



HAL
open science

La régulation épigénétique des éléments transposables dans les populations naturelles de *Drosophila simulans*

Benjamin Hubert

► **To cite this version:**

Benjamin Hubert. La régulation épigénétique des éléments transposables dans les populations naturelles de *Drosophila simulans*. Sciences agricoles. Université Claude Bernard - Lyon I, 2010. Français. NNT : 2010LYO10345 . tel-00815956

HAL Id: tel-00815956

<https://theses.hal.science/tel-00815956>

Submitted on 19 Apr 2013

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.

THESE DE L'UNIVERSITE DE LYON

Délivrée par

L'UNIVERSITE CLAUDE BERNARD LYON 1

L'ECOLE DOCTORALE E2M2

DIPLOME DE DOCTORAT

(Arrêté du 7 août 2006)

Soutenue publiquement le 17 décembre 2010

par

Monsieur Benjamin HUBERT

**La régulation épigénétique des éléments
transposables dans les populations naturelles de
*Drosophila simulans***

Directeur de thèse : Cristina VIEIRA

JURY :

Madame Malika AINOUCHE	Rapporteur
Madame Maria Pilar GARCIA-GUERREIRO	Rapporteur
Monsieur Pierre CAPY	Rapporteur
Monsieur Christian GAUTIER	Examineur
Monsieur Christophe TERZIAN	Examineur
Madame Cristina VIEIRA	Directeur

Laboratoire de Biométrie et Biologie Evolutive
UMR CNRS 5558, Université Claude Bernard Lyon 1
43 bd du 11 novembre 1918, 69622 Villeurbanne

RESUME

La méthylation de l'ADN et les modifications post-traductionnelles des histones sont des modifications épigénétiques qui interviennent dans la régulation des éléments transposables (ET) chez de nombreuses espèces. La proportion des ET dans les génomes varie selon les espèces considérées et pose la question des mécanismes de régulation de ces ET. Au sein de l'espèce *Drosophila simulans*, les populations naturelles présentent un polymorphisme unique dans le nombre de copies des ET, ce qui en fait un excellent modèle pour étudier cette question. L'étude de la méthylation d'ADN et des modifications post-traductionnelles des histones associées au rétrotransposon à LTR tirant dans la lignée germinale des populations naturelles a permis de montrer l'influence d'une copie d'ET sur la structure de la chromatine au site d'insertion. Dans un second volet, nous avons cherché à caractériser la méthylation de l'ADN chez la drosophile, chez laquelle la fonction est encore mal connue. Nous avons, par des approches spécifiques et globales, mesuré l'abondance de cette marque épigénétique chez la drosophile. Nous concluons que les taux de méthylation de l'ADN sont très faibles mais variables entre espèces. Notre travail n'a pas permis de mettre en évidence un rôle de la méthylation de l'ADN dans le contrôle des ET, toutefois, nous ne pouvons pas exclure ce système de régulation.

TITLE

Epigenetic régulation of transposable elements in natural populations of *Drosophila simulans*

ABSTRACT

Epigenetic modifications such as DNA methylation and post-translational histone modifications are involved in transposable elements (TE) silencing in many species. Their relative abundance in genomes ask the question of differences in regulation mecanism between species. Within the *Drosophila simulans* species, natural populations exhibits a unique polymorphism in TE copy number, providing a powerfull tool for the analysis of TE regulation in population from the same specie. We analyzed DNA methylation and post-translational histone modifications associated with the LTR retrotransposon tirant in the germline of natural populations and report the influence of this element on chromatine structure. DNA methylation is a wide-conserved epigenetic modification involved in gene regulation and TE silencing but its function in drosophila remains missunderstood. Using different methods, we determined the abundance of methylated cytosines in drosophila, and showed that methylation level are low and variable between species. Our results show low evidence for a TE regulation system involving DNA methylation but this cannot be so far excluded.

DISCIPLINE

E2M2 : Évolution Écosystèmes, Microbiologie, Modélisation

MOTS-CLES

Élément transposable, épigénétique, méthylation de l'ADN, modifications post-traductionnelles des histones, rétrotransposon à LTR, *Drosophila simulans*, populations naturelles

INTITULE ET ADRESSE DE L'U.F.R. OU DU LABORATOIRE :

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

Table des matières

INTRODUCTION	9
1. Les éléments transposables	11
1.1. Classification et transposition des ET	12
1.2. Relations avec le génome.....	14
2. La régulation des ET	17
2.1. Les régulations épigénétiques.....	17
2.2. Variabilité des régulations épigénétiques	22
3. Les populations naturelles de drosophiles pour comprendre la dynamique des ET	24
3.1. Variabilité en ET dans les populations naturelles de drosophiles.....	24
3.2. Epigénétique chez la drosophile.....	25
3.3. Le Modèle <i>tirant</i>	28
CHAPITRE 1. Étude des marques épigénétiques associées à <i>tirant</i> dans des populations naturelles de <i>D. simulans</i>	31
1. Détermination et analyse des profils de méthylation de la région régulatrice de <i>tirant</i>	33
1.1. Méthodes d'analyse de la méthylation de l'ADN	33
1.2. Profils de méthylation de la LTR 5' de <i>tirant</i>	43
1.3. Biais de la technique de bisulfite.....	50
2. Analyse des modifications d'histones associées à des insertions de <i>tirant</i>	53
2.1. Principe de la ChIP.....	53
2.2. Modifications d'histones associées à <i>tirant</i>	55
2.3. Influence de <i>tirant</i> sur la chromatine chez <i>D. simulans</i>	57
3. Conclusion sur la régulation épigénétique de <i>tirant</i> chez <i>D. simulans</i>	59
CHAPITRE 2. Analyse globale de la méthylation de l'ADN dans les génomes de drosophiles	61
1. Quantification des taux de méthylation de l'ADN dans les génomes de drosophiles	63
1.1. Introduction.....	63
1.2. Détection et quantification des 5-MedC par HPLC couplée à la spectrométrie de masse (MS)	65
1.3. Résultats.....	72
2. Exploration de la présence de cytosines méthylées au niveau des ET dans des populations naturelles de <i>D. melanogaster</i> et <i>D. simulans</i>	74
2.1. Utilisation de couples d'isoschizomères pour déterminer la présence de cytosines méthylées	74
2.2. Résultats.....	78
2.3. Conclusion.....	82
DISCUSSION - PERSPECTIVES	83
1. Problèmes d'interprétation des données produites par bisulfite	85
2. Conservation de la méthylation de l'ADN chez les eucaryotes et régulation des ET	89
3. La régulation des ET chez la drosophile : exemple de <i>tirant</i> dans les populations naturelles de <i>D. simulans</i>	92
PUBLICATIONS	95
MATÉRIELS ET MÉTHODES	143
1. Origine et élevage des souches de drosophiles	145
2. Récolte des embryons	145
3. Extraction d'ADN génomique	146
4. Traitement au Bisulfite	147
5. PCR de l'ADN traité au bisulfite de sodium	147
6. Clonage/Séquençage/Analyse des profils de méthylation	148

7. Digestions enzymatiques.....	149
8. Immunoprécipitation de la Chromatine (ChIP).....	151
9. qPCR.....	154
10. Hydrolyse de l'ADN et préparation des échantillons pour la HPLC.....	155
ANNEXES.....	157
1. Méthylation de l'ADN	159
2. HPLC et spectrométrie de masse.....	241
BIBLIOGRAPHIE	247

Table des illustrations

Figure 1. Classification des ET eucaryotes	13
Figure 2: Principales modifications post-traductionnelles des lysines de l'histone H3.	19
Figure 3: Structure du rétrotransposon à LTR <i>tirant</i>	29
Figure 4: Carte des zones d'échantillonnage des souches étudiées	30
Figure 5. Etapes de conversion de la cytosine en uracile par le bisulfite de sodium	35
Figure 6. Schéma de la technique de "Bisulfite Genomic Sequencing"	36
Figure 7. Amorces utilisées pour l'amplification par PCR de la LTR 5' de <i>tirant</i> après traitement au bisulfite de sodium	38
Figure 8. Profils de méthylation de la LTR 5' de <i>tirant</i> dans les ovaires de souches mutantes pour dDNMT2.....	41
Figure 9. Profils de méthylation de la LTR 5' de <i>tirant</i> dans des ovaires de la population de Makindu	42
Figure 10. Profils de méthylation de la LTR5' de <i>tirant</i> dans la population de Makindu (<i>D.</i> <i>simulans</i>)	44
Figure 11. Profils de méthylation de la LTR5' de <i>tirant</i> dans la population du Zimbabwe (<i>D.</i> <i>simulans</i>)	45
Figure 12. Profils de méthylation de la LTR5' de <i>tirant</i> dans la population de Chicharo (<i>D.</i> <i>simulans</i>)	46
Figure 13. Profils de méthylation de la LTR5' de <i>tirant</i> dans la population du Sénégal (<i>D.</i> <i>melanogaster</i>).....	47
Figure 14. Profils de méthylation des sites d'insertion des copies complètes M1, M2 et M3 de <i>tirant</i> de la population de Makindu	50
Figure 15. Efficacité du traitement au bisulfite selon les conditions expérimentales	51
Figure 16. Profils de méthylation du brin complémentaire de la LTR 5' de <i>tirant</i> dans la population de Makindu.....	52
Figure 17. Schéma de la technique de CHIP.....	54
Figure 18. Modifications post-traductionnelles de l'histone H3 associées à <i>tirant</i>	56
Figure 19. Influence de <i>tirant</i> sur la structure de la chromatine au niveau de deux sites d'insertion.....	58
Figure 20. Structures de dC et de 5-MedC.....	65
Figure 21. Détection des dC et 5-MedC par HPLC/DAD/ESI-MS.	66
Figure 22. Gammes étalon utilisées pour la quantification des dC et 5-MedC.....	68
Figure 23. Chromatogrammes des extractions d'ions correspondant aux fragments spécifiques de la dC (<i>m/z</i> 455) et de la 5-MedC (<i>m/z</i> 483)	69
Figure 24. Tests de quantification de nucléosides standards en lecture UV et en spectrométrie de masse	71
Figure 25. Détermination du taux de 5-MedC dans les génomes de drosophiles	72
Figure 26. Recherche de sites méthylés dans la région 5' de quatre ET.....	76
Figure 27. Illustration de la technique de MSRE.....	77
Figure 28. Détection de cytosines méthylées par enzymes de restriction dans la population de <i>Sénégal (D. melanogaster)</i>	78
Figure 29. Mise en évidence de sites méthylés dans les régions 5' de quatre ET de référence.....	79
Figure 30. Fréquence de dinucléotides méthylés dans les régions 5' de <i>Roo</i> , <i>F</i> et <i>412</i>	81
Figure 31. Mise en évidence de sites méthylés dans les régions codantes de <i>F</i>	81
Figure 32. Schéma du plan expérimental de la ChIP.	151

Tableau 1. Proportion en ET et taille des génomes eucaryotes de référence.	12
Tableau 2. Insertions des copies de type C de <i>tirant</i> identifiées dans les populations de Makindu, Chicharo et Zimbabwe (<i>D. simulans</i>)	43
Tableau 3. Proportion d'ET dans les 12 génomes séquencés de drosophiles (Drosophila 12 genomes consortium, 2007)	64
Tableau 4. Nombre moyen de copies des éléments <i>F</i> , <i>Roo</i> , <i>412</i> et <i>tirant</i> déterminé par hybridation <i>in situ</i> dans les populations naturelles de <i>D. melanogaster</i> et <i>D. simulans</i>	75
Tableau 5. Amorces PCR utilisées pour l'amplification de séquences d'ADN traitées au bisulfite de sodium.	148
Tableau 6. Enzymes de restriction utilisées.	150
Tableau 7. Paramètres utilisés pour les digestions enzymatiques en fonction des enzymes utilisées.....	150
Tableau 8. Amorces PCR utilisées pour diagnostiquer l'état de méthylation des sites de restriction.....	150
Tableau 9. Amorces utilisées en qPCR pour l'analyse des résultats de CHIP.....	154
Tableau 10. Gradient HPLC.....	156

Figures supplémentaires

Figure supplémentaire 1. Profil de méthylation des séquences non converties correspondant à la LTR 5' de <i>tirant</i>	159
Figure supplémentaire 2. Fréquence de méthylation des dinucléotides dans les séquences non-converties des populations de Makindu et Zimbabwe.	160
Figure supplémentaire 3. Schéma de la copie de <i>tirant</i> M2 de la population de Makindu, insérée dans l'intron du gène <i>tkv</i>	162
Figure supplémentaire 4. Élément <i>412</i>	163
Figure supplémentaire 5. Élément <i>F</i>	163
Figure supplémentaire 6. Élément <i>Roo</i>	163
Figure supplémentaire 7. Alignements des séquences traitées au bisulfite de sodium.	164
Figure supplémentaire 8. . Chromatogrammes en courant ionique total et UV obtenus par HPLC pour les 12 génomes séquencés de drosophiles.	242

Tableau supplémentaire 1. Résultats obtenus à partir d'ADN purifié extrait à partir d'ovaires.	160
Tableau supplémentaire 2. Résultats obtenus pour les lignées mutantes.	160
Tableau supplémentaire 3. Résultats obtenus pour les populations naturelles avec les amorces tir01 et tir52.	161
Tableau supplémentaire 4. Résultats complémentaires de vérification de la présence de biais expérimentaux.	162
Tableau supplémentaire 5. Fréquence de méthylation des dinucléotides pour les éléments <i>F</i> , <i>Roo</i> et <i>412</i> dans la population de Makindu.	162

Liste des Abréviations

5-MedC : 5-methyl-2'-déoxycytidine

5-OHMedC : 5-hydroxymethylcytosine

ADN : Acide Déoxyribonucléique

ARN : Acide Ribonucléique

BGS : Bisulfite Genomic Sequencing

ChIP : Immunoprécipitation de la Chromatine

DMI : Densely Methylated DNA Island

DNMT : DNA Methyl Transferase

dC : 2'-déoxycytidine

ET : Élément Transposable

HPCE : High Performance Capillary Electrophoresis

HPLC : High Performance Liquid Chromatography

LTR : Long Terminal Repeat

MeDIp : Methyl DNA ImmunoPrecipitation

MSRE : Methyl Sensitive Restriction Enzyme

ORF : Open Reading Frame

PCR : Polymerase Chain Reaction

pb : paire de bases

kb : kilo base

PCR : Polymerase Chain Reaction

RdDM : RNA directed DNA Methylation

INTRODUCTION

Le 25 avril 1953, Watson et Crick publient dans la revue scientifique *Nature* un article intitulé « *Molecular Structure Of Nucleic Acids* » qui bouleversa la génétique. Leurs travaux démontrent pour la première fois que la molécule d'ADN est composée de deux chaînes complémentaires et antisens d'acides nucléiques, structure permettant une copie des chromosomes, support de l'information génétique. Quelques décennies plus tard, en 2009, le génome de Watson lui-même sera séquencé.

1. Les éléments transposables

Quelques années avant la découverte de Watson et Crick, alors que les chromosomes sont définis comme des structures stables, B. McClintock publie le 8 avril 1950 « *The Origin And Behavior Of Mutable Loci In Maize* », article dans lequel elle démontre que des séquences d'ADN composées d'éléments régulateurs sont capables de « transposer », c'est-à-dire de se déplacer d'un endroit à un autre dans un génome. Cette notion de dynamique des génomes associée à des séquences mobiles mit plusieurs années pour être acceptée par la communauté scientifique. Les résultats de B. McClintock mettent en évidence que chez le maïs, les différences de pigmentation que l'on peut observer sur certains grains sont dues à la mobilisation des éléments *Ds* (Dissociateur), sous le contrôle de l'élément *Ac* (Activateur). En effet, le gène responsable de la pigmentation des grains possède une insertion de l'élément *Ds* et lors de sa mobilisation, activée par l'élément *Ac*, ce gène échappe à la répression induite par *Ds*, permettant ainsi la synthèse de pigments. Cette découverte, montrant d'une part que l'ADN n'est pas stable mais dynamique, et d'autre part que les ET influencent la régulation des gènes et donc le phénotype, fut difficile à faire accepter au sein de la communauté scientifique. De fait, B. McClintock fut contraint d'abandonner ses travaux. Ce n'est que quelques années plus tard que l'importance de ses travaux fut enfin reconnue et que B. McClintock se vit décerner, en 1983, le prix Nobel de médecine pour la découverte des ET. Aujourd'hui largement étudiés, les ET possèdent des caractéristiques étonnantes. Hormis chez *Plasmodium falciparum* et quelques bactéries endosymbiotiques, les ET sont présents en proportion variable dans tous les génomes (Tableau 1). Certaines de ces séquences d'ADN détiennent leurs propres séquences régulatrices, leur propre potentiel codant et sont autonomes tandis que d'autres dites déficientes, dépendent au contraire d'autres éléments pour se déplacer : ceci est notamment le cas pour des SINE (Small Intersperced Nuclear Element),

qui utilisent la machinerie moléculaire codée par les LINE (Long Intersperced Nuclear Element) pour transposer. Bien que les ET aient un cycle de répllication indépendant du cycle cellulaire, ils utilisent toutefois la machinerie transcriptionnelle de l'hôte. C'est la raison pour laquelle ils ont longtemps été qualifiés d' « ADN parasite ».

Tableau 1. Proportion en ET et taille des génomes eucaryotes de référence.

Espèce	Taille du génome (Mb)	% ET
<i>Zea mays</i>	2400	80
<i>Arabidopsis thaliana</i>	120	10
<i>Homo sapiens</i>	3000-3200	45
<i>Mus musculus</i>	2500	40
<i>Drosophila simulans</i>	180	2-3
<i>Drosophila melanogaster</i>	180	5-6
<i>Caenorhabditis elegans</i>	97	12
<i>Saccharomyces cerevisiae</i>	12	3-5

Les tailles de génomes ont été collectées dans la base de donnée « Genome Project » du NCBI (<http://www.ncbi.nlm.nih.gov/gate1.inist.fr/>) et les proportions en ET dans les articles de séquençage de référence (*Z. mays* (Schnable et al. 2009), *A. thaliana* (Arabidopsis Genome Initiative, 2000), *C. elegans* (C. elegans Sequencing Consortium, 1998), *H. sapiens* (Lander et al. 2001), *D. melanogaster* et *D. simulans* (Clark et al. 2007), *S. cerevisiae* (Goffeau, et al. 1997), *M. musculus* (Waterston et al. 2002)).

1.1. Classification et transposition des ET

Les ET peuvent être classés selon plusieurs critères. Une première classification proposée par Finnegan *et al* (1989) distingue les ET selon leur intermédiaire de transposition (Finnegan, 1989).

La classe 1 est composée de rétrotransposons à LTR (Long Terminal Repeat) et de rétrotransposons sans LTR, composés essentiellement de deux familles majeures : les LINE et les SINE (Figure 1). Leur mode de transposition fait intervenir un intermédiaire de type ARN, transcrit produit à partir d'une copie d'ET, puis rétrotranscrit en ADN pour former une nouvelle copie. La classe 2 est quant à elle composée de transposons qui ont un mode de transposition faisant intervenir un intermédiaire ADN. Dans ce cas, la transposition consiste en l'excision d'une copie d'ET et son insertion à un autre locus. Ce mode de transposition en « couper-coller » est non réplificatif, puisqu'il consiste en un déplacement de la même copie sur les chromosomes.

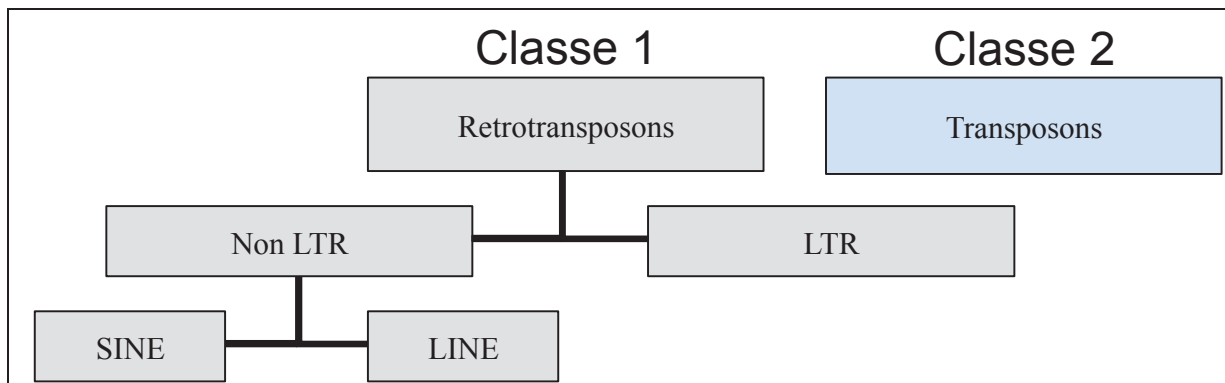


Figure 1. Classification des ET eucaryotes

Cette classification proposée par Finnegan divise les ET en deux familles, selon leur mode de transposition. Les ET de classe 1, qui sont transcrits puis rétrotranscrits pour aboutir à la formation d'une nouvelle copie, ont un mode de transposition en « copier-coller ». Les ET de classe 2, appelés aussi les transposons à ADN, codent pour une transposase qui va leur permettre de s'exciser et de se réinsérer ailleurs dans le génome ; il s'agit de la même copie qui se déplace, selon un mode de transposition en « couper-coller ». Cette classification très simplifiée n'illustre pas la totalité des familles d'ET qui sont répertoriées à ce jour.

Cette classification est la plus utilisée, elle permet de différencier la majeure partie des ET. Cependant, la grande diversité qui existe au sein des ET, mise en évidence durant les dernières années grâce aux quantités abondantes de données obtenues par les divers projets de séquençage, ne permet pas avec ce premier mode de classification de positionner toutes les familles d'ET. Certaines d'entre elles, comme par exemple les Miniature Inverted Tandem Repeat (MITE) sont difficilement caractérisables selon leur simple mode de transposition. En effet, les MITE transposent via un mécanisme de « copier-coller » sans passer par un intermédiaire ARN. Tous les ET connus sont référencés dans la base de données RepBase (<http://www.girinst.org/repbase/index.html>), dans laquelle les développeurs ont établi une nouvelle méthode pour leur classification. En prenant en compte l'enzymologie, les similarités de structure et les relations entre les séquences, les auteurs proposent une classification accompagnée d'une nouvelle nomenclature des ET, mentionnant notamment le nom de la superfamille, un identifiant de structure et un identifiant d'espèce (Kapitonov et Jurka, 2008). Une classification du même type est proposée en 2007 par Wicker et al., et intègre également la structure et la séquence des ET (Wicker et al. 2007). Il s'agit d'une classification unifiée, hiérarchique, qui permet de caractériser tous les ET connus, à partir de la classification établie par Finnegan (1989). Dans ce modèle de classification, les ET sont d'abord divisés en deux classes, selon la présence ou non d'un intermédiaire de transposition de type ARN, puis en sous-classes, ordres et familles. Dans la suite de mon manuscrit,

j'utiliserai la classification de Finnegan, représentée sur la Figure 1 (p.13), qui ne présente aucune ambiguïté concernant la classification des ET étudiés durant ma thèse.

1.2. Relations avec le génome

La transposition ou mobilisation d'un ET peut avoir des conséquences délétères pour un génome si il s'insère dans ou près d'un gène important. Le fait que les ET puissent être présents en grand nombre de copies dans un génome augmente aussi la probabilité de recombinaison pouvant ainsi engendrer des mutations, et altérer l'activité des gènes voisins (Kazazian, 2004). Chez l'homme, beaucoup de maladies liées aux ET ont été décrites : elles font principalement intervenir les ET de classe 1, qui sont les plus répandus dans ce génome. Malgré tout, il n'est pas toujours possible de conclure si l'insertion de l'ET est la cause directe de la maladie. Dans certains cas cependant, des insertions d'ET ont pu être identifiées comme déclencheurs de maladies. Par exemple, les gènes BRCA1 et BRCA2 sont tous deux impliqués dans des prédispositions au cancer du sein. En effet, une mutation de l'un de ces gènes est un facteur favorisant l'apparition de cancer chez le sujet porteur. Un cas d'insertion d'un élément *Alu* a été identifié dans l'exon 22 du gène BRCA2 d'une patiente atteinte d'un cancer du sein (Miki et al. 1996). La présence de cette insertion entraîne un décalage du cadre de lecture du gène, provoquant d'une part l'épissage de l'exon 22 et d'autre part l'apparition d'un codon stop prématuré, conduisant ainsi à la production d'une protéine tronquée.

Malgré ce caractère nuisible des ET, il n'en demeure pas moins que leur implication dans l'évolution des génomes est de plus en plus démontrée par la communauté scientifique. La vision « contemporaine » que nous avons des génomes, plus dynamique avec une « plasticité » certaine, confère aux ET un rôle central au cours de l'évolution qui est actuellement peu contesté. Leur fort pouvoir mutagène, associé au fait que ce soient des séquences multicopies, donc favorables aux recombinaisons, fait des ET des acteurs majeurs de la dynamique des génomes. De l'utilisation de leurs séquences régulatrices jusqu'à leur domestication, de nombreux exemples montrent que certaines insertions d'ET ont pu être sélectionnées par le génome hôte. Chez les vertébrés à mâchoire, les gènes RAG1 et RAG2 (Recombination Activating Genes) sont impliqués dans la recombinaison des segments V, D et J, servant à créer de la diversité d'immunoglobulines (Ig) et de TCR (T Cell Receptor) nécessaire à la réponse immunitaire spécifique à un antigène (Volff, 2006). Les clusters de gènes d'Ig et TCR sont composés des segments V, D et J flanqués de séquences inversées

appelées Recombination Signal Sequence (RSS), ce qui leur confère une structure de type Insertion Sequence (IS) procaryotique. Le complexe RAG1/RAG2 agit sur ces séquences RSS via RAG1, possédant l'activité recombinase pour induire la recombinaison V(D)J avec un mode d'action en couper-coller semblable à celui d'une transposase. L'analyse du gène RAG1 a montré qu'il possède une forte similarité de séquence au niveau de sa région Core avec les éléments de type transib, présents chez les insectes et nématodes et dont le mode de transposition est de type couper-coller (Fugmann, 2010). Une composante principale du système immunitaire des mammifères et plus largement des vertébrés à mâchoire proviendrait donc de la domestication d'ET (Feschotte, 2008). Chez l'Homme, le gène de la syncytine-1, impliqué dans la formation du placenta, plus précisément dans la fusion du trophoblaste, provient de la domestication du gène *env* d'un rétrovirus endogène de la famille *HERV-W* (Human Endogenous RetroVirus-W) (pour revue, Stoye, 2009). Chez la levure *Saccharomyces cerevisiae*, la famille de rétrotransposons à LTR *Ty* a largement contribué à l'évolution et à l'adaptation des levures (Lesage et Todeschini, 2005). Dans de nombreux stress, le rétrotransposon *Ty1* est réactivé transcriptionnellement avec arrêt des mécanismes de contrôle post-transcriptionnels, permettant sa transposition. En cas d'absence de nitrogène dans le milieu de croissance des levures (indispensable à leur métabolisme), les cellules vont former des filaments, composés de cellules juxtaposées, et vont ainsi pouvoir migrer dans le milieu de culture. Cette modification de la formation des colonies de levures leur permet d'envahir le milieu à la recherche de nitrogène (Lesage et Todeschini, 2005). Chez *Drosophila melanogaster*, les extrémités des chromosomes ne sont pas « allongées » après chaque mitose par une télomérase comme c'est le cas pour de nombreuses autres espèces, mais par l'insertion des rétrotransposons sans LTR, *Het-A* et *TART*. En effet, l'intégrité des télomères est préservée par la transposition successive de ces deux éléments, qui se fait surtout par la collaboration de leur protéine gag (Pardue et al. 2005). Chez les bactéries, les ET, ou Séquences d'Insertion (IS), peuvent intervenir dans l'échange de gènes de résistance aux antibiotiques entre bactéries. En effet, les IS peuvent être présents dans les bactériophages, dans les chromosomes ainsi que dans les plasmides, séquences d'ADN circulaires extrachromosomiques faisant partie du génome bactérien. Les plasmides interviennent dans le phénomène de conjugaison, qui n'est autre qu'un échange de matériel génétique entre deux bactéries, appelé transfert horizontal. Deux IS peuvent former un transposon composite, élément mobile comportant un gène qui, dans le cas de *Tn10*, est un gène de résistance à la tétracycline (Haniford, 2006). Le transposon *Tn7*, autre transposon composite, possède deux modes de transposition particulièrement astucieux. Un mode de

transposition vertical, au cours duquel des séquences spécifiques du chromosome bactérien sont la cible de la copie excisée, et un mode de transposition horizontal, au cours duquel les plasmides et les bactériophages seront à leur tour ciblés spécifiquement. Ces deux modes, qui font intervenir des partenaires moléculaires différents, ont des rôles essentiels dans le maintien du transposon *Tn7* dans les lignées de bactéries et dans sa propagation horizontale. Par ailleurs, comme dans le cas de *Tn10*, il a été montré que *Tn7* était impliqué dans le caractère multirésistant de certaines souches bactériennes comme *Acinetobacter baumannii*, qui portent plusieurs gènes de résistance à diverses drogues (Parks et Peters, 2009). La capacité de certains ET de la famille *Tn7* à transférer horizontalement des gènes de résistance permet aux bactéries de s'adapter et de survivre dans différents environnements. La relation entre les ET et le génome peut être qualifiée de relation hôte/parasite, mais peut aussi être considérée comme une interaction entre un écosystème (le génome) composé de niches écologiques dans lesquelles un ensemble d'espèces (ET, gènes etc.) interagissent (Le Rouzic et al. 2007; Venner et al. 2009).

2. La régulation des ET

Bien que les ET aient réussi à envahir certains génomes, ils sont soumis à des mécanismes de régulation de leur activité. Parmi ces mécanismes, les régulations transcriptionnelles et post-transcriptionnelles conduisent à la diminution de la transposition et, la recombinaison homologe entre deux copies du même ET conduit à une diminution de son nombre de copies. Ces dernières années, de nombreuses études ont montré l'implication des régulations épigénétiques dans le contrôle des ET, et ce, dans de nombreux organismes comme l'Homme, la plante, la drosophile, ou encore la levure.

2.1. Les régulations épigénétiques

Depuis les années 2000, de nombreux génomes ont été séquencés. Malgré la quantité abondante et précieuse d'informations que cela a apporté au monde scientifique, il a bien fallu se rendre à l'évidence : le code génétique, plus petit dénominateur commun de l'information génétique, ne permet pas à lui seul de comprendre le vivant. Quel déterminisme conduit une cellule souche embryonnaire à devenir potentiellement tout type cellulaire composant un organisme ? Comment expliquer que certains caractères phénotypiques « acquis » par un individu peuvent être hérités ? Ces questions trouvent des réponses depuis la découverte des régulations épigénétiques.

L'épigénétique, telle que définie par Riggs, Martienssen et Russo en 1996 dans la préface du livre « Epigenetic Mechanisms of Gene Regulation » (Edition Cold Spring Harbor Laboratory Press), est l'étude des changements héréditaires post-mitotiques ou post-méiotiques qui ne peuvent pas s'expliquer par des modifications dans la séquence de l'ADN. La chromatine, double hélice d'ADN enroulée autour de protéines de structure, telles que les histones, définit un degré de compaction des chromosomes dans le noyau des cellules. L'accessibilité de l'ADN aux facteurs de transcription dépend alors du niveau plus ou moins important de compaction de la chromatine. Les modifications post-traductionnelles des protéines d'histones, la méthylation des cytosines de l'ADN et les petits ARN interviennent dans l'établissement et le maintien de l'état chromatinien. Associés, ces trois mécanismes peuvent définir des zones euchromatiques, peu denses, associées à l'activation de la transcription, et des zones hétérochromatiques, denses, associées à la répression de la

transcription. Cette vision est toutefois simpliste, et plusieurs travaux démontrent que la dynamique associée à l'état chromatinien est en réalité plus complexe.

2.1.1. Les modifications post-traductionnelles des histones

Les histones sont des protéines très conservées au cours de l'évolution des eucaryotes. Elles sont organisées en nucléosome, structure définie comme un octamère autour duquel s'enroule l'ADN dans les noyaux des cellules eucaryotes. La succession des nucléosomes le long de la molécule d'ADN lui confère une structure en « collier de perle », premier niveau de compaction de la chromatine. Un octamère d'histone est un complexe protéique composé de doublets de molécules d'histone H2A, H2B, H3 et H4, par-dessus lesquelles l'histone H1 vient se fixer après enroulement de l'ADN. Les histones peuvent subir des modifications post-traductionnelles sur certains des acides aminés de leurs queues N et C terminales (Figure 2). Ces « marques d'histones », constituant un « code des histones », interviennent dans le degré de compaction de la chromatine (Strahl et Allis, 2000) et définissent des zones actives ou répressives de l'activité transcriptionnelle des gènes concernés. Ces modifications post traductionnelles des histones de type acétylation, méthylation, phosphorylation, sumoylation, ou encore ubiquitylation (Berger, 2007) sont réversibles. Les modifications les plus décrites dans la littérature concernent les lysines (K) de l'histone H3, qui peuvent être acétylées ou méthylées (mono, di ou triméthylées). D'autres modifications ont également été identifiées sur les parties N terminales des autres histones composant le nucléosome, comme par exemple la marque H4K20me3, dont le rôle dans la structure de l'hétérochromatine constitutive est largement conservé (Schotta et al. 2004). Certaines marques sont mutuellement exclusives, comme la triméthylation de la lysine 9 de l'histone H3, associée à une chromatine compacte et répressive, comme l'acétylation de ce même acide aminé, notée H3K9Ac, associée à une chromatine « ouverte » et active (Jenuwein et Allis, 2001). La méthylation n'est cependant pas toujours associée à une chromatine répressive puisque la marque H3K4me3 est associée aux régions 5' de gènes exprimés (Azuara et al. 2006). Ainsi, certaines marques d'histones sont directement associées à l'euchromatine.

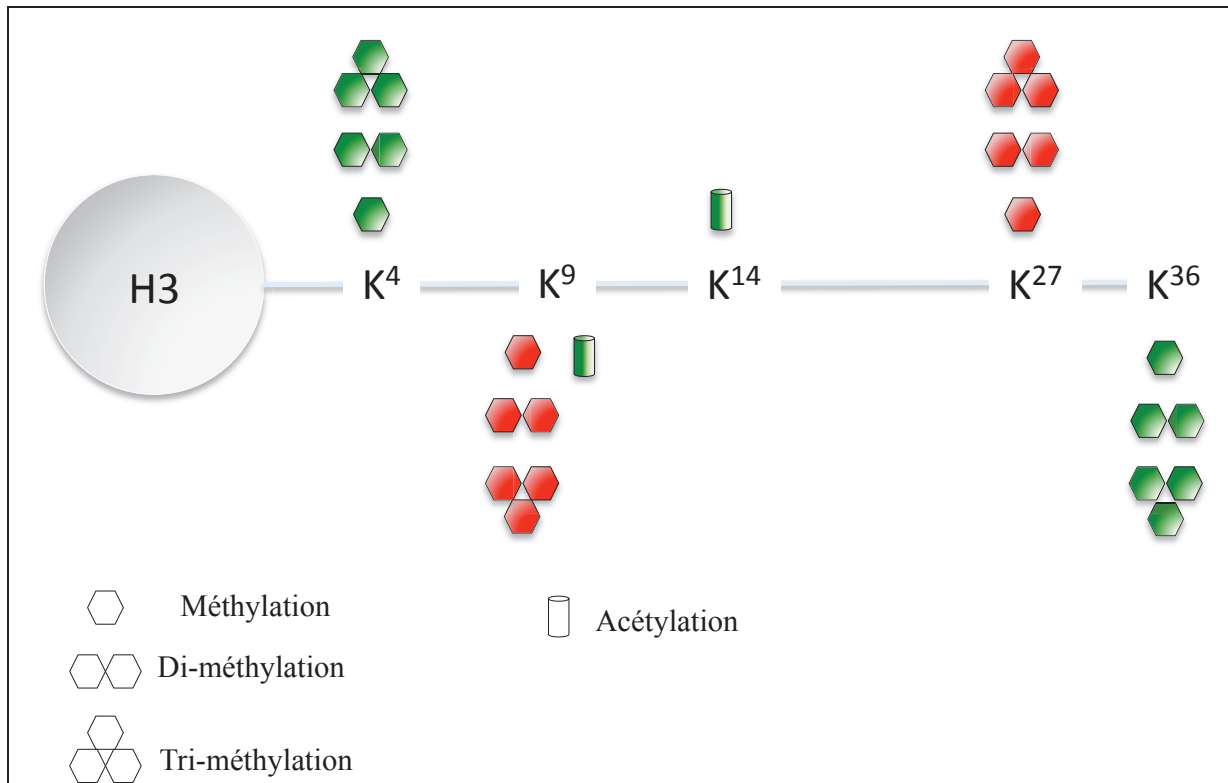


Figure 2: Principales modifications post-traductionnelles des lysines de l'histone H3.

Les lysines situées sur la queue de l'histone H3 (partie N terminale) peuvent subir une, deux ou trois méthylations et certaines d'entre elles peuvent être acétylées. Ces deux marques sont mutuellement exclusives et définissent des états chromatiniens différents. La méthylation (mono, di, ou tri) en position K4 et K36 et l'acétylation des positions K9 et K14 définissent un état de chromatine ouvert (vert) favorable à l'expression des gènes, tandis qu'elle est associée à un état répressif (rouge) pour les positions K9 et K27. D'après Ebert *et al* (2000).

2.1.2. La méthylation de l'ADN

La méthylation de l'ADN est une modification chimique des carbones 5 des cytosines de l'ADN. Elle est impliquée dans la régulation des gènes, dans la structure des génomes et dans la régulation des séquences répétées. Lorsqu'elle est localisée au niveau de la région promotrice des gènes, la méthylation des cytosines est associée à la répression de leur expression (Suzuki et Bird, 2008).

La méthylation des cytosines est catalysée par des enzymes appelées « DNA Methyl Transferases » (DNMT). Certaines de ces enzymes, comme les DNMT3a/b, dites *de novo*, sont responsables de la méthylation, tandis que d'autres, comme DNMT1, dites de maintien, ont un rôle de maintien des profils de méthylation. Le maintien s'effectue après la mitose, au moment où les molécules d'ADN sont des substrats hémiméthylés (Reik et al. 2001). Selon

les espèces, les DNMT peuvent méthyler les cytosines dans différents contextes nucléotidiques. Les cytosines méthylées peuvent impliquer des dinucléotides palindromiques CpG (ou CpNpG, où N = A, C, T, G) dits symétriques, ou des dinucléotides CpH (ou CpHpH ou H = A, C ou T) dits asymétriques. Pour les dinucléotides symétriques, une cytosine est portée sur chacun des deux brins complémentaires de la molécule d'ADN, tandis que pour les dinucléotides asymétriques un seul brin est porteur d'une cytosine. Après chaque division cellulaire, les DNMT de maintien reproduisent de façon symétrique, et à partir du brin matrice, le profil de méthylation sur le brin néosynthétisé, permettant ainsi un maintien des profils de méthylation au cours des divisions cellulaires. A l'inverse, dans le cas de dinucléotide asymétrique méthylé, seul l'un des deux brins est porteur d'une cytosine méthylée. Enfin, les cytosines méthylées sont fortement mutagènes et peuvent être spontanément converties en Thymine par déamination (perte du groupement amine NH₂) (Holliday et Grigg, 1993). Chez la plante *Arabidopsis thaliana*, la méthylation des cytosines est impliquée dans la régulation des gènes ainsi que dans le développement et le contrôle des ET. Ce sont les cytosines méthylées en contexte asymétrique qui sont fortement présentes au niveau des ET (Chan et al. 2005; Zhang et al. 2006). Ces profils de méthylation sont établis et maintenus par des DNMT spécifiques guidées par des petits ARN ciblant par homologie les séquences dont les cytosines sont à méthyler (Matzke et al. 2007). Ce type de méthylation ciblé par des petits ARN, appelé « RNA dependant DNA methylation » (RdDM), a également été mis en évidence chez la souris (Miyagawa et al. 2008).

Chez les mammifères, et plus particulièrement chez l'Homme, la méthylation des cytosines est indispensable au développement et à la survie des embryons par son implication dans des processus de régulation majeure tels que la régulation des gènes, l'inactivation du chromosome X et les empreintes génétiques (Morgan et al. 2005). Dès la fécondation et jusqu'au stade blastula, le génome du nouvel individu subit une déméthylation globale, à l'exception des régions centromériques et péri-centromériques, des rétrotransposons et des gènes à empreintes. Ces gènes qui portent des empreintes épigénétiques différentes (au niveau de la méthylation de l'ADN), établies dans la lignée germinale des parents, auront donc une expression différentielle tout au long de la vie du nouvel individu, selon qu'ils aient été hérités du père ou de la mère. Plus tard dans le développement, une mise à jour des empreintes, ou reprogrammation, aura lieu dans les cellules primordiales germinales du nouvel individu, de sorte que les gènes soumis aux empreintes des cellules de la lignée germinale portent des empreintes correspondant au sexe de l'individu qui les produit (Morgan et al.

2005). Des défauts dans la mise en place des empreintes peuvent avoir des conséquences graves sur le développement d'un individu. Chez l'Homme, un exemple très bien détaillé concerne la région d'ADN localisée en position 15q11-q13 (bandes 11 à 13 du bras long du chromosome 15), qui, si elle est soumise à un défaut d'empreinte, sera responsable des syndromes de Prader Willy ou d'Angelman. Le syndrome de Prader Willy est une pathologie associée à un défaut d'empreinte du chromosome 15 hérité du père, tandis que le syndrome d'Angelman est la pathologie associée à un défaut d'empreinte du chromosome 15 hérité de la mère (Cassidy et al. 2000). Des organismes comme *C. elegans* ou *S. cerevisiae* n'ont pas de méthylation de l'ADN, et aucune DNMT n'a pu être identifiée dans leurs génomes (Ponger et Li, 2005). L'absence de rôle fonctionnel de la méthylation de l'ADN dans ces organismes modèles laisse penser que cette modification épigénétique est principalement spécifique aux mammifères et aux plantes. Toutefois, la découverte d'un système complet de méthylation chez l'abeille *Apis mellifera* a permis de reconsidérer ce point de vue. En effet, l'abeille possède un système de méthylation fonctionnel dirigé sur les CpG, semblable à celui des mammifères (Wang et al. 2006). L'analyse des ratios CpG observés/attendus au niveau des gènes de l'abeille, suit une distribution bimodale, caractéristique de la présence de méthylation au niveau des dinucléotides CpG dans un génome. Les cytosines méthylées sont fortement mutagènes et leur déamination conduit à une déplétion en CpG pour les gènes concernés. La méthylation de l'ADN chez l'abeille est impliquée dans la régulation des gènes pendant le développement. Les auteurs proposent comme hypothèse que la méthylation de l'ADN est à la base de la différenciation entre les classes sociales chez l'abeille (Elango et al. 2009). Les gènes DNMT1, DNMT2 et DNMT3 ont également été identifiés chez l'insecte parasitoïde *Nasonia vitripennis* (Werren et al. 2010) et la méthylation des cytosines a été détectée chez de nombreux insectes.

2.1.3. Les petits ARN

Depuis leur découverte dans les années 1990, les mécanismes de régulation faisant intervenir des petits ARN ont révélé leur importance chez de nombreuses espèces. Ces petits ARN, de 20 à 29 nucléotides agissent en séquestrant ou en dégradant les ARN messagers, permettant une régulation de l'expression des gènes au niveau post-transcriptionnel (Siomi et Siomi, 2009). Impliqués dans des processus aussi variés que la mise en place de la lignée germinale et le contrôle des séquences répétées de type ET chez les mammifères, les drosophiles et les plantes, les petits ARN interviennent aussi pour cibler des modifications

épigénétiques sur certaines séquences répétées. Les premiers petits ARN découverts furent les « small interfering RNA » (siRNA) et de nouvelles familles sont régulièrement découvertes. Ainsi en plus des siRNA, les « micro RNA » (miRNA), les « PIWI-interacting RNA » (piRNA) ou encore les « endogenous-siRNA » (endo-siRNA) sont impliqués dans différentes voies de régulation. Ces petits ARN diffèrent par leur taille, par l'enzyme qui les clive et par la protéine qui les prend en charge. Les « piRNA » de 24 à 29 nucléotides, aussi appelés « repeat associated small interfering RNA » (rasiRNA), sont produits par des séquences répétées de type ET et ont été mis en évidence récemment dans la lignée germinale de la drosophile. L'implication de ces rasiRNA dans la régulation des ET sera détaillée ci-après.

2.2. Variabilité des régulations épigénétiques

Les mécanismes épigénétiques de régulation des gènes chez l'Homme et les plantes, et d'empreintes génomiques chez les mammifères sont bien décrits et nous avons vu précédemment les conséquences graves que peuvent avoir des défauts dans la mise en place ou le maintien des profils de méthylation. Cependant, quelques exemples montrent que les régulations épigénétiques et plus précisément la méthylation de l'ADN sont sensibles à l'environnement. Un exemple frappant concerne les jumeaux monozygotiques, par définition génétiquement identiques car issus de la division d'un zygote en deux cellules filles, qui peuvent dans certains cas avoir des phénotypes discordants et développer des maladies génétiques différentes. Chez l'Homme, l'étude des modifications épigénétiques a permis de montrer que des jumeaux monozygotiques adultes présentent des différences significatives au niveau des profils de méthylation des cytosines et d'acétylation des histones (Fraga et al. 2005). Au contraire, les jumeaux plus jeunes présentent relativement peu de différence de ces profils. Cela peut s'expliquer par le fait que les sujets jeunes vivent souvent dans le même environnement, tandis que les sujets adultes vivent plus éloignés l'un de l'autre. L'étude des écotypes Columbia (Col) et Landsberg (Ler) a permis de mettre en évidence chez la plante *Arabidopsis thaliana* des différences importantes dans les profils de méthylation de la molécule d'ADN sur laquelle les 5 méthylcytosines sont essentiellement localisées à l'intérieur des gènes sans entraîner de diminution de leur expression (Vaughn et al. 2007). En revanche, très peu de polymorphismes dans les profils de méthylation associés aux ET furent révélés entre ces deux écotypes, soulevant donc le rôle central que la méthylation de l'ADN possède dans le contrôle des ET chez cette plante. Des travaux menés chez l'abeille ont montré que l'alimentation est impliquée dans la différenciation entre la reine qui est fertile, et

les travailleuses qui sont stériles. En effet, la gelée royale est une substance spécialement synthétisée par les nourrices pour nourrir les larves destinées à devenir reine. La majorité des larves mutées pour DNMT3 par injection de siRNA donnent des reines, ce qui implique que la gelée royale contient des substances agissant sur l'établissement des profils de méthylation chez l'abeille (Kucharski et al. 2008). La relation entre la régulation des ET, la méthylation de l'ADN et les conditions de l'environnement suggère alors une implication en synergie de ces trois composantes dans l'évolution des espèces et la spéciation. Un relâchement de la répression des ET lors de changements des conditions de l'environnement peut fournir au génome hôte un avantage adaptatif certain (Rebollo et al. 2010).

3. Les populations naturelles de drosophiles pour comprendre la dynamique des ET

3.1. Variabilité en ET dans les populations naturelles de drosophiles

Les espèces *D. melanogaster* et *D. simulans* sont deux espèces très proches phylogénétiquement, dont la divergence est estimée à 2,5 millions d'années (par comparaison, la divergence entre l'Homme et le chimpanzé est estimée à 6 millions d'années (Goodman et al. 1998)). Malgré cette proximité évolutive, ces deux espèces ont, dans leurs génomes, des proportions très différentes de séquences répétées, de l'ordre de 5 % pour *D. simulans* et de 14 % pour *D. melanogaster* (Dowsett et Young, 1982). Le récent séquençage de ces deux génomes fait état de 2,73 % chez *D. simulans* et de 5,35 % chez *D. melanogaster* (Clark et al. 2007). Ces deux analyses reposent sur des techniques différentes. La première quantification fut réalisée en mesurant les cinétiques de dissociation-réassociation des séquences d'ADN chez ces deux espèces. La deuxième quantification fut quant à elle calculée à partir des parties séquencées des génomes de ces deux espèces, principalement de l'euchromatine, l'hétérochromatine étant difficilement séquençable et annotable. La proportion réelle en ET de ces deux génomes se situe probablement entre ces deux estimations puisque d'une part, la première méthode quantifie les séquences répétées au sens large, et donc pas exclusivement les ET et d'autre part, la deuxième méthode ne prend pas en compte l'hétérochromatine, particulièrement dense en ET. Bien que la valeur absolue de la quantité en ET des deux espèces puisse être discutée, la différence entre les deux espèces se confirme selon les estimations précitées. Ces deux espèces invasives originaires des régions est-africaines ont une histoire écologique différente et la colonisation du monde par *D. melanogaster* est plus ancienne que celle de *D. simulans* (Lachaise et Silvain, 2004). La colonisation d'un nouvel environnement est souvent associée à une augmentation de la transposition, comme le rapportent les études du nombre de copies qui est plus faible dans les populations africaines que celui observé dans les populations du reste du monde (Vieira et al. 1999). Le nombre de copie d'ET dans les populations de drosophiles serait maintenu au niveau de la lignée germinale par un équilibre entre la transposition et la sélection naturelle, comme le proposent différents modèles, et une régulation plus fine au niveau de chaque individu est apportée par les mécanismes épigénétiques (Lee et Langley, 2010). Par ailleurs, il ne semble pas y avoir de lien entre le nombre de copies des ET et leur taux de transposition chez ces deux espèces (Pasyukova et al. 1998; Vieira et al. 1997). L'ET de classe 2 *mariner* présente une différence

d'expression somatique entre les populations de *D. simulans* : l'expression est plus grande dans les jeunes populations invasives que dans les populations installées depuis plus longtemps (principalement africaines) (Picot et al. 2008). L'étude des populations naturelles démontre qu'il existe une grande variabilité du nombre de copies d'ET entre les populations de ces deux espèces (Vieira et Biéumont, 2004). Ces nombres de copies ont été déterminés pour plusieurs familles d'ET par hybridation *in situ* sur les chromosomes polytènes des glandes salivaires de larves de drosophiles. Cette technique, qui utilise une sonde spécifique à chaque ET, permet de compter les copies euchromatiques situées sur les bras des chromosomes. Les insertions d'ET présentes au niveau du chromocentre, ensemble fortement hétérochromatique composé des centromères des chromosomes polytènes, ne peuvent pas être dénombrées par cette technique. Ainsi, bien que plusieurs populations naturelles de *D. simulans* n'aient pas de copies euchromatiques de certains ET, toutes possèdent néanmoins des copies au niveau des chromocentres. Parmi les ET présentant un polymorphisme d'insertion important, le rétrotransposon *tirant* présente une variabilité du nombre de copies euchromatiques particulière chez *D. simulans*, puisqu'hormis deux populations est-africaines, aucune autre population n'en possède au niveau euchromatique.

3.2. Epigénétique chez la drosophile

3.2.1. Méthylation de l'ADN

Le taux de méthylation des cytosines chez *D. melanogaster* est compris entre 0,4 % et 1 % (Kunert et al. 2003; Lyko et al. 2000). Parmi les gènes connus codant des enzymes ayant une activité « DNA méthyltransferase », tous ont été perdus au cours de l'évolution des drosophiles à l'exception de DNMT2 (Ponger et Li, 2005). Aucune activité de type DNMT n'a pu être attribuée *in vitro* à dDNMT2 et sa fonction n'est connue que depuis quelques années. De nombreuses et récentes études montrent en effet que DNMT2 est une « RNA méthyltransferase » capable de transférer un groupement méthyl sur les cytosines des ARN de transfert (ARNt). Cette activité de méthylation des cytosines des ARN de transfert joue certainement un rôle biologique important car elle est conservée de l'Homme à la drosophile en passant par la levure *S. pombe* (Schaefer et Lyko, 2010; Goll et al. 2006). L'expression du gène dDNMT2, bien que faible, est maximale durant les stades larvaires du développement de la drosophile mais nulle au stade adulte (Lyko et al. 2000). Les mutations de DNMT2 sont viables et ne provoquent aucun phénotype, tandis que la surexpression de ce gène conduit à une augmentation de la longévité des mouches (Lin et al. 2005). Ceci conduit à penser que,

chez la drosophile, la méthylation de l'ADN n'intervient pas dans la régulation des gènes et des ET comme cela a été mis en évidence dans d'autres espèces. Cependant, l'utilisation d'enzymes de restriction et de traitement au bisulfite ont permis de mettre en avant la présence de sites méthylés chez *D. willistoni* (Garcia et al. 2007) et des séquences fortement méthylées au niveau de la région régulatrice du rétrotransposon à LTR *invader 4* durant les stades embryonnaires du développement de *D. melanogaster* (Phalke et al. 2009). Dans ce dernier exemple, la méthylation d'*invader 4* durant les stades précoces permettrait de recruter l'histone méthyl-transférase SUV4-20, pour ensuite triméthyliser H4K20. Un défaut des structures télomériques, associé à la transposition de l'élément *invader 4*, a été de plus constaté dans des mutants dDNMT2.

Enfin, il a récemment été montré par enzyme de restriction l'existence de profils de méthylation sexe-spécifique chez certaines espèces de drosophiles du sous-genre *Sophophora* (D'Ávila et al. 2010), suggérant ainsi une analogie avec les empreintes génétiques impliquées dans la régulation des gènes chez les mammifères. L'ensemble de ces résultats est contradictoire mais nous ne pouvons cependant pas exclure que la méthylation de l'ADN chez la drosophile soit impliquée dans la régulation des ET.

3.2.2. Modification des histones

Chez la drosophile, la méthylation de H3K9 est impliquée dans la régulation des gènes durant le développement, dans la structure des chromosomes et dans le contrôle des ET (Peng et Karpen, 2007). Le phénomène de « Position Effect Variegation » (PEV), mis en évidence chez la drosophile, est caractérisé par un événement de recombinaison ayant pour résultat de juxtaposer des régions euchromatiques à des régions d'hétérochromatine constitutive (Girton et Johansen 2008). Les gènes présents dans ces régions euchromatiques vont alors être hétérochromatinisés et devenir transcriptionnellement inactifs. Ces régions sont définies comme de l'hétérochromatine facultative. L'analyse des composants moléculaires intervenant dans le PEV a permis d'identifier de nombreuses protéines impliquées dans la formation de l'hétérochromatine, baptisées « Suppresseurs de Variegation » (SU(VAR)). Les SU(VAR) sont pour la plupart des enzymes assurant les modifications post-traductionnelle des histones. L'histone méthyltransférase SU(VAR)3-9, qui méthyle l'histone H3 sur sa lysine 9 (H3K9), a un rôle central dans l'établissement de l'hétérochromatine des chromocentres et du chromosome 4 de la drosophile (Schotta et al. 2002). De plus, il semblerait que la fonction de

cette enzyme soit extrêmement bien conservée puisqu'il est possible de restaurer un phénotype sauvage chez un mutant auquel l'enzyme SU(VAR)3-9 humaine est transfectée.

