2000). Ces particules semblaient principalement localisées dans les cellules folliculaires et très peu dans l'ovocyte. D'après les résultats obtenus durant ma thèse, nous savons désormais que la production de piARNs ciblant *ZAM* dans les cellules germinales de la lignée RevIH2 permet de dégrader post-transcriptionnellement les ARNs de *ZAM* provenant des cellules folliculaires. Ainsi, la lignée RevIH2 n'est probablement pas le meilleur modèle pour étudier le cycle rétroviral de *ZAM* puisque celui-ci semble être perturbé par la production de piARNs de *ZAM* dans les cellules germinales. Durant ma thèse, j'ai développé des lignées de drosophiles où *ZAM* est fortement dérégulé et envahi massivement l'ovocyte. J'ai donc utilisé ces modèles afin d'étudier plus en détail le cycle rétroviral de *ZAM*.

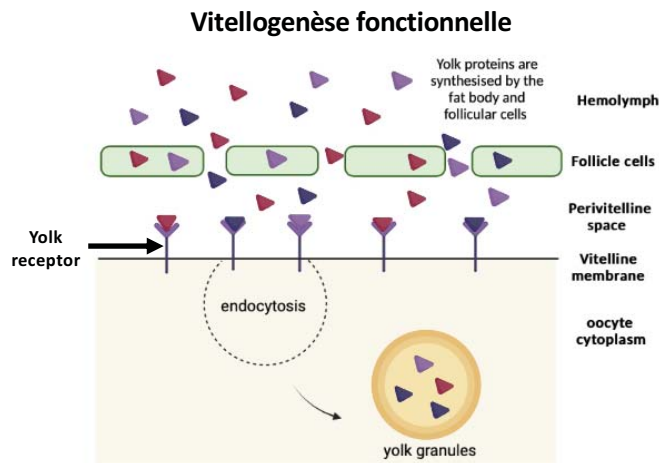
### I- Rôle du trafic de vitellogénine dans le transit de *ZAM*

J'ai dans un premier temps analysé la présence de pseudo-particules virales dans les ovaires par microscopie électronique. Pour réaliser cette expérience, nous avons fait appel au Centre d'Imagerie Cellulaire et Santé (CISC) qui a réalisé des coupes d'ovaires, permettant une caractérisation ultrastructurale du tissu reproducteur. Les ovaires de la lignée *vret*-sKD (RNAi dirigé contre Vreteno dans les cellules folliculaires) ont été analysés dans un premier temps puisqu'une très forte accumulation des ARNs de *ZAM* dans les cellules folliculaires, ainsi qu'une invasion de l'ovocyte, avaient été observées dans cette lignée. Un contrôle *white*-sKD, où aucune expression de *ZAM* n'est détectée par smFISH, a également été observé. J'ai recherché la présence de pseudo-particules de *ZAM* dans les stades 9-10 de l'ovaire, stades où *ZAM* est fortement exprimé dans les cellules folliculaires et envahit l'ovocyte. J'ai observé en condition *vret*-sKD la présence de particules sphériques d'environ 0,045  $\mu$ m (45 nm) dans les cellules folliculaires. Ces particules semblent être accumulées à proximité de la membrane vitelline qui sépare les cellules folliculaires de l'ovocyte (**Figure 1A**). Comme observé dans les études préalablement réalisées sur la lignée RevIH2, j'ai également détecté la présence de

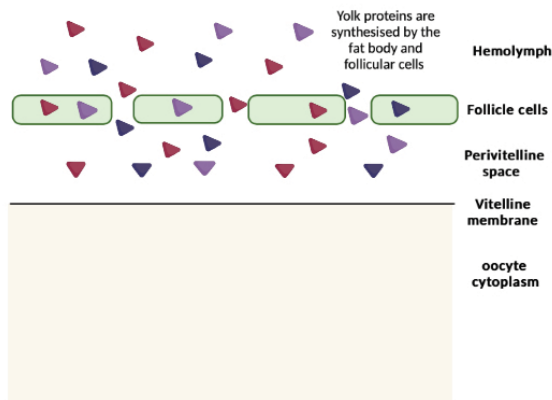


particules dans les granules Yolk localisées dans le cytoplasme de l'ovocyte (**Figure 1B**). Je n'ai pas détecté de particule bourgeonnante ou bien de particules libres dans le cytoplasme de l'ovocyte. Aucune accumulation de particule de la même forme et de la même taille ne semble présente dans les cellules folliculaires ou dans les granules Yolk en condition contrôle (*white-sKD*). Cependant, la détection de ces particules est particulièrement difficile car beaucoup d'éléments sphériques sont présents dans les ovaires. De plus, altérer l'expression de la protéine Vreteno affecte la voie des piARNs dans son ensemble et induit donc la dérégulation de multiples autres ERVs qui pourraient, eux aussi, former des pseudo-particules virales. Ainsi, afin de confirmer que les particules détectées sont des particules formées par le rétrovirus endogène *ZAM*, nous souhaitons coupler l'observation ultrastructurale avec un immunomarquage de la protéine Gag codée par *ZAM*, protéine formant la capsid de la pseudo-particule. Malheureusement, après plusieurs tentatives de mise au point du protocole, la plateforme de microscopie n'a pas réussi à développer cette technique de fixation permettant de réaliser l'immunomarquage et l'analyse ultrastructurale sur ovaire en même temps. Nous avons donc décidé d'utiliser un autre modèle : la lignée RevIH2 où la voie des piARNs est abolie dans les cellules germinales (RevIH2 *ago3-GLKD*, « Germ-Line Knock-Down »). Dans ces drosophiles, *ZAM* est dérégulé dans les cellules folliculaires et, en absence de piARNs produits par les cellules germinales, nous avons observé que *ZAM* envahit massivement l'ovocyte. Dans cette condition, la voie des piARNs dans les cellules folliculaires est fonctionnelle, la dérégulation de *ZAM* est due à la délétion de la partie proximale du cluster de piARNs *flamenco*. De fait, la dérégulation de la plupart des autres rétrovirus endogènes présents dans le génome ne devrait pas être affectée. En condition RevIH2 *ago3-GLKD*, des observations préliminaires ont permis de détecter la présence de particules dans les cellules folliculaires ainsi que dans les granules Yolk, comme observée précédemment dans la lignée *vret-sKD*. Cependant, je n'ai pas observé de différence significative avec une condition RevIH2 *white-GLKD* où les piARNs de *ZAM* produits dans les cellules germinales bloquent en principe le transit de *ZAM*. Il serait donc nécessaire de mettre au point la technique de microscopie électronique couplée à l'immunofluorescence afin de confirmer les résultats obtenus et conclure sur la présence de particules de *ZAM* dans les différentes conditions et dans les différents tissus. De plus, il serait intéressant de réaliser ces mêmes expériences en utilisant les lignées CRISPR-*flamΔZAM* obtenues récemment, possédant des copies actives de *ZAM* afin de suivre la formation de particules de *ZAM* dans ce contexte.





**Mutation du récepteur Yolkless permettant l'endocytose des protéines Yolk par l'ovocyte**



**Figure 2 : La mutation du récepteur Yolkless affecte la transmission des ARNs de ZAM à l'ovocyte.**

Images confocales d'ovarioles de drosophiles *vret-sKD* lorsque le trafic de vitellogénine est fonctionnel (panel du haut) ou en condition mutante pour le récepteur *yolkless* (*yl17*) qui est nécessaire pour l'endocytose des protéines Yolk par l'ovocyte (panel du bas). Beaucoup moins d'ARNs de *ZAM* sont retrouvés dans l'ovocyte lorsque le trafic de vitellogénine est bloqué.

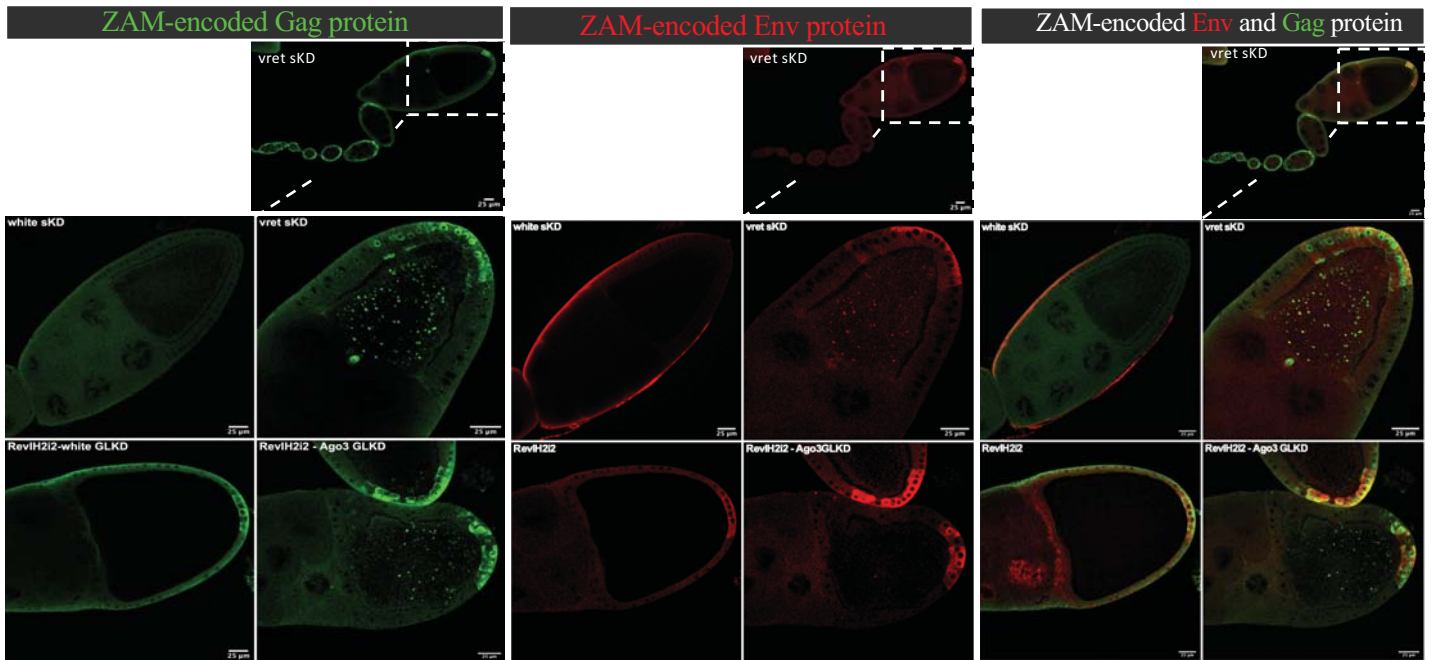
Malgré le caractère très préliminaire de ces observations, les résultats obtenus semblent montrer la présence de particules de *ZAM* dans les granules *Yolk* dans l'ovocyte. Cela confirmerait que *ZAM* utilise le trafic de vitellogénine pour transiter des cellules folliculaires jusqu'à l'ovocyte. Afin de confirmer cette hypothèse, j'ai généré des mouches *vret-sKD* où le récepteur permettant l'endocytose de la vitellogénine par l'ovocyte est muté (mutation du gène *yolkless yl<sup>17</sup>*). Cette mutation induit une stérilité des femelles, car les œufs pondus présentent un déficit important en vitellus et ne se développent pas en raison de la fragilité du chorion. J'ai analysé l'impact de cette mutation sur la capacité d'invasion de *ZAM* en réalisant un smFISH permettant de détecter l'ARNm de *ZAM*. Dans les mutants *yl<sup>17</sup>/yl<sup>17</sup> homozygotes*, en absence du récepteur de la vitellogénine, il n'y a pratiquement plus d'ARNm de *ZAM* dans l'ovocyte en contexte *vret-sKD* et les ARNs de *ZAM* s'accumulent fortement dans les cellules folliculaires (**Figure 2A**). Néanmoins, des ARNm de *ZAM* sont parfois encore détectés dans l'ovocyte, surtout dans les stades tardifs de l'ovogenèse (stade 10b). Ces résultats sont préliminaires, et d'autres expériences sont nécessaires pour confirmer le rôle du trafic de vitellogénine pour le transit de *ZAM* entre les cellules folliculaires et l'ovocyte dans des conditions où la voie des piARNs est perdue dans les cellules folliculaires. Il serait notamment intéressant d'analyser par microscopie électronique la présence et la localisation des pseudo-particules de *ZAM* dans cette condition mutante pour le récepteur *Yokless*. De plus, il n'est pas exclu qu'un autre processus existe pour permettre le transit de l'ARNm de *ZAM* des cellules folliculaires vers l'ovocyte.

## II- Rôle de la protéine Enveloppe dans le transit de *ZAM*

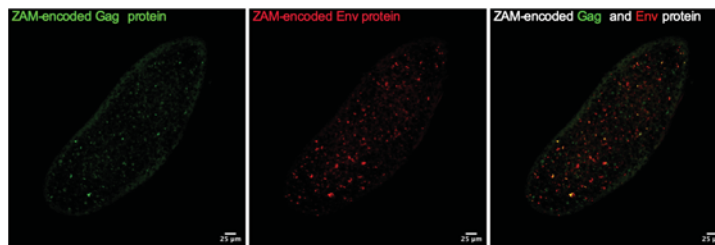
*ZAM* est un rétrovirus endogène qui code pour une protéine Enveloppe. La protéine Env chez les rétrovirus exogènes est nécessaire pour l'infection d'une cellule cible. Cette protéine permet la fixation du virus à son récepteur et son entrée dans la cellule hôte. Nous souhaitons étudier le rôle de l'enveloppe pour l'invasion des cellules germinales par le rétrovirus endogène *ZAM*.

J'ai réalisé un immunomarquage des protéines Env et Gag, protéines nécessaires à la formation de pseudo-particules enveloppées, dans des conditions où *ZAM* est fortement exprimé et envahi l'ovocyte : *vret-sKD* ou *RevI-H2 ago3-GLKD*. Dans ces deux conditions, la protéine Gag est retrouvée dans les cellules somatiques folliculaires de l'ovaire. Son expression commence précocement dans l'ovogenèse, dès la formation des cellules somatiques du germarium. Gag est ensuite retrouvée dans toutes les cellules folliculaires entourant les cellules germinales dans

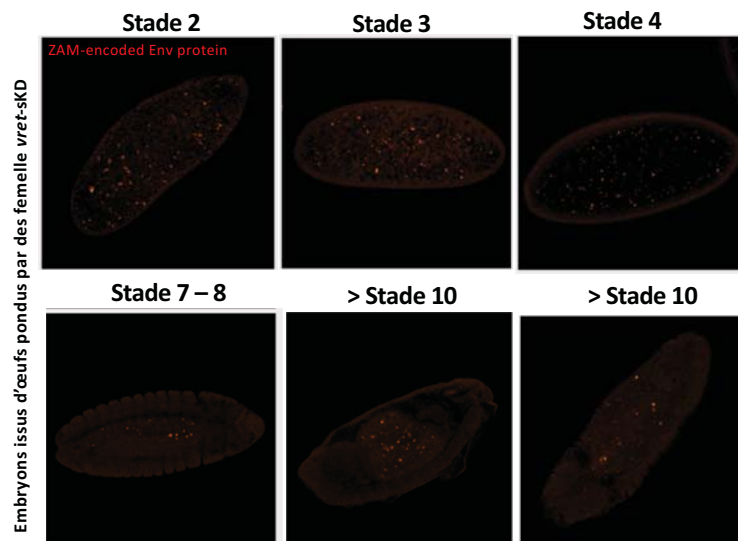
A



B

Embryon précoce issu d'œufs pondus par des femelle *vret-sKD* (< 2 h AEL)

C



### Figure 3 : Expression des protéines Gag et Env codées par ZAM dans les ovaires et embryons précoces en condition de dérégulation pour ZAM

A) Immunomarquage des protéines Gag (en vert) et Env (en rouge) codées par ZAM dans les ovarioles entières et les follicules de stade 10 de drosophiles *white-sKD* (lignée contrôle), *vret-sKD*, *RevIH2i2 white-GLKD* et *RevIH2i2 Ago3-GLKD*. Gag et Env sont retrouvées dans les cellules folliculaires dans les conditions où ZAM est dérégulé (*vret-sKD*, *RevIH2i2 white-GLKD* et *RevIH2i2 Ago3-GLKD*) et dans le cytoplasme de l'ovocyte en condition *vret-sKD* et *RevIH2i2 Ago3-GLKD*. B) Immunomarcages des protéines Gag (en vert) et Env (en rouge) codées par ZAM réalisés sur des embryons précoces, âgés de moins de 2 h (< 2 h *After Eggs Laying* (AEL)), pondus par des femelles *vret-sKD*. C) Immunomarcages de la protéine Env codée par ZAM réalisés sur des embryons pondus par des femelles *vret-sKD*. Le stade embryonnaire approximatif est indiqué pour chaque embryon.

les follicules de stades précoces, et se localise ensuite au niveau des cellules folliculaires situées au pôle postérieur des follicules plus tardifs (stade 9-10 de l'ovogenèse) (**Figure 3A – panel de gauche**). L'expression de la protéine Env semble plus localisée, un marquage est seulement détecté dans les cellules folliculaires des stades tardifs (à partir du stade 9-10 de l'ovogenèse), au niveau postérieur, dans les mêmes cellules où l'expression de Gag est détectée à ce stade (**Figure 3A – panel central**). Les expériences de smFISH réalisées précédemment ont révélé que l'ARNm de *ZAM* est présent dans le cytoplasme de l'ovocyte uniquement à partir du stade 9-10 de l'ovogenèse, suggérant que le transit de *ZAM* s'effectue uniquement à partir de ce stade. Il est donc intéressant de noter que c'est également à partir de ce stade que la protéine Env est détectée. Cependant, l'anticorps anti-Env étant assez ancien, et surtout très aspécifique, ces résultats restent à confirmer.

