N° d'ordre

Année 2012

THESE DE L'UNIVERSITE DE LYON

Délivrée par

L'UNIVERSITE CLAUDE BERNARD LYON 1

ECOLE DOCTORALE E2M2

DIPLOME DE DOCTORAT

(Arrêté du 7 août 2006)

Soutenue publiquement le 13 avril 2012

Par

Abdou AKKOUCHE

TITRE :

Régulation épigénétique d'un rétrovirus endogène, *tirant*, dans la lignée germinale de la drosophile

Directrice de thèse : Cristina VIEIRA-HEDDI

JURY :

Madame Dominique MOUCHIROUD	Présidente
Monsieur Pierre CAPY	Rapporteur
Monsieur Alain PELISSON	Rapporteur
Madame Laure TEYSSET	Examinatrice
Madame Cristina VIEIRA	Directrice
Madame Marie FABLET	Co-directrice

Laboratoire de Biométrie et Biologie Evolutive
UMR CNRS 5558, Université Claude Bernard Lyon1
43 bd du 11 novembre 1918-69622 Villeurbanne

UNIVERSITE CLAUDE BERNARD - LYON 1

Président de l'Université	**M. A. Bonmartin**
Vice-président du Conseil d'Administration	M. le Professeur G. Annat
Vice-président du Conseil des Etudes et de la Vie Universitaire	M. le Professeur D. Simon
Vice-président du Conseil Scientifique	M. le Professeur J-F. Mornex
Secrétaire Général	M. G. Gay

COMPOSANTES SANTE

Faculté de Médecine Lyon Est – Claude Bernard	Directeur : M. le Professeur J. Etienne
Faculté de Médecine et de Maïeutique Lyon Sud – Charles Mérieux	Directeur : M. le Professeur F-N. Gilly
UFR d'Odontologie	Directeur : M. le Professeur D. Bourgeois
Institut des Sciences Pharmaceutiques et Biologiques	Directeur : M. le Professeur F. Locher
Institut des Sciences et Techniques de la Réadaptation	Directeur : M. le Professeur Y. Matillon
Département de formation et Centre de Recherche en Biologie Humaine	Directeur : M. le Professeur P. Farge

COMPOSANTES ET DEPARTEMENTS DE SCIENCES ET TECHNOLOGIE

Faculté des Sciences et Technologies	Directeur : M. le Professeur F. De Marchi
Département Biologie	Directeur : M. le Professeur F. Fleury
Département Chimie Biochimie	Directeur : Mme le Professeur H. Parrot
Département GEP	Directeur : M. N. Siauve
Département Informatique	Directeur : M. le Professeur S. Akkouche
Département Mathématiques	Directeur : M. le Professeur A. Goldman
Département Mécanique	Directeur : M. le Professeur H. Ben Hadid
Département Physique	Directeur : Mme S. Fleck
Département Sciences de la Terre	Directeur : Mme le Professeur I. Daniel
UFR Sciences et Techniques des Activités Physiques et Sportives	Directeur : M. C. Collignon
	Directeur : M. B. Guiderdoni
Observatoire de Lyon	Directeur : M. P. Fournier
Ecole Polytechnique Universitaire de Lyon 1	Directeur : M. G. Pignault
Ecole Supérieure de Chimie Physique Electronique	Directeur : M. le Professeur C. Coulet
Institut Universitaire de Technologie de Lyon 1	Directeur : M. le Professeur J-C. Augros
Institut de Science Financière et d'Assurances	Directeur : M. R. Bernard
Institut Universitaire de Formation des Maîtres	

À mes parents

Remerciements

Je tiens d'abord à remercier en premier lieu Alain Pelisson et Pierre Capy d'avoir accepté la responsabilité de rapporter cette thèse. Je remercie également Laure Teysset d'avoir accepté de faire partie des membres du jury.

J'adresse ensuite un immense merci à Cristina Vieira pour m'avoir encadrée tout au long de cette thèse. Merci aussi pour sa disponibilité et sa gentillesse, elle a toujours eu des gestes d'amitié envers moi et je l'en remercie sincèrement.

Mes remerciements vont ensuite à Marie Fablet, merci de m'avoir consacré du temps pour enrichir la réflexion sur mon sujet de thèse, et ce malgré son emploi de temps ultra chargé.

J'exprime également toute ma reconnaissance à Dominique Mouchiroud pour m'avoir accueillie au sein de l'UMR de Biométrie et Biologie Evolutive.

Je remercie ensuite toutes les personnes qui m'ont aidé au quotidien dans la réalisation de ce travail, par leur aide technique et leurs conseils précieux et plus particulièrement : Nelly et Héléne.

Je tiens aussi à remercier particulièrement les membres de l'équipe TREEP ; Matthieu, Annabelle, Elias, Emmanuelle, Christian, Virginie, Laurent,

Ramy Zahran, Marwan Chamoun, Nadim Khazaka, AL Mahdy Hamieh pour avoir endossé le rôle de frères toujours prêts à m'aider quand j'avais besoin d'eux, et pour m'avoir encouragé dans les moments difficiles.

Et parce qu'il n'y a pas que la vie au labo pendant la thése, je remercie tous mes amis pour la décompression du samedi soir. John, Johny, Laure, Marwan, Yendy, Joseph, Marie Claire; Pauline, Wissam; Marion, Amer, Charbel, Chainez, Amanda.

A Samir et Françoise, qui m'ont accueilli comme un des leurs.

Cette thèse, j e la dédie à ma famille qui a lutté pour me soutenir pendant ces 5 ans de longue séparation. A ma mère, mon père, mes frères et ma sœur de qui je suis encore plus fier qu'ils ne le sont de moi. Pour leur courage et leur amour, Je veux qu'ils sachent qu'au fond de moi je ne cesse de me répéter que je suis l'homme le plus chanceux du monde d'être leur fils et leur frère.

RESUME

Une grande partie du génome des eucaryotes est constituée d'éléments transposables (ET). Ces séquences d'ADN répétées ont la capacité de se déplacer d'un site chromosomique à un autre et de multiplier le nombre de leurs copies, pouvant ainsi être la cause d'une instabilité génétique. Face à ce potentiel de mutagénèse, un certain nombre de systèmes ont été sélectionnés dans les génomes eucaryotes qui conduisent à une réduction de l'activité des ET. Dans la première partie de cette thèse, j'ai caractérisé un élément *tirant* actif dans les populations naturelles de *D. simulans*. Je me suis intéressé en particulier au gène de la protéine d'enveloppe (*env*), qui confère le caractère infectieux du rétrovirus. Dans la deuxième partie de mon travail, je me suis intéressé à l'étude de l'influence de *tirant* sur la structure de la chromatine au niveau de son site d'insertion et à son influence sur l'expression des gènes voisins. J'ai étudié trois modifications d'histones dans trois populations naturelles, dont une où *tirant* est inséré dans un intron du gène *tkv*. Enfin, je me suis intéressé à la régulation post-transcriptionnelle de *tirant* par les piARN.

TITLE: Epigenetic regulation of the *tirant* endogenous retrovirus in Drosophila germline.

ABSTRACT

Eukaryotic genomes harbor a wide variety of repeated sequences, such as transposable elements (TE). These sequences are able to move from one chromosomal site to another, to multiply their number of copies, and can be the cause of a genetic instability. Sophisticated genomic defenses have evolved to restrict their activity. In Drosophila, epigenetic modification such as post-translational histone modifications and RNAi interference are involved in TE silencing in reproductive tissues. The silencing of an LTR - like element, *tirant*, has been deeply analyzed in this work. Here, I first describe an active *tirant* element in natural populations of *D. simulans*. I have focused on the envelope protein gene (*env*), which confers the infectious behavior to the retrovirus. I then focused on the effects of TE insertions on chromatin structure and on its influence on the expression of the nearby genes. I studied three histone modification marks in three natural populations, in the locus in which *tirant* was inserted. Finally, I was interested in the post-transcriptional regulation of *tirant* involving the piRNA pathway.

DISCIPLINE
E2M2 : Evolution, Écosystèmes, Microbiologie, Modélisation

MOTS-CLES
Éléments transposables, épigénétique, rétrovirus endogène, modifications des histones, piARN**,** *Drosophila simulans*, populations naturelles

INTITULE ET ADRESSE DE L'U.F.R. OU DU LABORATOIRE
Laboratoire de Biométrie et Biologie Évolutive
UMR CNRS 5558, Université Claude Bernard Lyon1
43 bd du 11 novembre 1918-69622 Villeurbanne

Résumé :

Une grande partie du génome des eucaryotes est constituée d'éléments transposables (ET). Ces séquences d'ADN répétées ont la capacité de se déplacer d'un site chromosomique à un autre et de multiplier le nombre de leurs copies, pouvant ainsi être la cause d'une instabilité génétique. Face à ce potentiel de mutagénèse, un certain nombre de systèmes ont été sélectionnés dans les génomes eucaryotes qui conduisent à une réduction de l'activité des ET. Notamment, chez la drosophile, on a récemment mis en évidence des mécanismes de régulation impliquant les modifications d'histones, et une nouvelle classe de petits ARN, appelés piARN, qui contrôlent spécifiquement les éléments transposables dans les tissus reproducteurs.

tirant est un rétrotransposon à LTR de la drosophile, de type *Gypsy*, isolé au sein du laboratoire dans les populations naturelles de *D. simulans,* où le nombre de ses copies est variable entre populations. Cet ET possède la même structure génomique que les rétrovirus. Dans la première partie de cette thèse, j'ai caractérisé un élément *tirant* actif dans les populations naturelles de *D. simulans*. Je me suis intéressé en particulier au gène de la protéine d'enveloppe (*env*), qui confère le caractère infectieux du rétrovirus. La comparaison des transcrits et de la protéine du gène *env* entre populations de *D. simulans* a montré que *tirant* est actif dans une population, et cette activation est associée à sa mobilisation, alors que dans les autres populations *tirant* est présent, mais régulé.

Dans la deuxième partie de mon travail, je me suis intéressé à l'étude de l'influence de *tirant* sur la structure de la chromatine au niveau de son site d'insertion et à son influence sur l'expression des gènes voisins. J'ai étudié trois modifications d'histones dans trois populations naturelles, dont une où *tirant* est inséré dans un intron du gène *tkv*. Les résultats obtenus montrent que *tirant* est capable de modifier la structure de la chromatine au niveau de son site d'insertion, mais aussi en amont, par l'hétérochromatinisation d'un promoteur du gène *tkv,* en affectant ainsi son taux de transcription.

Enfin, je me suis intéressé à la régulation post-transcriptionnelle de *tirant* par les piARN. Par l'analyse de croisements intraspécifiques entre des souches contenant ou non des copies de *tirant* dans l'euchromatine, j'ai montré qu'une régulation post-transcriptionnelle par les piARN germinaux qui contrôle *tirant* dans les cellules folliculaires de l'ovaire. J'ai aussi pu montrer une expression variable entre populations des gènes de la voie piARN.

Abstract:

Eukaryotic genomes harbor a wide variety of repeated sequences, such as transposable elements (TE). These sequences are able to move from one chromosomal site to another, to multiply their number of copies, and can be the cause of a genetic instability. Sophisticated genomic defenses have evolved to restrict their activity. In Drosophila, epigenetic modification such as post-translational histone modifications and RNAi interference are involved in TE silencing in reproductive tissues. The silencing of an LTR like element, *tirant*, has been deeply analyzed in this work. *Tirant* is a *Gypsy* like element, isolated in our laboratory in natural populations of *D. simulans,* in which a high level of copy number variability is observed between strains.

Here, I first describe an active *tirant* element in natural populations of *D. simulans*. I have focused on the envelope protein gene (*env*), which confers the infectious behavior to the retrovirus. By comparison of *tirant* transcripts level and protein localization between natural populations of *D.simulans*, I showed that *tirant* is active in one population, and this activation is correlated with its mobilization.

I then focused on the effects of TE insertions on chromatin structure and in its influence on the expression of the nearby genes. I studied three histone modification marks in three natural populations, in the locus in which *tirant* was inserted. I show that *tirant* is associated with repressive marks and active marks, which explains the activity of the element. We also showed that *tirant* modifies the structure of the chromatin at the level of its site of insertion, but also upstream, by the heterochromatinization of the promoter of tkv gene, interfering with the level of transcription of the gene.

Finally, I was interested in the post-transcriptional regulation of *tirant* involving the piRNA pathway. By crossing *D.simulans* strains which contains different copy numbers of the *tirant* element, I showed that *tirant* is regulated in the follicular cells by the germ line piRNA pathway. I was also able to show a variable expression between populations of the proteins of the piRNA pathway.

Table des matières

1 Rappels bibliographiques 1

1.1 Les éléments transposables et les rétrovirus 1

1.2 Classification des éléments transposables 4

1.2.1 Les éléments à ARN (Classe I) 4

1.2.1.1 Les éléments à LTR et les rétrovirus endogènes (ou ERV pour « endogenous retrovirus ») 5

1.2.1.2 Les éléments sans LTR 6

1.2.2 Les éléments à ADN (classe II) 7

1.2.3 Autres 7

1.3 Cycle de rétrotransposition des éléments à LTR 8

1.4 Répartition des éléments transposables dans les génomes. 9

1.4.1 Un exemple de RDE chez Drosophila melanogaster : le locus flam/COM . 10

1.5 Relations avec le génome 11

1.5.1 Implication des Éléments transposables dans l'Évolution des génomes 11

1.6 Régulation épigénétique des éléments transposables 14

1.6.1 Méthylation de l'ADN 15

1.6.2 Modifications d'histones 16

1.6.3 Modifications post-traductionnelles des histones 17

1.6.3.1 Acétylation des histones 18

1.6.3.2 La méthylation des histones 18

1.6.4 Territoires chromatiniens 18

1.6.4.1 L'euchromatine 18

1.6.4.2 L'hétérochromatine 19

1.6.5 Historique de l'ARN interférence 20

1.6.5.1 Les microARN (miARN) 21

1.6.5.2 Les small interfering ARN (siARN) 22

1.6.5.3 Les Piwi-interacting RNA (piARN) 24

1.7 La drosophile comme modèle génétique. 40

1.7.1 Quelques modèles d'instabilité 41

1.7.1.1 Le facteur I 41

1.7.1.2 Les rétrovirus endogènes Gypsy, Idefix et ZAM 42

1.7.2 Les populations naturelles de drosophiles pour comprendre la dynamique des ET 44

1.7.2.1 Variabilité en ET dans les populations naturelles de drosophiles 44

1.7.2.2 Modèle d'étude Tirant 46

2 Résultats & discussion 49

2.1 *tirant*, un nouveau rétrovirus endogène actif chez *D. simulans* 50

2.1.1 Discussion 70

2.1.1.1 L'élément tirant : un modèle d'étude de prolifération de rétrovirus endogènes dans les populations naturelles de D. simulans. 71

2.1.1.2 L'expression de tirant dépend de la population. 72

2.2 Analyse des modifications d'histones associées à des insertions de *tirant* 74

2.2.1 Discussion 89

2.2.1.1 Modification d'histones associées au rétrovirus endogène tirant. ... 89

2.2.1.2 Influence de tirant sur la chromatine des gènes voisins. 90

2.3 Les piARNs germinaux déposés maternellement régulent le rétrovirus endogène *tirant* dans les cellules somatiques 91

2.3.1 Dicussion 105

2.3.1.1 Les piARN sont associés avec l'accumulation nucléaire des transcrits de tirant dans la lignée germinale femelle. 105

2.3.1.2 Les piARN déposé maternellement sont nécessaires à la régulation de tirant dans les cellules somatiques 106

2.3.1.3 Interaction entre les deux voies de piARN somatique et germinale 108

2.3.1.4 Rôle des piARN dans la mise en place du TGS 109

2.4 Polymorphisme des gènes de la voie piARN 112

2.4.1 Taux de transcrits des ET dans les populations naturelles de D. simulans et D. melanogaster 113

2.4.2 Polymorphisme de séquence et taux de transcrits des gènes de la voie piARN 116

3 Perspectives 119

Conclusion 123

Matériels et méthodes 124

Annexes 141

Publication 145

Bibliographie 153

Liste des figures

Figure 1. Proportion d'éléments transposables dans dix génomes eucaryotes. 2

Figure 2: Structure schématique des deux grandes familles d'éléments transposables. 3

Figure 3: Schéma simplifié des deux modes de transposition des ET. 4

Figure 4. Structure génomique des virus à ARN rétrotranscrit. 6

Figure 5. Cycle de rétrotransposition des rétrotransposons à LTR selon Sabot et Schulman 2006 8

Figure 6 : Distribution des éléments transposables le long des chromosomes de *D. melanogaster*. 9

Figure 7 : Principales modifications post-traductionnelles des histones. 17

Figure 8 : voie de miARN chez *Drosophila melanogaster*. 21

Figure 9 : Voie des siARN chez *D. melanogaster*. 24

Figure 10: Arbre phylogénétique des protéines PIWI chez la drosophile, la souris (mouse), le xénope et le poisson zèbre. D'après Saito & Siomi, 2010. 25

Figure 11. Localisation des protéines Argonautes de la sous classe PIWI dans les ovaires de drosophile. L'ADN est maqué en bleu, les protéines Argonautes en vert. D'après Brennecke et al, 2007. 27

Figure 12 : le biais relatif de composition en nucléotides pour chaque position dans tous les piARN obtenus à partir de complexes PIWI, AUB et AGO3. D'après Brennecke et al, 2007. 28

Figure 13: les locus unidirectionnels et bidirectionnels de production des piARN. 29

Figure 14: Mécanisme de biogenèse des piARN. D'après (Brennecke et al, 2007) ; (Gunawardane et al, 2007). 32

Figure 15: L'héritage maternel de la famille de protéines PIWI. ADN (bleu), Aubergine, Piwi et AGO3 (vert). D'après (Brennecke et al, 2008). 34

Figure 16 : Les piARN intervenant comme vecteur d'information épigénétique. 35

Figure 17: La biogénèse des piARN prend place dans des granules germinales. 39

Figure 18: Représentation schématique d'une paire d'ovaires et d'un ovariole. (A) D'après Huston et al, 2008 ; (B) D'après King 1970. 40

Figure 19: Mise en évidence de la répression post-transcriptionnelle du facteur *I* par la voie des piARN. 42

Figure 20: Détection des transcrits et de piARN antisens homologue aux transcrits de *Gypsy*, *ZAM* et *Idefix* 44

Figure 21: Structure du rétrotransposon à LTR *tirant* 47

Figure 22: Hybridation *in-situ* de l'ORF3 (env) du rétrovirus endogène *tirant* (rouge) dans les descendants des croisements A (Mâle Mayotte x Femelle Chicharo) et B (Mâle Chicharo x Femelle Mayotte). 107

Figure 23: Modèle de transmission des piARN germinaux vers les cellules somatiques 109

Figure 24: Quantification des transcrits des 4 ET dans les populations naturelles en noir : *D. simulans*, et en rouge *D. melanogaster*. Quantifications des transcrits de 4 ET (*412, tirant*,

Roo et *F*) par RT-PCR quantitative. La quantité de transcrits dans les ovaires est rapportée à la quantité de transcrits du gène rp49. 114

Figure 25: Mise en évidence de la présence de transcrits de l'élément *412* et *F* dans les ovaires par hybridation *in situ.* 115

Figure 26: Quantification des transcrits des gènes de la voie piARN dans les populations naturelles en noir : *D. simulans*, et en rouge *D. melanogaster.* 117

Figure 27: Schéma représentant le modèle expérimentale de PTT 135

Figure 28: Schéma du plan expérimental de la ChIP. 139

Figure 29: Analyse de l'expression de *tirant* par hybridation *in-situ* sur ovaires dans les populations naturelles de *D. melanogaster.* 142

Liste des tableaux

Tableau 1 : Mutants de la voie piARN et leurs effets. 36

Tableau 2. Variation des nombres de copies euchromatiques de l'élément *tirant* dans les populations naturelles. D'après Vieira et al, 2009. 46

Tableau 3: Valeurs de diversité nucléotidique (π). 116

Tableau 4: Quantités de transcrits détectés par RT-qPCR dans les ovaires (rapportées à la quantité de transcrits pour le gène de référence rp49)..... 117

Tableau 5: Liste des couples d'amorces utilisés en RT-PCR quantitative pour quantifier les transcrits des gènes de la voie piARN. 143

Tableau 6 : Données de charge en éléments transposables et de quantité de transcrits pour les gènes de la voie piARN, déterminées pour 12 populations naturelles de *D. simulans*. 144

Liste des abréviations

ADN : Acide désoxyribonucléique

ADNc : Acide désoxyribonucléique complémentaire

ARN: Acide ribonucléique

ARNdb: ARN double brin

ARNi: interference ARN

ARNm: Acide ribonucléique messager

ARNr: Acide ribonucléique ribosomal

ARNsb : ARN simple brin

ARNt : acide ribonucléique de transfert

ATP : Adénosine Tri-phosphate

°C : degré Celsius

ChIP : Chromatin Immuno-Precipitation

COM : Centre Organisateur de la Mobilisation

CTP : Cytosine Tri-Phosphate

DIG : Digoxigénine

dNTP : désoxy Nucléotide Tri-Phosphate

dnmt : DNA Méthyl Transférase

ET : Elément transposable

flam : locus flamenco

H3K4 : Lysine 4 de l'histone H3

H3K9 : Lysine 9 de l'histone H3

H3K27: Lysine 27 de l'histone H3

HAT: Histone Acétyl Transferase

HMT: Histone Methym Transferase

HP1: Heterochromatin protein 1/ protein Hétérochromatique 1

IN: Integrase

Kb: Kilo-bases

LINE: Long-Interspersed Nucleotidic Element

LTR: Long Terminal Repeat

miARN: micro-ARN

Miwi: Mouse Piwi

ORF: Open Reading Frame/ Cadre Ouvert de Lecture

Pb: paire de bases

PBS : Phosphate Buffer Saline

PCR : Polymerase Chain Reaction/Réaction de Polymerization en Chaine

piRNA : piwi-interacting RNA/ARN interagissant avec Piwi

piwi: « P element-Induced Wimpy testis »

polll. ARN polymerase III

PR: Protéase

PTGS : Post Trascriptional Gene Silencing/Silencing de Géne Post-Transcriptonnel

RDE : Régions Denses en Elements transposables

RH : Rnase H

RISC : RNA Induced Silencing Complex/ Complexe de Silencing Induit par ARN

RNAseIII: Ribonucléase III

RT: Rétrotranscriptase

scnRNA: scanRNA/petits ARNs interférants

siRNA : small-interfering RNA/petits ARNs interférants

SINE: Short Interspersed Nucleotidic Element

SU(VAR): Supressor of Variégation

TGS: Transcriptional Gene Silencing/Silencing de gène Transcriptinnel

Introduction générale

Introduction générale

Il est maintenant bien admis que le génome des êtres vivants est constitué d'une part importante de séquences moyennement répétées qui peuvent se déplacer sur les chromosomes. Ces séquences, appelées éléments transposables (ET), ont sérieusement bouleversé nos conceptions sur la stabilité du génome. Ces éléments mobiles ont été découverts par Barbara McClintock dans les années 1950, lors d'études concernant des phénomènes d'instabilité phénotypiques chez le maïs. Cette découverte a été couronnée par l'attribution du prix Nobel en Médecine en 1983.

La diversité des éléments transposables est extrêmement importante. Cette diversité est non seulement liée à leur structure, mais aussi à leur système de transposition, qui peut être à l'origine de différents types de copies lors de la transposition. Le nombre de copies de ces éléments au sein des génomes est très variable, et l'ensemble des éléments répétés compose souvent une part importante des génomes eucaryotes.

Chez la drosophile, les ET constituent environ 15% du génome et seraient responsables de 50 à 85% des mutations spontanées observées. Chez l'homme, ils représentent plus de 45% du génome mais seraient responsables de seulement 1 à 2% des mutations. La différence du nombre de mutations dues à ces éléments entre l'homme et la drosophile montre clairement une mise en place de mécanismes s'opposant à leur pouvoir envahissant et mutagène, dont l'efficacité semble être différente selon les espèces, et éventuellement liée à la charge globale en ET. La perte de ces mécanismes de contrôle génère une instabilité génétique importante, et la réactivation de nombreux éléments transposables.

L'objet de ce manuscrit est de comprendre les éléments de l'immunité intrinsèque mis en place par l'hôte pour lutter contre la mobilisation des ET. D'une façon générale, ces éléments sont peu actifs car ils échappent rarement aux mécanismes de défense de l'hôte. Par contre, il existe des modèles expérimentaux où ces mécanismes sont altérés, permettant ainsi l'expression et la mobilisation de ces éléments. Notre niveau de connaissance dans ce domaine est en constante progression. De nombreux travaux ont montré le rôle des mécanismes épigénétiques dans la régulation des ET, tant au niveau de la méthylation de cytosines dans la molécule d'ADN, qu'à celui des modifications des histones. Plus récemment des travaux ont mis en évidence, chez la drosophile, un mécanisme d'ARN interférence permettant d'éliminer les transcrits des ET. La drosophile se présente alors comme un bon modèle pour l'étude des mécanismes de contrôle épigénétique des ET, en raison de sa génétique bien connue.

1 Rappels bibliographiques

Un élément transposable (ET) est traditionnellement décrit comme étant un fragment d'ADN possédant la particularité d'être mobile et de se répliquer. Ces caractéristiques en font des composants atypiques des génomes. Leur découverte a été l'origine de fructueuses réflexions théoriques en biologie, de la régulation de l'expression des gènes jusqu'au niveau auquel la sélection naturelle est supposée agir. Pour autant, leur dynamique et leurs impacts sur les génomes ne sont toujours pas bien compris. Dans cette introduction je décris quelques points qui semblent importants pour la recherche sur les ET en mettant en relief les aspects plus proches de mon projet de recherche, tels que l'impact des ET sur le génome hôte et les systèmes de régulation d'ET. Ceci me permettra de poser le cadre conceptuel sur lequel repose mon travail.

1.1 Les éléments transposables et les rétrovirus

En 1944, lors d'un croisement visant à révéler la composition génique du bras court du chromosome 9 du maïs, Barbara McClintock déclencha ce que certains appelleront un « tremblement de terre génétique » (Jones, 2005). Dans les années qui suivirent, elle décrivit précisément les réarrangements chromosomiques observés à l'issue de ce fameux croisement. Plus intriguant, elle découvrit des locus mutables ne restant pas toujours à la même position le long des chromosomes. Elle parvint à les caractériser, distinguant deux classes (McClintock, 1950)

> "(1) those that require a separate activator factor for instability to be expressed and
>
> (2) those that are autonomous with respect to the factor that controls the onset of mutability."

Elle découvrit ainsi les premiers ET qu'elle nomma Ac pour « Activator » et Ds pour « Dissociation ». Poursuivant ses investigations, McClintock proposa alors dès 1956 de décrire ces locus mutables comme étant des éléments contrôlant l'action des gènes (McClintock, 1956).

"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)."

Dans cette dernière publication, elle mentionne également l'impact des éléments sur le contrôle des gènes localisés à une certaine distance de part et d'autre, *via* un effet de diffusion (« spreading »), ce qui serait maintenant interprété comme étant de l'hétérochromatinisation.

Aujourd'hui les éléments transposables sont largement étudiés vu leur ubiquité dans les génomes, mais aussi à cause de leur implication importante dans l'évolution des génomes hôtes (Biemont, 2010). Ces composants du génome ont été retrouvés dans presque toutes les espèces étudiées et leur proportion dans le génome est très variable. Ils peuvent représenter

peu de copies dans les génomes des bactéries ou composer plus de 90% du génome de certaines plantes.

La figure 1 présente un exemple de variation du pourcentage d'ET dans les génomes de certains eucaryotes. On peut voir que la proportion des ET chez les primates (homme, chimpanzé) est quasiment identique, autour de 45%. Cette proportion peut cependant varier entre deux espèces très proches comme c'est le cas entre *D. melanogaster*, qui a 15% d'ET, alors qu'une espèce sœur, *D. simulans,* n'en a que 5%. (Figure 1).

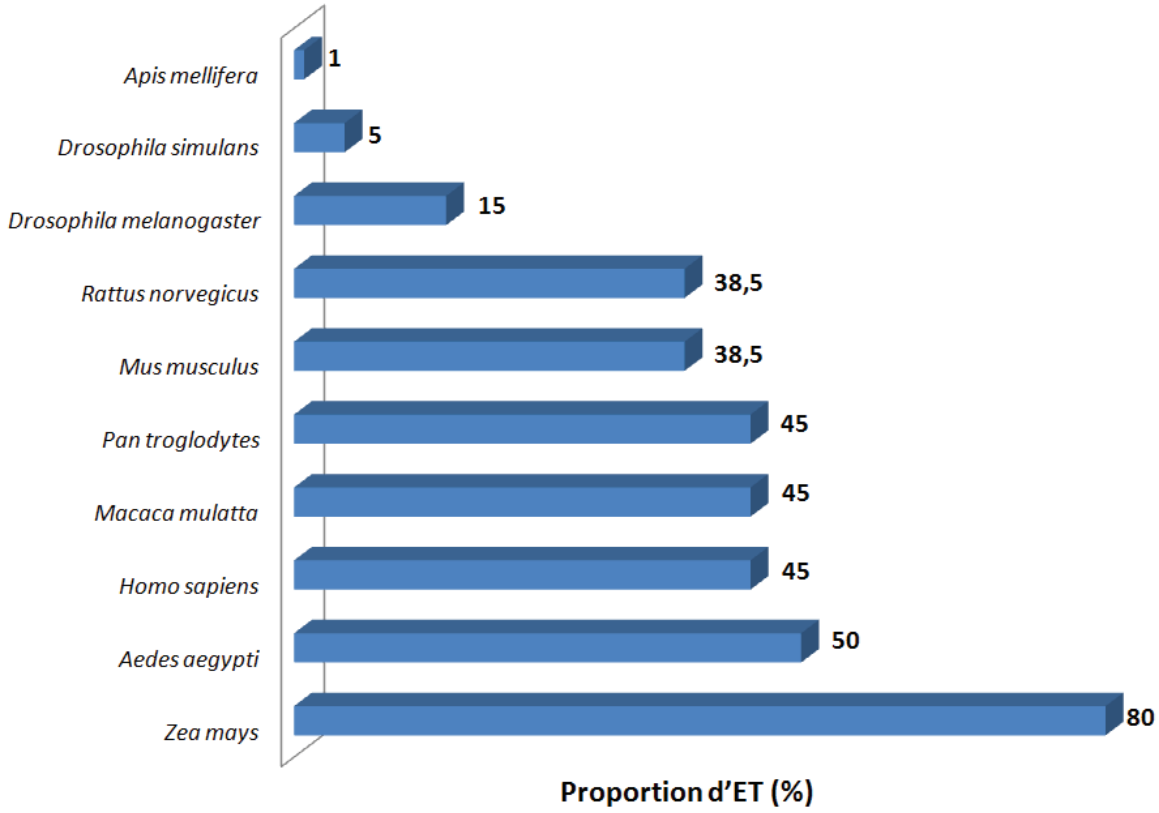


Figure 1. Proportion d'éléments transposables dans dix génomes eucaryotes.

Aedes aegypti [Nene et al, 2007], *Apis mellifera* [Honeybee Genome Sequencing Concortium, 2006], *Drosophila simulans* et *Drosophila melanogaster* [Drosophila 12 Genomes Consotrium, 2007], *Homo sapiens* [Lander et al., 2001], *Rattus norvegicus* [Gibbs et al.,2004], *Zea mays* [Schanble et al, 2009], *Mus musculus* [Mouse Genome Sequencing Consortrium, 2002], *Pan troglodytes* [Chimpanze Sequencing Constrium, 2002], *Macaca mulatta* [Rhesus Macaque Genome Sequencing and Analysis Consortium, 2007].

Depuis la découverte des ET, de nombreuses expériences ont été menées afin de caractériser leurs mécanismes de transposition et leur structure moléculaire. Ces expériences ont principalement été entreprises sur des organismes modèles comme la levure *Saccharomyces cerevisea* (Boeke et al., 1985) et la mouche *D. melanogaster* (O'Hare et Rubin 1983). Depuis, un grand nombre d'éléments ont été décrits, et une classification de ces éléments s'est avérée nécessaire.

Le but du prochain paragraphe est de décrire brièvement les différents types d'éléments, à partir de l'intermédiaire de transposition, sans avoir l'objectif d'être exhaustif. On distingue principalement deux grandes familles en fonction de leur structure générale et de leur intermédiaire de transposition. Il existe ainsi les éléments à ADN (classe II) et les éléments ARN (classe I) qui comprennent les LTR-rétrotransposons et les rétroposons (Finnegan, 1989) (Figure 2).

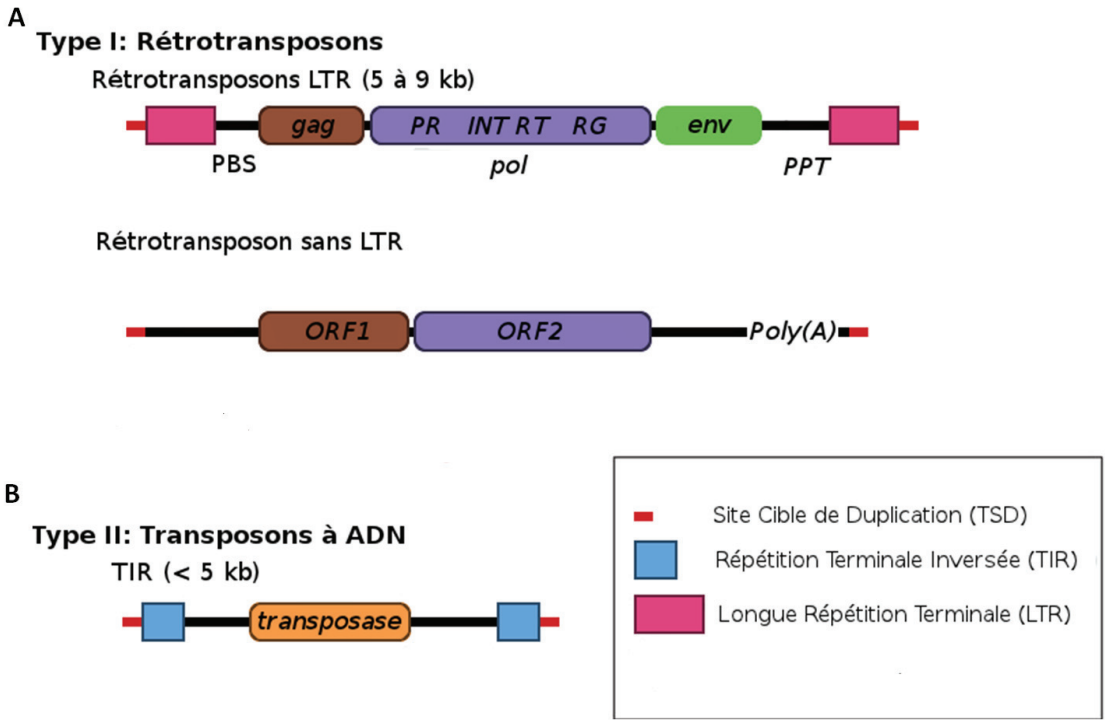


Figure 2: Structure schématique des deux grandes familles d'éléments transposables.

(A) Rétrotransposons à LTR. Ils sont composés de deux longues répétitions terminales directes (LTR, rectangle rose). Ces LTRs encadrent deux ORFs : gag (protéines de structure) et pol (protéines de réplication). Certains ont un troisième cadre de lecture noté env (protéine d'enveloppe). Rétrotransposons sans LTR. Ils codent une protéine de structure gag et une ORF2 pouvant coder diverses protéines nécessaires à la réplication. Ils sont terminés par une queue poly(A). (B) Transposons à ADN (Elément de classe II). Ils codent une transposase encadrée par deux répétitions inversées (carrés bleus). D'après Dufourt 2008.

1.2 Classification des éléments transposables

1.2.1 Les éléments à ARN (Classe I)

La classe I regroupe ce que l'on appelle les rétroéléments qui transposent *via* un intermédiaire ARN selon un mécanisme dit de « copier-coller » (Figure 3). Ils ont une origine commune avec les virus à ARN double brin et les virus à ARN simple brin (Ahlquist, 2006).

Les éléments complets codent une transcriptase inverse (RT) capable de synthétiser une molécule d'ADN complémentaire (ADNc) à partir d'une matrice d'ARN simple brin. Ils codent également une intégrase (IN) qui permet à cet ADNc de s'intégrer dans un nouveau site chromosomique dans le génome de son hôte. Le nombre de copies s'en trouve ainsi augmenté : on parle de mode de transposition réplicatif.

De nombreuses familles de rétroéléments ont été caractérisées chez toutes les espèces eucaryotes (Lynch & Tristem, 2003). Selon leur structure génomique, deux groupes de rétroéléments sont distingués : les rétrotransposons à LTR pour « Long Terminal Repeat » qui ont une structure de type proviral, et les rétrotransposons sans LTR.

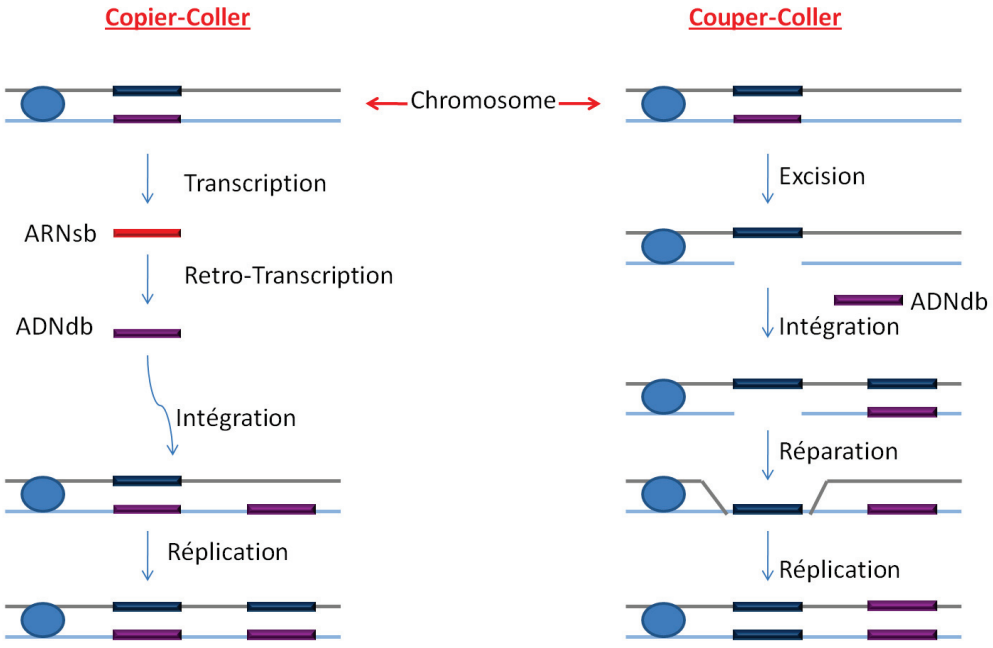


Figure 3: Schéma simplifié des deux modes de transposition des ET.

(A) Transposition en « copier-coller » des éléments transposables de classe I (rétrotransposon) : l'élément est transcrit en ARN simple brin (ARNsb), puis retro-transcrit en ADN double brin (ADNdb) et va s'insérer dans un nouveau locus génomique, sans excision de l'ET de départ. (B) mécanisme de transposition des éléments transposables de classe II (transposon) : l'élément s'excise sous forme d'ADN double brin (ADNdb) et va se réinsérer dans un site génomique cible. D'après Dufourt 2008

1.2.1.1 Les éléments à LTR et les rétrovirus endogènes (ou ERV pour « endogenous retrovirus »)

Les rétrotransposons à LTR sont des éléments transposables qui ont des homologies structurales avec les rétrovirus de vertébrés (Boeke, 1998a). En effet, deux longues répétitions terminales ou LTR encadrent les cadres ouverts de lecture *gag*, *pol* et *env*. (Figure 4). Les rétrotransposons à LTR peuvent donc transposer dans le génome de l'hôte *via* un mécanisme reproduisant le cycle viral avec ou sans étape infectieuse. La caractérisation chez *Saccharomyces cerevisiae* de la rétrotransposition de l'élément *Ty* apporte la première confirmation expérimentale d'un tel mécanisme (Boeke et al, 1985). Au niveau des insertions *de novo*, l'épissage précis d'un intron présent dans l'élément «donneur» ainsi que la régénération de copies natives à partir d'un élément sous le contrôle d'un promoteur inductible démontre une transposition selon un type rétroviral.

En parallèle, si l'intégration de la séquence ADN d'un rétrovirus a lieu dans la lignée germinale de l'hôte, le provirus sera transmis verticalement à la descendance, et se comportera alors comme un rétrovirus endogène. Ces séquences rétrovirales sont transmises de génération en génération de manière mendélienne, comme n'importe quelle séquence d'ADN présente dans l'ensemble du génome de leur hôte. Le processus conduisant à l'établissement (ou « colonisation ») des rétrovirus endogènes au sein des génomes cellulaire est nommé endogénisation (*endogenization*).

La plupart des rétrovirus endogènes de vertébrés descendent de rétrovirus infectieux. En effet, il est démontré qu'ils sont plus proches entre eux que des rétrotransposons à LTR de plantes, de champignons ou d'insectes (Malik & Eickbush, 2001). Enfin, certains rétrotransposons à LTR peuvent être à l'origine de l'apparition de nouveaux virus. Les rétrovirus de mammifères divergent ainsi à partir des rétrotransposons par l'acquisition d'un gène *env* (codant une protéine d'enveloppe) provenant d'autres virus ou par la captation d'un gène hypothétique d'une enveloppe cellulaire (Malik et al, 2001). Une nouvelle classification abolit la frontière diaphane qui sépare rétrovirus et LTR rétrotransposons en introduisant une classe de virus appelés « virus à ARN rétrotranscrit » (Boeke, 1998a). Les virus à ARN rétrotranscrit se divisent en trois familles : Retroviridae, Metaviridae et Pseudoviridae (Hull, 2001); (Xiong & Eickbush, 1990). Un gène *env* a été identifié uniquement chez les rétrotransposons à LTR du groupe des Metaviridea. La présence de ce gène traduit une convergence des rétrotransposons à LTR de Metaviridae vers les rétrovirus de mammifères par une acquisition indépendante et plus tardive de ce gène. (Figure 4).

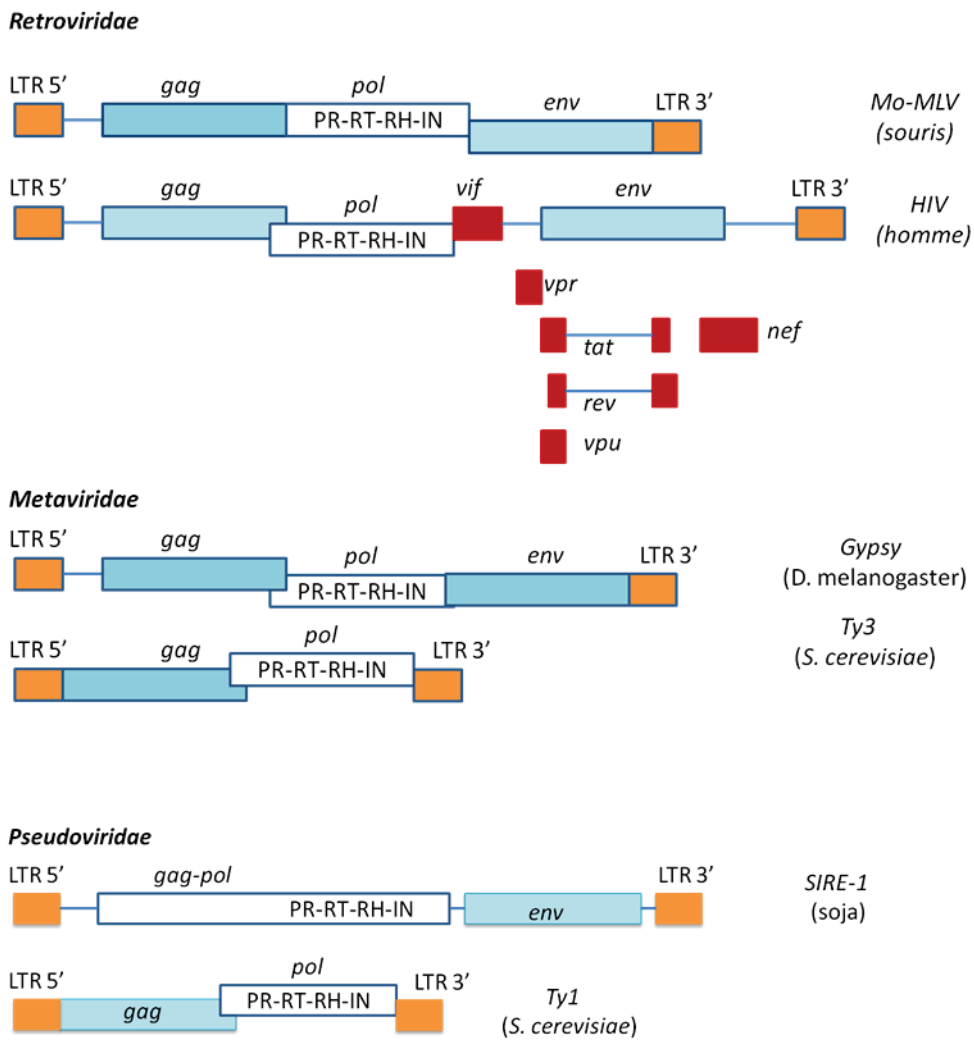


Figure 4. Structure génomique des virus à ARN rétrotranscrit.

Deux *retroviridae* : l'un simple (*Mo-MLV* : Moloney Leukemia Virus), l'autre complexe (*HIV* : *Human Immunodeficiency Virus*). Les protéines accessoires sont représentées en rouge. Deux *metaviridae* : l'un, avec enveloppe, un *errantivrius* (*Gypsy* chez les insectes), l'autre sans enveloppe, un *metavirus* (*Ty3* chez la levure). Deux pseudovirida : eux aussi avec ou sans enveloppe, (respectivement *SIRE-1* et *Ty1*). L'hôte de chaque élément est indiqué entre parenthèses. LTR : Long-terminal Repeat ; PR : protease ; RT rétrotranscriptase ; RH : RnaseH ; IN : intégrase. D'après Buchon 2006.

1.2.1.2 Les éléments sans LTR

Les rétroposons se caractérisent par l'absence de LTR. Ils utilisent un intermédiaire ARN dont la synthèse dépend de promoteurs internes situés en aval du site d'initiation de la transcription. Les rétrotransposons sans LTR peuvent être représentés par deux exemples historiques : les « Long Interspersed Nuclear Elements (LINE) qui codent leur propre RT et sont de ce fait dits autonomes, et les Short Interspersed Nuclear Elements (SINE) qui sont non autonomes et utilisent la machinerie des LINE pour pouvoir transposer (Kajiwaka & Okada 2002 ; Kramerov & Vassetzky, 2005). Leur mécanisme de transposition est différent de celui de rétrotransposons à LTR car la transcription inverse du transcrit se fait en même temps que l'intégration dans le génome.

1.2.2 Les éléments à ADN (classe II)

La classe II correspond aux transposons qui transposent selon un mode conservatif de « couper-coller », sans intermédiaire ARN (Figure 2). Leur transposition se compose donc de leur excision de leur site d'origine puis leur insertion dans un site génomique cible. Trois grandes familles chez les eucaryotes composent cette classe : la famille *P* constituée entre autres de l'élément *P* de *D. melanogaster* (Engels, 1989) ; la famille *hAT* avec *hobo* chez *D. melanogaster*, Ac (Activator) chez le maïs et *Tam3* chez le muflier *Antirrhium majus* (Calvi et al, 1991); la famille *Tc1/mariner* composée de l'élément *Tc1* de *Caenorhabditis elegans* (Emmons et al, 1983). Ils diffèrent par leurs sites cibles et la transposase codée par les éléments autonomes.

1.2.3 Autres

En plus des principales familles citées ci-dessus, il existe de nombreuses familles constamment découvertes et qui peuvent partager des caractéristiques communes avec un ou plusieurs groupes d'éléments transposables. Par exemple, un rétroélément caractérisé dans le génome humain présente une structure similaire aux rétrotransposons à LTR mais est encadré par deux séquences répétées en orientation inverse, particularité des éléments à ADN (Hughes & Coffin, 2002). La classification des éléments transposables est un exercice compliqué, puisque plusieurs caractères peuvent être utilisés qui sont souvent non exclusifs. Plusieurs tentatives ont été faites, et celle qui réunit le plus de consensus, après celle de Finnegan (1991) décrite ci-dessus, est celle de Wicker 2007, même si pas toujours simple d'utilisation.

1.3 Cycle de rétrotransposition des éléments à LTR

Les éléments de type rétrotransposons à LTR utilisent un cycle analogue à celui des rétrovirus, décrit par (Echalier, 1989). L'amorçage du cycle de la transcription inverse est assuré par la fixation d'un ARNt dans la partie U5 de la LTR 5'. La synthèse par la transcriptase inverse s'effectue jusqu'à l'extrémité 5' de la LTR, ce qui produit une molécule d'ADN d'environ 200 pb. La transcriptase inverse, par son activité RNAse H, dégrade alors l'ARN complémentaire jusqu'à la partie R. La partie 3' de la copie d'ADN ainsi libérée s'hybride alors sur la partie complémentaire de la LTR 3'. La synthèse du premier brin d'ADN est ainsi complète. L'activité ADN polymérase de la transcriptase inverse permet la synthèse du deuxième brin d'ADN. On obtient une copie d'ADN linéaire à deux LTR qui s'intégrera dans le génome. Dans certains cas, une copie circulaire à deux LTR est obtenue mais elle ne semble pas pouvoir se réintégrer. L'intégrase coupe de façon décalée au site d'insertion et assure la ligation des extrémités aux deux LTR, générant ainsi une duplication du site cible de part et d'autre des LTR. (Figure 5).

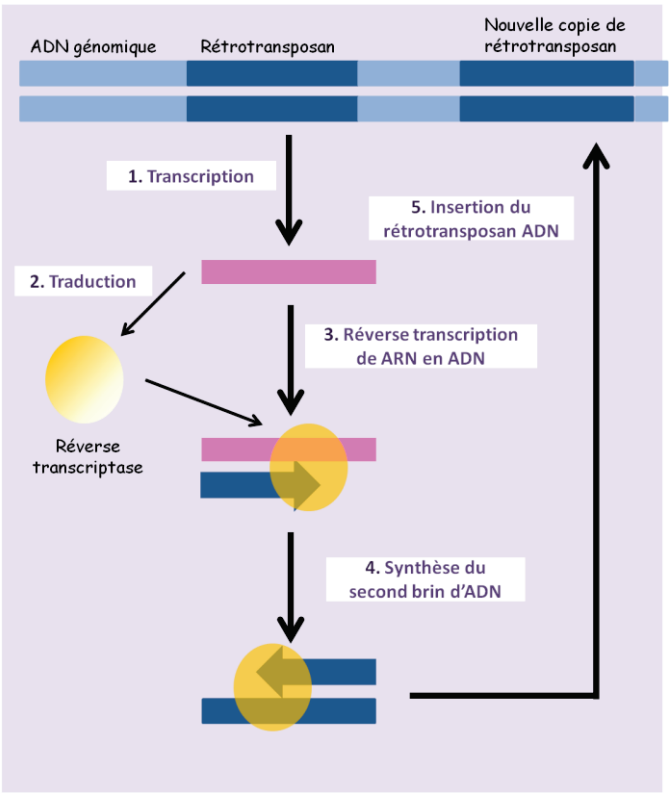


Figure 5. Cycle de rétrotransposition des rétrotransposons à LTR selon Sabot et Schulman 2006

Dans le cas des rétrovirus, deux molécules d'ARN peuvent être encapsidées dans un virion. Les protéines transcriptase inverse et intégrase sont incluses dans les particules virales formées. On peut noter que la recombinaison entre deux molécules est possible lors du cycle de réplication suivant et permettra la création de nouveaux rétrovirus avec de nouvelles fonctions.

1.4 Répartition des éléments transposables dans les génomes.

L'évolution des éléments transposables par l'apparition de mutations ou délétions internes conduit à une composition hétéroclite des génomes qui mélangent à la fois des copies complètes actives avec des vestiges parcellaires et dégénérés de ces éléments (Alonso-González et al, 2003; Cizeron & Biémont, 1999 ; Lerat et al, 2003, 2011).

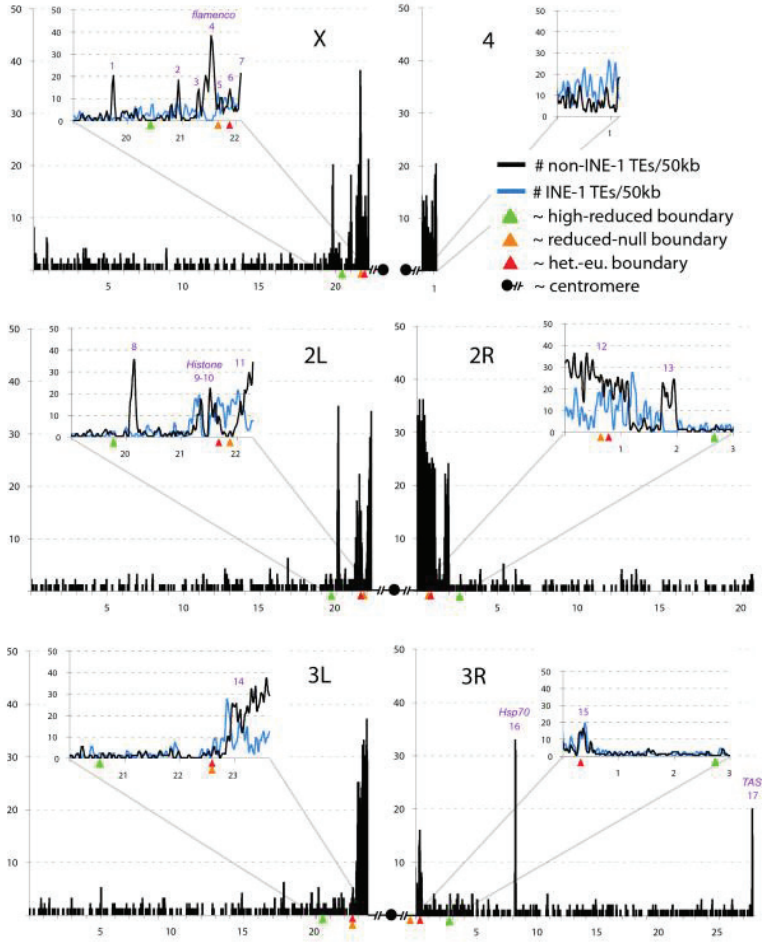


Figure 6 : Distribution des éléments transposables le long des chromosomes de *D. melanogaster*.

Le nombre d'éléments transposables par fenêtre de 50 kb est rapporté en fonction de leur position. La densité des éléments de la famille INE-1 est représentée en bleu. Celles des autres familles sont figurées en noir. Dans les régions péricentromériques la frontière cytologique qui sépare l'euchromatine de l'hétérochromatine est rapportée par des triangles rouges. De même, les frontières génétiques entre des régions chromosomiques à forte recombinaison et celles à recombinaison nulle, sont figurées par des triangles verts et bleus respectivement. D'après Bergman et al, 2006.

Par ailleurs, la sélection naturelle, la recombinaison et le propre mécanisme d'intégration des éléments, conduisent à une distribution différente des éléments entre espèces (comme mentionné ci-dessus) mais aussi entre les régions du génome. Certaines régions sont ainsi totalement dépourvues d'éléments transposables (Simons et al, 2006). Par exemple, chez *D. melanogaster*, le séquençage de la lignée isogénique ISO1a a permis d'identifier une

grande région sans éléments de plus de 850 kb et qui contient 106 gènes (Bergman et al, 2006). Aussi, l'analyse de ce génome, qui a une quantité relativement faible d'ET, a montré que seulement 6,41% des insertions d'ET avaient lieu dans des éléments déjà présents (Bergman et al, 2006). En parallèle, 20,5% des éléments sont complets donc putativement actifs. Par ailleurs, il a été montré que l'hétérochromatine péricentromérique, où le taux de recombinaison est faible voire nul, est constituée de régions denses en éléments transposables chez *Arabodopsis thaliana* et *D. melanogatser* (Vaury et al, 1989). En effet, Bergman et *al.,* ont caractérisé des régions localement très denses en éléments transposables (RDE). Elles se composent d'au moins 10 éléments (autres que INE1, élément très abondant et de petite taille) sur moins de 50 kb.

1.4.1 Un exemple de RDE chez Drosophila melanogaster : le locus flam/COM

Le locus 20A, situé dans l'hétérochromatine péricentromerique du chromosome X, a été caractérisé comme étant responsable du contrôle de la régulation de plusieurs ET (Desset et al, 1999, Prud'homme et al, 1995). Il a été montré que la mobilité de trois rétroéléments, *ZAM, Idefix* et *Gypsy*, était inhibée par la présence de déterminants génétiques au niveau de cette région appelée *flam/COM*. Le locus 20A est situé à la frontière de régions chromosomiques pouvant recombiner et des régions où le taux de recombinaison est faible voire nul, ce qui permet une accumulation d'ET, qui ne sont pas éliminés par la recombinaison. La présence d'un cluster d'éléments transposables au niveau de ce locus a été mise en évidence par hybridation *in situ* (Maside et al, 2005) et plus récemment, l'annotation des ET à partir du séquençage de l'hétérochromatine du génome de la drosophile a permis la caractérisation d'une région très dense en séquences répétées qui s'étend sur plus de 150 kilobases (kb) (Bergman et al, 2006). Le locus *flam/COM* est en effet composé d'un très grand nombre de copies défectives de séquences d'ET. La plupart de ces séquences hétérochromatiques présentent la particularité d'avoir été insérées sur le même brin. Elles forment ainsi une longue succession de séquences répétées de même orientation du centromère vers le télomère.

1.5 Relations avec le génome

La théorie de l'ADN égoïste (Doolittle & Sapienza, 1980 ; Orgel & Crick, 1980) suggère que les génomes sont constitués de deux types de séquences : de l'ADN « fonctionnel », avec une fonction dans la cellule, et de séquences qui se comportent comme des parasites, ne conférant aucun avantage à la cellule, et qui se maintiennent dans les génomes de par leur capacité réplicative, malgré leurs effets négatifs. Les ET sont des exemples d'ADN égoïste. En effet, dans des génomes compacts tels que ceux de *D. melanogaster, Arabidopsis thaliana* ou encore *Caenorhabditis elegans*, on n'observe des ET que dans les régions ou ils ne sont pas délétères car toutes les autres insertions sont éliminées par la sélection naturelle. Suite aux séquençages des grands génomes eucaryotes : homme, chimpanzé, plantes, il est apparu clairement que la partie codante ne représentait qu'une partie minime du génome alors que la majeure partie était représentée par de l'ADN non-codant ou ADN poubelle (« Junk DNA »), qui est en fait constitué majoritairement de diverses séquences répétées (satellites, minisatellites, microsatellites, éléments transposable).

Britten et Davidson ont été les premiers à émettre l'hypothèse selon laquelle l'exaptation des éléments transposables par le génome de l'hôte constitue un mécanisme important dans la structuration des génomes (Britten & Davidson, 1969). Ces séquences sont actuellement reconnues comme des acteurs majeurs dans l'évolution des génomes hôtes (Oliver & Greene, 2009, Biemont & Vieira, 2006), contribuant à la diversité phénotypique des organismes (Bohne et al, 2008). Au cours du chapitre suivant, nous verrons au travers de quelques exemples, l'importance de la domestication des éléments transposables dans l'édification des génomes actuels

1.5.1 Implication des Éléments transposables dans l'Évolution des génomes

La grande majorité des insertions d'ET présentes dans un génome sont sélectivement neutres ou faiblement délétères. Cependant, quelques insertions ont été « domestiquées » par les génomes et sont devenues des gènes ou des éléments de régulation des gènes. Le séquençage du génome humain a ainsi révélé qu'au moins 47 gènes dérivaient de la domestication d'un élément transposable (Lander et al, 2001). En théorie, toutes les fonctions codées par ces éléments (transposase, intégrase, transcriptase inverse, protéine de structure) peuvent être domestiquées.