3.2.3. Petits ARN

Quelques années auparavant, la découverte, chez la drosophile, des piRNA, initialement connus sous le nom de « repeat associated small interfering RNA » (rasiRNA) (Vagin et al. 2006), bouleversa nos connaissances en terme de régulation des ET. Les piRNA sont des éléments clés d'un système de régulation des ET à deux composantes : une composante génétique codée par des locus hétérochromatiques contenant des copies défectives d'ET d'une part, et une composante adaptative correspondant aux piRNA produits par les copies fonctionnelles des ET localisées dans l'euchromatine d'autre part. Les piRNA sont des petits ARN non codants, antisens, de 25 à 29 nucléotides, produits par des clusters localisés dans les régions péri-centromériques et télomériques des chromosomes, composés de séquences d'ET dégénérées (Saito et al. 2006). La sous-classe PIWI fait partie de la famille des protéines Argonaute, et comprend les protéines PIWI (« P-element Induced Wimpy testis »), Argonaute 3 (AGO3) et AUBergine (AUB). Les piRNA sont pris en charge par les protéines PIWI ou AUB, afin de cibler et dégrader les transcrits sens des ET pour lesquels ils sont complémentaires. Le clivage des transcrits sens des ET par les piRNA produit des petits ARN sens qui seront alors pris en charge par la protéine AGO3. Ce nouveau complexe ARN/AGO3 ciblera à son tour les transcrits antisens. Un modèle d'amplification de l'immunité induite par ces piRNA, appelé « ping-pong », est proposé par Brennecke et al (2007). De récentes études montrent que les protéines PIWI et AUB interviennent respectivement dans les régulations somatique et germinale des ET (Malone et al. 2009). Chez la drosophile, la dysgénésie des hybrides est un phénotype causé par la transposition massive de l'élément I ou de l'élément P. Le croisement de mâles inducteurs I (présence de copies actives de *I*) et de femelles réactives R (absence de copies actives de *I*) conduit à une stérilité de la descendance. Le croisement inverse impliquant des mâles R et des femelles I conduit à une descendance fertile. Chez les individus I, l'élément *I* est régulé et son expression est faible. Il a été montré que les facteurs cytoplasmiques transmis maternellement et permettant de réguler la transposition de l'élément *I*, correspondent à des piRNA spécifiques de la lignée germinale (Chambeyron et al. 2008; Brennecke et al. 2008). En effet, l'analyse de mutants spécifiques de la voie de régulation impliquant des piRNA AGO3-dépendant a montré une réactivation transcriptionnelle de l'élément *I* dans les souches

I. Les piRNA dirigés contre l'élément *I* et déposés maternellement assurent la régulation de cet ET et empêchent l'apparition du phénotype dysgénésique dans la descendance lors d'un croisement mâle R et femelle I. Un autre élément de la drosophile, *gypsy*, un rétrovirus endogène, permet d'illustrer la voie de régulation somatique impliquant des piRNA. Le locus flamenco (ou locus COM) (Desset et al. 2008), présent chez *D. melanogaster* et conservé chez *D. erecta* et *D. yakuba*, est un locus hétérochromatique composé d'ET dégénérés producteurs de piRNA antisens (Malone et al. 2009). Ces derniers, pris en charge par PIWI, sont impliqués dans la régulation somatique de *gypsy* dans les cellules folliculaires. La levée de la répression de *gypsy* a été montrée par l'utilisation de mutants de la voie PIWI somatique-dépendant ainsi que par des mutants du locus flamenco. Les auteurs proposent que ces deux systèmes distincts permettent de réguler les ET actifs au niveau somatique et germinale chez la drosophile.

3.3. Le Modèle *tirant*

D'une manière générale, pour un élément, le nombre de copies est plus élevé dans les populations naturelles de *D. melanogaster* présentant une variabilité inter-populationnelle de ce nombre faible, que chez les populations naturelles de l'espèce *D. simulans*, pour laquelle la variabilité inter-populationnelle est plus élevée (Vieira et al. 1999). L'ET *tirant* est un rétrotransposon à LTR de type *gypsy* et spécifique au sous-groupe *melanogaster* (Marsano et al. 2000). Sa structure canonique établie à partir de séquences de *D. melanogaster* est présentée sur la Figure 3 (Cañizares et al. 2000). Avec une structure caractéristique des rétrovirus endogènes, *tirant* est flanqué par deux LTR et possède les cadres ouverts de lecture *gag*, *pol* et *env* codant les enzymes indispensables à sa rétrotransposition. Dans les populations naturelles de *D. melanogaster* et *D. simulans*, cet élément possède une distribution du nombre de copies étonnante. D'une part, il a été mis en évidence par Southern blot que toutes les populations de *D. melanogaster* et *D. simulans* possèdent des copies de *tirant*, et que toutes les populations testées à l'exception de la population du Zimbabwe (*D. simulans*) contiennent au moins une copie complète, donc potentiellement active (Fablet et al. 2006). D'autre part, l'hybridation *in situ* sur chromosomes polytènes a montré que toutes les populations de *D. melanogaster* possèdent environ une dizaine de copies de *tirant*, tandis que chez *D. simulans*, la plupart des populations ne possèdent pas de copies de *tirant* sur les bras des chromosomes, à l'exception de deux populations de l'est africain, *Zimbabwe* et *Makindu* (Kenya) (Vieira et al. 1999).

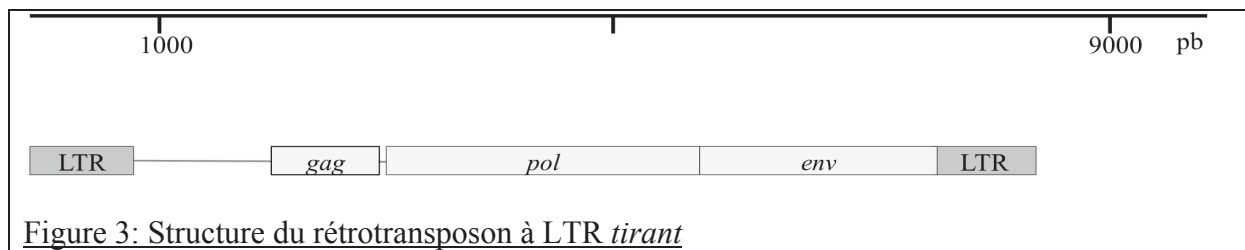


Figure 3: Structure du rétrotransposon à LTR *tirant*

Dans la suite de mon manuscrit, un intérêt particulier sera porté à quatre populations dont les caractéristiques sont les suivantes : la population de *D. melanogaster*, provenant du Sénégal, qui possède neuf copies euchromatiques de *tirant* ; les populations de *D. simulans* de Chicharo (Portugal), Zimbabwe et Makindu (Kenya), qui possèdent respectivement zéro, deux et cinq copies euchromatiques de *tirant* (Figure 4). L'analyse de l'expression de *tirant* a été déterminée par RT-PCR à partir d'ARN totaux, extraits à partir de femelles entières, d'ovaires et de testicules de ces quatre populations. Au niveau somatique, *tirant* est exprimé dans toutes les populations, tandis que dans la lignée germinale, *tirant* n'est exprimé que dans les ovaires et les testicules de la population de Makindu (*D. simulans*) (Fablet et al. 2006). Deux variants de *tirant* ont été identifiés selon des différences dans la séquence nucléotidique de leurs régions régulatrices, composée de la LTR 5' et de l'UTR 5'. Un premier variant présent chez les deux espèces a été baptisé type C ; un second, présent uniquement dans les populations naturelles de *D. simulans*, a été quant à lui baptisé type S. Des mesures d'activité promotrice des variants de type C et S avec un gène rapporteur en cultures cellulaires ont montré que tous deux sont capables *in vitro* d'initier la transcription (Fablet et al. 2009). En revanche, le séquençage des produits de RT-PCR a montré que seul le variant de type C est transcrit *in vivo* (Fablet et al. 2006).



Figure 4: Carte des zones d'échantillonnage des souches étudiées

La population échantillonnée au Sénégal appartient à l'espèce *D. melanogaster*, les populations de Chicharo, Zimbabwe et Makindu appartiennent à l'espèce *D. simulans*. Le nombre de copies euchromatiques de chaque population est indiqué.

Les sites d'insertion des copies de *tirant* dans les populations de Chicharo, Makindu et Zimbabwe (*D. simulans*) ont été déterminés par « genome walking », technique permettant d'identifier la séquence du génome flanquant l'insertion d'un ET. Ces populations naturelles de *D. simulans* offrent ainsi un système idéal pour étudier la dynamique de *tirant* chez la drosophile, et plus particulièrement les marques épigénétiques comme la méthylation de l'ADN et les modifications post-traductionnelles des histones.

CHAPITRE 1

Étude des marques épigénétiques associées à *tirant* dans des populations naturelles de *D. simulans*

1. Détermination et analyse des profils de méthylation de la région régulatrice de *tirant*

1.1. Méthodes d'analyse de la méthylation de l'ADN

Différentes approches, ne fournissant pas les mêmes informations, permettent d'analyser la méthylation de l'ADN des génomes. La détermination du taux de 5 méthylcytosine (5-MedC) par rapport au taux de cytosines (dC) totales de différents génomes (espèces, individus ou tissus différents) peut être réalisée par chromatographie dont le principe est détaillé dans le chapitre 2. Ce type d'analyse globale ne permet cependant pas d'identifier les séquences d'ADN méthylées. D'autres techniques sont utilisées pour identifier les séquences méthylées d'un génome et établir les profils de méthylation des cytosines correspondant à ces séquences. Elles reposent sur l'immunoprécipitation, les enzymes de restrictions ou encore, le traitement au bisulfite de sodium. L'amélioration des technologies de séquençage et l'apparition du séquençage à haut débit permettent maintenant de séquencer le méthylome d'un génome à partir de l'une de ces trois techniques (Pomraning et al. 2009).

1.1.1. L'immunoprécipitation de l'ADN méthylé (MeDiP)

L'utilisation d'un anticorps spécifique dirigé contre les 5-MedC sert à immunoprécipiter les fragments d'ADN méthylés. Il est ensuite possible de mesurer par q-PCR, pour un locus précis, un enrichissement relatif en séquences méthylées en comparant la fraction enrichie (immunoprécipitée par l'anticorps spécifique des 5-MedC), et la fraction témoin (immunoprécipitée par un anticorps aspécifique, généralement anti-IgG). Les séquences de la fraction enrichie en ADN méthylé peuvent aussi être identifiées par clonage et séquençage (Jacinto et al. 2008), ou par hybridation sur des puces à ADN (Zilberman et Henikoff, 2007).

1.1.2. Les enzymes de restriction

Certaines enzymes de restriction sont sensibles à la présence de 5-MedC sur leur site de restriction. Des techniques basées sur ces propriétés, appelées MSRE (Methyl Sensitive Restriction Enzymes) permettent de mettre en évidence la présence ou non de 5-MedC dans un génome. En effet, il est possible, après digestion d'un échantillon d'ADN par un couple

d'enzyme, sensible et non sensible à la méthylation de l'ADN, de visualiser sur gel d'électrophorèse les polymorphismes de digestion entre ces deux enzymes, signifiant alors la présence de sites de restriction méthylés sur leurs cytosines. De plus, la comparaison de deux profils de digestion de l'enzyme sensible à partir d'échantillons différents peut être utilisée pour détecter des profils de méthylation tissu ou sexe-spécifique. La résolution limitée de la méthode ne fait pas toujours ressortir des différences entre les enzymes sensibles et non sensibles. Il est néanmoins possible d'analyser les fragments d'ADN digérés par PCR, q-PCR (PCR quantitative), Southern-blot ou encore par hybridation sur des puces à ADN.

Au cours de ma thèse, j'ai utilisé la PCR avec des amorces flanquant le site de restriction pour mettre en évidence des sites de méthylation au niveau des ET; cette technique sera détaillée dans le chapitre 2.

1.1.3. Les traitements au bisulfite de sodium

Le « Bisulfite Genomic Sequencing » (BGS), permettant d'établir des profils de méthylation de séquences spécifiques, est très largement utilisé car il est très informatif. Par le biais de cette technique, il est possible de déterminer le profil de méthylation des cytosines sur des séquences d'ADN jusqu'à 1000 pb. Le BGS repose sur les propriétés chimiques du bisulfite de sodium qui convertit les cytosines non méthylées en uracile. La cinétique de réaction sur les cytosines non méthylées est très élevée, tandis qu'à l'inverse, elle est très faible sur les cytosines méthylées (Hayatsu 2008) (Figure 5).

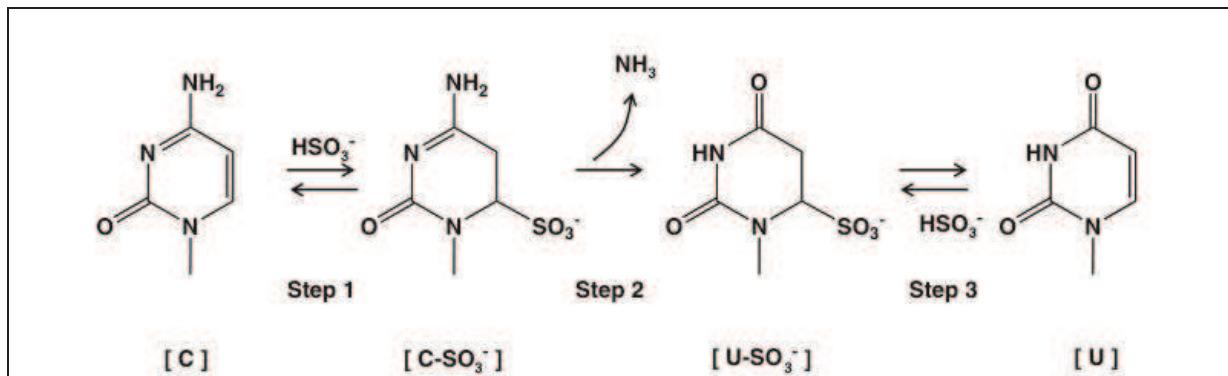


Figure 5. Etapes de conversion de la cytosine en uracile par le bisulfite de sodium

La conversion des cytosines non méthylées en uracile par le bisulfite de sodium se déroule en trois étapes. Tout d'abord, il y a addition d'ions hydrogénosulfates à la cytosine (Step 1), pour former des cytosines sulfurées (C-SO₃⁻) qui vont ensuite être déaminées par hydrolyse (Step 2) pour former des uraciles sulfurées (U-SO₃⁻). Enfin, un simple traitement alcalin (Step 3) permet de libérer les ions hydrogénosulfates (désulfonation) pour former l'uracile.

L'efficacité de cette réaction est optimale sur des molécules d'ADN purifiées et dénaturées. La conversion des cytosines non méthylées conduit à la formation de deux brins d'ADN qui ne sont plus complémentaires. Le profil de méthylation d'un locus d'intérêt peut être établi, après PCR, clonage et séquençage d'un ADN traité au bisulfite de sodium, par comparaison de la séquence traitée avec une séquence de référence non traitée (Figure 6). Après le traitement au bisulfite de sodium, les molécules d'ADN sont composées d'uraciles, obtenues après conversion des cytosines non méthylées, qui seront, lors de la PCR, remplacées par des thymines. Les cytosines méthylées, non converties lors du traitement, pourront être alignées avec les cytosines de la séquence de référence, tandis qu'un mésappariement T/C entre la séquence traitée et la séquence de référence sera caractéristique de la déamination, lors du traitement, d'une cytosine non méthylée. A partir de ce profil de méthylation, il est possible d'une part de calculer un taux de méthylation, et d'autre part d'étudier la distribution des cytosines méthylées en fonction du contexte dinucléotidique. Une méthode basée sur l'utilisation d'enzymes de restriction, sur un produit de PCR amplifié à partir d'un ADN génomique traité précédemment au bisulfite de sodium, a été développée et baptisée COBRA (Combine Bisulfite Restriction Analysis, (Eads et Laird, 2002)). Son principe est d'utiliser les propriétés du bisulfite pour modifier les sites de restriction d'endonucléases. Le site de restriction est préservé après le traitement au bisulfite de sodium si ses cytosines sont méthylées. Dans ce cas, l'enzyme pourra reconnaître son site de restriction et le cliver. A l'inverse, si les cytosines du site de restriction ne sont pas méthylées,

le traitement au bisulfite de sodium modifiera la séquence du site de restriction et l'enzyme ne pourra pas le cliver.

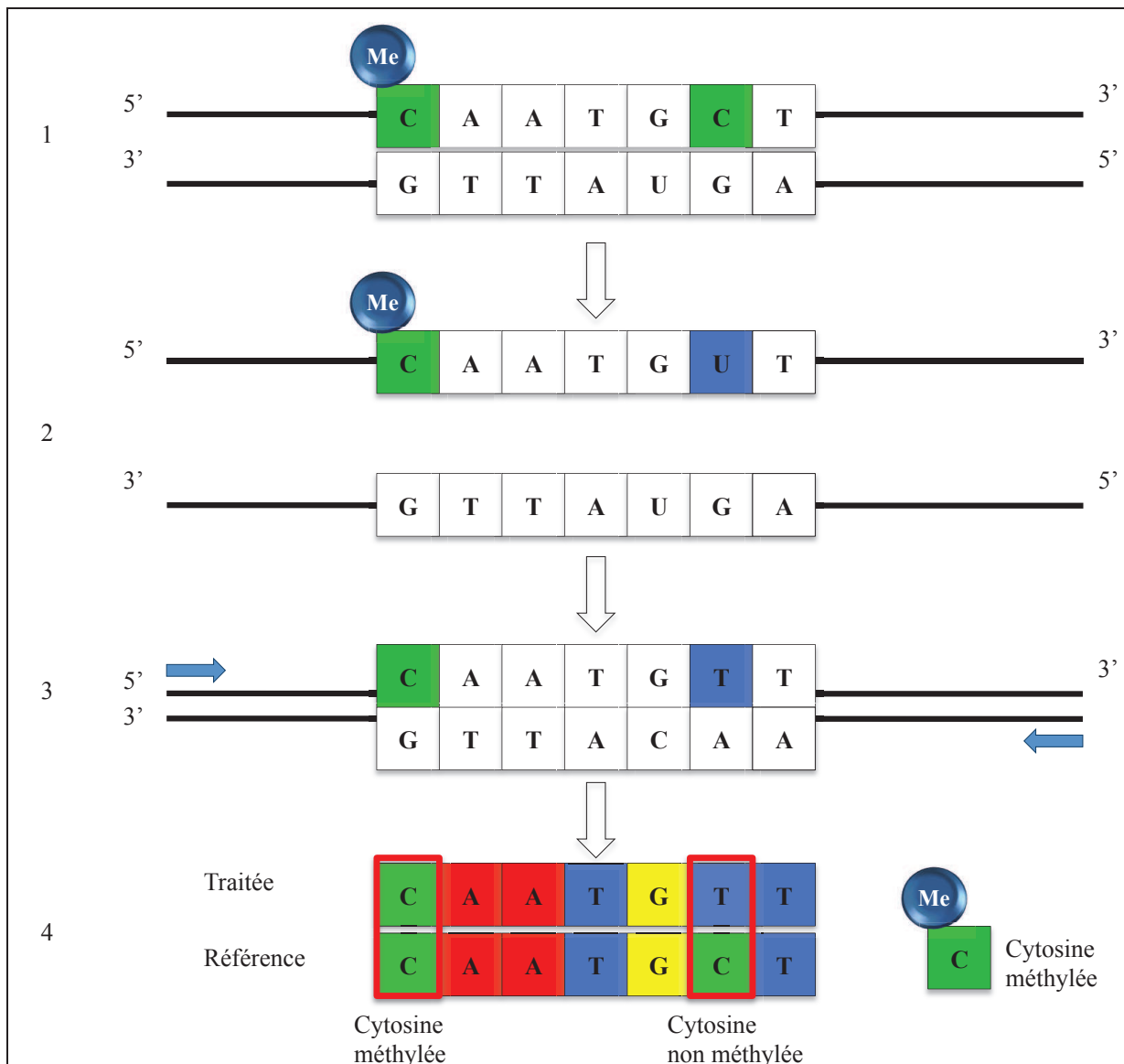
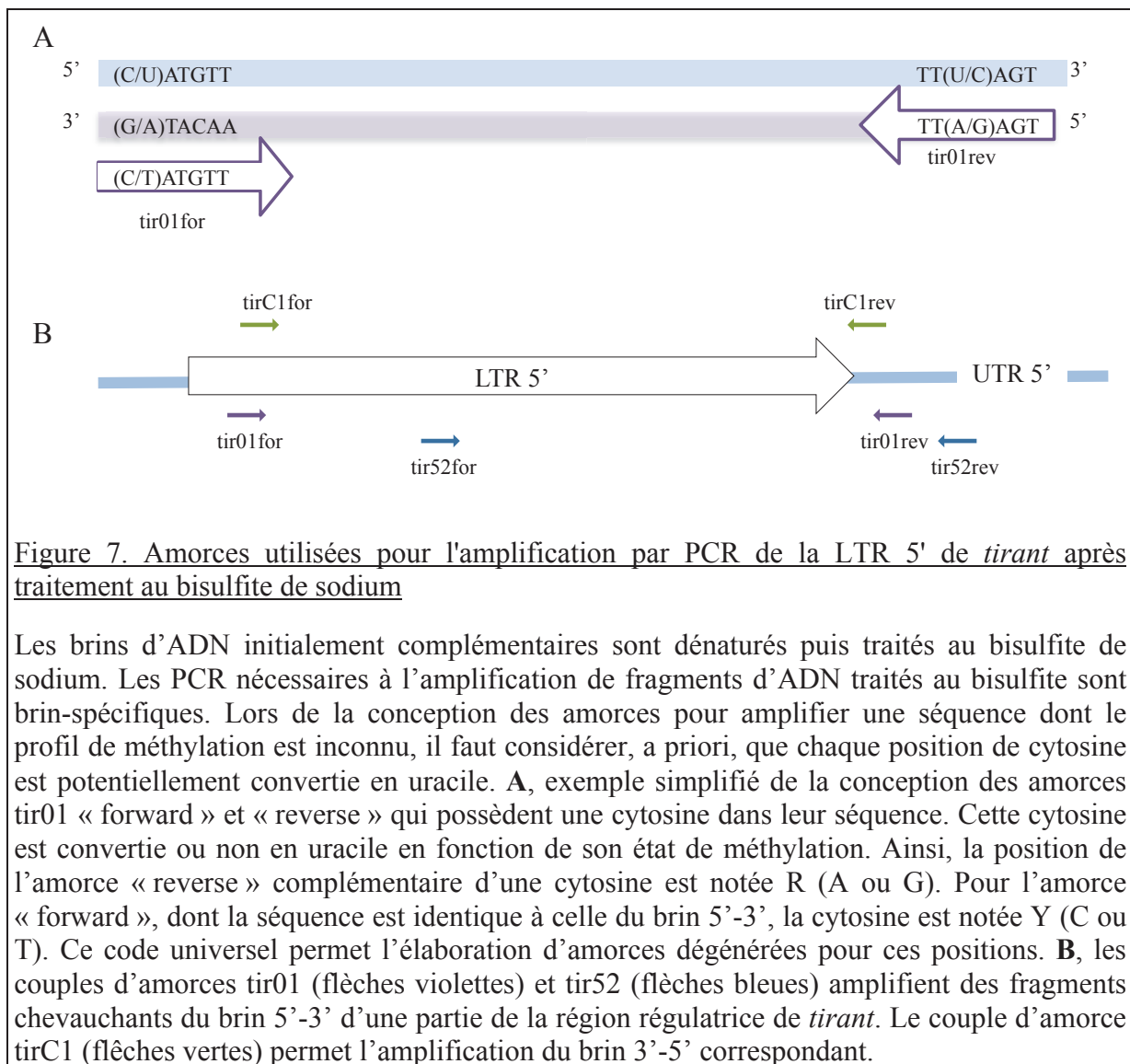


Figure 6. Schéma de la technique de "Bisulfite Genomic Sequencing"

1, une molécule d'ADN double brin porte une cytosine méthylée sur le brin 5'-3', localisée sur le dinucléotide CpA. La deuxième cytosine du même brin est, quant à elle, non méthylée. Le brin 3'-5' ne porte pas de cytosine méthylée. 2, une fois que l'ADN a été dénaturé, les cytosines non méthylées vont être converties en uracile. 3, le fragment d'intérêt est amplifié par PCR avec des amorces brin spécifique du brin 5'-3'. Lors de la PCR, des thymine sont incorporées à la place des uraciles. 4, Après clonage et séquençage, les séquences obtenues sont alignées avec la séquence de référence non traitée. L'alignement des cytosines permet d'identifier les cytosines méthylées tandis que l'alignement d'une cytosine avec une thymine permet d'identifier une cytosine non méthylée.

L'un des objectifs de ma thèse a été de déterminer si la région régulatrice de *tirant* est associée à des cytosines méthylées selon un profil spécifique et ce, afin d'expliquer son activité dans les populations naturelles. Dans ce but, j'ai déterminé les profils de méthylation de la région régulatrice de *tirant* par BGS. Cette technique comporte des artefacts liés au traitement au bisulfite, conduisant parfois à une interprétation erronée des résultats. Ces artefacts sont associés à certaines étapes critiques, telles que la dénaturation de l'ADN, la désulfonation (étape 3 du traitement au bisulfite, Figure 5) et l'amplification PCR des fragments traités. La conception des amorces nécessaires à l'amplification d'un fragment à partir d'ADN traité au bisulfite est aussi une étape critique. Il faut tenir compte du fait que l'amplification est brin spécifique, mais aussi que toutes les cytosines localisées au niveau de la zone d'hybridation entre l'amorce et l'ADN génomique sont susceptibles d'avoir été converties ou non en uracile. Chez l'homme, les cytosines méthylées sont exclusivement localisées au niveau des dinucléotides symétriques CpG, particularité importante pour la conception d'amorces et l'évaluation de l'efficacité du traitement puisque les cytosines en contexte asymétrique ne sont jamais méthylées. Lors de la conception des amorces, ces cytosines sont virtuellement converties en uracile et les amorces sont positionnées sur des séquences complémentaires dépourvues en CpG. Cette approche possède deux avantages : en premier lieu, l'utilisation d'amorces dégénérées (ensemble d'amorces contenant plusieurs combinaisons de séquences) n'est pas nécessaire, ce qui facilite la mise au point et la spécificité de la PCR ; en second lieu, seules les séquences totalement converties par le traitement sont amplifiées. De plus, toutes les cytosines hors CpG non converties par le traitement deviennent des contrôles internes qui permettent d'appréhender l'efficacité de la réaction de conversion. Ainsi, dans toutes les études de profils de méthylation symétriques, l'amplification par PCR brin spécifique, inhérente à la technique de BGS, offre un contrôle interne de l'efficacité du traitement. Chez la drosophile, la situation est plus complexe : les cytosines méthylées étant potentiellement localisées dans tous les contextes dinucléotidiques, la démarche décrite ci-dessus ne peut pas être envisagée. L'utilisation d'amorces dégénérées devient dès lors indispensable. Egalement, les conditions d'amplification par PCR (température d'hybridation, quantité d'ADN traité à amplifier, quantité de MgCl₂, cycle de PCR) nécessitent une longue période de mise au point. En effet, la conversion en uracile des cytosines des différentes copies d'un locus, comme dans le cas des copies d'un ET, sont susceptibles de générer plusieurs combinaisons de séquences cibles pour les amorces. Dans ces conditions, des gradients de température d'hybridation sont réalisés afin de déterminer celle qui permettra l'amplification d'une bande spécifique. Pour amplifier la région

régulatrice de *tirant*, j'ai conçu plusieurs couples d'amorces différents dont trois, nommés tir01, tir52 et tirC1, qui ont permis l'amplification de produit de PCR (Figure 7). Les deux couples d'amorces tir01 et tir52 amplifient des régions chevauchantes du brin 5'-3' des LTR-UTR5' de toutes les copies de type C de *tirant*. Elles ne permettent pas l'amplification de solo LTR mais privilégient les copies plus longues, voire complètes. Le couple d'amorces tirC1 amplifie le brin 3'-5' correspondant et sera utilisé comme contrôle pour vérifier les résultats obtenus par les autres amorces.



L'utilisation de différentes amorces pour amplifier de l'ADN traité au bisulfite est fortement conseillée car les biais d'amplification sont fréquents (Wojdacz et al. 2008). Les amorces tir01 et tir52 ont une composition en base et des longueurs différentes. Ainsi, les amorces tir01 et tir52 permettent respectivement l'amplification d'un produit de PCR

spécifique correspondant aux copies de type C de *tirant* à des températures d'hybridation de 43°C et 52,4°C. Des amorces amplifiant les copies de type S ont également été mises au point dans le but de vérifier si ces copies, exclusivement hétérochromatiques et non transcrites, possèdent des profils de méthylation spécifiques. La difficulté de cette expérience réside dans l'absence de contrôle interne permettant de valider l'efficacité du traitement. L'efficacité du traitement au bisulfite de sodium peut être réduite dans différentes conditions. La présence de protéines ou d'une dénaturation partielle de l'ADN peut empêcher la réaction de conversion des cytosines non méthylées ; ceci conduisant à une mauvaise interprétation des résultats et à une surestimation du taux de méthylation des locus analysés. Afin de palier au manque de contrôles internes de cette technique, j'ai intégré dans le plan expérimental des réplicats techniques et biologiques. J'ai aussi vérifié l'efficacité de la conversion des cytosines sur des ADN extraits à partir des 3 lignées différentes toutes mutantes pour le gène dDNMT2 (GS12412, DNMT2^{-/-} et R2), pour lesquelles le taux de méthylation déterminé doit correspondre au seuil d'efficacité de la technique. Les résultats, obtenus avec les amorces tir01 à partir d'ADN purifié extrait des lignées mutantes, sont présentés dans la Figure 8 sous la forme d'un alignement représentant les profils de méthylation des cytosines de la LTR 5' de *tirant*. Ces derniers ont été établis par comparaison avec la séquence de référence de *tirant*, obtenue pour chaque population ; les figures ont été produites par le logiciel Kismeth (Gruntman et al. 2008), disponible en ligne (<http://katahdin.mssm.edu/kismeth/revpage.pl>). Ce logiciel réalise ces profils à partir de deux fichiers soumis par l'utilisateur : un premier composé de toutes les séquences traitées, non alignées, et un second, contenant la séquence de référence. Les alignements correspondant aux figures présentées dans ma thèse sont disponibles en annexe. Dans la représentation des profils de méthylation donnée par le logiciel Kismeth, toutes les cytosines de la séquence sont mises bout à bout et un code couleur définit le contexte trinuécléotidique de chacune d'entre elles (CHH, CHG ou CG). Les cytosines non méthylées sont représentées par un rond vide et les cytosines méthylées par un rond rempli de la couleur correspondante au contexte trinuécléotidique. Les différents contextes trinuécléotidiques sont CHH, HG ou CG, dans lesquels H peut être A, C, ou T. Cet outil d'analyse des séquences d'ADN obtenues après traitement au bisulfite, initialement développé pour l'analyse de profils de méthylation chez les plantes, est particulièrement intéressant pour l'étude de la méthylation de l'ADN chez la drosophile, puisqu'il considère que toutes les cytosines peuvent être potentiellement méthylées indépendamment du contexte nucléotidique. La grande majorité des autres outils d'analyse des séquences traitées au bisulfite de sodium ne prennent en compte que les dinuécléotides CpG, masquant ainsi toutes

les autres cytosines. Les premiers résultats obtenus sur des mutants de dDNMT2 montrent que la majorité des cytosines de l'ensemble des clones a été convertie par le traitement, avec une efficacité identique à celle théorique de 99 %, mentionnée dans le kit utilisé. Deux séquences obtenues à partir du mutant GS12412 n'ont pas été converties, probablement à cause d'une mauvaise dénaturation de l'ADN. Dans les séquences obtenues à partir des mutants DNMT2^{-/-} et R2, nous distinguons deux types de séquences, en fonction de la répétition ou non d'un motif de 19 pb au début de la région analysée. Le type C de *tirant* présente en effet au niveau de sa région régulatrice une séquence de 19 pb qui peut être répétée ou non (Fablet et al. 2006). Les séquences de référence utilisées pour les alignements possèdent toutes une répétition de ce motif. Lorsque ce motif de 19 pb n'est pas répété dans les clones obtenus, cela se traduit, sur les profils de méthylation, par une délétion en début de séquence. Ces deux sous variants de *tirant* ont été identifiés dans les lignées DNMT2^{-/-} et R2, contrairement à la lignée GS12412, dans laquelle seul celui composé de deux motifs de 19 pb a été retrouvé dans mes résultats.

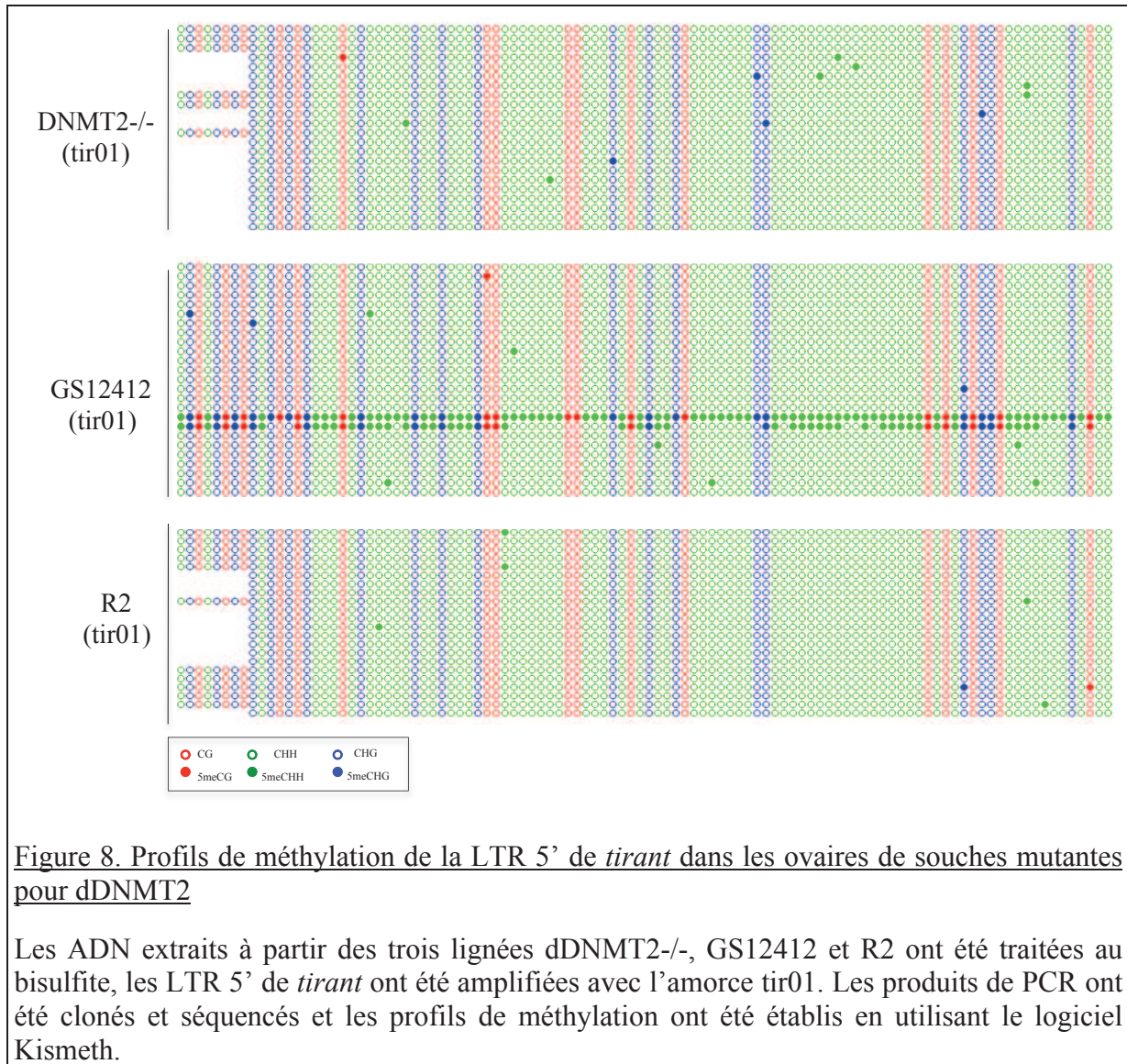


Figure 8. Profils de méthylation de la LTR 5' de *tirant* dans les ovaires de souches mutantes pour dDNMT2

Les ADN extraits à partir des trois lignées dDNMT2^{-/-}, GS12412 et R2 ont été traités au bisulfite, les LTR 5' de *tirant* ont été amplifiées avec l'amorce tir01. Les produits de PCR ont été clonés et séquencés et les profils de méthylation ont été établis en utilisant le logiciel Kismeth.

Ce premier résultat, qui permet notamment de valider la technique de BGS, montre que parfois, certaines séquences d'ADN présentent une résistance au bisulfite de sodium, indépendante de la méthylation des cytosines.

Durant la phase de mise au point de la technique, des profils de méthylation de la région régulatrice de *tirant* ont été obtenus à partir d'ADN purifié extrait d'ovaires de la population naturelle de Makindu (*D. simulans*), puis traité au bisulfite de sodium en utilisant indépendamment les amorces tir01 et tir52 (Figure 9).

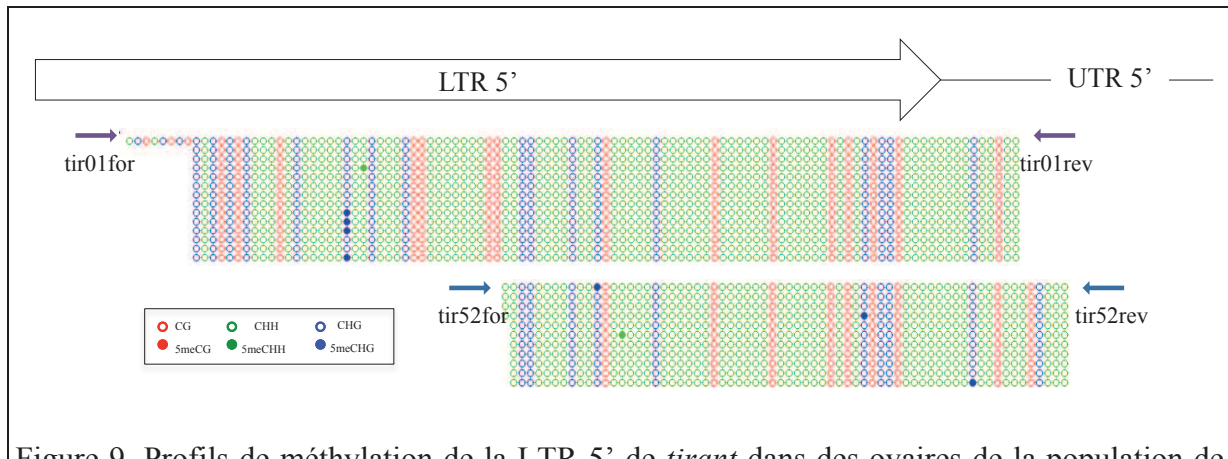


Figure 9. Profils de méthylation de la LTR 5' de *tirant* dans des ovaires de la population de Makindu

Des ADN purifiés extraits à partir d'ovaire de la population du Makindu ont été traités au bisulfite. Les deux couples d'amorces utilisés, tir01 for et tir01 rev (flèches violettes) et tir52 for et tir52 rev (flèches bleues) amplifient des fragments chevauchants de la région régulatrice de *tirant*.

Les résultats montrent que le taux de méthylation de la région régulatrice de *tirant* dans la population de Makindu est très faible et les cytosines sont méthylées sans position spécifique, comme observés chez les mutants. Une position méthylée commune entre quatre clones différents a été mise en évidence avec les amorces tir01. L'alignement de ces séquences disponibles en annexe ne permet cependant pas de détecter de polymorphisme entre ces quatre clones ; nous ne pouvons donc pas écarter l'hypothèse qu'il s'agisse du même produit de PCR cloné puis séquencé plusieurs fois. Par ailleurs, une substitution de thymine en cytosine en position 250 se retrouve uniquement dans ces quatre séquences, suggérant qu'il s'agit bien du même amplicon.

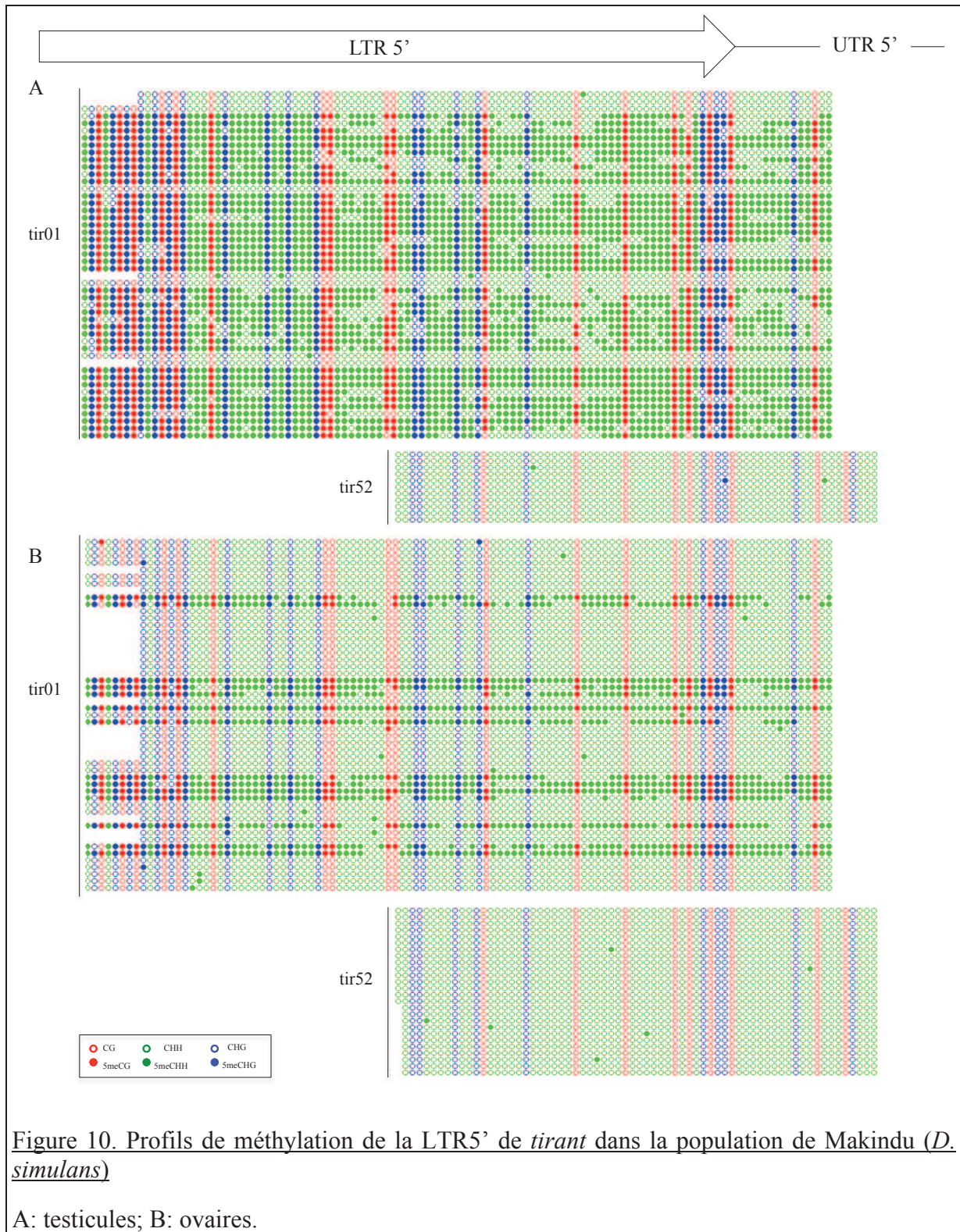
Les nouveaux traitements au bisulfite ne nécessitent pas de grandes quantités d'ADN, contrairement à ceux disponibles lorsque j'ai commencé mon projet de thèse. J'ai utilisé un kit similaire à celui testé précédemment, proposé par le même fabricant afin de traiter de très faibles quantités de tissus. Après broyage et traitement à la protéinase K, de faibles quantités de tissu peuvent être traitées sans étape d'extraction d'ADN. Ce choix a été motivé par le fait que l'extraction de quantités d'ADN de testicules suffisantes pour le traitement nécessitait de nombreuses heures de dissection. Ainsi, à partir de 35 paires de testicules ou de 20 paires d'ovaires broyées et traitées à la protéinase K, un traitement au bisulfite suivi d'une amplification PCR est possible.

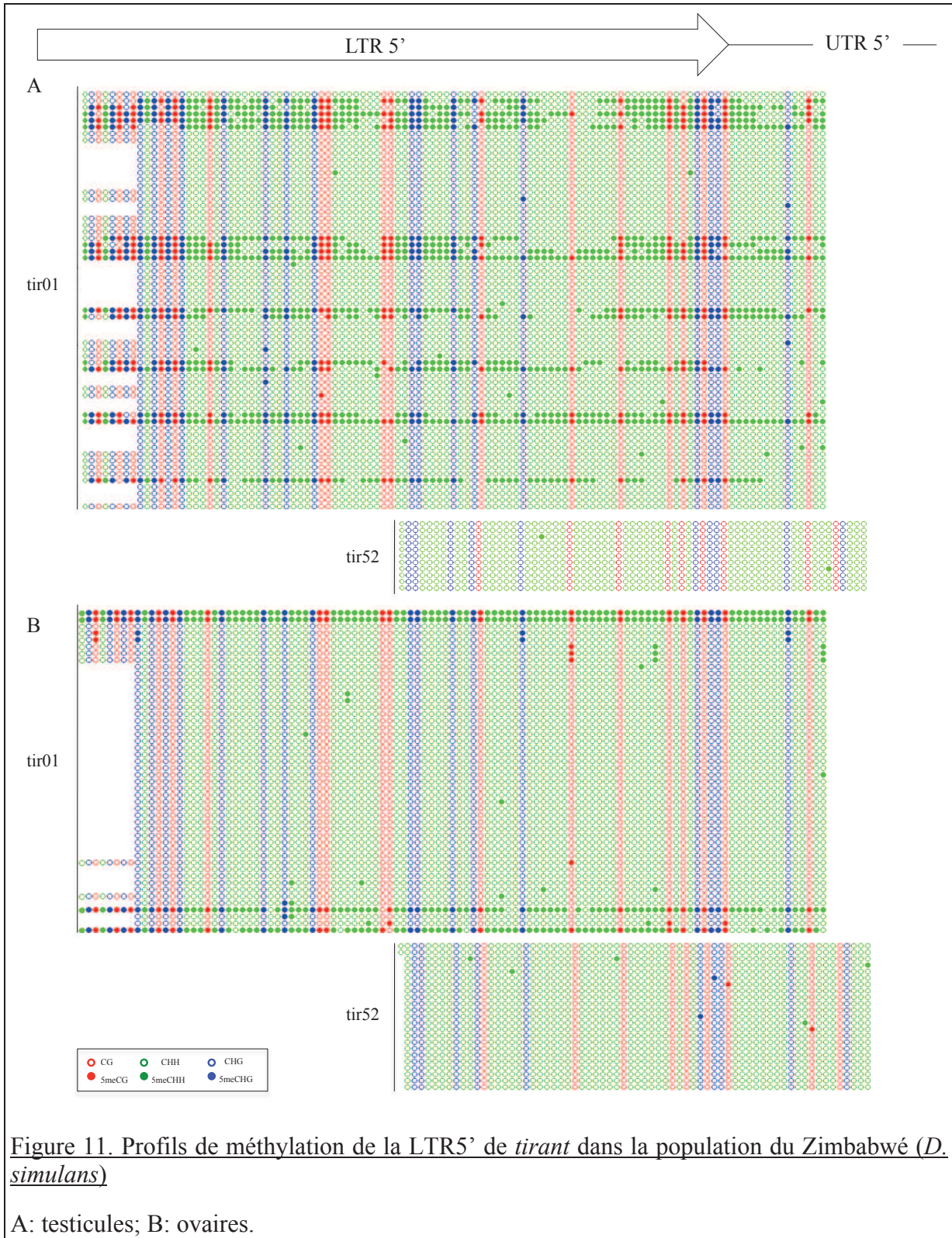
1.2. Profils de méthylation de la LTR 5' de *tirant*

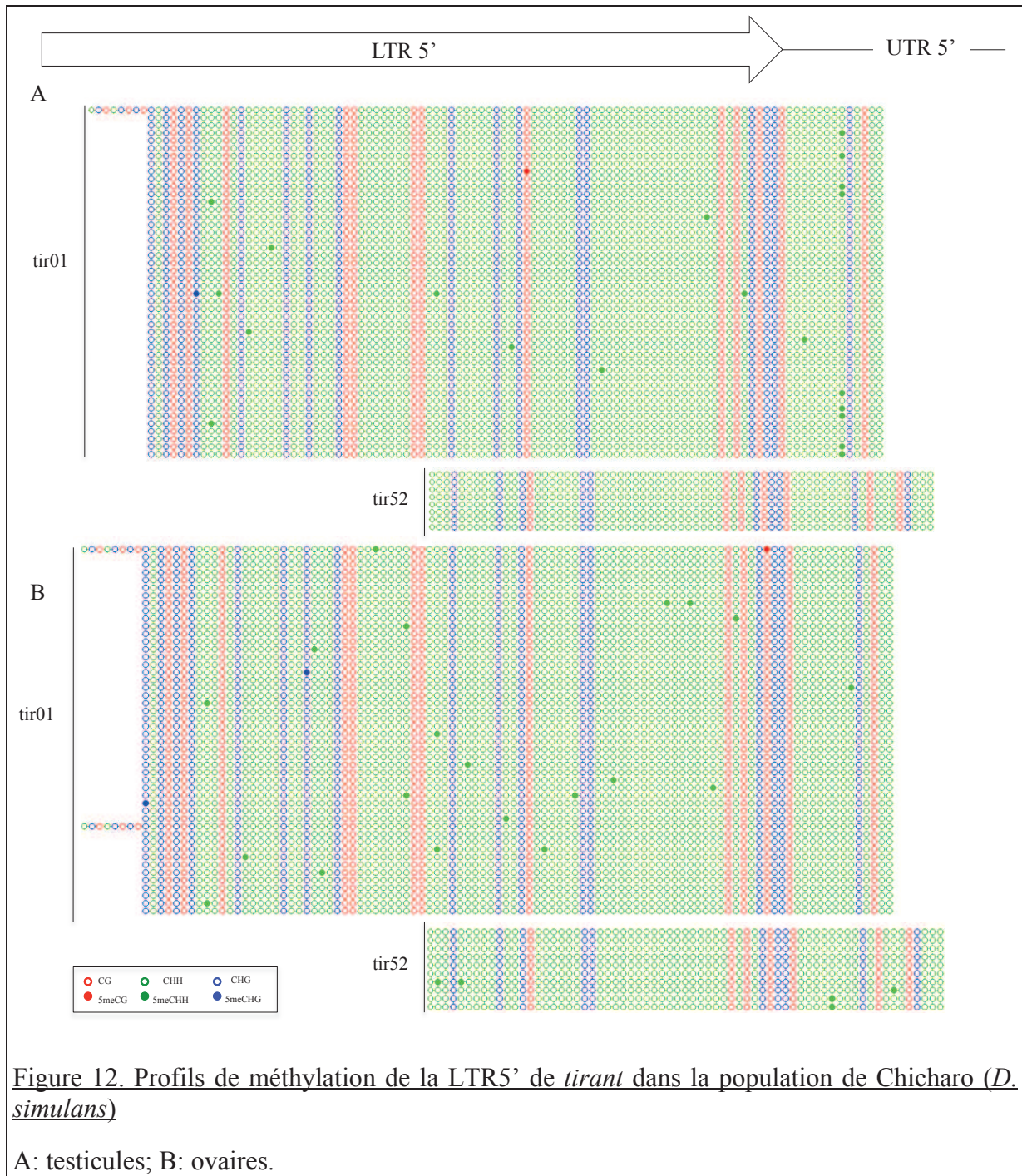
Les populations naturelles du Sénégal (*D. melanogaster*), de Chicharo, du Zimbabwe et de Makindu (*D. simulans*) possèdent respectivement neuf, zéro, deux et cinq copies euchromatiques de *tirant*. En outre, pour les populations de *D. simulans*, la connaissance des sites d'insertion des copies de *tirant* (Tableau 2) obtenues dans l'équipe par « Genome Walking » (Fablet et al. 2009) a permis une analyse site spécifique des modifications épigénétiques associées à *tirant*. Afin de savoir si la méthylation de l'ADN au niveau de la région régulatrice de *tirant* est associée au polymorphisme d'insertion dans les populations naturelles, j'ai réalisé une étude sur la lignée germinale des drosophiles, à partir d'ovaires et de testicules. La population de Chicharo est particulièrement intéressante car elle ne possède que des copies hétérochromatiques de *tirant*, permettant de détecter si celles-ci ont des profils de méthylation spécifiques de l'hétérochromatine. Les résultats obtenus avec les amorces tir01 et tir52 sont présentés dans les Figure 10 (Makindu), Figure 11 (Zimbabwe), Figure 12 (Chicharo) et Figure 13 (Sénégal).