De manière surprenante, les protéines Gag et Env sont également détectées dans le cytoplasme de l'ovocyte, à partir des stades 9-10 de l'ovogenèse dans les conditions où *ZAM* envahit l'ovocyte (*vret-sKD* et *RevIH2-i2 Ago3-GLKD*). Gag et Env semblent co-localiser dans le cytoplasme de l'ovocyte (**Figure 3A-panel de gauche**). Ces deux protéines sont ensuite transmises aux embryons précoces pondus par les femelles *vret-sKD* (**Figure 3B**) et la protéine Env est détectée jusqu'à des stades tardifs de différenciation (après le stade 10 de l'embryogenèse) (**Figure 3C**).

Les connaissances acquises grâce à l'étude du cycle d'infection des rétrovirus exogènes montrent que l'enveloppe permet la fixation du virus à sa cellule hôte, ce qui permet sa pénétration dans le cytoplasme, la destruction de la capsid virale et le relargage du contenu génétique. Ainsi, si *ZAM* infecte les cellules germinales selon le même processus que celui utilisé par les rétrovirus exogènes, nous ne nous attendions pas à retrouver les protéines Gag et Env dans le cytoplasme de l'ovocyte. Cependant, l'ARN de *ZAM* détecté dans le cytoplasme de l'ovocyte et dans les embryons peut potentiellement être traduit et ainsi permettre la formation de nouvelles particules virales. Malheureusement, effectuer un double marquage qui permettrait de détecter l'ARN de *ZAM* et la protéine Gag ou Env de *ZAM* est impossible pour le moment car les anticorps ciblant ces deux protéines ne fonctionnent pas avec le protocole de smFISH. Il est également possible que *ZAM* forme des pseudo-particules enveloppées mais que celles-ci ne transitent pas par un mécanisme d'infection actif - *via* la fixation à un récepteur - mais pourrait transiter grâce à un mécanisme plus passif, et notamment grâce au trafic de vitellogénine. Ainsi, le rôle de l'enveloppe dans le transit de *ZAM* reste à ce jour encore énigmatique.



# DISCUSSION



## **1- Les piARNs produits dans les cellules germinales peuvent contrer une invasion d'un rétrovirus endogène provenant des cellules somatiques adjacentes.**

Nous avons démontré que les piARNs produits par les cellules germinales de la lignée RevI-H2i2 permettent de contrer une invasion de *ZAM* provenant des cellules somatiques adjacentes. En effet, altérer la production de protéines impliquées dans la production des piARNs, telles que *Armi*, *Zuc* ou *Vret*, dans la lignée germinale induit une forte augmentation de l'invasion de *ZAM* dans l'ovocyte de la lignée RevI-H2i2. Ces résultats démontrent donc que la production de piARNs dans les cellules germinales est requise pour contrer l'invasion de *ZAM* s'il n'y a pas déjà une protection en amont, dans les cellules folliculaires.

Affecter la production de la protéine *Piwi*, acteur principal de la régulation transcriptionnelle des ETs, n'induit pas d'augmentation de l'invasion de *ZAM*. Cela suggère que les piARNs de *ZAM* ne sont pas impliqués dans la régulation transcriptionnelle de cet ET. En revanche, la protéine *Ago3*, impliquée dans la régulation post-transcriptionnelle des ETs, est nécessaire pour contrer l'invasion et dégrader les ARNs de *ZAM* dans les cellules germinales. Ces résultats révèlent que *ZAM* est contrôlé au niveau post-transcriptionnel dans les cellules germinales de la lignée RevI-H2i2.

Il est également important de souligner qu'altérer la production de la protéine *Aubergine*, facteur essentiel pour la répression post-transcriptionnelle des ETs dans les cellules germinales, affecte fortement la production des piARNs de *ZAM* et induit une perte de la signature ping-pong de ces piARNs. De manière surprenante, nous n'avons pas observé d'augmentation de l'invasion de *ZAM* dans les cellules germinales de la lignée RevI-H2i2 en condition *Aub-GLKD* (*Germinal Knock-Down*). Pour analyser ce résultat, il est important de constater que la déplétion de la protéine *Aub* dans les cellules germinales conduit à une atrophie sévère des ovaires, et que quasiment aucun follicule de stade 10 n'est formé dans cette condition. *Aub* s'accumule habituellement au pôle postérieur de l'ovocyte, structure appelée le « *germ plasm* », dans laquelle nous avons également constaté une forte accumulation des ARNs de *ZAM* lorsque cet ET transite vers l'ovocyte. Il a été montré que l'accumulation d'*Aub* au pôle postérieur de l'ovocyte permet le recrutement de protéines telles que *Wispy*, une protéine impliquée dans la polyadénylation des ARNm, plus particulièrement dans celle d'ARNm codant des « déterminants des cellules germinales », molécules nécessaires à la formation des cellules précurseurs de la lignée germinale (Dufourt et al., 2017). *Aub* est également impliquée dans la régulation de la traduction d'ARNm codant des protéines telles que *Oskar* et *Gurken*, deux morphogènes impliqués dans la spécialisation des cellules folliculaires et la polarisation de



l'ovocyte (Ephrussi and Lehmann, 1992; González-Reyes et al., 1995; Wilson et al., 1996; Nilson and Schüpbach, 1999; Tian et al., 2013; Ma et al., 2017). Ainsi, la déplétion d'*Aub* dans les cellules germinales conduit à des défauts morphologiques sévères qui pourraient potentiellement rendre le passage de *ZAM* vers l'ovocyte impossible. Les observations faites sur le transit de *ZAM* en condition *Aub*-GLKD reflètent probablement un mécanisme indépendant de la voie des piARNs. Globalement, les résultats obtenus révèlent un rôle important des piARNs produits dans les cellules germinales pour contrer l'invasion d'un ERV provenant des cellules somatiques adjacentes.

## **2- Comment les piARNs de *ZAM* sont-ils produits dans les cellules germinales de la lignée RevI-H2 ?**

Nous avons identifié quatre insertions de *ZAM* dans des clusters « dual-strand » localisés sur le chromosome X de la lignée RevI-H2i2 : trois insertions dans le cluster 9 et une insertion dans le cluster 13 - ou 56, nous n'avons pas réussi à la localiser clairement - (clusters de piARNs identifiés dans (George et al., 2015)). Nous suggérons que ces insertions dans des clusters « dual-strand » de piARNs pourraient permettre la production de piARNs de *ZAM* dans les cellules germinales de la lignée RevI-H2i2. Cependant, nous n'avons pas réussi à trouver une signature particulière sur les piARNs de *ZAM* qui permettrait de démontrer que ces derniers sont produits spécifiquement par les insertions de *ZAM* dans ces clusters de piARNs (c'est-à-dire des piARNs chimériques avec une partie de la séquence de *ZAM* et une partie de la séquence génomique du site d'insertion). De plus, il n'est pas exclu que les clusters de piARNs ne produisent pas, à eux seuls, la totalité des piARNs sens et antisens engagés dans le cycle ping-pong.

Un autre mécanisme a été proposé pour expliquer la production *de novo* de piARNs. En plus d'être produits à partir d'une insertion dans un cluster de piARNs, les piARNs peuvent également être produits par des insertions euchromatiques d'ETs converties en loci producteurs de piARNs, un phénomène appelé la paramutation (De Vanssay et al., 2012a; Mohn et al., 2014; Shpiz et al., 2014; Gebert et al., 2021). Plusieurs copies euchromatiques de *ZAM* ont été localisées sur tous les chromosomes de la lignée RevI-H2i2 (sauf sur le chromosome 4). Il a été montré que la paramutation de copies euchromatiques d'ETs requière la présence de piARNs complémentaires, déposés maternellement, qui provoquent la conversion d'insertions d'ETs en régions productrices de piARNs (De Vanssay et al., 2012b; Le Thomas et al., 2014; Mohn et al., 2014; Hermant et al., 2015). La source des piARNs qui permettent d'initier la paramutation reste inconnue, mais une possibilité serait que leur production soit déclenchée par une insertion

de l'ET envahisseur dans un cluster de piARNs. Ainsi, l'insertion de *ZAM* dans un cluster pourrait être l'évènement initial déclenchant la production de quelques piARNs initiaux, qui entraîneraient ensuite une réaction en chaîne où certaines insertions euchromatiques d'ETs pourraient être converties en loci producteurs de piARNs (Kofler et al., 2022).

Cependant, la paramutation nécessite que le locus soit transcrit, l'ARN naissant devant être reconnu par des piARNs afin d'induire le dépôt de marques H3K9me3 à l'origine du recrutement de la protéine Rhino. Or, l'expression de *ZAM* par son propre promoteur est impossible dans les cellules germinales car la présence du facteur de transcription Pointed2, uniquement exprimé dans les cellules somatiques folliculaires, est nécessaire (Meignin et al., 2004). L'expression des copies euchromatiques pourrait néanmoins être permise grâce à l'utilisation de promoteurs adjacents présents aux sites d'insertion. Nous avons notamment découvert que certaines insertions de *ZAM* présentes sur le chromosome X se trouvent dans ou proche de séquences d'ETs. Par exemple, une copie de *ZAM*, localisée sur le chromosome X de la lignée RevI-H2i2, est insérée à 820 pb en aval d'une copie récente du facteur *I*, ET connu pour être exprimé dans les cellules germinales. Une autre copie euchromatique de *ZAM* sur le chromosome X de la lignée RevI-H2i2, est insérée dans un ET *Jockey*. Dans les 2 cas, l'expression de *ZAM* pourrait alors être possible grâce à l'utilisation du promoteur du *facteur I*, actif dans les cellules germinales, ou de *Jockey* respectivement. Nous avons également montré que le chromosome X de la lignée RevI-H2, où les insertions dans les clusters de piARNs ont été détectées, est suffisant pour produire des piARNs sens et antisens avec une signature ping-pong. La présence du chromosome X est également suffisante pour induire la répression d'un transgène senseur contenant une séquence de *ZAM*. Les chromosomes II et III ne sont, eux, ni suffisants pour permettre la production de piARNs de *ZAM* avec une signature ping-pong, ni pour réprimer le transgène senseur, alors que ces deux chromosomes possèdent également des insertions euchromatiques de *ZAM*. Ces résultats montrent que les piARNs régulateurs sont produits par le chromosome X de la lignée RevI-H2i2, suggérant que les insertions dans les clusters de piARNs jouent un rôle majeur pour la production des piARNs de *ZAM*. Toutefois, la contribution des trois insertions euchromatiques, également présentes sur ce chromosome X, reste également possible.

Enfin, les transcrits sens engagés dans le cycle ping-pong pourraient provenir des ARNs de *ZAM*, issus des cellules folliculaires, qui ont envahi l'ovocyte. Les ARNs de *ZAM* seraient alors ciblés par les piARNs antisens produits dans les cellules germinales, puis maturés en piARNs sens afin d'alimenter le cycle ping-pong. Pour analyser le rôle des ARNs de *ZAM* provenant des cellules folliculaires dans la production des piARNs par les cellules germinales, nous

pourrions bloquer le passage de *ZAM* et analyser l'impact sur la production germinale des piARNs. Nous avons montré que, dans une condition *vret-sKD* où le gène codant le récepteur des protéines *Yolk* est muté, le transit de *ZAM* des cellules folliculaires vers l'ovocyte est fortement diminué. L'idéal serait donc de tester l'impact de cette mutation sur la localisation des ARNs de *ZAM* dans une condition *RevI-H2i2*, et analyser si la production de piARNs de *ZAM* par les cellules germinales est affectée.

Pour conclure, bien que les ETs ne soient pas forcément piégés de façon active dans les clusters de piARNs, nos données laissent à penser que l'insertion d'une copie dans un cluster de piARNs pourrait être une étape nécessaire pour réprimer efficacement l'invasion de la lignée germinale par un ET.

### **3- La lignée *RevI-H2* est une lignée stable, bien que *ZAM* soit toujours exprimé dans les cellules folliculaires.**

Nos résultats suggèrent que les piARNs produits par les cellules germinales de la lignée *RevI-H2i2* empêchent la transposition de *ZAM* dans le génome des cellules germinales. Cette hypothèse est confortée par le fait que le nombre de copies de *ZAM* présentes dans le génome de la lignée *RevI-H2* s'est stabilisé au moins au cours des dix dernières années. Cependant, *ZAM* est toujours exprimé dans les cellules folliculaires des ovaires de la lignée *RevI-H2i2*, démontrant que des copies fonctionnelles de *ZAM* sont présentes dans le génome. Même si *ZAM* ne transpose plus dans les cellules germinales, une transposition dans les cellules folliculaires pourrait être dangereuse. Ce phénomène pourrait en effet induire la mutation de gènes requis pour le bon déroulement de l'ovogenèse, et ainsi affecter indirectement le développement des cellules germinales.

Néanmoins, aucune atrophie des ovaires ou subfertilité n'a été relevée dans la lignée *RevI-H2i2*. Cette observation suggère que *ZAM* serait incapable de transposer dans le génome des cellules folliculaires, ou qu'un mécanisme réprimerait l'intégration de *ZAM* dans le génome des cellules folliculaires dans la lignée *RevI-H2i2*. De façon surprenante, l'expression de *ZAM* dans les cellules folliculaires de la lignée *RevI-H2i2* semble être plus faible et plus restreinte par rapport à l'expression de *ZAM* observée dans une condition *vret-sKD* ou *flamΔZAM*, suggérant qu'un mécanisme de répression s'exerce dans les cellules folliculaires de la lignée *RevI-H2i2*. Même si *ZAM* est absent de son cluster régulateur *flamenco* dans la lignée *RevI-H2i2*, il n'est pas exclu qu'un autre locus génomique induise la production de piARNs régulateurs de *ZAM* dans les cellules folliculaires.

Le cluster « unistrand » *20A* est un bon candidat ; ce cluster est présent sur le chromosome X de *Drosophila melanogaster* et a été décrit comme étant un cluster exprimé dans les cellules germinales et folliculaires (Brennecke et al., 2007; Klattenhoff et al., 2009; Mohn et al., 2014). En effet, des piARNs provenant du cluster *20A* ont été retrouvés dans les cellules OSS (lignée cellulaire dérivée des cellules souches folliculaires). Néanmoins, nous n'avons pas identifié d'insertion de *ZAM* dans le cluster *20A* dans le génome de la lignée RevI-H2i2. Étonnement, je n'ai visualisé aucune expression du cluster *20A* par smFISH dans les cellules folliculaires alors qu'une expression est observée dans les cellules germinales. Le cluster *20A* pourrait donc ne pas être exprimé dans les cellules somatiques folliculaires *in vivo*, ou bien être exprimé uniquement dans quelques cellules folliculaires. Finalement, nous n'avons identifié aucune insertion de *ZAM* dans des clusters de piARNs connus pour être exprimés dans les cellules folliculaires. Afin d'étudier plus spécifiquement la production de piARNs de *ZAM* par les cellules folliculaires, il serait nécessaire de trier spécifiquement ces cellules, grâce à une technique de FACS (*Fluorescence-activated cell sorting*) par exemple, et de séquencer les petits ARNs produits dans ces cellules.

Par ailleurs, j'ai constaté une expression plus forte de *ZAM* dans les cellules folliculaires de la lignée RevI-H2i2 lorsque la voie des piARNs est altérée uniquement dans les cellules germinales (RevI-H2i2 *Ago3*-GLKD versus RevI-H2i2 *white*-GLKD par exemple). Dans ce cas, le contexte génétique est similaire entre les deux lignées, la différence majeure réside dans la production ou non de piARNs de *ZAM* dans les cellules germinales. Ce résultat suggère que les piARNs produits dans les cellules germinales de la lignée RevI-H2i2 agiraient sur la régulation de *ZAM* dans les cellules folliculaires. Le transfert intercellulaire de piARNs pourrait s'effectuer grâce aux nombreuses communications existant entre les cellules germinales et les cellules somatiques de l'ovaire. Certaines études proposent notamment un rôle des exosomes et des jonctions gap dans le transfert intercellulaire de petits ARNs tels que les miARNs, les siARNs et les ARNt (Valadi et al., 2007; Dinger et al., 2008; Brink et al., 2012; Raposo and Stoorvogel, 2013; Tkach and Théry, 2016; Zong et al., 2016). Purifier les exosomes formés dans les ovaires de la lignée RevI-H2 et séquencer les petits ARNs présents dans ces vésicules permettraient d'analyser le rôle des exosomes dans le transit de piARNs. Nous pourrions également analyser l'impact d'un blocage des jonctions gap sur l'expression de *ZAM* dans les cellules folliculaires. La régulation de *ZAM* dans les cellules folliculaires de la lignée RevI-H2 reste donc une question ouverte.