Le cas le plus typique est fourni par les gènes rag1 et rag2 du système immunitaire des mammifères, qui dérivent d'un élément transposable à ADN (Sakano et al, 1979). Plusieurs études ont montré ainsi le parallèle qui existe entre la fonction de ces gènes et les mécanismes de transposition. Tout d'abord, les gènes rag1 et rag2 sont adjacents et sont dépourvus d'introns (Gellert, 1996). La protéine RAG1 arbore aussi un motif DDE consensuel (présent au niveau de certaines transposases). Enfin, les protéines RAG ont une activité de transposase *in vitro*, c'est-à-dire qu'elles fonctionnent comme une endonucléase introduisant des coupures double-brin proches des signaux spécifiques de recombinaison (Agrawal et al, 1998) (Hiom

et al, 1998). Ces données suggèrent que l'évolution du système immunitaire acquis chez certains vertébrés repose sur l'insertion d'un « transposon RAG » dans le génome de la lignée germinale d'un ancêtre commun.

Il a été de même montré qu'une transposase similaire à celle de l'élément *hAT* codée par le gène *Daysleeper* est essentielle pour le développement d'*Arabidopsis thaliana* (Bundock & Hooykaas, 2005) et pour sa réponse à certains stimuli lumineux (Lin et al, 2007), ce qui suggère que les protéines codées par les éléments transposables peuvent avoir un impact sur le fonctionnement des génomes. D'autres effets des éléments transposables ont été observés. La domestication des protéines de l'enveloppe des rétroéléments est essentielle à la mise en place du placenta chez les mammifères (Mallet et al, 2004). Deux protéines placentaires Syncitin-1 et Syncitin-2 dérivant de rétrovirus endogènes ont ainsi été identifiées chez l'homme (Blaise et al, 2003 ; Mi et al, 2000).

Un autre exemple d'impact des ET sur l'évolution des génomes se retrouve chez la souris. Peaston et al (2004) ont montré un contrôle de l'induction différentielle de l'expression de gènes, qui affecte les processus de développement des ovocytes et des jeunes embryons (Peaston et al, 2004), par les ET. Chez la souris également (Muotri et al, 2005), ont observé une modification de l'expression de gènes neuronaux par la transposition d'éléments LINE, dont l'activité pourrait être responsable de changements dans les interactions entre neurones.

Un autre exemple remarquable de domestication concerne l'élongation des extrémités des chromosomes chez la drosophile. En effet, à la fin de chaque division mitotique, une télomerase allonge les extrémités des chromosomes chez la plupart des espèces. Cependant chez *D. melanogaster*, l'érosion des extrémités chromosomiques est compensée par la rétrotransposition de deux rétroposons, *TART* et *HeTA* (Pardue & DeBaryshe, 2003).

Au niveau post-transcriptionnel, l'insertion d'un élément transposable peut aussi être responsable de l'épissage alternatif des gènes par l'apparition de codons stop et contribuer ainsi à l'homéostasie des ARN messagers. Ainsi, chez la levure, l'insertion des seules séquences LTR du retrotransposon *Ty1* dans le sens de la transcription provoque une inactivation de gène par une compétition entre les séquences TATA présentes dans la LTR du rétrotransposon et dans *HIS4*. Elle résulte en la synthèse d'un transcrit aberrant avec un codon d'initiation et plusieurs codons stop en amont du site d'initiation de *HIS4* (Wintson, 1992).

En plus de leur contribution structurale, les ET participent à la régulation des gènes qui sont dans leur voisinage. Dans le génome humain, près de 25% des promoteurs de gènes dérivent de séquences répétées (Jordan et al, 2003). Une étude similaire mais sur une distance plus grande à partir du point d'initiation de la transcription, a montré cette fois la présence d'éléments transposables dans environ 83% des gènes de l'homme. Des exemples *in vivo* illustrent par ailleurs l'influence des séquences répétées dans l'expression des gènes à proximité. Ainsi, l'activation du rétrotransposon à LTR *Wis2-1A* induit la régulation de l'expression d'un certain nombre de gènes chez le blé (Kashkush et al, 2003). En fonction de l'orientation du rétrotransposon, le gène sera alors exprimé ou réprimé. D'autre part, la forte

activité des transposons dans la région promotrice du gène CHS-D Ipomea est à l'origine de la diversité de pigmentation chez cette espèce (Clegg & Durbin, 2003).

Enfin les ET peuvent participer indirectement à la régulation des gènes *via* la production de micro-ARN (miRNAs). Les miRNA appartiennent à une classe d'ARN non codant qui réprime l'expression de gènes clés au cours du développement. Ce contrôle repose sur le même principe d'inhibition post-transcriptionnelle que la voie « classique » de l'ARNi (Ambros, 2004). La reconnaissance de l'ARNm cible est médiée par un petit ARN mature d'environ 22 nucléotides, présent sous forme simple brin dans le cytoplasme. Plus particulièrement, l'insertion des éléments transposables dans de nouveaux sites chromosomiques constitue une force substantielle dans la création des précurseurs de miRNA chez les plantes (Llave et al, 2002). D'autre part, une analyse *in silico*, réalisée chez les mammifères, a montré qu'au moins 13 miARN proviennent de séquences répétées d'éléments transposables dont la plupart sont conservées entre la souris, le rat et l'homme (Smalheiser & Torvik, 2005). Plus récemment, un total de 55 miARN dérivés d'un ET ont été caractérisés chez l'homme (Piriyapongsa & Jordan, 2007).

Un autre mécanisme par lequel les ET peuvent influencer l'évolution de leur génome hôte est la recombinaison. En effet, du fait de leur grand nombre de copies, des recombinaisons entre régions homologues d'ET peuvent avoir lieu. Ces recombinaisons peuvent se faire entre des éléments situés sur un même chromosome, engendrant ainsi une duplication, une délétion ou des inversions des séquences situées entre les ET, mais elles peuvent aussi avoir lieu entre des éléments situés sur des chromosomes différents ayant pour conséquence des translocations chromosomiques ou des réarrangements chromosomiques plus complexes. Bailey et al ont montré que les éléments *Alu* seraient à l'origine de duplications dans le génome humain (Bailey et al, 2003), et seraient responsables de la translocation entre les chromosomes 11 et 22 de l'homme (Hill et al, 2000). Néanmoins, des recombinaisons entre les autres types d'ET ont été rapportées, comme par exemple l'inversion sur le chromosome Y de l'homme médiée par une recombinaison entre deux LINE (Schwartz et al, 1998). Chez *Drosophila virilis*, deux éléments de type rétrotransposons, *Penelope* (sans LTR) et *Ulysses* (avec LTR) montrent une insertion préférentielle au niveau de sites d'inversions et sont ainsi à l'origine de divers réarrangements chromosomiques (inversions, translocations et délétions). Ces réarrangements induits par les ET semblent jouer un rôle important dans l'évolution des espèces du groupe *virilis* (Evgen'ev et al, 2000).

Les interactions entre le génome hôte et les ET sont nombreuses, et ici je n'ai mentionné que quelques exemples pour les illustrer. Les analyses génomiques à grande échelle vont sans doute continuer de mettre en évidence ce type d'exemple. Par ailleurs, la présence même d'ET dans un génome, indépendamment de leur mobilisation ou de leur séquence, peut conduire à des changements d'expression par le biais des systèmes de régulation épigénétique, ciblés sur les ET. Ces aspects font l'objet des prochains paragraphes.

.

1.6 Régulation épigénétique des éléments transposables

Le nombre d'insertions d'éléments transposables dans un génome est la conséquence des effets de la transposition, qui conduit à une augmentation du nombre de copies, et de la sélection (sous diverses formes) qui conduit à une élimination de ses insertions. On constate ainsi que le nombre d'insertions d'ET est non seulement fonction des espèces, comme mentionné ci-dessus, mais aussi fonction des populations, et aussi du type d'élément considéré. Ces dernières années ont vu un accroissement des études sur l'activité des éléments, et en particulier sur les systèmes qui permettent le « silencing » de telles séquences par les génomes hôte (Luchetta et al, 2005). Dans la plupart des organismes, ces éléments sont sous la dépendance de contrôles épigénétiques, autrement dit, leur expression dépend de modifications qui changent la structure de l'ADN elle-même, en absence de changement dans la séquence. Ces modifications, qui consistent en un « second code » par comparaison avec le code génétique classique, incluent la méthylation de certaines bases de la molécule d'ADN, l'interférence ARN et le remodelage de la chromatine par des modifications des protéines histones.

De nos jours, l'épigénétique fait référence à la fois aux changements héritables de l'activité et de l'expression de gènes dans la descendance d'une cellule ou d'un individu, et aux changements stables sur le long terme du programme transcriptionnel d'une cellule, sans pour autant être nécessairement transmis aux cellules filles. Ainsi, l'épigénétique représente l'ensemble des mécanismes permettant à une cellule de faire face aux différents événements de sa vie (différenciation, stress, etc..) en modulant son programme d'expression génétique (Bird et al, 2007).

Les modifications de l'information épigénétique sont à la fois stables et modulables par un grand nombre de facteurs cellulaires, reflétant un état physiologique ou pathologique de la cellule ainsi qu'une adaptabilité à son environnement. En comparaison par rapport au génome, qui représente l'ensemble des gènes d'une cellule, l'épigénome représente l'ensemble des marques et modifications épigénétiques d'une cellule. Cet épigénome est propre à chaque cellule à un instant *t*, et influe directement sur l'expression du génome et donc sur le programme génétique utilisé par la cellule. Il existe trois classes majeures d'information épigénétique : la méthylation de l'ADN, les petits ARN non codants qui permettent l'ARN interférence, et les modifications de la chromatine. Cette dernière classe comprend les modifications post-traductionnelles des histones, capables de conférer des spécificités aux nucléosomes où elles sont intégrées. Il est important de noter qu'il existe une grande connectivité entre ces différentes informations épigénétiques : de nombreux travaux mettent en évidence une synergie et un enchainement de modifications épigénétiques pour arriver *in fine* à la régulation spécifique de l'expression du génome. Ces différents mécanismes font appel à toute une batterie de facteurs et de complexes : les « écrivains » et les « effaceurs », capables de modifier les marques épigénétiques et cibler ainsi des régions spécifiques de la chromatine afin d'y recruter différents complexes protéiques ; et les « effecteurs », capables de remanier la chromatine et/ou d'entrainer différents processus

biologiques (Goldberg et al, 2007). Grâce aux études fondamentales réalisées à partir d'espèces modèles, il est maintenant clairement établi que l'épigénétique est au cœur de la stabilité des profils d'expression des gènes au cours du développement et de la différenciation cellulaire. De plus, il semble que la plupart des phénomènes épigénétiques sont sous-tendus chez les eucaryotes par des changements d'états de la chromatine.

Ces dernières années un intérêt accru a été donné à l'étude de l'importance de l'épigénétique dans la régulation des éléments transposables. La suite de ce chapitre s'efforcera de donner un aperçu des différentes marques et informations épigénétiques, mais, sachant que ce domaine est en évolution constante, il me sera difficile d'être exhaustif.

1.6.1 Méthylation de l'ADN

La méthylation de l'ADN est une modification chimique des carbones 5 des cytosines de l'ADN. Chez les plantes et les vertébrés la méthylation en 5 des cytosines (5metC) est une marque réversible caractéristique des séquences transcriptionnellement inactives (Klose & Bird, 2006). Lorsqu'elle est localisée au niveau de la région promotrice des gènes, la méthylation est associée à la répression de leur expression (Suzuki & Bird, 2008).

Bien que la méthylation de l'ADN soit un mécanisme de régulation épigénétique largement répandu, les génomes sont méthylés de façon différente selon les organismes. Chez les mammifères, en particulier l'homme et la souris pour lesquels on a la plupart des données, cette méthylation se fait principalement de façon symétrique sur les 2 brins complémentaires de l'ADN, sur les cytosines des dinucléotides CpG. Entre 70 et 80% des CpGs sont méthylés chez les mammifères. Le reste des CpGs non methylés ne se retrouve pas aléatoirement sur le génome et forme ce que l'on appelle les ilots CpG (1% du génome humain), principalement retrouvé dans 60% des promoteurs chez l'homme et chez la souris (Robertson & Wolffe, 2000). La méthylation de l'ADN chez les plantes est à la fois symétrique sur les cytosines des CG ou CHG (H étant A, T ou C), ou asymétrique dans un contexte CHH. Chez *Arabidopsis thaliana*, les proportions de cytosines méthylées au sein des nucléotides CG, CHG et CHH sont respectivement de 24, 6,7 et 1,7% (Cokus et al, 2008).

Les enzymes responsables de la méthylation des cytosines dans l'ADN sont représentés par la famille des DNA cytosine-5 méthyl transférase : Dnmt (Brenner & Fuks, 2006). Chez les vertébrés, il existe deux types d'enzymes selon leur mode de fonctionnement. On distingue en effet les Dnmt de maintien (Dnmt1) qui sont impliquées dans la méthylation post-réplicative des brins d'ADN hémi-méthylés et les Dnmt dites de *novo* (Dnmt3 a/b) qui sont responsables de la méthylation *de novo*. Ces deux enzymes participent essentiellement à une reprogrammation épigénétique du génome dans les premières étapes du développement (Dean et al, 2005). Une altération des profils de méthylation dans une cellule participe pour une grande part aux phénomènes de cancérisation qui ont été décrits. Le paradigme des cellules cancéreuses consiste en une hypométhylation du génome dans son ensemble associé à des hyperméthylation ponctuelles spécifiques de gènes, notamment des gènes de suppression de tumeurs, qui se retrouvent sous « silencing » (Szyf, 2005). Les ET de mammifères sont eux aussi soumis à un « silencing » par méthylation de l'ADN, associé aux autres systèmes de

régulation, notamment l'ARN interférence, bien que le mécanisme précis reste encore à déterminer (Aravin & Bourc'his, 2008) (Kuramochi-Miyagawa et al, 2008).

Chez la plante *Arabisopsis thaliana*, la méthylation des cytosines est impliquée dans la régulation des gènes ainsi que le contrôle des ET. La méthylation symétrique semble être spécifique des gènes, et ce sont les cytosines méthylées en contexte asymétrique qui sont fortement présentes au niveau des ET (Chan et al, 2005) (Zhang et al, 2006). Ces profils de méthylation sont établis et maintenus par des DNMT spécifiques, guidées par de petits ARN qui ciblent par homologie, les séquences dont les cytosines sont à méthyler (Matzke et al, 2007). Ce type de méthylation ciblée par des petits ARN, appelé « RNA dependant DNA Methylation » (RdDM), a également été mis en évidence chez la souris (Kuramochi-Miyagawa et al, 2008).

Cependant, la méthylation de l'ADN ne semble pas être fondamentale chez tous les eucaryotes. En effet chez la drosophile, le taux de méthylation des cytosines est très faible, compris entre 0.4% et 1% (Kunert et al, 2003); (Lyko et al, 2000). Parmi les gènes connus codant des enzymes ayant une activité « DNA methyltransferase », une seule Dnmt-*like* a été découverte dans cette espèce (Dnmt2) mais son rôle est encore controversé. Les mutants pour Dnmt2 sont viables et ne provoquent aucun phénotype, tandis que la surexpression de ce gène conduit à une augmentation de la longévité des mouches (Lin et al, 2005). Des études plus récentes sont une faveur d'une action de Dnmt2 comme une enzyme de méthylation de l'ARN, et il a été proposé que la méthylation observé chez la drosophile ne soit qu'un effet secondaire de cette activité enzymatique (Schaefer & Lyko, 2010). Ceci conduit à penser que chez la drosophile, la méthylation de l'ADN n'intervient pas dans la régulation des gènes et des ET comme cela a été mis en évidence dans d'autres espèces. Cependant, l'utilisation d'enzymes de restriction sensibles aux cytosines méthylées et l'analyse de séquences d'ADN après traitement au bisulfite, qui permet d'identifier les cytosines méthylées, ont permis de mettre en évidence la présence de sites méthylés chez *D. willistoni* (Garcia et al, 2007) et de séquences fortement méthylées au niveau de la région régulatrice du rétrotransposon à LTR *invader 4* durant les premiers stades embryonnaires de *D. melanogaster* (Phalke et al, 2009). Dans ce dernier exemple, la méthylation d'*invader 4* durant les premiers stades embryonnaires permettrait de recruter l'histone méthyl-transférase SUV4-20, permettant ensuite de triméthyler l'histone H4 au niveau de la lysine 20 (H4K20). Un défaut des structures télomériques, associé à la transposition de l'élément *invader* 4, a été de plus constaté dans des mutants dDnmt2.

1.6.2 Modifications d'histones

Le noyau des cellules eucaryotes est composé d'ADN et de protéines, l'ensemble constitue la chromatine. Les histones sont les protéines plus abondantes de la chromatine. Elles possèdent une masse moléculaire comprise entre 10 et 24 kD, elles sont très basiques, riches en résidus lysine et arginine et sont très conservées au cours de l'évolution. Il en existe cinq types qui sont présents en quantités plus ou moins variables et qui se répartissent en deux familles : les histones nucléosomiques (H2A, H2B, H3 et H4), qui sont responsables de

l'enroulement de l'ADN dans les nucléosomes, et les histones H1 qui sont des histones internucléosomiques.

Dans le noyau, la chromatine est formée d'une répétition de sous-unités appelées nucléosomes (Kornberg, 1974) (Luger et al.,1997). Ceux-ci sont constitués de 146 pb d'ADN enroulées deux fois autour d'un octamère d'histones. Chaque octamère comprend deux exemplaires des histones H2A, H2B, H3, H4. Deux nucléosomes sont séparés l'un de l'autre par une cinquantaine de paires de bases d'ADN nu et par l'histone internucléosomiques H1. Les nucléosomes sont ensuite assemblés en des structures de plus en plus complexes comme des spirales, des solénoïdes ou des hélices, jusqu'au niveau de compaction maximale : le chromosome.

1.6.3 Modifications post-traductionnelles des histones

Les histones possèdent un domaine globulaire carboxy-terminal formant le cœur du nucléosome et une queue amino-terminale exposée à la surface du nucléosome (Uberbacher & Bunick, 1985) et sujette à diverses modifications post-traductionnelles (Figure 7). Ce sont les modifications post-traductionnelles des histones qui confèrent à la chromatine deux états de compaction qui sont corrélés avec une activation de la transcription ou une répression. Les modifications d'histones participent à la régulation de l'expression des gènes, soit en induisant directement une modification de la structure de la chromatine, soit en servant de plateforme de liaison permettant le recrutement d'autres facteurs. Les modifications associées à une transcription active, telles que l'acétylation des histones H3 et H4 ou la di ou triméthylation de H3K4, sont considérées comme des marques de l'euchromatine. Les modifications fréquemment observées sur les gènes réprimés, comme H3K9me3 ou H3K27me3, sont référencées comme des modifications de l'hétérochromatine.

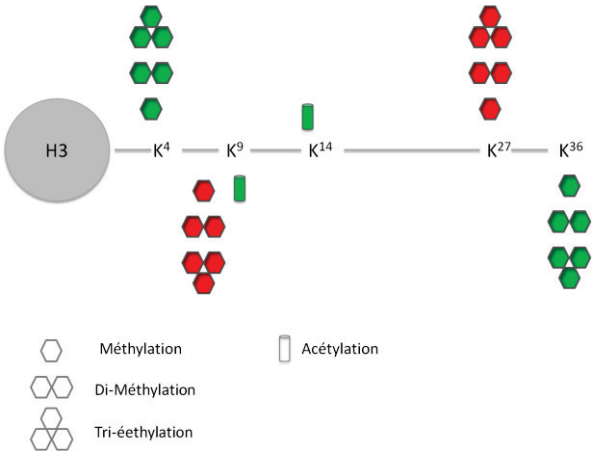


Figure 7 : Principales modifications post-traductionnelles des histones.

Les lysines situées sur la queue de l'histone H3 (partie N-terminale) peuvent subir une, deux ou trois méthylations et certaines d'entre-elles peuvent être acétylées. Ces deux marques sont mutuellement exclusives et définissent des états chromatiniens différents. La méthylation (mono, di, ou tri) des lysines en position 4 et 36 et l'acétylation des positions 9 et 14 définissent un état de chromatine ouverte favorable à l'expression des gènes

(Vert) , tandis qu'elle est associée a un état répressif (rouge) pour les lysines en position 9 et 27. D'après Ebert et al, 2000.

1.6.3.1 Acétylation des histones

L'acétylation des histones consiste en un transfert d'un groupement acétyl vers les résidus lysines des queues amino-terminales des histones. La neutralisation par le groupement acétyl des charges positives des résidus lysines conduit à une destabilisation de la structure chromatinienne par perte d'affinité des histones pour la double hélice d'ADN, il s'ensuit un relâchement de la structure de la chromatine permettant un accès plus facile à l'ADN par les facteurs de transcription. Le transfert du groupement acétyl est assuré par des protéines nucléaires possédant une activité histone acétyl-transférase (HAT). L'acétylation des lysines est associée à une chromatine accessible et une expression génique (Wang et al, 2008).

1.6.3.2 La méthylation des histones

La méthylation des histones touche les résidus lysine et arginine des queues amino-terminales des histones. Cette modification chimique post-traductionnelle est médiée par les protéines histones méthyl-transférases (HMT) (Zhang & Reinberg, 2001) qui assurent le transfert d'un, deux ou trois groupements méthyl vers les résidus cibles. La méthylation des histones peut aussi bien conduire à une activation ou à une répression de la transcription qui va dépendre de la nature du résidu modifié et de la combinaison de plusieurs modifications affectant les histones. A titre d'exemple, trois sites de méthylation sont impliqués dans l'activation de la transcription : H3K4, H3K36 et H3K79. Tandis que H3K4me se concentre à proximité du promoteur, H3K36me est exclusivement localisé sur l'ORF des gènes transcrits et s'accumule vers l'extrémité 3' (Barski et al, 2007) (Pokholok et al, 2005).

A l'opposé, la di-ou triméthylation des lysines 9 et 27 de l'histone 3 (H3K9me2 et me3, H3K27me2 et me3) est associée à une répression transcriptionnelle (Margueron et al, 2005) (Schotta et al, 2004).

1.6.4 Territoires chromatiniens

En fonction des modifications d'histones, la chromatine présente différentes conformations, et est organisée en régions plus au moins condensées. Ces territoires chromatiniens affectent l'expression des gènes ainsi que la stabilité des chromosomes. On considère que la chromatine se divise en deux compartiments distincts d'un point de vue structural, fonctionnel et cytologique : l'euchromatine, riche en gènes, et l'hétérochromatine, contenant moins de gènes et beaucoup de séquences répétées (Grewal & Elgin, 2007).

1.6.4.1 L'euchromatine

L'euchromatine est une forme relâchée de la chromatine. Certaines modifications d'histones sont observées chez l'homme et la drosophile avec une plus grande fréquence dans le compartiment euchromatinien, c'est le cas de la méthylation de la lysine 4 de l'histone H3 (H3K4me) (Santos-Rosa et al, 2002), de la méthylation de la lysine 36 de l'histone H3

(H3K36me) (Bernstein et al, 2005) et de l'hyperacétylation des histones H3 et H4 (H3ac, H4ac) (Grunstein, 1997).

1.6.4.2 L'hétérochromatine

L'hétérochromatine est la forme compactée de la chromatine et correspond à des régions peu ou pas transcrites. Il en existe deux formes :

Hétérochromatine constitutive :

L'hétérochromatine constitutive regroupe une grande partie des séquences répétées du génome. Elle se situe principalement au niveau des centromères et des régions péri-centromériques, des télomères et des organisateurs nucléolaires (Grewal & Elgin, 2007). L'hétérochromatine constitutive est nécessaire à l'intégrité du génome. Elle joue un rôle important dans la ségrégation des chromosomes et la protection des télomères contre leur dégradation (Bernard et al, 2001) (García-Cao et al, 2004).

L'hétérochromatine constitutive est un état stable et irréversible de la chromatine qui est transmis à chaque division cellulaire. Elle peut également affecter l'expression de gènes. En effet un gène situé près d'une région d'hétérochromatine constitutive voit son expression décroitre en fonction de la distance. Ce phénomène est décrit chez drosophile sous le nom de « variégation d'effet de position » (PEV) (Henikoff, 1990).

Hétérochromatine facultative :

L'hétérochromatine facultative correspond aux régions pouvant adopter de façon réversible les caractéristiques structurales et fonctionnelles de l'hétérochromatine constitutive en fonction de stades de développement particuliers ou dans des types cellulaires donnés. Un exemple bien caractérisé de formation d'hétérochromatine facultative est l'inactivation du chromosome X chez les mammifères aboutissant à une constitution différente de la chromatine des deux allèles pour un locus donné du chromosome X.

Le nombre de modifications d'histones différentes est extrêmement important et on ne connaît les effets que d'un faible nombre, qui sont bien étudiés. Par ailleurs, certaines modifications d'histones ne semblent pas avoir les mêmes effets dans différents organismes, ou en tout cas, ne semblent pas se localiser dans les mêmes régions (Ebert et al, 2006). Ceci implique qu'il faut rester prudent sur les effets des différentes modifications, et ne pas oublier les interactions possibles entre les modifications (Pietersen & van Lohuizen, 2008).

Les modifications d'histones sont un des mécanismes épigénétiques qui limite l'activité des éléments transposables. En effet la fixation de marques répressives au niveau des régions riches en ET provoque une condensation de la chromatine de cette région empêchant ainsi toute activité des ET. L'étude des modifications d'histones chez la souris, a montré que les nucléosomes associés aux ET sont enrichis en H3K9 monométhylé (H3K9me) qui est une marque répressive (Martens et al, 2005) et que cette modification, associée avec la triméthylation de H4K20 (H4K20me3), joue un rôle important dans la répression des

éléments répétés globalement et des ET plus spécifiquement (Mikkelsen et al, 2007). Des études similaires menées sur le génome humain ont permis de montrer un enrichissement en H3K9me2 au niveau des ET (Kondo & Issa, 2003). Plus globalement, dans le génome de la souris, les SINE montrent un enrichissement de la marque répressive H3K27me3 (Pauler et al, 2009). Enfin, il a été montré que la méthylation de H3K9 est impliquée dans la régulation des gènes, dans la structure des chromosomes et dans le contrôle des ET (Peng & Karpen, 2007). Toutes ces résultats montrent ainsi une répression de l'activité des ET par les modifications d'histones. Certains ET ne semblent cependant pas être soumis à cette répression. Récemment, une étude a montré que les ET impliqués dans des fonctions régulatrices chez l'homme sont enrichis en modifications d'histones activatrices (Huda et al, 2011).

Chez la drosophile la méthylation de H3K9 est impliquée dans la régulation des gènes durant le développement, dans la structure des chromosomes et dans le contrôle des ET (Peng & Karpen, 2007). L'une des protéines qui assure les modifications post-traductionnelles est l'histone méthyltransférase SU(VAR) 3-9 qui joue un rôle crucial dans la méthylation H3K9. Cette dernière permet le recrutement de protéines spécifiques des régions réprimées et de l'hétérochromatine constitutive *via* leur chromodomaines, telles que la protéine hétérochromatique 1 (HP1) (Bannister et al, 2001) (Lachner et al, 2001). En effet SU(VAR) 3-9 a un rôle central dans l'établissement de l'hétérochromatine des chromocentres de la drosophile (Schotta et al, 2002).

1.6.5 Historique de l'ARN interférence

Les démonstrations du rôle de l'acide ribonucléique (ARN) en biologie sont très nombreuses et constitue actuellement un des domaines de recherche les plus fertiles de la biologie. Depuis l'an 2000, de nombreuses études ont décrit l'existence d'une panoplie de types d'ARN aux rôles insoupçonnés jusqu'alors. La découverte récente d'une nouvelle technique fondée sur une inhibition naturelle de l'expression génique, l'ARN interférence (ARNi), a permis des avancées importantes dans la compréhension de la régulation de l'expression génique. Dans les prochains paragraphes nous allons décrire les différents types d'ARN interférents, leur rôle et leur importance dans la régulation des éléments transposables.

Le phénomène d'ARN interférence a été mis en évidence pour la première fois par l'équipe de Richard Jorgensen en 1990. Ce chercheur souhaitait renforcer la couleur pourpre des pétunias en introduisait dans l'organisme une copie supplémentaire du gène CHSA (chalcone synthase), responsable de la coloration des fleurs. De manière surprenante, certains pétunias devenaient partiellement ou totalement blancs**.**

Une analyse plus poussée du phénomène a révélé une diminution dramatique de l'ARN messager (ARNm) de *chs*A dans le cytoplasme, sans que la transcription des gènes correspondants soit affectée (Napoli et al, 1990). Dans un premier temps, ce mécanisme induisant l'absence conjointe des ARNm endogènes et du transgène fut nommé la Co-suppression. Par la suite beaucoup de travaux ont montrés de la co-supression, sans que pour autant on comprenne exactement le processus moléculaire.

Ce phénomène fut réobservé bien plus tard dans le règne animal par les études menées par le groupe d'Andrew Fire chez le nématode *C. elegans*. Identifié sous le nom de PTGS chez les plantes pour « Post-Transcriptional Gene Silencing », ce mécanisme d'extinction de l'expression d'un gène par l'introduction d'ARN homologue, fut alors appelé « ARN interférence ». Ce mécanisme a été retrouvé chez tous les eucaryotes pluricellulaires étudiés et il intervient dans de nombreux processus.

Les petits ARNs et la régulation des éléments transposables

1.6.5.1 Les microARN (miARN)

Ce sont de petits ARN de 20-24 pb dont la production est assurée par les enzymes DICER à partir de précurseurs transcrits de zones intergéniques, introniques ou mêmes de régions codantes. Les miARN vont guider les complexes RISC (RNA-Induced Silencing Complex) aux ARNm auxquels ils s'apparient avec une spécificité total ou, le plus souvent, relative (par appariement à un site spécifique du miARN, la « seed », de 6-8 pb). Le complexe ainsi formé conduit à une inhibition de la traduction ou une lyse de l'ARNm (Figure 8). Ces miARN jouent des rôles multiples tant dans le développement que l'adaptation au stress et à la signalisation hormonale, et sont aussi impliqués dans certaines types de cancer (Petillo et al, 2009)

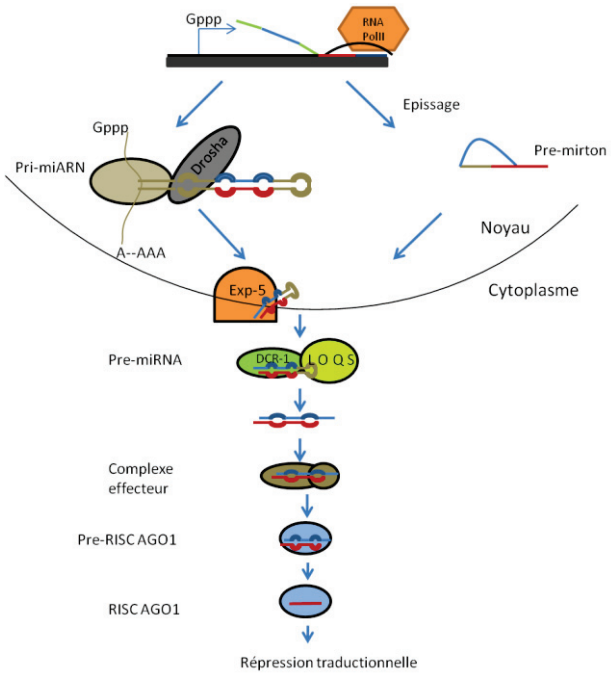


Figure 8 : voie de miARN chez *Drosophila melanogaster*.

Les micro-ARN codés par le génome sont transcrits en ARN partiellement double brin (Pri-miARN). Après leur reconnaissance par Drosha, ils sont clivés dans le noyau en pré-miARN. En parallèle, les précurseurs des miARN contenus dans les introns (Pré-mirtron) sont produits lors de l'épissage des messagers. Les pré-miARN sont ensuite transportés dans le cytoplasme pour être clivés en un duplex miARN/miARN par la protéine Dicer1. Le miARN guide associé à la protéine AGO1 va alors cibler le transcrit homologue et induire la répression traductionnelle de la cible. D'après Ghildiyal & Zamore, 2009.

1.6.5.2 Les small interfering ARN (siARN)

Les siARN, petits ARN d'environ 22 pb, sont synthétisés à partir de précurseurs ARN double brin traités par les enzymes DICER. Ces siARN peuvent avoir plusieurs origines différentes telles que les ARN viraux, des transgènes, des éléments transposables, ou encore des loci spécifiques. Ils peuvent être aussi synthétisés par des ARN polymérase ARN dépendantes (« RNA dependent RNA polymérase » (RdRp)) qui recopient des ARN simple brin (ce dernier mécanisme n'existe pas chez les mammifères). Une fois traités par les enzymes DICER, les siARN s'associent aux protéines Argonaute2 (AGO2) qui guident les ARN soit vers les transcrits ARN, soit vers l'ADN (Figure 9). Ces siARN jouent un rôle majeur dans la défense contre les virus chez la drosophile et les plantes notamment.

Comme chez les plantes, la voie des siRNA a été longtemps suspectée de jouer un rôle clé dans la défense antivirale chez les animaux, cette voie étant conservée du nématode jusqu'à l'homme. Récemment, des donnés étayent le fait que l'ARN interférence est capable de cibler l'expression des virus animaux, aussi bien en traitant des cultures cellulaires que des muqueuses avec des siARN homologues aux virus ciblés (Palliser et al, 2006) ; Tompkins et al, 2004 ; (Zhang et al, 2005). Chez la drosophile, il a été montré l'importance *in vivo*, d'un facteur clé de l'ARN interférence, Dicer2, dans le contrôle de l'expression, et donc dans la répression de trois types de virus à ARN : le Flock House Virus, le Sindbis (alpha virus), et le Drosophila C Virus (Galiana-Arnoux et al, 2006). Ils détectent également, suite à l'infection, des siARN qui leur sont homologues.

La voie d'endo-siARN

L'existence de siARN endogènes, c'est-à-dire codés par le génome a été tout d'abord mis en évidence chez les plantes et chez le nématode (Ambros, 2003), (Hamilton et al, 2002). La voie empruntée par les siARN était dès lors perçue comme un mécanisme utilisé par l'hôte pour assurer la dégradation de transcrits exogènes. Elle offrait ainsi un dispositif de défense efficace contre les virus qui, au cours de leur cycle de réplication, produisent des ARN double brin.

Néanmoins, il a été montré que l'ARNi était aussi responsable de l'inhibition de la réplication des éléments mobiles endogènes dans les tissus somatiques de la drosophile. En effet, les mutations des protéines Argonautes AGO1 et AGO2 dans les cellules S2 de drosophile augmentent l'expression des éléments transposables (Rehwinkel et al, 2006). De plus, dans des cellules KC de drosophile, des petits ARN homologues au transposon *1360* ont été détectés (Haynes et al, 2006).

Par conséquent, la découverte récente d'une nouvelle classe de siARN endogènes (endo-siARN ou esiARN) chez la souris (Tam et al, 2008) et chez *D. melanogaster* (Chung et al, 2008) (Ghildiyal et al, 2008) (Kawamura et al, 2008) n'est pas surprenante. L'équipe de Philip Zamore a ainsi détecté, *in vitro* et *in vivo*, des petits ARN d'environ 21 nucléotides dans les tissus somatiques de la drosophile (Ghilidiyal et al, 2008). La présence de ces petits

ARN a été confirmée par Greg Hanon et Julius Brennecke qui les ont par ailleurs caractérisés dans les tissus reproducteurs de la mouche (Czech et al, 2008).

La majorité des siARN endogènes caractérisés chez la drosophile dérive des éléments transposables. Leur quantité est variable selon les tissus (30% dans les têtes de mouches, 53% dans les ovaires et plus de 86% en cellules S2) (Czech et al, 2008) (Kawamura et al, 2008), et ils sont clairement impliqués dans la régulation des ET.

Les endo-siARN affichent des caractéristiques communes avec les siARN exogènes. En effet, ils se présentent sous une forme double brin d'environ 21 nucléotides avec une extrémité 3' modifiée (Kawamura et al, 2008). Leur mode de production et leur fonctionnement sont toutefois différents de la voie classique de l'interférence à ARN. En effet, leur synthèse, chez la drosophile, est dépendante des protéines DCR-2 et Loquacious, ayant été respectivement démontrées comme interagissant avec les exo-siARN et les miARN (Förstemann et al, 2005) (Saito et al, 2005). En effet la mutation des gènes de Dicer ou Argonaute augmente dramatiquement l'expression des éléments transposables étudiés tout en diminuant la quantité d'endo-siARN homologues (Chung et al, 2008) (Watanabe et al, 2006). Les trois éléments à LTR *297*, *1713* et *mdg1* contribuent à 62% des endo-siRNA dans les cellules S2.

Structures des loci producteurs de siARN

Plusieurs mécanismes sont à l'origine de la synthèse des siARN endogènes. Tout d'abord, leur production peut provenir des ET eux-mêmes. Par exemple le locus *flam/COM*, qui contrôle la mobilisation des rétrotransposons à LTR *Gypsy*, *ZAM* et *Idefix*, et contient un cluster d'endo-siARN chez *Drosophila melanogaster* (Ghildiyal et al, 2008). En revanche, le mécanisme à l'origine de la synthèse d'un ARN double brin qui pourra servir de substrat pour la production des siARN demeure inconnu. D'autre part, il existe des locus dont les séquences répétées forment, une fois transcrites, une structure en épingle à cheveux qui sert à produire directement les siARN. Les deux locus majeurs, esi1 et esi2, caractérisés chez la drosophile sont ainsi responsables de la synthèse des 20 principaux endo-siARN trouvés dans les ovaires et servent aussi à leur production dans les tissus somatiques des embryons, des larves et des adultes (Czech et al, 2008).

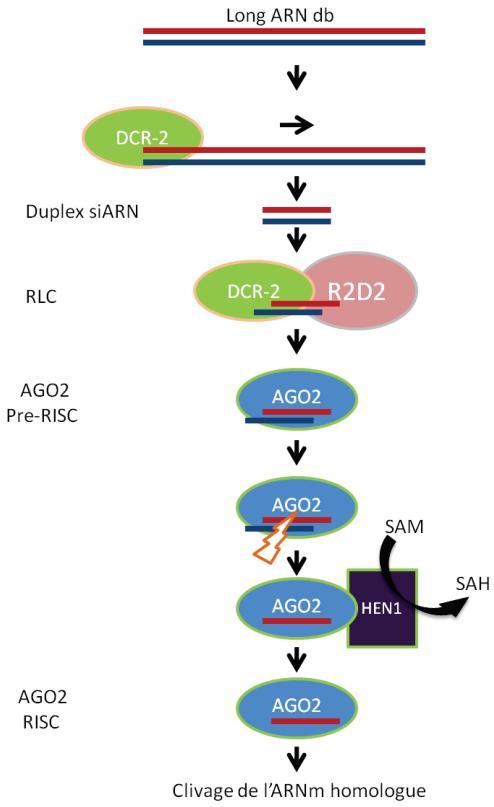


Figure 9 : Voie des siARN chez *D. melanogaster*.

Les précurseurs sous la forme d'ARN double brin sont découpés par DICER 2 (DCR-2) en un duplex de siARN. DICER-2 s'associe avec la protéine R2D2 pour former le complexe RLC qui va transférer le duplex de siARN sur la protéine Argonaute 2 (AGO2). Le brin passager (bleu) est détruit pendant que le brin guide (rouge), méthylé par HEN1, va permettre la reconnaissance par le complexe RISC des transcrits homologues. SAH : S-adénosyl homocystéine ; SAM : S-adénosyl méthionine. D'après Ghilidiyal & *Zam*ore, 2009.

1.6.5.3 Les Piwi-interacting RNA (piARN)

Nature des piARN et protéines associées

En étudiant des petits ARN dont l'action semblait différente de celle des microARN et des siARN, des chercheurs ont identifié des petits ARNs interférents d'environ 26 pb associés à des séquences répétées qui ont été isolés et dénommés *repeated associated small interfering RNAs* (rasiRNAs) (Vagin et al, 2006). Leur séquence correspond à des régions du génome de la drosophile où se situent des séquences répétées et des séquences d'ET (Brennecke et al, 2007). D'autres études ont ensuite mis en évidence une forte abondance de ces ARN de 26-30 pb dans les testicules des mammifères (Aravin & Hannon, 2008). D'autres expériences ont montré que ces petits ARNs co-précipitent avec des protéines appelées *P-element induced wimpy testis* (Piwi) d'où leur nom aujourd'hui adopté de piARN. Les protéines PIWI appartiennent à la famille des protéines Argonautes. Cette famille de protéines a été impliquée pour la première fois en 1998 dans le renouvellement autonome des cellules souches et le développement des cellules de la lignée germinale (Cox et al, 1998). La sous classe Piwi de la famille des protéines Argonautes, comprend PIWI, Aubergine (AUB) et AGO3 chez la

drosophile, MILI, MIWI et MIWI2 chez la souris et HILI, HIWI, HIWI2, HIWI3 chez l'homme (Figure 10).

Les mutations des gènes piwi, aub et ago3 conduisent à des perturbations majeures de la spermatogénèse chez l'homme et elles sont impliquées en premier lieu dans le développement de la lignée germinale et la gamétogenèse (Cox et al, 2000).

Ces protéines sont majoritairement exprimées dans la lignée germinale (Carmell et al, 2002) (Sasaki et al, 2003) où leur rôle est primordial. En effet, il a été démontré que leur inactivation bloque la production des piARN et entraîne une stérilité chez les mâles et les femelles. Par exemple, chez la souris, la mutation des gènes codants pour les protéines MILI et MIWI entraîne une baisse de la quantité de piARN et provoque des mutations dans le cycle de la spermatogénèse (Aravin et al, 2006) (Carmell et al, 2007) (Girard et al, 2006). Par ailleurs, la perte de fonction du gène aubergine chez la drosophile induit la levée de la répression du locus *stellate* et conduit à la stérilité mâle. Plus précisément, le locus *Supressor of Stellate* est à l'origine de la production de piARN antisens qui inhibent l'expression de Stellate par l'intermédiaire de la protéine AUB (Aravin et al, 2004). Le gène aubergine joue également un rôle important dans l'ovogénèse. Il a été identifié comme étant responsable de la formation des axes embryonnaires (Schüpbach & Wieschaus, 1991). Cette polarité est mise en place dans les ovaires par la localisation de morphogènes comme les facteurs de transcription *Gurken* et *Oskar* qui vont définir respectivement l'axe dorso-ventral et l'axe antéro-postérieur (Ghabrial & Schüpbach, 1999).

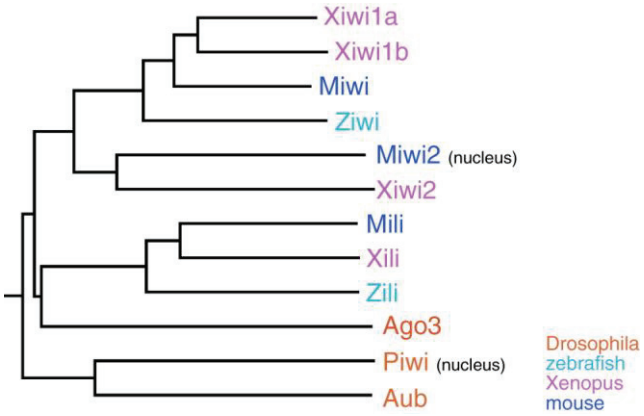


Figure 10: Arbre phylogénétique des protéines PIWI chez la drosophile, la souris (mouse), le xénope et le poisson zèbre. D'après Saito & Siomi, 2010.

Enfin, Les protéines MIWI2 chez la souris, et PIWI chez la drosophile sont impliquées dans le maintien des cellules souches à partir desquelles dérive la lignée germinale (Carmell et al, 2007) (Cox et al, 1998) (Cox et al, 2000).

La conservation de ces protéines suggère par ailleurs un mécanisme commun entre les espèces. Les protéines PIWI possèdent deux domaines : un domaine PAZ qui leur permet de se fixer sur la partie 3' des petits ARN, et un domaine PIWI qui a une structure proche des

RNAse H, et qui peut couper un brin d'ARN. Le brin coupé est complémentaire de l'ARN guide pris en charge par la molécule PIWI.

Bien que des miARNs et des siARNs soient aussi présents dans les cellules de la lignée germinale, les piARN y sont prédominants. Ces derniers ont été particulièrement étudiés chez la drosophile, le poisson, les nématodes (essentiellement chez *Caenorhabditis elegans* où ils sont appelés ARNs 21U) et la souris**.**

Les piARN sont de petits ARN simple brin légèrement plus longs que les siARNs endogènes. Par exemple, chez la drosophile, les piARN mesurent de 24 à 30pb et les siARNs ont une taille de 20 à 23 pb. Contrairement aux miARNs, l'extrémité 3' des piARN contient un groupe 2'-O-méthyl à l'extrémité 3'dont le rôle est de stabiliser les molécules d'ARN prises en charge (Horwich et al, 2007) (Kirino & Mourelatos, 2007) (Ohara et al, 2007) (Saito et al. 2007). Leurs séquences génomiques sont dispersées dans tout le génome et s'organisent en unités de transcription parfois très étendues, d'au moins plusieurs milliers de nucléotides, et pouvant chacune, proportionnellement à leur taille, générer entre 10 et 4500 piARN (Lau et al, 2006), (Watanabe et al, 2006).

L'étude de l'expression de MILI et MIWI, les orthologues de PIWI chez la souris, lors de la prophase dans des cellules de la ligné germinale, a montré que ces deux protéines suivent deux profils d'expression successifs en forme de cloche au cours du temps (Watanabe et al, 2006). Tandis que l'expression de MILI ne dépasse pas le pachytène (3éme phase de la prophase I), MIWI s'exprime dans les spermatides, c'est-à-dire les cellules de la lignée germinale issues de la deuxième division méiotique. Les quantités des piARN de 24-28 nt et des piARN de 29-32 nt suivent respectivement les profils de MILI et MIWI indiquant que chaque protéine interagit probablement avec sa propre catégorie de piARN.

Chez les mammifères, les piARN se partagent ainsi en deux catégories : des piARN dits « pré-pachytène » et des piARN dits « pachytène ». Comme leur nom l'indique, les piARN « pré-pachytène » sont exprimés avant le stade pachytène de la méiose. Ils proviennent de régions du génome riches en éléments transposables et en séquences répétées. Ces piARN interagissent avec les protéines PIWI, MILI, et MIWI2 et seraient l'objet du cycle ping-pong décrit plus loin.

Inversement, les piARN « pachytène » sont exprimés pendant le stade pachytène et sont associés aux protéines PIWI, MILI et MIWI. Plus de 80 000 piARN « pachytène » différents ont été décrits dans les spermatocytes de mammifères. Pourtant, leurs séquences ne donnent aucune idée de leur origine génomique (Kim et al, 2010).

Chez la drosophile, des criblages ont également été réalisés (Thomson & Lin, 2009). L'un d'eux a révélé l'existence d'environ 30 000 piARN répartis en deux populations : l'une est associée à une protéine Argonaute (AGO3), et l'autre à deux protéines à domaine PAZ et PIWI, les protéines PIWI et Aubergine (AUB) (Brennecke et al, 2007). Parmi les protéines PIWI, piwi est localisé dans le noyau (Cox et al, 2000), et sa fonction principale dans la régulation des ET semble être dans le noyau (Saito et al, 2010). Dans les ovaires, Piwi est

exprimée dans les cellules somatiques et les cellules germinales (Cox et al, 2000) (Saito et al, 2006) (Brennecke et al, 2007). Par contre AUB et AGO3 sont exprimées dans les cellules germinales mais sont absentes des cellules somatiques. Contrairement à PIWI, AUB et AGO3 sont localisées dans le cytoplasme et s'accumulent dans le nuage (Brennecke et al, 2007) (Gunawardane et al, 2007) (Li et al, 2009) (Figure 11). Ces observations indiquent que les protéines PIWI (Piwi, AUB et AGO3), ont des fonctions diverses.

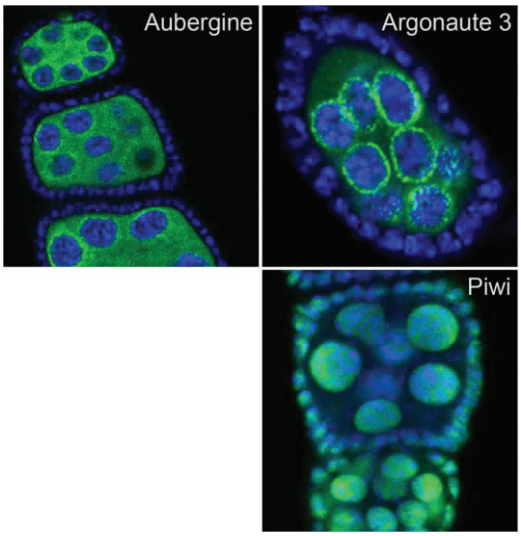


Figure 11. Localisation des protéines Argonautes de la sous classe PIWI dans les ovaires de drosophile. L'ADN est maqué en bleu, les protéines Argonautes en vert. D'après Brennecke et al, 2007.

Les piARN associés à PIWI/AUB correspondent souvent à des séquences complémentaires d'éléments transposables et leur premier nucléotide est très souvent une uridine (1U-piARN). Les piARN associés à AGO3 ont comme point commun une adénine en 10éme position (10A-piARN). Enfin, les 1U-piARN et les 10A-piARN correspondraient à des séquences complémentaires pouvant s'apparier (Brennecke et al, 2007) (Figure 12). Les piARN semblent jouer un rôle important dans le développement de la lignée germinale et le maintien de l'intégrité des génomes en contrôlant l'activité des éléments transposables.

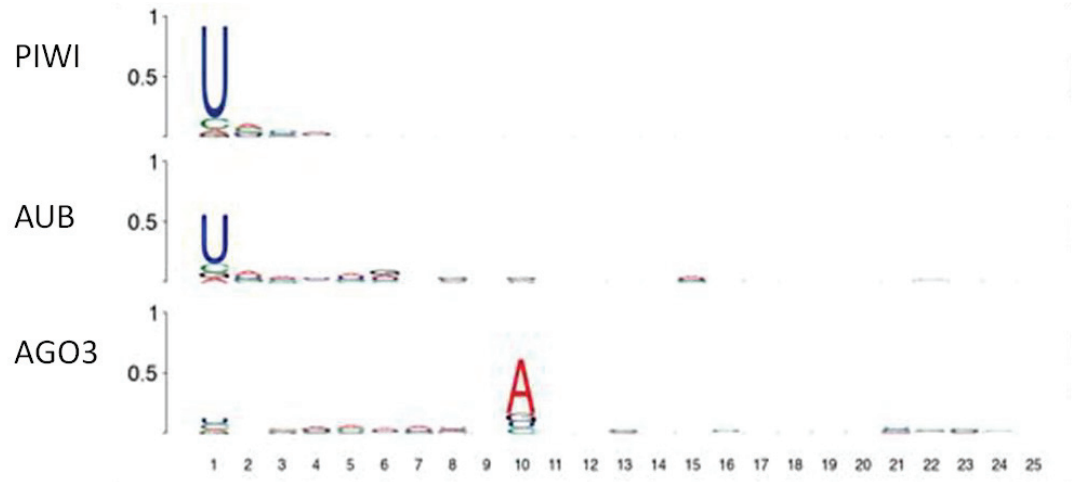


Figure 12 : le biais relatif de composition en nucléotides pour chaque position dans tous les piARN obtenus à partir de complexes PIWI, AUB et AGO3. D'après Brennecke et al, 2007.

Biosynthèse des piARN

Des analyses approfondies des piARN associés aux protéines PIWI, chez la drosophile et la souris, ont permis d'identifier les origines génomiques des piARN et conduit à la proposition de l'existence de deux voies de biogenèse : le « primary processing pathway » et le cycle d'amplification « ping-pong ». Les deux mécanismes sont importants pour contrôler l'activité des ET dans la lignée somatique et la lignée germinale. Ces deux voies sont conservées chez de nombreuses espèces animales ; elles ont été identifiées chez les rats, les grenouilles et les *Bombyx mori* (Lau et al, 2006) (Robine et al, 2009). Chez la drosophile, la biogenèse de piARN primaires se fait dans les cellules folliculaires ainsi que dans les cellules nourricières (germinales). Par contre la biogenèse des piARN secondaires par la voie d'amplification « ping-pong » est spécifique de la lignée germinale.

Comme décrit ci-dessus, les loci dans lesquels se trouve l'ADN codantr des piARN sont appelés des régions (clusters) piARN (Brennecke et al, 2007). Les clusters de piARN ont une taille qui varie entre 1 et 200 kb et chaque cluster contient de multiples séquences qui génèrent des piARN. Les piARN sont issus de la maturation de précurseurs de très grande taille correspondant à de longs ARN simple brin (Lau et al, 2006). Chez la drosophile, ils sont le plus souvent présents au niveau des régions pericentromériques et télomériques, qui sont composées de séquences d'éléments transposables défectueux (Brennecke et al, 2007). Les loci qui produisent les piARN se divisent en deux groupes : les loci qui fournissent des piARN sens et anti sens, qui sont bidirectionnels (Figure 13), comme le locus 42AB sur le chromosome 2 chez la drosophile, et les loci qui ne fournissent que des piARN du même sens, qui sont unidirectionnels (Figure 13), comme le locus flamenco (*flam*) sur le chromosome X de la drosophile (Brennecke et al, 2007).

Contrairement à celle des miARN et des siARN, la biogenèse des piARN ne fait pas appel à la protéine DICER et n'est pas effectuée à partir de précurseurs ARN double brin mais plutôt d'ARN simple brin (Vagin et al, 2006). Certaines observations obtenues à partir des RT-PCR et des bases de données soutiennent l'idée que les ARN simple brin seraient

effectivement transcrits à partir du cluster piARN (Brennecke et al, 2007) Les données génétiques chez *D. melanogaster* suggèrent que les longues molécules d'ARN qui englobent tout le cluster des piARN, sont transformées en piARN matures : par exemple, l'insertion de l'élément P à l'extrémité 5' du locus flamenco, qui interrompt la transcription de la longue molécule d'ARN, conduit également à l'élimination de tous les piARN de ce cluster (Malone et al, 2009) (Brennecke et al, 2007).

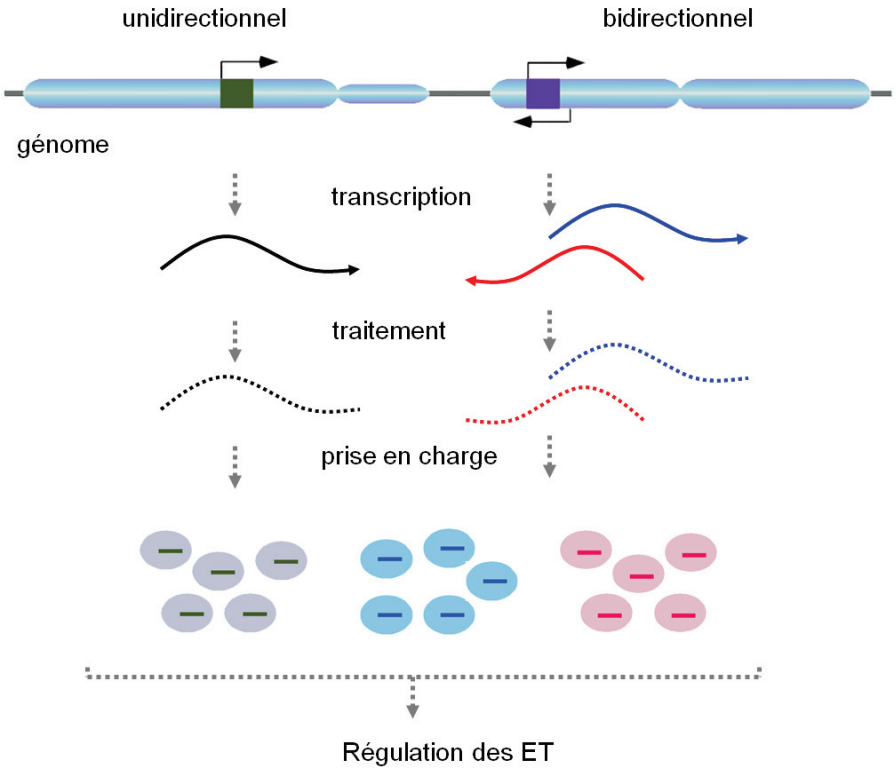


Figure 13: les locus unidirectionnels et bidirectionnels de production des piARN.

Les locus unidirectionnels sont transcrits dans une seule direction, par contre les locus bidirectionnels génèrent des piARN dans les deux sens. Les transcrits sens (bleu) et l'anti-sens (rouge) sont complémentaires l'un par rapport à l'autre. D'après Saito & Siomi, 2010.

Il reste néanmoins que le mécanisme précis de la biogenèse primaire reste inconnu. On en connait néanmoins quelques acteurs. Plusieurs protéines semblent impliquées dans la biogenèse des piARN primaires chez *D. melanogaster* comme l'ARN hélicase putative Armitage (Armi), la protéine YB (Fs(1) également connu comme Yb, et la nucléase putative Zucchini (ZUC). Des mutations dans l'une de ces trois protéines éliminent les piARN primaires associés avec la protéine PIWI dans les cellules somatiques de l'ovaire (Olivieri et al, 2010) (Haase et al, 2010) (Saito et al, 2010). ARMI et ZUC sont également requis pour l'accumulation des piARN dans les cellules germinales, alors que Yb est exclusivement exprimé dans les cellules folliculaires (Olivieri et al, 2010) (Haase et al, 2010). Zucchini interagit physiquement avec AUB (Pane et al, 2007) et la perte de fonction de zucchini entraine une diminution des piARN produits par les clusters (piARN primaires), mais n'affecte pas la production des piARN secondaires, ce qui montre bien que Zucchini

intervient dans la phase d'initiation de la biogenèse des piARN primaires (Haase et al, 2010) , (Olivieri et al, 2010).

L'un des exemples les mieux étudiés de production de piARN primaires vient de l'analyse du locus *flamenco*/COM chez *D. melanogaster*, qui constitue une source majeure de piARN anti-sens dans les cellules somatiques des ovaires (Brennecke et al, 2007).

En effet la majorité des fragments des transposons insérés dans ce locus correspondent aux rétrotransposons de la famille *Gypsy* (*ZAM*, *Idefix*, etc..), qui envahissent la lignée germinale *via* les cellules folliculaires (somatiques) (Sarot et al, 2004) (Desset et al, 2008). Le locus *flam*/COM (180 KB) localisé dans l'hétérochromatine péricentromérique du chomosome X serait responsable de 2,18% des piARN totaux et constituerait par conséquent le deuxième cluster génomique de piARN chez la drosophile. 90% des fragments des ET insérés dans ce locus sont orientés dans le sens inverse de la transcription, ce qui favorise la production de piARN anti-sens qui ciblent les transcrits sens des éléments (Brennecke et al, 2007).

Un soutien fort à ce modèle découle d'une analyse de fragments du rétrotransposon *ZAM* qui sont situés au sein du locus *flamenco*/COM : seuls les fragments de *ZAM* qu'on trouve dans flamenco produisent abondamment des piARN (Brennecke et al, 2007) (Senti & Brennecke, 2010). Par ailleurs, la lignée analysée ne possède pas des copies actives de *ZAM*, ce qui montre que la production de piARN dans le soma est indépendante de l'expression des éléments actifs. De même, dans le cas de l'élément *Gypsy*, les quantités des piARN somatiques qui sont dérivés de cet élément ne sont pas influencées par la présence de copies actives de *Gypsy* dans le génome (Pelisson et al, 2001). 94% des piARN produits par le locus flamenco sont associés à la protéine Piwi. Par ailleurs la perte de fonction du gène piwi induit la levée de répression du rétrovirus endogène *Gypsy*, qui s'accumule dans les cellules folliculaires des ovaires le long de la membrane apicale à la jonction avec l'ovocyte (Olivieri et al, 2010). Le locus flamenco, mis en évidence chez *D. melanogaster* est conservé chez *D. erecta* et *D. yakuba*, même si la qualité des séquences de cette région n'est pas très bonne (Malone et al, 2009).

Plusieurs études ont montré que des populations des piARN primaires de *D. melanogaster* pourraient également être impliquées dans la régulation d'un gène codant des protéines d'adhésion cellulaire de type immunoglobuline. Par exemple, la Fascicline 3 (Li et al, 2003). Les transcrits de ce gène seraient ciblés par des piARN dans les cellules somatiques des ovaires chez la drosophile. De même, les piARN associés à AUB dans les testicules pourraient cibler le gène vasa (Nishida et al, 2007). Certains piARN peuvent induire la dégradation des ARNm déposés maternellement, tels que Nanos, un morphogène postérieur de la drosophile au cours de la transition maternel–zygotique chez les embryons (Rouget et al, 2010). Ces résultats encore préliminaires, suggèrent que la fonction des piARN s'étend au-delà de régulation des ET.

