



HAL
open science

Contrôle génétique par l'hôte de deux rétrovirus endogènes, ZAM et Idefix, chez *Drosophila melanogaster*.

Sophie Desset

► To cite this version:

Sophie Desset. Contrôle génétique par l'hôte de deux rétrovirus endogènes, ZAM et Idefix, chez *Drosophila melanogaster*.. Organisation et fonctions cellulaires [q-bio.SC]. Université d'Auvergne - Clermont-Ferrand I, 2001. Français. NNT: . tel-01650125

HAL Id: tel-01650125

<https://uca.hal.science/tel-01650125>

Submitted on 28 Nov 2017

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

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

***ÉCOLE DOCTORALE
DES SCIENCES DE LA VIE ET DE LA SANTÉ***

N° ordre :

Thèse

Présentée à l'Université d'Auvergne
pour l'obtention du grade de Docteur d'Université

Spécialité
Biologie Cellulaire et Moléculaire

soutenue le 26 octobre 2001

Sophie DESSET

Contrôle génétique par l'hôte de deux rétrovirus endogènes, *ZAM* et *Idefix*,
chez *Drosophila melanogaster*.

JURY

Président : M. Bernard Dastugue
Examineur : M. Yves Courtois
Rapporteurs : M. Dominique Anxolabéhère
M. Alain Pélisson

Directeur de thèse : Mme Chantal Vaury

Année 2001

N° d'ordre :

***ÉCOLE DOCTORALE
DES SCIENCES DE LA VIE ET DE LA SANTÉ***

N° ordre :

Thèse

Présentée à l'Université d'Auvergne
pour l'obtention du grade de Docteur d'Université

Spécialité
Biologie Cellulaire et Moléculaire

soutenue le 26 octobre 2001

Sophie DESSET

Contrôle génétique par l'hôte de deux rétrovirus endogènes, *ZAM* et *Idefix*,
chez *Drosophila melanogaster*.

JURY

Président : M. Bernard Dastugue
Examineur : M. Yves Courtois
Rapporteurs : M. Dominique Anxolabéhère
M. Alain Péliçon

Directeur de thèse : Mme Chantal Vaury

Laboratoire Inserm UMR 384 – Interactions génétiques et cellulaires au cours de la différenciation.

REMERCIEMENTS

Yves Courtois et Bernard Dastugue m'ont accueillie dans leur laboratoire respectif en m'accordant la possibilité d'y préparer un doctorat. Je les remercie de la confiance qu'ils m'ont ainsi témoignée. Je suis particulièrement touchée de la présence d'Yves Courtois dans le jury, qui me permet d'associer l'ancienne unité 118 à mon travail ici et de conclure heureusement une histoire inachevée.

Merci à Dominique Anxolabéhère et Alain Pélisson, rapporteurs de ma thèse, de m'offrir le plus précieux cadeau : un peu de leur temps.

Maryvonne Laurent et Chantal Vaury ont été mes directeurs de thèse successifs. J'ai apprécié chez l'une et l'autre compétence, rigueur et qualités d'encadrement. Je garde une grande affection pour Maryvonne. Je voudrais rendre hommage à Chantal en une anecdote : un jour, un petit mâle aux yeux rouges que personne n'attendait s'est distingué sous sa loupe parmi tant d'autres aux yeux bruns, et elle a pris le temps et le risque de le garder, de le croiser, de l'étudier jusqu'à modifier son sujet de recherche. Dix ans plus tard, cinq thèses (soutenues ou en préparations), six publications, cinq articles soumis. Combien de généticiens auraient pourtant jeté ce mâle fondateur ?

Ma thèse est le résultat d'un travail d'équipe et je remercie chacun de sa participation, particulièrement Nathalie Gautier et Delphine Roudeix, pour leur aide efficace et enthousiaste. Je profite de ce que mes coéquipiers soient cités dans le manuscrit pour mentionner les participants habituellement anonymes: Fouzia Akhmassi, Virginie Cinié, Marie-Diane de la Morinerie, Agostinha de Souza, Catherine Duval, Françoise Pélissier et Christian Plévert,

Je remercie pour leur aide dans l'élaboration du manuscrit : Jean-Louis Couderc qui a toujours un article pour répondre à mes questions ; Carine Meignin, Olivier Bardot et Yannick Bidet pour leur indéfectible disponibilité (Ah ! Les pannes informatiques du vendredi soir !) ; Corinne Lours (un air de zen) pour sa relecture ; Geneviève Giraud pour son abondante bibliographie ; Adrien Brunet pour les références bibliographiques ; Valérie Calco, Caroline Conte, Corinne Lours (encore !) et Muriel Grammont pour leur soutien, et bien sûr, le très sympathique groupe drosos.

J'ai été moins disponible au labo depuis le début de la rédaction. Chacun l'a bien compris et a su respecter ma coquille temporaire. Peut-être à cause de mon statut particulier, j'ai l'impression que chacun a porté " le bébé " avec moi et que c'est une œuvre collective bien au-delà de l'équipe. Merci à tous de l'amitié dont vous m'honorez.

Je n'ai jamais compris pourquoi les candidats remercient leur famille, jusqu'à ce que je mette à contribution la mienne: merci à mes parents et nou-nous intérimaires préférées ; merci et pardon aux amis placés entre parenthèses et surtout à mes petits loups pour ces deux étés en pointillé; merci et bravo à François, pour son dévouement quotidien en deux points : patience et compréhension.

Mais avant tout, l'essentiel de ma reconnaissance est pour un petit mâle aux yeux rouges et l'ensemble de sa descendance. Bzz !Bzz !

SOMMAIRE

Rapport bibliographique.....	4
Introduction.....	5
I- Classification et nomenclature.....	7
II- Structures et cycles d'éléments transposables.....	23
III- Relations entre les éléments transposables et leurs hôtes.....	40
III-I Les différents éléments transposables dans les génomes d'aujourd'hui	41
III-II Co-adaptation des éléments transposables avec leur hôte : limitation de leur impact délétère sur l'hôte ou optimisation de leur effet bénéfique ?....	59
Le modèle.....	80
I- La drosophile, un modèle génétique.....	81
II- L'ovogenèse de <i>Drosophila melanogaster</i>	83
Matériels et méthodes.....	86
Résultats.....	89
I- Caractérisation de <i>ZAM</i> et d' <i>Idefix</i> , deux nouveaux <i>errantivirus</i> mobilisés dans une souche <i>Drosophila melanogaster</i> W ^{IR6REV} (RevI).....	91
II- Etude du cycle errantiviral de <i>ZAM</i> . (article III).....	119
III- Localisation des déterminants génétiques responsables de l'instabilité génétique dans la lignée W ^{IR6REV}	132
Discussion.....	141
I <i>ZAM</i> et <i>Idefix</i> sont deux <i>errantivirus</i>	142
II Identification d'un <i>locus</i> de l'hôte contrôlant l'expression de <i>ZAM</i> et d' <i>Idefix</i> : Centre Organisateur de la Mobilisation (C.O.M.).....	146
III- Conclusion.....	158
Liste des abréviations.....	159
Table des matières.....	160
Références bibliographiques.....	166



RAPPORT BIBLIOGRAPHIQUE

Introduction

Les éléments mobiles du génome, appelés éléments transposables ou transposons, sont des composants naturels et ubiquitaires de tous les génomes, des bactéries jusqu'aux vertébrés (Berg & Howe, 1989). Ils en sont même un constituant majeur, puisqu'on estime leur représentation à 15% chez la drosophile (Finnegan, 1992), 45% chez l'homme (International Human Genome Sequencing Consortium, 2001), plus de 50% pour le maïs (San Miguel *et al.*, 1996).

Chez les eucaryotes, ils constituent une partie des séquences répétées. Ils sont dispersés dans tout le génome, mais sont plus concentrés dans de grandes régions initialement décrites comme non transcrites, souvent sous forme fossile (incomplète ou inactive). La chromatine présente une forme condensée à l'interphase qui caractérise l'hétérochromatine. On distingue l'hétérochromatine α qui se trouve dans les centromères et les télomères, principalement composé d'ADN satellite, de l'hétérochromatine β , moins condensée, à la jonction des bras euchromatiques. Parce que l'hétérochromatine β contient peu de gènes et qu'elle est composée pour l'essentiel de ces éléments fossiles, certains auteurs lui donnent parfois le nom de "séquence poubelle" (garbage DNA et même Wasteland !).

Les premiers transposons ont été découverts en 1951 dans les chromosomes de maïs par Barbara Mac Clintock, prix Nobel de médecine en 1983 (Mac Clintock, 1955). Ces éléments ont la propriété de changer de localisation chromosomique. Ils sont capables de s'insérer dans l'ADN de l'hôte à de nouveaux sites chromosomiques et parfois de disparaître de leur localisation préalable. De nombreux transposons ont été décrits depuis cette première découverte. Ils ont été trouvés dans tous les génomes où on les a cherchés. Ils présentent une étonnante diversité dans leur structure, leur mécanisme de transposition, et leur distribution : Leur nombre de copies peut varier de quelques dizaines à quelques milliers selon l'élément et l'espèce considérés.

Chaque transposon contient les gènes requis pour sa propre mobilisation, mais il exploite les fonctions cellulaires de son hôte, à la façon d'un parasite, ce qui vaut aux éléments transposables une autre appellation tout aussi orientée et finaliste que celles désignant l'hétérochromatine de "séquences égoïstes" (selfish DNA).

La mobilisation des éléments transposables au sein du génome ne se fait pas toujours sans dommage pour la cellule ou l'organisme dans lequel elle se manifeste. Les transposons sont connus pour leurs effets délétères et sont à l'origine de pathologies. Les organismes ont développé de multiples stratégies pour que leur

génomique ne soit pas totalement envahi par les éléments transposables et pour se défendre de leur pouvoir mutagène. Pourtant, les éléments mobiles ont été maintenus au cours de l'évolution, grâce à leur capacité de réplication ou par la sélection (Finnegan, 1989; Charlesworth, Sniegowski & Stephan, 1994), et leur conservation au sein des génomes donne lieu à de farouches spéculations sur leur rôle éventuel et leur origine.

L'implication des éléments transposables dans l'évolution de leur hôte est fréquemment évoquée car leur activité est une source majeure de changements génétiques; ils pourraient jouer un rôle important pour le maintien des recombinaisons génétiques ou dans des processus comme la spéciation. De plus, il n'est plus très sûr que l'hétérochromatine ne soit qu'un cimetière d'éléments fossiles, résultant d'une sélection à long terme, car de nouvelles insertions massives y sont observées en une seule génération (Dimitri & Junakovic, 1999). Enfin, quelques équipes ont montré des domestications d'éléments, c'est à dire l'utilisation d'une fonction d'un élément par son hôte. Ces découvertes récentes ont provoqué l'apparition des mots "symbiotiques", "bénéfiques", "amis", "voisins tranquilles" et même "compagnons de voyage" pour désigner les éléments transposables et les revues sur leurs rôles éventuels abondent (Taruscio & Mantovani, 1998; Boecke & Devine, 1998; McDonald 1998; Dimitri & Junakovic, 1999; Sverdlov, 1999; Kidwell & Lisch, 2000; Pardue, 2000; Bennetzen, 2000).

D'une façon générale, les éléments transposables sont peu mobilisés. Les nouvelles insertions que l'on observe sont de rares échappements aux stratégies de défense mises en place dans les cellules. Il existe néanmoins des modèles expérimentaux permettant l'observation d'amplifications massives d'éléments: les conditions de stress peuvent être à l'origine de telles amplifications chez les plantes (Grandbastien, 1998) ou dans des cellules de mammifères en culture (Liu *et al.*, 1995; Kimura, Choudary & Schmid, 1999). Il arrive aussi que soit mise en évidence en laboratoire une instabilité génétique corrélée à des mobilisations d'élément(s). Ces modèles permettent d'étudier l'impact sur le génome de telles mobilisations et d'en rechercher le déterminisme génétique. Il existe peu d'exemples chez les vertébrés mais la drosophile, insecte facile à reproduire en laboratoire, dont la génétique est bien connue, est un organisme de choix pour ces études.

I- Classification et nomenclature des éléments transposables.

A) Les différents éléments mobiles et leur classification

Les éléments mobiles du génome ont en commun la propriété de changer de localisation chromosomique, mais ils présentent une grande diversité, tant structurale que mécanistique. On distingue deux grandes classes d'éléments, selon que l'intermédiaire de transposition est un ARN ou un ADN, et un deuxième niveau de classification rend compte de la variété de leurs structures (Finnegan, 1989). Cela permet de définir différentes familles qui portent souvent le nom du premier élément décrit. Mais au fil des nouvelles descriptions d'éléments, ou de l'avancée dans les connaissances de leurs cycles, nous verrons dans les prochains chapitres que certaines familles ont dû être reconsidérées :

1) La classe I

La classe I regroupe ce qu'on appelle les rétroéléments, parce que l'intermédiaire de transposition est un ARN. Les rétroéléments autonomes codent une rétrotranscriptase, capable de synthétiser une molécule d'ADN complémentaire (ADNc) à partir d'une matrice constituée de l'ARN génomique simple brin. Ils codent aussi une activité intégrase, qui permettra à cet ADNc de s'intégrer nouvellement dans le génome. Les éléments de classe I n'ont été décrits que dans les génomes eucaryotes.

Selon la structure génomique des rétroéléments, on distingue deux groupes : les rétrotransposons à LTR (Long Terminal Repeat) qui ont une structure de type provirale, et les rétroposons, qui ont une structure de type pseudogène.

Les rétrotransposons à LTR, décrits initialement chez les invertébrés et les plantes, présentent des analogies de structure et de séquences avec les rétrovirus, au point que les abréviations et la nomenclature pour les protéines ou autre caractère structural ont été adoptés pour leur description (Leis *et al.*, 1988). Depuis peu, ils sont regroupés avec les rétrovirus selon une nouvelle nomenclature (Boeke *et al.*, 1998 a,b). Comme l'objet de cette thèse porte sur la caractérisation de deux d'entre eux, *ZAM* et *Idefix*, et que cette nomenclature est récente, ils feront l'objet d'un chapitre particulier.

Le second groupe de rétrotransposons se distingue par l'absence de LTR, d'où la désignation "rétrotransposons sans LTR" utilisée parfois dans la littérature.

des éléments bactériens IS, ainsi qu'avec les intégrases des rétroéléments (figure 1) (Henikoff, 1992 ; Doak *et al.*, 1994).

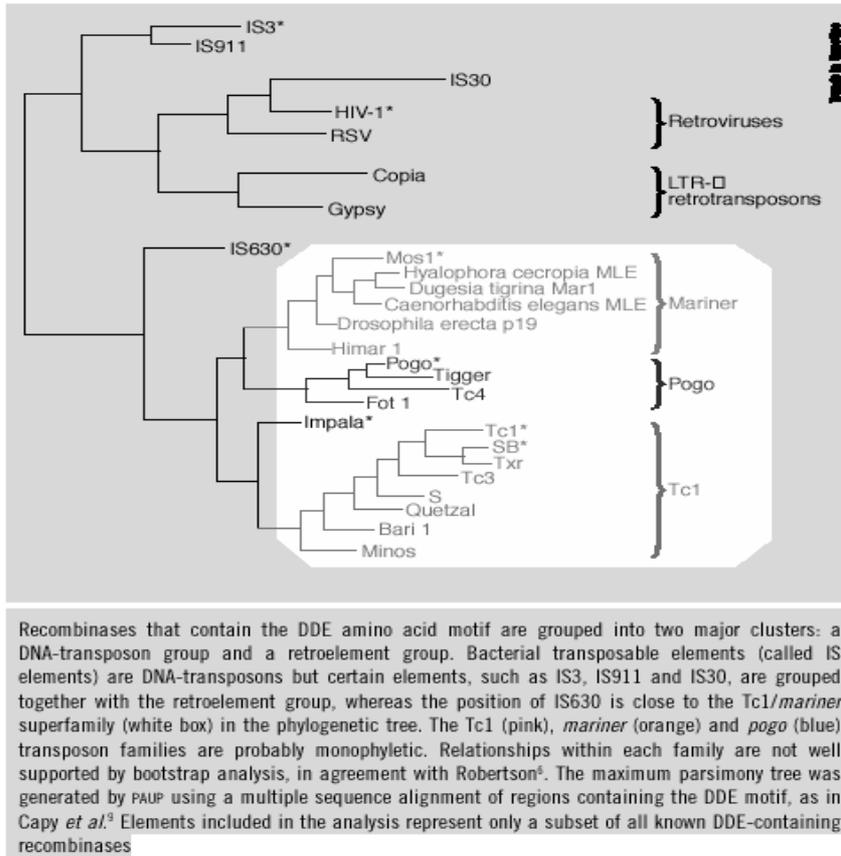


Figure 1: Arbre phylogénétique des intégrases de la superfamille Tc1/mariner. (Plasterk, Izvak & Ivics, 1999).

B) Les virus à ARN rétrotranscrit

a) Rétrovirus et LTR-rétrotransposons

Les rétrovirus appartiennent à la famille des virus enveloppés, et possèdent les propriétés infectieuses de tout virus. Ils ont en plus la faculté de modifier le génome de leur hôte en s'y insérant à la manière des éléments transposables. Ils sont donc transmis génétiquement à la descendance des cellules infectées sous la forme intégrée dite provirale. Quand les cellules infectées sont germinales, les provirus appartiennent alors au patrimoine génétique de leur hôte et sont transmis d'une génération à l'autre selon les lois de Mendel. C'est le cas pour les nombreux rétrovirus endogènes qui colonisent les génomes de vertébrés.

Les rétrovirus sont donc à l'intersection entre éléments transposables et virus, d'autant qu'en plus de cette propriété commune d'intégration dans le génome, les rétrovirus partagent les caractéristiques structurales des rétrotransposons à LTR. La différence essentielle entre les deux semblait résider dans la capacité d'infection

des rétrovirus. Cette capacité leur est conférée par une protéine que l'on croyait spécifique des rétrovirus, codée par le gène *env*. En s'en tenant à cette distinction, les rétrovirus endogènes défectifs pour *env* étaient qualifiés par certains auteurs de "LTR-rétrotransposons" (Kazazian 1998).

On peut noter une autre différence entre ces rétroéléments, plus historique que scientifique : les rétrovirus ont été initialement découverts chez les vertébrés, alors que les premiers rétrotransposons ont été décrits chez les invertébrés sous des formes provirales incomplètes (comprenez sans le gène *env*). La limite entre les deux groupes est devenue plus floue au fil des découvertes de nouveaux éléments, en particulier quand on a trouvé des LTR-rétrotransposons contenant le gène *env*. Certains auteurs s'en tenaient à la différence structurale (présence ou non du gène *env* complet), les autres au spectre d'hôtes (vertébrés ou invertébrés). Mais, comme le remarquaient Pantazidis *et al.* en 1999, depuis que l'on sait que des rétrovirus peuvent transposer sans étape extra-cellulaire (Tchenio & Heidman 1991), c'est à dire sans étape d'infection, et que les propriétés infectieuses d'un LTR-rétrotransposon ont été démontrées (Kim *et al.*, 1994 ; Péliesson *et al.*, 1994 et Song *et al.*, 1994), la classification de certains éléments en temps que rétrovirus ou LTR-rétrotransposons était devenue caduque ou pour le moins confuse.

Boecke *et al.*, ont mis fin à ces ambiguïtés en créant une nouvelle classe de virus appelée "virus à ARN rétrotranscrit", qui regroupe l'ensemble des rétrovirus et des LTR-rétrotransposons. Ils les répartissent en trois familles qui correspondent finalement aux anciennes appellations : la famille des *retroviridae* contient tous les rétrovirus de vertébrés, complets ou non ; les *metaviridae* sont les LTR-rétrotransposons de type *Ty3-gypsy* et les *pseudoviridae* les LTR-rétrotransposons de type *Ty1-copia*. Les trois familles étaient déjà ressorties d'une analyse phylogénique par Xiong et Eickbuch sur la base de leur domaine rétrotranscriptase en 1990. La nouveauté, si l'on peut dire, a été de définir des genres parmi les *metaviridae* et les *pseudoviridae* et d'accoler le suffixe "virus" à chacun d'eux. Cette classification pouvait laisser croire qu'elle respectait la dimension historique car elle recoupait les spectres d'hôtes préalablement définis, jusqu'à ce que les séquençages systématiques de génomes révèlent l'existence de *metaviridae* chez les vertébrés (Miller *et al.*, 1999).

Notons que pour les *metaviridae*, les genres créés se distinguent par l'absence (*metavirus*) ou la présence (*errantivirus*) du gène *env*. Nous allons voir que ce retour à la structure génomique des éléments atteint déjà ses limites.



b) Nomenclature des virus à ARN rétrotranscrit.

La nouvelle taxonomie est résumée dans le tableau I. Ce tableau ne permet pas de se retrouver dans toutes les appellations de la littérature, les plus anciennes ayant toujours cours, surtout pour ce qui concerne les *retroviridae*. C'est pourquoi je vais un peu détailler ce qu'il recouvre. D'autre part, il m'a paru utile de citer quelques exemples d'éléments dans chaque famille concernée et d'illustrer le texte avec des données historiques.

b1) Les *retroviridae*

Les premiers rétrovirus ont été découverts au début du XXème siècle (Ellerman & Baug, 1908). En 1970, l'existence de la fonction "rétrotranscriptase" mettait fin au dogme de l'unidirectionnalité de l'information génétique (Temin & Mizutani, 1970; Baltimore, 1970). Mais c'est seulement après l'identification des LTR et la découverte des provirus intégrés dans les chromosomes de l'hôte que le terme rétrovirus a été proposé (Gilboa *et al.*, 1979). Les rétrovirus ont d'abord été divisés en quatre groupes sur la base de leur morphologie, et nommés types A, B, C ou D (figure 2). Au fur et à mesure de la découverte de nouveaux rétrovirus, une classification tenant compte des similarités dans l'organisation des génomes, de la conservation de séquences nucléiques et peptidiques, et des modalités d'expression des gènes viraux s'est imposée pour définir d'abord trois genres, puis sept (tableau I). Du strict point de vue de l'organisation des génomes, on distingue les rétrovirus dits simples, possédant les trois cadres de lecture *gag*, *pol*, *env*, et les rétrovirus dits complexes qui contiennent des cadres de lecture supplémentaires, codant pour des protéines dites accessoires ou additionnelles.

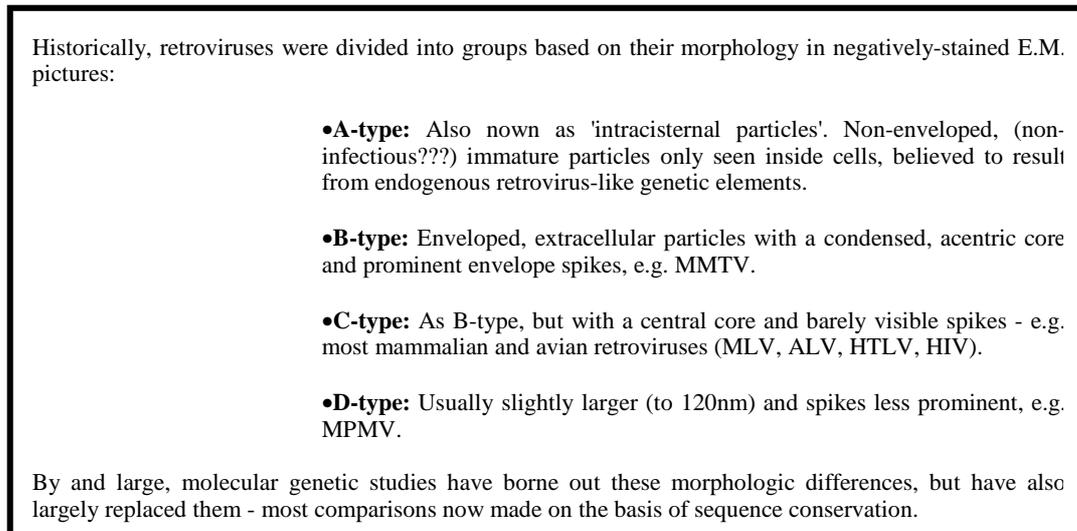


Figure 2 : Ancienne nomenclature des rétrovirus basée sur leurs morphologies. (<http://www-micro.msb.le.ac.uk/335/Retroviruses.html>)

- Les oncovirus

Le groupe des oncovirus contient les cinq premiers genres énumérés dans le tableau I. Les oncovirus ont en commun d'être responsables de tumeurs variées chez leurs hôtes. Ils peuvent être aussi à l'origine d'immunodéficiences et de dégénérescences neurologiques. Ces rétrovirus portent généralement des gènes cellulaires modifiés que l'on appelle "*v-onc*". Ces gènes ont été acquis par transduction à partir de gènes cellulaires dits "*c-onc*", impliqués dans la croissance ou la réplication cellulaire et confèrent leur pouvoir tumorigène (Bishop, 1983 ; Varmus, 1984 ; Cooper, 1990).

Tableau I : Taxonomie des virus à ARN rétrotranscrit telle qu'elle est publiée sur le site Internet: <http://www-micro.msb.le.ac.uk/335/Retroviruses.html>. J'ai laissé apparaître la colonne des spectres d'hôte qui est aujourd'hui totalement erronée en ce qui concerne les *metaviridae* et les *pseudoviridae*. En effet, ces deux familles sont présentes dans les plantes et chez certains vertébrés. Les cinq premiers genres des *retroviridae* (a,b,...,e) sont aussi regroupés sous l'appellation oncovirus.

Group VI : RNA Reverse Transcribing Viruses			
Family	Genus	Type Species	Hosts
<i>Retroviridae</i>	<i>Alpharetrovirus</i>	<i>Avian leukosis virus</i>	Vertebrates
	<i>Betaretrovirus</i>	<i>Mouse mammary tumor virus</i>	Vertebrates
	<i>Gammaretrovirus</i>	<i>Murine leukernia virus</i>	Vertebrates
	<i>Deltaretrovirus</i>	<i>Bovine leukemia virus</i>	Vertebrates
	<i>Epsilonretrovirus</i>	<i>Wallay dermal sarcoma virus</i>	Vertebrates
	<i>Lentivirus</i>	<i>Human immunodeficiency virus 1</i>	Vertebrates
	<i>Spumavirus</i>	<i>Human spumavirus</i>	Vertebrates
<i>Metaviridae</i>	<i>Metavirus</i>	<i>Saccharomyces cerevisiae</i> <i>Ty3 virus</i>	Fungi
		<i>Errantivirus</i>	<i>Drosophila melanogaster</i> <i>gypsy virus</i>
<i>Pseudoviridae</i>	<i>Pseudovirus</i>	<i>Saccharomyces cerevisiae</i> <i>Ty1 virus</i>	Invertebrates
		<i>Hemivirus</i>	<i>Drosophila melanogaster</i> <i>copia virus</i>

- Les lentivirus

Les lentivirus causent des maladies chroniques à évolution lente qui ne portent pas de composants néoplasiques. Les virus responsables chez les mammifères d'immunodéficiences acquises de plusieurs espèces appartiennent au

genre lentivirus : HIV (Human Immunodeficiency Virus), SIV (Simian), FIV (Feline), BIV (Bovine). Les lentivirus sont des rétrovirus complexes.

- Les spumavirus ou foamy virus

Les spumavirus ont un potentiel pathogénique incertain et controversé (Weiss, 1988). Ils sont fortement cytopathogènes en culture de cellules, mais infectent l'hôte de façon asymptomatique et persistante en présence d'anticorps neutralisants. Le seul cas de pathologie observé fut une immunodépression chez un lapin auquel on avait inoculé un spumavirus simien (Hooks & Detrick-Hooks, 1979).

b2) Les *metaviridae*

Les *metaviridae* sont les éléments les plus proches des *retroviridae*. Ils ont la même organisation du gène *pol* et leur domaine rétrotranscriptase est très conservé (figure 3). On distingue deux genres, selon la présence ou non d'un troisième cadre de lecture *env*. Les *metavirus* en sont dépourvus (ex : *Ty3*), tandis que les *errantivirus* le possèdent (ex : *gypsy*). Les propriétés infectieuses d'un *errantivirus* ont été démontrées pour l'élément *gypsy* (Kim *et al.*, 1994 ; Song *et al.*, 1994). Compte-tenu du tropisme des *errantivirus* pour les cellules germinales, ils ont donc toutes les propriétés d'un rétrovirus endogène (voir d- Rétrovirus exogènes et endogènes).

Les *metaviridae* (gypsy-Ty3 subgroup) sont plus proches des *retroviridae* (retroviruses) que des *pseudoviridae* (Copia-Ty1 subgroup). A noter, la position intermédiaire des éléments TAS, Pao, DIRS, et Pat au sein des virus à ARN rétrotranscrit.

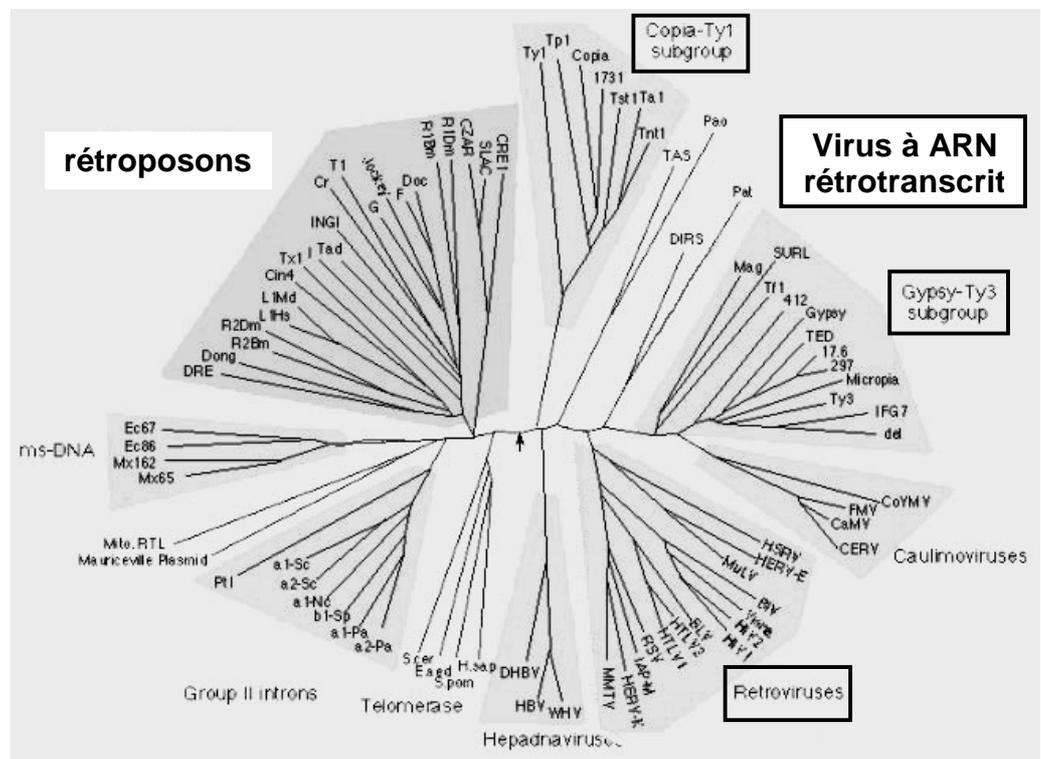
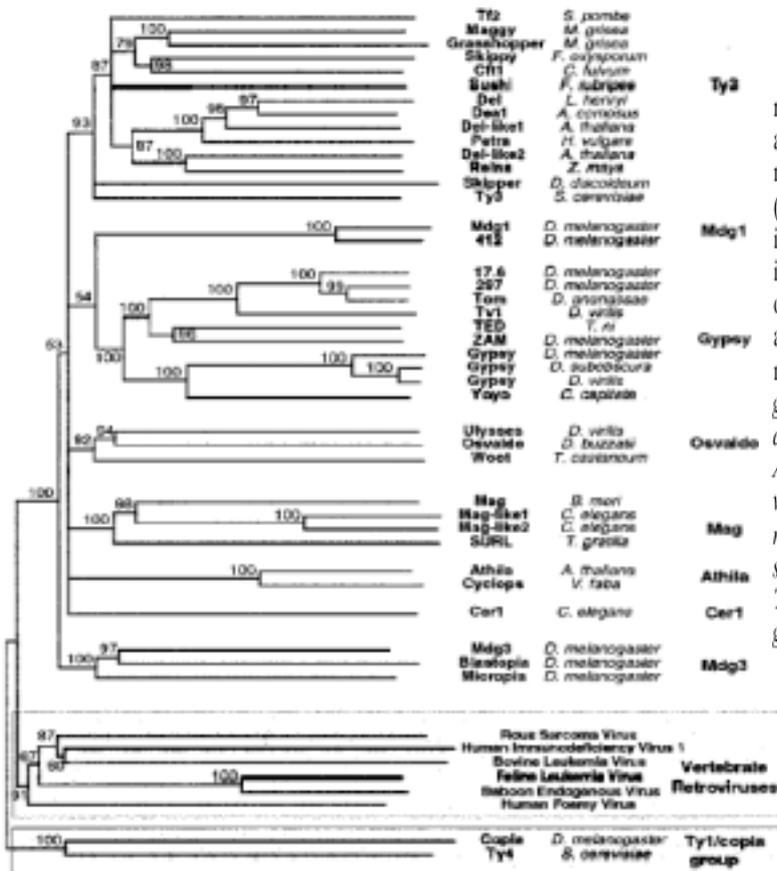


Figure 3 : Phylogénie des domaines rétrotranscriptase.

(<http://www.reochester.edu/College/BIO/labs/thelab/home.html>).

Malik & Eickbush (1999) ont défini huit clades dans un arbre phylogénétique établi en comparant chez les *metaviridae* trois domaines du gène *pol*: RT, RH, et IN (figure 4). L'étude plus exhaustive de Marin & Llorens en 2000 inclue des *metaviridae* identifiés dans les banques de données récentes de séquences génomiques de nombreux organismes. Le nouvel arbre construit sur la base du seul domaine RT possède un neuvième clade (*Grande/Tat*) (figure 5). Marin & Llorens proposent de créer le genre *chromovirus* pour désigner les éléments possédant un chromodomaine dans leur intégrase. Ils discutent de la position de deux éléments, *Ty3* et *Skipper* parce qu'ils n'ont pas la même dans l'arbre de Malik & Eickbush. Nous verrons qu'ils remettent en question la nomenclature de Boeckle *et al.* basée sur la présence de *env* pour distinguer les genres des *metaviridae* (voir c- La rétrotranscriptase au cœur de la nomenclature). Toutefois, comme sont rassemblés en un seul clade les éléments de type *gypsy* dans les deux études, j'appellerai les éléments de ce clade « *errantivirus* d'insectes ». Les *errantivirus* d'insectes décrits à ce jour sont listés dans le tableau II.



Phylogenetic tree of the Ty3/Gypsy group and related classes of LTR retrotransposons. The tree is based on an alignment of the sum of the amino acids in the reverse transcriptase, RNase H, and integrase domains (approximately 700 amino acid positions) and is rooted based on the Ty1/Copia group (29). Selected vertebrate retroviruses representing the diversity of this group are included as members of a sister group to the Ty3/Gypsy group. Numbers adjoining the branches indicate bootstrap values (percentages from 1,000 trials) based on results of the neighbor-joining (22) method. The various retroelements and associated host genomes are indicated. Each major lineage of the tree has been named after one of its elements. *S. pombe*, *Schizosaccharomyces pombe*; *M. grisea*, *Magnaporthe grisea*; *F. oxysporum*, *Fusarium oxysporum*; *C. fulvum*, *Cladosporium fulvum*; *F. rubripes*, *Fugu rubripes*; *L. henryi*, *Lilium henryi*; *A. comosus*, *Ananus comosus*; *A. thaliana*, *Arabidopsis thaliana*; *H. vulgare*, *Hyphomicrobium vulgare*; *Z. mays*, *Zea mays*; *D. discoideum*, *Dictyostelium discoideum*; *D. ananassae*, *Drosophila ananassae*; *D. virilis*, *Drosophila virilis*; *T. ni*, *Trichoplusia ni*; *D. subobscura*, *Drosophila subobscura*; *C. capitata*, *Ceratitidis capitata*; *T. castaneum*, *Tribolium castaneum*; *B. mori*, *Bombyx mori*; *C. elegans*, *Caenorhabditis elegans*; *T. granita*, *Tripneustes granitilla*; *V. faba*, *Vicia faba*.

Figure 4 : Arbre phylogénétique construit sur la comparaison des domaines RT, RH et IN du gène *pol*. Cet arbre distingue huit clades chez les *metaviridae* (Malik & Eickbush, 1999).

A partir des récentes données de séquences disponibles sur les génomes, Marin & Llorens (2000) évaluent la représentation des *metaviridae* : un à vingt *metaviridae* différents par espèce (respectivement chez *Saccharomyces cerevisiae*

et *Drosophila melanogaster*), mais aucun dans les génomes de mammifère. Il existe néanmoins des *metaviridae* chez les vertébrés, notamment chez les poissons, les amphibiens et les reptiles (Marin & Llorens, 2000 ; Miller *et al.*, 1999).

Tableau II : Liste des *errantivirus* d'insectes décrits comme appartenant au même clade défini par Malik & Eickbusch (1999), Malik, Henikoff & Eickbusch (2000) et Marin & Llorens (2000).

NOM	HOTE	REFERENCE
<i>gypsy</i>	<i>Drosophila melanogaster</i>	Marlor, Parkhurst & Corces, 1986
297	<i>Drosophila melanogaster</i>	Inouye, Yuki & Saigo, 1986
17.6	<i>Drosophila melanogaster</i>	Inouye <i>et al.</i> , 1986
<i>TED</i>	<i>Trichoplusia ni</i>	Friesen & Nissen , 1990
<i>Tom</i>	<i>Drosophila ananassae</i>	Tanda, Mullor & Corces, 1994
<i>Tv1</i>	<i>Drosophila virilis</i>	Andrianov & Shuppe, 1994
<i>Tirant</i>	<i>Drosophila melanogaster</i>	Molto <i>et al.</i> , 1996
<i>Burdock</i>	<i>Drosophila melanogaster</i>	Ponomarenko <i>et al.</i> , 1997
<i>ZAM</i>	<i>Drosophila melanogaster</i>	Leblanc <i>et al.</i> , 1997
<i>Nomad</i>	<i>Drosophila melanogaster</i>	Whalen & Grigliati, 1998
<i>yoyo</i>	<i>Ceratitis capitata</i>	Zhou & Haymer, 1998
<i>Idefix</i>	<i>Drosophila melanogaster</i>	Desset <i>et al.</i> , 1999
<i>Midline-jumper</i>	<i>Drosophila melanogaster</i>	Lammel & Klämbt, 2001

b3) Les *pseudoviridae*

Les *pseudoviridae* sont la troisième famille des virus à ARN rétrotranscrit. Leur gène *pol* présente ses domaines dans l'ordre : PR, IN, RT, RH et leur rétrotranscriptase est la plus divergente de celles des deux autres familles. A noter que les *pseudoviridae* ne sont pas restreints non plus aux invertébrés. Ils ont été décrit chez *Clupea harengus* (Flavell & Smith, 1992) mais aussi chez les plantes. Jusqu'à une date récente, ils se distinguaient par l'absence de gène *env* mais on a découvert chez le soja un élément qui possède ce gène, *SIRE-1* (Laten, Majumdar & Gaucher, 1998). Les *pseudoviridae* sont séparés en deux genres : les *pseudovirus* comme *Ty1* et les *hemivirus* comme *copia* (tableau I).

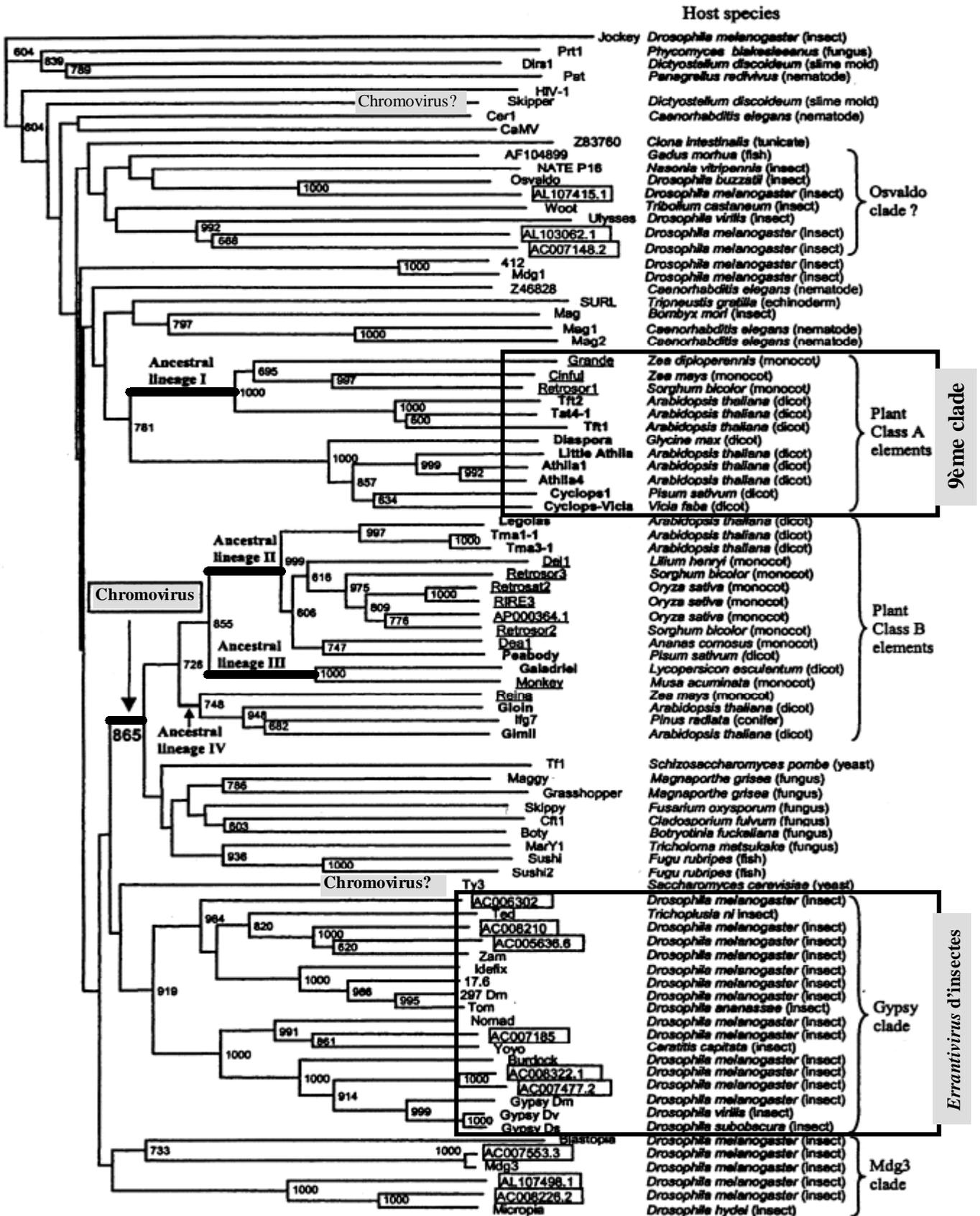


Figure 5 : Arbre phylogénétique construit sur la comparaison du seul domaine RT de tous les *metaviridae* publiés (Marin & Llorens, 2000). Cette étude comprend les séquences publiées par les consortiums impliqués dans les programmes de séquençages des génomes. Par rapport à l'arbre précédent (figure 4), cette étude définit un nouveau genre, celui des *chromovirus* et un neuvième clade, celui de *Grande/Tat*. A noter que le clade *gypsy* se distingue sans ambiguïté dans les deux arbres. J'appellerai ce clade "errantivirus d'insectes" dans la suite du texte.



c) La rétrotranscriptase au cœur de la nomenclature

La rétrotranscriptase est un enzyme commun à tous les rétroéléments. La nomenclature de Boeckle *et al.*, (1998 a,b) rend compte des comparaisons de séquences en acides aminés de différentes rétrotranscriptases établies précédemment par Xiong & Eickbuch en 1990 (figure 3) et permet de justifier scientifiquement le choix des familles qui a été proposé. Elle permet aussi de discuter des hypothèses sur l'évolution des rétroéléments.

Dès les années 80, certains auteurs ont étudié le lien évolutif qui pouvait exister entre rétrovirus et LTR-rétrotransposons (Temin, 1980 ; Flavell, 1981). En effet, en dehors de différences structurales évidentes comme l'absence du gène *env* dans les rétrotransposons qui étaient alors décrits, les similarités sont apparues rapidement pour les domaines codés par le gène *pol*, comme ceux de la RNAase H (RH), la rétrotranscriptase (RT), la protéase (PR), et l'intégrase (IN). En faisant l'hypothèse que ces similarités n'étaient pas une convergence, elles pouvaient refléter les séquences d'un ancêtre commun.

Temin supposait que les rétroéléments avaient évolué en même temps que le gène de la rétrotranscriptase à partir d'un ancêtre commun. D'anciens LTR-rétrotransposons auraient acquis le gène *env* par transduction(s) de gènes cellulaires codant pour des ligands de récepteurs membranaires et seraient à l'origine des rétrovirus. Cette acquisition n'aurait eu lieu que chez les vertébrés (Temin 1985). La découverte par Marlor, Parkhurst and Corces en 1986 d'un gène *env* chez un LTR-rétrotransposon d'invertébré a remis en question cette dernière hypothèse.

En 1990, alors que les données de séquences disponibles sur les rétroéléments étaient moins importantes qu'aujourd'hui, Xiong & Eickbuch ont établi un arbre phylogénétique en comparant les domaines de toutes les rétrotranscriptases décrites. Ils ont montré que toutes les rétrotranscriptases dérivent d'un ancêtre commun. Les rétrovirus et les LTR-rétrotransposons sont très proches et l'ensemble se démarque des autres rétrotransposons. Xiong & Eickbuch ont proposé un modèle évolutif où apparaissaient déjà les trois principales familles: rétrovirus, LTR-rétrotransposons de type *Ty1/copia*, et LTR-rétrotransposons de type *Ty3/gypsy* (figure 6). Dans la publication, les deux familles de rétrotransposons portaient encore le nom générique de LTR-rétrotransposons, mais les auteurs avaient bien mis en évidence que le groupe *Ty3-gypsy* était plus proche phylogénétiquement des rétrovirus. Ceci n'est pas très surprenant, car le groupe

Ty3-gypsy possède la même organisation du gène *pol* que l'on retrouve chez les rétrovirus : PR, RH, RT, IN. La figure 3 représente un arbre réactualisé depuis la publication, où les trois familles sont bien distinctes.

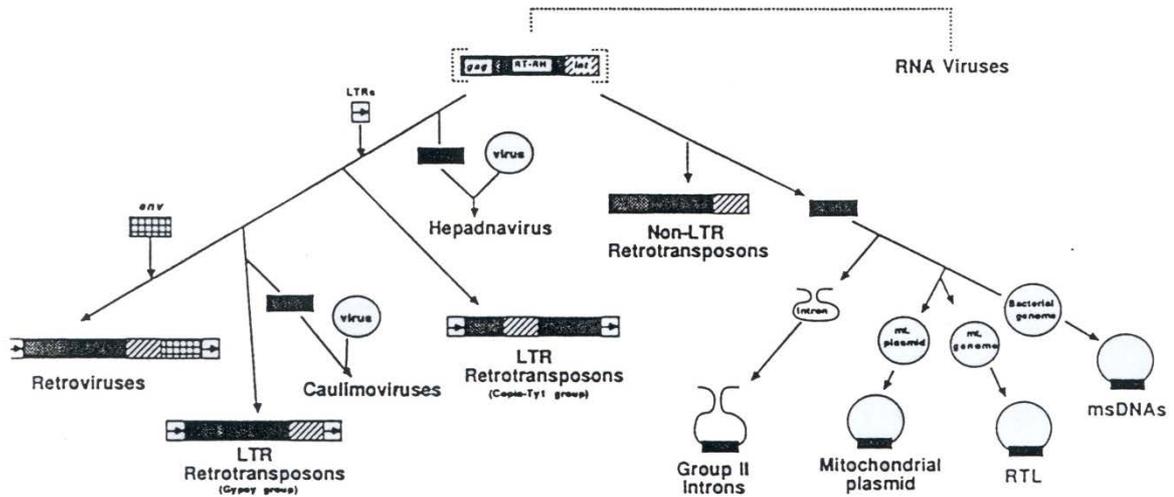


Figure 6 : Modèle évolutif des rétroéléments proposé à partir de l'arbre phylogénétique de la figure 3 (Xiong & Eickbush, 1990). On trouve les trois familles *retroviridae* (Retroviruses), *metaviridae* (LTR-rétrotransposons gypsy group) et *pseudoviridae* (LTR-retrotransposons copia-Ty1 group). Les deux dernières familles se distinguent par la position du domaine intégrase dans le gène *pol*. Une seule acquisition de gène *env* est proposée et concerne seulement les rétrovirus dans ce modèle.
 Rectangles: cadres de lecture ouverts principaux: en noir, domaine RT ; hâché en diagonal domaine IN ; grillagé, le domaine *env*. Les LTR sont repéré par des rectangles contenant des flèches.

Les gènes *env* de tous ces éléments ont aussi été comparés. Moins bien conservés que les gènes *pol*, on observe pourtant deux domaines communs entre les protéines d'enveloppe des *errantivirus* et celles des *retroviridae*, mais à des positions très différentes (Lerat & Capy, 1999). L'arbre obtenu en comparant ces domaines conservés est le même que celui établi par Xiong & Eickbush en 1990 avec le domaine RT, suggérant aux auteurs l'hypothèse d'acquisitions indépendantes par module des autres domaines d'enveloppe. A noter que ces domaines ne sont pas tous retrouvés dans la seule protéine d'enveloppe de *pseudoviridae* décrite, *SIRE-1* (Laten, Majumdar & Gaucher, 1998).

Pour autant, personne ne répond encore aujourd'hui à la question qui était posée par McClure en 1991 : Est-ce que l'ancêtre commun des virus à ARN rétrotranscrit a perdu son gène d'enveloppe dans certaines familles ou bien a-t-il acquis un gène d'enveloppe dans les autres ? L'histoire semble plus compliquée, ponctuée d'acquisitions indépendantes de gènes *env* différents. Malik, Henikoff & Eickbush (2000) ont montré par exemple des acquisitions distinctes de gènes *env*

(figure 7). Ils ont identifié dans les banques de données de séquences l'origine probable de certains de ces gènes : gène de *baculovirus* pour les *errantivirus* d'insectes, *herpesvirus* et *phlebovirus* pour le clade appelé ici *BEL*. Il sera difficile de répondre définitivement à la question tant que les auteurs ne trancheront pas sur le rôle des transferts horizontaux dans l'évolution des éléments, les uns en tenant compte (Xiong & Eickbush, 1990), les autres les ignorant (Marin & Llorens, 2000).