Tableau 2. Insertions des copies de type C de *tirant* identifiées dans les populations de Makindu, Chicharo et Zimbabwe (*D. simulans*)

Population	Localisation	Commentaire	Taille
Makindu			
M1	ADN centromérique	Ilots de Maupiti	Copie complète
M2	2L	Intron du gène <i>tkv</i>	Copie complète
M3	X	ADN non annoté	Copie complète
Zimbabwe			
Z1	ADN centromérique	Ilots de Maupiti	Fragment court
Z2	3R	ADN non annoté	Fragment court
Z3	3	Hétérochromatine	Fragment long
Chicharo			
C1	ADN centromérique	Ilots de Maupiti	Copie complète
C2		Hétérochromatine	Fragment long
C3		<i>diver2</i> LTR retrotransposon	Fragment court







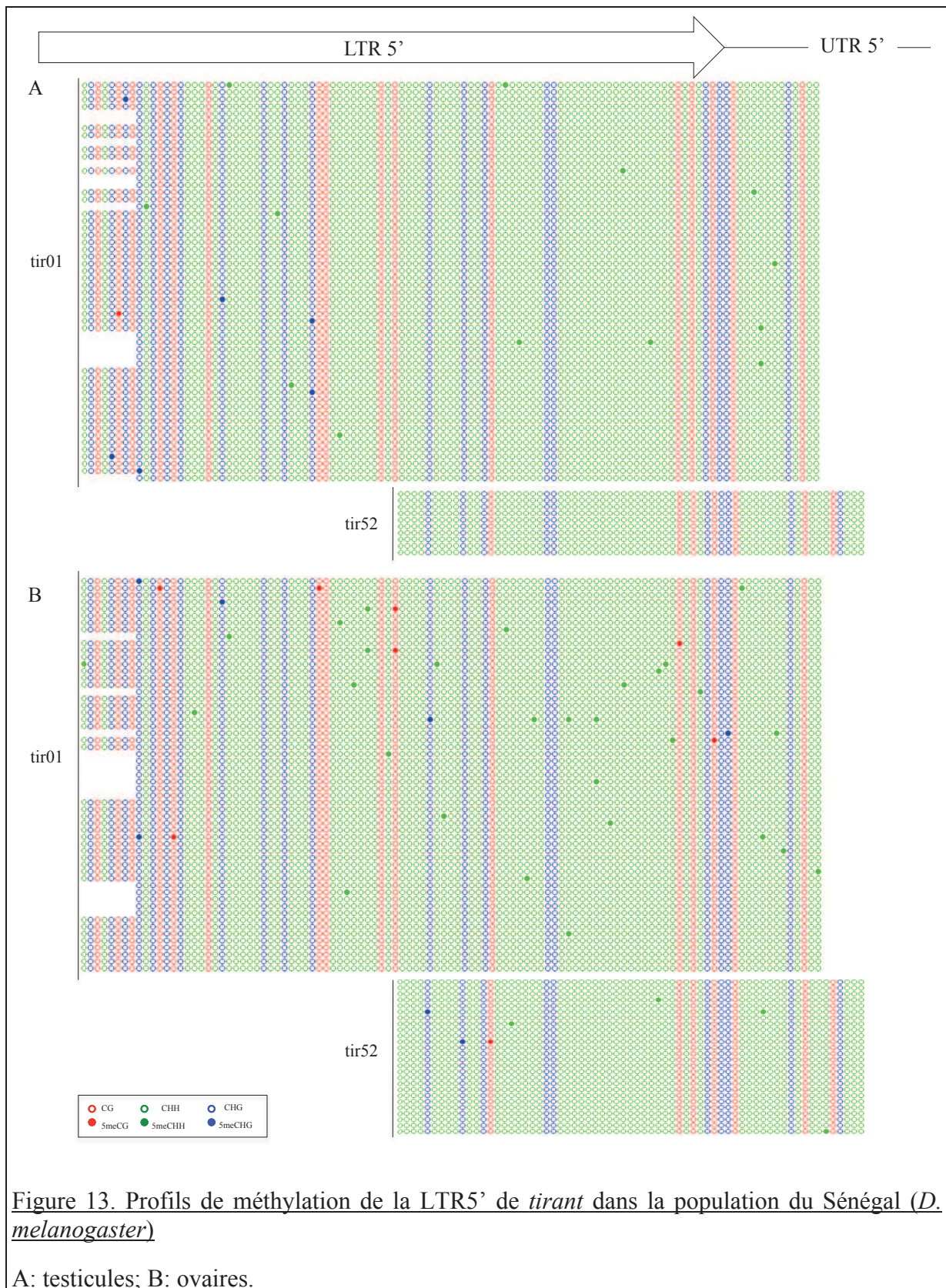


Figure 13. Profils de méthylation de la LTR5' de *tirant* dans la population du Sénégal (*D. melanogaster*)

A: testicules; B: ovaires.

Les séquences obtenues pour les populations du Sénégal (*D. melanogaster*) et de Chicharo (*D. simulans*) tant au niveau des ovaires qu'au niveau des testicules ont des niveaux très faibles de méthylation et aucun profil spécifique n'a été identifié. Lorsqu'on utilise les amorces tir01, les populations de Makindu et Zimbabwe montrent la présence de séquences dans lesquelles les cytosines sont non converties. Le taux de non conversion est élevé dans les réplicats de la population de Makindu mais plus faible dans la population du Zimbabwe. De plus, les séquences fortement méthylées ne concernent que les séquences de type C possédant deux motifs de 19 paires de bases répétées en début de LTR. En revanche, les séquences obtenues avec les amorces tir52 ne comportent que très peu de cytosines méthylées, dont la répartition tout au long de la séquence est sporadique. Nous avons vu, avec les résultats des mutants de DNMT2, que dans certains cas la LTR 5' de *tirant* peut être non convertie. Les résultats obtenus avec amorces tir01 montrent un biais d'amplification dans lequel les séquences non converties sont préférentiellement amplifiées et les amorces tir52 sont quant à elle idéales puisqu'elles n'amplifient que les séquences converties. Des séquences définies comme des « Densely Methylated DNA Island » (DMI) dont le taux de méthylation déterminé par BGS est compris entre 50 et 100 %, ont été mises en évidence il y a une quinzaine d'années chez plusieurs espèces dont l'Homme (Selker et al. 1993; Tasheva et Roufa, 1995; Tasheva et Roufa, 1994). Le profil de méthylation de ces séquences ne semble pas avoir de préférence dinucléotidique. Ces résultats ont été discrédités par d'autres études des mêmes régions, dans les mêmes génomes, dans des conditions optimales de traitement au bisulfite (Rein et al. 1997). La non-dénaturation, la dénaturation partielle ou encore l'absence de traitement à la protéinase K favorisent l'obtention de séquences non converties (Warnecke et al. 2002). Nous reviendrons sur les artefacts liés à la technique du BGS dans le paragraphe suivant mais il est important de noter que les profils de méthylation des séquences identifiées comme méthylées par Selker et al. et par Tasheva et al. sont du même type que ceux que nous avons obtenus avec les amorces tir01 dans les populations de Makindu et Zimbabwe. Pour la suite de mon manuscrit, les séquences ayant un taux de méthylation élevé, sans préférence dinucléotidique et sans position conservée entre les clones, seront donc considérées comme des séquences non converties. Il reste à noter cependant que les séquences non converties ont été retrouvées systématiquement et uniquement dans les réplicats techniques et biologiques des populations de Makindu et Zimbabwe. Les raisons de la reproductibilité de cette non conversion, spécifique à ces populations, restent à déterminer, et ce d'autant plus que nous observons des profils de cytosines non converties forts ressemblants entre diverses populations, avec certaines positions très conservées (Figure supplémentaire 1, p. 161).

Sans tenir compte de ces séquences, le taux de méthylation moyen observé dans les autres clones des différentes populations est comparable à celui observé dans les lignées mutantes pour dDNMT2. L'absence de positions méthylées conservées entre les différents clones est un argument majeur permettant d'affirmer que les cytosines identifiées comme méthylées dans mes résultats, ne sont pas impliquées dans la régulation de l'expression de *tirant*. Toutefois, l'analyse que j'ai réalisée concerne la méthylation de l'ensemble des LTR5' des copies de *tirant* des génomes de chaque espèce. Or les biais d'amplification PCR peuvent conduire à une représentation biaisée des différents profils de méthylation existants au sein de chaque génome. Afin de vérifier si certains profils sont caractéristiques de certaines régions, j'ai réalisé une analyse site-spécifique de la LTR de trois copies complètes M1, M2 et M3 (Tableau 2, p.43), présentes dans le génome de la population de Makindu, à des compartiments chromatiniens différents. Les résultats des profils de méthylation de ces insertions sont représentés sur la Figure 14.

Les résultats obtenus avec le couple d'amorces tir151 pour la copie M3 insérée sur le chromosome X montrent une forte proportion de séquences non converties. En revanche, l'utilisation d'un deuxième couple d'amorces, tir152, pour cette même insertion permet d'amplifier des séquences converties. Pour chacun des trois sites étudiés, aucun ne présente des taux de méthylation supérieurs à ceux présents chez les mutants et aucune position méthylée n'est retrouvée parmi les clones. Ainsi, aucun profil de méthylation spécifique des copies complètes de *tirant* insérées dans l'hétérochromatine ou dans l'euchromatine ne peut être mis en évidence.

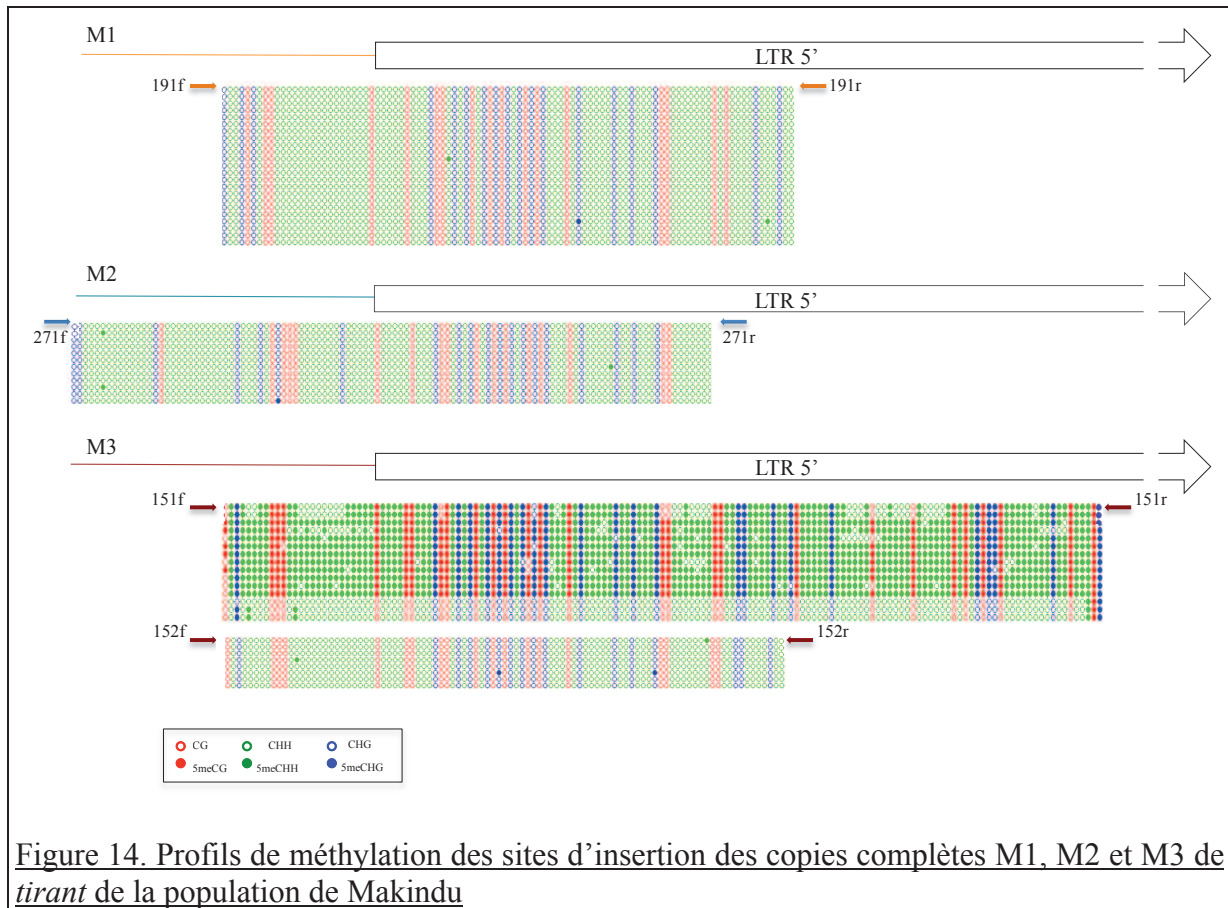


Figure 14. Profils de méthylation des sites d'insertion des copies complètes M1, M2 et M3 de *tirant* de la population de Makindu

1.3. Biais de la technique de bisulfite

Pour vérifier si les séquences non converties sont dues à la présence de cytosines méthylées, j'ai réalisé des expériences complémentaires visant à déterminer les profils de méthylation du brin 5'-3' dans différentes conditions, ainsi que du brin complémentaire de la région régulatrice de *tirant* dans la population de Makindu, population dans laquelle la proportion de séquences non converties est la plus forte. Les facteurs ayant la capacité de se fixer à l'ADN empêchent la conversion des cytosines lors du traitement au bisulfite et sont soit de type protéiques soit de type nucléiques. De nombreuses populations de petits ARN ont été mises en évidence ces dernières années, principalement des piRNA, produits à partir des ET. La méthode que j'ai utilisée pour le traitement au bisulfite à partir des tissus ne comporte pas d'étape de dégradation des ARN. Nous pouvons aussi envisager que, lors du traitement, dans lequel la renaturation des brins d'ADN est une étape critique, des petits ARN complémentaires de *tirant* présents augmentent la cinétique de réassociation des molécules. Différents traitements au bisulfite ont été réalisés, dont un premier dans les conditions standards, un deuxième précédé d'un traitement plus long à la protéinase K, et enfin un troisième précédé d'un traitement à la RNase H (réalisé après le traitement à la protéinase K)

dégradant ainsi les ARN appariés à l'ADN. Les résultats présentés sur la Figure 15 confirment, une fois de plus, que les conditions standards conduisent à l'obtention de séquences non converties. Le traitement réalisé avec un supplément de protéinase K permet quant à lui d'obtenir une séquence non convertie. Enfin, les séquences obtenues à partir d'un traitement avec supplément en protéinase K, suivi d'une incubation d'une heure à 37°C avec la RNaseH ont toutes été converties. Ainsi, le traitement des échantillons à la RNaseH et à la protéinase K optimise la conversion des cytosines par traitement au bisulfite. Ceci n'explique cependant pas pourquoi des séquences non converties sont retrouvées avec les amorces tir01, uniquement dans les populations de Makindu et Zimbabwe, et que ces séquences ne concernent que les copies de *tirant* de type C comportant une répétition du motif de 19 paires de bases.

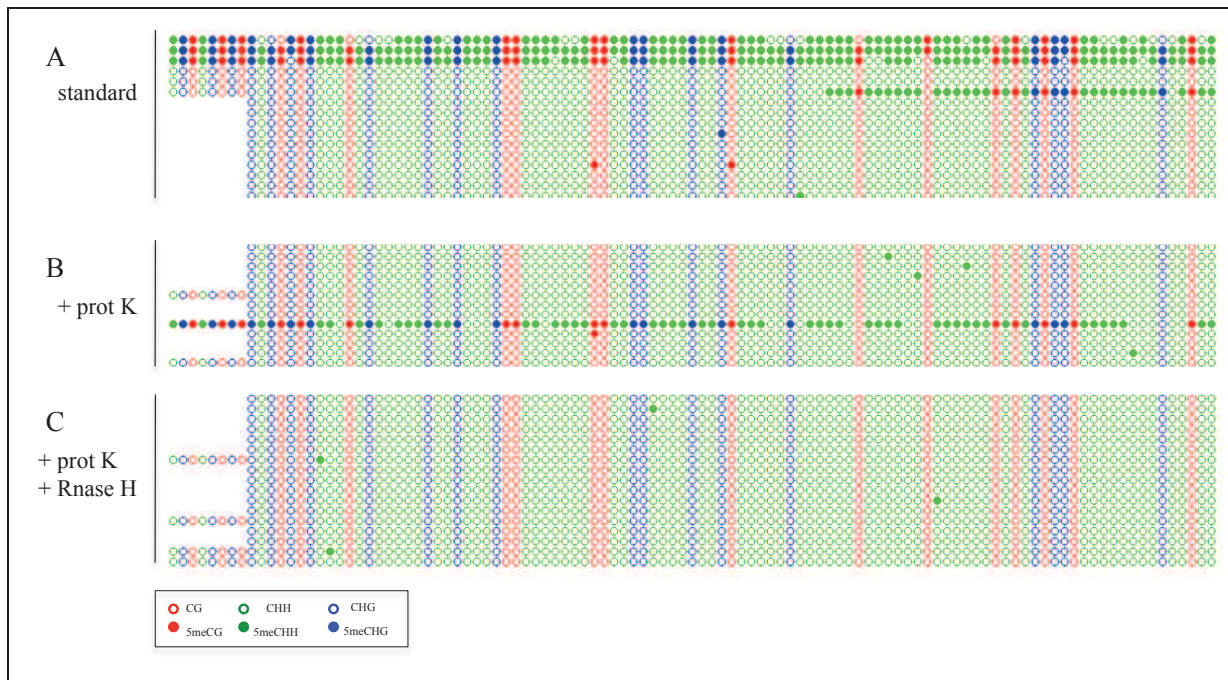
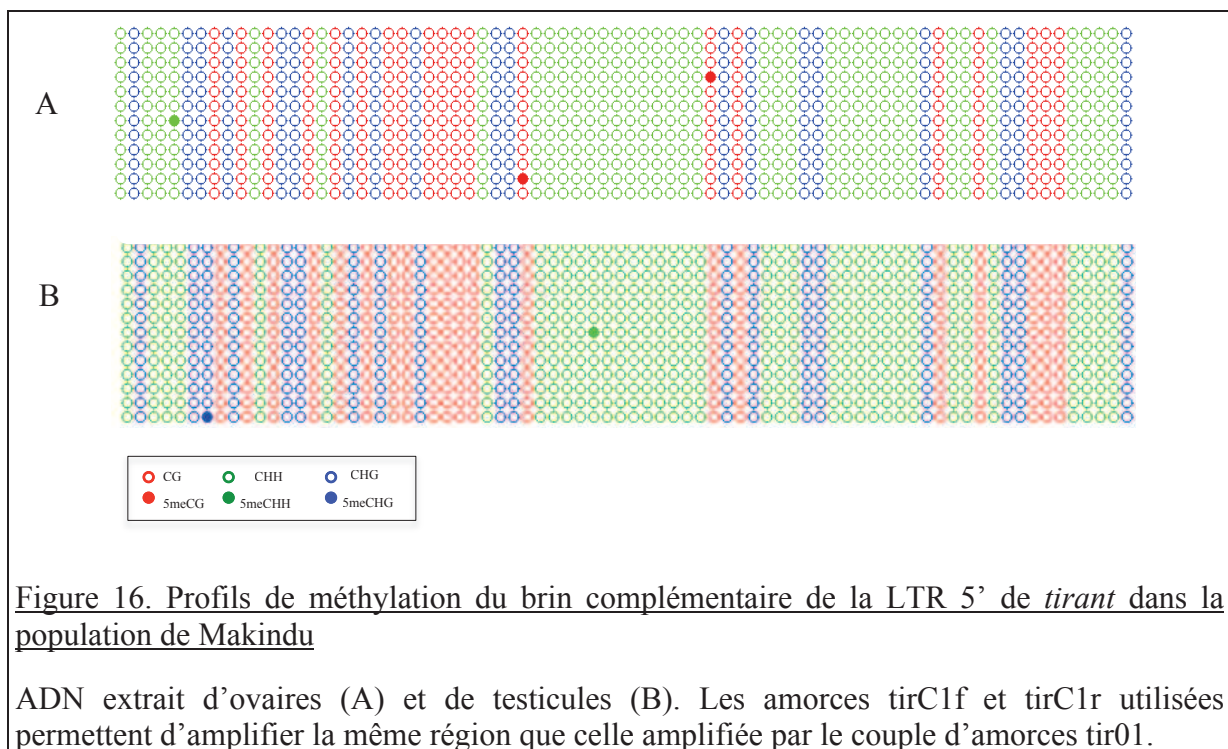


Figure 15. Efficacité du traitement au bisulfite selon les conditions expérimentales

Des traitements au bisulfite de sodium ont été réalisés à partir d'ovaires broyées de la population de Makindu. Le couple d'amorces tir01 a été utilisé pour amplifier la LTR 5' de *tirant* à partir de l'ADN traité. A: ADN traité dans les conditions standards. B: Ajout de protéinase K au tissu broyé. C: Ajout de protéinase K suivi d'un traitement à la Rnase H.

Les cytosines non converties du brin 5'-3' présentes dans les séquences obtenues à partir d'échantillons de la population de Makindu font intervenir tous les types de dinucléotides, tant symétriques qu'asymétriques. Les cytosines méthylées des dinucléotides CpG devraient également être détectées méthylées sur le brin complémentaire dans le cas de vraie méthylation symétrique. Ainsi, la détermination du profil de méthylation du brin

complémentaire valide les résultats observés au niveau du brin sens (Warnecke et al. 2002). Les amorces tirC1 sont des amorces dégénérées qui permettent d'amplifier après traitement au bisulfite de sodium une partie complémentaire des séquences déterminées avec les amorces tir01. Les profils de méthylation du brin complémentaire de la LTR 5' de *tirant* représentés sur la Figure 16 ont été établis à partir d'ovaires et de testicules de la population de Makindu. Seules quelques cytosines méthylées ont été mises en évidence sur le brin complémentaire, dont deux CpG méthylés dans les ovaires. Ce faible taux de méthylation symétrique sur le brin complémentaire est un argument supplémentaire démontrant que les séquences fortement méthylées n'ont pas de réalité biologique.



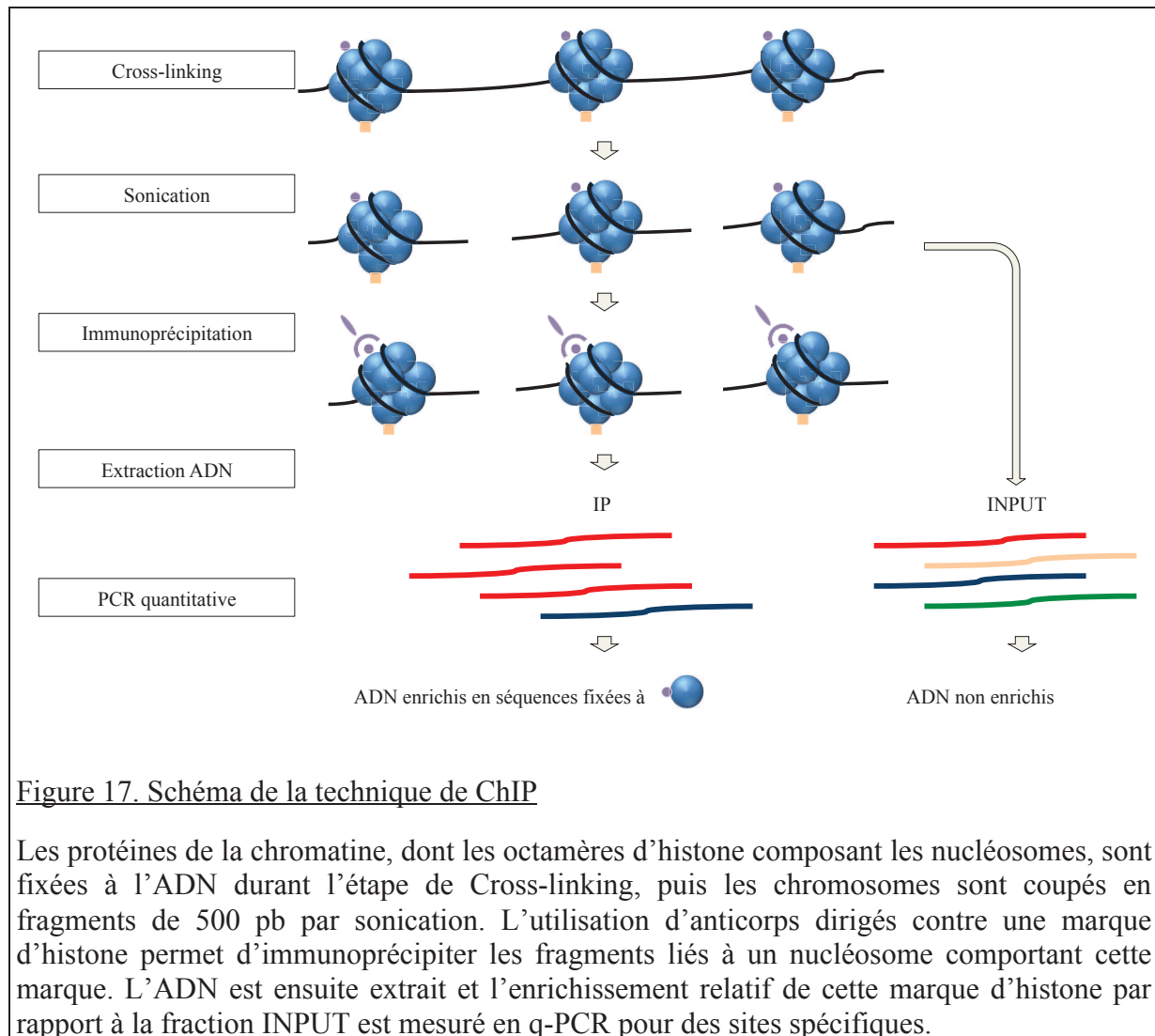
2. Analyse des modifications d'histones associées à des insertions de *tirant*

Comme nous l'avons vu dans l'introduction, certaines marques d'histones sont associées à de l'hétérochromatine constitutive, d'autres à de l'euchromatine. Afin de caractériser le contexte chromatinien des insertions de *tirant*, j'ai choisi des marques caractéristiques de l'euchromatine, comme H3K4me2 et de l'hétérochromatine, comme H3K9me2 et H3K27me3. En effet, la chromatine associée à des séquences répétées présente de très faible taux de H3K4me2 et est enrichie en H3K9me2 (Peng et Karpen, 2007). Afin d'établir le contexte chromatinien des copies de *tirant* chez *D. simulans*, j'ai ensuite recherché par Immunoprécipitation de la Chromatine (ChIP), parmi ces marques d'histones, lesquelles étaient le plus fortement représentées.

2.1. Principe de la ChIP

La ChIP est une immunoprécipitation de la chromatine via un anticorps spécifique d'une marque d'histone. Cette méthode vise à obtenir une fraction d'ADN enrichie en cette marque (Figure 17). Différentes analyses peuvent alors être réalisées, selon la façon dont cette fraction d'ADN enrichie sera analysée. La première étape de la ChIP consiste à solidariser l'ADN de la chromatine en fixant les protéines liées à l'ADN. Cette étape, appelée « cross linking », préserve les liaisons entre l'ADN et les protéines impliquées dans la structure de la chromatine. L'utilisation d'anticorps dirigés contre les marques d'histones d'intérêt permettra d'obtenir une fraction appelée Immunoprécipitat (IP) d'ADN enrichi en cette marque par rapport à la fraction d'ADN non enrichie appelée INPUT (Figure 32, matériels et méthodes, p.153). La chromatine est extraite des noyaux par lyse mécanique et chimique des cellules, puis est soniquée (cassure par ultrasons) pour casser les chromosomes en petits fragments de 500 paires de bases et permettre une bonne résolution de la ChIP. Ces fragments sont ensuite immunoprécipités par les anticorps spécifiques des marques d'histones recherchées. Un témoin d'immunoprécipitation, réalisé avec un anticorps anti IgG, sera utilisé comme contrôle négatif, mesurant l'aspécificité de l'immunoprécipitation. La récupération des complexes ADN/protéines immunoprécipités s'obtient par des lavages. La chromatine est ensuite traitée à la RNase et à la protéinase K (ou « reverse cross-link »), puis l'ADN est extrait au phénol-chloroforme. L'enrichissement relatif des différentes copies de *tirant* pour chaque marque d'histone est déterminé par qPCR par rapport à la fraction INPUT, non

enrichie. Les modifications d'histones étudiées sont toutes liées à l'histone H3. Les résultats d'enrichissement des différentes marques ont été normalisés par l'enrichissement en H3 total (H3T) de chaque site analysé, ce qui permet de s'affranchir du biais dû à l'abondance en nucléosomes pour chaque site analysé. Ceci est nécessaire car les régions d'insertion des copies ne sont pas associées de façon identique à H3. Le gène euchromatique RP49 a été utilisé comme contrôle de ChIP.



L'analyse par qPCR de l'ADN immunoprécipité doit concerner des copies uniques afin d'être la plus résolutive possible. Puisque les sites d'insertion de *tirant* dans les populations de Makindu, Zimbabwe et Chicaro sont connus, une analyse site spécifique des modifications d'histones associées à *tirant* vérifiera que l'état chromatinien observé corresponde à celui attendu, relatif à la position de la copie sur le chromosome. Les sites d'insertion de *tirant* dans les populations étudiées sont rappelés dans le Tableau 2 (p.43).

2.2. Modifications d'histones associées à *tirant*

Les résultats des marques d'histones associées aux différentes copies de *tirant* sont représentés dans la Figure 18. Ces résultats montrent que rp49 est majoritairement associé à la marque H3K4me2, et que l'enrichissement en H3K9me2 et H3K27me3 est très faible. Ceci signifie que le gène rp49 est associé à une région euchromatique, favorable à sa transcription, ce qui confirme que celui-ci est un gène de ménage. Les copies de *tirant* M1, Z1 et C1 sont toutes insérées au même site centromérique dans les ilots de Maupiti. Ces copies sont enrichies en H3K27me3 dans les trois populations et présentent un enrichissement variable en H3K9me2 et H3K4me2. En effet, la marque H3K4me2 se retrouve enrichie uniquement au niveau des copies Z1 et M1, contrairement à la copie C1. Les autres copies de la population de Chicharo ont des enrichissements en H3K27me3 uniquement.

La marque euchromatique H3K4me2 se retrouve fortement associée aux copies M2, M3 et Z2 tandis que la marque H3K27me3 est associée à la copie M2. Les copies M3, copie complète, et Z2, fragment court de *tirant*, ont des profils d'histones similaires, ce qui montre qu'il n'y a pas d'influence de la taille de la copie sur les modifications d'histones qui lui seront associées. Cela s'observe aussi pour les copies de Chicharo, dont la copie C1 est une copie complète, tandis que les autres sont délétées. Enfin, la copie Z3 présente un profil original d'enrichissement, principalement en H3K9me2, avec de faibles valeurs d'enrichissement pour les marques H3K4me2 et H3K27me3. La présence de marques euchromatiques et hétérochromatiques associées à la même séquence, ou bivalence, concernant les copies M1 et M2, a déjà été mise en évidence pour *tirant* dans les tissus somatiques de *D. simulans* (Fablet et al. 2009) ainsi que chez l'Homme, dans les cellules souches embryonnaires avec les marques H3K4me2 et H3K27me3 (Bernstein et al. 2006).

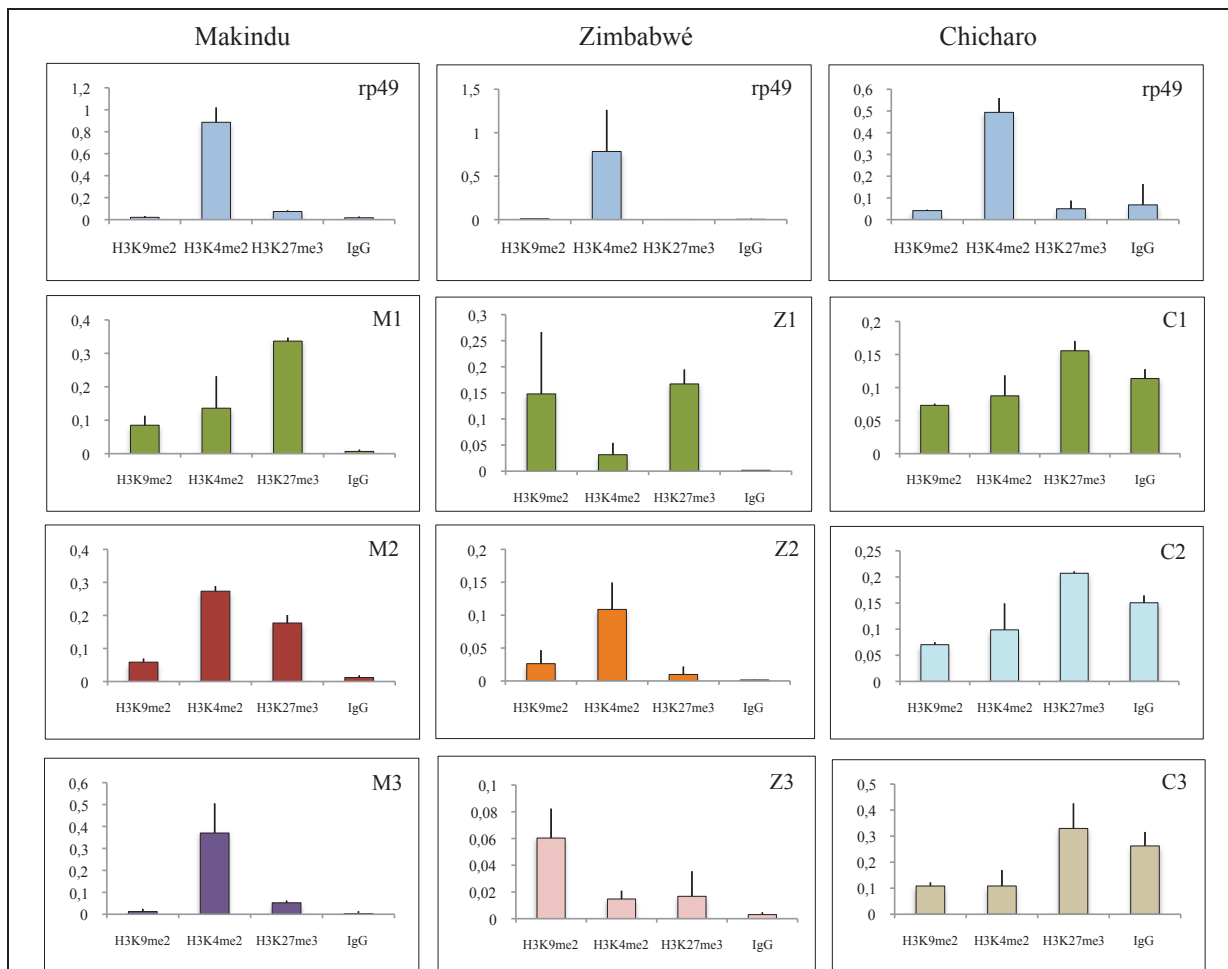


Figure 18. Modifications post-traductionnelles de l'histone H3 associées à *tirant*

Les enrichissements relatifs pour chaque marque d'histone analysée ont été normalisés par l'histone H3 totale. Le gène *rp49* a été utilisé comme contrôle euchromatique. Les marques H3K9me2, H3K4me2 et H3K27me3 ont été analysées. L'IgG a été utilisé comme contrôle de l'aspécificité des IP.

Les profils d'histones associés aux copies *Z3* et *Z2* de *tirant* montrent que la copie *Z3* est dans un contexte hétérochromatique alors que la copie *Z2* est dans un contexte euchromatique. Ce résultat est tout à fait compatible avec le fait que la copie *Z2* est insérée sur le bras du chromosome, tandis que la copie *Z3* est insérée dans une région hétérochromatique. Pour la population de Chicharo, les copies *C2* et *C3* ne peuvent pas à ce jour être positionnées dans le génome séquencé de *D. simulans* à cause de la mauvaise qualité de son séquençage ainsi que les problèmes d'annotations qui lui sont liés. Les enrichissements des copies *C2* et *C3* sont exclusivement hétérochromatiques, ce qui est en accord avec le fait que la population de Chicharo ne possède pas de copies de *tirant* sur les bras des chromosomes. L'ensemble des résultats présentés ci-dessus montre que les profils de

modifications d'histones identifiés pour les copies de *tirant* sont cohérents avec leurs localisations dans les génomes des populations étudiées. L'expression de *tirant* dans la population de Makindu peut être expliquée par le fait que la copie M3 est associée majoritairement à une modification d'histone caractéristique de l'euchromatine favorable à la transcription. A l'inverse, les copies M1 et M2 de *tirant*, localisées dans des régions euchromatiques, portent des marques d'hétérochromatine facultative. La copie M2 est insérée dans le gène *tkv*, qui code pour un récepteur de decapentaplegic (dpp), impliqué dans le développement embryonnaire de la drosophile (Terracol et Lengyel, 1994; Brummel et al. 1994). Lorsqu'une nouvelle copie d'un ET s'insère à un locus du génome, elle peut adopter la structure chromatinienne de son site d'insertion, ou être la cible de nouvelles modifications épigénétiques. En conséquence, l'insertion de *tirant* dans ce gène provoquera, le cas échéant, un bouleversement de la structure chromatinienne de ce gène et modifiera ainsi son expression. Nous avons ensuite utilisé le polymorphisme d'insertion de *tirant* au sein des populations naturelles de *D. simulans* pour voir l'effet de cet élément sur la structure de la chromatine au niveau de son site d'insertion.

2.3. Influence de *tirant* sur la chromatine chez *D. simulans*

Les sites M2 et M3 sont des sites d'insertion de *tirant* spécifiques de la population de Makindu. Ces sites n'ont pas d'insertion de *tirant* dans les populations de Chicharo et Zimbabwe. Nous avons donc analysé les enrichissements relatifs en H3K9me2, H3K27me3 et H3K4me2, aux deux sites d'insertion des copies M2 et M3 dans les populations de Chicharo et Zimbabwe (Figure 19), afin de savoir si *tirant* est capable de modifier la structure de la chromatine à son site d'insertion. Nos résultats montrent que les sites d'insertion des copies M1 et M3 ne sont associés qu'à la marque euchromatique H3K4me2. Or, chez Makindu, en plus de la présence de la marque euchromatique H3K4me2, les marques hétérochromatiques H3K9me2 et H3K27me3 sont associées à cette copie de *tirant*. Ainsi, les sites d'insertion des copies M1 et M3 ne sont pas hétérochromatiques lorsque *tirant* est absent, et les marques hétérochromatiques associées à *tirant* dans la population de Makindu pour ces deux copies ont probablement comme influence la répression de ces copies. Une insertion de *tirant* dans un génome peut provoquer une modification de la conformation de la chromatine au niveau des sites d'insertion. Nous nous sommes ainsi posés la question de savoir si un « spreading » (étendue) de ces marques pouvait être mis en évidence et associé à des modifications de l'expression des gènes proches. Pour répondre à cette question, nous avons regardé si les

marques hétérochromatiques associées à la copie M2 s'étendent jusqu'au promoteur le plus proche. Le gène *tkv* produit quatre variants nommés RA, RB, RC et RD, initiés par quatre promoteurs différents (Figure supplémentaire 3, p.164). Le promoteur du variant RD est situé en amont de l'insertion de *tirant*, et nous avons déterminé les marques d'histones associées à ce promoteur dans les populations de Makindu, Chicharo et Zimbabwe (Figure 19).

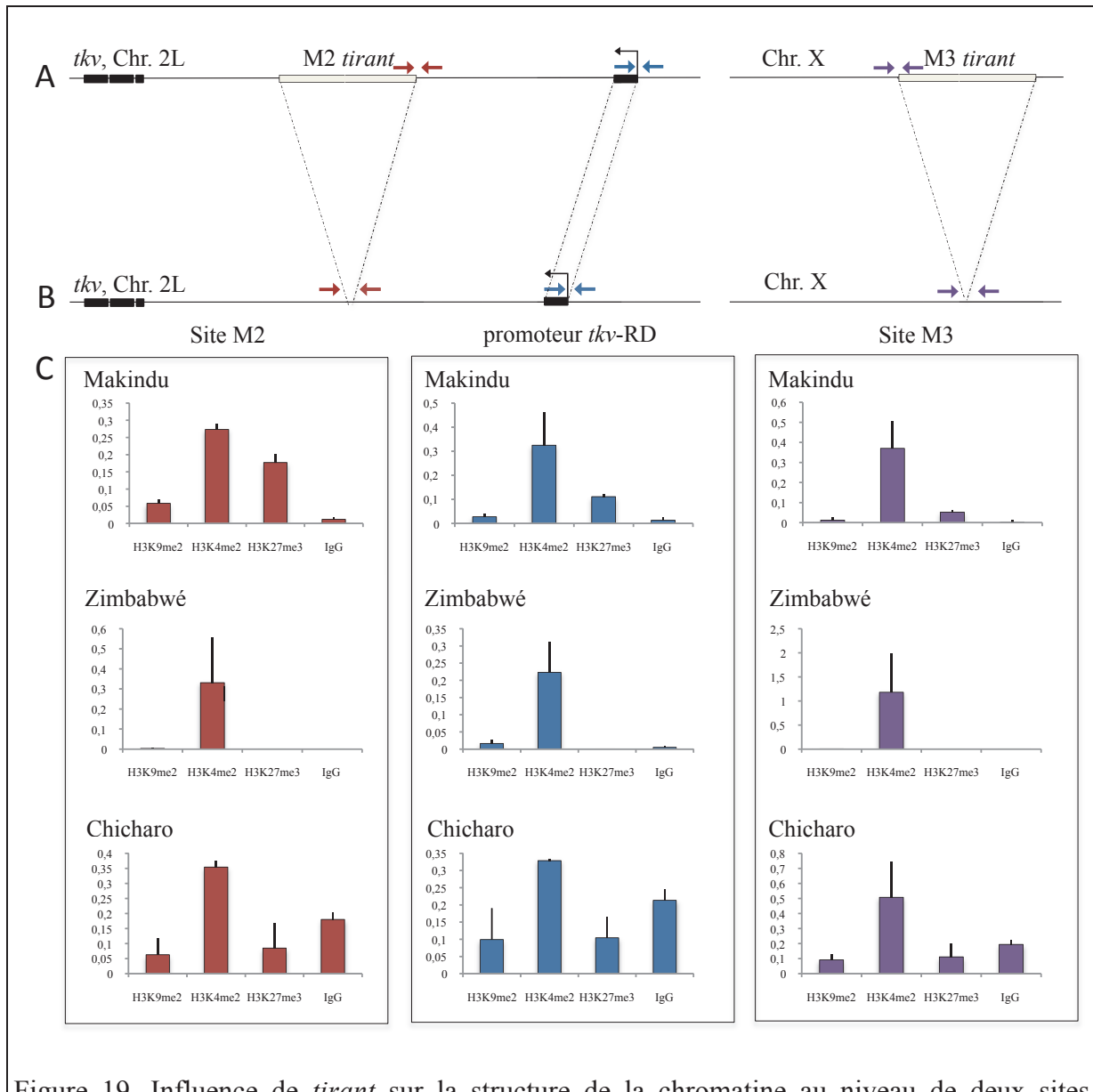


Figure 19. Influence de *tirant* sur la structure de la chromatine au niveau de deux sites d'insertion

Sites d'insertions des copies M2 (dans le gène *tkv*) et M3 (dans le chromosome X) dans la population de Makindu (A) et sites vides correspondants dans les populations de Chicharo et Zimbabwe (B). Les marques d'histones associées à ces sites ont été analysées par ChIP (C). Les amorces utilisées en qPCR sont positionnées (flèches).

Mes résultats montrent que le promoteur du variant RD du gène *tkv* est associé à la marque hétérochromatique H3K27me3 dans la population de Makindu, tandis que dans les populations de Chicharo et Zimbabwe, il est associé uniquement à H3K4me2. Un « spreading » des marques hétérochromatiques associées à *tirant* existe donc dans la population de Makindu au niveau de la copie M2, entraînant une modification de la structure chromatinienne du promoteur du variant RD du gène *tkv*. Cependant, la distance entre l'insertion de *tirant* et le promoteur RD, qui est de 2571 pb chez *D. melanogaster*, n'est pas connue chez *D. simulans*. En effet, certaines régions du génome séquencé de *D. simulans*, dont une partie du gène *tkv*, présentent des « gaps » pour lesquels la séquence nucléotidique n'a pas pu être déterminée. L'évaluation par PCR de la distance entre le site d'insertion de la copie de *tirant* et le promoteur RD est en cours, et nous pourrions évaluer plus précisément cette distance chez *D. simulans*. Ainsi, au delà de l'effet délétère qu'une nouvelle insertion d'ET peut engendrer, *tirant* est capable de bouleverser la structure chromatinienne de son site d'insertion. L'expression du gène *tkv* dans ces trois populations a été mesurée par qPCR et montre que l'insertion de la copie M2 dans le gène *tkv* est associée à une diminution de son expression. Ces résultats sont présentés dans un article en préparation disponible dans la partie Publications, p.95.

3. Conclusion sur la régulation épigénétique de *tirant* chez *D. simulans*

L'analyse des marques épigénétiques associées à *tirant* dans les populations naturelles de *D. simulans* montre que les modifications des histones semblent prédominer sur la méthylation de l'ADN. En effet, il paraît peu probable que les très faibles taux de méthylation associés à *tirant* sans position spécifiquement méthylée soient impliqués dans un mécanisme de contrôle de *tirant*. Ces résultats, obtenus de façon globale au niveau des LTR5' de *tirant*, ont été renforcés par l'absence de profils sites spécifiques lors de l'analyse des copies M1, M2 et M3 de la population de Makindu. Ces copies présentent des profils de modifications d'histones différents bien que chaque insertion de *tirant* soit caractérisée par la présence des marques hétérochromatiques H3K9me2 et H3K27me3, ce qui suggère un contrôle de l'activité de cet élément induit par des modifications de la structure de la chromatine. En outre, la copie complète M3, qui échappe à cette règle, pourrait correspondre à la copie exprimée dans la population de Makindu.

CHAPITRE 2

Analyse globale de la méthylation de l'ADN dans les génomes de drosophiles

1. Quantification des taux de méthylation de l'ADN dans les génomes de drosophiles

1.1. Introduction

Un nucléotide est formé d'une base azotée, purique (A,G) ou pyrimidique (C,T), associée à un 2'désoxyribose et à un groupe 5' phosphate. Le groupe 5' phosphate est impliqué dans la liaison phosphodiester permettant de relier entre eux les nucléotides dans le sens 5'-3' et de former ainsi la molécule d'ADN. Un nucléotide qui a perdu son groupe 5' phosphate est appelé un nucléoside et ne peut plus former de liaison phosphodiester. Les cytosines de l'ADN peuvent être méthylées sur leur carbone numéro 5 et les nucléosides non méthylés et méthylés correspondant sont appelés respectivement 2'-déoxycytidine (dC) et 5-methyl-2'-déoxycytidine (5-MedC) (Figure 20). Les taux de méthylation dans un génome peuvent être importants. C'est le cas chez les mammifères, de nombreux insectes et la plante *A. thaliana*. Chez les drosophiles, des estimations faites pour les espèces *D. simulans*, *D. melanogaster* et *D. pseudoobscura* décrivent des niveaux très faibles de 5-MedC (< 1%) (Marhold et al. 2004). Alors que certains insectes comme l'abeille ou le bombyx ont des taux de méthylation de leur génome élevés (Xiang et al. 2010; Elango et al. 2009), aucune étude plus large de quantification des 5-MedC au sein des différentes espèces de drosophiles n'a, à ce jour, été réalisée. Afin de mieux comprendre l'importance et la fonction de la méthylation de l'ADN chez les drosophiles, nous avons quantifié les taux de 5-MedC des génomes séquencés de 12 espèces de drosophiles (Drosophila 12 genomes consortium 2007) pour lesquels la charge en ET est notée dans le Tableau 3 (p.64). Ce travail a été réalisé en collaboration avec Floriant Bellvert du Centre d'Etude des Substances Naturelles (CESN) de Lyon, dirigé par Gilles Comte.

Tableau 3. Proportion d'ET dans les 12 génomes séquencés de drosophiles (*Drosophila 12 genomes consortium, 2007*)

Espèce	% ET
<i>D. melanogaster</i>	5,35
<i>D. simulans</i>	2,73
<i>D. sechellia</i>	3,67
<i>D. yakuba</i>	12,04
<i>D. erecta</i>	6,97
<i>D. ananassae</i>	24,93
<i>D. pseudoobscura</i>	2,76
<i>D. persimilis</i>	8,47
<i>D. willistoni</i>	15,57
<i>D. virilis</i>	13,96
<i>D. mojavensis</i>	8,92
<i>D. grimshawi</i>	2,84

Les techniques de quantification des cytosines méthylées sont de différents types. Des techniques basées sur la combinaison de traitement au bisulfite avec du séquençage à haut débit permettent de cartographier des profils de méthylation tout le long des chromosomes et de calculer le nombre total de cytosines méthylées (Bibikova et Fan, 2010). Ces techniques, bien qu'automatisées et très informatives, sont toutefois très coûteuses et nécessitent un lourd travail bioinformatique d'annotation des séquences obtenues et sont surtout dépendantes des biais de conversion possibles. D'autres techniques d'identification basées sur la séparation des nucléosides selon leurs propriétés physico-chimiques comme les électrophorèses en capillaire à haute performance (HPCE) et les chromatographies liquides à haute performance (HPLC) peuvent également être utilisées, (Berdasco et al. 2009). Les appareils à HPCE et HPLC sont alors équipés d'un détecteur à barrette de diode (DAD) ou d'un spectromètre de masse (MS) afin d'identifier les composés analysés par chromatographie. C'est cette dernière méthode (HPLC couplée à un DAD et à la MS, l'appareillage étant alors noté HPLC/DAD/ESI-MS) que nous avons utilisée pour estimer les taux de 5-MedC dans les 12 génomes de drosophiles. Ainsi, l'HPLC sépare les molécules entre une phase stationnaire (la colonne) et une phase mobile (les solvants). La colonne utilisée ici est de type phase inversée C18 (chaîne de 18 atomes de carbone fixés sur des billes de silice). Les composés sont ainsi séparés dans une large gamme de polarité. La chromatographie utilisant cette colonne est une chromatographie dite « d'affinité ». Le détecteur DAD permet d'obtenir les

chromatogrammes à différentes longueurs d'ondes. Nos composés ayant une absorbance maximale à 280 nm, nous avons analysé les chromatogrammes à la longueur d'onde de 280 nm (Figure 21 A). Le DAD est très robuste au niveau quantitatif mais son seuil de détection n'est pas suffisant pour détecter les 5-MedC dans certains de nos échantillons. C'est pour cela que nous avons eu besoin de coupler à l'HPLC la spectrométrie de masse (MS).

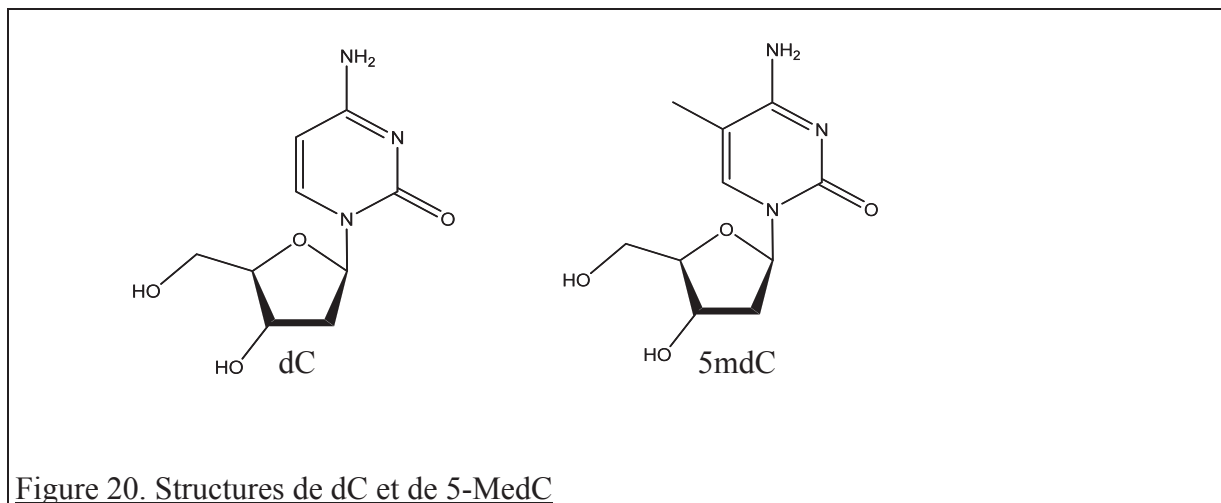


Figure 20. Structures de dC et de 5-MedC

1.2. Détection et quantification des 5-MedC par HPLC couplée à la spectrométrie de masse (MS)

La MS enregistre l'ionisation des molécules pour des masses allant de m/z 100 jusqu'à m/z 900 (ou m/z = masse/charge) et fournit un chromatogramme de courant ionique total (TIC) (Figure 21 B et Figure 21 C).

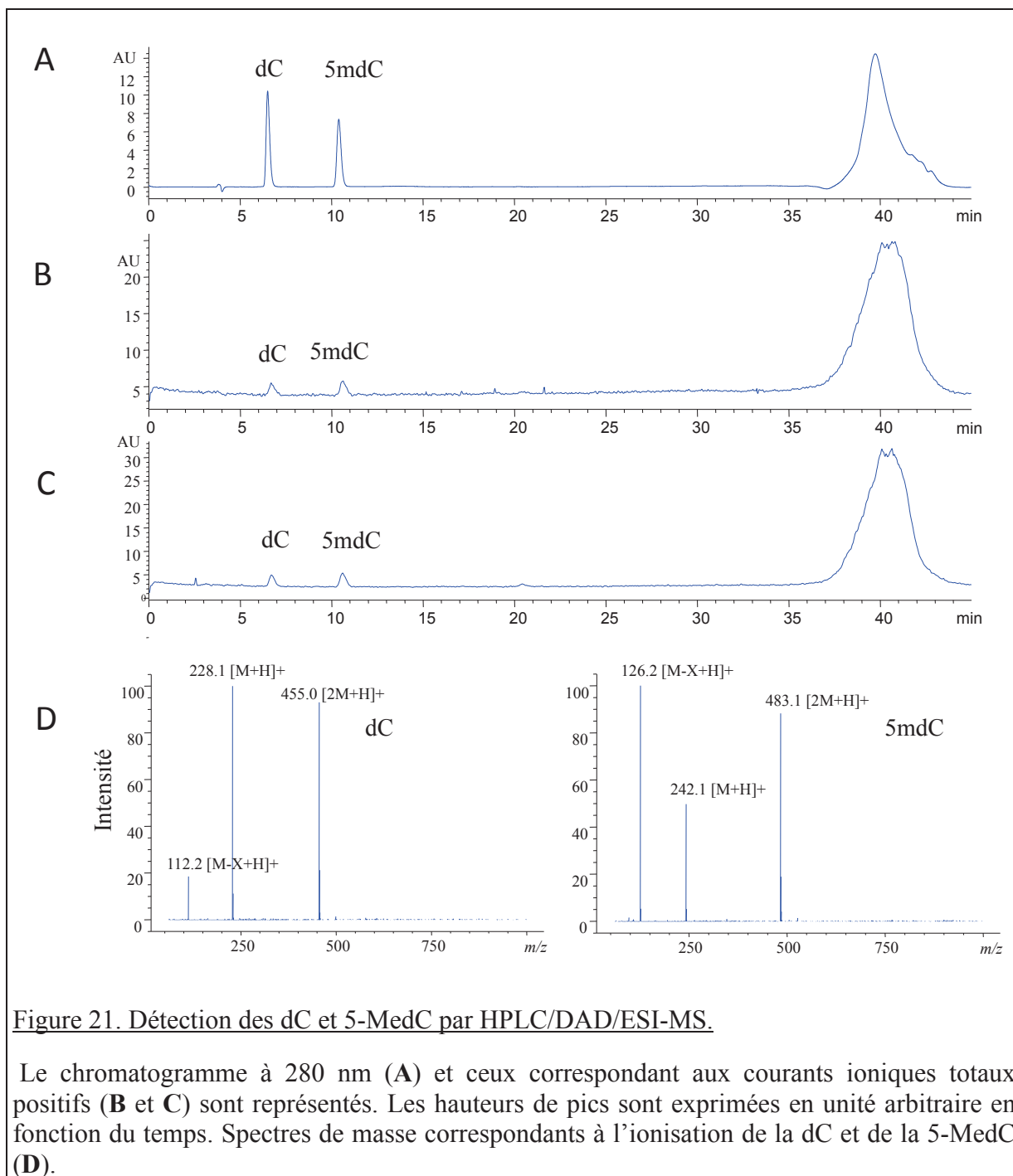


Figure 21. Détection des dC et 5-MedC par HPLC/DAD/ESI-MS.

Le chromatogramme à 280 nm (**A**) et ceux correspondant aux courants ioniques totaux positifs (**B** et **C**) sont représentés. Les hauteurs de pics sont exprimées en unité arbitraire en fonction du temps. Spectres de masse correspondants à l'ionisation de la dC et de la 5-MedC (**D**).