Dans certains types de cellules, la biogenèse des piARN primaires est le seul mécanisme qui génère des piARN. En particulier tous les piARN produits dans les cellules somatiques

folliculaires des ovaires chez la drosophile et les piARN pachytènes exprimés durant la méiose dans les spermatocytes chez la souris sont générés par ce processus primaire (*primary processing*) (Aravin & Hannon, 2008). Cependant dans d'autres types cellulaires, tels que les cellules de la lignée germinale femelle chez *D. melanogaster* et les spermatogonies pré-méiotiques chez la souris, il existe un autre mécanisme appelé le ping-pong qui amplifie des séquences spécifiques qui ont été générées par la voie de la biogenèse des piARN primaires.

Amplification post-transcriptionnelle des piARN : locus germinal et le modèle Ping-pong

Chez la drosophile, les piARN que l'on retrouve dans la lignée germinale proviennent des différents clusters des piARN, qui ont été décrits ci-dessus, mais peuvent également provenir de transcrits d'éléments actifs, selon une voie appelée « ping-pong ». Le système le mieux connu concerne l'élément *I* (un rétroélément de la famille des LINEs), pour lequel les éléments actifs qui sont présents dans le génome ont divergé suffisamment pour permettre de les différencier sur la base de leurs séquences nucléotidiques (Brennecke et al, 2008). De façon similaire aux clusters à expression somatique, les clusters à expression germinale sont fortement enrichis en fragments d'ET (Brennecke et al, 2007). Il n'est pas surprenant que les clusters «germinaux» contiennent une collection plus diverse de fragments d'éléments transposables puisque l'activité de ces derniers, si elle n'est pas contrainte dans la lignée germinale, est délétère (Malone et al, 2009). Le cluster situé en position 42AB est un cluster spécifique de la lignée germinale, qui produit des piARN à partir des deux brins d'ADN, indiquant ainsi une transcription bidirectionnelle (Figure 13) (Brennecke et al, 2007). Un autre cluster germinal a été mis en évidence dans la région 20A. Ce dernier est le seul cluster de piARN qui ressemble à flamenco puisqu'il est transcrit de façon unidirectionnelle et contient seulement des fragments anti-sens des ET (Brennecke et al, 2007) (Malone et al, 2009). Enfin les clusters germinaux de piARN sont également trouvés au niveau des télomères, où un nombre abondant de piARN dérive d'arrangements télomériques des éléments *HeTA, TART* et *TAHRE* ainsi que des répétitions satellites subtélomériques (Brennecke et al, 2007). Ces derniers clusters sont moins étudiés que le cluster 42AB, mais montrent que plusieurs clusters sont présents dans les génomes.

Cycle d'amplification « Ping-pong »

Chez la drosophile, les piARN primaires qui sont associés avec PIWI et AUB sont anti-sens par rapport aux transcrits actifs des ET. Par contre, le clonage des piARN associés à la protéine AGO3 a montré que ces séquences ont une orientation sens (Gunawardane et al, 2007). Il est supposé que la transcription des locus péricentromériques aboutit à la formation des petits ARN anti-sens (les 1U-piRNA), qui correspondent donc au piARN primaires. Ces piARN sont pris en charge par les protéines Piwi et Aubergine et ciblent les ARN sens qui leurs sont homologues (les transcrits des ET, par exemple). Elles les clivent 10 pb à partir de l'extrémité 5' du duplex piARN/ARNm (Gunawardane et al, 2007) (Saito et al, 2006), créant ainsi des petits ARN sens (piARN secondaires) (les 10A-piARN, qui ont une adénine en position 10) sont alors pris en charge par la protéine AGO3. Ce nouveau complexe piARN/AGO3 cible à son tour les transcrits issus des régions péricentromériques créant ainsi

une boucle d'amplification appelé modèle « ping-pong » (Figure 14) (Brennecke et al, 2007). Les piARN produits par le cycle ping-pong sont nommés piARN secondaires.

Les acteurs centraux dans ce cycle de «Ping-pong » sont AUB et AGO3, qui sont localisés dans le cytoplasme des cellules nourricières autour du noyau. Le mécanisme ping-pong fait probablement intervenir l'activité RNase H (slicer) de ces deux protéines mais pas de la protéine PIWI.

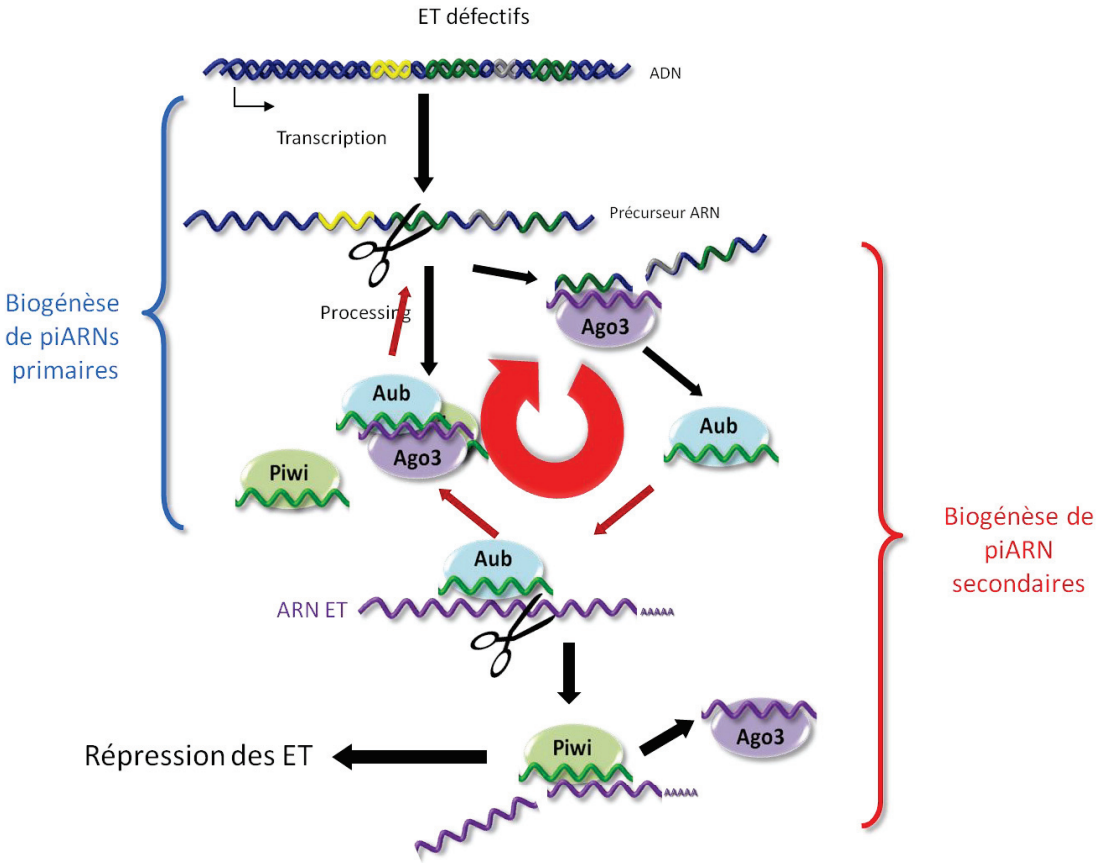


Figure 14: Mécanisme de biogenèse des piARN. D'après (Brennecke et al, 2007) (Gunawardane et al, 2007).

Néanmoins, le lien entre la biogenèse des piARN primaires et le cycle de « ping-pong » reste mal connu. Le signal de ping-pong est principalement associé aux couples de protéines AUB-AGO3 et AUB-AUB (Brenencke et al, 2007). Dans les mutants Aub le cycle de ping-pong est supprimé (Brennecke et al, 2007), alors que pour les mutants AGO3 les niveaux de piARN ne sont que peu diminués (Li et al, 2007). Par contre dans les mutants Piwi, le cycle de ping-pong n'est pas affecté, ce qui montre que la protéine Piwi n'intervient pas dans cette voie. L'amplification des piARN a également pu être mise en évidence chez les mammifères chez qui le modèle ping-pong a pu être en partie validé par le clonage des protéines MILI et MIWI2 qui pourraient correspondre respectivement aux protéines de la drosophile AUB/Piwi et AGO3 (Aravin & Hannon, 2008). Au cours du cycle ping-pong durant la spermatogenèse chez la souris, les piARN primaires sont synthétisés à partir des transcrits des ET et des transcrits issus des clusters. Ils sont alors associés à des protéines MILI (appelées aussi PIWI-like protein 2 (Aravin & Hannon, 2008). Ces piARN sens associés aux protéines MILI clivent les transcrits anti-sens générant alors des piARN secondaires qui s'associent à des protéines

MIWI2 (appelées aussi PIWI-like protein 4). MIWI2 à son tour clive les transcrits sens et génère de nouveaux transcrits piARN sens qui s'associent alors aux protéines MILI (Aravin & Hannon, 2008).

Les piARN sont des éléments clés d'une sorte de système immunitaire à deux composantes : une composante génétique codée par des loci hétérochromatiques contenant des copies défectives d'ET (piRNA clusters) produisant des piARN anti-sens, et une composante adaptative correspondant aux piARN produits par les copies fonctionnelles des ET localisées dans l'euchromatine.

En conclusion, les piARN sont produits par deux voies différentes. Les piARN primaires dérivent d'un long précurseur simple brin présent dans un cluster composé d'un très grand nombre de copies défectives d'ET. Ces piARN en association avec PIWI ou Aubergine ciblent les transcrits des ET qui sont homologues complémentaires, et produisent ensuite des piARN secondaires. Cela signifie que les piARN peuvent amplifier le signal de silencing et en même temps réguler les transcrits des ET.

L'importance de la boucle d'amplification ping-pong dans la lutte contre l'envahissement de la lignée germinale femelle par les ET est bien démontrée par l'étude du mécanisme de dysgénésie des hybrides chez *D. melanogaster*.

Transmission maternelle des piARN et dysgénésie des hybrides

La drosophile est un excellent modèle pour l'étude des mécanismes qui contrôlent l'expression des rétroéléments. Les travaux réalisés chez cet organisme ont permis de caractériser différentes instabilités liées aux éléments transposables, dont la plus connue est la dysgénésie des hybrides. Ce syndrome se caractérise par une stérilité femelle, des réarrangements chromosomiques, une haute fréquence de mutations dans la lignée germinale ainsi qu'une distorsion de ségrégation, apparaissant lors de certains types de croisements (Kidwell 1997). La dysgénésie des hybrides fut d'abord décrite pour le facteur *I* et l'élément P chez *D. melanogaster* (Bucheton et al, 1976) (Picard, 1976) (Picard et al, 1978) (Kidwell et al, 1977), avant d'être mise en évidence pour d'autres éléments et espèces tels que *Penelope* chez *D. virilis* (Petrov, 1995).

Ces symptômes, liés à une forte mobilisation des éléments, apparaissent dans la lignée germinale d'individus issus du croisement entre des mâles possédant des éléments actifs et des femelles qui ne possèdent pas d'élément (Kidwell et al, 1977). Malgré leurs différences de fécondité, les descendants des croisements dysgéniques (stériles) et non-dysgéniques (fertile) sont génétiquement identiques, ce qui suggère par ailleurs la participation d'une composante cytoplasmique des femelles qui assure le maintien de la répression au travers des générations.

La première preuve du rôle des petits ARN transmis maternellement dans la régulation des ET a été obtenue chez *D. virilis*, où la dysgénésie des hybrides est caractérisée par la mobilisation de plusieurs familles d'ET y compris les rétrotransposons de la famille Penelope (Blumenstiel & Hartl, 2005). Il n'en reste pas moins que le décorticage des voies de

régulation par les petits ARN n'a été fait que plus récemment, chez *D. melanogaster* (Brennecke et al, 2008) (Chambeyron et al, 2008) avec les outils génétiques que l'on connaît.

Pendant le développement des ovaires, les protéines AUB et Piwi s'accumulent dans le pôle postérieur de l'ovocyte, où les déterminants des cellules germinales sont localisés (Rongo & Lehmann, 1996). Les deux protéines AUB et PIWI déposées maternellement sont incorporées dans les cellules germinales primordiales et sont ensuite directement transmises à la génération suivante (Figure 15).

En revanche AGO3 ne semble pas être transmis de la mère à sa descendance, ce qui suggère que les piARN antisens peuvent être sélectivement hérités, tandis que les piARN sens ne le sont pas. En parallèle du modèle ping-pong proposé par Brennecke et al, 2007, il est avancé l'hypothèse que de petits ARN non codants présents dans le cytoplasme maternel seraient transmis à la descendance et pourraient réguler les éléments transposables. Ces piARN maternels forment une couche protectrice contre les ET, même avant l'initiation de la transcription dans les futures cellules germinales. L'étude de Brennecke et al, (2008) se concentre sur deux systèmes de dysgénésie bien caractérisés chez *D. melanogaster*, le système I-R et le système P-M décrit ci-dessus, relatifs respectivement à la dérépression de l'élément *I* et du transposon *P* (Figure16).

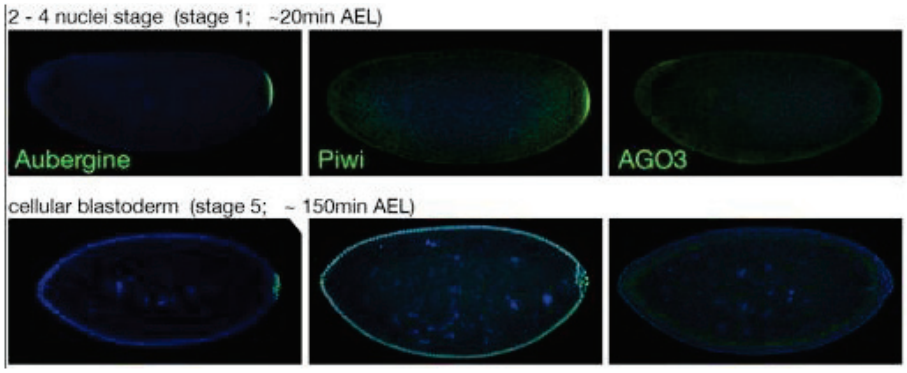


Figure 15: L'héritage maternel de la famille de protéines PIWI. ADN (bleu), Aubergine, Piwi et AGO3 (vert). D'après (Brennecke et al, 2008).

Brennecke et al (2008) ont montré que les piARN déposés maternellement sont nécessaires pour initier la boucle d'amplification ping-pong dans la génération F1. Deux possibilités peuvent expliquer cette observation : (1) les piARN maternels sont nécessaires pour lancer le cycle de ping-pong et (2) les piARN maternels ont une influence sur l'état de la chromatine des clusters des piARN et /ou des éléments transposables. Leur absence, en dépit de la présence des transcrits sens et anti-sens des ET, provoque des échecs à la fois dans l'accumulation des piARN et la répression de l'élément. Cette exigence créé une énigme par rapport à ce qui arrive en premier : le ping-pong des piARN ou les piARN déposés maternellement nécessaires pour le démarrage du ping-pong. Il est vraisemblable que ceci soit dû à l'accumulation progressive, à la fois, du nombre de copies de l'élément *I*, et des piARN nécessaires pour réprimer l'expression de restes d'un ancien élément *I* dans le cluster de piARN 42AB. Il reste quand même difficile de savoir quelle est la première étape dans

l'adaptation au contrôle d'un nouvel élément : d'abord *via* les piARN dans les ovaires d'un adulte ou le franchissement d'un seuil avec le dépôt de piARN suffisants pour démarrer le contrôle. De toute façon, il semble clair que l'engagement dans la voie doit être précédé par la transposition d'un élément dans un cluster de piARN (Senti & Brennecke 2010).

La pénétrance du phénotype dysgénique peut également être influencée par des facteurs externes, y compris la température à laquelle la femelle est élevée et son âge (Bucheton et al, 1976). Bien qu'elle doive encore être prouvée, une hypothèse selon laquelle l'environnement peut influencer le contenu en petits ARN maternels, et donc modifier le phénotype de la descendance d'une manière héréditaire en utilisant les petits ARN en tant que vecteur pour transmettre l'information épigénétique, est envisageable (Saito & Siomi, 2010).

Jusqu'à présent la transmission des caractères par le biais de petits ARN n'a été observée que chez la drosophile. Toutefois, les petits ARN ou leurs partenaires s'accumulent dans les ovocytes d'autres espèces (Houwing et al, 2007) (Watanabe et al, 2006), suggérant la possibilité de rôles identiques pour les voies des petits ARN avec des effets maternels sur les phénotypes de leurs descendances.

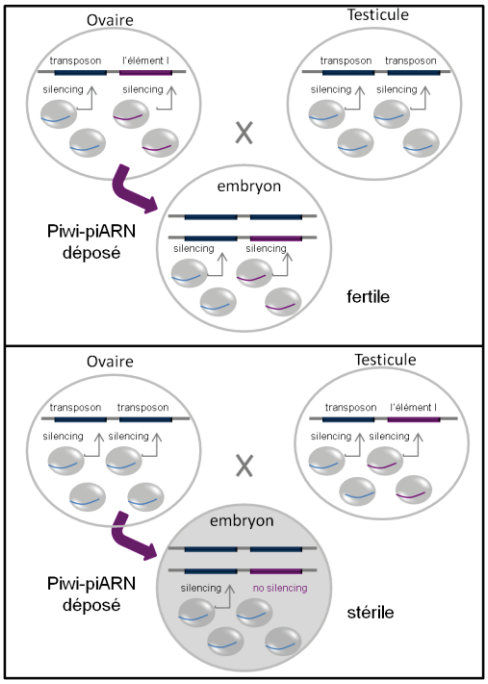


Figure 16 : Les piARN intervenant comme vecteur d'information épigénétique.

Un phénomène, appelé dysgénésie des hybrides, est observé lorsque des souches de drosophiles contenant un ensemble différent d'ET sont croisées. Les descendances sont fertiles, quand les piARN de l'ET (comme l'élément *I*) sont déposés par la mère (A). Toutefois, lorsque les ET sont hérités du père, la descendance est stérile (B). Le manque de dépôt maternel des complexes Piwi-piARN est apparemment la cause principale des dysgénésies des hybrides.

Les différentes protéines de la voie piARN.

Il est difficile de faire une liste exhaustive de toutes les protéines qui interviennent directement ou indirectement dans la voie des piARN. Ci-après j'en décris quelques-unes pour lesquelles les effets sont relativement bien étudiés chez la Drosophile (tableau 1).

Mutant	Fonction	Expression	Biogénèse des piARN primaires	Ping-pong	Derépression des ET
spindle-E	**ARN hélicase**	germinale	+++	---	Deux éléments télomériques germinaux, *TART* et *HeT-*A, et les éléments *mdg 1* , *Beagle* , *gate*
armitage		somatique + germinale	+ /-	+++	Retrotransposons somatiques, *Gypsy* , *Zam* , *Idefix* , et tabor
vasa		germinale	+++	---	Derépression des ET dans la lignée germinale
yb	**Domaine Tudor**	somatique +germinale	---	+++	Retrotransposons somatiques *Gypsy, Zam,Idefix* , et *Tabor*
vreteno					
tejas		germinale	+++	+/-	Elément *I* dans la lignée germinale ainsi que *TART* et *HeT-A* et *roo*
krimper				---	
zucchini	**Putative nucléase**	somatique + germinale	+/-	+++	Retrotransposons somatiques *Gypsy*, *Zam* , *Idefix* et *Tabor*
mael	**?**	germinale	+++	---	Eléments H*MS-Beagle* et *Gate*

Tableau 1 : Mutants de la voie piARN et leurs effets.

(+++): Fonctionnel/ avec production des piARN, (---) : Non fonctionnel, pas des piARN ; (+/-) peu des piARN.

La perte de fonction de zucchini (*zuc*) et de squash (*squ*) affecte la production des piARN et entraine la dérépression des ET (Pane et al, 2007). Les protéines *Zuc* et *squ* sont des nucléases qui interagissent directement avec la protéine AUB. Il est possible que ces nucléases soient responsables de la maturation des piARN à la place des protéines *DICER* (Pane et al, 2007).

Armitage et spindle-E (spn-E) sont des gènes qui codent des hélicases putatives qui sont nécessaires pour la production des piARN (Aravin et al, 2004) (Vagin et al, 2006). Dans les mutants spindle-E, il a été observé une réduction de l'amplification des piARN et de la production de piARN provenant du cluster 42AB (Malone et al, 2009). Il a été avancé que ces deux protéines seraient responsables de la séparation des intermédiaires doubles brins formés lors de la production des piARN, au moment de la reconnaissance de la cible ARN ou lors de son clivage.

Enfin *Maelstrom (Mael)* et *cutoff* (*cuff)* sont deux gènes qui jouent un rôle dans la voie des piARN (Chen et al, 2007) (Findley et al, 2003). L'inactivation de la protéine *Cuff* reproduit le phénotype ovarien observé lors de la mutation d'aubergine. En revanche, l'absence de la protéine *cutoff* n'interdit pas la présence de piARN homologues au rétrotransposon à LTR *roo*. *Cuff* est donc impliqué en aval de la synthèse des piARN sans doute en facilitant l'action des complexes de protéines PIWI.

Les protéines PIWI, ne sont pas les seuls facteurs qui ont été mis en évidence dans la voie des piARN chez *D. melanogaster*. L'analyse des modifications post-traductionnelles des protéines PIWI a démontré qu'elles possèdent un domaine symétrique dimethylarginine (sDMA) qui lui permet d'interagir avec le domaine protéique Tudor (Kirino et al, 2009), (Nishida et al, 2009) (Liu et al, 2010).

Le domaine protéique TUDOR est un motif se fixant symétriquement sur des arginines diméthylées (domaine sDMA, symetrically dimethylated arginines) (Boswell & Mahowald, 1985) (Thomson & Lasko, 2004) (Arkov et al, 2006). Les protéines possédant des domaines TUDOR ont un rôle majeur dans la localisation des complexes PIWI/piARNs au niveau des ovocytes et des cellules spermatiques, mais leur rôle précis n'a pas encore été élucidé (Vagin et al, 2009). Chez la souris et la drosophile l'absence des gènes tudor aboutit à une diminution de la synthèse des piARN et de la stabilité des protéines PIWI (Nishida et al, 2009). Par ailleurs, les protéines TUDOR jouent un rôle important dans le cycle ping-pong, en interaction avec d'autres complexes protéiques. Il existe donc une interaction complexe entre les protéines PIWI et TUDOR, qui est encore mal connue.

Localisation cellulaire des voies de synthèse des piARN

Les Yb-bodies

Chez la drosophile, la synthèse des piARN primaires associés à la répression somatique des ET et au cluster *flam*/COM prend place dans les cellules somatiques ovariennes. Beaucoup de protéines sont nécessaires pour la voie des piARN germinaux, comme SPN-E, VASA, MAEL, TUD et KRIMP sont aussi indispensables pour la production des piARN primaires dans les cellules somatiques, ce qui montre un chevauchement des deux voies dans l'utilisation de ces effecteurs (Siomi et al, 2010). Des études récentes ont démontrées que la synthèse des piARN primaires dans les cellules folliculaires somatiques est associée à des granules particulières appelées corps Yb (Figure 17C), dont le principal composant est une protéine dénommée FS(1)YB et dans lesquels on trouve aussi la protéine Piwi. Les précurseurs des piARN sont internalisés dans ces corps Yb et transformés en piARN matures en présence notamment de FS(1)YB, qui s'associe avec la protéine ARMI (Figure 17B). La protéine ARMI, qui est aussi fondamentale dans la biogenèse des piARN secondaires (voir la suite), serait ici présente avec une isoforme qui est spécifique de la lignée somatique (Olivieri et al, 2010).

Les deux protéines Armi et Yb contiennent un domaine ARN hélicase et sont donc impliquées dans le métabolisme de l'ARN. Une mutation dans Armi entraine une baisse

sévère dans l'accumulation des piARN dans les ovaires (Vagin et al, 2006). Il a été proposé que, dans le cas de perte de fonction d'Armi ou Yb dans les cellules somatiques de l'ovaire, la protéine Piwi ne prend plus en charge les piARN, ce qui empêche la formation de complexe Piwi/piARN et la protéine Piwi reste dans le cytoplasme plutôt que de s'accumuler dans le noyau, ce qui explique la perte observé de piARN produits dans le soma et l'activation de l'expression des rétrotransposons (Haase et al, 2010) ; (Olivieri et al, 2010). Ceci suggère que Armi et Yb sont des modulateurs de biogenèse des piARN somatiques, ces 2 protéines stabilisant ou guidant le complexe Piwi/piARN pour dégrader les transcrits des rétrotransposons comme *Gypsy, ZAM* et *Idefix.*

Le rôle du gène armi dans la voie de synthèse des piARN primaires de la voie somatique a été aussi démontré par la suppression du gène armi par RNAi dans les cellules OSS (Ovarian Somatic Sheet Cell line). Cette mutation s'accompagne d'une perte de régulation des rétrotransposons de type *Gypsy* (*Gypsy*, *blood* et *Idefix*) qui est caractérisée par une augmentation du taux de transcrits étudié par qPCR (Haase et al, 2010). Cette dérégulation pour les rétrotransposons de type *Gypsy* a été montrée aussi pour les mutants du gène zucchini pour lesquels on observe 20 fois plus de transcrits de *ZAM*, *Idefix* et *Gypsy* dans les mutants nuls par rapport aux mutants hétérozygotes (Haase et al, 2010).

Le nuage

Les protéines de la famille PIWI sont localisées dans le cytoplasme et dans le noyau de la cellule. Dans le cytoplasme, les protéines PIWI ainsi que d'autres complexes protéiques sont co-localisés dans des structures cytoplasmiques granulaires qui sont présentes dans la lignée germinale chez différentes espèces de métazoaires. Chez la drosophile, les protéines Aubergine et AGO3 (impliquées dans le cycle ping-pong) sont localisées dans le « nuage » (Chen et al, 2007) (Lim & Kai, 2007) (Pane et al, 2007). Le nuage est une structure périnucléaire localisée dans le cytoplasme des cellules nourricières de l'ovaire (Figure 17A). Le nuage est impliqué dans la voie des piARN, car la mutation de la protéine Krimper, qui contient un domaine Tudor et qui est exprimée dans le cytoplasme des cellules nourricières, bloque son assemblage, ce qui provoque une altération de la production des piARN et de la répression de l'expression des ET (Lim & Kai, 2007). En plus des protéines AUB et AGO3, les cellules germinales de l'ovaire expriment aussi la protéine PIWI qui est localisée dans le noyau. Ces observations suggèrent donc une compartimentation de la production des piARN et de leur fonction (Lim & Kai, 2007). Dans ce modèle, les ARN sont transcrits dans le noyau à partir des clusters de piARN et sont ensuite exportés dans le nuage.

Outre les protéines PIWI, d'autres protéines (faisant intervenir les domaines TUDOR), comme les nucléases *zucchini* et *squash*, les hélicases *armitage* et *spindle-E*, Maelstrom et Cutoff, sont co-localisées dans le nuage (Figure 17B). Une quantité insuffisante de ces protéines bloque la production des piARN. L'un de ces complexes protéiques y compris SPN-E et KRIMP possèdent un domaine TUDOR (Lim & Kai, 2007), tandis que d'autres comme AUB et AGO3 possèdent un sDMAs « symmetrical dimethylarginines » qui sont reconnus par les domaines Tudor (Kirino et al, 2009) (Nishida et al, 2009) (Liu et al, 2010) (Siomi et al, 2010). Par conséquent une interaction protéine-protéine au sein du nuage est souvent médiée

par l'association des domaines Tudor-sDMA. Le nuage serait un site de biogenèse des piARN et/ou de destruction des piARN (Lim & Kai, 2007).

Chez la souris on retrouve le même type de compartimentation : la production des piARN dans les cellules somatiques a lieu dans le corps P (P-bodies, processing bodies) et la production de piARN germinaux, dans des corps chromatoides (CB).

Ces données indiquent que les structures granulaires cytoplasmiques sont indispensables pour les mécanismes liés aux piARN, à la fois dans les cellules germinales et dans les cellules somatiques folliculaires, bien que les compositions des granules soient différentes dans ces deux types cellulaires**.**

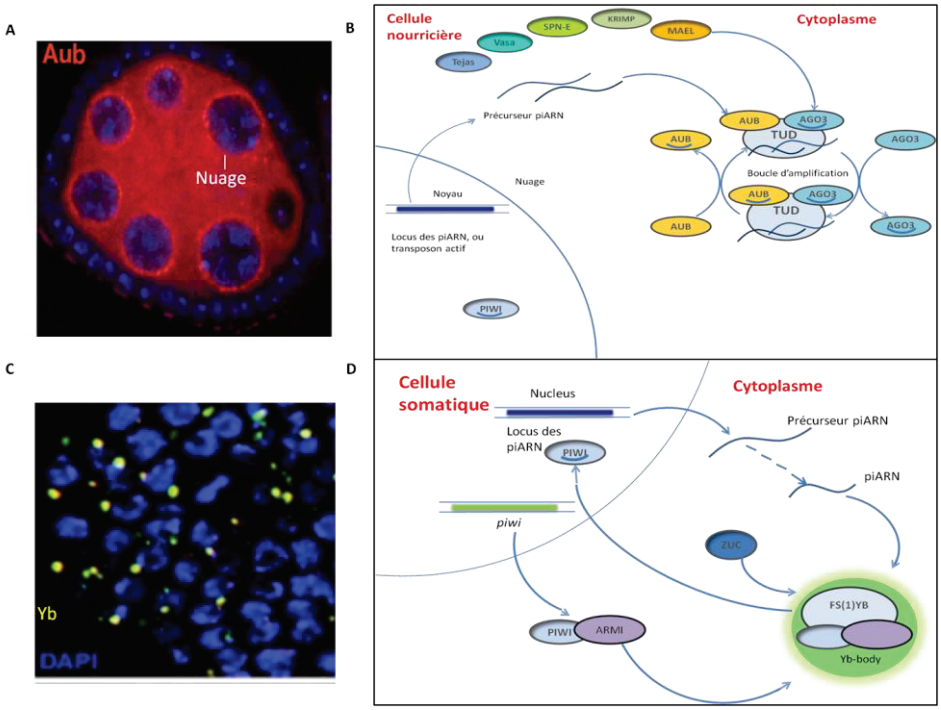


Figure 17: La biogénèse des piARN prend place dans des granules germinales.

(A) Localisation cellulaire de protéine AUB (rouge), qui montre clairement l'accumulation d'AUB dans le nuage. (B) Le nuage est probablement le site où la boucle d'amplification a eu lieu. Dans les cellules nourricières, les facteurs de piARN tels qu'AUB et AGO3 sont localisés dans le nuage. TUDOR (TUD) s'associe avec AUB et AGO3 simultanément dans les cellules germinales. Grâce à cette association Tudor pourrait activer la boucle d'amplification. TUDOR peut aussi avoir un rôle dans le recrutement des précurseurs de piARN vers le complexe de « silencing » (piRISCs). Vasa, spindle-E (SPN-E), Krimper (KRIMP), Tejas et Maelstrom (MAEL) sont aussi localisés dans le nuage, ce qui suggère qu'ils sont impliqués dans cette voie. Toutefois, leur fonction dans la biogenèse des piARN demeure peu claire. (C) Localisation cellulaire de la protéine YB dans les cellules somatiques des ovaires. L'ADN est marqué en bleu et la protéine YB en jaune. (D) Le YB body est le site de biogenèse de piARN primaires dans les cellules folliculaires de l'ovaire où la protéine PIWI est exprimée (mais pas AUB ni AGO3). La protéine Armitage (ARMI) s'associe avec Piwi et se localise dans le corps Yb . La composante principale des Yb est la protéine FS (1) YB. Les piARN intermédiaires, qui sont partiellement transformés à partir des précurseurs primaires, sont chargés sur le complexe comprenant Armi, Piwi et FS(1) YB dans YB et transformés en piARN matures. La nucléase putative ZUC est également requise pour la production de piARN primaires et se localise à proximité des Yb.

1.7 La drosophile comme modèle génétique.

Le système reproducteur femelle chez la drosophile est composé des deux ovaires (King, 1970), chacun étant constitué de seize à vingt-cinq ovarioles disposés parallèlement (Figure 18).

Un ovariole est formé d'une chaîne de six à sept follicules ordonnés suivant leur stade de développement, du plus précoce à l'œuf mature, selon l'axe antéropostérieur (Figure 18). Le germarium, en antérieur, contient les cellules souches et les follicules précoces en formation. Un follicule est formé d'un ciste de 16 cellules germinales, entouré d'une monocouche de cellules folliculaires somatiques. Au cours des stades suivant, le germarium effectue la maturation de l'ovocyte, cellule choisie parmi les seize cellules germinales, les quinze autres devenant des cellules nourricières. La maturation se termine à l'extrémité postérieure de l'ovaire. Des connections existent entre les cellules germinales et les cellules folliculaires, ce qui permet l'échange de signaux de développement ou de substances nutritives.

Le génome de l'ovocyte est ainsi susceptible d'être affecté par des facteurs internes ou externes, notamment par une transposition d'ET. Il a été démontré qu'un certain nombre d'ET s'expriment dans les tissus reproducteurs, comme mentionné ci-dessus. L'élément *I* et l'élément P sont deux exemples d'expression limitée à la lignée germinale. D'autre part, trois rétrotransposons à LTR appartenant au groupe de *Metaviridae* : *ZAM, Idefix* et *Gypsy* s'expriment spécifiquement dans les cellules somatiques qui entourent l'ovocyte chez la drosophile, le tissu somatique qui est en contact avec la lignée germinale. Il a été précédemment suggéré que les particules virales de *Gypsy*, formées dans ces cellules, sont transférées au pôle postérieur de l'ovocyte. L'historique des travaux réalisés chez cet organisme a permis de caractériser différentes instabilités liées à ces éléments.

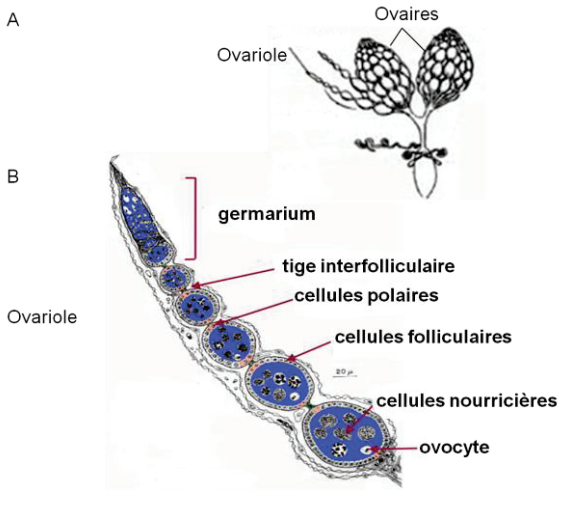


Figure 18: Représentation schématique d'une paire d'ovaires et d'un ovariole. (A) D'après Huston et al, 2008 ; (B) D'après King 1970.

1.7.1 Quelques modèles d'instabilité

1.7.1.1 Le facteur I

Le facteur *I* est un rétrotransposon sans LTR présent dans le génome de *D. melanogaster* (Bucheton et al, 1984). L'invasion du génome de cette espèce par des copies actives est récente, ce qui entraine l'existence de deux types de souches de drosophiles : les souches Inductrices (I), qui contiennent un nombre limité de copies euchromatiques actives et des copies plus anciennes dans l'hétérochromatine, et les souches Réactives (R), qui ont échappé à l'invasion, et ne possèdent que des copies hétérochromatiques. Les facteurs *I* sont réprimés dans les souches inductrices, où la transposition reste un événement rare. En revanche, les facteurs *I* transposent avec une haute fréquence dans la lignée germinale des femelles hybrides, dites SF (« sterile females ») et issues du croisement entre un mâle I et une femelle R. Cette transposition massive entraine une dysgénésie hybride, et les femelles SF sont plus ou moins stériles. La présence d'une séquence signal dans l'ORF2 des copies actives des facteurs *I* permet le transport des transcrits de pleine taille issus de copies fonctionnelles vers l'ovocyte (Van De Bor et al, 2005). L'inhibition de l'expression des facteurs *I* fonctionnels nécessite les protéines effectrices de la voie piARN, comme décrit ci-dessus. En effet, la perte de fonction d'aubergine, squash, zucchini, armitage, spindle-E ou piwi entraine une expression du facteur *I* dans les souches inductrices (Chambeyron et al, 2008).

La répression du facteur *I* nécessite une composante cytoplasmique maternelle. En effet, dans les femelles isogéniques, appelées RSF et issues du croisement entre une femelle I et un mâle R, les facteurs *I* transposent avec une faible fréquence et leurs transcrits sont moins abondants (Chaboissier et al, 1990) (Van De Bor et al, 2005). En revanche, les piARN détectés dans les ovaires des femelles RSF sont plus abondants que ceux trouvés dans les ovaires des femelles SF (Figure 19C) (Chambeyron et al, 2008) (Brennecke et al, 2008). Le dépôt de piARN dans le cytoplasme des ovocytes des femelles I peut expliquer leur abondance dans la descendance RSF. Ces piARN sont déposés dans l'embryon, en absence de toute transcription du zygote durant les 2 premières heures.

Une comparaison des populations des piARN des ovaires entre les souches I et R a révélé de fortes similitudes dans leurs contenus. La différence la plus prononcée a été la quantité de piARN spécifiques de l'élément *I*, qui était 20 fois plus faible dans la souche R que dans la souche I (Brennecke et al, 2008). Cette différence est maintenue dans les embryons correspondants. Ces données indiquent clairement que ce sont les piARN liés aux protéines PIWI qui fournissent l'héritage maternel qui conduit à la suppression de l'élément *I,* et que cet héritage est réalisé grâce à la transmission directe de piARN maternels *via* le matériel génétique incorporé dans les cellules germinales embryonnaires.

De manière surprenante, des transcrits sens du facteur *I* ont été découverts dans le noyau des cellules nourricières des femelles I ce qui suggère au moins un mécanisme de répression post-transcriptionnelle (PTGS). A la différence des femelles SF, ces transcrits ne sont pas transportés dans le noyau de l'ovocyte (Figure 18A). En revanche, ils s'accumulent dans les cellules nourricières où ils colocalisent avec les piARN (Chambeyron et al, 2008). Le

mécanisme exact qui permet aux piARN de maintenir les transcrits du facteur *I* dans les cellules nourricières reste cependant à élucider.

Les souches inductrices I et les souches réactives R possèdent des copies du facteur *I* défectives localisées dans l'hétérochromatine péricentromérique (Crozatier et al, 1988). Il est avancé que la transcription de ces copies défectives est à l'origine de la production de piARN homologues aux transcrits du facteur *I* (Figure 18B).

Dans les souches réactives et inductrices à la fois, le cluster 42AB représente une source majeure de piARN, ciblant une variété des ET (Brennecke et al, 2007). Dans la souche inductrice, la majorité des piARN de l'élément *I* dérive des fragments hétérochromatiques, fragments ancestraux au sein du cluster 42AB. Les piARN de la souche réactive sont également générés par le cluster 42AB. (Brennecke et al, 2008).

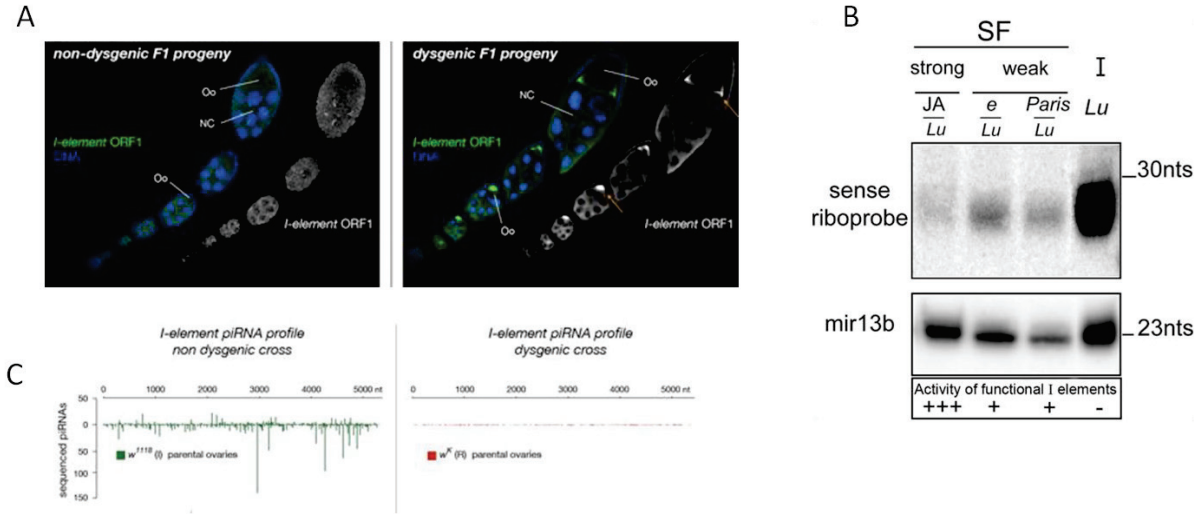


Figure 19: Mise en évidence de la répression post-transcriptionnelle du facteur *I* par la voie des piARN.

La présence de transcrits du facteur I est détectée dans les ovaires de drosophile F1 dysgénique mais pas dans les femelles issues des croisements non-dysgéniques. Les transcrits sens du facteur I sont marqués en vert. L'ADN est marqué en bleu. (B) Détection par northern blot de la présence de piARN homologues au facteur I. Dans une souche inductrice, où le facteur I est réprimé, des piARN homologues à cet élément sont produits. La quantité de piARN détectés dans la descendance SF d'un croisement entre une femelle R et un mâle I est inversement proportionnelle au degré de réactivité présent chez la mère et à l'expression des I fonctionnels dans les ovaires de la descendance (Chambeyron et al, 2008). (C) Comparaison des quantités de piARN entre les croisements dysgéniques et non dysgéniques (Brennecke et al, 2008).

1.7.1.2 Les rétrovirus endogènes Gypsy, Idefix et ZAM

Gypsy a été identifié chez *D. melanogaster* et a été le premier rétrovirus endogène décrit chez les invertébrés (Ilyin et al, 1980). *ZAM* et *Idefix* sont deux autres exemples de rétrovirus endogènes décrits depuis dans cette espèce (Leblanc et al, 1997). D'autres rétrovirus ont été décrits chez les invertébrés, notamment *tom* chez D. ananassae (Tanda et al, 1988) et *tirant* chez *D. simulans* (Akkouche et al, 2012).

Le génome de *D. melanogaster* contient des copies inactives de *Gypsy*, de *ZAM* et d'*Idefix* au niveau de l'hétérochromatine péricentromérique. Ces éléments présentent peu de copies euchromatiques, mais ont conservé leur capacité de transposition dans le génome. Cependant, la mobilisation de ces éléments transposables est un événement rare car il a été montré que leur expression est réprimée dans les tissus où ils s'expriment.

Cependant, certaines souches de *D. melanogaster* présentent un fort taux de transposition pour ces éléments (Desset et al, 1999) (Kim et al, 1990); (Prud'homme et al, 1995). Dans le cas de l'élément *Gypsy*, le suivi de la transposition dans une lignée mutante appelée ovoD, a permis de mettre en évidence une insertion particulièrement fréquente au niveau de ce locus *ovo*. La mutation portée par l'allèle dominant ovo D1 génère des problèmes de stérilité femelle mais l'insertion de *Gypsy* permet la réversion de ce phénotype.

En parallèle, lors de l'étude de la souche de drosophile wIR6**,** une mutagénèse par transposition de l'élément P réalisée dans le laboratoire de Chantal Vaury a entrainé la mobilisation des rétroéléments *ZAM* et *Idefix* dans l'une des descendances. Cette nouvelle souche, baptisée Rev, est dite instable pour l'expression de *ZAM* et d*'Idefix*. En effet, ces deux rétroéléments peuvent être exprimés et mobilisés, alors qu'ils sont fortement réprimés dans la lignée parentale dite stable. Des expériences d'hybridation *in situ* ont montré que ce contrôle s'exerce sur l'expression virale : les transcrits de *ZAM* sont détectés dans les cellules folliculaires aux pôles postérieurs des chambres d'œufs de souches instables mais pas dans les souches stables. De même, les transcrits d'*Idefix* sont détectés dans les stades précoces de l'ovogénèse (germarium) uniquement dans la souche instable (Figure 20A).

La répression de l'expression des copies euchromatiques actives de *Gypsy*, *ZAM* et *Idefix* fait intervenir la voie piARN dans les cellules folliculaires, les tissus somatiques de l'appareil reproducteur, où ils s'expriment (Desset et al, 2008). En effet, des expériences de protection RNAse ont montré la présence de petits ARN de 28-30 nucléotides correspondant au brin antisens de chacun des rétroéléments (Figure 20B). D'autre part, la mutation de la protéine Piwi entraine la levée de la répression de *Gypsy*, *ZAM* et *Idefix* dans les tissus reproducteurs (Desset et al, 2008) (Sarot et al, 2004) (Figure 20C).

Les déterminants génétiques de la répression des rétroéléments *Gypsy*, *ZAM* et *Idefix* ont été caractérisés. La mobilité est en effet contrôlée par le locus *flam/COM,* situé dans l'hétérochromatine péricentromérique du chromosome X en 20A (Desset et al, 1999) (Prud'homme et al, 1995). Plusieurs classes d'allèles sont décrites pour ce locus : l'allèle restrictif *flamR*, est dominant et réprime la transposition de l'élément *Gypsy,* l'allèle permissif *flamP* permet un haut niveau de transposition de cet élément. L'allèle stable COM, réprime l'expression des éléments *ZAM* et *Idefix* alors que l'allèle instable COM, induit une levée de l'inhibition de ces éléments.

Récemment, il a été montré par le laboratoire de Greg Hanon, que le locus *flam/COM* était un cluster de piARN et d'endo-siARN. Ainsi, 79% des piARN homologues à *ZAM* et 30% de ceux qui ciblent *Idefix* sont synthétisés à partir de ce locus. La mutation dans ce locus entraine la derépression de ces trois rétroéléments (Brennecke et al, 2007).

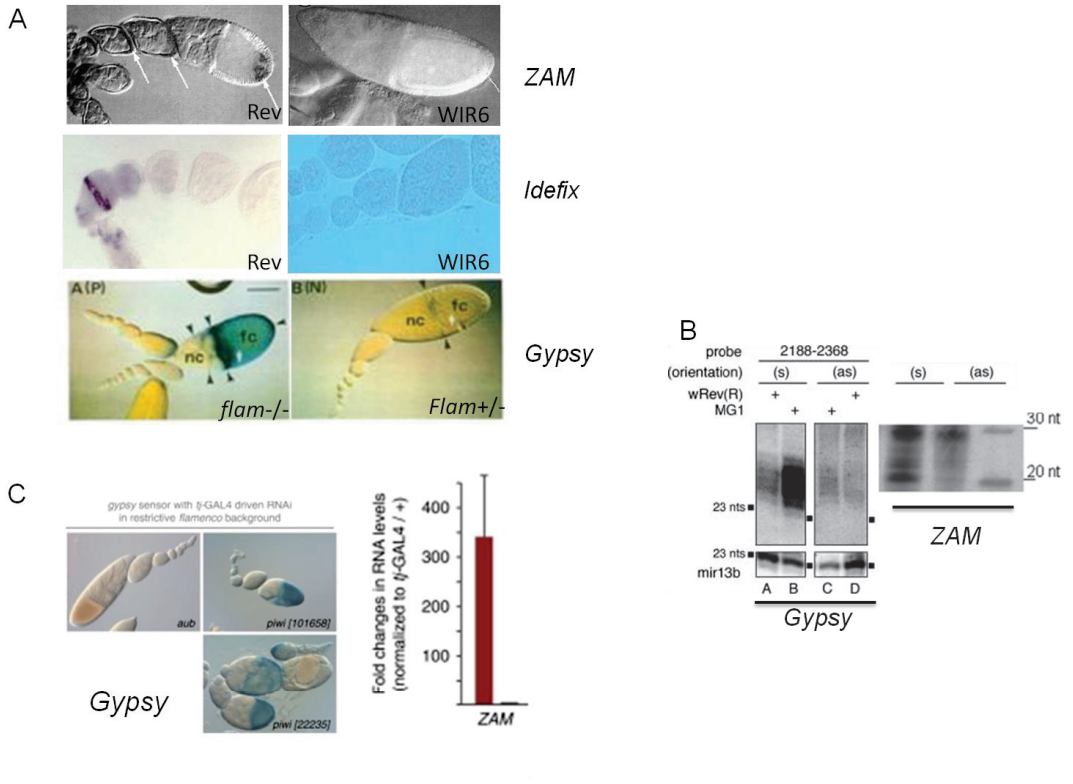


Figure 20: Détection des transcrits et de piARN antisens homologues aux transcrits de *Gypsy*, *ZAM* et *Idefix*

(A) Hybridation *in situ* pour détecter les transcrits de *ZAM*, *Idefix* et *Gypsy* dans les ovaires. Les transcrits de *ZAM* et *Idefix* sont détectés uniquement dans la lignée Rev de manière tissu-spécifique, et les transcrits de *Gypsy* sont détectés dans la lignée *flam-/-*. *ZAM* est exprimé en position postérieure au niveau des cellules folliculaires. *Idefix* est exprimé dans le germanium (Desset et al, 1999). *Gypsy* est exprimé le long de la membrane apicale en contact avec l'ovocyte (Pélisson et al, 1994). **(B)** Caractérisation par protection RNase, de piARN anti-sens de 26 à 30 nucléotides homologues à *Gypsy* (Pélisson et al, 2007) et *ZAM* (Desset et al, 2008). **(C)** La perte de fonction de piwi lève la répression de *Gypsy* dans les ovaires, ces transcrits sont accumulés dans l'épithélium des cellules folliculaires (bleu) pour *Gypsy*. On observe également une augmentation des quantités de transcrits par RT-qPCR pour *ZAM* dans les mutants. Les quantités des transcrits dans les mutants : Piwi en rouge et Aub en noir (Olivieri et al, 2010).

1.7.2 Les populations naturelles de drosophiles pour comprendre la dynamique des ET

1.7.2.1 Variabilité en ET dans les populations naturelles de drosophiles

Les espèces *D. melanogaster* et *D. simulans* sont deux espèces très proches phylogénétiquement, dont la divergence est estimée à environ 2,5 millions d'années. Ces deux espèces diffèrent pour leur quantité d'ET et la taille de leur génome (Biemont & Vieira, 2005). *D. simulans* a 3 fois moins de copies d'ET que *D. melanogaster,* de l'ordre de 5% pour *D. simulans* et de 14 % pour *D. melanogaster* (Dowsett & Young, 1982). Une telle différence

peut être expliquée par les généticiens des populations comme liée à la taille efficace des espèces. En effet, dans les espèces à faible taille efficace, la sélection naturelle est globalement moins efficace, ce qui expliquerait pourquoi les ET sont moins efficacement éliminés. Or jusqu'à présent, les données concernant les tailles efficaces respectives de *D. melanogaster* et *D. simulans* sont contradictoires. Une autre façon d'essayer de comprendre ces différences est de considérer que les deux espèces ont des histoires de « régulation » des ET qui sont différentes. Une telle différence pourrait s'expliquer a) par une capacité du génome de *D.simulans* à réguler ses ET de façon plus stricte que *D. melanogaster*, b) par une perte des copies d'ET qui seraient devenues inactives rapidement chez *D. simulans*, c) par le fait que le génome de *D. simulans* ait été relativement protégé jusqu'à présent de l'invasion par les ET, ce qui peut être lié à son écologie. Cette dernière hypothèse est soutenue par la grande variabilité du nombre de copies d'ET que l'on trouve entre les différentes populations naturelles de *D. simulans* (Vieira et al, 1999b). Ces deux espèces invasives originaires des régions de l'Est africain ont une histoire écologique différente et la colonisation du monde par *D. melanogaster* est plus ancienne que celle de *D. simulans* (Lachaise & Silvain, 2004). La colonisation d'un nouvel environnement peut être associée à une augmentation du nombre de copies, comme le rapportent les études qui montrent un plus faible nombre de copies dans les populations africaines par rapport à celui observé dans les populations du reste du monde (Vieira et al, 1999a). D'autres études associant transposition et colonisation ont été réalisées pour les espèces *D. buzatti* et *D. kepferae* (Labrador et al, 1999). Le nombre de copies d'ET observé dans les populations de drosophiles doit être le résultat des effets entre la transposition et la sélection naturelle, comme le proposent différentes modèles, et une régulation plus fine au niveau de chaque individu qui est apportée par les mécanismes épigénétiques (Lee & Langely, 2010) (Fablet & Vieira, 2011).

L'étude des populations naturelles démontre qu'il existe une grande variabilité du nombre de copies d'ET entre les populations chez *D. simulans* (Biémont & Vieira, 2004). Ces nombres de copies ont été déterminés pour plusieurs familles d'ET par hybridation *in situ* sur les chromosomes polytènes des glandes salivaires de larves de drosophiles. Cette technique, qui utilise une sonde spécifique à chaque ET, permet de compter les copies euchromatiques situées sur les bras des chromosomes. Les insertions d'ET présentes au niveau du chromocentre, ensemble hétérochromatique composé des centromères des chromosomes polytènes, ne peuvent pas être dénombrées par cette technique. Ainsi, bien que beaucoup de populations naturelles de *D. simulans* n'aient pas de copies euchromatiques de certains ET, toutes possèdent néanmoins des copies au niveau des chromocentres. Dans le cas des populations naturelles de *D. melanogaster*, bien qu'une différence de nombre copies euchromatiques ait été identifiée, elle n'est en aucune mesure de la même ampleur que celle retrouvé pour *D. simulans* (Vieira et al, 1999a) (Borie et al, 2002).

Au vu de ce que j'ai décrit ci-dessus sur les systèmes de régulation d'ET, les populations naturelles de *D. simulans* sont un modèle d'excellence pour décortiquer les systèmes de régulations fins, puisque nous avons, pour un grand nombre de lignées, pour un ET donné, des souches dites vides et des souches avec des copies euchromatiques, potentiellement fonctionnelles. C'est aussi le matériel idéal pour chercher à mettre en

évidence des différences d'efficacité des effecteurs de la régulation par des comparaisons populationnelles.

Parmi les ET présentant un polymorphisme d'insertion important, le rétrotransposon *tirant*, décrit ci-après, a été le sujet de ce travail.

1.7.2.2 Modèle d'étude Tirant

Notre laboratoire travaille depuis plusieurs années sur le contrôle de l'expression du rétrotransposon *tirant* chez *D. simulans*. Cet ET est un rétrotransposon à LTR de type *Gypsy* (Marsano et al, 2000). La copie complète de référence de *tirant* est longue de 8526 pb et présente les trois cadres de lecture ouverts *gag*, *pol* et *env* caractéristiques des rétrotransposons à LTR de type *Gypsy* (Figure 21). Ces trois ORFs codent les enzymes indispensables à la rétrotransposition. Marsano et al (2000) ont analysé *tirant* dans quatre souches de *D. melanogaster* : Oregon-R, Charolles, W80 Bakup et FS (PM11-19). Ces quatre souches possèdent au moins une copie complète de *tirant* ainsi que des copies délétées. Une analyse des éléments transposables du génome séquencé de *D. melanogaster* menée par Lerat et al, 2003)a mis en évidence 19 copies de *tirant*, présentant toutes plus de 99% de similarité avec la séquence de référence. Trois de ces copies montrent d'importantes délétions internes, tandis que les autres sont toutes au moins longues de 7800 pb, avec plusieurs copies présentant une séquence potentiellement fonctionnelle. Les études d'expression réalisées par Marsano et al, (2000), montrent aussi une expression de *tirant* par northern blot, dans certains stades du développement.

Dans les populations naturelles de *D. melanogaster* et *D. simulans,* cet élément possède une distribution de nombre de copies étonnante. D'une part, il été mis en évidence par southern blot que toutes les populations de *D. melanogaster* et *D. simulans* possèdent des copies de *tirant* et que toutes les populations testées à l'exception de la population de Zimbabwé de *D. simulans*, contiennent au moins une copie complète, donc potentiellement active (Fablet et al, 2006).

Populations	Mayotte	Makindu	Zimbabwé	Chicharo	Sénégal
Espéce	*D. simulans*				*D. melanogaster*
Nombre des copies euchromatiques	2	5	2	0	9

Tableau 2. Variation des nombres de copies euchromatiques de l'élément *tirant* dans les populations naturelles. D'après Fablet et al. 2009.

D'autre part, l'hybridation *in-situ* sur chromosomes polytènes a montré que toutes les populations de *D. melanogaster* possèdent environ une dizaine de copies de *tirant*, tandis que chez *D. simulans,* la plupart des populations ne possèdent pas de copies de *tirant* sur les bras des chromosomes, à l'exception de quelques populations de l'Est de l'Afrique et de l'Ouest de l'Océan Indien dont Zimbabwé, Makindu (Kenya) et la population Mayotte (Vieira et al,

1999a), que nous avons étudiées en détail. Le nombre de copies des populations est présenté dans le tableau 2.

Deux variant de *tirant* ont été identifiés, chez *D. simulans*, selon des différences dans la séquence nucléotidique de la région régulatrice, composée de la LTR 5' et de l'UTR 5'. Un premier variant présent chez les deux espèces a été baptisé type C ; un second, présent uniquement dans les populations naturelles de *D. simulans* et spécifique à l'hétérochromatine, a été quant à lui baptisé type S. Le séquençage des produits de RT-qPCR a montré que seul le variant de type C est transcrit *in vivo* (Fablet et al, 2006).

Par ailleurs, la détermination précise des sites d'insertion de *tirant* par « genome walking », y compris les copies hétérochromatiques, a été réalisée pour trois lignées de *D. simulans*.

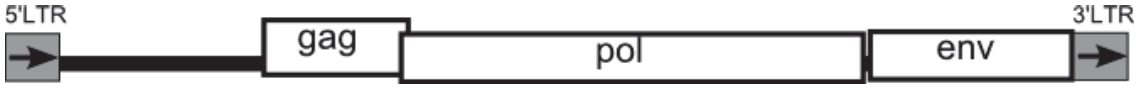


Figure 21: Structure du rétrotransposon à LTR *tirant*.

Objectif de la thèse et stratégies employées.

L'objectif principal de cette thèse est de caractériser un rétrovirus endogène présent chez *D. simulans*, et ainsi d'en comprendre les mécanismes de régulation dans la lignée germinale. La première stratégie a eu pour but de caractériser des populations naturelles de *D. simulans* capables de produire une protéine d'enveloppe du rétrotransposon *tirant.* Plusieurs approches ont été utilisées pour cela. Premièrement : j'ai effectué l'analyse du polymorphisme nucléotidique du gène d'enveloppe de *tirant* dans quatre populations naturelles de *D. simulans*. Deuxièmement, j'ai comparé la quantité de transcrits du gène d'enveloppe par RT-PCR quantitative entre les populations naturelles de *D. simulans.* Puis j'ai caractérisé l'expression de la protéine ENV de *tirant* dans les ovaires par immunohistochimie. L'approche menée sur ce rétrotransposon est présentée sous forme d'article dans la première partie des résultats.

La deuxième stratégie a pour but d'explorer la régulation épigénétique de *tirant* par la chromatine (modification des histones) ainsi que, au niveau post-transcriptionnel, par la voie piARN dans des croisements intraspécifiques de *D. simulans*. En premier lieu j'ai déterminé les types d'histones associés aux copies de *tirant* par la technique d'immunoprécipitation de la chromatine (CHIP), et j'ai mis en évidence un effet de l'insertion de *tirant* sur les modifications d'histones et sur l'expression du géne *tkv* (article 2). Ensuite, en profitant d'un système avec des souches vides et des souches avec des éléments actifs, nous avons établi le rôle des piARN dans la régulation de *tirant*, et nous avons montré que la régulation somatique de *tirant* est dépendante des piARN germinaux (article 3).

Enfin, la troisième stratégie a pour but la caractérisation de la variabilité nucléotidique des gènes impliqués dans la voie piARN, et également la variation de leurs niveaux de transcription, dans les populations naturelles de *D. simulans*, que nous avons pu associer à des niveaux de transcrits de plusieurs ET (Partie 4 des résultats).