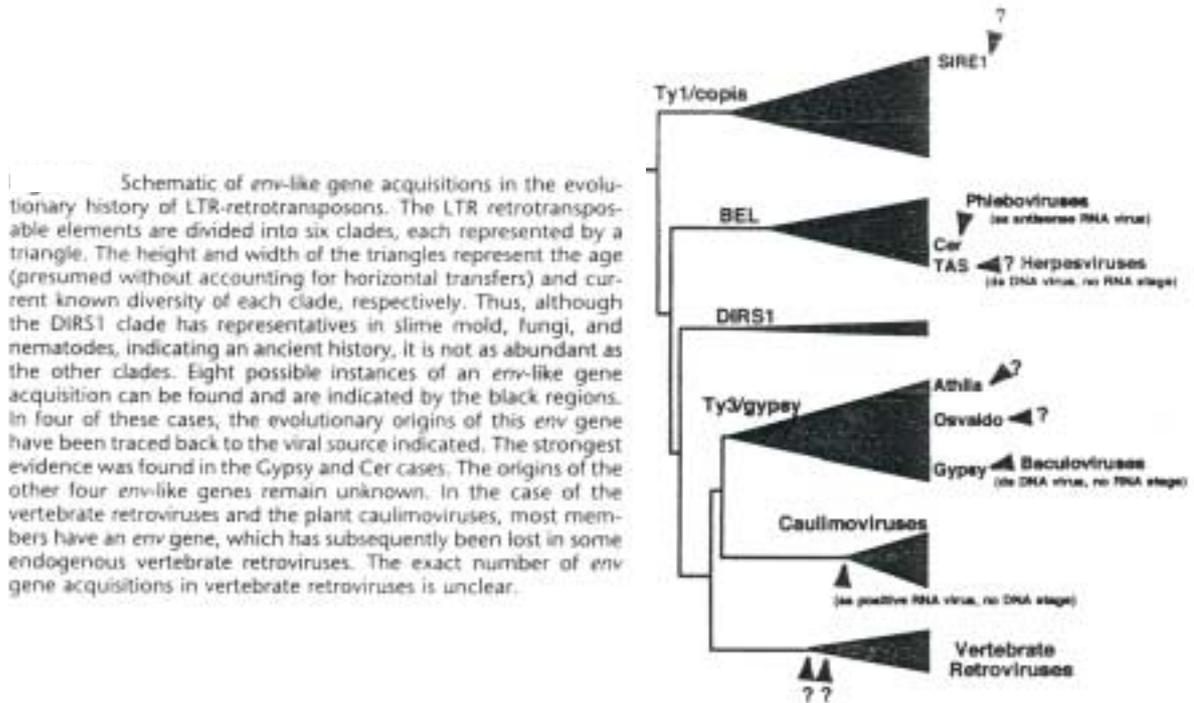


Figure 7 : Acquisition des différents gènes de type *env* par les virus à ARN rétrotranscrit, encore appelés LTR-rétrotransposons par les auteurs (Malik, Henikoff & Eickbush, 2000).

Parmi les *metaviridae*, huit clades différents ont été définis par Malik & Eickbush en 1999 sur la base de trois domaines du gène *pol* (figure 4). Ces clades ne regroupent pas dans un seul lignage les éléments initialement définis comme *errantivirus* (contenant *env*). En effet, les *metaviridae* possédant le gène *env* sont répartis dans quatre clades non monophylétiques : *gypsy*, *Osvaldo*, *Athila* et *Cer-1*. Ceci suggère des étapes successives de perte/acquisition du gène *env* ou des acquisitions indépendantes. Cette dernière hypothèse prévaut depuis les découvertes d'un *pseudoviridae* possédant un gène *env*, *SIRE-1* (Laten, Majumdar & Gaucher, 1998), et d'un élément intermédiaire entre *pseudoviridae* et *metaviridae* possédant lui aussi un gène *env*, *Cer-7* (Bowen & McDonald, 1999). Si les acquisitions sont indépendantes, cela remet en question la nouvelle taxonomie qui distingue deux genres sur la base de la présence ou non du gène *env*.

On a donc atteint les limites d'une classification basée sur l'organisation structurale et on se retourne à nouveau sur la comparaison des gènes *pol* pour faire la part des choses, en particulier le domaine intégrase : Malik & Eickbuch ont mis en évidence un chromodomaine dans certaines intégrases (1999). Marin & Llorens (2000) ont fait une analyse plus exhaustive des *metaviridae*, en ne comparant au départ de l'étude que le domaine RT, et en incluant les dernières données des séquences génomiques de différents organismes. Ils ont mis en évidence un groupe d'éléments monophylétiques qui ont en commun de porter un chromodomaine dans leur intégrase (sauf pour quatre d'entre eux). Ils proposent donc de créer le genre *chromovirus* pour désigner ces éléments. L'ancêtre commun des *chromovirus* serait antérieur à la séparation entre les plantes et les métazoaires puisqu'on en trouve dans les deux ordres. L'acquisition de ce domaine serait unique et les quatre éléments de ce clade qui en sont dépourvus l'auraient perdu au cours de l'évolution (figure 5).

Il semble donc que les classifications basées sur l'organisation structurale des éléments soient insuffisantes et s'accommodent mal de l'augmentation des données, alors que la comparaison du domaine très conservé de la rétrotranscriptase permet de classer les nouveaux venus. Mais on le voit, les genres sont encore discutables et discutés !

d) Rétrovirus exogènes et endogènes

Les appellations de rétrovirus exogènes ou endogènes méritent un commentaire car elles sont très utilisées dans la littérature. Contrairement à ce que l'on pourrait croire, les deux termes ne sont pas exclusifs ; c'est pourquoi ces appellations entraînent parfois une certaine confusion.

Historiquement, les rétrovirus endogènes ont été découverts au hasard de criblages de banques génomiques avec des sondes du gène *pol* de rétrovirus déjà caractérisés. Le terme « provirus » serait donc plus approprié pour désigner ces éléments. Ils ont été identifiés sur la base d'analogies avec des séquences rétrovirales, mais ils ne répondent pas obligatoirement aux critères taxonomiques qui définissent un rétrovirus : Pour Coffin, Hughes & Varmus (1997) les rétrovirus sont une famille de virus à ARN enveloppés possédant des caractères taxonomiques communs, comprenant la structure et la composition de leur particule, leur organisation génomique, et leur cycle. Le plus souvent, les rétrovirus endogènes sont incomplets, inactifs et ne possèdent aucun de ces caractères. On retiendra que ce sont des séquences du génome de l'hôte (endogènes), présentant

des analogies de séquence avec les provirus, sans présager de leur activité, tant pour leur expression que pour leur éventuelle mobilité dans le génome.

Du point de vue évolutif, on considère que les éléments les plus anciens sont ceux que l'on trouve à l'état fossile tandis que les plus complets, présentant encore un tropisme pour les cellules germinales, sont arrivés récemment dans l'espèce. On appelle ces deux catégories de rétrovirus endogènes respectivement "anciens" ou "modernes".

Une nomenclature est utilisée pour distinguer les nombreuses familles de rétrovirus endogènes (*HERV* pour *Human Endogen RetroVirus*). Elle fait référence à la répllication des rétrovirus. Chaque rétrovirus amorce sa rétrotranscription avec un ARN de transfert (ARNt) spécifique de sa séquence (voir le chapitre sur le cycle des rétrovirus). Les ARNt sont spécifiques d'un acide aminé. Aussi, on ajoute parfois la lettre symbolisant cet acide aminé au nom *HERV* pour désigner une famille d'éléments utilisant l'ARNt correspondant. ex: la famille *HERV-K* désigne les rétrovirus utilisant l'ARNt spécifique de la Lysine pour initier leur rétrotranscription.

Les rétrovirus exogènes se caractérisent par leur capacité de transfert horizontal entre les individus de l'espèce hôte ou parfois entre espèces, aptitude qui dépend de son tropisme pour un tissu cellulaire donné. Ce tropisme dépend de nombreux facteurs de l'élément comme de ceux de l'hôte, en particulier de la protéine codée par le gène *env* et de son affinité pour le récepteur cellulaire cible. Les rétrovirus exogènes sont parfois classés selon leur spectre d'hôtes. Les termes indiquent le tropisme d'un rétrovirus pour une espèce d'hôte donnée : On distingue pour une espèce les rétrovirus écotropiques, qui n'infectent **que** cette espèce, les rétrovirus xénotropiques, qui n'infectent **pas** cette espèce, et les rétrovirus polytropiques qui infectent l'espèce considérée **et** d'autres. Ces caractères sont parfois variables parmi les individus d'une même espèce (voir III-).

On a vu que le classement taxonomique place les *errantivirus* dans une famille distincte de celle des *retroviridae*. Toutefois, il faut noter que les *errantivirus* sont des séquences provirales dans le génome de l'hôte correspondant à des rétrovirus endogènes d'insectes. Terzian, Péliesson and Bucheton (2001) ont d'ailleurs proposé d'adopter une nomenclature similaire à celle des rétrovirus endogènes de vertébrés pour les *errantivirus* d'insectes: *IERV* (*Insect ERrantiVirus* ou *Insect Endogenous RetroVirus*) suivi de la lettre symbolisant l'acide aminé porté par l'ARNt amorçant la rétrotranscription de l'élément.



Les *errantivirus* sont un excellent modèle pour étudier les relations d'un rétrovirus avec son hôte, car ce type d'étude est plus facile à mener chez un insecte comme la drosophile, dont la génétique est bien connue, que chez les vertébrés. Inversement, les connaissances déjà établies sur les rétrovirus, sur leur mode de réplication notamment, sont une base de données bien utile pour l'étude du cycle des *errantivirus*.

II- Structures et cycles d'éléments transposables

Je n'évoquerai ici que certains éléments modèles nécessaires à la compréhension du chapitre suivant sur les relations éléments-hôtes. Je détaillerai davantage la structure et le cycle des rétrovirus, qui ont été la base de nos recherches pour caractériser *ZAM* et *Idefix*.

A) Eléments de classe II sur un exemple : Structure et cycle de l'élément *P*

L'élément *P* a été isolé lors de la caractérisation d'une mutation au locus *white* chez *Drosophila melanogaster* (Rubin, Kidwell & Bingham, 1982). Il s'agit d'un élément dont l'intermédiaire de transposition est un ADN (Engels, 1989).

La taille de l'élément canonique complet est de 2907 pb. Il possède des séquences répétées inversées de 31 pb à ses extrémités pouvant lier la protéine de drosophile IRBP (Inverse Repeat Binding Protein). Ces séquences encadrent quatre exons numérotés de 0 à 3. Ces exons peuvent être traduits en deux protéines différentes selon l'épissage subi par le messager (figure 8)(O'Hare & Rubin, 1983). L'ARNm totalement épissé contient quatre exons et permet la traduction de la transposase, une protéine de 87 kDa, tandis que le messager partiellement épissé possède l'intron IVS3, situé entre l'exon 2 et l'exon 3, et code un répresseur de 66 kDa (Karess & Rubin, 1984; Rio, Laski & Rubin, 1986 ; Misra & Rio, 1990 ; Laski, Rio & Rubin, 1986). Enfin, deux régions sont nécessaires à la transposition, situées en 5' et 3' de l'élément, qu'on appelle les pieds de *P*. La présence de ces séquences permet de mobiliser en *trans* des éléments défectifs pour les régions codantes (Engels, 1984). Cette propriété est exploitée dans les vecteurs de transgénèse chez la drosophile qui contiennent les pieds de *P* (Spradling & Rubin, 1982 ; Rubin & Spradling, 1982).

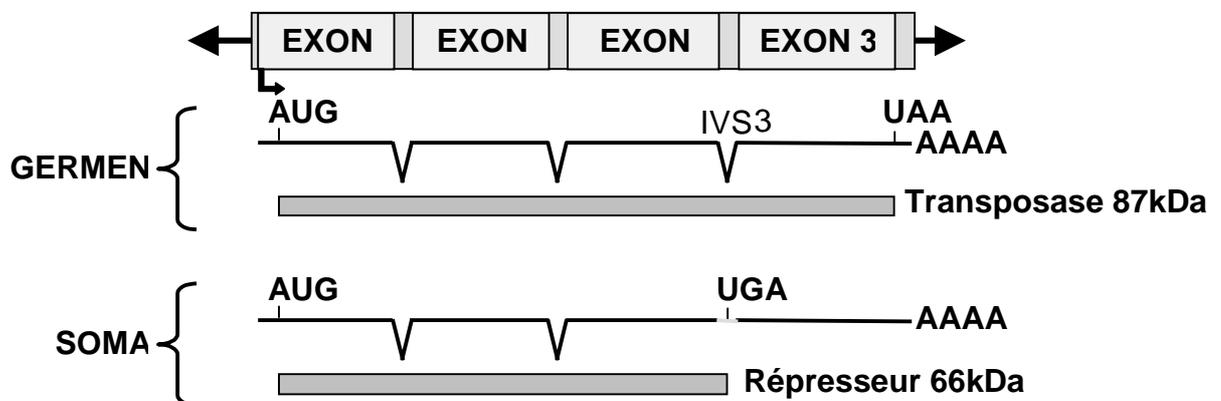


Figure 8 : Structure génomique et produits de l'élément *P*. L'épissage de l'intron IVS3 est spécifique des cellules germinales et permet la production de la transposase. Cet intron possède un codon stop, ce qui conduit à la production du répresseur en l'absence d'épissage.

L'élément *P* est transcrit dans tous les tissus mais subit un épissage alternatif germinale du troisième intron. Cet épissage explique la spécificité germinale de la transposition, puisque c'est le seul tissu où la transposase est produite (Laski, Rio et Rubin, 1986). Dans le soma, le troisième intron, dit IVS3 pour "intervening sequence 3", n'est pas épissé, conduisant à la traduction du répresseur à cause de la présence d'un codon stop dans l'intron.

Sur la base d'expériences génétiques, Engels *et al.* (1990) ont proposé que la transposition de *P* se produise selon un mécanisme conservatif "couper-coller".

B) Eléments de classe I

1) Les rétroposons de type *LINE* et leur mécanisme de transposition.

Les rétroposons *LINEs* ont d'abord été identifiés comme de longues séquences répétitives (>4 kb) dans les génomes de mammifères (Singer, 1982). La structure d'un élément *LINE* complet est schématisé figure 9. Chaque *LINE* possède une région 5' non traduite (5'UTR : 5' UnTranslated Region) contenant un promoteur interne, deux phases ouvertes de lecture (ORF1 et ORF2) et une région 3' non traduite (3'UTR) terminée par une queue polyA (élément *L1* ; Scott *et al.*, 1987) ou riche en A (répétitions TAA du facteur *I* chez *D. melanogaster* ; Fawcett *et al.*, 1986).

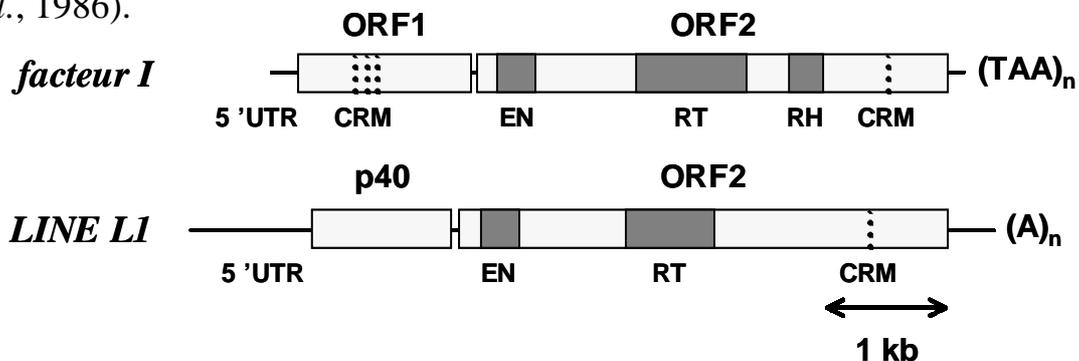


Figure 9 : Structure génomique des *LINEs*. 5'UTR: région 5' non traduite / CRM: motif riche en cystéine / EN: endonucléase / RT: rétrotranscriptase / RH: RNAaseH.

L'ORF1 du *LINE L1* code une protéine possédant une activité ARNsb-binding, la p40 (Holmes, Singer et Swergold, 1992). Elle présente des homologies avec les protéines Gag de rétrovirus (Kolosha & Martin, 1997). La fonction de la p40 n'est pas connue, mais on sait qu'elle n'est pas indispensable à la rétrotranscription (Moran *et al.*, 1996). Il a été montré que l'activité de liaison à l'ARNsb de la protéine p40 est spécifique du transcrit *L1*, dont elle reconnaît la partie 5' de la région codant l'ORF2 (Hohjoh & Singer, 1997). La p40 forme des

complexes ribonucléoprotéiques (RNP) contenant ses transcrits (Hohjoh & Singer, 1996 ; Martin, 1991).

L'ORF2 code pour une rétrotranscriptase possédant une activité endonucléase (Mathias *et al.*, 1991; Feng, Moran & Kazazian, 1996). Le domaine endonucléase/ rétrotranscriptase est très conservé entre les éléments *L1*. Il représente environ 300 acides aminés à l'extrémité N-terminale. En C-terminal, on trouve parfois un domaine RNAaseH. La rétrotranscriptase peut lier l'ARN de *L1* et l'accompagner dans le noyau mais elle est plus difficilement détectée que la p40.

Le mécanisme de rétrotransposition proposé est basé sur l'étude de l'élément *R2Bm* du ver à soie réalisée *in vitro* (Luan *et al.*, 1993) : Une activité endonucléase coupe un brin de l'ADN génomique de l'hôte et libère une extrémité 3'OH. Cette extrémité sert d'amorce à l'activité polymérase de la reverse transcriptase. La rétrotranscription peut s'arrêter prématurément, ce qui explique les nombreux éléments *LINES* tronqués en 5'. Le second brin de la cible est clivé en aval de la première coupure, ce qui aboutira à la duplication de la cible à chaque extrémité. La distance entre les deux coupures est variable et détermine la taille des répétitions directes flanquantes générées (figure 10). Le second brin d'ADN est ensuite synthétisé.

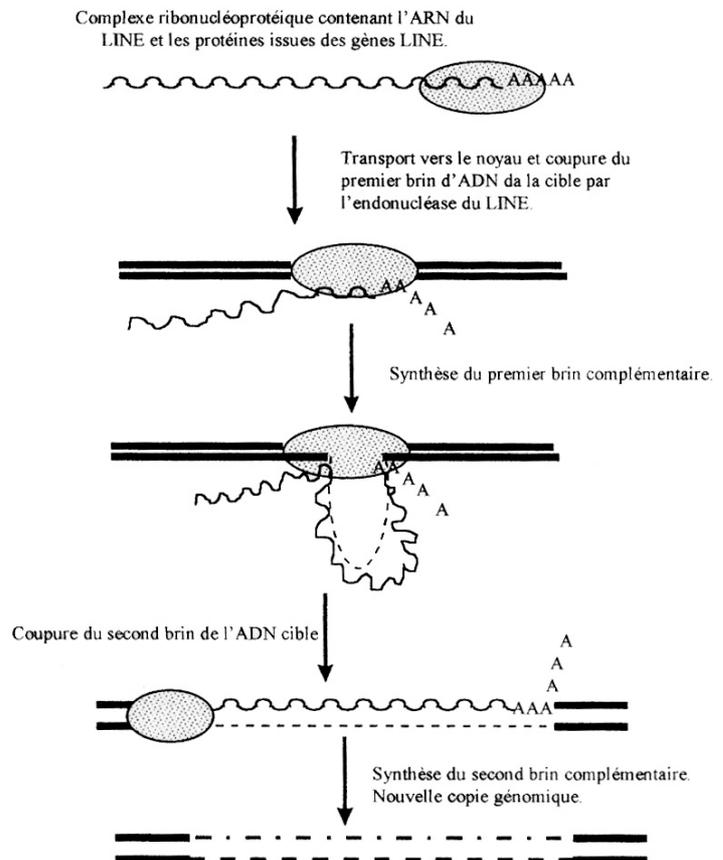


Figure 10 : Mécanisme de rétrotransposition TPRT (Target Primed Reverse Transcription). (Arnaud (2000) d'après Luan *et al.*, 1993).

Ce mécanisme nécessitant une coupure simple brin préalable à la rétrotranscription est appelé Target Primed Reverse Transcription (TPRT). L'activité endonucléase de la protéine codée par l'ORF2 a été testée *in vitro* et suggère que le mécanisme TPRT est généralisable à l'ensemble des éléments *LINEs* (Feng *et al.*, 1996). La cible préférentielle des *LINEs* est (Py)_n/(Pu)_n, "/" indiquant le site de coupure (Feng *et al.*, 1996). Les protéines d'un élément *LI* donné lieraient préférentiellement les ARN transcrits à partir d'un élément *LI* de même origine (Boeke, 1997 ; Naas *et al.*, 1998) voire le transcrit à partir duquel les protéines *LI* ont été traduites (Esnault, Maestre & Heidman, 2000). Cette propriété est appelée *cis*-préférence. Si l'activité TPRT semble bien liée au mécanisme de rétrotransposition des éléments *LINEs*, la reverse transcriptase pourrait fonctionner dans le cytoplasme avec n'importe quel messenger cellulaire comme matrice, avec une moindre efficacité compte-tenu de la *cis* préférence (Holmes *et al.*, 1994; Moran *et al.*, 1996). C'est pourquoi il a été proposé que les *LINEs* seraient à l'origine des pseudogènes (Dhelin, Maestre et Heidmann, 1997). Jurka a d'ailleurs identifié la même année un pseudogène possédant une séquence flanquante très homologue aux séquences flanquantes des éléments *LI*.

2) Les rétroposons *SINEs*

Les *SINEs* sont des éléments de petite taille (100-500pb) se terminant en 3' par un stretch d'adenines ou encore des poly ATT. Ils dérivent tous d'un ARNt (Okada, 1991 ; Ohshima et Okada, 1994), sauf les familles *Alu* et *B1* qui dérivent de l'ARN 7SL (Ullu et Tschudi, 1984 ; Weiner, 1980). Ils possèdent un promoteur interne reconnu par l'ARN polymérase III, avec les deux boîtes très conservées A et B distantes de 30 à 60 nucléotides (figure 11). Contrairement aux *LINEs*, les *SINEs* sont toujours des séquences pleine longueur (Weiner, Deininger & Efstrahadis, 1986).

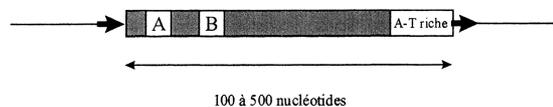


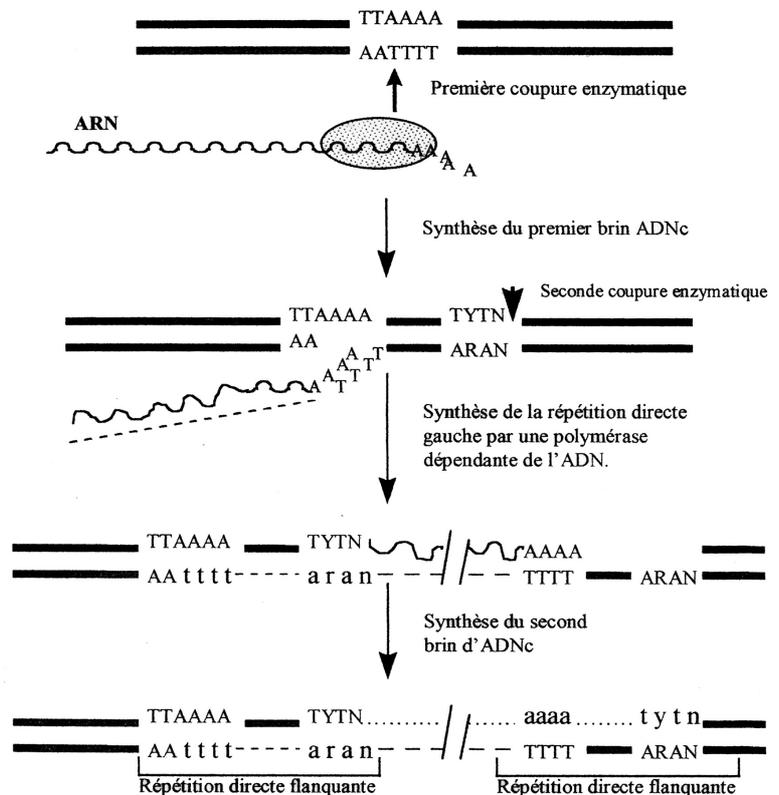
Figure 11 : Structure génomique des *SINEs* (Arnaud, 2000). Les flèches indiquent les répétitions directes de la séquence cible (jusqu'à une vingtaine de nucléotides). A et B sont des séquences conservées de promoteurs reconnus par l'ARN polymérase polIII. La région AT riche est de taille variable selon les éléments (quelques dizaines de nucléotides).

Les *SINEs* ne codent aucune protéine. Leur rétrotransposition nécessite une complémentation en *trans*. Les éléments *LINEs* sont de bons candidats pour cette complémentation. Jurka a recensé de nombreuses jonctions d'insertions et séquencé les cibles des rétroposons (1997-1998). Il observe que les rétroposons ont le même consensus TT/AAAA suggérant un mécanisme commun de rétrotransposition. Les

mêmes observations ont été faites chez les plantes (Tatout, Lavie & Deragon, 1998).

On estime la représentation de la seule famille *Alu* chez l'homme à $1,1 \cdot 10^6$ copies par génome, celle des *LINEs LI* à $0,5 \cdot 10^6$ copies (International Human Genome Sequencing Consortium). Ces éléments non autonomes ont donc envahi le génome humain 2 fois plus que les éléments *LINEs* alors qu'ils exploiteraient les activités reverse transcriptase et endonucléase de ces derniers ! L'expansion spectaculaire des *SINEs*, si elle est bien due à l'exploitation de la machinerie des *LINEs*, peut sembler contradictoire avec la *cis* préférence des protéines *LINEs* pour leurs messagers. Boeke l'explique par la probable proximité des messagers des *SINEs* et des protéines de *L1* sur les ribosomes (1997). Il s'appuie sur la forte affinité du domaine nucléique 7SL des *Alu* pour des protéines liant la grande sous-unité des ribosomes (Sarrowa, Chang et Maraia, 1997). Le complexe placerait les queues polyA des *Alu* à proximité des reverse transcriptases de *L1* (figure 12). D'autre part, l'équipe d'Okada a mis en évidence que plusieurs familles de *SINE* apparentées aux ARNt présentent en 3' une forte homologie avec les régions 3' des *LINEs* dans le même organisme, suggérant des couples *SINE-LINE* (Okada *et al.*, 1997).

Figure 12 : Modèle de rétrotranscription et d'intégration d'un *SINE* (*Alu*, *BI* ou *ID*) utilisant les protéines codées par un élément *LINE* (*L1*). (Arnaud (2000) d'après Jurka, 1997).

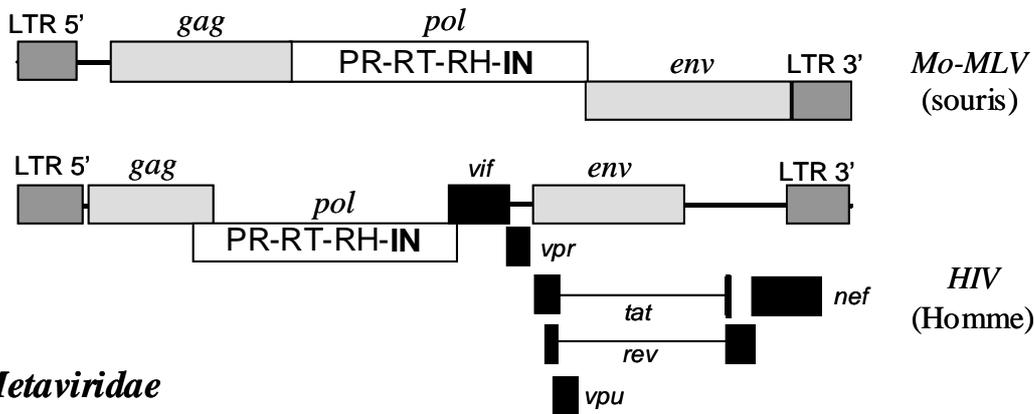


3) Les virus à ARN rétrotranscrit

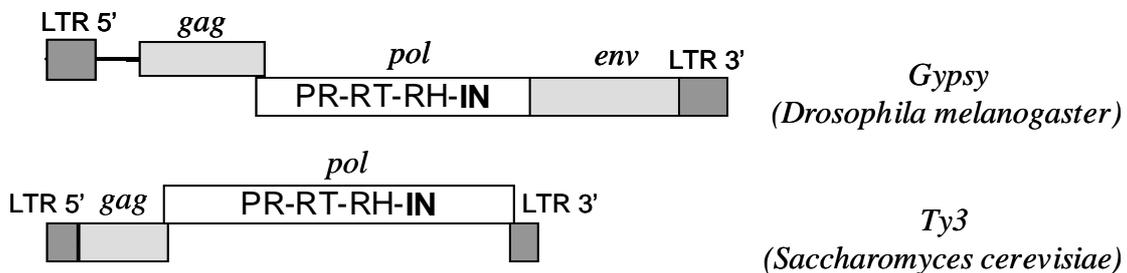
a) Structure des génomes intégrés

Le point commun de tous les virus à ARN rétrotranscrit est la présence aux extrémités de deux régions directement répétées, appelées LTR (Long Terminal Repeat) lesquelles encadrent des séquences non codantes et deux phases ouvertes de lecture *gag* et *pol*. Nous avons vu que certains éléments possèdent un troisième cadre de lecture, dit *env*, situé entre *pol* et le LTR3'. Les rétrovirus complexes ont la particularité de posséder des cadres de lecture supplémentaires codant pour des protéines dites accessoires. En figure 13, sont schématisés les génomes d'éléments représentatifs de chaque famille dans leur forme provirale (ou intégrée). En figure 14A est représenté un génome type avec troisième cadre de lecture.

Retroviridae



Metaviridae



Pseudoviridae

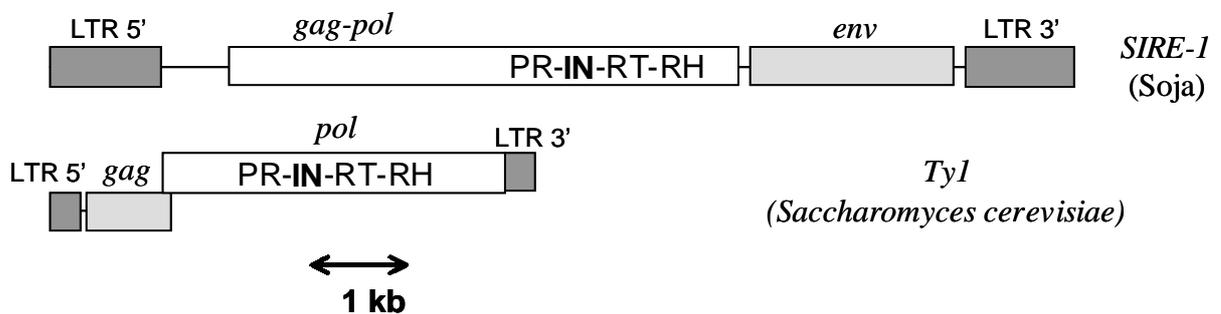


Figure 13 : Structures génomiques des virus à ARN rétrotranscrit : deux *retroviridae*: l'un simple (*Mo-MLV*: *Moloney Murine Leukemia Virus*), l'autre complexe (*HIV*: *Human Immunodeficiency Virus*); les protéines accessoires sont figurées en noir / deux *metaviridae*: l'un, avec enveloppe, un *errantivirus* d'insecte (*gypsy*), l'autre, sans enveloppe, un *metavirus* (*Ty3*) / deux *pseudoviridae*: 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: RNAase H / IN: intégrase.

- Les régions non codantes

**Les Long Terminal Repeat (LTR) sont deux longues régions directement répétées localisées aux extrémités des génomes. Chaque LTR se subdivise en trois parties, U3, R, et U5, qui sont définies selon leurs frontières, mais qui ne présentent pas forcément des homologies entre éléments. Chacune de ces régions a une taille très variable selon les virus, mais le plus souvent, la région centrale R est la plus petite (figure 14A).

Les frontières entre ces différentes régions sont définies par le début de transcription qui commence entre U3 et R, et le site de coupure et de polyadénylation qui se situe entre R et U5. Les autres frontières peuvent se déduire en comparant les deux LTR 5' et 3'. Elles correspondent aux sites d'initiation de la synthèse des brins "plus" et "moins" de l'ADNc. Dans l'ARN génomique, on trouve les régions R-U5 en 5' et U3-R en 3' (figure 14B). La région centrale R est donc la seule qui soit répétée dans l'ARN. C'est au cours de la synthèse de l'ADNc que sont générés les LTR complets du provirus.

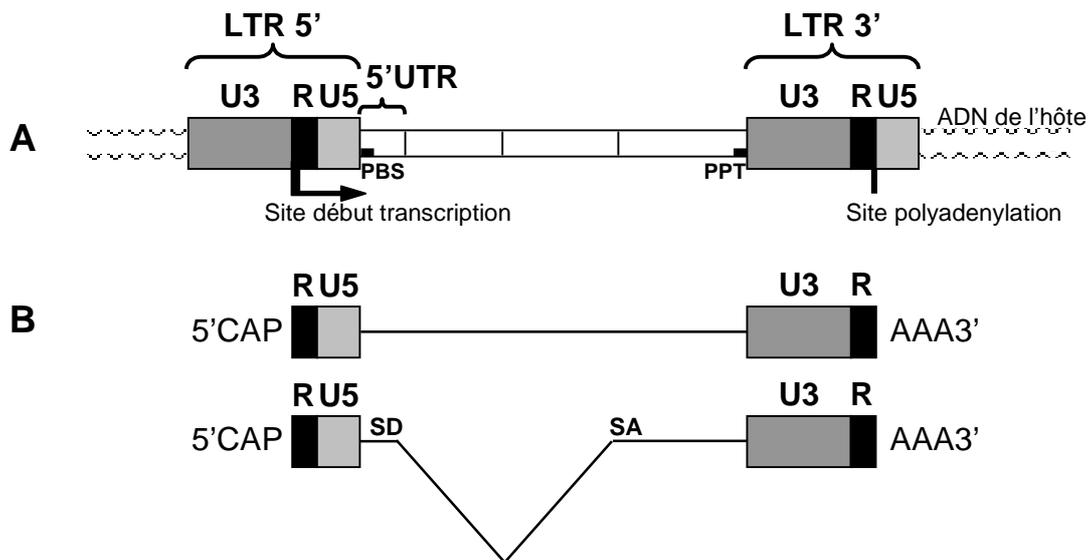


Figure 14 : Représentations schématiques d'un provirus et des transcrits correspondants.

A- Schéma d'un provirus avec trois cadres de lecture et une région 5' non traduite (5'UTR). Les deux LTR (Long Terminal Repeat) sont détaillés. Le début de U3 et la fin de U5 sont définis par rapport à la cible d'insertion. Les frontières U3/R et R/U5 sont définies respectivement par le début de la transcription et le site de polyadénylation. U3 contient quelques centaines à mille nucléotides, R une douzaine à une centaine, et U5 environ deux cents. L'ensemble U3/R/U5 compose le LTR. Deux petits rectangles noirs représentent le PBS (Poly Binding Site) dans le 5'UTR et le PPT (Poly Purine Tract) dans le cadre *env*.

B- Structure des messagers correspondant au provirus décrit en A.

CAP ou *coiffe 5'* / *SD* et *SA*: sites donneur et accepteur d'épissage / *AAA*: queue polyAdenyl.

La région U3 contient la plupart des éléments impliqués dans le contrôle transcriptionnel du provirus : le promoteur et les séquences *enhancer* responsables de la spécificité tissulaire de l'expression de l'élément.

**La région PBS (Primer Binding Site) est située juste en aval du LTR 5'. Cette séquence est complémentaire sur 18 nucléotides à l'extrémité 3' d'un ARN de transfert (ARNt) et permet d'amorcer la rétrotranscription.

**La région PPT (Poly Purine Tract) est plus petite, 9 nucléotides, composée d'adenines et guanines. Elle permet l'initiation de la synthèse du second brin d'ADN, dit brin "plus".

**La région 5' non traduite (5' UnTranslated Region (5' UTR), encore appelée séquence leader) est localisée entre LTR 5' et *gag*. Elle contient le PBS, la séquence Psi (Ψ) qui permet l'encapsulation des ARN génomiques dans les particules virales, et souvent des sites donneurs d'épissage utilisés lors de la maturation des ARN messagers (ARNm). Certains *errantivirus* contiennent dans leur 5'UTR d'autres séquences utilisées par l'élément comme les IRES (Internal Ribosome Entry Site) (Meignin *et al.*, soumis¹) mais aussi des séquences ayant une incidence sur le génome de l'hôte comme des séquences barrières (Geyer & Corces, 1992) ou des *enhancers* (Conte *et al.*, 2000 ; Conte, Dastugue & Vaury, 2002).

- Les régions codantes



Figure 15 : Représentation et organisation type du gène *gag* des *retroviridae*. Myr: site de myristylation / MA: matrice / CA: capside / NC: nucléocapside.

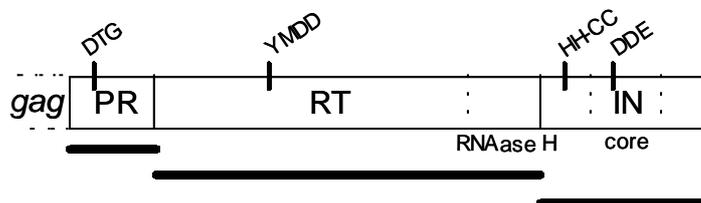


Figure 16 : Représentation et organisation des domaines du gène *pol* du *MLV* (*Murine Leukemia Virus*), que l'on retrouve chez les *metaviridae* et les *retroviridae*. Les traits gras figurent les protéines clivées. Les sites catalytiques de chaque domaine sont indiqués au dessus du schéma. Le domaine C-terminal de l'intégrase est mieux conservé chez les *metaviridae*. C'est dans ce domaine que se situe le motif GPY/F des *chromovirus* (Marin & Llorens, 2000).

¹ Meignin C., Bailly J.L., Dastugue B. and Vaury C. " The 5' untranslated region of a *Drosophila* retroelement, *Idefix*, promotes an Internal translation initiations."

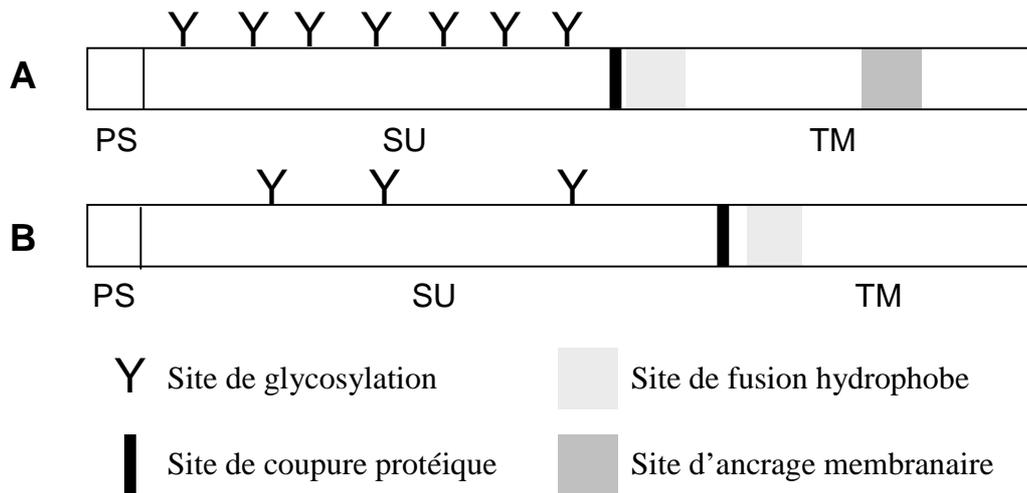


Figure 17 : Organisation des protéines d'enveloppe des virus à ARN rétrotranscrit. Dans chacun des éléments représentés, les deux traits verticaux séparent les trois domaines: PS: peptide signal / SU: sous-unité extra-cellulaire / TM: domaine transmembranaire.

A- Structure type d'une enveloppe de *retroviridae* et d'*errantivirus*. Certains éléments possèdent des sites de glycosylation dans leur domaine TM. Le site de coupure protéique est RXXR. Les errantivirus d'insectes possèdent un peptide de fusion en aval du site de coupure protéique.

B- Structure prédictive de la protéine d'enveloppe du *pseudoviridae SIRE-1*.

Les informations sur les régions codantes concernant les *retroviridae* sont extraites du livre "Retroviruses" (Coffin, Hughes, & Varmus, 1997).

****Le cadre de lecture *gag*** est le premier cadre de lecture rencontré après la région leader (figure 15). Il code une polyprotéine qui sera clivée en plusieurs sous-unités matures par la protéase du rétrovirus, toutes impliquées dans la structure interne de la particule virale ; dans l'ordre Nterm-Cterm : la matrice (MA pour membrane associated), la capside (CA) et la nucléocapside (NC). Les unités MA, CA, NC sont toujours assemblées respectivement de l'extérieur vers l'intérieur des particules (figure 18).

****Les gènes *prot/pol***. La majorité des virus à ARN rétrotranscrit ont un gène *pol* qui code pour trois activités enzymatiques : la protéase, (PR), la rétrotranscriptase/RNAase H (RT/RH) et l'intégrase (IN) dans cet ordre (figure 16). Seuls les *pseudoviridae* ont un domaine IN précédant celui de la rétrotranscriptase (figure 13). Il existe des exceptions: le domaine PR peut être codé dans un autre gène, à la fin du gène *gag* (ASLV, RSV), avant le domaine NC (Cer12, Bowen & McDonald, 1999) ou par un gène supplémentaire nommé *prot*, placé dans ce cas en amont de *pol* (HTLV-1, BLV, MMTV).

Les enzymes forment des domaines dans le précurseur Gag-Pro ou Gag-Pro-Pol mais ne sont pas toujours clivés en des protéines matures bien séparées. La

maturation a lieu au moment de l'assemblage des particules dans le cytoplasme ou même plus tard après la sortie de la cellule, selon les éléments.

La **protéase** est responsable du clivage de la polyprotéine Gag en MA, CA, NC et de la polyprotéine Pol en PR, RT/RH, et IN (figure 16).

L'activité **rétrotranscriptase** synthétise l'ADN complémentaire (ADNc) à partir de l'ARN génomique matrice, tandis que la **RNAase H** détruit cette matrice après la synthèse pour permettre ensuite la synthèse de l'autre brin, dit brin "plus". Le site actif RT est reconnaissable au motif YMDD (YMMD ou YVDD) (Xiong & Eickbush, 1988). Dans RNAase H, le "H" signifie hybride parce que le substrat de la RNAase H est l'hybride ADN/ARN. Par convention, rétrotranscriptase suppose la présence des deux activités RT et RH.

L'**intégrase** permet l'intégration du provirus dans le génome de l'hôte. Elle contient trois domaines fonctionnels essentiels dans les *retroviridae* qui sont conservés chez les *metaviridae* (Bowen & McDonald, 1999). En N-terminal, un motif HH-CC, qui ressemble à un doigt de zinc, serait responsable de la formation d'un complexe stable entre l'intégrase et le LTR de la séquence virale. La partie centrale, appelée core, contient le motif D, D₃₅-E qui est la partie catalytique de l'enzyme. Le dernier domaine est le moins conservé, sauf chez les *metaviridae* (Malik & Eickbush, 1999). Localisé en C-terminal, il présente de nombreuses charges positives et lierait l'ADN cible (Nymark-McMahon & Sandmeyer, 1999). Enfin, Malik et Eickbush, 1999 ont mis en évidence un chromodomaine dans certains *metaviridae*, avec un motif appelé GPY/F que l'on retrouve dans un lignage d'éléments appelés *chromovirus* par Marin & Llorens (2000). Le chromodomaine aurait un rôle dans l'adressage de l'élément.

****Le gène *env* :** La protéine d'enveloppe est traduite à partir d'un ARNm épissé, dit subgénomique, qui élimine les premiers cadres de lecture (figure 14B). C'est une protéine partiellement excrétée, possédant une séquence signal en N-terminal. Comme Gag et Pol, elle sera clivée, mais cette fois, par des protéases cellulaires, de type furine, pour donner deux sous-unités qui resteront liées par des liaisons non covalentes ou par des ponts disulfure chez ASLV. Le site consensus de coupure par la furine est de type RX(K/R)R. Un schéma de la structure primaire d'une protéine d'enveloppe type est représenté figure 17A.

La sous-unité N-terminal est une protéine de **surface (SU)**, très glycosylée et responsable de l'interaction avec le récepteur présent à la surface des cellules cible de l'infection. On trouve encore dans la littérature une nomenclature pour les sous-

unités SU correspondant à leur poids moléculaire apparent sur SDS-PAGE précédé des lettres gp pour glycoprotéine : ex: gp 120 est la sous-unité SU de HIV.

Le précurseur est dirigé pendant sa synthèse dans le lumen du réticulum endoplasmique grâce au peptide signal. Il reste fiché dans la membrane grâce à un domaine d'ancrage en C-terminal de la deuxième sous-unité, dite **transmembranaire (TM)**. Suit derrière ce domaine d'ancrage une queue C-terminale de la sous-unité. Le clivage entre les deux sous-unités a lieu dans l'appareil de Golgi. La sous-unité TM possède un domaine d'ancrage membranaire et un domaine hydrophobe en N-terminal responsable de la fusion membranaire entre l'enveloppe du virus et la membrane de la cellule cible au moment de l'infection. L'absence de clivage n'empêche pas la formation de la particule enveloppée mais la fusion membranaire lors de l'infection.

On retrouve dans les *metaviridae* le peptide signal, les sous-unités SU et TM, les sites de glycosylation dans SU, et le domaine de coupure RXXR. Une étude comparative des protéines d'enveloppe entre *retroviridae* et *metaviridae* a permis d'identifier deux autres motifs conservés, KRG et LTPL. Ce dernier est absent chez les *lentivirus* caprins et ovins. La position de ces deux motifs est très variable selon les éléments (Lerat & Capy, 1999). D'autre part, les enveloppes errantivirales contiendraient le motif de fusion membranaire des paramixovirus (Terzian, Péliesson & Bucheton, 2000).

La protéine Env du *pseudoviridae SIRE-1* semble posséder une séquence signal, peu de sites de glycosylation, des cystéines, mais semble-t-il, ni domaine d'ancrage ni site de coupure RX(K/R)R (figure 17B). Ils possèdent néanmoins un site de coupure putatif composé de quatre acides aminés comprenant deux acides aminés basiques (Laten, Majumdar & Gaucher, 1998).

****Les protéines accessoires :** Les *retroviridae* complexes possèdent parfois des cadres de lecture supplémentaires. Ils codent pour de petites protéines régulatrices. Chez les *metaviridae*, de telles protéines n'ont pas été décrites, même si leur existence est parfois suspectée, au vu de séquences nucléiques ou de certains transcrits.

b) Structure des particules de *retroviridae* enveloppées

La structure de la particule rétrovirale est donnée comme l'une des trois définitions d'un rétrovirus (Coffin, Hughes & Varmus, 1997). Elle est modélisée figure 18. Le génome rétroviral est constitué de deux molécules d'ARN, qui sont

deux brins physiquement liés par des liaisons hydrogène. Chaque ARN génomique est un brin "plus" présentant toutes les caractéristiques des messagers : chaque ARN génomique possède une coiffe en 5' (cap) et une queue polyadénylée. Les particules sont constituées avec les protéines structurales décrites plus haut codées par le gène *gag* et possèdent tous les éléments nécessaires à la rétrotranscription et à l'intégration : les enzymes codés par le gène *pol*, l'ARNt spécifique pour la répllication et des acides nucléiques

Les rétrovirus sont les seuls virus diploïdes. Leur génome est entièrement produit par la machinerie de transcription cellulaire sans l'intervention de polymérase codée par le virus. Ce sont les seuls virus qui requièrent un ARN spécifique cellulaire (ARNt) pour leur répllication. Enfin, leur ARN génomique n'est pas utilisé comme messenger après l'infe

Schematic cross section through a retroviral particle. The viral envelope is formed by a cell-derived lipid bilayer into which proteins encoded by the *env* region of the viral genome are inserted. These consist of the transmembrane (TM) and the surface (SU) components linked together by disulfide bonds. Internal nonglycosylated structural proteins are encoded by the *gag* region of the viral genome. They are the matrix (MA) protein, capsid (CA) protein, and nucleocapsid (NC) protein. The suggested icosahedral structure of the retroviral capsid is not definitely established (Chapter 2). Major products of the *pol*-coding region are reverse transcriptase (RT) and integrase (IN). The protease (PR) is derived from the *pro* gene between *gag* and *pol*.

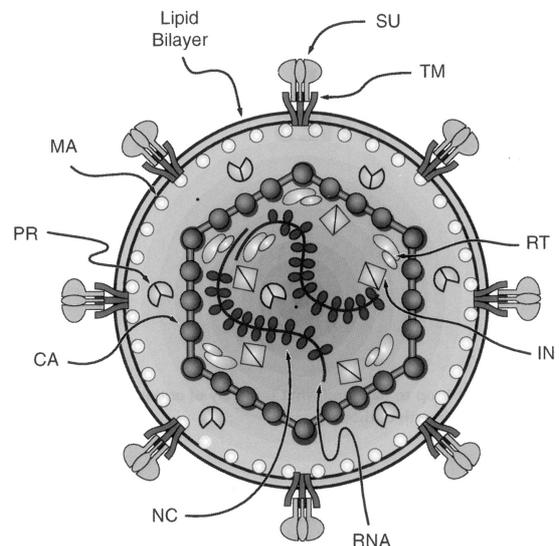


Figure 18 :

Structure schématique d'une particule virale mature enveloppée (Coffin, Hughes & Varmus, 1997).

c) Cycle des virus à ARN rétrotranscrit

Tous les virus à ARN rétrotranscrit ont un mécanisme semblable pour certaines étapes, l'étape extracellulaire étant propre aux éléments enveloppés. Précisons toutefois que ce sont surtout les cycles des rétrovirus qui ont été étudiés.

Comme je me suis intéressée aux rétrovirus endogènes, je commencerai le cycle à partir d'un provirus intégré. Un cycle rétroviral complet commence par une expression du provirus, c'est à dire sa transcription et sa traduction en protéines. Les particules virales sont ensuite assemblées dans le cytoplasme. Certains éléments retournent au noyau et évitent l'étape extracellulaire qui va de la sortie des particules dans le milieu extracellulaire, à une nouvelle infection de cellule hôte. La rétrotranscription commence dès l'entrée du virion dans la cellule, et l'ARN génomique est totalement répliqué en une molécule d'ADN provirale. La dernière

étape est la nouvelle intégration de l'ADNc dans le génome de l'hôte. Tous les virus à ARN rétrotranscrit suivent ce cycle, décrit figure 19A. Le détail des étapes est résumé à partir du livre "Retroviruses" (Coffin, Hughes & Varmus, 1997).

c1) Expression des provirus

****Transcription** : Le provirus est intégré dans le génome de l'hôte et donc, subit une régulation dépendant de son site d'intégration comme tous les gènes cellulaires. Lorsqu'il est transcrit, il est transcrit à partir de son promoteur dans les tissus d'intérêt pour son cycle et sa propagation, en particulier dans les tissus germinaux en ce qui concerne les rétrovirus endogènes.

Le provirus est transcrit par la machinerie de l'hôte. La transcription est initiée par l'ARN polymérase de type II au début de la région R. Le transcrit de pleine longueur se termine par une queue polyadénylée à la jonction R-U5 du LTR 3'.

Il semble qu'il existe deux populations distinctes d'ARN pleine longueur, la première comprenant les ARN génomiques qui seront encapsidés dans les particules virales, la seconde les messagers qui seront traduits en protéines.