Lors de l'ionisation, il y a ajout d'un proton sur la molécule M qui devient alors un ion positif noté $[M+H]^+$. Un dimère de la molécule se forme alors, suivi de l'ajout d'un proton noté $[2M+H]^+$. Ces ions passent ensuite au niveau du fragmenteur qui engendre des cassures formant les ions fragments notés $[M-X+H]^+$ où, dans le cas de nos molécules, X correspond à l'unité de sucre. La proportion relative des trois formes $[M+H]^+$, $[2M+H]^+$ et $[M-X+H]^+$ dépend de l'intensité de l'ionisation et du fragmenteur. Ainsi, chaque molécule de dC et de 5-

MedC sera caractérisée par des fragments de masse spécifique (Figure 21 D). A partir du chromatogramme de courant ionique total, il est possible d'obtenir un chromatogramme correspondant uniquement au courant ionique d'une masse spécifique. Ces chromatogrammes sont appelés « chromatogrammes d'extraction d'ions ». La quantification des molécules est déterminée à partir des chromatogrammes TIC et des chromatogrammes d'extraction d'ions des masses des fragments spécifiques de la dC et de la 5mdC. La spécificité et la sensibilité de la quantification sont majorées en utilisant les chromatogrammes d'extraction d'ions. Les temps de rétention de la dC et de la 5-MedC sont déterminés grâce à l'injection de standards. Ils sont respectivement de 6 et 10 minutes dans nos conditions analytiques. Chacun de ces fragments, représentatif d'un composé, est utilisé pour quantifier ce composé à partir de l'aire sous la courbe définie par le pic sur le chromatogramme correspondant au temps de rétention du composé. Pour chaque détecteur et chaque molécule, nous avons établi des gammes étalon permettant de convertir l'aire sous la courbe en concentration, exprimée en nano molaire (nM) et ainsi faire le rapport des concentrations de ces deux composés présents dans le même échantillon. Le taux de 5-MedC dans les génomes analysés est défini comme suit : $\% \text{ 5-MedC} = \frac{[\text{5-MedC (nM)}]}{[\text{dC (nM)} + \text{5-MedC (nM)}]} * 100$. Les gammes réalisées en HPLC/DAD donnent des courbes de régression dont le coefficient de linéarité est très satisfaisant (0,999). En revanche, le seuil de détection des molécules ne permet pas de déceler de faibles concentrations en 5-MedC (<0,21 μM). Si les courbes des gammes obtenues à partir des chromatogrammes d'extraction d'ions ont des coefficients de linéarité moins bons, cette méthode a toutefois un seuil de détection plus bas. Parmi les ions-fragments représentatifs de la dC et de la 5-MedC, ceux qui permettent l'établissement de gammes étalon les plus linéaires (dont le coefficient de linéarité est supérieur à 0,99) sont respectivement les ions-fragments m/z 455 et m/z 483, correspondant tous deux aux formes $[2\text{M}+\text{H}]^+$ des composés. Les gammes établies pour la dC et la 5-MedC à partir de ces extractions d'ions, ainsi qu'à partir des chromatogrammes DAD, sont présentées sur la Figure 22. Nous avons testé la validité de notre méthode de quantification en mesurant les taux de méthylation de deux oligonucléotides de même séquence (). Le premier, l'oligonucléotide méthylé (OM), comporte cinq cytosines méthylées sur neuf. Le second est l'oligonucléotide non méthylé (ONM). Cette méthode de quantification a été validée par l'obtention des taux de 5-MedC attendus pour l'oligonucléotide méthylé ($\% \text{ 5-MedC} = 44\%$) ainsi que pour l'oligonucléotide non méthylé (pour lequel nous n'avons pas détecté de 5-MedC).

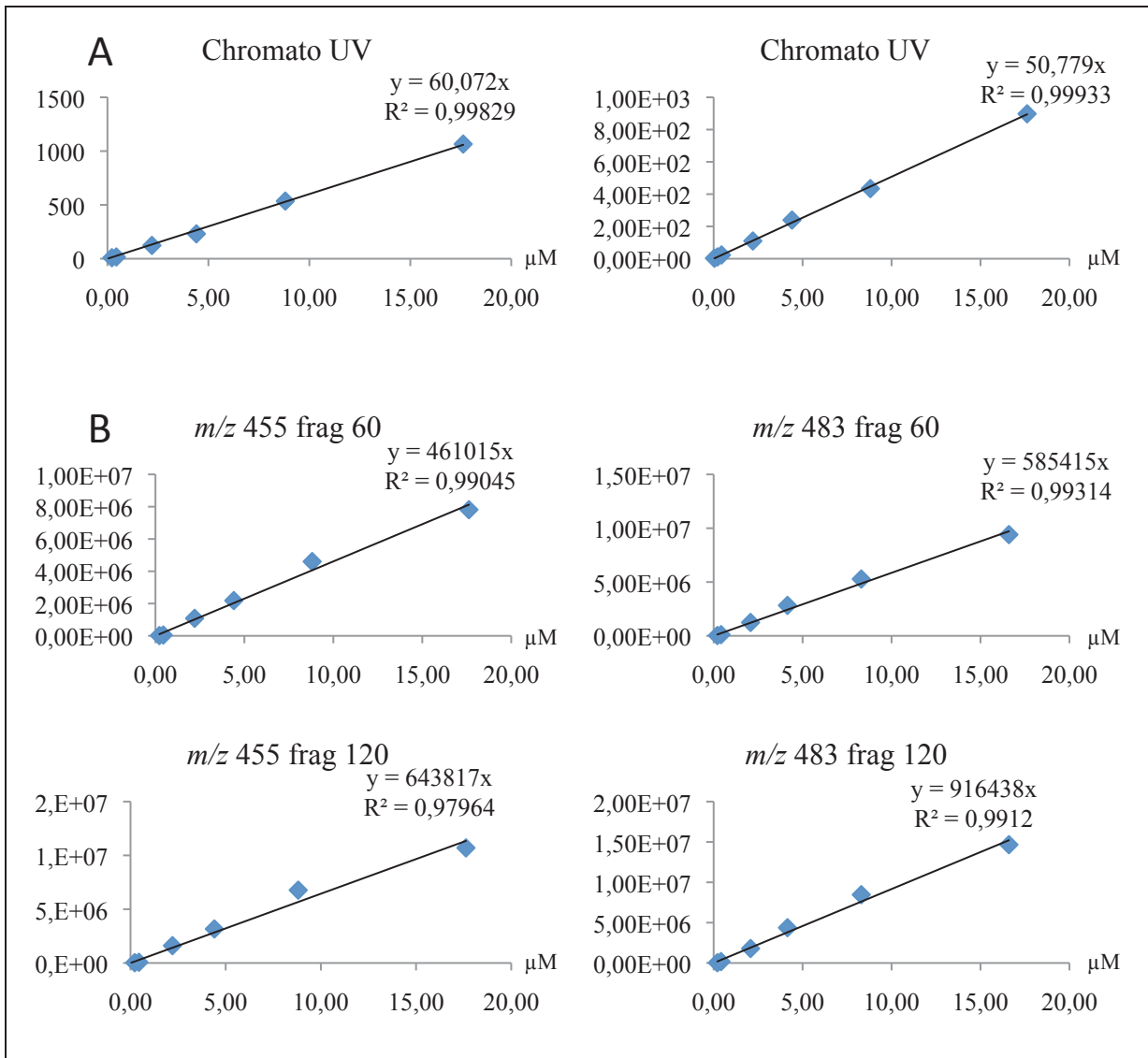
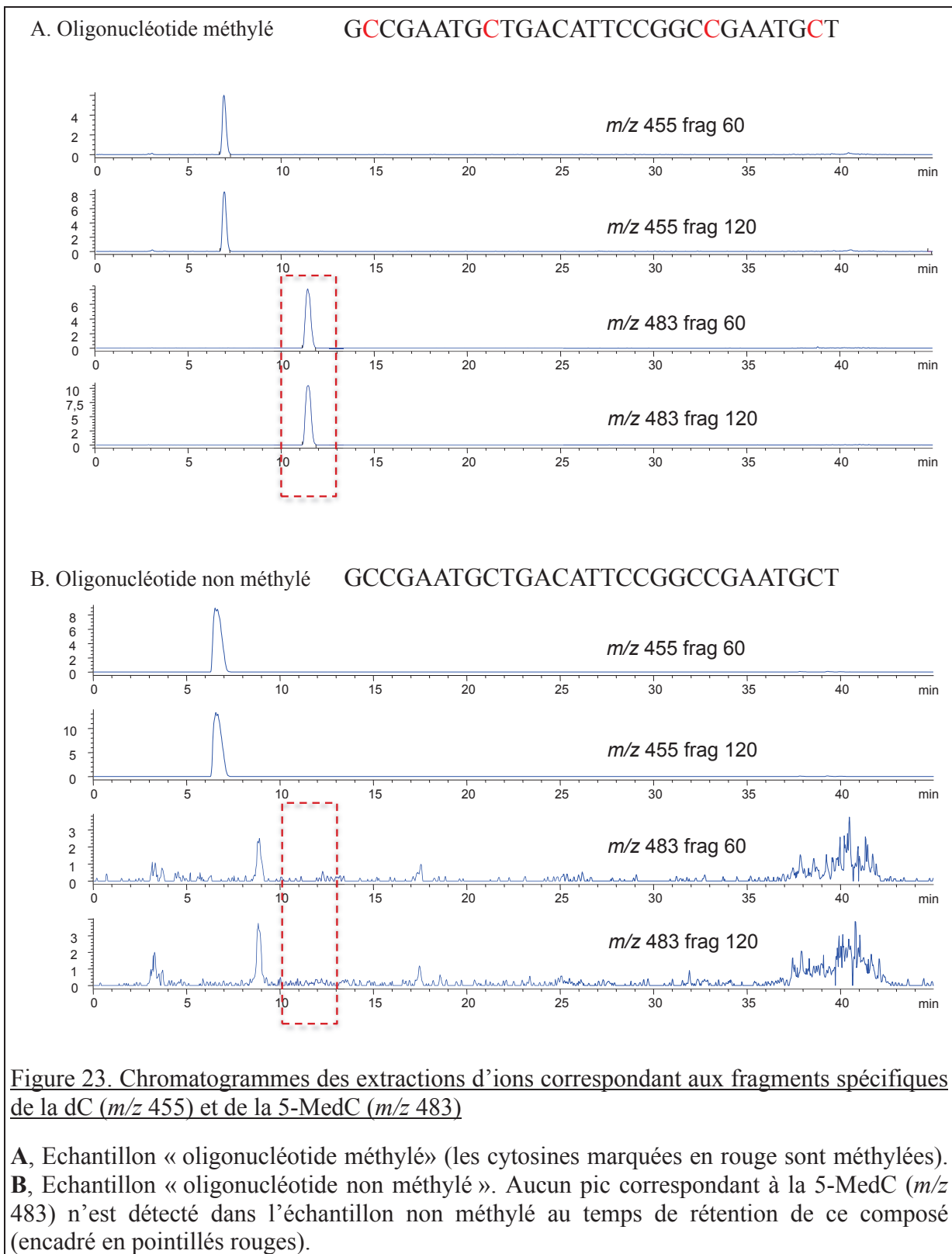


Figure 22. Gammes étalon utilisées pour la quantification des dC et 5-MedC

Ces gammes ont été établies en mesurant l'aire sous la courbe (axe des ordonnées) du pic correspondant aux fragments m/z 455 et m/z 483 (avec les fragmenteurs 60 et 120), à 280nm pour des concentrations croissantes en nM de dC et 5-MedC. Les coefficients directeurs de chacune des courbes ainsi que le coefficient de linéarité R^2 sont renseignés.



Afin d'évaluer la reproductibilité de notre méthode de quantification, nous avons ensuite réalisé six injections différentes, des jours différents, d'un mélange équimolaire de dC et 5-MedC, à des concentrations de 1000ng/mL (4,4 μ M) dont le taux de 5-MedC est de 50 %.

Les quantifications absolues des dC et 5-MedC sont mesurées pour chaque échantillon en DAD à 280 nm (Figure 24A) et en MS avec les ions-fragments 455, et 483 avec les fragmenteurs 60 et 120 (Figure 24B). La quantification absolue en DAD à 280 nm de ces échantillons est reproductible alors qu'elle présente une variabilité importante en MS. Néanmoins, ces deux outils de mesure sont robustes et hautement reproductibles pour l'étude de la quantification relative du taux de 5-MedC par rapport aux dC totales, (vérifié par ANOVA1, $F_{obs}=2,38$; $F_{seuil}=3,68$; $p=0,127$) (Figure 24C). Nous avons également quantifié les taux 5-MedC dans le thymus de veau et les cellules humaines Hela, fréquemment utilisés comme contrôle de quantification et pour lesquels les valeurs publiées sont respectivement 2,3 % et 6,4 % (Brown et al. 2007; Stach et al. 2003). Une autre étude a montré que le taux de méthylation de l'ADN de thymus de veau peut être compris entre 4,75 % et 8,1 % (Kok et al. 2007). Un groupe de scientifique anglo-saxon a publié dans « *Journal of Chromatography B* » la méthode de quantification des taux de 5-MedC par HPLC et spectrométrie de masse que nous étions en train de mettre au point (Sandhu et al. 2009). Les taux de 5-MedC mesurés pour les cellules Hela et le thymus de veau sont respectivement de 3,02 % et 6,26 %. Une nouvelle mesure des 5-MedC sera réalisée prochainement afin de vérifier ces résultats.

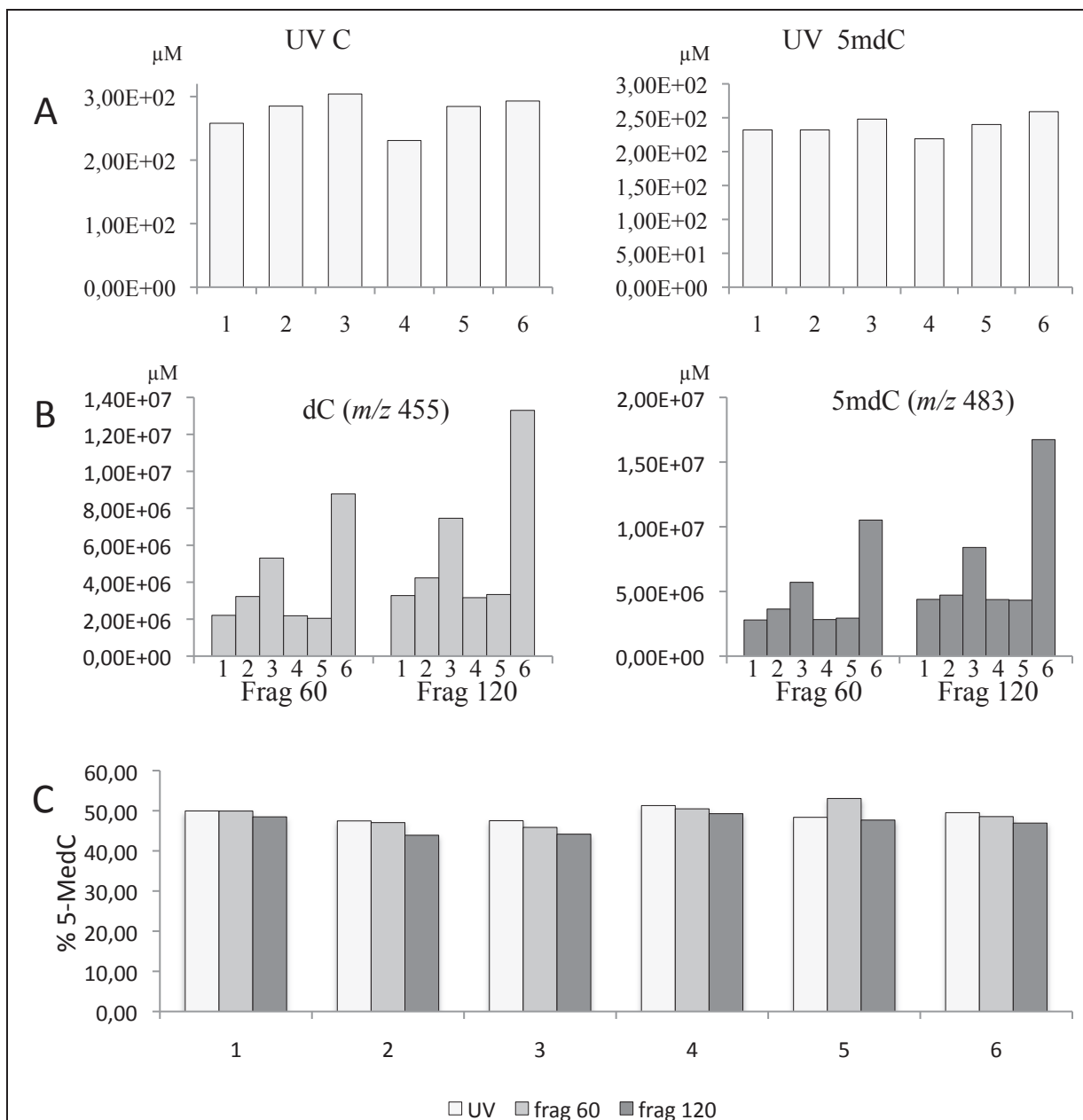


Figure 24. Tests de quantification de nucléosides standards en lecture UV et en spectrométrie de masse

Afin de vérifier la robustesse de la quantification et la reproductibilité des résultats, nous avons réalisé une série de 6 injections sur différents jours de la même solution équimolaire de dC et 5-MedC. Les quantifications absolues de la dC et de la 5-MedC ont été déterminées en lecture UV (A) et en spectrométrie de masse (B) en utilisant les ions fragments m/z 455 pour la dC et m/z 483 pour la 5-MedC (fragmenteurs 60 et 120). La quantification relative du taux de 5-MedC de la solution a été effectuée (C) avec les mesures déterminées en lecture UV et en spectrométrie de masse avec les coefficients directeurs correspondants.

1.3. Résultats

Les ADN des génomes des 12 espèces de drosophiles, fournis par le *Drosophila* STOCK CENTER de San Diego (Californie), sont extraits à partir d'embryons de moins de 16 heures. Nous avons ajouté à cette étude des ADN extraits à partir d'embryons de moins de deux heures des populations naturelles de Makindu (*D. simulans*) et du Sénégal (*D. melanogaster*). Les résultats des taux de méthylation, représentés sur la Figure 25, sont préliminaires et la réalisation d'un réplicat biologique permettra la validation statistique des résultats.

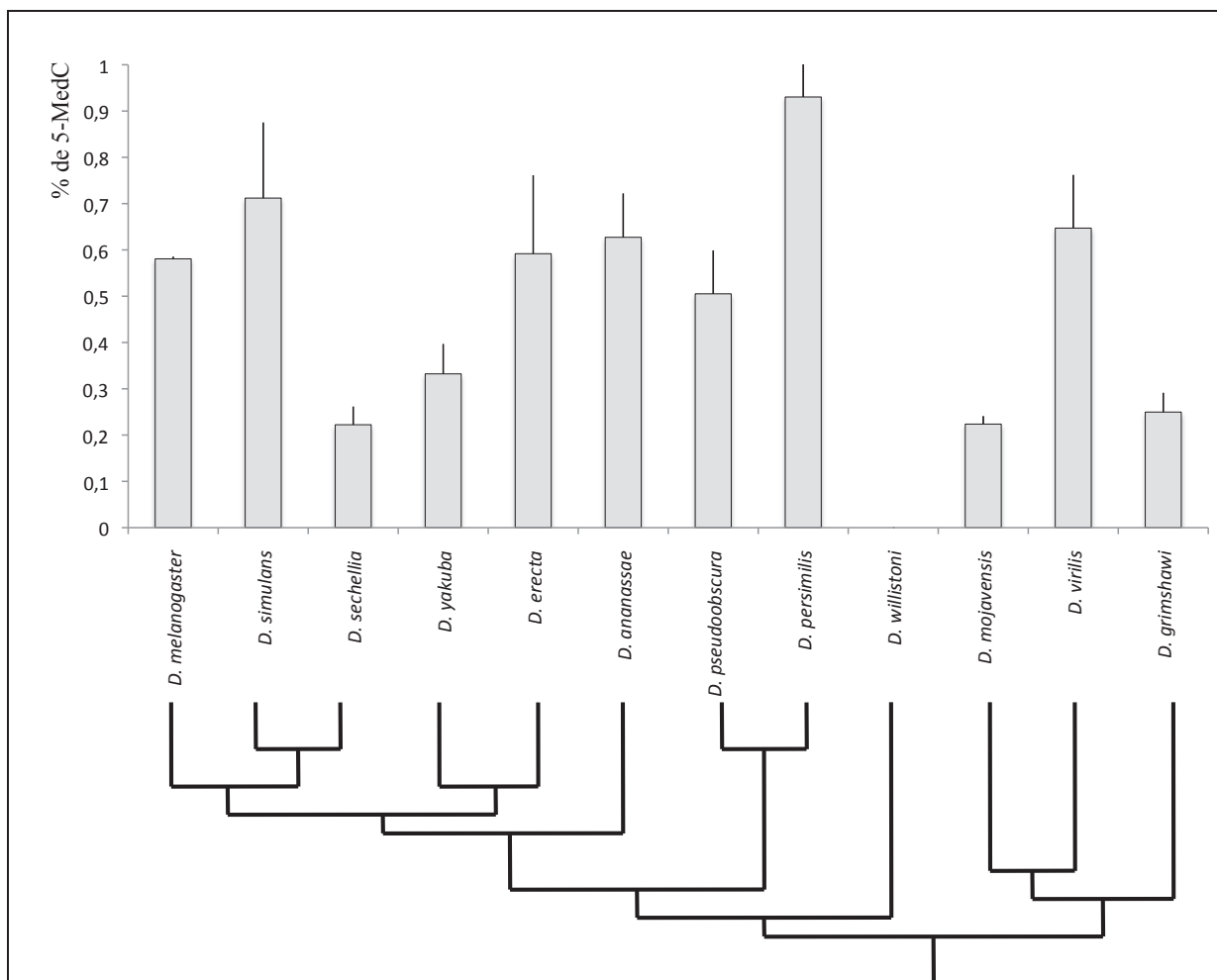


Figure 25. Détermination du taux de 5-MedC dans les génomes de drosophiles

La déviation standard, représentée pour chaque échantillon (barre verticale), a été établie à partir des valeurs obtenues par les fragmenteurs 60 et 120.

Nos résultats montrent que les taux de méthylation dans les génomes de drosophiles sont globalement faibles et inférieurs à 1 %. Les espèces *D. sechellia*, *D. yakuba*, *D. mojavensis* et *D. grimshawi* possèdent des taux de méthylation de l'ADN inférieurs à 0,50 % tandis qu'il est de 0,71 % pour *D. simulans*. Nos résultats ne sont pas en accord avec les taux de méthylation d'ADN d'embryons de moins de deux heures précédemment publiés (Marhold et al. 2004) pour les espèces *D. melanogaster* (0,37 %) et *D. pseudoobscura* (0,21 %) pour lesquelles nous observons respectivement 0,58 % et 0,50 % de 5-MedC. Aucun niveau de 5-MedC n'a pu être détecté pour l'espèce *D. willistoni*, pour laquelle la quantité en dC est équivalente à celle mesurée chez *D. ananasae* dont le taux de 5-MedC est pourtant de 0,67 %. La présence de méthylation sexe-spécifique au niveau des séquences répétées d'ADN ribosomiaux chez des adultes de *D. willistoni* a cependant été montrée par enzymes de restriction et Southern-blot (Garcia et al. 2007). Ainsi, les taux de méthylation de l'ADN des embryons de cette espèce sont probablement faibles et inférieurs au seuil de détection de notre technique.

Les taux de méthylation des embryons de moins de deux heures, obtenus pour les populations naturelles du Sénégal (*D. melanogaster*) et de Makindu (*D. simulans*), sont eux aussi très faibles. En effet, le taux de 5-MedC est de 0,04% pour la population naturelle de Makindu et non détectable pour celle du Sénégal. Les taux de 5-MedC mesurés sont largement inférieurs à ceux obtenus avec les génomes des souches séquencées des espèces correspondantes, dont les ADN ont été extraits à partir d'embryons âgés de moins de seize heures. Bien que ces derniers échantillons sont composés d'embryons d'âges différents, les taux quasi nuls de 5-MedC obtenus sur les populations naturelles ne sont pas en accord avec les données bibliographiques, selon lesquelles les taux de méthylation sont plus élevés dans les embryons de moins de deux heures que dans ceux de moins de seize heures (Lyko et al. 2000). De fait, nous pouvons alors envisager que les populations naturelles présentent des caractéristiques épigénétiques différentes de celles des souches de laboratoires quant à la méthylation de l'ADN. Une deuxième répétition de ces mesures, en cours, et permettra de vérifier ces résultats.

2. Exploration de la présence de cytosines méthylées au niveau des ET dans des populations naturelles de *D. melanogaster* et *D. simulans*

2.1. Utilisation de couples d'isoschizomères pour déterminer la présence de cytosines méthylées

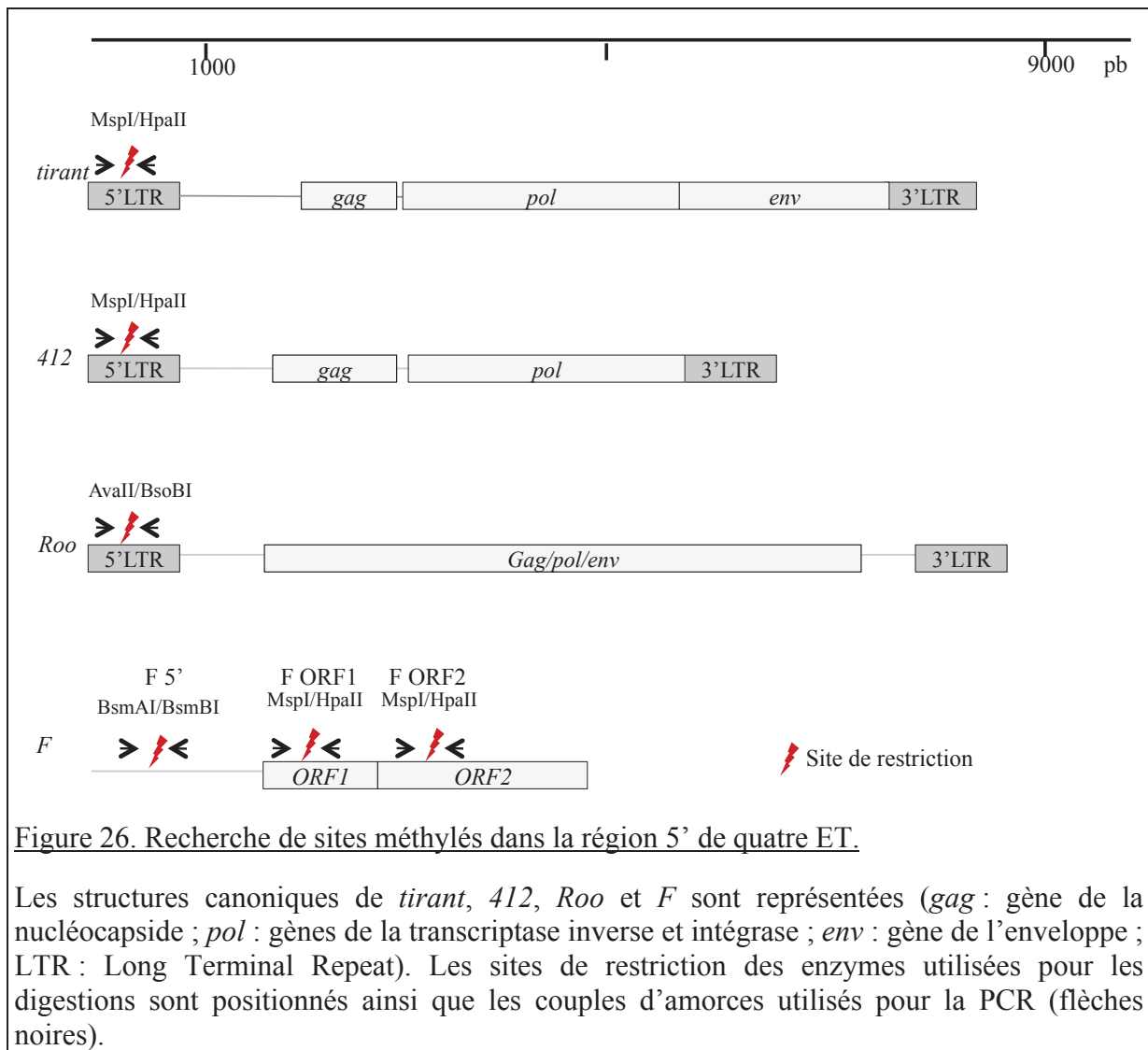
Dans certains génomes comme celui d'*A. thaliana*, les ET sont une cible privilégiée des DNMT, et plus globalement des systèmes de régulation épigénétique impliqués dans leur régulation et le maintien de leur stabilité. Il a été observé chez la drosophile que des séquences d'ET sont retrouvées dans des fractions enrichies en 5-MedC par immunoprécipitation de l'ADN méthylé (Salzberg et al. 2004). Le taux de méthylation chez *D. simulans* par HPLC et spectrométrie de masse étant de 0,70 %, nous avons voulu savoir si les ET y étaient méthylés dans la lignée germinale des populations naturelles de cette espèce. Pour cela, j'ai utilisé une approche par enzymes de restriction, dans le but de mettre en évidence la présence de sites de restriction méthylés dans des séquences d'ET par la technique de MSRE combinée à de la PCR. Les endonucléases sont des enzymes possédant un site de restriction propre, l'activité de clivage du site de restriction par l'enzyme pouvant, dans certains cas, être inhibée. Ceci est notamment le cas lorsqu'une ou plusieurs cytosines de ce site sont méthylées. L'utilisation d'enzymes clivant le même site de restriction (ou isoschizomères) est particulièrement intéressante puisque certaines d'entre elles sont sensibles, ou non, à la présence de 5-MedC au site de restriction. Ainsi, il est possible d'identifier un site de restriction localisé dans la région régulatrice d'un ET, associé à deux isoschizomères dont l'un sera sensible à la présence de 5-MedC et l'autre pas. Des amorces de PCR positionnées de part et d'autre du site de restriction permettent de vérifier l'amplification dans les cas de digestion avec une enzyme sensible et une enzyme non sensible. J'ai étudié les rétrotransposons à LTR de références *Roo*, *tirant*, et *412* ainsi que l'élément LINE *F*. Les populations étudiées proviennent du Sénégal et de Chicharo (Portugal) pour l'espèce *D. melanogaster* et de Chicharo, Makindu (Kenya), Zimbabwe, Grand Ferrade, Papetee, Canberra, Réunion, Can River, Amieu et Anjiro pour l'espèce *D. simulans*. Pour toutes ces populations, les nombres de copies euchromatiques des quatre ET analysés ont été estimés par hybridation *in situ*, et sont représentés dans le Tableau 4 (p.75).

Tableau 4. Nombre moyen de copies des éléments *F*, *Roo*, *412* et *tirant* déterminé par hybridation *in situ* dans les populations naturelles de *D. melanogaster* et *D. simulans*.

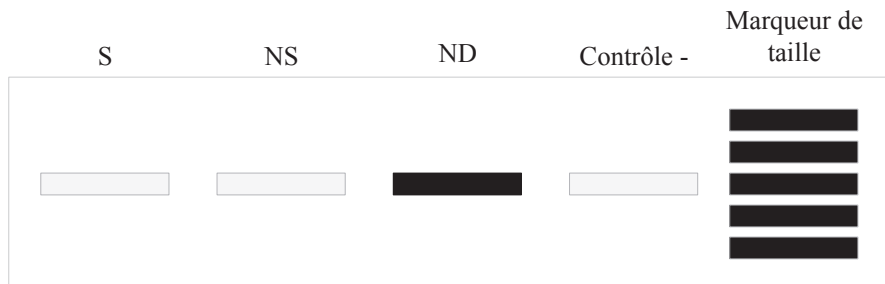
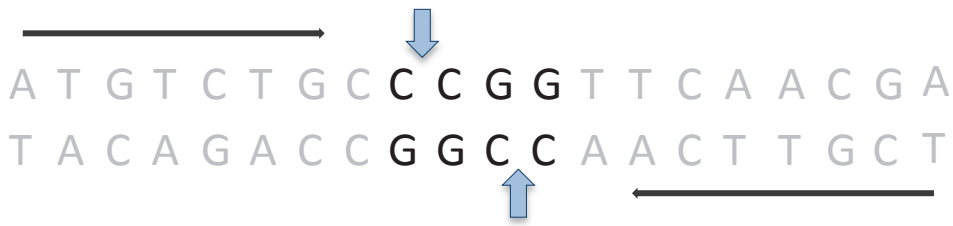
	<i>D. melanogaster</i>		<i>D. simulans</i>									
	Sén	Chi	Chi	Mak	Zim	Gra	Pap	Can	Réu	CaR	Ami	Anj
<i>F</i>	16	32,5	1	0	0	37	13/0	10,5	0	3	1	ND
<i>Roo</i>	44,5	65	40	40,5	45	41,5	33,5	47,5	43,5	40	28	ND
<i>412</i>	24,5	33,5	16,5	2,5	9	11	20	58	5,5	5	7	ND
<i>Tirant</i>	9	15	0	4	2,5	0	2	0,5	2	0	0	ND

Sén: *Sénégal*, Chi: *Chicharo*, Mak: *Makindu*, Zim: *Zimbabwe*, Gra: *Grand Ferade*, Pap: *Papetee*, Can: *Canberra*, CaR: *Can River*, Ami: *Amieu*, Anj: *Anjiro*.

A partir des séquences de référence des quatre ET établies chez *D. melanogaster* et *D. simulans*, j'ai déterminé dans leur région 5', des sites de restrictions pouvant être clivés par des enzymes sensibles ou non à la méthylation via le logiciel NEB CUTTER (<http://tools.neb.com/NEBcutter2/>; (Vincze et al. 2003). La structure canonique de chacun de ces ET est représentée sur la Figure 26 sur laquelle sont également positionnés les sites de restriction des enzymes ainsi que les amorces utilisées pour la PCR. Puisque la méthylation de l'ADN au niveau des régions codantes des gènes a été mise en évidence chez *D. melanogaster* (Salzberg et al. 2004), j'ai intégré à mon étude des sites de restriction localisés dans les parties codantes de l'élément *F*. Pour les éléments *tirant* et *412*, ainsi que pour les sites *F ORF1* et *F ORF2* localisés dans les parties codantes de *F*, les couples *MspI/HpaII*, dont *HpaII* est l'enzyme sensible, ont été utilisés. Pour les éléments *Roo* et la région 5' de *F*, les couples *AvaII/BsoBI* et *BsmAI/BsmBI* ont été respectivement utilisés, *AvaI* et *BsmBI* étant les enzymes sensibles. Les sites de restriction de ces enzymes contiennent tous un CpG sur lequel la présence d'une méthylation empêchera l'enzyme sensible de cliver le site. Un schéma illustrant cette technique avec les enzymes *MspI* et *HpaII* est représenté sur la Figure 27 et les sites de restriction de chaque enzyme sont détaillés dans la partie Matériels et Méthodes.



A: site non méthylé



B: site méthylé

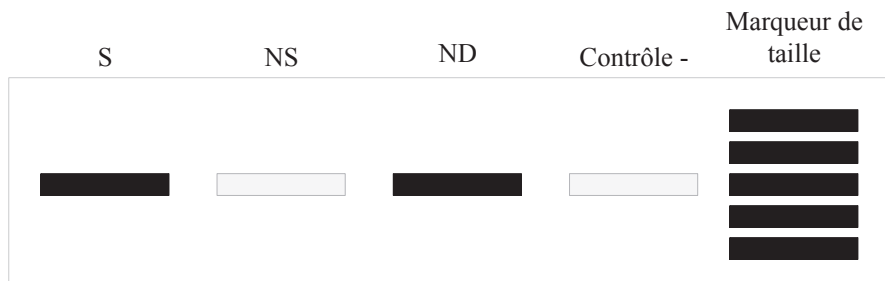
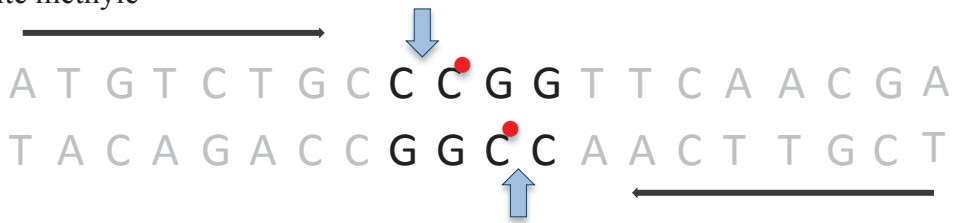
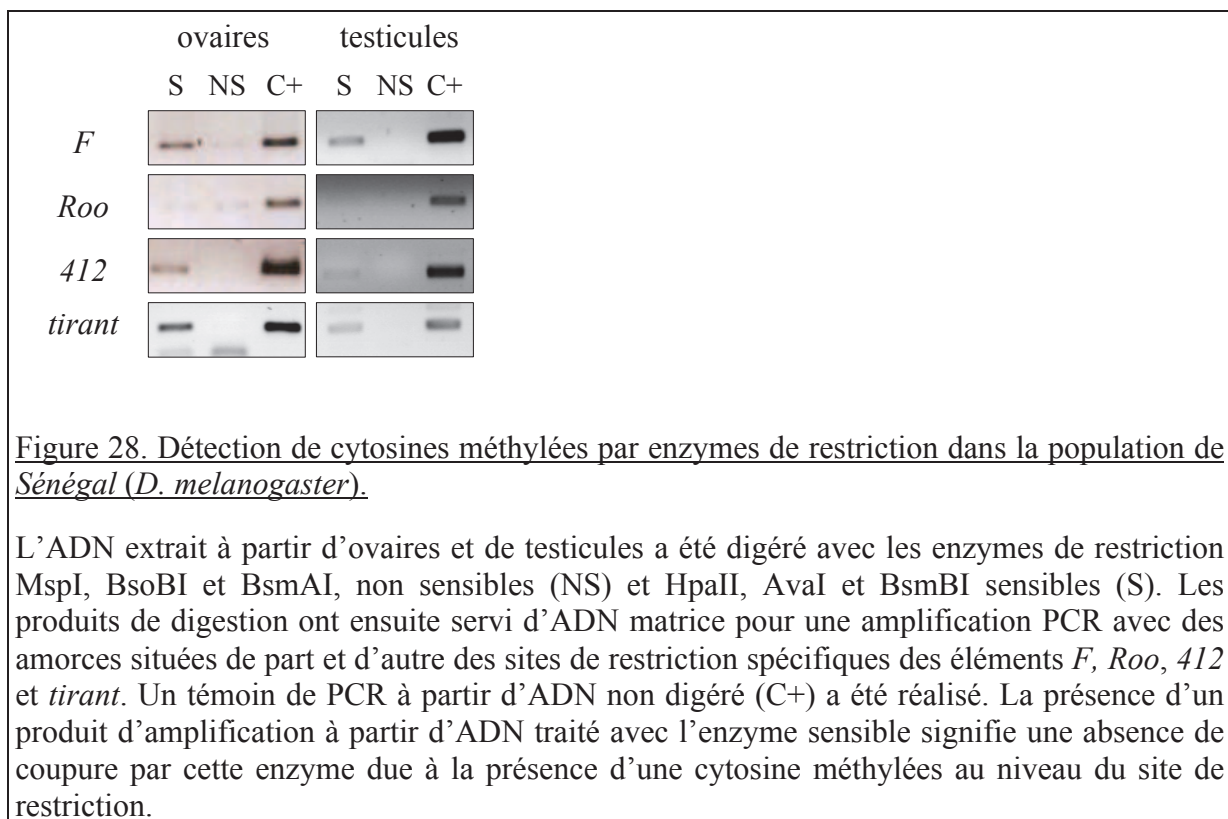


Figure 27. Illustration de la technique de MSRE.

A: un fragment d'ADN non méthylé. **B:** un fragment d'ADN méthylé, au niveau du dinucléotide CpG du site de restriction du couple MspI/ HpaII dont la séquence nucléotidique est CCGG. L'enzyme sensible (S) est HpaII et l'enzyme non sensible (NS) est MspI. Après digestion, une amplification PCR avec les amorces situées de part et d'autre du site de restriction permet de visualiser sur gel d'agarose si le fragment est clivé (pas d'amplification = rectangle vide) ou non (amplification = rectangle plein). Le contrôle non digéré (ND) permet de visualiser la bande attendue si le fragment n'est pas clivé, et le contrôle négatif (contrôle -) permet de vérifier que la digestion est totale.

2.2. Résultats

L'amplification par PCR des fragments correspondant aux ET analysés diagnostique la présence ou non de cytosines méthylées au niveau du site de restriction des enzymes utilisées. Les produits de PCR sont déposés sur gel d'agarose, comme représenté sur la Figure 28 pour la population du *Sénégal* (*D. melanogaster*). Les résultats montrent que l'enzyme sensible n'a pas clivé les sites de restriction correspondant aux régions 5' des éléments *F* et *tirant*, dans les ovaires et les testicules, ce qui signifie que ces sites de restriction sont méthylés dans ces tissus. L'élément *Roo* n'est pas méthylé dans les ovaires et testicules de cette population.



Le polymorphisme nucléotidique qui peut exister au sein de la séquence des ET conduit à des digestions partielles si les sites polymorphes sont localisés au niveau des sites de restriction des enzymes utilisés. Afin de contrôler la digestion de ces sites dans toutes les populations, j'ai vérifié la digestion par les enzymes non sensibles à la méthylation de l'ADN des produits de PCR de chaque couple d'amorces utilisées. Dans tous les cas, les produits de PCR sont digérés, signifiant qu'aucun polymorphisme n'existe dans les échantillons au niveau des sites étudiés.

Dans la suite de mon manuscrit, l'ensemble des résultats sera simplifié et chaque site sera représenté par un cercle rempli en noir s'il est méthylé, et vide s'il ne l'est pas. Ces résultats sont représentés sur la Figure 29. De nombreux sites méthylés ont pu être mis en évidence par cette méthode dans les populations de *D. melanogaster* et *D. simulans* pour l'ensemble des ET étudiés à l'exception de *Roo*. En effet, pour cet ET, seul un site méthylé a pu être mis en évidence dans l'ADN des ovaires de la population de Chicharo (*D. simulans*).

Pour *F* et *412*, les sites de restriction analysés sont méthylés dans les ovaires dans un grand nombre de populations. En revanche, dans les testicules, ces mêmes sites ne sont pas méthylés dans la plupart des populations (sauf Sénégal (*D. melanogaster*) et Amieu (*D. simulans*)).

Ovaires	<i>F</i>	<i>Roo</i>	<i>tirant</i>	<i>412</i>
<i>D. melanogaster</i>				
Sénégal	●	○	●	●
Chicharo	●	○	●	○
dDNMT2-/-	○	○	●	●
<i>D. simulans</i>				
Chicharo	○	●	●	○
Makindu	●	○	●	○
Zimbabwe	●	○	●	●
Grand Ferrade	●	○	○	○
Papeete	●	○	●	●
Canberra	●	○	●	○
Réunion	●	○	●	●
Can river	●	○	●	●
Amieu	●	○	●	●
Anjiro	●	○	●	○
Testicules	<i>F</i>	<i>Roo</i>	<i>tirant</i>	<i>412</i>
<i>D. melanogaster</i>				
Sénégal	●	○	●	○
Chicharo	○	○	●	○
dDNMT2-/-	○	○	●	○
<i>D. simulans</i>				
Chicharo	○	○	●	○
Makindu	○	○	●	○
Zimbabwe	○	○	○	○
Grand Ferrade	○	○	○	○
Papeete	○	○	○	○
Canberra	○	○	○	○
Réunion	○	○	○	○
Can river	○	○	●	○
Amieu	○	○	●	●
Anjiro	○	○	○	○

Figure 29. Mise en évidence de sites méthylés dans les régions 5' de quatre ET de référence

Pour *tirant*, le site de restriction est méthylé au niveau des ovaires dans toutes les populations sauf Grand Ferrade (*D. simulans*) et au niveau des testicules dans les populations de Sénégal et Chicharo de *D. melanogaster*, et Chicharo, Makindu, Can River et Amieu de *D. simulans*.

Les sites de restriction localisés dans les LTR 5' des éléments *Roo*, *412* et *tirant*, ainsi que dans la région non codante 5' de *F* sont donc méthylés dans la plupart des populations au niveau de la lignée germinale femelle. Ces mêmes sites ont aussi été identifiés comme méthylés dans la lignée mutante dDNMT2^{-/-}, utilisée comme contrôle.

Les données de BGS obtenues lors de l'analyse de *tirant* sont en contradiction avec nos données obtenues par enzyme de restriction. En effet, le site de restriction qui apparaît méthylé n'est pas retrouvé dans les résultats présentés dans le premier chapitre de ma thèse, concernant les profils de méthylation de la LTR5' de *tirant* dans les ovaires et les testicules de la population de Makindu.

La méthode de détection par simple PCR des fragments non digérés par l'enzyme sensible peut amplifier des séquences minoritaires non retrouvées dans les clones séquencés du premier chapitre. Afin de vérifier cela pour les autres ET, j'ai déterminé par BGS les profils de méthylation des régions 5' de *F*, *Roo* et *412* comprenant les sites de restriction étudiés, à partir d'ADN d'ovaires de la population de Makindu (Figure supplémentaire 4, Figure supplémentaire 5, Figure supplémentaire 6). Les sites de restriction analysés étant sensibles à la méthylation des cytosines des dinucléotides CpG, j'ai calculé la fréquence de méthylation des cytosines en fonction de leur contexte dinucléotidique à partir des profils de méthylation de *F*, *Roo* et *412* (Figure 30). Le dinucléotide CpG du site de restriction des enzymes MspI et HpaII (CCGG) a été retrouvé méthylé dans un clone sur 29 pour l'élément *412*. Aucune méthylation des sites de restriction analysés pour *F* et de *Roo* n'a pu être identifiée par BGS. Les cytosines méthylées de la région 5' de *F* et *412* concernent tous les dinucléotides avec une préférence au niveau des dinucléotides CpA et CpG pour *F* et CpA, CpG et CpT pour *412*. Les cytosines méthylées de la région 5' de *Roo* sont exclusivement localisées au niveau de dinucléotides CpA et CpT, pour lesquels aucune méthylation symétrique n'a été identifiée par bisulfite. Ce résultat est cohérent avec les résultats obtenus avec l'enzyme AvaI, sensible à la présence de CpG méthylés.

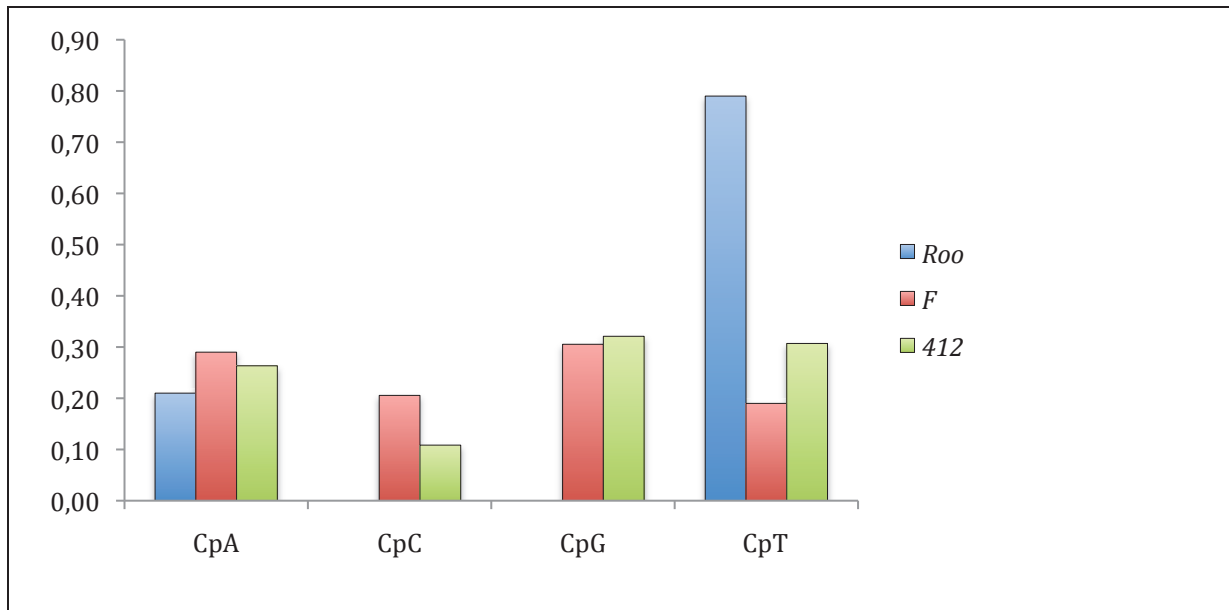


Figure 30. Fréquence de dinucléotides méthylés dans les régions 5' de *Roo*, *F* et *412*

ADN extrait à partir d'ovaires de la population de Makindu (*D. simulans*), puis traité au bisulfite de sodium.

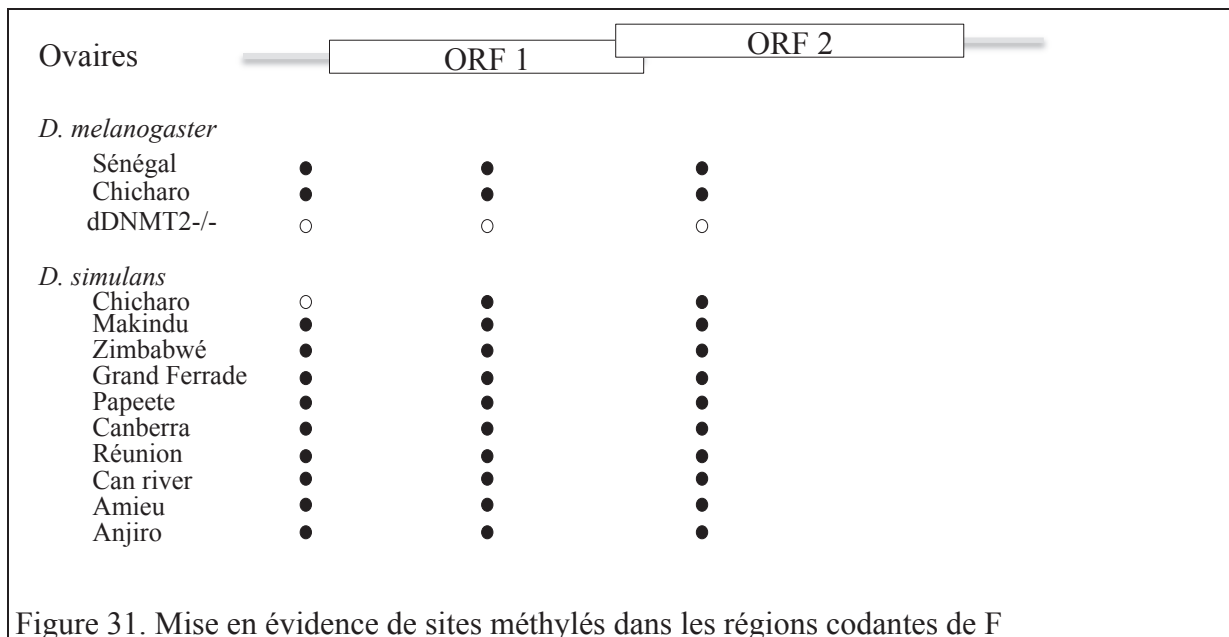


Figure 31. Mise en évidence de sites méthylés dans les régions codantes de F

Enfin, les sites F ORF1 et F ORF2 localisés dans les parties codantes de l'élément F sont méthylés dans l'ensemble des populations analysées (Figure 31) sauf dans la lignée mutante dDNMT2-/-.

2.3. Conclusion

La méthylation de l'ADN chez les 12 génomes de drosophiles est faible avec un niveau de 5-MedC inférieur à 1%. Nous ne pouvons pas établir de corrélation entre la proportion en 5-MedC d'un génome et sa quantité en ET, ce qui suggère que la méthylation des cytosines chez la drosophile n'est pas principalement dirigée contre les ET. Chez *D. simulans*, nous avons cependant mis en évidence des cytosines méthylées au niveau de dinucléotides CpG pour les régions 5' de *F*, *412* et *tirant*. Ce travail exploratoire doit être approfondi et il est nécessaire de pouvoir vérifier ces résultats par des techniques différentes. Pour cela, plusieurs méthodes sont envisageables. Il serait très enrichissant de réaliser des « Methyl DNA ImmuniPrécipitation » (MeDIP) dans les populations naturelles de *D. simulans* et de cloner et séquencer l'ADN enrichi en 5-MedC. Aussi, il serait opportun de déterminer le profil de méthylation sur tout le long de la séquence d'un ET, afin de voir si des profils spécifiques de méthylation peuvent être mis en évidence au niveau des cytosines dans les parties codantes de ces ET. Les résultats obtenus au niveau des parties codantes de l'élément F favorisent l'hypothèse d'une méthylation de l'ADN à l'intérieur des gènes, et cela devra être étudié pour les autres ET. Enfin, l'existence d'une différence de méthylation entre les lignées germinales mâle et femelle est à confirmer.