2 Résultats & discussion

PARTIE 1

2.1 tirant, un nouveau rétrovirus endogène actif chez *D. simulans*

Les génomes eucaryotes sont envahis par un nombre très important de «parasites génétiques» ayant la capacité de se déplacer et de se multiplier au sein de l'ADN qui les héberge. Ces éléments mobiles du génome appartiennent à la famille extrêmement complexe des éléments transposables découverts par Barbara McClintock dans les années 1950. La diversité de ces éléments transposables est extrêmement riche : leur capacité à se déplacer d'un site à l'autre et à générer des copies mobiles fait appel à des mécanismes variés. Dans les années 1980, des éléments ont été identifiés comme étant capables de coder des protéines rétrovirales. Il s'avère que la très grande majorité de ces éléments ont conservé la capacité à coder une transcriptase inverse (RT, Reverse Trancriptase), d'où le terme de rétroéléments utilisé pour nommer ces éléments. Une partie de ces rétroéléments présente une très forte similarité de structure avec les rétrovirus exogène (XRVs), et sont nommés rétrovirus endogènes (*Endogenous Retroviruses, ERVs*). Les ERV possèdent une structure très proche des rétrovirus exogènes. Ils possèdent généralement trois gènes similaires aux gènes *gag, pol* et *env* de rétrovirus, bordés par deux longues répétitions terminales (LTR). La troisième ORF code des glycoprotéines virales d'enveloppe (Env) qui ont une propriété fusogénique. La protéine d'enveloppe favorise l'assemblage de virions à la membrane plasmique et polarise la production virale à l'extérieur de cellules épithéliales (Garoff et al, 1998) (Owens et al,1991). Ces rétrovirus sont des virus dont la réplication s'effectue par synthèse (transcription inverse) d'un ADN bicaténaire de leur génome ARN monocaténaire, l'ADN viral étant intégré dans le génome de la cellule infectée (provirus). La manière dont les provirus s'intègrent dans la lignée germinale reste peu connue, puisque les mécanismes d'infection de la lignée germinale par des rétrovirus exogènes ne sont pas encore élucidés. Le processus conduisant à l'établissement (« ou colonisation ») des rétrovirus endogènes au sein de génomes cellulaires est nommé endogénisation. Le génome de tous les vertébrés est ainsi envahi par de telles structures et le séquençage systématique d'un grand nombre de génomes, dont ceux de l'homme et de la souris, montre que les rétrovirus endogènes représentent près de 8% du matériel génétique de ces deux espèces (Belshaw et al, 2004).

Les rétrovirus endogènes identifiés chez la drosophile sont pour la plupart actifs, capables de se répliquer de manière autonome. La nomenclature adoptée par l'ICTV « International Committee on Taxonomy of Viruses » a conduit à nommer ces éléments des

errantivirus. Parmi les errantivirus, il a été montré expérimentalement que *Gypsy* est un errantivirus infectieux chez *D. melanogaster* capable de se transmettre d'un individu à un autre (Terzian et al, 2001) (Kim et al, 1990). Il est probable que d'autres errantivirus soient également infectieux si l'on considère leurs niveaux de similarité avec *Gypsy*. Une particularité des errantivirus provient du fait que leur gène d'enveloppe a sans doute été acquis à partir d'un gène présent chez un baculovirus et codant une protéine de fusion (Malik & Eickbush, 2001) (Pearson et al, 2002).

Au début de ma thèse, aucun ERV n'était décrit comme actif chez *D. simulans*, bien qu'il était décrit un polymorphisme du nombre de copies de l'élément *tirant* (Fablet et al, 2006). Mon hypothèse de départ était la suivante : un élément actif doit coder l'ensemble de ses protéines, et donc sa protéine d'enveloppe. Mes travaux ont donc commencé par la caractérisation du gène d'enveloppe du rétrotransposon *tirant*, ce qui a donné lieu à un article publié dans « Journal of Virology », et présenté ci-après.

Nous avons étudié en détail *tirant* dans quatre populations naturelles de *D. simulans* et une population naturelle de *D. melanogaster.* Les populations de *D. simulans* ont été choisies de façon à représenter une gamme du nombre de sites d'insertion de *tirant* déterminée par hybridation *in situ* sur chromosome polytènes : Makindu (Kenya) avec cinq sites, Zimbabwé et Mayotte avec deux sites et Chicharo (Portugal) qui ne présente aucun site d'insertion sur les bras des chromosomes.

Nous avons évalué l'intégrité du gène d'enveloppe dans nos souches par la technique du PTT « protein truncation test », qui consiste en l'amplification de séquences génomiques par PCR, suivie par une transcription puis une traduction *in vitro*. Nos résultats montrent la présence d'une ORF complète de l'enveloppe de *tirant* dans toutes les populations étudiées. Ensuite nous avons isolé et séquencé les gènes d'enveloppe de *tirant* des 4 populations.

Des transcrits de *tirant* sont détectés par RT-PCR quantitative chez les individus issus de Makindu et de Mayotte. Ils sont présents plus abondamment dans la population Makindu. Par contre aucun transcrit n'a pu être détecté dans les deux autres souches de *D. simulans,* Zimbabwe et Chicharo ainsi que la souche Sénégal de *D. melanogaster*. Ces résultats ont été confirmés en parallèle par *hybridation in situ* ARN sur des ovaires en utilisant une ribosonde (1kb) correspondant au gène d'enveloppe de *tirant*.

Dans les ovaires d'individus de la population naturelle Makindu, ces transcrits sont localisés dans le noyau des cellules folliculaires qui entourent l'ovocyte ainsi que dans les cellules nourricières. Par contre chez les individus de la population Mayotte, ils sont localisés dans le cytoplasme des cellules folliculaires, et aucun transcrit n'a pu être détecté dans les cellules nourricières. Ils sont par ailleurs détectés à tous les stades de l'ovogenèse, du germarium à l'œuf mature. Aucun transcrit sens n'est détecté dans les trois autres lignées Zimbabwé, Chicharo et Sénégal.

Nos résultats montrent que *tirant* a une expression qui est dépendante de la population analysée, mais aussi que cet élément peut soit avoir à une expression somatique, soit une expression somatique et germinale.

Nous avons ensuite caractérisé l'expression de la protéine ENV de *tirant* dans les ovaires des populations naturelles de *D. simulans*. Nous avons observé la présence d'ENV dans les ovaires de femelles Mayotte, l'intensité du marquage étant important dans les cellules folliculaires. En revanche, aucun marquage n'est observé dans les ovaires des autres souches.

Par la suite, en utilisant la technique du transposon display nous avons mesuré le taux de transposition de *tirant* dans les populations étudiées et nos résultats montrent que la transposition du retroélément *tirant* s'effectue dans la souche Mayotte alors qu'elle n'a pas lieu dans la souche Makindu.

L'article « *tirant*, a Newly discovered active endogenous retrovirus of *Drosophila simulans* », Akkouche et al 2012 publié dans « Journal of Virology » est présenté ici.

Tirant, a newly discovered active endogenous retrovirus in *Drosophila simulans*

Abdou Akkouche[1], Rita Rebollo[1,3], Nelly Burlet[1], Caroline Esnault[1,4], Sonia Martinez[1], Barbara Viginier[2], Christophe Terzian[2], Cristina Vieira[1,5], Marie Fablet[1]*

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

[2]Université de Lyon, F-69000, Lyon; Université Lyon 1; INRA EPHE, UMR754, Retrovirus et Pathologie Comparée, F-69366, Lyon, France

[3]Present address: Terry Fox Laboratory, British Columbia Cancer Agency, Vancouver, British Columbia, Canada

[4]Present address: National Institute of Child Health and Human Development, National Institute of Health, Bethesda, Maryland 20892, USA

[5]Institut Universitaire de France

*Address for correspondence:

Marie Fablet

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

Tel: (+33) 4 72 43 29 16

Fax: (+33) 4 72 43 13 88

marie.fablet@univ-lyon1.fr

Abstract

Endogenous retroviruses have the ability to become permanently integrated into the genomes of their host, and they are generally transmitted vertically from parent to progeny. With the exception of *Gypsy*, few endogenous retroviruses have been identified in insects. In this study, we describe the *tirant* endogenous retrovirus in a subset of *Drosophila simulans* natural populations. By focusing on the *envelope* gene, we show that the entire retroviral cycle (transcription, translation and retrotransposition) can be completed for *tirant* within one population of this species.

Introduction

Endogenous retroviruses are genomic sequences that are widely dispersed throughout the host genome. These sequences constitute 8% of the human genome and represent the remnants of ancient infections by retroviruses (1). Some of these retroviruses were domesticated and generated cellular neo-genes, such as the *syncytin* gene (9, 14). The structure of a canonical endogenous retrovirus consists of three open reading frames (ORFs), which are bordered by long terminal repeats (LTRs). The third ORF encodes the envelope protein (Env), which possesses fusogenic properties and is responsible for the infectious behavior of exogenous retroviruses. In insects, several endogenous retroviruses have been described, and most of them possess a complete retroviral structure. Representative insect endogenous retroviruses (IERVs) or insect ErantiViruses (24) include the following: *Gypsy*, *ZAM*, *Idefix*, *tirant*, *17.6*, *297* and *nomad* in *Drosophila melanogaster*; *tom* in *Drosophila ananassae*; *Tv1* in *Drosophila virilis*; *TED* in *Trichoplusia ni*; *Osvaldo* in *Drosophila buzzatii* and *Yoyo* in *Ceratitis capitata* (see (19) for a review). With the exception of *nomad* and *Tv1*, each of these displays a complete *env* ORF. Thus far, only the well-known *Gypsy* element of *D. melanogaster* has been shown to possess infectious properties (10), as MoMLV pseudotypes with the *Gypsy* Env protein were shown to infect insect cells (25).

The *tirant* LTR retrotransposon of *D. melanogaster* was previously described to share sequence similarity with the fusion proteins of certain baculoviruses (13, 16, 20) (see Fig. 1 for *tirant* structure). In a phylogenetic study of numerous elements from insects, Terzian et al. (24) proposed that the *tirant* element from *D. melanogaster* (accession number X93507) belongs to the IERVs. Thus, the *tirant* element was placed into this clade of retroviruses, which uses a tRNA-Ser binding site to prime reverse transcription. In this report, we examined this ERV family in *Drosophila simulans* using a collection of strains with variable numbers of genomic insertions (7), and we demonstrated that this family was able to produce Env proteins in the ovaries of the host. We found that *tirant* was capable of performing the first step required for infection, *i.e.,* the production of a functional Env protein, which suggests that it could be classified as an active endogenous retrovirus of *D. simulans*. Furthermore, we demonstrated that the complete retrotransposition cycle of *tirant* occurs within a particular natural strain of *D. simulans*, which indicates that the endogenization of *tirant* in the natural populations of this species is an ongoing process. In addition, this experimental system is of particular interest because our panel of natural strains displayed three distinct states in regards to *tirant* dynamics, and thus will be ideal for deciphering the fine regulatory mechanisms: (i) the absence of transcription in most strains, (ii) transcription

but the absence of translation and (iii) transcription and translation both occurring in a single strain.

Materials and methods

Drosophila strains

We analyzed samples of flies collected from several geographically distinct natural populations of *D. simulans*. These populations were maintained in the laboratory at 25°C as isofemale lines or as small-mass cultures with approximately 50 pairs in each generation.

The *D. simulans* strains used were previously described by Fablet et al. (7). They differed in the number of euchromatic copies of *tirant* that could be detected by *in situ* hybridization on polytene chromosomes. These strains and their respective copies of *tirant* were as follows: Chicharo (Portugal, 0 copies), Makindu (Kenya, 5 copies), Zimbabwe (2 copies), Mayotte (Indian Ocean, 2 copies) and Brazzaville (Congo, 2 copies). Each population is known to display specific, unique euchromatic insertion sites, although bands representing heterochromatic sites may appear be shared between populations in transposon display experiments.

As controls, we used a strain of *Drosophila pseudoobscura*, which is a species devoid of any *tirant* copies (17), that was kindly provided by F. Lemeunier from Gif-sur-Yvette, France and a *D. melanogaster* isolate of a natural population from Senegal, which was grown in the laboratory and contains 9 euchromatic insertions of *tirant* that can be detected by *in situ* hybridization on polytene chromosomes (2). It should also be noted that *D. melanogaster* populations display much less variability of the *tirant* copy number than *D. simulans* populations (2, 7).

Analysis of genomic copies of tirant env

Screening for *env* full-length ORFs in natural populations of *D. simulans*

To search for full-length ORFs for *tirant env* in natural populations of *D. simulans*, we used the protein truncation test (PTT) protocol, which consists of PCR amplification of the *env* gene followed by *in vitro* transcription and translation. A T7-coupled reticulocyte lysate system (Promega) was used for the PTT analysis, according to the protocol recommended by the manufacturer (the TNT quick-coupled transcription/translation system and the Transcend non-radioactive translation detection system from Promega). The primers used for *env* amplification were designed from the *tirant* sequence (accession number AC0054444 positions 50,203-58,729) and consisted of the following sequences: Forward (Fw) 5' GGA TCC TAA TAC GAC TCA CTA TAG GGA GCC ACC ATG GAA CAA TTC 3' and Reverse (Rv) 5' CCT GAA TTT TTT TTT GGG TTA GAG GCG 3'. The upstream primer was positioned at the beginning of ORF3, which allowed transcription to begin at the underlined ATG. We also used *D. pseudoobscura*, a species devoid of *tirant* insertions, as a negative control.

Analysis of *tirant env* sequence polymorphism

We isolated *tirant env* sequences from a subset of the pool of natural populations by PCR using the following primers: Fw 5' TGT ATG GAA CAA TTC CAT CTG AC 3' and Rv: 5' TTG TTT AGA GGC GTG GGG GT 3'. These PCR products were then cloned (Invitrogen TOPO TA Cloning Kit) and sequenced (Sanger method, GenoScreen France). The sequences were also deposited in the GenBank database with the following accession numbers: JN786085 to JN786091 for Makindu, JN786092 to JN786098 for Zimbabwe, JN786099 to JN786104 for Chicharo and JQ219965 to JQ219971 for Mayotte. The DNA sequences were aligned and clustered using the Seaview software (version 4, (8); [http://pbil.univ-lyon1.fr/software/seaview.html]), and DNA polymorphisms were measured using the DnaSP software (version 5, (21)).

Analysis of tirant env transcription

Isolation of *tirant* mRNAs and synthesis of cDNA

Total RNA was extracted from pools of 20-30 ovaries from adult females using the RNeasy kit (Qiagen) and subsequently treated with DNAse (Ambion DNAfree kit). One microgram of the total RNA was then converted into cDNA using the Thermoscript Invitrogen kit primed with a mix of oligo $(dT)_{20}$ and random primers.

Quantification of the env transcripts of *tirant* by qPCR

The cDNAs were diluted 50 times and then quantified using SYBR Green qPCR in a LightCycler (Roche diagnostics) with primers specific to the *tirant env* gene (Fw 5' ACA CGT TCC CTG AAC AGA CG 3' and Rv 5' GAA CGT TAC CAA TCC GAG CA 3') (see Fig. 1 for the position of the amplicon). The transcript quantity was estimated relative to that of the *rp49* gene, for which we used the following primers: Fw 5' CGG ATC GAT ATG CTA AGC TGT 3' and Rv 5' GCG CTT GTT CGA TCC GTA 3'. The *rp49* gene demonstrated equivalent amounts of transcripts for all of the strains analyzed, which enabled the amount of *env* transcripts to be compared. Measurements were obtained from two independent experiments, *i.e.*, RNAs were extracted from two different pools of ovaries. For each experiment, the reactions were performed in duplicate, and the standard curves were calculated using serial dilutions of genomic DNA. The primers were tested for all the strains, and an efficiency value greater than 1.9 was maintained.

Detection of tirant env transcripts by in situ hybridization in ovaries

Ovaries were dissected in PBT (PBS 1x, 0.2% Tween 20), fixed for 20 to 40 minutes, washed in PBT and then incubated in PBS 1x containing 3.7% formaldehyde for 10 min at room temperature. Hybridization (overnight) and washing were performed at 65°C. The *tirant* antisense riboprobe was a 1-kb fragment corresponding to the *env* gene and included a T3 promoter site in the reverse primer (Fw 5' ACG GGG TTT GAT TAA CG 3' and Rv (T3) 5' GGA ATT AAC CCT CAC TAA AGG GAG TGT CCA GGT GTG CTG 3'). The riboprobe was labeled by the *in vitro* transcription of T3 using the DIG RNA labeling mix (Roche). DIG-labeled RNA probes were detected using anti-DIG Fab antibody fragments (1/50 dilution, Roche) and fluorescence amplification (Tyramide Signal Amplification kit,

PerkinElmer). An RNaseH control was included to prevent RNA-DNA hybridization in the experiments. Specifically, after the hybridization and washing steps, the ovaries were treated with RNaseH (0.3 U/µL) for 1 h at 37°C and then evaluated by immunodetection. To stain the DNA, the ovaries were incubated in PBS containing Sytox Green (Invitrogen), and visualization was performed with a Zeiss LSM510 Meta confocal microscope.

Analysis of tirant Env protein production

Detection of *tirant* Env proteins in ovary preparations using immunochemistry

Antibodies

A polyclonal antibody against the *tirant* Env protein was produced *via* the DNA vaccination of mice. The Env cDNA sequence was synthesized and cloned into a pVax plasmid designed using ICAntibodies Technology (developed by In Cell Art for *in vivo* gene expression), and this plasmid was injected into a Swiss mouse.

Immunochemistry

Ovaries were dissected in PBS on ice and fixed with 4% paraformaldehyde in PBS with 0.2% Triton X-100 at room temperature for 20 min. The tissue samples were then washed three times for 2 min each with PBS. After blocking in PBS containing 0.2% Triton X-100 and 0.1% BSA for 2 h at room temperature, the ovaries were incubated overnight at 4°C with the primary antibody at a 1/50 dilution in PBS containing 0.2% Triton X-100 and 0.1% BSA. After three washes in PBT, the ovaries were incubated with the secondary antibody (1/600 dilution of Alexa 488-conjugated anti-mouse IgG H+L, Molecular Probes) for 2 h at room temperature. For the negative controls, the ovaries were incubated with a 1/500 dilution of mouse pre-immune serum for the corresponding primary antibody. The DNA was stained with propidium iodide, and confocal images were acquired using a LSM510 microscope (Carl Zeiss).

Detection of tirant retrotransposition by transposon display

Total genomic DNA was isolated from individual flies using a standard phenol-chloroform extraction procedure after proteinase K digestion. The transposon display was performed using a modified version of the protocol employed by (6). Briefly, genomic DNA from individual adult flies was digested with 10 units of *Mse*I for 3 h at 37°C, and the adaptors consisted of the following sequences: (AdaptMseI+) 5’ AAC AGC TGG ACG ATG AGT CCT GAG A 3’ and (AdaptMseI-) 5’ TAT CTC AGG AGT GTA 3’. The ligation of the adaptors was performed at 37°C for 1 h with 2.5 units of T4 DNA ligase. During the first round of amplification, 0.625 units of Dream Taq Polymerase (Fermentas), 0.2 µM of dNTP, 0.4 µM of the adaptor-specific primer (5' AAC AGC TGG ACG ATG AGT 3') and 0.2 µM of the *tirant*-specific primer (5’ GTC TTC CCG GTT GAG TGT 3’) were used. During the second amplification round, 0.2 µM of the adaptor primer, 0.25 µM of the nested *tirant*-specific primer with 5' HEX fluorescent labeling (5’ TAG AGG CGT GGG GGT TTA 3’) and 0.625 units of Dream Taq Polymerase (Fermentas) were used. The PCR reaction was performed for 35 cycles with amplification steps that lasted for 45 seconds. Negative controls

were included that used the adaptor-specific primer or the element-specific primer alone. One microliter of the PCR products was loaded onto a 3730XL capillary DNA analyzer (Applied Biosystems) with a GS600LIZ standard size marker (Applied Biosystems). The raw data were analyzed using the Gene Mapper software (Applied Biosystems).

We measured the number of novel bands that were obtained in the progeny (7 to 10 males and females) of five independent pairs for the Makindu population and four independent pairs for the Mayotte population. The bands were considered novel only when they were absent from both parental samples and were present in no more than one individual in the progeny.

Results

D. simulans strains display potentially functional env genes

We evaluated the integrity of the *tirant env* genes in our strains using a protein truncation test (PTT), which consists of the amplification of genomic DNA sequences by PCR, followed by *in vitro* transcription and translation. In this technique, subsequent western blotting reveals whether the *env* gene contains a full-length ORF. A premature stop codon or internal deletions at the genomic level appear as bands of smaller mass compared with those of the expected full-length protein. Our assay revealed the presence of complete ORFs for *tirant env* in the tested natural strains of *D. simulans,* which demonstrates that these sequences could potentially be functional *in vitro* (Fig. 2). Similarly, the *tirant env* ORFs of *D. melanogaster* were found to be full-length ORFs. Indeed, western blotting after *in vitro* translation indicated two specific bands that corresponded to approximately 56 kDa and 52 kDa. As the expected product would give a band corresponding to 56.2 kDa, we assumed that the observed signal represented a complete and potentially functional Env protein. The shorter products may correspond to the use of a weaker downstream AUG signal (located 177 bp downstream) or to the presence of a premature stop codon or deletions in certain *tirant* copies that conserve the reading frame.

We isolated and analyzed in detail the sequences of *tirant env* from a subset of these populations (Makindu, Mayotte, Zimbabwe and Chicharo). We did not observe length polymorphism in the PCR products, although the sequencing of a few clones revealed the existence of two sequence variants (Fig. 3), both of which belonged to the transcriptionally active C subfamily of the *tirant* copies (7). One of the sequence variants (referred to here as variant A) was found in all four tested strains (accession numbers: JN786085 to JN786091 for Makindu, JQ219965 to JQ219969 for Mayotte, JN786092 to JN786095 and JN786097 for Zimbabwe and JN786103 and JN786104 for Chicharo), and the sequences were highly similar (nucleotide diversity $\pi = 0.0079$). This variant presumably corresponds to the translated sequence from the PTT assay and is thus a putatively functional sequence. The other variant (referred to here as variant B) was found in all the samples except that of the Makindu strain (accession numbers: JQ219970 and JQ219971 for Mayotte, JN786096 and JN786098 for Zimbabwe and JN786099 to JN786102 for Chicharo), and this variant also displayed high levels of sequence similarity among the strains (nucleotide diversity $\pi = 0.0059$). However, this variant presented a premature stop codon. The calculation of π for the consensus of each

variant resulted in a value of 0.0429. The sequencing of the *env* cDNAs that were obtained from the total mRNA in the Mayotte sample revealed that both variants are transcribed.

Tirant produces env transcripts in the ovaries of certain D. simulans strains

We teted four *D. simulans* strains for the presence of *tirant env* transcripts. Our RT-qPCR experiments showed that only two strains, Makindu and Mayotte, had significant levels of *tirant env* transcripts in the ovaries (Fig. 4A). Even the *D. melanogaster* strain, which harbors significantly more insertions within the euchromatin, did not contain *tirant* transcripts in the ovaries.

To visualize the localization of the *tirant* transcripts in the germline, we performed RNA fluorescent *in situ* hybridization for the ovaries. As expected, no transcripts were detected in the strains that showed no RT-qPCR signal, namely Zimbabwe and Chicharo, or in the *D. melanogaster* strain (Fig. 4B). We observed that *tirant* transcripts accumulated in the mature egg chambers of the ovaries of the Makindu and Mayotte strains (a mature egg chamber consists of the oocyte and nurse cells, which both constitute the germline, surrounded by the somatically derived follicle cells) (Fig. 4B). However, the patterns of transcript localization were different between these two strains; *tirant* transcripts were detected in both the follicle and nurse cells for the Makindu strain, whereas these transcripts accumulated substantially in the follicle cells and were absent from the nurse cells for the Mayotte strain. This observation may also explain why the transcript levels, as determined by RT-qPCR, were higher in the Makindu strain than the Mayotte strain.

Tirant env transcription results in protein production in one D. simulans strain

We evaluated the production of *tirant* Env proteins using immunochemistry, which was performed on whole ovaries using the *tirant* Env antibody. We observed staining for *tirant* Env in the ovaries of the Mayotte strain, whereas no clear signal for Env was detected in the other samples (Fig. 4C). Our immunofluorescence analysis revealed the presence of the Env protein at a late stage (stage 9) of oogenesis in the follicle cells of the Mayotte strain.

Tirant is active in the Mayotte strain

The insertion of a retrotranscribed copy into a new genomic site would represent the final step of the endogenous retroviral cycle of *tirant*. To assess whether this step occurs in the Mayotte population, we performed a transposon display. This technique consists of digesting genomic DNA, ligating adaptors to the obtained fragments and performing a nested PCR with one primer specific to the adaptor and the other primer specific to *tirant*. The observation of the amplified fragment profile allows the detection of insertions of *tirant* in a given genome, and the comparison of the transposon display profiles among the mother, the father and the progeny subsequently allows the identification of new bands, which are only present in one unique descendant and are absent from both parents. The analysis of four pairs of individuals from the Mayotte strain revealed the presence of novel *tirant* bands in the progeny, which were numbered between two and 22, depending on the pair analyzed (Table 1). This finding indicates that *tirant* does retrotranspose in this population.

Furthermore, we repeated the *in situ* hybridization on polytene chromosomes, which we had performed approximately 10 years ago (2). Whereas *tirant* had previously been estimated to display two insertions in the Mayotte strain, the current study was able to identify seven of these insertions in the same strain. It should be noted that transposon display and *in situ* hybridization on polytene chromosomes are not expected to provide identical results. Indeed, heterochromatic copies were revealed by the transposon display technique only, which explains the greater number of bands observed in the transposon display experiments.

Interestingly, in the Makindu population, we identified between zero to five novel bands by the transposon display, which suggests the existence of either a *tirant* provirus amplification process that is independent of Env production or chromosomal rearrangement events.

Discussion

Tirant, a newly discovered active endogenous retrovirus of Drosophila simulans

Similar to other IERVs, *tirant* was previously shown to contain copies with a complete third ORF (4, 15, 27). Short transcripts that potentially correspond to the subgenomic spliced form of *env* were detected by northern blotting in young *D. melanogaster* embryos (15), but no substantial evidence was found to indicate the potential activity of this gene. In the present study, we have demonstrated at least two natural strains of *D. simulans* in which the *tirant env* gene is transcribed, and we found that the corresponding Env protein is produced in one of these strains (Table 2). Therefore, *tirant* appears to be a newly identified active endogenous retrovirus of *D. simulans*, a species in which no other active endogenous retrovirus has been reported thus far. Furthermore, we found that this IERV was not active in *D. melanogaster*, which suggests that the same endogenous retrovirus can display different behaviors in different but closely related species. This is not the first report to describe such differences between *D. melanogaster* and *D. simulans*. For example, one recent study revealed that the respective contents of the transposable elements in these species are quite different; *D. melanogaster* harbors mainly complete copies, whereas *D. simulans* displays many deleted and degenerated copies of elements within the same family (12).

These findings indicate that each strain of *D. simulans* studied possesses at least one potentially complete functional *tirant env* gene in its genome. A detailed sequence analysis revealed that each of these populations had highly similar sequences for *tirant env*. Moreover, another variant of *env* that contains a premature stop codon was also found in most of the studied populations (in all except Makindu).

In this study, we found one *D. simulans* strain in which *tirant env* was both transcribed and translated. In this strain, the staining for Env proteins was observed in the follicle cells of the ovaries during the later stages of development. Other IERVs have also been shown to display activity in the ovaries of certain strains that are referred to as "permissive" or "unstable". In these strains, for example, Tanda et al. (22) found that *tom* RNAs accumulate in the germarium, in nurse cells (germline cells) and in the oocyte nucleus. *Idefix* was shown to be expressed in very specific early follicle cells (somatic cells) within the ovaries of unstable

strains (23). *Gyspy*, which is the only infectious IERV identified to date, was shown to produce Env proteins that accumulated in the follicle cells of permissive females at stage 10 of oogenesis (18). Leblanc et al. (11) studied *ZAM*, which is the IERV most closely related to *tirant*, and found *ZAM* Env proteins in the follicle cells surrounding the posterior portion of the oocytes in *D. melanogaster* ovaries at stages 9 and 10.

For *Gypsy* and *ZAM*, the somatic expression of the *env* genes was thus demonstrated. However, only a transfer from the soma to the oocyte (germline cell) can result in an increase in the copy number of these IERVs in the following generation. This mechanism was demonstrated for *Gypsy* (5) and *ZAM* (3), and it likely involves the yolk protein traffic machinery that operates between the follicle cells and the oocyte (3). In the Mayotte strain, *tirant* transcripts are located in the follicle cells, which is also the case for *Gypsy* and *ZAM*. Our transposon display results, together with the more recent *in situ* hybridization experiments on polytene chromosomes, show that *tirant* does transpose in this strain. Therefore, we may assume that the infection of an oocyte by *tirant* viral particles occurs and that these particles are produced in the surrounding follicle cells as they are for *Gypsy* and *ZAM*.

The tirant element: a model for the study of ERV proliferation in natural populations

As described for other endogenous retroviruses of *Drosophila*, such as *ZAM* and *Gypsy*, most strains are classified as "restrictive" or "stable", which means that the activity and copy number of the endogenous retroviruses are limited. In addition, there are certain "permissive" or "unstable" strains for which this inhibition is no longer detected. Our study assessed the original features of *tirant* in a collection of natural populations, and we detected the following three categories of strains that differed according to their *tirant* dynamics: (i) the “Chicharo” category had zero (or very few, as in the case of Zimbabwe) euchromatic copies of *tirant* and no detectable production of transcripts; (ii) the “Mayotte” category, which displays the qualities of a permissive strain, possessed *tirant* copies that were able to undergo the entire retrotransposition cycle and (iii) the “Makindu” category had an intermediate state, where *tirant* was transcribed but not translated. Most natural populations worldwide behave similarly to those within the Chicharo category and present no detectable *tirant* activity. Our results indicate that these strains have potentially functional *tirant env* genes, which are, however, silenced.

Therefore, at least in the Mayotte strain, *tirant* is able to complete the retroviral cycle and can therefore be considered an active endogenous retrovirus of D. simulans. However, because we do not have the sequences of all the *tirant* entire copies, we cannot exclude the possibility that the expression of the other ORFs may originate from different copies and that the retroviral cycle is achieved through complementation. It remains unclear why *tirant* could not be translated in the Makindu strain. However, we can postulate that a post-transcriptional control of *tirant* occurs in the Makindu strain, which is not the case in the Mayotte strain. Our collection of natural populations, which provides examples of different *tirant* dynamics, is an ideal tool for deciphering such mechanisms.

Furthermore, it is tempting to propose that these different natural strains illustrate different time periods of the history of the *tirant* endogenous retrovirus in the *D. simulans* genome (26). When an active copy of *tirant* is inserted into a favorable genomic site, it can be transcribed and translated, and can retrotranspose to other sites, thereby increasing its copy number. This was likely the situation for the Mayotte strain. In the context of the "arms race" hypothesis, the *D. simulans* genome likely develops strategies to combat this retroviral invasion, such as post-transcriptional locking, which appears to be the case in the Makindu strain. Such locking may also occur in *D. melanogaster* populations, where the number of euchromatic insertions is relatively high compared with *D. simulans*, although no significant transcript levels can be detected in these populations. However, because both species display very different dynamics related to transposable elements (12), we may assume that *tirant* regulation is under the control of many distinct mechanisms.

Acknowledgements

This work was supported by the ANR grant GENEMOBILE from the CNRS, FINOVI, IUF and CIBLE 2008 (Région Rhône Alpes).

We would like to thank Christian Biémont for his useful discussions and Geoffrey Hutinet for his technical help. The English quality was certified by American journal experts.

References

Belshaw, R., V. Pereira, A. Katzourakis, G. Talbot, J. Paces, A. Burt, and M. Tristem. 2004. Long-term reinfection of the human genome by endogenous retroviruses. Proc. Natl. Acad. Sci. U. S. A. **101:**4894-4899.

Biémont, C., C. Nardon, G. Deceliere, D. Lepetit, C. Loevenbruck, and C. Vieira. 2003. Worldwide distribution of transposable element copy number in natural populations of Drosophila simulans. Evolution **57:**159-167.

Brasset, E., A. R. Taddei, F. Arnaud, B. Faye, A. M. Fausto, M. Mazzini, F. Giorgi, and C. Vaury. 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:**25.

Cañizares, J., M. Grau, N. Paricio, and M. D. Moltó. 2000. *Tirant* is a new member of the *Gypsy* family of retrotransposons in Drosophila melanogaster. Genome **43:**9-14.

Chalvet, F., L. Teysset, C. Terzian, N. Prud'homme, P. Santamaria, A. Bucheton, and A. Pélisson. 1999. Proviral amplification of the *Gypsy* endogenous retrovirus of

Drosophila melanogaster involves env-independent invasion of the female germline. EMBO J. **18:**2659-2669.

Esnault, C., M. Boulesteix, J. B. Duchemin, A. A. Koffi, F. Chandre, R. Dabiré, V. Robert, F. Simard, F. Tripet, M. J. Donnelly, D. Fontenille, and C. Biémont. 2008. High genetic differentiation between the M and S molecular forms of Anopheles gambiae in Africa. PloS One 3:e1968.

Fablet, M., J. F. McDonald, C. Biémont, and C. Vieira. 2006. Ongoing loss of the *tirant* transposable element in natural populations of Drosophila simulans. Gene **375:**54-62.

Galtier, N., M. Gouy, and C. Gautier. 1996. SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comp. Appl. Biosci. **12:**543-548.

Heidmann, O., C. Vernochet, A. Dupressoir, and T. Heidmann. 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.

Kim, A., C. Terzian, P. Santamaria, A. Pélisson, N. Prud'homme, and A. Bucheton. 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.

Leblanc, P., S. Desset, F. Giorgi, A. R. Taddei, A. M. Fausto, M. Mazzini, B. Dastugue, and C. Vaury. 2000. Life cycle of an endogenous retrovirus, *ZAM*, in Drosophila melanogaster. J. Virol. 74:10658-10669.

Lerat, E., N. Burlet, C. Biémont, and C. Vieira. 2011. Comparative analysis of transposable elements in the melanogaster subgroup sequenced genomes. Gene **473:**100-109.

Malik, H. S., S. Henikoff, and T. H. Eickbush. 2000. Poised for contagion: evolutionary origins of the infectious abilities of invertebrate retroviruses. Genome Res. **10:**1307-1318.

Mallet, F., O. Bouton, S. Prudhomme, V. Cheynet, G. Oriol, B. Bonnaud, G. Lucotte, L. Duret, and B. Mandrand. 2004. The endogenous retroviral locus ERVWE1 is a bona

fide gene involved in hominoid placental physiology. Proc. Natl. Acad. Sci. U. S. A. **101:**1731-1736.

Marsano, R. M., R. Moschetti, C. Caggese, C. Lanave, P. Barsanti, and R. Caizzi. 2000. The complete *Tirant* transposable element in Drosophila melanogaster shows a structural relationship with retrovirus-like retrotransposons. Gene **247:**87-95.

Misseri, Y., G. Labesse, A. Bucheton, and C. Terzian. 2003. Comparative sequence analysis and predictions for the envelope glycoproteins of insect endogenous retroviruses. Trends Microbiol. **11:**253-256.

Moltó, M. D., N. Paricio, M. A. López-Preciado, V. F. Semeshin, and M. J. Martínez-Sebastián. 1996. *Tirant*: a new retrotransposon-like element in Drosophila melanogaster. J. Mol. Evol. **42:**369-375.

Pélisson, A., S. U. Song, N. Prud'homme, P. A. Smith, A. Bucheton, and V. G. Corces. 1994. *Gypsy* transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the Drosophila flamenco gene. EMBO J. **13:**4401-4411.

Pelisson, A., L. Mejlumian, V. Robert, C. Terzian, and A. Bucheton. 2002. Drosophila germline invasion by the endogenous retrovirus *Gypsy*: involvement of the viral env gene. Insect Biochem. Mol. Biol. **32:**1249-1256.

Rohrmann, G. F., and P. A. Karplus. 2001. Relatedness of baculovirus and *Gypsy* retrotransposon envelope proteins. BMC Evol. Biol. **1:**1.

Rozas, J., and R. Rozas. 1995. DnaSP, DNA sequence polymorphism: an interactive program for estimating population genetics parameters from DNA sequence data. Comp. Appl. Biosci. **11:**621-625.

Tanda, S., J. L. Mullor, and V. G. Corces. 1994. The Drosophila tom retrotransposon encodes an envelope protein. Mol. Cell. Biol. **14:**5392-5401.

Tcheressiz, S., V. Calco, F. Arnaud, L. Arthaud, B. Dastugue, and C. Vaury. 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.

Terzian, C, A. Pélisson, and A. Bucheton. 2001. Evolution and phylogeny of insect endogenous retroviruses. BMC Evol. Biol. **1:**3.

Teysset, L., J. C. Burns, H. Shike, B. L. Sullivan, A. Bucheton, and C. Terzian. 1998. A Moloney murine leukemia virus-based retroviral vector pseudotyped by the insect retroviral *Gypsy* enveloppe can infect Drosophila cells. J. Virol. **72:**853-856.

Vieira, C., M. Fablet, and E. Lerat. 2009. Infra- and transspecific clues to understanding the dynamics of transposable elements. p. 115-123. *In* Transposons and the dynamic genome, Genome Dynamics and Stability. Springer-Verlag, Berlin and Heidelberg.

Viggiano, L., C. Caggese, P. Barsanti, and R. Caizzi. 1997. Cloning and characterization of a copy of *Tirant* transposable element in Drosophila melanogaster. Gene **197:**29-35.

Tables

Table 1 Summary of the transposon display results.

Strain	Pair #	Number of progeny	Number of novel bands	Number of progeny without novel bands
		male / female	male / female	male / female
Makindu	1	10 / 9	0 / 0	10 / 9
	2	10 / 9	0 / 0	10 / 9
	3	9 / 10	1 / 1	9 / 9
	4	9 / 10	3 / 1	8 / 9
	5	10 / 10	4 / 1	8 / 9
Mayotte	1	8 / 10	3 / 4	6 / 8
	2	10 / 10	0 / 2	10 / 8
	3	10 / 7	8 / 10	7 / 4
	4	10 / 10	7 / 15	8 / 3

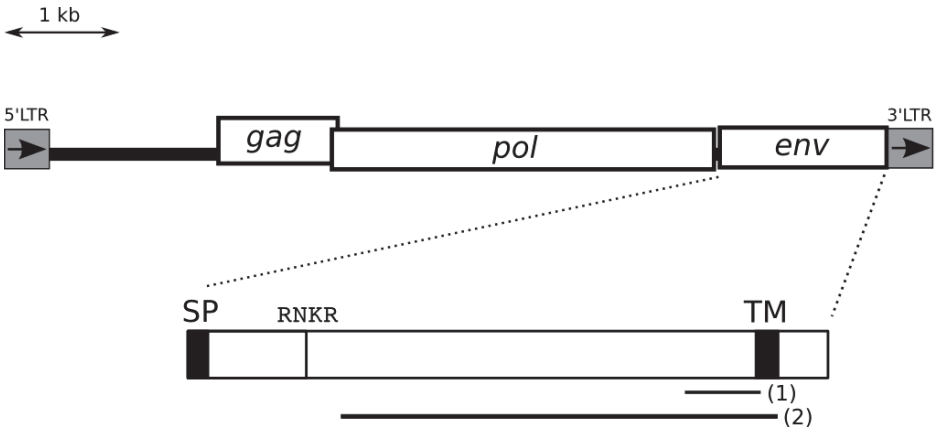
Table 2 Summary of the results obtained with the different strains.

Population	Makindu	Mayotte	Zimbabwe	Chicharo	Senegal (D. melanogaster)
Number of euchromatic copies [a]	5	2	2	0	9
mRNA level [b]	1.53	1	0.02	4.10^{-4}	3.10^{-5}
In situ localization of mRNAs	Nurse cells + Follicle cells +	Follicle cells +++	ND	ND	ND
PTT	Complete ORF	Complete ORF	Complete ORF	Complete ORF	Complete ORF
Env proteins	ND	Follicle cells	ND	ND	ND

[a] as determined by *in situ* hybridization on polytene chromosomes by (2).
[b] relative to the Mayotte level
ND: not detected

Figure legends

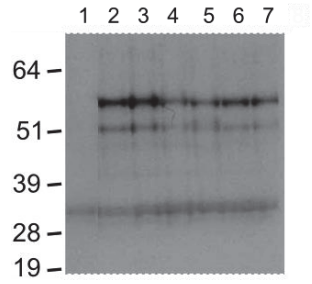
Figure 1



Structure of *tirant env*.

SP: putative signal peptide, as predicted by (4). TM: transmembrane domain, as predicted by (27). RNKR: peptide sequence of the furin cleavage site, as determined by (16). Fragment (1): RT-qPCR amplicon. Fragment (2): probe for *in situ* hybridization of the transcripts.

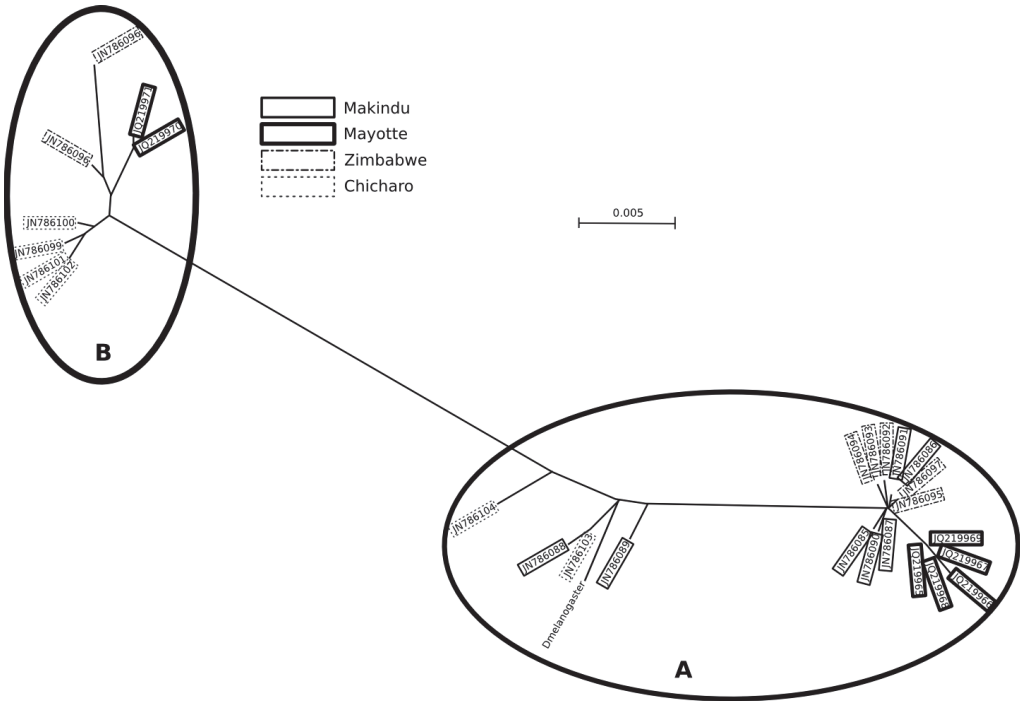
Figure 2



Protein truncation test for *tirant env* in *D. simulans* natural strains.

Western blotting of the products from *in vitro* transcription and translation. (1) *D. pseudoobscura* negative control, (2) *D. melanogaster* strain from Senegal and *D. simulans* strains from (3) Mayotte, (4) Brazzaville, (5) Zimbabwe, (6) Makindu and (7) Chicharo. The molecular mass ladder (in kDa) is included on the left side of the image. All of the tested *D. simulans* natural populations displayed identical profiles that corresponded to a potentially functional *env* ORF. The fainter, lower 33 kDa band is not *tirant*-specific because we also found this band in the *D. pseudoobscura* negative control (a species devoid of *tirant* insertions).

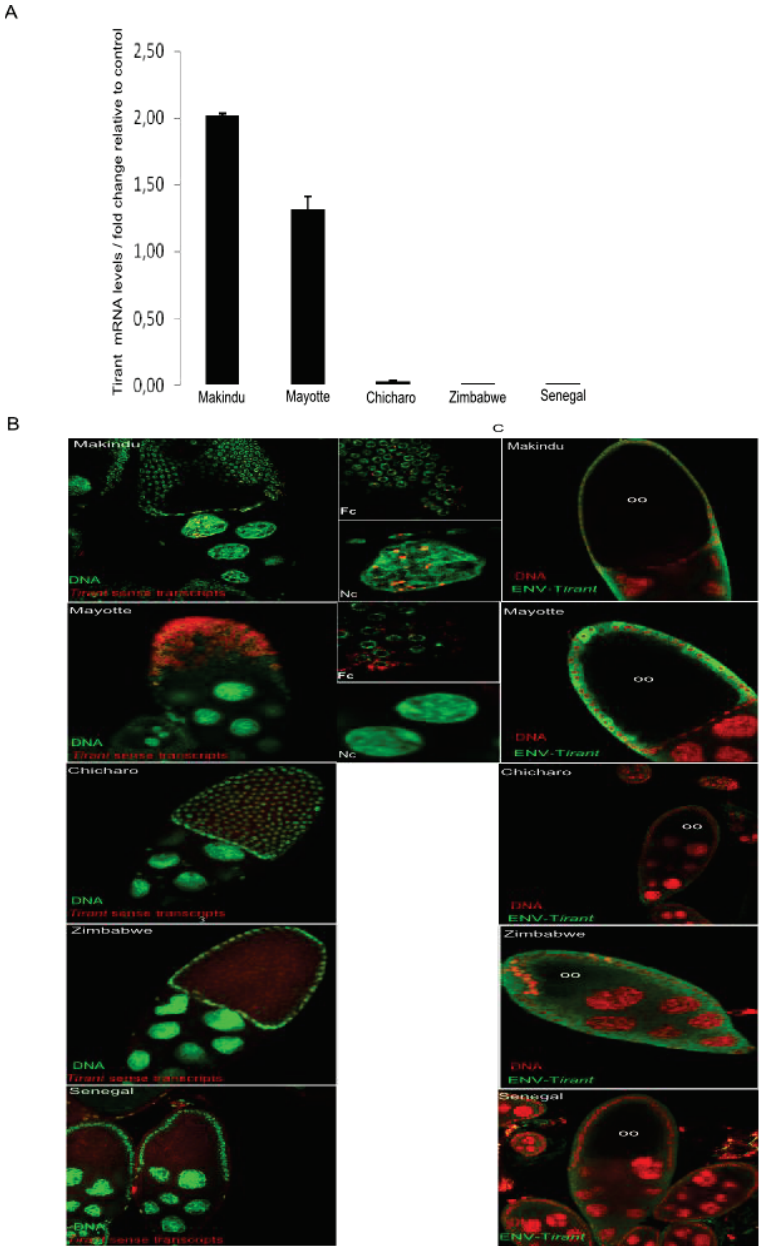
Figure 3



Neighbor-joining reconstruction of the relationships between *tirant env* sequences.

Included is a reference from a *tirant* sequence from the *D. melanogaster* genome, which was extracted from contig AC005444 at positions 50,203 to 58,729. Variant A, which was found in all four of the tested strains, is the canonical sequence for the *tirant env* gene, whereas variant B contains a premature stop codon.

Figure 4



Expression of the *tirant env* gene in the ovaries of natural strains of *D. simulans*.

A. RT-qPCR of the ovaries. The transcript levels were estimated relative to those of the *rp49* gene. Significant levels of *tirant env* transcripts were detected in the Makindu and Mayotte populations. *D. melanogaster* Senegal was used as a control strain. B. *In situ* hybridization of *tirant* transcripts (red). DNA is labeled in green (Sytox Green). Fc: follicle cells; Nc: nurse cells. Left panel: view of ovary chambers (x40 magnification); right panel: details of the left panel. *Tirant* transcripts were found in the nurse cells and follicle cells of the Makindu strain and in the follicle cells of the Mayotte strain. No staining was observed in the other strains. C. Detection of *tirant* Env protein by immunochemistry (green) in the Mayotte strain. DNA is stained in red (propidium iodide). Oo: oocyte. The green halos in the other strains do not represent Env staining but rather correspond to autofluorescence, as they were also observed following staining with pre-immune serum

2.1.1 Discussion

Une vaste étude de comparaison des séquences d'acides nucléiques des éléments rétroides (« retroid éléments ») suggère que les rétrovirus ont évolué à partir des rétrotransposons à LTR par l'acquisition d'un troisième gène (Xiong & Eickbush, 1990). Ce troisième cadre de lecture ouvert (ORF3) des rétrovirus est transcrit en ARN et épissé pour produire la protéine d'enveloppe membranaire, qui joue un rôle dans la capacité des rétrovirus à envahir leur hôte. La structure de *tirant* est très semblable à celles des rétrovirus de vertébrés, avec deux longues répétitions terminales (LTR) encadrant trois cadres de lecture ouverts, gag, pol, et env, chacun codant des protéines homologues aux protéines virales. L'alignement des séquences de *tirant* avec d'autres éléments transposables montre que *tirant* appartient à la famille *Gypsy*, l'élément dont il est le plus proche étant *ZAM* (Marsano et al, 2000).

Chez la drosophile, les rétrovirus endogènes décrits comme actifs sont peu nombreux : *Gypsy* et *ZAM* chez *D. melanogaster* (Pelisson et al, 2002), *Tom* chez *D. ananassae* (Tanda et al, 1994). Ces rétrovirus sont actifs dans les ovaires de certaines lignées permissives ou instables, où ils réalisent leur cycle de réplication, conduisant à la multiplication de leur génome dans les cellules germinales. Les protéines de l'enveloppe de ces rétrovirus ont été détectées dans les cellules folliculaires (Leblanc et al, 2000) (Pélisson et al, 1994). Les particules virales de ces rétrovirus sont produites dans les cellules folliculaires et se concentrent majoritairement le long de la membrane apicale en contact avec l'ovocyte.

En plus d'améliorer notre connaissance sur les rétrovirus endogènes chez la drosophile, l'étude présentée ici nous a permis de caractériser un nouveau rétrovirus endogène actif chez *D. simulans.*

Nous avons détecté la protéine d'enveloppe de ce rétrotransposon chez les femelles d'une souche particulière (Mayotte) où la transposition de *tirant* est détectée (Akkouche et al, 2012). Dans cette population, la protéine ENV est présente principalement à la membrane des cellules folliculaires, comme pour *Gypsy* et *ZAM* (Leblanc et al, 2000) (Pélisson et al, 1994). La présence d'une copie de *tirant* qui code des protéines d'enveloppe potentiellement fonctionnelles dans la population Mayotte soutient l'idée que des particules virales infectieuses peuvent être assemblées dans les ovaires de cette population. Il a été proposé que les particules virales de *ZAM* empruntent le réseau de microtubules de la cellule germinale pour accéder à son noyau, intégrer le génome viral dans les chromosomes de l'hôte et ainsi le transmettre à la descendance (Leblanc et al, 2000).

Dans la population Mayotte, les résultats d'immunomarquage anti-ENV et d'hybridation *in situ* de l'ARN de *tirant* indiquent que la protéine et l'ARN sont localisés au niveau de la membrane apicale. On peut donc envisager que le site d'assemblage des particules de *tirant* est localisé à la membrane apicale. Il faut noter que ce type de localisation polarisée dans une cellule épithéliale est également observé pour de nombreuses

glycoprotéines virales telles que celles codées par le virus influenza et le VIH (Tanentzapf & al, 2000). Toutes ces données suggèrent que les rétrovirus et les « retrovirus–like elements » ne sont pas liés simplement phylogénétiquement, mais aussi fonctionnellement.

Le moment de l'infection de l'ovocyte de drosophile par les particules virales produites dans les cellules folliculaires semble suivre un schéma précis : après l'étape 10 de l'ovogenèse, l'ovocyte devient complètement entouré par une membrane vitelline épaisse imperméable, ce qui constitue une barrière à l'infection virale. Dans le cas de *tirant*, nous observons que l'accumulation de la protéine d'ENV est faible dans des stades précoces de l'ovogénèse et augmente pendant le stade 9. Par conséquent, *tirant* à une fenêtre de temps pour former des particules virales et infecter l'ovocyte.

2.1.1.1 L'élément tirant : un modèle d'étude de prolifération de rétrovirus endogènes dans les populations naturelles de D. simulans.

L'expression des copies actives de *ZAM* et *Gypsy* est inhibée dans la plupart des lignées de drosophile ; ces lignés sont qualifiées de lignées « restrictives » ou stables. En revanche, dans certaines lignées cette répression (inhibition) est perdue, ce qui permet aux éléments de transposer et ainsi de produire de nouvelles insertions génomiques (Desset et al, 1999). Ces lignées sont appelées permissives ou instables.

Notre modèle biologique, consistant en une collection de populations naturelles différant par leur nombre de copies de *tirant*, s'est révélé idéal pour nous permettre de comprendre la dynamique de ce rétrotransposon, en déterminant des statuts différents de *tirant* dans trois des populations étudiées. (1) Dans la souche Chicharo qui ne contient pas de copies euchromatiques, aucun transcrit n'a été détecté, ce qui est attendu. (2) Dans la population de Mayotte, les transcrits de *tirant* ainsi que la protéine d'enveloppe correspondante sont localisés dans les cellules folliculaires qui entourent l'ovocyte, *tirant* est actif dans les ovaires de cette lignée, où il réalise son cycle de réplication, conduisant à la multiplication de son génome dans les cellules germinales. Donc on peut considérer Mayotte comme une souche permissive de *tirant*. (3) Dans la population de Makindu, *tirant* est exprimé mais la transposition reste un événement rare. Ceci implique une régulation post-transcriptionelle, comme il le sera montré dans la troisiéme partie du chapitre 2.

Dans la majorité des populations naturelles, l'élément *tirant* est inactif comme dans la population Chicharo, malgré le fait que nos résultats montrent que ces populations contiennent les gènes d'enveloppe de *tirant* complets.

2.1.1.2 L'expression de tirant dépend de la population.

L'analyse de la distribution de *tirant* dans le génome hôte a montré que le nombre de copies est variable entre populations naturelles de *D. simulans*. Trois populations (Mayotte, Makindu et Zimbabwé) ont des insertions euchromatiques de *tirant* dont le nombre varie entre 2 et 5, et une population (Chicharo) ne contient pas de copies euchromatiques.

Toutes les populations étudiées de *D. simulans* ont des copies hétérochromatiques de *tirant* dans leur génome (Fablet et al, 2006, 2009). Cette distribution de *tirant* est similaire à la distribution de *Gypsy* chez *D. melanogaster*. Toutes les populations contiennent des copies de *Gypsy* localisées dans l'hétérochromatine péricentromérique, et quelques populations contiennent des copies de *Gypsy* sur les bras chromosomiques (Bucheton, 1995). La présence de copies de *Gypsy* sur les bras est corrélée avec la mobilisation de l'élément.

tirant a des copies euchromatiques sur les bras dans les deux populations Mayotte et Makindu. Dans ces deux populations, *tirant* à une localisation cellulaire différente, ce qui suppose une dynamique différente. Les transcrits de *tirant* sont détectés dans le cytoplasme des cellules folliculaires dans les ovaires de la population Mayotte, par contre cette localisation folliculaire est nucléaire dans la population Makindu. De manière surprenante un marquage pour des transcrits sens de *tirant* a aussi été observé dans le noyau des cellules nourricières des femelles Makindu, à la différence des femelles Mayotte, où on ne détecte pas de transcrits dans les cellules nourricières (Figure 4).

La localisation des transcrits dans le noyau des cellules folliculaires a été décrite récemment pour l'élément *mdg1* (Sokolova et al, 2011). La localisation des transcrits dans les noyaux des cellules nourricières a été également observée pour l'élément *I* (Chambeyron et al, 2008). Cette accumulation nucléaire suggère au moins un mécanisme de répression post-transcriptionnelle (PTGS) (Chambeyron et al, 2008). L'accumulation des transcrits de *tirant* dans le noyau des ovaires de la population Makindu ainsi que l'absence de transposition suggère la mise en place d'un système de régulation qui contrôle la mobilisation de *tirant* dans la lignée germinale. Ainsi la régulation de *Gypsy* dans les cellules somatiques est contrôlée par le locus péricentromérique flamenco (Lécher et al, 1997).

Cette idée est renforcée par plusieurs travaux au cours de ma thèse (voir en particulier Akkouche et al, en préparation (article 3)).

Bien que les ET aient réussi à envahir certains génomes, ils sont soumis à des mécanismes de régulation de leur activité. Ces mécanismes contrôlent leur expression, en évitant leur remobilisation massive qui serait délétère (Saito & Siomi, 2010). Parmi ces mécanismes, les régulations transcriptionnelles et post-transcriptionnelles conduisent à la diminution de la transposition. Ces dernières années, de nombreuses études ont montré l'implication des régulations épigénétiques dans le contrôle des ET, et ce, dans de nombreux organismes comme l'homme, la plante, la drosophile ou encore la levure. Les mécanismes épigénétiques qui permettent des modifications transmissibles et réversibles de l'expression

des gènes sans modification de la séquence nucléotidique, comme les modifications post-traductionnelle des histones et les *piwi-interacting RNA* (piARN), sont des voies utilisées par l'hôte pour contrôler l'expression des éléments transposables.

PARTIE 2

2.2 Analyse des modifications d'histones associées à des insertions de *tirant*

Un éventail important de modifications covalentes des histones est aujourd'hui bien caractérisé et il existe dans la plupart des cas une corrélation étroite entre ces modifications et un état transcriptionel activé ou réprimé. Certaines marques d'histones sont associées à de l'hétérochromatine, d'autres à de l'euchromatine. Si un ET se trouve dans une région ouverte de la chromatine, il pourra être transcrit, et l'envahissement du génome par cet élément devient possible. Au contraire, si un ET se trouve dans une région à chromatine compacte, la transcription est plus difficile, et les chances qu'il envahisse le génome sont plus faibles dues à un manque d'activité. Par ailleurs, la présence d'insertion d'ET à un site génomique peut aussi conduire à un recrutement de marques d'histone qui conduira en générale à une chromatine silencieuse (Girton & Johansen, 2008) (Lippman & Martienssen, 2004). L'objectif de l'étude présentée ci-après, est d'identifier les marques d'histones qui sont associées à des insertions spécifiques de *tirant*, et d'évaluer l'impact des se.s insertions sur les gènes voisins.

Puisque les sites d'insertions de *tirant* dans les populations de Makindu, Zimbabwé et Chicharo sont connus, une analyse des modifications d'histones au niveau de chaque site d'insertion de *tirant* a été faite. Afin de caractériser le contexte chromatinien des insertions de *tirant*, j'ai choisi des marques caractéristiques de l'euchromatine, comme H3K4me2, et de l'hétérochromatine, comme H3K9me2 et H3K27me3. Notons qu'il est connu que la chromatine associée à des séquences répétées présente de très faible taux de H3K4me2 et est enrichie en H3K9me2 (Peng & Karpen, 2007). La conformation chromatinienne des copies de *tirant* chez *D.simulans* a été analysée par immunoprécipitation de la chromatine (ChIP).

L'analyse des conformations de la chromatine montre que *tirant* est systématiquement associé à des marques hétérochromatiques. Nous avons montré que des régions euchromatiques sont associées à des marques hétérochromatiques lorsqu'une copie de cet élément y est insérée. Nous avons montré que *tirant* est capable de modifier la structure de la chromatine au niveau de son site d'insertion, mais aussi au voisinage. En effet, l'insertion d'une copie complète de *tirant* dans l'intron du gène *tkv* dans la population de Makindu est responsable de la présence de la marque hétérochromatique H3K27me3 au niveau de son promoteur. Ceci montre une diffusion (« spreading ») de la marque hétérochromatique

H3K27me3 qui entraine une modification de la structure chromatinienne du promoteur du gène *tkv,* avec une forte réduction du taux d'expression de ce gène. En revanche, dans les populations de Chicharo et Zimbabwe, *tirant* est associé uniquement à H3K4me2.

L'ensemble de ces résultats est présenté et discuté dans l'article qui suit. « A Retrotransposon-induced chromatin modification affects nearby gene in natural population of *Drosophila simulans* ».

A RETROTRANSPOSON-INDUCED CHROMATIN MODIFICATION AFFECTS NEARBY GENE IN NATURAL POPULATIONS OF DROSOPHILA SIMULANS

Abdou Akkouche[1], Benjamin Hubert[1], Rita Rebollo[1, 2], Nelly Burlet[1], Marie Fablet[1], Cristina Vieira[1, 3]

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

[2]Present address: Terry Fox Laboratory, British Columbia Cancer Agency, Vancouver, British Columbia, Canada

[3]Institut Universitaire de France

Running head: Retrotransposon induced chromatin modifications

Adresse correspondence to : Cristina Vieira, CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, 43 boulevard du 11 novembre 1918, 69622 Villeurbanne cedex, France ; telephone: 00 33 472448198 ; Fax: 00 33 472431388 ; E-mail: cristina.vieira@univ-lyon1.fr

Abstract

Transposable elements (TEs) are major components of all genomes, in which they are submitted to tight regulatory mechanisms that prevent them from invading the genomes. Among these mechanisms, chromatin modifications are widespread epigenetic effectors of TE regulation in Eukaryotic genomes. The way the epigenetic marks that target TEs affect nearby DNA is however still not clearly established. Using the insertion polymorphism of the *tirant* LTR retrotransposon in natural populations of Drosophila simulans, we show that *tirant* influences the local chromatin structure and triggers heterochromatic marks on its insertion sites. These chromatin modifications spread to the neighboring gene tkv and affect its activity, strongly reducing the expression levels.