#### **4- La délétion de ZAM de son cluster de piARNs *flamenco* engendre une subfertilité des femelles.**

Nous avons montré que la délétion de *ZAM* du cluster de piARNs *flamenco* (lignée *flam* $\Delta$ *ZAM* générée grâce à la technique CRISPR-Cas9) conduit à une forte dérépression de *ZAM* dans les cellules folliculaires et à une invasion de l'ovocyte. Les femelles *flam* $\Delta$ *ZAM* présentent des ovaires atrophiés et sont quasiment stériles, la ponte et l'éclosion des œufs sont affectées. Ce phénotype ressemble à celui observé chez les femelles *flam*<sup>KG</sup> ou *flam*<sup>BG</sup> pour lesquelles l'expression du cluster *flamenco* est totalement abolie. Les femelles *flam*<sup>KG</sup> et *flam*<sup>BG</sup> sont totalement stériles et possèdent des ovarioles rudimentaires : l'épithélium folliculaire est complètement désorganisé, et la différenciation et la division des cellules germinales sont également affectées (Robert et al., 2001; Mével-Ninio et al., 2007). Les rares œufs matures présents dans les ovaires des femelles montrent des défauts morphologiques, notamment au niveau de la formation de l'appendice dorsal, et ils ne sont jamais pondus. Une analyse plus détaillée des phénotypes induits par la mutation *flam* $\Delta$ *ZAM* serait nécessaire afin de comparer ces conditions et *flam* $\Delta$ *ZAM*. Une différence majeure entre ces deux modèles repose sur la production de piARNs. La délétion *flam* $\Delta$ *ZAM* induit uniquement une perte de production de piARNs au niveau des régions internes de *ZAM*, la production de piARNs par les régions localisées en amont et en aval de la délétion n'étant pas affectée. Cela suggère que la régulation des autres ETs n'est pas impactée dans la lignée *flam* $\Delta$ *ZAM*, contrairement aux lignées *flam*<sup>KG</sup>/*flam*<sup>BG</sup>. Ainsi, la dérépression d'un seul élément transposable dans les cellules somatiques folliculaires de l'ovaire affecte directement la fertilité. Ces résultats montrent pour la première fois une corrélation directe entre l'expression des éléments transposables somatiques et la fertilité, et suggèrent que les phénotypes observés chez les femelles *flam*<sup>KG</sup>/*flam*<sup>BG</sup> pourraient aussi être directement causés par l'expression d'ETs dans les cellules somatiques folliculaires.

Des tests de stérilité réalisés sur les femelles *vret*-sKD et *yb*-sKD ont montré que la fertilité des femelles est également affectée dans ces conditions. Le nombre d'œufs pondus par les femelles *vret*-sKD et *yb*-sKD est similaire à celui des femelles contrôle *white*-sKD. En revanche, la moitié des œufs pondus n'éclorent pas, même si, dans ces deux conditions, les ovaires ne sont pas atrophiés. Nous avons également observé que *ZAM* est fortement déréprimé dans les cellules folliculaires de l'ovaire en condition *vret*-sKD ou *yb*-sKD. Par conséquent, si les phénotypes observés chez les femelles *flam* $\Delta$ *ZAM* étaient réellement causés par une expression de *ZAM* fonctionnels dans les cellules folliculaires, nous nous attendions à observer les mêmes phénotypes entre une condition *vret*-sKD ou *yb*-sKD et une condition *flam* $\Delta$ *ZAM*. Or, ce n'est

pas le cas. Plusieurs hypothèses peuvent être émises pour expliquer les différences phénotypiques observées :

1 - Nous avons seulement réalisé un « *knock-down* » des protéines Yb et Vret dans les cellules folliculaires, la production de piARNs n'est donc pas totalement abolie dans ces conditions. Les piARNs toujours produits pourraient être suffisants pour protéger les cellules folliculaires d'une transposition massive de *ZAM*.

2 - Les défauts observés pourraient être causés par une expression de *ZAM* dans des tissus autres que les cellules somatiques folliculaires.

En effet, l'expression des protéines Vret et Yb a été altérée uniquement et spécifiquement dans les cellules folliculaires des ovaires grâce au sKD. Ainsi, les effets observés sont uniquement dus à un phénomène se produisant dans les cellules folliculaires, ou dans les cellules germinales en cas d'invasion par des rétrotransposons comme *ZAM*. La mutation induite dans la lignée *flamΔZAM* affecte quant à elle toutes les cellules de l'organisme. Certaines études soulignent un rôle potentiel de la voie des piARNs dans les tissus somatiques autres que l'ovaire et, chez la drosophile, les protéines PIWI ont été détectées dans des tissus autres que les gonades, comme dans les glandes salivaires, le corps gras ou encore le cerveau (Brower-Toland et al., 2007; Perrat et al., 2013; Jones et al., 2016). Certains travaux ont également révélés la production de piARNs dans la tête des drosophiles, organe où une expression du cluster de piARNs *flamenco* a été détecté (Yan et al., 2011; Mirkovic-Hösle and Förstemann, 2014; Jones et al., 2016; Giordani et al., 2022). Enfin, il a été montré que *ZAM*, et d'autres ETs tels que *gypsy* et *Idefix*, seraient exprimés dans la tête de drosophiles lorsque *flamenco* n'est plus exprimé (mutants *flam<sup>kc</sup>/flam<sup>bc</sup>*) (Giordani et al., 2022). Ainsi, nous pourrions imaginer que les phénotypes obtenus pourraient être dus à une expression de *ZAM* dans d'autres tissus chez les femelles *flamΔZAM*, notamment au niveau du cerveau.

L'expression d'ERVs dans le cerveau a déjà été observée chez des patients humains atteints de troubles neurologiques tels que la maladie d'Alzheimer, la sclérose en plaque ou la maladie de Charcot (sclérose latérale amyotrophique) (Décrit dans la revue (Tam et al., 2019a)). La mobilisation d'ERVs n'a pour le moment pas été détectée dans les tissus de patients, mais l'utilisation d'organismes modèles, tels que la drosophile, a permis de démontrer la pathogénicité causée par l'expression des ERVs. Les acides nucléiques des ERVs ou les protéines rétrovirales peuvent être reconnus par les systèmes de détection de l'immunité innée et induire une inflammation locale ou chronique. L'expression du rétrotransposon *gypsy* a

notamment été observée dans les cellules gliales du cerveau dans un modèle drosophile de sclérose latérale amyotrophique (Krug et al., 2017). Les facteurs de transcription requis pour l'expression de *gypsy* pourraient alors être exprimés dans le cerveau de drosophile où un promoteur adjacent au site d'insertion pourrait être utilisé pour l'expression de *Gypsy*. L'expression de *gypsy* dans le cerveau active les voies de dommages à ADN, induisant l'apoptose des cellules gliales et des neurones, en plus d'une inflammation chronique. Il est intéressant de noter que le déclenchement de cette inflammation conduit chez la drosophile à une altération générale des organes, incluant les ovaires, les muscles, le tissu adipeux et les reins (Revue dans : (Lu and Vogel, 2009)). Ce syndrome, appelé « *organ wasting* », entraîne un gonflement abdominal dû à l'accumulation de fluides ; le corps gras et les ovaires de ces mouches sont atrophiés, les femelles sont stériles et elles présentent une durée de vie réduite. Dans notre modèle *flamΔZAM*, j'ai pu constater plusieurs de ces caractéristiques : les femelles présentent des ovaires atrophiés et un abdomen gonflé, elles sont quasiment stériles et leur durée de vie semble être diminuée. Il serait donc intéressant d'analyser si elles présentent également d'autres caractéristiques spécifiques du syndrome « *d'organ wasting* ».

Ce phénotype peut éventuellement être causé par l'expression d'un ET dans le cerveau. Nous pourrions donc étudier si la mutation *flamΔZAM* induit une expression de *ZAM* dans le cerveau. Néanmoins, plusieurs études ont montré que l'expression de rétrovirus endogènes dans les tissus somatiques de manière générale pouvait aussi déclencher des mécanismes d'inflammation (Hurst and Magiorkinis, 2015; Mu et al., 2016; Tam et al., 2019a; Jönsson et al., 2021). Les infections bactériennes intestinales sont également bien connues pour déclencher de telles inflammations chroniques, *via* l'activation de la voie NFκB, et peuvent engendrer une altération générale des organes (Zugasti et al., 2020). Une expression de *ZAM* dans le cerveau n'est donc pas la seule explication possible aux phénotypes observés. Il serait de ce fait intéressant d'analyser l'expression de *ZAM* dans différents tissus autres que les ovaires. Nous avons également prévu d'analyser l'activation de la réponse immunitaire innée dans la lignée *flamΔZAM* afin de voir si une réactivation de *ZAM* conduit au déclenchement d'une réaction inflammatoire. Finalement, la lignée *flamΔZAM* est un modèle unique pour étudier l'impact de la réactivation d'un rétrovirus endogène à l'échelle moléculaire, cellulaire, tissulaire et sur l'organisme entier.

3) Une autre explication concerne l'efficacité d'expression et de transposition des différentes copies de *ZAM*, ainsi que l'influence du nombre de copies génomiques. Le séquençage du génome de référence de *Drosophila melanogaster* (lignée Iso1A) a révélé que seulement deux

copies entières de *ZAM* sont présentes dans le génome de cette lignée. Le génome des lignées *vret-sKD* et *yb-sKD* n'ont pas été séquencés, nous ne connaissons donc pas encore le nombre de copies de *ZAM* insérées dans le génome de ces lignées, ni même la nature de ces copies. Les expériences de smFISH ont révélé une quantité importante d'ARNs de *ZAM* dans ces deux conditions, aussi bien dans les cellules folliculaires que dans les cellules germinales, montrant la présence d'au moins une copie transcriptionnellement active dans le génome de ces deux lignées. Cependant, il est important de dissocier l'activité transcriptionnelle de la transposition : une copie d'ET peut être transcrite afin de former un ARN, détecté par smFISH ou par RT-qPCR, mais cet ARN peut ne jamais être rétrotranscrit ou ne jamais aboutir à une nouvelle intégration génomique de l'ET.

Le génome de la lignée *flamΔZAM* ne contenait initialement aucune copie transcriptionnellement active de *ZAM*. J'ai donc apporté par croisement avec la lignée RevI-H2i2 des copies de *ZAM* localisées sur différents chromosomes. Nous avons constaté que différents variants de *ZAM* sont présents dans le génome de la lignée RevI-H2i2, et plusieurs copies de chaque variant sont retrouvées dans le génome, suggérant que ces variants sont fonctionnels pour la transposition. Les tests de stérilité ont été réalisés sur la lignée *flamΔZAM* où deux copies de *ZAM*, localisées sur le chromosome X de la lignée RevI-H2i2, ont été apportées. L'une de ces deux copies présente une forte homologie de séquence avec la séquence du *ZAM* consensus identifié précédemment (Leblanc et al., 1997). L'autre insertion est un variant de *ZAM* qui possède une délétion de 626 pb dans la partie C-terminale du gène *pol*, délétion qui se trouve dans une région ne codant aucun domaine protéique connu. Les capacités de transposition de ce variant restent aujourd'hui à explorer. Nous planifions de faire les tests de stérilité avec les lignées *flamΔZAM* où les chromosomes II ou III de la lignée RevI-H2i2, possédant respectivement cinq et trois insertions de *ZAM*, seront apportés. Cette étude nous permettra par la suite d'étudier l'influence du nombre d'insertion *ZAM* et de la nature des copies sur les capacités d'invasion et de transposition de *ZAM*, ainsi que sur la réponse cellulaire induite.

## **5- La délétion d'un ET d'un cluster de piARNs : une approche prometteuse**

La technique CRISPR/Cas9 nous a permis d'éliminer l'élément transposable *ZAM* du cluster *flamenco*. Nous sommes les premiers à réussir une telle délétion ciblée dans un cluster de piARNs. Toutefois, seules les séquences internes de *ZAM-flam* ont été éliminées, un solo-LTR de 454 pb est encore présent au site initial d'insertion de *ZAM*. Des piARNs ciblant le LTR de



*ZAM* sont donc toujours produits dans les ovaires de la lignée *flam* $\Delta$ *ZAM* mais ne sont pas suffisants pour réprimer des copies actives de *ZAM*. Une séquence d'ET insérée dans *flamenco* d'une taille supérieure à 454 pb est donc nécessaire pour permettre une répression efficace de cet ET dans les cellules folliculaires. Cependant, nous ne pouvons pas exclure que les piARNs produits permettent tout de même une faible répression de *ZAM*. Afin d'éliminer entièrement les séquences de *ZAM* du cluster *flamenco*, nous pourrions essayer une stratégie différente. En effet, il semble que, dans les différentes lignées mutantes obtenues, les cassures double brin générées par l'endonuclease Cas9 ont été réparées par recombinaison homologue (HDR) en utilisant une région homologue localisée à proximité du site de cassure comme matrice. Nous pourrions donc fournir une matrice complémentaire différente, *via* un autre plasmide - comme lors de la réalisation d'un *knock-in* (intégration d'une séquence donnée à un locus ciblé) - qui serait utilisée pour la réparation homologue (Banan, 2020). Dans ce cas, la matrice contiendrait le site de *flamenco* ciblé sans l'insertion *ZAM-flam*. Après réparation, nous obtiendrions ainsi le site d'insertion « vide », sans *ZAM-flam*.

Pour la suite, nous prévoyons aussi d'éliminer d'autres ETs de *flamenco* afin d'analyser l'impact sur leur régulation et leur expression. Pour cela, il est nécessaire de cibler des ETs insérés uniquement dans *flamenco* et dans aucun autre cluster de piARNs, tel que *gtwin*. Il serait également intéressant de retirer des ETs localisés dans des clusters de piARNs « dual-strand » germinaux afin de comparer le rôle des clusters somatiques et germinaux dans la répression des ETs. En effet, une étude récente a montré que la délétion des trois clusters de piARNs majoritaires dans les cellules germinales - à savoir *42AB*, *38C* et *20A* - engendre aucune dérégulation d'ETs, ni de stérilité (Gebert et al., 2021). Cependant, il existe un nombre beaucoup plus important de clusters exprimés dans les cellules germinales, induisant une forte redondance avec plusieurs clusters impliqués dans la répression d'un même ET. De plus, il est nécessaire de vérifier la présence de copies fonctionnelles de l'ET considéré dans le génome de la lignée utilisée pour l'étude. En effet, nous avons par exemple constaté dans notre cas que, dans la lignée *flam* $\Delta$ *ZAM*, aucune copie fonctionnelle de *ZAM* n'était initialement présente. Ceci révèle une diversité importante dans le fond génétique des différentes populations de drosophile. De manière générale, le contenu en ETs varie considérablement entre les espèces et les populations, reflétant une dynamique de transposition spécifique à chaque population (Lerat et al., 2019; Luo et al., 2020). La variabilité du fond génétique est une réelle difficulté à laquelle il est nécessaire d'être particulièrement vigilant pour toutes les études comparatives. De manière générale, éliminer de façon ciblée un ET d'un cluster de piARNs permet d'analyser le rôle des différents clusters de piARNs dans le contrôle de chaque ET, ainsi que l'impact de

la dérégulation d'éléments transposables individuels à l'échelle moléculaire, cellulaire et tissulaire.

## **6- Analyse de la transmission intercellulaire de *ZAM* et de son intégration dans le génome hôte**

À ce stade de notre étude, le mécanisme précis utilisé pour le transit de *ZAM* des cellules folliculaires jusqu'à l'ovocyte reste inconnu. Le rôle de l'enveloppe dans ce processus reste particulièrement énigmatique. Pour étudier sa fonction potentielle dans la transmission intercellulaire de l'ERV *gypsy*, Richard M Keegan et ses collaborateurs ont généré des transgènes rapporteurs contenant la séquence de *gypsy*, avec ou sans le gène codant l'enveloppe, fusionnée à un gène rapporteur (Keegan et al., 2021). Ces constructions ont été injectées dans un système cellulaire et ont permis de suivre l'expression de *gypsy* grâce au gène rapporteur afin d'analyser le transfert intercellulaire en présence et en absence de l'enveloppe. Nous souhaitons effectuer des constructions similaires avec *ZAM* ; je n'ai cependant jamais réussi à cloner les séquences de *ZAM* entièrement, la présence de ces séquences dans les bactéries semblait empêcher toute croissance bactérienne. Nous avons donc fait appel à une compagnie externe (Genewiz) pour réaliser ce clonage. Ils ont essayé différentes approches et différentes souches bactériennes, mais n'ont à leur tour jamais réussi à cloner la séquence entière de *ZAM*. Il semblerait que l'ORF de *ZAM* codant la RNase H soit toxique pour les bactéries. Nous envisageons donc de changer de stratégie et d'utiliser une technique qui ne nécessite pas de passer par des bactéries pour l'amplification plasmidique (amplification avec la protéomérase TelN). Une autre possibilité consiste à réaliser une modification directement sur un *ZAM* génomique, par CRISPR/Cas9, afin d'éliminer l'ORF *env*, et ensuite analyser l'impact de cette délétion sur ses facultés de transmission. Pour cela, il est nécessaire d'être dans un contexte où une seule copie fonctionnelle de *ZAM* est présente dans le génome et où il n'y a pas de production de piARNs ciblant *ZAM* afin de pouvoir suivre la réactivation de cette unique copie de *ZAM* mutée. La lignée *flamΔZAM*, qui ne possède pas de copie fonctionnelle de *ZAM* dans son génome initial, semble idéale pour mener cette étude puisqu'il est simple d'apporter une seule copie fonctionnelle de *ZAM*.