****Epissage** : L'ARN pleine longueur est plus ou moins épissé pour générer des messagers de différentes tailles qui permettront la traduction de différentes protéines. Chaque rétrovirus possède un seul site donneur d'épissage localisé dans le LTR 5', dans la région 5' non traduite, ou au début du cadre de lecture *gag*. En revanche, on peut trouver plusieurs sites accepteurs qui génèrent chacun un transcrit subgénomique. Le transcrit d'enveloppe, totalement épissé pour *gag* et *pol* est caractéristique de tous les virus à ARN rétrotranscrit possédant le gène *env* (excepté le *pseudoviridae SIRE-1* dont le transcrit d'enveloppe n'a pas été décrit). Les autres transcrits n'ont pas toujours été formellement identifiés. Certains sont des transcrits codant les protéines accessoires des rétrovirus complexes, ou encore des transcrits spécifiques de *pol* dans lesquels *gag* est totalement épissé, comme dans le cas des spumavirus.

****Traduction** : La traduction peut s'initier après un consensus de Kosak ou après un IRES si une telle séquence est présente dans l'élément. A partir du messenger de pleine longueur, on peut traduire la protéine Gag. Un "frame-shift" (décalage de la phase de lecture), ou la suppression de la terminaison de traduction, ou un messenger épissé spécifique de *pol*, permettent de traduire Pol dans 5% des molécules synthétisées, sous la forme du précurseur Gag-Pro-Pol. Tous ces

mécanismes conduisent à l'obtention des protéines Pol et Gag en proportion définie, en général de 1 pour 20 respectivement. Cette proportion doit être respectée pour permettre un assemblage correct des particules.

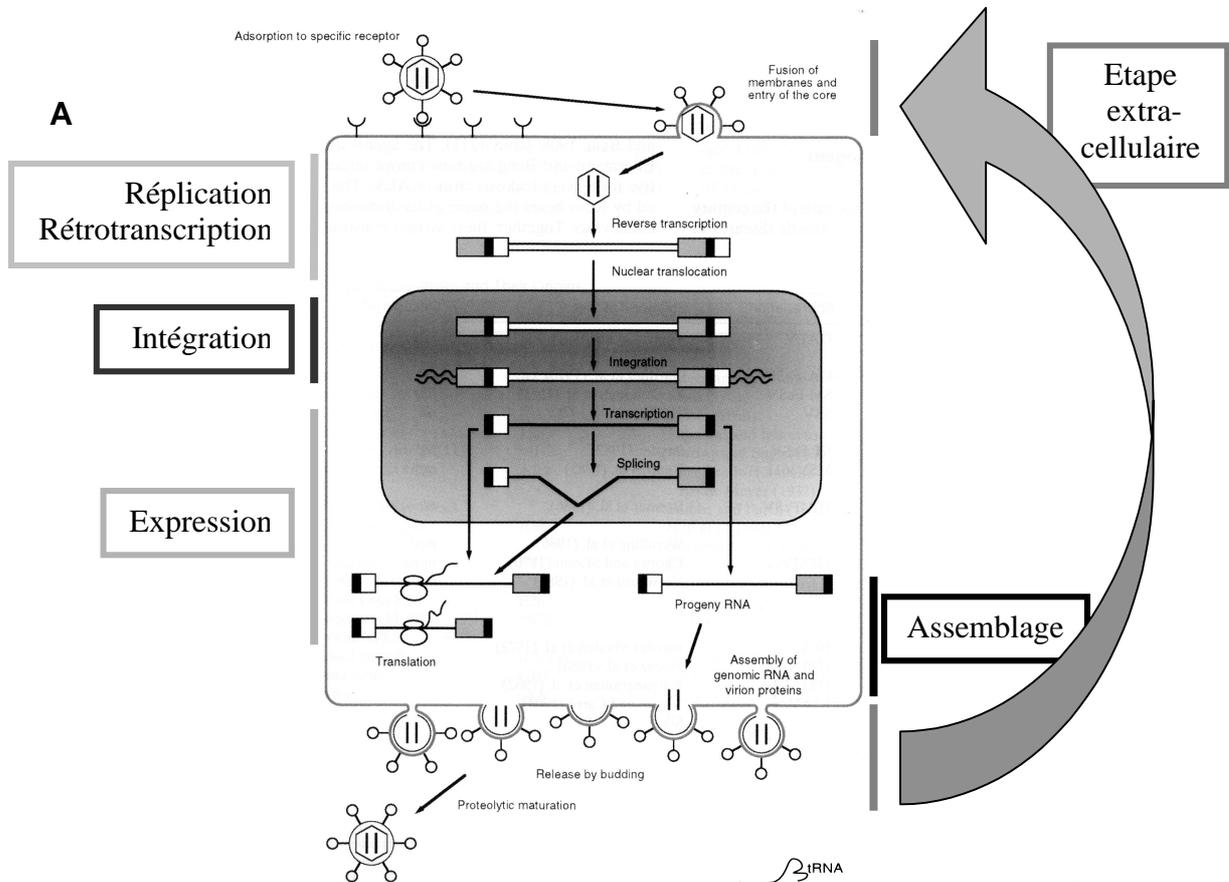
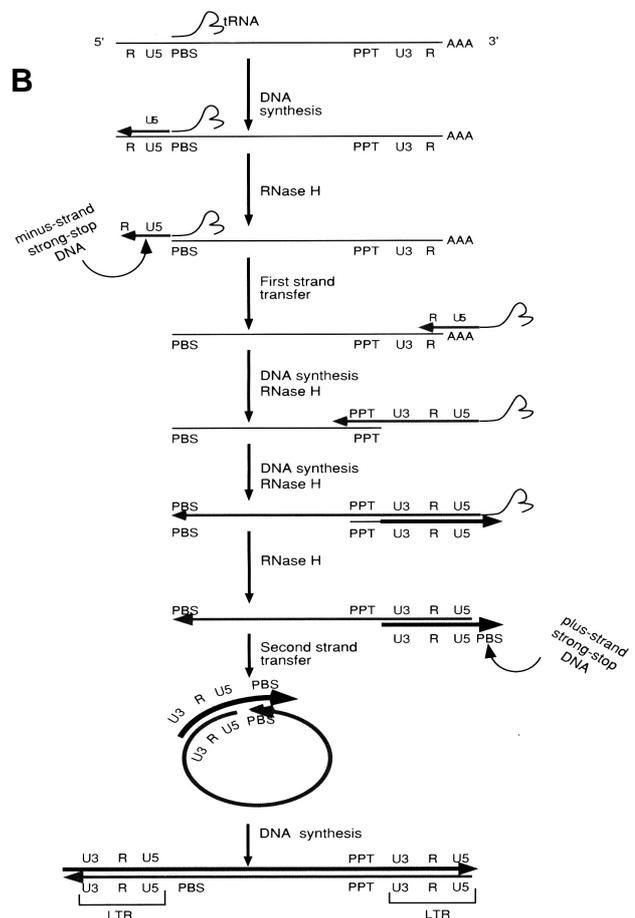


Figure 19 : Cycle des virus à ARN rétrotranscrit (d'après Coffin *et al.*, 1997).

A- Cycle complet.

L'étape extra-cellulaire concerne seulement les éléments possédant le gène *env*. Les éléments sans gènes *env*, ainsi que certains *retroviridae*, peuvent réaliser un cycle entièrement intra-cellulaire : les virions assemblés dans le cytoplasme sont alors dirigés vers le noyau.

B- Réplication de l'ARN génomique en ADN double brin.



 **Modifications post-traductionnelles : La protéine d'enveloppe subit des modifications post-traductionnelles dans la lumière du réticulum endoplasmique où elle est transloquée grâce à son peptide signal. Elle est glycosylée, elle se polymérise, grâce à l'intervention de protéines chaperonnes cellulaires. La glycosylation continue dans l'appareil de Golgi où l'enveloppe sera coupée au consensus R-X-K/R-R en deux sous-unités SU et TM.

c2) Assemblage des particules virales

Les particules se forment sous la face interne de la membrane plasmique où les protéines Gag et Pol sont transportées après leur synthèse. C'est la protéine Gag qui dirige et contrôle le processus d'assemblage, elle est même appelée "machine à faire les particules". Au moment de l'assemblage mais aussi après le bourgeonnement, les précurseurs Gag, Gag-Pro-Pol sont clivés dans la particule pour donner les sous-unités décrites précédemment.

c3) Etape extracellulaire

**La sortie de la particule : Le bourgeonnement décrit la phase de sortie des particules virales de la cellule. La particule sort en se recouvrant de la membrane plasmique, probablement par l'intermédiaire de la reconnaissance de la sous-unité MA de Gag et TM de l'enveloppe. En microscopie électronique, on observe effectivement des bourgeonnements de la membrane de la taille des particules produites. Le mécanisme de bourgeonnement n'est pas observé pour tous les virus à ARN rétrotranscrit.

**L'infection : Les particules ainsi produites peuvent infecter une nouvelle cellule. Pour cela, la sous-unité SU de la protéine d'enveloppe se fixe sur un récepteur cellulaire. Puis les membranes fusionnent grâce au domaine hydrophobe de la sous-unité TM. Le virion est alors libéré dans le cytoplasme.

Certains rétrovirus, sensibles au pH, pénètrent dans la cellule après fixation au récepteur par un mécanisme d'endocytose avec l'enveloppe. Dans ce cas, la membrane virale fusionne avec celle de l'endosome avant le relargage du virion.

Cette étape n'est évidemment pas possible pour les éléments dépourvus de gène d'enveloppe qui passent directement à l'étape suivante, la rétrotranscription, dans la cellule où ils ont été produits. Il a été montré que les particules rétrovirales non encore enveloppées peuvent aussi avoir un destin intracellulaire et s'insérer dans leur cellule d'origine (Tchénié *et al.*, 1991). Ce phénomène a été observé plus

récemment pour le foamy virus, suggérant que ce cycle entièrement intracellulaire peut-être utilisé par de nombreux rétrovirus (Linial, 1999).

c4) La réplication de l'ARN génomique

Le processus de rétrotranscription débute dès l'entrée de la particule virale dans le cytoplasme de la cellule infectée. L'ARN génomique (brin positif) est présent en double exemplaire dans le virion ce qui est indispensable pour que la rétrotranscription reconstitue un provirus complet, c'est à dire un ADN double brin avec un LTR complet à chaque extrémité.

La synthèse de l'ADNc, brin négatif, commence par le LTR5'. Elle est amorcée en effet par un ARN de transfert spécifique complémentaire sur un vingtaine de bases à la région PBS. La rétrotranscription aboutit à la synthèse des régions R et U5 présentes dans le LTR5'. Ce court ADNc est appelé "strong stop strand-minus DNA" (ADNsss).

L'activité RNAase H dégrade l'ARN génomique hybridé avec l'ADNsss. L'ADNsss est alors transféré au LTR 3' dont il est complémentaire au niveau de la région R. Ce transfert fait intervenir la sous-unité NC. La rétrotranscription permet la synthèse du brin "moins" total, jusqu'à une partie au moins du PBS, permettant par hybridation la circularisation du brin "moins". L'ARN génomique est totalement dégradé, sauf la région PPT qui servira d'amorce à la synthèse du brin positif.

La synthèse du brin plus est amorcée par la région PPT non encore dégradée de l'ARN génomique, à la limite de la région U3. La totalité du LTR est copié à partir du brin moins. Cette partie du brin plus nouvellement synthétisé est appelée "strong stop plus DNA". L'ARNt est alors dégradé, libérant la région PBS du brin moins. La circularisation du brin moins au niveau du PBS permet le transfert de l'ADNsss plus au niveau du PBS et l'achèvement de la synthèse du brin plus.

Cette étape complète aboutit théoriquement à la production d'une copie linéaire double brin extra chromosomique du rétrovirus, susceptible de s'intégrer lors de l'étape suivante (figure 19B). En fait, dans le cytoplasme, on trouve des molécules circulaires avec un ou deux LTR et des molécules linéaires avec deux LTR. Seules les molécules possédant deux LTR sont infectieuses. On ignore comment se circularisent les provirus.

c5) Intégration du provirus dans le génome de l'hôte.

L'ADN proviral parvient dans le noyau sous la forme d'un complexe nucléoprotéique appelé complexe de préintégration (PIC= preintegration complexe). Dans le cas du *HIV-1*, il est constitué de l'ADN proviral, de l'intégrase, de la matrice, de la nucléocapside, de la reverse transcriptase et d'une protéine accessoire. De tels complexes ont été isolés à partir de cellules infectées et sont suffisants pour catalyser une intégration du provirus dans une séquence cible *in vitro*. Cependant, l'intégrase du *HIV-1* est stimulée par la protéine cellulaire *in-1* *in vitro* et d'autres facteurs semblent augmenter l'efficacité de l'intégration. *In vivo*, des facteurs cellulaires semblent requis pour l'intégration, comme la protéine HMG I(Y), requise par *HIV-1* (Farnet & Bushman, 1997) et *MoMuLV* (Li *et al.*, 1998).

L'intégrase catalyse l'insertion du provirus qui se trouve sous forme linéaire, et non circulaire. Deux nucléotides sont enlevés à l'extrémité 3' de chaque LTR ce qui génère deux extrémités 3' hydroxyles. Ces extrémités attaquent l'ADN cible sur les liaisons phosphodiester et participent à une réaction de transestérification qui aboutit à la liaison entre la cible et l'extrémité 3' de chaque brin proviral (Farnet & Bushman, 1997). Les jonctions manquantes sont ensuite réparées par synthèse d'ADN, mais on ignore le mécanisme et l'origine des protéines intervenant dans cette étape.

III- Relations entre les éléments transposables et leurs hôtes

Les éléments transposables (ET) ont la propriété de s'insérer dans les chromosomes de leur hôte à de nouvelles localisations. Ces insertions sont des événements mutagènes qui peuvent être délétères, neutres, ou avantageux pour l'hôte. Inversement, le site d'insertion a un impact sur l'expression de l'élément nouvellement intégré et, à long terme, sur la persistance de cet élément au site considéré. Hôtes et ET exercent donc l'un sur l'autre des contraintes auxquelles ils se sont adaptés au cours de l'évolution, avec, comme contrainte supplémentaire pour l'élément, le fait que son maintien dans l'hôte soit entièrement dépendant de la survie de ce dernier. Aujourd'hui, nous observons des situations d'équilibre dans lesquelles l'élément et l'hôte ont développé des stratégies visant à limiter l'impact délétère des ET, et "cohabitent" dans une relation de type hôte-parasite, évoquant parfois une relation symbiotique.

III-I Les différents éléments transposables dans les génomes d'aujourd'hui

Nous avons vu dans les chapitres précédents que l'on pouvait classer les éléments transposables selon leurs cycles et leurs structures. Mais quand on s'intéresse à leur impact sur le génome de l'hôte et aux relations qui se sont établies, il est plus judicieux de les classer selon leur ancienneté dans l'espèce, leurs représentation et distribution dans l'hôte ou encore leur taux d'activité.

Dans les génomes des espèces actuelles, on détecte à la fois des insertions dont le site est polymorphe dans les populations, témoignant d'une mobilisation récente (postérieure à l'apparition de l'espèce), et les vestiges d'insertions dont la localisation est conservée dans l'espèce et parfois dans des espèces proches. Les premières sont souvent identifiées à cause de leurs effets délétères et des pathologies qu'elles entraînent, et ne seront probablement pas maintenues dans l'espèce. Les secondes dérivent d'insertions anciennes qui ont eu un impact neutre ou avantageux sur leur hôte. Dans ce dernier cas, la fonction conférant un avantage a été sélectionnée au cours de l'évolution.

Comment s'est établi cet équilibre? Il existe des situations de dérégulation ou des cas de transferts horizontaux d'éléments nouveaux pour l'espèce, dans lesquelles des éléments sont fortement amplifiés. De telles situations permettent l'étude des relations hôte-éléments au début d'une invasion.

A) Les insertions fixées dans le génome

Les "insertions fixées" désignent les éléments ou leur "signature" dans un site chromosomique donné chez tous les individus d'espèces proches ou d'une espèce, voire d'une population. Certaines analogies de mécanismes ou de séquences protéiques suggèrent pour certains gènes cellulaires un ancêtre commun avec des éléments transposables. Même si l'origine ET n'est pas toujours démontrée, j'évoquerai ici les exemples se rapportant à de tels cas.

1) Domestication de régions codantes

Le mot domestication moléculaire a été proposé par Miller *et al.*, pour désigner la conversion d'une séquence codante dérivée de l'élément *P* en un gène intégré de l'hôte (1992; 1995). Il désigne aujourd'hui l'utilisation par l'hôte d'une fonction quelconque d'un élément transposable, qu'elle soit une activité protéique, ou une séquence nucléique ayant un effet depuis son site d'insertion. Les régions

domestiquées sont souvent les seules traces persistantes de l'élément, le reste n'étant plus soumis à pression de sélection.

On découvre des cas de domestication (souvent par hasard) quand on identifie dans une espèce ou des espèces proches une région conservée et active provenant d'un élément défectif, mais on ignore le plus souvent le rôle dans la cellule de la fonction domestiquée.

- Les néogènes de l'élément *P*

La première description d'une séquence codante dérivée d'un élément transposable en gène intégré de l'hôte concernait une protéine dérivée de l'élément *P* dans trois espèces de drosophile du groupe *Drosophila obscura* (Paricio *et al.*, 1991; Miller *et al.*, 1992). Les espèces de drosophiles sont devenues un modèle de choix pour rechercher les fonctions domestiquées dérivées de l'élément *P*.

Dans le sous-groupe de drosophiles *obscura*, chez *D. guanche*, *D. madeirensis* et *D. subobscura*, on trouve des éléments *P* tronqués en 5' et 3' qui n'ont conservé que la région codant le répresseur 66kDa (figure 8) et donc appelé "repressor-like". Ces éléments sont organisés en clusters de 10 à 50 copies au même site 85C, ce qui indique leur fixation ancienne dans le groupe du transposon ancestral, estimé à au moins quatre millions d'années (Kobayashi *et al.*, 1993). Chaque copie a la capacité de coder le répresseur-like qui présente 70% de similitude avec le répresseur 66kDa trouvé chez *D. melanogaster*. Les trois domaines fonctionnels de la protéine sont conservés. Si l'on fait l'hypothèse que le répresseur a conduit à la dégénérescence de cet ancêtre dans le génome, il est difficile d'expliquer que ce répresseur soit encore sous pression de sélection, d'où l'hypothèse qu'il a une autre fonction, indépendante de l'élément *P* (Miller *et al.*, 1992).

Un autre exemple de répresseur-like fixé a été identifié chez *D. tsacasi* du groupe *montium* (Nouaud & Anxolabéhère, 1997). Là encore, la fonction domestiquée a probablement un autre rôle que celui de répresseur, car il n'y a plus de *P* actif dans cette espèce. Les fonctions capturées dans les deux groupes *obscura* et *montium*, bien que dérivant chacune du répresseur de *P*, sont peut-être différentes dans chacun de leur hôte car elles ne présentent entre elles qu'une similitude de 69% et n'ont pas le même promoteur (Anxolabéhère, Nouaud & Miller, 2000).

- Les protéines d'enveloppe des rétrovirus ont-elles un rôle physiologique chez leur hôte?

Les protéines d'enveloppe des rétrovirus ont des propriétés de liaison à des récepteurs cellulaires et de fusion membranaire qui sont nécessaires à l'entrée du virus dans la cellule. Or, on détecte la présence de nombreuses protéines d'enveloppes rétrovirales dans les placentas de mammifères, tissus où se produisent des fusions cellulaires (Larsson, Andersson & Nilsson, 1994; Venables *et al.*, 1995). L'idée que les propriétés de fusion membranaire pourraient médier la fusion trophoblastique est donc ancienne (Boyd *et al.*, 1993).

En 1999, Blond *et al.*, identifiaient une protéine homologue d'une protéine Env d'un *HERV-W* à l'issue d'un crible de protéines sécrétées par les cellules trophoblastiques. Cette protéine est fortement exprimée dans la couche des syncytiotrophoblastes. L'année suivante, Mi *et al.* transfectaient cette enveloppe dans différents types cellulaires, ce qui provoquait la formation de *syncytia* géants, d'où le nom de *Syncytin* donné à cette protéine. La première équipe réalisait le même type d'expériences, mais montrait que les fusions impliquaient une interaction avec un récepteur de rétrovirus de mammifères de type D (Blond *et al.*, 2000). L'ensemble de ces expériences suggère fortement un rôle physiologique de la *Syncytin* dans la formation du placenta chez l'homme.

L'homologue de l'élément *HERV-W* est absent chez la souris. Mais il existe des différences notables dans la placentation des rongeurs. D'autres part, il est concevable que d'autres *ERV* de rongeurs jouent le rôle de *HERV-W*.

- Les protéines provirales interagissent avec les rétrovirus exogènes

A la différence de la *Syncytin*, il semble que certaines protéines *Env* aient été domestiquées pour leur capacité de compétition, plutôt que pour leur fonction propre. La présence de ces protéines dans le génome confère en effet une résistance à la surinfection par des virus exogènes utilisant les mêmes récepteurs. Ceci a été observé chez le poulet (Payne *et al.*, 1971) et la souris où deux *loci* ont été décrits, *Rmcf* et *Fv4* (Buller *et al.*, 1987; Limjoco *et al.*, 1993). *Fv4*, malgré une arrivée ancienne dans le génome, n'est pas fixé dans toutes les populations sauvages, suggérant une interférence négative avec la fonction normale du récepteur, et une pression de sélection sur les populations en contact avec le virus (Boeke & Stoye, 1997).

Le locus *Fv1* confère une résistance au rétrovirus *MLV* chez la souris (Lilly & Pincus, 1973). Cette protéine est un néogène issu de l'assemblage de plusieurs éléments mobiles, qui présente 60% d'homologie avec les domaines MA et CA du *HERV-L* (Best *et al.*, 1996). La protéine *Fv1* interagit avec la capsid du complexe de préintégration du virus, et donc interfère avec lui bien après l'infection. L'activité de ce gène n'a été détectée que dans des lignées de souris contenant des *MLV* endogènes, suggérant une évolution en réponse à la présence de ce virus. C'est aussi un exemple de domestication dans une population soumise à une pression de sélection.

D'une façon générale, les fonctions résiduelles de rétrovirus endogènes interfèrent avec l'infection de rétrovirus exogènes de même famille: blocage du cycle rétroviral comme on l'a vu, mais aussi modulation de la réponse immunitaire aux virus exogènes, recombinaisons ou complémentations pouvant conduire à la formation de nouveaux virus ou à des pseudotypes. Toutes ces interactions ne sont pas toujours au bénéfice de l'hôte! Ajoutons que le risque de générer des pseudotypes est un frein au développement des techniques de xénogreffes ou de thérapie génique utilisant des vecteurs rétroviraux.

2) Domestication de séquences régulatrices

Toutes les séquences des ET sont porteuses d'une fonction susceptible d'être utilisées par l'hôte, et les domestications de fonctions protéiques, plus spectaculaires, sont en fait des exceptions parmi les cas de domestications. Parfois, l'insertion initiale est neutre, mais des modifications ultérieures lui permettent d'apporter une nouvelle fonction comme dans l'exemple de la création d'introns par les *Alu* (Makalowski *et al.*, 1994).

- Séquences contrôlant la transcription de gènes cellulaires voisins

Les éléments transposables sont connus pour contenir des *enhancers*, des *silencers* et des sites d'attachement pour des facteurs de transcription (voir le chapitre II sur la structure des éléments). Ces séquences régulatrices sont nécessaires à la biologie de l'élément, mais elles peuvent aussi modifier la régulation de gènes cellulaires voisins. Une vingtaine d'exemples de capture de telles séquences est passée en revue dans le génome humain (Deragon & Capy, 2000).

Le premier exemple de capture d'un *enhancer* décrit concerne le gène de souris *Sex-limited protein (Slp)*: le signal responsable de sa régulation par les

androgènes se situe dans le LTR d'un élément proviral (Stavenhagen et Robins, 1988; Adler *et al.*, 1991). Les séquences *Alu*, particulièrement riches en sites potentiels de facteurs de transcription, peuvent générer de nouveaux *enhancers* (Norris *et al.*, 1995; Babich *et al.*, 1999).

- Séquences transcrites avec le gène cible

On trouve la trace d'insertions d'éléments, complets ou non, dans les parties transcrites de gènes, y compris dans les régions codantes. Ces séquences transcrites sont avantageuses ou neutres. Les éléments *SINEs* peuvent apporter de nouveaux signaux de polyadénylation, qui peuvent modifier la région 3' non traduite du gène cible (chez le lapin, Krane et Hardison, 1990), ou qui seront utilisés de façon alternative comme pour le gène codant pour une γ -phosphorylase kinase du muscle chez la souris (Maichele, Farwell et Chamberlain, 1993).

Des auteurs ont quantifié la présence d'éléments *Alu* dans les ARN messagers humains en les recherchant dans les banques de données d'ADNc. Les *Alu* sont détectés dans 5% des messagers humains, localisés dans les régions 5' et 3' non traduites, ainsi que dans les introns (Yulug, Yulug and Fisher, 1995).

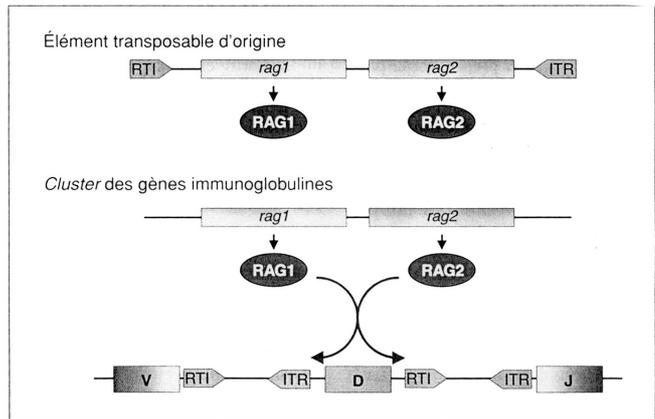
3) Maintien de la fonction ou du mécanisme de transposition

- Système immunitaire des vertébrés à mâchoire

Le système immunitaire de certains vertébrés exploite le potentiel de diversité pouvant être induit par les éléments transposables. Le développement des cellules souches de l'immunité passe par des étapes de recombinaisons entre les domaines *VDJ* des gènes codant pour les anticorps et aboutit aux milliers de clones différents capables de reconnaître chacun spécifiquement un épitope antigénique. Le mécanisme de ces recombinaisons est tout à fait homologue à celui des éléments de classe II *TcI* et fait intervenir des recombinases et de l'ADN mobile (Agrawal, Eastman & Schatz, 1998). Il implique deux protéines, *RAG1* et *RAG2*, qui reconnaissent spécifiquement des signaux nucléotidiques au voisinage des segments *V*, *D*, et *J* et peuvent cliver ces segments au voisinage de ces signaux (figure 20).

Origine du système immunitaire des mammifères par cooptation de l'activité transposase d'un élément transposable. Les éléments transposables à l'origine de la recombinaison V(D)J devaient posséder des répétitions terminales inversées (RTI) comme ceux des transposons de type II, et deux gènes (rag1 et rag2) codant les deux propriétés essentielles d'une transposase: la reconnaissance des sites de coupures et l'activité endonucléasique. Selon le modèle, les recombinases RAG1 et RAG2 et les motifs répétés du système V(D)J dérivent de ces anciens éléments transposables.

Figure 20 : Mécanisme de recombinaison des gènes V,D,J avec les protéines RAG1 et RAG2. (Anxolabéhère, Nouaud & Miller, 2000).



- Télomères

Les télomères sont les extrémités des chromosomes des eucaryotes. Le mécanisme de réplication de l'ADN nécessite des amorces complémentaires du brin matrice. Les télomérases sont les enzymes qui préviennent les pertes d'information des extrémités chromosomiques des eucaryotes en synthétisant des amorces. Ces enzymes sont des reverse transcriptases de la même famille que celle des rétrotransposons (figure 3). Elles synthétisent de courtes répétitions à l'extrémité des chromosomes à partir de la matrice ARN qui leur est associée. La sous-unité catalytique de la télomérase présente de grandes similarités avec l'ORF2 des *LINES*, suggérant la capture ancienne d'un rétrotransposon pour une fonction cellulaire essentielle. On ignore si les rétrotransposons sont bien à l'origine des télomérases mais l'arbre phylogénétique des rétrotranscriptases indique que les rétrotransposons sont antérieurs à l'apparition des eucaryotes et de leurs télomérases (Malik & Eickbusch, 1999; Nakamura & Cech, 1998).

Chez *Drosophila melanogaster*, on ne trouve pas de télomérase. La domestication du mécanisme de transposition est encore plus remarquable puisque les éléments eux-mêmes ont été conservés. Les télomères de la drosophile sont constitués par les rétrotransposons *HeT-A* et *TART*. Seuls les *TART* codent pour leur propre rétrotranscriptase. La transposition de ces éléments, qui ne se produit qu'à l'extrémité des chromosomes, assure le maintien de l'intégrité des chromosomes de drosophiles. En fait, tous les auteurs ne s'accordent pas sur l'origine de ces deux éléments : certains pensent que ce sont d'anciens rétrotransposons domestiqués et contrôlés par la machinerie cellulaire (Biessmann & Masson, 1992), d'autres que d'anciens gènes de l'hôte de type télomérase auraient été transformés en ADN mobile (Pardue *et al.*, 1996). Mais toutes ces protéines ont un ancêtre commun qui pourrait être un rétroélément.



- Centromères

La protéine centromérique *CENP-B* est une protéine très conservée, dérivant d'une *transposase* de type *pogo* et se liant à un motif appelé *CENP-B box*. On trouve ce motif dans les séquences satellites centromériques. La protéine *CENP-B* a une forte homologie avec la *transposase* de l'élément *Tigger*, et le motif correspond aux régions inversées de cet élément. La protéine *CENP-B* permettrait des recombinaisons homologues et participerait à l'homogénéisation des répétitions de satellites, ce qui est une caractéristique de l'évolution des répétitions centromériques. Cela montre la participation à l'évolution des centromères d'une protéine de même famille que les transposases des ET (Kipling & Warburton, 1997).

Les éléments fixés dans le génome ont été sélectionnés au cours de l'évolution pour leurs effets bénéfiques. Les premières captures décrites étaient considérées comme opportunistes, car les ET n'étaient encore (re)connus que pour leurs effets mutagènes. Mais le nombre croissant de descriptions accrédite l'idée que les domestications sont un mécanisme plus général d'évolution. D'autre part, une meilleure connaissance des fonctions et du rôle des régions hétérochromatiques comme les centromères et les télomères met à jour l'imbrication entre les génomes eucaryotes et les éléments transposables, où la relation hôte-parasite semble symbiotique.

B) Les insertions polymorphes du génome

Les éléments transposables ont été découverts à cause de leur pouvoir mutagène qui s'observait sur la couleur de grains de maïs (Mac Clintock, 1955). Les ET sont un agent mutagène important. On estime par exemple qu'ils sont responsables de 80% des mutations spontanées observées chez la drosophile (Capy *et al.*, 1997). Ils sont potentiellement mutagènes par leurs propriétés de néoinsertion: ils peuvent rompre la phase codante d'un gène, modifier sa régulation spatio-temporelle, perturber l'épissage ou créer de nouveaux sites... Les effets mutagènes possibles sont nombreux et entraînent parfois des pathologies.

1) Les insertions polymorphes ont été d'abord identifiées pour leurs effets délétères

- Chez l'homme

C'est en 1988 que Kazazian *et al.* mettaient en évidence pour la première fois l'implication d'un événement récent de rétrotransposition dans une pathologie

humaine. Dans le gène codant pour le facteur VIII, ils ont identifié deux insertions d'éléments *LI* tronqués chez 2 patients sans lien de parenté atteints d'hémophilie A. Ces insertions sont responsables de la maladie.

Depuis, six autres cas de néoinsertion de l'élément *LI* entraînant une pathologie ont été décrits (pour revue, Kazazian et Moran, 1998). Ces insertions se sont produites dans les cellules germinales ou au cours de l'embryogenèse précoce, sauf dans un cas de rétrotransposition somatique, à l'origine d'un cancer du colon: le génome des cellules de la tumeur possédait une insertion *LI* dans le gène *APC* (Adenomatous Polyposis Coli) au contraire des autres tissus du patient (Miki *et al.*, 1992).

L'autre élément dont l'insertion est responsable de pathologies dans l'espèce humaine est un *SINE*, l'élément *Alu*. Deininger et Batzer estiment à 0,1% les désordres génétiques générés par des insertions *Alu* (1999). Le plus souvent, la cible décrite se situe sur le chromosome X. Ca ne signifie pas que ce chromosome soit un point chaud d'insertion mais plutôt que l'X est un "chromosome chaud" de détection à cause de sa situation d'hémizygote chez le mâle.

En 1997, on comptait au total 19 événements de rétrotransposition responsables d'une pathologie chez l'homme sur les 12700 mutations répertoriées (11 dus à *Alu* et 8 à *LI*, donc des rétroposons uniquement) (Kazazian, 1998). Cinq autres insertions *Alu* entraînant une pathologie ont été décrites en 1999. Le nombre total est probablement sous-estimé: d'une part, la technique pour les rechercher utilise la PCR qui est limitée par la taille des insertions. D'autre part, tous les événements génétiques qui pourraient être à l'origine de tumeurs ou d'avortements spontanés précoces ne sont pas systématiquement recherchés. Or, au moins un cas de mutation somatique a été montré pour un *Alu*, et on sait que les ET s'expriment au cours de l'embryogenèse précoce, où leur mobilisation pourrait entraîner des mutations dominantes létales. Leurs implications réelles sont donc probablement sous-estimées.

- Chez la drosophile

Chez la drosophile, les éléments transposables sont mobilisés à des taux supérieurs à ceux des vertébrés, et leur pouvoir mutagène a été, est encore, largement exploité dans les laboratoires, pour créer des mutants d'intérêt. La drosophile est donc devenue un organisme modèle dans l'étude des effets mutagènes des éléments transposables. Tous les cas de mutation insertionnelle par un élément transposable ont été observés. Parmi ceux observés au laboratoire,

citons la perturbation de la transcription du gène cible par l'apport d'un *enhancer* ou par compétition avec le promoteur de l'élément (Conte *et al.*, 2000; Conte, Dastugue & Vaury, 2002 ; Conte, Dastugue & Vaury, ¹soumis) et la perturbation de l'épissage (Lajoinie *et al.*, 1995). De nouvelles mutations engendrées par les ET peuvent être encore régulièrement observées dans cet organisme et entretiennent l'idée que les transpositions n'ont que des effets délétères.

Les insertions fixées dans le génome, examinées précédemment, n'ont-elles que des effets bénéfiques fortuits? Les éléments transposables ont-ils majoritairement des effets délétères? Des études plus générales sur le polymorphisme des populations et leurs conséquences sur le *fitness* de l'hôte viennent bouleverser certaines idées reçues et répondre partiellement à ces questions. La démarche dans ces études ne consiste plus à rechercher les insertions causes d'une pathologie, mais à comparer les conséquences d'un polymorphisme d'insertions entre plusieurs populations.

2) Polymorphisme des insertions et potentiel adaptatif de l'hôte

a) Polymorphisme des provirus endogènes chez la souris

La cartographie génomique de nombreuses souches naturelles de souris a été réalisée pour différents provirus endogènes (Frankel *et al.*, 1990). La distribution provirale est très diversifiée dans cette espèce, où l'on observe un polymorphisme important entre les souches. Très peu, sinon aucun provirus n'est commun à l'ensemble des souches, montrant une colonisation de la lignée germinale postérieure à la spéciation. Chaque souche possède ses propres insertions fixées, au point que ces dernières permettent l'identification et l'origine de populations murines issues de différents croisements au laboratoire (Naggert *et al.*, 1995). Ce polymorphisme reflète l'adaptation des populations aux différents virus exogènes qu'elles ont rencontrées (voir l'exemple de *Fv1* et *Fv4* dans le chapitre III-I-A sur les insertions fixées).

En dehors de tels exemples de protection aux virus exogènes, les éléments transposables contribuent-ils au potentiel adaptatif des cellules? Si les ET ont un rôle positif dans l'évolution, on attend que leur transposition ou les réarrangements génomiques qu'ils induisent apportent une amélioration visible sur leur hôte. Le *fitness* représente les effets combinés de toutes les propriétés phénotypiques d'un génotype donné sur sa capacité de survie et de reproduction dans un

¹ Conte C., Dastugue B. and Vaury C. "Promoter competition as a mechanism of transcriptional interference mediated by retrotransposons."

environnement particulier (Lenski, 1991). Pour évaluer d'éventuelles améliorations, on étudiera donc les conséquences des insertions sur le *fitness* de l'hôte.

b) Impact du polymorphisme d'insertions sur le *fitness* de l'hôte

- Chez *Escherichia coli*

Les études visant à mesurer les conséquences du polymorphisme sur le *fitness* de l'hôte sont menées sur des micro-organismes: leur croissance rapide permet de suivre des populations sur des milliers de générations, et leur congélation permet de comparer des individus de générations différentes dans des conditions variables.

Les génomes de populations issues d'un même ancêtre *Escherichia coli* ont été comparés par la technique RFLP (Restriction Fragment Length Polymorphism) en utilisant pour sondes les *IS* (*Insertion Sequences*), éléments procaryotes de classe II. La première étude analysait la variation génomique dans une culture "stab" conservée pendant 30 ans. En moyenne, les clones obtenus différaient de leur ancêtre putatif par une douzaine de modifications (Naas *et al.*, 1994; Naas *et al.*, 1995). Mais l'expérience ne permettait pas de savoir si ces conditions de culture avaient provoqué ou maintenu cette diversité.

L'étude suivante a porté sur l'évolution de populations d'*E. coli* issues d'un seul ancêtre, cultivées pendant 10 000 générations dans un milieu minimum et repiquées tous les 1,5 jours. La congélation de prélèvements dans ces populations à intervalles réguliers a permis la comparaison de différents paramètres pour évaluer les variations du *fitness* et du génome. L'évolution phénotypique est assez rapide pendant les 2000 premières générations puis se ralentit. En moyenne, le *fitness* est amélioré de 50% par rapport à l'ancêtre (Lenski *et al.*, 1991; Lenski Travisano, 1994). L'analyse moléculaire des génomes corrèle ces résultats avec une divergence génétique qui augmente au fil des générations. Mais surtout, parmi les mutations observées, certaines sont fixées, c'est à dire présentes dans toute la descendance d'une population (mutations "pivots"). Ces résultats suggèrent que de telles mutations sont responsables d'un bénéfice pour l'hôte (Papadopoulos *et al.*, 1999).

Sur l'ensemble des mutations analysées, les *IS* sont majoritairement impliquées. Les mutations pivots sont toutes dues à des transpositions ou à des recombinaisons entre ces éléments. L'analyse de différents *loci* affectés par des

mutations pivots permet d'extrapoler sur l'avantage sélectif attendu au regard de la mutation et des conditions de culture (Schneider *et al.*, 2000). Toutefois, pour être totalement convaincants, ces résultats devront être confirmés par des expériences de compétition entre des souches isogéniques ne différant que par une mutation donnée.

- Chez *Saccharomyces cerevisiae*

Les résultats chez la levure sont un peu plus contrastés que chez *E. coli*. Les souches de *Saccharomyces cerevisiae* cultivées contiennent 25-35 copies du *pseudoviridae Ty1*, éléments plutôt stables au laboratoire comme en industrie. Au contraire, les populations naturelles sont très hétérogènes pour leur nombre de copies avec une moyenne par génome plus basse. Bien que l'impact moyen des transpositions de *Ty1* soit délétère, des populations initiées avec un clone contenant un *Ty1* unique accumulent des éléments en 1000 générations (Wilke, Maimer & Adams, 1992).

Les mêmes auteurs ont réalisé des compétitions entre différentes populations pour tester si cette augmentation du nombre de copies était avantageuse. Le *fitness* a été évalué selon le rendement et le taux de croissance des cultures, et rapporté au nombre de copies de *Ty1* dans chaque population. En partant de populations très hétérogènes quant à leur nombre de copies *Ty1*, ils observent des effets négatifs sur le *fitness*: la densité cellulaire est plus faible en phase stationnaire et corrélée à un nombre de *Ty1* plus élevé. Pourtant, après 100 générations, les clones sans *Ty1* sont pratiquement indétectables, et les clones contenant des mutations pivots prédominent. Malgré un effet paraissant négatif sur le *fitness*, des événements de transposition de *Ty1* ont généré des mutations adaptatives qui prédominent dans la population (Wilke & Adams, 1992).

- Chez la drosophile

L'espèce *Drosophila melanogaster* a subi l'invasion d'un nouvel élément il y a une cinquantaine d'années, après un transfert horizontal (chapitre III-II). Il s'agit d'un élément de classe II, l'élément *P*. Malgré les effets délétères des insertions de *P*, toutes les populations naturelles actuelles de *D. melanogaster* possèdent cet élément dans leur génome. La situation peut sembler paradoxale, car toutes les souches naturelles ont sélectionné diverses stratégies pour réprimer *P*, reflétant le caractère délétère de *P* contre lequel elles se sont défendues, mais aucune population sans *P* n'a survécu, suggérant un avantage aux nouvelles insertions de *P*. En fait, cette invasion spectaculaire du génome *D. melanogaster* par l'élément *P*

est attribuée à la capacité de transposition de l'élément *P* lui-même. *P* transpose de façon non répliquative, laissant une coupure ADN double brin à son point de départ, coupure qui sera réparée par la cellule. La réparation utilise souvent la chromatide sœur qui possède le matériel génétique homologue à la chromatide cassée. L'élément *P* de la chromatide sœur est alors recopié sur le site vide. Ce gain net d'une copie de *P* à chaque transposition est un mécanisme d'invasion très efficace à travers une population que la sélection naturelle est incapable de prévenir. Dans les petites populations, l'invasion rapide des éléments *P* conduit à l'extinction du stock, mais dans les grandes populations, il y a souvent un échappement, probablement grâce à un mécanisme contrôlant négativement l'activité de transposition de *P* (Preston & Engels, 1989).

De façon générale, il semble que le génome de *D. melanogaster* ait augmenté en taille dans les régions du monde récemment colonisées par cette espèce. L'augmentation de la taille du génome est corrélée à l'augmentation du nombre de copies moyen d'ET euchromatiques. L'accroissement de la taille du génome refléterait les adaptations nécessaires à un nouvel environnement, et serait dû à de nouvelles insertions d'éléments transposables (Vieira-Heddi *et al.*, 2001)

c) Polymorphisme et réponse aux stress cellulaires

L'expression des éléments transposables est très souvent augmentée à la suite d'un stress cellulaire (modification brutale de l'environnement). C'est un constat qui est fait pour de nombreuses familles d'éléments et dans de nombreuses espèces. La question est de savoir si c'est un avantage pour l'élément ou pour l'hôte ou bien si chacun y trouve son compte: l'élément profiterait de remaniements protéiques dans la cellule et de redistributions de protéines liant l'ADN pour échapper à une répression; l'hôte pourrait trouver un avantage à subir de nouvelles insertions pour s'adapter.

Les populations de l'ancêtre sauvage de l'orge, *Hordeum spontaneum*, sont réparties dans différents habitats. Le génome de cette plante abrite une famille de rétrotransposons actifs en plusieurs milliers d'exemplaires. Le nombre de copies de *BARE-1* est corrélé positivement avec la taille du génome et l'aridité du climat, semblant indiquer un rôle de la mobilisation de cet élément dans l'adaptation des populations à l'aridité (Vicient *et al.*, 1999). Ceci est cohérent avec la présence d'éléments de réponse à l'acide abscissique dans le promoteur de l'élément *BARE-1* qui sont typiques des gènes de réponse à la sécheresse (Manninen & Schulman, 1993).

La même équipe a voulu savoir si l'élément participait effectivement à la diversification des individus. Elle a exploité pour cela une situation géographique particulière: un canyon bénéficiant d'un microclimat donné, dans lequel l'ensoleillement et l'aridité varient fortement selon l'altitude. Le génome des individus récoltés le long d'une coupe nord-sud dans ce canyon a été analysé: Sur la face sud, la plus aride, le nombre de copies de pleine taille est plus grand que dans les populations du nord, et la perte d'éléments est plus faible. Ces données suggèrent une adaptation sélective par l'augmentation de la taille du génome *via* un rétrotransposon (Kalendar *et al.*, 2000).

C) Les éléments transposables sont plus ou moins répétés dans le génome

Les ET sont généralement présents en plusieurs exemplaires dans le génome (de quelques copies à plus d'un million), ce qui leur confère des propriétés communes avec les séquences répétées. Un élément très répété sera plus fréquemment impliqué dans des phénomènes de recombinaison illégitime, c'est à dire entre deux localisations chromosomiques non alléliques. De tels événements conduisent à des duplications ou à des délétions.

Les séquences répétées ont un rôle dans la structure des chromosomes, et les éléments transposables ont parfois été sélectionnés en tant que répétition pour cette fonction. C'est grâce à ces propriétés de répétition que l'on trouve le plus d'exemples illustrant le rôle des ET dans l'évolution des chromosomes de l'hôte et la thèse d'une relation symbiotique.

1) Effets délétères des recombinaisons illégitimes

- Implication des *SINEs*

Les *SINEs* sont de bons candidats pour la recombinaison homologe de par leur importante représentation (un million et demi de *SINEs* par génome haploïde humain). Les premiers cas de pathologie entraînés par de tels événements ont été rapportés pour les gènes codant pour la *globine* et le *récepteur LDL*, dont les délétions sont responsables respectivement de la thalassémie et de l'hypercholestérolémie. Les deux gènes contiennent des séquences *Alu* dans leurs introns et leur région 3' non traduite, aux sites de rupture des délétions (Ottolenghi *et al.*, 1982; Jagadeeswaran *et al.*, 1982; Lehrman *et al.*, 1986).

La recombinaison est favorisée par la présence d'une séquence de 26 pb très conservée, considérée comme un point chaud de recombinaison. Actuellement,

treinte-trois maladies génétiques et seize cas de cancers sont attribués à ce type de recombinaisons inégales entre des éléments *Alu*, soit une participation de 0,3% aux désordres génétiques chez l'homme! C'est trois fois plus que les pathologies engendrées par de nouvelles insertions du même élément *Alu*.

- Implication d'un élément de classe II

En identifiant un nouvel élément de type *mariner* chez l'homme, qu'ils appellent *MITE* pour *mariner insect transposon-like element*, Reiter *et al.* (1996) ont montré pour la première fois la présence d'un élément de classe II à proximité d'un point chaud de recombinaison dans deux pathologies humaines: Charcot-Marie-Tooth de type 1A (CMT1A) et une autre neuropathie, Hereditary Neuropathy Liability to Pressure Palsies (HNPP). Les mutations dans ces deux neuropathies sont le résultat de réarrangements d'ADN qui sont les produits réciproques de recombinaisons inégales sur le chromosome 17 (Chance *et al.*, 1994).

Les auteurs ne s'arrêtent pas à l'implication éventuelle de ce *MITE* dans l'origine de la recombinaison homologue au site incriminé. Ils avancent l'hypothèse qu'une *transposase* de type *mariner* serait responsable de l'événement de recombinaison lui-même. Ils s'appuient sur les résultats expérimentaux concernant l'activité de recombinaison homologue de la *transposase* de l'élément *P*, sur des données de séquences montrant la présence d'ADNc homologues à 80% avec les *transposases* d'insecte dans les gamètes mâles où s'observent justement les recombinaisons illégitimes. L'identification d'une *transposase* active manque à ces travaux pour valider cette hypothèse, or, cinq ans après leur publication, l'éventuelle *transposase* joue toujours les arlésiennes...

2) Avantages des recombinaisons illégitimes

- Duplications de gènes

De même que les mutations insertionnelles peuvent apporter un bénéfice, les recombinaisons inégales peuvent avoir des conséquences avantageuses pour leur hôte. En effet, les éléments transposables seraient parfois à l'origine de duplications de gènes conservées. Dans le cas du gène codant pour la *glycophorine* chez l'homme, des éléments *Alu* seraient à l'origine de son expansion en trois copies (Labuda, Zietkiewicz & Mitchell, 1995). De même pour les gènes *α 1-acid glycoprotein* et *opsine* qui sont dupliqués (Merritt, Eastal & Board, 1990; Dulai *et al.*, 1999). Les rétroéléments ont aussi participé à l'évolution et à la diversité des

gènes dupliqués de type *HLA-B* et *HLA-C* (Kulski *et al.*, 1997). Enfin, un très bel exemple d'insertion rétrovirale chez un ancêtre des primates serait la cause de la duplication du gène de l'*amylase* : cette duplication serait à l'origine du régime alimentaire à base de céréales de ses descendants, notamment de l'homme (Meisler & Ting, 1993).

- Maintien de la taille des séquences péricentromériques

Chez l'homme, les séquences péricentromériques sont de l'ADN α satellite constitué de la répétition de monomères divergeant fortement des monomères centromériques. La quantité de ces séquences est très variable pour un chromosome donné au sein de l'espèce, de quelques centaines de kb à plusieurs Mb (Marçais, *et al.*, 1991). Cette quantité varie en fonction de recombinaisons inégales ou après leur amplification massive (amplifications saltatoires). L'augmentation de la taille de telles séquences pourrait être sans fin. L'élément *LI* aurait pour rôle de la contrôler. En effet, des familles d'éléments *LINE-1* sont fréquemment insérées dans les séquences α satellites péricentromériques, et totalement absentes des séquences centromériques. La présence d'un élément donné est très polymorphe: ce ne sont pas les mêmes éléments qui occupent un même site selon les individus (Laurent *et al.*, 1997; Puechberty *et al.*, 1999). On suppose que cela permet des recombinaisons intra-chromosomiques entre les éléments *LI* péricentromériques de part et d'autre du centromère, évitant ainsi "l'obésité" des régions α satellite.

D) Les éléments transposables actifs: quels sont les éléments encore mobiles dans les génomes d'aujourd'hui?

Parmi les insertions polymorphes examinées précédemment, toutes ne sont pas des éléments actifs: Par exemple, les *LINEs* achèvent rarement leur rétrotranscription en 5' au moment de leur insertion, et les nouvelles insertions sont souvent dépourvues de promoteur. Ces nouvelles insertions sont donc des éléments défectifs. Pourtant, elles proviennent d'éléments actifs. Je m'intéresse ici à ces éléments capables de générer de nouvelles insertions.

1) Eléments actifs de drosophile

Chez la drosophile, les éléments actifs sont des éléments de classe II comme *P* et *hobo*, et de classe I, comme les LTR-rétrotransposons (*copia/pseudoviridae* et *gypsy/metaviridae*), et les rétroposons (*facteur I*). Les rétroposons de type *SINE* sont totalement absents de cet organisme. Le suivi des insertions *de novo* dans le *locus sepia* dans 182 lignées isogéniques pendant 80 générations a permis

d'estimer le taux de transposition à 10^{-5} par site et par génération (Dominguez & Albornoz, 1996). En revanche, on sait déclencher dans cet organisme des instabilités génétiques associées à la mobilisation d'un élément (voir III-II).

Le séquençage du génome permet de rechercher de façon plus exhaustive les éléments transposables non encore identifiés, grâce à la mise au point de nouveaux algorithmes. Un élément de 3,4 kb de la famille *hoppel* a ainsi été découvert. Il code dans sa région centrale pour une protéine de 361 acides aminés, présentant 50% de similitude avec la transposase de l'élément *P*, *PS2*. L'élément *hoppel* répond à une interrogation qui portait sur la distribution de l'élément *P* dans les espèces de drosophiles: *P* est présent dans tous les génomes de *drosophilidae*, sauf, dans le sous-groupe *melanogaster*. L'évolution de *P* aurait conduit dans ce groupe à des séquences très divergentes, devenues fossiles, ou au contraire à un nouvel élément mobile. *Hoppel* serait cet élément, complet et actif: Il est polymorphe dans les populations de *D. melanogaster*, suggérant qu'il est toujours mobile. D'autre part, on le trouve dans toutes les espèces du sous-groupe *melanogaster*, (sauf dans l'espèce *D. erecta*) (Reiss *et al.*, 2001). Cette découverte montre que le génome de drosophile peut encore abriter des éléments mobiles actifs dont les effets délétères sont si rares qu'ils ont totalement échappé aux généticiens après plus de cinquante années de recherches acharnées!