Introduction

Analyses of whole genomes in recent decades have revealed that transposable elements (TEs) are major components of the non-coding fraction of all genomes, and that these repeated sequences can influence genome function (Kazazian 2004). Initially it was suggested that TEs were simply “junk DNA” capable of displaying mutagenic effects when they disrupted coding or regulatory sequences (Ohno 1972). Subsequently, some TE sequences were found to have been "domesticated" by genomes, meaning that they had been partially or entirely recruited as genes or promoters, working in tandem with endogenous sequences (Volff 2006). More recently, it has been shown that TEs are targets of epigenetic modifications (Kondo and Issa 2003; Lippman et al. 2004; Yao et al. 2004; Huda and Jordan 2009; Rebollo et al. 2011). The epigenetic modifications or marks that regulate the dynamics of TEs in a genome could also interfere with the transcriptional and post-transcriptional regulation of the host sequences.

Another striking fact about TEs is that their relative amounts vary hugely in different species, even when closely-related species are considered. For instance, the total numbers of TEs in *Drosophila melanogaster* and *D. simulans*, which diverged recently - around 2.5 million years ago (Lachaise and Silvain 2004), have been shown by various methods to differ two to three-fold (Vieira et al. 1999; Vieira and Biemont 2004; Clark et al. 2007). In addition, even within the *D. simulans* species, a high level of copy number variability is observed between natural populations (Vieira et al. 1999; Vieira, Piganeau, and Biemont 2000). Understanding the causes and consequences of such differences in TE content within and between species is therefore of fundamental interest, given the wide range of effects that TEs can have on the genome functioning (Biemont and Vieira 2006). In this study, we focused on *tirant*, an LTR retrotransposon found in *D. simulans.* In this species, *tirant* is present on the chromosome arms of samples from East Africa and Indian Ocean, and its transcription appears to be restricted to a couple of populations (Fablet et al. 2006, Akkouche et al. 2012). This naturally-occurring system involving differing insertion sites in different strains provides an ideal tool for investigating the effects of TE insertions on chromatin structure. In *Drosophila,* like in other eucaryotes, the histone modifications constitute a "histone code" combining enhancing and inhibiting marks (Strahl and Allis 2000) that are responsible for the chromatin structure. The dimethylation of histone 3 at lysine 9 (H3K9me2) is typical of constitutive heterochromatin, H3K27me3 is found in facultative heterochromatin, whereas euchromatin is characterized by H3K4me2 (Schotta et al. 2002; Ebert et al. 2004; Schotta et al. 2004; Ebert et al. 2006).

In this study we take advantage of the characterization of the TE content of natural populations of *Drosophila* to analyze the impact of particular insertions of *tirant*. A previous study had revealed the existence of two *tirant* subfamilies in *D. simulans*, designated C and S. The C subfamily is the only one with detectable levels of mRNA in *D. simulans* populations, and as it displays population insertion site and copy number differences (Fablet et al.2006, Fablet et al. 2009, Akkouche et al. 2012), we focused on this particular subfamily. In the subsequent sections, any reference to *tirant* should be interpreted as referring to the C subfamily. Also, as new TE insertions are transmitted to the progeny only when they occur in

the germline, we focused our analysis on the female gonads of *Drosophila*. We investigated several histone-3 modifications associated with specific copies of *tirant* by ChIP (Chromatin ImmunoPrecipitation), and quantified by qPCR. By comparing empty genomic sites for *tirant* with filled genomic sites with *tirant*, we show that *tirant* triggers heterochromatic marks, and that these marks spread to neighboring regions. Finally, we demonstrate that the heterochromatic marks of a particular *tirant* copy located downstream of the promoter of the *tkv* gene, spread to the neighboring host sequences, and interferes with the level of transcription of the gene.

Experimental procedures

Drosophila strains

We worked on fly samples collected from several geographically-distinct natural populations of D. simulans: Chicharo (Portugal), Zimbabwe, and Makindu (Kenya). These strains were maintained in the laboratory at 18°C as isofemale lines. We had previously fully characterized these three strains with regard to their *tirant* copies, and so know the exact sequence of their insertion sites (Fablet et al. 2009). To ensure consistency, we used the same numbering system for *tirant* insertions as in (Fablet et al. 2009) (Table 1).

Chromatin ImmunoPrecipitation (ChIP)

The extraction of chromatin from ovaries and the immunoprecipitation (IP) procedures were adapted from Sandmann *et al*. (Sandmann, Jakobsen, and Furlong 2006). Each chromatin extraction was performed on 100 pairs of ovaries. The cell lysis buffer was replaced by 5 mM PIPES pH8, 85 mM KCl, 0.5% NP40 supplemented with protease inhibitors. Chromatin was sheared with a Bioruptor sonicator water bath (Diagenode, Belgium) for 6 x 30s on / 30s off cycles at high power in order to obtain random fragments of between 500 and 1000 bp. Sheared chromatin was incubated overnight at 4°C with antibodies recognizing H3K9me2 (Millipore 07441), H3K27me3 (Millipore 07449), H3K4me2 (Millipore 07030), H3 (Abcam ab1791), and rabbit IgG (Sigma-Aldrich I5006), respectively. The antigen-antibody complexes were washed as described before (Sandmann, Jakobsen, and Furlong 2006), but a modified second washing solution was used: TE 2X, 500 mM NaCl, 1% Triton, 0.1% SDS.

In order to quantify each IP, real-time PCR was performed using SYBR Green 1 mix (Roche Diagnostics) on a LightCycler 480 (Roche Diagnostics). Reactions were performed in duplicate, and standard curves were calculated from series of input chromatin. To evaluate the relative enrichment of *tirant* insertion sites after IP, we calculated the difference in cycles between the IP-enriched sample and the input DNA for *tirant* copies, using the rp49 gene as a control. We sequenced qPCR products for each set of primers to confirm that the signals were specific. The primers used for each of the *tirant* insertion sites are shown in Supplementary Table 1. In analyzing the data we used the following normalization method: we subtracted the signal obtained with IgG, which corresponds to noise, and normalized with the signal obtained with H3, which reflects the nucleosomes present in the sample.

Characterization of a copy of tirant inserted into an intron of the tkv gene

From a previous analysis, we know the sequences of all the *tirant* insertion sites in the genomes of the Makindu, Zimbabwe and Chicharo populations of *D. simulans* (Fablet et al. 2009). In particular, in the Makindu population, one full-length copy of *tirant* is inserted into an intron of the *tkv* gene (CG14026, FBgn0003716). We focused on this insertion to investigate the impact of the presence of *tirant* on the neighboring genomic sequences. We compared the transcriptional level of the *tkv* gene in populations with or without the insertion of *tirant* by RT-qPCR, using the following protocol. Twenty-five pairs of ovaries were dissected in PBS 1x. Total RNA from dissected tissues was extracted using RNeasy-Kit (Qiagen). Complementary DNAs were produced by reverse transcription of 1 µg of RNA using the ThermoScript RT-PCR system (Invitrogen) and oligo(dT) primers according to the standard procedure, and the resulting cDNAs were used as templates for real-time PCR. cDNAs were diluted 50 fold, and then quantified by qPCR with SYBR Green (Roche Diagnostics) on a LightCycler 480 (Roche Diagnostics) according to the manufacturer's procedure, using primers specific to *tkv* gene (tkvD_F: 5' CATCAGCCTTTGGCCCA 3', tkvD_R: 5' CACGGACAGGGAGATGATGA 3'). We were only able to amplify in *D. simulans*, the tkv isoformeD. The amount of transcripts was estimated relative to those of the *rp49* housekeeping gene, using the following primers Fw: 5' CGGATCGATATGCTAAGCTGT 3', and Rv: 5' GCGCTTGTTCGATCCGTA 3'. Measurements were done in two independent experiments, *i.e.*, RNAs were extracted from two different ovary pools. For each experiment, reactions were performed in duplicate, and standard curves were calculated on serial dilutions of genomic DNA. In the same populations, we analyzed the chromatic environment by ChIP (protocol as described above) around the *tkv* promoter that was identified as being nearest to the *tirant* insertion. Primers are shown in Supplementary material.

In order to identify the copy from which *tirant* transcripts had originated in the Makindu population, we cloned a PCR product obtained on the 5' region of the previously synthesized cDNAs (Forward primer: 5' CCCTTCATACGTATACACATCTTATCT 3' (position 355), Reverse primer: 5' GCAGTGTGCGTTTTGAGTGT 3' (position 1411)), and sequenced 13 of these clones (Sanger sequencing, Genoscreen). We also cloned and sequenced the most 5' 1.4 kb of genomic insertions #1 and #2, and identified six position differences between them that could be used to assign the transcripts to a specific copy.

Results

We worked with three populations of *D. simulans* with differing euchromatic loads of *tirant*: Makindu (Kenya, 5 copies), Zimbabwe (2 copies) and Chicharo (Portugal, 0 copy). We had previously mapped the insertions of *tirant* in these three populations and we know the exact genomic sequence insertion site for all the copies. We also know that full-length copies of *tirant* are found in the heterochromatin of the Makindu and Chicharo populations, but not in Zimbabwe (Fablet et al. 2006; Fablet et al. 2009). Transcripts of *tirant* are only detected in ovaries from the Makindu populations (Akkouche et al. 2012).

Histone modifications of tirant sequences.

Histones are the proteins around which the DNA molecule is twisted in chromatin. As a result of chemical modifications of their N-terminal tails, histones can modulate the structure of chromatin, which ranges from a very open conformation in euchromatin, which is usually required for gene expression, to a very densely compacted conformation in constitutive heterochromatin, which is thought to be less accessible to the transcriptional machinery.

We investigated chromatin conformation around *tirant* promoter regions by site-specific ChIP of histone modifications. We compared the modifications associated with the insertion of specific *tirant* copies into different genomic compartments, which are representative of the insertions found in the natural populations analyzed. Although numerous modifications have been listed, we focused on three major ones, which have been unambiguously reported in *Drosophila* as being marks of euchromatin (H3K4me2), of constitutive heterochromatin (H3K9me2), and of facultative heterochromatin (H3K27me3), respectively (Ebert et al. 2006).

As previously established (Fablet et al. 2006; Fablet et al. 2009), copies found in euchromatic regions contain H3K4me2 either alone or accompanied by H3K27me3, as for the insertion #2 of *tirant* in the Makindu population (Figure 3A). Since the Makindu population is the only one for which transcripts were detected, we wanted to determine which full-length insertions were expressed. Sequencing of the 5' region of transcripts (1.4 kb) from the Makindu population revealed that the sequences corresponded to both insertions #1 and #2 (see Table 1 for details of the insertions), and thus indicating that transcription is independent of the H3K27 enrichment.

Copies present in heterochromatic regions are more enriched in repressive marks, such as H3K27me3 and H3K9me2. One copy in the Zimbabwe population displays a significant marked H3K4me2 enrichment, but does not correspond to a full-length element.

Impact of tirant insertions on chromatin conformation.

The next issue we questionned at was whether the genomic insertions of *tirant* entered regions that were already H3K27me3 enriched, or whether *tirant* triggered this mark on its sequence after the insertion event had occurred. We focused here on two full-length *tirant* insertions in the Makindu population, which are located within an intron of the *tkv* gene (insertion #2), and on the euchromatic X chromosome (insertion #1), respectively. These Makindu-specific insertions are not found in the genomes of the other populations we looked at. The ChIP experiments revealed H3K27me3 enrichment in the Makindu population (Figure 1A), which was not found in the orthologous empty sites of the other populations studied (Figure 1B). If we parsimoniously assume that the chromatin conformation at this locus was the same in the different populations prior to the insertion event, then this finding clearly demonstrates that it is the insertion of *tirant* that triggers a heterochromatic mark, which changes the local chromatin structure.

One of these full-length Makindu-specific insertions is located with in an intron of the *tkv* gene (insertion #2). In order to find out whether the H3K27me3 enrichment of *tirant* could spread to neighboring sequences, we analyzed the histone modifications associated with the nearest promoter of *tkv* (see supplementary Figure 1).We observed significant H3K27me3 enrichment in this sequence compared to that in populations devoid of this insertion of *tirant* (Figure 1C). We also demonstrated by RT-qPCR that this heterochromatinization of the nearest promoter of the *tkv* gene in the Makindu population is associated with a 10-fold reduction in the transcription level compared to the other populations (Figure 1D).

Discussion

Tirant is a *Gypsy*-like LTR retrotransposon of Drosophila, and is found in a variable number of euchromatic insertions in natural populations of D. simulans. While most populations worldwide do not display any euchromatic site, *tirant* is found at low copy numbers in East African populations, with up to 5 euchromatic insertions in the Makindu population of Kenya. This population is the only one for which significant *tirant* expression has been detected by RT-qPCR and by in situ hybridization in ovaries, indicating that the transcription level observed is of biological relevance (Akkouche et al. 2012). Natural populations of D. simulans provide a powerful tool for the analysis of TE-specific regulation, and also to understand how TE sequences affect the chromatin at their insertion sites. Our system both involves both genomes without any euchromatic copies of *tirant* (most D. simulans populations), and one genome with transcribed euchromatic copies (D. simulans from Makindu). In order to understand the dynamics of *tirant*, and the extent to which the insertion of a TE can impact on its genomic neighborhood, we focused on the chromatin conformation at its insertion sites.

Tirant is transcriptionally regulated by heterochromatin.

The transcription of a genomic sequence depends on its chromatin structure, and it is generally admitted that genes located in open chromatin are transcriptionally active. The level of chromatin compaction is partly "encoded" by various chemical modifications on the N-terminal tail of the histones, around which DNA is twisted in eukaryotic genomes (Jenuwein and Allis 2001).

We show here that *tirant* insertions are associated with enrichment in H3K27me3 in all these populations, which could explain why overall levels of transcription are low. In addition, in the Makindu population, two full-length copies are enriched with the H3K4me2 activating mark, which is consistent with the transcriptional activity observed in this population, as we previously reported for somatic (Fablet et al. 2009) and germline tissues (Akkouche et al. 2012). We found that the sequences of *tirant* transcripts correspond to both of these copies. Although bivalence for chromatin marks has only been reported in mouse (Bernstein et al. 2006; Jiang et al. 2011), and studies of *Drosophila* cell cultures have not shown this effect so far (Gan et al. 2010), our findings once again tend to suggest that bivalent chromatic marks, *i.e.* H3K27me3 and H3K4me2 occurring at the same locus, may also be found in *Drosophila*. Also, a recent study has shown that TEs in euchromatic regions can be enriched with both ‘‘silent’’ and ‘‘active’’ marks in *Drosophila* (Riddle et al. 2011). It has

also been shown that retrotransposons are enriched with H3K9me2 in the heterochromatin of *D. melanogaster* (Ebert et al. 2006; Riddle et al. 2011). However, *tirant* shows remarkably little enrichment with this mark, suggesting that it is able to escape from the inhibition of this chromatin modification. Some evidence has suggested that H3K4me2 and H3K9me2 are mutually exclusive H3 modifications (Rea et al. 2000). Our results for the centromeric insertion #4 in Makindu show non-significantly different levels of enrichment for both of these marks, however it is possible either that either two distinct nucleosomes could overlap in the region, or that we were observing different kinds of cells.

The presence of tirant induces a change in gene expression mediated by heterochromatinization.

By comparing the chromatin state of orthologous loci in different populations with or without an insertion of *tirant*, we demonstrate that *tirant* triggers facultative heterochromatin at its insertion sites. In addition, we show that the Makindu insertion #2 — inside an intron of the *tkv* gene — downregulates the transcription level of *tkv*, a gene which is a decapentaplegic receptor involved in development in *Drosophila* (Brummel et al. 1994; Terracol and Lengyel 1994). We were not able to amplify in *D. simulans* the other described isoformes of the tkv gene in the genomes of *D. melanogaster*. In the Makindu population, *tirant* insertion #2 is the target of H3K27me3 heterochromatic marks, and this does not prevent its transcription, since we were able to detect copy-specific transcripts. However, such repressive targeting clearly does interfere with the transcription of *tkv*, since the promoter of the gene is enriched in H3K27me3, and the consequently transcription rate of the gene is lower than in the natural populations without this *tirant* insertion. The heterochromatinization of the *tirant* insertion therefore spreads to neighboring sequences, over a distance of at least up to 2.6 kb, which is the distance between the *tirant* insertion and the nearest *tkv* promoter. We sequenced this region in the Makindu population, and checked for the presence of other TEs than *tirant* (data not shown), that could contribute to this silencing, but found no other TE insertion. Also, we did not identify any specific phenotype associated with this reduced expression of the *tkv* gene. This may have been due to the fact that the *tkv* gene can have other alternative promoters, which in this case could possibly compensate for the depletion of this particular mRNA.

In this study, we took advantage of the insertion site polymorphism of TEs in natural populations to show how the insertion of a TE can interfere with the endogenous genes by epigenetic means, mediated by post-transcriptional histone modifications. As far as we are aware, this is the first time that the chromatin modifications associated with a single insertion of a TE have been shown to influence the expression of a near-by gene in *Drosophila*. This finding reinforces the necessity to analyze TE insertions at a population level, since this makes it possible to access natural polymorphism and to measure the impact of TEs on neighboring genes.

References

Bernstein BE, Mikkelsen TS, Xie X, et al. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125:315-326.

Biemont C, Vieira C. 2006. Genetics: junk DNA as an evolutionary force. Nature 443:521-4.

Brummel TJ, Twombly V, Marqués G, Wrana JL, Newfeld SJ, Attisano L, Massagué J, O'Connor MB, Gelbart WM. 1994. Characterization and relationship of Dpp receptors encoded by the saxophone and thick veins genes in Drosophila. Cell 78:251-261.

Clark AG, Eisen MB, Smith DR, et al. 2007. Evolution of genes and genomes on the Drosophila phylogeny. Nature 450:203-18.

Ebert A, Lein S, Schotta G, Reuter G. 2006. Histone modification and the control of heterochromatic gene silencing in Drosophila. Chromosome Res 14:377-92.

Ebert A, Schotta G, Lein S, Kubicek S, Krauss V, Jenuwein T, Reuter G. 2004. Su(var) genes regulate the balance between euchromatin and heterochromatin in Drosophila. Genes Dev 18:2973-2983.

Fablet M, Lerat E, Rebollo R, et al. 2009. Genomic environment influences the dynamics of the *tirant* LTR retrotransposon in Drosophila. Faseb J 23:1482-9.

Fablet M, McDonald JF, Biemont C, Vieira C. 2006. Ongoing loss of the *tirant* transposable element in natural populations of Drosophila simulans. Gene 375:54-62.

Gan Q, Schones DE, Ho Eun S, Wei G, Cui K, Zhao K, Chen X. 2010. Monovalent and unpoised status of most genes in undifferentiated cell-enriched Drosophila testis. Genome Biol 11:R42.

Huda A, Jordan IK. 2009. Epigenetic Regulation of Mammalian Genomes by Transposable Elements. Annals of the New York Academy of Sciences 1178:276-284.

Jenuwein T, Allis CD. 2001. Translating the histone code. Science 293:1074-1080.

Jiang H, Shukla A, Wang X, Chen W-yi, Bernstein BE, Roeder RG. 2011. Role for Dpy-30 in ES cell-fate specification by regulation of H3K4 methylation within bivalent domains. Cell 144:513-525.

Kazazian HH. 2004. Mobile elements: drivers of genome evolution. Science 303:1626-1632.

Kondo Y, Issa J-PJ. 2003. Enrichment for histone H3 lysine 9 methylation at Alu repeats in human cells. J. Biol. Chem 278:27658-27662.

Lachaise D, Silvain J-F. 2004. How two Afrotropical endemics made two cosmopolitan human commensals: the Drosophila melanogaster-D. simulans palaeogeographic riddle. Genetica 120:17-39.

Lippman Z, Gendrel AV, Black M, et al. 2004. Role of transposable elements in heterochromatin and epigenetic control. Nature 430:471-6.

Ohno S. 1972. So much "junk" DNA in our genome. Brookhaven Symp. Biol 23:366-370.

Rea S, Eisenhaber F, O'Carroll D, et al. 2000. Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature 406:593-599.

Rebollo R, Karimi MM, Bilenky M, et al. 2011. Retrotransposon-induced heterochromatin spreading in the mouse revealed by insertional polymorphisms. PLoS Genet. 7:e1002301.

Riddle NC, Minoda A, Kharchenko PV, et al. 2011. Plasticity in patterns of histone modifications and chromosomal proteins in Drosophila heterochromatin. Genome Res 21:147-163.

Sandmann T, Jakobsen JS, Furlong EE. 2006. ChIP-on-chip protocol for genome-wide analysis of transcription factor binding in Drosophila melanogaster embryos. Nature protocols 1:2839-55.

Schotta G, Ebert A, Krauss V, Fischer A, Hoffmann J, Rea S, Jenuwein T, Dorn R, Reuter G. 2002. Central role of Drosophila SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. EMBO J 21:1121-1131.

Schotta G, Lachner M, Sarma K, Ebert A, Sengupta R, Reuter G, Reinberg D, Jenuwein T. 2004. A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. Genes Dev 18:1251-1262.

Strahl BD, Allis CD. 2000. The language of covalent histone modifications. Nature 403:41-45.

Terracol R, Lengyel JA. 1994. The thick veins gene of Drosophila is required for dorsoventral polarity of the embryo. Genetics 138:165-178.

Vieira C, Biemont C. 2004. Transposable element dynamics in two sibling species: Drosophila melanogaster and Drosophila simulans. Genetica 120:115-23.

Vieira C, Lepetit D, Dumont S, Biemont C. 1999. Wake up of transposable elements following Drosophila simulans worldwide colonization. Mol Biol Evol 16:1251-5.

Vieira C, Piganeau G, Biemont C. 2000. High copy numbers of multiple transposable element families in an Australian population of Drosophila simulans. Genet Res 76:117-9.

Volff J-N. 2006. Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. Bioessays 28:913-922.

Yao S, Sukonnik T, Kean T, Bharadwaj RR, Pasceri P, Ellis J. 2004. Retrovirus silencing, variegation, extinction, and memory are controlled by a dynamic interplay of multiple epigenetic modifications. Mol. Ther 10:27-36.

Tables

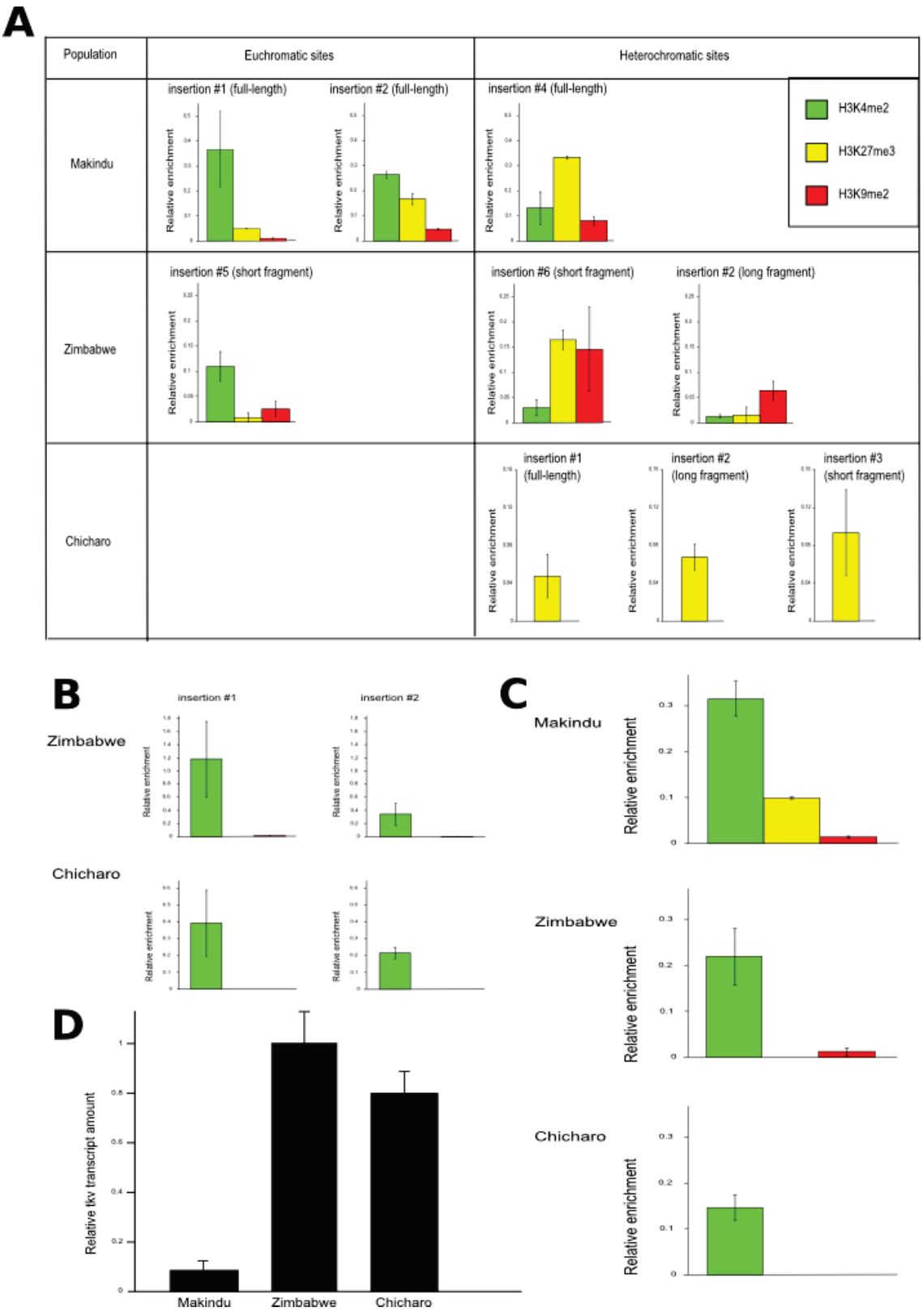
Table 1. *Tirant* insertions in the natural populations studied (from Fablet et al. 2009)

Population	Insertion	Location	Structure
Makindu	1	Chr. X	Full-length
	2	Chr. 2L, *tkv* gene intron	Full-length
	4	Centromeric DNA, Maupiti islands	Full-length
Zimbabwe	4	Chr. 2L, heterochromatin	Short fragment
	5	Chr. 3R	Short fragment
	6	Centromeric DNA, Maupiti islands	Short fragment
Chicharo	1	Centromeric DNA, Maupiti islands	Full-length
	2	Heterochromatin	Long fragment
	3	*Diver2* LTR retrotransposon	Short fragment

Figure legends

Figure 1: Akkouche et al.

A. Chromatin state at *tirant* insertion sites analyzed by ChIP in natural populations of *D. simulans*. The state of each insertion is shown in brackets. For the Chicharo population, only H3K27me3 was detected. Insertions are described in the text and in 24. B. Chromatin states at loci without *tirant* insertions, analyzed by ChIP, in the Zimbabwe and Chicharo populations. Insertion #1 is located on the euchromatic X chromosome, and insertion #2 is found inside an intron on the tkv gene (chromosome 2), as previously described (24). Both insertions are full-length elements in the Makindu population. C. Analysis of the promoter of the tkv gene. The promoter of the tkv gene is located 2.6 kb upstream of the *tirant* insertion. The presence of *tirant* is associated with enrichment of the H3K27me3 mark. See Supplementary Figure 1 for the description of tkv structure. D. tkv expression by RTqPCR. Results correspond to the transcription level of the tkvD isoform (see supplementary Figure 1), normalized with the rp49 reference gene. *Tirant* is inserted 2.6-kb downstream of the corresponding promoter only in the Makindu population. All data were normalized as described in the Material and Method section. The numbering of the insertions and their classification as euchromatic or heterochromatic correspond to what was reported in a previous study by Fablet et al. (24), and is presented in Table 1.


A
Population
Euchromatic sites
Heterochromatic sites
Makindu
Zimbabwe
Chicharo
insertion #1 (full-length)
insertion #2 (full-length)
insertion #4 (full-length)
insertion #5 (short fragment)
insertion #6 (short fragment)
insertion #2 (long fragment)
insertion #1 (full-length)
insertion #2 (long fragment)
insertion #3 (short fragment)
Relative enrichment
H3K4me2
H3K27me3
H3K9me2
B
Zimbabwe
Chicharo
insertion #1
insertion #2
C
Makindu
Zimbabwe
Chicharo
Relative enrichment
D
Relative tkv transcript amount
Makindu
Zimbabwe
Chicharo

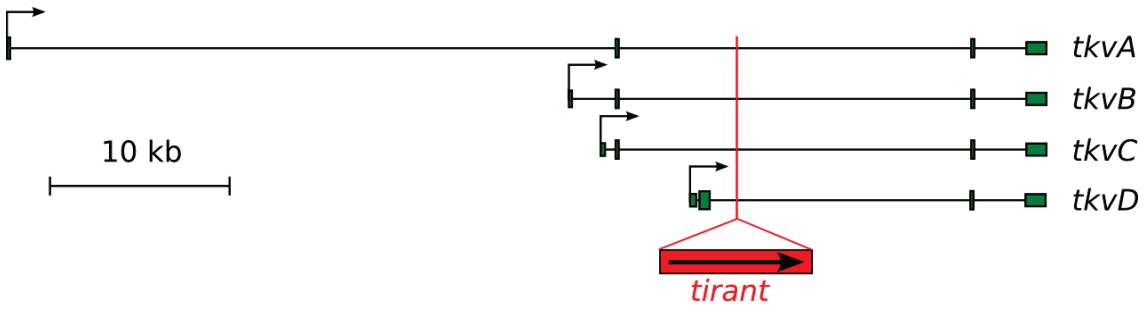
Supplementary Table 1

Primers used for ChIP-qPCR

Population	Locus	Forward Primer	Reverse Primer
Makindu	C4 insertion	caccgagtgaaactgctgaa	atgttgcagtgcgacttttg
	C2 insertion	ctgagcacttgatttgggcttag acaggc	gtgttccagttgccgtcttc
	C1 insertion	tgtgcggatttctactgttttc	agcataatgaacatgccgact
Zimbabwe, Chicharo	C1 empty site	gcatccaatgcgaacaagaa	aggtccgcagttccacagtt
Zimbabwe, Chicharo	C2 empty site	gccttttgtgtgcccaact	gcgttgctctaagggga
Makindu, Zimbabwe, Chicharo	Rp49	ctggtttccggcaaggtatgt	cagttcaactcaaaaccgccaaa
Makindu, Zimbabwe, Chicharo	TkvRD	tttaaacgccagcaagtgtg	acgactcgcgaagaacgtat

Supplementary Figure 1.

Structure of the *tkv* gene and the predicted transcripts. The position of *tirant* insertion in the Makindu population is indicated. The distance from the promoter was verified by PCR, cloning, and sequencing.



2.2.1 Discussion

Le nombre variable de copies de *tirant* dans les génomes des populations naturelles de *D. simulans*, ainsi que les différences de profils d'expression mis en évidence, suggèrent des différences possibles dans la régulation de l'élément. Les modifications de la chromatine, par le biais de modifications post-traductionnelles des histones, sont l'un des mécanismes possibles pour cela. A nouveau, notre système biologique, constitué de populations naturelles présentant des statuts différents quant à la dynamique de *tirant*, s'avère idéal pour une telle étude.

Les différents états de chromatine, plus ou moins compacte, sont fonction, entre autres, des modifications post-traductionnelles des queues amino-terminales des histones qui dépassent du corps globulaire. Nous avons analysé la structure de la chromatine au niveau des sites d'insertions de *tirant* dans trois populations différentes, en étudiant en particulier une marque caractéristique de l'euchromatine, H3K4me2, et deux marques hétérochromatiques, H3K27me3 et H3K9me2

2.2.1.1 Modifications d'histones associées au rétrovirus endogène tirant.

L'analyse de la conformation de la chromatine par immunoprécipitation (ChIP) à l'aide d'un anticorps dirigé contre les protéines H3K27me3 (marque de l'hétérochromatine constitutive), H3K9me2 (marque de l'hétérochromatine facultative) et H3K4me2 (marque de l'euchromatine) montre que les copies de *tirant* sont enrichies en H3K27me3 dans les trois populations étudiées. Ceci explique alors le fait que *tirant* présente un taux de transcrits relativement faible par rapport aux gènes de référence. Par contre dans la population Makindu, 2 copies complètes sont fortement associées à la marque euchromatique H3K4me2, qui favorise la transcription. Ceci explique la présence des transcrits qui ont été décrits auparavant dans les embryons (Fablet et al, 2009) et dans la lignée germinale femelle (ovaires) (Akkouche et al, 2012).

La présence de marques euchromatiques et hétérochromatiques associées à la même séquence, appelée bivalence, a été mise en évidence dans les cellules souches embryonnaires chez l'homme avec la présence des deux marques H3K4me2 et H3K27me3 (Brenstein et al., 2006), ainsi que chez la souris (Breinstein et al, 2006) (Jiang et al, 2011). Nos résultats suggèrent une bivalence des 2 marques H3K27me3 and H3K4me2 au niveau des insertions de copies complètes de *tirant* dans les ovaires de la drosophile. La bivalence des marques euchromatiques et hétérochromatiques sur le même site d'insertion de *tirant* avait déjà été mise en évidence dans les tissus somatiques de *D. simulans* (Fablet et al, 2009).

Nos résultats sont confortés par des données récentes qui montrent que les séquences d'ET de régions euchromatiques peuvent être enrichies à la fois en marques activatrices et répressives (Riddle et al, 2011). Ces chercheurs ont aussi montré que les rétrotransposons sont associés à H3K9me2 dans l'hétérochromatine de *D. melanogaster*. C'est aussi ce que nous observons pour le rétrotransposon *tirant* quand il est localisé dans les régions

hétérochromatiques. Par contre, nous détectons un faible enrichissement pour cette marque sur les copies complètes localisées dans l'euchromatine, ce qui peut etre à l'origine des transcrits mis en évidence dans la population de Makindu (Akkouche et al, 2012).

2.2.1.2 Influence de tirant sur la chromatine des gènes voisins.

La présence d'une copie complète de *tirant* enrichie en H3K27me3 dans un intron du gène *tkv*, nous a permis de rechercher si une diffusion de ces marques était possible, et d'étudier leur impact sur l'expression des gènes voisins. Nos résultats montrent que *tirant* est capable de modifier la structure de la chromatine au niveau de son site d'insertion, mais aussi en amont. En effet, l'insertion d'une copie complète de *tirant* dans un intron du gène *tkv* dans la population Makindu est responsable de la présence de la marque hétérochromatique H3K27me3 au niveau d'un promoteur de *tkv*. La distance entre le début de la LTR 5' de cette copie de *tirant* et le promoteur du gène *tkv,* établie à partir du génome séquencé de *D. melanogaster* et vérifiée expérimentalement pour les populations de *D. simulans*, est estimée à 2.6 kb.

La conséquence de l'hétérochromatinisation du promoteur du gène *tkv* par l'insertion de *tirant* est une diminution du taux d'expression de ce gène dans la population Makindu, mais pas dans les deux autres populations, où cette région est dépourvue de *tirant*. Aucun phénotype associé à des mutations du gène *tkv* n'a été mis évidence dans la population Makindu. Pour aller un peu plus loin dans l'analyse de l'interaction entre l'insertion de l'élément et le gène tkv, nous cherchons à quantifier la production de la protéine tkv dans les trois populations par western blot. Ces résultats sont en cours d'obtention.

Ce cas particulier n'est pas unique dans la littérature, mais reste rare dans le cas des systèmes naturels. Par exemple, l'influence qu'un ET peut avoir sur la structure de la chromatine au niveau de son environnement génomique a été décrit avec l'élément P, dont l'insertion peut conduire au phénomène de « Position Effect Variegation » (PEV), caractérisé par une hétérochromatinisation de gènes euchromatiques (Girton & Johansen, 2008). De plus, l'inhibition épigénétique provoquée par l'insertion d'un ET dans un gène ou à proximité a été mise évidence chez *A. thaliana* (Lippman & Martienssen, 2004).

PARTIE 3

2.3 Les piARNs germinaux déposés maternellement régulent le rétrovirus endogène *tirant* dans les cellules somatiques

Récemment, une machinerie spécifique (impliquant des facteurs de type Argonaute 3, Aubergine et PIWI), requise pour la prise en charge d'une famille de petits ARNi présents uniquement dans la lignée germinale, appelé PIWI–interacting RNA (piARN), a été mise en évidence (Brennecke et al, 2007). Chez la drosophile et la souris, des mutations affectant les facteurs impliqués dans cette voie entrainent une expression massive d'un grand nombre des ET dans la lignée germinale (Vagin et al, 2006) (Lim & Kai, 2007). La mutation du gène piwi entraine la derépression de *Gypsy*, *ZAM* et *Idefix* ainsi que d'autres rétrotransposons dans les cellules folliculaires (Olivieri et al, 2010) (Desset et al, 2008) (Pélisson et al, 2007). La mutation du gène aub entraine une expression de l'élément *I* dans l'ovocyte de la lignée germinale (Chamberyon et al, 2008). Outre les protéines Piwi, d'autres facteurs ont été décrits comme partenaires dans cette voie de régulation comme l'hélicase spn-E qui est exprimée dans le nuage (Malone et al, 2009) (Shpiz et al, 2008). La mutation de spn-E entraine la dérégulation de plusieurs éléments comme les deux éléments télomériques *TART* et *Het-A* (Shpiz et al, 2011), ainsi que d'autres éléments comme les rétrotransposons *MDG1* et *HMS-Beagle* (Sokolova et al, 2011).

Chez la drosophile, ce système de régulation permet l'inhibition des ET dans l'ovaire par deux voies de régulation différentes. Les ET qui ont une expression somatique dans les cellules folliculaires, sont régulés par des piARN primaires (c'est à dire directement transcrits à partir du génome) (Olivieri et al, 2010), où la protéine Piwi joue un rôle crucial vue leur localisation somatique (Brennecke et al, 2007) et les ET à expression germinale (le facteur *I* par exemple) qui sont régulés par une voie appelée « ping-pong », qui nécessite parfois une composante cytoplasmique maternelle pour être déclenchée (Chambeyron et al, 2008) (Brennecke et al, 2008). Différentes protéines interviennent dans cette voie comme SPN-E, AUB, KRIMP, MAEL (Malone et al, 2009) (Olivieri et al, 2010). Le contrôle des ET par cette voie est transmissible de génération en génération.

Les travaux présentés dans la littérature vont dans le sens d'une indépendance des deux voies somatiques et germinales. Ceci pose alors la question du dialogue et de l'interaction

entre ces deux voies de régulation. Comment les éléments qui ont une expression mixte « somatique et germinale » sont-ils régulés ?

A partir de nos connaissances portant sur le contrôle des ET (l'élément *I* dans le germen et *Gypsy* dans le soma), nous avons étudié la biogénèse des piARN, le mode d'action des piARN dans la répression de *tirant*, et les mécanismes épigénétiques de transmission héréditaire de ce mode de répression.

Nous avons voulu tester l'hypothèse que des éléments qui ont une expression germinale et somatique utilisent les deux machineries de biogénèse des piARN dans le soma et dans le germen pour être régulés, en essayant de mettre en évidence un lien entre ces deux voix de régulation.

Dans le travail qui suit, j'exposerai une analyse détaillée des piARN pour comprendre le mécanisme de régulation de *tirant*, Pour cela, nous avons réalisé cette étude sur des hybrides intraspécifiques de *D. simulans*, obtenus par croisements entre une souche vide (Chicharo) et une souche contenant des copies euchromatiques de *tirant* (Makindu). Nous savons que les souches Makindu et Chicharo possèdent des copies de *Tirant* défectives localisées dans l'hétérochromatine (Fablet et al, 2009), et que *tirant* est réprimé dans la souche Makindu où la transposition reste un événement rare (Akkouche et al, 2012). Les transcrits de *tirant* ont été mis en évidence dans le noyau des cellules folliculaires ainsi que dans les cellules nourricières des femelles Makindu. La présence de piARN similaires aux copies actives a été mise en évidence et leur quantité est inversement proportionnelle à l'accumulation des transcrits de *tirant*, ce qui indique que la répression de *tirant* dans la souche Makindu nécessite la voie piARN.

Ces travaux indiquent que la répression du rétrovirus endogène *tirant* nécessite une composante cytoplasmique maternelle. En effet dans les femelles issues du croisement entre une femelle Makindu et un mâle Chicharo, aucun transcrit dans les cellules somatique n'est détecté. En revanche, les piARN détectés dans les embryons (0-2h) issus de ces croisements sont en plus grande quantité que ceux trouvés dans les embryons des femelles issus du croisement entre une femelle Chicharo et un mâle Makindu, où on détecte une dérégulation de *tirant* dans les cellules somatiques. Le dépôt de piARN dans le cytoplasme des ovocytes des femelles Makindu peut expliquer leur abondance dans les embryons issus du croisement entre femelle Makindu et mâle Chicharo. L'absence de piARN déposé maternellement dans les croisements issus d'un mâle Makindu et femelle Chicharo entraîne une augmentation de l'expression de *tirant* dans les cellules folliculaires qui entourent l'ovocyte. Nous montrons donc au sein de populations naturelles que la régulation somatique de *tirant* est dépendante de la présence de petits ARN déposés maternellement.

Le modèle présenté propose que les piARN secondaires germinaux produits par le mécanisme ping-pong sont transmis vers les cellules folliculaires et jouent un rôle dans la répression de *tirant*. Il s'agit ici de la première caractérisation de l'existence du passage des piARN germinaux des cellules nourricières vers les cellules somatiques. Ces travaux sont

exposés dans le manuscrit qui suit, «Germline maternally deposited piRNA regulate tirant retrotransposon in somatic cells».

To be submitted

Germline maternally deposited piRNA regulate a retrotransposon in somatic cells

Abdou Akkouche[1], Thomas Grentzinger[2], Marie Fablet[1], Claudia Armenise[2], Nelly Burlet[1], Séverine Chambeyron[2], Cristina Vieira[1, 3*]

[1]Université de Lyon, Université Lyon 1, CNRS UMR5558, Laboratoire de Biométrie et Biologie Evolutive, F-69622 Villeurbanne, France,
[2] Institut de Génétique Humaine, Centre National de la Recherche Scientifique, 34396 Montpellier Cedex 5, France
[3]Institut Universitaire de France
*To whom correspondence should be addressed: cristina.vieira@univ-lyon1.fr

Running head: Maternal piRNA regulate retrotransposon in somatic cells

Abstract

Transposable elements (TEs), whose propagation can result in severe damage to the host genome, are silenced in somatic and germ line cells. Silencing in the somatic cells is achieved by primary piRNA, independently of the germ line biogenesis of secondary piRNA.. Working with the *tirant* retrotransposon in Drosophila, we show here that the somatic silencing of this TE requires secondary piRNAs originated from the activation of the ping pong germline specific pathway, indicating that these two pathways are not independent.

Transposable elements (TE) are major structural elements of essentially all eukaryotic genomes (Biémont and Vieira 2006). Active TEs are highly mutagenic and their silencing is necessary to ensure genetic stability (Malone and Hanon 2009; Chiu and Greene 2008). Recent research on the PIWI-interacting RNA (piRNA) pathway in Drosophila has shown that Piwi-Argonaute proteins act in concert with TE- derived small RNAs (piRNAs) to target and silence active TEs in male and female gonads (Aravin AA et al, 2007 , Klattenhoff C and Theurkof W 2008). Two distinct piRNA pathways act within the ovary. The germ line piRNA pathway, or piRNA ping-pong pathway, is strictly associated with the germ line lineage and with the deposition of maternal piRNAs in embryos, which ensures the silencing of TEs in the offspring developing germ line (Brennecke et al, 2007, 2008, Chambeyron et al, 2008). Two classes of piRNA are implicated in the germ line regulation; the primary piRNAs that are produced from genomic clusters, and the secondary piRNAs which biogenesis is dependent on the maternal deposition of piRNA and on the amplification loop (Brennecke et al, 2007; 2008). The somatic pathway, which is ping-pong independent, controls TEs in somatic follicle cells (Malone et al, 2009; Olivieri et al, 2010). Only primary piRNAs are known to be implicated in the somatic regulation of TEs.

Well-known examples of TEs that fit with either of the pathways are the *I* factor with germ line regulation (Bucheton et al, 1990) and the gypsy endogenous retrovirus regulated by the somatic pathway (Chalvet el al, 1999). When the silencing pathways are disrupted, both types of TEs are able to infect the germ line and new insertions are detected on the chromosome of the individuals of the next generation.

Whether and how the secondary piRNAs play a role in somatic silencing of TEs is unknown. To tackle this question, we studied the regulation of the *Drosophila simulans tirant* endogenous retrovirus. We used an empty strain (Chicharo) that is free of functional *tirant* copies and a strain (Makindu) which contains *tirant* copies dispersed over the chromosome arms and which displays RNA label in both nurse and somatic cells, but show no transposition (Akkouche et al, 2012).

Considering that *tirant* is a retrovirus like element, we could expect that it's regulation was similar to those of previously described elements of *D. melanogaster*, such as *ZAM* and *Gypsy* (Pelisson et al, 1994; Chavlet et al, 1999; Leblanc et al, 2000; Desset el al, 2003; Li et al, 2009; Malone et al, 2009), and should thus be only dependent on the primary piRNA pathway on the somatic cells. Therefore, it was expected that offspring from reciprocal crosses between Chicharo and Makindu populations should present the same RNA expression levels and same cellular patterns. Under this scenario, no maternally deposited RNA was implicated in the regulation of the element. Using FISH RNA hybridization with a *tirant* probe on ovaries from parents and F1 daughters, we showed that derepression of *tirant* in follicular cells occurred when crossing females from the Chicharo strain (devoid of *tirant* copies) with males from the Makindu population, while silencing of the *tirant* was observed in the reciprocal crosses (Figure 1). We have named these crosses "**N**on **R**egulated **T**irant" (NRT) when the female from the Chicharo strain was involved, and "**R**egulated **T**irant" (RT), when the female from the Makindu strain was involved. This result suggests a maternal regulation of *tirant* somatic expression. Quantitative RT-PCR analysis further confirmed the

higher level of *tirant* transcripts in the F1 daughters from the NRT crosses, which were much higher than in the parental lines (Figure 1 C). Using a polyclonal antibody against the *tirant* envelope gene, we further showed that the derepression of *tirant* was associated with the envelope protein production, which localized in the somatic cells of the F1 daughters in which *tirant* transcripts were detected (Figure 1 D). This reveals the absence of posttranscriptional regulation of *tirant* element.

To confirm the role of secondary maternally inherited piRNAs in *tirant* retrovirus–like element silencing in the offspring of both RT and NRT crosses, we sequenced piRNAs from 0-2 h embryos, in which zygotic transcription is not yet activated. We show that Chicharo mothers deposited 5-fold fewer *tirant* piRNAs than Makindu mothers (Figure 2 A, SM1). The piRNAs maternally deposited show the ping-pong signature, confirming that ping-pong is active in the Makindu maternal strain (Figure 2 B, SM1). These differences in the amount of piRNAs were mirrored in the F1 adult gonads, where a deficit in piRNAs allows *tirant* to be active in somatic cells of the offspring of the NRT crosses.

Our data support a maternal transmission of piRNAs in *D. simulans* as previously observed in *D. melanogaster* (Brennecke et al, 2008) and *D. virilis* (Rozhkov et al, 2010). They also support the idea that maternally inherited piRNA populations promote effective responses in progeny induction of the ping-pong cycle for some TE families. However, our study reveals a critical role of secondary maternally deposited *tirant* piRNAs in the silencing of the retrovirus-like element *tirant* in somatic cells. We can speculate that secondary piRNA mediate chromatin conformation of the somatic piRNA clusters, magnifying the production of primary piRNA, and thus increasing the silencing of somatic elements. We now need to understand the "quantitative" interaction between the two pathways and how the germline secondary piRNAs interact with the primary piRNAs in the somatic cells. Our results open new questions on the variability of piRNA pathways in natural populations and contribute to a deeper comprehension of the maintenance of TEs in genomes and populations.

Acknowledgments.

This work was supported by ANR grant GENEMOBILE ANR-09-BLAN-0103, the CNRS, FINOVI, Institut Universitaire de France, and CIBLE 2008 (Région Rhône Alpes)….. We would like to thank Christian Biémont and Christophe Terzian for useful discussion, and Sonia Martinez for technical help, V. James for the English corrections.

References

Abdou Akkouche, Rita Rebollo, Nelly Burlet, Caroline Esnault, Sonia Martinez, Barbara Viginier, Christophe Terzian, Cristina Vieira, and Marie Fablet. *tirant*, a Newly Discovered Active Endogenous Retrovirus in, *Drosophila simulans* 2012, 7, jvi07146-11.

Aravin, A.,D.Gaidatzis, S.Pfeffer, M. Lagod-Quintana, P. Landgraf, N. Iovino, P Morris, MJ. Brownstein, S. Kuramochi-Miyagawa, T.Nakano, et al 2006. A novel class of small RNAs bind

Aravin AA, Hannon Gj, Brennecke J. Science 2007; 135:3 (Pubmed : 17975059).

Brennecke, J. et al (2007) Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* 128, 1089-1103.

Brennecke, J. et al (2008) An epigenetic role for maternally inherited piRNas in transposon silencing. Science 322, 1387-1392.

Bucheton, A. (1990) I transposable elements and I-R hybrid dysgenesis in *Drosophila*. *Trends Genet*. 6, 16-21

Chambeyron S, Popkova A, Payen-Groschene G, Brun C, Laouini D et al. (2008) piRNA mediated nuclear accumulation of retrotransposon transcripts in the Drosophila female germline. Proc Natl Acad Sci USA 105: 14964-14969.

Chalvet F, Teysset L, Terzian C, Prud'homme N, Santamaria P, Bucheton A, Pelisson A (1999) Proviral amplification of the Gypsy endogenous retrovirus of Drosophila melanogaster involves env-independent invasion of the female germline. *EMBO J* 18: 2659-2669

Chiu YL, Greene WC. The APOBEC3 cytidine deaminases: an innate defensive network opposing exogenous retroviruses and endogenous retroelements. Annu Rev Immunol. 2008; 26:317–353

Desset S, Meignin C, Dastugue B, Vaury C (2003) Com, A heterochromatic locus governing the control of independent endogenous retroviruses from Drosophila melanogaster. Genetics 164: 501-509.

Gilboa L, Lehmann R How different is Venus from Mars? The genetics of germ-line stem cells in Drosophila females and males. Development. 2004 Oct;131(20):4895-905.

Klattenhoff C, Theurkof W. Development 2008; 135:3 (pubmed: 18032451).

Leblanc P, Desset S, Giorgi F, Taddei AR, Fausto AM, Mazzini M, Dastugue B, Vaury C (2000) Life cycle of an endogenous retrovirus, ZAM, in Drosophila melanogaster. J. Virol 74: 10658-10669.

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

Malone, C.D. and Hannon, G.J. (2009) Molecular Evolution of piRNA and transposon control pathways in *Drosophila*. Cold Spring Harb Symp Quant Biol. 2009 ; 74: 225–234

Malone CD, Hannon GJ. Small RNAs as guardians of the genome. Cell. 2009; 136:656–668. [PubMed: 19239887]

Malone CD, Brennecke J, Dus M, Stark A , McCombie WR, Sachidanandam R Hannon Gj (2009) Specialized piRNA pathway act in germline and somatic tissues of the Drosophila ovary. Cell 137: 522-535.

McClintock, B. 1953 Induction of instability at selected loci in Maize. Genetics 38: 579-599

Olivieri D,Sykora MM, Sachidanandam R, Mechtler K, Brenencke J An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in Drosophila. EMBO J. 2010 29(19): 3301–3317.

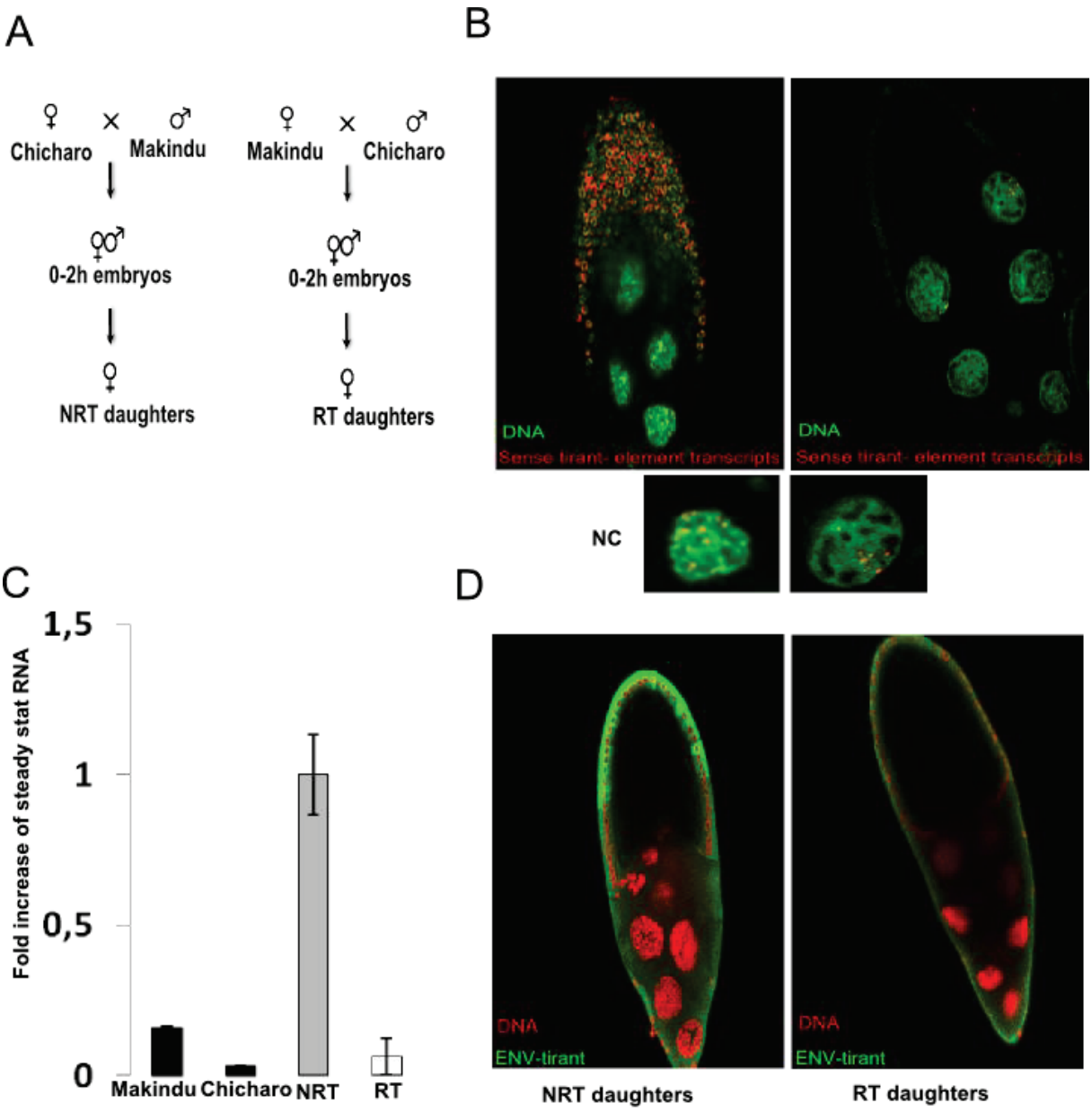
Pelisson A, song SU , Prud'homme N, Smith PA, Bucheton A, Corces VG (1994) Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue –specific control of the Drosophila flamenco gene. *EMBO J* 13: 4401-4411.

Rozhkov, N.V. et al. (2010) Small RNA-based silencing strategies for transposons in the process of invanding Drosophila species. RNA 16, 1634-1645.

Supporting Online Material

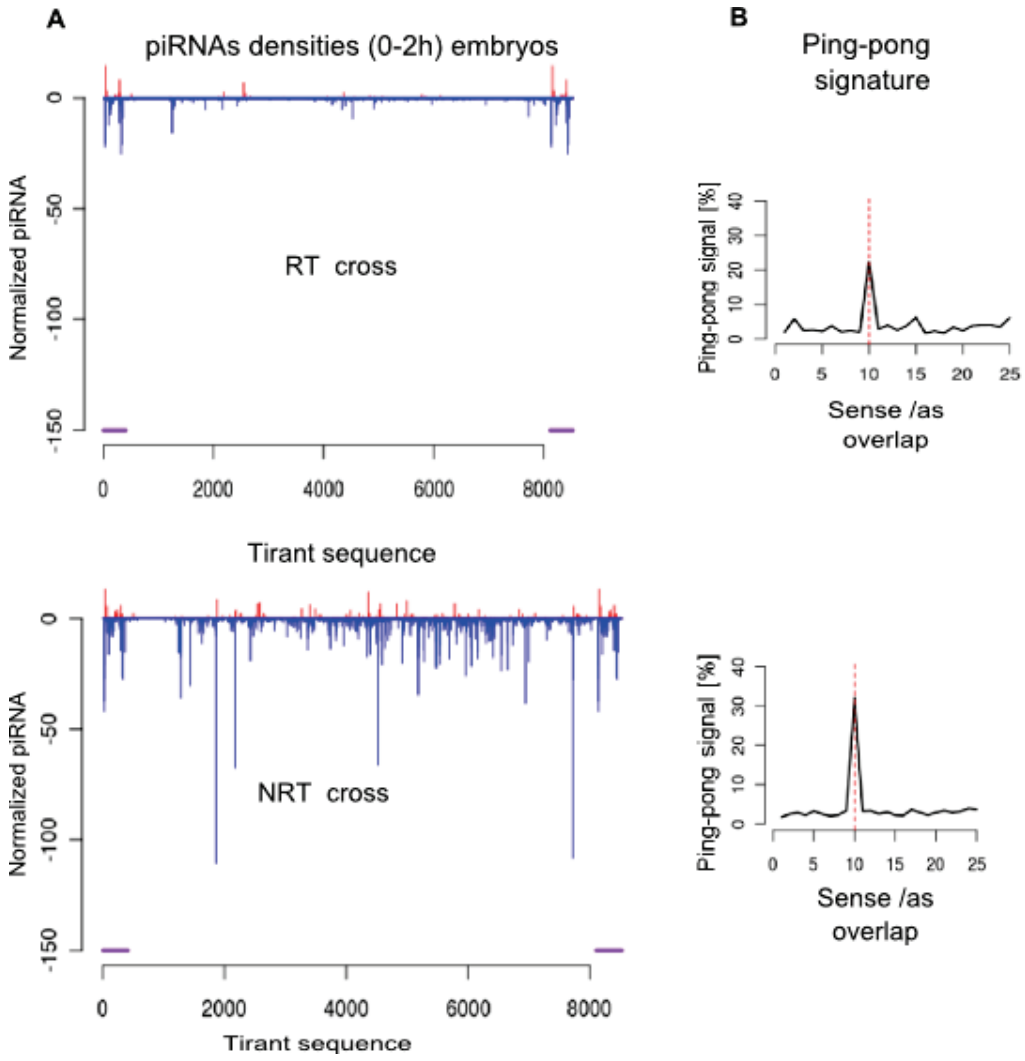
Figure legends

. **Figure 1, Akkouche et al.**



A. Crosses between the two drosophila strains, NRT (Non Regulated *Tirant* cross) and RT (Regulated *Tirant* cross) B. RNA FISH on ovaries from the F1 offspring daughters of both crosses, C. *tiran*t transcript analysis by RTqPCR in the ovaries of parental lines (black) and the F1 (gray) offspring for the NRT and RT cross, showing upregulation of *tirant* transcripts in ovaries of F1 daughters D. Immunohistochemistry with an antibody against env protein in the F1 offspring of both type of crosses.

Figure 2, Akkouche et al.



A. Density profiles of piRNAs matching the *tirant* sequence (up to 4 mismatches) split into sense piRNAs (red) and antisense piRNAs (blue). B. ping pong signature of *tirant* in both types of crosses.

Materials and Methods

Drosophila strains

We worked on fly samples from two natural populations of *D. simulans:* Chicharo (Portugal) and Makindu (Kenya). These strains were maintained in the laboratory at 18°C as isofemales lines. We previously fully characterized these strains for their *tirant* copies, and

precisely know the sequence of their insertion sites (Fablet et al. 2009) and their expression patterns (Akkouche et al 2012).

In-situ RNA hybridization and immunohistochemistry

In situ RNA analysis was carried out according to the previously described procedure (Akkouche et al 2012) using dioxigenin (DIG) – labeled strand specific *tirant* riboprobes (Akkouche et al 2012 for primers) and Fab fragments (diluted at 1/50) anti-DIG antibodies (Roche). Protein hybridization was performed according to the previous procedure (Akkouche et al, 2012).