Nous avons montré qu'éliminer *ZAM* de son cluster de piARNs, ou abolir la voie des piARNs dans les cellules folliculaires, induit une dérégulation massive de *ZAM* dans les cellules folliculaires et une invasion de l'ovocyte par celui-ci. Nous avons cependant très peu de données sur l'intégration génomique de *ZAM*. D'autres études ont été réalisées *via* l'utilisation

d'une lignée de drosophile où la protéine Piwi est éliminée des cellules folliculaires de manière inductible entre les stades pupal et adulte à toutes les générations (*piwi-sKD* inductible) (Barckmann et al., 2018). Un taux d'ARN de *ZAM* important dans les ovaires a été constaté dans cette condition, corrélé avec la présence de forme circulaires de l'ADN rétroviral de *ZAM* (eccDNA : *extra-chromosomal circular DNA*) dans les embryons précoces pondus par les femelles *piwi-sKD*. Certaines études suggèrent que les eccDNA pourraient servir d'indicateur de transpositions car ils seraient des « *by-product* » formés pendant l'intégration du rétrovirus (Sloan and Wainberg, 2011; Thomas and Pritham, 2015; Bin Hamid et al., 2016). Cependant, une autre hypothèse propose que ces eccDNA proviennent d'une intégration abortive de l'ET dans le génome et reflèteraient plutôt que celui-ci ne s'est pas inséré dans le génome (Munir et al., 2013). Ainsi, la présence d'eccDNA dans les embryons n'est pas nécessairement un bon indicateur d'un cycle de réplication finalisé mais témoigne tout de même que la majorité des étapes du cycle, incluant l'invasion des cellules germinales et la transcription inverse, ont été menées à bien. De plus, de nouvelles insertions génomiques de *ZAM* ont été retrouvées dans le génome des embryons F2 pondus par les filles des femelles *piwi-sKD*, témoignant que *ZAM* a transposé dans le génome des cellules germinales suite à la déplétion de Piwi dans les cellules folliculaires (Barckmann et al., 2018). Cependant, l'analyse après soixante-treize générations sur la lignée *piwi-sKD* révèle que ces insertions n'ont pas été fixées dans la population (Mohamed et al., 2020). Notre analyse sur la lignée RevI-H2i2 a quant à elle permis de montrer la présence de multiples copies de *ZAM* qui ont été fixées dans la population, puisque celles-ci étaient déjà présentes dans le génome de la lignée RevI-H2 séquencé 10 ans auparavant. Ces résultats confirment que *ZAM* transpose bien dans le génome de la lignée germinale lorsqu'aucun mécanisme n'est présent pour éviter l'intégration génomique. Dix-sept de ces insertions ont été fixées dans la population, néanmoins un nombre beaucoup plus important d'insertions a probablement dû se produire à l'échelle individuelle. Certaines études suggèrent qu'à la suite d'une invasion par un élément transposable, un « *burst* » de transposition se produit, induisant une augmentation rapide du nombre d'insertions génomiques jusqu'à la mise en place d'une réponse adaptative (Le Rouzic and Capy, 2005). Les insertions neutres ou bénéfiques ont ensuite davantage de chance d'être fixées dans la population (Revue dans : (Capy, 2021)). De manière intéressante, nous avons trouvé quatre insertions de *ZAM* dans des clusters de piARNs « dual-strand » dans le génome de la lignée RevI-H2i2. Cette observation ne signifie pas nécessairement que *ZAM* s'insère préférentiellement dans les clusters de piARNs mais suggère plutôt que ces insertions, fixées dans le génome, jouent un rôle avantageux pour la survie de l'espèce.

Des études réalisées au sein de l'équipe ont montré que la machinerie d'intégration de *ZAM* reconnaît des séquences particulières et conduit à l'insertion de *ZAM* préférentiellement dans des séquences du type 5'-GCGCGCg-3' (Leblanc et al., 1999). La lignée *flamΔZAM* est un modèle d'étude puissant pour conduire une étude plus poussée sur les sites d'insertion génomiques de *ZAM*. Il serait notamment intéressant d'analyser l'environnement chromatinien initial des régions ciblées pour l'intégration de *ZAM* (présence de marques répressives telles que H3K9me3, H3K9me2 et H3K27me3, ou de marques activatrices telles que H3K4me3 et H3K36me3). L'impact d'une nouvelle intégration de *ZAM* sur l'environnement chromatinien pourra également être examiné par des expériences de ChIP-seq ou de CUT&RUN sur ovaires. Par ailleurs, le séquençage ARN nous permettra d'étudier l'impact des nouvelles insertions de *ZAM* sur l'expression des gènes de manière générale. Cette étude permettra donc, de manière plus large, d'analyser l'impact de la réactivation d'un ERV sur le génome dans des conditions où la voie des piARNs est complètement fonctionnelle.

Enfin, le stade développemental au cours duquel *ZAM* s'intègre dans le génome germlinal reste à déterminer. Nous avons visualisé la présence d'ARNs de *ZAM* dans l'ovocyte lorsque *ZAM* est dérégulé dans les cellules folliculaires. Ainsi, l'insertion de *ZAM* pourrait directement se produire dans le génome de l'ovocyte. Cependant, l'ovocyte est bloqué en prophase I de la méiose, phase pendant laquelle les chromosomes sont fortement condensés, ce qui représente une contrainte importante pour l'expression des gènes et pour l'intégration potentielle des éléments transposables. De façon intéressante il a été montré que le génome de l'ovocyte est transcriptionnellement inactif pendant une grande partie de la prophase I mais l'expression de certains gènes est réactivée 13 h environ avant la reprise de la méiose, au stade 9 de l'ovogenèse. Cette réactivation transcriptionnelle est associée à un remodelage de la chromatine et à une décompaction des caryosomes (Navarro-Costa et al., 2016). La déméthylase KDM5 régule l'établissement d'un niveau approprié de marques H3K4me3 dans le noyau de l'ovocyte au début de l'ovogenèse, ce qui définira à la fin de la prophase I les régions réactivées pour la transcription. Ainsi, durant les stades 9-10 de l'ovogenèse, stades où *ZAM* envahit l'ovocyte, une partie du génome de l'ovocyte est décompactée. Ce remodelage chromatinien permettrait potentiellement à *ZAM* de s'insérer dans le génome de l'ovocyte. Cependant, nous avons aussi montré que les ARNs de *ZAM* sont transmis aux embryons précoces et s'accumulent au pôle postérieur des embryons, où les futures cellules germinales vont se cellulariser. Plusieurs éléments transposables, comme l'élément *P*, ont la capacité de transposer dans le génome des cellules germinales primordiales (PGCs) des embryons (Moon et al., 2018). Les nouvelles insertions sont ensuite présentes dans toutes les cellules germinales de l'individu et ainsi

transmises aux descendants. Un senseur de transposition a été utilisé pour étudier les cellules ciblées par la transposition de différents ETs (Jensen and Heidmann, 1991; Wang et al., 2018). Le clonage de *ZAM* étant toujours impossible à l'heure actuelle, nous ne pouvons pour l'instant pas utiliser cette technique afin de répondre à cette question. Néanmoins, nous pourrions effectuer une hybridation *in situ* de *ZAM* sur chromosomes polytènes, ou un séquençage d'ADNgénomique par la technique Nanopore pour détecter les nouvelles insertions de *ZAM* dans le génome. Ces expériences pourraient être réalisées sur la descendance de femelles où l'expression de *ZAM* est réactivée (F1 *flamΔZAM*) afin de voir si de nouvelles insertions sont présentes dans les cellules somatiques des descendants, témoignant que celles-ci se sont produites directement dans le génome de l'ovocyte. Enfin, une autre question qui reste encore non résolue concerne la capacité de *ZAM* à transposer de manière intracellulaire, c'est-à-dire directement dans les cellules où il est exprimé, à savoir les cellules somatiques de l'ovaire.

#### 7- Observations en laboratoire versus évènements naturels

Pour finir, je tenais à souligner que les observations faites dans nos études sont probablement différentes de ce qui peut réellement se produire dans les populations sauvages. En effet, nous nous plaçons toujours dans des conditions extrêmes : nous créons des mutations radicales, nous abolissons la production d'une protéine à l'échelle d'un tissu entier, et tout ça toujours à l'échelle d'une population entière. Les phénotypes générés sont souvent drastiques, engendrant notamment une stérilité. La mutation *flamΔZAM* à l'état homozygote, par exemple, induit une dépression massive de *ZAM* dans les cellules folliculaires, une invasion de l'ovocyte et une stérilité quasiment totale. Il est intéressant de noter que cette même mutation à l'état hétérozygote induit seulement une légère dépression de *ZAM* dans les cellules folliculaires : *ZAM* n'est que très faiblement exprimé dans quelques cellules folliculaires localisées en postérieur des follicules tardifs, et je n'ai pas détecté d'ARN de *ZAM* dans l'ovocyte. Aucune atrophie des ovaires est visible à l'état hétérozygote et la fertilité de ces femelles ne semble pas être affectée.

Lorsqu'une telle mutation se produit dans une population naturelle, celle-ci est tout d'abord présente à l'état hétérozygote au sein d'un seul individu, avant d'être transmise à seulement quelques descendants dans un premier temps. Ainsi, la mutation persiste à l'état hétérozygote au sein de la population pendant un certain temps, permettant aux organismes de développer des mécanismes adaptatifs de défense progressivement. Même si aucune invasion des cellules germinales n'a été détectée par smFISH chez les individus hétérozygotes *flamΔZAM*, *ZAM* étant exprimé dans les cellules folliculaires, il est possible qu'une faible invasion de l'ovocyte

se produise en condition hétérozygote. Ce phénomène permettrait ainsi aux cellules germinales de développer une réponse adaptative efficace avant que la mutation ne soit fixée dans la population et ne soit présente à l'état homozygote.

Bien que les conditions d'expérimentation en laboratoire soient différentes des conditions naturelles, les études que nous menons nous permettent de comprendre les mécanismes moléculaires qui peuvent être développés suite à l'apparition de mutations et nous offrent l'opportunité de suivre de manière accélérée des événements qui se déroulent en un laps de temps beaucoup plus important dans la nature.

## CONCLUSION

Les résultats obtenus durant ma thèse montrent que les clusters de piARNs permettent une adaptation constante des cellules germinales face à l'activité mutagène des éléments transposables. Il est maintenant admis que la transposition des ETs peut être avantageuse sur bien des points, c'est une réelle source de modification génomique qui conduit à des innovations génétiques. Cependant, la transposition est également une source importante de dangers puisqu'elle peut engendrer l'apparition de mutations délétères, risquées pour la survie même des espèces. C'est pourquoi il est nécessaire pour l'organisme de maintenir une balance entre répression et expression des ETs afin de permettre l'évolution des espèces d'une part et la préservation des génomes de l'autre.

Les organismes ont donc développé des stratégies sophistiquées pour contrôler la transposition des ETs, plus particulièrement dans les cellules germinales qui sont les seules cellules de l'organisme à transmettre leur patrimoine génétique à la génération suivante. Les études menées jusqu'à présent ont principalement porté sur les mécanismes de contrôle présent à l'échelle cellulaire : comment les cellules germinales contrôlent-elles l'expression et la transposition des ETs exprimés au sein de leur propre génome ?

Dans les cellules germinales, la voie des piARNs permet de réprimer efficacement la transposition des ETs exprimés dans ces cellules. Cependant, il est maintenant admis que les cellules germinales sont aussi sujettes à des invasions d'ETs provenant de l'extérieur. C'est par exemple le cas lors des transferts horizontaux. Lorsqu'un ET arrive pour la première fois dans un organisme, celui-ci doit atteindre les cellules germinales et s'intégrer dans le génome germinale afin d'être ensuite transmis verticalement aux descendants. Les cellules germinales doivent dans ce cas développer une réponse immédiate et/ou adaptative pour se défendre efficacement contre l'invasion de cet envahisseur mutagène. Les mécanismes de défense mis en place par les cellules germinales en réponse à l'invasion d'un ET restent encore mal connus. Nous avons développé un modèle d'étude dans lequel les cellules germinales sont envahies par un ET provenant des cellules somatiques adjacentes. En effet, la délétion du rétrovirus endogène *ZAM* de son cluster de piARNs *flamenco* induit une dérépression massive de *ZAM* dans les cellules somatiques adjacentes aux cellules germinales et une invasion de l'ovocyte. Au moment de l'invasion, aucun système de défense permettant de protéger les cellules germinales face à la transposition de *ZAM* n'est présent dans celles-ci. Cette condition peut ainsi être comparée à ce qui se produit en cas de transfert horizontal, lorsqu'un ET envahit pour

la première fois les cellules germinales. L'analyse de la réponse des cellules germinales après une invasion par *ZAM* est donc un modèle unique pour étudier la réponse des cellules germinales à la suite d'une invasion par un ET envahisseur. Mes travaux de thèse ont démontré pour la première fois que, dans les cellules germinales, la voie des piARNs permet une réponse adaptative efficace pour contrôler la transposition d'ETs provenant des cellules somatiques adjacentes. De ce fait, la voie des piARNs n'est pas seulement efficace pour contrôler les ETs à l'échelle cellulaire, mais permet aussi le contrôle des ETs provenant de l'extérieur, à l'échelle tissulaire. Cependant, la réponse par la voie des piARNs ne semble pas être immédiate, suggérant qu'une phase de « susceptibilité » permet aux ETs de transposer dans le génome germinale avant la mise en place d'une réponse adaptative. La transposition dans les cellules germinales permet aux ETs d'être conservés dans les génomes et de générer de potentielles diversités génétiques.

Une fois la réponse adaptative mise en place, les cellules germinales sont protégées face à de nouvelles invasions : le rétrovirus endogène envahisseur a été ajouté à la liste des ETs à réprimer dans les cellules germinales. C'est probablement grâce à cette coévolution permanente entre les ETs et les systèmes de défense de l'hôte que les ETs cohabitent dans le génome de toutes les espèces depuis maintenant déjà bien longtemps à l'échelle de l'évolution.





## LA SCIENCE AVANCE PLUS VITE AVEC DU REcul

Je voudrais finir cette thèse par une conclusion beaucoup plus générale.

Comme beaucoup de projet, celui de ma thèse ne se termine pas vraiment aujourd'hui mais amorce juste le début d'une autre histoire. Durant ces 4 années de thèse j'ai donné le meilleur de moi-même et j'espère avoir apporté une petite pierre à l'édifice scientifique et fait naître quelques idées nouvelles. Cette aventure m'a permis de plonger au cœur du monde de la recherche, un univers fait d'innombrables montagnes russes scientifiques et émotionnelles. On m'a toujours dit que pour être un bon chercheur il était nécessaire d'être excessivement curieux et je me rends compte aujourd'hui à quel point c'est vrai. Mais je dirais que la persévérance et la patience ont également été plus que nécessaires, deux qualités que j'ai dû apprendre pendant cette thèse... Durant 4 ans, j'ai eu l'impression de mener une enquête, où je récoltais chaque jour des indices supplémentaires pour aller dans une direction plutôt qu'une autre. Ce fût un périple à la fois passionnant, stimulant mais aussi semé d'embûches et de faux espoirs. Et le problème de cette enquête, c'est qu'elle n'a pas vraiment de fin... Pourtant un jour il a fallu prendre tous les éléments qu'on avait en notre possession et construire une histoire cohérente. Ce fût une étape difficile à franchir de mon côté, car elle nécessite une sacrée confiance en ses résultats, en soi. De nombreuses fois j'ai eu des doutes, sur mes expériences, sur mes conclusions et sur mes connaissances aussi... Mais j'ai vite appris la nécessité de douter et, de ce fait, l'impossibilité de tenir une chose pour acquise, de tout savoir ou de tout expliquer.

Aujourd'hui les résultats que je vous partage ne sont pas toujours noir ou blanc, c'est souvent le cas je crois. Comme tout projet, le mien présente des faiblesses et nombreuses des conclusions que nous dressons pourront probablement être remises en question. C'est pourquoi, je tenais à insister sur le fait que les conclusions établies dans ce manuscrit sont celles que nous avons pu imaginer avec les informations que nous avons récolté à cet instant t. En lisant les thèses écrites par les étudiants passés, notamment celle de Sophie DESSET qui a été une réelle source d'inspiration, je me suis rendue compte à quel point les projets s'imbriquent et se construisent les uns grâce aux autres.

Donc au final cette histoire, n'est pas non plus comme un morceau plastique (dixit Barbara McClintock) : elle est, comme les organismes, vouée à constamment évoluer.

A vous qui lisez cette thèse peut-être plusieurs années après son écriture, voilà où nous en étions en 2022. J'ai déjà hâte de savoir comment les connaissances vont évoluer dans quelques années...