2) Eléments actifs chez les mammifères

Chez l'homme, on ne connaît pas d'élément de classe II actif alors que ces éléments occupent 3,1% du génome (International Human Genome Sequencing Consortium, 2001). De même, aucune observation ne permet d'affirmer que les rétrovirus endogènes (8% du génome) sont encore mobiles, puisqu'on n'a pas décrit à ce jour de nouvelle insertion provirale. Seule, la famille *HERV-K* est décrite comme potentiellement active car elle produit des particules virales enveloppées extra-cellulaires (Boeke & Stoye, 1997). De plus, il a été montré que le gène *pol* dans cette famille code pour une reverse transcriptase active (Berkhout, Jebbink & Zsiros, 1999). Cependant, les seules copies pleine taille *HERV-K* du génome humain portent des mutations et aucun ne code l'ensemble des protéines actives. Si transposition il y a, ce ne peut-être que par complémentation (Tönjes, Czauderna & Kurth, 1999). Cette famille pourrait générer des pseudogènes (Berkhout, Jebbink & Zsiros, 1999). Il existe dans cette famille des éléments de séquence et de localisation spécifiques à l'espèce humaine, montrant une amplification propre à l'espèce (Medstrand & Mager, 1998). L'ensemble de ces données laisse suggérer aux auteurs que la famille *HERV-K* est peut-être encore potentiellement active.

Sur les 500 000 copies de *LINE L1* du génome humain , on estime à seulement trois ou quatre mille le nombre d'éléments complets (95% des *L1* sont tronqués en 5') et parmi eux, seulement 30 à 60 seraient capables de rétrotransposer (Sassaman *et al.*, 1997), 300 chez la souris (Naas *et al.*, 1998).

Les éléments SINEs comptent un million cinq cent mille copies (13% du génome) et poursuivent leur amplification dans le génome humain. On estime qu'il se produit une nouvelle insertion *Alu* pour 200 naissances (Deininger et Batzer, 1999).

Chez la souris, de nombreux *ERV* sont xénotropiques , c'est à dire qu'ils ne reconnaissent plus les récepteurs de leur hôte. On peut penser qu'une mutation du récepteur a été sélectionnée pour limiter l'amplification de tels rétrovirus après l'infection initiale. En fait, il apparaît que les *ERV* xénotropiques peuvent néanmoins être mobilisés, même si c'est très faiblement. D'une part, il peut se produire des complémentations en trans, soit pour la fonction SU d'un autre *ERV* (utilisant un récepteur différent), soit pour une autre activité de fusion cellulaire (Blobel *et al.*, 1992) : Ce co-empaquetage est décrit pour les *MLV* co-exprimés et conduit à ce que l'on appelle des pseudotypes. Ce concept permet d'expliquer les mouvements des *IAP* et des *VL30*. D'autre part, si l'infection initiale exige la protéine d'enveloppe, les amplifications ultérieures peuvent être de simples rétrotranspositions, pourvu qu'elles se produisent en lignée germinale ou chez l'embryon précoce.

Les provirus endogènes sont plutôt stables dans cette espèce, avec de faibles changements apparents d'une génération à l'autre. On a pu observer de nouvelles insertions de *MLV* dans des souches de laboratoire consanguines, au pedigree bien défini (souche AKR) par *Southern blot*. Il se produit une nouvelle insertion en 50-100 générations pour un *MLV* écotropique, en 350-700 générations pour un *MLV* non écotropique. Sur 160 allèles spontanés responsables de pathologies, 17 ont été la cible d'une rétrotransposition, soit environ 10% (cette étude portait sur 86 loci) (Kazazian, 1998). Cette fréquence est 60 fois moins élevée chez l'homme. Selon R. Löwer, cette différence entre l'homme et la souris serait une conséquence de l'élevage et ne refléterait pas forcément la situation naturelle. Dans les populations consanguines en effet, la forte répression des *ERV* est relâchée. Mais l'éventualité d'une plus grande mobilité des *IAP* par rapport aux *ERVs* est aussi avancée. Les données sur les populations naturelles de souris sont insuffisantes pour trancher (Löwer 1999).

En conclusion, les rétroposons sont les seuls éléments dont on observe la

mobilisation chez l'homme. Chez la souris, s'ajoutent à ces éléments les rétrovirus endogènes. Plus généralement, l'ensemble des données actuelles indique que les rétroéléments sont les seuls éléments transposables encore actifs chez les mammifères. Si l'on observe quelques rétrotranspositions aujourd'hui dans les génomes de vertébrés, on constate néanmoins qu'elles sont exceptionnellement délétères. En effet, si on considère une nouvelle insertion *Alu* pour 200 naissances humaines, on peut les estimer à 30 millions dans la population humaine (sans compter celles des précédentes générations). C'est beaucoup d'insertions pour seulement seize qui sont décrites comme responsables d'une pathologie, même si le chiffre d'événements délétères est sous-estimé.

Comment expliquer cette tolérance des génomes vis à vis de ces nouvelles insertions? L'idée d'une co-adaptation entre l'hôte et l'élément visant à limiter les effets délétères des insertions peut expliquer cette tolérance.

III-II Co-adaptation des éléments transposables avec leur hôte : limitation de leur impact délétère sur l'hôte ou optimisation de leur effet bénéfique ?

A) Contrôles transcriptionnels

Le cycle de tout élément transposable passe par une étape de transcription. C'est pourquoi la stratégie largement utilisée pour limiter les effets délétères liés à la transposition d'éléments mobiles consiste à réprimer les éléments transposables, et ainsi, empêcher toute étape du cycle de transposition.

1) La méthylation

La méthylation de certaines bases d'ADN en des sites particuliers du génome est un procédé de répression très utilisé chez les plantes et les vertébrés. La cytosine est la cible la plus fréquente des méthylases. L'ADN ainsi modifié est couplé à un complexe répressif, capable d'inhiber la transcription, dans sa phase d'initiation chez les eucaryotes supérieurs (Eden & Cedar, 1994) ou au cours de l'élongation chez les champignons filamenteux (Rountree & Selker, 1997).

La méthylation est impliquée au cours du développement des mammifères et des plantes, mais sa principale fonction serait le contrôle transcriptionnel des éléments transposables (Yoder, Walsh & Bestor, 1997). En effet, les 5-méthylcytosines, produits des méthylations, ne sont pas distribuées au hasard dans les génomes de mammifères et de plantes mais se concentrent principalement dans les séquences moyennement et hautement répétées (Bennetzen, 1996; Yoder, Walsh & Bestor, 1997; Finnegan *et al.*, 1998).

Plusieurs expériences ont montré le lien entre la méthylation des ET et leur répression chez les vertébrés : Levée de répression des ET chez un mutant nul (gène K.O.) de *méthylase Dnmt1* de souris (Walsh, Chaillet & Bestor, 1998), hypométhylation corrélée à des néoinsertions chez des hybrides de kangourou (O'Neil *et al.*, 1998), inhibition *in vitro* et *in vivo* de la transcription d'éléments *Alu* par la méthylation (Liu & Schmid, 1993; Liu *et al.*, 1995). De plus, la méthylation protégerait le génome des événement de recombinaison mitotique entre séquences répétées, ce qui va encore dans le sens d'une protection contre les

ET (Chen *et al.*, 1998). Le même lien entre méthylation et répression des éléments transposables est décrit chez les plantes (revue : Bender, 1998).

La méthylation est un mécanisme épigénétique, c'est à dire qu'elle persiste dans la descendance des cellules où elle est intervenue, alors que la séquence de l'ADN n'est pas modifiée. Elle ne serait qu'un maintien de l'état répressif, qui serait mis en place par d'autres mécanismes impliquant des reconnaissances ADN-ADN, ADN-ARN, utilisant des voies similaires à celles impliquées dans les phénomènes de *silencing* (voir plus loin).

Chez les mammifères, une vague de méthylation *de novo* a lieu dans les cellules somatiques entre le moment de l'implantation de l'embryon et celui de la gastrulation. Le génome du zygote est donc déméthylé entre la fécondation et le stade blastocyte : les cellules germinales, quant à elles, restent déméthylées jusqu'à la gamétogenèse. Si la méthylation a pour fonction la protection de l'hôte vis à vis des ET, on peut s'étonner que cette protection n'intervienne pas lors des stades clefs d'amplification des ET dans le génome d'un individu, c'est à dire dans l'embryon précoce et les cellules germinales primordiales.

Il est possible qu'une modification épigénétique répressive persiste durant ces stades. Effectivement, Morgan *et al.* ont observé une telle modification qui se transmet dans l'embryon par la lignée germinale et qui est corrélée à une méthylation ultérieure de cet élément en somatique (1999). L'état répressif serait donc à la fois maintenu dans les génomes transitoirement déméthylés, par un autre mécanisme, et transmissible d'une génération d'individu à l'autre.

Mais ceci n'est pas toujours vrai. Plusieurs études indiquent que des rétroéléments s'expriment effectivement durant la gamétogenèse et l'embryogenèse. Ainsi d'une famille d'éléments *Alu* spécifiquement exprimée dans les spermatozoïdes (Hellmann-Blumberg *et al.*, 1993; Kochanek, Renz & Doerfler, 1993; Chesnokov & Schmid, 1995) ; une nouvelle insertion germinale mâle de cet élément a d'ailleurs été observée (Wallace *et al.*, 1991). On trouve la même expression pour des *IAP* de souris (Dupressoir & Heidmann, 1996). Soit les éléments ont trouvé un moyen d'échappement au mécanisme répressif relais pendant l'embryogenèse, soit il est avantageux pour l'hôte de maintenir des éléments actifs pour leur potentiel évolutif. La protection somatique est avantageuse pour l'hôte, mais aussi pour l'élément s'il veut être transmis à la génération suivante. Ce mécanisme est donc bien en place pour limiter les effets délétères des ET, non pour empêcher leur expansion dans le génome.

La méthylation n'existe pas dans tous les organismes. En particulier, on ne trouve pas de méthylase chez les invertébrés, ni chez *C. elegans*, ni chez *Drosophila melanogaster*. Parmi les champignons et les protistes, la méthylation n'est présente que dans certaines espèces (Colot & Rossignol, 1999). Enfin, il existe un chordée, *Ciona intestinalis*, où ce sont les gènes qui sont méthylés et les ET déméthylés ! (Simmen *et al.*, 1999).

Chez la drosophile, on trouve néanmoins un mécanisme de maintien de la répression grâce aux protéines de la famille des Polycombes (*Pc*). D'autres mécanismes répressifs existent : découverts de façon indépendante, ils ne sont pas encore totalement élucidés mais semblent avoir des points communs entre eux. Leur fonction biologique pourrait être la répression des ET (figure 21)(Birchler, Pal Bhadra & Bhadra, 2000) :

Diagrammatic representation of three types of silencing of repeated genes. (a) Repeated sequences in tandem (left) are often recognized by the cell and bound with proteins characteristic of heterochromatin (right). (b) Homologous transgenes (triangles) and the endogenous gene (rectangles) (left) might associate transiently (center), which triggers a repressive chromatin complex accumulation that is maintained in subsequent cell divisions (right). (c) Homologous transgenes and the endogenous genes contribute to a common mRNA pool (left). At a certain threshold, dsRNA may be formed that targets the homologous RNAs for destruction (right).

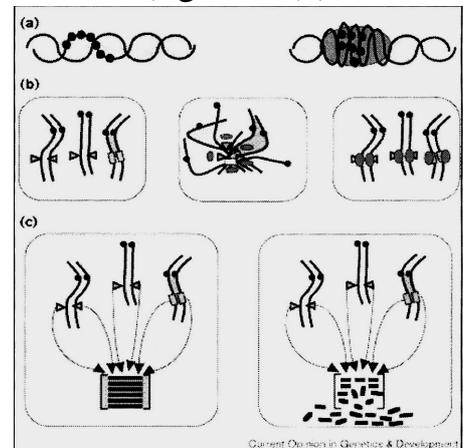


Figure 21 : Répression transcriptionnelle (Birchler, Pal Bhadra & Bhadra, 2000).

* *Repeat induced gene silencing* ou hétérochromatisation. Elle intervient localement sur des gènes ou des séquences transgéniques, en tandem ou multicopies et conduit à rendre le gène ainsi "hétérochromatinisé" silencieux (figure 21A).

* *Dispersed gene silencing*. Il s'agit encore de transgènes en plusieurs exemplaires, mais dispersés dans le génome. Ce phénomène a été appelé aussi *homologie dependant gene silencing*, *post-transcriptional gene silencing*, *transcriptional silencing*, ce qui en dit long sur la méconnaissance du mécanisme. Le terme "co-suppression", plus neutre est le mieux adapté. La co-suppression désigne l'extinction croissante avec le nombre de copies introduites de l'expression d'un transgène et de son homologue endogène (figure 21B).

* Dégradation post-transcriptionnelle de messagers. Quand cette dégradation suit l'injection expérimentale d'ARN double brin, on parle d'*interférence ARN* (figure 21C).

2) Répression d'éléments transposables et interférence ARN (RNA-i)

Les molécules d'ARN double brin (ARNdb) peuvent conduire à la dégradation des transcrits présentant une homologie avec ce double brin, ce qu'on appelle l'*interférence* ARN. Ce phénomène, initialement décrit chez *C.elegans* (Mongomery & Fire, 1998), a été reproduit chez la drosophile (Kennerdell & Carthew, 1998), le trypanosome (Ngo *et al.*, 1998) et les vertébrés (Li *et al.*, 2000; Wianny & Zernicka-Goetz, 2000). Il s'agit d'une régulation post-transcriptionnelle transmissible à la descendance cellulaire, au moins sur quelques divisions, conduisant à l'extinction totale du ou des gènes ciblés puisqu'on ne détecte plus aucun transcrit correspondant. Chez la drosophile, les produits de dégradation des transcrits seraient de petits ARN d'environ 25 nucléotides (Hammond *et al.*, 2000; Zamore *et al.*, 2000).

Des mutagenèses conduites chez *Neurospora crassa* et *Caenorhabditis elegans* ont permis d'identifier des gènes cellulaires intervenant dans l'*interférence* ARN (Cogoni & Macino, 1999; Ketting *et al.*, 1999; Tabara *et al.*, 1999). Certains d'entre eux sont les homologues de gènes codant des RdRP (RNA directed RNA polymerase) impliqués dans les phénomènes de régulation post-transcriptionnelle (*PTGS*) chez *Arabidopsis thaliana* (Dalmay *et al.*, 2000; Mourrain *et al.*, 2000). D'autre part, l'étude génétique de quatre gènes a permis de distinguer les "initiateurs", impliqués dans la reconnaissance du signal ARNdb (*rde-1* et *rde-4*), et les "effecteurs", impliqués dans l'*interférence* elle-même (*mut-7* et *rde-2*) (Grishok, Tabara & Mello, 2000).

Chez certains de ces mutants qui ont perdu leur compétence pour l'ARN *interférence*, de *C. elegans*, les familles d'éléments *Tc* se trouvent brusquement déréprimés (Tabara *et al.*, 1999). Inversement, à l'issue d'une mutagenèse visant à déréguler les éléments *Tc*, vingt-deux mutants sont devenus incompetents pour l'*interférence* ARN parmi les trente qui avaient été obtenus (Ketting *et al.*, 1999). L'identification de ces mutants est en cours. Il apparaît en tout cas que les gènes définis comme initiateurs *rde-1* et *rde-4*, ne sont pas impliqués dans le contrôle des transposons, contrairement aux effecteurs *mut-7* et *rde-2*.

L'ensemble de ces résultats montre que les ET peuvent être régulés par un mécanisme commun avec ceux de l'*interférence* ARN ou du *PTGS*, qui sont des régulations post-transcriptionnelles aboutissant à la destruction des transcrits. Toutefois, l'équivalent des gènes initiateurs pour les ET n'est pas connu.

3) Modèles d'instabilités génétiques chez la drosophile

Plusieurs instabilités génétiques associées à la mobilisation d'éléments ont été caractérisées. Les premières ont été identifiées grâce à un syndrome qui leur est associé, le syndrome de dysgénésie hybride. Ce syndrome a été décrit pour la première fois chez *Drosophila melanogaster* où il a conduit à l'identification du *facteur I* (Picard, 1976). Des phénomènes similaires mais totalement indépendants ont permis d'identifier l'*élément P* (Kidwell, Kidwell & Sved, 1977; Rubin, Kidwell & Bingham, 1982) et *hobo* dans la même espèce (Blackman *et al.*, 1987; Yannopoulos *et al.*, 1987), puis *Penelope* chez *D. virilis* (Petrov *et al.*, 1995). Depuis, d'autres mobilisations massives ont été décrites pour les éléments *gypsy* et *copia* sans qu'elles s'accompagnent du même syndrome.

La dysgénésie des hybrides se caractérise par une stérilité agamétique, des réarrangements chromosomiques, une distorsion de ségrégation et une haute fréquence de mutations qui sont restreintes à la lignée germinale (Kidwell, Kidwell & Sved, 1977).

a) L'*élément P* et le *trans-silencing*

L'*élément P* a d'abord été caractérisé chez *D. melanogaster* puis il a été identifié chez d'autres *drosophilidae* (Clark, Maddison & Kidwell, 1994; Kidwell, 1994). Les *éléments P* provenant des différentes souches de *D. melanogaster* ne diffèrent que par deux substitutions nucléotidiques. On ne détecte aucune copie complète dans les souches anciennes de laboratoire (O'Hare *et al.*, 1992). De plus, on ne détecte qu'une substitution entre les *éléments P* de *D. melanogaster* et *D. sophophora* (Daniels *et al.*, 1990). Ces observations ont étayé l'hypothèse d'une invasion de *D. melanogaster* par l'*élément P* il y a 50 ans suite à un transfert horizontal provenant de l'espèce *D. willistoni* (sous-genre de *sophophora*) (Kidwell, 1983). On explique la possibilité d'un tel transfert par l'intermédiaire d'un parasite commun aux deux espèces, qui infeste les embryons.

On distingue les souches *P*, contenant des *éléments P*, et les souches *M* qui en sont dépourvues. Le syndrome de dysgénésie, due à l'activité de l'*élément P*, s'observe dans la descendance d'un mâle *P* et d'une femelle *M*, mais pas dans le croisement réciproque. Après une période transitoire de dysgénésie, différents mécanismes de répression se sont développés, qui, selon les populations, sont transmis maternellement - cytotype *P*- (Engels, 1979; 1989) ou par les deux parents - *P* susceptibility- (Kidwell, 1985; Black *et al.*, 1987; Simmons *et al.*, 1990).

Les femelles P ont un cytotype P, capable de réprimer la transposition des *éléments P* dans la lignée germinale de leur descendance. Le cytotype P a une composante extra-chromosomique qui permet la transmission de cet état répressif à la génération suivante, appelée pré-cytotype P (Ronsseray, Lemaître & Coen, 1993). Pour le maintien de l'état répressif dans les générations suivantes, le cytotype P nécessite l'héritage des *éléments P* chromosomiques (Engels, 1983; 1989). La composante extra-chromosomique du cytotype P est donc une autorégulation par les *éléments P*.

Les souches P contiennent généralement 40 à 50 éléments dont un tiers seulement est de taille pleine longueur. Les populations naturelles de *D. melanogaster* semblent avoir développé autant de variantes dans leur mécanisme de répression de *P* qu'il y a eu de rencontres avec cet élément. La capacité de répression d'un élément donné dépend de sa structure et de son site d'insertion. Le cytotype P serait une combinaison entre les produits codés par l'*élément P* et des interactions chromatine-chromatine (Ronsseray *et al.*, 1998).

Les premiers mécanismes élucidés sont liés à la régulation de l'élément P, c'est à dire à la spécificité tissulaire d'un épissage alternatif permettant la synthèse d'un répresseur en somatique, et de la transposase en germinal (chapitre II). Certains éléments défectifs, comme l'*élément KP* (Robertson & Engels, 1989; Misra & Rio, 1990), codent des répresseurs ou des répresseurs-like constitutivement dans la lignée germinale. De tels éléments ont été fixés dans les populations naturelles (Black *et al.*, 1987; Simmons *et al.*, 1990; Rasmusson, Raymond & Simmons, 1993).

La souche naturelle $\pi 2$ ne contient pas d'*éléments KP* mais de nombreux *éléments P*. Une répression se serait mise en place au fil des générations et serait entretenue un faible taux de transcription en germinal. Les transcrits seraient en si faible quantité que l'efficacité de l'épissage en serait très réduite, ce qui expliquerait l'observation d'une synthèse accrue de répresseur en lignée germinale au détriment de la transposase (Roche, Schiff & Rio, 1995).

Le deuxième type de mécanisme semble être un *trans silencing* exercé par un élément P en position télomérique, dépendant de la structure chromatiniennne (Ronsseray *et al.*, 1996 ; Ronsseray *et al.*, 1998). De nombreuses souches naturelles possèdent des *éléments P*, complets ou non, en position télomérique 1A suggérant la sélection de ce site d'insertion (Ajioka & Eanes, 1989; Ronsseray, Lehmann & Anxolabéhère, 1989; Biéumont *et al.*, 1994). Cette répression est

exercée aussi par des transgènes de type *P-LacZ* inséré en 1A sur d'autres transgènes, pourvu qu'ils possèdent en commun une région transcrite d'homologie (Roche & Rio, 1998). La répression implique la transcription de l'élément inséré en 1A, mais pas l'usage du promoteur de l'élément *P* lui-même (Marin *et al.*, 2000).

Deux mécanismes sont proposés pour rendre compte du *trans silencing* homologie dépendant. Dans le cas du tabac, le locus *silencer* reconnaît la région homologue après un scan du génome et l'inactiverait (Matzke *et al.*, 1994). Dans la plante, l'inactivation de la cible est une méthylation (Vaucheret, 1993;1994). Dans la drosophile, le *trans silencing* serait une trans-hétérochromatisation (Roche & Rio, 1998). L'autre mécanisme proposé est le *PTGS* (*Post transcriptional Gene Silencing*), également découvert chez les plantes (Park *et al.*, 1996; Vaucheret *et al.*, 1996) : Le locus *silencer* produit des ARN qui conduiraient à la dégradation d'ARN homologues. Le pré-cytotype P pourrait être médié par de tels ARN.

b) Le *facteur I* et la co-suppression

Le *facteur I* est un rétroposon ancien de *Drosophila melanogaster* dont le génome héberge 20 à 30 copies défectives dans des sites péri-centromériques (Bucheton *et al.*, 1984). Dans la nature, on trouve des populations contenant des *facteurs I* complets euchromatiques, appelées *inductrices* (I), et des populations *réactives* (R), qui en sont dépourvues. Apporter des *facteurs I* actifs, paternels ou transgéniques, dans une souche R entraîne une dysgénésie hybride dans la descendance femelle (Picard, 1976). De telles femelles, dites SF, pondent des œufs qui n'éclosent pas avec une fréquence variable (Lavigne *et al.*, 1986) La fréquence de mort embryonnaire définit le taux de réactivité des mères R, qui lui-même est fonction de la souche (Bucheton *et al.*, 1976). Ce taux peut-être modifié de façon réversible par des facteurs individuels liés à l'âge des mères ou environnementaux comme la température d'élevage (Bregliano & Kidwell, 1983).

Le taux de réactivité est directement corrélé au taux de transposition du *facteur I* dans les survivants (Picard, 1978). Il est donc utilisé comme critère indirect pour évaluer la fréquence de transposition du *facteur I* (Lachaume & Pinon, 1993; de La Roche Saint André & Bregliano, 1998). Dans les deux générations qui suivent un croisement dysgénique, le nombre de *facteurs I* augmente jusqu'à atteindre 15 à 17 copies euchromatiques par génome haploïde, puis il est progressivement réprimé par autorégulation. La stérilité femelle persiste au-delà de la mise en place de cette répression (Pélisson & Bregliano, 1987).

Comme pour *P*, la dysgénésie ne s'observe que dans un sens de croisement, mais la mobilisation du *facteur I* peut s'observer dans la descendance du croisement réciproque c'est à dire le croisement de mâles R avec des femelles I. Les filles d'un tel croisement sont appelées RSF et ne présentent aucun signe de stérilité. Le taux de transposition est moindre que dans le croisement dysgénétique correspondant. On détecte les transcrits pleine taille du *facteur I* dans la lignée germinale des femelles SF et RSF (Chaboissier *et al.*, 1990), ainsi que dans leur descendance (Lachaume & Pinon, 1993).

A la façon des *éléments P*, ce sont les *facteurs I* qui sont responsables de la répression de leur transcription dans les souches I. Il suffit en effet d'introduire des *facteurs I* complets dans une souche R pour obtenir une souche I, c'est à dire une souche avec 10-15 copies dans laquelle le *facteur I* ne s'exprime plus. (Péllisson & Bregliano, 1987; Pritchard *et al.*, 1988). Mais, contrairement à *P*, on ne connaît aucun répresseur qui soit codé par le *facteur I*. C'est le nombre de copies qui dirige la répression. C'est pourquoi les premières études de régulation cherchaient à mettre en évidence chez l'hôte des protéines activatrices qui auraient été titrées par le *facteur I*.

Deux équipes ont montré un effet répressif exercé par différentes parties du *facteur I* dont l'efficacité augmente avec le nombre de copies des transgènes utilisés (Jensen, Gassama & Heidmann, 1999a; Chaboissier, Bucheton & Finnegan, 1998). Jensen *et al.*, ont introduit la région codant l'ORF1 de *I* sous le contrôle du promoteur de choc thermique *hsp*. Chaboissier *et al.*, ont testé le 5'UTR du *facteur I* qui possède le promoteur et l'*enhancer* spécifiques de la lignée germinale femelle. Cette dernière équipe a montré que la répression était indépendante de la transcription du transgène. Au contraire, la première équipe perdait la répression en délétant le promoteur *hsp*.

Malinsky, Bucheton & Busseau (2000) ont confirmé tous les résultats de Jensen *et al.* (1999a ;1999b) avec des constructions utilisant le promoteur *hsp*. Bien que la régulation dépende du nombre de copies et de leur transcription (Jensen, Gassama, & Heidmann, 1999a ; 1999b) elle n'est pourtant pas corrélée à la quantité de transcrits (Malinsky, Bucheton & Busseau 2000). Si un ARN médie l'effet répressif, il est certainement insuffisant. Mais les résultats obtenus avec le promoteur *hsp* ne sont pas totalement reproduits dans les expériences avec d'autres promoteurs, comme le 5'UTR de *I*, qui est plus proche de la situation biologique du *facteur I*. Tous ces résultats indiquent que le mécanisme de répression du *facteur I* fait probablement intervenir encore un phénomène de *trans-silencing homologie-dependant*, mais il est loin d'être totalement élucidé.

c) *Gypsy* et *flamenco*

Gypsy (aussi appelé mdg4) a été identifié par deux équipes chez *Drosophila melanogaster* (Bayev *et al.*, 1984; Modolell, Bender & Meselson, 1983). Cet *errantivirus* a été entièrement séquencé en 1986 par Marlor, Parkhurst & Corces. Le génome de *D. melanogaster* contient des copies inactives de *gypsy* dans l'hétérochromatine péri-centromérique, dont la moitié est fixée dans l'espèce. Quelques souches possèdent en plus des insertions polymorphes dans l'euchromatine (pas plus de cinq copies), ce qui suggère une répression de *gypsy* dans ces souches (Modolell *et al.*, 1983; Bayev *et al.*, 1984; Peifer & Bender, 1988; Vaury *et al.*, 1989).

Quatre instabilités génétiques associées à la mobilisation de *gypsy* ont été décrites dans des souches de laboratoires, dans lesquelles on dénombre plus de vingt copies euchromatiques. Les souches *Uc* et *tuh1:tuh3* (Kuhn, Woods & Andrew, 1981; Lim *et al.*, 1983; Peifer & Bender, 1988 ; Sheen, Lim & Simmons, 1993) ont aujourd'hui perdu leur instabilité. Les souches *MS* (Kim *et al.*, 1990) et *MG* (Mével-Ninio, Mariol & Gans, 1989; Prud'homme *et al.*, 1995) permettent toujours la mobilisation de *gypsy* lorsqu'elles possèdent des copies actives de l'élément.

Les femelles mutantes portant l'allèle dominant *ovo^{D1}* sont stériles. L'insertion de *gypsy* dans l'allèle *ovo^{D1}* permet de rétablir la fertilité des femelles (figure 22) (Mével-Ninio, Mariol & Gans, 1989). Or, *Gypsy* possède un site chaud d'insertion au locus *ovo* (Busson *et al.*, 1983). Prud'homme *et al.*, (1995) ont mis au point un test expérimental dans lequel le taux de réversion *ovo^{D1}* est corrélé au taux de transposition de *gypsy*. A l'aide de ce test, ils ont localisé le déterminant génétique majeur de la mobilisation de *gypsy* sur le chromosome X et l'ont nommé *flamenco* (*flam*). Ils distinguent les allèles *flam* "permissifs" (*flam^P*) ou "restrictifs" (*flam⁺*) qui permettent respectivement la mobilisation ou non (Pélisson *et al.*, 1994).

Le gène *flam* a été localisé en 20A1-A3, entre les loci *eo* et *wap* (Prud'homme *et al.*, 1995). Le génome dans cette région a été analysé par Valérie Robert ce qui a permis de compléter les données de séquences du consortium de séquençage du génome de *drosophila melanogaster* (Robert, 2000 ; Adam *et al.*, 2000). Cette région, située à la jonction entre l'euchromatine et l'hétérochromatine centromérique, contient des séquences moyennement répétées et peu de séquences uniques. Le gène *dip1* se situe à cette localisation et était un bon candidat pour être

le gène *flam*. (Mukhopadhyay *et al.*, soumis¹). L'équipe d'Alain Bucheton a tenté par transgénèse plusieurs expériences de sauvetage du phénotype *flam^P*, en vain (Robert *et al.*, 2001). A ce jour, le gène *flam* n'est donc toujours pas caractérisé.

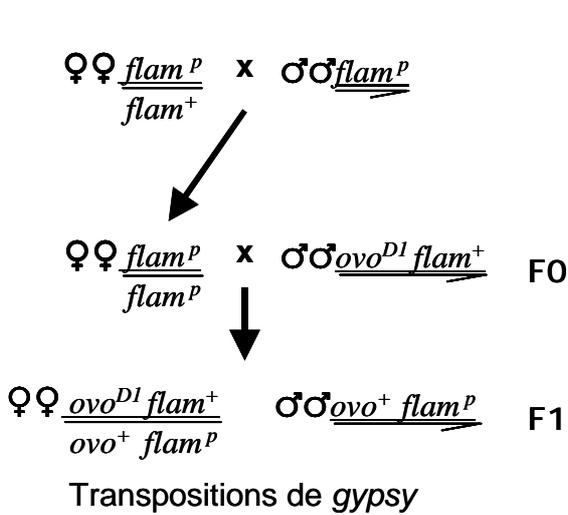


Figure 22 : Test *ovo^{D1}*. Crible d'insertion de l'*errantivirus gypsy* permettant d'évaluer sa mobilité et le taux de permissivité R d'un allèle *flamenco*. L'allèle dominant *ovo^{D1}* est responsable d'une stérilité femelle dans laquelle les ovaires ne se développent pas. Les femelles F1 décrites ici sont donc stériles. L'élément *gypsy* transpose dans la lignée germinale de la descendance F1 des femelles homozygotes *flam^P* (*flamenco* permissif), quel que soit l'allèle *flam* du père F0. *Gypsy* a un point chaud d'insertion au locus *ovo*. Quand il s'insère à ce site dans une cellule germinale souche, il permet la réversion de l'allèle *ovo^{D1}*. Un ovaire peut alors se développer dans la femelle F1. On évalue le taux R en rapportant le nombre n d'ovaires développés au nombre total d'ovaires.

La régulation par *flam* semble être un contrôle transcriptionnel de *gypsy* : Les transcrits et la protéine d'enveloppe sont détectés dans les cellules folliculaires somatiques des ovaires de femelles homozygotes *flam^P/flam^P* mais pas dans les femelles hétérozygotes (Péligsson *et al.*, 1994). Les allèles *flam^P* sont récessifs car une seule copie de *flam^P* est insuffisante pour mobiliser *gypsy*. La mobilisation de *gypsy* se produit dans la lignée germinale des filles de femelles homozygotes *flam^P/flam^P*, quel que soit le génotype du père. Pour réaliser le test *ovo^{D1}*, on analyse donc les ovaires des petites filles (figure 22). On peut classer les différents allèles permissifs selon le taux de transposition évalué par le test *ovo^{D1}* (Péligsson *et al.*, 1997).

Les allèles permissifs sont très instables en présence d'éléments *gypsy* actifs et présentent parfois des stérilités femelles. Plusieurs cas de fixations d'allèles restrictifs ont été observés au cours de l'entretien de souches permissives, qui ont conduit à l'arrêt de la transposition dans le stock. Dans un cas au moins, il s'agit d'une néomutation de l'allèle *flam^P* car la population avait été établie à partir d'un seul chromosome X (Robert, 2000). Les allèles *flam^P* sont donc instables en présence de *gypsy* actifs. Le maintien de l'allèle *flam^P* dans une lignée hétérozygote évite la mobilisation de *gypsy* et permet de préserver la souche MG (Prud'homme *et al.*, 1995).

¹ Mukhopadhyay M., Pelka P., DeSouza D., Robert V., Péligsson A., Bucheton A & Campos R. « Identification and characterization of the *Drosophila* dip1 (disco interacting-1 protein) gene, a new member of the Staufen family of RNA binding protein. »

Le système *gypsy/flamenco* présente des différences fondamentales avec les systèmes P/M ou I/R. Tout d'abord, l'expression de *gypsy* n'est détectée que dans les cellules somatiques des tissus germinaux, non dans les cellules germinales elles-mêmes. Cette spécificité d'expression est possible parce que les rétrovirus endogènes ont une étape extra-cellulaire leur permettant d'avoir quand même pour cible la lignée germinale. *Gypsy* échappe ainsi aux systèmes de répression qui semblent se mettre en place dans la lignée germinale pour *P* et *I*, systèmes spécifiques des séquences de *P* ou *I* mais faisant intervenir un mécanisme préexistant chez l'hôte. D'autre part, le contrôle de l'expression de *gypsy* est assuré par un locus de l'hôte unique *flamenco*, indépendant du nombre de copies de *gypsy* (Prud'homme et al., 1995).

B) Régulations post-transcriptionnelles

1) Limitation du taux de transpositions

Certains éléments sont maintenus actifs dans la cellule, c'est à dire qu'ils s'expriment, et que tous les produits nécessaires à leur cycle sont présents. Ces éléments actifs sont souvent stables dans le génome, suggérant une régulation post-transcriptionnelle pour limiter leur expansion, (post-transcriptionnelle au sens large, c'est à dire de toute étape du cycle après la transcription). Afin de comprendre les mécanismes mis en jeu pour empêcher la transposition de tels éléments, on a recherché des situations dans lesquelles se produisent des mobilisations plus fréquentes, par exemple en criblant des mutants dont le phénotype serait une augmentation du taux de transposition.

a) Détournement d'une fonction générale de l'hôte

Ty1 est un *pseudoviridae* actif de *Saccharomyces cerevisiae*. Il est présent dans toutes les souches de levure, majoritairement sous forme complète, suggérant une relation symbiotique (Boeke, 1996). Quelle qu'en soit la raison, le maintien d'un rétrotransposon actif dans son génome ne peut-être toléré par la levure que si le taux de transposition reste limité. C'est effectivement ce qu'on constate :

Malgré l'abondance des transcrits *Ty1* (1% des ARN cellulaires totaux), on détecte peu d'ADNc correspondant : environ une copie d'ADNc pour 40 000 transcrits par cellule (Conte, 1998), et les événements de transposition sont rares. Ce ne sont pourtant pas les éléments défectifs qui sont responsables de la dormance des *Ty1* (Curcio & Garfinkel, 1994). Il semble donc que la cellule

exerce un contrôle post-transcriptionnel sur *Ty1* dont la conséquence est de limiter les transpositions.

Deux mutants de levure présentent un taux de transposition de *Ty1* plus élevé. Ils ont été isolés grâce à un marqueur phénotypique de transposition permettant de détecter tous les événements d'intégration de *Ty1* dans le génome.

Le premier mutant concerne la protéine *Fus3p*, une *MAP kinase* (*Membrane Associated Protein kinase*) intervenant dans la transduction du signal de la phéromone impliquée dans les accouplements. Deux hypothèses sont avancées par les auteurs pour expliquer cette fonction de *Fus3p* dans le contrôle de la transposition de *Ty1*:

Dans la première, *Ty1-GAG* serait une cible pour la kinase *Fus3p*. *Ty1-GAG* est, en effet, une glycoprotéine, et sa stabilité ou sa capacité d'association dans les VLP (Virus Like Particles) peuvent dépendre de son état de phosphorylation. Cette hypothèse est argumentée par une ancienne observation montrant une corrélation entre l'hyperphosphorylation de *Ty1-GAG* et l'inhibition de la transposition lors d'un accouplement (Xu & Boeckle, 1991). L'autre hypothèse envisage un rôle plus indirect de *Fus3p* qui inhiberait un activateur de transposition. *Fus3p* régule négativement des gènes impliqués dans la croissance du filament (Madhani, Styles & Fink, 1997). Justement, l'expression de *Ty1* est dépendante de facteurs de transcription présents lors de cette croissance. Les deux hypothèses ne sont pas forcément exclusives. On peut envisager que la cellule ferait d'une pierre deux coups en conditionnant sa réponse au stress.

L'autre protéine est *ssl2p*, une *hélicase* appartenant au complexe *TFIIH*, un complexe contenant la *RNA polymerase II*. Plusieurs mutants de la protéine *ssl2* sont responsables de l'augmentation de la transposition de *Ty1* (Lee *et al.*, 1998). Dans tous les cas, la quantité d'ADNc augmente, mais pour certains mutants, sans corrélation avec une augmentation de l'expression de *Ty1*. Certains mutants d'une autre *hélicase* du complexe *TFIIH*, la protéine *RAD3*, stimulent également la transposition de *Ty1*. Ces observations sont cohérentes avec le rôle putatif de *TFIIH* dans la dégradation d'intermédiaires défectifs de recombinaison comme le sont les ADNc de *Ty1*.

Des fonctions cellulaires générales ont donc été adaptées par l'hôte pour prévenir des rétrotranspositions tout en maintenant cet élément actif. On peut s'étonner que de tels détournements ne soient pas décrits pour d'autres éléments ou dans d'autres espèces. Cela traduit sans doute la difficulté de les mettre en

évidence, car plus la fonction détournée a un rôle général dans la cellule, et plus nombreuses sont les perturbations lorsque cette fonction est mutée. Sur ce point, il n'est pas étonnant que les premiers résultats parus concernent un organisme unicellulaire comme la levure.

b) Résistances cellulaires à des rétrovirus

Les *MLV* ont été classés en trois catégories selon leurs capacités d'infection de cellules murines, capacités dépendant de leur protéine *Env*: Ces virus sont dits écotropiques s'ils ne peuvent infecter que des cellules murines, xénotropiques s'ils ne les infectent pas, et polytropiques s'ils peuvent infecter d'autres espèces en plus des souris. Ces différentes catégories correspondent à des récepteurs de l'hôte qui ont été identifiés (tableau III), et qui, par extension, prennent les mêmes noms. Dans les souris consanguines de laboratoire, le récepteur des *MLV* xénotropiques semble absent, alors qu'il est présent dans de nombreuses populations naturelles sensibles au virus (Kozak, 1985). Cela paraît contradictoire avec un rôle physiologique du récepteur de l'hôte. Plus probablement, ce récepteur est muté dans les souches de laboratoire et cette mutation les a rendues résistantes à l'infection par les *MLV* xénobiotiques (Kozak, 1985). De telles mutations ont été décrites pour le récepteur écotropique de cellules de *Mus dunni*: un simple changement d'acide aminé a rendu ces cellules résistantes au *Mo-MLV*, mais pas à d'autres sous-types viraux écotropiques (Eiden *et al.*, 1993). D'autres modifications, comme des altérations post-traductionnelles du récepteur écotropique *RD114* chez le chat ont rendu ce récepteur xénobiotique (Dunn *et al.*, 1993).

Tableau III : Récepteurs cellulaires correspondant à chaque catégorie de virus endogène chez la souris (Coffin, Hughes & Varmus, 1997).

Nom	Tropisme cellulaire	Récepteur de souris
Écotropique	souris seulement	<i>Rec I</i>
Xénotropique	non souris seulement	-
Polytropique	souris et non souris	<i>Rmc f</i>

Parmi les résistances aux rétrovirus exogènes, on peut rappeler la capture par l'hôte de fonctions provirales capables de compétition, comme *Fv1* et *Fv4* (chapitre III-I-A). On peut imaginer que d'autres fonctions provirales conservées dans

différentes espèces, chez l'homme en particulier, ont un rôle de protection contre des infections de rétrovirus exogènes, à la façon des *loci Fv1* et *Fv4* chez la souris.

c) Spécificité tissulaire de l'activité des ET

Limiter strictement les transpositions aux tissus germinaux peut-être un bon moyen d'éviter des mutations en somatique: Les mutations somatiques nuisent au potentiel de survie de l'hôte sans augmenter le développement de l'élément dans le génome puisque les nouvelles insertions dans ces tissus ne seront pas transmises aux individus de la génération suivante. Ce moyen a été adopté par les éléments *I* et *P* de *Drosophila melanogaster*, qui chacun ne s'expriment que dans les cellules germinales. Le *facteur I* contient un enhancer situé juste après le promoteur, ciblant son expression dans les cellules germinales précoces de l'ovaire adulte (Udomkit *et al.*, 1996). Concernant l'élément *P*, l'intron 2-3 n'est pas épissé dans les tissus somatiques. Il en résulte la synthèse d'un répresseur de 66 kDa spécifique du promoteur *P* et l'absence de transposase dans ces tissus. Cette régulation est très intéressante et montre l'interaction avec l'hôte, puisque la machinerie d'épissage tissu-spécifique requiert des facteurs de l'hôte (Bingham, *et al.*, 1988; Tseng *et al.*, 1991).

d) Autorégulation du nombre de copies

Les éléments qui envahissent massivement et brutalement le génome dans les cas d'instabilités génétiques atteignent en général un nombre de copies maximum dans le génome qui varie selon l'élément. Cela signifie que le taux de transposition diminue au fil des générations pour devenir quasiment nul. Le contrôle du nombre de copies semble propre à chaque élément, même lorsqu'on a atteint le nombre de copies critique pour l'un, on peut mobiliser un élément d'une autre famille. D'autre part, ce nombre maximum varie d'un élément à l'autre, d'où l'idée d'une autorégulation de l'élément, encore à démontrer.

On peut citer les exemples de l'élément *Ac* du maïs, dont le taux de transposition décroît avec l'augmentation du nombre de copies (McClintock, 1955), de l'élément *P* qui ne dépasse pas 50 copies par génome, et du *facteur I* dont la fréquence de transposition diminue quand il a atteint 15-17 copies par génome. Ces deux éléments sont effectivement responsables de leur répression. Ainsi, après son amplification, l'élément limite lui-même ses transpositions, préservant ainsi les intérêts de l'hôte. Dans le cas du *facteur I*, chaque fois que le génome de l'hôte est dilué par des croisements avec des souches R, il recommence à transposer jusqu'à atteindre son nombre de copies limite.

2) Limitation des conséquences des transpositions

a) Insertions "fantômes"

Certains ET insérés dans des régions transcrites ont la faculté d'être épissés, ce qui les rend théoriquement invisibles pour la machinerie de traduction et ne perturbe pas l'expression du gène cible. En fait, ces épissages sont plutôt imprécis et de telles insertions sont mutagènes, mais elles n'empêchent pas totalement l'expression du gène cible. Des insertions qui auraient pu être létales peuvent ainsi être tolérées. On trouve de tels exemples dans le maïs (Wessler, 1989; Varagona, Purugganan & Wessler 1992) et la drosophile (Fridell, Pret & Searles, 1990; Pret & Searles, 1991).

Les *LINEs* ont un faible signal de polyadénylation ou de fin de transcription. Même insérés dans des régions transcrites, ils n'interrompent pas la transcription. Cette propriété, couplée à un adressage fréquent des *LINEs* dans les introns, peut rendre une insertion *LINE* transparente pour l'expression du gène cible (Bingham & Zachar, 1989).

b) Adressage du site d'insertion

- De la limitation de l'impact délétère...

Avoir des cibles d'insertion préférentielles peut-être très avantageux pour un ET qui peut "choisir" des sites sans conséquence mutationnelle trop importante pour l'hôte.

Ainsi, l'élément *P* chez *Drosophila melanogaster* s'insère dans des régions potentiellement régulatrices plutôt que dans des exons (Spradling *et al.*, 1995). La préférence de cet élément pour ces régions avait été déjà observée (Tsubota, Ashburner & Schedl, 1985; Kelley *et al.*, 1987). En étudiant de façon systématique les sites d'insertion obtenus après une mutagenèse, Spradling *et al.*, ont confirmé cette préférence, et ont montré que les rares insertions en dehors de ces régions se font plutôt dans la partie 5' non codante des gènes, ou plus rarement encore dans le premier intron. Sur 56 insertions, 4 seulement se trouvaient dans une région codante. Les compilations de 500 expériences dites "enhancer trap" ont montré que l'élément *P* s'insérait bien dans des régions régulatrices car le gène rapporteur utilisé dans ces expériences garde le profil d'expression spatio-temporel du gène cible (Bellen *et al.*, 1989). On peut parler aussi du site télomérique 1A, qui semble être un point chaud d'insertion pour des éléments *P-LacZ* (Karpen & Spradling,

1992). En fait, cet adressage n'est pas observé par Ajioka & Eanes (1989), sauf dans certaines souches, suggérant que les télomères 1A auraient des aptitudes variables selon les souches à recevoir des *éléments P*. Cet adressage en 1A est bénéfique pour l'hôte, puisque nous avons vu que les *P* insérés au 1A ont pour propriété de réguler fortement négativement les autres *P* du génome.

On observe la même préférence de l'élément *Ty1* de *Saccharomyces cerevisiae* pour s'insérer dans les régions non codantes (Sandmeyer, Hansen & Chalker, 1990). Sur plus de 100 nouvelles insertions de *Ty1* observées dans le chromosome III, pratiquement toutes sont localisées dans ou près d'un gène ARNt ou dans des LTR préexistantes. Seules 3% d'entre elles se trouvent dans des cadres de lecture (Ji *et al.*, 1993).

- A des ciblages de plus en plus spécifiques...

Certains rétroposons codent pour une endonucléase site-spécifique de site. Sur les onze lignages de rétroposons établis par Malik, Burke & Eickbush (1999), trois codent une endonucléase site spécifique présentant des similitudes avec les enzymes de restriction (Yang, Malik & Eickbush) et s'intègrent en des sites spécifiques (tableau IV).

Tableau IV : Sites d'insertion préférentiels de *LINEs*.

NOM	HOTE	SITE D'INSERTION	REFERENCE
<i>Tx1</i>	Xenope	Autres éléments mobiles	Garrett, Knutzon & Carroll, 1989
<i>CRE/SLACS</i>	Trypanosome	Exons leader épissés	Aksoy <i>et al.</i> , 1990
<i>R4/Dong</i>	nématodes	Entre les unités de transcription des gènes ARNr	Xiong & Eickbusch, 1993
<i>R4/Dong</i>	Insectes	ARNr 26s	Burke, Müller & Eickbusch, 1995
<i>Zepp</i>	<i>Chlorella</i>	Copies préexistantes de lui-même	Higashiyama <i>et al.</i> , 1997
<i>R1/R2</i>	Arthropodes	ARNr 28s	Burke <i>et al.</i> , 1998

Les huit autres lignages, apparus postérieurement à la divergence des trois précédents, auraient perdu ce domaine au profit d'un domaine nommé *AP-like* (Yang, Malik & Eickbush, 1999). Le domaine *AP-like* présente des similarités avec celui des endonucléases apurinique/apyrimidique et se situe avant le domaine rétrotranscriptase, dans l'*ORF 2*. Il est responsable de l'activité endonucléase lors de l'intégration et il est relativement peu spécifique (Feng *et al.*, 1996; Feng, Schumann & Boeckle, 1998). La diversification des rétroposons aurait accompagné l'augmentation de la taille des génomes eucaryotes et la perte de la spécificité

d'insertion aurait permis d'exploiter une plus grande opportunité d'insertions au hasard sans s'accompagner d'effet délétère (Malik, Burke & Eickbush, 1999). La spécificité d'insertion n'est donc pas un avantage sélectif pour tous les éléments, ou à tout moment dans le génome, même si quelques uns parmi les huit l'ont conservée (tableau IV). Ce qu'il faut remarquer, c'est que chaque fois qu'il y a spécificité d'insertion, le site ciblé permet de minimiser l'impact sur le génome.

Les *metaviridae* possèdent souvent une spécificité de site d'insertion plus précise que le ciblage d'une région non codante. *Ty3* par exemple, s'insère 5 pb avant le début de transcription de gènes transcrits par l'ARN *pol III* (Chalker & Sandmeyer, 1992). Cette polymérase fonctionne avec des promoteurs internes. Aussi, une insertion en amont est supposée avoir peu ou pas d'impact sur le taux d'expression du gène cible (Kinsey & Sandmeyer, 1995). Kirchner, Connolly & Sandmeyer ont montré en 1995 que cette spécificité impliquait des interactions protéine-protéine entre les complexes d'intégration de *Ty3* et la machinerie *PolIII* avec les facteurs de transcription *TFII-B* et *-C*. Concernant les *errantivirus*, on ne connaît pas les partenaires intervenant dans la spécificité du site d'intégration, mais des points chauds d'insertion ont été décrits pour *gypsy* (Dej *et al.*, 1998) et *Idefix* (Desset *et al.*, 2000). A noter que *ZAM* présente une spécificité d'intégration remarquable parmi les *errantivirus* et les *retroviridae* dont on a étudié les séquences flanquantes d'insertion, avec un consensus palindromique de 6 bases GCGCGC (Leblanc, Dastugue & Vaury, 1999).

- En passant par une fonction des ET

Concernant les *MITE*, transposons de classe II, leur distribution dans les plantes et les animaux montre une préférence d'insertion dans les régions 5' non traduites (Wessler, Bureau & White, 1995; Tu, 1997). Toutefois, comme il ne s'agit pas de nouvelles insertions, on ne sait pas si cette distribution reflète effectivement une préférence de l'élément pour ces régions ou la sélection de ces sites par l'hôte. Même question pour le *facteur I* de *Drosophila melanogaster* dont on trouve les vestiges de nombreuses copies accumulés dans l'hétérochromatine β (Vaury, Bucheton & Pélisson, 1989). S'agit-il d'un adressage des insertions dans l'hétérochromatine pour limiter les effets délétères ou cette accumulation est-elle une conséquence à long terme de l'absence de recombinaison dans ces régions? Il semble que cet élément soit bien adressé préférentiellement dans l'hétérochromatine dans cette espèce, mais pas pour limiter les effets délétères! Curieusement, les effets délétères sont plus fréquents dans l'hétérochromatine que dans l'euchromatine (Dimitri *et al.*, 1997). Si cette observation contredit l'idée que le ciblage de l'insertion permet de diminuer le risque de mutation, elle devient un

argument pour une fonction des ET dans l'hétérochromatine (Dimitri & Junakovic, 1999).