Quantitative RT- PCR analysis

In the first-strand cDNA synthesis, 1µg of total RNA were used as a template for oligonucleotide DT primer reverse transcription using the Thermoscript RT-PCR system (invitrogen) and an oligo DT primer according to manufacturer's instructions.

Real time RT-PCR was performed in the Lightcycler 480 (Roche diagnostics). Relative abundance of the different transcripts in ovaries tissues was determined using fluo Cycle for SYBR green (Roche diagnostics) according to manufacturer's protocol using primers specific to *tirant* enveloppe gene (Akkouche et al 2012). Each reaction was performed in duplicate (triplicate?). The PCR primers were Env_f: 5' ACACGTTCCCTGAACAGACG 3', Env_r: 5' GAACGTTACCAATCCGAGCA 3 and Fw: 5' CGGATCGATATGCTAAGCTGT 3', and Rv: 5' GCGCTTGTTCGATCCGTA 3'.corresponding to the rp49 gene used to normalize the results.

Germinal small RNA purification

Ovaries of appropriated flies stocks were manually dissected in PBS 0,3 Tween, then the buffer were removed and the samples stocked at -80°C. Embryos collections where realized every 1, 75 hours on petri dishes poured with 2, 7 % (w/v) agar, 14, 4 % (v/v) vinegar, methyl 4 hydroxybenzoate 6, 3 mM, colored by neutral red and dapped with fresh yeast medium. After isolation on filter and intensive washing with PBS (REF), embryos where dechorionned during 3 minutes in 4% bleach, then washed vigorously with PBS 0,3% Tween and stocked dry at -80°C. All of the following steps are carried out on ice or at 4 °C. 100 to 150 ovary pairs were crushed in 400 µL of binding buffer (20mM Hepes / KOH pH = 7.9, 10% glycerol, 0.2 mM EDTA, 1,5 mM MgCl2, 1 mM DTT, 0,2 mM PMSF, 100 mM potassium acetate) and centrifuged at 10 000 g, 30 minutes, then supernatant was get aside. A 1 mL HitrapQ HP anion exchange column (GE healthcare, Uppsala, Belgique) was regenerated with 5 column volume (CV) of elution buffer (same as binding buffer but with 1 M potassium acetate) and equilibrated with 7 CV of binding buffer on ÄKTA purifier FPLC system. 400 µL of total ovaries clarified lysate was injected in the column and washed with 3 CV of binding buffer, flowthrough was collected in 1 mL fractions. The elution was performed using a mix of 22 % of elution buffer in binding buffer during 6 CV and collected

in 1 mL fractions. All of the fractions generated were pooled in 50 mL Falcon tube and deproteinized with 5 mL of acidic phenol / chlorophorm pH = 4.5 (Ambion, Austin, TX, USA), then the aqueous phase was re-extracted with 5 mL of chlorophorm (VWR, Fontenay sous bois, France) to avoid phenol contaminations. The nucleic acids were precipitated by adding 5 µL of 15 mg/mL dyed glycogen (Glycoblue, Ambion, Austin, TX, USA), 400 mM sodium acetate pH = 5.5 (Ambion, Austin, TX, USA) and 30 mL of absolute ethanol, and incubated over night at -80°C. The sample was transferred in Ultraclear centrifuge tube (Beckman, Palo Alto, CA, USA) and centrifugated 4 hours at 20 000 g in swi 32 Ti rotor in Beckman optima XL 100 K ultracentrifuge. Supernatant was discarded and the pellet was resuspended in 50 µL DEPC water. Small RNA were then purified on 15 % acrylamide, 8 M urea gel at 300 volts using TBE 1x as migration buffer. A radiolabbeled RNAs (Decade markers, Ambion, Austin, TX, USA) ladder was used to cut between 15 and 45 nucleotides, the strip of gel was crushed in 3 volumes of 500 mM NaCl (Ambion, Austin, TX, USA) and incubated on a wheel over night at 4 °C, then acrylamide fragments were discarded using 0.45 µm Spin X filter (Corning, Wilkes Barre, PA, USA). Small RNAs were recovered by precipitation with 1 µL of dyed Glycogen and 1 volume of isopropanol, after overnight incubation at -80°C, nucleic acids were recovered by centrifugation at 20 000 g for 45 minutes. The supernatant was discarded and the pellet was resuspended in 10 µL DEPC water. Quality and quantity of small RNAs were controled on BioAnalyser profile (Agilent technologies, Santa Clara, CA, USA) prior to the library construction.

Bioinformatics analysis of small RNA libraries

Sequenced reads were stripped of the adapter in the 3'-end and the retrieving small RNA reads were mapped to the genome sequence of *Drosophila simulans* (release 1.3) according to sequencing quality using Novoalign (www.novocraft.com). Only reads perfectly matching the fly genome were analyzed (one mismatch was nevertheless allowed on low-quality nucleotides).

Libraries were then annotated according to reference databases containing ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), micro RNAs (miRNAs), transcripts, small nuclear RNAs (snRNAs) and transposable elements sequences. rRNAs, tRNAs and snRNAs were retrieved from modEncode (http://www.modencode.org/ (Celniker et al., 2009)), miRNAs from miRBase (http://www.mirbase.org/ (Kozomara & Griffiths-Jones, 2011)), and transcripts from Flybase (http://flybase.org/). As *Drosophila simulans* transposable elements have not yet been exhaustively sequenced, libraries are annotated according to *Drosophila melanogaster* transposable elements from Repbase (http://www.girinst.org/repbase/index.html (Jurka et al., 2005).

After subtracting reads matching abundant cellular species such as rRNAs, tRNAs and snRNA, the remainders are called bonafide and can be split into siRNAs (21nt), miRNAs (22nt) and piRNAs (23-29nt). For the piRNA identification, we selected bonafide reads strictly exceeded 22 nt in length and not annotated as miRNAs.

In order to control for sequencing depth and to perform differential analyses between libraries, small RNA counts were normalized to one million bonafide reads (total amount of

microRNAs, siRNAs and piRNAs). All calculations underlying heatmaps, histograms and plots were based on the normalized libraries.

Analysis of tirant piRNAs

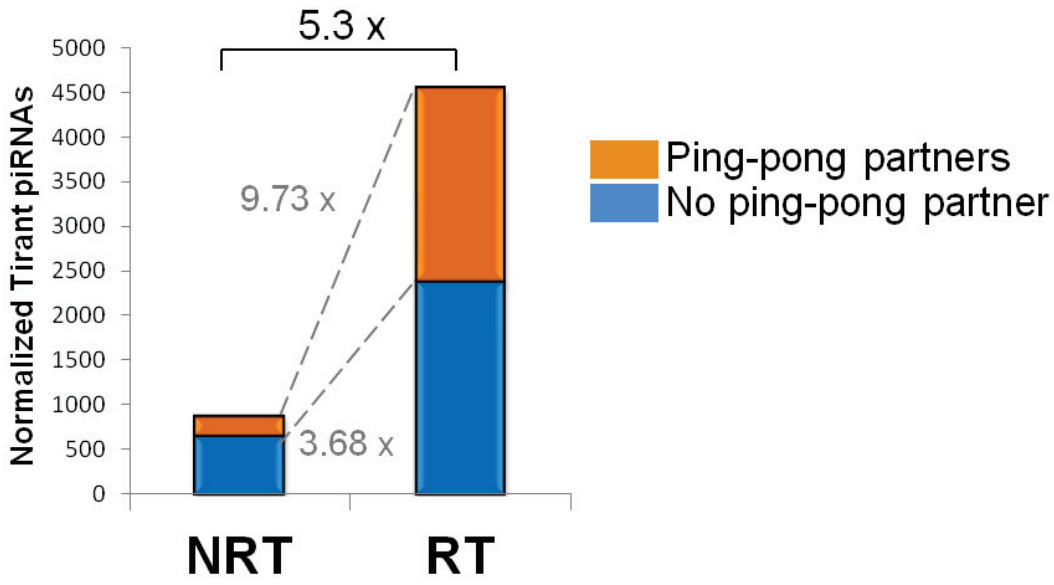
The *tirant* sequence used as consensus is the Repbase entry *TIRANT*_DM. For piRNA density plots, piRNAs were mapped onto the *TIRANT*_DM sequence allowing up to 4 mismatches. piRNAs were then collapsed to their 5' ends and plotted along the *Tirant* sequence.

Ping-pong signature and identification of ping-pong partners

The ping-pong signature aims to determine the likelihood that a piRNA has a complementary partner with a 10nt overlap and has been previously described by Brennecke et al. in 2008. Each annotated *tirant* piRNA was reduced to its 5'end. For each piRNA, that we can call "target", we determined how many annotated *tirant* piRNAs, called "partner", matched to the opposite strand in a distance from 0nt to 24nt (imperfect overlap between "target" and "partner" from 1nt to 25nt). For each "target" piRNA, the histogram of "partners" was then normalized to 100 % over all "partners". We then combined all these histograms (one histogram per annotated *tirant* piRNA), weighting each "target" histogram's contribution by the abundance of its "target" piRNA.

Ping-pong partners were defined as piRNAs which have at least one "partner" piRNA matching to the opposite strand in a distance of 9 nt (imperfect overlap of 10 nt between ping-pong partners).

Figure SM1



Profiles Comparaison of Secondary and primary piRNA (23-30nt small RNAs) obtained from embryos of F1 NRT (Male Makindu x Female chicharo) and RT (Male Chicharo x Female makindu) mapping to *Tirant* sequences.

2.3.1 Dicussion

2.3.1.1 Les piARN sont associés avec la repression des transcrits de tirant dans les cellules folliculaires

Les analyses de la dynamique de *tirant* dans les populations naturelles de *D. simulans* présentées dans les chapitres précédents ont permis la mise en évidence d'une souche dans laquelle *tirant* est réprimé post-transcriptionnellement. Les transcrits sens du rétrotransposon sont détectés dans le noyau des cellules somatiques et des cellules nourricières, mais nous n'observons aucune production de protéine (Akkouche et al, 2012). La détection de transcrits dans les cellules nourricières de souche réprimée a aussi été observée dans le cas de l'élément *I* (Chamberyon et al, 2008). Nous pouvons penser que ces transcrits sont en cours de dégradation par les piARN.

Dans l'ovaire de drosophile, les piARN sont un médiateur essentiel dans le contrôle des éléments transposables (Vagin et al, 2006, Pélisson et al, 2007). Afin de mettre en évidence un rôle éventuel de ces piARN dans la régulation de *tirant*, j'ai réalisé des croisements intraspécifiques entre la souche Makindu et la souche Chicharo, dans laquelle on n'a pas des copies actives de *tirant*. Mes résultats montrent que la répression de *tirant* est liée à la présence de piARN anti-sens capables de s'hybrider aux transcrits de *tirant*. Ceci est montré par la quantité des piARN détectés dans les embryons du croisement issu de la femelle Makindu.

Nos résultats ne montrent pas une action des piARN secondaires dans la régulation de *tirant* dans les cellules nourricières. Les transcrits de tirant ont été détectés dans les cellules nourriciéres dans les deux sens de croisements (NRT et RT) (Figure 1B Article). Ces résulats sont surprenants, puisque que l'on pourrait s'attendre à une dérepression de tirant dans les cellules nourricière des femelles du croisement NRT.

Les principales différences entre les femelles F1 (issues d'une femelle Chicharo, croisement NRT) et les femelles Makindu de la souche parentale sont la quantité et la localisation des transcrits de *tirant* dans les cellules folliculaires. Dans les femelles F1, ils sont exportés en dehors des noyaux des cellules somatiques (ils sont donc cytoplasmiques), tandis que dans la lignée de Makindu, les transcrits de *tirant* sont accumulés dans les noyaux des cellules folliculaires (Akkouche et al, 2012). Cette accumulation est corrélée avec la présence de piARN antisens de *tirant* dans les ovaires de la population Makindu, et cette accumulation disparait lorsque la quantité de piARN est réduite, dans l'hybride réciproque. Par conséquent les piARN semblent être responsables de l'accumulation nucléaire des transcrits sens dans les cellules somatiques synthétisés par les copies euchromatiques de *tirant* dans la population Makindu. Ceci semble être aussi le cas des transcrits de l'élément *I* dans les cellules nourricières de la souche I ; dans les descendants dysgéniques ces transcrits sont transportés dans l'ovocyte (Chambeyron et al, 2008).

2.3.1.2 Les piARN déposés maternellement sont nécessaires à la régulation de tirant dans les cellules somatiques

La production des piARN par le modèle de ping-pong implique la transcription des brins sens et anti-sens des copies fonctionnelles et défectives des ET. Ceci semble être le cas de l'élément *I* où la présence des copies euchromatiques actives est nécessaire pour le déclenchement du ping-pong et la production des piARN (Chamberyon et al, 2008, Brennecke et al, 2008). Toutefois, ce modèle ne s'applique pas à la régulation du rétrovirus endogène *Gypsy*. En effet les piARN anti-sens de *Gypsy* proviennent de la transcription d'une copie défective de *Gypsy* située dans le locus hétérochromatique *flamenco*, et sont produits même en absence de copies euchromatiques actives (Pélisson et al, 2007). La différence dans la mise en place du système de régulation entre *Gypsy* et l'élément *I* pourrait résulter du fait que l'élément I est spécifique à la lignée germinale femelle, alors que le contrôle de *Gypsy* se produit dans les cellules folliculaires somatiques. Ceci suggère que la biogénèse des piARN diffère selon les lignées de cellules ovariennes et diffère aussi selon les éléments (Malone et al, 2009, Olivieri D et al, 2010). Que se passe-t-il dans le cas des éléments qui, comme *tirant*, ont une double expression (somatique et germinale) ?

La quantité de piARN produits n'est pas seulement déterminée par les piARN produits par les clusters ou par les piARN produits par les copies actives des ET, mais dépend aussi des piARN transmis par la mère. Le dépôt des piARN dans les embryons précoces a déjà été décrit (Aravin et al, 2003 ; Chung wj et al, 2008), et la quantité de piARN déposés dans les embryons est corrélée à la répression de la dysgénésie des hybrides chez *D. virilis* (pour l'élément *penelope*) et chez *D. melanogaster* (pour l'élément *I*) (Blumenstiel, 2005) (Rozhkov et al, 2010) (Chamberyon et al, 2008) (Brennecke et al, 2008).

Nous avons vérifié que les piARN de *tirant* sont également présents dans les embryons précoces dans les deux sens de croisement. Le dépôt de tels piARN pourrait être responsable du déclenchement de la boucle d'amplification ping-pong ainsi que de la régulation de l'élément. Nos résultats montrent que la surexpression de *tirant* dans les cellules somatiques du croisement (NRT) est corréle avec la diminution de piARN déposés par la mère.

Un aspect surprenant de nos résultats est la dérégulation de *tirant* dans les cellules somatiques des ovaires des femelles adultes, en absence de piARN déposés maternellement dans les croisements (NRT). Ceci nous fait émettre l'idée de la connexion de ces deux voies de régulation, et du lien entre les piARN secondaires et les piARN primaires dans les cellules somatiques.

Par ailleurs, la nécessité des piARN déposés maternellement dans le processus de régulation de *tirant* a été testée par des croisements entre la souche Mayotte, où *tirant* est exprimé uniquement dans le soma (Akkouche et al, 2012) et la souche vide (Chicharo). Notre hypothèse est que l'absence de transcrits dans les cellules nourricières aboutit à un ping-pong

inactif, c'est-à-dire une absence des piARN germinaux déposés par la femelle Mayotte et qui ne seront donc pas transmis à la descendance. Notre résultat montre une dérégulation de *tirant* dans les deux sens de croisement A (Mâle Mayotte x Femelle Chicharo) et B (Mâle Chicharo x Femelle Mayotte) (Figure 22), en accord avec notre hypothèse.

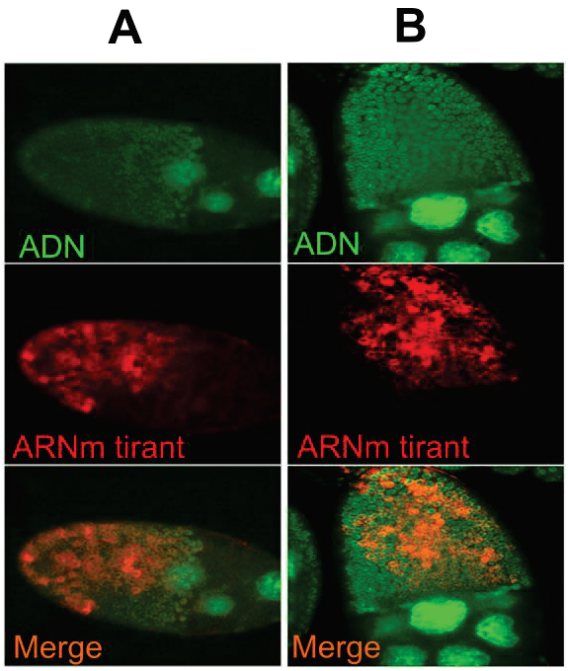


Figure 22: Hybridation *in-situ* de l'ORF3 (env) du rétrovirus endogène *tirant* (rouge) dans les descendants des croisements A (Mâle Mayotte x Femelle Chicharo) et B (Mâle Chicharo x Femelle Mayotte).

L'absence de contrôle de *tirant* dans les cellules folliculaires dans les deux sens de croisement valide notre modèle. 1) Les piARN déposés par la mère Makindu dans les descendants F1 déclenchent le mécanisme ping-pong, ce qui entraine une production de piARN germinaux secondaires, et ces piARN sont transmis vers les cellules folliculaires pour réguler *tirant*. 2) Dans le deuxième sens de croisement (Mâle Makindu x femelle Chicharo), l'absence de piARN maternels déposés par la femelle Chicharo entraine une dérégulation de *tirant* dans les cellules folliculaires.

Vu que *tirant* est un rétrovirus endogène comme *Gypsy* et *ZAM*, et le passage des transcrits vers l'ovocyte se fait à partir des cellules folliculaires, la dérégulation somatique de *tirant* est accompagnée d'une production de la protéine d'ENV dans les cellules folliculaires (Figure1D du manuscrit). La colocalisation de l'ARN avec l'ENV de *tirant* dans les cellules folliculaires suggère un mécanisme de passage des particules virales des cellules folliculaires vers l'ovocyte.

Enfin, nos données confirment la transmission maternelle des populations de piARN chez *D. simulans*, comme cela avait été précédemment observé chez *D. melanogaster*

(Chambeyron et al, 2008, Brennecke et al, 2008) et *D. virilis* (Rozhkov et al, 2010). Nos résultats sont également compatibles avec l'hypothèse que les piARN transmis par la mère déclenchent dans les descendants l'amorçage du cycle d'amplification ping-pong puisque tous les descendants qui proviennent de la mère Makindu ont une répression de *tirant* due à la présence de piARN.

2.3.1.3 Interaction entre les deux voies de piARN somatique et germinale

L'ovaire de la drosophile est constitué de cellules germinales et de cellules somatiques (les cellules folliculaires). Ces dernières jouent un rôle dans le maintien et la protection des cellules germinales. Ces deux types de cellules échangent des signaux de développement et de substances nutritives. La possibilité de passage des ARN de plusieurs rétrotransposons à LTR comme par exemple Gypsy et ZAM à partir des cellules folliculaires vers les cellules germinales a déjà été observée (Chalvet et al, 1999 ; Leblanc et al, 2000). Par contre le passage des transcrits des ET des cellules germinales aux cellules folliculaires n'a jamais été décrit. Est-ce que les piARN peuvent aussi êtres échangés entre type cellulaire ?

Chez la drosophile, chaque ovaire est composé de 16-18 unités appelées ovarioles. Dans la partie antérieure de chaque ovariole, deux ou trois cellules souches germinales « Germline Stem Cells » (GSCs) interagissent avec les cellules somatiques et affectent leur établissement et la différenciation (Gilboa et al, 2004). (Figure 23A). Plusieurs voies de signalisation réciproques existent entre les cellules souches germinales (GSC) et les cellules folliculaires qui utilisent des jonctions gap (Gilboa et lehman, 2004).

Les transcrits de *tirant* sont présents dans les GSC (Figure 23B), par conséquent un passage des piARN pourrait être possible entre les cellules germinales et les cellules folliculaires (Figure 23). Cette hypothèse reste à valider. On peut trouver dans la littérature quelques exemples de transferts de petits ARN comme chez la Paramécie, entre le micro et le macronucleus.

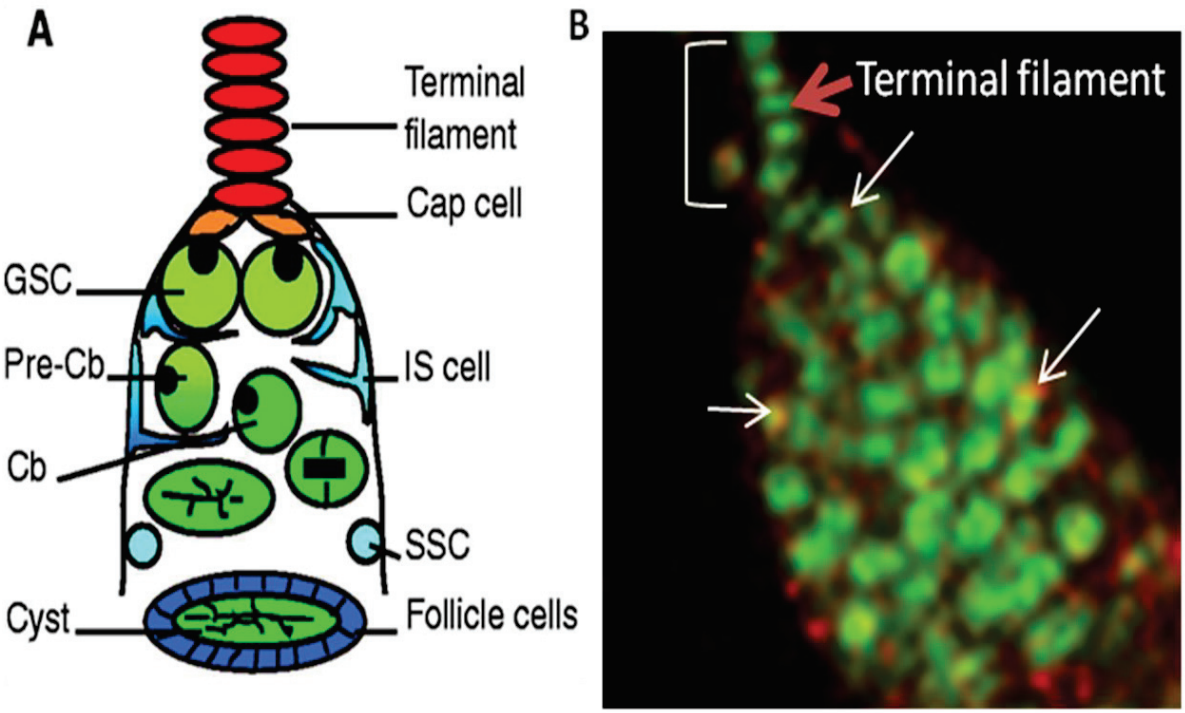


Figure 23: Modèle de transmission des piARN germinaux vers les cellules somatiques

A : Représentation schématique du plan antérieur d'un ovariole de drosophile. GSC (*Germline Stem Cells*), Cb les filles de GSC, [les cellules Cap, IS (Inner Sheats), SSC (Somatic Stem Cells) sont des cellules folliculaires]. D'après Gilboa & Lehman , 2004 . **B :** Hybridation *in situ* qui montre la présence des transcrits de *tirant* dans les GSC (en rouge : les transcrits de *tirant* et en vert : l'ADN).

Les organismes unicellulaires du genre *Tetrahymena* et *Paramecium* sont des protozoaires ciliés qui possèdent deux noyaux différents au sein de la même cellule, le micronucleus et le macronucleus (Karrer, 2000). Le micronucleus est diploïde est correspond à la lignée germinale diploïde et le macronucléus est polyploïde est correspond à la lignée somatique. Dans le micronucleus des ARN double brin sont formés par la complémentation entre deux transcrits homologues. Ils sont ensuite clivés en dehors du micronoyau par TW11, l'homologue de la protéine Piwi chez *D. melanogaster*. Il en découle la synthèse de petits ARN d'environ 28 nucléotides appelés scanARN ou scnARN. Ces scnARN sont responsables de la mise en place du macronucléus à chaque génération, par élimination de séquences repétées au niveau de l'ADN (Mochizuki et al, 2002). Chez la drosophile, les piARNsecondaires produits dans les cellules germinales pourraient être comparés aux scanARN caractéristiques des protozoaires ciliés.

2.3.1.4 Rôle des piARN dans la mise en place du TGS

En 2002, il a été montré chez *Saccharomyces pombe* que certains locus centromériques riches en séquences répétées sont ciblés par des siARN qui permettent le recrutement de plusieurs protéines dont SWI6 (homologue de la protéine hétérochromatique 1 (HP1) et Clr4 (protéine responsable de la méthylation de la lysine 9 sur l'histone H3) (Volpe et al, 2002). Des mutations dans la voie siARN aboutissent à une décondensation de la

chromatine, corrélée à une diminution de la méthylation H3K9 et donc à une délocalisation de SWI6. Des études génétiques ont de plus précisé l'action d'un complexe dévolu à la mise en place de l'hétérochromatine : le complexe RITS, formé des protéines Argonaute Ago1, Chp1 et tas3 et permettant le ciblage des protéines de type hétérochromatiques telles que SWI6 (Cam et al, 2005) (Noma et al, 2004; Verdel et al, 2004) (Verdel & Moazed, 2005). Le lien entre PTGS (*Post-Transcriptional Gene Silencing*) et TGS (*Transcriptional Gene Silencing*) a permis de révéler un nouveau rôle des voies de l'ARNi dans la structuration de la chromatine, et donc dans l'établissement d'un état épigénétique.

Malgré de nombreuses études décrivant la répression des rétrotransposons chez la Drosophile, seuls quelques travaux suggèrent un mécanisme de répression qui associe les modifications de la chromatine par le biais des modifications des histones et les piARN. Un lien entre la protéine Piwi et la protéine HP1 dans la régulation des ET a été mis en évidence récemment. Dans la lignée germinale, la protéine Piwi recrute la protéine HP1 à certains sites des séquences des ET pour induire la formation d'un état hétérochromatique (Wang & Elgin, 2011).

Les histones méthyl-transférases (HMT) permettent le recrutement de protéines spécifiques des régions hétérochromatiques via leur chromodomaine, telles que la protéine HP1 (Bannister et al, 2001). L'HMT de type SU(VAR)3-9 joue un rôle crucial dans la méthylation de H3K9. La perte de régulation de certains ET est corrélée avec la perte de HP1a (isoforme de HP1) et d'H3K9me2 au niveau des sites des ET; c'est le cas, par exemple, de l'élément HeT-A (Shpiz et al, 2011). Mais on sait aussi que la mutation de Piwi dans la lignée germinale est accompagnée par une réduction du niveau d'enrichissement de H3K9me2 sur la région 5' du promoteur de HeT-A. Le mécanisme par lequel Piwi recrute HP1a est encore mal élucidé. Ces résultats suggèrent une régulation des ET dans la lignée germinale par un mécanisme Piwi-HP1a dépendent, par condensation de la chromatine. Ces données confirment les résultats publiés par Klenov et al, en (2007), qui montrent que la mutation du gène spn-E de la voie piARN a un impact sur la structure de la chromatine au niveau des ET. Le débat sur les rôles de la voie piARN dans la mise en place du TGS ne fait probablement que commencer.

Une autre question que l'on peut alors poser porte sur l'origine des transcrits précurseurs, issus des clusters, et qui sont à l'origine des piARN primaires. Quel est le lien entre la conformation de la chromatine et les protéines de la voie piARN ?

Récemment, Rangan et al 2011, montrent un lien entre la production de piARN et la formation de l'hétérochromatine (Rangan et al, 2011) Dans cette étude les auteurs ont montré que dSETDB1 (HMT responsable de la methylation de H3K9) est requis pour le contrôle des ET dans les cellules somatiques et germinales des ovaires. La présence de la marque H3K9me3 au niveau des clusters piARN conduit à une activation de la transcription. La mutation de dSETDB1 entraine une réduction de quantités de piARN générés par le locus germinal 42AB et le locus somatique flamenco. Cette réduction est due à une diminution de

transcription des précurseurs de piARN dans ces deux locus. La perte dSETDB1 est ainsi corrélée à une augmentation du taux de transcrits des éléments germinaux comme *HeT-A* et *TART*, et des éléments somatiques comme *Gypsy* et *ZAM*.

Nous avons montré que la présence des piARN secondaires était nécessaire à la régulation de tirant dans la lignée somatique. On peut ainsi se poser la question sur l'interaction entre les piARN secondaires et la transcription des clusters responsables de la production des transcrits précurseurs des piARN primaires. Dans le cas de notre modèle, cette question est difficile à étudier, puisque nous ne connaissons pas la localisation génomique des clusters dans les lignées de *D. simulans* que nous avons utilisées.

Un deuxième point concerne le lien entre les piARN secondaires et les marques d'histones associées aux copies d'ET. Nous avons émis l'hypothèse que les piARN secondaires pourraient cibler dans les tissus somatiques l'élément *tirant*, en déposant des marques hétérochromatiques d'histones comme H3K9me2/3 ou H3K27me3, ce qui entrainerait leur répression. Les expériences sont en cours pour éprouver cette hypothèse. Ceci nous permettra de mieux comprendre le lien entre PTGS et TGS dans la régulation du rétrovirus endogène *tirant*.

PARTIE 4

2.4 Polymorphisme des gènes de la voie piARN

Note : Un des objectifs de ma thèse était d'avoir une mesure de la variabilité des systèmes de régulations épigénétique par les piARN dans les populations naturelles de drosophile. Cette variabilité, si détectée pourrait expliquer la variabilité observée dans la dynamique des ET. Dans ce chapitre je présente les données qui ont été acquises jusqu'à aujourd'hui. Leur analyse et leur interprétation est encore à l'état initial, et ce travail sera terminé après la soutenance de thèse. Les aspects de génétique de populations ont été pris en charge par Marie Fablet.

Dans les populations naturelles de *D. simulans* on observe une forte variation dans le nombre de copies de *tirant*. Ceci est aussi le cas pour de nombreux autres éléments comme cela a été décrit auparavant (Vieira et al, 1999a). Par ailleurs, les travaux de l'équipe ont montré que les populations naturelles de drosophile avaient des « charges » en ET variables, et que cette charge en ET était liée à la situation géographique des populations, les populations africaines (berceau de l'espèce) ayant une charge en ET plus faible. Nous avons mis en évidence dans ce travail le fait que *tirant* est régulé par la voie piARN. En prenant en compte les données de la littérature on s'attend à que la plupart des ET soient pris en charge par la voie des piARN. Afin de déterminer si la variabilité de nombre de copies et d'expression observée pour *tirant* et pour les autres éléments dans les populations naturelles peut être reliée à une variabilité dans la voie piARN, nous avons étudié le polymorphisme de sept gènes de cette voie ainsi que leurs niveaux d'expression dans les populations naturelles. J'ai aussi déterminé les niveaux d'expression par RTqPCR pour trois autres ET : *412*, *F* et *roo*.

Je me suis intéressé à sept gènes qui semblent avoir un rôle important dans la voie de biogénèse de piARN, soit dans la biogénèse des piARN primaires soit des piARN secondaires. Trois protéines PIWI de la famille Argonaute (Piwi, AUB, AGO3) jouent un rôle clé dans la voie des piARN : PIWI dans la biogénèse des piARN primaires dans les cellules somatiques et germinales, AUB et AGO3 dans la biogénèse des piARN secondaires.

ARMI et ZUC sont nécessaires pour la production des piARN primaires dans les cellules somatiques (Haase et al, 2010); (Olivieri et al, 2010) (Saito et al, 2010) et dans les cellules germinales (Malone et al, 2009). SPN-E est un hélicase qui intervient dans la production de piARN secondaires (Malone et al, 2009). La perte de fonction de squash (squ) affecte la production des piARN (Pane et al, 2007). Les facteurs qui interviennent dans la voie des piARN primaires dans le soma sont indispensables pour la régulation des rétrotransposons de la famille *Gypsy*, et les gènes qui interviennent dans la voie secondaire germinale régulent des ET comme l'élément *I*, *TART*, et *HeT-A* (Brennecke et al, 2009) (Schipz et al ,2009 ; 2011). L'ensemble des gènes analysés a été présenté dans l'introduction, tableau 1.

2.4.1 Taux de transcrits des ET dans les populations naturelles de D. simulans et D. melanogaster

J'ai déterminé le taux de transcrits de quatre éléments transposables *(roo*, *412*, *F* et *tirant*) dans 5 populations de *D. simulans* et 2 populations de *D. melanogaster*. Les résultats de la RT-PCR quantitative sur ovaires sont présentés dans la Figure 24. Nos résultats montrent une forte variabilité dans l'expression des ET entre populations de *D. simulans*, et aussi entre espèces. L'élément transposable *tirant* est exprimé dans la population de Makindu, mais ce n'est pas le cas dans les autres populations. Les éléments *roo* et *F* présentent une variabilité dans l'expression entre des différentes populations et l'élément *F* est fortement réprimé chez les populations de Makindu et Papette. L'élément *412* est fortement exprimé dans la population Canberra, en accord avec les données de la littérature (Borie et al, 2002), et il est moins exprimé dans les autres populations. A noter que les taux de transcrits sont très différents selon les éléments.

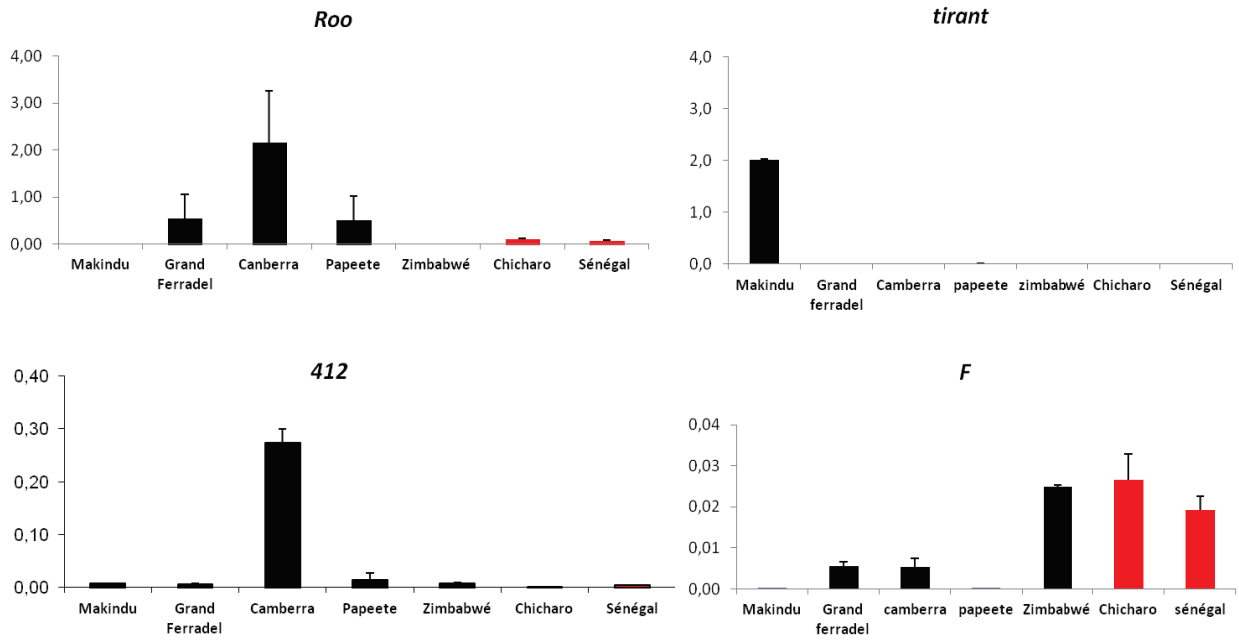


Figure 24: Quantification des transcrits des 4 ET dans les populations naturelles en noir : *D. simulans*, et en rouge *D. melanogaster*. Quantifications des transcrits de 4 ET (*412, tirant*, *Roo* et *F*) par RT-PCR quantitative. La quantité de transcrits dans les ovaires est rapportée à la quantité de transcrits du gène rp49.

En effet, *roo* et *tirant* ont des taux de transcrits bien plus importants que les éléments *412* et *F*. Nous avons voulu nous assurer que les transcrits détectés pour *412* et *F* était visibles au niveau des ovaires par FISH. Nous avons observé des transcrits de *412* dans les ovaires de la population de Canberra, mais pas de marquage pour l'élément *F* dans la population Zimbabwé (*D. simulans*) (Figure 25). Ceci nous laisse penser que l'élément *F* est probablement réprimé dans toutes les populations des *D. simulans* analysées. Nous avons aussi utilisé deux population de *D. melanogaster*, pour lesquelles un marquage à été observé.

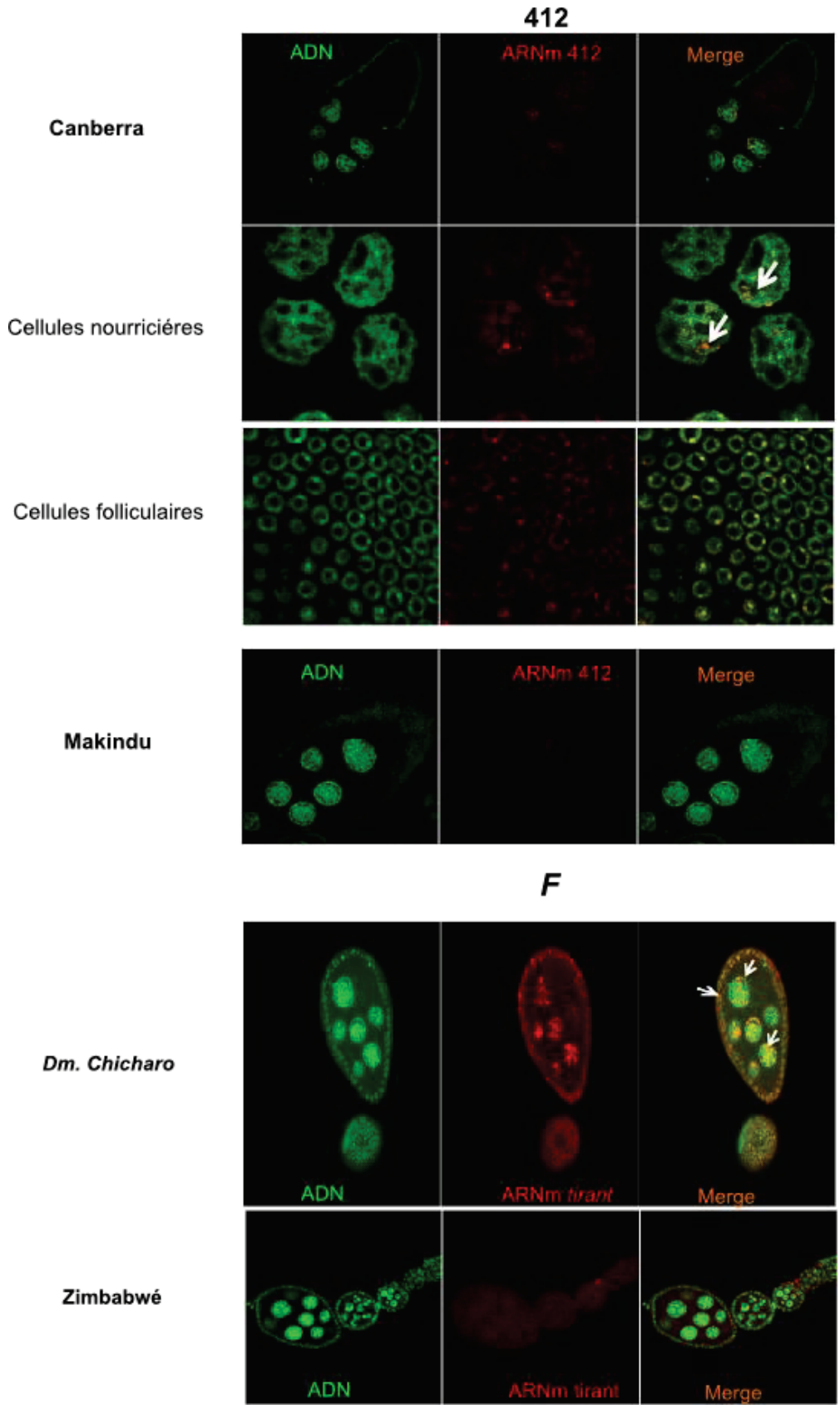


Figure 25: Mise en évidence de la présence de transcrits de l'élément *412* et *F* dans les ovaires par hybridation *in situ.*

2.4.2 Polymorphisme de séquence et taux de transcrits des gènes de la voie piARN

Nous avons voulu savoir si les différences dans les taux de transcrits des différents éléments pouvaient être liées à la présence de différents allèles des gènes de la voie piARN dans les différentes populations naturelles. Les différences de nombre de copies et d'activité des différentes familles peuvent-elles être associées à une variabilité du système de régulation, et en particulier à une variabilité des gènes impliqués dans la voie des piARN ? Existe-t-il une corrélation entre le nombre copies et /ou l'activité des éléments et le polymorphisme et/ou le niveau de transcription des gènes de la voie piARN ?

Nous avons isolé par PCR et séquencé les régions codantes des sept gènes de la voie piARN dans les mêmes 5 populations naturelles de *D. simulans* et 2 populations naturelles de *D. melanogaster*. Pour chaque population, nous avons pu identifier un ou plusieurs allèles pour chaque gène. Nous avons calculé les valeurs de diversité nucléotidique pour chaque gène à la fois chez *D. melanogaster* et *D. simulans.* (Tableau 3).

Tableau 3: Valeurs de diversité nucléotidique (π).

gène	aub	ago3	piwi	zuc	squ	armi	spnE
D. simulans	0.594%	0.595%	1.374%	1.494%	0.994%	0.787%	0.733%
D. melanogaster	0.447%	0.338%	0.445%	0.149%	0.352%	0.303%	0.314%

La tendance générale est que la diversité nucléotidique est, pour tous les gènes, plus élevée chez *D.simulans* que chez *D. melanogaster*. A partir de ces données de séquences, nous avons déterminé pour chaque gène des primers de qPCR pouvant amplifier tous les allèles mis en évidence. Nous avons ensuite mesuré par RT-qPCR, les niveaux de transcrits des différents gènes de cette voie dans 14 populations naturelles de *D. simulans* pour lesquelles nous connaissons la charge globale en ET (déterminée par hybridation in situ sur chromosomes polytènes, par famille d'ET (Vieira et al., 1999)). La mesure des taux de transcrits a été réalisée sur ovaires (trois répliquats par population).

Nos résultats montrent une grande variabilité de la quantité de transcrits mis en évidence dans les différentes populations naturelles, pour tous les gènes, comme cela est illustré par les graphiques (Figure 26) et par les valeurs très élevées des coefficients de variation (écart-type rapporté à la moyenne) (Tableau 4).

Tableau 4: Quantités de transcrits détectés par RT-qPCR dans les ovaires (rapportées à la quantité de transcrits pour le gène de référence rp49).

gène	ago3	armi	aub	piwi	spnE	squ	zuc
moyenne	0.074	0.054	0.011	0.052	0.101	0.033	0.028
écart-type	0.068	0.087	0.014	0.064	0.180	0.040	0.034
Coefficient de variation	92.5%	160.9%	120.7%	121.6%	178.9%	121.0%	123.4%

Les moyennes, écart-type et coefficients de variation sont calculés sur les données obtenues pour l'ensemble des populations analysées.

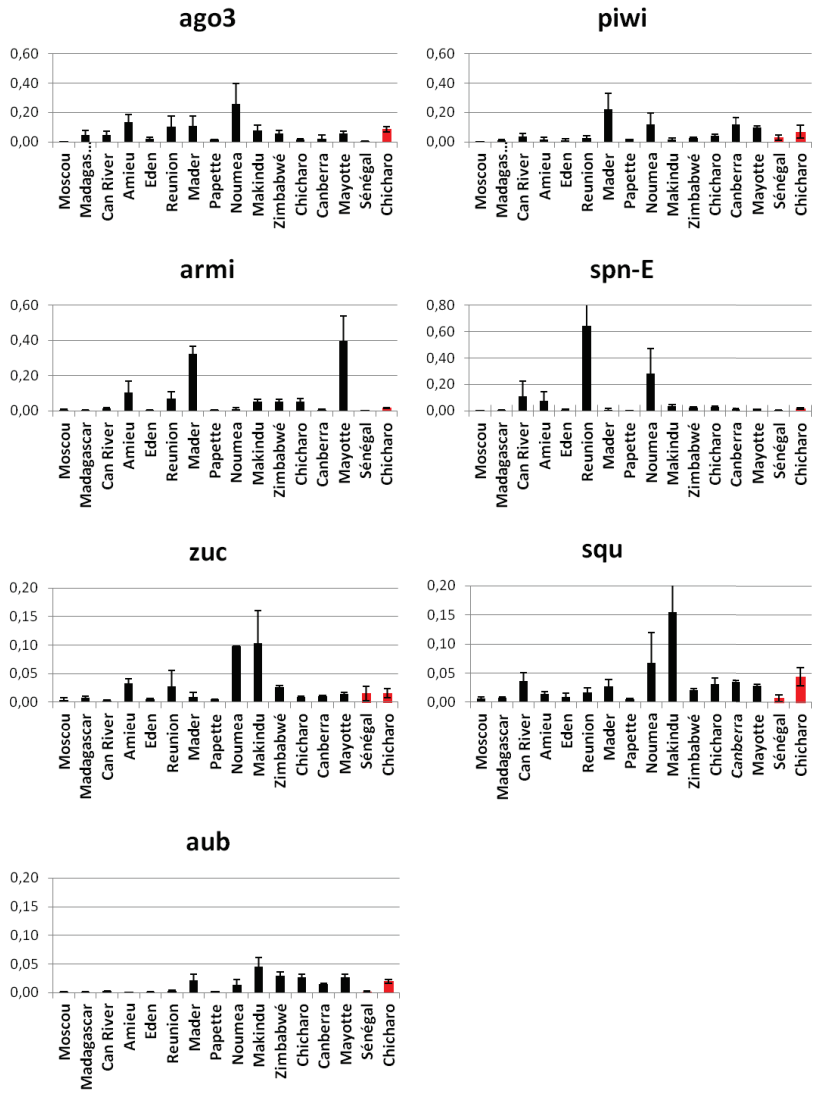


Figure 26: Quantification des transcrits des gènes de la voie piARN dans les populations naturelles en noir : *D. simulans*, et en rouge *D. melanogaster*.

La quantité de transcrits dans les ovaires est rapportée à la quantité de transcrits du gène rp49.

Nous avons ainsi pu mettre en évidence de la variabilité dans les quantités de transcrits produits par les gènes de la voie piARN. Est-il alors possible d'établir des corrélations avec la variabilité observée pour la charge en ET ?

Les données obtenues dans ce chapitre sont encore en cours de finalisation et d'analyse. Ils ont été produit en collaboration avec Marie Fablet pour les aspects analyse de séquence. Néanmoins, nous avons déjà pu mettre en évidence une corrélation positive significative entre le niveau d'expression du gène piwi et la charge globale en éléments de type LINE (ρ de Spearman = 0.73, p-value= 0.0074) ainsi qu'une corrélation négative entre le niveau d'expression de ago3 et la charge globale en éléments de type LTR (ρ de Spearman = -0.69, p-value= 0.0126). Ainsi, et même si c'est préliminaire, il semble que l'on puisse en partie relier la variabilité de la charge en ET à la variabilité d'un des partenaires de régulation par les piARN. Néanmoins, des analyses fonctionnelles plus poussées sont nécessaires pour aller plus loin dans les interprétations de ces données. Ces travaux sont actuellement en cours (voir tableau 6 Annexe C).

3 Perspectives

L'analyse *in vivo* du gène d'enveloppe impliquant la rétrotransposition de *tirant* a été largement abordée au début de cette thèse. Dans notre premier article, nous montrons l'augmentation du nombre des copies de *tirant in vivo* dans les descendants de la population qui produit la protéine d'enveloppe (Mayotte). Des travaux devront à l'avenir évaluer la portée réelle des fonctions liées au gène d'enveloppe. Une stratégie possible serait de décortiquer de façon plus approfondie le cycle de rétrotransposition de *tirant in vitro* dans les cultures de cellules de drosophiles S2.

Nous avons déterminé que *tirant* est un rétrovirus endogène actif chez *D. simulans*, mais qu'en est-il de son potentiel infectieux ? La réponse à cette question nécessitera des analyses plus approfondies. Par exemple il serait nécessaire de mettre en évidence des particules virales (VLP) pour *tirant*, et de déterminer s'il y a un passage de ces particules des cellules somatiques à l'ovocyte comme pour les rétrovirus endogènes déjà décrits. Pour comprendre ce mécanisme de passage ainsi que l'assemblage des particules virales, il sera indispensable de construire les outils permettant de suivre *in vitro* le trafic des protéines rétrovirales GAG et ENV dans les cellules. Cette description permettra ensuite d'analyser l'assemblage des particules virales de *tirant*. En effet, il est important de bien connaitre les modalités de cet assemblage pour ensuite aborder les problèmes du passage de *tirant* dans l'ovocyte. Ces constructions nous permettront d'étudier avec précision le trafic des protéines GAG et ENV et de l'ARN de *tirant* et de préciser les localisations subcellulaires à l'aide de co-marquages, par exemple avec des marqueurs d'endosomes tardifs et précoces, ainsi que de lysosomes dont le rôle dans le trafic des protéines virales a été montré.

Par ailleurs, les résultats de mes analyses montrent que la dynamique de *tirant* semble très différente entre les deux espèces sœurs *D. melanogaster* et *D. simulans*. Nous disposons maintenant d'un certain nombre d'éléments pour la compréhension de la régulation de *tirant* chez *D. simulans*, mais une analyse plus poussée des populations de *D. melanogaster* pourrait élargir notre vision des mécanismes généraux permettant à un génome de contrôler les rétrovirus endogènes qu'il héberge. Mon analyse de l'expression de *tirant* n'a porté que sur deux populations de *D. melanogaster*, et nos résultats montrent une répression de *tirant* dans ces deux populations naturelles. Il est donc indispensable de vérifier le profil d'expression de *tirant* chez d'autres populations de *D. melanogaster*. Est ce que la répression de *tirant* chez *D melanogaster* est générale à toutes les populations ? Quelques résultats préliminaires sur 10 populations ont été obtenus récemment par hybridation *in situ* et montrent que *tirant* est transcrit dans certaines populations de *D. melanogaster*. Les résultats obtenus sont présentés dans l'annexe 1.

Dans notre article « *tirant*, a newly discovered active endogenous retrovirus in *Drosophila simulans* », nous montrons que *tirant* est actif dans la population Mayotte, mais malheureusement on ne connait pas les sites d'insertions de *tirant* dans cette population. Il sera intéressant de déterminer ceux ci par genome walking. Cela nous permettra de déterminer la(les) copie(s) responsable(s) de l'activité de *tirant* et de voir si les tendances observées pour les trois modifications d'histones seront identiques dans la population Mayotte. Aussi, cela pourrait nous donner des pistes pour comprendre pourquoi l'expression de *tirant* est uniquement dans les cellules somatiques dans cette population.

Par la suite, je me suis intéressé à trois modifications d'histones associées à *tirant*. Les résultats de cette analyse indiquent également un effet important des ET sur l'expression des gènes voisins. L'insertion de *tirant* est responsable de la présence de marques hétérochromatiques associées au promoteur de ce gène. L'analyse de l'expression du gène tkv a permis de montrer que la présence de *tirant* est corrélée avec une diminution de l'expression de ce gène. Cet effet des ET doit être décortiqué de façon plus détaillée afin de comprendre l'influence des ET sur les gènes, et les interactions qui peuvent exister entre ces séquences.

Finalement, nous avons vu dans la partie 3 du chapitre 2 que la régulation de *tirant* dans les cellules somatiques est dépendante des piARN secondaires germinaux. On sait bien que, dans les cellules folliculaires, les ET sont régulés par les piARN primaires qui sont produits par des loci hétérochromatiques. Comment et par quel mécanisme les piARN secondaires interagissent-ils avec les piARN primaires ou les loci qui produisent ces piARN ?

Ce cas particulier devra être étudié afin de comprendre le lien entre les piARN secondaires et les piARN primaires, vu que c'est une question sujette à débat aujourd'hui. Là encore, notre modèle biologique apparaît idéal pour l'étude de telles questions.

D'un point de vue moléculaire, les travaux sur la régulation de *tirant* par les piARN posent aussi des questions de fond sur le mode de transmission des piARN secondaires des cellules nourricières vers les cellules folliculaires. Comment et quand ces piARN secondaires sont-ils transmis ? Mon hypothèse est que les piARN déposés maternellement déclenchent la boucle d'amplification « ping-pong », qui favorise la production des piARN secondaires, qui sont ensuite transmis vers les cellules folliculaires pour interagir avec les piARN primaires et régule les ET. Il sera nécessaire de poursuivre les recherches pour essayer de mieux comprendre comment se passe cette interaction.

L'analyse présentée ici n'a porté que sur le rétrotransposon *tirant*, pour lequel nous disposons d'un système d'étude idéal. Les résultats obtenus sont-ils particuliers à *tirant* ou peuvent-ils être généralisés à d'autres ET (pour lesquels on ne dispose pas nécessairement d'un système d'étude aussi performant) ? Il est donc indispensable de vérifier les résultats obtenus pour d'autres ET présentant aussi une double expression somatique et germinale. Il faut noter que j'ai commencé à étudier l'élément *412*, un rétrotransposon pour lequel on dispose de beaucoup de données dans l'équipe. Des résultats préliminaires par hybridation *in situ* ARN ont été obtenus récemment, et montrent une expression double de *412* dans les cellules folliculaires et les cellules nourricières des ovaires de certaines populations naturelles de *D. simulans*. Il serait donc très important de vérifier si les tendances observées dans la régulation de *tirant* peuvent être généralisées à d'autres éléments.

On sait aujourd'hui que les mécanismes épigénétiques chez la drosophile comportent deux composantes majeures (les modifications d'histones et les petits ARN). Il est donc important de considérer ces deux voies simultanément pour déterminer leurs liens dans la régulation des ET.

Outre la régulation post transcriptionnelle, les ET peuvent également être régulés transcriptionnellement. Une gamme de modifications de la structure de la chromatine réprime leur transcription, dont des modifications des histones (Slotkin and Martienssen, 2007).

Nos résultats montrent que les piARN secondaires (germinaux) régulent *tirant* dans les cellules somatiques. Récemment des études ont montré un lien entre la voie des piARN et les modifications des histones (Shpiz et al, 2011) (Wang & Elgin, 2011). C'est le cas par exemple de la régulation somatique d'*Idefix* (un retrotransposons à LTR), qui est initiée par une régulation post-transcriptionnelle (PTGS) suivie par une régulation transcriptionelle induite par les marques H3K27me3 et H3K9me3 (TGS *transcriptional gene silencing*) (Dufourt et al, 2011). Il est donc important de considérer ces deux voies ensemble pour comprendre les mécanismes de régulation de *tirant* dans les cellules somatiques. L'hypothèse est la suivante : plus la quantité de piARN secondaires déposés est importante, plus l'enrichissement en marques répressives sur *tirant* augmente. On pourra tester cette hypothèse par ChIP site spécifique de *tirant* avec des marques répressives comme H3K9me3 et H3K27me3, qui sont connues pour leur importance dans la répression des ET (Dufourt et al, 2011). Ceci serait réalisé sur des embryons et des ovaires, en comparant les deux sens de croisements (RT et NRT), qui se distinguent par des quantités différentes en piARN déposés maternellement.

Conclusion

La compréhension de l'évolution des génomes hôtes passe par une bonne connaissance de la dynamique de l'ADN répété et notamment des éléments transposables. De par leur capacité de déplacement, leur nature répétée et encore leur impact sur l'épigénome, les ET restent un facteur de variabilité (épi)génétique important (Fablet & Vieira, 2011). En absence d'un contrôle de l'activité par le génome hôte, les conséquences de la transposition peuvent être importantes, avec des effets délétères, comme c'est le cas dans les phénomènes de dysgénésie des hybrides chez la Drosophile. Les systèmes de régulation post transcriptionnelle des ET sont en place aussi bien dans la lignée germinale que dans la lignée somatique (Saito & Siomi, 2010) et constituent des voies moléculaires complexes mises à jour ces dernières années (Senti & Brennecke, 2010).

Au cours de cette thèse je me suis intéressé à un ET de type rétroviral (*tirant*), que j'ai pu décrire pour son activité et pour sa régulation post transcriptionnelle par les petits ARN de type piARN.

Notre analyse de l'activité de *tirant* dans différentes populations naturelles montre des réponses variées. *tirant* semble inégalement régulé au niveau transcriptionnel selon les populations et même selon les espèces. Ceci montre une variabilité dans les populations naturelles de la mise en place de la régulation.

Certains ET ont notamment un impact direct sur l'expression des gènes voisins. Notamment, une copie de *tirant* insérée dans un intron du gène *tkv*, et associée à des marques hétérochromatiques, conduit à une baisse de l'expression de ce gène, du fait de la diffusion des marques répressives de *tirant* à un promoteur du gène. Ces études mettent en évidence l'effet d'un ET sur les gènes voisins et montrent une diffusion des marques hétérochromatiques.

Au cours de cette thèse, j'ai tenté de trouver le lien entre la biogenèse des piARN primaires dans les cellules somatiques et les piARN secondaires de la lignée germinale. Bien que les multiples rôles des piARN secondaires soient progressivement mis à jour, et que leurs fonctions essentielles dans la régulation des ET dans la lignée germinale apparaissent évidentes, le lien entre les piARN secondaires et la régulation des ET dans les cellules somatiques n'a pas été mis en évidence. L'ET *tirant* présente l'avantage d'avoir un profil d'expression somatique et germinal, en fonction des populations. Nous avons ainsi un système idéal pour étudier cette question. Nous avons mis en évidence que la production de piARN secondaires est essentielle pour la régulation de *tirant* dans la lignée somatique. Nos résultats mettent en évidence le lien entre les deux voies et surtout mettent en évidence la variabilité populationnelle dans les systèmes de régulation.

Matériels et méthodes

Matériels et méthodes

Lignées de drosophiles

Les populations naturelles capturées dans la nature sont élevées au laboratoire. Ces souches sont des lignées isofemelles établies à partir de femelles uniques fécondées dans leur milieu naturel. Ces lignées isofemelles sont fortement consanguines. La lignée mutante pour le gène *spindle-E* nous a été amicalement donnée par Benjamin Loppin. Les souches de drosophiles sont maintenues à 25°C sur milieu levure maïs agar. (Table).

Récolte des embryons

Des mouches adultes ont été mises à pondre dans des boites contenant de la gélose composée de jus de raisin +3% d'agar, et partiellement recouverte de levure de boulanger. Deux heures après, les embryons sont récoltés grâce a un tamis, par rinçage de la gélose au PBS 1X. Les embryons ainsi récoltés sont collectés dans des tubes Eppendorf contenant du PBS 1X, déchorionés par des lavages à l'eau de javel 1%.puis l'ADN ou l''ARN sont extraits en suivant les protocoles des extractions décris ci-dessous

Extraction de l'ADN génomique

1- broyer le tissu dans 170 µL de tampon de broyage (Tris-HCl pH8 50mM+EDTA 50m + SDS 1%)

2- Ajouter 10 µl PK à 10 µg/ µL puis agiter

3- Incuber à 4ºC, 15 minutes

4- Ajouter 36 µl de KO acetate 5M, vortexer

6- Incuber à 4ºC pendant 15 minutes

7- Centrifuger (12000 rpm, 15 minutes, 4ºC)

8- Récupérer le surnageant (SN)

9- Ajouter 36 µl de KO acétate 5M, puis agiter

10- Incuber à 4ºC, 15 minutes

11- Centrifuger (12000 rpm, 15 minutes, 4ºC)

12- ajouter 1 volume de Phenol/ Chloroforme/AIA, puis vortexer 15''

13-Centrifuger (12 000 rpm, 5 minutes, 4ºC)

14- Récupérer le SN et mettre dans un nouveau tube et recommencer une deuxième fois

15- Précipitation de l'ADN :

- 1 µL de glycogène
- 1/10 volume d'acétate de sodium 3M (pH=5,2)
- 2.5 volume de l'éthanol 95% froid

16- Homogénéiser

17- Incuber toute la nuit à -20ºC

18- Centrifuger (14000 rpm, 30 minutes, 4ºC)

19- Jeter le surnageant et laver le culot avec 1ml d'ETOH 70% froid

20- Centrifuger (14000 rpm, 15 minutes, 4ºC)

21- Jeter le surnageant et laisser sécher le culot d'ADN à température ambiante pendant 2{Citation} heures

22- Reprendre le culot dans 45 µl H2O

23- Incuber 1h, 37ºC

24- Déposer sur gel d'agarose 1 % pour vérifier la qualité de l'ADN

25- Quantifier l'ADN

PCR

La plupart des réactions de PCR ont été réalisées avec la polymérase BlueTaq de Eurobio. Les amplifications PCR de la région étudiée à partir de l'ADN ont été réalisées avec les amorces répertoriées dans le tableau 6 (Annexe C). Pour chaque PCR, 2µl d'ADN ont été amplifiés dans un volume final de réaction de 50µl. Le cycle de PCR utilisé est le suivant :

Programme :

1-	94°C	5 min	
2-	94°C	30''	
3-	59°C	45''	**x 30 Cycles**
4-	68°	1 min 30''	
5-	68°C	10 min{Citation}	
6-	10°-		

Clonage et séquençage des produits de PCR

Les produits de PCR ont été clonés avec le kit TOPO-TA cloning pour séquençage (Invitrogen) suivant les instructions du fournisseur. Les bactéries chimiocompétentes transformées sont étalées en boite de Pétri contenant du milieu de culture L.B. plus ampicilline [pour 100 ml de milieu : 1,0 g peptone + 0,5 g extrait de levure + 1,0 g NaCl + 1,5 g Agar + 0,1 mg.ml-1 ampicilline), incubées à 37°C pendant 16 heures. La confirmation des colonies positives est réalisée par PCR en utilisant les amorces universelles M13 forward et reverse, qui permettent d'amplifier les fragments d'ADN insérés dans le vecteur.