## BIBLIOGRAPHIE

- Acheson, N. H. (2011). *Fundamentals of Molecular Virology*. doi:10.1002/pauz.200790112.
- Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., et al. (2000). The genome sequence of *Drosophila melanogaster*. *Science* 287, 2185–2195. doi:10.1126/science.287.5461.2185.
- Akkouche, A., Mugat, B., Barckmann, B., Varela-Chavez, C., Li, B., Raffel, R., et al. (2017). Piwi Is Required during *Drosophila* Embryogenesis to License Dual-Strand piRNA Clusters for Transposon Repression in Adult Ovaries. *Mol. Cell* 66, 411–419.e4. doi:10.1016/j.molcel.2017.03.017.
- Andersen, P. R., Tirian, L., Vunjak, M., and Brennecke, J. (2017). A heterochromatin-dependent transcription machinery drives piRNA expression. *Nature* 549, 54–59. doi:10.1038/nature23482.
- Antel, M., and Inaba, M. (2020). Modulation of Cell-Cell Interactions in *Drosophila* Oocyte Development. *Cells* 9, 2–3. doi:10.3390/cells9020274.
- Aravin, A. A., Hannon, G. J., and Brennecke, J. (2007). The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* 318, 761–764. doi:10.1126/science.1146484.
- Aravin, A. A., Lagos-Quintana, M., Yalcin, A., Zavolan, M., Marks, D., Snyder, B., et al. (2003). The small RNA profile during *Drosophila melanogaster* development. *Dev. Cell* 5, 337–350. doi:10.1016/s1534-5807(03)00228-4.
- Aravin, A. A., Naumova, N. M., Tulin, A. V., Vagin, V. V., Rozovsky, Y. M., and Gvozdev, V. A. (2001). Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline. *Curr. Biol.* 11, 1017–1027. doi:10.1016/S0960-9822(01)00299-8.
- Aravin, A., Gaidatzis, D., Iovino, N., Morris, P., Brownstein, M. J., Kuramochi-miyagawa, S., et al. (2006). A novel class of small RNAs bind to MILI protein in mouse testes. 442, 203–207. doi:10.1038/nature04916.
- Babich, V., Aksenov, N., Alexeenko, V., Oei, S. L., Buchlow, G., and Tomilin, N. (1999). Association of some potential hormone response elements in human genes with the Alu family repeats. *Gene* 239, 341–349. doi:10.1016/s0378-1119(99)00391-1.
- Balaj, L., Lessard, R., Dai, L., Cho, Y.-J., Pomeroy, S. L., Breakefield, X. O., et al. (2011). Tumour microvesicles contain retrotransposon elements and amplified oncogene sequences. *Nat. Commun.* 2, 180. doi:10.1038/ncomms1180.
- Banan, M. (2020). Recent advances in CRISPR/Cas9-mediated knock-ins in mammalian cells. *J. Biotechnol.* 308, 1–9. doi:10.1016/j.jbiotec.2019.11.010.
- Bannert, N., and Kurth, R. (2006). The evolutionary dynamics of human endogenous retroviral families. *Annu. Rev. Genomics Hum. Genet.* 7, 149–173. doi:10.1146/annurev.genom.7.080505.115700.
- Barckmann, B., El-Barouk, M., Péliisson, A., Mugat, B., Li, B., Franckhauser, C., et al. (2018). The somatic piRNA pathway controls germline transposition over generations. *Nucleic Acids Res.* 46, 9524–9536. doi:10.1093/nar/gky761.
- Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., et al. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709–1712. doi:10.1126/science.1138140.
- Bastock, R., and St Johnston, D. (2008). *Drosophila* oogenesis. *Curr. Biol.* 18, 1082–1087. doi:10.1016/j.cub.2008.09.011.
- Batki, J., Schnabl, J., Wang, J., Handler, D., Andreev, V. I., Stieger, C. E., et al. (2019). The nascent RNA binding complex SFiNX licenses piRNA-guided heterochromatin formation. *Nat. Struct. Mol. Biol.* 26, 720–731. doi:10.1038/s41594-019-0270-6.
- Baumgartner, L., Handler, D., Platzer, S., Duchek, P., and Brennecke, J. (2022). The *Drosophila* ZAD zinc finger protein Kipferl guides Rhino to piRNA clusters. *bioRxiv*, 2022.05.09.491178. doi:10.1101/2022.05.09.491178.
- Bedwell, G. J., and Engelman, A. N. (2021). You can keep your coat on—High-resolution imaging techniques reveal new insights into the actions of the retrovirus HIV-1 inside host cells. *Elife* 10, 1–3. doi:10.7554/ELIFE.64776.
- Bejerano, G., Lowe, C. B., Ahituv, N., King, B., Siepel, A., Salama, S. R., et al. (2006). A distal enhancer and an ultraconserved exon are derived from a novel retroposon. *Nature* 441, 87–90. doi:10.1038/nature04696.
- Belshaw, R., Pereira, V., Katzourakis, A., Talbot, G., Pačes, J., Burt, A., et al. (2004). Long-term reinfection of the human genome by endogenous retroviruses. *Proc. Natl. Acad. Sci. U. S. A.* 101, 4894–4899. doi:10.1073/pnas.0307800101.
- Benson, S. J., Ruis, B. L., Fadly, A. M., and Conklin, K. F. (1998). The unique envelope gene of the subgroup J avian leukosis virus derives from ev/J proviruses, a novel family of avian endogenous viruses. *J. Virol.* 72, 10157–10164. doi:10.1128/JVI.72.12.10157-10164.1998.
- Berg, D. E., and Howe, M. M. (1989). *Mobile DNA*. American Society for Microbiology, Washington, D.C.

- Biémont, C., Ronsseray, S., Anxolabéhère, D., Izaabel, H., and Gautier, C. (1990). Localization of P elements, copy number regulation, and cytotype determination in *Drosophila melanogaster*. *Genet. Res.* 56, 3–14. doi:10.1017/s0016672300028822.
- Bier, E. (2005). *Drosophila*, the golden bug, emerges as a tool for human genetics. *Nat. Rev. Genet.* 6, 9–23. doi:10.1038/nrg1503.
- Biessmann, H., and Mason, J. M. (1992). Genetics and molecular biology of telomeres. *Adv. Genet.* 30, 185–249. doi:10.1016/s0065-2660(08)60321-1.
- Bin Hamid, F., Kim, J., and Shin, C.-G. (2016). Cellular and viral determinants of retroviral nuclear entry. *Can. J. Microbiol.* 62, 1–15. doi:10.1139/cjm-2015-0350.
- Blaise, S., de Parseval, N., Bénil, L., and Heidmann, T. (2003). Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. *Proc. Natl. Acad. Sci. U. S. A.* 100, 13013–13018. doi:10.1073/pnas.2132646100.
- Blond, J. L., Lavillette, D., Cheynet, V., Bouton, O., Oriol, G., Chapel-Fernandes, S., et al. (2000). An envelope glycoprotein of the human endogenous retrovirus HERV-W is expressed in the human placenta and fuses cells expressing the type D mammalian retrovirus receptor. *J. Virol.* 74, 3321–3329. doi:10.1128/jvi.74.7.3321-3329.2000.
- Boeke, J. D., and Stoye, J. P. (1997). “Retrotransposons, Endogenous Retroviruses, and the Evolution of Retroelements,” in eds. J. M. Coffin, S. H. Hughes, and H. E. Varmus (Cold Spring Harbor (NY)).
- Bohrmann, J., and Zimmermann, J. (2008). Gap junctions in the ovary of *Drosophila melanogaster*: Localization of innexins 1, 2, 3 and 4 and evidence for intercellular communication via innexin-2 containing channels. *BMC Dev. Biol.* 8, 1–12. doi:10.1186/1471-213X-8-111.
- Brasset, E., Bantignies, F., Court, F., Cheresiz, S., Conte, C., and Vaury, C. (2007). Idefix insulator activity can be modulated by nearby regulatory elements. *Nucleic Acids Res.* 35, 2661–2670. doi:10.1093/nar/gkm140.
- Brasset, E., Taddei, A. R., Arnaud, F., Faye, B., Fausto, A. M., Mazzini, M., et al. (2006). Viral particles of the endogenous retrovirus ZAM from *Drosophila melanogaster* use a pre-existing endosome/exosome pathway for transfer to the oocyte. *Retrovirology* 3, 1–9. doi:10.1186/1742-4690-3-25.
- Bregliano, J. C., Picard, G., Bucheton, A., Pelisson, A., Lavige, J. M., and L’Heritier, P. (1980). Hybrid dysgenesis in *Drosophila melanogaster*. *Science* 207, 606–611. doi:10.1126/science.6766221.
- Brennecke, J., Aravin, A. A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., et al. (2007). Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* 128, 1089–1103. doi:10.1016/j.cell.2007.01.043.
- Brennecke, J., Malone, C. D., Aravin, A. A., Sachidanandam, R., Stark, A., and Hannon, G. J. (2008). An Epigenetic Role for Maternally Inherited piRNAs in Transposon Silencing. *Science (80-. )*. 1387, 1387–1392. doi:10.1126/science.1165171.
- Brink, P. R., Valiunas, V., Gordon, C., Rosen, M. R., and Cohen, I. S. (2012). Can gap junctions deliver? *Biochim. Biophys. Acta - Biomembr.* 1818, 2076–2081. doi:10.1016/j.bbamem.2011.09.025.
- Brower-Toland, B., Findley, S. D., Jiang, L., Liu, L., Yin, H., Dus, M., et al. (2007). *Drosophila* PIWI associates with chromatin and interacts directly with HP1a. *Genes Dev.* 21, 2300–2311. doi:10.1101/gad.1564307.
- Bucheton, A. (1995). The relationship between the flamenco gene and gypsy in *Drosophila*: how to tame a retrovirus. 11, 349–353.
- Burns, K. H. (2017). Transposable elements in cancer. *Nat. Rev. Cancer* 17, 415–424. doi:10.1038/nrc.2017.35.
- Calvi, B. R., Hong, T. J., Findley, S. D., and Gelbart, W. M. (1991). Evidence for a common evolutionary origin of inverted repeat transposons in *Drosophila* and plants: hobo, Activator, and Tam3. *Cell* 66, 465–471. doi:10.1016/0092-8674(81)90010-6.
- Capy, P. (2021). Taming , Domestication and Exaptation : Trajectories of Transposable Elements in Genomes.
- Casier, K., Autaa, J., Gueguen, N., Delmarre, V., Marie, P. P., Ronsseray, S., et al. (2022). The histone demethylase KDM3 prevents auto-immune piRNAs production in *Drosophila*. *bioRxiv*, 9–25.
- Cavarec, L., Jensen, S., Casella, J. F., Cristescu, S. A., and Heidmann, T. (1997). Molecular cloning and characterization of a transcription factor for the copia retrotransposon with homology to the BTB-containing lola neurogenic factor. *Mol. Cell. Biol.* 17, 482–494. doi:10.1128/mcb.17.1.482.
- Cecere, G., and Grishok, A. (2014). A nuclear perspective on RNAi pathways in metazoans. *Biochim. Biophys. Acta - Gene Regul. Mech.* 1839, 223–233. doi:10.1016/j.bbagr.2013.11.009.
- Chalvet, F., Teyssset, L., Terzian, C., Prud’homme, N., Santamaria, P., Bucheton, A., et al. (1999). Proviral amplification of the Gypsy endogenous retrovirus of *Drosophila melanogaster* involves env-independent invasion of the female germline. *EMBO J.* 18, 2659–2669. doi:10.1093/emboj/18.9.2659.
- Chang, T. H., Mattei, E., Gainetdinov, I., Colpan, C., Weng, Z., Zamore, P. D., et al. (2019). Maelstrom Represses Canonical Polymerase II Transcription within Bi-directional piRNA Clusters in *Drosophila melanogaster* Article Maelstrom Represses Canonical Polymerase II Transcription within Bi-directional piRNA Clusters in *Drosophila melanogaster*. *Mol. Cell* 73, 291-303.e6. doi:10.1016/j.molcel.2018.10.038.
- Chen, P. Y., Mannings, H., Slanchev, K., Chien, M., Russo, J. J., Ju, J., et al. (2005). The developmental miRNA

- profiles of zebrafish as determined by small RNA cloning. *Genes Dev.* 19, 1288–1293. doi:10.1101/gad.1310605.
- Chenais, B. (2015). Transposable elements in cancer and other human diseases. *Curr. Cancer Drug Targets* 15, 227–242. doi:10.2174/1568009615666150317122506.
- Chénais, B. (2013). Transposable elements and human cancer: A causal relationship? *Biochim. Biophys. Acta - Rev. Cancer* 1835, 28–35. doi:10.1016/j.bbcan.2012.09.001.
- Cheng, Z.-J., and Murata, M. (2003). A centromeric tandem repeat family originating from a part of Ty3/gypsy-retroelement in wheat and its relatives. *Genetics* 164, 665–672. doi:10.1093/genetics/164.2.665.
- Cherkasova, E., Scrivani, C., Doh, S., Weisman, Q., Takahashi, Y., Harashima, N., et al. (2016). Detection of an immunogenic HERV-E envelope with selective expression in clear cell kidney cancer. *Cancer Res.* 76, 2177–2185. doi:10.1158/0008-5472.CAN-15-3139.
- Chung, W. J., Okamura, K., Martin, R., and Lai, E. C. (2008). Endogenous RNA Interference Provides a Somatic Defense against *Drosophila* Transposons. *Curr. Biol.* 18, 795–802. doi:10.1016/j.cub.2008.05.006.
- Coffin, J. M., Hughes, S. H., and Varmus, H. E. (1997). “The Interactions of Retroviruses and their Hosts.” in, eds. J. M. Coffin, S. H. Hughes, and H. E. Varmus (Cold Spring Harbor (NY)).
- Conte, C., Dastugue, B., and Vaury, C. (2002). Coupling of Enhancer and Insulator Properties Identified in Two Retrotransposons Modulates Their Mutagenic Impact on Nearby Genes. *Mol. Cell. Biol.* 22, 1767–1777. doi:10.1128/mcb.22.6.1767-1777.2002.
- Cox, D. N., Chao, A., Baker, J., Chang, L., Qiao, D., and Lin, H. (1998). A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. *Genes Dev.* 12, 3715–3727. doi:10.1101/gad.12.23.3715.
- Czech, B., Malone, C. D., Zhou, R., Stark, A., Schlingeheyde, C., Dus, M., et al. (2008). An endogenous small interfering RNA pathway in *Drosophila*. *Nature* 453, 798–802. doi:10.1038/nature07007.
- Datta, N., Hedges, R. W., Shaw, E. J., Sykes, R. B., and Richmond, M. H. (1971). Properties of an R factor from *Pseudomonas aeruginosa*. *J. Bacteriol.* 108, 1244–1249. doi:10.1128/jb.108.3.1244-1249.1971.
- De Parseval, N., and Heidmann, T. (2005). Human endogenous retroviruses: From infectious elements to human genes. *Cytogenet. Genome Res.* 110, 318–332. doi:10.1159/000084964.
- de Parseval, N., Lazar, V., Casella, J.-F., Benit, L., and Heidmann, T. (2003). Survey of human genes of retroviral origin: identification and transcriptome of the genes with coding capacity for complete envelope proteins. *J. Virol.* 77, 10414–10422. doi:10.1128/jvi.77.19.10414-10422.2003.
- De Vanssay, A., Bougé, A. L., Boivin, A., Hermant, C., Teyssset, L., Delmarre, V., et al. (2012a). Paramutation in *Drosophila* linked to emergence of a piRNA-producing locus. *Nature* 490, 112–115. doi:10.1038/nature11416.
- De Vanssay, A., Bougé, A. L., Boivin, A., Hermant, C., Teyssset, L., Delmarre, V., et al. (2012b). Paramutation in *Drosophila* linked to emergence of a piRNA-producing locus. *Nature* 490, 112–115. doi:10.1038/nature11416.
- Deininger, P. L., and Batzer, M. A. (2002). Mammalian retroelements. *Genome Res.* 12, 1455–1465. doi:10.1101/gr.282402.
- Delviks-Frankenberry, K., Galli, A., Nikolaitchik, O., Mens, H., Pathak, V. K., and Hu, W. S. (2011). Mechanisms and factors that influence high frequency retroviral recombination. *Viruses* 3, 1650–1680. doi:10.3390/v3091650.
- Dennis, C., Brasset, E., Sarkar, A., and Vaury, C. (2016). Export of piRNA precursors by EJC triggers assembly of cytoplasmic Yb-body in *Drosophila*. *Nat. Commun.* 7, 1–12. doi:10.1038/ncomms13739.
- Dennis, C., Zanni, V., Brasset, E., Eymery, A., Zhang, L., Mteirek, R., et al. (2013). “Dot COM”, a Nuclear Transit Center for the Primary piRNA Pathway in *Drosophila*. *PLoS One* 8. doi:10.1371/journal.pone.0072752.
- Desset, S., Buchon, N., Meignin, C., Coiffet, M., and Vaury, C. (2008). In *Drosophila melanogaster* the COM locus directs the somatic silencing of two retrotransposons through both Piwi-dependent and -independent pathways. *PLoS One* 3. doi:10.1371/journal.pone.0001526.
- Desset, S., Conte, C., Dimitri, P., Calco, V., Dastugue, B., and Vaury, C. (1999). Mobilization of two retroelements, ZAM and Idefix, in a novel unstable line of *Drosophila melanogaster*. *Mol. Biol. Evol.* 16, 54–66. doi:10.1093/oxfordjournals.molbev.a026038.
- Desset, S., Meignin, C., Dastugue, B., and Vaury, C. (2003). COM, a heterochromatic locus governing the control of independent endogenous retroviruses from *Drosophila melanogaster*. *Genetics* 164, 501–509.
- Dewannieux, M., Blaise, S., and Heidmann, T. (2005). Identification of a functional envelope protein from the HERV-K family of human endogenous retroviruses. *J. Virol.* 79, 15573–15577. doi:10.1128/JVI.79.24.15573-15577.2005.
- Ding, S. W. (2010). RNA-based antiviral immunity. *Nat. Rev. Immunol.* 10, 632–644. doi:10.1038/nri2824.
- Dinger, M. E., Mercer, T. R., and Mattick, J. S. (2008). RNAs as extracellular signaling molecules. *J. Mol.*