Une autre façon de limiter l'impact insertionnel des éléments dans l'hôte est de cibler d'autres éléments transposables. Dans le maïs, dont le génome contient plus de 50% de rétroéléments, ceux-ci sont localisés pour la plupart dans les régions en dehors des gènes, et plus de la moitié des insertions analysées se situent à l'intérieur d'autres éléments (San Miguel *et al.*, 1996). Nous avons vu déjà l'exemple des rétroposons *Tx1* et *Zepp* qui s'intègrent dans des ET préexistants. Si ces observations permettent d'argumenter l'hypothèse de l'évolution des ET vers une limitation de l'impact insertionnel, ils vont aussi dans le sens du modèle de Dimitri & Junakovic dans lequel les ET sont adressés vers les parties du génome où ils sont susceptibles de jouer un rôle. Des études plus approfondies, en particulier sur les rétroposons, seraient nécessaire pour savoir si ce modèle peut-être généralisé.

L'élément *Ty5* apporte un argument dans ce sens. Dans le génome de *Saccharomyces cerevisiae*, on trouve des *Ty5* incompetents pour la transcription, fixées dans l'espèce au locus mat HMR et en position subtélomérique. En utilisant un élément actif d'une espèce proche, *S. paradoxus*, Zou, Kim & Voytas (1996) ont montré que *Ty5* s'insère préférentiellement dans la chromatine silencieuse (18 insertions sur 19). En comparant les télomères des deux espèces de levure, ils montrent un rôle de *Ty5* dans des événements de recombinaison qui ont conduit à des réorganisations chromosomiques, et ainsi un rôle des ET en tant que séquences répétées dans l'architecture des chromosomes.

Le ciblage de l'insertion est donc très répandu parmi les ET. Depuis les régions intergéniques des *éléments P* aux sites séquences spécifiques des *errantivirus*, les mécanismes permettant ce ciblage et les intervenants cellulaires sont encore mal connus. Quant aux raisons de ces adressages, elles sont encore discutées. Mais, quelles que soient les interprétations, l'adressage indique toujours une évolution de l'élément vers une meilleure adaptation à l'hôte.

C) Les ET ont-ils un rôle fonctionnel?

La représentation ubiquitaire des ET dans tout le monde vivant suggère que les ET ont un rôle intrinsèque dans l'hôte. Certaines études tentent de mettre en évidence une fonction indispensable à la vie de l'organisme qui serait assurée par un élément potentiellement actif. La démarche génétique habituelle consisterait à

supprimer la fonction pour en étudier les conséquences et en déterminer le rôle. L'idéal serait donc de disposer d'organismes sans élément. Mais ces modèles n'existent pas naturellement et seraient difficiles à obtenir compte tenu de la répétition des éléments d'une même famille et de la redondance possible d'une fonction par plusieurs familles. Certains auteurs ont contourné ce problème en créant dans un organisme des conditions de crible pour mettre en évidence une fonction vitale.

Par exemple, deux équipes ont mis en évidence l'intervention de la reverse transcriptase des éléments *Ty1* dans la réparation de certaines coupures double brin de l'ADN de leur hôte, suggérant un rôle fonctionnel de *Ty1* dans la réparation chez *S. cerevisiae* (Teng, Kim & Gabriel, 1996; Moore & Haber, 1996). Pour cela, elles ont utilisé des levures mutées pour la réparation et recherché "la signature" de *Ty1* à la jonction de l'ADN cassé. Mais le fait de montrer que l'utilisation de l'activité rétrotranscriptase de l'élément est possible chez un mutant ne prouve pas qu'elle assure bien cette fonction dans les levures sauvages.

Un autre exemple concerne les *SINEs*. La transcription des *SINEs* se trouve augmentée en condition de stress (Liu *et al.*, 1995; Kimura, Choudary & Smid, 1999), tout comme augmente la quantité de la forme active d'une protéine cellulaire, la *PKR*, qui inhibe alors la synthèse protéique. Chu *et al.* ont montré en 1998 que les transcrits *Alu* inhibaient l'activation de la *PKR*, l'empêchant alors d'assurer son rôle d'inhibition de la synthèse protéique. Pour cela, ils utilisent un gène rapporteur, la luciférase, et montrent par transfection, que la surexpression d'*Alu* permet l'augmentation de la synthèse de luciférase. Les *SINEs* auraient donc une fonction dans la régulation de la synthèse protéique, antagoniste de celui de la *PKR* et auraient pour rôle de mettre fin à l'inhibition de la traduction à la suite d'un stress (Chu *et al.*, 1998; Schmid, 1998).

Finalement, on se rapproche un peu dans ces exemples du concept de domestication, car une seule fonction de l'élément est requise, même si, dans le premier cas, l'activité rétrotranscriptase nécessite aussi la matrice ARN de l'élément *Ty1* pour synthétiser un ADNc. Pourquoi alors l'élément est-il maintenu actif plutôt que la fonction d'intérêt domestiquée? On peut imaginer qu'un même élément aurait plusieurs fonctions, non encore identifiées. Le contrôle des transpositions de *Ty1* par des protéines ayant un rôle central dans la levure (chapitre III-II-B) suggère en tout cas que cet élément puisse avoir une fonction importante dans les cellules de levure.

Le rôle que jouent les ET dans l'évolution de leur hôte peut-il être considéré comme une fonction? En clair, les ET subissent-ils une sélection positive du fait de leur rôle dans l'évolution? En 1994, Charlesworth, Sniegowski & Stephan concluait, au vu de la faible fréquence d'ET fixés dans le génome de *D. melanogaster*, que les ET avaient peu ou pas d'importance pour l'évolution de leur hôte. Pourtant, cette conclusion ne tient pas pour les mammifères où les mutations d'ET fixées sont nombreuses et laisse la porte ouverte à l'hypothèse d'une sélection positive (Deininger, 1989). Nous avons examiné la relation entre les ET et le fitness de l'hôte (chapitre III-I-). Savoir si le potentiel adaptatif des ET est suffisamment sollicité pour compenser leur maintien dans tout le règne vivant n'est pas encore tranché. Pour l'instant, on ne peut que constater la corrélation entre l'augmentation du nombre de copies d'un élément et l'avantage sélectif qui semble en résulter pour l'hôte.

Un exemple semblant répondre aux critères d'une fonction intrinsèque se trouve sans doute dans les télomères de drosophiles, bien que l'origine des éléments *HeT-A* et *TART* soit discutée. Plus généralement, les ET semblent très impliqués dans le maintien de structures chromatinienne propres aux eucaryotes comme les centromères et les télomères, et certains auteurs supposent que les ET auraient contribué à la macro-évolution, comme la transition procaryotes-eucaryotes ou l'apparition des vertébrés (Mc Donald, 1999; Wolffe & Matzke, 1999). Les éléments transposables ont pu être recrutés pour leurs propriétés de liaison à certaines protéines afin de participer à la formation de l'hétérochromatine (Bureau, Ronald & Wessler, 1996). Longtemps, l'abondance des ET ne semblait pas s'accompagner d'un rôle visible. Cela était interprété comme une absence de rôle dans l'évolution. Henikoff *et al.* ont proposé au contraire en 1997 qu'elle témoignait des effets bénéfiques de la flexibilité des génomes au cours de l'évolution. Pour autant, compte-tenu de la grande variabilité du nombre de copies des ET dans les espèces eucaryotes, ce n'est peut-être pas le nombre de copies lui-même qui est sujet à une sélection positive malgré sa capacité d'augmenter la flexibilité génomique (Kidwell & Lisch, 2000).

Si les ET sont quelques fois délétères pour un individu, ils ont été jusqu'à présent le partenaire incontournable des espèces survivantes et semblent même avoir joué un rôle moteur au cours de l'évolution. L'imbrication des ET avec le génome est telle qu'elle suffit à expliquer leur maintien sans qu'on leur cherche une fonction. Certains auteurs suggèrent même que la distinction entre génome de l'hôte et éléments génomiques égoïstes serait dépassée au regard de l'origine des génomes modernes (McDonald, 1999 ; Anxolabéhère, Nouaud & Miller, 2000). Aussi, "compagnons de voyage" me paraît l'expression la plus appropriée pour les



éléments transposables, car elle ne présume pas de la nature de la relation avec l'hôte et peut donc rendre compte de toutes les situations que nous avons rencontrées dans ce chapitre: Tous les voyageurs non solitaires savent bien qu'on peut-être accompagné par le pire ou le meilleur, et que ces considérations dépendent en plus des circonstances!...



LE MODELE

I- La drosophile, un modèle génétique

1) Fiche BiOiB

BiOiB, base terminologique en Biologie moléculaire et cellulaire.

Inserm U492, CHU Henri Mondor, 94 010 Créteil

développement et diffusion « La Biothèque », 119 rue de Montreuil, 75 011 Paris.

Le texte qui suit est extrait de la base de donnée BiOiB pour laquelle je rédige des fiches:

La drosophile est une petite mouche de 2-3 mm de long, communément appelée "mouche du vinaigre" en français et "fruit fly" en anglais. *Drosophila melanogaster* est l'espèce la plus étudiée.

En 1910, Morgan observe pour la première fois un phénotype mutant de la drosophile. Peu après, il trouve dans sa collection une drosophile aux yeux blancs, au lieu du rouge brique trouvé habituellement chez les mouches sauvages. Il nomme le gène muté à partir du phénotype mutant, "white". Morgan et son équipe ont ainsi décrit des dizaines de mutations et établi les premières cartes génétiques. Ils ont introduit la notion de distance génétique dont l'unité est le Morgan. Le prix Nobel de médecine a été attribué à Morgan en 1933 pour sa théorie proposant les chromosomes comme support de l'hérédité .

Grâce à son temps de génération court (10 jours à 25°C), à sa facilité d'élevage et ses quatre petits chromosomes (150 Mb) aisément cartographiables dans leur forme polyténisée, la drosophile est devenue un modèle de choix pour les généticiens. Le premier animal transgénique obtenu est une drosophile, grâce aux travaux de Spradling et Rubin qui, en 1982, ont réalisé la première transgénèse en utilisant les propriétés de l'élément transposable, *P*.

La drosophile est aussi un modèle remarquable pour étudier la biologie du développement. Le prix Nobel de médecine 1995 a été attribué collégialement à trois pionniers dans ce domaine:

-Nüsslein-Volhard et Wieschaus ont mis au point les cribles qui ont permis l'identification des gènes responsables des étapes précoces du développement: les gènes à effet maternel, les gènes *gap*, *pair rule* et de polarité segmentaire.

-Lewis avait découvert auparavant les gènes homéotiques, impliqués plus tardivement dans le développement, et responsables de l'identité des segments. Ces gènes ont en commun une homeoboîte, codant pour un homeodomaine. Ces gènes ont été retrouvés très conservés dans tous les animaux où on les a cherchés.

Au jourd'hui, la communauté scientifique dispose de très nombreux mutants, de plusieurs bases de données contenant tous les gènes décrits et leurs allèles (Berkeley *Drosophila* Genome Project-<http://www.fruitfly.org/>, European *Drosophila* Genome Project-<http://edgp.ebi.ac.uk/cgi-bin/progress.pl>, Fly Base-<http://flybase.bio.indiana.edu/>), leur localisation génétique et chromosomique, les éléments transposables connus, des livres très complets sur la description minutieuse de toutes les étapes du développement de l'embryon, de la larve ou de la puppe, les trois étapes de développement communes aux insectes. Enfin, le génome de *Drosophila melanogaster* est pratiquement entièrement séquencé.

La drosophile a toutefois quelques inconvénients:

-La conservation d'une lignée d'intérêt exige un élevage continu en laboratoire à une température minimale de 18°C (on ne peut congeler puis décongeler les drosophiles à aucune étape de leur développement contrairement aux gamètes et embryons de mammifères). Des Stock Centers élèvent ainsi une collection importante de drosophiles.

-La transgénèse nécessite l'utilisation d'éléments transposables (*P* ou *hobo*) ne permettant pas de cibler le site d'insertion. On ne peut donc pas obtenir un mutant knock out comme chez la souris.

Les domaines de recherche impliquant la drosophile aujourd'hui sont tous ceux nécessitant un outil génétique puissant: biologie du développement bien sûr, mais aussi évolution, éléments mobiles du génome, et à leur conjonction, la structure chromatiniennne et sa régulation. Depuis quelques années, de nouveaux domaines émergent, comme les études portant sur l'immunité, le vieillissement, l'apoptose, le comportement.

Remerciements

Jean-Louis Couderc

Bibliographie

The Genome of *Drosophila melanogaster*
Dan L.Lindsley and Georgianna G. Zimm
Academic Press, INC 1990

Drosophila: a laboratory handbook
Michael Ashburner
Cold Spring Harbor Laboratory Press,1989

La drosophile: des chromosomes aux molécules
Jean Deutsch
Medecine Science,1994

The making of a fly.
The genetics of animal design
Peter Lawrence
Blackwell Scientific Publications Ltd,1992

2) Actualisation

Le génome de la drosophile est aujourd'hui entièrement séquencé (Adam *et al.*, 2000). Il demeure néanmoins des "trous" ou des incertitudes, en particulier dans les zones répétées. Il est regrettable que ces incertitudes ne soient pas identifiées comme telles dans les banques de données. Le projet de séquençage ne prévoit pas de lever toutes ces incertitudes et il est probable que la version trois (la « release 3 » qui sera publiée en octobre 2001) soit la dernière.

La réalisation de "knock out" de gène est aujourd'hui possible chez la drosophile grâce à une technique de recombinaison mise au point par Rong et Golic (2000). Cette technique, délicate et laborieuse, est encore rarement citée dans la littérature scientifique. L'utilisation de l'interférence ARN (RNA-*i*), plus simple à réaliser, lui fait une concurrence certaine, puisqu'elle permet d'obtenir des phénotypes équivalents à des mutants nuls par simple transgénèse, et surtout, elle peut éteindre l'expression de gènes redondants pourvu qu'ils soient suffisamment homologues au transgène.

II- L'ovogenèse de *Drosophila melanogaster*

Les éléments transposables peuvent s'exprimer dans de nombreux tissus, mais, dans le cadre de l'étude de leur cycle et de l'instabilité génétique, nous nous sommes intéressés aux tissus germinaux. Nous avons étudié le cycle de mobilisation de *ZAM* en recherchant ses lieux d'expression, chez le mâle et la femelle à différents stades du développement, et c'est dans les ovaires que nous avons pu en observer quelques étapes.

Le système reproducteur femelle est composé de deux ovaires, des canaux génitaux, d'un utérus et de ses annexes (figure 23) (King *et al.*, 1970). Chaque ovaire est constitué de seize à vingt-cinq ovarioles disposées parallèlement.

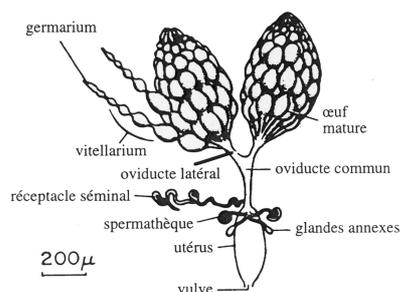


Figure 23 : Organe reproducteur femelle de *Drosophila melanogaster* (Grammont, 1998 d'après King *et al.*, 1970).

Un ovariole est un enchaînement linéaire de six à sept follicules ordonnés suivant leur stade de développement, du plus précoce à l'œuf mature, selon l'axe antéro-postérieur. Le germarium, en antérieur, contient les cellules souches et les follicules précoces en formation. Un follicule est formé d'un cyste de 16 cellules germinales, entouré d'une monocouche de cellules folliculaires somatiques. Au cours des stades suivants s'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, juste avant les oviductes latéraux (figure 24).

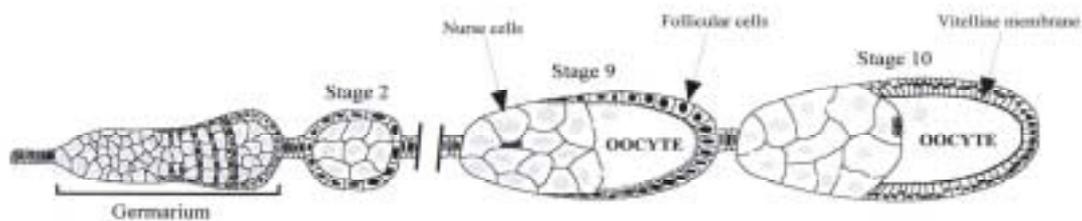
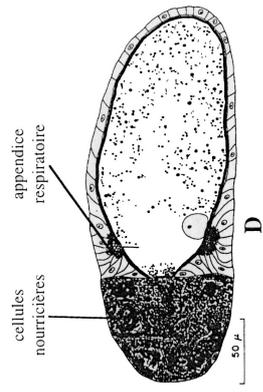


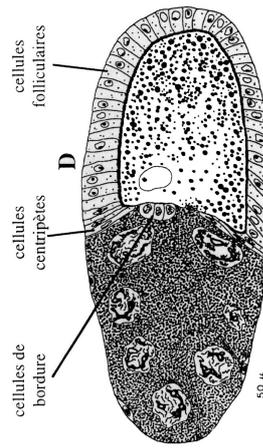
Figure 24 : Schéma d'un ovariole (Grammont, 1998).

Le développement des follicules a été divisé en 14 stades bien définis selon des critères morphologiques (King, 1970). Les stades prévitellogéniques (1-7) correspondent à la mise en place de l'axe antéro-postérieur du follicule. Dès le stade 6, l'ovocyte est localisé en postérieur du follicule. Les stades vitellogéniques (8-14) sont caractérisés par l'établissement de l'axe dorso-ventral, la migration de cellules somatiques, le transfert du cytoplasme des cellules nourricières vers l'ovocyte et la synthèse de la membrane de vitelline (figure 25).

Les stades d'intérêt dans notre étude sont le germarium et les stades 8-10. Lors de ces derniers stades, on peut observer la migration de cellules dites de bordure à travers les cellules nourricières puis leur positionnement face au noyau de l'ovocyte (position antéro-dorsal du noyau).

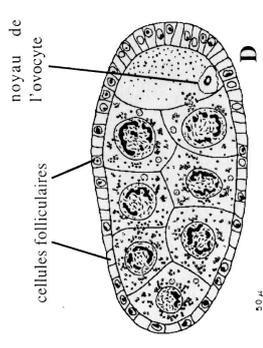
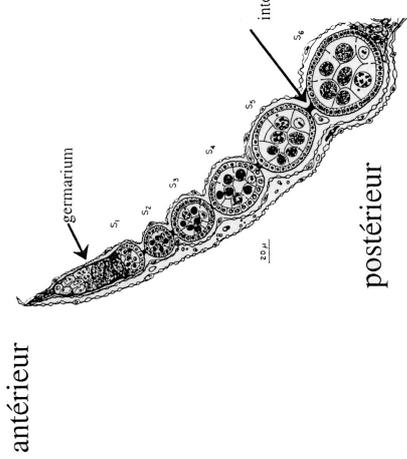


Stade 11: Des cellules folliculaires situées antérieurement se rassemblent pour former les deux appendices respiratoires. Les cellules nourricières se fragmentent et leur contenu cytoplasmique se déverse dans l'ovocyte.

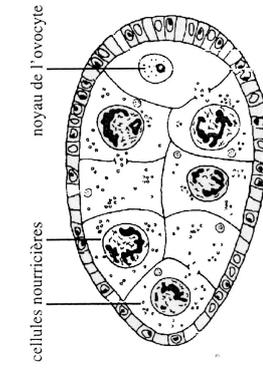


Stade 10B: Toutes les cellules folliculaires situées antérieurement commencent une migration centripète de manière à entourer complètement l'ovocyte. Les cellules de bordure se placent en face du noyau de l'ovocyte.

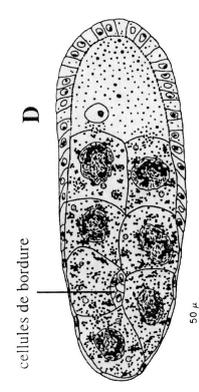
Stades 1 à 6: Les chambres d'œufs se présentent sous une forme ovoïde. L'ovocyte se situe au pôle postérieur de chaque chambre d'œuf. Les stades 1 à 6 sont définis en fonction de la taille des chambres d'œufs.



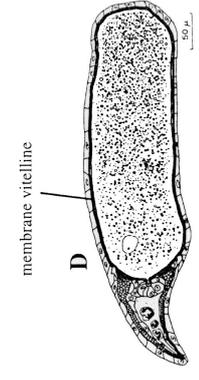
Stade 8: Le noyau de l'ovocyte adopte une position antéro-dorsale dans l'ovocyte. La taille de l'ovocyte devient plus importante que celles des cellules nourricières. Cette différence de taille s'intensifie jusqu'au stade 13.



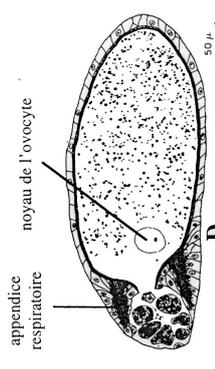
Stade 7: La chambre d'œuf devient plus allongée.



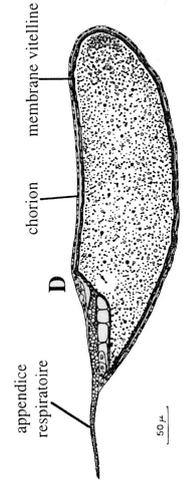
Stade 9: Certaines cellules folliculaires, appelées cellules de bordure, situées à l'extrémité antérieure de la chambre d'œuf, commencent à migrer entre les cellules nourricières pour atteindre la membrane plasmique de l'ovocyte.



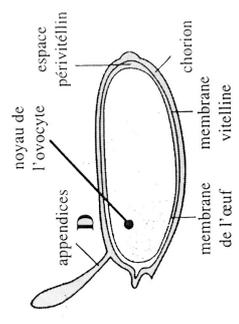
Stade 13: Une membrane vitelline synthétisée par les cellules folliculaires, entoure l'ovocyte.



Stade 12: Les filaments des deux appendices progressent antérieurement. Des fragments des cellules folliculaires sont encore visibles. L'ovocyte atteint son volume maximal.



Stade 14: La formation des appendices respiratoires est terminée. Le chorton, synthétisé par les cellules folliculaires, entourent la membrane vitelline.



Représentation schématique d'un œuf mature de *Drosophila melanogaster*. (Möhler, 1992)

Représentation schématique des différents stades de développement des chambres d'œuf: (King, 1970). Pour les stade 7 à 14, l'extrémité antérieure de la chambre d'œuf est à gauche et l'extrémité postérieure est à droite. Pour chaque schéma, la position du dos est indiquée par la lettre D.

Figure 25 : Description des stades folliculaires au cours de l'ovogenèse (King *et al.*, 1970).



MATERIALS AND METHODS



RESULTATS



Résultats

Quand je suis arrivée au laboratoire, j'ai tout de suite été intriguée par la nouvelle instabilité génétique que venait de mettre en évidence l'équipe de Chantal Vaury. Cette instabilité se manifestait dans une lignée établie au laboratoire de *Drosophila melanogaster* après une mutagenèse P/M. Les résultats que je vais présenter ici ont eu pour objet la recherche des déterminants génétiques de cette instabilité.

L'équipe de Chantal étudiait l'effet mutagène du *facteur I* sur un modèle utilisant le gène *white*. Ce gène code pour une protéine impliquée dans la couleur de l'œil chez la drosophile. Dans l'allèle w^{IR6} , le gène *white* est interrompu par l'insertion du *facteur I* dans le premier intron. Un nouveau site d'épissage est créé au site d'insertion. On observe deux types de messagers dans les mouches de la lignée W^{IR6} : le messager sauvage, correctement épissé, un messager non épissé, et un messager aberrant. Comme les messagers sauvages sont les seuls qui soient traduits en protéine active, cette mutation aboutit à une moindre production de la protéine White et donc à une modification de la couleur des yeux: Les yeux des mouches de la lignée W^{IR6} sont brun-orangé, au lieu du rouge brique des mouches sauvages (Lajoinie *et al*, 1995).

Dans le but d'identifier des modificateurs de ce phénotype, Chantal avait établi plusieurs lignées mutantes après une mutagenèse P/M sur la souche parentale W^{IR6} . C'est au cours de l'entretien de l'une de ces lignées qu'elle observa un mâle revertant, avec des yeux rouge brique comme le sauvage...

Chantal a établi une lignée avec ce mâle révertant, qu'elle a appelée W^{IR6REV} (article I). Cette lignée s'est révélée génétiquement instable car au cours de son entretien, de nouveaux mutants apparaissent sporadiquement.

La première partie de ce travail a consisté en l'identification moléculaire des mutations observées. C'est ainsi qu'ont été décrits ZAM et Idefix, deux rétrovirus endogènes de la famille des *errantivirus*. Les études de distribution génomique ont montré que ZAM et Idefix ont été mobilisés dans la souche W^{IR6REV} alors qu'ils sont stables dans la souche W^{IR6} (articles I et II). Chercher la cause de l'instabilité revenait à chercher la cause de la mobilisation de ces deux éléments. Dans la seconde partie, nous avons étudié les différentes étapes du cycle errantiviral dans la lignée W^{IR6REV} , où ZAM est amplifié, en comparaison avec la lignée W^{IR6} où

ZAM est en faible nombre de copies, afin de voir quelle était l'étape affectée. ZAM ayant été caractérisé avec un peu d'avance sur *Idefix*, je n'ai participé qu'à l'étude du cycle de ZAM (article III). L'ensemble des résultats de ces trois articles nous ont permis de construire les outils moléculaires nécessaires à la recherche des déterminants génétiques de l'instabilité, qui fera l'objet de la dernière partie. La méthodologie qui s'y rapporte est indiquée dans le chapitre Matériels et méthodes.

I- Caractérisation de ZAM et d'Idefix, deux nouveaux *errantivirus* mobilisés dans une souche *Drosophila melanogaster* WIR6REV(RevI)

Les analyses génétiques et moléculaires de la lignée WIR6REV ont rapidement montré que la nouvelle mutation responsable du phénotype révertant affectait le locus du gène *white*. D'après des résultats de *Southern* génomiques, la mutation était une nouvelle insertion d'environ 9 kb qui se situait 3 kb en amont du début de transcription de *white* (figure 1 article I).

A mon arrivée au laboratoire, Pascal Leblanc avait partiellement cloné et séquencé cette insertion. Les banques de données consultées indiquaient des homologies de cette séquence avec les gènes *env* d'éléments mobiles qui étaient encore appelés à l'époque des rétrotransposons à LTR.

Pascal n'a jamais pu cloner l'insertion entière (ni personne encore au laboratoire!). Il a baptisé ce nouvel élément "ZAM" par dépit... (Z'en Ai Marre!). J'ai préféré utiliser la technique de clonage "shot gun", qui, petit à petit, a permis le séquençage complet de ZAM et de révéler sa structure. Dans l'article I, nous montrons que ZAM a la structure d'un *errantivirus* complet et potentiellement actif (figures 2-3-7, article I).

Pascal a réalisé les *Southern* génomiques et les hybridations *in situ* sur chromosomes polytènes qui révèlent que l'ensemble des souches de *D. melanogaster* testées contiennent un faible nombre de copies de ZAM, exceptées deux souches de laboratoire: WIR6REV, et la souche Charolles issue d'une population naturelle (CHA) (figures 4 et 5, article I). Ces copies sont essentiellement sur les bras euchromatiques (figure 5). Ces résultats étaient confirmés au sein de l'équipe, en collaboration avec Patrizio Dimitri, par la technique FISH (Fluorescence *In Situ* Hybridization) sur les chromosomes polytènes (Baldrich *et al.*, 1997). L'augmentation du nombre de copies de ZAM entre la souche parentale WIR6 et la souche instable WIR6REV montre que ZAM a été fortement mobilisé dans cette dernière souche.

Par *Northern blots* à partir de mouches entières, Pascal a corrélié l'augmentation du nombre de copies de *ZAM* dans *WIR6REV* à l'accumulation de ses transcrits. A ce stade, on ignorait si cette accumulation était la conséquence de l'augmentation de l'élément *ZAM* dans *WIR6REV* ou si elle avait précédé son invasion du génome. Cette observation montre en tout cas un différentiel entre les souches *WIR6* et *WIR6REV* au niveau transcriptionnel. Cette modification de la quantité de transcrits de *ZAM* peut être due à une augmentation de la transcription et/ou de la stabilité des ARN. D'autre part, Pascal a identifié des transcrits épissés subgénomiques suggérant que l'expression de l'élément pouvait conduire à la synthèse de la protéine Env (figure 6, article I).

Pendant ce temps...

Au cours de l'entretien de la souche *WIR6REV*, Chantal isolait, de façon indépendante, de nouveaux mutants pour la couleur de l'œil et elle établissait de nouvelles lignées. Tous les mutants secondaires dérivant de la souche *WIR6REV* avaient la même couleur brun-orangé et ont permis l'établissement de lignées dites Rev II; la souche mère *WIR6REV* est devenue Rev I. Les mutations apparues dans les souches Rev II ont tout naturellement conduit à des lignées qui furent appelées Rev III. Un seul mutant Rev IV a été isolé à partir d'une lignée Rev III. L'ensemble des lignées dérivant du premier mâle *WIR6REV* a été désigné sous l'appellation générique de "souche instable Rev" (figure 1, article II).

Les analyses génétiques et moléculaires de chacun des 11 premiers mutants Rev II ont révélé que le responsable de la mutation était toujours le même élément, inséré à chaque fois au même site, à la base près, dans une seule orientation (figures 2,3 et 5, article II). Depuis, des insertions d'*Idefix* dans l'orientation inverse ont été isolées (Conte, Dastugue & Vaury¹). Laury Arthaud puis Valérie Calco ont cloné et séquencé cet élément qui ne pouvait prendre que le nom d'*Idefix* (Mais ne le répétez pas à la famille Uderzo!). *Idefix* possède comme *ZAM* une structure errantivirale avec un gène *env* (figures 3 et 4 article II). Quant aux trois mutations à l'origine des lignées Rev III, aucune n'affecte le locus *white*. Avec Caroline Conte, nous avons localisé génétiquement ces mutations, puis, en collaboration avec Patrizio Dimitri, nous avons montré par la technique *FISH* que de nouvelles insertions de *ZAM* ou d'*Idefix* étaient localisés dans ces *loci*. Dans le mutant RevIV, Caroline a montré par PCR qu'un deuxième *Idefix* était inséré au

¹ Conte C., Dastugue B. and Vaury C. "Promoter competition as a mechanism of transcriptional interference mediated by retrotransposons."

locus *white*, au même site que le premier mais en orientation inverse (figure 3, 6, article II).

Idefix est un élément répété présent dans toutes les souches testées (figure 7A). Un fragment interne particulier de l'élément complet n'a été détecté que dans la lignée parentale et la lignée Rev, suggérant qu'il est sous une forme incomplète dans les autres souches (figure 7B). Ces copies incomplètes sont localisées dans le chromocentre des chromosomes polytènes (figure 8). Le fragment particulier d'*Idefix* détecté par *Southern* augmente son nombre de copies dans les lignées Rev, révélant sa mobilisation récente dans ces lignées. Les résultats de *FISH* que j'ai obtenus sur les chromosomes polytènes confirment la mobilisation d'*Idefix*, qui passe de quatre copies euchromatiques dans la lignée parentale WIR⁶ à une vingtaine dans la lignée Rev.



Article I pp94-105

Invertebrate retroviruses: ZAM a new candidate in
D.melanogaster.

P.LebLANC, S.DessET, B.Dastugue and C.Vaury.

The EMBO Journal Vol.16 No.24 pp.7521-7531, (1997)

PMID: 9405380 PMCID: PMC1170351

DOI: 10.1093/emboj/16.24.7521

























Article II pp 106-118:

Mobilization of Two Retroelements, *ZAM* and *Idefix*, in a Novel Unstable Line of *Drosophila melanogaster*.

Sophie Desset, Caroline Conte, Patrizio Dimitri, Valérie Calco, Bernard Dastugue, and Chantal Vaury

Mol. Biol. Evol. 16(1):54-66. 1999

PMID: 10331252

























II- Etude du cycle errantiviral de ZAM.(article III)

Afin d'étudier le cycle de ZAM, nous avons réalisé différentes études d'expression de cet élément en privilégiant les tissus germinaux, sièges de la mobilisation de ZAM dans le génome des lignées Rev. Les premières hybridations *in situ* que Pascal a réalisées avec une sonde *env* ont montré que ZAM subit un contrôle transcriptionnel dans les ovaires et les gonades des embryons. Dans ces tissus en effet, les transcrits de ZAM ne sont détectés que dans la lignée Rev (figure 2). De plus, aucun transcrit n'a été détecté dans les testicules des mouches Rev ou WIR⁶ (non montré).

ZAM est transcrit dans tous les follicules dès les stades précoces de l'ovogénèse, dans les cellules somatiques folliculaires. Au cours du développement des follicules, l'expression se restreint à quelques cellules folliculaires en position postérieure et le taux de transcription culmine pendant le stade 10 (figure 2).

Nous avons ensuite recherché les protéines de ZAM dans les mêmes tissus, grâce aux anticorps spécifiques obtenus contre les protéines recombinantes Gag et Env. J'ai pris en charge la recherche de la protéine Env tandis que Pascal s'occupait de Gag. Nous n'avons détecté ces protéines que dans les ovaires (figure 3). Par microscopie confocale, j'ai montré dans des expériences de double immunomarquage que les protéines Gag et Env sont co-produites dans les cellules folliculaires dès le stade 9 (figure 4).

Tous les composants nécessaires à la formation de particules étaient présents dans les cellules folliculaires en postérieur des follicules des stades 9-10. Nous avons donc entrepris des détections de microscopie électronique en collaboration avec l'équipe de Franco Giorgi. Des particules virales, coïncidant avec la localisation de la protéine Gag, ont été observées dans les cellules folliculaires de la lignée Rev. Ces particules sont absentes de la lignée WIR⁶ (figures 5-6-7). Bien que nous n'ayons pu observer d'images de "budding", il semble que les particules soient transmises à l'ovocyte, puisque quelques unes sont détectées dans son cytoplasme (figures 6 et 7). La protéine Env ne partage pas entièrement la même localisation que les particules, suggérant que les particules ne sont pas recouvertes de la protéine Env (figure 8).























III- Localisation des déterminants génétiques responsables de l'instabilité génétique dans la lignée W^{IR6}REV.

1) Le LTR est suffisant pour diriger l'expression de *ZAM* dans les ovaires de la lignée Rev.

Afin de comprendre comment les déterminants génétiques de l'instabilité contrôlent l'expression de l'errantivirus *ZAM*, nous avons construit un plasmide contenant le LTR de *ZAM* et un gène rapporteur pouvant mimer l'expression de *ZAM in vivo*. Pour cela, nous avons amplifié un fragment de *ZAM* entre les nucléotides 1 et 473 et fusionné ce fragment au gène *LacZ* d'*Escherichia coli* dans le vecteur p7 (figure 26). Le fragment amplifié contient la totalité du LTR dans lequel se trouvent des séquences régulatrices putatives et le promoteur de *ZAM* (Leblanc *et al.*, 1997). Cette construction, appelée pZAM, a été injectée dans des mouches de la souche *yw*. Cette souche contient un faible nombre de copies de *ZAM* (non montré). Quatre lignées transformantes ont été rendues homozygotes pour le transgène et appelées *yw-pZAM*.

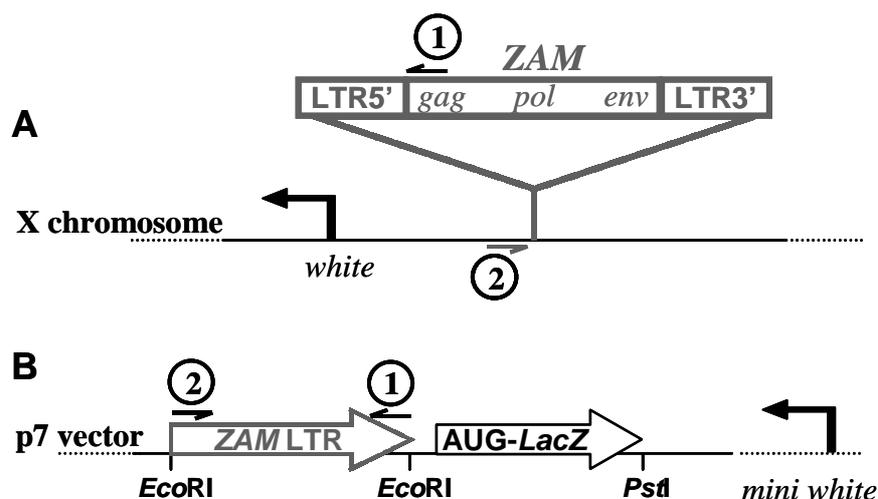


Figure 26 : Cloning of *ZAM* LTR in p7 vector.

A- Amplification of the *ZAM* LTR from the *white* locus of the RevI line. Primers 1 and 2, and genomic DNA are described in Materials and Methods.

B- Diagram of LTRZAM-*LacZ* construct in p7 vector. This construct called pZAM was used to generate the transgenic flies discussed in the text.

Nous avons recherché l'activité enzymatique β -galactosidase par un test de détection *in situ* dans les ovaires des lignées transgéniques dans différents contextes génétiques: nous attendions que les mouches portant le transgène ne l'expriment pas dans les contextes stables pour *ZAM* (lignée parentale W^{IR6}, lignée injectée *yw*) et que le profil d'expression de *ZAM* soit retrouvé dans un contexte Rev. Nous avons croisé des mâles transgéniques *yw-pZAM* avec des femelles de

lignées *yw*, WIR6, ou Rev. Les quatre lignées transgéniques ont ainsi été croisées. Les croisements réciproques ont été réalisés. Les ovaires des lignées F1 ont ensuite été disséqués, fixés et testés pour l'activité *LacZ*. Aucune coloration révélant l'activité *LacZ* n'a été détectée dans les femelles F1 issues des croisements avec *yw* (non montré) ou WIR6 (figure 27A). L'activité du gène rapporteur n'a été trouvée que dans les cellules folliculaires des ovaires de femelles F1 contenant un génome hétérozygote Rev, avec une localisation postérieure à partir des stades 6-7 (figure 27B). Le même résultat a été obtenu avec les quatre lignées transgéniques et dans les deux sens de croisement. Le LTR de *ZAM* possède donc toute l'information nécessaire pour diriger spécifiquement l'expression de l'élément dans les cellules folliculaires postérieures des ovaires et répondre au contrôle spécifique de souche du taux de transcrits.

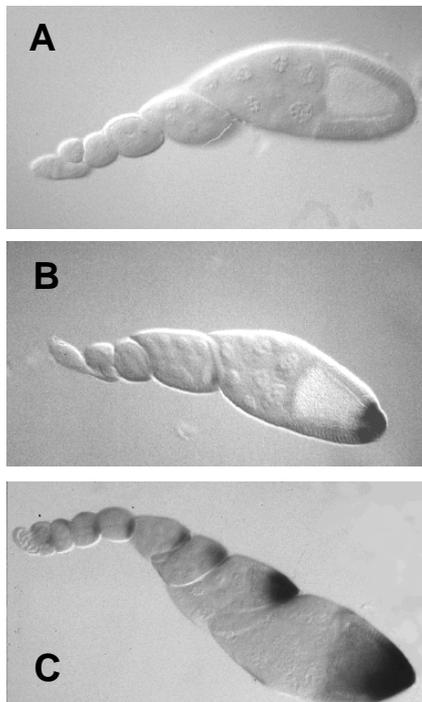


Figure 27 : Expression pattern of the *Escherichia coli* *LacZ* gene fused to *ZAM* 5'LTR in ovaries as a function of genetic determinants present in different lines.

A- Ovariolar with W^{IR6} genetic background stained for *LacZ* expression. No staining is detected. Dissected flies are $[X^{IR6}/X^{yw}; pZAM/+; +/+]$.

B- Ovariolar with a heterozygous Rev genetic background stained for *LacZ* expression. A typical *ZAM* expression pattern is observed in the follicle cells. Dissected flies are $[X^{Rev}/X^{yw}; pZAM/+; +/+]$.

C- Ovariolar with two X^{Rev} chromosomes stained for *LacZ* expression. Typical *ZAM* expression pattern is observed in the follicle cells. Dissected flies are $[X^{Rev}/X^{Rev}; pZAM/pZAM; +/+]$.

2) Le chromosome X provenant de la lignée Rev (X^{Rev}) permet l'accumulation des transcrits de *ZAM*.

Nous avons recherché quel chromosome était responsable du contrôle du taux de transcrits de *ZAM*. A l'aide d'une souche portant des chromosomes balanceurs, nous avons établi des lignées contenant le chromosome X de la lignée Rev (X^{Rev}) et les autosomes de la lignée parentale WIR6 (II^{IR6} ; III^{IR6}), lignée appelée $\langle X^{Rev}; II^{IR6}; III^{IR6} \rangle$. Nous avons aussi établi la lignée réciproque, $\langle X^{IR6}; II^{Rev}; III^{Rev} \rangle$, possédant les autosomes de la lignée Rev (II^{Rev} ; III^{Rev}) et le chromosome X de la lignée parentale (X^{IR6}) (figure 28A). Chacune des quatre

lignées transgéniques *yw-pZAM* a été croisée avec les lignées $\langle X^{Rev}; II^{IR6}; III^{IR6} \rangle$ ou $\langle X^{IR6}; II^{Rev}; III^{Rev} \rangle$ pour réaliser un test *LacZ* tel que décrit figure 28B. Seules les femelles F1 ayant reçu le chromosome X^{Rev} présentent un profil d'expression typique de *ZAM* dans les ovaires colorés pour l'expression *LacZ* (figure 28C). Aucune coloration n'a été observée dans les ovaires de femelles issues des croisements avec les lignées portant les autosomes $II^{Rev}; III^{Rev}$ (figure 28C). Cela montre que les déterminants génétiques responsables du contrôle du taux des transcrits de *ZAM* sont portés par le chromosome X.

3) Le chromosome X^{Rev} a un effet dose sur la transcription de *ZAM*.

Afin de placer le transgène *pZAM* dans les conditions de transcription de la lignée Rev, et pour simplifier des croisements ultérieurs, nous avons établi une lignée homozygote à la fois pour le transgène et le chromosome X^{Rev} , à partir d'une souche transgénique et d'une lignée Rev nommée Rev III7y. La couleur jaune des yeux de cette dernière lignée permet de sélectionner dans la descendance les mouches possédant le transgène, car le marqueur de sélection du transgène est le gène *white*. Nous avons nommé la lignée obtenue Rev-*pZAM*. Nous avons testé l'activité enzymatique *LacZ* dans les ovaires de la lignée Rev-*pZAM* en même temps et dans les mêmes conditions que des ovaires de femelles F1 hétérozygotes $\langle X^{Rev}/X^{yw}; pZAM/+ \rangle$. Les ovaires de la lignée Rev-*pZAM* ont une coloration due à l'activité *LacZ* plus intense que les ovaires des mouches contenant un seul chromosome X^{Rev} et un seul transgène (figure 27C). L'activité *LacZ* est intense dans tous les follicules, y compris dans les stades les plus précoces. Pour obtenir la même intensité de coloration dans les ovaires de mouches hétérozygotes, il faut poursuivre la coloration toute la nuit (non montré).

Nous avons testé si cette différence de coloration était due au nombre de copies du transgène. Pour cela, nous avons croisé des mâles homozygotes de la lignée Rev-*pZAM* avec des femelles Rev III7y, sans transgène, ou des femelles de la lignée transgénique *yw-pZAM*, sans chromosome X^{Rev} . Les femelles F1 qui possèdent deux copies du chromosome X^{Rev} et un seul transgène ont des ovaires avec une coloration maximale (coloration intense dans tous les follicules) lorsqu'ils sont testés pour l'activité *LacZ*, comme les ovaires de la figure 27C. Les femelles F1 possédant deux transgènes et un seul chromosome X^{Rev} ont des ovaires moins colorés, de même intensité que dans la figure 27B. L'ensemble de ces résultats montre un effet dose du chromosome X^{Rev} sur l'expression de *ZAM*.

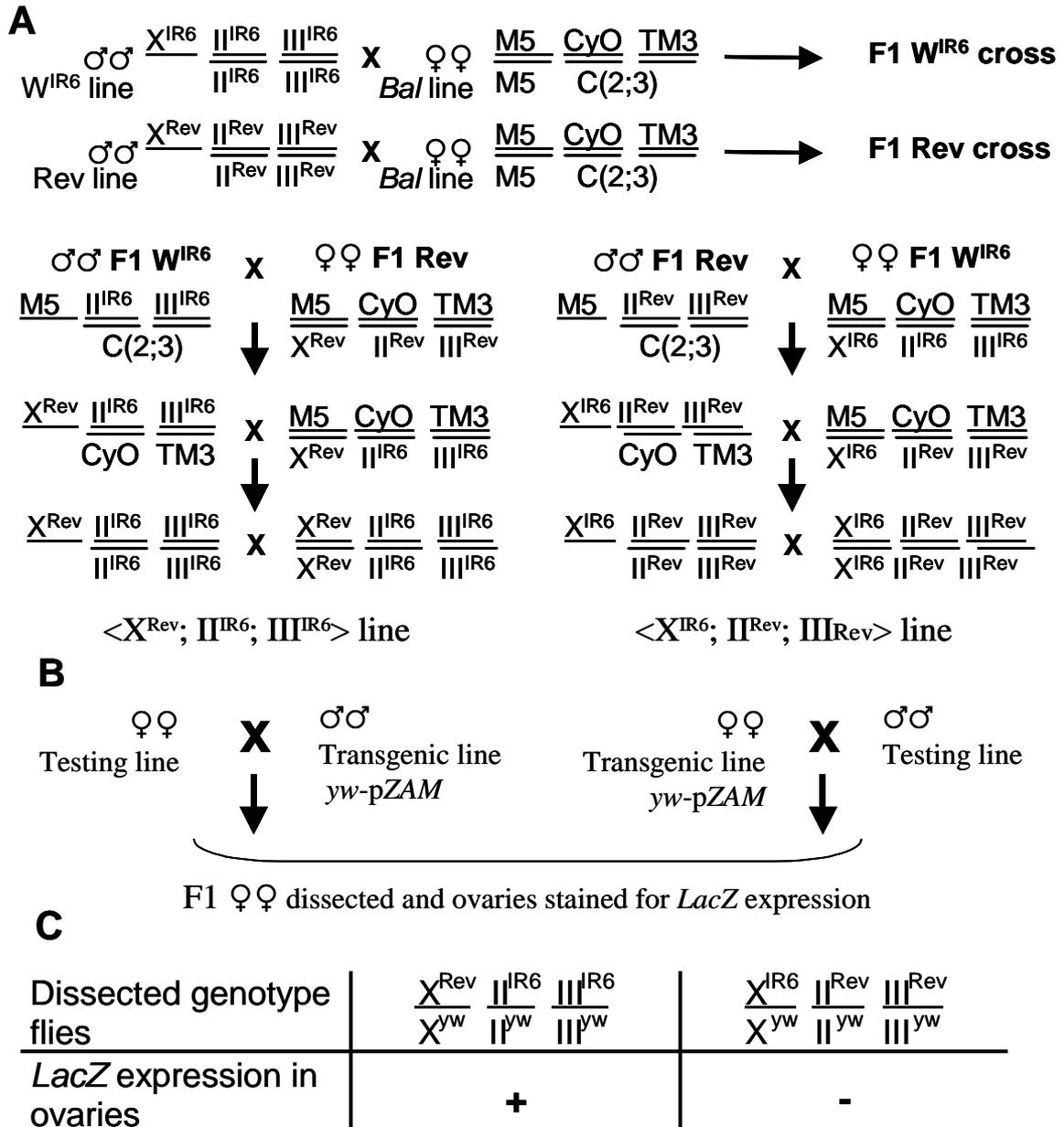


Figure 28 : Chromosomal location of genetic determinants controlling *ZAM* expression. **A-** Crosses performed to obtain two lines $\langle X^{Rev}, II^{IR6}, III^{IR6} \rangle$ and $\langle X^{IR6}, II^{Rev}, III^{Rev} \rangle$ with sexual or autosomal chromosomes from the Rev line respectively.

B- *LacZ* assay experiment.

C- Results of *LacZ* assay experiments described in B- carried out with each line established in A. The *LacZ* assay experiments was carried out with four independent transgenic lines. pZAM transgene is present as a single copy in the hemi-genome $X^{yw} II^{yw} III^{yw}$. Its location is not indicated since it is variable depending on the transgenic line used for the experiment. (+): staining observed in the follicle cells. (-): absence of staining.

4) Localisation des déterminants génétiques contrôlant la transcription de *ZAM* en 20A2-20A3.

Nous avons établi des lignées recombinantes possédant un chromosome X mosaïque contenant des portions du chromosome X^{Rev}, provenant de la lignée Rev, et des portions du chromosome X de la lignée *yellow2*. La lignée *yellow2* possède six marqueurs phénotypiques dispersés sur le chromosome X pour identifier l'origine génétique de chaque portion du chromosome X (figure 29A). Nous avons croisé des mâles *yellow2* avec des femelles Rev, puis les individus F1 entre eux pour permettre des recombinaisons méiotiques des femelles hétérozygotes. Les mâles F2 recombinants ont été croisés individuellement avec des femelles contenant les chromosomes X balanciers *FM6/FM7* pour établir les lignées recombinantes. Une cinquantaine de lignées a été ainsi établie. Nous avons ensuite recherché la présence des déterminants génétiques dans chaque lignée recombinante. Pour cela, nous avons réalisé un test *LacZ* tel que décrit figure 28B: nous avons croisé des femelles de chaque lignée recombinante avec des mâles de la lignée transgénique *yw-pZAM*, puis coloré les ovaires de la descendance pour l'activité *LacZ*. Le test est positif si les ovaires sont colorés et désigne alors les lignées possédant les déterminants génétiques responsables de la transcription du transgène.

Les résultats du test de chaque lignée ont été examinés séparément pour chaque marqueur. Un seul marqueur phénotypique, localisé près du centromère, ségrège avec les déterminants que nous recherchons: les 24 lignées présentant un phénotype *forked* sauvage, qui possèdent donc la région du chromosome X^{Rev} située entre le centromère et ce marqueur, ont un test positif, c'est à dire que les ovaires sont colorés. Inversement, 20 lignées sur 24 possédant le marqueur *forked* muté, indiquant que la région centromère-*forked* provient du chromosome X^{*yellow2*}, ont un test négatif (pas de coloration). Les 4 lignées positives possédant le marqueur *forked* muté proviennent sans doute des rares événements de recombinaison intervenus entre *forked* et le centromère, qui sont séparés de 9,3 cM. Les déterminants génétiques se situent donc entre les positions cytogénétiques 15F et 20F (figure 29B).

Afin de poursuivre la localisation des déterminants, nous avons utilisé des lignées possédant des délétions du chromosome X situées entre 15F et le centromère, en choisissant d'abord les lignées proposées dans le kit de délétion du Stock Center de Bloomington. Chacune des lignées déficientes *Df(1)* a été croisée avec la lignée homozygote Rev-pZAM possédant deux copies du transgène pZAM

et deux copies du chromosome X^{Rev} . Les ovaires des femelles F1 de génotype $\langle X^{Rev}/Df(1); pZAM/+ \rangle$ ont été disséqués puis colorés pour l'activité *LacZ*. Nous avons fait l'hypothèse que la délétion couvrant le locus des déterminants génétiques, délétion correspondant à un mutant nul de ces déterminants, permettrait une coloration maximale comme observée avec le génotype $X^{Rev}/X^{Rev} pZAM/+$ (figure 27C). Effectivement, la lignée *Df(1) DCB1-35-b* possédant une déficience située entre les positions 19F et 20F permet cette coloration maximale des ovaires. Les autres lignées testées ont des ovaires moins colorés comme ceux des mouches hétérozygotes X^{Rev} . Ces résultats désignent la région 19F-20F comme contenant les déterminants responsables de la transcription de *ZAM*. De plus, ils nous donnent une indication sur le contrôle spécifique de souche: Si le génotype $X^{Rev}/Df(1)$ augmente le taux de transcrits du transgène par rapport au contexte hétérozygote X^{Rev} , cela signifie que la déficience *Df(1)* est une perte de fonction d'un système répressif (c'est à dire diminuant le taux de transcrits) et que les déterminants portés par X^{Rev} ne sont donc pas des activateurs (augmentant le taux de transcrits).