Pour extraire les plasmides, j'utilise le Kit NucleoSpin de Macherey-Nagel pour les mini-préparations suivant les recommandations du fabriquant ; en éluant dans de l'eau distillée. Les produits de PCR, clonés dans des plasmides bactériens, ont été séquencés par GenoScreen et BioFidal (méthode de Sanger). Les séquences ont été alignées par (http://www.ebi.ac.uk/Tools/msa/clustalw2/).

Extraction des ARN

Les extractions d'ARN sont réalisées à partir de 30 ovaires à l'aide du kit RNeasy Protect Mini de Qiagen. Dans un premier temps, les ovaires sont disséqués sur la glace dans un tampon PBS (1x).

● Déposer les ovaires dans 100 µl de Tampon RLT (auquel on aura probablement ajouté 1% de B-mercaptoéthanol). Placer sur la glace. Broyer, tout en maintenant sur la glace.

● Ajouter 250 µl du même tampon. Resuspendre au vortex 10 sec. Centrifuger 15 sec à 4°C 10 000 rpm. Transférer le surnageant dans un tube propre.

● Ajouter 350µl d'éthanol 100%. Bien mélanger.

● Transférer les 700 µl dans la colonne. Centrifuger 15 sec à 4°C. vider le liquide recueilli.

● Ajouter 700 µl de tampon RW1 dans la colonne. Centrifuger 15 sec à 4°C. Vider le liquide recueilli.

● Ajouter 500 µl de tampon RPE dans la colonne. Centrifuger 15 sec à 4°C. vider le liquide recueilli.

● Centrifuger 2 min supplémentaires à 4°C 10000 rpm. Ceci afin de se débarrasser de tout le tampon RPE.

L'ARN est finalement repris dans 30µl d'eau ultra-pure puis centrifugé à 10000 rpm pendant une minute à 4°C. Le surnageant contenant l'ARN est conservé à -80°C. La quantité et la qualité des ARN totaux extraits sont mesurées à l'aide d'un Nanodrop.

Afin d'éviter la contamination par l'ADN les ARN extraits subissent un traitement à la DNAse, grâce au kit et au protocole DNA-free AMBION(1906) : à 10µl d'ARN sont ajoutés 1µl de tampon 10x puis 1µl de DNAse. Après mélange et incubation à 37°C, 1µl de réactif d'inactivation est ajouté, mélangé puis incubé 2 min à température ambiante. Une centrifugation de 2 min à vitesse maximale permet de séparer le réactif d'inactivation du surnageant, qui contient l'ARN et qui est récupéreé. Le surnageant contenant l'ARN est conservé à -80°C. La quantité et la qualité des ARN totaux extraits sont mesurées à l'aide d'un Nanodrop.

Synthèse des ADNc (Kit Thermoscript RT-PCR System, Invitrogen)

Toutes les étapes doivent être réalisées dans la glace

Composition du mix de synthèse :

	Réactions (Vf= 10µl)	Témoin négatif
Tampon 5x	4µl	4µl
DTT 0.1M	1µl	1µl
RNase OUT	1µl	1µl
H2O DEPC	1µl	2µl
dNTP 10mM	2µl	2µl
Thermoscript	1µl	-

Composition du mix de réaction ARN / amorces :

	Réactions (Vf= 10µl)
Oligo dT	0.5µl
Random hexamers	0.5µl
ARN (0.5µg)	xµl
H2O DEPC	qsq 10µl

- Incuber le mix ARN /amorces 5 min à 65°C

- Transférer immédiatement dans la glace pour 2 à 5 min

- Ajouter 10µl de mix de synthèse

55°C 1h30

85°C 5 min

- Ajouter 1µl de RNAse H par réaction, incuber 20 min à 37°C

- Diluer les ADNc au 1/10eme

- Stocker à -20°C

PCR quantitative sur ADNc

La PCR quantitative est réalisée avec du syber green (roche). Les ADN sont amplifiés à l'aide de couples uniques d'amorces amplifiant une taille comprise entre 80 et 200 nucléotides. Les séquences des amorces sont données en annexe dans le tableau (). La quantité des transcrits est rapportée à la quantité de messagers du gène contrôle. Plusieurs gènes de références destinés à normaliser les mesures de transcrits d'ET pour s'affranchir des variations de concentration d'ADNc d'un échantillon à l'autre ont été testés ; ils doivent en effet avoir une transcription identique chez tous les individus et toutes les populations testés. Le gène rp49 a été retenu.

- Purifier les ADN utilisés pour la gamme étalon suivant les recommandations du Kit Nucleospin Extract II, Macherey Nagel

- Diluer en cascade les ADN pour la gamme étalon :

[C] (en ng/µl)	1.10^{-1}	1.10^{-3}	$1,10^{-5}$	1.10^{-7}	1.10^{-9}
Volumes	1 µl ADNc +x µl H2O	5 µl ADNc + 495 µl H2O	5µl + 495 µl H2O	5µl + 495 µl H2O	5µl + 495 µl H2O

Les mix réactionnels contenaient :

	x1
Mix Sybergreen 2x (Roche)	5 µl
amorce F	0.3 µl
amorce R	0.3µl
H2O bidistillé	2.4 µl
ADNc $_{1/10eme}$	2 µl

- Déposer les échantillons sur une plaque pour PCR quantitative

- Centrifuger la plaque 1 min

- La PCR a été réalisée grâce à « *Light Cycler ® 480 Real-Time PCR* » avec le programme suivant :

95°C	5 min
95°C	15 sec
60°C	10 sec x 50 cycles
72°C	10 sec
Puis :	
95°C	1 min
70°C	20 sec
4°C	-

Marquage des ribosondes

La ribosonde ARN est synthétisée *in vitro* à partir d'un fragment ADN contenant la séquence du gène *env* de *Tirant* ainsi que la séquence du promoteur de la T3 polymérase. Ce Fragment ADN a été obtenu par une amplification par PCR en utilisant un couple d'amorces spécifiques du géne env (Fw 5' ACG GGG TTT GAT TAA CG 3' et Rv (T3) 5' GGA ATT AAC CCT CAC TAA AGG GAG TGT CCA GGT GTG CTG 3') dont l'amorce Rv contient en 5' le promoteur de la T3 polymérase. Ainsi, cet ADN permet de produire in vitro l'ARN anti sens d'env sous contrôle de T3.

La sonde utilisée pour l'hybridation *in situ* est marquée à l'aide de 3,5 mM d'UTP-DIG, un résidu uridine triphosphate couplé à la DIG (Roche), selon les instructions du kit de transcription ARN couplé à la DIG (Roche). La sonde utilisée en Northern Blot est marquée à l'aide 50µCI de γ P^{32} CTP, un isotope radioactif (3000 ci/mmol) selon les instructions du KIT T7 ou T3 (Roche).

La synthèse des ribosondes a été effectuée avec un mélange contenant :0.5µg d'ADN, 2µl de dNTP labelling mix DIG (Roche), 2µl de Tampon 10X (Roche), 0.5µl RNASine, 2µl de DTT (100mM), 2µl de T3 ARN polymérase et H2O qsp 20µl. Apres 2 heures d'incubation à 37°C, les matrices ADN qui ont servi lors de la transcription sont ensuite dégradées par l'addition de 20 U de DNAse I (Roche) pendant 30 min a 37°C puis la transcription est stoppée par addition de 2µl de EDTA 0.5M. Ensuite, les sondes sont purifiées en les précipitant 2 heures à -20°C par 5µl de AcNH4 5M, les sondes sont reprises dans de l'eau ultra-pure et elles sont conservées à -20°C jusqu'à utilisation.

Hybridation in situ ovaires

Des ovaires sont disséqués dans du [PBT PBS1X, 2%]. Tween sur la glace puis fixés 20 à 30 min dans tampon fixation [Formaldéhyde 3.7%, PBT, Heptane]. Les ovaires sont rincés 3 fois dans du PBT, puis post-fixés 10 min dans du tampon de post-fixation [Formaldéhyde 3.7%, PBS 1X]. Les ovaires sont ensuite lavés 5 fois dans du PBT.

L'étape de pré-hybridation est alors réalisée par l'incubation pendant 1 heure, des ovaires dans une solution d'hybridation HYB [Deionized formamid 50%, SSC5x, 0.1% Tween 20 PH=6.5]. L'hybridation des tissus avec la ribosonde marquée à la DIG est pratiquée sur la nuit à 65°C dans la solution suivante [Deionized formamid 50%, SSC5x, 0.1% Tween 20, 50mg/ml tRNA, 50mg/ml héparine, PH=6.5]. Après l'hybridation, une série de lavages est accomplie. Elle débute par 2 lavages successifs de 20' à 65°C dans la solution d'hybridation, puis d'un lavage de 20 min dans une solution composée de 50% de solution d'hybridation et 50% PBT à 65°C et enfin 4 lavages de 20 min dans du PBT à température ambiante. La sonde hybridée est détectée par le kit TSA (Perkin) utilisé selon les instructions des fournisseurs. Ce Kit permet d'amplifier le signal d'hybridation de la sonde ARN en utilisant un anticorps anti-DIG, couplé à la Peroxydase et laissé sous agitation pendant 1 heure reconnu par un anticorps secondaire couplé au fluochrome CY3. La réaction est effectuée à l'obscurité pendant 5 min. L'arrêt de la réaction est réalisé par 3 rinçages dans une solution de PBT. Les ovaires sont finalement montés entre lame et lamelle dans du milieu Vectashield (*Vector laboratories*). Les lames sont conservées à 4°C à l'abri de la lumière.

Transposon Display :

(*Zam*picinini et al. Genome 2004. 47 : 1154-1163)

1/ Préparation adaptateurs

(Site de restriction spécifique en rouge)

- MseI (*Tirant*)

AdaptMseI+ : 5' AAC AGC TGG ACG ATG AGT CCT GAG A 3'

AdaptMseI- : 5' TAT CTC AGG AGT GTA 3'

- Mélange à 10µM de chaque adaptateur dans l'eau (10µl chaque + 80µl H_2O)

- Dénaturation à 100°C, 10min (eau bouillante)

- Renaturation à température ambiante

2/ Digestion enzymatique

	Concentration finale	Volume final
Tampon enzyme 10X (Invitrogen)	1X	2µl
Enzyme (Invitrogen)	10U	2µl
DNA		5µl
H2O		11µl
V(f)		20µl

-Digestion pendant 3h à 37°C, inactivation enzyme 15min à 65°C puis 5min à 4°C

3/ Ligation

	Concentration finale	Volume final
Tampon ligase 5X (Invitrogen)	1X	8µl
T4 DNA ligase (Invitrogen)	2.5U	2.5µl
Mix adapteurs	0.2µM	0.8µl
H2O		10.7µl
Produit ligation		20µl
V(f)		42µl

-Incubation 1h à 24°C

4/ Amplification externe

Primers (vert : spécifique de l'adaptateur, bleu : spécifique de l'élément transposable et externe)

PrimerMseI : 5' AAC AGC TGG ACG ATG AGT 3' Tm = 52°C

*Tirant*Ext67- : 5' GTC TTC CCG GTT GAG TGT 3' Tm = 52°C

	Concentration finale	Volume final
Tampon 10X (Fermentas)	1X	2.5µl
dNTPs (10mM)	0.2µM	0.5µl
PrimerMseI (10µM)	0.4µM	1µl
PrimerExt67- (20µM)	2µM	1µl
Dream Taq (Fermentas)	0.625U (0.025U/µl)	0.125µl
H2O		17.5µl
Digestion/ligation produit		2µl
V(f)		25µl

Programme PCR: 94°C - 2min

94°C - 1min
52°C - 1min
72°C - 1min
} x

72°C - 4min

4/ Amplification interne

Primers (vert : spécifique de l'adaptateur, orange : spécifique de l'élément transposable interne et marqué par le fluorochrome HEX en 5'.

PrimerMseI : 5' AAC AGC TGG ACG ATG AGT 3' Tm = 52°C

*Tirant*Nes36- : 5'* TAG AGG CGT GGG GGT TTA 3' Tm = 52°C

	Concentration finale	Volume final
Tampon 10X (Fermentas)	1X	2µl
dNTPs (10mM)	0.2µM	0.5µl
PrimerMseI (10µM)	0.2µM	0.4µl
PrimerExt67- (20µM)	0.25µM	0.1µl
Dream Taq (Fermentas)	0.625U (0.025U/µl)	0.1µl
H2O		15.5µl
Produit PCR1 dilué au 1/20		1.5µl
V(f)		20µl

Programme PCR : 94°C - 1min

94°C - 45sec
52°C - 45sec
72°C - 45sec
} X 35 cycles

72°C - 3min

<u>Le Test de Troncation des Protéines (Protein Truncation test) (PTT).</u>

Le test de troncation des protéines (PTT, *protein truncation test*) s'avère particulièrement précieux pour la détection spécifique des mutations qui décalent le cadre de lecture de l'ARNm et provoquent une terminaison de traduction prématurée (Hogervorst, 1995). A la différence des autres techniques de criblage, les mutations ne sont pas détectées au niveau de l'ADN, mais au niveau des produits polypeptidiques obtenus après transcription *in vitro* des fragments d'ADN génomiques ou d'ADNc amplifiés à partir des régions d'intérêt du gène étudié. Le principe de base de la technique est la différence de taille (et par conséquent de migration éléctrophorétique) entre une protéine normale et un variant tronqué résultant par exemple de la présence d'un codon stop prématuré. Intérêt pour l'étude des éléments répétés : en une seule expérience, on obtient une vision globale de l'intégrité d'une ORF donnée pour toutes les copies d'une famille d'ET.

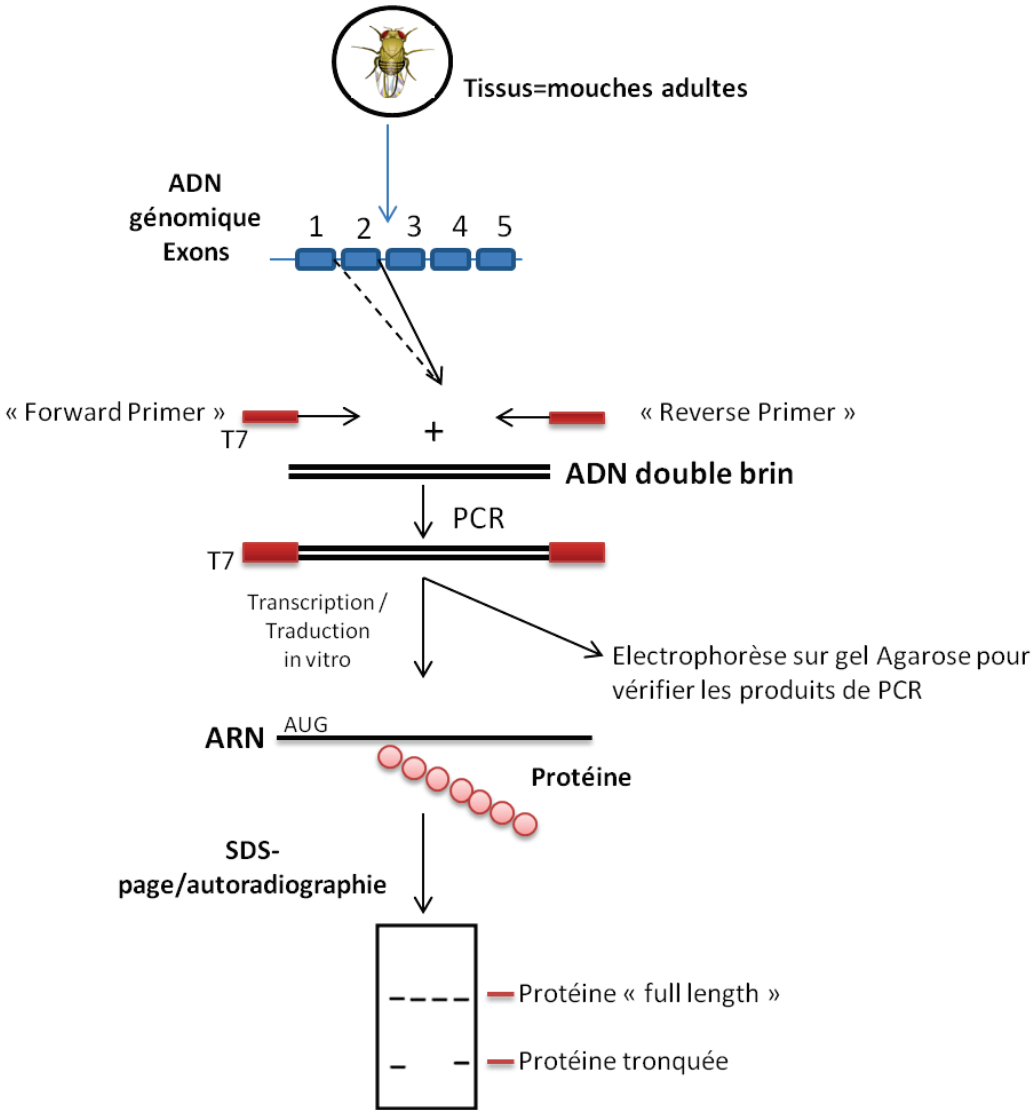


Figure 27: Schéma représentant le modèle expérimentale de PTT

Les produits PCR contenant une séquence de promoteur T7 ARN polymerase sont transcrits et traduits *in vitro* selon les instructions du Kit « TNT Quick Coupled Transcription/Translation Systems » (Promega, L1171) et le « Transcend non-radioactive translation detection system » (Promega, L5080).

- PCR1 :

Mix :

Tampon 10X	5µl
dNTP (10mM)	1µl
MgSO4(50mM)	2µl

Platinium Taq Dna Polymerase	0.2µl
EnvORFtir F (10µM)	1µl
EnvORFtir R (10µM)	1µl
H_2O	qsp 50µl
ADN	1µl

Le Tm pour les amorces utilisées est 62°C, l'hybridation se fait pendant 45s, l'étape d'élongation se fait à 68°C pendant 1min30 et les cycles sont répétés 30 fois.

EnvORFtir F : 5' TGT ATG GAA CAA TTC CAT CTG AC 3'

EnvORFtir R : 5' TTG TTT AGA GGC GTG GGG GT 3'

- PCR2 :

Mix :

Tampon 10X	5µl
dNTP (10mM)	2µl
MgSO4(50mM)	3µl
Platinium Taq Dna Polymerase	0.5µl
EnvORFtir F (10µM)	2µl
EnvORFtir R (10µM)	2µl
H_2O	qsp 50µl
ADN (dilué au 1/20)	1µl

Le Tm pour les amorces utilisées est 56°C, l'hybridation se fait pendant 1min, l'étape d'élongation se fait à 68°C pendant 1min et les cycles sont répétés 40 fois.

En 5' de chaque amorce, des sites nécessaires à la transcription et traduction en ajoutés (site de restriction, site d'initiation de la transcription, promteur T7 polymerase, polyA)

PTTenv F : 5' GGATCC TAA TAC GAC TCA CTA TAG GGA GCC ACC ATG GAA CAA TTC TGT ATG GAA CAA TTC CAT CTG AC 3'

PTTEnv R : 5’ CCT GAA TTT TTT TTT GGG TTA GAG GCG TTG TTT AGA GGC GTG GGG GT 3’

- Transcription/Traduction *In vitro* :
- kit « TNT Quick Coupled transcription/Translation Systems » (Promega, L1171)

	Volume
TNT T7 Quick Master Mix	40µl
Méthionine (1mM)	1µl
ADN (Produit de PCR)	2.5µl
TNT PCR Enhancer T7	1µl
Transcend Biotin-Lysyl-tRNA	1.5µl
H2O	4µl

- Incubation 1h30 à 30°C

- Analyse des protéines par SDS-PAGE :

2µL de produit de PCR biotinilés sont dénaturés avec 15µL de LDS Sample Buffer (Invitrogen, NP0007) pendant 15min à 70°C. Après migration sur gel NuPAGE 4-12% Bis-Tris (Invitrogen, NP0323BOX), et transfert des protéines sur membrane de nitrocellulose (Intvitrogen, LC2009), les protéines ont révélées avec le kit « Transcend non-radioactive translation detection system » (Promega, L5080). L’anticorps du kit Streptavidine-HRP dilué au 1/10000 est incubé 1h à température ambiante avec la membrane puis la membrane est révélée par addition du substrat du kit (mix 50% réactif et

50% réactif B) pendant 1 min. Puis la membrane est placée dans une cassette avec un film autoradiographique BioMax MS Films Kodak pendant 2 min (Sigma-Aldrich, Z363030-50EA) et le film est fixé et révélé par Kodak® processing chemicals for autoradiography films (Sigma-Aldrich, P7042-1GA).

Extraction des petits ARN

Les petits ARNs des ovaires de drosophiles sont extraits a l’aide du kit « mirVana miRNA Isolation (Ambion) ». Dans un premier temps, 200 paires d’ovaires sont disséquées sur la glace dans du PBS (1X). Les ovaires sont ensuite broyés dans 2 ml de tampon de lyse

(*Lysis buffer*). Après 200 µl du tampon « *homogenite additive* » sont ajoutés. Ensuite les tubes sont vortexés et ensuite sont incubés 10 min sur la glace. 2ml de Phenol Chloroforme acide (Ambion) sont ajoutés, puis les tubes sont vortexés 30 à 60 sec. Les tubes sont centrifugés pendant 5 min à 10000 g. Deux phases se distinguent alors, la phase aqueuse, qui contient les ARN, est alors récupérée. On y ajoute 1/3 volume d'éthanol 100%. Les tubes sont vortexés, ensuite la phase surnageant + éthanol est transférée sur une colonne (miRvana kit miRNA Isolation), qui est centrifugé 15 sec à 10000 g. 2/3 volume d'éthanol 100% sont de nouveau ajoutés sur le surnageant et ils sont ensuite transférés sur une deuxième colonne (mirVana kit miRNA Isolation), puis centrifugés 15 sec à 10000 g. Le surnageant est enlevé et les colonnes subissent trois lavages successifs :

1- 700µl de miRNA Wash solution 1, Centrifuger 15 sec à 10000 g

2- 500µl de Wash Solution 2/3, puis centrifuger 15 sec à 10000 g

L'étape de lavage avec la solution de lavage 2/3 est alors répétée. Ensuite, la colonne est séchée par centrifugation 2 min à 1000 g. Enfin l'ARN est repris dans 100µl de tampon d'élution (Elution Solution fournie par le Kit) à 95°C.

Les ARN extraits sont ensuite traités à la DNAse pour éliminer toute contamination par l'ADN (Voir protocole de l'extraction d'ARN total).

Marquage immunohistochimique des ovaires

Les ovaires deDes femelles âgées de 2 à 3 jours sont disséqués à froid dans une solution PBS 1X, puis fixés 20 min à température ambiante dans un tampon [PBS 1X, formaldehyde 4%]. Les ovaires sont lavés deux fois au PBS 1X à température ambiante puis incubés 1h dans du PBT contenant du tampon PBS 1X et 0,5% de TritonX- 100. Ils sont ensuite incubés une nuit à 4C dans du PBT, 0.2% BSA et l'anticorps primaire de souris dilué au 1/200 sous agitation. Les ovaires sont ensuite lavés trois fois au PBT, puis incubés 2h à température ambiante dans du PBT, 0.2% BSA et l'anticorps secondaire adéquat. Les ovaires sont ensuite lavés et sont incubés 15 min avec de l'iodure de propidium pour marquer l'ADN. Les ovaires sont ensuite montés entre lame et lamelle dans du milieu de montage Vectashield (*Vector Laboratories*).

Immunprécipitation de la Chromatine (ChIP)

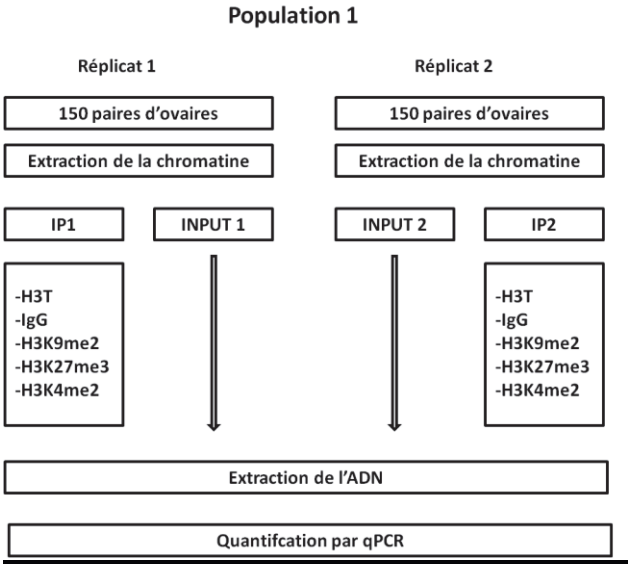


Figure 28: Schéma du plan expérimental de la ChIP.

Extraction de la chromatine

Les ovaires de Drosophile sont disséqués dans une solution de PBS 1x (Phosphate Buffer Saline). La chromatine est figée, après « cross-link » entre l'ADN et les protéines, par une incubation de 10 minutes à température ambiante dans une solution de [50mM Hepes pH8, 1mM EDTA.Na2, 0.5mM EGTA, 100mM NaCL]. La réaction est rapidement neutralisée par l'ajout de 10 ml de glycine (125mM), PBT (agitation 5 minutes à température ambiante). La réaction est ensuite centrifugée à 4000g pendant 5 minutes à 4°C. Le précipité est rincé dans 7,5 ml de PBT froid + 300µl d'inhibiteurs de protéases. Les tissus sont ensuite homogénéisés avec un pilon Potter homogénéisateur, puis centrifugés à 400g pendant une minute à 4°C. Les surnageant sont récupérés, puis centrifugés à 1100g pendant 10 minutes à 4°C. Le précipité est rincé avec 7,5 ml de « cell Lysis Buffer » [PIPES 5mM pH8, KCL 85mM, N/IGEPAL 0.5%] froid et 300µl d'inhibiteurs de protéases (Roche) et la solution est homogénéisée avec un pilon Potter Homogénéisateur, puis centrifugée à 2000g pendant 4 minutes à 4°C. Le précipité est rincé avec 1 ml de « Nuclear Lysis Buffer» [10mM EDTA, 0.5% N-lauroylsarcosine, 50mM HEPES, pH8] froid et 60µl d'inhibiteurs de protéases. La solution est incubée à température ambiante pendant 20 minutes. Le matériel cross-linké est alors repris dans 1 ml de « Nuclear Lysis Buffer» [10mM EDTA, 0.5% N-lauroylsarcosine, 50mM HEPES, pH8], dans lequel a été rajouté 60µl d'inhibiteurs de protéases.

La chromatine est fragmentée en morceaux de 400 à 600 pb par sonication (Diagenode bioruptor), 7cycles (30 secondes ON, 30 secondes OFF). Le matériel est centrifugé à 20 000 g

pendant 10 minutes à 4°C. Le surnageant est conservé et 1 ml de solution de [TritonX100 1%, 0.1% sodium deoxycholate, 0.1% SDS, 140 mM NaCL, TE1x] est ajouté. L'ADN est ensuite dosé et aliquoté. Les aliquots sont conservés à – 80°C jusqu'à utilisation.

Préparation des billes :

Peser 160 mg de protéine A Sépharose et ajouter 8 ml de PBS 1X puis incuber 1 heure à 4°C sur un plateau rotatif. La solution est centrifugée à 400g pendant 3 minutes à 4°C puis le surnageant est éliminé. Ensuite 720µl de solution, PBS 1X, BSA 0.1% sont ajoutés.

Immunoprécipitation de la chromatine (CHIP).

150µl de solution de billes de séphadex sont ajoutés sur des aliquots d'ADN. La solution est ensuite centrifugée à 1000 g pendant une minute à 4°C. A partir de cette solution, 500 µl sont conservés pour servir d'Input. La protéine d'intérêt est alors précipitée en ajoutant l'anticorps au surnageant et en laissant agiter pendant toute la nuit à 4°C : les anticorps: anti-H3K9-di-méthylé, anti-H3K27-tri-méthylé, anti-H3K4-di-méthylé ont été fournis par Millipore, et l'anti-histone H3 (ChIP Grade) a été fourni par AbCam.

Après l'incubation, l'anticorps est précipité par ajout de 80 µl de solution contenant des billes et laissé sous agitation pendant 3 heures à 4°C sur un plateau rotatif. Les échantillons sont centrifugés à 1000 g pendant 2 minutes à 4°C. Le surnageant est enlevé et les billes précipitées sont reprises dans 1 ml de solution à faible concentration en sels [TritonX100 1%, SDS 0.1%, 500mM NaCl et TE1x]. Les échantillons sont laissés sous agitation pendant 5 minutes à 4°C avant d'être centrifugés à 1000 rpm pendant 1 minute à 4°C. Le surnageant est enlevé et les billes précipitées sont resuspendues ensuite dans 1 ml de solution à forte concentration en sels [triton X100 1%, Deoxycholate 0.1%, SDS 0.1%, 140mM NaCl, 1 mM PMSF et TE1x]. Les échantillons sont laissés sous agitation pendant 5 minutes à 4°C avant d'être centrifugés à 1000 rpm pendant 1 minute à 4°C. Le surnageant est enlevé et les billes sont rincées dans 1 ml de solution de Tris-EDTA 1x (TE 1x). Les échantillons sont laissés sous agitation pendant 5 minutes à 4°C avant d'être centrifugés à 10000 rpm pendant une minute à 4°C. L'étape de rinçage dans une solution de TE 1x est alors répétée. Ensuite 100µl de TE 1x + 1µl de RNAse A (10mg/ml) sont ajoutés aux billes. La solution est incubée pendant une heure à 37°C. Les protéines associées à l'ADN sont dégradées par l'ajout de protéinase K (0.5 mg/ml) et 10 µl de SDS 10% et les échantillons sont incubés 3 heures à 55°C sous agitation. L'ADN est ensuite extrait en ajoutant un volume de phénol/ chloroforme. La solution est centrifugée à 12000 rpm pendant 5 minutes à 4°C. La phase aqueuse est transférée dans un nouveau tube et l'ADN est précipité en ajoutant 25 µl de NaAC 3M pH 5,3 et 1 µl de glycogène et 550 µl d'éthanol absolu froid. Les échantillons sont conservés 2 heure à -20°C avant d'être centrifugés à 18000 g pendant 15 minutes à 4°C.

L'ADN est lavé en ajoutant 200µl d'éthanol 70%, et centrifugé à 18000g pendant 20 min à 4°C. L'ADN est alors séché, repris dans 40µl de Tris-HCl 10mM pH8 et conservé à -20°C. Enfin l'ADN est quantifié au nanodrop et amené à une concentration de à 6ng/µl.

Annexes

Annexe A

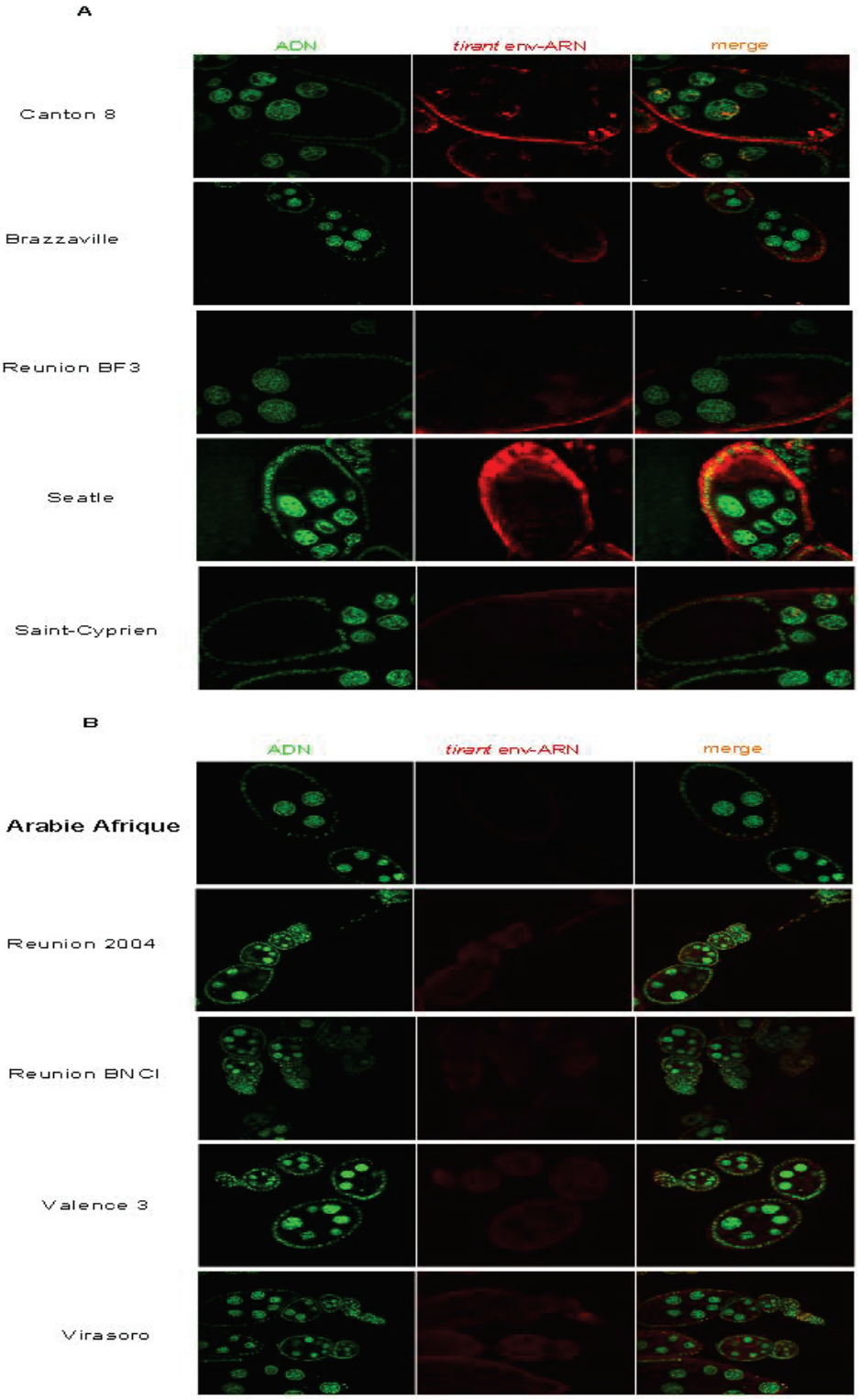


Figure 29: Analyse de l'expression de *tirant* par hybridation *in-situ* sur ovaires dans les populations naturelles de *D. melanogaster.*

(A) Les populations naturelles dont lesquelles on détecte des transcrits de *tirant* (B**)** les populations naturelles de *D. melanogaster* dans lesquelles on ne détecte pas des transcrits de *tirant.*

Annexe B

Tableau 5: Liste des couples d'amorces utilisés en RT-PCR quantitative pour quantifier les transcrits des gènes de la voie piARN.

Nom oligo		Séquence
piwiDs	for	AGGCGTCCACTTAACGAAGA
piwiDs	rev	GCAGCCATGGATGTGTGGT
aubDs	for	GACCAGCAAGAAAGGAGTGG
aubDs	rev	ACGAGGAAAGCCAGTTTGTG
ago3Ds	for	CGAGGACGTGCTAAGCTTAT
ago3Ds	rev	CAGAAAGTGATTCAGCCACA
squashDs	for	CCAGCCCTATCGCAAAGTAG
squashDs	rev	ATTCTTCTTTCTGGGAATCT
armiDs	for	GCGCAAGGAATTAAGCTATC
armiDs	rev	TGAACGCACCGTGGAGATGA
zucDs	for	AGGTGATTTGGAAGCTGGTC
zucDs	rev	GGCACTCCTTTGTGTGTGAA
spnDs	for	GCCCCAATACATTTTGGACG
spnDs	rev	CCCGTTTACTTAAGCCCACA

=

Annexe C

Tableau 6 : Données de charge en éléments transposables et de quantité de transcrits pour les gènes de la voie piARN, déterminées pour 12 populations naturelles de *D. simulans*.

	Charges en ET					Quantité de transcrits pour les gènes						
Population	chTot	chLTR	chLINE	chRetro	chDNA	ago3	armi	aub	piwi	spnE	squ	zuc
Moscou	202.5	101.5	40.5	142.0	60.5	0.00368	0.00589	0.00080	0.00351	0.00463	0.00667	0.00373
Madagascar	163.5	66.5	34.5	101.0	62.5	0.04416	0.00271	0.00097	0.01194	0.00560	0.00772	0.00823
Can-River	195.5	83.5	43.0	126.5	69.0	0.04800	0.01351	0.00256	0.03667	0.11247	0.03610	0.00272
Amieu	201.0	62.0	33.5	95.5	105.5	0.13164	0.10519	0.00045	0.01787	0.07432	0.01403	0.03329
Eden	210.0	97.0	34.0	131.0	79.0	0.02275	0.00418	0.00121	0.01252	0.00699	0.00933	0.00495
Reunion	179.0	81.0	36.0	117.0	62.0	0.10089	0.06911	0.00246	0.02466	0.64277	0.01626	0.02776
Madére	196.0	89.5	47.0	136.5	59.5	0.10984	0.32454	0.02122	0.22296	0.00879	0.02665	0.00855
Papeete	224.0	99.5	42.5	142.0	82.0	0.01342	0.00446	0.00285	0.02194	0.00129	0.00537	0.00490
Noumea	197.0	76.5	50.0	126.5	70.5	0.25975	0.01131	0.01364	0.11735	0.27976	0.06804	0.09737
Canberra	285.0	154.5	59.0	213.5	71.5	0.01960	0.00596	0.01513	0.12006	0.01390	0.03497	0.01045
Makindu	205.0	86.5	28.5	115.0	90.0	0.07576	0.05378	0.04477	0.01535	0.03448	0.15445	0.10384
Zimbabwé	200.0	100.5	36.5	137.0	63.0	0.05642	0.05120	0.02933	0.02433	0.02401	0.02024	0.02591

Les données de charge pour les différents types d'ET sont des données de l'équipe obtenues pour un grand nombre de familles d'éléments transposables par hybridation *in situ* sur chromosomes polytènes. Elles sont exprimées en nombre de copies. J'ai obtenu les données de quantités de transcrits pour les gènes de la voie piARN par RT-qPCR (quantités rapportées à la quantité de transcrits pour rp49).

Publication

JVI
Journals.ASM.org

tirant, a Newly Discovered Active Endogenous Retrovirus in *Drosophila simulans*

AQ: au Abdou Akkouche,[a] Rita Rebollo,[a]* Nelly Burlet,[a] Caroline Esnault,[a]* Sonia Martinez,[a] Barbara Viginier,[b] Christophe Terzian,[b] Cristina Vieira,[a,c] and Marie Fablet[a]

AQ: A Université de Lyon, Lyon, and Université Lyon 1, CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, Villeurbanne, France[a]; Université de Lyon, Lyon, and Université Lyon 1, INRA EPHE, UMR754, Retrovirus et Pathologie Comparée, Lyon, France[b]; and Institut Universitaire de France, Paris, France[c]

Endogenous retroviruses have the ability to become permanently integrated into the genomes of their host, and they are generally transmitted vertically from parent to progeny. With the exception of *gypsy*, few endogenous retroviruses have been identified in insects. In this study, we describe the *tirant* endogenous retrovirus in a subset of *Drosophila simulans* natural populations. By focusing on the envelope gene, we show that the entire retroviral cycle (transcription, translation, and retrotransposition) can be completed for *tirant* within one population of this species.

Endogenous retroviruses are genomic sequences that are widely dispersed throughout the host genome. These sequences constitute 8% of the human genome and represent the remnants of ancient infections by retroviruses (1). Some of these retroviruses were domesticated and generated cellular neogenes, such as the *syncytin* gene (9, 14). The structure of a canonical endogenous retrovirus consists of three open reading frames (ORFs), which are bordered by long terminal repeats (LTRs). The third ORF encodes the envelope protein (Env), which possesses fusogenic properties and is responsible for the infectious behavior of exogenous retroviruses. In insects, several endogenous retroviruses have been described, and most of them possess a complete retroviral structure. Representative insect endogenous retroviruses (IERVs) or insect ErantiViruses (24) include the following: *gypsy*, *ZAM*, *Idefix*, *tirant*, *17.6*, *297*, and *nomad* in *Drosophila melanogaster*; *tom* in *Drosophila ananassae*; *Tv1* in *Drosophila virilis*; *TED* in *Trichoplusia ni*; *Osvaldo* in *Drosophila buzzatii*; and *Yoyo* in *Ceratitis capitata* (see reference 19 for a review). With the exception of *nomad* and *Tv1*, each of these displays a complete *env* ORF. Thus far, only the well-known *gypsy* element of *D. melanogaster* has been shown to possess infectious properties (10), as Moloney murine leukemia virus pseudotypes with the *gypsy* Env protein were shown to infect insect cells (25).

The *tirant* LTR retrotransposon of *D. melanogaster* was previously described to share sequence similarity with the fusion proF1 teins of certain baculoviruses (13, 16, 20) (see Fig. 1 for the *tirant* structure). In a phylogenetic study of numerous elements from insects, Terzian et al. (24) proposed that the *tirant* element from *D. melanogaster* (GenBank accession number X93507) belongs to the IERVs. Thus, the *tirant* element was placed into this clade of retroviruses, which uses a tRNA-Ser binding site to prime reverse transcription (RT). In the study described in this report, we examined this ERV family in *Drosophila simulans* using a collection of strains with variable numbers of genomic insertions (7), and we demonstrated that this family was able to produce Env proteins in the ovaries of the host. We found that *tirant* was capable of performing the first step required for infection, i.e., the production of a functional Env protein, which suggests that it could be classified as an active endogenous retrovirus of *D. simulans*. Furthermore, we demonstrated that the complete retrotransposition cycle of *tirant* occurs within a particular natural strain of *D. simulans*, which indicates that the endogenization of *tirant* in the natural populations of this species is an ongoing process. In addition, this experimental system is of particular interest because our panel of natural strains displayed three distinct states in regard to *tirant* dynamics and thus will be ideal for deciphering the fine regulatory mechanisms: (i) the absence of transcription in most strains, (ii) transcription but the absence of translation, and (iii) transcription and translation both occurring in a single strain.

MATERIALS AND METHODS

***Drosophila* strains.** We analyzed samples of flies collected from several geographically distinct natural populations of *D. simulans*. These populations were maintained in the laboratory at 25°C as isofemale lines or as small-mass cultures with approximately 50 pairs in each generation.

The *D. simulans* strains used were previously described by Fablet et al. (7). They differed in the number of euchromatic copies of *tirant* that could be detected by *in situ* hybridization on polytene chromosomes. These strains and their respective copy numbers of *tirant* were as follows: Chicharo (Portugal, 0 copies), Makindu (Kenya, 5 copies), Zimbabwe (2 copies), Mayotte (Indian Ocean, 2 copies), and Brazzaville (Congo, 2 copies). Each population is known to display specific, unique euchromatic insertion sites, although bands representing heterochromatic sites may appear to be shared between populations in transposon display experiments.

As controls, we used a strain of *Drosophila pseudoobscura*, which is a species devoid of any *tirant* copies (17), that was kindly provided by F. Lemeunier from Gif-sur-Yvette, France, and a *D. melanogaster* isolate of a natural population from Senegal, which was grown in the laboratory and contains 9 euchromatic insertions of *tirant* that can be detected by *in situ* hybridization on polytene chromosomes (2). It should also be noted that

Received 18 December 2011 Accepted 10 January 2012
Published ahead of print •••
Address correspondence to Marie Fablet, marie.fablet@univ-lyon1.fr.
* Present address: Rita Rebollo, Terry Fox Laboratory, British Columbia Cancer Agency, Vancouver, British Columbia, Canada. Caroline Esnault, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland, USA.
Supplemental material for this article may be found at http://jvi.asm.org/.
Copyright © 2012, American Society for Microbiology. All Rights Reserved.
doi:10.1128/JVI.07146-11

AUTHOR: Publication of this article cannot proceed without the signature of the person who read and corrected the proof on behalf of all the authors: _____ signature _____ date

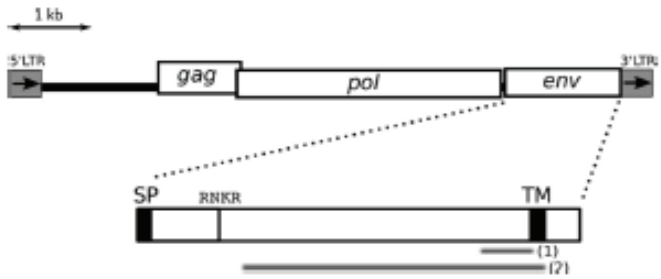


FIG 1 Structure of *tirant env*. SP, putative signal peptide, as predicted previously (4); TM, transmembrane domain, as predicted previously (27); RNKR, peptide sequence of the furin cleavage site, as determined previously (16); fragment (1), RT-qPCR amplicon; fragment (2), probe for *in situ* hybridization of the transcripts.

D. melanogaster populations display much less variability of the *tirant* copy number than *D. simulans* populations (2, 7).

Analysis of genomic copies of *tirant env*. (i) Screening for *env* full-length ORFs in natural populations of *D. simulans*. To search for full-length ORFs for *tirant env* in natural populations of *D. simulans*, we used the protein truncation test (PTT) protocol, which consists of PCR amplification of the *env* gene followed by *in vitro* transcription and translation. A T7-coupled reticulocyte lysate system (Promega) was used for the PTT analysis, according to the protocol recommended by the manufacturer (the TNT quick-coupled transcription/translation system and the Transcend nonradioactive translation detection system from Promega). The primers used for *env* amplification were designed from the *tirant* sequence (GenBank accession number AC0054444, positions 50203 to 58729) and consisted of the following sequences: forward (Fw) primer 5′ GGA TCC TAA TAC GAC TCA CTA TAG GGA GCC ACC ATG GAA CAA TTC 3′ and reverse (Rv) primer 5′ CCT GAA TTT TTT TTT GGG TTA GAG GCG 3′. The upstream primer was positioned at the beginning of ORF3, which allowed transcription to begin at the underlined ATG. We also used *D. pseudoobscura*, a species devoid of *tirant* insertions, as a negative control.

AQ: B

(ii) Analysis of *tirant env* sequence polymorphism. We isolated *tirant env* sequences from a subset of the pool of natural populations by PCR using the following primers: Fw primer 5′ TGT ATG GAA CAA TTC CAT CTG AC 3′ and Rv primer 5′ TTG TTT AGA GGC GTG GGG GT 3′. These PCR products were then cloned (TOPO TA cloning kit; Invitrogen) and sequenced (Sanger method; GenoScreen, France). The DNA sequences were aligned and clustered using the Seaview software (version 4) (8; http://pbil.univ-lyon1.fr/software/seaview.html), and DNA polymorphisms were measured using the DnaSP software (version 5) (21). The alignment of the sequences is provided in the supplemental material.

Analysis of *tirant env* transcription. (i) Isolation of *tirant* mRNAs and synthesis of cDNA. Total RNA was extracted from pools of 20 to 30 ovaries from adult females using an RNeasy kit (Qiagen) and subsequently treated with DNase (DNA-free kit; Ambion). One microgram of the total RNA was then converted into cDNA using a Thermoscript Invitrogen kit primed with a mix of oligo$(dT)_{20}$ and random primers.

(ii) Quantification of *env* transcripts of *tirant* by qPCR. The cDNAs were diluted 50 times and then quantified using SYBR green quantitative PCR (qPCR) in a LightCycler apparatus (Roche Diagnostics) with primers specific to the *tirant env* gene (Fw primer 5′ ACA CGT TCC CTG AAC AGA CG 3′ and Rv primer 5′ GAA CGT TAC CAA TCC GAG CA 3′) (see Fig. 1 for the position of the amplicon). The transcript quantity was estimated relative to that of the *rp49* gene, for which we used the following primers: Fw primer 5′ CGG ATC GAT ATG CTA AGC TGT 3′ and Rv primer 5′ GCG CTT GTT CGA TCC GTA 3′. The *rp49* gene demonstrated equivalent amounts of transcripts for all of the strains analyzed, which enabled the amount of *env* transcripts to be compared. Measurements were obtained from two independent experiments; i.e., RNAs were extracted from two different pools of ovaries. For each experiment, the reactions were performed in duplicate, and the standard curves were calculated using serial dilutions of genomic DNA. The primers were tested for all the strains, and an efficiency value greater than 1.9 was maintained.

(iii) Detection of *tirant env* transcripts by *in situ* hybridization in ovaries. Ovaries were dissected in PBT (1× phosphate-buffered saline [PBS], 0.2% Tween 20), fixed for 20 to 40 min, washed in PBT, and then incubated in 1× PBS containing 3.7% formaldehyde for 10 min at room temperature. Hybridization (overnight) and washing were performed at 65°C. The *tirant* antisense riboprobe was a 1-kb fragment corresponding to the *env* gene and included a T3 promoter site in the reverse primer (Fw primer 5′ ACG GGG TTT GAT TAA CG 3′ and Rv primer T3-5′ GGA ATT AAC CCT CAC TAA AGG GAG TGT CCA GGT GTG CTG 3′). The riboprobe was labeled by the *in vitro* transcription of T3 using a digoxigenin (DIG)-RNA labeling mix (Roche). DIG-labeled RNA probes were detected using anti-DIG Fab antibody fragments (1/50 dilution; Roche) and fluorescence amplification (Tyramide signal amplification kit; PerkinElmer). An RNase H control was included to prevent RNA-DNA hybridization in the experiments. Specifically, after the hybridization and washing steps, the ovaries were treated with RNase H (0.3 U/μl) for 1 h at 37°C and then evaluated by immunodetection. To stain the DNA, the ovaries were incubated in PBS containing Sytox green (Invitrogen), and visualization was performed with a Zeiss LSM510 Meta confocal microscope.

Analysis of *tirant* Env protein production by detection of *tirant* Env proteins in ovary preparations using immunochemistry. (i) Antibodies. A polyclonal antibody against the *tirant* Env protein was produced via the DNA vaccination of mice. The Env cDNA sequence was synthesized and cloned into a pVax plasmid designed using the ICAntibodies technology (developed by In Cell Art for *in vivo* gene expression), and this plasmid was injected into a Swiss mouse.

(ii) Immunochemistry. Ovaries were dissected in PBS on ice and fixed with 4% paraformaldehyde in PBS with 0.2% Triton X-100 at room temperature for 20 min. The tissue samples were then washed three times for 2 min each with PBS. After blocking in PBS containing 0.2% Triton X-100 and 0.1% bovine serum albumin (BSA) for 2 h at room temperature, the ovaries were incubated overnight at 4°C with the primary antibody at a 1/50 dilution in PBS containing 0.2% Triton X-100 and 0.1% BSA. After three washes in PBT, the ovaries were incubated with the secondary antibody (1/600 dilution of Alexa 488-conjugated anti-mouse IgG [H+L]; Molecular Probes) for 2 h at room temperature. For the negative controls, the ovaries were incubated with a 1/500 dilution of mouse preimmune serum for the corresponding primary antibody. The DNA was stained with propidium iodide, and confocal images were acquired using an LSM510 microscope (Carl Zeiss).

Detection of *tirant* retrotransposition by transposon display. Total genomic DNA was isolated from individual flies using a standard phenol-chloroform extraction procedure after proteinase K digestion. The transposon display was performed using a modified version of the protocol employed previously (6). Briefly, genomic DNA from individual adult flies was digested with 10 units of MseI for 3 h at 37°C, and the adaptors consisted of the following sequences: 5′ AAC AGC TGG ACG ATG AGT CCT GAG A 3′ (AdaptMseI+) and 5′ TAT CTC AGG AGT GTA 3′ (AdaptMseI−). The ligation of the adaptors was performed at 37°C for 1 h with 2.5 units of T4 DNA ligase. During the first round of amplification, 0.625 units of Dream *Taq* polymerase (Fermentas), 0.2 μM deoxynucleoside triphosphate, 0.4 μM the adaptor-specific primer (5′ AAC AGC TGG ACG ATG AGT 3′), and 0.2 μM the *tirant*-specific primer (5′ GTC TTC CCG GTT GAG TGT 3′) were used. During the second amplification round, 0.2 μM the adaptor primer, 0.25 μM the nested *tirant*-specific primer with 5′ hexachloro-6-carboxyfluorescein (HEX) fluorescent labeling (5′ TAG AGG CGT GGG GGT TTA 3′), and 0.625 units of Dream *Taq* polymerase (Fermentas) were used. The PCR was performed for 35 cycles with amplification steps that lasted for 45 s. Negative controls that used the adaptor-specific primer or the element-specific primer alone were

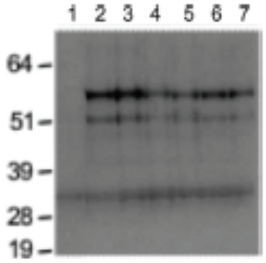


FIG 2 Protein truncation test for *tirant env* in *D. simulans* natural strains. Western blotting of the products from *in vitro* transcription and translation. Lanes: 1, *D. pseudoobscura* negative control; 2, *D. melanogaster* strain from Senegal; 3 to 7, *D. simulans* strains from Mayotte, Brazzaville, Zimbabwe, Makindu, and Chicharo, respectively. The molecular mass ladder (in kDa) is included on the left side of the image. All of the tested *D. simulans* natural populations displayed identical profiles that corresponded to a potentially functional *env* ORF. The fainter, lower 33-kDa band is not *tirant* specific because we also found this band in the *D. pseudoobscura* negative control (a species devoid of *tirant* insertions).

included. One microliter of the PCR products was loaded onto a 3730XL capillary DNA analyzer (Applied Biosystems) with a GS600LIZ standard size marker (Applied Biosystems). The raw data were analyzed using the GeneMapper software (Applied Biosystems).

We measured the number of novel bands that were obtained in the progeny (7 to 10 males and females) of five independent pairs for the Makindu population and four independent pairs for the Mayotte population. The bands were considered novel only when they were absent from both parental samples and were present in no more than one individual in the progeny.

Nucleotide sequence accession numbers. The nucleotide sequences were deposited in the GenBank database with the following accession numbers: JN786085 to JN786091 for Makindu, JN786092 to JN786098 for Zimbabwe, JN786099 to JN786104 for Chicharo, and JQ219965 to JQ219971 for Mayotte.

RESULTS

***D. simulans* strains display potentially functional *env* genes.** We evaluated the integrity of the *tirant env* genes in our strains using a PTT, which consists of the amplification of genomic DNA sequences by PCR, followed by *in vitro* transcription and translation. In this technique, subsequent Western blotting reveals whether the *env* gene contains a full-length ORF. A premature stop codon or internal deletions at the genomic level appear as bands of smaller mass compared with those of the expected full-length protein. Our assay revealed the presence of complete ORFs for *tirant env* in the tested natural strains of *D. simulans*, which demonstrates that these sequences could potentially be functional *in vitro* (Fig. 2). Similarly, the *tirant env* ORFs of *D. melanogaster* were found to be full-length ORFs. Indeed, Western blotting after *in vitro* translation indicated two specific bands that corresponded to approximately 56 kDa and 52 kDa. As the expected product would give a band corresponding to 56.2 kDa, we assumed that the observed signal represented a complete and potentially functional Env protein. The shorter products may correspond to the use of a weaker downstream AUG signal (located 177 bp downstream) or to the presence of a premature stop codon or deletions in certain *tirant* copies that conserve the reading frame.

We isolated and analyzed in detail the sequences of *tirant env* from a subset of these populations (Makindu, Mayotte, Zimbabwe, and Chicharo) (see the supplemental material for sequence alignment). We did not observe length polymorphism in the PCR products, although the sequencing of a few clones revealed the existence of two sequence variants (Fig. 3). A previous study of *tirant* sequences identified two subfamilies in natural populations of *D. simulans*, based on differences in the sequences of the regulatory regions of the element: the C subfamily is transcriptionally active and found in both euchromatin and heterochromatin, and the S subfamily is silenced and exclusively heterochromatic (7). Both of the variants identified here belonged to the transcriptionally active C subfamily of the *tirant* copies. One of the sequence variants (referred to here as variant Ca) was found in all four tested strains (GenBank accession numbers JN786085 to JN786091 for Makindu, JQ219965 to JQ219969 for Mayotte, JN786092 to JN786095 and JN786097 for Zimbabwe, and JN786103 and JN786104 for Chicharo), and the sequences were highly similar (nucleotide diversity [π] = 0.0079). This variant presumably corresponds to the translated sequence from the PTT assay and is thus a putatively functional sequence. The other variant (referred to here as variant Cb) was found in all the samples except that of the Makindu strain (GenBank accession numbers JQ219970 and JQ219971 for Mayotte, JN786096 and JN786098 for Zimbabwe, and JN786099 to JN786102 for Chicharo), and this variant also displayed high levels of sequence similarity among the strains (π = 0.0059). However, this variant presented a premature stop codon. The calculation of π for the consensus of each variant resulted in a value of 0.0429. The sequencing of the *env* cDNAs that were obtained from the total mRNA in the Mayotte sample revealed that both variants are transcribed.

***tirant* produces *env* transcripts in the ovaries of certain *D. simulans* strains.** We tested four *D. simulans* strains for the presence of *tirant env* transcripts. Our RT-qPCR experiments showed that only two strains, Makindu and Mayotte, had significant levels of *tirant env* transcripts in the ovaries (Fig. 4A). Even the *D. melanogaster* strain, which harbors significantly more insertions within the euchromatin, did not contain *tirant* transcripts in the ovaries.

To visualize the localization of the *tirant* transcripts in the germ line, we performed RNA fluorescent *in situ* hybridization for the ovaries. As expected, no transcripts were detected in the strains that showed no RT-qPCR signal, namely, Zimbabwe and Chicharo, or in the *D. melanogaster* strain (Fig. 4B). We observed that *tirant* transcripts accumulated in the mature egg chambers of the ovaries of the Makindu and Mayotte strains (a mature egg chamber consists of the oocyte and nurse cells, which both constitute the germ line, surrounded by the somatically derived follicle cells) (Fig. 4B). However, the patterns of transcript localization were different between these two strains; *tirant* transcripts were detected in both the follicle and nurse cells for the Makindu strain, whereas these transcripts accumulated substantially in the follicle cells and were absent from the nurse cells for the Mayotte strain. This observation may also explain why the transcript levels, as determined by RT-qPCR, were higher in the Makindu strain than the Mayotte strain.

***tirant env* transcription results in protein production in one *D. simulans* strain.** We evaluated the production of *tirant* Env proteins using immunochemistry, which was performed on whole ovaries using the *tirant* Env antibody. We observed staining for *tirant* Env in the ovaries of the Mayotte strain, whereas no clear signal for Env was detected in the other samples (Fig. 4C). Our immunofluorescence analysis revealed the presence of the Env protein at a late stage (stage 9) of oogenesis in the follicle cells of the Mayotte strain.