- Endocrinol.* 40, 151–159. doi:10.1677/JME-07-0160.
- Dis Duke, D. J., and Aiken, C. (2006). Evidence for a functional link between uncoating of the human immunodeficiency virus type 1 core and nuclear import of the viral preintegration complex. *J. Virol.* 80, 3712–3720. doi:10.1128/JVI.80.8.3712-3720.2006.
- Dönertas, D., Sienski, G., and Brennecke, J. (2013). Drosophila Gtsf1 is an essential component of the Piwi-mediated transcriptional silencing complex. *Genes Dev.* 27, 1693–1705. doi:10.1101/gad.221150.113.
- Duc, C., Yoth, M., Jensen, S., Mounié, N., Bergman, C. M., Vaury, C., et al. (2019). Trapping a somatic endogenous retrovirus into a germline piRNA cluster immunizes the germline against further invasion. *Genome Biol.* 20, 1–14. doi:10.1186/s13059-019-1736-x.
- Dufourt, J., Bontonou, G., Chartier, A., Jahan, C., Meunier, A. C., Pierson, S., et al. (2017). PiRNAs and Aubergine cooperate with Wispy poly(A) polymerase to stabilize mRNAs in the germ plasm. *Nat. Commun.* 8, 1–12. doi:10.1038/s41467-017-01431-5.
- Dupressoir, A., Lavialle, C., and Heidmann, T. (2012). From ancestral infectious retroviruses to bona fide cellular genes: role of the captured syncytins in placentation. *Placenta* 33, 663–671. doi:10.1016/j.placenta.2012.05.005.
- Dupressoir, A., Marceau, G., Vernochet, C., Bénéit, L., Kanellopoulos, C., Sapin, V., et al. (2005). Syncytin-A and syncytin-B, two fusogenic placenta-specific murine envelope genes of retroviral origin conserved in Muridae. *Proc. Natl. Acad. Sci. U. S. A.* 102, 725–730. doi:10.1073/pnas.0406509102.
- Echalier, G. (1989). Drosophila retrotransposons: interactions with genome. *Adv. Virus Res.* 36, 33–105. doi:10.1016/s0065-3527(08)60582-5.
- ElMaghraby, M. F., Andersen, P. R., Pühringer, F., Hohmann, U., Meixner, K., Lendl, T., et al. (2019). A Heterochromatin-Specific RNA Export Pathway Facilitates piRNA Production. *Cell* 178, 964–979.e20. doi:10.1016/j.cell.2019.07.007.
- Emmons, S. W., Yesner, L., Ruan, K. S., and Katzenberg, D. (1983). Evidence for a transposon in *Caenorhabditis elegans*. *Cell* 32, 55–65. doi:10.1016/0092-8674(83)90496-8.
- Ephrussi, A., and Lehmann, R. (1992). Induction of germ cell formation by oskar. *Nature* 358, 387–392. doi:10.1038/358387a0.
- Fabiet, M., Lerat, E., Rebollo, R., Horard, B., Burlet, N., Martinez, S., et al. (2009). Genomic environment influences the dynamics of the tirant LTR retrotransposon in *Drosophila*. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* 23, 1482–1489. doi:10.1096/fj.08-123513.
- Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* 9, 397–405. doi:10.1038/nrg2337.
- Feschotte, C., and Pritham, E. J. (2007). DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* 41, 331–368. doi:10.1146/annurev.genet.40.110405.090448.
- Finnegan, D. J. (1989). Eukaryotic transposable elements and genome evolution. *Trends Genet.* 5, 103–107. doi:10.1016/0168-9525(89)90039-5.
- Gebert, D., Neubert, L. K., Lloyd, C., Gui, J., Lehmann, R., and Teixeira, F. K. (2021). Large *Drosophila* germline piRNA clusters are evolutionarily labile and dispensable for transposon regulation. *Mol. Cell*, 1–14. doi:10.1016/j.molcel.2021.07.011.
- George, P., Jensen, S., Pogorelnik, R., Lee, J., Xing, Y., Brasslet, E., et al. (2015). Increased production of piRNAs from euchromatic clusters and genes in *Anopheles gambiae* compared with *Drosophila melanogaster*. *Epigenetics and Chromatin* 8. doi:10.1186/s13072-015-0041-5.
- Ghildiyal, M., Seitz, H., Horwich, M. D., Li, C., Du, T., Lee, S., et al. (2008). Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science* 320, 1077–1081. doi:10.1126/science.1157396.
- Ghildiyal, M., and Zamore, P. D. (2009). Small silencing RNAs: An expanding universe. *Nat. Rev. Genet.* 10, 94–108. doi:10.1038/nrg2504.
- Giordani, G., Cavaliere, V., Gargiulo, G., Lattanzi, G., and Andrenacci, D. (2022). Retrotransposons down- and up-regulation in aging somatic tissues. *Cells* 11, 3–11. doi:10.3390/cells11010079.
- Girard, A., Sachidanandam, R., Hannon, G. J., and Carmell, M. A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* 442, 199–202. doi:10.1038/nature04917.
- Gonzalez-Cao, M., Iduma, P., Karachaliou, N., Santarpia, M., Blanco, J., and Rosell, R. (2016). Human endogenous retroviruses and cancer. *Cancer Biol. Med.* 13, 483–488. doi:10.20892/j.issn.2095-3941.2016.0080.
- González-Reyes, A., Elliott, H., and St Johnston, D. (1995). Polarization of both major body axes in *Drosophila* by gurken-torpedo signalling. *Nature* 375, 654–658. doi:10.1038/375654a0.
- Goodier, J. L., and Kazazian, H. H. (2008). Retrotransposons Revisited: The Restraint and Rehabilitation of Parasites. *Cell* 135, 23–35. doi:10.1016/j.cell.2008.09.022.
- Goriaux, C., Desset, S., Renaud, Y., Vaury, C., and Brasslet, E. (2014a). Transcriptional properties and splicing of the flamenco piRNA cluster. *EMBO Rep.* 15, 411–418. doi:10.1002/embr.201337898.

- Goriaux, C., Théron, E., Brasset, E., and Vaury, C. (2014b). History of the discovery of a master locus producing piRNAs: The flamenco/COM locus in *Drosophila melanogaster*. *Front. Genet.* 5, 1–8. doi:10.3389/fgene.2014.00257.
- Grivna, S. T., Beyret, E., Wang, Z., and Lin, H. (2006). A novel class of small RNAs in mouse spermatogenic cells. *Genes Dev.* 20, 1709–1714. doi:10.1101/gad.1434406.
- Gunawardane, L. S., Saito, K., Nishida, K. M., Miyoshi, K., Kawamura, Y., Nagami, T., et al. (2007). A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* 315, 1587–1590. doi:10.1126/science.1140494.
- Halo, J. V., Pendleton, A. L., Jarosz, A. S., Gifford, R. J., Day, M. L., and Kidd, J. M. (2019). Origin and recent expansion of an endogenous gammaretroviral lineage in domestic and wild canids. *Retrovirology* 16, 6. doi:10.1186/s12977-019-0468-z.
- Han, B. W., Wang, W., Li, C., and Weng, Z. (2015). piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. 348, 817–822.
- Hancks, D. C., and Kazazian, H. H. (2016). Roles for retrotransposon insertions in human disease. *Mob. DNA* 7. doi:10.1186/s13100-016-0065-9.
- Handler, D., Meixner, K., Pizka, M., Lauss, K., Schmied, C., Gruber, F. S., et al. (2013). The genetic makeup of the drosophila piRNA pathway. *Mol. Cell* 50, 762–777. doi:10.1016/j.molcel.2013.04.031.
- Hartl, D. L., Lozovskaya, E. R., Nurminsky, D. I., and Lohe, A. R. (1997). What restricts the activity of mariner-like transposable elements. *Trends Genet.* 13, 197–201. doi:10.1016/s0168-9525(97)01087-1.
- Hayashi, R., Schnabl, J., Handler, D., Mohn, F., Ameres, S. L., and Brennecke, J. (2016). Genetic and mechanistic diversity of piRNA 3'-end formation. *Nature* 539, 588–592. doi:10.1038/nature20162.
- Heidmann, O., Vernochet, C., Dupressoir, A., and Heidmann, T. (2009). Identification of an endogenous retroviral envelope gene with fusogenic activity and placenta-specific expression in the rabbit: a new “syncytin” in a third order of mammals. *Retrovirology* 6, 107. doi:10.1186/1742-4690-6-107.
- Hermant, C., Boivin, A., Teyssset, L., Delmarre, V., Asif-Laidin, A., Van Den Beek, M., et al. (2015). Paramutation in drosophila requires both nuclear and cytoplasmic actors of the piRNA pathway and induces cis-spreading of piRNA production. *Genetics* 201, 1381–1396. doi:10.1534/genetics.115.180307.
- Homolka, D., Pandey, R. R., Goriaux, C., Brasset, E., Vaury, C., Sachidanandam, R., et al. (2015a). PIWI Slicing and RNA Elements in Precursors Instruct Directional Primary piRNA Biogenesis. *Cell Rep.* 12, 418–428. doi:10.1016/j.celrep.2015.06.030.
- Homolka, D., Pandey, R. R., Goriaux, C., Brasset, E., Vaury, C., Sachidanandam, R., et al. (2015b). PIWI Slicing and RNA Elements in Precursors Instruct Directional Primary piRNA Biogenesis. *Cell Rep.* 12, 418–428. doi:10.1016/j.celrep.2015.06.030.
- Horwich, M. D., Li, C., Matranga, C., Vagin, V., Farley, G., Wang, P., et al. (2007). The *Drosophila* RNA Methyltransferase, DmHen1, Modifies Germline piRNAs and Single-Stranded siRNAs in RISC. *Curr. Biol.* 17, 1265–1272. doi:10.1016/j.cub.2007.06.030.
- Huang, C. R. L., Burns, K. H., and Boeke, J. D. (2012). Active Transposition in Genomes. *Annu. Rev. Genet.* doi:10.1146/annurev-genet-110711-155616.
- Hur, J. K., Luo, Y., Moon, S., Ninova, M., Marinov, G. K., Chung, Y. D., et al. (2016). Splicing-independent loading of TREX on nascent RNA is required for efficient expression of dual-strand piRNA clusters in *Drosophila*. *Genes Dev.* 30, 840–855. doi:10.1101/gad.276030.115.
- Hurst, T. P., and Magiorkinis, G. (2015). Activation of the innate immune response by endogenous retroviruses. *J. Gen. Virol.* 96, 1207–1218. doi:10.1099/jgv.0.000017.
- Ikeda, H., and Odaka, T. (1984). A cell membrane “gp70” associated with Fv-4 gene: immunological characterization, and tissue and strain distribution. *Virology* 133, 65–76. doi:10.1016/0042-6822(84)90426-4.
- Ivics, Z., Hackett, P. B., Plasterk, R. H., and Izsvák, Z. (1997). Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell* 91, 501–510. doi:10.1016/s0092-8674(00)80436-5.
- Jehn, J., Gebert, D., Pipilescu, F., Stern, S., Kiefer, J. S. T., Hewel, C., et al. (2018). PIWI genes and piRNAs are ubiquitously expressed in mollusks and show patterns of lineage-specific adaptation. *Commun. Biol.* 1, 137. doi:10.1038/s42003-018-0141-4.
- Jenkins, N. A., and Copeland, N. G. (1985). High frequency germline acquisition of ecotropic MuLV proviruses in SWR/J-RF/J hybrid mice. *Cell* 43, 811–819. doi:10.1016/0092-8674(85)90254-5.
- Jensen, S., and Heidmann, T. (1991). An indicator gene for detection of germline retrotransposition in transgenic *Drosophila* demonstrates RNA-mediated transposition of the LINE I element. *EMBO J.* 10, 1927–1937.
- Johnson, W. E. (2019). Origins and evolutionary consequences of ancient endogenous retroviruses. *Nat. Rev. Microbiol.* 17, 355–370. doi:10.1038/s41579-019-0189-2.
- Jones, B. C., Wood, J. G., Chang, C., Tam, A. D., Franklin, M. J., Siegel, E. R., et al. (2016). A somatic piRNA pathway in the *Drosophila* fat body ensures metabolic homeostasis and normal lifespan. *Nat. Commun.* 7,



13856. doi:10.1038/ncomms13856.
- Jones, R. N. (2005). McClintock's controlling elements: The full story. *Cytogenet. Genome Res.* 109, 90–103. doi:10.1159/000082387.
- Jönsson, M. E., Garza, R., Sharma, Y., Petri, R., Södersten, E., Johansson, J. G., et al. (2021). Activation of endogenous retroviruses during brain development causes an inflammatory response. *EMBO J.* 40, e106423. doi:10.15252/embj.2020106423.
- Jordan, I. K., Rogozin, I. B., Glazko, G. V., and Koonin, E. V. (2003). Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends Genet.* 19, 68–72. doi:10.1016/s0168-9525(02)00006-9.
- Kachroo, P., Leong, S. A., and Chattoo, B. B. (1994). Pot2, an inverted repeat transposon from the rice blast fungus *Magnaporthe grisea*. *Mol. Gen. Genet.* 245, 339–348. doi:10.1007/BF00290114.
- Kapitonov, V. V., and Jurka, J. (2005). RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. *PLoS Biol.* 3, e181. doi:10.1371/journal.pbio.0030181.
- Kapitonov, V. V., and Jurka, J. (2008). A universal classification of eukaryotic transposable elements implemented in Repbase. *Nat. Rev. Genet.* 9, 411–2; author reply 414. doi:10.1038/nrg2165-c1.
- Kawamura, Y., Saito, K., Kin, T., Ono, Y., Asai, K., Sunohara, T., et al. (2008). Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* 453, 793–797. doi:10.1038/nature06938.
- Kawamura, Y., Sanchez Calle, A., Yamamoto, Y., Sato, T. A., and Ochiya, T. (2019). Extracellular vesicles mediate the horizontal transfer of an active LINE-1 retrotransposon. *J. Extracell. Vesicles* 8. doi:10.1080/20013078.2019.1643214.
- Keegan, R. M., Talbot, L. R., Chang, Y. H., Metzger, M. J., and Dubnau, J. (2021). Intercellular viral spread and intracellular transposition of *Drosophila gypsy*. *PLoS Genet.* 17, 1–22. doi:10.1371/journal.pgen.1009535.
- Kidwell, M. G., Kidwell, J. F., and Sved, J. A. (1977). Hybrid dysgenesis in *Drosophila melanogaster*: a syndrome of aberrant traits including mutation, sterility and male recombination. *Genetics* 86, 813–833. doi:10.1093/genetics/86.4.813.
- Kim, A., Terzian, C., Santamaria, P., Péliisson, A., Prud'homme, N., and Bucheton, A. (1994). Retroviruses in invertebrates: The gypsy retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U. S. A.* 91, 1285–1289. doi:10.1073/pnas.91.4.1285.
- Klattenhoff, C., Xi, H., Li, C., Lee, S., Xu, J., Khurana, J. S., et al. (2009). The *Drosophila* HP1 Homolog Rhino Is Required for Transposon Silencing and piRNA Production by Dual-Strand Clusters. *Cell* 138, 1137–1149. doi:10.1016/j.cell.2009.07.014.
- Kneuss, E., Munafò, M., Eastwood, E. L., Deumer, U. S., Preall, J. B., Hannon, G. J., et al. (2019). Specialization of the *Drosophila* nuclear export family protein Nxf3 for piRNA precursor export. *Genes Dev.* 33, 1208–1220. doi:10.1101/gad.328690.119.
- Kobayashi, H., and Tomari, Y. (2016). RISC assembly: Coordination between small RNAs and Argonaute proteins. *Biochim. Biophys. Acta* 1859, 71–81. doi:10.1016/j.bbagr.2015.08.007.
- Kofler, R., Nolte, V., and Schlötterer, C. (2022). The transposition rate has little influence on the plateauing level of the P-element. *Mol. Biol. Evol.* 39, 1–16. doi:10.1093/molbev/msac141.
- Kong, Q., Quan, X., Du, J., Tai, Y., Liu, W., Zhang, J., et al. (2019). Endo-siRNAs regulate early embryonic development by inhibiting transcription of long terminal repeat sequence in pig $\dagger$ . *Biol. Reprod.* 100, 1431–1439. doi:10.1093/biolre/ioz042.
- Kozak, C. A., Gromet, N. J., Ikeda, H., and Buckler, C. E. (1984). A unique sequence related to the ecotropic murine leukemia virus is associated with the Fv-4 resistance gene. *Proc. Natl. Acad. Sci. U. S. A.* 81, 834–837. doi:10.1073/pnas.81.3.834.
- Krug, L., Chatterjee, N., Borges-Monroy, R., Hearn, S., Liao, W. W., Morrill, K., et al. (2017). *Retrotransposon activation contributes to neurodegeneration in a Drosophila TDP-43 model of ALS*. doi:10.1371/journal.pgen.1006635.
- Kuramochi-Miyagawa, S., Kimura, T., Yomogida, K., Kuroiwa, A., Tadokoro, Y., Fujita, Y., et al. (2001). Two mouse piwi-related genes: miwi and mili. *Mech. Dev.* 108, 121–133. doi:10.1016/s0925-4773(01)00499-3.
- Lajoinie, O., Drake, M. E., Dastugue, B., and Vauray, C. (1995). Aberrant pre-mRNA maturation is caused by LINE insertions into introns of the white gene of *Drosophila melanogaster*. *Nucleic Acids Res.* 23, 4015–4022. doi:10.1093/nar/23.20.4015.
- Lakshmanan, V., Sujith, T. N., Bansal, D., Shivaprasad, P. V., Palakodeti, D., and Krishna, S. (2021). Comprehensive annotation and characterization of planarian tRNA and tRNA-derived fragments (tRFs). *RNA* 27, 477–495. doi:10.1261/rna.077701.120.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921. doi:10.1038/35057062.
- Langin, T., Capy, P., and Daboussi, M. J. (1995). The transposable element impala, a fungal member of the Tc1-mariner superfamily. *Mol. Gen. Genet.* 246, 19–28. doi:10.1007/BF00290129.
- Lau, N. C., Seto, A. G., Kim, J., Kuramochi-Miyagawa, S., Nakano, T., Bartel, D. P., et al. (2006).