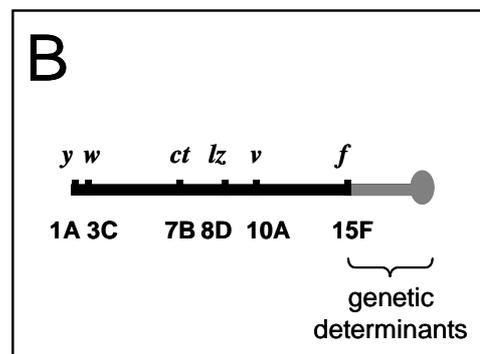
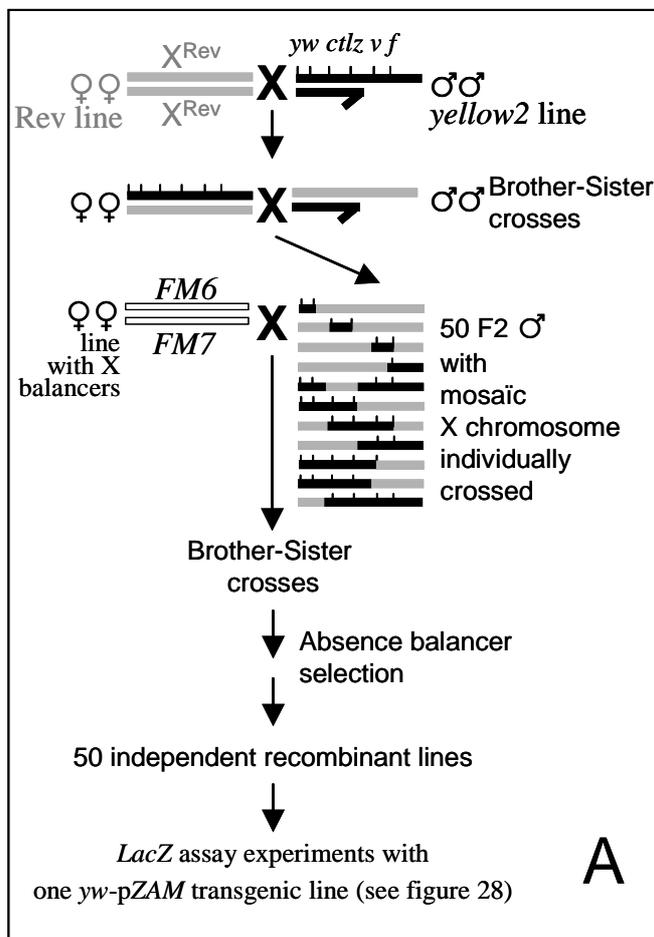


Figure 29 : Localisation of the genetic determinants present on the X^{Rev} chromosome and responsible for *ZAM* transcription.

A- Recombinant lines establishment.

B- Localisation of the genetic determinants deduced from *LacZ* assay experiments carried out with each recombinant line. X chromosome is represented with phenotypical markers and their respective position. Centromere is indicated by disc.

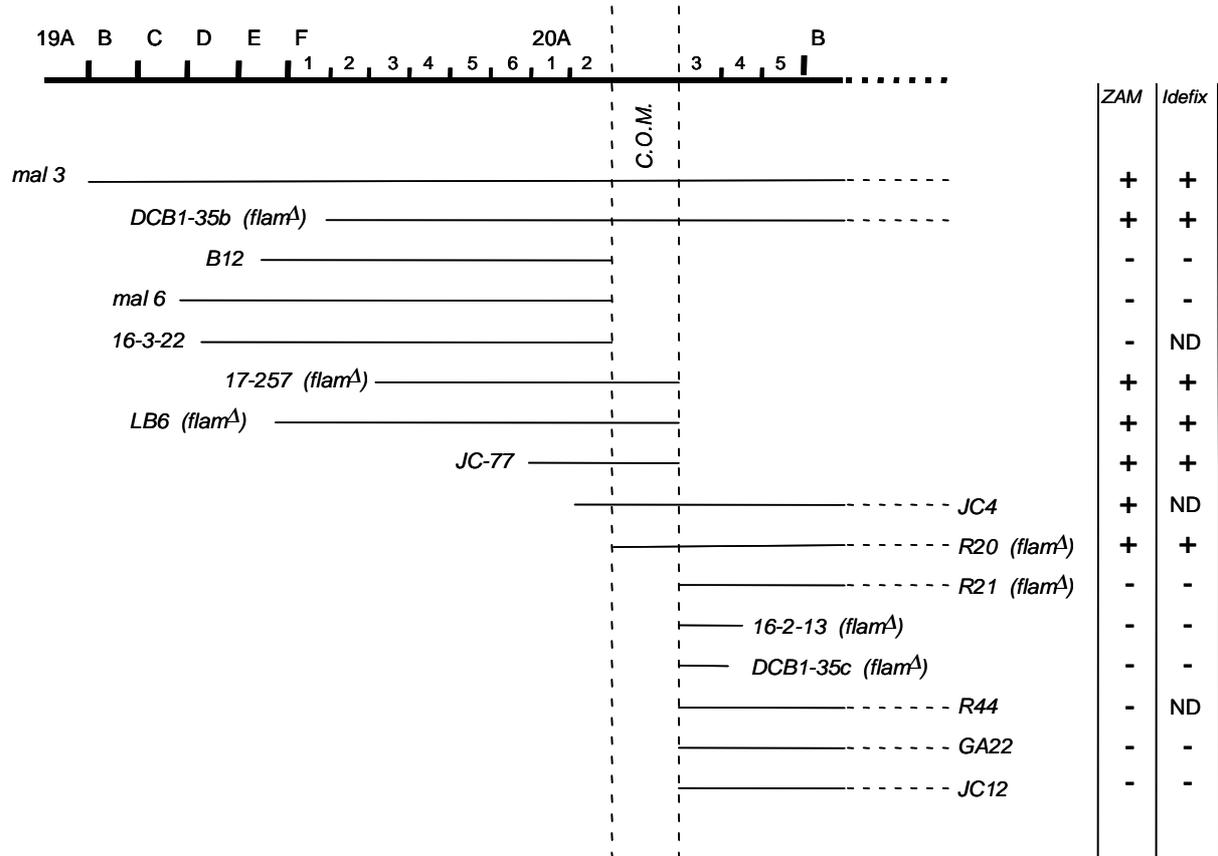


Figure 30 : Structure of deficiencies used for cytogenetic mapping of C.O.M. The organisation of the chromosome is represented at the top. Lines correspond to deficiencies. Dissected flies have a $X^{Rev}/Df(1)$ genotype and one copy of transgene (pZAM or pIde, as indicated in columns). Ovaries were stained for *LacZ* activities. Deficiencies designed *flam^A* are *flam* permissive. (+): Intense staining as in figure 2C. (-): Weak staining due to one copy of the X^{Rev} chromosome and one copy of the transgene as in figure 2B.

Nous avons alors répété cette expérience avec d'autres lignées déficientes pour préciser davantage le locus. L'ensemble des lignées testées est représenté figure 30. La synthèse des résultats obtenus permet de définir la région du chromosome X^{Rev} située en 20A2-20A3 comme porteur des déterminants génétiques responsables de l'expression de ZAM.

5) La mobilisation d'un deuxième *errantivirus*, *Idefix*, est contrôlée par le chromosome X^{Rev} .

Idefix est un *errantivirus* qui, comme ZAM, est mobilisé dans la lignée Rev (article II). *Idefix* possède un site chaud d'insertions au locus *white*. Nous avons isolé un nouveau mutant dans la lignée précédemment établie $\langle X^{Rev}; IIIR6; IIIIR6 \rangle$ qui possède les autosomes de la lignée $WIR6$, dans laquelle *Idefix* est en faible

nombre de copies, et le chromosome X de la lignée Rev (figure 28A). Ce mutant a les yeux de couleur brun-orangé comme les mutants RevII. L'analyse génétique de ce mutant a montré que la mutation se trouve sur le chromosome X^{Rev} au locus *white*. L'analyse moléculaire a montré qu'il s'agit d'une insertion *de novo* d'*Idefix* au site précédemment décrit (article II), dans l'orientation 5'-3' par rapport au gène *white* (non montré).

Nous avons voulu savoir si *Idefix* avait été fortement amplifié dans cette lignée <X^{Rev}; II^{IR6}; III^{IR6}> comme il l'a été dans la lignée Rev. Nous avons réalisé des expériences *FISH* avec une sonde spécifique d'*Idefix* sur des chromosomes polytènes de glandes salivaires de troisième stade larvaire provenant de la lignée où a été isolé ce mutant RevII. Une vingtaine de copies d'*Idefix* est répartie sur tous les bras euchromatiques de cette lignée où le chromosome X^{Rev} a été introduit (figure 31). Les autosomes qui dérivent de la lignée parentale W^{IR6} ne possédaient auparavant que trois copies d'*Idefix* sur les bras euchromatiques (figure 8, article II). Comme attendu, le chromosome X possède des copies euchromatiques d'*Idefix* qui étaient déjà présentes sur le chromosome X^{Rev}. Ces résultats montrent qu'*Idefix* a été récemment amplifié dans cette nouvelle lignée, et que le chromosome X^{Rev} est suffisant pour permettre cette mobilisation.

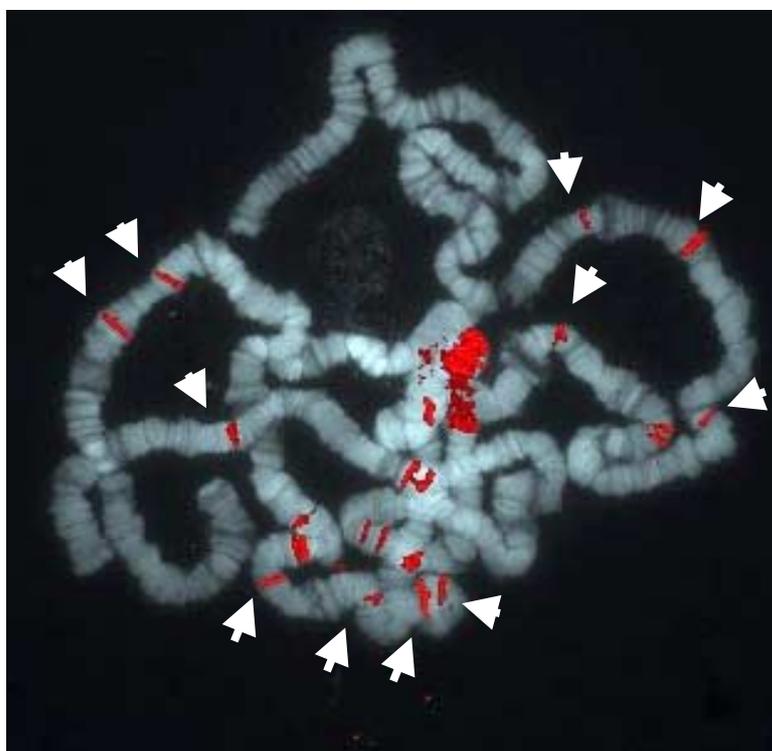


Figure 31 : *FISH* mapping of *Idefix* insertions on polytene chromosomes from the salivary glands of the <X^{Rev}; II^{IR6}; III^{IR6}> established in figure 3A. Hybridization signals of *Idefix* on the autosomes have increased in copy number in this line. The signals are merged with DAPI staining of salivary gland chromosomes. Some signals are indicated with arrows.

6) Un Centre Organisateur de Mobilisation (C.O.M.) localisé en 20A2-A3 contrôle trois *errantivirus*

Idefix est exprimé comme *ZAM* dans les cellules folliculaires des ovaires, mais uniquement dans les stades les plus précoces, c'est à dire dans le *germarium* (Tcheressiz *et al.*, 2002). Comme pour *ZAM*, les transcrits d'*Idefix* ne sont détectés que dans la lignée Rev. Des lignées de mouches transgéniques contenant une construction avec le gène rapporteur *LacZ* sous le contrôle du LTR d'*Idefix* ont été établies, ainsi que des lignées homozygotes $\langle X^{Rev}/X^{Rev}; pIde/ pIde \rangle$ appelées Rev-*pIde*. Ces lignées ont été testées pour leur activité *LacZ* dans les ovaires:

- L'expression du gène rapporteur dépend du contexte génétique; les ovaires ne sont colorés dans les *germaria* qu'en contexte Rev.
- La coloration des *germaria* n'est détectée que lorsque les mouches possèdent au moins une copie du chromosome X^{Rev} . Elle est plus intense quand le chromosome X^{Rev} est homozygote (Tcheressiz *et al.*, 2002).

Afin de savoir si le contrôle du taux des transcrits d'*Idefix* était assuré par les mêmes déterminants génétiques que ceux contrôlant *ZAM*, nous avons testé les déficiences du chromosome X situées entre 19F et 20F. Nous avons croisé des individus de la lignée Rev-*pIde* avec des individus de chaque lignée *Df(1)*, puis analysé les femelles F1 de génotype $\langle X^{Rev}/Df(1); pIde/ + \rangle$. Les ovaires de ces femelles ont été disséqués puis colorés pour l'activité *LacZ*. Les résultats de colorations sont strictement corrélés à ceux préalablement obtenus pour *ZAM* (figure 30). Ils désignent la même région 20A2-A3 comme responsable de l'expression d'*Idefix* sur le chromosome X^{Rev} .

Le gène *flamenco* (*flam*), responsable du contrôle transcriptionnel et de la mobilisation de l'*errantivirus gypsy* a été localisé dans la même région 20A2-A3 (Prud'homme *et al.*, 1995). Trois *errantivirus* subissent donc un contrôle transcriptionnel par cette région que nous nommons C.O.M. pour Centre Organisateur de la Mobilisation.



DISCUSSION

I *ZAM* et *Idefix* sont deux *errantivirus*

1) Structure des deux insertions *ZAM* et *Idefix* au locus *white*.

Deux éléments transposables, *ZAM* et *Idefix*, se sont nouvellement insérés au locus *white* chez *Drosophila melanogaster*. Ces deux éléments ont été clonés et séquencés (articles I et II). Ils ont une structure génomique complète d'*errantivirus*: Deux LTR encadrent trois cadres de lecture ouverts *gag*, *pol* et *env* susceptibles de coder des protéines actives de type Gag, Pol, Env (figure 2, article I ; figure 3, article II). Au moment de leur découverte, seulement cinq *errantivirus* complets avaient été décrits : *gypsy* (Marlor *et al.*, 1986), *17.6* (Saigo *et al.*, 1984), *297* (Inouye *et al.*, 1986), *Tom* (Tanda *et al.*, 1988) et *TED* (Friesen & Nissen, 1990). Ils étaient encore appelés LTR-rétrotransposons avec gènes d'enveloppe.

ZAM et *Idefix* ont la structure d'éléments actifs. Ils se sont amplifiés dans le génome de la lignée Rev où nous les avons identifiés. La découverte d'un rétrogène d'enveloppe de *ZAM* dans la souche Charolles montre que cet élément a la capacité d'être rétrotranscrit et que son ADNc peut être inséré dans les chromosomes de la lignée germinale (figure 8, article I). L'analyse des séquences de chacun des cadres de lecture dans les deux éléments met en évidence des séquences protéiques prédictives contenant les domaines spécifiques des rétrovirus (figure 3, article I ; figure 4, article II) :

- Dans le gène *pol*, les quatre domaines PR, RT, RH, IN sont présents dans cet ordre qui est propre aux *retroviridae* et aux *metaviridae*.
- Chacun des gènes *env* possède un peptide signal de sécrétion, un domaine SU (N-terminal) et un domaine TM (C-terminal) ainsi qu'un site potentiel de clivage entre ces deux sous-unités. On trouve aussi des sites potentiels de N-glycosylation (figure 7, article I).
- Le gène *gag* ne possède pas les caractéristiques des gènes *gag* de *retroviridae*, mais on constate des homologies avec les protéines Gag des autres *errantivirus*.

Les régions U5, R et U3 composant les LTR ont été définies pour les deux *errantivirus* par analyse des séquences, et confirmées expérimentalement : dans le cas de *ZAM* (article I), dans le cas d'*Idefix* par Tcheressiz *et al.*, (2002). Les régions 5'UTR de *ZAM* et d'*Idefix* contiennent chacune le site de liaison au primer d'initiation de la rétrotranscription (PBS). Les PBS sont tous deux homologues à un ARNt de serine. Ces régions 5' non traduites sont particulièrement longues : 1300 pb pour *ZAM* et 400 pb pour *Idefix*. Nous avons identifié des séquences répétées dans les deux éléments et une duplication directe de 300 pb dans *ZAM* (figure 32).

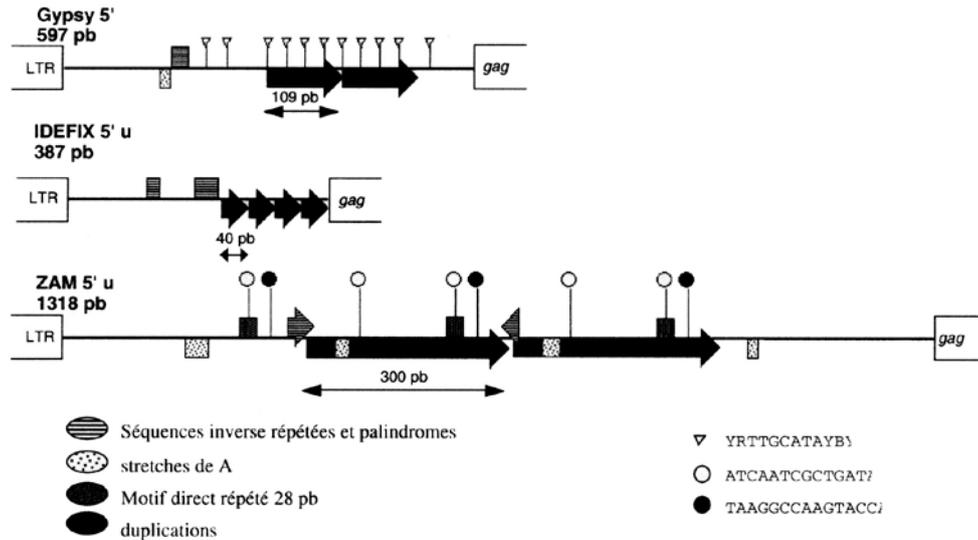


Figure 32 : Structure des régions 5' non traduites de trois *errantivirus* : *gypsy*, *ZAM* et *Idefix*.

Chez les *retroviridae*, ces régions, beaucoup plus courtes, contiennent des sites de liaison à des protéines régulatrices cellulaires ou accessoires. Dans l'élément *gypsy*, la région 5'UTR a une taille moyenne et contient une séquence barrière (Geyer & Corces, 1992). Cette barrière est constituée d'un motif répété douze fois. Parmi ces douze motifs, huit sont localisés dans une duplication de 109 pb, le tout étant un site de liaison à la protéine *Su(Hw)*. Caroline Conte a étudié l'impact mutagène de la région 5' non traduite de *ZAM* et mis en évidence ses propriétés *enhancer* dans les yeux (Conte, Dastugue & Vaury, 2002). D'autre part, Carine Meignin a mis en évidence un IRES (Internal Ribosome Entry Site) dans chacun des deux éléments (Meignin *et al.*, soumis¹). Pour l'instant, le rôle de ces régions dans la biologie de *ZAM* et d'*Idefix* n'est pas connu.

ZAM et *Idefix* ont un PBS présentant une séquence complémentaire avec les ARNt serine. Par analogie avec la classification utilisée pour les *ERV*, ils appartiennent à une même famille que l'on peut nommer *IERV-S* (*Insect ErrantiVirus*) selon Terzian, Péliesson & Bucheton (2001), tandis que *gypsy*, utilisant l'ARNt lysine est un *IERV-K*.

¹Meignin C., Bailly J.L., Dastugue B. and Vaury C. " The 5' untranslated region of a *Drosophila* retroelement, *Idefix*, promotes an internal translation initiations. "

2) Cycle errantiviral de *ZAM*

Des particules de *ZAM* sont détectées dans les cellules folliculaires de stades 9-10.

Les étapes du cycle errantiviral ont été étudiées pour *ZAM* en comparant l'expression de cet élément dans la lignée Rev et la lignée parentale WIR⁶, depuis les transcrits jusqu'aux particules. Tous les composants nécessaires au cycle rétroviral sont présents dans les ovaires de mouches de la lignée Rev, et sont absents dans la lignée WIR⁶. Nous observons des particules virales spécifiques de *ZAM* dans les cellules folliculaires postérieures de cette même lignée dans les follicules des stades 9-10 (article III).

Les particules de *ZAM* passent dans l'ovocyte par une étape extra-cellulaire.

Les insertions de *ZAM* étant transmissibles verticalement, on attendait effectivement d'observer l'expression de cet élément dans un tissu germinale de la lignée où il est mobilisé. Cette expression n'est détectée que dans les cellules somatiques de l'ovaire, ce qui suppose une étape extra-cellulaire avant d'atteindre l'ovocyte. Des particules sont détectées dans l'ovocyte, suggérant que cette étape a bien lieu. Mais on ignore encore comment se fait le passage.

Nous n'avons pas observé d'image de bourgeonnement ni de particules enveloppées entre l'ovocyte et les cellules folliculaires. La même absence avait été remarquée lors de la recherche de particules de l'élément *gypsy* dans le même tissu (Lécher, Bucheton & Péliesson, 1997). Nos données montrent que les particules pourraient utiliser le trafic de vitellogénine pour entrer dans l'ovocyte (discussion, article III) : Des infections indépendantes de la protéine d'enveloppe ont été décrites chez le poulet où les protéines Gag ont la propriété de s'assembler avec la protéine Env, mais aussi avec des protéines cellulaires glycosylées, pour former des particules (Dong, Roth, & Hunter, 1992; Young *et al.*, 1990). La vitellogénine étant glycosylée, on peut penser qu'elle s'associe avec Gag.

Si nous n'avons pu mettre en évidence de particules enveloppées ou d'image de bourgeonnement, rien n'indique que ce bourgeonnement n'ait pas lieu ou que la protéine d'enveloppe n'ait pas de rôle dans la biologie de l'élément, puisque nous n'avons pu tester son absence. Dans le cas de *gypsy*, des expériences ont montré que la transposition était indépendante de la protéine d'enveloppe (Chalvet *et al.*, 1999). Cependant, les auteurs ont détecté, après la publication de leurs travaux, des protéines présentant une immuno-réactivité croisée à l'anticorps anti-Env de

gypsy dans la souche où les expériences ont été réalisées (Christophe Terzian, communication personnelle). De telles protéines immuno-réactives auraient pu compléter les éléments gypsy défectifs qui étaient utilisés. Les résultats de Chalvet *et al.* concernant la protéine Env de gypsy sont donc à considérer avec toute la réserve nécessaire.

Quel est le destin des particules dans l'ovocyte ainsi infecté ?

Si nous observons bien quelques particules dans l'ovocyte, nous ignorons comment l'ADNc est conduit jusqu'au noyau et à quel moment se produisent les nouvelles intégrations. Dans le cas de gypsy, la production de particules virales est observée dans le contexte génétique *flam^p/flam^p*, mais les intégrations ont lieu dans la descendance de telles mouches, quel que soit le génotype de la génération suivante. C'est un effet maternel des femelles homozygotes *flam^p*. Si, comme on le croit, les particules gypsy infectant l'ovocyte sont à l'origine des nouvelles insertions, l'intégration n'a pas lieu dans l'individu issu de cet ovocyte au stade embryonnaire précoce, mais bien plus tard. Seules les intégrations dans les cellules germinales de cet individu sont détectées. Plusieurs frères et sœurs portent parfois la même insertion *de novo*, suggérant que ces événements insertionnels précèdent la méiose.

Dans le cas de *ZAM* et d'*Idefix*, chaque mutant apparu dans la lignée Rev a toujours été trouvé seul. Dans les deux cas, l'intégration est donc post-méiotique, à un moment situé entre la gamétogenèse tardive et l'embryogenèse précoce. La présence d'éléments *ZAM* sur le chromosome Y (Baldrich *et al.*, 1997) suggère que l'intégration n'a pas lieu dans l'ovocyte où sont détectées les particules, mais plus tard, au moins après la fécondation. Soit l'intégration est précoce: les insertions auraient alors lieu dans l'embryon issu de l'ovocyte avant les premières divisions. Soit l'intégration est tardive à la manière de gypsy: les insertions se produiraient dans ce cas dans les cellules germinales de l'adulte mâle ou femelle issu de l'ovocyte où sont détectées les particules.