DISCUSSION - PERSPECTIVES

1. Problèmes d'interprétation des données produites par bisulfite

Dans divers organismes, de nombreuses études utilisant le traitement au bisulfite ont mis en évidence la présence de « Densely Methylated DNA Island » (DMI), qui présentent des taux de méthylation supérieur à 50% (Wang et al. 2001; Tasheva et Roufa, 1995; Tasheva et Roufa, 1994). Dans le cadre de l'étude de la régulation somatique du rétrotransposon *Invader 4* chez *D. melanogaster*, des séquences fortement méthylées, correspondant à la LTR5' de cet élément, ont été rapportées (Phalke et al. 2009). Ces résultats, obtenus par bisulfite sur de l'ADN extrait à partir d'embryons de moins de deux heures, montrent des profils de méthylation des cytosines sans préférence dinucléotidique, avec des taux de 5-MedC très élevés (>50%). Les amorces utilisées pour l'amplification après traitement ne sont pas dégénérées et sont donc favorables à l'amplification préférentielle des séquences non converties. Les conditions de traitement, caractérisées par une courte phase de déamination d'une heure avec une solution de bisulfite de sodium de faible molarité (3,1 M), ont été remises en question quelques mois plus tard (Schaefer et Frank Lyko, 2010). De plus, il a été montré que la conversion des cytosines non méthylées par des solutions de bisulfite de sodium de faible molarité (< 5,5 M) nécessite une phase de déamination plus longue pour obtenir une efficacité de conversion de l'ordre de 95 % (Genereux et al. 2008). Ainsi, j'ai considéré que de telles séquences, obtenues durant ma thèse, devaient être considérées comme des artefacts, les expériences complémentaires réalisées par la suite ne montrant pas un tel type de profil de méthylation. Chez *A. thaliana*, il est admis que les taux de méthylation sont élevés et aussi bien symétriques qu'asymétriques. Récemment, Slotkin et al (2009) proposent à ce titre un modèle de la régulation des ET dans le pollen reposant principalement sur la présence de ces séquences fortement méthylées. L'expression des ET dans le pollen est supérieure à celle trouvée dans les autres tissus de la plante et est associée à une déméthylation de leurs régions régulatrices. Le grain de pollen contient un noyau végétatif et un noyau germinal (sperme), et les résultats de ces auteurs montrent que les séquences d'ET correspondant au noyau végétatif présentent une déméthylation des trinuécléotides CHH, tandis que les séquences d'ET du sperme sont hyperméthylées au niveau de ces mêmes positions. Les auteurs proposent *via* leur modèle que les transcrits d'ET produits dans le noyau végétatif sont clivés en petits ARN allant cibler la méthylation des sites correspondant dans le génome du sperme et ainsi réguler l'expression des ET. Ce modèle repose en partie

sur des résultats de bisulfite montrant de longues chaînes de cytosines méthylées sans préférence dinucléotidique sur les brins sens et antisens de fragments d'ADN du noyau germinal. Ces profils de méthylation ressemblent fortement à ceux mis en évidence dans mes résultats préliminaires sur les populations de Makindu et Zimbabwe de *D. simulans* et doivent être interprétés avec précaution. En effet, dans les travaux de cartographie de la méthylation de l'ADN chez *A. thaliana*, il a été considéré que les séquences contenant plus de trois cytosines méthylées consécutives dans un contexte CHH sont non converties (Cokus et al. 2008). Dans cette étude, l'efficacité du traitement au bisulfite est estimée à 99,97 %, en prenant comme référence le taux de conversion du génome non méthylé du chloroplaste. Malgré toutes ces précautions, des séquences fortement méthylées ont été obtenues, dont le profil de méthylation présente une périodicité de 150 pb conservée entre les clones. Cette périodicité est définie comme l'alternance de régions fortement méthylées et de régions plus faiblement méthylées. Selon ces auteurs, cette distance de 150 pb correspond à la distance entre deux nucléosomes qui seraient positionnés au niveau des régions faiblement méthylées, et l'ADN enroulé autour du nucléosome serait moins accessible aux DNMT (Cokus et al. 2008). Pour les populations de Makindu et Zimbabwe, les séquences non converties ont des profils fortement corrélés les uns avec les autres (Figure supplémentaire 1, p.159), ce qui suggère que ces profils ne sont pas dus à des artefacts aléatoires le long de la séquence d'ADN, mais plutôt à des composants moléculaires ou protéiques qui empêcheraient la conversion des cytosines non méthylées à certaines positions. Ces profils présentent également une périodicité de l'ordre de 150 pb comparable à celle observée dans les séquences fortement méthylées du génome d'*A. thaliana* (Cokus et al. 2008). De plus, c'est dans la population de Makindu, dans laquelle *tirant* est transcrit, que la proportion de séquences non converties est la plus forte. La relation entre l'ARN et la méthylation de l'ADN fait intervenir le mécanisme de « RNA directed DNA Methylation » (RdDM) chez la plante, et permet la mise en place de profils de méthylation spécifiques au niveau des séquences d'ET hétérochromatiques (Mathieu et Bender 2004). Cependant, rien n'explique de manière évidente comment le mécanisme de RdDM permet la méthylation de toutes les cytosines d'une séquence. Toujours chez la plante, et après avoir testé plusieurs méthodes différentes de traitement au bisulfite, d'autres auteurs ont réalisé deux traitements consécutifs permettant d'obtenir une efficacité de conversion de 99,14 % (taux de conversion du génome du chloroplaste) (Lister et al. 2008). Dans cette étude, pour laquelle les petits ARN ont été totalement séquencés, les auteurs ont mis en évidence la correspondance entre les séquences de ces derniers et les séquences fortement méthylées avant d'en conclure que les petits ARN

sont impliqués dans la mise en place de ces profils de méthylation. Durant ma thèse, j'ai mis en avant que l'absence de séquences non converties était corrélée avec l'absence d'hétéroduplexes ARN/ADN. Le traitement au bisulfite de sodium convertit les cytosines des molécules d'ADN dénaturées. Par conséquent, les homoduplexes ADN/ADN ainsi que les hétéroduplexes ARN/ADN peuvent empêcher la conversion des cytosines de ces séquences. Le traitement au bisulfite se fait dans un thermocycleur à une température de 64°C et favorise ainsi des hybridations ADN/ARN sur certaines régions. Un article sur la présence d'hétéroduplexes ARN/ADN et leur implication dans les artefacts liés au traitement au bisulfite est actuellement en préparation.

Henderson et al (2010) a publié un article sur les conditions expérimentales à appliquer pour obtenir des résultats fiables issus de traitements au bisulfite de sodium chez les plantes, dans lequel il rappelle que de longues chaînes de cytosines consécutives méthylées correspondent à l'amplification par PCR d'ADN non converti. Ainsi, tout résultat obtenu après traitement au bisulfite qui montre de fortes quantités en 5-MedC dans une séquence doit être éprouvé par une autre approche permettant de vérifier ces profils, notamment l'utilisation d'enzymes de restriction. Pour ce faire, j'ai donc vérifié cela par l'utilisation d'enzymes de restriction sensibles et non sensibles à la méthylation de l'ADN. Cette approche, qui m'a permis de montrer que *tirant* n'est pas hyperméthylé dans les populations naturelles de *D. simulans*, a révélé que le génome de ces populations n'est pas dépourvu de 5-MedC. En effet, de nombreux sites de restriction localisés dans des séquences d'ET sont méthylés dans un grand nombre de populations. De plus, l'identification de certains sites de restriction méthylés dans la lignée mutante dDNMT2 suggère que d'autres enzymes capables de méthyler les cytosines de l'ADN peuvent être présentes chez cette espèce.

Récemment, la spectrométrie de masse a permis la découverte d'une nouvelle modification épigénétique associée aux cytosines. Les 5-hydroxyméthylcytosines (5-OHMedC) résultent de l'ajout d'un groupe hydroxyle sur les 5-MedC par les enzymes de la famille Tet, très conservées chez les mammifères. Les 5-OHMedC sont particulièrement enrichies dans les cellules de Purkinje du cerveau (Kriaucionis et Heintz, 2009), et absentes des cellules cancéreuses. Des mutations dans des protéines de la famille Tet ont de plus été associées à des leucémies myéloïdes (Tahiliani et al. 2009), suggérant ainsi que les 5-OHMedC influeraient la régulation de certains gènes. Ces 5-OHMedC, contrairement aux 5-MedC, ne permettent pas la fixation de la « Methyl Binding Protein » MeCP2 (Valinluck et

al. 2004). Enfin, les 5-OHMedC sont converties de la même manière que les 5-MedC par le bisulfite de sodium. Cette technique ne distingue donc pas les 5-MedC des 5-OHMedC dans une séquence (Huang et al. 2010). Bien qu'un orthologue de la famille TET ait été identifié chez *D. melanogaster* (Tahiliani et al. 2009), nous n'avons pas détecté de 5-OHMedC par spectrométrie de masse dans toutes les espèces pour lesquelles nous avons estimé le taux de méthylation. Nous suggérons donc que cette modification des 5-MedC n'existe pas chez la drosophile, ou à des niveaux très faibles situés en dessous de la limite de détection de notre appareillage.

Les analyses de méthylation de l'ADN par BGS génèrent des quantités abondantes d'informations mais cette technique comporte de nombreux biais dont la reproductibilité a été montrée lors de ma thèse. Les biais de conversion et les biais d'amplification peuvent être mis en évidence par l'utilisation d'enzymes de restriction et en utilisant différents jeux d'amorces pour amplifier une même région.

2. Conservation de la méthylation de l'ADN chez les eucaryotes et régulation des ET

Bien que fortement conservée, la méthylation de l'ADN n'a à ce jour pas pu être détectée dans les génomes de *Saccharomyces cerevisiae* et *Caenorhabditis elegans*. Ces deux espèces d'eucaryotes possèdent d'autres systèmes de régulation post-transcriptionnelle des ET impliquant notamment chez le nématode la voie des RNAi (Lesage et Todeschini 2005; Bessereau 2006). Le champignon *Neurospora crassa* possède une unique DNMT appelée DIM-2, responsable de la méthylation des cytosines en contextes symétriques et asymétriques (Kouzminova et Selker, 2001). Chez cette espèce, un système de mutagenèse dirigé contre les ET appelé « Repeat Induced Point-mutation » (RIP), détecte les séquences répétées de plus de 400 pb et induit des transition C:G en T:A. Ce mécanisme est si efficace que peu d'ET sont actifs chez *N. crassa*. Les séquences méthylées chez cette espèce, identifiées par MedIP, correspondent à des ET riches en AT qui sont sous le contrôle de la RIP (Selker et al. 2003). Chez les insectes, la fonction de la méthylation des cytosines est impliquée, selon les espèces, dans des fonctions aussi diverses que la régulation des gènes, les empreintes génomique, la régulation des ET et la stabilité du génome (Field et al. 2004). De nombreux insectes possèdent des systèmes de méthylation comparables à ceux des vertébrés avec dans leur génome un ou plusieurs orthologues des différentes DNMT. L'abeille *Apis mellifera*, le ver à soie *Bombyx mori*, et la guêpe *Nasoni vitripentis* en font partie. Le génome de l'abeille, dans lequel les gènes DNMT1, DNMT2 et DNMT3 sont présents, possède un système de méthylation de l'ADN dirigé contre les dinucléotides CpG intervenant dans la régulation des gènes (Wang et al. 2006). De plus, les cytosines méthylées sont principalement localisées dans les régions codantes des gènes et absentes au niveau des ET (Field et al. 2004). Le phasme *Medauroidea extradentada* possède quant à lui une forte proportion de CpG méthylés, ainsi que des CpA et CpT localisés au niveau des séquences répétées et à l'intérieur des gènes (Krauss et al. 2009). Le génome du ver à soie, *B. mori*, contient 0,67% de 5-MedC, principalement localisées au niveau des dinucléotides CpG, à l'intérieur des gènes et absente au niveau des ET (Xiang et al. 2010). Étonnamment, aucun lien n'a pu être établi chez cette espèce entre la méthylation des promoteurs et la répression de la transcription. Au contraire, une corrélation a été mise en évidence entre le taux de méthylation des régions codantes d'un gène et son expression. Cette même corrélation a été mise en évidence chez *A. thaliana*, chez laquelle la méthylation des cytosines dans le corps des gènes est un événement également

fréquent. Selon le dogme de la méthylation des cytosines, la présence de 5-MedC au niveau de la région promotrice d'un gène empêche son expression, et de nombreuses données récentes montrent que les parties codantes des gènes exprimés sont fortement méthylées. Il a été proposé que la méthylation de l'ADN dans les parties codantes des gènes qui s'expriment (donc non méthylés au niveau de leur région promotrice) empêcherait une initiation de la transcription ailleurs que sur le promoteur (Suzuki et Bird, 2008).

La vaste conservation au cours de l'évolution de la méthylation de l'ADN suggère un rôle important et nous avons vu que son implication dans la régulation des gènes n'est pas la même selon l'espèce considérée. La méthylation des cytosines dans le corps des gènes est un événement répandu au cours de l'évolution des génomes eucaryotes (Zemach et al. 2010). A l'inverse, les systèmes de régulation des ET par la méthylation de l'ADN sont peu conservés. Une étude chez l'Homme a montré que la majorité des sites CpG sont méthylées (Edwards et al. 2010), et que le taux de méthylation des séquences est corrélé avec leur densité en dinucléotides CpG. Les séquences très denses en CpG, comme les régions promotrices des gènes, sont à l'inverse non méthylées. Selon les auteurs, l'état des cytosines est, par défaut, méthylé et que les promoteurs seraient préservés de cette méthylation par leur forte densité en dinucléotides CpG. L'implication de la méthylation de l'ADN dans la structure et le maintien de la stabilité du génome est appuyée par plusieurs exemples, dans lesquels l'absence de méthylation est associée à des anomalies chromosomiques ainsi qu'à une levée de la répression des ET.

« The genome is an ecological niche, and it was inevitable that the resources available there should come to be exploited by specialized replicating entities ».

Dans leur article, dont la citation ci-dessus a été extraite, Bestor et al (1997) indiquent que la méthylation des cytosines est un mécanisme de préservation de l'intégrité du génome des mammifères et des plantes en réprimant les ET. Ils considèrent que le génome de la drosophile, qui possède beaucoup d'éléments actifs et peu de méthylation de l'ADN, ne possède pas ce système de répression. Chez la drosophile, la seule enzyme capable de méthyler l'ADN est DNMT2, qui est aussi la « DNA methyl transférase » la plus conservée des génomes eucaryotes (Ponger et Li, 2005). Bien qu'elle possède tous les domaines catalytiques spécifiques des DNMT, l'activité DNMT de cette enzyme est très faible (Hermann et al. 2003) et comme nous l'avons vu dans le chapitre 2, les taux de 5-MedC dans

les génomes de drosophiles sont faibles. Des résultats récents ont montré que son activité principale est de méthyler les ARNtAsp, (Acide Aspartique) chez l'Homme, les plantes, la souris et la drosophile (Goll et al. 2006). L'enzyme DNMT2 est une « RNA Methyl Transferase » parmi d'autres, dont la conservation extrêmement importante des procaryotes jusqu'aux eucaryotes pourrait toutefois suggérer que les DNMT des mammifères et des plantes en dérivent (Pavlopoulou et Kossida, 2009). La localisation de DNMT2 nucléaire et cytoplasmique, ainsi que la faible activité DNMT, peuvent expliquer les faibles taux de 5-MedC observés chez la drosophile. Bien que le gène dDNMT2 semble être impliqué chez la drosophile dans la mise en place de marques d'histones hétérochromatiques au niveau des ET (Gou et al. 2010; Phalke et al. 2009), le lien entre la méthylation de l'ADN des séquences d'ET et son rôle dans leur régulation transcriptionnelle n'est, à ce jour, pas établi. Ceci suggère que le gène dDNMT2 pourrait être impliqué dans un mécanisme de régulation des gènes et des ET qui ne ferait pas intervenir la méthylation des cytosines de l'ADN.

3. La régulation des ET chez la drosophile : exemple de *tirant* dans les populations naturelles de *D. simulans*

Le phénotype associé à des mutants pour l' « histone methyl transferase » SU(VAR)3-9, qui méthyle l'histone H3 sur la lysine 9, est caractérisé par des translocations chromosomiques, défauts de structure de la chromatine et perte d'hétérozygotie, aussi bien dans la lignée germinale que dans la lignée somatique (Peng et Karpen, 2009). Une dérégulation des ET peut conduire au même phénotype (Bourc'his et Bestor 2004; Miura et al. 2001), et, comme nous avons vu au cours du chapitre 1, la structure de la chromatine est impliquée dans la régulation de l'ET *tirant*. Les ET peuvent être considérés comme des éléments essentiels de la stabilité des génomes. Les séquences d'ET seraient donc impliquées dans l'établissement des régions hétérochromatiques et euchromatiques, indispensables pour la stabilité du génome. Mes résultats ont montré que *tirant* était capable de modifier la structure de la chromatine au niveau de son site d'insertion, mais aussi en amont. En effet, l'insertion d'une copie complète de *tirant* dans l'intron du gène *tkv* dans la population de Makindu est responsable de la présence de la marque hétérochromatique H3K27-me3 au niveau de son promoteur. Etablie à partir du génome séquencé de *D. melanogaster*, la distance entre le début de la LTR5' de cette copie de *tirant* et le promoteur du gène *tkv* est estimé à 2 571 pb. L'influence qu'un ET peut avoir sur la structure de la chromatine au niveau de son environnement génomique a été décrit avec l'élément *P*, dont l'insertion peut conduire au phénomène de « Position Effect Variegation » (PEV) caractérisé par une hétérochromatinisation de gènes euchromatiques (Girton et Johansen, 2008; Weiler et Wakimoto, 1995). Le spreading, sur plusieurs kb, de la marque H3K27me3 a été rapporté chez la *Drosophile* (Suganuma et Workman, 2010) et chez la *A. thaliana*, chez laquelle le « silencing » épigénétique provoqué par l'insertion d'un ET dans ou à proximité d'un gène a de plus été mis en évidence (Schubert et al. 2006; Lippman et al. 2004). Des résultats préliminaires obtenus dans l'équipe montrent que le gène *tkv* est plus faiblement exprimé dans la population de Makindu que dans les autres populations analysées, ce qui confirme l'influence de la présence de la marque H3K27me3 au niveau de son promoteur. Le phénotype associé à des mutations du gène *tkv* chez la drosophile est caractérisé par des ailes dont les nervures sont fines, n'a cependant pas été mis en évidence dans la population de Makindu. Ce cas particulier devra être étudié afin de comprendre l'influence des ET sur les gènes et les interactions qui peuvent exister entre ces séquences.

Avec l'exemple de *tirant*, nous avons montré que des régions euchromatiques sont associées à des marques hétérochromatiques lorsqu'une copie de cet élément y est insérée. Comment expliquer que *tirant* soit exprimé uniquement dans la population de Makindu ? Les copies M3 et Z2 sont associées à des marques euchromatiques, mais seule la copie M3 est complète, ce qui peut expliquer que des transcrits de *tirant* n'aient pu être mis en évidence que dans la population de Makindu. Afin d'identifier la copie de *tirant* qui est exprimé, il faudra étudier par CHIP l'association de la copie M3 avec d'autres marqueurs caractéristiques de l'hétérochromatine et de l'euchromatine, ainsi qu'avec des marqueurs d'expression comme par exemple, la polymérase II, qui peu être utilisée en ChIP pour identifier les séquences transcrites (Sun et al. 2010). Le séquençage des transcrits de la population de Makindu pourra également nous permettre d'identifier la copie transcrite à partir des polymorphismes de séquences observés.

Des résultats récents d'hybridation *in situ* de *tirant* sur ovaires des populations de Sénégal, Makindu, Chicharo et Zimbabwe, réalisées par Abdou Akkouche (Thèse de doctorat), ont permis de localiser les transcrits de type C de *tirant* de la population de Makindu au niveau des cellules nourricières appartenant à la lignée germinale. De nombreux transferts d'ARN ont lieu des cellules nourricières vers l'ovocyte lors de la maturation de l'ovocyte, cependant, aucun transcrit de *tirant* n'a été localisé dans l'ovocyte. Cela suggère d'une part qu'un mécanisme de régulation de *tirant* empêche son expression dans l'ovocyte, et d'autre part, que les ARN de *tirant* localisés dans les cellules nourricières sont séquestrés dans des structures particulières empêchant leurs transferts vers l'ovocyte. Dans la voie de régulation de la lignée germinale des ET par les piRNA, il a été montré récemment chez la drosophile que les transcrits d'ET régulés sont séquestrés dans les Yb-Bodies, structures localisées dans le nuage nucléaire (Olivieri et al. 2010). L'ensemble de ces données suggère que les copies exprimées de *tirant* sont régulées par la voie piRNA dans la lignée germinale de la population de Makindu. Cette voie de régulation, également impliquée dans le phénotype de dysgénésie des hybrides (Brennecke et al. 2008), fait intervenir des populations de petits ARN dont la composition peut être différente entre plusieurs souches, comme cela a été montré au sein de l'espèce *D. virilis* (Rozhkov et al. 2010). Ce travail ouvre différentes perspectives dont une approche spécifique d'analyse de la régulation de *tirant* par la chromatine et par les piRNA, et une approche globale d'identification des piRNA dans les populations naturelles étudiées ici.

Nous avons vu que le modèle de *tirant* est idéal pour étudier l'influence d'un ET sur la structure de la chromatine et les conséquences au niveau de l'expression des gènes situés à proximité. Le rôle fonctionnel des protéines impliquées dans la voie de régulation piRNA devra être confirmé par la mise en évidence de la colocalisation et de l'interaction des protéines PIWI avec les ARN de *tirant*. Dans notre hypothèse, les transcrits de *tirant* séquestrés dans les Yb-bodies proviendraient de la copie M3.

Enfin, nous avons vu dans le chapitre 2 que les taux de méthylation chez les drosophiles, déterminés par HPLC, sont faibles, mais que l'analyse par enzymes de restriction montre que les ET présentent des sites méthylés. Chez de nombreuses espèces d'insectes, de mammifères ou de plantes, les ET sont la cible privilégiée de la méthylation des cytosines. Les résultats préliminaires obtenus à partir d'ADN extrait d'embryons de moins de 2 heures des populations de Makindu (*D. simulans*) et Sénégal (*D. melanogaster*) montrent des taux de 5-MedC très faibles. Ces résultats devront être confirmés et l'analyse devra être étendue à d'autres populations naturelles de *D. simulans*, à différents stades de développement.

PUBLICATIONS

Article publié en 2010:

Rebollo R, Horard B, **Hubert B**, Vieira C, “Jumping genes and epigenetics: Towards new species”, *Gene*, 454. 1-7

Article en révision :

Ken Kraaijeveld^{1,2*}, Brechtje Zwanenburg¹, Padu Franco¹, **Benjamin Hubert**³, Cristina Vieira³, Sylvia de Pater¹, Richard Stouthamer⁴, Peter de Knijff² & Jacques J. M. van Alphen⁵, “Rampant proliferation of a transposable element in an asexual wasp”. *En révision dans la revue Science. J'ai réalisé dans le cadre de ce travail les analyses de méthylation d'un ET par BGS.*

Articles en préparation:

Hubert B, Akkouche A, Fablet M, Vieira C, “Tirant transposable element induced change in chromatin structure in natural populations of *D. simulans*” (journal pour soumission: FASEB)

Hubert B, Bellvert F, Comte G, Vieira C, “Variation of 5methylcytosine proportion among drosophilids phylogeny and natural populations”



Review

Jumping genes and epigenetics: Towards new species

Rita Rebollo^a, Béatrice Horard^b, Benjamin Hubert^a, Cristina Vieira^{a,*}^a CNRS, UMR5538, Laboratoire de Biométrie et Biologie Évolutive, Université Lyon 1, Villeurbanne, France^b Laboratoire de Biologie Moléculaire de la Cellule-UMR 5239 CNRS, ENS LYON, Université LYON 1, HCL, 165, Chemin du Grand Saconnet, 69635 Pierre Bénite or 46, allée d'Italie, 69304 Lyon cedex 07, France

ARTICLE INFO

Article history:

Received 24 November 2009

Received in revised form 6 January 2010

Accepted 19 January 2010

Available online 25 January 2010

Received by Prescott Deininger

Keywords:

Transposable elements

Evolution

Rapid speciation

Natural populations

Epigenetic control

ABSTRACT

Transposable elements (TEs) are responsible for rapid genome remodelling by the creation of new regulatory gene networks and chromosome restructuring. TEs are often regulated by the host through epigenetic systems, but environmental changes can lead to physiological and, therefore, epigenetic stress, which disrupt the tight control of TEs. The resulting TE mobilization drives genome restructuring that may sometimes provide the host with an innovative genetic escape route. We suggest that macroevolution and speciation might therefore originate when the host relaxes its epigenetic control of TEs. To understand the impact of TEs and their importance in host genome evolution, it is essential to study TE epigenetic variation in natural populations. We propose to focus on recent data that demonstrate the correlation between changes in the epigenetic control of TEs in species/populations and genome evolution.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Speciation can be a slow process in which genetic differences between individuals become fixed, either as a result of changes in their fitness, ecological specialization, or simply the induction of genetic incompatibility, followed in both cases by micro-population isolation. Speciation is also thought to happen quickly in the context of “non-genetic” speciation, i.e., when important karyotypic differences between individuals of the same species lead to sexual isolation. Our exploration of the dynamics of transposable elements (TEs) in natural populations led us to ask how TE mobilization is involved in speciation (Vieira et al., 1999; Rebollo et al., 2008; Fabier et al., 2009). In this short-review we argue that TEs are able to induce speciation through chromosomal rearrangements since, 1) chromosomal rearrangements are able to induce speciation (Noor et al., 2001; Baird et al., 2009; Greig, 2009), 2) bursts of TE transposition can cause chromosomal rearrangements (Geurts et al., 2006; Weil, 2009; Zhang et al., 2009), and 3) bursts of TE transposition may be driven by the selective release of active elements as the result of an epigenetic response to the environment (Lisch, 2009). We also discuss the importance of TE-induced speciation compared to that of other speciation mechanisms.

2. Transposable elements and the genome: a partnership on the move

Genome-sequencing programs have provided new clues explaining the lack of correlation between phenotypic complexity and genome size—which is known as the “C value” paradox—by revealing that most of the differences in genome size between species reside in the non-coding parts (reviewed in Biemont and Vieira, 2006). For instance, the human genome is composed of ~98% of non-coding DNA (International Human Genome Sequencing Consortium, 2004), whereas the fruit fly, *Drosophila melanogaster*, has a very compact genome with far fewer sequences of this type (Dowsett and Young, 1982; Hoskins et al., 2002; Kaminker et al., 2002; Clark et al., 2007).

This variable part of the genome consists mostly of repetitive sequences, such as satellite DNAs and TEs. We will focus mainly on the latter in this short review (see Fig. 1 for a classification of eukaryotic TEs). The evolutionary importance of TEs is no longer open to question. In general, all families of DNA repeats could potentially have an impact on genome organization, either by generating genome instability, since multicopy elements are known to be powerful recombinogenic substrates (Hedges and Deininger, 2007), or as being components of essential chromosomal domains, such as centromeres and telomeres in many species (Wong and Choo, 2004; Lamb et al., 2007). It is clear that TE replication might induce genetic mutations via transposition, as reported for some maize lineages, where *Ac/Ds* alternative transposition (from the ends of two different elements) is directly responsible for major chromosomal rearrangements (translocations, duplications, inversions...) (Zhang et al., 2009). The immediate effects

Abbreviation: TE, Transposable element.

* Corresponding author. Tel.: +33 4 72 43 29 18; fax: +33 4 72 43 13 88.
E-mail address: vieira@biomserv.univ-lyon1.fr (C. Vieira).

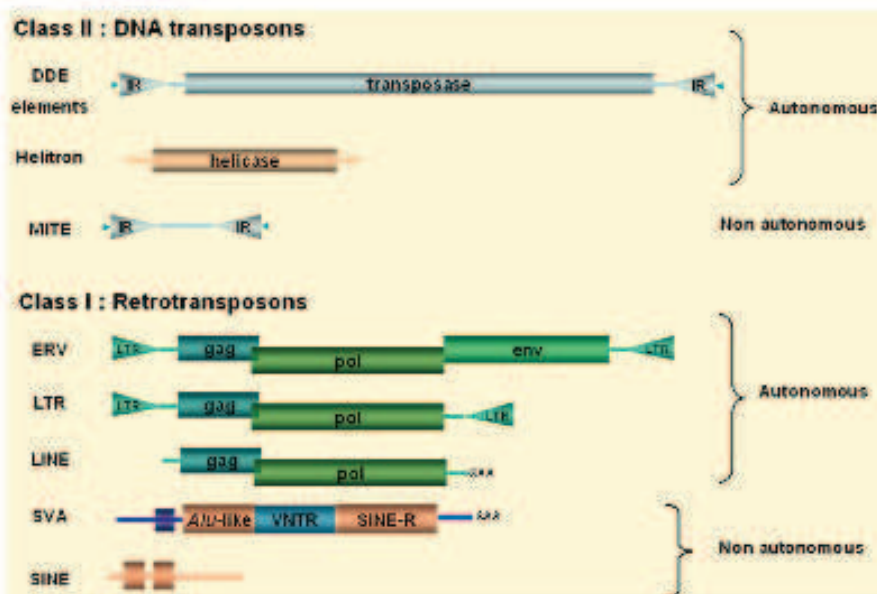


Fig. 1. General classification of eukaryotic transposable elements. TEs are abundant and ubiquitous mobile sequences capable of jumping inside the genome. TEs are divided into two major classes on the basis of differences in their transposition mechanisms: Class I Retrotransposons “copy and paste” through an RNA intermediate, whereas Class II DNA transposons just “cut and paste” their own molecule. Autonomous retrotransposons harbor long terminal repeats in their ends (LTR) or not (LINE-like), and can be infectious agents (endogenous retroviruses). Non-autonomous retrotransposons, such as SINEs, are dependent on autonomous elements to be “copied and pasted” in trans. The same dependency is observed among DNA transposons, where MITEs need a full-length transposase coded by autonomous DNA transposons to be “cut and pasted” in trans. Full-length helitrons, recently identified Class II DNA transposons, play an important role in exon shuffling thanks to their “rolling circle” replication mechanism. For a recent classification of eukaryotic TEs, please refer to *Wicker et al., 2007*. Boxes represent open reading frames, triangles are either inverted repeats (IR) in blue, or long terminal repeats (LTR) in green, and small blue arrows correspond to duplicated insertion site representations. DDE elements: transposases carrying the aspartate (D), aspartate (D), glutamate and (E) motif. MITE: miniature inverted repeated elements; ERV: endogenous retrovirus; LINE: long interspersed nuclear element; SVA: composite element composed of parts of SINE (short interspersed nuclear element), VNTR (variable number of tandem repeats) and Alu repeats—the first box represents CCCTCT hexamer repeats; SINE red boxes indicate a diagnostic feature; Gag, Pol, Env: retroviral-like proteins coded by TE open reading frames.

of transposition can be detrimental, as illustrated by several human diseases (reviewed in *Callinan and Batzer, 2006*). Despite these damaging effects, TEs have been maintained in almost all genomes either as full-length or truncated copies. Full-length copies have kept their ability to mutate the genome as a result of transposition, while the truncated copies have often lost their capacity to transpose. However, the truncated versions might also be recruited by the genome. Indeed, recent reports have proposed a scenario of co-evolution of TEs and hosts, in which TEs (often as truncated copies) are involved in complex genomic processes such as post transcriptional gene regulation, gene protein translation enhancement, etc. (*Muotri et al., 2007; Sinzelle et al., 2009*). Truncated copies may also act as recombinogenic substrates for other truncated or full-length copies, inducing genome rearrangements. TE copies have been shown to give rise to new regulatory sequences, alternative splice sites, polyadenylation signals (*Marino-Ramirez et al., 2005*), and new transcription-factor binding sites (*Polavarapu et al., 2008*). TEs also enhance genome regulation as, for example, when they give rise to microRNAs, which are able to regulate gene expression (*Hasler et al., 2007; Piriyaongsa et al., 2007*). Since TEs are widespread in the genome and have so many different influences on gene regulation, several authors have suggested that TEs may play a vital role in creating, remodelling and regulating gene networks (*McClintock, 1984; Feschotte, 2008*). We therefore hypothesize that bursts of TE activity may have profound impacts on genome structure and gene regulation.

3. Bursts of TE transposition drive speciation

TEs have been observed in all sequenced genomes analyzed to date, and comparative genomics gives us a broad insight into the

variability of TE content between genomes of different species. For instance, TEs represent 85% of the maize genome (*Schnable, 2009*), but only 14% of *Arabidopsis thaliana*'s (*The Arabidopsis Genome Initiative, 2000*). Variations in genome size between closely related species are often related to differences in the amount of TEs. For instance, in the *D. melanogaster* species subgroup, larger genomes are partly attributable to high TE-like sequence content, as estimated from the amount of reverse transcriptase-related sequences determined by dot blot (*Boulesteix et al., 2006*). In cotton, differences in genome size between species, ranging from 40 to 65%, have been associated with a particular TE subfamily, the gypsy-like Gorge3 element (*Hawkins et al., 2006*). Differences in the relative proportions of TEs can also occur within a species, as observed for copies of several TE families when counted in the euchromatic arms on polytene chromosomes in natural populations of *D. simulans* (*Vieira et al., 1999*).

TE abundance, TE-derived genomic features and chromosomal rearrangements involving TE sequences are frequently lineage specific and, therefore, suggest that TEs have contributed to the process of speciation, either as a cause, or an effect (*Marino-Ramirez et al., 2005; Bohne et al., 2008*). One should note that correlating TE transposition consequences and speciation is rather tricky, since it is difficult to determine the exact timing of TE bursts and natural species diversification. We hypothesize that TE burst of transposition might induce rapid speciation, but the debate is still open. A variety of factors such as gene transfers and losses, mutations affecting “speciation genes”, endosymbiotic interactions, antirecombination mechanisms during meiosis, and introgression could also account for reproductive isolation (*Presgraves et al., 2003; Scannell et al., 2006; Lexer and Widmer, 2008; Lowry et al., 2008; Greig, 2009*). In addition, an influence of TEs on speciation apparently without transposition bursts was reported in

some species. For instance, the formation of the recent invasive species *Spartina anglica* through natural allopolyploidization was accompanied by major structural and epigenetic remodelling (CpG methylation changes) in the vicinity of TEs (Parisod et al., 2009), but bursts of transposition have not been detected (Baumel et al., 2002). Heterochromatin, which is highly enriched in TEs and other repeats, also seems to play a role in speciation, as observed by Ferree and Barbash in *Drosophila* hybrids (Ferree and Barbash, 2009; Hughes and Hawley, 2009). Furthermore, bursts of transposition are not always associated with rapid speciation. It should be kept in mind that *P* and *I* elements have recently invaded natural populations of *D. melanogaster*, but no speciation has been observed. Also, in a few rice strains, nearly 40 new copies of the DNA transposon *mPing* are observed per plant generation. For the present, the only consequence of this burst of transposition has been slight transcriptional deregulation of a few genes that have *mPing* copies inserted into their 5' flanking region. However, it is interesting to note that a few of the genes containing *mPing* in their flanking regions are salt- and cold-inducible as a result of the presence of the DNA transposon, suggesting that a few insertions may play an adaptive role (Naito et al., 2009). No massive restructuring of the genome has been observed in rice despite the high rate of transposition, and no speciation seems to be happening in this species. We hypothesize that in order for bursts of transposition to induce rapid speciation, massive genome restructuring or mutations in "speciation genes" have to occur. Several researchers have found examples of concordant timing between bursts of transposition or massive TE extinction and speciation (Table 1). While this short-review was in revision, two bibliographic analyses of the punctuated equilibrium theory, and the general impact of TEs in genomes were published, and both reinforce the hypothesis that TEs may drive macroevolution via bursts of transposition (Oliver and Greene, 2009; Zeh et al., 2009). Significant TE activity is observed in several species, often during periods of radiation, suggesting that massive speciation and massive TE activity may be associated. The genetic distance between two organisms is calculated as a function of their genetic divergence, so every episode that creates divergence, such as lineage-specific transposition events, could contribute to the reproductive isolation of those organisms. TE patterns that differ between individuals of the same species, whether as a cause or a consequence of genetic differentiation, may not only provide genetic markers for researchers, but also constitute evidence of a speciation process occurring within the species concerned (Esnault et al., 2008). For instance, significant TE insertion site polymorphism can be observed in the Japonica and Indica cultivars of rice, and accounts for 14% of their genetic differences (Huang et al., 2008).

Table 1
TE transposition bursts concomitant with radiation periods.

TE events/Species history	References
Decreased LI and SINE accumulation during emergence of African apes (14–15 Mya).	Conortian (2002)
Contraction of LI subfamilies in less than 0.3 Mya concomitant with intense speciation in <i>Rattus norvegicus</i> .	Vernau et al. (1998)
The timing of La family (LI ancestral family) amplification is close in time to the murine radiation.	Pascali et al. (1999)
Rapid speciation in the genus <i>Leptis</i> (gerbil) occurred, and massive transposition of <i>Fla</i> in new lineages was observed.	Dohigny et al. (2004)
DNA elements were extremely active during the Mycine radiation.	Ray et al. (2008)
DNA transposon transposition bursts are concomitant with speciation events in pseudotetraploid salmonids and occurred after genome duplication.	de Boer et al. (2007)
Acquisition and consequent transposition of an endogenous retrovirus element in <i>Escherichia histolytica</i> , and lineage specific enrichment in TEs might affect speciation and pathogenicity.	Lorenzi et al. (2008)

Since the exact evolutionary history of a species is difficult to determine, interspecies crosses provide useful macroevolution study models. Indeed, interspecies hybrids are classic examples of bursts of transposition that have caused severe dysfunctions, and could potentially induce rapid speciation (reviewed in Fontdevila, 2005; Michalak, 2009) (Table 2). Three independent hybrids of the sunflower species have a genome that is 50% larger than the parental lines as the result of a massive TE transposition, and they are thought to have undergone rapid speciation (in fewer than 60 generations in one case) (Ungerer et al., 1998, 2006). Also, in dosage-dependent crosses between *A. thaliana* and *A. arenosa* (crosses where the amount of the maternal and paternal genomes are variable, and so may be different), high expression of the paternal *A. arenosa* *Athila* element in the hybrid is correlated with seed lethality (Josefsson et al., 2006). Such observations are essential for understanding hybrid "compatibility", since *A. arenosa* and *A. thaliana* hybridization has been successful at least once in nature (Jakobsson et al., 2005). In insects, *D. buzzatii* and *D. koepferae* are still able to interbreed, and share common TE families that are maintained in both genomes. Crosses between these two species considerably induce the transposition of *Oswaldo* in the resulting hybrids, even though it is repressed in both parental genomes (Labrador et al., 1999). In wallabies, interspecies hybrids contain variable centromeres, composed of satellite repeats and newly replicated TE copies (Metcalfe et al., 2007). McFadden and Knowles have designed an algorithm in order to model evolution in asexual digital organisms with or without TEs. They conclude that transposon-mediated mutations were associated with punctuated bursts of rapid evolution and appearance of new adaptive peaks, in contrast to the stasis trap observed in organisms without transposons (McFadden and Knowles, 1997). All these examples suggest that bursts of TE transposition occurring during hybrid speciation may induce important karyotypic changes because of the ability of TEs to induce major chromosomal rearrangements and ectopic recombination (Hedges and

Table 2
Hybrid analysis: epigenetic remodeling and TE activation.

Experimental conclusions	References
<i>May musculus</i> and <i>M. caroli</i> crosses induce retrotransposon hypomethylation on chromosome 10, the substrate of double minute chromosome formation in interspecies hybrids.	Brown et al. (2008)
Intraspecific crosses of <i>D. melanogaster</i> can result in hybrid dysgenesis, associated with modification of <i>P</i> or <i>I</i> elements, dependent on <i>trfA</i> RNA production in the germline cell line and causing several abnormalities (such as female sterility).	Brennecke et al. (2000)
Crosses between <i>D. buzzatii</i> and <i>D. koepferae</i> activate <i>Oswaldo</i> copies in the hybrid.	Labrador et al. (1999)
Interspecific macropodid hybrids (<i>Macropus rufogriseus</i> and <i>M. agilis</i>) present centromeric instability due to TEs and satellite replication, probably inducing karyotypic isolation from the parental species.	Metcalfe et al., 2007
Genome-wide hypomethylation and centromeric expansion, due to TE activation, are observed in <i>M. eugeni</i> or Wallaby <i>bicolor</i> hybrids.	O'Neill et al., 1998
In <i>A. thaliana</i> and <i>A. arenosa</i> dosage-dependent crosses, the usually silenced paternal <i>Athila</i> elements are activated concomitantly with the deregulation of polycomb-complex-dependent gene regulation.	Josefsson et al. (2006)
Wheat allotetraploid formation is accompanied by TE activation, DNA methylation and gene expression alterations.	Kashikish et al., 2002
<i>Helianthus annuus</i> and <i>H. petiolaris</i> hybrids have a genome 50% larger than parental individuals due to TE amplification.	Ungerer et al. (1998), Ungerer et al. (2006)
DNA introgression in <i>Zizania latifolia</i> causes TE activation through modifications in DNA methylation and morphological deviations from the parental line. Note that de novo stable silencing of TEs is observed in the introgressed lines.	Liu and Wendel (2003)

Deininger, 2007; Weil, 2009). In this way, TE mobilization can result in novel phenotypes followed by the ecological isolation of micro-populations, which are the components required for rapid and divergent evolution.

Both transcription and transposition activities of TEs are controlled via a variety of mechanisms. The expression of TEs is dependent on the presence of transcription factors, as illustrated by the evolution of the L1 lineage in humans. Indeed, the recruitment of regulatory regions in new L1 subfamilies harboring new transcription-factor binding sites is essential for L1 expression (Khan et al., 2006). Furthermore, cellular inhibitors may influence TE transposition post-transcriptionally, as has been observed for some members of the APOBEC family, which are capable of reducing HERV-K infectivity (50 fold) (Lee and Bieniasz, 2007), and block Alu transposition in a manner independent of ORF1p L1 (Hulme et al., 2007). Moreover, transposition of mobile elements induces DNA breaks, suggesting that an interaction occurs between the host DNA repair machinery and TEs. ERCC1-XPF heterodimers are implicated in DNA repair processes and limit L1 insertion (Gasior et al., 2008). Apart from cellular inhibitors and transcription factor dependency, TEs are also transcriptionally and post-transcriptionally regulated through epigenetic pathways (Lisch, 2009). However, we know that epigenetic mechanisms are labile in response to environmental changes, and so TEs may occasionally escape silencing, and in some cases could cause genome rearrangements. In order to understand how speciation occurs as a result of transposition bursts, it is therefore essential to understand epigenetic reprogramming.

4. Epigenetic reprogramming of TEs

Epigenetic regulation of TEs involves interdependent pathways, such as chromatin remodelling factors, DNA methylation and non-coding small RNAs (Lisch, 2009; Obbard et al., 2009) (Table 3 for an overview of TE epigenetic regulation). In rice, for instance, specific mutants of histone H3K9 methyltransferase induce DNA demethylation of *Tos17* (copin-like retrotransposon) and, consequently, lead to transposition (Ding et al., 2007). In plants, RNA-dependent DNA methylation (RDDM) of TEs and genes is often observed. This is reversible, since it is dependent on the presence of small interfering RNAs (Matzke et al., 2007). Recent investigations have highlighted the central role of RNA in controlling TE activity: such a system was probably present in a common eukaryote ancestor as it is well conserved between species; and it may act as an immunological system against non-self RNAs (Obbard et al., 2009). Also, small RNAs allow for target specificity of DNA methylation or histone modification in a given sequence. For instance, epigenetic instability in long-term cultured cells of *A. thaliana* evolves into the hypomethylation of specific TEs and subsequent activation (Tanurdzic et al., 2008). Indeed, *Athila* or *copin* elements are hypomethylated, regardless of their location, whereas no change is observed for gypsy class elements (Tanurdzic et al., 2008). Such specificity is possibly due to the fact that siRNAs are produced differently in TE families subjected to stress of this type, varying from 21 nt and 24 nt for hypomethylated activated TEs, but with only 24 nt for silenced gypsy class elements (Tanurdzic et al., 2008).

The TE epigenetic regulation system is in fact rather efficient. It is general in nature, because the TE families capable of invasion are multiple and divergent but, at the same time, it also appears to be specific, and targets single TE families through sequence-specific small RNAs. Each pathway in the epigenetic regulation of TEs seems, therefore, to be both essential and extremely rigorous. Naturally the question arises as to how TEs can possibly invade a genome if they are trapped in an inviolable prison. In fact, we know that TEs often do translocate at a very low rate, suggesting that the prison is after all, somewhat permeable. Indeed, it has recently been suggested that small RNAs can be linked with the total or partial silence of elements,

Table 3

General view of TE epigenetic regulation.

Histone modifications	Position effect variegation (PEV) is the mechanism behind variation in the transcription of a given gene, and is correlated to its chromatin localization. Mutations in <i>Su(var)</i> genes responsible for such variegation are often accompanied by TE amplification. The major function of this gene family is to post-translationally modify histone N-terminal ends. Usually histone methylation in lysine residues (H3K9me, H3K27me, and H4K20me) typically occurs in a closed chromatin conformation, in contrast to the acetylation of histones and methylation in H3K4, which are often observed in open chromatin structures. TEs are closely associated with repressive marks, like H3K9me2 in humans, H4K20me3 in <i>Drosophila</i> and H3K9me2 in plants.
DNA methylation	In plants and mammals, DNA methylation plays an important role in silencing TEs. In insects, DNA methylation is observed as a silencing process in genes and TEs.
Non-coding RNAs	Post-translational gene silencing (PTGS) via small interfering RNAs (siRNA) processed by the AGO, DICER, RISC complex is another mechanism that can be used to silence TEs. Indeed siRNAs derived from TE copy transcripts can target full-length and positively active TE transcripts, thus preventing TE transposition. Piwi-related RNAs (piRNAs or snoRNAs, standing for repeated associated small interfering RNAs) in <i>Drosophila</i> are processed via the Piwi/Aub/AGO3 pathway, are 24–30 nt and are known to silence TEs in the germline, subgenomic siRNA (endogenous small interfering RNAs) processed by DICER2, AGO2 are 21 nt, and are capable of somatic silencing TEs. Germlinal and somatic silencing are therefore possible thanks to non-coding RNAs. However, the presence and the transcription of a TE copy in the genome are essential to engage PTGS (post-translational transposable element silencing). The idea that this constitutes an immune system is therefore appropriate, since having non-coding RNAs of a given TE family will protect the genome from further invasions.

as observed in *Drosophila* hybrid dysgenesis. Indeed, intraspecific *Drosophila* crosses may cause hybrid dysgenesis of *P* and *I* elements, resulting in very seriously deleterious effects, such as female sterility or chromosomal abnormalities (Bucheton et al., 1984; Castro and Carareto, 2004). In these crosses, individuals of the same species have different amounts of TEs since one of the parents has an “empty” genome. A deficit in a small interfering RNA (piRNA) in the maternal gamete allows originally silenced TEs to translocate in the hybrids (Brennecke et al., 2008; Chambeyron et al., 2008).

The study of natural populations and the observation of the natural variability that exists in epigenetic host control can explain TE-induced macroevolution. Epigenetic variation in hybrids, in allopolyploid species, and in single individuals could arouse the TEs, induce a burst of transposition and, as described above, increase karyotypic changes followed by ecological isolation. TE epigenetic regulation has been reported both in somatic tissues (Barbot et al., 2002; Malone et al., 2009) and in germline tissues (Malone et al., 2009). Both types of regulation can influence population behavior by creating potentially heritable phenotypic variations. Variation in TE epigenetic regulation has been observed, for instance, in the LINE-like element *Sadhu*, that displays epigenetic variation (DNA methylation and different silencing states) in three different *A. thaliana* ecotypes (Rangwala et al., 2006). Other epialleles, or differences in the epigenetic regulation of a given sequence in different tissues and/or individuals belonging to the same population, have been reported, mostly in plants and mice. However, further progress in population epigenetics is still necessary, along with ecological epigenetic studies, if we are fully to understand natural population variation in epigenetic regulation (Bossdorf et al., 2008; Johannes et al., 2008; Richards, 2008). TE

epigenetic regulation is, therefore, a variable and flexible mechanism that can induce massive TE transposition in the germline, and consequent chromosomal rearrangements.

The gibbon species has rapidly accumulated chromosomal rearrangements and, hence, offers an interesting model for karyotypic evolution and speciation. Carbone et al. recently reported an example of differences in the epigenetic regulation of *Alu* elements in humans and gibbons that is associated with breakpoints between the species (Carbone et al., 2009). They observed that CpG content was higher in the gibbon *Alu* elements near the breakpoints (typical of active elements), and that these elements were undermethylated relative to human *Alu*. *Alu* elements present in the breakpoints are probably active and responsible, in part, for the rapid chromosomal remodeling in the gibbon. The authors propose that “the association between undermethylation and chromosomal rearrangement in gibbons suggests a correlation between epigenetic state and structural genome variation in evolution”.

Conjugating two different genomes in the same organism, as in hybrids or in allopolyploids, may require significant adaptations of all the regulatory mechanisms, including TE epigenetic regulation (reviewed in Michalak, 2009) (Table 2). In wallabies, interspecies crosses cause a burst of transposition of a retrotransposon, together with genome-wide hypomethylation (O'Neill et al., 1998). Such a burst of transposition targets a single parental genome, and results in extended centromeres, suggesting rapid karyotype differentiation from the parents (O'Neill et al., 1998). The authors also analyzed some other natural crosses, and found that hypomethylation of the hybrids was always observed as *de novo* chromosomal changes. In allopolyploidization, TE transposition may also be concomitant with genome-wide epigenetic changes (Liu and Wendel, 2003). These examples show how genome remodeling could occur after epigenetic variation in TE copies. However, we need to identify the causes of genome-wide epigenetic modifications and subsequent TE activation. Interspecies crosses induce genomic stress, i.e. changes in genomic stability (chromatin changes, density of repeats...) and organization (DNA recombination, TE replication, retroposed or duplicated genes...), that could indeed have an impact on epialleles and provoke TE activation. Genome-wide epigenetic changes might play a role in genome adaptation to environmental changes. One can readily imagine that TE arousal occurs due to epigenetic changes, and that these changes originate in one individual in response to specific environmental changes. The ecological outcomes of TE mobilization due to environmental changes may be numerous, including things such as survival of the host, increase of host fitness, micro-population isolation etc. The subsequent spread of these factors within populations could lead to sexual isolation and speciation.

5. The environment induces epigenetic reprogramming

Several studies have demonstrated that modifications in the environment can induce epigenetic modifications and, therefore, transcription state changes (Jaenisch and Bird, 2003). Such transcriptional changes are a source of phenotypic variability that may be exploited by organisms to increase the “adaptive potential” of the host. Indeed, diet changes, temperature variation, stress etc. all have an impact on gene regulation (Waterland and Jirtle, 2004; Copley et al., 2006; Gibert et al., 2007; Chinnusamy and Zhu, 2009). In addition, diet changes, temperature variations, stress etc. could all affect TE transposition (El-Sawy et al., 2005; Hashida et al., 2006; Ebina and Levin, 2007; Cho et al., 2008). Consequently, activation of TEs could result from the relaxation of epigenetic control induced by environmental changes.

There is a huge amount of literature relating the activation of TEs to environmental stress, but only a few examples suggest a link between environmental epigenetic instability and the activation of TEs. Early nutrition has an impact on the epigenetic regulation of

TEs, especially via DNA methylation, as reviewed by Waterland and Jirtle, 2004. The *agouti* gene controls hair color in mice (brown in wild type), and the insertion of an IAP retrotransposon in the first exon induces ectopic and variable expression of *agouti*. LTR from IAP elements are regulated by DNA methylation, which varies between individuals. Dietary supplementation (with methyl donors) shifts the phenotype to the wild type brown colour, which is indicative of higher DNA methylation in the IAP element (Waterland and Jirtle, 2003).

In *D. melanogaster*, both heat treatment and aging induce the transcription of older heterochromatic I copies and, hence, the production of small interfering RNAs (siRNA) that repress active I elements in the germline (Dramard et al., 2007). DNA methylation of L1 and *Alu* I elements is decreased in individuals exposed to the pollutant benzene (Bollati et al., 2007) and, similarly, benzo(a)pyrene increases retrotransposition of L1 elements in HeLa cells (Strimbskis and Ramos, 2006). In mice, a long-term peroxisome proliferating diet induces hypomethylation of satellites, IAP and L1/L2 elements (Pogribny et al., 2007).

6. Conclusion

The fact that TE copies are subject to epigenetic regulation has two main consequences: 1) the environment can have a direct influence on TE activity through epigenetic instability and 2) TE sequences are present in the host genome in a “harmless” state. Since bursts of transposition have been observed in several species it is tempting to suggest that their defense systems have, at least temporarily, broken down. However, this failure is transient, and the host may rapidly silence any *de novo* TE copies produced. Although the benefit is not immediate, transposition might have a long term advantage. Indeed, transposition bursts have numerous consequences, resulting in a renewal of genetic diversity, which is the major prerequisite for genome evolution and selection to occur. Genetic diversity is fundamental for gene networks to be renewed, allowing new species to emerge. Each environmental change indirectly creates an increase in host genetic variability, which means that selection can act over a larger repertoire of genetic information. Epigenetic instability of TEs would lead to significant genetic variability, and the subsequent selection of the best adapted organism.

Acknowledgements

Owing to space limitations, we sincerely regret the omission of many outstanding publications in this field. We thank Dr. M. Fablet, Dr. D. Reiss and two anonymous reviewers that contributed significantly to improve the manuscript. We thank Monika Gosh for correcting the English of this article. This work was supported by ANR-09-BLANC-0103-01, FINOVI and CIBLE2008 (Région Rhône Alpes).

References

- Baird, A.B., Hillis, D.M., Patton, J.C., Bickham, J.W., 2009. Speciation by monobrachial centric fusions: a test of the model using nuclear DNA sequences from the bat genus *Rhyngon*. *Mol. Phylogenet. Evol.* 50, 256–267.
- Barbot, W., Dupressoir, A., Lazar, V., Heidmann, T., 2002. Epigenetic regulation of an IAP retrotransposon in the aging mouse: progressive demethylation and de-silencing of the element by its repetitive induction. *Nucleic Acids Res.* 30, 2365–2373.
- Bammel, A., Ainsuete, M., Kalendar, B., Schulman, A.H., 2002. Retrotransposons and genomic stability in populations of the young allopolyploid species *Spartina anglica* C.E. Hubbard (Poaceae). *Mol. Biol. Evol.* 19, 1218–1227.
- Bernmont, C., Vieira, C., 2006. Genetics: junk DNA as an evolutionary force. *Nature* 443, 521–524.
- Bohne, A., Brunet, F., Galiana-Arnoux, D., Schuitheis, C., Voifit, J.N., 2008. Transposable elements as drivers of genomic and biological diversity in vertebrates. *Chromosome Res.* 16, 203–215.
- Bollati, V., et al., 2007. Changes in DNA methylation patterns in subjects exposed to low-dose benzene. *Cancer Res.* 67, 876–880.
- Bossdorf, D., Richards, C.L., Pigliucci, M., 2008. Epigenetics for ecologists. *Ecol. Lett.* 11, 106–115.
- Bouletoux, M., Weiss, M., Bernmont, C., 2006. Differences in genome size between closely related species: the *Desophila melanogaster* species subgroup. *Mol. Biol. Evol.* 23, 162–167.