- Characterization of the piRNA complex from rat testes. *Science* (80- ). 313, 363–367. doi:10.1126/science.1130164.
- Lavialle, C., Cornelis, G., Dupressoir, A., Esnault, C., Heidmann, O., Vernochet, C., et al. (2013). Paleovirology of “syncytins”, retroviral env genes exapted for a role in placentation. *Philos. Trans. R. Soc. London. Ser. B, Biol. Sci.* 368, 20120507. doi:10.1098/rstb.2012.0507.
- Le Rouzic, A., and Capy, P. (2005). The first steps of transposable elements invasion: Parasitic strategy vs. genetic drift. *Genetics* 169, 1033–1043. doi:10.1534/genetics.104.031211.
- Le Thomas, A., Stuwe, E., Li, S., Du, J., Marinov, G., Rozhkov, N., et al. (2014). Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. *Genes Dev.* 28, 1667–1680. doi:10.1101/gad.245514.114.
- Leblanc, P., Dastugue, B., and Vaury, C. (1999). The Integration Machinery of ZAM, a Retroelement from *Drosophila melanogaster*, Acts as a Sequence-Specific Endonuclease. *J. Virol.* 73, 7061–7064. doi:10.1128/jvi.73.8.7061-7064.1999.
- Leblanc, P., Dasset, S., Dastugue, B., and Vaury, C. (1997). Invertebrate retroviruses: ZAM a new candidate in *D. melanogaster*. *EMBO J.* 16, 7521–7531. doi:10.1093/emboj/16.24.7521.
- Leblanc, P., Dasset, S., Giorgi, F., Taddei, A. R., Fausto, A. M., Mazzini, M., et al. (2000). Life Cycle of an Endogenous Retrovirus, ZAM, in *Drosophila melanogaster*. *J. Virol.* 74, 10658–10669. doi:10.1128/jvi.74.22.10658-10669.2000.
- Lee, Y. N., and Bieniasz, P. D. (2007). Reconstitution of an infectious human endogenous retrovirus. *PLoS Pathog.* 3, e10. doi:10.1371/journal.ppat.0030010.
- Lerat, E., and Capy, P. (1999). Retrotransposons and retroviruses: analysis of the envelope gene. *Mol. Biol. Evol.* 16, 1198–1207. doi:10.1093/oxfordjournals.molbev.a026210.
- Lerat, E., Goubert, C., Guirao-Rico, S., Merenciano, M., Dufour, A.-B., Vieira, C., et al. (2019). Population-specific dynamics and selection patterns of transposable element insertions in European natural populations. *Mol. Ecol.* 28, 1506–1522. doi:10.1111/mec.14963.
- Lewis, S. H., Quarles, K. A., Yang, Y., Tanguy, M., Frézal, L., Smith, S. A., et al. (2018). Pan-arthropod analysis reveals somatic piRNAs as an ancestral defence against transposable elements. *Nat. Ecol. Evol.* 2, 174–181. doi:10.1038/s41559-017-0403-4.
- Li, C., Vagin, V. V., Lee, S., Xu, J., Ma, S., Xi, H., et al. (2009). Collapse of Germline piRNAs in the Absence of Argonaute3 Reveals Somatic piRNAs in Flies. *Cell* 137, 509–521. doi:10.1016/j.cell.2009.04.027.
- Lim, A. K., and Kai, T. (2007). Unique germ-line organelle, nuage, functions to repress selfish genetic elements in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U. S. A.* 104, 6714–6719. doi:10.1073/pnas.0701920104.
- Lingel, A., Simon, B., Izaurralde, E., and Sattler, M. (2003). Structure and nucleic-acid binding of the *Drosophila* Argonaute 2 PAZ domain. *Nature* 426, 465–469. doi:10.1038/nature02123.
- Löwer, R. (1999). The pathogenic potential of endogenous retroviruses: facts and fantasies. *Trends Microbiol.* 7, 350–356. doi:10.1016/s0966-842x(99)01565-6.
- Lu, B., and Vogel, H. (2009). *Drosophila* models of neurodegenerative diseases. *Annu. Rev. Pathol.* 4, 315–342. doi:10.1146/annurev.pathol.3.121806.151529.
- Luo, S., Zhang, H., Duan, Y., Yao, X., Clark, A. G., and Lu, J. (2020). The evolutionary arms race between transposable elements and piRNAs in *Drosophila melanogaster*. *BMC Evol. Biol.* 20, 1–18. doi:10.1186/s12862-020-1580-3.
- Luo, Y., He, P., Kanrar, N., Toth, K. F., and Aravin, A. A. (2022). Maternally inherited siRNAs initiate piRNA cluster formation. *bioRxiv*, 2022.02.08.479612. doi:10.1101/2022.02.08.479612.
- Ma, X., Zhu, X., Han, Y., Story, B., Do, T., Song, X., et al. (2017). Aubergine Controls Germline Stem Cell Self-Renewal and Progeny Differentiation via Distinct Mechanisms. *Dev. Cell* 41, 157-169.e5. doi:10.1016/j.devcel.2017.03.023.
- Malik, H. S., Henikoff, S., and Eickbush, T. H. (2000). Poised for contagion: evolutionary origins of the infectious abilities of invertebrate retroviruses. *Genome Res.* 10, 1307–1318. doi:10.1101/gr.145000.
- Malone, C. D., Brennecke, J., Dus, M., Stark, A., McCombie, W. R., Sachidanandam, R., et al. (2009). Specialized piRNA Pathways Act in Germline and Somatic Tissues of the *Drosophila* Ovary. *Cell* 137, 522–535. doi:10.1016/j.cell.2009.03.040.
- Mangency, M., and Heidmann, T. (1998). Tumor cells expressing a retroviral envelope escape immune rejection in vivo. *Proc. Natl. Acad. Sci. U. S. A.* 95, 14920–14925. doi:10.1073/pnas.95.25.14920.
- Marco, A., and Marín, I. (2009). CGIN1: a retroviral contribution to mammalian genomes. *Mol. Biol. Evol.* 26, 2167–2170. doi:10.1093/molbev/msp127.
- Marlor, R. L., Parkhurst, S. M., and Corces, V. G. (1986). The *Drosophila melanogaster* gypsy transposable element encodes putative gene products homologous to retroviral proteins. *Mol. Cell. Biol.* 6, 1129–1134. doi:10.1128/mcb.6.4.1129-1134.1986.
- Matreyek, K. A., and Engelman, A. (2013). Viral and cellular requirements for the nuclear entry of retroviral

- preintegration nucleoprotein complexes. *Viruses* 5, 2483–2511. doi:10.3390/v5102483.
- McClintock, B. (1950). The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. U. S. A.* 36, 344–355. doi:10.1073/pnas.36.6.344.
- McClintock, B. (1953). Induction of Instability at Selected Loci in Maize. *Genetics* 38, 579–99. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17247459> [Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1209627](http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1209627).
- Meignin, C., Dastugue, B., and Vaury, C. (2004). Intercellular communication between germ line and somatic line is utilized to control the transcription of ZAM, an endogenous retrovirus from *Drosophila melanogaster*. *Nucleic Acids Res.* 32, 3799–3806. doi:10.1093/nar/gkh708.
- Mével-Ninio, M., Pelisson, A., Kinder, J., Campos, A. R., and Bucheton, A. (2007). The flamenco Locus Controls the gypsy and ZAM Retroviruses and Is Required for *Drosophila* Oogenesis. *Genetics* 1624, 1615–1624. doi:10.1534/genetics.106.068106.
- Mi, S., Lee, X., Li, X., Veldman, G. M., Finnerty, H., Racie, L., et al. (2000). Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* 403, 785–789. doi:10.1038/35001608.
- Miki, Y., Katagiri, T., Kasumi, F., Yoshimoto, T., and Nakamura, Y. (1996). Mutation analysis in the BRCA2 gene in primary breast cancers. *Nat. Genet.* 13, 245–247. doi:10.1038/ng0696-245.
- Miki, Y., Nishisho, I., Horii, A., Miyoshi, Y., Utsunomiya, J., Kinzler, K. W., et al. (1992). Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Res.* 52, 643–645.
- Miller, W. J., Hagemann, S., Reiter, E., and Pinsker, W. (1992). P-element homologous sequences are tandemly repeated in the genome of *Drosophila guanche*. *Proc. Natl. Acad. Sci. U. S. A.* 89, 4018–4022. doi:10.1073/pnas.89.9.4018.
- Mirkovic-Hösle, M., and Förstemann, K. (2014). Transposon defense by endo-siRNAs, piRNAs and somatic piRNAs in *Drosophila*: contributions of Loqs-PD and R2D2. *PLoS One* 9, e84994. doi:10.1371/journal.pone.0084994.
- Mohamed, M., Dang, N. T. M., Ogyama, Y., Bulet, N., Mugat, B., Boulesteix, M., et al. (2020). A Transposon Story: From TE Content to TE Dynamic Invasion of *Drosophila* Genomes Using the Single-Molecule Sequencing Technology from Oxford Nanopore. *Cells* 9, 1–23. doi:10.3390/cells9081776.
- Mohn, F., Handler, D., and Brennecke, J. (2015). piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. *Publ. Sci.* 15, 4–10. doi:10.1126/science.aaa1039.
- Mohn, F., Sienski, G., Handler, D., and Brennecke, J. (2014). The Rhino-Deadlock-Cutoff complex licenses noncanonical transcription of dual-strand piRNA clusters in *Drosophila*. *Cell* 157, 1364–1379. doi:10.1016/j.cell.2014.04.031.
- Moon, S., Cassani, M., Lin, Y. A., Wang, L., Dou, K., and Zhang, Z. Z. (2018). A Robust Transposon-Endogenizing Response from Germline Stem Cells. *Dev. Cell* 47, 660–671.e3. doi:10.1016/j.devcel.2018.10.011.
- Mu, X., Ahmad, S., and Hur, S. (2016). Endogenous Retroelements and the Host Innate Immune Sensors. *Adv. Immunol.* 132, 47–69. doi:10.1016/bs.ai.2016.07.001.
- Mugat, B., Akkouche, A., Serrano, V., Armenise, C., Li, B., Brun, C., et al. (2015). MicroRNA-Dependent Transcriptional Silencing of Transposable Elements in *Drosophila* Follicle Cells. *PLoS Genet.* 11, 1–23. doi:10.1371/journal.pgen.1005194.
- Munir, S., Thierry, S., Subra, F., Deprez, E., and Delelis, O. (2013). Quantitative analysis of the time-course of viral DNA forms during the HIV-1 life cycle. *Retrovirology* 10, 87. doi:10.1186/1742-4690-10-87.
- Navarro-Costa, P., McCarthy, A., Prudêncio, P., Greer, C., Guilgur, L. G., Becker, J. D., et al. (2016). Early programming of the oocyte epigenome temporally controls late prophase I transcription and chromatin remodelling. *Nat. Commun.* 7, 12331. doi:10.1038/ncomms12331.
- Nilson, L. A., and Schüpbach, T. (1999). EGF receptor signaling in *Drosophila* oogenesis. *Curr. Top. Dev. Biol.* 44, 203–243. doi:10.1016/s0070-2153(08)60471-8.
- Norris, J., Fan, D., Aleman, C., Marks, J. R., Futreal, P. A., Wiseman, R. W., et al. (1995). Identification of a new subclass of Alu DNA repeats which can function as estrogen receptor-dependent transcriptional enhancers. *J. Biol. Chem.* 270, 22777–22782. doi:10.1074/jbc.270.39.22777.
- O’Neill, E. M., Rebay, I., Tjian, R., and Rubin, G. M. (1994). The activities of two Ets-related transcription factors required for *Drosophila* eye development are modulated by the Ras/MAPK pathway. *Cell* 78, 137–147. doi:10.1016/0092-8674(94)90580-0.
- Ohnishi, Y., Totoki, Y., Toyoda, A., Watanabe, T., Yamamoto, Y., Tokunaga, K., et al. (2010). Small RNA class transition from siRNA/piRNA to miRNA during pre-implantation mouse development. *Nucleic Acids Res.* 38, 5141–5151. doi:10.1093/nar/gkq229.
- Okamura, K., Chung, W. J., Ruby, J. G., Guo, H., Bartel, D. P., and Lai, E. C. (2008). The *Drosophila* hairpin RNA pathway generates endogenous short interfering RNAs. *Nature* 453, 803–806. doi:10.1038/nature07015.

- Okamura, K., and Lai, E. C. (2008). Endogenous small interfering RNAs in animals. *9*, 673–678. Available at: [www.nature.com/reviews/molcellbio](http://www.nature.com/reviews/molcellbio).
- Olivieri, D., Sykora, M. M., Sachidanandam, R., Mechtler, K., and Brennecke, J. (2010). An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in *Drosophila*. *EMBO J.* *29*, 3301–3317. doi:10.1038/emboj.2010.212.
- Olovnikov, I. a., and Kalmykova, a I. (2013). piRNA clusters as a main source of small RNAs in the animal germline. *Biochem. Biokhimiia* *78*, 572–84. doi:10.1134/S0006297913060035.
- Ozata, D. M., Gainetdinov, I., Zoch, A., O’Carroll, D., and Zamore, P. D. (2019). PIWI-interacting RNAs: small RNAs with big functions. *Nat. Rev. Genet.* *20*, 89–108. doi:10.1038/s41576-018-0073-3.
- Paces, J., Pavlíček, A., and Paces, V. (2002). HERVd: database of human endogenous retroviruses. *Nucleic Acids Res.* *30*, 205–206. doi:10.1093/nar/30.1.205.
- Palmarini, M., Gray, C. A., Carpenter, K., Fan, H., Bazer, F. W., and Spencer, T. E. (2001). Expression of endogenous betaretroviruses in the ovine uterus: effects of neonatal age, estrous cycle, pregnancy, and progesterone. *J. Virol.* *75*, 11319–11327. doi:10.1128/JVI.75.23.11319-11327.2001.
- Pandey, R. R., Homolka, D., Chen, K. M., Sachidanandam, R., Fauvarque, M. O., and Pillai, R. S. (2017). Recruitment of Armitage and Yb to a transcript triggers its phased processing into primary piRNAs in *Drosophila* ovaries. *PLoS Genet.* *13*, 1–20. doi:10.1371/journal.pgen.1006956.
- Pane, A., Jiang, P., Zhao, D. Y., Singh, M., and Schüpbach, T. (2011). The Cutoff protein regulates piRNA cluster expression and piRNA production in the *Drosophila* germline. *EMBO J.* *30*, 4601–4615. doi:10.1038/emboj.2011.334.
- Pardue, M.-L., and DeBaryshe, P. G. (2003). Retrotransposons provide an evolutionarily robust non-telomerase mechanism to maintain telomeres. *Annu. Rev. Genet.* *37*, 485–511. doi:10.1146/annurev.genet.38.072902.093115.
- Parisot, N., Pelin, A., Gasc, C., Polonais, V., Belkorchia, A., Panek, J., et al. (2014). Microsporidian genomes harbor a diverse array of transposable elements that demonstrate an ancestry of horizontal exchange with metazoans. *Genome Biol. Evol.* *6*, 2289–2300. doi:10.1093/gbe/evu178.
- Péllisson, A., Song, S. U., Prud’homme, N., Smith, P. A., Bucheton, A., and Corces, V. G. (1994a). 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*, 4401–4411.
- Péllisson, A., Song, S. U., Prud’homme, N., Smith, P. A., Bucheton, A., and Corces, V. G. (1994b). 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*, 4401–4411. doi:10.1002/j.1460-2075.1994.tb06760.x.
- Perrat, P. N., DasGupta, S., Wang, J., Theurkauf, W., Weng, Z., Rosbash, M., et al. (2013). Transposition-driven genomic heterogeneity in the *Drosophila* brain. *Science* *340*, 91–95. doi:10.1126/science.1231965.
- Picard, G. (1976). Non mendelian female sterility in *Drosophila melanogaster*: hereditary transmission of I factor. *Genetics* *83*, 107–123. doi:10.1093/genetics/83.1.107.
- Polak, P., and Domany, E. (2006). Alu elements contain many binding sites for transcription factors and may play a role in regulation of developmental processes. *BMC Genomics* *7*, 133. doi:10.1186/1471-2164-7-133.
- Preston, C. R., and Engels, W. R. (1989). Spread of P transposable elements in inbred lines of *Drosophila melanogaster*. *Prog. Nucleic Acid Res. Mol. Biol.* *36*, 71–85. doi:10.1016/s0079-6603(08)60162-2.
- Rangan, P., Malone, C. D., Navarro, C., Newbold, S. P., Hayes, P. S., Sachidanandam, R., et al. (2011). PiRNA production requires heterochromatin formation in *Drosophila*. *Curr. Biol.* *21*, 1373–1379. doi:10.1016/j.cub.2011.06.057.
- Raposo, G., and Stoorvogel, W. (2013). Extracellular vesicles: exosomes, microvesicles, and friends. *J. Cell Biol.* *200*, 373–383. doi:10.1083/jcb.201211138.
- Reddien, P. W., Oviedo, N. J., Jennings, J. R., Jenkin, J. C., and Sánchez Alvarado, A. (2005). SMEDWI-2 is a PIWI-like protein that regulates planarian stem cells. *Science* *310*, 1327–1330. doi:10.1126/science.1116110.
- Rehwinkel, J., Natalin, P., Stark, A., Brennecke, J., Cohen, S. M., and Izaurralde, E. (2006). Genome-Wide Analysis of mRNAs Regulated by Drosha and Argonaute Proteins in *Drosophila melanogaster*. *Mol. Cell Biol.* *26*, 2965–2975. doi:10.1128/mcb.26.8.2965-2975.2006.
- Reiss, D., Josse, T., Anxolabéhère, D., and Ronsseray, S. (2004). aubergine mutations in *Drosophila melanogaster* impair P cytotypic determination by telomeric P elements inserted in heterochromatin. *Mol. Genet. Genomics* *272*, 336–343. doi:10.1007/s00438-004-1061-1.
- Reuss, F. U., Frankel, W. N., Moriwaki, K., Shiroishi, T., and Coffin, J. M. (1996). Genetics of intracisternal-A-particle-related envelope-encoding proviral elements in mice. *J. Virol.* *70*, 6450–6454. doi:10.1128/JVI.70.9.6450-6454.1996.
- Ribet, D., Harper, F., Dupressoir, A., Dewannieux, M., Pierron, G., and Heidmann, T. (2008). An infectious