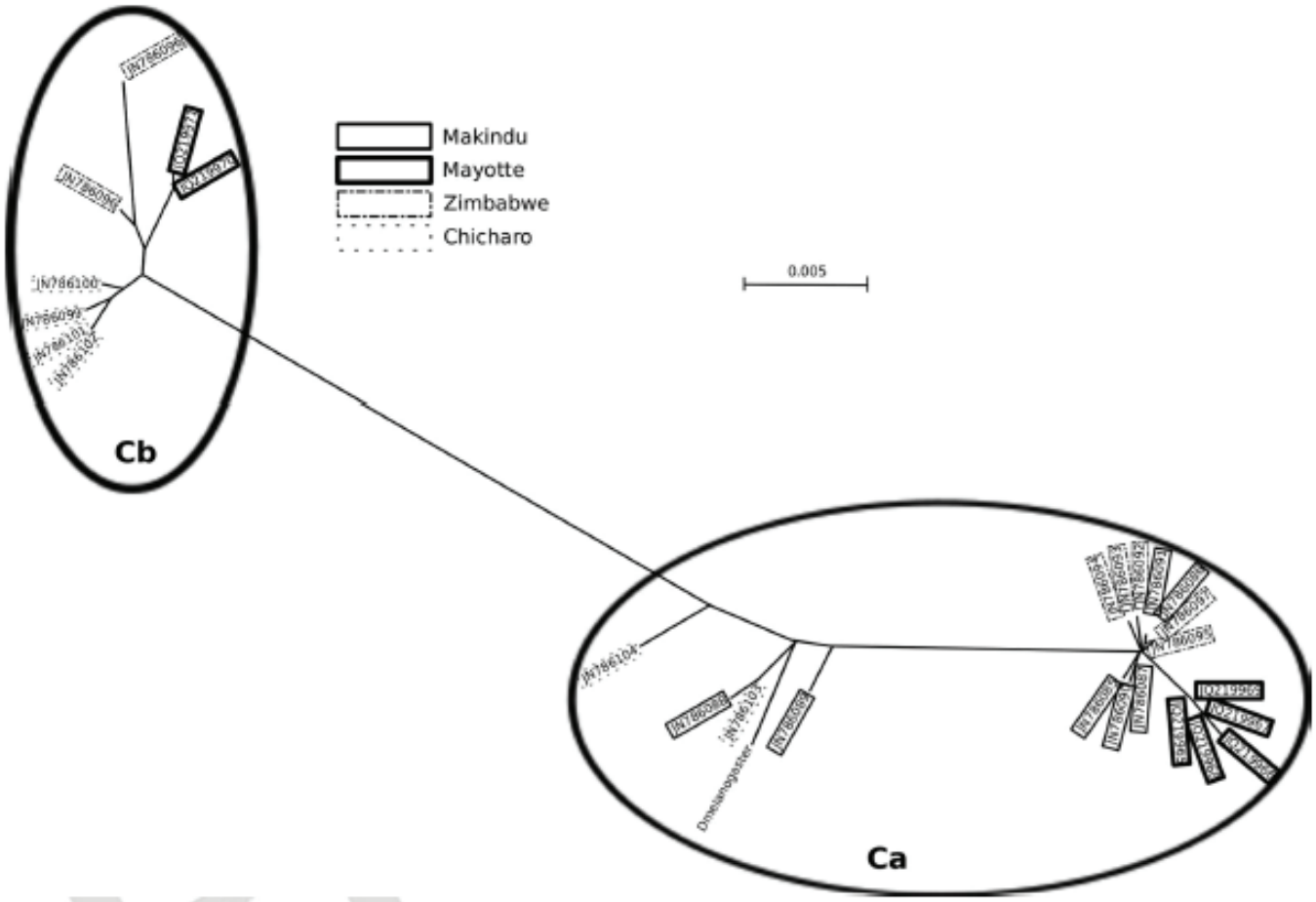


FIG 3 Neighbor-joining reconstruction of the relationships between *tirant env* sequences. Included is a reference from a *tirant* sequence from the *D. melanogaster* genome, which was extracted from contig AC005444 at positions 50203 to 58729. Variant Ca, which was found in all four of the tested strains, is the canonical sequence for the *tirant env* gene, whereas variant Cb contains a premature stop codon.

***tirant* is active in the Mayotte strain.** The insertion of a retrotranscribed copy into a new genomic site would represent the final step of the endogenous retroviral cycle of *tirant*. To assess whether this step occurs in the Mayotte population, we performed a transposon display. This technique consists of digesting genomic DNA, ligating adaptors to the obtained fragments, and performing a nested PCR with one primer specific to the adaptor and the other primer specific to *tirant*. The observation of the amplified fragment profile allows the detection of insertions of *tirant* in a given genome, and the comparison of the transposon display profiles among the mother, the father, and the progeny subsequently allows the identification of new bands, which are present in only one unique descendant and are absent from both parents. The analysis of four pairs of individuals from the Mayotte strain revealed the presence of novel *tirant* bands in the progeny, which were numbered between 2 and 22, depending on the pair analyzed (Table 1). This finding indicates that *tirant* does retrotranspose in this population. T1

Furthermore, we repeated the *in situ* hybridization on polytene chromosomes, which we had performed approximately 10 years ago (2). Whereas *tirant* had previously been estimated to display two insertions in the Mayotte strain, the current study was able to identify seven of these insertions in the same strain. It should be noted that transposon display and *in situ* hybridization on polytene chromosomes are not expected to provide identical results. Indeed, heterochromatic copies were revealed by the transposon display technique only, which explains the greater number of bands observed in the transposon display experiments.

Interestingly, in the Makindu population, we identified between zero and five novel bands by the transposon display, which suggests the existence of either a *tirant* provirus amplification process that is independent of Env production or chromosomal rearrangement events.

DISCUSSION

***tirant*, a newly discovered active endogenous retrovirus of *Drosophila simulans*.** Similar to other IERVs, *tirant* was previously shown to contain copies with a complete third ORF (4, 15, 27). Short transcripts that potentially correspond to the subgenomic spliced form of *env* were detected by Northern blotting in young *D. melanogaster* embryos (15), but no substantial evidence was found to indicate the potential activity of this gene. In the present study, we have demonstrated at least two natural strains of *D. simulans* in which the *tirant env* gene is transcribed, and we found that the corresponding Env protein is produced in one of these strains (Table 2). Therefore, *tirant* appears to be a newly identified active endogenous retrovirus of *D. simulans*, a species in which no other active endogenous retrovirus has been reported thus far. T2/AQ:D

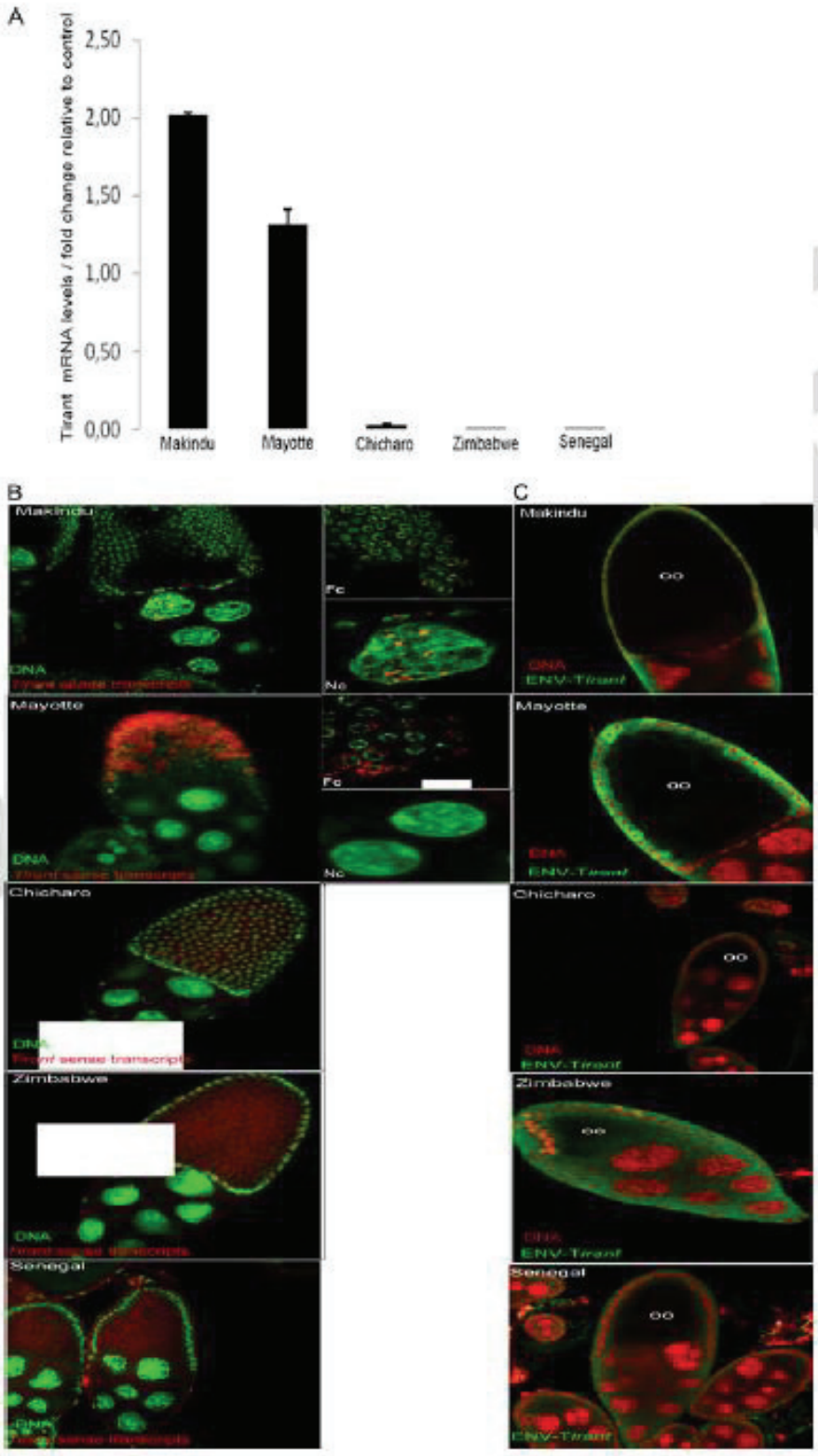


FIG 4 Expression of the *tirant env* gene in the ovaries of natural strains of *D. simulans*. (A) RT-qPCR of the ovaries. The transcript levels were estimated relative to those of the *rp49* gene. Significant levels of *tirant env* transcripts were detected in the Makindu and Mayotte populations. *D. melanogaster* Senegal was used as a control strain. (B) *In situ* hybridization of *tirant* transcripts (red). DNA is labeled in green (Sytox Green). Fc, follicle cells; Nc, nurse cells. (Left) View of ovary chambers (magnification, ×40); (right) details of the left panel. *tirant* transcripts were found in the nurse cells and follicle cells of the Makindu strain and in the follicle cells of the Mayotte strain. No staining was observed in the other strains. (C) Detection of *tirant* Env protein by immunochemistry (green) in the Mayotte strain. DNA is stained in red (propidium iodide). oo, oocyte. The green halos in the other strains do not represent Env staining but rather correspond to autofluorescence, as they were also observed following staining with preimmune serum.

COLOR

TABLE 1 Transposon display results

Strain	Pair no.	No. of male/no. of female progeny	No. of male/no. of female novel bands	No. of male/no. of female progeny without novel bands
Makindu	1	10/9	0/0	10/9
	2	10/9	0/0	10/9
	3	9/10	1/1	9/9
	4	9/10	3/1	8/9
	5	10/10	4/1	8/9
Mayotte	1	8/10	3/4	6/8
	2	10/10	0/2	10/8
	3	10/7	8/10	7/4
	4	10/10	7/15	8/3

Furthermore, we found that this IERV was not active in *D. melanogaster*, which suggests that the same endogenous retrovirus can display different behaviors in different but closely related species. This is not the first report to describe such differences between *D. melanogaster* and *D. simulans*. For example, one recent study revealed that the respective contents of the transposable elements in these species are quite different; *D. melanogaster* harbors mainly complete copies, whereas *D. simulans* displays many deleted and degenerated copies of elements within the same family (12).

These findings indicate that each strain of *D. simulans* studied possesses at least one potentially complete functional *tirant env* gene in its genome. A detailed sequence analysis revealed that each of these populations had highly similar sequences for *tirant env*. Moreover, another variant of *env* that contains a premature stop codon was also found in most of the studied populations (in all except Makindu).

In this study, we found one *D. simulans* strain in which *tirant env* was both transcribed and translated. In this strain, the staining for Env proteins was observed in the follicle cells of the ovaries during the later stages of development. Other IERVs have also been shown to display activity in the ovaries of certain strains that are referred to as "permissive" or "unstable." In these strains, for example, Tanda et al. (22) found that *tom* RNAs accumulate in the germarium, in nurse cells (germ line cells), and in the oocyte nucleus. *idefix* was shown to be expressed in very specific early follicle cells (somatic cells) within the ovaries of unstable strains (23). *gyspy*, which is the only infectious IERV identified to date, was shown to produce Env proteins that accumulated in the follicle cells of permissive females at stage 10 of oogenesis (18). Leblanc et al. (11) studied *ZAM*, which is the IERV most closely related to *tirant*, and found *ZAM* Env proteins in the follicle cells surrounding the posterior portion of the oocytes in *D. melanogaster* ovaries at stages 9 and 10.

For *gypsy* and *ZAM*, the somatic expression of the *env* genes was thus demonstrated. However, only a transfer from the soma to the oocyte (germ line cell) can result in an increase in the copy number of these IERVs in the following generation. This mechanism was demonstrated for *gypsy* (5) and *ZAM* (3), and it likely involves the yolk protein traffic machinery that operates between the follicle cells and the oocyte (3). In the Mayotte strain, *tirant* transcripts are located in the follicle cells, which is also the case for *gypsy* and *ZAM*. Our transposon display results, together with the more recent *in situ* hybridization experiments on polytene chromosomes, show that *tirant* does transpose in this strain. Therefore, we may assume that the infection of an oocyte by *tirant* viral particles occurs and that these particles are produced in the surrounding follicle cells as they are for *gypsy* and *ZAM*.

The *tirant* element: a model for the study of ERV proliferation in natural populations. As described for other endogenous retroviruses of *Drosophila*, such as *ZAM* and *gypsy*, most strains are classified as "restrictive" or "stable," which means that the activity and copy number of the endogenous retroviruses are limited. In addition, there are certain "permissive" or "unstable" strains for which this inhibition is no longer detected. Our study assessed the original features of *tirant* in a collection of natural populations, and we detected the following three categories of strains that differed according to their *tirant* dynamics: (i) the Chicharo category had zero (or very few, as in the case of Zimbabwe) euchromatic copies of *tirant* and no detectable production of transcripts; (ii) the Mayotte category, which displays the qualities of a permissive strain, possessed *tirant* copies that were able to undergo the entire retrotransposition cycle; and (iii) the Makindu category had an intermediate state, where *tirant* was transcribed but not translated. Most natural populations worldwide behave similarly to those within the Chicharo category and present no detectable *tirant* activity. Our results indicate that these strains have potentially functional *tirant env* genes, which are, however, silenced.

Therefore, at least in the Mayotte strain, *tirant* is able to complete the retroviral cycle and can therefore be considered an active endogenous retrovirus of *D. simulans*. However, because we do not have the sequences of all the *tirant* entire copies, we cannot exclude the possibility that the expression of the other ORFs may originate from different copies and that the retroviral cycle is achieved through complementation. It remains unclear why *tirant* could not be translated in the Makindu strain. However, we can postulate that a posttranscriptional control of *tirant* occurs in the Makindu strain, which is not the case in the Mayotte strain. Our

TABLE 2 Results obtained with the different strains

Population	No. of euchromatic copies[a]	mRNA level[b]	*In situ* localization of mRNAs	PTT result	Env protein
Makindu	5	1.53	Nurse cells +, follicle cells +	Complete ORF	ND[c]
Mayotte	2	1	Follicle cells +++	Complete ORF	Follicle cells
Zimbabwe	2	0.02	ND	Complete ORF	ND
Chicharo	0	4.10^{-4}	ND	Complete ORF	ND
Senegal (*D. melanogaster*)	9	3.10^{-5}	ND	Complete ORF	ND

[a] As previously determined by *in situ* hybridization on polytene chromosomes (2).
[b] Relative to the Mayotte level.
[c] ND, not detected.

collection of natural populations, which provides examples of different *tirant* dynamics, is an ideal tool for deciphering such mechanisms.

Furthermore, it is tempting to propose that these different natural strains illustrate different time periods of the history of the *tirant* endogenous retrovirus in the *D. simulans* genome (26). When an active copy of *tirant* is inserted into a favorable genomic site, it can be transcribed and translated and can retrotranspose to other sites, thereby increasing its copy number. This was likely the situation for the Mayotte strain. In the context of the "arms race" hypothesis, the *D. simulans* genome likely develops strategies to combat this retroviral invasion, such as posttranscriptional locking, which appears to be the case in the Makindu strain. Such locking may also occur in *D. melanogaster* populations, where the number of euchromatic insertions is relatively high compared with *D. simulans*, although no significant transcript levels can be detected in these populations. However, because both species display very different dynamics related to transposable elements (12), we may assume that *tirant* regulation is under the control of many distinct mechanisms.

ACKNOWLEDGMENTS

This work was supported by the ANR grant GENEMOBILE from the CNRS, FINOVI, IUF, and CIBLE 2008 (Région Rhône Alpes).

We thank Christian Biémont for his useful discussions and Geoffrey Hutinet for his technical help. The English quality was certified by American journal experts.

AQ: ref

REFERENCES

1. **Belshaw R, et al.** 2004. Long-term reinfection of the human genome by endogenous retroviruses. Proc. Natl. Acad. Sci. U. S. A. 101:4894–4899.
2. **Biémont C, et al.** 2003. Worldwide distribution of transposable element copy number in natural populations of Drosophila simulans. Evolution 57:159–167.
3. **Brasset E, 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 3:25.
4. **Cañizares J, Grau M, Paricio N, Moltó MD.** 2000. Tirant is a new member of the gypsy family of retrotransposons in Drosophila melanogaster. Genome 43:9–14.
5. **Chalvet F, 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.
6. **Esnault C, et al.** 2008. High genetic differentiation between the M and S molecular forms of Anopheles gambiae in Africa. PLoS One 3:e1968.
7. **Fablet M, McDonald JF, Biémont C, Vieira C.** 2006. Ongoing loss of the tirant transposable element in natural populations of Drosophila simulans. Gene 375:54–62.
8. **Galtier N, Gouy M, Gautier C.** 1996. SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comp. Appl. Biosci. 12:543–548.
9. **Heidmann O, Vernochet C, Dupressoir A, 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.
10. **Kim A, et al.** 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.
11. **Leblanc P, et al.** 2000. Life cycle of an endogenous retrovirus, ZAM, in Drosophila melanogaster. J. Virol. 74:10658–10669.
12. **Lerat E, Burlet N, Biémont C, Vieira C.** 2011. Comparative analysis of transposable elements in the melanogaster subgroup sequenced genomes. Gene 473:100–109.
13. **Malik HS, Henikoff S, Eickbush TH.** 2000. Poised for contagion: evolutionary origins of the infectious abilities of invertebrate retroviruses. Genome Res. 10:1307–1318.
14. **Mallet F, et al.** 2004. The endogenous retroviral locus ERVWE1 is a bona fide gene involved in hominoid placental physiology. Proc. Natl. Acad. Sci. U. S. A. 101:1731–1736.
15. **Marsano RM, et al.** 2000. The complete tirant transposable element in Drosophila melanogaster shows a structural relationship with retrovirus-like retrotransposons. Gene 247:87–95.
16. **Misseri Y, Labesse G, Bucheton A, Terzian C.** 2003. Comparative sequence analysis and predictions for the envelope glycoproteins of insect endogenous retroviruses. Trends Microbiol. 11:253–256.
17. **Moltó MD, Paricio N, López-Preciado MA, Semeshin VF, Martínez-Sebastián MJ.** 1996. Tirant: a new retrotransposon-like element in Drosophila melanogaster. J. Mol. Evol. 42:369–375.
18. **Pélisson A, et al.** 1994. Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the Drosophila flamenco gene. EMBO J. 13:4401–4411.
19. **Pelisson A, Mejlumian L, Robert V, Terzian C, Bucheton A.** 2002. Drosophila germline invasion by the endogenous retrovirus gypsy: involvement of the viral env gene. Insect Biochem. Mol. Biol. 32:1249–1256.
20. **Rohrmann GF, Karplus PA.** 2001. Relatedness of baculovirus and gypsy retrotransposon envelope proteins. BMC Evol. Biol. 1:1.
21. **Rozas J, Rozas R.** 1995. DnaSP, DNA sequence polymorphism: an interactive program for estimating population genetics parameters from DNA sequence data. Comput. Appl. Biosci. 11:621–625.
22. **Tanda S, Mullor JL, Corces VG.** 1994. The Drosophila tom retrotransposon encodes an envelope protein. Mol. Cell. Biol. 14:5392–5401.
23. **Tcheressiz S, et al.** 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.
24. **Terzian C, Pélisson A, Bucheton A.** 2001. Evolution and phylogeny of insect endogenous retroviruses. BMC Evol. Biol. 1:3.
25. **Teysset L, et al.** 1998. A Moloney murine leukemia virus-based retroviral vector pseudotyped by the insect retroviral gypsy envelope can infect Drosophila cells. J. Virol. 72:853–856.
26. **Vieira C, Fablet M, Lerat E.** 2009. Infra- and transspecific clues to understanding the dynamics of transposable elements, p 115–123. *In* Transposons and the dynamic genome, genome dynamics and stability. Springer-Verlag, Berlin, Germany.
27. **Viggiano L, Caggese C, Barsanti P, Caizzi R.** 1997. Cloning and characterization of a copy of tirant transposable element in Drosophila melanogaster. Gene 197:29–35.

PROOF

Bibliographie

Ahlquist P (2006) Parallels among positive-strand RNA viruses, reverse-transcribing viruses and double-stranded RNA viruses. *Nature Reviews Microbiology* 4: 371-382

Akkouche A, Rebollo R, Burlet N, Esnault C, Martinez S, Viginier B, Terzian C, Vieira C, Fablet M (2012) Tirant, a newly discovered active endogenous retrovirus in Drosophila simulans. *Journal of Virology*

Alonso-González L, Domínguez A, Albornoz J (2003) Structural heterogeneity and genomic distribution of Drosophila melanogaster LTR-retrotransposons. *Molecular Biology and Evolution* 20: 401-409

Ambros V (2003) MicroRNA pathways in flies and worms: growth, death, fat, stress, and timing. *Cell* 113: 673-676

Ambros V (2004) The functions of animal microRNAs. *Nature* 431: 350-355

Aravin AA, Bourc'his D (2008) Small RNA guides for de novo DNA methylation in mammalian germ cells. *Genes & Development* 22: 970-975

Aravin AA, Hannon GJ (2008) Small RNA silencing pathways in germ and stem cells. *Cold Spring Harbor Symposia on Quantitative Biology* 73: 283-290

Aravin AA, Klenov MS, Vagin VV, Bantignies F, Cavalli G, Gvozdev VA (2004) Dissection of a natural RNA silencing process in the Drosophila melanogaster germ line. *Molecular and Cellular Biology* 24: 6742-6750

Arkov AL, Wang J-YS, Ramos A, Lehmann R (2006) The role of Tudor domains in germline development and polar granule architecture. *Development (Cambridge, England)* 133: 4053-4062

Bailey JA, Liu G, Eichler EE (2003) An Alu transposition model for the origin and expansion of human segmental duplications. *American Journal of Human Genetics* 73: 823-834

Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas JO, Allshire RC, Kouzarides T (2001) Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* 410: 120-124

Barski A, Cuddapah S, Cui K, Roh T-Y, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K (2007) High-resolution profiling of histone methylations in the human genome. *Cell* 129: 823-837

Belshaw R, Pereira V, Katzourakis A, Talbot G, Paces J, Burt A, Tristem M (2004) Long-term reinfection of the human genome by endogenous retroviruses. *Proceedings of the National Academy of Sciences of the United States of America* 101: 4894-4899

Bergman CM, Quesneville H, Anxolabéhère D, Ashburner M (2006) Recurrent insertion and duplication generate networks of transposable element sequences in the Drosophila melanogaster genome. *Genome Biology* 7: R112-R112

Biemont C (2010) A Brief History of the Status of Transposable Elements: From Junk DNA to Major Players in Evolution. *Genetics* 186: 1085-1093

Biémont C, Vieira C (2004) [The influence of transposable elements on genome size]. *Journal De La Société De Biologie* 198: 413-417

Biemont C, Vieira C (2005) What transposable elements tell us about genome organization and evolution: the case of Drosophila. *Cytogenetic and Genome Research* 110: 25-34

Biemont C, Vieira C (2006) Genetics - Junk DNA as an evolutionary force. *Nature* 443: 521-524

Blaise S, de Parseval N, Bénit L, Heidmann T (2003) Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. *Proceedings of the National Academy of Sciences of the United States of America* 100: 13013-13018

Blumenstiel JP, Hartl DL (2005) Evidence for maternally transmitted small interfering RNA in the repression of transposition in Drosophila virilis. *Proceedings of the National Academy of Sciences of the United States of America* 102: 15965-15970

Bohne A, Brunet F, Galiana-Arnoux D, Schultheis C, Volff JN (2008) Transposable elements as drivers of genomic and biological diversity in vertebrates. *Chromosome Research* 16: 203-215

Borie N, Maisonhaute C, Sarrazin S, Loevenbruck C, Biémont C (2002) Tissue-specificity of 412 retrotransposon expression in Drosophila simulans and D. melanogaster. *Heredity* 89: 247-252

Boswell RE, Mahowald AP (1985) tudor, a gene required for assembly of the germ plasm in Drosophila melanogaster. *Cell* 43: 97-104

Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ (2007) Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* 128: 1089-1103

Brennecke J, Malone CD, Aravin AA, Sachidanandam R, Stark A, Hannon GJ (2008) An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science (New York, NY)* 322: 1387-1392

Brenner C, Fuks F (2006) DNA methyltransferases: facts, clues, mysteries. *Current Topics in Microbiology and Immunology* 301: 45-66

Britten RJ, Davidson EH (1969) Gene regulation for higher cells: a theory. *Science (New York, NY)* 165: 349-357

Bucheton A (1995) The relationship between the flamenco gene and gypsy in Drosophila: how to tame a retrovirus. *Trends in Genetics: TIG* 11: 349-353

Bucheton A, Lavige JM, Picard G, L'Heritier P (1976) Non-mendelian female sterility in Drosophila melanogaster: quantitative variations in the efficiency of inducer and reactive strains. *Heredity* 36: 305-314

Bucheton A, Paro R, Sang HM, Pelisson A, Finnegan DJ (1984) The molecular basis of I-R hybrid dysgenesis in Drosophila melanogaster: identification, cloning, and properties of the I factor. *Cell* 38: 153-163

Bundock P, Hooykaas P (2005) An Arabidopsis hAT-like transposase is essential for plant development. *Nature* 436: 282-284

Calvi BR, Hong TJ, Findley SD, Gelbart WM (1991) Evidence for a common evolutionary origin of inverted repeat transposons in Drosophila and plants: hobo, Activator, and Tam3. *Cell* 66: 465-471

Cam HP, Sugiyama T, Chen ES, Chen X, FitzGerald PC, Grewal SIS (2005) Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome. *Nature Genetics* 37: 809-819

Carmell MA, Xuan Z, Zhang MQ, Hannon GJ (2002) The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes & Development* 16: 2733-2742

Chaboissier MC, Busseau I, Prosser J, Finnegan DJ, Bucheton A (1990) Identification of a potential RNA intermediate for transposition of the LINE-like element I factor in Drosophila melanogaster. *The EMBO Journal* 9: 3557-3563

Chambeyron S, Popkova A, Payen-Groschêne G, Brun C, Laouini D, Pelisson A, Bucheton A (2008) piRNA-mediated nuclear accumulation of retrotransposon transcripts in the Drosophila female germline. *Proceedings of the National Academy of Sciences of the United States of America* 105: 14964-14969

Chan SWL, Henderson IR, Jacobsen SE (2005) Gardening the genome: DNA methylation in Arabidopsis thaliana. *Nature Reviews Genetics* 6: 351-360

Chen Y, Pane A, Schüpbach T (2007) Cutoff and aubergine mutations result in retrotransposon upregulation and checkpoint activation in Drosophila. *Current Biology: CB* 17: 637-642

Cizeron G, Biémont C (1999) Polymorphism in structure of the retrotransposable element 412 in Drosophila simulans and D. melanogaster populations. *Gene* 232: 183-190

Clegg MT, Durbin ML (2003) Tracing floral adaptations from ecology to molecules. *Nature Reviews Genetics* 4: 206-215

Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* 452: 215-219

Cox DN, Chao A, Baker J, Chang L, Qiao D, Lin H (1998) A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. *Genes & Development* 12: 3715-3727

Cox DN, Chao A, Lin H (2000) piwi encodes a nucleoplasmic factor whose activity modulates the number and division rate of germline stem cells. *Development (Cambridge, England)* 127: 503-514

Crozatier M, Vaury C, Busseau I, Pelisson A, Bucheton A (1988) Structure and genomic organization of I elements involved in I-R hybrid dysgenesis in Drosophila melanogaster. *Nucleic Acids Research* 16: 9199-9213

Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, Dus M, Perrimon N, Kellis M, Wohlschlegel JA, Sachidanandam R, Hannon GJ, Brennecke J (2008) An endogenous small interfering RNA pathway in Drosophila. *Nature* 453: 798-802

Desset S, Buchon N, Meignin C, Coiffet M, 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: e1526-e1526

Desset S, Conte C, Dimitri P, Calco V, Dastugue B, Vaury C (1999) Mobilization of two retroelements, ZAM and Idefix, in a novel unstable line of Drosophila melanogaster. *Molecular Biology and Evolution* 16: 54-66

Doolittle WF, Sapienza C (1980) Selfish genes, the phenotype paradigm and genome evolution. *Nature* 284: 601-603

Dowsett AP, Young MW (1982) Differing levels of dispersed repetitive DNA among closely related species of Drosophila. *Proceedings of the National Academy of Sciences of the United States of America* 79: 4570-4574

Dufourt J, Brasset E, Desset S, Pouchin P, Vaury C (2011) Polycomb group-dependent, heterochromatin protein 1-independent, chromatin structures silence retrotransposons in somatic tissues outside ovaries. *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes* 18: 451-461

Echalier G (1989) Drosophila retrotransposons: interactions with genome. *Advances in Virus Research* 36: 33-105

Emmons SW, Yesner L, Ruan KS, Katzenberg D (1983) Evidence for a transposon in Caenorhabditis elegans. *Cell* 32: 55-65

Engels WR (1989) P Elements in Drosophila. *Genetics*

Evgen'ev MB, Zelentsova H, Poluectova H, Lyozin GT, Veleikodvorskaja V, Pyatkov KI, Zhivotovsky La, Kidwell MG (2000) Mobile elements and chromosomal evolution in the virilis group of Drosophila. *Proceedings of the National Academy of Sciences of the United States of America* 97: 11337-11342

Fablet M, Lerat E, Rebollo R, Horard B, Burlet N, Martinez S, Brasset E, Gilson E, Vaury C, Vieira C (2009) Genomic environment influences the dynamics of the tirant LTR retrotransposon in Drosophila. *The FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology* 23: 1482-1489

Fablet M, McDonald JF, Biémont C, Vieira C (2006) Ongoing loss of the tirant transposable element in natural populations of Drosophila simulans. *Gene* 375: 54-62

Fablet and Vieira (2011) Evolvability, epigenetics and transposable elements. BioMol Concepts, Vol. 2 (2011), 333–341

Findley SD, Tamanaha M, Clegg NJ, Ruohola-Baker H (2003) Maelstrom, a Drosophila spindle-class gene, encodes a protein that colocalizes with Vasa and RDE1/AGO1 homolog, Aubergine, in nuage. *Development (Cambridge, England)* 130: 859-871

Förstemann K, Tomari Y, Du T, Vagin VV, Denli AM, Bratu DP, Klattenhoff C, Theurkauf WE, Zamore PD (2005) Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. *PLoS Biology* 3: e236-e236

Galiana-Arnoux D, Dostert C, Schneemann A, Hoffmann JA, Imler J-L (2006) Essential function in vivo for Dicer-2 in host defense against RNA viruses in drosophila. *Nature Immunology* 7: 590-597

Garcia RN, D'Avila MF, Robe LJ, Loreto ELdS, Panzera Y, de Heredia FO, Valente VLdS (2007) First evidence of methylation in the genome of Drosophila willistoni. *Genetica* 131: 91-105

García-Cao M, O'Sullivan R, Peters AHFM, Jenuwein T, Blasco MA (2004) Epigenetic regulation of telomere length in mammalian cells by the Suv39h1 and Suv39h2 histone methyltransferases. *Nature Genetics* 36: 94-99

Garoff H, Hewson R, Opstelten DJ (1998) Virus maturation by budding. *Microbiology and Molecular Biology Reviews: MMBR* 62: 1171-1190

Gellert M (1996) A new view of V(D)J recombination. *Genes to Cells: Devoted to Molecular & Cellular Mechanisms* 1: 269-275

Ghabrial A, Schüpbach T (1999) Activation of a meiotic checkpoint regulates translation of Gurken during Drosophila oogenesis. *Nature Cell Biology* 1: 354-357

Ghildiyal M, Seitz H, Horwich MD, Li C, Du T, Lee S, Xu J, Kittler ELW, Zapp ML, Weng Z, Zamore PD (2008) Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. *Science (New York, NY)* 320: 1077-1081

Ghildiyal M, Zamore PD (2009) Small silencing RNAs: an expanding universe. *Nature Reviews Genetics* 10: 94-108

Girard A, Sachidanandam R, Hannon GJ, Carmell MA (2006) A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* 442: 199-202

Girton JR, Johansen KM (2008) Chromatin structure and the regulation of gene expression: the lessons of PEV in Drosophila. *Advances in Genetics* 61: 1-43

Goldberg AD, Allis CD, Bernstein E (2007) Epigenetics: a landscape takes shape. *Cell* 128: 635-638

Grewal SIS, Elgin SCR (2007) Transcription and RNA interference in the formation of heterochromatin. *Nature* 447: 399-406

Grunstein M (1997) Histone acetylation in chromatin structure and transcription. *Nature* 389: 349-352

Gunawardane LS, Saito K, Nishida KM, Miyoshi K, Kawamura Y, Nagami T, Siomi H, Siomi MC (2007) A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in Drosophila. *Science (New York, NY)* 315: 1587-1590

Haase AD, Fenoglio S, Muerdter F, Guzzardo PM, Czech B, Pappin DJ, Chen C, Gordon A, Hannon GJ (2010) Probing the initiation and effector phases of the somatic piRNA pathway in Drosophila. *Genes & Development* 24: 2499-2504

Hamilton A, Voinnet O, Chappell L, Baulcombe D (2002) Two classes of short interfering RNA in RNA silencing. *The EMBO Journal* 21: 4671-4679

Haynes KA, Caudy AA, Collins L, Elgin SCR (2006) Element 1360 and RNAi components contribute to HP1-dependent silencing of a pericentric reporter. *Current Biology: CB* 16: 2222-2227

Henikoff S (1990) Position-effect variegation after 60 years. *Trends in Genetics: TIG* 6: 422-426

Hill AS, Foot NJ, Chaplin TL, Young BD (2000) The most frequent constitutional translocation in humans, the t(11;22)(q23;q11) is due to a highly specific alu-mediated recombination. *Human Molecular Genetics* 9: 1525-1532

Hiom K, Melek M, Gellert M (1998) DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. *Cell* 94: 463-470

Horwich MD, Li C, Matranga C, Vagin V, Farley G, Wang P, Zamore PD (2007) The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Current Biology: CB* 17: 1265-1272

Houwing S, Kamminga LM, Berezikov E, Cronembold D, Girard A, van den Elst H, Filippov DV, Blaser H, Raz E, Moens CB, Plasterk RHA, Hannon GJ, Draper BW, Ketting RF (2007) A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell* 129: 69-82

Huda A, Bowen NJ, Conley AB, Jordan IK (2011) Epigenetic regulation of transposable element derived human gene promoters. *Gene* 475: 39-48

Hull R (2001) Classifying reverse transcribing elements: a proposal and a challenge to the ICTV. International Committee on Taxonomy of Viruses. *Archives of Virology* 146: 2255-2261

Ilyin YV, Chmeliauskaite VG, Ananiev EV, Georgiev GP (1980) Isolation and characterization of a new family of mobile dispersed genetic elements, mdg3, in Drosophila melanogaster. *Chromosoma* 81: 27-53

Jiang Z-f, Croshaw DA, Wang Y, Hey J, Machado CA (2011) Enrichment of mRNA-like noncoding RNAs in the divergence of Drosophila males. *Molecular biology and evolution* 28: 1339-1348

Jordan IK, Rogozin IB, Glazko GV, Koonin EV (2003) Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends in Genetics* 19: 68-72

Kapitonov VV, Jurka J (2003) Molecular paleontology of transposable elements in the Drosophila melanogaster genome. *Proceedings of the National Academy of Sciences of the United States of America* 100: 6569-6574

Kashkush K, Feldman M, Levy AA (2003) Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nature Genetics* 33: 102-106

Kawamura Y, Saito K, Kin T, Ono Y, Asai K, Sunohara T, Okada TN, Siomi MC, Siomi H (2008) Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* 453: 793-797

Kidwell MG, Kidwell JF, Sved JA (1977) HYBRID DYSGENESIS IN DROSOPHILA-MELANOGASTER - SYNDROME OF ABERRANT TRAITS INCLUDING MUTATION, STERILITY AND MALE RECOMBINATION. *Genetics* 86: 813-833

Kim AI, Belyaeva ES, Aslanian MM (1990) Autonomous transposition of gypsy mobile elements and genetic instability in Drosophila melanogaster. *Molecular & General Genetics: MGG* 224: 303-308

Kim Y-K, Heo I, Kim VN (2010) Modifications of small RNAs and their associated proteins. *Cell* 143: 703-709

King RC (1970) The meiotic behavior of the Drosophila oocyte. *International Review of Cytology* 28: 125-168

Kirino Y, Kim N, de Planell-Saguer M, Khandros E, Chiorean S, Klein PS, Rigoutsos I, Jongens TA, Mourelatos Z (2009) Arginine methylation of Piwi proteins catalysed by dPRMT5 is required for Ago3 and Aub stability. *Nature Cell Biology* 11: 652-658

Kirino Y, Mourelatos Z (2007) Mouse Piwi-interacting RNAs are 2'-O-methylated at their 3' termini. *Nature Structural & Molecular Biology* 14: 347-348

Klose RJ, Bird AP (2006) Genomic DNA methylation: the mark and its mediators. *Trends in Biochemical Sciences* 31: 89-97

Klenov, Sergey A. Lavrov, Anastasia D. Stolyarenko, Sergey S. Ryazansky, Alexei A. Aravin, Thomas Tuschl,and Vladimir A. Gvozdev (2007) Repeat-associated siRNAs cause chromatin silencing of retrotransposons in the *Drosophila melanogaster* germline. ucleic Acids Res.; 35(16): 5430–5438.

Kondo Y, Issa J-PJ (2003) Enrichment for histone H3 lysine 9 methylation at Alu repeats in human cells. *The Journal of Biological Chemistry* 278: 27658-27662

Kornberg RD (1974) Chromatin structure: a repeating unit of histones and DNA. *Science (New York, NY)* 184: 868-871

Kramerov DA, Vassetzky NS (2005) Short retroposons in eukaryotic genomes. *International Review of Cytology* 247: 165-221

Kunert N, Marhold J, Stanke J, Stach D, Lyko F (2003) A Dnmt2-like protein mediates DNA methylation in Drosophila. *Development (Cambridge, England)* 130: 5083-5090

Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Totoki Y, Toyoda A, Ikawa M, Asada N, Kojima K, Yamaguchi Y, Ijiri TW, Hata K, Li E, Matsuda Y, Kimura T, Okabe M, Sakaki Y, Sasaki H, Nakano T (2008) DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes & Development* 22: 908-917

Labrador M, Farré M, Utzet F, Fontdevila a (1999) Interspecific hybridization increases transposition rates of Osvaldo. *Molecular biology and evolution* 16: 931-937

Lachaise D, Silvain J-F (2004) How two Afrotropical endemics made two cosmopolitan human commensals: the Drosophila melanogaster-D. simulans palaeogeographic riddle. *Genetica* 120: 17-39

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, ., et al. (2001) Initial sequencing and analysis of the human genome. *Nature* 409: 860-921

Lau NC, Seto AG, Kim J, Kuramochi-Miyagawa S, Nakano T, Bartel DP, Kingston RE (2006) Characterization of the piRNA complex from rat testes. *Science (New York, NY)* 313: 363-367

Leblanc P, Desset S, Dastugue B, Vaury C (1997) Invertebrate retroviruses: ZAM a new candidate in D.melanogaster. *The EMBO Journal* 16: 7521-7531

Leblanc P, Desset S, Giorgi F, Taddei AR, Fausto AM, Mazzini M, Dastugue B, Vaury C (2000) Life cycle of an endogenous retrovirus, ZAM, in Drosophila melanogaster. *Journal of Virology* 74: 10658-10669

Lerat E, Burlet N, Biémont C, Vieira C (2011) Comparative analysis of transposable elements in the melanogaster subgroup sequenced genomes. *Gene* 473: 100-109

Lerat E, Rizzon C, Biemont C (2003) Sequence divergence within transposable element families in the Drosophila melanogaster genome. *Genome Research* 13: 1889-1896

Lécher P, Bucheton A, Pélisson A (1997) Expression of the Drosophila retrovirus gypsy as ultrastructurally detectable particles in the ovaries of flies carrying a permissive flamenco allele. *The Journal of General Virology* 78 (Pt 9): 2379-2388

Li XC, Barringer BC, Barbash DA (2009) The pachytene checkpoint and its relationship to evolutionary patterns of polyploidization and hybrid sterility. *Heredity* 102: 24-30

Lim AK, Kai T (2007) Unique germ-line organelle, nuage, functions to repress selfish genetic elements in Drosophila melanogaster. *Proceedings of the National Academy of Sciences of the United States of America* 104: 6714-6719

Lin M-J, Tang L-Y, Reddy MN, Shen CKJ (2005) DNA methyltransferase gene dDnmt2 and longevity of Drosophila. *The Journal of Biological Chemistry* 280: 861-864

Lippman Z, Martienssen R (2004) The role of RNA interference in heterochromatic silencing. *Nature* 431: 364-370

Liu H, Wang J-YS, Huang Y, Li Z, Gong W, Lehmann R, Xu R-M (2010) Structural basis for methylarginine-dependent recognition of Aubergine by Tudor. *Genes & Development* 24: 1876-1881

Llave C, Xie Z, Kasschau KD, Carrington JC (2002) Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. *Science (New York, NY)* 297: 2053-2056

Lyko F, Ramsahoye BH, Jaenisch R (2000) DNA methylation in Drosophila melanogaster. *Nature* 408: 538-540

Lynch C, Tristem M (2003) A co-opted gypsy-type LTR-retrotransposon is conserved in the genomes of humans, sheep, mice, and rats. *Current Biology: CB* 13: 1518-1523

Malik HS, Eickbush TH (2001) Phylogenetic analysis of ribonuclease H domains suggests a late, chimeric origin of LTR retrotransposable elements and retroviruses. *Genome Research* 11: 1187-1197

Mallet F, Bouton O, Prudhomme S, Cheynet V, Oriol G, Bonnaud B, Lucotte G, Duret L, Mandrand B (2004) The endogenous retroviral locus ERVWE1 is a bona fide gene involved in hominoid placental physiology. *Proceedings of the National Academy of Sciences of the United States of America* 101: 1731-1736

Malone CD, Brennecke J, Dus M, Stark A, McCombie WR, Sachidanandam R, Hannon GJ (2009) Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. *Cell* 137: 522-535

Margueron R, Trojer P, Reinberg D (2005) The key to development: interpreting the histone code? *Current Opinion in Genetics & Development* 15: 163-176

Marsano RM, Moschetti R, Caggese C, Lanave C, Barsanti P, Caizzi R (2000) The complete Tirant transposable element in Drosophila melanogaster shows a structural relationship with retrovirus-like retrotransposons. *Gene* 247: 87-95

Martens JHA, O'Sullivan RJ, Braunschweig U, Opravil S, Radolf M, Steinlein P, Jenuwein T (2005) The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *The EMBO Journal* 24: 800-812

Maside X, Assimacopoulos S, Charlesworth B (2005) Fixation of transposable elements in the Drosophila melanogaster genome. *Genetical Research* 85: 195-203

Matzke M, Kanno T, Huettel B, Daxinger L, Matzke AJM (2007) Targets of RNA-directed DNA methylation. *Current Opinion in Plant Biology* 10: 512-519

McClintock B (1950) The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Sciences of the United States of America* 36: 344-355

McClintock B (1956) Controlling elements and the gene. *Cold Spring Harbor Symposia on Quantitative Biology* 21: 197-216

Mi S, Lee X, Li X, Veldman GM, Finnerty H, Racie L, LaVallie E, Tang XY, Edouard P, Howes S, Keith JC, Jr., McCoy JM (2000) Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* 403: 785-789

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim T-K, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, Russ C, Xie X, Meissner A, Wernig M, Jaenisch R, Nusbaum C, Lander ES, Bernstein BE (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448: 553-560

Muotri AR, Chu VT, Marchetto MCN, Deng W, Moran JV, Gage FH (2005) Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* 435: 903-910

Napoli C, Lemieux C, Jorgensen R (1990) Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. *The Plant Cell* 2: 279-289

Nishida KM, Okada TN, Kawamura T, Mituyama T, Kawamura Y, Inagaki S, Huang H, Chen D, Kodama T, Siomi H, Siomi MC (2009) Functional involvement of Tudor and dPRMT5 in the piRNA processing pathway in Drosophila germlines. *The EMBO Journal* 28: 3820-3831

Nishida KM, Saito K, Mori T, Kawamura Y, Nagami-Okada T, Inagaki S, Siomi H, Siomi MC (2007) Gene silencing mechanisms mediated by Aubergine piRNA complexes in Drosophila male gonad. *RNA (New York, NY)* 13: 1911-1922

Noma K-i, Sugiyama T, Cam H, Verdel A, Zofall M, Jia S, Moazed D, Grewal SIS (2004) RITS acts in cis to promote RNA interference-mediated transcriptional and post-transcriptional silencing. *Nature Genetics* 36: 1174-1180

Ohara T, Sakaguchi Y, Suzuki T, Ueda H, Miyauchi K, Suzuki T (2007) The 3' termini of mouse Piwi-interacting RNAs are 2'-O-methylated. *Nature Structural & Molecular Biology* 14: 349-350

Oliver KR, Greene WK (2009) Transposable elements: powerful facilitators of evolution. *Bioessays* 31: 703-714

Olivieri D, Sykora MM, Sachidanandam R, Mechtler K, Brennecke J (2010) An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in Drosophila. *The EMBO Journal* 29: 3301-3317

Orgel LE, Crick FH (1980) Selfish DNA: the ultimate parasite. *Nature* 284: 604-607

Palliser D, Chowdhury D, Wang Q-Y, Lee SJ, Bronson RT, Knipe DM, Lieberman J (2006) An siRNA-based microbicide protects mice from lethal herpes simplex virus 2 infection. *Nature* 439: 89-94

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

Pardue M-L, DeBaryshe PG (2003) Retrotransposons provide an evolutionarily robust non-telomerase mechanism to maintain telomeres. *Annual Review of Genetics* 37: 485-511

Pauler FM, Sloane MA, Huang R, Regha K, Koerner MV, Tamir I, Sommer A, Aszodi A, Jenuwein T, Barlow DP (2009) H3K27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone banding pattern on a mouse autosomal chromosome. *Genome Research* 19: 221-233

Pearson MN, Russell RLQ, Rohrmann GF (2002) Functional analysis of a conserved region of the baculovirus envelope fusion protein, LD130. *Virology* 304: 81-88

Peaston AE, Evsikov AV, Graber JH, de Vries WN, Holbrook AE, Solter D, Knowles BB (2004) Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Developmental Cell* 7: 597-606

Pelisson A, Mejlumian L, Robert V, Terzian C, Bucheton A (2002) Drosophila germline invasion by the endogenous retrovirus gypsy: involvement of the viral env gene. *Insect Biochemistry and Molecular Biology* 32: 1249-1256

Pélisson A, Sarot E, Payen-Groschêne G, Bucheton A (2007) A novel repeat-associated small interfering RNA-mediated silencing pathway downregulates complementary sense gypsy transcripts in somatic cells of the Drosophila ovary. *Journal of Virology* 81: 1951-1960

Pélisson A, Song SU, Prud'homme N, Smith PA, Bucheton A, Corces VG (1994) Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the Drosophila flamenco gene. *The EMBO Journal* 13: 4401-4411

Peng JC, Karpen GH (2007) H3K9 methylation and RNA interference regulate nucleolar organization and repeated DNA stability. *Nature Cell Biology* 9: 25-35

Petillo D, Kort EJ, Anema J, Furge KA, Yang XJ, Teh BT (2009) MicroRNA profiling of human kidney cancer subtypes. *International Journal of Oncology* 35: 109-114

Petrov DA (1995) Diverse Transposable Elements are Mobilized in Hybrid Dysgenesis in Drosophila virilis. *Proceedings of the National Academy of Sciences* 92: 8050-8054

Phalke S, Nickel O, Walluscheck D, Hortig F, Onorati MC, Reuter G (2009) Retrotransposon silencing and telomere integrity in somatic cells of Drosophila depends on the cytosine-5 methyltransferase DNMT2. *Nature Genetics* 41: 696-702

Picard G (1976) Non-mendelian female sterility in Drosophila melanogaster: hereditary transmission of I factor. *Genetics* 83: 107-123

Picard G, Bregliano JC, Bucheton A, Lavige JM, Pelisson A, Kidwell MG (1978) Non-mendelian female sterility and hybrid dysgenesis in Drosophila melanogaster. *Genetical Research* 32: 275-287

Pietersen AM, van Lohuizen M (2008) Stem cell regulation by polycomb repressors: postponing commitment. *Current Opinion in Cell Biology* 20: 201-207

Piriyapongsa J, Jordan IK (2007) A family of human microRNA genes from miniature inverted-repeat transposable elements. *PloS One* 2: e203-e203

Pokholok DK, Harbison CT, Levine S, Cole M, Hannett NM, Lee TI, Bell GW, Walker K, Rolfe PA, Herbolsheimer E, Zeitlinger J, Lewitter F, Gifford DK, Young RA (2005) Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* 122: 517-527

Prud'homme N, Gans M, Masson M, Terzian C, Bucheton A (1995) Flamenco, a gene controlling the gypsy retrovirus of Drosophila melanogaster. *Genetics* 139: 697-711

Rangan P, Malone CD, Navarro C, Newbold SP, Hayes PS, Sachidanandam R, Hannon GJ, Lehmann R (2011) piRNA production requires heterochromatin formation in Drosophila. *Current Biology: CB* 21: 1373-1379

Rehwinkel J, Natalin P, Stark A, Brennecke J, Cohen SM, Izaurralde E (2006) Genome-wide analysis of mRNAs regulated by Drosha and Argonaute proteins in Drosophila melanogaster. *Molecular and Cellular Biology* 26: 2965-2975

Riddle NC, Minoda A, Kharchenko PV, Alekseyenko AA, Schwartz YB, Tolstorukov MY, Gorchakov AA, Jaffe JD, Kennedy C, Linder-Basso D, Peach SE, Shanower G, Zheng H, Kuroda MI, Pirrotta V, Park PJ, Elgin SCR, Karpen GH (2011) Plasticity in patterns of histone modifications and chromosomal proteins in Drosophila heterochromatin. *Genome Research* 21: 147-163

Robertson KD, Wolffe AP (2000) DNA methylation in health and disease. *Nature Reviews Genetics* 1: 11-19

Robine N, Lau NC, Balla S, Jin Z, Okamura K, Kuramochi-Miyagawa S, Blower MD, Lai EC (2009) A broadly conserved pathway generates 3'UTR-directed primary piRNAs. *Current Biology: CB* 19: 2066-2076

Rongo C, Lehmann R (1996) Regulated synthesis, transport and assembly of the Drosophila germ plasm. *Trends in Genetics: TIG* 12: 102-109

Rouget C, Papin C, Boureux A, Meunier A-C, Franco B, Robine N, Lai EC, Pelisson A, Simonelig M (2010) Maternal mRNA deadenylation and decay by the piRNA pathway in the early Drosophila embryo. *Nature* 467: 1128-1132

Rozhkov NV, Aravin AA, Zelentsova ES, Schostak NG, Sachidanandam R, McCombie WR, Hannon GJ, Evgen'ev MB.(2010) Small RNA-based silencing strategies for transposons in the process of invading Drosophila species. **RNA**. 2010 Aug;16(8):1634-45.

Saito K, Ishizu H, Komai M, Kotani H, Kawamura Y, Nishida KM, Siomi H, Siomi MC (2010) Roles for the Yb body components Armitage and Yb in primary piRNA biogenesis in Drosophila. *Genes & Development* 24: 2493-2498

Saito K, Ishizuka A, Siomi H, Siomi MC (2005) Processing of pre-microRNAs by the Dicer-1-Loquacious complex in Drosophila cells. *PLoS Biology* 3: e235-e235

Saito K, Siomi MC (2010) Small RNA-mediated quiescence of transposable elements in animals. *Developmental Cell* 19: 687-697

Sakano H, Hüppi K, Heinrich G, Tonegawa S (1979) Sequences at the somatic recombination sites of immunoglobulin light-chain genes. *Nature* 280: 288-294

Santos-Rosa H, Schneider R, Bannister AJ, Sherriff J, Bernstein BE, Emre NCT, Schreiber SL, Mellor J, Kouzarides T (2002) Active genes are tri-methylated at K4 of histone H3. *Nature* 419: 407-411

Sarot E, Payen-Groschêne G, Bucheton A, Pélisson A (2004) Evidence for a piwi-dependent RNA silencing of the gypsy endogenous retrovirus by the Drosophila melanogaster flamenco gene. *Genetics* 166: 1313-1321

Sasaki T, Shiohama A, Minoshima S, Shimizu N (2003) Identification of eight members of the Argonaute family in the human genome small star, filled. *Genomics* 82: 323-330

Schaefer M, Lyko F (2010) Lack of evidence for DNA methylation of Invader4 retroelements in Drosophila and implications for Dnmt2-mediated epigenetic regulation. *Nature Genetics* 42: 920-921; author reply 921-920-921; author reply 921

Schotta G, Ebert A, Krauss V, Fischer A, Hoffmann J, Rea S, Jenuwein T, Dorn R, Reuter G (2002) Central role of Drosophila SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. *The EMBO Journal* 21: 1121-1131

Schotta G, Lachner M, Sarma K, Ebert A, Sengupta R, Reuter G, Reinberg D, Jenuwein T (2004) A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes & Development* 18: 1251-1262

Schüpbach T, Wieschaus E (1991) Female sterile mutations on the second chromosome of Drosophila melanogaster. II. Mutations blocking oogenesis or altering egg morphology. *Genetics* 129: 1119-1136

Schwartz A, Chan DC, Brown LG, Alagappan R, Pettay D, Disteche C, McGillivray B, de la Chapelle A, Page DC (1998) Reconstructing hominid Y evolution: X-homologous block, created by X-Y transposition, was disrupted by Yp inversion through LINE-LINE recombination. *Human Molecular Genetics* 7: 1-11

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

Shpiz S, Olovnikov I, Sergeeva A, Lavrov S, Abramov Y, Savitsky M, Kalmykova A (2011) Mechanism of the piRNA-mediated silencing of Drosophila telomeric retrotransposons. *Nucleic Acids Research* 39: 8703-8711

Simons C, Pheasant M, Makunin IV, Mattick JS (2006) Transposon-free regions in mammalian genomes. *Genome Research* 16: 164-172

Siomi MC, Mannen T, Siomi H (2010) How does the Royal Family of Tudor rule the PIWI-interacting RNA pathway? *Genes & Development* 24: 636-646

Smalheiser NR, Torvik VI (2005) Mammalian microRNAs derived from genomic repeats. *Trends in Genetics: TIG* 21: 322-326

Sokolova OA, Iakushev EI, Stoliarenko AD, Mikhaleva EA, Gvozdev VA, Klenov MS (2011) [The interplay of transposon silencing genes in the Drosophila melanogaster germline]. *Molekuliarnaia Biologiia* 45: 633-641

Suzuki MM, Bird A (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nature Reviews Genetics* 9: 465-476

Szyf M (2005) Therapeutic implications of DNA methylation. *Future Oncology (London, England)* 1: 125-135

Tam OH, Aravin AA, Stein P, Girard A, Murchison EP, Cheloufi S, Hodges E, Anger M, Sachidanandam R, Schultz RM, Hannon GJ (2008) Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* 453: 534-538

Tanda S, Mullor JL, Corces VG (1994) The Drosophila tom retrotransposon encodes an envelope protein. *Molecular and Cellular Biology* 14: 5392-5401

Tanda S, Shrimpton AE, Chueh LL, Itayama H, Matsubayashi H, Saigo K, Tobari YN, Langley CH (1988) Retrovirus-like features and site specific insertions of a transposable element, tom, in Drosophila ananassae. *Molecular & General Genetics: MGG* 214: 405-411

Tanentzapf G, Smith C, McGlade J, Tepass U (2000) Apical, lateral, and basal polarization cues contribute to the development of the follicular epithelium during Drosophila oogenesis. *The Journal of Cell Biology* 151: 891-904

Terzian C, Pélisson A, Bucheton A (2001) Evolution and phylogeny of insect endogenous retroviruses. *BMC Evolutionary Biology* 1: 3-3

Thomson T, Lasko P (2004) Drosophila tudor is essential for polar granule assembly and pole cell specification, but not for posterior patterning. *Genesis (New York, NY: 2000)* 40: 164-170

Thomson T, Lin H (2009) The biogenesis and function of PIWI proteins and piRNAs: progress and prospect. *Annual Review of Cell and Developmental Biology* 25: 355-376

Uberbacher EC, Bunick GJ (1985) X-ray structure of the nucleosome core particle. *Journal of Biomolecular Structure & Dynamics* 2: 1033-1055

Vagin VV, Sigova A, Li C, Seitz H, Gvozdev V, Zamore PD (2006) A distinct small RNA pathway silences selfish genetic elements in the germline. *Science (New York, NY)* 313: 320-324

Van De Bor V, Hartswood E, Jones C, Finnegan D, Davis I (2005) gurken and the I factor retrotransposon RNAs share common localization signals and machinery. *Developmental Cell* 9: 51-62

Vaury C, Bucheton A, Pelisson A (1989) The beta heterochromatic sequences flanking the I elements are themselves defective transposable elements. *Chromosoma* 98: 215-224

Verdel A, Jia S, Gerber S, Sugiyama T, Gygi S, Grewal SIS, Moazed D (2004) RNAi-mediated targeting of heterochromatin by the RITS complex. *Science (New York, NY)* 303: 672-676

Verdel A, Moazed D (2005) RNAi-directed assembly of heterochromatin in fission yeast. *FEBS Letters* 579: 5872-5878

Vieira C, Lepetit D, Dumont S, Biemont C (1999a) Wake up of transposable elements following Drosophila simulans worldwide colonization. *Molecular Biology and Evolution* 16: 1251-1255

Vieira C, Lepetit D, Dumont S, Biémont C (1999b) Wake up of transposable elements following Drosophila simulans worldwide colonization. *Molecular Biology and Evolution* 16: 1251-1255

Volpe TA, Kidner C, Hall IM, Teng G, Grewal SIS, Martienssen RA (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science (New York, NY)* 297: 1833-1837

Wang SH, Elgin SCR (2011) Drosophila Piwi functions downstream of piRNA production mediating a chromatin-based transposon silencing mechanism in female germ line. *Proceedings of the National Academy of Sciences of the United States of America* 108: 21164-21169

Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, Cui K, Roh T-Y, Peng W, Zhang MQ, Zhao K (2008) Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature Genetics* 40: 897-903

Watanabe T, Takeda A, Tsukiyama T, Mise K, Okuno T, Sasaki H, Minami N, Imai H (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 & Development* 20: 1732-1743

Xiong Y, Eickbush TH (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. *The EMBO Journal* 9: 3353-3362

Zhang W, Yang H, Kong X, Mohapatra S, San Juan-Vergara H, Hellermann G, Behera S, Singam R, Lockey RF, Mohapatra SS (2005) Inhibition of respiratory syncytial virus infection with intranasal siRNA nanoparticles targeting the viral NS1 gene. *Nature Medicine* 11: 56-62

Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SWL, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, Ecker JR (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell* 126: 1189-1201

Zhang Y, Reinberg D (2001) Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes & Development* 15: 2343-2360