- Brunerck, J., Malone, C.D., Aravin, A.A., Sachidanandam, R., Stark, A., Hamton, G.J., 2008. An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* 322, 1367–1369.
- Brown, J.D., Goldin, D., O'Neill, R.J., 2008. Methylation perturbations in retroelements within the genome of a *Mus* interspecific hybrid correlate with double minute chromosome formation. *Genomics* 91, 267–273.
- Bucheton, A., Paris, R., Saig, H.M., Peñsant, A., Finnegan, D.J., 1984. The molecular basis of I-R hybrid dysgenesis in *Drosophila melanogaster*: identification, cloning, and properties of the I factor. *Cell* 38, 153–163.
- Callinan, P.A., Botzer, M.A., 2000. Retrotransposable elements and human disease. *Genome Dyn.* 1, 104–115.
- Carbone, L. et al., 2009. Evolutionary breakpoints in the gibbon suggest association between cytosine methylation and karyotype evolution. *PLoS Genet.* 5, e1000538.
- Gastro, J.P., Carareto, C.M., 2004. *Drosophila melanogaster* P transposable elements: mechanisms of transposition and regulation. *Genetica* 121, 107–118.
- Chambeyron, S. et al., 2008. piRNA-mediated nuclear accumulation of retrotransposon transcripts in the *Drosophila* female germline. *Proc. Natl. Acad. Sci. U.S.A.* 105, 14964–14969.
- Chinnusamy, V., Zhu, J.K., 2009. Epigenetic regulation of stress responses in plants. *Curr. Opin. Plant Biol.* 12, 133–139.
- Cho, K., Lee, Y.K., Greenhalgh, D.G., 2008. Endogenous retroviruses in systemic response to stress signals. *Shock* 30, 105–110.
- Clark, A.G. et al., 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450.
- Consortium, M.G.S., 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* 420, 520–562.
- Cornley, J.E., Suter, C.M., Beckman, K.B., Martin, D.I., 2006. Germ-line epigenetic modification of the murine *A* *vy* allele by nutritional supplementation. *Proc. Natl. Acad. Sci. U.S.A.* 103, 17308–17312.
- de Boer, J.G., Yazawa, R., Davidson, W.S., Koop, B.F., 2007. Bursts and horizontal evolution of DNA transposons in the speciation of pseudotetraploid salmonids. *BMC Genomics* 8, 422.
- Ding, Y. et al., 2007. SDG714, a human H3K9 methyltransferase, is involved in Tss17 DNA methylation and transposition in rice. *Plant Cell* 19, 9–22.
- Dobryz, G., Ossuf-Costaz, C., Waters, P.D., Benillo, C., Coutancova, J.P., Volobouev, V., 2004. LINE-1 amplification accompanies explosive genome repatterning in rodents. *Chromosome Res.* 12, 787–793.
- Dowson, A.P., Young, M.W., 1982. Differing levels of dispersed repetitive DNA among closely related species of *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* 79, 4570–4574.
- Dramard, X., Heidmann, T., Jensen, S., 2007. Natural epigenetic protection against the I-factor, a *Drosophila* LINE retrotransposon, by remnants of ancestral invasions. *PLoS ONE* 2, e304.
- Ebina, H., Levin, H.L., 2007. Stress management: how cells take control of their transposons. *Mol. Cell* 27, 180–181.
- E-Sawy, M. et al., 2005. Nickel stimulates L1 retrotransposition by a post-transcriptional mechanism. *J. Mol. Biol.* 354, 249–257.
- Emault, C. et al., 2008. High genetic differentiation between the M and S molecular forms of *Anopheles gambiae* in Africa. *PLoS ONE* 3, e1998.
- Fabiet, M. et al., 2009. Genomic environment influences the dynamics of the transposon L1 in *Drosophila*. *FASEB J.* 23, 1482–1489.
- Ferree, P.M., Barbash, D.A., 2009. Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in *Drosophila*. *PLoS Biol.* 7, e1000234.
- Feschotte, C., 2008. Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* 9, 397–405.
- Fonseca, A., 2005. Hybrid genome evolution by transposition. *Cytogenet. Genome Res.* 110, 49–55.
- Gavin, S.L., Roy-Engel, A.M., Deininger, P.L., 2008. ERCC1/XPF limits L1 retrotransposition. *DNA Repair (Amst.)* 7, 981–989.
- Geurts, A.M. et al., 2000. Gene mutations and genomic rearrangements in the mouse as a result of transposon mobilization from chromosomal concatemers. *PLoS Genet* 2, e150.
- Gibert, J.M., Peronnet, F., Schlotterer, C., 2007. Phenotypic plasticity in *Drosophila* pigmentation caused by temperature sensitivity of a chromatin regulator network. *PLoS Genet* 3, e10.
- Grig, D., 2009. Reproductive isolation in *Soicorhynchus*. *Heredity* 102, 39–44.
- Hashida, S.N., Uchiyama, T., Morita, C., Kishima, Y., Sano, Y., Mikami, T., 2005. The temperature-dependent change in methylation of the antihuman transposon Tss17 is controlled by the activity of its transposase. *Plant Cell* 18, 104–118.
- Hadler, J., Samuelsson, T., Strub, K., 2007. Useful 'junk': Alu RNAs in the human transcriptome. *Cell. Mol. Life Sci.* 64, 1791–1800.
- Hawkins, J.S., Kim, H., Nason, J.D., Wing, R.A., Wendel, J.F., 2005. Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. *Genome Res.* 15, 1252–1261.
- Hedger, D.J., Deininger, P.L., 2007. Inviting instability: transposable elements, double-strand breaks, and the maintenance of genome integrity. *Mutat. Res.* 616, 46–59.
- Hoskins, K.A. et al., 2002. Heterochromatic sequences in a *Drosophila* whole-genome shotgun assembly. *Genome Biol.* 3, RESEARCH0085.
- Huang, X., Lu, C., Zhao, Q., Liu, X., Han, B., 2008. Genome-wide analysis of transposon insertion polymorphisms reveals intraspecific variation in cultivated rice. *Plant Physiol.* 148, 25–40.
- Hughes, S.E., Hawley, R.S., 2009. Heterochromatin: a rapidly evolving species barrier. *PLoS Biol.* 7, e1000233.
- Hulme, A.E., Bogerd, H.P., Cullen, B.R., Moran, J.V., 2007. Selective inhibition of Alu retrotransposition by APOBEC3G. *Gene* 390, 199–205.
- International Human Genome Sequencing Consortium, 2004. Finishing the euchromatic sequence of the human genome. *Nature* 431, 931–945.
- Jamisch, R., Buit, A., 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.* 33, 245–254.
- Jakobsson, M. et al., 2006. A unique recent origin of the allotetraploid species *Arabidopsis suecica*: evidence from nuclear DNA markers. *Mol. Biol. Evol.* 23, 1217–1231.
- Johannes, F., Colot, V., Jansen, R.C., 2008. Epigenome dynamics: a quantitative genetics perspective. *Nat. Rev. Genet.* 9, 883–890.
- Juarez, C., Dilkes, B., Comai, L., 2006. Parent-dependent loss of gene silencing during interspecies hybridization. *Curr. Biol.* 16, 1322–1328.
- Kamraker, J.S. et al., 2002. The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. *Genome Biol.* 3, RESEARCH0084.
- Kashkush, K., Feldman, M., Levy, A.A., 2002. Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. *Genetics* 160, 1651–1659.
- Khan, H., Smit, A., Boissinot, S., 2000. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Res.* 10, 78–87.
- Labrador, M., Farré, M., Uzet, F., Fontdevila, A., 1999. Interspecific hybridization increases transposition rates of *Osvado*. *Mol. Biol. Evol.* 16, 931–937.
- Lamb, J.C., Yu, W., Han, F., Birchler, J.A., 2007. Plant chromosomes from end to end: telomeres, heterochromatin and centromeres. *Curr. Opin. Plant Biol.* 10, 110–122.
- Lee, Y.N., Boerger, P.D., 2007. Reconstitution of an infectious human endogenous retrovirus. *PLoS Pathog.* 3, e10.
- Lopez, C., Widmer, A., 2008. Review: The genic view of plant speciation: recent progress and emerging questions. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 3023–3030.
- Lisch, D., 2009. Epigenetic regulation of transposable elements in plants. *Annu. Rev. Plant Biol.* 60, 43–66.
- Liu, B., Wendel, J.F., 2003. Epigenetic phenomena and the evolution of plant allopolyploids. *Mol. Phylogenet. Evol.* 29, 365–379.
- Lorenal, H., Thangarajan, M., Haas, B., Wortman, J., Hall, N., Cater, E., 2008. Genome wide survey, discovery and evolution of repetitive elements in three *Entomobrya* species. *BMC Genomics* 9, 395.
- Lowry, D.B., Modliszewski, J.L., Wright, K.M., Wu, C.A., Willis, J.H., 2008. Review: The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 363, 3009–3021.
- Malone, C.D. et al., 2009. Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell* 137, 522–535.
- Mariano-Ramirez, L., Lewis, K.C., Landisman, D., Jordan, I.K., 2005. Transposable elements donate lineage-specific regulatory sequences to host genomes. *Cytogenet. Genome Res.* 110, 333–341.
- Martiz, M., Kanno, T., Huettel, B., Dasgupta, I., Marzán, A.J., 2007. Targets of RNA-directed DNA methylation. *Curr. Opin. Plant Biol.* 10, 512–519.
- McClintock, B., 1984. The significance of responses of the genome to challenge. *Science* 226, 792–801.
- McFadden, J., Knowles, C., 1997. Escape from evolutionary stasis by transposon-mediated deleterious mutations. *J. Theor. Biol.* 186, 441–447.
- Metcalfe, C.J. et al., 2007. Genomic instability within centromeres of interspecific mammalian hybrids. *Genetics* 177, 2507–2517.
- Michalak, P., 2009. Epigenetic, transposon and small RNA determinants of hybrid dysfunction. *Heredity* 102, 45–50.
- Moitt, A.R., Marchetto, M.C., Coufal, N.G., Gage, F.H., 2007. The necessary junk: new functions for transposable elements. *Hum. Mol. Genet.* 16 (Spec No. 2), R159–R167.
- Nano, K. et al., 2009. Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature* 461, 1130–1134.
- Noor, M.A., Gratta, K.L., Bertucci, L.A., Reiland, J., 2001. Chromosomal inversions and the reproductive isolation of species. *Proc. Natl. Acad. Sci. U.S.A.* 98, 12084–12088.
- Obbard, D.J., Gordon, K.H., Buck, A.H., Jiggins, F.M., 2008. The evolution of RNAi as a defence against viruses and transposable elements. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 99–115.
- Oliver, K.R., Greene, W.K., 2009. Transposable elements: powerful facilitators of evolution. *BioEssays* 31, 703–714.
- O'Neill, R.J., O'Neill, M.J., Graves, J.A., 1998. Undermethylation associated with retroelement activation and chromosome remodeling in an interspecific mammalian hybrid. *Nature* 393, 68–72.
- Parsons, C., Salmon, A., Zorjal, T., Tenailon, M., Grandbastien, M.A., Alouche, M., 2009. Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in *Spartina*. *New Phytol.* 184, 1003–1015.
- Pescado, E., Valle, E., Furano, A.V., 1990. Amplification of an ancestral mammalian L1 family of long interspersed repeated DNA occurred just before the murine radiation. *Proc. Natl. Acad. Sci. U.S.A.* 87, 9481–9485.
- Priyapongsa, J., Mariano-Ramirez, L., Jordan, I.K., 2007. Origin and evolution of human microRNA from transposable elements. *Genetics* 176, 1323–1337.
- Pogribny, I.P., Tryndyak, V.P., Woods, C.G., Witt, S.E., Rusyn, I., 2007. Epigenetic effects of the continuous exposure to peroxisome proliferator Wy-14,143 in mouse liver are dependent upon peroxisome proliferator activated receptor alpha. *Mutat. Res.* 625, 62–71.
- Polavarapu, N., Mariano-Ramirez, L., Landisman, D., McDonald, J.F., Jordan, I.K., 2008. Evolutionary rates and patterns for human transcription factor binding sites derived from repetitive DNA. *BMC Genomics* 9, 22E.
- Pregraves, D.C., Balagopal, L., Almay, S.M., Orr, H.A., 2003. Adaptive evolution drives divergence of a hybrid instability gene between two species of *Drosophila*. *Nature* 423, 713–718.
- Rangwala, S.H., Dhumal, B., Vamir, C., Ozkan, H., Gobrecht, D.W., Richards, E.J., 2006. Meiotically stable natural epialleles of *Sathu*, a novel *Arabidopsis* retrotransposon. *PLoS Genet.* 2, e30.

- Ray, D.A., et al., 2008. Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*. *Genome Res.* 18, 717–728.
- Robolo, B., Levat, F., Kleine, L.L., Bismont, C., Vieira, C., 2008. Losing helena: the extinction of a drosophila line-like element. *BMC Genomics* 9, 148.
- Richards, E.J., 2008. Population epigenetics. *Curr. Opin. Genet. Dev.* 18, 221–226.
- Scannell, D.R., Byrne, K.P., Gordon, J.L., Wong, S., Wolfe, K.H., 2006. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature* 440, 341–345.
- Schnable, P.S., 2009. The B73 maize genome: complexity, diversity, and dynamics. *Science* 326, 1112–1115.
- Simzelle, L., Izsvak, Z., Ivics, Z., 2009. Molecular domestication of transposable elements: from detrimental parasites to useful host genes. *Cell. Mol. Life Sci.* 60, 1073–1093.
- Strimskis, V., Ramos, K.S., 2008. Activation of human long interspersed nuclear element 1 retrotransposition by benzo(a)pyrene, an ubiquitous environmental carcinogen. *Cancer Res.* 68, 2816–2820.
- Tanorduc, M., et al., 2008. Epigenomic consequences of immortalized plant cell suspension culture. *PLoS Biol.* 6, 2880–2895.
- The Arabidopsis Genome Initiative, 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815.
- Ungerer, M.C., Strakoski, S.C., Zhen, Y., 2006. Genome expansion in three hybrid sunflower species is associated with retrotransposon proliferation. *Curr. Biol.* 16, 8372–8373.
- Vernou, O., Cazetelli, F., Furano, A.V., 1998. Determining and dating recent rodent speciation events by using L1 (LINE-1) retrotransposons. *Proc. Natl. Acad. Sci. U.S.A.* 95, 11284–11289.
- Vieira, C., Lepetit, D., Dumont, S., Bismont, C., 1999. Wake up of transposable elements following *Drosophila simulans* worldwide colonization. *Mol. Biol. Evol.* 16, 1251–1255.
- Waterland, R.A., Jirtle, R.L., 2003. Transposable elements: targets for early nutritional effects on epigenetic gene regulation. *Mol. Cell. Biol.* 23, 5203–5300.
- Waterland, R.A., Jirtle, R.L., 2004. Early nutrition, epigenetic changes at transposons and imprinted genes, and enhanced susceptibility to adult chronic diseases. *Nutrition* 20, 63–68.
- Weil, C.F., 2008. Too many ends: aberrant transposition. *Genes Dev.* 22, 1032–1036.
- Wicker, T., et al., 2007. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8, 973–982.
- Wing, J.H., Choo, K.H., 2004. Evolutionary dynamics of transposable elements at the centromere. *Trends Genet.* 20, 611–616.
- Zeh, D.W., Zeh, J.A., Ishida, Y., 2009. Transposable elements and an epigenetic basis for punctuated equilibria. *BioEssays* 31, 715–726.
- Zhang, J., et al., 2009. Abundance of retrotransposons in the genome of *Arabidopsis thaliana*. *PLoS Biol.* 7, e155.

Title:

Population analyses reveal chromatin modifications induced by a retrotransposon

Authors:

Benjamin Hubert*, Abdou Akkouche*, Rita Rebollo[‡], Nelly Burlet, Marie Fablet, Cristina Vieira

*These authors equally contributed to this work.

[‡] Present adress: Terry Fox Laboratory, British Columbia Cancer Agency, Vancouver, British Columbia, Canada

Affiliations:

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

Key words:

Retrotransposon, Chromatin, DNA methylation, Germline, *Drosophila*,

ABSTRACT: Transposable elements (TE) are major components of genomes, in which they can be found in different proportions. Despite their impact on genome evolution, regulation mechanisms were identified in many species, preventing for TE mutagenic effects. Epigenetic modifications, such as DNA methylation and histone modifications are conserved effectors of TE repression in eukaryotic genomes. In *D. simulans*, the TE proportion is globally low compared to close *Drosophilids* species, and their copy number is variable in natural populations. We analyzed the epigenetic modification associated with the retrotransposon *tirant* and used the polymorphism provided by natural population to show the influence of a TE on the chromatin structure of its insertion site. Our results clearly demonstrate that heterochromatic modifications are triggered by *tirant* insertions, even if located in euchromatic regions. The analysis of a specific insertion of *tirant* in the *tkv* gene shows that this newly heterochromatic/euchromatic boundary leads to the spreading of heterochromatin on the promoting region of the *tkv* gene, correlated with downregulation of its transcriptional activity.

Introduction

Whole genome analyses in the last decades revealed that transposable elements (TEs) are major components of the non-coding fraction of all genomes, and the influence of these repeated sequences on the functioning of a genome has been constantly demonstrated (Kazazian 2004). Indeed, the first consequence of a TE insertion to be proposed was its mutagenic effect, when disrupting coding or regulatory sequences (Kazazian 2004). Then, some TE sequences were found to be "domesticated" by the genomes, meaning that they were partially or totally recruited as genes or promoters, working together with the endogenous sequences (Volff 2006). Lately, it was shown that TEs are targets of epigenetic modifications (Huda & Jordan 2009). Thus, these epigenetic marks regulating the dynamics of a TE in a genome can interfere with the regulation of the host sequences, at the transcriptional and post-transcriptional levels.

What is also striking about TEs is that their relative amounts are highly variable across the branches of the tree of life, even when closely related species are considered. For instance, *Drosophila melanogaster* and *Drosophila simulans*, which diverged recently (2,5 million years ago), different methods show a two to three-fold difference in the global amount of their transposable elements (Drosophila 12 genomes consortium 2007, Vieira et al. 1999). In addition, even within the *D. simulans* species, a high level of copy number variability is observed between natural populations (Vieira et al. 1999, Biémont et al. 2003, Vieira & Biémont 2004). The understanding of the causes and consequences of such differences in TE content, within and between species, is therefore of fundamental interest, given the wide panel of effects TEs can have on the functioning of a genome. In this perspective, we focused on *tirant*, a retrotransposon found in *D. melanogaster* and *D. simulans*, which, in the later, is only present on the chromosome arms of East-African populations, and which transcription is restricted to a Kenyan population named Makindu (Fablet et al. 2006). This natural system, with different insertions between strains, is thus ideal to investigate the regulation of a

transposable element and the effects of the insertions in the chromatin structure. In order to understand the dynamics of *tirant* in these genomes, we analyzed its germline regulation by epigenetic marks, which are known to be frequently involved in TE inhibition such as described (Schotta et al. 2004; Schotta et al. 2002; Ebert et al. 2006; Phalke et al. 2009).

Epigenetic regulation in *Drosophila* occurs via three different mechanisms: DNA methylation, histone modifications and production of small RNAs. DNA methylation, an inhibitory mark located at cytosines, is assumed to be rare in *Drosophila*, restricted to a very short window of time in development (Lyko et al. 2000); it is involved in retrotransposon silencing in *Drosophila* (Phalke et al. 2009). Histone modifications are diverse and constitute a "histone code", combining enhancing and inhibiting marks (Strahl & Allis 2000). In *Drosophila*, the dimethylation of histone 3 at lysine 9 (H3K9me2) is typical of constitutive heterochromatin, the trimethylation of histone 3 at lysine 27 (H3K27me3) is found in facultative heterochromatin, and euchromatin is characterized by the dimethylation of histone 3 at lysine 4 (H3K4me2) (Ebert et al. 2006). Small interfering RNAs are involved in post-transcriptional gene silencing, and TEs are frequently found to be regulated by a particular class of such RNAs, called rasiRNAs (repeat associated small interfering RNAs) (Malone et al. 2009; Brennecke et al. 2008).

In this work we have analysed the impact of specific *tirant* insertions, taking advantage of natural populations of *Drosophila*. A previous study revealed the existence of two *tirant* subfamilies in *D. simulans*, called C and S. Since the C subfamily is the only with detectable levels of mRNA in *D. simulans* populations (Fablet et al. 2006), we focused this work on this particular subfamily. In the subsequent sections, any reference to *tirant* should be interpreted as a reference to the C subfamily. Also, as new TE insertions are transmitted to the progeny only when they occur in the germline, we focussed our analysis on *Drosophila* female gonads. We investigated the levels of DNA methylation and the chromatin conformation, on and around *tirant* copies. We analyzed DNA methylation of the 5'LTR of

tirant by bisulfite sequencing, and found a very low amount of 5-methylcytidine with no conserved methylated positions in clones from all populations, regardless *tirant* expression. Histone 3 post-translational modifications associated with insertion sites of *tirant* copies were investigated by ChIP (Chromatin ImmunoPrecipitation) and quantified by qPCR. Two different insertions in the Makindu and Zimbabwe populations were nonambiguously enriched with the activating H3K4me2 mark, and can correspond to the active copies of the element. The others *tirant* insertions were always enriched with the heterochromatic H3K27me3 modification. The regulation of *tirant* in *D. simulans* therefore appears to depend on heterochromatinization by histone modifications, but not on DNA methylation. In addition, we take advantage of this natural system, with polymorphism between populations in the presence of copies, to study the impacts of *tirant* insertions on chromatin conformation, and demonstrate that the heterochromatic marks found in *tirant* copies spread in the neighbouring host sequences, interfering with the regulation of endogenous genes.

Materials and Methods

Drosophila strains

We worked on fly samples collected from several geographically distinct natural populations *D. simulans* Chicharo (Portugal), Zimbabwe, and Makindu (Kenya)). These strains were maintained in the laboratory at 18°C as isofemales lines. We previously fully characterized these three strains for their *tirant* copies, and precisely know the sequence of their insertion sites (Fablet et al. 2009).

Analysis of *tirant* expression

Twenty-five pairs of ovaries were dissected in PBS. Total RNA of dissected tissues was extracted using Rneasy-Kit (Qiagen). Complementary DNAs were produced by reverse transcription of 1µg of RNA using the thermoscript RT-PCR system (Invitrogen) and an oligo dT primer according to the standard procedure. The resulting cDNAs were used as templates for real-time PCR. cDNAs were diluted 50 times and then quantified by SYBR Green 1 mix on a LightCycler 480 (Roche diagnostics) using primers specific to *tirant* genes (Env_f:ACACGTTCCCTGAACAGACG, Env_r:GAACGTTACCAATCCGAGCA, Gag_f:CCTTTCGAGCACCTTCACAT, Gag_r:GGGGGTTTGTCTGGGATAGT). The amount of transcripts was estimated relative to those of the housekeeping gene, RP49, using the following primers Fw: CGGATCGATATGCTAAGCTGT, and Rv: GCGCTTGTTGATCCGTA. Measurements were done in two independent experiments, *i.e.*, RNAs were extracted from two different pools of ovaries. For each experiment, reactions were done in duplicates and standard curves were calculated on serial dilutions of genomic DNA.

DNA methylation analysis

Bisulfite treatment

For each of the studied natural strains, ovaries and testes were dissected, from 20 four day-old adult females and 35 four day-old adult males, respectively, and DNA was

extracted. Each extraction was split into two samples (technical replicates), which were independently treated with the EZ DNA Methylation Direct kit (Zymo Research), according to the manufacturer's instructions. The procedure was performed twice for each strain. After bisulfite treatment, we amplified *tirant* 5'LTRs by PCR using the following degenerated primers: TirBT_Fw 5' GAAAAYGGAGGAYTYGAGTAG 3', and TirBT_Rv 5' CATTTRRFACTTCACAACCTTCAATT 3'. The PCR was a 3-step reaction, run in 30 cycles, with an annealing temperature of 43°C. PCR products were cloned (Invitrogen TOPO TA Cloning kit) and sequenced (Genoscreen). As a control, the same procedure was performed on the actine gene, with the following primers: Act5cBT_Fw 5' GGYGGATTTGGYGGAGAYGATG 3', and Act5cBT_Rv 5' CCARATCTTCTCCATATCRTCCTCC 3' (annealing temperature: 56°C).

Sequence analyses

We obtained reference sequence for each studied population using the same TirBT_Fw and TirBT_Rv primers on untreated DNA. We sequenced 10 products per population. A consensus was built for each population. Sequences obtained after bisulfite treatment were compared to these references.

In situ hybridization of tirant RNAs

Ovaries were dissected in PBT (PBS 1X, 0.2% Tween 20), fixed for 20–40 minutes, washed in PBT (PBS 1X, 0.2% Tween20), and then incubated in PBS 1X, 3.7% formaldehyde for 10 min at room temperature. Hybridization (overnight) and washing were performed at 65°C. The *tirant* antisense riboprobe was a 1-kb fragment corresponding to the *env* gene (ENVT7_f: GGTAATACGACTCACTATAGGAACGGGGTTTGATTAACG, ENV3_r: GGAATTAACCCTCACTAAAGGGAGTGTCCAGGTGTGCTG) labelled by T3 *in vitro* transcription with the DIG RNA labelling mix (Roche). DIG-labeled RNA probes were detected using the anti-DIG Fab fragments antibody (1/50 dilution, Roche), with fluorescence amplification (tyramide signal amplification kit, PerkinElmer).

An RNaseH control was done to exclude the possibility of RNA-DNA hybridization in the experiments: after hybridization and washing, ovaries were treated with RNaseH (0.3U/ μ l) for 1h at 37°C, and then submitted to immunodetection. To stain DNA, ovaries were incubated in PBS containing Sytox Green. Visualization was performed with the confocal microscope Zeiss LSM510 Meta.

Chromatin ImmunoPrecipitation (ChIP)

Extraction of chromatin from ovaries and immunoprecipitation (IP) were adapted from Sandmann *et al.* (2006). Each chromatin extraction was performed on 100 ovary pairs. Cell lysis buffer was changed to 5mM PIPES pH8, 85mM KCl, 0,5 % NP40 supplemented with protease inhibitors. Chromatin was sheared with a Bioruptor sonicator water bath (Diagenod, Belgium) for 6 x 30s on / 30s off cycles at high power in order to have random fragments from 500 to 1000 bp. Sheared chromatin was incubated overnight at 4°C with antibodies recognizing H3K9me2 (Millipore 07441), H3K27me3 (Millipore 07449), H3K4me2 (Millipore 07030), H3 (Abcam ab1791), and rabbit IgG (Sigma-Aldrich I5006). The antigen-antibody complexes were washed as described before (Sandmann *et al.* 2006) but second washing solution was modified as follows: TE 2X, 500 mM NaCl, 1 % Triton, 0,1 % SDS.

In order to quantify each IP, real time PCR was performed using QuantiTect™ SYBR Green PCR kit (Qiagen, France) on a MXP3000P PCR system (Stratagene, France). Reactions were done in duplicates and standard curves were calculated on serial of input chromatin. To evaluate the relative enrichment of *tirant* insertion sites after IP, we calculated the difference in cycles between the IP-enriched sample and the input DNA for *tirant* copies and the RP49 gene was used as a control. We sequenced qPCR products for each set of primers to be confident that the signals are specific. The primers used for each *tirant* insertion site are given in Supplementary Table.

For data analysis, we used the following normalization method: we subtracted the signal obtained with IgG, which corresponds to noise, and normalized with the signal obtained

with H3, which reflects the amount of nucleosomes present in the sample.

Characterization of the insertion of tirant inside tkv intron

From a previous analysis, we know the sequences of all insertion sites of *tirant* in the genomes of the Makindu, Zimbabwe and Chicharo populations of *D. simulans* (Fablet et al. 2009). In particular, in the Makindu population, one full-length copy of *tirant* is inserted into the intron of the *tkv* gene (CG14026, FBgn0003716). We decided to focus on this insertion to study the consequences of the presence of *tirant* on the genomic neighbouring.

We compared the transcriptional level of the *tkv* gene in populations with or without the insertion of *tirant* by RT-qPCR, with the same protocol as described above, and the following primers: *tkvD_F*: CATCAGCCTTTGGCCCA, *tkvD_R*: CACGGACAGGGAGATGATGA.

In the same populations, we analyzed the chromatic environment by CHIP (protocol as described above) around the *tkv* promoter that was identified nearest to *tirant* insertion. Primers are given in supplementary material.

Results

***Tirant* is transcribed in the Makindu population**

The first step in the mobilization of a retrotransposon is the transcription of its sequence into an RNA molecule. Then, if retrotransposition occurs in the germline, it can be vertically transferred to the progeny, and potentially become fixed in a population. Therefore, we looked for *tirant* transcription in the ovaries of three populations of *D. simulans*, differing in their euchromatic load of *tirant*: Makindu (Kenya, 5 copies), Zimbabwe (2 copies) and Chicharo (Portugal, 0 copy). In a previous work, we showed however that full-length copies of *tirant* were found in heterochromatin of all of these three populations, except Zimbabwe (Fablet et al. 2006; Fablet et al. 2009). The production of *tirant* mRNAs was quantified by RT-qPCR with amplicons from the three coding regions (Figure 1). Significant levels of transcription are only detected in the Makindu population with an increase fold up to 400 depending on the region analyzed, relative to the other strains (figure 2).

We located *tirant* transcripts in the three populations by *in situ* hybridization experiments in ovaries using a *tirant* element antisense riboprobe. As expected with the RT-qPCR results, *tirant* sense transcripts were only detected in the Makindu population, located in the follicle and nurse cells at different stages and cannot be detected in the oocyte (Figure 3). The observed localization for *tirant* transcripts in the germline is consistent previous published results for the I element in inductive lines (Chambeyron et al. 2008).

DNA methylation of *tirant* sequences in the germline

While the inhibiting effects on gene expression of DNA methylation is clearly established in mammals and plants, the extent and impacts of this process is vigorously debated in *D. melanogaster* (Weyrich et al. 2008; Gou et al. 2010). According to some studies, fly DNA is showed to be very weakly methylated, from 0,1% of methylated cytosines in adults to 0,4% in embryos, and DNA methylation is assumed to be devoid of any functional role

(Schaefer & Lyko 2010), while in other reports, it is demonstrated to be quantitatively significant and to play a role in retrotransposon silencing and gene silencing in somatic tissues (Phalke et al. 2009; Gou et al. 2010).

We investigated the DNA methylation profile of *tirant* copies in ovaries tissues by bisulfite analysis. The primers amplified a region stretching from the 5'LTR to the internal 5'UTR, excluding solo LTRs from the analysis (figure 1). In each sampled population, we found weak levels of DNA methylation, consistent with the currently considered levels, and no obvious pattern for the position of the methylated cytosines (figure 4). Even if a role for DNA methylation in TE silencing cannot be ruled out in *Drosophila*, it does not seem to act on the regulation of *tirant* in the ovaries.

Histone modifications on *tirant* sequences

Histones are the proteins around which the DNA molecule is twisted in chromatin. According to the chemical modifications of their N-term tails, histones can modulate the structure of chromatin, from a very open conformation in euchromatin, usually required for gene expression, to a very dense compaction in constitutive heterochromatin, which is considered to be genetically inactive.

We investigated the conformation of chromatin around *tirant* promoter regions by site-specific Immunoprecipitation of chromatin of histone modifications and then compare modifications associated with *tirant* copies inserted in different genomic compartments, which are representative of the insertions found in the natural populations analyzed. Although multiple modifications are listed, we focused on three major ones, which are unambiguously described in *Drosophila* to be marks of euchromatin (H3K4me2), constitutive heterochromatin (H3K9me2), and facultative heterochromatin (H3K27me3) (Ebert et al. 2006). The most prevalent enrichment for each *tirant* insertion corresponds to H3K27me3, a mark of facultative heterochromatin (Figure 5). In addition, in the Makindu population, two insertions are enriched with H3K4me2, a signature for active genetic sequences, which is not

found for *tirant* copies in the other populations (one copy in the Zimbabwe population shows a strong enrichment for H3K4me2, but it does not correspond to a full-length element).

Impacts of *tirant* insertions on chromatin conformation

The question then arises whether *tirant* preferentially inserts into H3K27me3-enriched region or triggers this mark on its sequence after the insertion event. In this respect, we focussed on two full-length *tirant* insertions in the Makindu population, which are located in one intron of the *tkv* gene and on the euchromatic X chromosome. These Makindu-specific insertions are not found in the genomes of the other populations analyzed in this study. The ChIP experiments of the insertion sites in Makindu show the enrichment in H3K27me3 which could not be found in the homologous empty sites in the other populations (Figure 6). If we parsimoniously consider that the chromatin conformation was the same at this locus in the different populations before the insertion event, then this result clearly demonstrates that the insertion of *tirant* itself triggers a heterochromatic mark, changing the chromatin structure.

Since one of the Makindu insertions is located in the intron of the *tkv* gene, we analyzed the histones modifications associated with the proximal promoter of this gene (see supplementary Figure 2) and showed that significant levels of H3K27me3 are associated with this sequence (FIGURE 7).

As *tirant* insertion impacts on the chromatin conformation of the *tkv* gene promoter we analyzed by RT-qPCR its expression in the different populations, and found that the presence of *tirant* is correlated with 10 fold reduction in the expression of the *tkv* gene (Figure 8).

Discussion

Tirant is a gypsy-like LTR retrotransposon of *Drosophila*, found in a variable number of euchromatic insertions in *D. simulans* natural populations. While most worldwide populations do not display any euchromatic site, *tirant* is found at low copy number in East African populations, with a maximum of 5 euchromatic insertions in the Makindu population of Kenya. This population is the only one for which *tirant* transcription is significant by RT-qPCR. In addition, the observation of *tirant* transcripts by *in situ* hybridization in ovaries indicates that the observed transcription level is of biological relevance. The presence of hybridization signal in the follicle cells is comparable to what was described for the gypsy and ZAM LTR retrotransposons in permissive strains (Leblanc et al. 2000; Péliesson et al. 1994), and in the nurse cell with the I element (non-LTR retrotransposon) in inductive lines (Chambeyron et al. 2008). Thus, the Makindu population shows some similarities with these lines of *Drosophila*, and crosses between this natural strain and strains with no euchromatic copies of *tirant* should therefore be investigated to search for hybrid dysgenic phenotypes.

Natural populations of *D. simulans* provide a powerful tool for the analysis of a TE own regulation, but also to understand how such sequences affect the chromatin of the insertion site. Our system handles with genomes without euchromatic copies of *tirant* (most of the *D. simulans* populations), and a genome with transcribed euchromatic copies (*D. simulans* from Makindu). In order to understand the dynamics of *tirant*, and the extent to which the insertion of a TE can impact the genome neighbourhood, we focused on the methylation of *tirant* sequences and the chromatin conformation at its insertion sites.

DNA methylation does not participate in tirant regulation in the ovaries.

Cytosine methylation of the DNA sequence is an inhibiting epigenetic mark, clearly demonstrated to act on the silencing of TEs in mammals and plants (Chan et al. 2005; Huda &

Jordan 2009). Although controversial in *Drosophila*, a role of DNA methylation in controlling TE transcription in the somatic tissues from embryos was proposed for the *invader4* element (Phalke et al. 2009). In addition, a recent study of the SETDB1 histone H3 methyl-transferase in *Drosophila* shows that DNA methylation is promoted by the H3K9-me3 modification and involved in retrotransposon silencing in the same tissue (Gou et al. 2010). In the present work, the amount of DNA methylation on *tirant* 5'LTR sequences was low and not significantly different between populations, whether they show or not transcripts of *tirant*. We conclude that DNA methylation does not participate in the transcriptional regulation of *tirant*.

Tirant is transcriptionally regulated by heterochromatinization

The transcription of a genomic sequence is under the dependence of its chromatin structure (Jenuwein & Allis 2001) and it is generally admitted that genes that are in open chromatin are transcriptionally active. The level of chromatin compaction is "encoded" by various chemical modifications on the N-terminal tail of the histones, around which DNA is twisted in eukaryotic genomes. The dimethylation of histone 3 at lysine 9 (H3K9me2) is typical of the strongly compacted constitutive heterochromatin, the trimethylation of histone 3 at lysine 27 (H3K27me3) is found in facultative heterochromatin, and relaxed euchromatin is characterized by the dimethylation of histone 3 at lysine 4 (H3K4me2) (Ebert et al. 2004; Ebert et al. 2006).

We show here that, in every population *tirant* insertions are associated with enrichment in H3K27me3 modifications which could explain low levels of transcription in these populations. In addition, in the Makindu population, two copies are enriched with H3K4me2 activating marks, which is consistent with the transcriptional activity revealed in this population. One or both of these insertions can thus be the origin of the transcripts detected in the population by RT-qPCR. The copy located on the X chromosome, that also has lower enrichment of H3K27me3, when compared to the insertion on the intron of the *tkv* gene,

is probably the only expressed copy in the Makindu population. This is consistent with what we previously showed for *tirant* in somatic tissues (Fablet et al. 2009a), and we demonstrate again that bivalent chromatic marks, H3K27me3 and H3K4me2 at the same locus, are found in *Drosophila*.

Bivalence for chromatin marks has only been described until now in mouse (Bernstein et al. 2006) and studies on *Drosophila* cells culture couldn't show this effect (Gan et al. 2010).

It was also shown that retrotransposons are highly enriched in H3K9me2 in *D. melanogaster*, however, *tirant* is remarkably weakly enriched in this mark (Yasuhara & Wakimoto 2008), suggesting that *tirant* can escape from inhibition of this chromatin modification. Some evidences showed that H3K4me2 and H3K9me2 are mutually exclusives H3 modification (S Rea et al. 2000). Our results for the centromeric insertion #4 show significant equal level of enrichment for both of these marks, but the possibility that two distinct nucleosomes cover this region cannot be excluded.

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

By comparing the chromatin state of the same locus in different populations with or without an insertion of *tirant*, we demonstrate that *tirant* triggers facultative heterochromatin on its insertions, rather than is inserted preferentially into heterochromatin. In addition, we show that the insertion #2 in Makindu population inside the intron of the *tkv* gene downregulates its transcription level. This gene is a decapentaplegic receptor involved in development in *Drosophila* (Brummel et al. 1994; Terracol & Lengyel 1994). We know that *tirant* is transcribed in the Makindu population, and we can assume that insertions #1 is the only one expressed copy of *tirant* in this population. All together, these results show again

how a TE can interfere with the host genome, by way of modifying chromatin conformation, and provide strong evidences for the identification of the expressed copy in the Makindu population.

A post-transcriptional repression of *tirant* in the Makindu population?

Heterochromatinization then appears to repress the expression of *tirant* in most populations worldwide, even those with euchromatic insertions. At least two copies escape from this silencing in the Makindu population, and are potentially responsible for the observed transcription. By *in situ* hybridization on ovaries, we show that *tirant* RNAs in the Makindu population are localizing in the nurses cells nuclei in the germline but not detected the oocyte. The mechanisms controlling TE at the post-transcriptional level in *Drosophila* are involving piRNA and the identification of such regulators in the natural population of *D. simulans* would provide a better understanding of the regulation of TE in this species.

Tables and Figures

1. *Tirant* structure
2. *Tirant* expression in ovaries of natural populations analyzed by RT-qPCR
3. *Tirant* RNA *in situ* hybridization ovaries
4. *Tirant* DNA methylation by bisulfite sequencing
5. Chromatin state at *tirant* insertion sites analyzed by ChIP in natural populations of *D. simulans*
6. Comparison of chromatin states at loci with or without *tirant* insertions, analyzed by ChIP
7. Analysis of the *tirant* insertion in the *tkv* gene
8. *tkv* gene expression
9. Supplementary:
 1. Supplementary table 1: Complete list of primers used for this study
 2. *Supplementary figure 1: Structure of the tkv-tirant* insertion

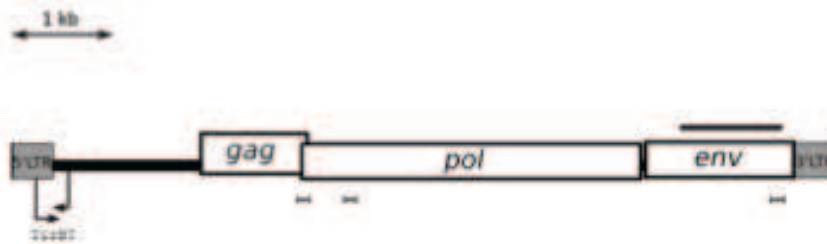


Figure 1. *Tirant* structure

This structure corresponds to a full-length reference copy of *tirant* extracted from the sequenced genome of *D. melanogaster* (Fablet et al. 2006). Below the chart are the positions of the amplicons obtained by RT-qPCR. Above the chart is position of the *env* probe used in the RNA *in situ* hybridization experiment.

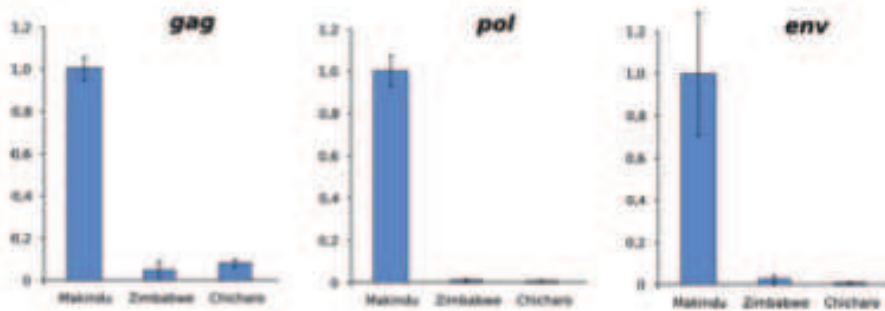


Figure 2. *Tirant* expression in ovaries of natural populations analyzed by RT-qPCR

The positions of the corresponding amplicons are given in Figure 1. The results are given after normalization by the *rp49* reference gene. *Tirant* appears to produce RNAs at a significant level only in the Makindu population.

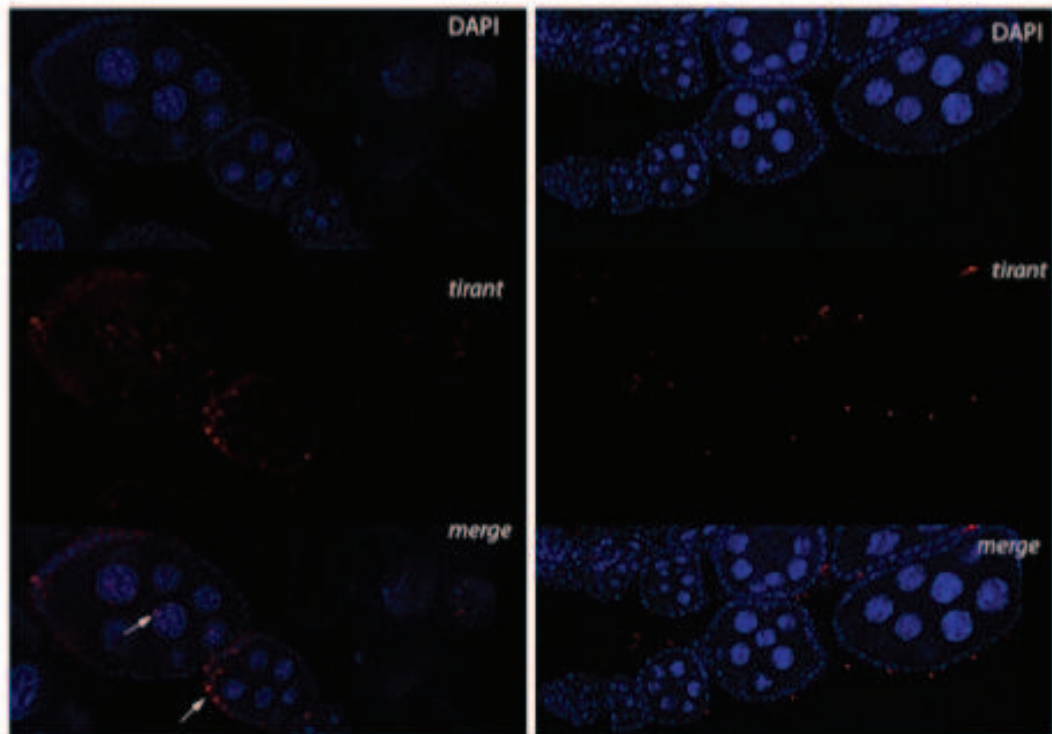


Figure 3. *Tirant* RNA *in situ* hybridization ovaries

Tirant sense transcripts accumulate in Makindu strains. The results were obtained by *in situ* RNA hybridization with *tirant* riboprobes. *Tirant* retroelement sense transcripts were detected by a rhodamine-conjugated antibody (red) against a digoxigenin-labeled antisense riboprobe. DNA was stained with Dapi (blue). Ovaries were dissected from Makindu and Chicharo populations. The arrows show the accumulation of sense transcripts in the follicle and nurse cells.

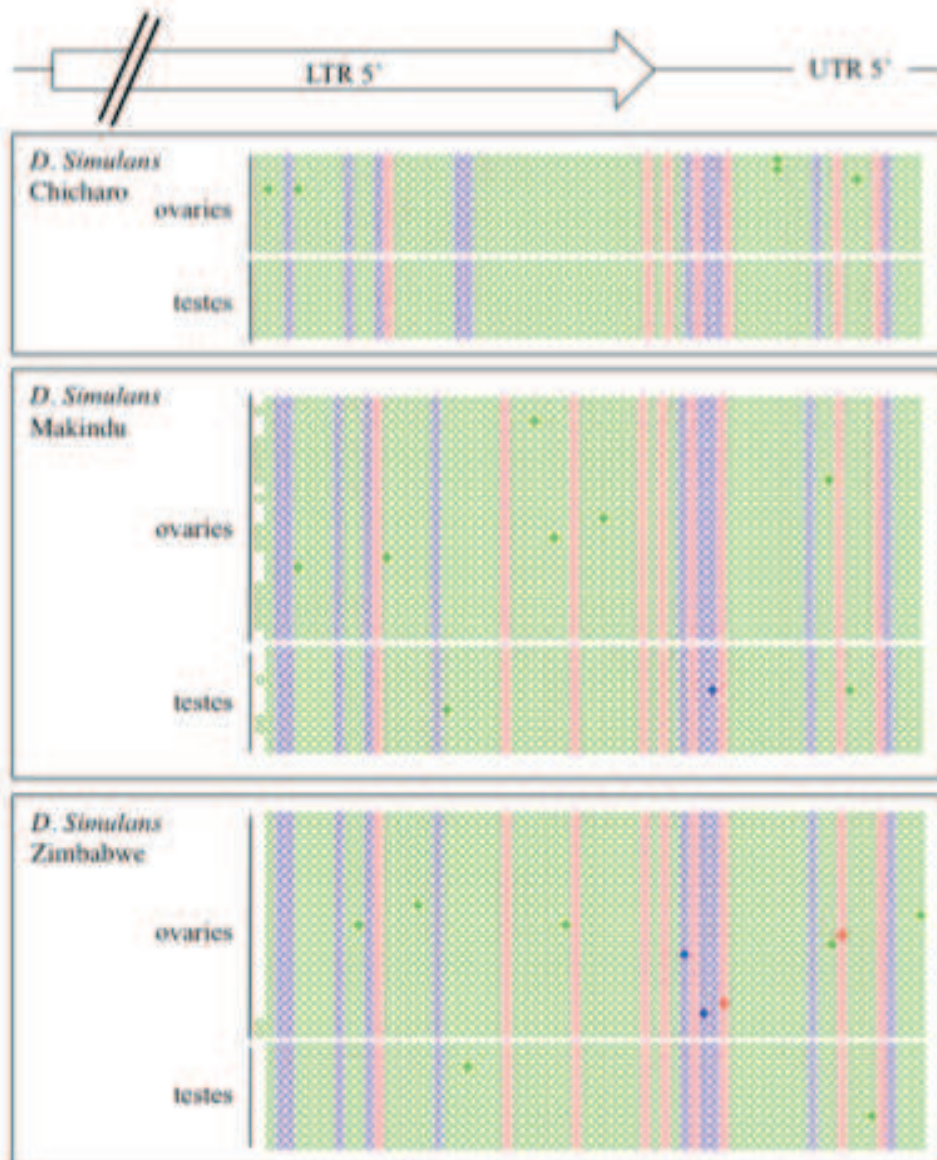


Figure 4. Analysis of tirant DNA methylation by bisulfite sequencing. DNA from Ovaries and testes was bisulfite treated, PCR amplified and clones were sequenced, then compare with reference sequences. These alignments of the cytosines of all the sequenced clones is represented here, and were established using Kismeth (<http://katahdin.mssm.edu/kismeth/revpage.pl>). Each circle represent a single cytosine in trinucleotide context CHH (green), CHG (blue) and CG (red). If methylated, the circle is filled.

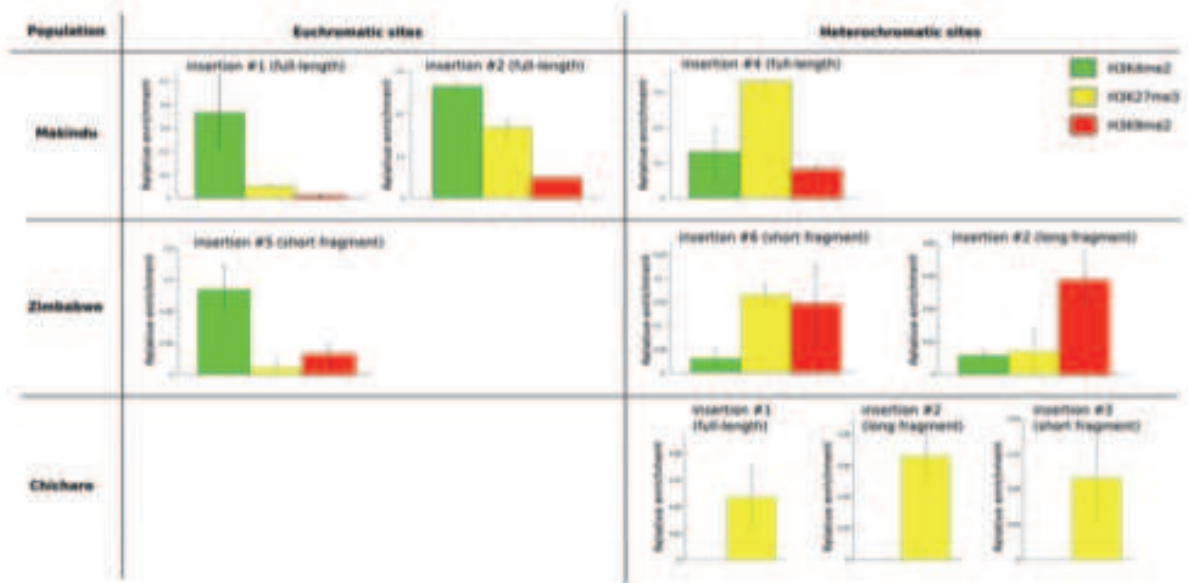


Figure 5. Chromatin state at *tirant* insertion sites analyzed by CHIP in natural populations of *D. simulans*

Presented results are relative enrichments. Data were normalized as described in the Material and Method section. The numbering of the insertions and their classification as euchromatic or heterochromatic correspond to what was presented in the previous study by Fablet et al. (2009). The length of each insertion is given in brackets.

All insertion sites are enriched with the H3K27me3 mark, characteristic of facultative heterochromatin. This mark is predominant for most *tirant* insertions identified as heterochromatic by Fablet et al. (2009), while a more prevalent enrichment in H3K4me2 is observed for *tirant* insertions identified as euchromatic by the former study.

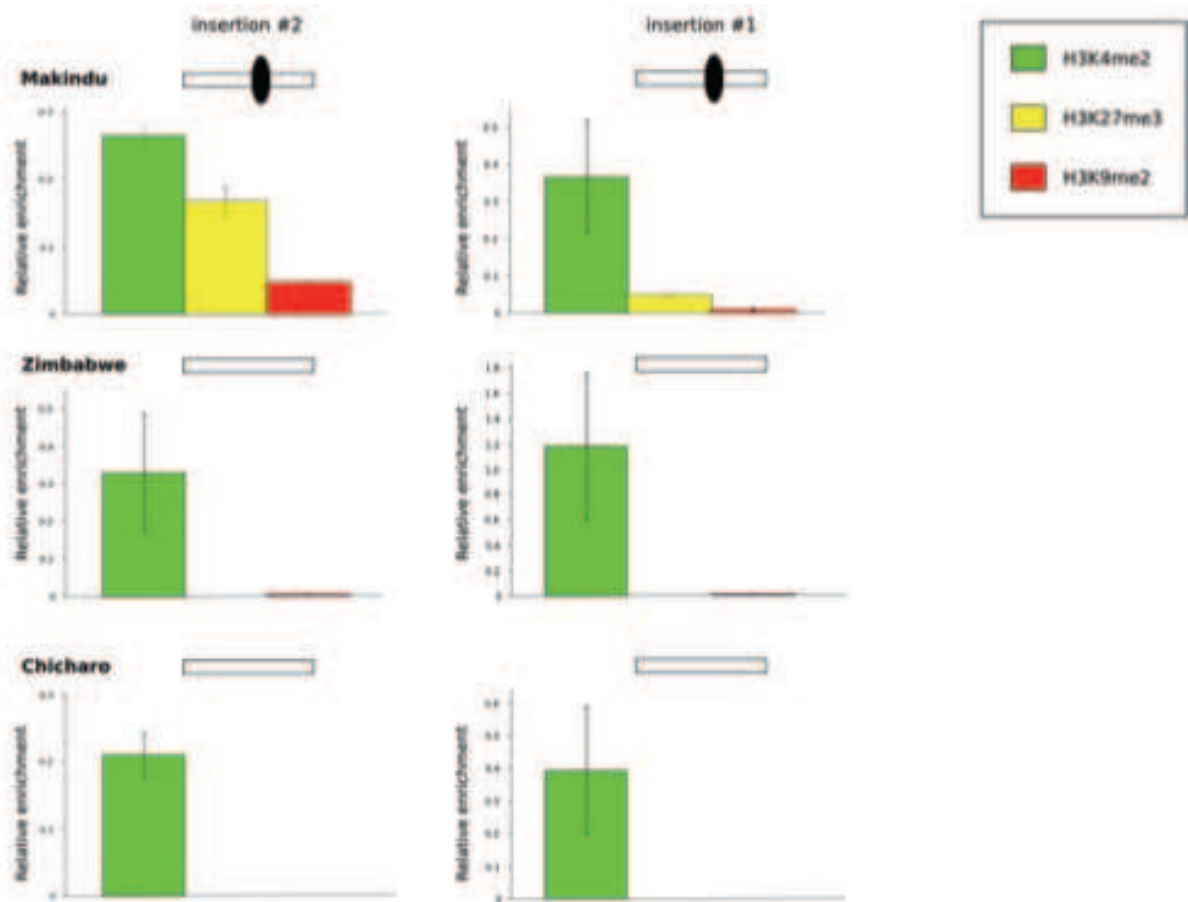


Figure 6. Comparison of chromatin states at loci with or without *tirant* insertions, analyzed by ChIP

Presented results are relative enrichments. Data were normalized as described in the Material and Method section. The numbering of the insertions corresponds to what was presented in the previous study by Fablet et al. (2009): insertion #1 is located on the euchromatic X chromosome, and insertion #2 is found inside an intron on the *tkv* gene (chromosome 2). Both insertions are full-length elements in the Makindu population. The white bars represent the studied loci, and the black ovals stand for *tirant* insertions.

The presence of *tirant* at a given locus is associated to an enrichment in H3K27me3 marks, compared to the euchromatic empty sites, which only show the H3K4me2 mark.