Dans l'hypothèse d'une intégration tardive, les particules ne sont donc peut-être pas dirigées vers le noyau de l'ovocyte infecté. Ceci est à rapprocher d'une observation commune aux deux études de microscopie électronique dans les ovaires, tant pour gypsy (Lécher, Bucheton & Péliesson, 1997), que pour *ZAM* (article III) : Les particules ont la même morphologie que les particules immatures de *HIV* dans lesquelles la protéine Gag n'a pas encore été clivée (Goto, Nakai & Ikuta, 1998). Bien qu'aucun exemple ne soit décrit (mais aucun autre exemple de transmission d'un rétrovirus produit par la mère aux gamètes de ses enfants n'a été

~~~~~  
décrit à ma connaissance), on peut imaginer qu'une maturation des particules intervient postérieurement à leur arrivée dans l'ovocyte.

## **II Identification d'un locus de l'hôte contrôlant l'expression de *ZAM* et d'*Idefix* : Centre Organisateur de la Mobilisation (C.O.M.).**

### 1) Localisation de C.O.M.

*ZAM* subit un contrôle du taux de transcrits spécifique du contexte génétique: ses transcrits sont détectés dans la lignée Rev où il a été récemment amplifié, mais pas dans la lignée parentale W<sup>IR6</sup> où il est en faible nombre de copies. Afin d'étudier ce contrôle nous avons établi des lignées de mouches transgéniques portant un gène rapporteur dirigé par le LTR de *ZAM*. Dans la figure 27, nous montrons que le LTR de *ZAM* contient toute l'information nécessaire pour diriger spécifiquement l'expression du gène rapporteur dans les cellules folliculaire d'ovaires où *ZAM* s'exprime, ainsi que le contrôle spécifique de souche: L'expression du rapporteur n'est détecté que dans le contexte génétique de la lignée Rev comme c'est le cas pour l'élément *ZAM*.

A l'aide des lignées de mouches transgéniques, nous avons montré que le chromosome X de la lignée Rev (X<sup>Rev</sup>) possède les déterminants génétiques responsables de l'expression de *ZAM* (figure 28). En utilisant des lignées avec des marqueurs phénotypiques du chromosome X, puis des délétions de ce chromosome, nous avons localisé ces déterminants dans une région génomique de l'hôte, à la position cytogénétique 20A2-A3 (figures 29 et 30).

On dispose de peu d'information sur la région 20A2-A3 car les données de séquences fournies par le consortium de séquençage ne semblent pas fiables pour cette région : Un contig de 100 kb situé entre les positions cytogénétiques 19F et 20A5 a été publié dans la release 2 en mars 2001 (AE003574.2). Dans ce contig, on ne retrouve pas les séquences uniques identifiées par Valérie Robert qui a fait une marche sur le chromosome dans cette région (Robert, 2000). Il est probable que les séquences publiées par le consortium proviennent de clones réarrangés. En effet, une microdissection de la région 20A à partir de chromosomes polytènes avait montré que 60% des clones contiennent de l'ADN moyennement répété (Miklos et al, 1988 ; Miklos & Cotsell, 1990). D'après le consortium (Adam *et al.*, 2000) l'information manquante entre les deux contigs les plus proches, positionnés en 19F et 20A5, serait de 600kb.

Le chromosome X<sup>Rev</sup> contrôle la mobilisation d'*Idefix*, un autre *errantivirus* mobilisé dans la lignée Rev (figure 31). Comme *ZAM*, *Idefix* subit un contrôle transcriptionnel spécifique de souche dans les cellules folliculaires d'ovaires (Tcheressiz *et al.*, 2002). Nous avons montré que les déterminants génétiques contrôlant *Idefix* sont aussi localisés en 20A2-20A3 (figure 30). De plus, le gène *flamenco* contrôle la transcription et la mobilisation d'un autre *errantivirus*, *gypsy* (Péligon *et al.*, 1994 ; Prud'homme *et al.*, 1995). *Flamenco* a été localisé, grâce à un test de transposition de *gypsy*, à la même position cytogénétique 20A2-A3 (Prud'homme *et al.*, 1995). C'est pourquoi nous proposons d'appeler cette région Centre Organisateur de la Mobilisation (C.O.M.). La région COM de la lignée Rev sera notée COM<sup>Rev</sup>.

Des délétions du chromosome X avaient aussi été utilisées pour localiser *flamenco* par Prud'homme *et al.* (1995) : Les délétions affectant le locus *flamenco* (*flam*<sup>Δ</sup>) permettent de nouvelles insertions de *gypsy* dans la descendance des mouches *flam*<sup>1</sup>/*flam*<sup>Δ</sup>. Les dix délétions *flam*<sup>Δ</sup> communes à notre étude sont indiquées dans la figure 30. La plupart de ces délétions *flam*<sup>Δ</sup> permettent la transcription de *ZAM* et d'*Idefix* dans le contexte COM<sup>Rev</sup>/*flam*<sup>Δ</sup> sauf trois, localisées du côté proximal par rapport au centromère : *16-2-13*, *R21* et *DCB1-35c*. Cela indique que les déterminants génétiques contrôlant la transcription de *ZAM* et d'*Idefix* sont séparables du gène *flamenco*, et localisés plus distalement. Mais on ne peut exclure que ces trois délétions soient en dehors du gène *flamenco* et possèdent un allèle permissif pour la mobilisation de *gypsy*, comme cela est décrit pour la délétion *B12* (Robert, 2000). COM est défini comme le locus contrôlant au moins trois *errantivirus*, mais sa position proximale au centromère reste à préciser.

## 2) Caractérisation de l'instabilité génétique COM<sup>Rev</sup>

### **Origine des *errantivirus* *ZAM* et *Idefix* et des déterminants génétiques responsables de l'instabilité dans la lignée Rev.**

Les instabilités génétiques liées à la mobilisation d'éléments transposables chez *Drosophila melanogaster* sont toujours la rencontre d'un élément potentiellement actif avec un génome compétent pour sa mobilisation. La souche Rev est probablement le résultat d'une telle rencontre, puisque *ZAM* et *Idefix* sont amplifiés dans cette lignée, mais dans aucune autre lignée utilisée au cours de la mutagenèse P/M. On peut donc supposer qu'un ou plusieurs génomes ont apporté les éléments *ZAM* et *Idefix* complets, et qu'un autre a apporté la compétence. La compétence peut aussi être une conséquence imprévue de la mutagenèse.

Concernant l'origine de *ZAM* et d'*Idefix*, nous avons montré que la lignée parentale  $W^{IR6}$  contient deux copies euchromatiques de *ZAM* et quatre d'*Idefix*. Nous ne disposons d'aucune information sur la qualité de ces copies : nous ignorons si elles sont complètes et si elles peuvent être à l'origine des nouvelles insertions. Toutefois, les Southern génomiques réalisés semblent indiquer l'existence d'au moins un élément *ZAM* pleine longueur dans le génome  $W^{IR6}$  (Leblanc, Dastugue & Vaury, 1999). Nous n'avons pas analysé les chromosomes polytènes des autres souches utilisées pour la mutagenèse pour savoir si elles contiennent des copies euchromatiques.

Nous avons montré que la compétence du génome Rev pour la mobilisation de *ZAM* et d'*Idefix* sont les déterminants génétiques  $COM^{Rev}$  du chromosome X qui permettent l'accumulation de transcrits de *ZAM* et d'*Idefix*. Il faut donc établir si ces déterminants  $COM^{Rev}$  préexistaient sur un chromosome X transmis à la lignée Rev pendant la mutagenèse ou s'ils sont le résultat d'une modification moléculaire intervenue au *locus*  $COM^{IR6}$ . A l'issue de la mutagenèse, le chromosome X des mutants mâles sélectionnés était dérivé du seul chromosome X parental  $W^{IR6}$ . Or, nous avons montré que les déterminants  $COM^{Rev}$  sont absents de ce chromosome, puisque le LTR de *ZAM* et celui d'*Idefix* ne permettent pas l'expression du gène rapporteur dans le contexte génétique  $W^{IR6}$ . Après la mutagenèse, une souche contenant des autosomes balanceurs a été utilisée pour rendre homozygote les lignées mutantes. Des recombinaisons méiotiques du chromosome X ont pu se produire entre le chromosome X de la lignée  $W^{IR6}$  et cette souche. Pour savoir si les déterminants  $COM^{Rev}$  ont pu être transmis par le chromosome X de cette souche, il nous faudra tester sa capacité de régulation de *ZAM* et d'*Idefix*.

Les déterminants génétiques peuvent aussi être une conséquence de la mutagenèse. Patrizio Dimitri a recherché la présence d'*éléments P* dans la lignée Rev sur des chromosomes polytènes. Aucune trace de *P* n'a été mise en évidence au site 20A. Cependant, des mutations issues d'une mutagenèse P/M ne portant pas la signature d'un *élément P* ont déjà été décrites (Bingham, Kidwell & Rubin, 1982). D'autres mutations comme des réarrangements chromosomiques ont pu se produire au cours de la mutagenèse. Dans la région COM, les allèles *flamenco* permissifs sont décrits comme instables en présence d'éléments *gypsy* actifs (Prud'homme *et al.*, 1994). Un cas de réversion *de novo* du gène *flam* a été rapporté dans une population *flam<sup>p</sup>* établie à partir d'un seul chromosome X (Robert, 2000). Le génome  $W^{IR6}$  possède justement un allèle *flam* permissif (Alain Pélisson, communication personnelle). Une telle réversion du locus  $COM^{IR6}$  a pu se produire au cours de la mutagenèse pendant que le génome subissait des

insertions *de novo* de *P*. Les déterminants génétiques COM<sup>Rev</sup> que nous recherchons dans la lignée Rev seraient alors une modification moléculaire intervenue sur le chromosome X de la lignée W<sup>IR6</sup>.

### **L'instabilité COM<sup>Rev</sup> permet la mobilisation de deux *errantivirus*.**

Dans la lignée Rev, deux *errantivirus*, *ZAM* et *Idefix*, ont été récemment mobilisés et amplifiés. Peu de cas de co-mobilisations ont été décrits dans la littérature : *hobo* et *gypsy* dans la souche *Uc* (Lim *et al.*, 1983), *copia* et *Doc* dans une lignée isogénique (Pasyukova & Nuzdhin, 1993), enfin, un système de dysgénésie hybride chez *D. virilis* a été décrit dans lequel quatre éléments sont mobilisés (Petrov *et al.*, 1995). Contrairement à *ZAM* et *Idefix* qui sont deux *errantivirus*, ces cas ne concernent pas des éléments de même famille : *hobo* et *gypsy* sont de classe II et I respectivement, *copia* est un *pseudoviridae* alors que *Doc* est un rétroposon. Dans le système de dysgénésie de *D. virilis* les quatre éléments sont *Paris* (superfamille *mariner*), *Helena* (*LINE-like*), *Penelope* (un rétrotransposon de type *DIRS-I*) et *Ulysses* (un *metavirus*). Ces co-mobilisations suggèrent un mécanisme de régulation présentant des points communs.

Il existe une souche, nommée Charolles (CHA), issue d'une population naturelle, dans laquelle on dénombre une dizaine d'insertions euchromatiques de *ZAM* (Baldrich *et al.*, 1997 ; article I) alors que des *Southern* génomiques montrent que cette souche contient un faible nombre de copies d'*Idefix* (non montré). Cela suggère que *ZAM* a été mobilisé seul dans la lignée CHA. Aujourd'hui, *ZAM* est stable dans la souche CHA. On ne détecte aucun transcrite *ZAM* dans les ovaires et le gène rapporteur *LacZ* sous le contrôle du LTR de *ZAM* n'est pas exprimé dans le contexte génétique CHA (non montré). Trois hypothèses permettent d'expliquer le comportement différent d'*Idefix* dans les lignées Rev et CHA :

- *ZAM* a pu être mobilisé dans CHA selon un mécanisme et un contrôle différent de celui que nous observons dans Rev. Ce mécanisme pourrait être indépendant d'*Idefix*.

- Les déterminants génétiques permettant l'accumulation de transcrits de *ZAM* et d'*Idefix* seraient des allèles équivalents dans les lignées Rev et CHA. Dans ce cas, peut-être n'y avait-il pas d'élément *Idefix* complet susceptible d'être mobilisé dans CHA.

- Les déterminants génétiques dans la souche Rev affecteraient deux *loci* proches mais distincts qui seraient chacun spécifique de l'expression d'un élément, *ZAM* ou *Idefix*. La souche CHA n'aurait porté que des déterminants responsables de la mobilité de *ZAM*.

Pour les cas de co-mobilisations que nous avons cités, il faut préciser que ces éléments peuvent être mobilisés indépendamment dans d'autres instabilités génétiques. Les mêmes hypothèses peuvent sans doute être proposées à ces cas.

### **La région COM contrôle l'expression d'au moins trois *errantivirus*.**

Les résultats que nous venons de présenter et ceux de Prud'homme *et al.* (1995) montrent que la région COM contrôle l'expression de trois *errantivirus* : *gypsy*, *ZAM*, et *Idefix*. Mais il semble que ce ne sont pas les mêmes allèles qui permettent l'expression de l'un et des autres. Les allèles *flamenco* capables de mobiliser *gypsy* sont dits permissifs, les autres sont dits restrictifs. Un test de transposition de *gypsy* au locus *ovo<sup>D1</sup>* permet de connaître le statut de *flamenco* d'une souche quelconque (Péligsson *et al.*, 1997). La lignée Rev ne permet pas la mobilisation de *gypsy* selon les résultats du test *ovo<sup>D1</sup>* (Alain Péligsson, communication personnelle) ; la région COM<sup>Rev</sup> contient donc un allèle *flam* restrictif. Inversement, l'allèle permissif *flam<sup>l</sup>* ne permet pas l'expression du rapporteur *LacZ* sous le contrôle du LTR de *ZAM* ou de celui d'*Idefix* dans le contexte *flam<sup>l</sup>/flam<sup>l</sup>* (non montré). D'autre part, le test *ovo<sup>D1</sup>* réalisé sur la lignée parentale W<sup>IR6</sup> a montré que cette souche porte un allèle permissif de *flam*. Dans l'hypothèse d'une modification moléculaire de la région COM pendant la mutagenèse, de W<sup>IR6</sup> à Rev, cette modification aurait abouti à la fois à la perte de l'allèle *flam<sup>p</sup>* et à l'acquisition des déterminants COM<sup>Rev</sup>.

On suppose que le génome de drosophiles était permissif pour les *errantivirus* jusqu'à l'arrivée de ces éléments (Péligsson *et al.*, 1997). Toutes les populations ont rencontré l'élément *gypsy* puisqu'il est présent dans tous les génomes analysés (Modolell, Bender & Meselson, 1983 ; Bayev *et al.*, 1984 ; Peifer & Bender, 1988; Vaury *et al.*, 1989). Dans la nature aujourd'hui, on trouve différents allèles *flam* restrictifs qui ont été sélectionnés selon un schéma évolutif propre à chaque population (Péligsson *et al.*, 1997). La même histoire peut être imaginée pour *ZAM* et *Idefix*. L'originalité du locus COM est que la même région de l'hôte soit responsable de la transcription dans les mêmes tissus folliculaires de trois *errantivirus* distincts, alors qu'ils sont arrivés probablement à des moments différents dans le génome.

La permissivité que l'on observe parfois dans des lignées instables comme MG pour *gypsy* ou Rev pour *ZAM* et *Idefix*, ne semble pas concerner les trois *errantivirus* à la fois, ce qui pourrait rendre compte de l'histoire de chaque population et donc de chaque locus COM. Il faudrait tester d'autres régions COM de statut *flam<sup>p</sup>* pour leur capacité d'expression de *pZAM* ou de *pIde* afin de savoir



en place dans la lignée Rev d'un contrôle post-transcriptionnel de *ZAM* et d'*Idefix*. On peut s'étonner par exemple de la rareté des particules de *ZAM* présentes dans l'ovocyte, en comparaison du nombre de celles produites par les cellules folliculaires (article III). Le fait que l'on observe encore des mutants après dilutions du génome Rev suggère que la compétence pour la mobilisation de la lignée Rev est maintenue. Cependant, on peut penser que dans le cycle errant viral de *ZAM* et d'*Idefix*, il existe un contrôle postérieur à celui sur le taux de transcrits empêchant de nouvelles insertions, contrôle qui serait lié au nombre de copies présentes dans le génome.

### **COM<sup>Rev</sup> est-il dominant pour la mobilisation de *ZAM* et d'*Idefix* ?**

Si l'on a bien mis en évidence une capacité de transcription de *ZAM* et d'*Idefix* propre à la souche Rev, on ignore si le taux de transcrits est relié au taux de mobilisation de ces éléments. Dans le cas de *gypsy*, le LTR permet l'expression du gène rapporteur *LacZ* dans les ovaires de mouches hétérozygotes *flam<sup>p</sup>* (Sarot *et al.*, 2001), mais il a été montré que de telles mouches hétérozygotes ne permettent pas la mobilisation de *gypsy* (Prud'homme *et al.*, 1995). Dans les mouches *flam<sup>p</sup>/flam<sup>p</sup>*, le taux d'expression du rapporteur est très augmenté, et il est corrélé à la mobilisation de *gypsy*, ce qui signifie que *flam<sup>p</sup>* est récessif (Prud'homme *et al.*, 1995). Nous observons aussi que le taux de transcription de *ZAM* et d'*Idefix* augmente avec le nombre de copies de COM<sup>Rev</sup>, mais nous ignorons son incidence sur la mobilisation de ces éléments. Il manque à notre système un test de transposition, comme le test *ovo<sup>D1</sup>* de *gypsy* ou un élément traceur, qui permettrait d'étudier directement la mobilisation de *ZAM* et d'*Idefix* et sa relation avec le taux de transcription. L'absence de ces données nous empêche de savoir si COM<sup>Rev</sup> est dominant ou récessif pour la mobilisation de *ZAM* et d'*Idefix*, mais il est semi-dominant pour la transcription de ces éléments.

3) Mécanisme du contrôle du taux de transcrits.

### **Le LTR de *ZAM* possède l'information nécessaire pour diriger une transcription tissulaire spécifique et subir un contrôle spécifique de souche.**

Nous avons montré que le LTR de *ZAM* possède tout à la fois l'information suffisante pour diriger spécifiquement le rapporteur *LacZ* dans les cellules folliculaires postérieures des ovaires et pour répondre au contrôle spécifique de souche. La région U3 du LTR des *retroviridae* est connue pour contenir les *enhancers* responsables de la spécificité tissulaire des éléments, ainsi que leur

région promotrice. C'est ce que Carine Meignin s'attache à démontrer au laboratoire.

**Les transcrits de *ZAM* et d'*Idefix* sont absents dans les contextes stables, en grande quantité dans la lignée Rev.**

Deux hypothèses peuvent expliquer l'accumulation de transcrits de *ZAM* et d'*Idefix* dans la lignée Rev : Un nouveau système activateur est mis en place dans la lignée Rev (activation du taux de transcription ou messagers stabilisés par recrutement de nouveaux facteurs), ou bien un système répressif est levé (levée d'une répression de la transcription ou dégradation des messagers stoppée par perte de facteurs). Les deux hypothèses rendent compte de l'effet dose des déterminants COM<sup>Rev</sup> sur la transcription (figure 27).

La délétion de la région COM (COM<sup>Δ</sup>) ne devrait pas augmenter le taux de transcription de *ZAM* et d'*Idefix* dans le cas d'un nouveau système activateur. Dans l'hypothèse d'un système répressif, COM<sup>Δ</sup> correspond à un allèle perte de fonction et devrait renforcer le phénotype, c'est à dire augmenter la quantité de transcrits. C'est ce que l'on observe : le contexte COM<sup>Rev</sup> / COM<sup>Δ</sup> permet une accumulation accrue des messagers exprimés à partir des transgènes p*ZAM* et p*Ide*, comme dans le contexte COM<sup>Rev</sup> / COM<sup>Rev</sup> (figure 30). Le contrôle spécifique de souche du taux de transcrits est donc un système répressif dans la lignée parentale W<sup>IR6</sup>, et une absence de ce système dans Rev. Ces expériences ne nous permettent pas de trancher entre une action pré ou post-transcriptionnelle, mais pour simplifier la suite du texte, les termes « répression » et « contrôle transcriptionnel » désigneront ce contrôle négatif sur le taux de transcrits de *ZAM* et d'*Idefix*. Par analogie avec le système *flamenco/gypsy*, les déterminants COM<sup>Rev</sup> correspondent à des allèles permissifs pour *ZAM* et *Idefix*, COM<sup>IR6</sup> à des allèles restrictifs.

**Par quel mécanisme *ZAM* et *Idefix* sont-ils réprimés dans la lignée W<sup>IR6</sup> ?**

Le fait que le contrôle transcriptionnel de trois *errantivirus* soit porté par la même région COM suggère qu'il existe un mécanisme commun de répression. Les mécanismes de répression décrits sont assurés soit par des répresseurs protéiques qui fixent une séquence cible et empêchent la transcription, soit par *trans-silencing* entre une séquence hétérochromatique et une séquence euchromatique homologue selon un mécanisme inconnu.

## Modèle protéique

Les rétrovirus possèdent dans la région U3 de leur LTR, non seulement des sites *enhancers*, mais aussi des sites de fixation pour des répresseurs protéiques. Par exemple, un récepteur nucléaire des hormones appelé ELP (Embryonal LTR-Binding Protein) réprime spécifiquement le *MoMuLV* dans les cellules embryonnaires de souris (Tsukiyama & Niwa, 1992 ; Tsukiyama *et al.*, 1992). Les facteurs de transcription peuvent jouer le rôle de répresseur : Chez *Drosophila melanogaster*, la protéine Lola réprime le *pseudoviridae copia* alors que la protéine Lola-like de *D. hydei*, espèce sans élément *copia*, est un activateur de *copia* (Cavarec *et al.*, 1997). Le modèle *lola/copia* suggère une évolution du gène *lola* dans *D. melanogaster* sous la pression de sélection des transpositions *copia*.

On peut imaginer semblable évolution d'une autre protéine qui serait spécifique des *errantivirus* et localisée au locus COM. Cela supposerait qu'une séquence ou une structure nucléique suffisamment commune aux *errantivirus* puisse être reconnue par cette protéine putative. L'action des répresseurs peut nécessiter des partenaires protéiques ou des cofacteurs. Ils peuvent aussi être relayés par d'autres protéines qui maintiennent un état réprimé. Les gènes codant de telles protéines sont des candidats pour le locus COM.

Le schéma évolutif pour obtenir cet éventuel nouveau répresseur est difficile à concevoir, puisque tous les éléments ne sont pas arrivés en même temps dans le génome, mais que chaque *errantivirus* a pu suffisamment s'amplifier avant d'être réprimé pour qu'on trouve sa trace dans le génome. Cela signifie que ce répresseur aurait évolué à chaque arrivée d'un nouvel élément. Or il a en plus comme contrainte de rester actif pour sa fonction première dans l'hôte. La modification d'une fonction protéique est donc une hypothèse peu probable. Plus simplement, la propriété de réprimer spécifiquement les *errantivirus* a pu être apportée pour chaque élément par une modification du profil d'expression de l'éventuel répresseur sans que soit modifiée la protéine elle-même : l'expression d'un tel répresseur dans les cellules folliculaires aurait pour nouvelle fonction la répression des *errantivirus* ou bien la quantité produite de ce répresseur serait modifiée. En effet, le système de répression apporté par la région COM restrictif semble dépendre de sa dose.

Ce modèle d'une région COM portant un répresseur peut-il rendre compte des données ? Si on considère que les allèles restrictifs correspondent à une production importante de ce répresseur dans toutes les cellules folliculaires des follicules ovariens, réprimant ainsi tous les *errantivirus*, chaque allèle permissif

pourrait correspondre à un moindre niveau d'expression dans certaines cellules. Par exemple, dans l'allèle COM<sup>Rev</sup>, la baisse d'expression du répresseur serait dans le germarium (où s'exprime *Idefix*), et dans les cellules postérieures (où s'exprime *ZAM*), mais elle serait maintenue dans les autres cellules folliculaires, notamment dans les cellules antérieures (où *gypsy* s'exprime). Ou encore, la quantité de répresseur serait diminuée partout, mais *gypsy* serait plus sensible que *ZAM* et *Idefix* aux faibles doses de répresseur.

Comment imaginer ce modèle du point de vue moléculaire ? La région 20A est hétérochromatique et contient des éléments transposables (Miklos et al, 1988 ; Miklos & Cotsell, 1990). On peut imaginer que les *errantivirus* ont un point chaud d'insertion à cette localisation. Chaque élément inséré pourrait apporter son propre LTR comme enhancer du répresseur. Ainsi, ce répresseur serait exprimé précisément dans les cellules où s'exprime habituellement l'élément. Chaque fois que la région portant le LTR serait modifiée (délétion ou autre), l'expression du répresseur dans les cellules correspondantes s'en trouverait diminuée. De telles modifications conduiraient à de nouveaux allèles permissifs, spécifiques des seuls éléments touchés, et à l'expression de leurs homologues présents dans le reste du génome.

Comment valider ce modèle ? Si ce répresseur existe, on doit pouvoir identifier les séquences cibles dans les trois *errantivirus gypsy*, *ZAM* et *Idefix*, responsables de la spécificité de souche, ainsi que la protéine reconnaissant ces séquences.

### **Modèle de *trans-silencing***

Le *trans-silencing* a été décrit pour l'élément *P*. Un élément *P* inséré dans un site subtélomérique peut réprimer les éléments *P* euchromatiques du génome par un effet de position. L'efficacité de cette répression dépend de la structure de l'élément, de son site d'insertion télomérique. Plusieurs éléments *P* télomériques peuvent avoir des effets additionnels (Ronsseray et al., 1996 ; Ronsseray et al., 1998).

Le site 20A pourrait avoir le même rôle vis à vis des *errantivirus* que les télomères pour l'élément *P*. On sait qu'il est le siège d'une accumulation d'éléments transposables. Des séquences suffisamment homologues à chaque *errantivirus* pourraient être responsables de son extinction transcriptionnelle. Si la position 20A semble être importante pour la répression, les éléments pourraient être dispersés sans ordre précis sur une grande région. Chaque allèle COM serait la combinaison aléatoire de plusieurs séquences errantivirales et serait propre à une

population. Dans ce modèle aussi, on comprend bien qu'une modification partielle d'un allèle COM restrictif puisse toucher un ou plusieurs *errantivirus*.

Comment distinguer ce modèle du modèle protéique ? Dans le modèle protéique, on attend que la séquence cible des *errantivirus* soit une séquence régulatrice comme le LTR ou l'UTR. Dans le modèle de trans-silencing, on attend que toute séquence homologue entre l'*errantivirus* cible et la séquence en 20A permette la répression, indépendamment des séquences régulatrices de l'*errantivirus*. Des séquences errantivirales pourraient être introduites en amont d'un gène rapporteur dans des constructions possédant un promoteur et un *enhancer* non errantiviraux, permettant une expression en cellules folliculaires. De telles séquences devraient permettre la répression du gène rapporteur dans les contextes restrictifs, et l'expression en contexte permissif (expériences en cours).

### Rôle de l'hétérochromatine

Dans les deux mécanismes envisagés, on peut attendre un rôle de la structure chromatinienne sur l'efficacité de la répression puisque la région 20A est hétérochromatique. J'ai voulu tester ce rôle avec des mutants suppresseurs ou *enhancers* de variéation fournis par Stephane Ronsseray. En contexte hétérozygote pour le chromosome X<sup>Rev</sup> et le transgène pZAM, il semble que les mutants suppresseurs de variéation *HPI* nul, *HPI* tronqué et le mutant *Orc2* permettent d'augmenter l'expression de ZAM dans les ovaires des femelles, par rapport à leurs sœurs sans mutation (non montré). Cette augmentation n'est pas observée dans toutes les expériences et son amplitude semble dépendre de paramètres que je n'ai pas identifiés, mais quand il y a une différence de coloration entre les mutantes et leurs sœurs, c'est toujours dans le même sens. De plus, j'observe une coloration intense des ovaires quand les mouches Rev-pZAM sont élevées à 25°C et une très faible coloration quand les mouches sont élevées à 18°C. Ces résultats indiquent un rôle répressif de l'hétérochromatine sur l'expression de ZAM.

Les lignées portant les mutants *enhancers* de variéation contiennent toutes le chromosome X avec l'inversion *In(1)w<sup>m4</sup>*. Je n'observe aucune coloration *LacZ* dans les ovaires de femelles *In(1)w<sup>m4</sup> / X<sup>Rev</sup>*, indépendamment de la présence des mutations. Les premières expériences que j'ai réalisées avec ces lignées ne me permettent donc pas de vérifier si les *enhancers* de variéation ont un effet sur la transcription de ZAM. Cette inversion *In(1)w<sup>m4</sup>* place la région 20A dans un contexte hétérochromatique différent ce qui peut expliquer un effet sur la transcription de ZAM.

J'ai établi une lignée homozygote  $X^{IR6}$  ; pZAM afin de tester si la répression par l'allèle  $COM^{IR6}$  est maintenue en contexte mutant suppresseur de variévation.

### Arguments en faveur d'un modèle ou de l'autre.

Du point de vue évolutif, le *trans-silencing* semble plus facile à mettre en place, car il n'interfère pas avec la biologie de l'hôte. Ce système répressif serait alors assez comparable avec les mécanismes d'autorégulation du *facteur I* ou de l'*élément P*, ne dépendant que des éléments eux-mêmes. On peut s'étonner que cette répression ne s'exerce que depuis le site 20A, et pas depuis un autre site hétérochromatique. Cette exclusivité de site est plus en faveur d'un gène codant un répresseur, mais elle n'est pas démontrée, puisqu'on ne connaît que deux exemples d'instabilité génétique avec mobilisation d'*errantivirus*. D'autre part, le mécanisme de *trans-silencing* n'est pas connu. Il existe peut-être un lien entre la position hétérochromatique et le tissu cible dans lequel s'exerce la répression. La position des *P* télomériques permet bien la répression de *P* dans les cellules germinales femelles, mais pas dans les tissus somatiques ni la lignée germinale mâle (Stephane Ronsseray, communication personnelle). La position des séquences en 20A dirige peut-être la répression dans les cellules folliculaires ovariennes.

Les trois *errantivirus gypsy*, *ZAM* et *Idefix* appartiennent à des familles différentes et présentent des divergences anciennes comme le montre l'arbre phylogénique construit par Christophe Terzian à partir des gènes *pol* et *env* (figure 33). Il serait étonnant qu'ils aient conservé une séquence de régulation commune les réprimant, alors que leurs séquences régulatrices LTR et UTR sont plus divergentes que leurs séquences codantes. Récemment, l'équipe d'Alain Bucheton a cloné des éléments *gypsy* d'autres espèces de drosophiles et testé l'expression de ces *gypsy* dans des constructions avec un gène *LacZ* par transgénèse chez *D. melanogaster* (Mejlumian *et al.*, 2001<sup>2</sup>). Curieusement, les constructions possédant les éléments *gypsy* de *D. virilis* et de *D. subobscura* ne subissent pas de répression en contexte *flam* restrictif de *D. melanogaster*. Les éléments *gypsy* sont pourtant plus proches entre eux au sein du groupe drosophiles que de *ZAM* et d'*Idefix* (figure 33). Dans le modèle de *trans-silencing*, la divergence existant entre ces *gypsy* peut expliquer la perte de répression par une perte d'homologie avec les séquences présentes au site 20A.

<sup>2</sup> Mejlumian L, Péliouon A, Bucheton A, Terzian C. « Phylogenetic and functional studies of *Drosophila* species invasion by the *gypsy* endogenous retrovirus.





## LISTE DES ABREVIATIONS

|              |                                         |
|--------------|-----------------------------------------|
| ADN          | Acide deoxyribonucléique                |
| ADNc         | ADN complémentaire                      |
| ARN          | Acide ribonucléique                     |
| ARNdb        | ARN double brin                         |
| ARNm         | ARN messenger                           |
| ARNr         | ARN ribosomal                           |
| ARNsb        | ARN simple brin                         |
| ARNt         | ARN de transfert                        |
| BLV          | Bovine Leukemia Virus                   |
| CA           | Capside                                 |
| CHA          | Charolles                               |
| CM           | centi Morgan                            |
| EN           | Endonucléase                            |
| ERV          | Endogenous RetroVirus                   |
| ET           | Elément transposable                    |
| KDa          | kilo Dalton                             |
| HERV         | Human Endogenous RetroVirus             |
| HIV          | Human Immunodeficiency Virus            |
| HTLV         | Human T-cell Leukemia Virus             |
| IAP          | Internal Associated Particles           |
| IN           | Intégrase                               |
| LINE         | Long INterspersed repetitive Element    |
| MA           | Matrice                                 |
| MLV          | Murine Leukemia Virus                   |
| MMTV         | Mouse Mammary Tumor Virus               |
| MoMuLV/MoMLV | Moloney Murine Leukemia Virus           |
| NC           | Nucleocapside                           |
| ORF          | Open Reading Frame                      |
| PBS          | Primer Binding Site                     |
| PCR          | Polymerase Chain Reaction               |
| PR           | Protease                                |
| PPT          | PolyPurine Tract                        |
| RH           | RNase H                                 |
| RNA-I        | RiboNucleic acid interference           |
| RNP          | RiboNucleo Protein                      |
| RT           | Rétrotranscriptase                      |
| SDS-PAGE     | SDS- PolyAcrylamide Gel Electrophoresis |
| SINE         | Short INterspersed repetitive Element   |
| SU           | Sous Unité                              |
| TM           | TransMembranaire                        |
| TPRT         | Target Primed Reverse Transcription     |
| UTR          | UnTranslated Region                     |
| VLP          | Virus Like Particle                     |



# **TABLE DES MATIERES**



## TABLE DES MATIERES

|                                                                                       |           |
|---------------------------------------------------------------------------------------|-----------|
| <b>Rapport bibliographique.....</b>                                                   | <b>4</b>  |
| Introduction.....                                                                     | 5         |
| <b>I- Classification et nomenclature.....</b>                                         | <b>7</b>  |
| A) Les différents éléments mobiles et leur classification.....                        | 7         |
| 1) La classe I .....                                                                  | 7         |
| 2) La classe II .....                                                                 | 8         |
| B) Les virus à ARN rétrotranscrit.....                                                | 9         |
| a) Rétrovirus et LTR-rétrotransposons.....                                            | 9         |
| b) Nomenclature des virus à ARN rétrotranscrit.....                                   | 11        |
| b1) Les <i>retroviridae</i> .....                                                     | 11        |
| - Les oncovirus .....                                                                 | 12        |
| - Les lentivirus.....                                                                 | 12        |
| - Les spumavirus ou foamy virus.....                                                  | 13        |
| b2) Les <i>metaviridae</i> .....                                                      | 13        |
| b3) Les <i>pseudoviridae</i> .....                                                    | 15        |
| c) La rétrotranscriptase au cœur de la nomenclature.....                              | 17        |
| d) Rétrovirus exogènes et endogènes.....                                              | 20        |
| <b>II- Structures et cycles d'éléments transposables.....</b>                         | <b>23</b> |
| A) Eléments de classe II sur un exemple : Structure et cycle de l' <i>élément P23</i> |           |
| B) Eléments de classe I.....                                                          | 24        |
| 1) Les rétroposons de type <i>LINE</i> et leur mécanisme de transposition.....        | 24        |
| 2) Les rétroposons <i>SINEs</i> .....                                                 | 26        |
| 3) Les virus à ARN rétrotranscrit.....                                                | 27        |
| a) Structure des génomes intégrés.....                                                | 27        |
| - Les régions non codantes.....                                                       | 29        |
| - Les régions codantes.....                                                           | 30        |
| b) Structure des particules de <i>retroviridae</i> enveloppées.....                   | 33        |
| c) Cycle des virus à ARN rétrotranscrit.....                                          | 34        |
| c1) Expression des provirus.....                                                      | 35        |
| c2) Assemblage des particules virales.....                                            | 37        |
| c3) Etape extracellulaire.....                                                        | 37        |
| c4) La réplication de l'ARN génomique.....                                            | 38        |
| c5) Intégration du provirus dans le génome de l'hôte.....                             | 39        |

III- Relations entre les éléments transposables et leurs hôtes..... 40

**III-I Les différents éléments transposables dans les génomes d'aujourd'hui..... 41**

        A) Les insertions fixées dans le génome..... 41

            1) Domestication de régions codantes..... 41

                - Les néogènes de l'élément *P*..... 42

                - Les protéines d'enveloppe des rétrovirus ont-elles un rôle physiologique chez leur hôte?..... 43

                - Les protéines provirales interagissent avec les rétrovirus exogènes..... 43

            2) Domestication de séquences régulatrices..... 44

                - Séquences contrôlant la transcription de gènes cellulaires voisins..... 44

                - Séquences transcrites avec le gène cible..... 45

            3) Maintien de la fonction ou du mécanisme de transposition..... 45

                - Système immunitaire des vertébrés à mâchoire..... 45

                - Télomères..... 46

                - Centromères..... 47

        B) Les insertions polymorphes du génome..... 47

            1) Les insertions polymorphes ont été d'abord identifiées pour leurs effets délétères..... 47

                - Chez l'homme..... 47

                - Chez la drosophile..... 48

            2) Polymorphisme des insertions et potentiel adaptatif de l'hôte..... 49

                a) Polymorphisme des provirus endogènes chez la souris..... 49

                b) Impact du polymorphisme d'insertions sur le *fitness* de l'hôte ..... 50

                    - Chez *Escherichia coli*..... 50

                    - Chez *Saccharomyces cerevisiae*..... 51

                    - Chez la drosophile..... 51

                c) Polymorphisme et réponse aux stress cellulaires..... 52

        C) Les éléments transposables sont plus ou moins répétés dans le génome..... 53

            1) Effets délétères des recombinaisons illégitimes..... 53

                - Implication des *SINEs*..... 53

                - Implication d'un élément de classe II..... 54

            2) Avantages des recombinaisons illégitimes..... 54

                - Duplications de gènes..... 54

                - Maintien de la taille des séquences péricentromériques..... 55

        D) Les éléments transposables actifs: quels sont les éléments encore mobiles dans les génomes d'aujourd'hui?..... 55

            1) Eléments actifs de drosophile..... 55

            2) Eléments actifs chez les mammifères..... 56

|                                                                                                                                                                      |           |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| <p>III-II Co-adaptation des éléments transposables avec leur hôte : limitation de leur impact délétère sur l'hôte ou optimisation de leur effet bénéfique ?.. 59</p> |           |
| A) Contrôles transcriptionnels.....                                                                                                                                  | 59        |
| 1) La méthylation.....                                                                                                                                               | 59        |
| 2) Répression d'éléments transposables et interférence ARN (RNA-i).....                                                                                              | 62        |
| 3) Modèles d'instabilités génétiques chez la drosophile.....                                                                                                         | 63        |
| a) L'élément <i>P</i> et le <i>trans-silencing</i> .....                                                                                                             | 63        |
| b) Le <i>facteur I</i> et la co-suppression.....                                                                                                                     | 65        |
| c) <i>Gypsy</i> et <i>flamenco</i> .....                                                                                                                             | 67        |
| B) Régulations post-transcriptionnelles.....                                                                                                                         | 69        |
| 1) Limitation du taux de transpositions.....                                                                                                                         | 69        |
| a) Détournement d'une fonction générale de l'hôte.....                                                                                                               | 69        |
| b) Résistances cellulaires à des rétrovirus.....                                                                                                                     | 71        |
| c) Spécificité tissulaire de l'activité des ET.....                                                                                                                  | 72        |
| d) Autorégulation du nombre de copies.....                                                                                                                           | 72        |
| 2) Limitation des conséquences des transpositions.....                                                                                                               | 73        |
| a) Insertions "fantômes".....                                                                                                                                        | 73        |
| b) Adressage du site d'insertion .....                                                                                                                               | 73        |
| - De la limitation de l'impact délétère.....                                                                                                                         | 73        |
| - A des ciblage de plus en plus spécifiques.....                                                                                                                     | 74        |
| - En passant par un rôle potentiel des ET.....                                                                                                                       | 75        |
| C) Les ET ont-ils un rôle fonctionnel?.....                                                                                                                          | 76        |
| <b>Le modèle.....</b>                                                                                                                                                | <b>80</b> |
| <b>I- La drosophile, un modèle génétique.....</b>                                                                                                                    | <b>81</b> |
| 1) Fiche BiOiB .....                                                                                                                                                 | 81        |
| 2) Actualisation.....                                                                                                                                                | 83        |
| <b>II- L'ovogenèse de <i>Drosophila melanogaster</i>.....</b>                                                                                                        | <b>83</b> |
| <b>Matériels et méthodes.....</b>                                                                                                                                    | <b>86</b> |
| <b>Drosophila strains.....</b>                                                                                                                                       | <b>87</b> |
| <b>PCR amplification of <i>ZAM</i> LTR.....</b>                                                                                                                      | <b>87</b> |
| <b>Plasmid construction and <i>P element</i> mediated germ-line transformations.....</b>                                                                             | <b>87</b> |
| <b>Histochemical analysis of <math>\beta</math>-galactosidase.....</b>                                                                                               | <b>88</b> |
| <b><i>In situ</i> Hybridization on polytene chromosomes.....</b>                                                                                                     | <b>88</b> |

|                                                                                                                                                                  |            |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|
| <b>Résultats.....</b>                                                                                                                                            | <b>89</b>  |
| <b>I- Caractérisation de ZAM et d'Idefix, deux nouveaux errantivirus mobilisés dans une souche <i>Drosophila melanogaster</i> W<sup>IR6REV</sup> (RevI).....</b> | <b>91</b>  |
| ARTICLE I.....                                                                                                                                                   | 94         |
| ARTICLE II.....                                                                                                                                                  | 106        |
| <b>II- Etude du cycle errantiviral de ZAM. (article III).....</b>                                                                                                | <b>119</b> |
| ARTICLE III.....                                                                                                                                                 | 120        |
| <b>III- Localisation des déterminants génétiques responsables de l'instabilité génétique dans la lignée W<sup>IR6REV</sup>.....</b>                              | <b>132</b> |
| 1) Le LTR est suffisant pour diriger l'expression de ZAM dans les ovaires de la lignée Rev... 132                                                                | 132        |
| 2) Le chromosome X provenant de la lignée Rev (X <sup>Rev</sup> ) permet la transcription de ZAM..... 133                                                        | 133        |
| 3) Effet dose du chromosome X <sup>Rev</sup> sur la transcription de ZAM..... 134                                                                                | 134        |
| 4) Localisation des déterminants génétiques contrôlant la transcription de ZAM en 20A2-20A3..... 136                                                             | 136        |
| 5) La mobilisation d'un deuxième errantivirus, Idefix, est contrôlée par le chromosome X <sup>Rev</sup> ..... 138                                                | 138        |
| 6) Un Centre Organisateur de Mobilisation (C.O.M.) localisé en 20A2-A3 contrôle trois errantivirus..... 140                                                      | 140        |
| <b>Discussion.....</b>                                                                                                                                           | <b>141</b> |
| <b>I ZAM et Idefix sont deux errantivirus.....</b>                                                                                                               | <b>142</b> |
| 1) Structure des deux insertions ZAM et Idefix au locus <i>white</i> ..... 142                                                                                   | 142        |
| 2) Cycle errantiviral de ZAM..... 144                                                                                                                            | 144        |
| <b>Des particules de ZAM sont détectées dans les cellules folliculaires de stades 9-10 ..... 144</b>                                                             | <b>144</b> |
| <b>Les particules de ZAM passent dans l'ovocyte par une étape extra-cellulaire..... 144</b>                                                                      | <b>144</b> |
| <b>Quel est le destin des particules dans l'ovocyte ainsi infecté ?..... 145</b>                                                                                 | <b>145</b> |
| <b>II Identification d'un locus de l'hôte contrôlant l'expression de ZAM et d'Idefix : Centre Organisateur de la Mobilisation (C.O.M.).....</b>                  | <b>146</b> |
| 1) Localisation de C.O.M..... 146                                                                                                                                | 146        |
| 2) Caractérisation de l'instabilité génétique COM <sup>Rev</sup> ..... 147                                                                                       | 147        |
| <b>Origine des errantivirus ZAM et Idefix et des déterminants génétiques responsables de l'instabilité dans la lignée Rev.....</b>                               | <b>147</b> |
| <b>L'instabilité COM<sup>Rev</sup> permet la mobilisation de deux errantivirus.....</b>                                                                          | <b>149</b> |
| <b>La région COM contrôle l'expression d'au moins trois errantivirus.....</b>                                                                                    | <b>150</b> |

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |  |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
| <p>           Tolérance du génome pour COM<sup>Rev</sup> : <i>ZAM</i> et <i>Idefix</i> sont-ils toujours mobilisés dans la lignée Rev ?.....151<br/>           COM<sup>rev</sup> est-il dominant pour la mobilisation de <i>ZAM</i> et d'<i>Idefix</i> ?.....152<br/>           3) Mécanisme du contrôle transcriptionnel.....152<br/>           Le LTR de <i>ZAM</i> possède l'information nécessaire pour diriger une transcription tissulaire spécifique et subir un contrôle spécifique de souche.....152<br/>           La transcription de <i>ZAM</i> et d'<i>Idefix</i> est réprimée dans les contextes stables, déréprimée dans la lignée Rev.....153<br/>           Par quel mécanisme <i>ZAM</i> et <i>Idefix</i> sont-ils réprimés dans la lignée W<sup>IR6</sup> ? .....153<br/>               Modèle protéique.....154<br/>               Modèle de <i>trans-silencing</i>.....155<br/>               Rôle de l'hétérochromatine.....156<br/>               Arguments en faveur d'un modèle ou de l'autre.....156<br/>           III- Conclusion.....158<br/>           Liste des abréviations.....159<br/>           Table des matières.....160<br/>           Références bibliographiques.....166         </p> |  |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|



## **REFERENCES BIBLIOGRAPHIQUES**

1- Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, George RA, Lewis SE, Richards S, Ashburner M, Henderson SN, Sutton GG, Wortman JR, Yandell MD, Zhang Q, Chen LX, Brandon RC, Rogers YH, Blazej RG, Champe M, Pfeiffer BD, Wan KH, Doyle C, Baxter EG, Helt G, Nelson CR, Gabor GL, Abril JF, Agbayani A, An HJ, Andrews-Pfannkoch C, Baldwin D, Ballew RM, Basu A, Baxendale J, Bayraktaroglu L, Beasley EM, Beeson KY, Benos PV, Berman BP, Bhandari D, Bolshakov S, Borkova D, Botchan MR, Bouck J, Brokstein P, Brottier P, Burtis KC, Busam DA, Butler H, Cadieu E, Center A, Chandra I, Cherry JM, Cawley S, Dahlke C, Davenport LB, Davies P, de Pablos B, Delcher A, Deng Z, Mays AD, Dew I, Dietz SM, Dodson K, Doup LE, Downes M, Dugan-Rocha S, Dunkov BC, Dunn P, Durbin KJ, Evangelista CC, Ferraz C, Ferreira S, Fleischmann W, Fosler C, Gabrielian AE, Garg NS, Gelbart WM, Glasser K, Glodek A, Gong F, Gorrell JH, Gu Z, Guan P, Harris M, Harris NL, Harvey D, Heiman TJ, Hernandez JR, Houck J, Hostin D, Houston KA, Howland TJ, Wei MH, Ibegwam C, Jalali M, Kalush F, Karpen GH, Ke Z, Kennison JA, Ketchum KA, Kimmel BE, Kodira CD, Kraft C, Kravitz S, Kulp D, Lai Z, Lasko P, Lei Y, Levitsky AA, Li J, Li Z, Liang Y, Lin X, Liu X, Mattei B, McIntosh TC, McLeod MP, McPherson D, Merkulov G, Milshina NV, Mobarry C, Morris J, Moshrefi A, Mount SM, Moy M, Murphy B, Murphy L, Muzny DM, Nelson DL, Nelson DR, Nelson KA, Nixon K, Nusskern DR, Pacleb JM, Palazzolo M, Pittman GS, Pan S, Pollard J, Puri V, Reese MG, Reinert K, Remington K, Saunders RD, Scheeler F, Shen H, Shue BC, Siden-Kiamos I, Simpson M, Skupski MP, Smith T, Spier E, Spradling AC, Stapleton M, Strong R, Sun E, Svirkas R, Tector C, Turner R, Venter E, Wang AH, Wang X, Wang ZY, Wassarman DA, Weinstock GM, Weissenbach J, Williams SM, Woodage T, Worley KC, Wu D, Yang S, Yao QA, Ye J, Yeh RF, Zaveri JS, Zhan M, Zhang G, Zhao Q, Zheng L, Zheng XH, Zhong FN, Zhong W, Zhou X, Zhu S, Zhu X, Smith HO, Gibbs RA, Myers EW, Rubin GM, Venter JC.

The genome sequence of *Drosophila melanogaster*.  
 Science. 2000 : 287(5461):2185-95.

2- Adler AJ, Scheller A, Hoffman Y, Robins DM.

Multiple components of a complex androgen-dependent enhancer.  
 Mol Endocrinol. : 5(11):1587-96.

3- Agrawal A, Eastman QM, Schatz DG.

Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system.  
 Nature. 1998 : 394(6695):744-51.

4- Ajioka JW, Eanes WF.

The accumulation of P-elements on the tip of the X chromosome in populations of *Drosophila melanogaster*.  
 Genet Res. 1989 : 53(1):1-6.

5- Aksoy S, Williams S, Chang S, Richards FF.

SLACS retrotransposon from *Trypanosoma brucei gambiense* is similar to mammalian LINEs.  
 Nucleic Acids Res. 1990 : 18(4):785-92.

6- Andrianov BV, Shuppe NG.

Tv1--a new family of *Drosophila virilis* retrotransposons.  
 Genetika. 1994 :30(4):437-44. Russian.

7- Anxolabéhère D, Nouaud D, Miller W

Éléments transposables et nouveautés génétiques chez les eucaryotes  
 Médecine sciences 2000 :16(11) I-IX

8- Arnaud P.

Repression de transcription du rétroposon S1 par la méthylation chez *Brassica napus*. Mécanismes et impacts. 159p

Thèse : spécialité Biologie cellulaire et moléculaire : Clermont-Ferrand. 2000

9- Babich V, Aksenov N, Alexeenko V, Oei SL, Buchlow G, Tomilin N.

Association of some potential hormone response elements in human genes with the Alu family repeats.  
 Gene. 1999 : 239(2):341-9.

10- Baldrich E, Dimitri P, Desset S, Leblanc P, Codipietro D, Vaury C.

Genomic distribution of the retrovirus-like element ZAM in *Drosophila*.  
 Genetica. 1997;100(1-3):131-40.

11- Baltimore, D.

RNA-dependent DNA polymerase in virions of RNA tumor viruses.  
 Nature 1970 : 226, 1209-1211.

12- Bayev AA Jr, Lyubomirskaya NV, Dzhumagaliev EB, Ananiev EV, Amiantova IG, Ilyin YV.

Structural organization of transposable element mdg4 from *Drosophila melanogaster* and a nucleotide sequence of its long terminal repeats.

Nucleic Acids Res. 1984 :12(8):3707-23.

- 13- **Bellen HJ, O'Kane CJ, Wilson C, Grossniklaus U, Pearson RK, Gehring WJ.**  
P-element-mediated enhancer detection: a versatile method to study development in *Drosophila*.  
*Genes Dev.* 1989 ;3(9):1288-300.
- 14- **Bender J.**  
Cytosine methylation of repeated sequences in eukaryotes: the role of DNA pairing.  
*Trends Biochem Sci.* 1998 ;23(7):252-6.
- 15- **Bennetzen JL.**  
Transposable element contributions to plant gene and genome evolution.  
*Plant Mol Biol.* 2000 ; 42(1):251-69.
- 16- **Bennetzen JL.**  
The Mutator transposable element system of maize.  
*Curr Top Microbiol Immunol.* 1996: 204:195-229.
- 17- **Berg D.E., Howe M.M.**  
Mobile DNA.  
1989 Washington, DC: American Society for microbiology. pp437-484
- 18- **Berkhout B, Jebbink M, Zsiros J.**  
Identification of an active reverse transcriptase enzyme encoded by a human endogenous HERV-K retrovirus.  
*J Virol.* 1999 ; 73(3):2365-75.
- 19- **Best S, Le Tissier P, Towers G, Stoye JP.**  
Positional cloning of the mouse retrovirus restriction gene Fv1.  
*Nature.* 1996 ; 382(6594):826-9.
- 20- **Biemont C, Lemeunier F, Garcia Guerreiro MP, Brookfield JF, Gautier C, Aulard S, Pasyukova EG.**  
Population dynamics of the copia, mdg1, mdg3, gypsy, and P transposable elements in a natural population of *Drosophila melanogaster*.  
*Genet Res.* 1994 ; 63(3):197-212.
- 21- **Biessmann H, Mason JM.**  
Genetics and molecular biology of telomeres.  
*Adv Genet.* 1992;30:185-249.
- 22- **Bingham PM, Chou TB, Mims I, Zachar Z.**  
On/off regulation of gene expression at the level of splicing.  
*Trends Genet.* 1988 ; 4(5):134-8
- 23- **Bingham PM, Kidwell MG, Rubin GM.**  
The molecular basis of P-M hybrid dysgenesis: the role of the P element, a P-strain-specific transposon family.  
*Cell.* 1982 ; 29(3):995-1004.
- 24- **Bingham PM, Zachar.**  
Retrotransposons and the FB transposon from *Drosophila melanogaster*.  
In *Mobile DNA 1989* (ed. D.E. Berg and M.M. Howe), pp. 485-502. American Society for Microbiology, Washington, D.C.
- 25- **Birchler JA, Bhadra MP, Bhadra U.**  
Making noise about silence: repression of repeated genes in animals.  
*Curr Opin Genet Dev.* 2000 ; 10(2):211-6.
- 26- **Bishop, J. M.**  
Cellular oncogenes and retroviruses.  
*Annu. Rev. Biochem.* 1983 ; 52, 301-354.
- 27- **Black DM, Jackson MS, Kidwell MG, Dover GA.**  
KP elements repress P-induced hybrid dysgenesis in *Drosophila melanogaster*.  
*EMBO J.* 1987 ; 6(13):4125-35.
- 28- **Blackman RK, Grimaila R, Koehler MM, Gelbart WM.**  
Mobilization of hobo elements residing within the decapentaplegic gene complex: suggestion of a new hybrid dysgenesis system in *Drosophila melanogaster*.  
*Cell.* 1987 ; 49(4):497-505.
- 29- **Blobel CP, Wolfsberg TG, Turck CW, Myles DG, Primakoff P, White JM.**  
A potential fusion peptide and an integrin ligand domain in a protein active in sperm-egg fusion.  
*Nature* 1992 ; 356:248-252
- 30- **Blond JL, Beseme F, Duret L, Bouton O, Bedin F, Perron H, Mandrand B, Mallet F.**  
Molecular characterization and placental expression of HERV-W, a new human endogenous retrovirus family.  
*J Virol.* 1999 ;73(2):1175-85.

- 31- **Blond JL, Lavillette D, Cheynet V, Bouton O, Oriol G, Chapel-Fernandes S, Mandrand B, Mallet F, Cosset FL.**  
An envelope glycoprotein of the human endogenous retrovirus HERV-W is expressed in the human placenta and fuses cells expressing the type D mammalian retrovirus receptor.  
J Virol. 2000 : 74(7):3321-9.
- 32- **Boeke JD.**  
LINEs and Alus--the polyA connection.  
Nat Genet. 1997 : 16(1):6-7.
- 33- **Boeke JD.**  
DNA repair. A little help for my ends.  
Nature. 1996 383(6601):579, 581.
- 34- **Boeke JD, Devine SE.**  
Yeast retrotransposons: finding a nice quiet neighborhood.  
Cell. 1998 : 93(7):1087-9.
- 35- **Boeke JD, Eickbush T, Sandmeyer SB, Voytas DF.**  
Metaviridae.  
*In* F. A. Murphy, 1998 a: ed Virus taxonomy: ICTV VIIIth report. Springer-Verlag, New York.
- 36- **Boeke JD, Eickbush T, Sandmeyer SB, Voytas DF.**  
Pseudoviridae.  
*In* F. A. Murphy, 1998 b : ed Virus taxonomy: ICTV VIIIth report. Springer-Verlag, New York.
- 37- **Boeke JD, Stoye JP.**  
Retrotransposons, endogenous retroviruses, and the evolution of retroelements.  
*In* Retroviruses.  
1997 Edited by J. M. Coffin, S. H. Hughes and H. E. Varmus. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- 38- **Bowen NJ, McDonald JF.**  
Genomic analysis of *Caenorhabditis elegans* reveals ancient families of retroviral-like elements.  
Genome Res. 1999 : 9(10):924-35.
- 39- **Boyd MT, Bax CM, Bax BE, Bloxam DL, Weiss RA.**  
The human endogenous retrovirus ERV-3 is upregulated in differentiating placental trophoblast cells.  
Virology. 1993 : 196(2):905-9.
- 40- **Bregliano JC, Kidwell MG.**  
Hybrid dysgenesis determinants.  
*In* "Mobile DNA" 1995 ed Berg DE and Howe. American Society for microbiology, Washington, DC.
- 41- **Bucheton A, Lavigne JM, Picard G, L'Heritier P.**  
Non-mendelian female sterility in *Drosophila melanogaster*: quantitative variations in the efficiency of inducer and reactive strains.  
Heredity. 1976 : 36(3):305-14.
- 42- **Bucheton A, Paro R, Sang HM, Pelisson A, Finnegan DJ.**  
The molecular basis of I-R hybrid dysgenesis in *Drosophila melanogaster*: identification, cloning, and properties of the I factor.  