- progenitor for the murine IAP retrotransposon: emergence of an intracellular genetic parasite from an ancient retrovirus. *Genome Res.* 18, 597–609. doi:10.1101/gr.073486.107.
- Robert, V., Prud'homme, N., Kim, A., Bucheton, A., and Pélisson, A. (2001). Characterization of the flamenco region of the *Drosophila melanogaster* genome. *Genetics* 158, 701–713. doi:10.1093/genetics/158.2.701.
- Robinson, H. L., Astrin, S. M., Senior, A. M., and Salazar, F. H. (1981). Host Susceptibility to endogenous viruses: defective, glycoprotein-expressing proviruses interfere with infections. *J. Virol.* 40, 745–751. doi:10.1128/JVI.40.3.745-751.1981.
- Ronsseray, S., Lehmann, M., and Anxolabehere, D. (1991). The maternally inherited regulation of P elements in *Drosophila melanogaster* can be elicited by two P copies at cytological site 1A on the X chromosome. *Genetics* 129, 501–512. doi:10.1093/genetics/129.2.501.
- Ronsseray, S., Lehmann, M., and Anxolabéhère, D. (1989). Copy number and distribution of P and I mobile elements in *Drosophila melanogaster* populations. *Chromosoma* 98, 207–214. doi:10.1007/BF00329685.
- Roovers, E. F., Rosenkranz, D., Mahdipour, M., Han, C.-T., He, N., Chuva de Sousa Lopes, S. M., et al. (2015). Piwi proteins and piRNAs in mammalian oocytes and early embryos. *Cell Rep.* 10, 2069–2082. doi:10.1016/j.celrep.2015.02.062.
- Rubin, G. M., Kidwell, M. G., and Bingham, P. M. (1982). The molecular basis of P-M hybrid dysgenesis: The nature of induced mutations. *Cell* 29, 987–994. doi:10.1016/0092-8674(82)90462-7.
- Saito, K., Ishizu, H., Komai, M., Kotani, H., Kawamura, Y., Nishida, K. M., et al. (2010). Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in *Drosophila*. *Genes Dev.* 24, 2493–2498. doi:10.1101/gad.1989510.
- Sakano, H., Hüppi, K., Heinrich, G., and Tonegawa, S. (1979). Sequences at the somatic recombination sites of immunoglobulin light-chain genes. *Nature* 280, 288–294. doi:10.1038/280288a0.
- Sarot, E., Bucheton, A., and Pe, A. (2004). Evidence for a piwi-Dependent RNA Silencing of the gypsy Endogenous Retrovirus by the *Drosophila melanogaster* flamenco Gene. *Genetics* 1321, 1313–1321.
- Saunders, C., and Cohen, R. S. (1999). The role of oocyte transcription, the 5'UTR, and translation repression and derepression in *Drosophila* gurken mRNA and protein localization. *Mol. Cell* 3, 43–54. doi:10.1016/s1097-2765(00)80173-2.
- Savitsky, M., Kwon, D., Georgiev, P., Kalmykova, A., and Gvozdev, V. (2006). Telomere elongation is under the control of the RNAi-based mechanism in the *Drosophila* germline. *Genes Dev.* 20, 345–354. doi:10.1101/gad.370206.
- Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. *Science* 326, 1112–1115. doi:10.1126/science.1178534.
- Senti, K. A., and Brennecke, J. (2010). The piRNA pathway: A fly's perspective on the guardian of the genome. *Trends Genet.* 26, 499–509. doi:10.1016/j.tig.2010.08.007.
- Senti, K. A., Jurczak, D., Sachidanandam, R., and Brennecke, J. (2015). piRNA-guided slicing of transposon transcripts enforces their transcriptional silencing via specifying the nuclear piRNA repertoire. *Genes Dev.* 29, 1747–1762. doi:10.1101/gad.267252.115.
- Shapiro, J., Machattie, L., Eron, L., Ihler, G., Ippen, K., and Beckwith, J. (1969). Isolation of pure lac operon DNA. *Nature* 224, 768–774. doi:10.1038/224768a0.
- Shpiz, S., Ryazansky, S., Olovnikov, I., Abramov, Y., and Kalmykova, A. (2014). Euchromatic Transposon Insertions Trigger Production of Novel Pi- and Endo-siRNAs at the Target Sites in the *Drosophila* Germline. *PLoS Genet.* 10. doi:10.1371/journal.pgen.1004138.
- Sienski, G., Dönertas, D., and Brennecke, J. (2012). Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell* 151, 964–980. doi:10.1016/j.cell.2012.10.040.
- Sloan, R. D., and Wainberg, M. A. (2011). The role of unintegrated DNA in HIV infection. *Retrovirology* 8, 52. doi:10.1186/1742-4690-8-52.
- Smit, A. F., and Riggs, A. D. (1996). Tiggers and DNA transposon fossils in the human genome. *Proc. Natl. Acad. Sci. U. S. A.* 93, 1443–1448. doi:10.1073/pnas.93.4.1443.
- Sokolova, O. A., Mikhaleva, E. A., Kharitonov, S. L., Abramov, Y. A., Gvozdev, V. A., and Klenov, M. S. (2020). Special vulnerability of somatic niche cells to transposable element activation in *Drosophila* larval ovaries. *Sci. Rep.* 10. doi:10.1038/s41598-020-57901-2.
- Song, J.-J., Smith, S. K., Hannon, G. J., and Joshua-Tor, L. (2004). Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* 305, 1434–1437. doi:10.1126/science.1102514.
- Song, S. U., Gerasimova, T., Kurkulos, M., Boeke, J. D., and Corces, V. G. (1994). An env-like protein encoded by a *Drosophila* retroelement: evidence that gypsy is an infectious retrovirus. *Genes Dev.* 8, 2046–2057. doi:10.1101/gad.8.17.2046.
- Spradling, A. C., and Rubin, G. M. (1982). Transposition of cloned P elements into *Drosophila* germ line chromosomes. *Science* 218, 341–347. doi:10.1126/science.6289435.
- Spradling, A. C., Stern, D. M., Kiss, I., Roote, J., Laverty, T., and Rubin, G. M. (1995). Gene disruptions using P

- transposable elements: an integral component of the *Drosophila* genome project. *Proc. Natl. Acad. Sci. U. S. A.* 92, 10824–10830. doi:10.1073/pnas.92.24.10824.
- Sudarsan, V., Pasalodos-Sanchez, S., Wan, S., Gampel, A., and Skaer, H. (2002). A genetic hierarchy establishes mitogenic signalling and mitotic competence in the renal tubules of *Drosophila*. *Development* 129, 935–944. doi:10.1242/dev.129.4.935.
- Sugimoto, J., Sugimoto, M., Bernstein, H., Jinno, Y., and Schust, D. (2013). A novel human endogenous retroviral protein inhibits cell-cell fusion. *Sci. Rep.* 3, 1462. doi:10.1038/srep01462.
- Sun, Y. H., Zhu, J., Xie, L. H., Li, Z., Meduri, R., Zhu, X., et al. (2020). Ribosomes guide pachytene piRNA formation on long intergenic piRNA precursors. *Nat. Cell Biol.* 22, 200–212. doi:10.1038/s41556-019-0457-4.
- Syomin, B. V., Fedorova, L. I., Surkov, S. A., and Ilyin, Y. V. (2001). The endogenous *Drosophila melanogaster* retrovirus gypsy can propagate in *Drosophila hydei* cells. *Mol. Gen. Genet.* 264, 588–594. doi:10.1007/s004380000344.
- Tabara, H., Sarkissian, M., Kelly, W. G., Fleenor, J., Grishok, A., Timmons, L., et al. (1999). The *rde-1* gene, RNA interference, and transposon silencing in *C. elegans*. *Cell* 99, 123–132. doi:10.1016/s0092-8674(00)81644-x.
- Tam, O. H., Aravin, A. A., Stein, P., Girard, A., Murchison, E. P., Cheloufi, S., et al. (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* 453, 534–538. doi:10.1038/nature06904.
- Tam, O. H., Ostrow, L. W., and Gale Hammell, M. (2019a). Diseases of the nERVous system: Retrotransposon activity in neurodegenerative disease. *Mob. DNA* 10, 1–14. doi:10.1186/s13100-019-0176-1.
- Tam, O. H., Rozhkov, N. V., Shaw, R., Kim, D., Hubbard, I., Fennessey, S., et al. (2019b). Postmortem Cortex Samples Identify Distinct Molecular Subtypes of ALS: Retrotransposon Activation, Oxidative Stress, and Activated Glia. *Cell Rep.* 29, 1164–1177.e5. doi:10.1016/j.celrep.2019.09.066.
- Tcheressiz, S., Calco, V., Arnaud, F., Arthaud, L., Dastugue, B., and Vaury, C. (2002). Expression of the Idefix retrotransposon in early follicle cells in the germarium of *Drosophila melanogaster* is determined by its LTR sequences and a specific genomic context. *Mol. Genet. Genomics* 267, 133–141. doi:10.1007/s00438-002-0641-1.
- Théron, E., Dennis, C., Brasset, E., and Vaury, C. (2014). Distinct features of the piRNA pathway in somatic and germ cells: From piRNA cluster transcription to piRNA processing and amplification. *Mob. DNA* 5, 1–11. doi:10.1186/s13100-014-0028-y.
- Thomas, J., and Pritham, E. J. (2015). Helitrons, the Eukaryotic Rolling-circle Transposable Elements. *Microbiol. Spectr.* 3. doi:10.1128/microbiolspec.MDNA3-0049-2014.
- Tian, A.-G., Tamori, Y., Huang, Y.-C., Melendez, N. T., and Deng, W.-M. (2013). Efficient EGFR signaling and dorsal-ventral axis patterning requires syntaxin dependent Gurken trafficking. *Dev. Biol.* 373, 349–358. doi:10.1016/j.ydbio.2012.10.029.
- Tkach, M., and Théry, C. (2016). Communication by Extracellular Vesicles: Where We Are and Where We Need to Go. *Cell* 164, 1226–1232. doi:10.1016/j.cell.2016.01.043.
- Tsichlis, P. N., and Coffin, J. M. (1980). Recombinants between endogenous and exogenous avian tumor viruses: role of the C region and other portions of the genome in the control of replication and transformation. *J. Virol.* 33, 238–249. doi:10.1128/JVI.33.1.238-249.1980.
- Vachias, C., Fritsch, C., Pouchin, P., Bardot, O., and Mirouse, V. (2014). Tight coordination of growth and differentiation between germline and soma provides robustness for *Drosophila* egg development. *Cell Rep.* 9, 531–541. doi:10.1016/j.celrep.2014.09.035.
- Vagin, V. V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., and Zamore, P. D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* 313, 320–324. doi:10.1126/science.1129333.
- Vagin, V. V., Yu, Y., Jankowska, A., Luo, Y., Wasik, K. A., Malone, C. D., et al. (2013). Minotaur is critical for primary piRNA biogenesis. *RNA* 19, 1064–1077. doi:10.1261/rna.039669.113.
- Valadi, H., Ekström, K., Bossios, A., Sjöstrand, M., Lee, J. J., and Lötvall, J. O. (2007). Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat. Cell Biol.* 9, 654–659. doi:10.1038/ncb1596.
- Wang, L., Dou, K., Moon, S., Tan, F. J., and Zhang, Z. Z. (2018). Hijacking Oogenesis Enables Massive Propagation of LINE and Retroviral Transposons. *Cell* 174, 1082–1094.e12. doi:10.1016/j.cell.2018.06.040.
- Watanabe, T., Takeda, A., Tsukiyama, T., Mise, K., Okuno, T., Sasaki, H., et al. (2006). Identification and characterization of two novel classes of small RNAs in the mouse germline: Retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes Dev.* 20, 1732–1743. doi:10.1101/gad.1425706.
- Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., et al. (2008).

- Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* 453, 539–543. doi:10.1038/nature06908.
- Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., et al. (2007). A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8, 973–982. doi:10.1038/nrg2165.
- Wierzbicki, F., Kofler, R., and Signor, S. (2021). Evolutionary dynamics of piRNA clusters in *Drosophila*. *Mol. Ecol.*, 1–17. doi:10.1111/mec.16311.
- Wilson, J. E., Connell, J. E., and Macdonald, P. M. (1996). Aubergine enhances oskar translation in the *Drosophila* ovary. *Development* 122, 1631–1639. doi:10.1242/dev.122.5.1631.
- Yamashita, M., and Engelman, A. N. (2017). Capsid-Dependent Host Factors in HIV-1 Infection. *Trends Microbiol.* 25, 741–755. doi:10.1016/j.tim.2017.04.004.
- Yan, K. S., Yan, S., Farooq, A., Han, A., Zeng, L., and Zhou, M.-M. (2003). Structure and conserved RNA binding of the PAZ domain. *Nature* 426, 468–474. doi:10.1038/nature02129.
- Yan, Z., Hu, H. Y., Jiang, X., Maierhofer, V., Neb, E., He, L., et al. (2011). Widespread expression of piRNA-like molecules in somatic tissues. *Nucleic Acids Res.* 39, 6596–6607. doi:10.1093/nar/gkr298.
- Yoth, M., Jensen, S., and Brasset, E. (2022). The Intricate Evolutionary Balance between Transposable Elements and Their Host: Who Will Kick at Goal and Convert the Next Try? *Biology (Basel)*. 11. doi:10.3390/biology11050710.
- Zanni, V., Eymery, A., Coiffet, M., Zytnicki, M., Luyten, I., Quesneville, H., et al. (2013). Distribution, evolution, and diversity of retrotransposons at the flamenco locus reflect the regulatory properties of piRNA clusters. *Proc. Natl. Acad. Sci.* 110, 19842–19847. doi:10.1073/pnas.1313677110.
- Zhang, F., Wang, J., Xu, J., Zhang, Z., Koppetsch, B. S., Schultz, N., et al. (2012). UAP56 couples piRNA clusters to the perinuclear transposon silencing machinery. *Cell* 151, 871–884. doi:10.1016/j.cell.2012.09.040.
- Zhang, G., Tu, S., Yu, T., Zhang, X. O., Parhad, S. S., Weng, Z., et al. (2018). Co-dependent Assembly of *Drosophila* piRNA Precursor Complexes and piRNA Cluster Heterochromatin. *Cell Rep.* 24, 3413–3422.e4. doi:10.1016/j.celrep.2018.08.081.
- Zhang, Z., Wang, J., Schultz, N., Zhang, F., Parhad, S. S., Tu, S., et al. (2014). The HP1 homolog Rhino anchors a nuclear complex that suppresses piRNA precursor splicing. *Cell* 157, 1353–1363. doi:10.1016/j.cell.2014.04.030.
- Zhou, F., Li, M., Wei, Y., Lin, K., Lu, Y., Shen, J., et al. (2016). Activation of HERV-K Env protein is essential for tumorigenesis and metastasis of breast cancer cells. *Oncotarget* 7, 84093–84117. doi:10.18632/oncotarget.11455.
- Zila, V., Margiotta, E., Turoňová, B., Müller, T. G., Zimmerli, C. E., Mattei, S., et al. (2021). Cone-shaped HIV-1 capsids are transported through intact nuclear pores. *Cell* 184, 1032–1046.e18. doi:10.1016/j.cell.2021.01.025.
- Zong, L., Zhu, Y., Liang, R., and Zhao, H. B. (2016). Gap junction mediated miRNA intercellular transfer and gene regulation: A novel mechanism for intercellular genetic communication. *Sci. Rep.* 6, 1–9. doi:10.1038/srep19884.
- Zugasti, O., Tavignot, R., and Royet, J. (2020). Gut bacteria-derived peptidoglycan induces a metabolic syndrome-like phenotype via NF- $\kappa$ B-dependent insulin/PI3K signaling reduction in *Drosophila* renal system. *Sci. Rep.* 10, 14097. doi:10.1038/s41598-020-70455-7.
- Zytnicki, M., Coiffet, M., Zanni, V., Quesneville, H., Luyten, I., Jensen, S., et al. (2013). Distribution, evolution, and diversity of retrotransposons at the flamenco locus reflect the regulatory properties of piRNA clusters. *Proc. Natl. Acad. Sci.* 110, 19842–19847. doi:10.1073/pnas.1313677110.