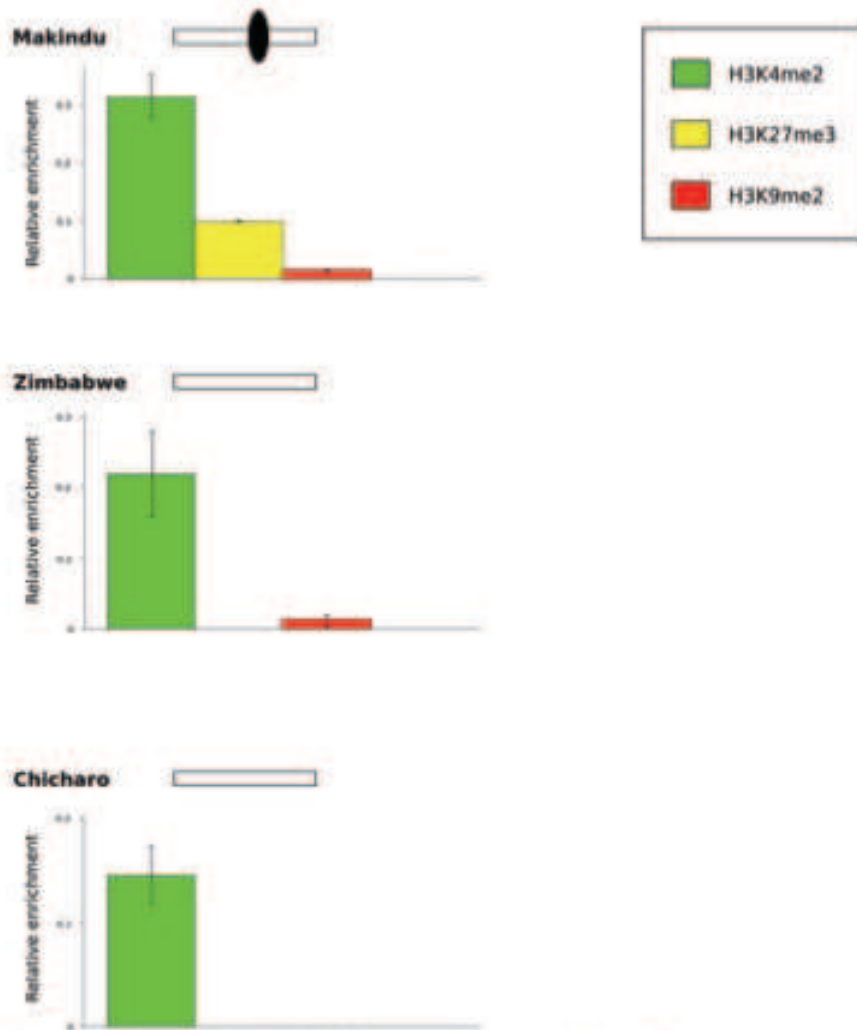


Figure 7. Analysis of the promoter of the *tkv* gene

Presented results are relative enrichments. Data were normalized as described in the Material and Method section. The promoter of the *tkv* gene is located 2,5 kb upstream the *tirant* insertion. The presence of *tirant* is associated with the presence of the H3K27me3 mark. Both insertions are full-length elements in the Makindu population. The white bars represent the studied loci, and the black ovals stand for *tirant* insertions.

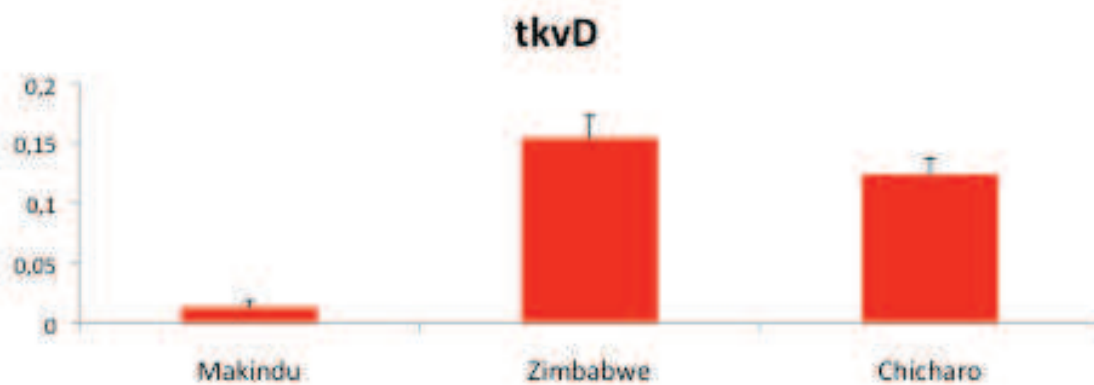


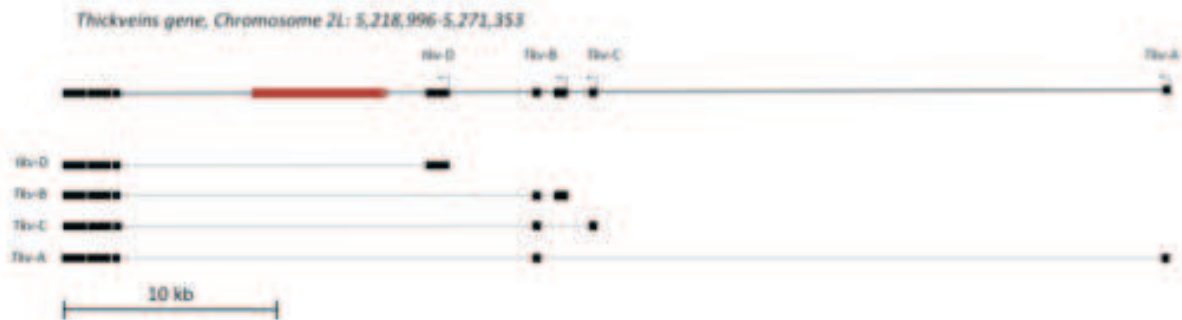
Figure 8. *tkv* gene expression by RTqPCR. The transcription level of the *tkv* RD gene was determined by RT-qPCR from three independent RNA extractions, and was normalized with the RP49 reference gene. Error bars represent SD.

Supplementary

Supplementary Table 1
Primers used for ChIP-qPCR

Population	Locus	Forward Primer	Reverse Primer
Makindu	C4 insertion	caccgagtgaaaactgctgaa	atgttgcaagtgcgacttttg
	C2 insertion	ctgagcacttgatttgggcttagac aggc	gtgtccagttgccgtcttc
	C1 insertion	tgtgcggatttctactgttttc	agcataatgaacatgccgact
Zimbabwe, Chicharo	C1 empty site	gcatccaatgcgaacaagaa	aggccgcagttccacagtt
Zimbabwe, Chicharo	C2 empty site	gcctttttgtgtgcccaact	gcgttgctctaagggga
Makindu, Zimbabwe, Chicharo	Rp49	ctggtttccggcaaggtatgt	cagttcaactcaaaaccgcaaaa
Makindu, Zimbabwe, Chicharo	TkvRD	tttaaaccgccagcaagtgtg	acgactcgcaagaacgtat

Supplementary figure 2. Structure of the *tkv/tirant* insertion.



References

- Drosophila 12 genomes consortium, 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature*, 450(7167), p.203-218.
- Bernstein, B.E. et al., 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, 125(2), p.315-326.
- Biémont, C. et al., 2003. Worldwide distribution of transposable element copy number in natural populations of *Drosophila simulans*. *Evolution; International Journal of Organic Evolution*, 57(1), p.159-167.
- Brennecke, J. et al., 2008. An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science (New York, N.Y.)*, 322(5906), p.1387-1392.
- Brummel, T.J. et al., 1994. Characterization and relationship of Dpp receptors encoded by the saxophone and thick veins genes in *Drosophila*. *Cell*, 78(2), p.251-261.
- Chambeyron, S. et al., 2008. piRNA-mediated nuclear accumulation of retrotransposon transcripts in the *Drosophila* female germline. *Proceedings of the National Academy of Sciences of the United States of America*, 105(39), p.14964-14969.
- Chan, S.W., Henderson, I.R. & Jacobsen, S.E., 2005. Gardening The Genome: DNA Methylation in *Arabidopsis thaliana*. *Nat Rev Genet*, 6(7), p.590.
- Ebert, A. et al., 2006. Histone modification and the control of heterochromatic gene silencing in *Drosophila*. *Chromosome Research*, 14(4), p.377-392.
- Ebert, A. et al., 2004. *Su(var)* genes regulate the balance between euchromatin and heterochromatin in *Drosophila*. *Genes & Development*, 18(23), p.2973-2983.
- Fablet, M. et al., 2009a. Genomic environment influences the dynamics of the tirant LTR retrotransposon in *Drosophila*. *The FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology*, 23(5), p.1482-1489.
- Fablet, M. et al., 2009b. Genomic environment influences the dynamics of the tirant LTR

retrotransposon in *Drosophila*. *The FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology*, 23(5), p.1482-1489.

- Fablet, M. et al., 2006a. Ongoing loss of the tirant transposable element in natural populations of *Drosophila simulans*. *Gene*, 375, p.54-62.
- Fablet, M. et al., 2006b. Ongoing loss of the tirant transposable element in natural populations of *Drosophila simulans*. *Gene*, 375, p.54-62.
- Gan, Q. et al., 2010. Monovalent and unpoised status of most genes in undifferentiated cell-enriched *Drosophila* testis. *Genome Biology*, 11(4), p.R42.
- Gou, D. et al., 2010. SETDB1 Is Involved in Postembryonic DNA Methylation and Gene Silencing in *Drosophila*. *PLoS ONE*, 5(5), p.e10581.
- Huda, A. & Jordan, I.K., 2009. Epigenetic regulation of Mammalian genomes by transposable elements. *Annals of the New York Academy of Sciences*, 1178, p.276-284.
- Jenuwein, T. & Allis, C.D., 2001. Translating the histone code. *Science (New York, N.Y.)*, 293(5532), p.1074-1080.
- Kazazian, H.H., 2004. Mobile elements: drivers of genome evolution. *Science (New York, N.Y.)*, 303(5664), p.1626-1632.
- Leblanc, P. et al., 2000. Life cycle of an endogenous retrovirus, ZAM, in *Drosophila melanogaster*. *Journal of Virology*, 74(22), p.10658-10669.
- Lyko, F., Ramsahoye, B.H. & Jaenisch, R., 2000. DNA methylation in *Drosophila melanogaster*. *Nature*, 408(6812), p.538-540.
- Malone, C.D. et al., 2009. Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell*, 137(3), p.522-535.
- Pélisson, A. et al., 1994. Gypsy transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the *Drosophila* flamenco gene. *The EMBO Journal*, 13(18), p.4401-4411.
- Phalke, S. et al., 2009. Retrotransposon silencing and telomere integrity in somatic cells of *Drosophila* depends on the cytosine-5 methyltransferase DNMT2. *Nature Genetics*, 41(6), p.696-702.
- Rea, S. et al., 2000. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature*, 406(6796), p.593-599.
- Sandmann, T., Jakobsen, J.S. & Furlong, E.E.M., 2006. ChIP-on-chip protocol for genome-wide analysis of transcription factor binding in *Drosophila melanogaster* embryos. *Nature Protocols*, 1(6), p.2839-2855.
- Schaefer, M. & Lyko, F., 2010. Solving the Dnmt2 enigma. *Chromosoma*, 119(1), p.35-40.
- Schotta, G. et al., 2002. Central role of *Drosophila* SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. *The EMBO Journal*, 21(5), p.1121-1131.

- Schotta, G. et al., 2004. A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes & Development*, 18(11), p.1251-1262.
- Strahl, B.D. & Allis, C.D., 2000. The language of covalent histone modifications. *Nature*, 403(6765), p.41-45.
- Terracol, R. & Lengyel, J.A., 1994. The thick veins gene of *Drosophila* is required for dorsoventral polarity of the embryo. *Genetics*, 138(1), p.165-178.
- Vieira, C. et al., 1999. Wake up of transposable elements following *Drosophila simulans* worldwide colonization. *Molecular Biology and Evolution*, 16(9), p.1251-1255.
- Volff, J., 2006. Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 28(9), p.913-922.
- Weyrich, A. et al., 2008. Mammalian DNMTs in the male germ line DNA of *Drosophila*. *Biochemistry and Cell Biology = Biochimie Et Biologie Cellulaire*, 86(5), p.380-385.
- Yasuhara, J.C. & Wakimoto, B.T., 2008. Molecular Landscape of Modified Histones in *Drosophila* Heterochromatic Genes and Euchromatin-Heterochromatin Transition Zones. *PLoS Genet*, 4(1), p.e16.

Rampant proliferation of a transposable element in an asexual wasp

Ken Kraaijeveld^{1,2*}, Brechtje Zwanenburg¹, Padu Franco¹, Benjamin Hubert³, Cristina Vieira³, Sylvia de Pater¹, Richard Stouthamer⁴, Peter de Knijff² & Jacques J. M. van Alphen⁵

¹Institute of Biology, Leiden University, PO Box 9505, 2300 RA Leiden, The Netherlands

²Department of Human Genetics, Leiden University Medical Center S4-P, P.O. Box 9600, 2300 RC Leiden, The Netherlands

³Université de Lyon, F-69000, Lyon; Université Lyon 1; CNRS, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, F-69622, Villeurbanne, France

⁴Entomology, University of California, Riverside, CA 92521, USA

⁵UMR 6553 ECOBIO, Université de Rennes I, Campus de Beaulieu, Avenue du Général Leclerc, 35 042 Rennes cedex, France

*To whom correspondence should be addressed. E-mail: [\(Rozhkov et al. 2010\)](#)

One-sentence summary: Asexual lineages of a parasitoid wasp have accumulated large numbers of copies of a retrotransposon as predicted by theory.

ABSTRACT

Asexual taxa are predicted to face early extinction because they lack an effective means of containing the proliferation of transposable elements (TEs). We report that asexual strains of a parasitoid wasp harbour over four times as many copies of a *gypsy*-like retrotransposon as sexual strains from the same species. The asexual strains represent at least eight distinct clonal lineages that have been accumulating TEs independently. The difference in copy number was not the result of the infection with parthenogenesis-inducing *Wolbachia*. The accumulation of TE copies may eventually create a significant genetic load and drive the extinction of asexual populations.

The widespread occurrence of sex is one of the most elusive problems in evolutionary biology. The inefficiency of purifying selection in asexual lineages can potentially lead to unchecked proliferation of vertically transmitted transposable elements (TEs), causing their demise (1,2). To test this prediction, we established 25 asexual and 10 sexual isofemale lines of the parasitoid wasp *Leptopilina clavipes*. Asexual populations are infected with a parthenogenesis-inducing *Wolbachia*, while sexual populations are not infected. Sexual and asexual field populations were variable at both microsatellite (16 loci) and mtDNA loci (1516 bp; Fig. 1A & B). In the asexuals, the number of mtDNA nucleotide differences between strains correlated to the pair-wise genetic relatedness estimated from the microsatellite data (Mantel test $r = -0.57$, $P = 0.002$). This shows that multiple female lineages have become infected with *Wolbachia*. Since the sequences for 6 *Wolbachia* genes were identical for all strains, the most likely explanation for is horizontal transmission of *Wolbachia* during its spread.

We identified a partial sequence (1110 bp) of a reverse transcriptase in the genome of *L. clavipes*, showing significant similarity to the *gag-pol* polyprotein domain of retrotransposons from the *gypsy*-Ty3 superfamily. The sequences for two sexual and two asexual strains were identical. Quantitative real-time PCR (qPCR) revealed that the asexual strains had greater than four-fold higher copy numbers than the sexual strains (Fig. 1C). Southern blot analysis produced a single band at approximately 700 bp, which was substantially more intense in the asexual strains compared to the sexual strains ($F_{1,18} = 103.82$, $P < 0.00000001$; fig. S1). Genetically related strains resembled each other in copy number of the *gypsy*-like element (qPCR results vs. microsatellite data; Mantel test:

$r_{\text{asexual}} = -0.38$, $P_{\text{asexual}} = 0.05$, $r_{\text{sexual}} = -0.39$, $P_{\text{sexual}} = 0.02$), indicating that asexual lineages have accumulated copies of this element independently.

To investigate whether the accumulation of TE copies was due to a direct effect of *Wolbachia* infection, we repeated the qPCR analysis using asexual females from which *Wolbachia* had been removed using antibiotics. The difference in copy number between sexual and asexuals was very similar to that in untreated females (Fig. 1C) and the Cp ratios were highly correlated to those of the untreated samples of the same strain ($r = 0.84$, $P < 0.0001$). *Wolbachia* may affect the methylation state of the host genome (3), thereby re-activating methylated TEs. We tested for such an effect by applying bisulfite sequencing to eight samples. There was no evidence that the *gypsy*-like element in *L. clavipes* is methylated in either sexual or asexual strains (fig. S2).

The time taken to accumulate TE copies can be estimated using the formula $y = (1+t)^g$, where y is the proportional increase in copy number, t is the transposition rate and g is the number of generations (4). Assuming a TE transposition rate of 10^{-4} per element per generation, the observed four-fold increase would have taken 14,000 generations in the absence of purifying selection. The asexual strains differed from the sexual strains by an average of 3.35 nucleotides on a total of 275 synonymous mtDNA sites, resulting in an estimated divergence time of 60,000 generations (mutation rate assumed to be 2.1×10^{-7} per base per generation; 5). Latent assortative mate preferences show that the ancestor of the asexual populations remained sexual for a considerable part of this divergence time (6). The time left would have been enough to accumulate a four-fold difference in TE copy number, but only if purifying selection on the asexuals was inefficient. No signs of fitness reduction in asexual strains of *L. clavipes* have so far been detected, but if the asexual strains continue to gain TE copies at this rate, negative fitness effects would be expected in the future.

References and Notes

1. E. S. Dolgin, B. Charlesworth, *Genetics* 174, 817-827 (2006).
2. I. Arkhipova, M. Meselson, *Bioessays* 27, 76-85 (2005).
3. I. Negri, *et al.*, *Proc. R. Soc. B* 276, 2485-2491 (2009).
4. A. Burt, R. Trivers, *Genes in conflict: the biology of selfish genetic elements*. (Belknap Press, Cambridge, 2006).
5. R. Raychoudhury *et al.*, *Heredity* 104, 318-326 (2010).
6. K. Kraaijeveld, P. Franco, B. M. Reumer, J. J. M. van Alphen, *Evolution* 63, 3085-3069 (2009).
7. We thank F. Kraaijeveld-Smit, A. Goudswaard, R. Vossen, N. Pul, B. Zwaan, R. Glas, T. de Jong, M. van der Zee and J. Bast for help and discussions. This study was supported by a NWO-Veni-grant to KK.

Supporting Online Material

Materials and Methods

Figs. S1 to S4

Tables S1 to S2

References

Author contributions

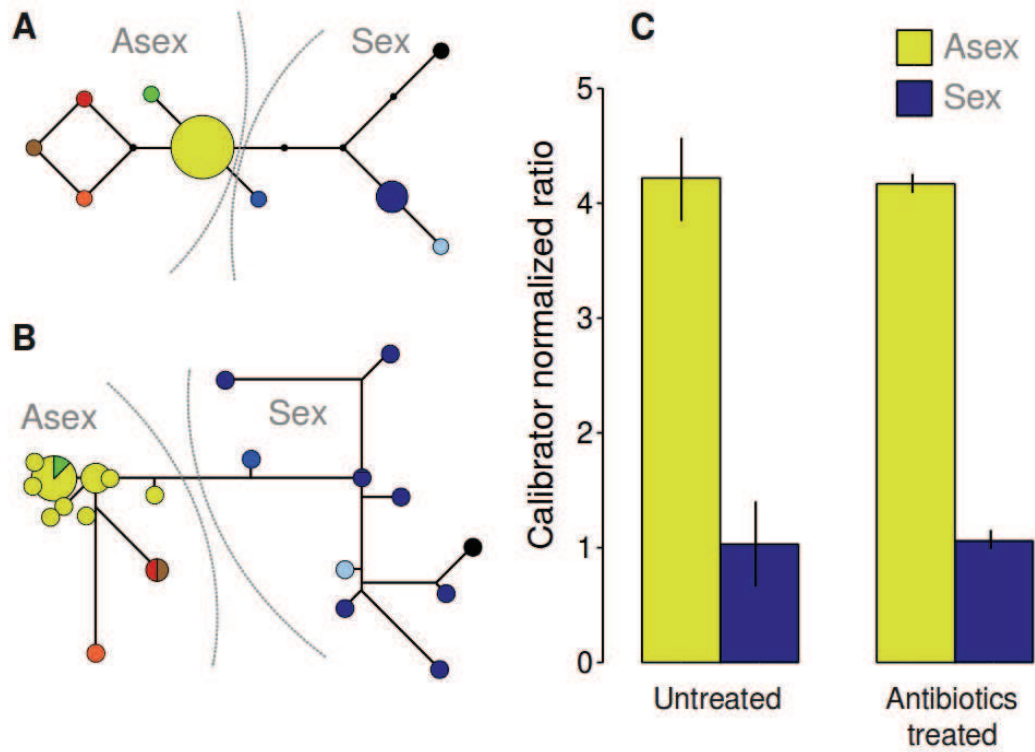


Fig. 1. (A) mtDNA haplotype network of *L. clavipes* strains. Branch lengths are proportional to genetic distance and dot size to the number of strains. (B) Microsatellite genotype network, with colors representing mtDNA haplotypes (as in A). (C) Relative difference in copy number of the *gypsy*-like retrotransposon as measured by qPCR for 10 asexual and 9 sexual strains. Error bars are SEM.

Supporting online material

Materials and Methods

Wasp strains and *Wolbachia* removal. Sexual and asexual strains of *L. clavipes* were collected and cultured as described in (6). The removal of *Wolbachia* using antibiotics is described in (6).

Genetic diversity structure of sexual and asexual populations of *L. clavipes*. We assessed genetic diversity at three levels: nuclear, mitochondrial and *Wolbachia*. We commissioned a genomic library enriched for three types of tandem repeats (CA, GA and ATG; Genetic Identification Services). One hundred colonies were sequenced, from which we selected 16 loci. Loci were amplified using loci-specific primers (table S2; all $T_a = 55^\circ\text{C}$) and run on a MegaBase 1000 (Amersham). Genotype profiles were scored manually. Asexual genotypes were homozygous due to gamete duplication by *Wolbachia*. All but two sexual genotypes were homozygous due to prolonged inbreeding in the lab. The diploid genotypes were therefore treated as haploid in Fig. 1. We sequenced 1057 bp of the mitochondrial gene cytochrome oxidase subunit 1 (CO1) using the primers described in (S1), except the reverse primer for the second part, which was re-designed for this study: TCATCTAAAAATTTTAATCCAGT. We also sequenced 459 bp ND1 using the primers described in (S2). The sequences for the two mitochondrial genes were concatenated using BioEdit and analysed using DNAsp. Haplotype networks were drawn using Network. Last, we sequenced six *Wolbachia* genes (*coxA*, *hcpA*, *ftsZ*, *fbpA*, *gatb*, *wsp* (S3)).

Identification of the *gypsy*-like element. A short fragment of sequence was obtained using the nested set of degenerate PCR primers described in (S4). DNA of the sexual strain DC was used as template for this reaction. The sequence showed significant similarity to *gypsy*-like elements in *Nasonia vitripennis* (accession XM_001601092.1) and *Bombyx mori* (accession AB032718.1). Based on the similarity between these two sequences, new degenerate primers were designed: AARYTNTAYGCNGCNAAY and RTARAARTTNACCATNCC. These were used in PCR reactions (25 μl total volume) with the following conditions: 94°C for 4 min, then 45 cycles of 94°C for 40 s, 50°C for 40 s and 72°C for 30 s and then 72°C for 3 min. The PCR product was cleaned (Promega wizard kit) and cloned with the TOPO-TA cloning kit (Invitrogen). Positive colonies were picked and amplified using standard M13 primers. Clones containing inserts of the expected size (~ 1000 bp) were sequenced. These contained the target sequence within a total of 1110 bp. To compare the sequence between strains, we designed specific primers (TTTACGCTGCAAATGGTTCA and TTGAATCGCTTCGACCTTTT) and used these to sequence the fragment in two sexual and two asexual strains.

Identification of the forkhead gene. The transcription factor *forkhead* was chosen as a control gene to standardize the qPCR signal, as this gene is highly conserved and single-copy in *Nasonia vitripennis*. Sequences for *Nasonia vitripennis* (accession XM_001601750.1), *Trichogramma kaykai* (accession EU650782.1) and *Cirrospilus coachellae* (accession EF137269.1) were aligned using ClustalW. The following degenerate primers were picked using the software Genefisher: RTSACCCTCAACGGCA and AGCCGTTKTCRAACA, cloning and sequencing were performed as described above. Species-specific primers were developed (GGATGCAGAGTCCAGAAGGA and TTGGCAAATTCATTAGGC) for use in qPCR.

Quantitative PCR. For qPCR, we designed a set of primers spanning a 109 bp region in the reverse transcriptase (CGTTCGGTCTGTTTCAATTT and GCGAAACAGAAGTCCAATCC). Genomic DNA was extracted from one recently emerged female per strain (Qiagen blood and tissue kit). Relative copy numbers of the *gypsy*-like TE and forkhead sequences were quantified using the Lightcycler LC-480 qPCR system (Roche). The qPCR reactions contained 5 μl qPCR MasterMix for SYBR Green (BioRad), the equivalent of 60 ng DNA and 300nM of each primer, and were run in duplicates, using the following two-step cycling program: 95°C for 15 s, 60°C for 1 min for 40 cycles. Efficiency of the PCR reactions was estimated from the samples separately for the two genes using the software LinRegPCR. Data from qPCR were analysed using the second derivative maximum method. Here, the C_p -value represents the cycle at which the increase of fluorescence is highest and where the logarithmic phase of a PCR begins. The mean C_p of the duplicates for the TE was calibrated on one of the sexual samples ($\text{PCR efficiency}^{\Delta C_p}$) and then normalized on the control gene.

Southern blot analysis. Genomic DNA was extracted from ten recently emerged females per strain using phenol/chloroform. Ten μg DNA per strain was digested with EcoRI and electrophorized on a 0.7% agarose gel. The DNA was transferred to hybond+ membrane (Amersham Biosciences), which was then probed with the DIG-labelled 1110 bp *gypsy*-like sequence described above. Hybridization and detection were performed using the DIG labelling protocol (Roche Applied Sciences). The intensity of each band was quantified using ImageJ and normalized on the intensity of the marker band. The relative intensities were ln-transformed and compared using ANOVA.

Bisulfite sequencing. Bisulfite treated samples were PCR amplified with primers GGTAATGAAGAAATGYAAAGTAT and CAATCACCRAAAATATTRTTTTTCC to yield a PCR product of 373 bp. The PCR product was cleaned and sequenced following standard procedures. Bisulfite sequencing data were analysed using Kismeth (S5).

Supplementary Figures

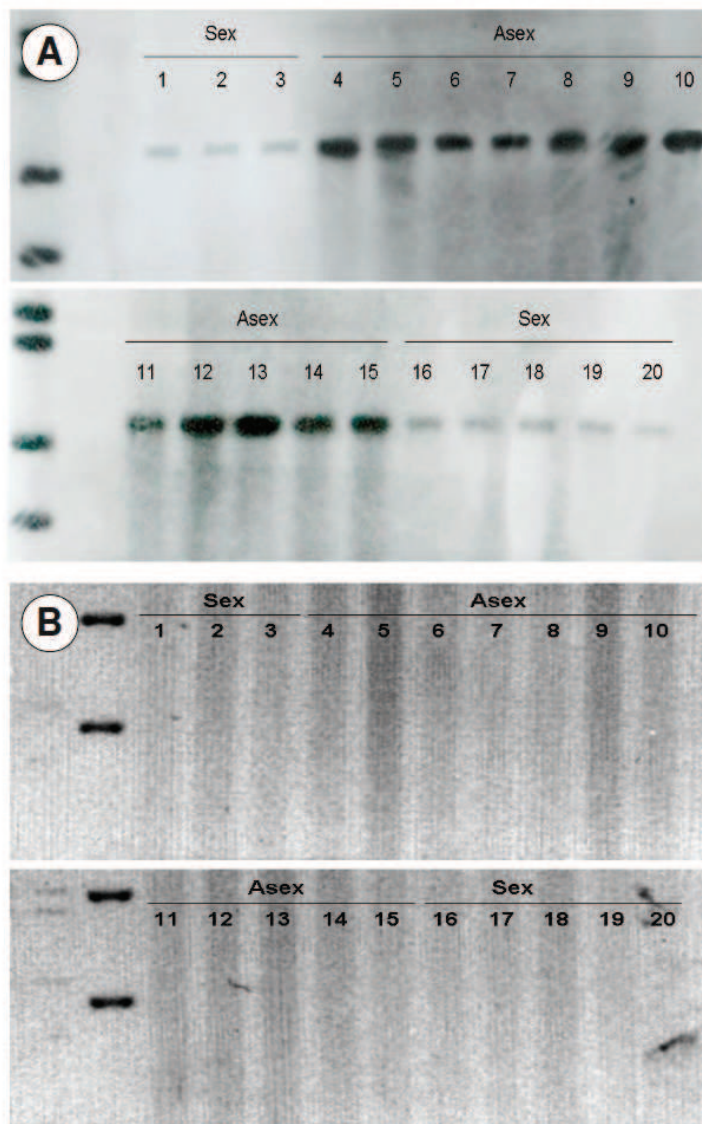


Figure S1: (A) Southern blot of genomic DNA digested with EcoRI and probed with the *gypsy*-like sequence. The first lane contains the size marker (HindIII-digested lambda). Each sample lane contains DNA extracted from ten female wasps from a single strain. (B) Ethidium bromide-stained agarose gel used for the southern blot, showing the relative amounts of DNA loaded in each lane.

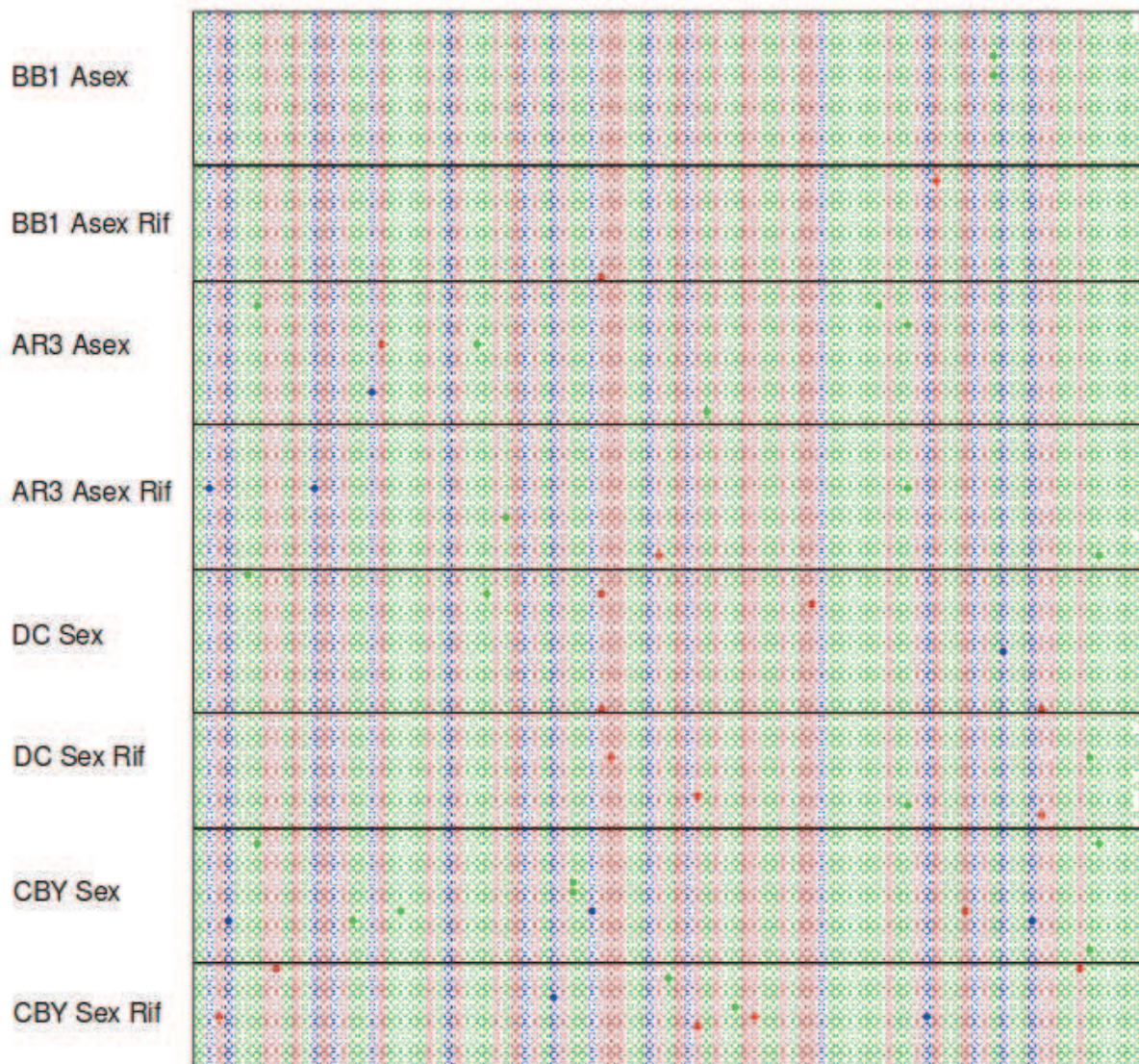


Figure S2: Cytosine methylation patterns of the *gypsy*-like transposable element of *L. clavipes* as determined by bisulfite sequencing. Each line represents a single clone and each circle a cytosine (red: CG, blue: CHG, green: CHH). The circle is filled when the cytosine is methylated. Two to fifteen clones were sequenced per wasp strain. Two asexual strains (AR3 and BB1) and two sexual strains (CBY and DC) were used. Females that had been cured of their *Wolbachia* infection using antibiotics are denoted 'Rif'.

Supplementary Tables

Table S1: Genbank accession numbers.

Locus	Accession numbers
Gypsy-like retrotransposon	HM999657
Microsatellite loci	HM999630 - HM999648
COI (mtDNA)	HM999658 - HM999666
ND1 (mtDNA)	HM999649 - HM999650
<i>Wolbachia</i> coxA	HM999651
<i>Wolbachia</i> hcp	HM999652
<i>Wolbachia</i> ftsZ	HM999653
<i>Wolbachia</i> fbp	HM999654
<i>Wolbachia</i> gatB	HM999655
<i>Wolbachia</i> wsp	HM999656

Table S2: Primers used for microsatellite genotyping.

Locus name	Forward	Reverse
E101	GAGAGTCGTCCTTTGAGTCAC	CGTCGTCGTCAGTGGAGATACT
E108	CGTTAAACCAATCTGTGACG	AATGTGGATGCAAATCAGAG
E3a	CTCTCACTCCAAACTTGTCTG	GAACGAATTAACCCATTGTG
F101	AAACGAAGCTCAGAATACAAGC	GTTGCGTAAAATTGTCAACTTG
F102	TTACCCCTTATTTTCCTCTGTC	TGGGCTCTCTATTCTATTCTCC
F103	ATTCCGTGAACCCAACAC	CGTAATGAAGGAGAGTCTGC
F107	GCGAGAAACGTCTACTAGCTC	TCACCACCTCATTCAATCAC
F112	GGATTGATTTACGACAGACG	CCCAACTCCTGACATTGAG
F118	GGAACAGCCTTTACTGCAC	AATCGTAAGAAAACGCCTAAC
F119	GCATTCTCATTCAAGCATAAAC	CGGTGTATTTCCCAAAGAC
F124	CGGTTTATCGCAGTGATACA	GGACGTTACATTTGTCTCTTCA
F129	CGCCAACAAAGAGATGTC	CGTGTGCGAATAAAATACG
F130	CTGGTTCCAATCTCGTGTG	GCCATCATCATCATCAAGG
F2	GGAATGGGATGGGATGAG	GCTCGGAGAGAGATTACGC
F5	GCACGCGATAACACCAATAG	GCCGATGTTCTTGATGTTTG
G10	ACAGGTCCGACGAACATAAC	AGGGGATTTCTGAACTGGAT

Supplemental References

- S1. S. J. Scheffer, E. E. Grissell, *Mol. Ecol.* 12, 415-421 (2003).
- S2. P. T. Smith, S. Kambhampati, W. Völkl, M. Mackauer, *Mol. Phylogenet. Evol.* 11, 236-245 (1999).
- S3. L. Baldo *et al.*, *Appl. Environ. Microb.* 72, 7098-7110 (2006).
- S4. I. Arkhipova, M. Meselson, *Proc. Natl. Acad. Sci. USA* 97, 14473-14477 (2000).
- S5. E. Gruntman *et al.*, *BMC Bioinformatics* 9, 371 (2008).

Author Contributions

K.K. conceived the project, participated in all aspects and wrote the manuscript. B.Z. identified the *gypsy* transposable element. P.F. sequenced the mtDNA. B.H. and C.V. performed the bisulfite sequencing. S.P. performed the Southern blot analysis. R.S. conceived the phylogenetic part. P.K. analyzed the population genetic data. J.A. supervised the project. All authors discussed and commented on the manuscript.

MATÉRIELS ET MÉTHODES

1. Origine et élevage des souches de drosophiles

Les populations naturelles de *D.melanogaster* et *D.simulans* étudiées durant ma thèse ont été capturées dans la nature et sont élevées au laboratoire. Ces souches sont des lignées isofemelle établies à partir de femelles uniques fécondées dans leur milieu naturel où elles ont été récoltées à différentes dates et par différents expérimentateurs. Les lignées isofemelles sont fortement consanguines. Les lignées mutantes pour le gène dDNMT2 sont dérivées de *D. melanogaster* et ont été amicalement fournies par Gunter Reuter (dDNMT2-/-), Kent Golic (R2), et C.K. James Shen (GS12412).

Les souches sont maintenues au laboratoire dans des étuves à 25°C. Le milieu de culture se prépare de la façon suivante :

Quantité pour 1,2 kg, soit une soixantaine de tubes :

15 g agar-agar + 400 mL d'eau du robinet ont été portés à ébullition en remuant constamment.

110 g de farine de maïs, 115 g de levure de bière sèche, et 500 mL d'eau du robinet ont été mixés jusqu'à obtention d'un mélange homogène, puis versés dans l'agar cuit.

A ce mélange ont été ajoutés 6 g de nipagine dissous dans 47,5 mL d'éthanol 95 %, l'ensemble a été ajusté à 1,2 kg avec de l'eau chaude, et bien mélangé. Stérilisés à l'autoclave pendant 30 minutes à 130°C, le milieu est ensuite homogénéisé et coulé dans les tubes d'élevage.

2. Récolte des embryons

Des mouches adultes ont été mises à pondre dans des boîtes contenant de la gélose, composée de jus de raisin + 3 % d'agar, et partiellement recouverte de levure de boulanger. Deux heures après, les embryons sont récoltés, grâce à un tamis, par rinçage de la gélose au PBS 1X. Les embryons ainsi récoltés sont collectés dans des tubes Eppendorf de 1,5mL contenant du PBS 1X, déchorionés par des lavages à l'eau de javel 1%, puis l'ADN est extrait en suivant le protocole d'extraction d'ADN par phénol chloroforme décrit ci-dessous.

3. Extraction d'ADN génomique

1. Broyer le tissu dans 170 μ L de tampon de broyage (Tris-HCl pH8 50mM + EDTA 50mM + SDS 1%)
2. Ajouter 10 μ L Pk à 10 μ g / μ L puis agiter
3. Incuber pendant 3h à 55°C (puis ajouter 5 μ L RNase A 10 μ g/%L, incuber 5 minutes à 37°C)
4. Incuber à 4°C, 15 minutes
5. Ajouter 36 μ L de KO acétate 5M, vortexer
6. Incuber a 4°C pendant 15 minutes
7. Centrifuger (12 000 rpm, 15 minutes, 4°C)
8. Récupérer le surnageant (SN)
9. Ajouter 36 μ L de KO acétate 5M, agiter (STEP1)
10. Incuber à 4°C, 15 minutes
11. Centrifuger (12 000 rpm, 15 minutes, 4°C)
12. Ajouter 1 volume de Phénol/Chloroforme/IAA, vortexer 15''
13. Centrifuger (12 000 rpm, 5 minutes, 4°C)
14. Récupérer le SN et mettre dans un nouveau tube et recommencer une deuxième fois
15. Précipitation de l'ADN :
 - 1 μ L of glycogène
 - 1/10 vol d'acétate de sodium 3M (pH=5.2)
 - 2.5Vol d'EtOH 95% froid
16. Homogénéiser
17. Incuber toute la nuit à -20°C
18. Centrifuger (14 000 rpm, 30 minutes, 4°C)
19. Jeter le surnageant et laver le culot avec 1 mL d'EtOH 70 % froid
20. Centrifuger (14 000 rpm, 15 minutes, 4°C)
21. Jeter le surnageant et laisser sécher le culot d'ADN à température ambiante pendant 2 heures
22. Reprendre le culot dans 45 μ L H₂O
23. Incuber 1h, 37°C
24. Déposer sur gel d'agarose 1 % pour vérifier la qualité de l'ADN
25. Quantifier l'ADN.

4. Traitement au Bisulfite

Le traitement de l'ADN au bisulfite de sodium est utilisé pour déterminer le profil de méthylation de séquences d'ADN. Les cytosines non méthylées vont être converties en uracile tandis que les cytosines méthylées ne sont pas converties. Une fois l'ADN traité, les séquences d'intérêt sont amplifiées par PCR en utilisant des amorces dégénérées pour toutes les positions de cytosines. Durant cette étape, les uraciles seront remplacés par des thymines. Les produits de PCR sont clonés et séquencés. Pour toutes les expériences de traitement au bisulfite de sodium, des réplicats techniques ont été réalisés, et des réplicats biologiques et techniques ont été réalisés pour les résultats obtenus avec les amorces tir01.

Préparation des échantillons :

- Disséquer à partir d'individus adultes de 4 jours, dans du PBS 1X :
 - o 20 paires d'ovaires pour les femelles
 - o 35 paires de testicules pour les males
- Broyer et aliquoter les ovaires et testicules des volumes respectifs de 3 µL et 9 µL
- Traiter à la protéinase K, en suivant pour les ovaires et pour les testicules, respectivement, la procédure A et la procédure B (de EZ DNA Methylation-Direct™ KIT)
- Traiter au bisulfite de sodium et extraire l'ADN en suivant les conditions standard recommandées par le fabricant (EZ DNA Methylation-Direct™ KIT, ZYMO RESEARCH).

5. PCR de l'ADN traité au bisulfite de sodium

Les amplifications PCR de la région régulatrice de *tirant* à partir de l'ADN traité ont été réalisées avec les amorces répertoriées dans le Tableau 5. Pour chaque PCR, 4 µL d'ADN traité ont été amplifiés dans un volume final de réaction de 50 µL. Le cycle de PCR utilisé est le suivant :

1. Dénaturation (94 °C, 10 minutes)
2. Dénaturation (94 °C, 30 secondes)
3. Hybridation (T_m selon amorce, 30 secondes)
4. Elongation (72 °C, 30 secondes), puis retour étape 2, 30 cycles
5. Finition (72 °C, 10 minutes)

Tableau 5. Amorces PCR utilisées pour l'amplification de séquences d'ADN traitées au bisulfite de sodium.

Bisulfite Genomic Sequencing				
ET/gène	amorce	Séquence (5'-3')	Tm (en °C)	Taille du produit de PCR (pb)
<i>tirant</i> (type C)	Tir01 f	gga-aga-ygg-yaa-ytg-g	43	451
	Tir01 r	car-tct-aca-tct-trc		
	Tir52 f	gaa-aay-gga-gga-yty-gag-tag	51,5	
	Tir52 r	cat-ttr-rta-ctt-cac-aac-ttc-aat-t		
	TirC1 f	agg-tag-gta-att-att-aat-agg-tag-g	52,4	
	TirC1 r	cac-ccc-cta-aac-ccc-cac-rcc-tct-a		
Copie M1	191 f	ayg-gag-tga-aay-tgy-tga-att-g	52,4	328
	191 r	cca-ata-aar-aat-acr-act-rac-ttt-c		
Copie M2	271 f	gaa-aaa-aaa-aag-gga-gag-aga-aaa-a	52,4	408
	271 r	tcc-tcc-rtt-ttc-atc-tta-tac-ata-a		
Copie M3	151f	gga-gya-tgg-agy-aya-agg	59	585
	151r	caa-trt-trc-art-rcr-ac		
	152f	tta-agt-agg-aaa-gtt-ttt-gta-aag-g	52,4	
	152r	cca-ata-aar-aat-acr-act-rac-ttt-c		
<i>tirant</i> (Type S)	TypeSb f	tga-agt-ggy-atg-gaa-att-atg-ttt-a	48,5	463
	TypeSb r	aca-ctt-rca-ttr-tcr-aat-aaa-aat-t		
<i>Roo</i>	Roobisu f	gag-agy-gat-aaa-tta-tat-tta-gga-t	45,3	563
	Roobisu r	tat-tta-tac-tct-rtt-cct-tac-ctt-t		
<i>F</i>	Fbisu f	ggt-tgt-ttt-gag-tga-agt-gaa	43,5	527
	Fbisu r	acc-car-ctt-ttt-trt-tta-crt-ttt-c		
412	412bisu f	aag-gya-aaa-agt-ata-agt-gya-tgg-t	48,5	487
	412bisu r	cca-tac-tca-ata-raa-ctc-tac-tca-c		
Actine 5C	Act5Cf	ggy-ygg-att-tgy-ygg-aga-yg	58	161
	Act5Cr	cca-rat-ctt-ctc-cat-atc-rtc		

6. Clonage/Séquençage/Analyse des profils de méthylation

Après traitement au bisulfite de sodium, les produits de PCR ont été clonés avec le Kit TOPO-TA cloning pour séquençage (Invitrogen) suivant les recommandations du fabricant. Les bactéries transformées sont étalées en boîte de Pétri contenant du milieu de culture L.B. plus ampicilline, incubées à 37 °C pendant 16 heures. La confirmation des colonies positives est réalisée par PCR en utilisant les amorces universelles M13 forward et reverse, qui permettent d'amplifier les fragments d'ADN insérés dans le vecteur. Les produits de PCR correspondant à la taille attendue ont été séquencés par Genoscreen. Les séquences ont été alignées avec CLUSTALW2 (Larkin et al. 2007). Les profils de méthylation ont été analysés avec le logiciel KISMETH (<http://katahdin.mssm.edu/kismeth/revpage.pl>) (Gruntman et al. 2008). Les fréquences de méthylation des dinucléotides ont été calculées grâce à des outils développés (codes pythons) par Hussein Mortada. Le premier script python, appelé Cpos.py

prend comme entrée le fichier texte d'alignement des séquences traitées. Dans ces séquences, le nombre de C va être compté pour chaque position permettant ainsi de calculer le pourcentage de cytosines non converties qui est donc égal, pour chaque position, au rapport du nombre de cytosines divisé par le nombre total de séquences alignées. Le deuxième, appelé Racio.py, prend comme entrée le fichier texte de la séquence de référence (fichier1) et le fichier texte d'alignement des séquences traitées (fichier2). A partir de ces deux fichiers, le pourcentage de C pour chaque séquence du fichier2 est calculé (rapport du nombre de C dans chaque séquence du fichier1 par le nombre de C de la séquence de référence) ainsi que celui des dinucléotides CpA, CpT, CpG et CpC.

Solution : Milieu de culture L.B. (Luria-Bertani) plus ampicilline :

- peptone 1 g
- Extraits de levure 0,5 g
- NaCl 1 g
- Agar 1,5 g
- Ampicilline 0,1 mg/mL
- H₂O qsp 100 mL

7. Digestions enzymatiques

Les ADN génomiques ont été digérés pendant 16 heures à 37°C ou 55°C selon l'enzyme considérée, en accord avec les instructions du fabricant (BioLabs). Les paramètres de digestion utilisés pour chaque enzyme sont indiqués dans le tableau 3. Les couples d'enzymes sensibles/non sensibles correspondant à chaque ET étudié sont indiqués dans le Tableau 6 et les paramètres de digestion dans le Tableau 7. Les amorces PCR utilisées pour diagnostiquer l'état de méthylation de chacun des sites sont référencées dans le Tableau 8.

Tableau 6. Enzymes de restriction utilisées.

Elément transposable	Enzyme sensible	Enzyme non sensible
<i>F</i> site 5'	BsmBI	BsmAI
<i>F</i> ORF1	HpaII	MspI
<i>F</i> ORF2	HpaII	MspI
<i>Roo</i>	AvaI	BsoBI
<i>412</i>	HpaII	MspI
<i>tirant</i>	HpaII	MspI

Tableau 7. Paramètres utilisés pour les digestions enzymatiques en fonction des enzymes utilisées.

Enzyme	Site de restriction	T°C	Nombre d'unités
BsmAI	5'-CGTCTCN-3'	55	50
BsmBI	5'- C GTCTCN-3'	55	10
MspI	5'-CCGG-3'	37	500
HpaII	5'- C CGG-3'	37	150
AvaI	5'-CY C GRG-3'	37	200
BsoBI	5'-CYCGRG-3'	37	40

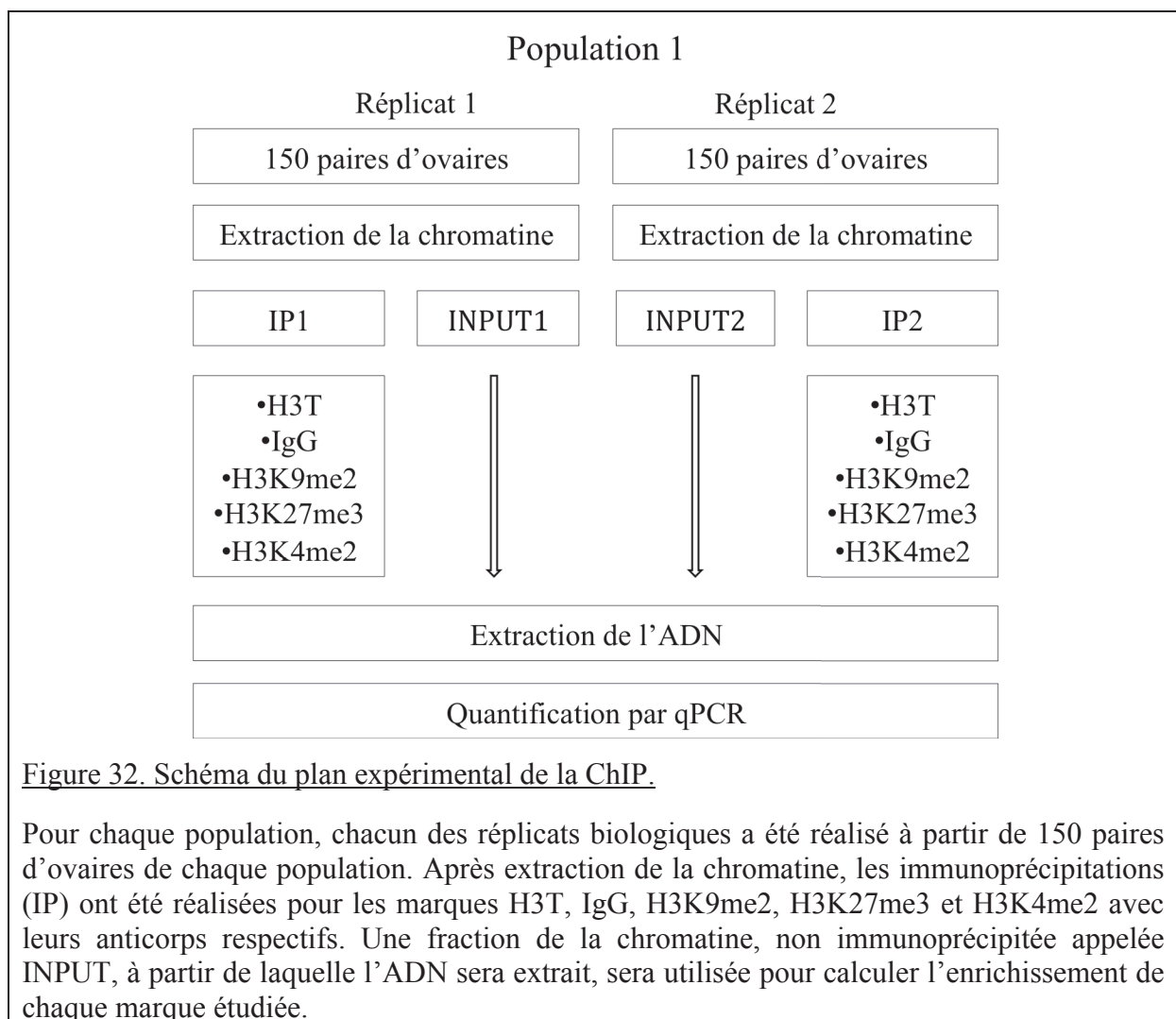
Les cytosines qui bloquent l'activité de clivage de l'enzyme si elles sont méthylées (colorées rouge) sont toutes localisées au niveau de dinucléotides CpG. La température (T°C) ainsi que le nombre d'unités enzymatiques utilisées pour la digestion sont renseignées.

Tableau 8. Amorces PCR utilisées pour diagnostiquer l'état de méthylation des sites de restriction.

Enzymes de restriction				
ET	Amorce	Séquence (5'-3')	Tm (en °C)	Taille du produit de PCR (pb)
<i>F</i> site 5'	FR1	atg-tct-gcg-ttt-ggg-att-gt	58	239
	FF1	tga-gtg-aag-tga-acg-cca-aa		
<i>F</i> ORF1	FR1site1	tct-cgc-atg-cat-ttt-ctt-tg	57	194
	FF1site1	caa-act-ccagct-cct-ttt-gc		
<i>F</i> ORF2	FR1site2	cgt-aac-tgc-gaa-gtc-gat-ca	62	207
	FF1site2	gca-gta-ggg-cca-cac-ttc-at		
<i>Roo</i>	Roo r1	ata-gtc-ccc-gcc-tta-tcg-ag	58	250
	Roo f1	ttg-ggctcc-gtt-cat-atc-tt		
<i>412</i>	412 r1	tag-gct-aat-ccg-cgt-agc-tg	58	219
	412 f1	aac-atg-atc-acc-cac-tcg-aag		
<i>tirant</i> (type C)	Tir r1	ttg-ctt-cgt-ttc-tta-caa-tgg	62,5	153
	Tir f1	ggg-gaa-acc-taa-aac-cct-tc		
	Tir r1bis	gct-gaa-act-taa-aaa-cct-tca-tac-a	60	
	Tir f1bis	agt-ttg-ttt-cgt-ttc-tta-cat-gtt-t		

8. Immunoprécipitation de la Chromatine (ChIP)

La ChIP est une technique qui permet d'extraire la chromatine des cellules puis de l'immunoprécipiter avec des anticorps spécifiques dirigés contre des protéines constitutives de la chromatine, comme par exemple les histones. Il est ainsi possible de récupérer des fractions d'ADN enrichies en H3 totale (H3T), H3K9me2, H3K27me3 et H3K4me2. Les anticorps anti-diméthyl histone H3 (Lys4), anti-diméthyl histone 3 (Lys9) et anti-triméthyl histone 3 (Lys27) ont été fournis par Millipore et l'anti-histone H3 (ChIP Grade) a été fournie par AbCam. Une région faiblement enrichie en H3T aura des valeurs d'enrichissement faible pour les marques étudiées, il sera donc important de normaliser les résultats obtenus en fonction de l'enrichissement en nucléosome, rapporté par l'enrichissement en H3T. L'enrichissement en IgG permet de mesurer l'aspécificité de l'immunoprécipitation. Le plan expérimental utilisé pour cette technique est représenté sur la Figure 32.