Cell. 1984 : 38(1):153-63.
- 43- **Buller RS, Ahmed A, Portis JL.**  
Identification of two forms of an endogenous murine retroviral env gene linked to the Rmcf locus.  
J Virol. 1987 : 61(1):29-34.
- 44- **Bureau TE, Ronald PC, Wessler SR.**  
A computer-based systematic survey reveals the predominance of small inverted-repeat elements in wild-type rice genes.  
Proc Natl Acad Sci U S A. 1996 : 93(16):8524-9.
- 45- **Burke WD, Malik HS, Lathe WC 3rd, Eickbush TH.**  
Are retrotransposons long-term hitchhikers?  
Nature. 1998 : ;392(6672):141-2.
- 46- **Burke WD, Muller F, Eickbush TH.**  
R4, a non-LTR retrotransposon specific to the large subunit rRNA genes of nematodes.  
Nucleic Acids Res. 1995 : 23(22):4628-34.
- 47- **Busseau I, Malinsky S, Balakireva M, Chaboissier MC, Teninges D, Bucheton A.**  
A genetically marked I element in *Drosophila melanogaster* can be mobilized when ORF2 is provided in trans.  
Genetics. 1998 Jan;148(1):267-75.

- 48- **Busson D, Gans M, Komitopoulou K, Masson M.**  
Genetic analysis of three dominant female-sterile mutations located on the X chromosome in *Drosophila melanogaster*.  
*Genetics* 1983 : 105:309-325
- 49- **Calvi BR, Hong TJ, Findley SD, Gelbart WM.**  
Evidence for a common evolutionary origin of inverted repeat transposons in *Drosophila* and plants: hobo, Activator, and Tam3.  
*Cell*. 1991 : 66(3):465-71.
- 50- **Capy P, Langin T, Higuete D, Maurer P, Bazin C.**  
Do the integrases of LTR-retrotransposons and class II element transposases have a common ancestor?  
*Genetica*. 1997;100(1-3):63-72.
- 51- **Cavarec L, Jensen S, Casella JF, Cristescu SA, Heidmann T.**  
Molecular cloning and characterization of a transcription factor for the copia retrotransposon with homology to the BTB-containing lola neurogenic factor.  
*Mol Cell Biol*. 1997 : 17(1):482-94.
- 52- **Chaboissier MC, Bucheton A, Finnegan DJ.**  
Copy number control of a transposable element, the I factor, a LINE-like element in *Drosophila*.  
*Proc Natl Acad Sci U S A*. 1998 : 95(20):11781-5.
- 53- **Chaboissier MC, Busseau I, Prosser J, Finnegan DJ, Bucheton A.**  
Identification of a potential RNA intermediate for transposition of the LINE-like element I factor in *Drosophila melanogaster*.  
*EMBO J*. 1990 9(11):3557-63.
- 54- **Chalker DL, Sandmeyer SB.**  
Ty3 integrates within the region of RNA polymerase III transcription initiation.  
*Genes Dev*. 1992 : 6(1):117-28.
- 55- **Chalvet F, Teysset L, Terzian C, Prud'homme N, Santamaria P, Bucheton A, Pelisson A.**  
Proviral amplification of the Gypsy endogenous retrovirus of *Drosophila melanogaster* involves env-independent invasion of the female germline.  
*EMBO J*. 1999 : 18(9):2659-69.
- 56- **Chance PF, Abbas N, Lensch MW, Pentao L, Roa BB, Patel PI, Lupski JR.**  
Two autosomal dominant neuropathies result from reciprocal DNA duplication/deletion of a region on chromosome 17.  
*Hum Mol Genet*. 1994 : 3(2):223-8.
- 57- **Charlesworth B, Sniegowski P, Stephan W.**  
The evolutionary dynamics of repetitive DNA in eukaryotes.  
*Nature*. 1994 : 371(6494):215-20.
- 58- **Chen RZ, Pettersson U, Beard C, Jackson-Grusby L, Jaenisch R.**  
DNA hypomethylation leads to elevated mutation rates.  
*Nature*. 1998 : 395(6697):89-93.
- 59- **Chesnokov IN, Schmid CW.**  
Specific Alu binding protein from human sperm chromatin prevents DNA methylation.  
*J Biol Chem*. 1995 : 270(31):18539-42.
- 60- **Chu WM, Ballard R, Carpick BW, Williams BR, Schmid CW.**  
Potential Alu function: regulation of the activity of double-stranded RNA-activated kinase PKR.  
*Mol Cell Biol*. 1998 : 18(1):58-68.
- 61- **Clark JB, Maddison WP, Kidwell MG.**  
Phylogenetic analysis supports horizontal transfer of P transposable elements.  
*Mol Biol Evol*. 1994 : 11(1):40-50.
- 62- **Coffin, J. M., S. H. Hughes, H. E. Varmus, P. K. Vogt, V. M. Vogt, E. Hunter, A. Telesnitsky, S. P. Goff, P. O Brown, A. B. Rabson, B. J. Graves, R. Swanstrom, J. W. Wills, J. D. Boeke, J. P. Stoye.**  
*Retroviruses*.  
1997 Edited by J. M. Coffin, S. H. Hughes and H. E. Varmus. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- 63- **Cogoni C, Macino G.**  
Gene silencing in *Neurospora crassa* requires a protein homologous to RNA-dependent RNA polymerase.  
*Nature*. 1999 : 399(6732):166-9.
- 64- **Colot V, Rossignol JL.**  
Eukaryotic DNA methylation as an evolutionary device.  
*Bioessays*. 1999 : 21(5):402-11.

- 65- **Conte C, Calco V, Desset S, Leblanc P, Dastugue B, Vaury C.**  
Impact of multiple insertions of two retroelements, ZAM and Idefix at an euchromatic locus.  
*Genetica*. 2000;109(1-2):53-9.
- 66- **Conte C, Dastugue B, Vaury C.**  
Coupling of enhancer and insulator properties identified in two retrotransposons modulate their mutagenic impact on nearby genes.  
*Mol Cell Biol*. 2002 : in press
- 67- **Conte D Jr, Barber E, Banerjee M, Garfinkel DJ, Curcio MJ.**  
Posttranslational regulation of Ty1 retrotransposition by mitogen-activated protein kinase Fus3.  
*Mol Cell Biol*. 1998 18(5):2502-13.
- 68- **Cooper JA.**  
Oncogenes and anti-oncogenes.  
*Curr Opin Cell Biol*. 1990 : 2(2):285-95. Review.
- 69- **Curcio MJ, Garfinkel DJ.**  
Heterogeneous functional Ty1 elements are abundant in the *Saccharomyces cerevisiae* genome.  
*Genetics*. 1994 : 136(4):1245-59.
- 70- **Dalmay T, Hamilton A, Rudd S, Angell S, Baulcombe DC.**  
An RNA-dependent RNA polymerase gene in *Arabidopsis* is required for posttranscriptional gene silencing mediated by a transgene but not by a virus.  
*Cell*. 2000 : 101(5):543-53.
- 71- **Daniels SB, Peterson KR, Strausbaugh LD, Kidwell MG, Chovnick A.**  
Evidence for horizontal transmission of the P transposable element between *Drosophila* species.  
*Genetics*. 1990 : 124(2):339-55.
- 72- **Deininger PL.**  
SINES: short interspersed repeated DNA elements in higher eukaryotes.  
In "Mobile DNA."  
1989 Washington, DC: American Society for microbiology. pp 619-636
- 73- **Deininger PL, Batzer MA.**  
Alu repeats and human disease.  
*Mol Genet Metab*. 1999 : 67(3):183-93.
- 74- **Deininger PL, Batzer MA.**  
SINE master genes and population biology.  
In "The impact of short interspersed element (SINES) on the host genome. 1995 : (Maraia RJ ed). pp 43-60, RG Landes company, Springer, Austin, Texas, USA.
- 75- **Dej KJ, Gerasimova T, Corces VG, Boeke JD.**  
A hotspot for the *Drosophila* gypsy retroelement in the ovo locus.  
*Nucleic Acids Res*. 1998 : 26(17):4019-25.
- 76- **de La Roche Saint Andre C, Bregliano JC.**  
Evidence for a multistep control in transposition of I factor in *Drosophila melanogaster*.  
*Genetics*. 1998 : 148(4):1875-84.
- 77- **Deragon JM, Capy P.**  
Impact of transposable elements on the human genome.  
*Ann Med*. 2000 : 32(4):264-73.
- 78- **Desset S, Conte C, Dimitri P, Calco V, Dastugue B, Vaury C.**  
Mobilization of two retroelements, ZAM and Idefix, in a novel unstable line of *Drosophila melanogaster*.  
*Mol Biol Evol*. 1999 : 16(1):54-66.
- 79- **Dhellin O, Maestre J, Heidmann T.**  
Functional differences between the human LINE retrotransposon and retroviral reverse transcriptases for in vivo mRNA reverse transcription.  
*EMBO J*. 1997 : 16(21):6590-602.
- 80- **Dimitri P, Arca B, Berghella L, Mei E.**  
High genetic instability of heterochromatin after transposition of the LINE-like I factor in *Drosophila melanogaster*.  
*Proc Natl Acad Sci U S A*. 1997 ; 94(15):8052-7.
- 81- **Dimitri P, Junakovic N.**  
Revising the selfish DNA hypothesis: new evidence on accumulation of transposable elements in heterochromatin.  
*Trends Genet*. 1999 : 15(4):123-4.
- 82- **Doak TG, Doerder FP, Jahn CL, Herrick G.**  
A proposed superfamily of transposase genes: transposon-like elements inciliated protozoa and a common "D35E" motif.  
*Proc Natl Acad Sci U S A*. 1994 : 91(3):942-6.

**83- Dominguez A, Albornoz J.**

Rates of movement of transposable elements in *Drosophila melanogaster*.

Mol Gen Genet. 1996 : 251(2):130-8.

**84- Dong J, Roth MG, Hunter E.**

A chimeric avian retrovirus containing the influenza virus hemagglutinin gene has an expanded host range.

J Virol. 1992 : 66(12):7374-82.

**85- Dulai KS, von Dornum M, Mollon JD, Hunt DM.**

The evolution of trichromatic color vision by opsin gene duplication in New World and Old World primates.

Genome Res. 1999 : 9(7):629-38.

**86- Dunn KJ, Yuan CC, Blair DG.**

A phenotypic host range alteration determines RD114 virus restriction in feline embryonic cells.

J Virol. 1993 : 67(8):4704-11.

**87- Dupressoir A, Heidmann T.**

Germ line-specific expression of intracisternal A-particle retrotransposons in transgenic mice.

Mol Cell Biol. 1996 : 16(8):4495-503.

**88- Eden S, Cedar H.**

Role of DNA methylation in the regulation of transcription.

Curr Opin Genet Dev. 1994 : 4(2):255-9.

**89- Eiden MV, Farrell K, Warsowe J, Mahan LC, Wilson CA.**

Characterization of a naturally occurring ecotropic receptor that does not facilitate entry of all ecotropic murine retroviruses.

J Virol. 1993 : 67(7):4056-61.

**90- Ellerman, V. and O. Bbang.**

Experimentelle Leukämie bei Hühnern.

Zentralbl. Bakteriologie. Parasitenkd. Infektionskr. 1908 Hyg. Abt. Orig. 46, 595-609.

**91- Emmons SW, Yesner L, Ruan KS, Katzenberg D.**

Evidence for a transposon in *Caenorhabditis elegans*.

Cell. 1983 : 32(1):55-65.

**92- Engels WR.**

P elements in *Drosophila*.

In mobile DNA. 1989 : ed Berg DE and Howe MM. Washington, DC: American Society for microbiology. pp 437-484

**93- Engels WR.**

A trans-acting product needed for P factor transposition in *Drosophila*.

Science. 1984 : 226(4679):1194-6.

**94- Engels WR.**

The P family of transposable elements in *Drosophila*.

Annu Rev Genet. 1983;17:315-44.

**95- Engels WR.**

Extrachromosomal control of mutability in *Drosophila melanogaster*.

Proc Natl Acad Sci U S A. 1979 a: 76(8):4011-5.

**96- Engels WR, Johnson-Schlitz DM, Eggleston WB, Sved J.**

High-frequency P element loss in *Drosophila* is homolog dependent.

Cell. 1990 : 62(3):515-25.

**97- Esnault C, Maestre J, Heidmann T.**

Human LINE retrotransposons generate processed pseudogenes.

Nat Genet. 2000 : 24(4):363-7.

**98- Farnet CM, Bushman FD.**

HIV-1 cDNA integration: requirement of HMG I(Y) protein for function of preintegration complexes in vitro.

Cell. 1997 : ;88(4):483-92.

**99- Fawcett DH, Lister CK, Kellett E, Finnegan DJ.**

Transposable elements controlling I-R hybrid dysgenesis in *D. melanogaster* are similar to mammalian LINES.

Cell. 1986 : 47(6):1007-15.

**100- Feng Q, Moran JV, Kazazian HH Jr, Boeke JD.**

Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition.

Cell. 1996 : 87(5):905-16.

**101- Feng Q, Schumann G, Boeke JD.**

Retrotransposon R1Bm endonuclease cleaves the target sequence.

Proc Natl Acad Sci U S A. 1998 : 95(5):2083-8.

**102- Finnegan DJ.**

Transposable elements.  
Curr Opin Genet Dev. 1992 : 2(6):861-7.

**103- Finnegan DJ.**

Eukaryotic transposable elements and genome evolution.  
Trends Genet. 1989 : 5(4):103-7

**104- Finnegan EJ, Genger RK, Kovac K, Peacock WJ, Dennis ES.**

DNA methylation and the promotion of flowering by vernalization.  
Proc Natl Acad Sci U S A. 1998 : 95(10):5824-9.

**105- Flavell A.**

Did retroviruses evolve from transposable elements?  
Nature. 1981 289(5793):10-1.

**106- Flavell AJ, Smith DB.**

A Ty1-copia group retrotransposon sequence in a vertebrate.  
Mol Gen Genet. 1992 : 233(1-2):322-6.

**107- Frankel WN, Stoye JP, Taylor BA, Coffin JM.**

A linkage map of endogenous murine leukemia proviruses.  
Genetics. 1990 : 124(2):221-36.

**108- Fridell RA, Pret AM, Searles LL.**

A retrotransposon 412 insertion within an exon of the *Drosophila melanogaster* vermilion gene is spliced from the precursor RNA.

Genes Dev. 1990 : 4(4):559-66.

**109- Friesen PD, Nissen MS.**

Gene organization and transcription of TED, a lepidopteran retrotransposon integrated within the baculovirus genome.

Mol Cell Biol. 1990 Jun;10(6):3067-77.

**110- Garrett JE, Knutson DS, Carroll D.**

Composite transposable elements in the *Xenopus laevis* genome.  
Mol Cell Biol. 1989 : 9(7):3018-27.

**111- Gatti M, Bonaccorsi S, Pimpinelli S.**

Looking at *Drosophila* mitotic chromosomes.  
Methods Cell Biol. 1994;44:371-91.

**112- Geyer PK, Corces VG.**

DNA position-specific repression of transcription by a *Drosophila* zinc finger protein.

Genes Dev. 1992 : 6(10):1865-73.

**113- Gilboa E, Goff S, Shields A, Yoshimura F, Mitra S, Baltimore D.**

In vitro synthesis of a 9 kbp terminally redundant DNA carrying the infectivity of Moloney murine leukemia virus.

Cell. 1979 16(4):863-74.

**114- Goto T, Nakai M, Ikuta K.**

The life-cycle of human immunodeficiency virus type 1.

Micron. 1998 Apr-Jun;29(2-3):123-38.

**115- Grammont M.**

Toucan: gene nécessaire pour la formation et le développement de la chambre d'oeuf chez *Drosophila melanogaster*. 100p

Thèse. Spécialité Biologie cellulaire et moléculaire. Clermont-Ferrand 1997

**116- Grandbastien, M. A.**

Activation of plant retrotransposons under stress conditions.

*Trends Plant Sci.* 1998 : 3, 181-187.

**117- Grishok A, Tabara H, Mello CC.**

Genetic requirements for inheritance of RNAi in *C. elegans*.

Science. 2000 : 287(5462):2494-7.

**118- Hammond SM, Bernstein E, Beach D, Hannon GJ.**

An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells.

Nature. 2000 : 404(6775):293-6.

**119- Hartl DL, Lohe AR, Lozovskaya ER.**

Modern thoughts on an ancient mariner: function, evolution, regulation.

Annu Rev Genet. 1997: 31:337-58.

- Developmental differences in methylation of human Alu repeats.  
Mol Cell Biol. 1993 : 13(8):4523-30.
- 120- Hellmann-Blumberg U, Hintz MF, Gatewood JM, Schmid CW.**
- Detection of Caenorhabditis transposon homologs in diverse organisms.  
New Biol. 1992 : 4(4):382-8.
- 121- Henikoff S.**
- Gene families: the taxonomy of protein paralogs and chimeras.  
Science. 1997 : 278(5338):609-14.
- 122- Henikoff S, Greene EA, Pietrokovski S, Bork P, Attwood TK, Hood L.**
- Zepp, a LINE-like retrotransposon accumulated in the Chlorella telomeric region.  
EMBO J. 1997 : 16(12):3715-23.
- 123- Higashiyama T, Noutoshi Y, Fujie M, Yamada T.**
- Sequence-specific single-strand RNA binding protein encoded by the human LINE-1 retrotransposon.  
EMBO J. 1997 : 16(19):6034-43.
- 124- Hohjoh H, Singer MF.**
- Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA.  
EMBO J. 1996 : 15(3):630-9.
- 125- Holmes SE, Dombroski BA, Krebs CM, Boehm CD, Kazazian HH Jr.**
- A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q produces a chimaeric insertion.  
Nat Genet. 1994 : 7(2):143-8.
- 126- Holmes SE, Singer MF, Swergold GD.**
- Studies on p40, the leucine zipper motif-containing protein encoded by the first open reading frame of an active human LINE-1 transposable element.  
J Biol Chem. 1992 : 267(28):19765-8.
- 127- Holmes SE, Singer MF, Swergold GD.**
- Simian foamy virus-induced immunosuppression in rabbits.  
J Gen Virol. 1979 : 44(2):383-90.
- 128- Hooks JJ, Detrick-Hooks B.**
- Complete nucleotide sequence and genome organization of a Drosophila transposable genetic element, 297.  
Eur J Biochem. 1986 : 154(2):417-25.
- 129- Inouye S, Yuki S, Saigo K.**
- Sequence-specific insertion of the Drosophila transposable genetic element 17.6.  
Nature. 1984 : ;310(5975):332-3.
- 130- Inouye S, Yuki S, Saigo K.**
- Nature 2001 : 409:824-958
- 131- International Human Genome Sequencing Consortium**
- Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells.  
Cell. 1997 : 91(4):501-10.
- 132- Ivics Z, Hackett PB, Plasterk RH, Izsvak Z.**
- Molecular structure of a somatically unstable transposable element in Drosophila.  
Proc Natl Acad Sci U S A. 1986 : 83(22):8684-8.
- 133- Jacobson JW, Medhora MM, Hartl DL.**
- A gene deletion ending at the midpoint of a repetitive DNA sequence in one form of hereditary persistence of fetal haemoglobin.  
Nature. 1982 : 296(5856):469-70
- 134- Jagadeeswaran P, Tuan D, Forget BG, Weissman SM.**
- Taming of transposable elements by homology-dependent gene silencing.  
Nat Genet. 1999 a : 21(2):209-12.
- 135- Jensen S, Gassama MP, Heidmann T.**
- Cosuppression of I transposon activity in Drosophila by I-containing sense and antisense transgenes.  
Genetics. 1999 b : 153(4):1767-74.
- 136- Jensen S, Gassama MP, Heidmann T.**
- Hotspots for unselected Ty1 transposition events on yeast chromosome III are near tRNA genes and LTR sequences.  
Cell. 1993 : 73(5):1007-18.
- 137- Ji H, Moore DP, Blomberg MA, Braiterman LT, Voytas DF, Natsoulis G, Boeke JD.**

**138- Jurka J.**

Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons.  
Proc Natl Acad Sci U S A. 1997 : 94(5):1872-7.

**139- Kachroo P, Leong SA, Chattoo BB.**

Pot2, an inverted repeat transposon from the rice blast fungus *Magnaporthe grisea*.  
Mol Gen Genet. 1994 : 245(3):339-48.

**140- Kalendar R, Tanskanen J, Immonen S, Nevo E, Schulman AH.**

Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence.  
Proc Natl Acad Sci U S A. 2000 : 97(12):6603-7.

**141- Karess RE, Rubin GM.**

Analysis of P transposable element functions in *Drosophila*.  
Cell. 1984 : 38(1):135-46.

**142- Karpen GH, Spradling AC.**

Analysis of subtelomeric heterochromatin in the *Drosophila* minichromosome Dp1187 by single P element insertional mutagenesis.  
Genetics. 1992 : 132(3):737-53.

**143- Kazazian HH Jr.**

Mobile elements and disease.  
Curr Opin Genet Dev. 1998 : 8(3):343-50.

**144- Kazazian HH Jr, Moran JV.**

The impact of L1 retrotransposons on the human genome.  
Nat Genet. 1998 : 19(1):19-24.

**145- Kazazian HH Jr, Wong C, Youssoufian H, Scott AF, Phillips DG, Antonarakis SE.**

Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man.  
Nature. 1988 : 332(6160):164-6.

**146- Kelley MR, Kidd S, Berg RL, Young MW**

Restriction of P-element insertions at the notch locus of *Drosophila melanogaster*.  
Mol Cell Biol. 1987 : 7(4):1545-1548

**147- Kennerdell JR, Carthew RW.**

Use of dsRNA-mediated genetic interference to demonstrate that frizzled and frizzled 2 act in the wingless pathway.  
Cell. 1998 : 95(7):1017-26.

**148- Ketting RF, Haverkamp TH, van Luenen HG, Plasterk RH.**

Mut-7 of *C. elegans*, required for transposon silencing and RNA interference, is a homolog of Werner syndrome helicase and RNaseD.  
Cell. 1999 : 99(2):133-41.

**149- Kidwell MG.**

The Wilhelmine E. Key 1991 Invitational Lecture. The evolutionary history of the P family of transposable elements.  
J Hered. 1994 : 85(5):339-46.

**150- Kidwell MG.**

Hybrid dysgenesis in *Drosophila melanogaster*: nature and inheritance of P element regulation.  
Genetics. 1985 : 111(2):337-50.

**151- Kidwell MG.**

Evolution of hybrid dysgenesis determinants in *Drosophila melanogaster*.  
Proc Natl Acad Sci U S A. 1983 : 80(6):1655-9. No abstract available.

**152- Kidwell MG, Kidwell JF, Sved JA**

Hybrid dysgenesis in *Drosophila melanogaster*: a syndrome of aberrant traits including mutation, sterility, and male recombination.  
Genetics 1997 : 86:813-833

**153- Kidwell MG, Lisch DR.**

Transposable elements and host genome evolution.  
Trends Ecol Evol. 2000 : 15(3):95-99.

**154- Kim AI, Belyaeva ES, Aslanian MM.**

Autonomous transposition of gypsy mobile elements and genetic instability in *Drosophila melanogaster*.  
Mol Gen Genet. 1990 : 224(2):303-8.

- 115- **Kim A, Terzian C, Santamaria P, Pelisson A, Purd'homme N, Bucheton A.**  
Retroviruses in invertebrates: the gypsy retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*.  
Proc Natl Acad Sci U S A. 1994 : 91(4):1285-9.
- 116- **Kimura RH, Choudary PV, Schmid CW.**  
Silk worm Bm1 SINE RNA increases following cellular insults.  
Nucleic Acids Res. 1999 : 27(16):3380-7.
- 117- **King RC.**  
Ovarian development in *Drosophila melanogaster*.  
1970 New York, Academic Press
- 118- **Kinsey PT, Sandmeyer SB.**  
Ty3 transposes in mating populations of yeast: a novel transposition assay for Ty3.  
Genetics. 1995 : 139(1):81-94.
- 119- **Kipling D, Warburton PE.**  
Centromeres, CENP-B and Tigger too.  
Trends Genet. 1997 : 13(4):141-5.
- 120- **Kirchner J, Connolly CM, Sandmeyer SB.**  
Requirement of RNA polymerase III transcription factors for in vitro position-specific integration of a retroviruslike element.  
Science. 1995 : 267(5203):1488-91.
- 121- **Klemenz R, Weber U, Gehring WJ.**  
The white gene as a marker in a new P-element vector for gene transfer in *Drosophila*.  
Nucleic Acids Res. 1987 : 15(10):3947-59.
- 122- **Kobayashi S, Kitamura T, Sasaki H, Okada M**  
Two types of pole cells are present in the *Drosophila* embryo, one with and one without splicing activity for the third P-element intron.  
Development 1993 117(3):885-893
- 123- **Kochanek S, Renz D, Doerfler W.**  
DNA methylation in the Alu sequences of diploid and haploid primary human cells.  
EMBO J. 1993 : 12(3):1141-51.
- 124- **Kolosha VO, Martin SL.**  
In vitro properties of the first ORF protein from mouse LINE-1 support its role in ribonucleoprotein particle formation during retrotransposition.  
Proc Natl Acad Sci U S A. 1997 : 94(19):10155-60.
- 125- **Kozak CA.**  
Susceptibility of wild mouse cells to exogenous infection with xenotropic leukemia viruses: control by a single dominant locus on chromosome 1.  
J Virol. 1985 : 55(3):690-5.
- 126- **Krane DE, Hardison RC.**  
Short interspersed repeats in rabbit DNA can provide functional polyadenylation signals.  
Mol Biol Evol. 1990 : 7(1):1-8.
- 127- **Kuhn DT, Woods DF, Andrew DJ.**  
Deletion analysis of the tumorous-head (tuh-3) gene in *Drosophila melanogaster*.  
Genetics. 1981 : 99(1):99-107.
- 128- **Kulski JK, Gaudieri S, Bellgard M, Balmer L, Giles K, Inoko H, Dawkins RL.**  
The evolution of MHC diversity by segmental duplication and transposition of retroelements.  
J Mol Evol. 1997 Dec;45(6):599-609.
- 129- **Labuda, Zietkiewicz , Mitchell**  
Alu elements as a source of genomic variation : Deleterious effects and evolutionary novelties.  
In "The impact of short interspersed elements (SINEs) on the host genome 1995 (Maraia ed). pp 1-22 RG Landes Compagny, Springer, Austin, Texas USA.
- 130- **Lachaume P, Pinon H.**  
Germ-line expression of the I factor, a functional LINE from the fruit fly *Drosophila melanogaster*, is positively regulated by reactivity, a peculiar cellular state.  
Mol Gen Genet. 1993 : 240(2):277-85.
- 131- **Lajoinie O, Drake ME, Dastugue B, Vaury C.**  
Aberrant pre-mRNA maturation is caused by LINE insertions into introns of the white gene of *Drosophila melanogaster*.  
Nucleic Acids Res. 1995 : 23(20):4015-22.
- 132- **Lammel U, Klambt C.**  
Specific expression of the *Drosophila* midline-jumper retro-transposon in embryonic CNS midline cells.  
Mech Dev. 2001 : 100(2):339-42.

- 173- **Langin T, Capy P, Daboussi MJ.**  
The transposable element impala, a fungal member of the Tc1-mariner superfamily.  
Mol Gen Genet. 1995 : 246(1):19-28.
- 174- **Larsson E, Andersson AC, Nilsson BO.**  
Expression of an endogenous retrovirus (ERV3 HERV-R) in human reproductive and embryonic tissues--  
evidence for a function for envelope gene products.  
Ups J Med Sci. 1994 : 99(2):113-20.
- 175- **Laski FA, Rio DC, Rubin GM.**  
Tissue specificity of Drosophila P element transposition is regulated at the level of mRNA splicing.  
Cell. 1986 : 44(1):7-19.
- 176- **Laten HM, Majumdar A, Gaucher EA.**  
SIRE-1, a copia/Ty1-like retroelement from soybean, encodes a retroviral envelope-like protein.  
Proc Natl Acad Sci U S A. 1998 : 95(12):6897-902.
- 177- **Laurent AM, Puechberty J, Prades C, Gimenez S, Roizes G.**  
Site-specific retrotransposition of L1 elements within human alphoid satellite  
sequences.  
Genomics. 1997 : 46(1):127-32.
- 178- **Lavigne JM**  
I-R system of hybrid dysgenesis in Drosophila melanogaster : further data on the arrest of development  
of the embryos from SF females.  
Biol Cell 1986 : 56:207-216
- 179- **Leblanc P, Dastugue B, Vaury C.**  
The integration machinery of ZAM, a retroelement from Drosophila melanogaster,  
acts as a sequence-specific endonuclease.  
J Virol. 1999 : 73(8):7061-4.
- 180- **Leblanc P, Desset S, Dastugue B, Vaury C.**  
Invertebrate retroviruses: ZAM a new candidate in D.melanogaster.  
EMBO J. 1997 : 16(24):7521-31.
- 181- **Leblanc P, Desset S, Giorgi F, Taddei AR, Fausto AM, Mazzini M, Dastugue B, Vaury C.**  
Life cycle of an endogenous retrovirus, ZAM, in Drosophila melanogaster.  
J Virol. 2000 : 74(22):10658-69.
- 182- **Lecher P, Bucheton A, Pelisson A.**  
Expression of the Drosophila retrovirus gypsy as ultrastructurally detectable particles in the ovaries of  
flies carrying a permissive flamenco allele.  
J Gen Virol. 1997 : 78 ( Pt 9):2379-88.
- 183- **Lee BS, Lichtenstein CP, Faiola B, Rinckel LA, Wysock W, Curcio MJ, Garfinkel DJ.**  
Posttranslational inhibition of Ty1 retrotransposition by nucleotide excision repair/transcription factor  
TFIIH subunits Ssl2p and Rad3p.  
Genetics. 1998 : 148(4):1743-61.
- 184- **Lee TN, Singer MF.**  
Structural organization of alpha-satellite DNA in a single monkey chromosome.  
J Mol Biol. 1982 : 161(2):323-42.
- 185- **Lehrman MA, Russell DW, Goldstein JL, Brown MS.**  
Exon-Alu recombination deletes 5 kilobases from the low density lipoprotein receptor gene, producing a  
null phenotype in familial hypercholesterolemia.  
Proc Natl Acad Sci U S A. 1986 : 83(11):3679-83.
- 186- **Leis J, Baltimore D, Bishop JM, Coffin J, Fleissner E, Goff SP, Oroszlan S, Robinson H,  
Skalka AM, Temin HM, et al.**  
Standardized and simplified nomenclature for proteins common to all retroviruses.  
J Virol. 1988 : 62(5):1808-9.
- 187- **Lenski RE.**  
Quantifying fitness and gene stability in microorganisms.  
Biotechnology. 1991;15:173-92..
- 188- **Lenski RE, Travisano M.**  
Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations.  
Proc Natl Acad Sci U S A. 1994 : 91(15):6808-14.
- 189- **Lerat E, Capy P.**  
Retrotransposons and retroviruses: analysis of the envelope gene.  
Mol Biol Evol. 1999 : 16(9):1198-207.
- 190- **Li L, Farnet CM, Anderson WF, Bushman FD.**  
Modulation of activity of Moloney murine leukemia virus preintegration complexes by host factors in vitro.  
J Virol. 1998 : 72(3):2125-31.

- 191- **Li YX, Farrell MJ, Liu R, Mohanty N, Kirby ML.**  
 Double-stranded RNA injection produces null phenotypes in zebrafish.  
 Dev Biol. 2000 : 217(2):394-405.
- 192- **Lilly, F. and T. Pincus.**  
 Genetic control of murine viral leukemogenesis.  
 Adv. Cancer Res. 1973 : 17, 231-277.
- 193- **Lim JK, Simmons MJ, Raymond JD, Cox NM, Doll RF, Culbert TP.**  
 Homologue destabilization by a putative transposable element in *Drosophila melanogaster*.  
 Proc Natl Acad Sci U S A. 1983 : 80(21):6624-7.
- 194- **Limjoco TI, Dickie P, Ikeda H, Silver J.**  
 Transgenic Fv-4 mice resistant to Friend virus.  
 J Virol. 1993 : 67(7):4163-8.
- 195- **Linial ML**  
 Foamy viruses are unconventional retroviruses  
 J Virol. 1999 : 73(3):1747-1755
- 196- **Lindsley DL, Zimm GG.**  
 The genome of *Drosophila melanogaster*.  
 1992 London, Academic Press, Incorporated
- 197- **Liu WM, Chu WM, Choudary PV, Schmid CW.**  
 Cell stress and translational inhibitors transiently increase the abundance of mammalian SINE transcripts.  
 Nucleic Acids Res. 1995 : 23(10):1758-65.
- 198- **Liu WM, Schmid CW.**  
 Proposed roles for DNA methylation in Alu transcriptional repression and mutational inactivation.  
 Nucleic Acids Res. 1993 : 21(6):1351-9.
- 199- **Luan DD, Korman MH, Jakubczak JL, Eickbush TH.**  
 Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition.  
 Cell. 1993 : 72(4):595-605.
- 200- **Lower R.**  
 The pathogenic potential of endogenous retroviruses: facts and fantasies.  
 Trends Microbiol. 1999 : 7(9):350-6. Review.
- 201- **Madhani HD, Styles CA, Fink GR.**  
 MAP kinases with distinct inhibitory functions impart signaling specificity during yeast differentiation.  
 Cell. 1997 : 91(5):673-84.
- 202- **Maichele AJ, Farwell NJ, Chamberlain JS.**  
 A B2 repeat insertion generates alternate structures of the mouse muscle gamma-phosphorylase kinase gene.  
 Genomics. 1993 : 16(1):139-49.
- 203- **Makalowski W, Mitchell GA, Labuda D.**  
 Alu sequences in the coding regions of mRNA: a source of protein variability.  
 Trends Genet. 1994 : 10(6):188-93.
- 204- **Malik HS, Eickbush TH.**  
 Modular evolution of the integrase domain in the Ty3/Gypsy class of LTR retrotransposons.  
 J Virol. 1999 : 73(6):5186-90.
- 205- **Malik HS, Burke WD, Eickbush TH.**  
 The age and evolution of non-LTR retrotransposable elements.  
 Mol Biol Evol. 1999 : 16(6):793-805.
- 206- **Malik HS, Henikoff S, Eickbush TH.**  
 Poised for contagion: evolutionary origins of the infectious abilities of invertebrate retroviruses.  
 Genome Res. 2000 : 10(9):1307-18.
- 207- **Malinsky S, Bucheton A, Busseau I.**  
 New insights on homology-dependent silencing of I factor activity by transgenes containing ORF1 in *Drosophila melanogaster*.  
 Genetics. 2000 : 156(3):1147-55.
- 208- **Manninen I, Schulman AH.**  
 BARE-1, a copia-like retroelement in barley (*Hordeum vulgare* L.).  
 Plant Mol Biol. 1993 : 22(5):829-46.
- 209- **Marçais B, Bellis M, Gérard A, Pagès M, Boublik Y, Roizès G.**  
 Structural organization and polymorphism of the alpha satellite DNA sequences of chromosome 13 and 21 as revealed by pulse field electrophoresis.  
 Hum. Genet 1991 : 86:311-316

- 210- **Marin I, Llorens C.**  
Ty3/Gypsy retrotransposons: description of new *Arabidopsis thaliana* elements and evolutionary perspectives derived from comparative genomic data.  
Mol Biol Evol. 2000 : 17(7):1040-9.
- 211- **Marin L, Lehmann M, Nouaud D, Izaabel H, Anxolabehere D, Ronsseray S.**  
P-Element repression in *Drosophila melanogaster* by a naturally occurring defective telomeric P copy.  
Genetics. 2000 : 155(4):1841-54.
- 212- **Marlor RL, Parkhurst SM, Corces VG.**  
The *Drosophila melanogaster* gypsy transposable element encodes putative gene products homologous to retroviral proteins.  
Mol Cell Biol. 1986 : 6(4):1129-34.
- 213- **Martin SL.**  
Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells.  
Mol Cell Biol. 1991 : 11(9):4804-7.
- 214- **Mathias SL, Scott AF, Kazazian HH Jr, Boeke JD, Gabriel A.**  
Reverse transcriptase encoded by a human transposable element.  
Science. 1991 : 254(5039):1808-10.
- 215- **Matzke MA, Moscone EA, Park YD, Papp I, Oberkofler H, Neuhuber F, Matzke AJ.**  
Inheritance and expression of a transgene insert in an aneuploid tobacco line.  
Mol Gen Genet. 1994 : 245(4):471-85.
- 216- **Mc Clintock B.**  
Induction of instability at selected loci in maize.  
Genetics 1955 : 38:579-599
- 217- **McClure MA**  
Evolution of retrotransposons by acquisition or deletion of retrovirus-like genes.  
Mol. Biol. Evol. 1991 8/835-856
- 218- **McDonalk JF.**  
Transposable elements, gene silencing and macroevolution.  
Trends Ecol. Evol. 1998 : 13:94-95
- 219- **McDonald JF.**  
Genomic imprinting as a coopted evolutionary character.  
Trends Ecol Evol. 1999 : 14(9):359.
- 220- **Medstrand P, Mager DL.**  
Human-specific integrations of the HERV-K endogenous retrovirus family.  
J Virol. 1998 : 72(12):9782-7.
- 221- **Meisler MH, Ting CN.**  
The remarkable evolutionary history of the human amylase genes.  
Crit Rev Oral Biol Med. 1993 : 4(3-4):503-9.
- 222- **Merritt CM, Easteal S, Board PG.**  
Evolution of human alpha 1-acid glycoprotein genes and surrounding Alu repeats.  
Genomics. 1990 : 6(4):659-65.
- 223- **Mével-ninio, M., C. Mariol and M. Gans.**  
Mobilization of *gypsy* and *copia* retrotransposons in *Drosophila melanogaster* induces reversion of the *ovo<sup>D</sup>* dominant female-sterile mutations :molecular analysis of revertant alleles.  
EMBO J. 1989 : 8, 1549-1558.
- 224- **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.**  
Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis.  
Nature. 2000 : 403(6771):785-9.
- 225- **Miki Y, Nishisho I, Horii A, Miyoshi Y, Utsunomiya J, Kinzler KW, Vogelstein B, Nakamura Y.**  
Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer.  
Cancer Res. 1992 : 52(3):643-5.
- 226- **Miklos GL, Cotsell JN.**  
Chromosome structure at interfaces between major chromatin types: alpha- and beta-heterochromatin.  
Bioessays. 1990 : 12(1):1-6.
- 227- **Miller K, Lynch C, Martin J, Herniou E, Tristem M.**  
Identification of multiple Gypsy LTR-retrotransposon lineages in vertebrate genomes.  
J Mol Evol. 1999 : 49(3):358-66.
- 228- **Miller WJ, Hagemann S, Reiter E, Pinsker W.**  
P-element homologous sequences are tandemly repeated in the genome of *Drosophila guanche*.  
Proc Natl Acad Sci U S A. 1992 : 89(9):4018-22.

- 229- **Miller WJ, Paricio N, Hagemann S, Martinez-Sebastian MJ, Pinsker W, de Frutos R.**  
Structure and expression of clustered P element homologues in *Drosophila subobscura* and *Drosophila guanche*.  
*Gene*. 1995 : 156(2):167-74.
- 230- **Misra S, Rio DC.**  
Cytotype control of *Drosophila* P element transposition: the 66 kd protein is a repressor of transposase activity.  
*Cell*. 1990 : 62(2):269-84.
- 231- **Modolell J, Bender W, Meselson M.**  
*Drosophila melanogaster* mutations suppressible by the suppressor of Hairy-wing are insertions of a 7.3-kilobase mobile element.  
*Proc Natl Acad Sci U S A*. 1983 : 80(6):1678-82.
- 232- **Molto MD, Paricio N, Lopez-Preciado MA, Semeshin VF, Martinez-Sebastian MJ.**  
Tirant: a new retrotransposon-like element in *Drosophila melanogaster*.  
*J Mol Evol*. 1996 : 42(3):369-75.
- 233- **Montgomery MK, Fire A.**  
Double-stranded RNA as a mediator in sequence-specific genetic silencing and co-suppression.  
*Trends Genet*. 1998 : 14(7):255-8.
- 234- **Moore JK, Haber JE.**  
Capture of retrotransposon DNA at the sites of chromosomal double-strand breaks.  
*Nature*. 1996 : 383(6601):644-6.
- 235- **Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, Kazazian HH Jr.**  
High frequency retrotransposition in cultured mammalian cells.  
*Cell*. 1996 : 87(5):917-27.
- 236- **Morgan HD, Sutherland HG, Martin DI, Whitelaw E.**  
Epigenetic inheritance at the agouti locus in the mouse.  
*Nat Genet*. 1999 : 23(3):314-8.
- 237- **Mourrain P, Beclin C, Elmayer T, Feuerbach F, Godon C, Morel JB, Jouette D, Lacombe AM, Nikic S, Picault N, Remoue K, Sanial M, Vo TA, Vaucheret H.**  
*Arabidopsis* SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance.  
*Cell*. 2000 : 101(5):533-42.
- 238- **Naas T, Blot M, Fitch WM, Arber W.**  
Dynamics of IS-related genetic rearrangements in resting *Escherichia coli* K-12.  
*Mol Biol Evol*. 1995 : 12(2):198-207.
- 239- **Naas T, Blot M, Fitch WM, Arber W.**  
Insertion sequence-related genetic variation in resting *Escherichia coli* K-12.  
*Genetics*. 1994 : 136(3):721-30.
- 240- **Naas TP, DeBerardinis RJ, Moran JV, Ostertag EM, Kingsmore SF, Seldin MF, Hayashizaki Y, Martin SL, Kazazian HH.**  
An actively retrotransposing, novel subfamily of mouse L1 elements.  
*EMBO J*. 1998 : 17(2):590-7.
- 241- **Naggert JK, Mu JL, Frankel W, Bailey DW, Paigen B.**  
Genomic analysis of the C57BL/Ks mouse strain.  
*Mamm Genome*. 1995 : 6(2):131-3.
- 242- **Nakamura TM, Cech TR.**  
Reversing time: origin of telomerase.  
*Cell*. 1998 : 92(5):587-90.
- 243- **Ngo H, Tschudi C, Gull K, Ullu E.**  
Double-stranded RNA induces mRNA degradation in *Trypanosoma brucei*.  
*Proc Natl Acad Sci U S A*. 1998 : 95(25):14687-92.
- 244- **Norris J, Fan D, Aleman C, Marks JR, Futreal PA, Wiseman RW, Iglehart JD, Deininger PL, McDonnell DP.**  
Identification of a new subclass of Alu DNA repeats which can function as estrogen receptor-dependent transcriptional enhancers.  
*J Biol Chem*. 1995 Sep 29;270(39):22777-82.
- 245- **Nouaud D, Anxolabehere D.**  
P element domestication: a stationary truncated P element may encode a 66-kDarepressor-like protein in the *Drosophila montium* species subgroup.  
*Mol Biol Evol*. 1997 : 14(11):1132-44.

**246- Nymark-McMahon MH, Sandmeyer SB.**

Mutations in nonconserved domains of Ty3 integrase affect multiple stages of the Ty3 life cycle.  
*J Virol.* 1999 : 73(1):453-65.

**247- O'Hare K, Driver A, McGrath S, Johnson-Schiltz DM.**

Distribution and structure of cloned P elements from the *Drosophila melanogaster* P strain pi2.  
*Genet Res.* 1992 : 60(1):33-41.

**248- O'Hare K, Rubin GM.**

Structures of P transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome.  
*Cell.* 1983 : 34(1):25-35.

**249- Ohshima K, Okada N.**

Generality of the tRNA origin of short interspersed repetitive elements (SINEs). Characterization of three different tRNA-derived retroposons in the octopus.  
*J Mol Biol.* 1994 : 243(1):25-37.

**250- O'Neill RJ, O'Neill MJ, Graves JA.**

Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid.  
*Nature.* 1998 : 393(6680):68-72.

**251- Okada N.**

SINEs.  
*Curr Opin Genet Dev.* 1991 : 1(4):498-504.

**252- Okada N, Hamada M, Ogiwara I, Ohshima K.**

SINEs and LINEs share common 3' sequences: a review.  
*Gene.* 1997 : 205(1-2):229-43.

**253- Ottolenghi S, Giglioni B.**

The deletion in a type of delta 0-beta 0-thalassaemia begins in an inverted Alu repeat.  
*Nature.* 1982 : 300(5894):770-1.

**254- Pantazidis A, Labrador M, Fontdevila A.**

The retrotransposon *Oswaldo* from *Drosophila buzzatii* displays all structural features of a functional retrovirus.  
*Mol Biol Evol.* 1999 : 16(7):909-21.

**255- Papadopoulos D, Schneider D, Meier-Eiss J, Arber W, Lenski RE, Blot M.**

Genomic evolution during a 10,000-generation experiment with bacteria.  
*Proc Natl Acad Sci U S A.* 1999 : 96(7):3807-12.

**256- Pardue ML.**

Transposable elements: friends, foes, or merely fellow travelers?  
*Trends Genet.* 2000 : 16(4):155-6.

**257- Pardue ML, Danilevskaya ON, Lowenhaupt K, Slot F, Traverse KL.**

*Drosophila* telomeres: new views on chromosome evolution.  
*Trends Genet.* 1996 : 12(2):48-52.

**258- Paricio N, Perez-Alonso M, Martinez-Sebastian MJ, de Frutos R.**

P sequences of *Drosophila subobscura* lack exon 3 and may encode a 66 kd repressor-like protein.  
*Nucleic Acids Res.* 1991 : 19(24):6713-8.

**259- Park YD, Papp I, Moscone EA, Iglesias VA, Vaucheret H, Matzke AJ, Matzke MA.**

Gene silencing mediated by promoter homology occurs at the level of transcription and results in meiotically heritable alterations in methylation and gene activity.  
*Plant J.* 1996 : 9(2):183-94.

**260- Pasyukova EG, Nuzhdin SV.**

Doc and copia instability in an isogenic *Drosophila melanogaster* stock.  
*Mol Gen Genet.* 1993 : 240(2):302-6.

**261- Payne LN, Pani PK, Weiss RA.**

A dominant epistatic gene which inhibits cellular susceptibility to RSV(RAV-O).  
*J Gen Virol.* 1971 : 13(3):455-62.

**262- Peifer M, Bender W.**

Sequences of the gypsy transposon of *Drosophila* necessary for its effects on adjacent genes.  
*Proc Natl Acad Sci U S A.* 1988 : 85(24):9650-4.

**263- Pelisson A, Bregliano JC.**

Evidence for rapid limitation of the I element copy number in a genome submitted to several generations of I-R hybrid dysgenesis in *Drosophila melanogaster*.  
*Mol. Gen. Genet.* 1987 : 207 : 306-313

- 264- **Pelisson A, Song SU, Prud'homme N, Smith PA, Bucheton A, Corces VG.**  
Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the *Drosophila flamenco* gene.  
EMBO J. 1994 : 13(18):4401-11.
- 265- **Pelisson A, Teyssset L, Chalvet F, Kim A, Prud'homme N, Terzian C, Bucheton A.**  
About the origin of retroviruses and the co-evolution of the gypsy retrovirus with the *Drosophila flamenco* host gene.  
Genetica. 1997;100(1-3):29-37.
- 266- **Petrov DA, Schutzman JL, Hartl DL, Lozovskaya ER.**  
Diverse transposable elements are mobilized in hybrid dysgenesis in *Drosophila virilis*.  
Proc Natl Acad Sci U S A. 1995 : 92(17):8050-4.
- 267- **Picard G.**  
Non-mendelian female sterility in *Drosophila melanogaster*: hereditary transmission of I factor.  
Genetics. 1976 : 83(1):107-23.
- 268- **Picard G, Bregliano JC, Bucheton A, Lavigne JM, Pelisson A, Kidwell MG.**  
Non-mendelian female sterility and hybrid dysgenesis in *Drosophila melanogaster*.  
Genet Res. 1978 : 32(3):275-87.
- 269- **Plasterk RH, Izsvak Z, Ivics Z.**  
Resident aliens: the Tc1/mariner superfamily of transposable elements.  
Trends Genet. 1999 : 15(8):326-32.
- 270- **Ponomarenko NA, Bannikov VM, Anashchenko VA, Tchurikov NA.**  
Burdock, a novel retrotransposon in *Drosophila melanogaster*, integrates into the coding region of the cut locus.  
FEBS Lett. 1997 : 413(1):7-10.
- 271- **Preston CR and Engels WR.**  
Spread of P transposable elements in inbred lines of *Drosophila melanogaster*.  
In Progress in Nucleic Acid Research and Molecular Biology: Hollaender Symposium Proceeding, 1989 : edited by W Cohn and K Moldave. Academic Press, San Diego. Vol 36: pp 71-85
- 272- **Pret AM, Searles LL.**  
Splicing of retrotransposon insertions from transcripts of the *Drosophila melanogaster* vermilion gene in a revertant.  
Genetics. 1991 : 129(4):1137-45.
- 273- **Pritchard MA, Dura JM, Pelisson A, Bucheton A, Finnegan DJ.**  
A cloned I-factor is fully functional in *Drosophila melanogaster*.  
Mol Gen Genet. 1988 : 214(3):533-40.
- 274- **Prud'homme N, Gans M, Masson M, Terzian C, Bucheton A.**  
Flamenco, a gene controlling the gypsy retrovirus of *Drosophila melanogaster*.  
Genetics. 1995 : 139(2):697-711.
- 275- **Puechberty J, Laurent AM, Gimenez S, Billault A, Brun-Laurent ME, Calenda A, Marcais B, Prades C, Ioannou P, Yurov Y, Roizes G.**  
Genetic and physical analyses of the centromeric and pericentromeric regions of human chromosome 5: recombination across 5cen.  
Genomics. 1999 : 56(3):274-87.
- 276- **Rasmusson KE, Raymond JD, Simmons MJ.**  
Repression of hybrid dysgenesis in *Drosophila melanogaster* by individual naturally occurring P elements.  
Genetics. 1993 : 133(3):605-22.
- 277- **Reis D, Quesneville H, Nouaud D, Andrieu O, Anxolabéhère D.**  
Un nouvel élément de la super famille P : L'élément hoppel.  
X ème colloque « Eléments transposables », 2001, Lyon.
- 278- **Reiter LT, Murakami T, Koeuth T, Pentao L, Muzny DM, Gibbs RA, Lupski JR.**  
A recombination hotspot responsible for two inherited peripheral neuropathies is located near a mariner transposon-like element.  
Nat Genet. 1996 : 12(3):288-97.
- 279- **Rio DC, Laski FA, Rubin GM.**  
Identification and immunochemical analysis of biologically active *Drosophila* Pelement transposase.  
Cell. 1986 : 44(1):21-32.
- 280- **Robert V.**  
Caractérisation de la région du génome de *Drosophila melanogaster* contenant la fonction flamenco responsable de la répression de l'errantivirus gypsy.  
Thèse. Génétique cellulaire et moléculaire. Versailles 2000
- 281- **Robert V, Prud'homme N, Kim A, Bucheton A, Pelisson A.**

- Characterization of the flamenco Region of the *Drosophila melanogaster* Genome.  
Genetics. 2001 : 158(2):701-13.
- 282- Robertson HM, Engels WR.**  
Modified P elements that mimic the P cytotype in *Drosophila melanogaster*.  
Genetics. 1989 : 123(4):815-24.
- 283- Roche SE, Rio DC.**  
Trans-silencing by P elements inserted in subtelomeric heterochromatin involves the *Drosophila* Polycomb group gene, Enhancer of zeste.  
Genetics. 1998 : 149(4):1839-55.
- 284- Roche SE, Schiff M, Rio DC.**  
P-element repressor autoregulation involves germ-line transcriptional repression and reduction of third intron splicing.  
Genes Dev. 1995 : 9(10):1278-88.
- 285- Rogers MF**  
Oncogene chromosome breakpoints and alu sequences.  
Nature 1985 : 317(6037) 559
- 286- Rong YS, Golic KG.**  
Gene targeting by homologous recombination in *Drosophila*.  
Science. 2000 : 288(5473):2013-8.
- 287- Ronsseray S, Lehmann M, Anxolabehere D.**  
Copy number and distribution of P and I mobile elements in *Drosophila melanogaster* populations.  
Chromosoma. 1989 : 98(3):207-14.
- 288- Ronsseray S, Lehmann M, Nouaud D, Anxolabehere D.**  
P element regulation and X-chromosome subtelomeric heterochromatin in *Drosophila melanogaster*.  
Genetica. 1997;100(1-3):95-107
- 289- Ronsseray S, Lehmann M, Nouaud D, Anxolabehere D.**  
The regulatory properties of autonomous subtelomeric P elements are sensitive to a Suppressor of variegation in *Drosophila melanogaster*.  
Genetics. 1996 Aug;143(4):1663-74.
- 290- Ronsseray S, Lemaitre B, Coen D.**  
Maternal inheritance of P cytotype in *Drosophila melanogaster*: a "pre-P cytotype" is strictly extra-chromosomally transmitted.  
Mol Gen Genet. 1993 : 241(1-2):115-23.
- 291- Ronsseray S, Marin L, Lehmann M, Anxolabehere D.**  
Repression of hybrid dysgenesis in *Drosophila melanogaster* by combinations of telomeric P-element reporters and naturally occurring P elements.  
Genetics. 1998 : 149(4):1857-66.
- 292- Rountree MR, Selker EU.**  
DNA methylation inhibits elongation but not initiation of transcription in *Neurospora crassa*.  
Genes Dev. 1997 : 11(18):2383-95.
- 293- Rubin GM, Kidwell MG, Bingham PM.**  
The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations.  
Cell. 1982 : 29(3):987-94.
- 294- Rubin GM, Spradling AC.**  
Vectors for P element-mediated gene transfer in *Drosophila*.  
Nucleic Acids Res. 1983 : 11(18):6341-51.
- 295- Rubin GM, Spradling AC.**  
Genetic transformation of *Drosophila* with transposable element vectors.  
Science. 1982 : 218(4570):348-53.
- 296- Saigo K, Kugimiya W, Matsuo Y, Inouye S, Yoshioka K, Yuki S.**  
Identification of the coding sequence for a reverse transcriptase-like enzyme in a transposable genetic element in *Drosophila melanogaster*.  
Nature. 1984 : 312(5995):659-61.
- 297- Sandmeyer S.**  
Targeting transposition: at home in the genome.  
Genome Res. 1998 : 8(5):416-8.
- 298- SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL.**  
Nested retrotransposons in the intergenic regions of the maize genome.  
Science. 1996 : 274(5288):765-8.

- 299- **Sarot E, Payen-Groschene G, Bucheton A, Pélisson A**  
Recherche de séquences de l'élément gypsy impliqués dans la régulation par le gène flamenco.  
Xème colloque « éléments transposables » 2001, Lyon,
- 300- **Sarrowa J, Chang DY, Maraia RJ.**  
The decline in human Alu retroposition was accompanied by an asymmetric decrease in SRP9/14 binding to dimeric Alu RNA and increased expression of small cytoplasmic Alu RNA.  
Mol Cell Biol. 1997 : 17(3):1144-51.
- 301- **Sassaman DM, Dombroski BA, Moran JV, Kimberland ML, Naas TP, DeBerardinis RJ, Gabriel A, Swergold GD, Kazazian HH Jr.**  
Many human L1 elements are capable of retrotransposition.  
Nat Genet. 1997 : 16(1):37-43.
- 302- **Schmid CW**  
Does SINE evolution preclude Alu function?  
Nucleic Acids Res 1998 : 26 (20) 4541-4550
- 303- **Schneider D, Duperchy E, Coursange E, Lenski RE, Blot M.**  
Long-term experimental evolution in Escherichia coli. IX. Characterization of insertion sequence-mediated mutations and rearrangements.  
Genetics. 2000 : 156(2):477-88.
- 304- **Scott AF, Schmeckpeper BJ, Abdelrazik M, Comey CT, O'Hara B, Rossiter JP, Cooley T, Heath P, Smith KD, Margolet L.**  
Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence.  
Genomics. 1987 : 1(2):113-25.
- 305- **Sheen F, Lim JK, Simmons MJ.**  
Genetic instability in Drosophila melanogaster mediated by hobo transposable elements.  
Genetics. 1993 : 133(2):315-34.
- 306- **Simmen MW, Leitgeb S, Charlton J, Jones SJ, Harris BR, Clark VH, Bird A.**  
Nonmethylated transposable elements and methylated genes in a chordate genome.  
Science. 1999 : 283(5405):1164-7.
- 307- **Simmons MJ, Raymond JD, Rasmusson KE, Miller LM, McLarnon CF, Zunt JR.**  
Repression of P element-mediated hybrid dysgenesis in Drosophila melanogaster.  
Genetics. 1990 : 124(3):663-76.
- 308- **Singer MF.**  
SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes.  
Cell. 1982 : 28(3):433-4.
- 309- **Smit AF, Riggs AD.**  
Tiggers and DNA transposon fossils in the human genome.  
Proc Natl Acad Sci U S A. 1996 Feb 20;93(4):1443-8.
- 310- **Smit AF, Riggs AD.**  
MIRs are classic, tRNA-derived SINEs that amplified before the mammalian radiation.  
Nucleic Acids Res. 1995 : 23(1):98-102.
- 311- **Song SU, Gerasimova T, Kurkulos M, Boeke JD, Corces VG.**  
An env-like protein encoded by a Drosophila retroelement: evidence that gypsy is an infectious retrovirus.  
Genes Dev. 1994 : 8(17):2046-57.
- 312- **Spradling AC, Rubin GM.**  
Transposition of cloned P elements into Drosophila germ line chromosomes.  
Science. 1982 : 218(4570):341-7.
- 313- **Spradling AC, Stern DM, Kiss I, Roote J, Laverly T, Rubin GM.**  
Gene disruptions using P transposable elements: an integral component of the Drosophila genome project.  
Proc Natl Acad Sci U S A. 1995 : 92(24):10824-30.
- 314- **Stavenhagen JB, Robins DM.**  
An ancient provirus has imposed androgen regulation on the adjacent mouse sex-limited protein gene.  
Cell. 1988 : 55(2):247-54.
- 315- **Sverdlov ED.**  
Retroviruses and primate evolution.  
Bioessays. 2000 : 22(2):161-71.
- 316- **Tabara H, Sarkissian M, Kelly WG, Fleenor J, Grishok A, Timmons L, Fire A, Mello**  
The rde-1 gene, RNA interference, and transposon silencing in C. elegans.  
Cell. 1999 Oct 15;99(2):123-32.

- 317- **Tanda S, Shrimpton AE, Chueh LL, Itayama H, Matsubayashi H, Saigo K, Tobar YN, Langley CH.**  
 Retrovirus-like features and site specific insertions of a transposable element, tom, in *Drosophila ananassae*.  
 Mol Gen Genet. 1988 Nov;214(3):405-11.
- 318- **Taruscio D, Mantovani A.**  
 Human endogenous retroviral sequences: possible roles in reproductive physiopathology.  
 Biol Reprod. 1998 : 59(4):713-24.
- 319- **Tatout C, Lavie L, Deragon JM.**  
 Similar target site selection occurs in integration of plant and mammalian retroposons.  
 J Mol Evol. 1998 : 47(4):463-70.
- 320- **Tchenio T, Heidmann T.**  
 Defective retroviruses can disperse in the human genome by intracellular transposition.  
 J Virol. 1991 : 65(4):2113-8.
- 321- **Tcheressiz S., Calco V., Arnaud F., Arthaud L., Dastugue B., Vaury C**  
 Idefix retrotransposon expression in the early follicle cells in the germarium is determined by its LTR sequences and a specific genomic context.  
 Mol Gen. Genet. 2002 in press.
- 322- **Temin HM.**  
 Reverse transcription in the eukaryotic genome: retroviruses, pararetroviruses, retrotransposons, and retrotranscripts.  
 Mol Biol Evol. 1985 : 2(6):455-68.
- 323- **Temin HM.**  
 Origin of retroviruses from cellular moveable genetic elements.  
 Cell. 1980 : 21(3):599-600.
- 324- **Temin, H. M. and S. Mizutani.**  
 RNA-directed DNA-polymerase in virions of Rous sarcoma virus.  
 Nature 1970 : 226, 1211-1213.
- 325- **Teng SC, Kim B, Gabriel A.**  
 Retrotransposon reverse-transcriptase-mediated repair of chromosomal breaks.  
 Nature. 1996 : 383(6601):641-4.
- 326- **Terzian C, Pélisson A, Bucheton A**  
 Evolution and phylogeny of Insect endogenous retroviruses  
 BMC Evol Biol. 2001 : 1(1):3  
<http://www.biomedcentral.com/1471-2148/1/3>
- 327- **Terzian C, Pélisson A, Bucheton A**  
 Evolution and phylogeny of Insect errantiviruses  
 IX ème colloque « éléments transposables » 2000, Strasbourg
- 328- **Tonjes RR, Czauderna F, Kurth R.**  
 Genome-wide screening, cloning, chromosomal assignment, and expression of full-length human endogenous retrovirus type K.  
 J Virol. 1999 : 73(11):9187-95.
- 329- **Tseng JC, Zollman S, Chain AC, Laski FA.**  
 Splicing of the *Drosophila* P element ORF2-ORF3 intron is inhibited in a human cell extract.  
 Mech Dev. 1991 : 35(1):65-72.
- 330- **Tsubota S, Ashburner M, Schedl P.**  
 P-element-induced control mutations at the r gene of *Drosophila melanogaster*.  
 Mol Cell Biol. 1985 : 5(10):2567-74.
- 331- **Tsukiyama T, Niwa O.**  
 Isolation of high affinity cellular targets of the embryonal LTR binding protein, an undifferentiated embryonal carcinoma cell-specific repressor of Moloney leukemia virus.  
 Nucleic Acids Res. 1992 : 20(7):1477-82.
- 332- **Tsukiyama T, Ueda H, Hirose S, Niwa O.**  
 Embryonal long terminal repeat-binding protein is a murine homolog of FTZ-F1, a member of the steroid receptor superfamily.  
 Mol Cell Biol. 1992 : 12(3):1286-91.
- 333- **Tu Z.**  
 Three novel families of miniature inverted-repeat transposable elements are associated with genes of the yellow fever mosquito, *Aedes aegypti*.  
 Proc Natl Acad Sci U S A. 1997 : 94(14):7475-80.

- 334- **Udomkit A, Forbes S, McLean C, Arkhipova I, Finnegan DJ.**  
Control of expression of the I factor, a LINE-like transposable element in *Drosophila melanogaster*.  
EMBO J. 1996 : 15(12):3174-81.
- 335- **Ullu E, Murphy S, Melli M.**  
Human 7SL RNA consists of a 140 nucleotide middle-repetitive sequence inserted in an alu sequence.  
Cell. 1982 : 29(1):195-202.
- 336- **Ullu E, Tschudi C.**  
Alu sequences are processed 7SL RNA genes.  
Nature. 1984 : 312(5990):171-2.
- 337- **Varagona MJ, Purugganan M, Wessler SR.**  
Alternative splicing induced by insertion of retrotransposons into the maize waxy gene.  
Plant Cell. 1992 : 4(7):811-20.
- 338- **Varmus HE.**  
The molecular genetics of cellular oncogenes.  
Annu Rev Genet. 1984;18:553-612.
- 339- **Vaucheret H.**  
Promoter dependant trans-inactivation in transgenic tobacco plant : kinetic aspects of gene silencing and gene reactivation.  
C.R. Acad. Sci. Paris 1994 : 317-310-323
- 340- **Vaucheret H.**  
Identification of a general silencer for 19s and 35S promoters in a transgenic tobacco plant : 90 pb of homology in the promoter sequence are sufficient for a trans-inactivation.  
C.R. Acad. Sci. Paris 1993 : 316 :1471-1483
- 341- **Vaucheret H, Elmayan T, Mourrain Ph, Palauqui JC.**  
Analysis of tobacco transgene locus that triggers both transcriptional and posttranscriptional silencing.  
In epigenetic mechanisms of gene regulation. 1996 : ed Russo VEA, Martinssen RA, Riggs AD. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- 342- **Vaury C, Bucheton A, Pelisson A.**  
The beta heterochromatic sequences flanking the I elements are themselves defective transposable elements.  
Chromosoma. 1989 : 98(3):215-24.
- 343- **Venables PJ, Brookes SM, Griffiths D, Weiss RA, Boyd MT.**  
Abundance of an endogenous retroviral envelope protein in placental trophoblasts suggests a biological function.  
Virology. 1995 : 211(2):589-92.
- 344- **Vicient CM, Suoniemi A, Ananthawat-Jonsson K, Tanskanen J, Beharav A, Nevo E, Schulman AH.**  
Retrotransposon BARE-1 and Its Role in Genome Evolution in the Genus *Hordeum*.  
Plant Cell. 1999 : 11(9):1769-1784.
- 345- **Vieira-Heddi C, Arpin C, Nardon C, Biémont C**  
Evolution de la taille du génome chez la drosophile. La part des éléments transposables.  
X ème colloque « éléments transposables » 2001 Lyon
- 346- **Wallace MR, Andersen LB, Saulino AM, Gregory PE, Glover TW, Collins FS.**  
A de novo Alu insertion results in neurofibromatosis type 1.  
Nature. 1991 : 353(6347):864-6.
- 347- **Walsh CP, Chaillet JR, Bestor TH.**  
Transcription of IAP endogenous retroviruses is constrained by cytosine methylation.  
Nat Genet. 1998 : 20(2):116-7..
- 348- **Weiner AM.**  
An abundant cytoplasmic 7S RNA is complementary to the dominant interspersed middle repetitive DNA sequence family in the human genome.  
Cell. 1980 : 22(1 Pt 1):209-18.
- 349- **Weiner AM, Deininger PL, Efstratiadis A.**  
Nonviral retroposons: genes, pseudogenes, and transposable elements generated by  
Annu Rev Biochem. 1986 : 55:631-61
- 350- **Weiss RA.**  
Foamy retroviruses. A virus in search of a disease.  
Nature. 1988 : 333(6173):497-8.
- 351- **Wessler SR.**  
The splicing of maize transposable elements from pre-mRNA--a minireview.  
Gene. 1989 : 82(1):127-33.

**352- Whalen JH, Grigliatti TA.**

Molecular characterization of a retrotransposon in *Drosophila melanogaster*, nomad, and its relationship to other retrovirus-like mobile elements.

Mol Gen Genet. 1998 : 260(5):401-9.

**353- Wianny F, Zernicka-Goetz M.**

Specific interference with gene function by double-stranded RNA in early mouse development.

Nat Cell Biol. 2000 : 2(2):70-5.

**354- Wilke CM, Adams J.**

Fitness effects of Ty transposition in *Saccharomyces cerevisiae*.

Genetics. 1992 : 131(1):31-42.

**355- Wilke CM, Maimer E, Adams J.**

The population biology and evolutionary significance of Ty elements in *Saccharomyces cerevisiae*.

Genetica. 1992 : 86(1-3):155-73.

**356- Wolffe AP, Matzke MA.**

Epigenetics: regulation through repression.

Science. 1999 : 286(5439):481-6.

**357- Xiong Y, Eickbush TH.**

Dong, a non-long terminal repeat (non-LTR) retrotransposable element from *Bombyx mori*.

Nucleic Acids Res. 1993 : 21(5):1318.

**358- Xiong Y, Eickbush TH.**

Origin and evolution of retroelements based upon their reverse transcriptase sequences.

EMBO J. 1990 : 9(10):3353-62.

**359- Xiong Y, Eickbush TH.**

Similarity of reverse transcriptase-like sequences of viruses, transposable elements, and mitochondrial introns.

Mol Biol Evol. 1988 : 5(6):675-90.

**360- Xu H, Boeke JD.**

Inhibition of Ty1 transposition by mating pheromones in *Saccharomyces cerevisiae*.

Mol Cell Biol. 1991 : 11(5):2736-43.

**361- Yang J, Malik HS, Eickbush TH.**

Identification of the endonuclease domain encoded by R2 and other site-specific, non-long terminal repeat retrotransposable elements.

Proc Natl Acad Sci U S A. 1999 Jul 6;96(14):7847-52.

**362- Yannopoulos G, Stamatis N, Monastirioti M, Hatzopoulos P, Louis C.**

hobo is responsible for the induction of hybrid dysgenesis by strains of *Drosophila melanogaster* bearing the male recombination factor 23.5MRF.

Cell. 1987 : 49(4):487-95.

**363- Yoder JA, Walsh CP, Bestor TH.**

Cytosine methylation and the ecology of intragenomic parasites.

Trends Genet. 1997 : 13(8):335-40.

**364- Young JA, Bates P, Willert K, Varmus HE.**

Efficient incorporation of human CD4 protein into avian leukosis virus particles.

Science. 1990 : 250(4986):1421-3.

**365- Yulug IG, Yulug A, Fisher EM.**

The frequency and position of Alu repeats in cDNAs, as determined by database searching.

Genomics. 1995 : 27(3):544-8.

**366- Zamore PD, Tuschl T, Sharp PA, Bartel DP.**

RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals.

Cell. 2000 : 101(1):25-33.

**367- Zhou Q, Haymer DS.**

Molecular structure of yoyo, a gypsy-like retrotransposon from the Mediterranean fruit fly, *Ceratitis capitata*.

Genetica. 1997-98;101(3):167-78.

**368- Zou S, Kim JM, Voytas DF.**

The *Saccharomyces* retrotransposon Ty5 influences the organization of chromosome ends.

Nucleic Acids Res. 1996 : 24(23):4825-31.

---

RESUME : Les rétrovirus endogènes colonisent les génomes de tous les vertébrés et sont parfois à l'origine de pathologies. Peu de données sont disponibles sur les relations qui se sont établies entre les rétrovirus et leur hôte car les études génétiques chez les vertébrés sont difficiles à mettre en oeuvre.

*ZAM* et *Idefix* sont deux rétrovirus endogènes d'un invertébré, *Drosophila melanogaster*, qui ont été chacun identifiés dans notre laboratoire. Ils sont mobilisés dans une lignée génétiquement instable où leur nombre de copies a été amplifié. Nous avons recherché les déterminants génétiques qui contrôlent la mobilisation de *ZAM* et d'*Idefix* dans cette lignée appelée Rev.

Nous avons d'abord caractérisé l'élément *ZAM* : *ZAM* possède la structure génomique complète d'un *errantivirus*, c'est à dire d'un rétrovirus endogène d'insecte. Il exprime tous les constituants nécessaires à son cycle dans les cellules somatiques des ovaires adultes de la lignée Rev, mais dans aucune autre des lignées de drosophile que nous avons testées. *ZAM* subit un contrôle transcriptionnel spécifique de souche. Nous avons recherché les déterminants génétiques responsables de ce contrôle à l'aide de drosophiles transgéniques qui possèdent un gène rapporteur dirigé par le LTR de *ZAM*. Le LTR contient en effet les séquences nécessaires à l'expression de l'élément. Nous montrons que le LTR possède également toute l'information pour répondre au contrôle transcriptionnel de souche et nous localisons les déterminants génétiques responsables de ce contrôle sur le chromosome X de l'hôte, au site cytogénétique 20A. Une étude parallèle sur la caractérisation d'*Idefix* menée au laboratoire a également mis en évidence un contrôle transcriptionnel spécifique de souche. Nous montrons que les déterminants contrôlant *Idefix* ont la même localisation 20A. Cette région 20A est déjà connue pour contrôler un autre *errantivirus*, *gypsy*. Nous l'avons donc appelée « COM » pour « Centre Organisateur de la Mobilisation ».

---

TITRE anglais : Host genetic control of two endogenous retroviruses, *ZAM* and *Idefix*, in *Drosophila melanogaster*.

---

---

DISCIPLINE: Biologie cellulaire et moléculaire

MOTS-CLES: *metaviridae*, *errantivirus*, LTR-retrotransposon, *gypsy*-like, relation hôte/rétrovirus, contrôle transcriptionnel, lignée Rev.

---

INSERM UMR384, BP38, 28 place Henri Dunant, 63001 CLERMONT-FERRAND CEDEX