Protocole de ChIP :

- 1- 150 paires d'ovaires collectées à partir de femelles âgées de 3 jours pour chaque population ont été disséquées, et lavées au **PBT**.
- 2- Ajouter 10 mL de « **Crosslink solution** ». Agiter pendant 10 minutes à température ambiante.
- 3- Enlever la phase aqueuse, ajouter 10 mL de « **STOP solution** » puis agiter 5 minutes à température ambiante.
- 4- Enlever le surnageant et resuspendre les ovaires dans 7,5 mL de PBT froid + 300 µL d'inhibiteurs de protéases).
- 5- Isolation des cellules : Homogénéiser les tissus avec un pilon Potter homogénéisateur, puis centrifuger (1 minute, 400 g, 4°C).
- 6- Récupérer le surnageant puis centrifuger (10 minutes, 1100 g, 4°C).
- 7- Resuspendre le culot avec 7,5 mL de « **Cell Lysis Buffer** » froid + 300 µL d'**inhibiteurs de protéases**.
- 8- Isolement des noyaux : Homogénéiser les tissus avec un pilon Potter homogénéisateur, puis centrifuger (4 minutes, 2000 g, 4°C).
- 9- Resuspendre le culot avec 1 mL de « **Nuclear Lysis Buffer** » froid + 60µL d'inhibiteurs de protéases. Incuber à température ambiante pendant 20 minutes.
- 10- Ajouter 1mL de « **Nuclear Lysis Buffer** » froid + 60µL d'**inhibiteurs de protéases**.
- 11- Sonication de la chromatine (7 cycles, 30 secondes ON, 30 secondes OFF) puis vérifier sur gel d'agarose la taille des fragments (500 pb).
- 12- Centrifuger (10 minutes, 20 000 g, 4°C).
- 13- Récupérer le surnageant, puis ajouter le tampon 1 (qsp 6 mL).
- 14- Ajouter 150 µL de solution de billes de séphadex puis incuber à 4°C pendant 1 heure.
- 15- Centrifuger (1 minute, 1000 g, 4°C).
- 16- Aliquoter la chromatine (1 mL pour l'IP, 500 µL pour l'INPUT).
- 17- Ajouter 5 µL d'anticorps (1 mg/mL) puis incuber pendant la nuit à 4°C sur un plateau rotatif.
- 18- Ajouter 80 µL de solution de billes puis incuber 3 heures à 4°C sur un plateau rotatif.
- 19- Centrifuger (2 minutes, 400 g, 4°C).
- 20- Faire les lavages suivants : 2 x 4 minutes avec la **solution d'IP lavage 1**, puis 2 fois x 4 minutes avec la **solution IP lavage 2**, puis 2 x 2 minutes avec du TE.
- 21- Ajouter 100 µL de TE + 1 µL de RNase A (10 mg/mL) pendant 1 heure à 37°C.

- 22- Ajouter de la protéinase K (0,5 mg/mL) + 10 µL de SDS 10% et incuber 3 heures à 55°C.
- 23- Ajouter 1 volume de Phénol/Chloroforme/IAA, vortexer 15 secondes
- 24- Centrifuger (12 000 rpm, 5 minutes, 4°C)
- 25- Récupérer le SN et mettre dans un nouveau tube
- 26- Ajouter 25µL de NaAc 3M pH5,3 + 1µL de glycogène + 550µL d'éthanol 100.
- 27- Incuber 1 heure à -20°C.
- 28- Centrifuger (18 000 g, 15 minutes, 4°C), laver le culot à l'éthanol 70, puis centrifuger encore une fois.
- 29- Reprendre le culot dans 40 µL de TRIS 10mM pH8 pour les IP, et dans 100 µL pour l'INPUT.
- 30- Quantifier les échantillons au nanodrop et ramener leur concentration à 6 ng/µL.

Solutions à préparer avant de commencer:

- **PBT:** PBS 1X, 0,1% Triton X100
- **Stop solution:** 125 mM Glycine, PBT
- **Crosslink Solution:** 50 mM Hepes pH8, 1 mM EDTA.Na2, 0,5 mM EGTA, 100 mM NaCl.
- **Préparation des billes:**
Peser 160 mg de protéine A Sépharose et ajouter 8 mL de PBS 1X puis incuber 1 heure à 4°C sur un plateau rotatif. Centrifuger (3 minutes, 400g, 4°C) puis enlever le surnageant. Ajouter 2mL de PBS 1X, BSA 0,1% puis incuber toute la nuit à 4°C sur un plateau rotatif. Centrifuger (3 minutes, 400g, 4°C) puis enlever le surnageant. Ajouter 720 µL de PBS 1X, BSA 0,1%.
- **Solution IP lavage 1:** Triton X100 1%, SDS 0,1%, 500 mM NaCl, TE1X
- **Solution IP lavage 2:** Triton X100 1%, Deoxycholate 0,1%, SDS 0,1%, 140 mM NaCl, 1 mM PMSF, TE1X.
- **« Cell Lysis Buffer » :** (PIPES 5mM pH8, KCL 85mM, NP40/IGEPAL 0,5%)
- **« Nuclear Lysis Buffer » :** 10 mM EDTA, 0,5% N-lauroylsarcosine, 50 mM HEPES, pH8
- **Tampon 1:** 0,1% Triton X100, 0,1% sodium deoxycholate, 0,1% SDS, 140 mM NaCl, TE1X.
- **Inhibiteurs de protéases :** « Complete Protease Inhibitor Cocktail tablets in glass vials » (Roche).

9. qPCR

Les ADN ont été dilués et ramenés à la concentration de 6 ng/μL, puis quantifiés par qPCR dans un lightcycler (Roche diagnostics) en utilisant les amorces du Tableau 9. Le taux d'enrichissement a été calculé par différence de Ct par rapport à l'INPUT de chacune des populations et les résultats ont été normalisés par H3T. Les mesures ont été réalisées à partir de deux expériences indépendantes.

Tableau 9. Amorces utilisées en qPCR pour l'analyse des résultats de CHIP.

qPCR TM=62°C		
site	amorce	Séquence (5'-3')
promoteur <i>tkv</i> RD	Rdprochip f	ttt-aaa-cgc-cag-caa-gtg-tg
	Rdprochip r	acg-act-cgc-gaa-gaa-cgt-at
M1, C1, Z1	ilot f1	cac-cga-gtg-aaa-ctg-ctg-aa
	ilot r1	atg-ttg-cag-tgc-gac-ttt-tg
M2	CflanqR	ctg-agc-act-tga-ttt-ggg-ctt-aga-cag-gc
	C2tir r	gtg-ttc-cag-ttg-ccg-tct-tc
M3	15qf2	tgt-gcg-gat-ttc-tac-tgt-ttt-c
	15qr2	agc-ata-atg-aac-atg-ccg-act
C2	3122 f	cgc-cag-aaa-aca-cca-aca-ga
	3122 r	ggg-tta-gag-ggt-ggg-gtg-gt
C3	451 f	tgt-gga-cat-tag-gga-ctg-ga
	452 r	tag-agg-cgt-ggg-ggt-tta-g
Z2	idl2_2 f	cca-gag-caa-gca-aac-att-ga
	idl2_2 r	tag-agg-cgt-ggg-ggt-tta-g
Z3	idl3_1 f	cgt-atc-gtt-cgt-gcc-att-ta
	idl3_1 r	ggg-tga-gtg-tcc-gac-gat-tt
15notir	15notir f	gca-tcc-aat-gcg-aac-aag-aa
	15notir r	agg-tcc-gca-gtt-cca-cag-tt
27notir	27notir f	gcc-ttt-ttg-tgt-gcc-caa-ct
	27notir r	gcg-ttg-ctc-taa-ggg-ga
RP49	rp49 f	ct-ggt-ttc-cgg-caa-ggt-atg-t
	rp49 r	ca-gtt-caa-ctc-aaa-acc-gcc-aaa

10. Hydrolyse de l'ADN et préparation des échantillons pour la HPLC

Les échantillons d'ADN génomique des 12 espèces de drosophiles dont les génomes ont été séquencés ont été extraits à partir d'embryons âgés de moins de 16 heures et fournis par le Drosophila Species Stock Center (University of California, San Diego, <http://stockcenter.ucsd.edu>). L'ADN de thymus de veau a été commandé chez Invitrogen, les oligonucléotides méthylés (5'-GCC-GAA-TGC-TGA-CAT-TCC-GGC-CGA-ATG-CT-3') et non méthylés (5'-GC-5MedC-GAA-TG-5MedC-TGA-CAT-TCC-GGC-5MedC-GA-ATG-5MedC-T-3') ont été produits par EUROGENTEC. Les nucléotides standards 2'-deoxyuridine 5'monophosphate (dU), 2'-deoxyguanosine 5'monophosphate (dG), 2'-deoxycytidine 5'monophosphate (dC), 2'-deoxyadénosine 5'monophosphate (dA), 5-méthyl 2'-cytidine 5'monophosphate (5-MedC), et 2'-deoxythymidine 5'monophosphate (dT) ont été fournis par SIGMA-ALDRICH.

Tous les échantillons ont été traités à la RNase A (concentration finale de 20 µg/µL) pendant 1 heure à 37°C. L'ADN a ensuite été précipité avec 1 volume d'isopropanol froid, puis centrifugé à 11 000 g pendant 10 minutes. Le culot d'ADN a ensuite été lavé à l'éthanol 70%, centrifugé 5 minutes à 11 000 g, repris dans de l'eau et stocké à + 4°C.

Pour chaque hydrolyse, j'ai utilisé 6,5 µg d'ADN dans un volume de 10 µL. Les échantillons ont été dénaturés 2 minutes à 100°C puis mis rapidement dans la glace pendant 5 minutes. La réaction d'hydrolyse a été initiée par l'ajout de 10 µL de nucléase P1 et 2 µL de sulfate de zinc, et les échantillons ont été incubés 16 heures à 37°C. Une fois la réaction d'hydrolyse terminée, les échantillons ont été incubés 2 heures à 37°C avec 0,75 µL de phosphatase alcaline + 1,25 µL de Tris-HCL (pH8,3) afin d'enlever les groupements phosphates des nucléotides. Les échantillons ont été centrifugés et le surnageant a été récupéré et conservé à + 4°C.

Les taux de méthylation chez la drosophile sont faibles, et afin d'avoir des quantités en 5-MedC détectables pour chacun des échantillons, les produits de quatre hydrolyses ont été rassemblés, dont 70 % ont été injectés pour la HPLC.

Pour cette étude, les analyses sont réalisées sur une HPLC HP 1100 series (Agilent technologies) couplée à un détecteur à barrette de diode (DAD) (G1315 ; Agilent technologies) et un spectromètre de masse simple quadripôle avec une source d'ionisation par électrospray HP MSD 1100 series (G1946 ; Agilent technologies). L'ensemble de l'appareillage est piloté par le logiciel chemstation (Agilent technologies). Ce système est équipé d'une colonne en phase inverse Eclipse XDB[®] C18 (250 × 4.6 mm, 5 μm ; Agilent technologies). Une solution d'H₂O avec 0.1 % d'acide formique (solvant A) et une solution d'H₂O/méthanol (50/50) avec 0.1 % d'acide formique (solvant B) sont utilisées pour l'élution. Le gradient est indiqué dans le tableau 1 et le débit utilisé est de 0.8 mL/min. La température de la colonne est de 25°C. Le DAD enregistre les données entre 200 et 600 nm et la longueur d'onde utilisée pour la quantification est à 280 nm. Pour le spectromètre de masse, l'enregistrement se fait uniquement en mode positif avec deux niveaux de fragmenteur 60 et 120; Le gaz de nébulisation est à une température de 350°C avec un débit de 10L/min et à une pression de 30 p.s.i ; le voltage du capillaire est de 3500 V et la température du quadripole à 300°C. L'enregistrement des spectres se fait entre les masses *m/z* 100 et *m/z* 900.

Le volume d'injection de l'échantillon est de 35 μL ou 70 μL selon les matrices. L'eau a été distillée et filtrée extemporanément, et le méthanol a été distillé et filtré.

Tableau 10. Gradient HPLC.

Temps en minute	% solvant A H ₂ O + 1% acide formique	% solvant B H ₂ O/Methanol + 1% acide formique
0	95	5
3	95	5
30	84	16
32	75	25
35	0	100
37	0	100
38	95	5
45	95	5

ANNEXES

1. Méthylation de l'ADN

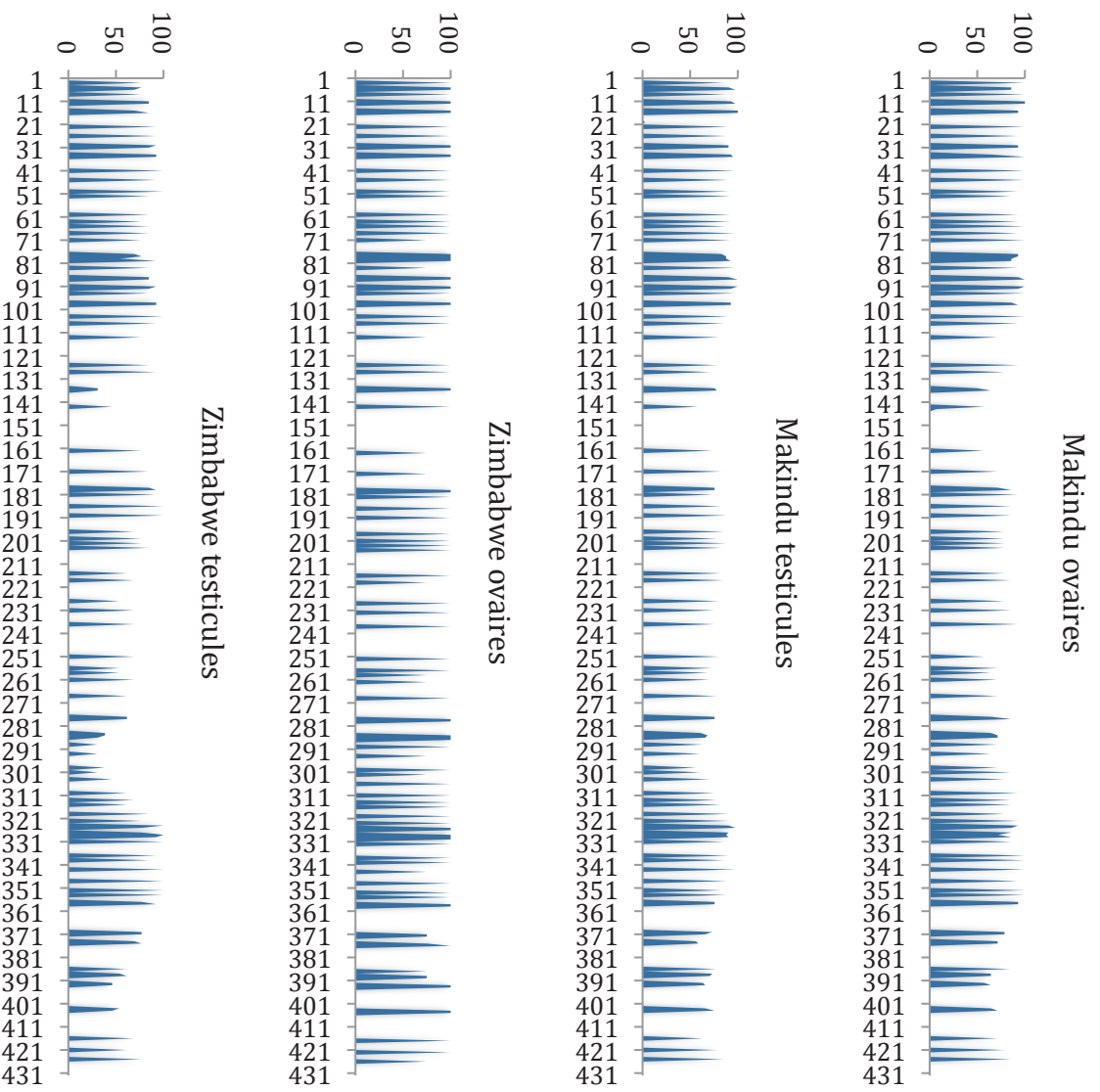
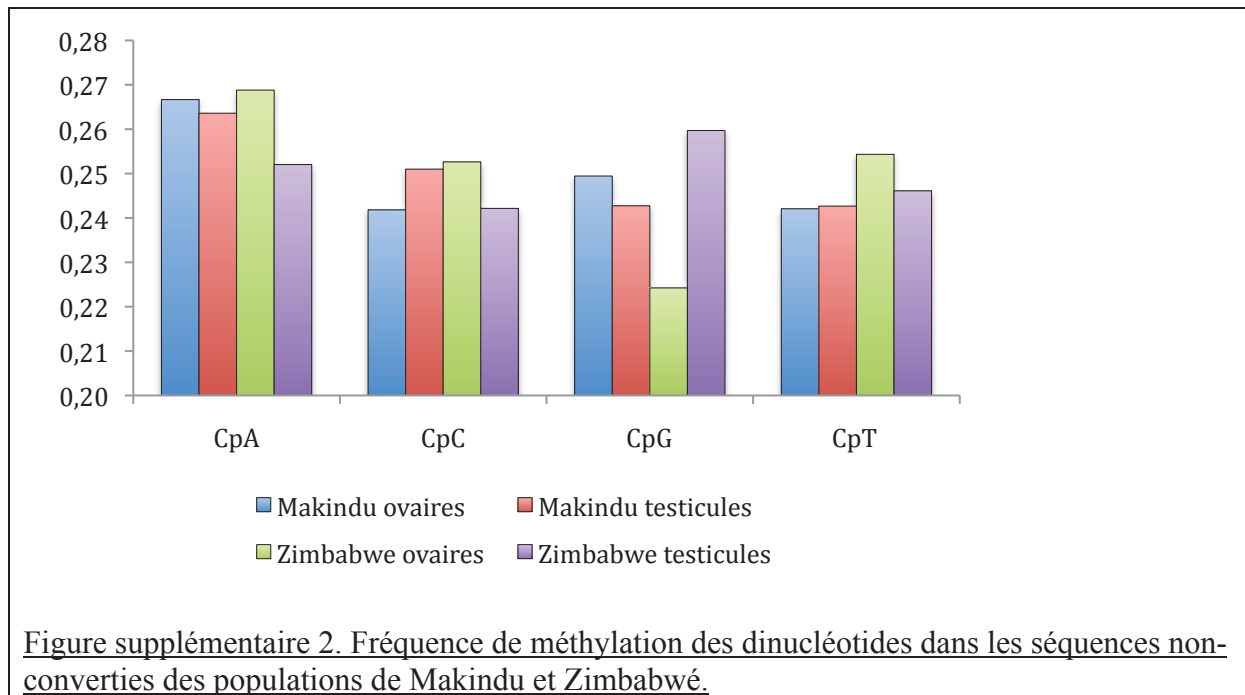


Figure supplémentaire 1. Profil de méthylation des séquences non converties correspondant à la LTR 5' de *tirant*.

Les positions de cytosine dans la séquence sont représentées en abscisse et le taux de méthylation en %, en ordonnée, a été calculé pour chaque position de cytosine à partir des séquences non converties. Ces profils de méthylation ont été établis à partir de résultats obtenus pour des ovaires et des testicules des populations de Makindu et Zimbabwe. Ces trois histogrammes présentent des régions fortement méthylés et des régions faiblement méthylées. De plus, les profils de méthylation obtenus pour ces séquences non converties sont identiques (les coefficients de corrélation de Spearman entre ces profils de méthylation sont significativement différents de zéro).



Tableaux supplémentaires 1 à 3 : Fréquence de méthylation des dinucléotides calculées à partir des résultats de BGS.

Tableau supplémentaire 1. Résultats obtenus à partir d'ADN purifié extrait à partir d'ovaires.

Tissu	Population	amorces	nombre de séquences	5-MedC (%)	CpA	CpC	CpG	CpT
ovaires	Makindu	tir01	converties n=14	0,67	0,71	0,13	0,16	0,00
			non converties n=0					
		tir52	converties n=11	0,56	0,28	0,00	0,00	0,72
			non converties n=0					

Tableau supplémentaire 2. Résultats obtenus pour les lignées mutantes.

Tissu	Population	amorces	nombre de séquences	5-MedC (%)	CpA	CpC	CpG	CpT
ovaires	dDNMT2	tir01	converties n=22	0,52	0,25	0,33	0,13	0,29
			non converties n=0					
	GS12412		converties n=27	0,48	0,21	0,19	0,24	0,36
			non converties n=2	81,78	0,27	0,25	0,25	0,23
	R2		converties n=18	0,31	0,22	0,40	0,25	0,13
			non converties n=0					
Makindu	Act5	converties n=19		0,00	0,00	1,00	0,00	
		non converties n=0						

Tableau supplémentaire 3. Résultats obtenus pour les populations naturelles avec les amorces tir01 et tir52.

Tissus	Population	Amorces	Nb de séquences	5-MedC (%)	CpA	CpC	CpG	CpT
testicules	Chicharo	tir01	converties n=46	0,49	0,15	0,60	0,18	0,07
			non converties n=0					
		tir52	converties n=8	0,00				
			non converties n=0					
	Makindu	tir01	converties n=8	0,70	0,58	0,42	0,00	0,00
			non converties n=38	82,32	0,26	0,25	0,24	0,24
		tir52	converties n=11	0,53	0,18	0,67	0,00	0,14
			non converties n=0					
	Sénégale	tir01	converties n=56	0,33	0,21	0,46	0,08	0,25
			non converties n=0					
		tir52	converties n=10	0,00				
			non converties n=0					
	Zimbabwe	tir01	converties n=50	0,47	0,39	0,11	0,07	0,43
			non converties n=15	73,42	0,25	0,24	0,26	0,25
		tir52	converties n=11	0,49	0,35	0,65	0,00	0,00
			non converties n=0					
ovaires	Chicharo	tir01	converties n=48	0,45	0,33	0,30	0,08	0,30
			non converties n=0					
		tir52	converties n=10	0,75	0,79	0,00	0,00	0,21
			non converties n=0					
	Makindu	tir01	converties n=37	0,53	0,31	0,00	0,16	0,53
			non converties n=14	84,62	0,27	0,24	0,25	0,24
		tir52	converties n=25	0,35	0,87	0,00	0,00	0,13
			non converties n=0					
	Sénégale	tir01	converties n=57	0,70	0,18	0,21	0,27	0,34
			non converties n=0					
		tir52	converties n=26	0,46	0,14	0,19	0,48	0,19
			non converties n=0					
	Zimbabwe	tir01	converties n=44	0,85	0,16	0,14	0,32	0,38
			non converties n=4	95,80	0,27	0,25	0,22	0,25
		tir52	converties n=23	0,66	0,46	0,00	0,40	0,14
			non converties n=0					

Tableau supplémentaire 4. Résultats complémentaires de vérification de la présence de biais expérimentaux.

Tissus	Amorces	Conditions	Nombre de séquences	5-MedC (%)	CpA	CpC	CpG	CpT
testicules	ctir	Standard	converties n=14	0,51	0,26	0,00	0,21	0,53
			non converties n=0					
converties n=14			0,75	0,00	0,00	0,78	0,22	
non converties n=0								
ovaires	tir01	Standard	converties n=14	0,40	0,25	0,18	0,44	0,12
			non converties n=4	75,47	0,27	0,23	0,24	0,26
	+ prot K	converties n=6	0,62	0,45	0,33	0,00	0,22	
		non converties n=1	81,31	0,23	0,24	0,26	0,27	
		+ prot K + Rnase H	converties n=17	0,23	0,42	0,00	0,37	0,21
			non converties n=0					

Tableau supplémentaire 5. Fréquence de méthylation des dinucléotides pour les éléments *F*, *Roo* et *412* dans la population de Makindu.

Tissu	Population	amorces	nombre de séquences	5-MedC (%)	CpA	CpC	CpG	CpT
ovaires	Makindu	Roobisu	converties n=40	0,69	0,21	0,00	0,00	0,79
			non converties n=0					
		Fbisu	converties n=40	0,72	0,29	0,21	0,31	0,19
			non converties n=0					
		412bisu	converties n=29	0,55	0,26	0,11	0,32	0,31
			non converties n=0					

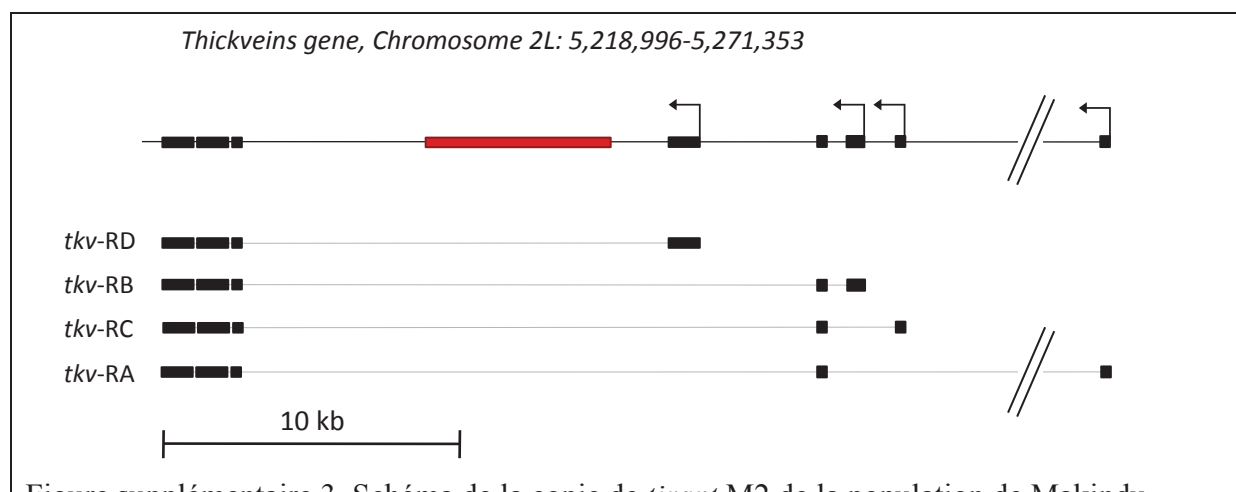


Figure supplémentaire 3. Schéma de la copie de *tirant* M2 de la population de Makindu, insérée dans l'intron du gène *tkv*.

Le gène *tkv* et la copie de *tirant* sont insérés dans le même sens sur le chromosome 2L. Les exons sont représentés par des rectangles noirs. Ce gène de 52 kb produit quatre transcrits A, B, C et D. Les résultats de ChIP présentés dans le chapitre 2 concernent la région promotrice du transcrit D, localisée à 2,5 kb de l'insertion de *tirant*.

Figure supplémentaires 4 à 6 : Profils de méthylation des éléments *412*, *F* et *Roo* dans les ovaires de la population de Makindu.

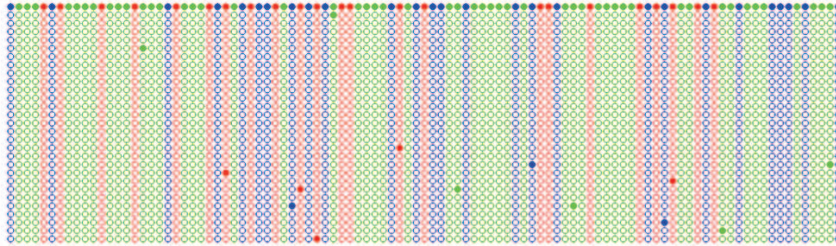


Figure supplémentaire 4. Élément *412*.

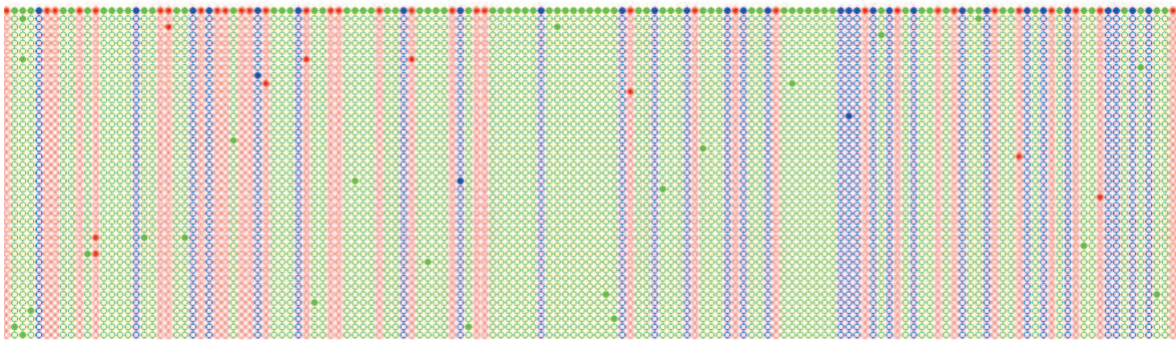


Figure supplémentaire 5. Élément *F*.

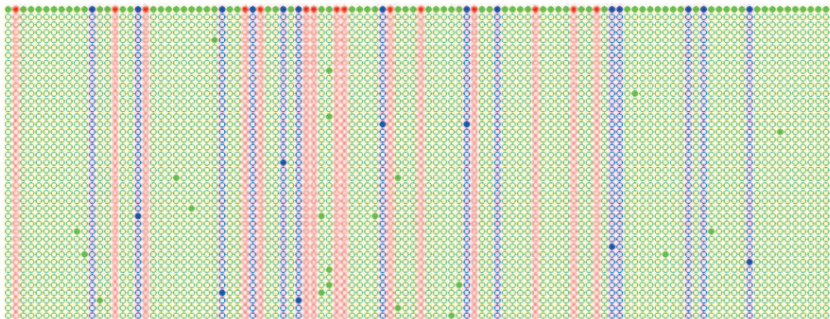


Figure supplémentaire 6. Élément *Roo*.

Figure supplémentaire 7. Alignements des séquences traitées au bisulfite de sodium.

Alignment: /Users/benjaminhubert/Desktop/thèse/data/DNA_methylation/STRAINS/MAKINDU/OWARIES/TIR01/MAKINDU.txt
 Seaview [blocks=10 fontsize=8 A4] on Tue Oct 12 17:01:23 2010

```

1
reference AACACCCGCAT CCGGCCGAAT GCTGACATTC CGGCCGAATG CTGACATTGC ACTAAAAGTC GCAC TGCAAC ATTGTCCCA GCTAGCCAGC
M01.1.1 AATATTGTAT TTGGTTGAAT GTTGATATTT TGGTTGAATG TTGATATTGT ATTAAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M01.1.2 AATATTGTAT TTGGTTGAAT GCTGATATTT TGGTTGAATG TTGATATTGT ATTAAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M01.1.3 AACACCCGCAT CCGGCCGAAT GCTGACATTC CGGCCGAATG CTGACATTGC ACTAAAAGTC GCAC TGCAAC ATTGTCCCA GCTAGCCAGC
M01.1.4 AACATTGCAT CCGGCCGAAT GCTGACATTC CGGCCGAATG CTGACATTGC ACTAAAAGTC GCAC TGCAAC ATTGTCCCA GCTAGCCAGC
M01.1.5 AACACCCGCAT CCGGCCGAAT GCTGACATTC CGGCCGAATG CTGACATTGC ACTAAAAGTC GCAC TGCAAC ATTGTCCCA GCTAGCCAGC
M01.1.6 -----AAT ATTG-TATTT -GGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M01.1.7 AATATTGTAT TTGGTTGAAT GTTGATATTT TGGTTGAATG TTGATATTGT ATTAAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M01.1.8 -----AAT ATTG-TATTT -GGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M01.1.9 AATATTGTAT TTGGTTGAAT GTTGATATTT TGGTTGAATG TTGATATTGT ATTAAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M01.1.10 -----AAT ATTG-TATTT -GGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M01.1.11 AATATTGTAT TTGGTTGAAT GTTGATATTT TGGTTGAATG TTGACATTGT ATTAAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M01.2.1 AACACCCGCAT CCGGCCGAAT GCTGACATTC CGGTCGAATG CTGACATTGC ACTAAAAGTC GCAC TGCAAC ATTGTCCCA GCTAGCCAGC
M01.2.2 AACACCCGCAT CCGGCCGAAT GCTGACATTT TGGTCGAATG CTGACATTGC ACTAAAAGTC GCAC TGCAAC ATTGTCCCA GCTAGCCAGC
M01.2.3 AACACCCGCAT CCGGCCGAAT GCTGACATTC CGGCCGAATG CTGACATTGC ACTAAAAGTC GCAC TGCAAC ATTGTCCCA GCTAGCCAGC
M01.2.4 AATATTGTAT TTGGTTGAAT GTTGATATTT TGGTTGAATG TTGATATTGT ATTAAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M01.2.5 AATATTGTAT TTGGTTGAAT GTTGATATTT TGGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M01.2.6 -----AAT ATTG-TATTT -GGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M01.2.7 -----AAT ATTG-TATTT -GGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M01.2.8 AATATTGTAT TTGGTTGAAT GTTGATATTT -GGTTGAATG TTGATATTGT ATTAAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M01.2.9 AATATTGTAT TTGGTTGAAT GTTGATATTT TGGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M01.2.10 AACATCGCAT CCGGCCGAAT GCTGACATTC CGGCCGAATG CTGACATTGC ACTAAAAGTC GCAC TGCAAC ATTGTCCCA GCTAGCCAGC
M02.2.1 -----AAT ATTG-TATTT -GGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M02.2.2 -----AAT ATTG-TATTT -GGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M02.2.3 -----AAT ATTG-TATTT -GGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M02.2.4 -----AAT ATTG-TATTT -GGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M02.2.5 -----AAT ATTG-TATTT -GGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M02.2.6 -----AAT ATTG-TATTT -GGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M02.2.7 -----AAT ATTG-TATTT -GGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M02.2.8 -----AAT ATTG-TATTT -GGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M02.2.9 -----AAT ATTG-TATTT -GGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M02.2.10 -----GAT ATTG-TATTT -GGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M02.2.11 -----AAT ATTG-TATTT -GGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M02.2.12 -----AAT ATTG-TATTT -GGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M02.2.13 AATATCGTAT TTGGTTGAAT GTTGATATTT TGGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M02.2.14 AATATTGTAT TTGGTTGAAT GTTGATATTT TGGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M02.2.15 AATATTGTAT TTGGTTGAAT GTTGATATTT TGGTTGAATG TTGATATTGT ATTAAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M02.2.16 AATATTGTAT TTGGTTGAAT GTTGATATTT TGGTTGAATG TTGATATTGT ATTAAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M02.2.17 AATATTGTAT TTGGTTGAAT GTTGATATTT TGGTTGAATG TTGATATTGT ATTAAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M02.2.18 AGTATTGTAT TTGGTTGAAT GCTGATATTT TGGTTGAATG TTGATATTGT ATTAAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M02.2.19 AACACCCGCAT CCGGCCGAAT GCTGACATTC CGGCCGAATG CTGACATTGC ACTAAAAGTC GCAC TGCAAC ATTGTCCCA GCTAGCCAGC
M02.2.20 AACACTGCAT CCGGCTGAAT GCTGATATTC CGGTCGAATG CTGACATTGC ACTAAAAGTC GCAC TGCAAC ATTGTCCCA GCTAGCCAGC
M02.1.1 -----AAT ATTG-TATTT -GGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M02.1.2 -----AAT ATTG-TATTT -GGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M02.1.3 -----AAT ATTG-TATTT -GGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M02.1.4 -----AAT ATTG-TATTT -GGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M02.1.5 AATATTGTAT TTGGTTGAAT GTTGATATTT TGGTTGAATG TTGATATTGT ATTAAAAGTT GTATTGTAAT ATTGTTTTA GTTAGTTAGT
M02.1.6 AACACCCGCAT CCGGCCGAAT GCTGACATTC CGGTCGAATG CTGACATTGT ATTAAAAGTT GTACTGCAAC ATTGTCCCA GCTAGCCAGC
M02.1.7 AACACCCGCAT CCGGCCGAAT GCTGACATTC CGGCCGAATG CTGACATTGC ACTAAAAGTC GCAC TGCAAC ATTGTCCCA GCTAGCCAGC
M02.1.8 AACACCCGCAT CCGGCCGAAT GCTGACATTC CGGCCGAATG CTGACATTGC ACTAAAAGTC GCAC TGCAAC ATTGTCCCA GCTAGCCAGC
M02.1.9 AACACCCGCAT CCGGTCGAAT GCTGACATTC CGGTCGAATG CTGACATTGC ACAAAAGTC GCAC TGCAAC ATTGTCCCA GCTAGCCAGC
M02.1.10 AACACCCGCAT CCGGCCGAAT GCTGACATTC CAGCCGAATG CTGACATTGC ACTAAAAGTC GCAC TGCAAC ATTGTCCCA GCTAGCCAGC
  
```


271

reference	GACTGAAAGT	GACTAGGAAC	AAATTTTAT	AAAACAA
M01.1	GATTGAAAGT	GATTAGCAAT	AAATTTTAT	AAAATAA
M01.2	GATTGAAAGT	GAATAAGCAAT	AAATTTTAT	AAAAAAA
M01.3	GATTGAAAGT	GATTAGCAAT	AAATTTTAT	AAAATAA
M01.4	GATTGAAAGT	GATTAGCAAT	AAATTTTAT	AAAATAA
M01.5	GATTGAAAGT	GAATAAGCAAT	AAATTTTAT	AAAAAAA
M01.6	GATTGAAAGT	GAATAAGCAAT	AAATTTTAT	AAAAAAA-
M01.7	GATTGAAAGT	GAATAAGCAAT	AAATTTTAT	AAAAAAA
M01.8	GATTGAAAGT	GAATAAGCAAT	AAATTTTAT	AAAAAAA-
M01.9	GATTGAAAGT	GAATAAGCAAT	AAATTTTAT	AAAAAAA-
M01.10	GATTGAAAGT	GATTAGCAAT	AAATTTTAT	AAAATAA
M02.1	GATTGAAAGT	GAATAAGCAAT	AAATTTTAT	AAAAAAA-
M02.2	GATTGAAAGT	GATTAGCAAT	AAATTTTAT	AAAATAA
M02.3	GATTGAAAGT	GATTAGCAAT	AAATTTTAT	AAAATAA
M02.4	GATTGAAAGT	GAATAAGCAAT	AAATTTTAT	AAAAAAA-
M02.5	GATTGAAAGT	GAATAAGCAAT	AAATTTTAT	AAAAAAA-
M02.6	GATTGAAAGT	GAATAAGCAAT	AAATTTTAT	AAAAAAA-
M02.7	GATTGAAAGT	GATTAGCAAT	AAATTTTAT	AAAATAA
M02.8	GATTGAAAGT	GATTAGCAAT	AAATTTTAT	AAAATAA
M02.9	GATTGAAAGT	GATTAGCAAT	AAATTTTAT	AAAATAA
M02.10	GATTGAAAGT	GAATAAGCAAT	AAATTTTAT	AAAAAAA-
M02.11	GATTGAAAGT	GAATAAGCAAT	AAATTTTAT	AAAAAAA-
M02.12	GATTGAAAGT	GAATAAGCAAT	AAATTTTAT	AAAAAAA-
M02.13	GATTGAAAGT	GAATAAGCAAT	AAATTTTAT	AAAAA--
M02.14	GATTGAAAGT	GAATAAGCAAT	AAATTTTAT	AAAAAAA-
M02.15	GATTGAAAGT	GATTAGCAAT	AAATTTTAT	AAAATAA

Alignment: /Users/benjaminhubert/Desktop/thèse/data/DNA_methylation/STRAINS/MAKINDU/TESTES/TIR01/MAKINDU.txt
Seaview [blocks=10 fontsize=8 A4] on Tue Oct 12 17:02:07 2010

```
1
reference AACACCCGCAT CCGGCCGAAT GCTGACATTC CGGCCGAATG CTGACATTGC ACTAAAAGTC GCACTGCAAC ATTGTCOCCA GCTAGCCAGC
MT1.1.1 AACACCCGCAT CCGGCCGAAT GCTGACATTC CAACCCGAATG CTGACATTGC ACTAAAAGTC GCACTGCAAC ATTGTCOCCA GCTAGCCAGC
MT1.1.2 AACACCCGCAT CCGGCCGAAT GCTGACATTC CGACCCGAATG CTGACATTGC ACTAAAAGTC GCACTGCAAC ATTGTCOCCA GCTAGCCAGC
MT1.1.3 AACACCCGCAT CCGGCCGAAT GCTGACATTC CGGCCGAATG CTGACATTGC ACTAAAAGTC GCACTGCAAC ATTGTCOCCA GCTAGCCAGC
MT1.1.4 AATACCGTAT CCGGCCGAAT GCTGACATTC CGGCCGAATG CTGACATTGC ACTAAAAGTC GCACTGCAAC ATTGTTTTTA GCTAGCCAGC
MT1.1.5 AACACCCGCAT CCGGCCGAAT GCTGACATTC CGGCCGAATG CTGATATTGC ACTAAAAGTT GCACTGCAAC ATTGTCOCCA GCTAGCCAGC
MT1.1.6 AACACCCGCAT CCGGCCGAAT GCTGACATTC CGGCCGAATG CTGACATTGC ACTAAAAGTC GCACTGCAAC ATTGTCOCCA GCTAGCCAGC
MT1.1.7 AACACCGTAT CCGGCCGAAT GCTGACATTC TGTTGSAATG TTGATATTGC ACTAAAAGTC GCACTGCAAC ATTGTCOCCA GCTAGCCAGC
MT1.1.8 AACACCCGCAT CCGGCCGAAT GCTGACATTC CGGCCGAATG CTGACATTGC ACTAAAAGTC GCACTGCAAC ATTGTCOCCA GCTAGCCAGC
MT1.1.9 AACACCCGCAT CCGGCCGAAT GCTGACATTC CGGCCGAATG CTGACATTGC ATTTAAAAGTC GCACTGCAAC ATTGTCOCCA GCTAGCCAGC
MT1.1.10 AACACCCGCAT CCGGCCGAAT GTTGACATTC CGGCCGAATG CTGACATTGC ACTAAAAGTC GCACTGCAAC ATTGTCOCCA GCTAGCCAGC
MT1.2.1 AACACCGTAT TCGGCCGAAT GCTGACATTC CGGCCGAATG CTGACATTGC ACTAAAAGTC GCACTGCAAC ATTGTCOCCA GCTAGCCAGC
MT1.2.2 AACACCGTAT TCGGCCGAAT GCTGACATTC CGGCCGAATG CTGACATTGC ACTAAAAGTC GCACTGCAAC ATTGTCOCCA GCTAGCCAGC
MT1.2.3 AACACCCGCAT CCGGCCGAAT ACTGACATTC CGGCCGAATG CTGACATTGC ACAAAAAGTC GCACTGCAAC ATTGTCOCCA GCTAGCCAGC
MT1.2.4 AACACCCGCAT CCGGCCGAAT GCTGATATTG CGGCCGAATG CTGACATTGT ATAAAAAGTT GTATTGTAAT ATTGTCCTCA GCTAGCCAGC
MT1.2.5 AACACCCGCAT CCGGCCGAAT ACTGACATTC CGGCCGAATG CTGACATTGC ACAAAAAGTC GCACTGCAAC ATTGTCOCCA GCTAGCCAGC
MT1.2.6 AACACCCGCAT CCAACCGAAT GCTGACATTC CGGCCGAATG CTGACATTGC ACTAAAAGTC GCACTGCAAC ATTGTCOCCA GCTAGCCAGC
MT1.2.7 AACACCCGCAT CCGGCCGAAT GCTGACATTC CGGCCGAATG CTGACATTGC ACTAAAAGTC GCACTGCAAC ATTGTCOCCA GCTAGCCAGC
MT1.2.8 AACACCCGCAT CCGGCCGAAT GTTGATATTG TGGTGAATG TTGATATTGC ACTAAAAGTC GCACTGCAAC ATTGTCOCCA GCTAGCCAGC
MT1.2.9 AATATTGTAT TTGGTTGAAT GTTGATATTG TGGTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
MT1.2.10 AACACCCGCAT CCGGCCGAAT GTTGATATTG TGGCCGAATG CTGACATTGC ACTAAAAGTC GCACTGCAAC ATTGTCOCCA GCTAGCCAGC
MT1.2.11 AACACCCGCAT CCGGCCGAAT GCTGACATTC CGGCCGAATG CTGACATTGC ACTAAAAGTC GCACTGCAAC ATTGTCOCCA GCTAGCCAGC
MT1.2.12 AACACCCGCAT CCGGCCGAAT GCTGACATTC CGGCCGAATG CTGACATTGC ACTAAAAGTC GCACTGCAAC ATTGTCOCCA GCTAGCCAGC
MT2.1.1 -----AAT AITG-TATTT -GGTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
MT2.1.2 -----AAT AITG-TATTT -GGTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
MT2.1.3 AATATTGTAT TTGGTTGAAT GTTGATATTG TGGTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
MT2.1.4 AATATTGTAT TTGGTTGAAT GTTGATATTG TGGTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
MT2.1.5 AACACCCGCAT CCGGCCGAAT GCTGACATTC CGGCCGAATG CTGACATTGC ACTAAAAGTC GCACTGCAAC ATTGTTTTTA GCTAGCCAGC
MT2.1.6 AACACCCGCAT CCGGCCGAAT GTTGATATTG TGGCCGAATG CTGATATTGC ACTAAAAGTC GCACTGCAAC ATTGTCCTCA GCTAGCCAGC
MT2.1.7 AATATCGCAT CCGGCCGAAT GCTGACATTC CGGCTGAATG CTGACATTGC ACTAAAAGTC GCACTGCAAC ATTGTCOCCA GCTAGCCAGC
MT2.1.8 AACACCCGCAT CCGGCCGAAT GCTGATATTG CGGCCGAATG CTGACATTGC ATTTAAAAGTC GCACTGCAAC ATTGTCOCCA GCTAGCCAGC
MT2.1.9 AACACTGTAT TTGGCCGAAT GCTGACATTC CGGCCGAATG CTGATATTGT ACTAAAAGTT GCACTGCAAC ATTGTCOCCA GCTAGCCAGC
MT2.1.10 AACACCCGCAT CCGGCCGAAT GCTGACATTC CGGCCGAATG CTGACATTGC ACTAAAAGTC GTATTGCAAC ATTGTCOCCA GCTAGTCAGC
MT2.1.11 AACACCCGCAT CCGGCCGAAT GCTGACATTC CGGCCGAATG CTGACATTGC ACTAAAAGTC GTATTGCAAC ATTGTCOCCA GCTAGTCAGC
MT2.1.12 AATATCGCAT CCGGCCGAAT GCTGACATTC CGGCCGAATG CTGACATTGT ACTAAAAGTC GCACTGCAAC ATTGTTTTTA GTTAGCCAGC
MT2.1.13 AACATCGCAT CCGGCCGAAT GCTGACATTC CGGCCGAATG CTGACATTGC ACTAAAAGTC GCACTGCAAC ATTGTCOCCA GCTAGCCAGC
MT2.2.2 -----AAT AITG-TATTT -GGTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
MT2.2.3 -----AAT AITG-TATTT -GGTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
MT2.2.4 AATATTGTAT TTGGTTGAAT GTTGATATTG TGGTGAATG TTGATATTGT ATTTAAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
MT2.2.5 AATATCGCAT CCGGCCGAAT GCTGACATTC CGGCCGAATG CTGACATTGC ACTAAAAGTC GCACTGCAAC ATTGTCOCCA GCTAGCCAGC
MT2.2.6 AACACCCGCAT CCGACCGAAT GCTGACATTC CGGCCGAATG CTGACATTGC ACTAAAAGTC GCACTGCAAC ATTATCCOCCA GCTAGCCAGC
MT2.2.7 AACACCCGCAT CCGGCCGAAT GCTGACATTC CGGCCGAATG CTGACATTGC ACTAAAAGTC GCACTGCAAC ATTGTCOCCA GCTAGCCAGC
MT2.2.8 AATACCGCAT CCGGCCGAAT GCTGACATTC CGGTCGAATG CTGACATTGC ACTAAAAGTC GCATTGCAAC ATTGTCOCCA GCTAGTCAGC
MT2.2.9 AACACCCGCAT CCAACCAAAC GCTGACATTC CAGCCGAATG CTGACATTGC ACTAAAAGTC GCACTGCAAC ATTGTCOCCA GCTAGCCAGC
MT2.2.10 AACACCCGCAT CCGGCCGAAT GCTGACATTC CGGCCGAATG CTGACATTGC ACTAAAAGTC GCACTGCAAC ATTATCCOCCA ACTAGCCAGC
MT2.2.11 AACACCCGCAT CCGGCCGAAT GCTGACATTC CGGCCGAATG CTGACATTGT ACTAAAAGTC GCACTGCAAC ATTGTCCTCA GCTAGCCAGC
MT2.2.12 AACACCCGCAT CCGGCCGAAT GCTGACATTC CGGCCGAATG CTGACATTGC ACTAAAAGTC GCACTGCAAC ATTGTCOCCA GCTAGCCAGC
MT2.2.13 AACACCCGCAT CCGGCCGAAT GCTGACATTC CGGCCGAATG CTGACATTGC ACTAAAAGTC GCACTGCAAC ATTGTCOCCA GCTAGCCAGC
MT2.2.14 AATACCCGCAT CCGGCCGAAT GCTGACATTC CGGTCGAATG CTGACATTGC ACTAAAAGTC GCATTGCAAC ATTGTCOCCA GCTAGTCAGC
```


reference	AACCTAAAAC	CCTTCATACG	TATACACATC	TTATCITTCAC	ATACGGCTCC	ACCCTOGAGT	GCACGGACTT	AACTGGCGCA	GCCGGTAGGA
MT1.1.1	AACCTAAAAC	CCTTCATACG	TATACACATC	TTATCITTCAC	ATACGGCTCC	ACCCTOGAGT	GCACGGACTT	AACTGGCGCA	ACCGGTAAAA
MT1.1.2	AACCTAAAAC	CCTTCATACG	TATACACATC	TTATCITTCAC	ATACGGCTCC	ACCCTOGAGT	GCACGGACTT	AACTGGCGCA	GCCGGTAGGA
MT1.1.3	AACCTAAAAC	CCTTCATACG	TATACACATC	TTATCITTCAC	ATACGGCTCC	ACCCTOGAGT	GCACGGACTT	AACTGGCGCA	GCCGGTAGGA
MT1.1.4	AAITTTAAAAT	TTTTTATATG	TATATATATT	TTATTTTAT	ATACGGCTCC	ACCCTOGAGT	GCACGGATT	AATTGGCGCA	GCCGGTAGGA
MT1.1.5	AACCTAAAAC	CCTTCATATG	TATATACATC	TTATCITTCAC	ATACGGCTCC	ACCCTOGAGT	GCACGGACTT	AATTGGTGT	GTGGTAGGA
MT1.1.6	AACCTAAAAC	CCTTCATACG	TATACACATC	TTATCITTCAC	ATACAACTCC	ACCCTOGAGT	GCACGGACTT	AACTGGCGCA	GCCGGTAGGA
MT1.1.7	AACCTAAAAC	CCTTCATACG	TATACACATC	TTATCITTCAC	ATACGGCTCC	ACCCTOGAGT	GCACGGACTT	AACTGGCGCA	GCCGGTAGGA
MT1.1.8	AACCTAAAAC	CCTTCATACG	TATACACATC	TTATCITTCAC	ATACGGCTCC	ACCCTOGAGT	GCACGGACTT	AACTGGCGCA	GCCAGTAAAA
MT1.1.9	AACCTAAAAC	TCITTTATATG	TATATACATC	TTATCITTCAT	ATATGGCTCC	ACCCTOGAGT	GCACGGACTT	AACTGGTGT	GCCGGTAGGA
MT1.1.10	AAITTTAAAAT	TTTTTATATG	TATATATATT	TTATCITTCAC	ATACGGCTCC	ACCCTOGAGT	GCATGGACTT	AACTGGCGCA	GCCGGTAGGA
MT1.2.1	AAITTTAAAAT	CCTTCATACG	TATACATATT	TTATTTTAC	ATACGGCTCC	ACCCTOGAGT	GTACGGACTT	AACTGGTGT	GCCGGTAGGA
MT1.2.2	AAITTTAAAAT	CCTTCATACG	TATACATATT	TTATTTTAC	ATACGGCTCC	ACCCTOGAGT	GTACGGACTT	AACTGGTGT	GCCGGTAGGA
MT1.2.3	AACCTAAAAC	CCTTCATACG	TATACACATC	TTATCITTCAC	ATACAACTCC	ACCCTOGAGT	GCACGGACTT	AACTGGCGCA	GCCGGTAGGA
MT1.2.4	AACCTAAAAC	CCTTCATACG	TATACACATC	TTATCITTCAC	ATACGGCTCC	ACCCTOGAGT	GCACGGACTT	AACTGGCGCA	GCCGGTAGGA
MT1.2.5	AACCTAAAAC	CCTTCATACG	TATACACATC	TTATCITTCAC	AAACGGCTCC	ACCCTCAAGT	GCACGGACTT	AACTGGCGCA	GCCGATAAGA
MT1.2.6	AACCTAAAAC	CCTTCATACG	TATACACATC	TTATCITTCAC	ATACAGCTCC	ACCCTOGAGT	ACACGAATT	AACTGGCGCA	GCCGGTAGGA
MT1.2.7	AAITTTAAAAT	TTTTTATATG	TATATATATC	TTATCITTCAC	ATATGGTTTC	ACCCTTNGAGT	GCACGGACTT	AACTGGCGCA	GCCGGTAGGA
MT1.2.8	AACCTAAAAC	CCTTCATACG	TATACACATC	TTATTTTAT	ATACGGCTCC	ACCCTOGAGT	GCACGGACTT	AATTGGTGT	GTGGTAGGA
MT1.2.9	AAITTTAAAAT	TTTT-ATATA	TTTATATATT	ATGTTTTTAT	AAAAGGTTTT	ATTTTTGATT	--ATGGATT	AATTGGTGT	GTGGTAGGA
MT1.2.10	AACCTAAAAC	CCTTCATACG	TATACACATC	TTATTTTAT	ATACGGCTCC	ACCCTOGAGT	GCACGGACTT	AATTGGTGT	GTGGTAGGA
MT1.2.11	AAITTTAAAAT	CCTTCATACG	TATACATATT	TTATTTTAC	ATACAACTCC	ACCCTOGAAT	GCACGGACTT	AACTGGCGCA	GCCGGTAGGA
MT1.2.12	AACCTAAAAC	CCTTCATACG	TATACACATC	TTATCITTCAC	ATACAACTCC	ACCCTOGAAT	GCACGGACTT	AACTGGCGCA	GCCGGTAGGA
MT2.1.1	AAITTTAAAAT	TTTT-ATATA	TTTATATATT	ATATTTTAT	AAAAGGTTTT	ATTTTTGATT	--ATGGATT	AATTGGTGT	GTGGTAGGA
MT2.1.2	AACTTTGAGT	TTTG-----	--TATATATT	ATATTTTAT	AAAAGGTTTT	ATTTTTGATT	--ATGGATT	AATTGGTGT	GTGGTAGGA
MT2.1.3	AAITTTAAAAT	TTTT-ATATA	TTTATATATT	ATATTTTAT	AAAAGGTTTT	ATTTTTGATT	--ATGGATT	AATTGGTGT	GTGGTAGGA
MT2.1.4	AAITTTAAAAT	TTTT-ATATA	TTTATATATT	ATATTTTAT	AAAAGGTTTT	ATTTTTGATT	--ATGGATT	AATTGGTGT	GTGGTAGGA
MT2.1.5	AAITTTAAAAT	TTTTTATATG	TATATACATT	TTATTTTAT	ATATGGTTTT	ATTTTTGATT	GCACGGACTT	AACTGGCGCA	GCTGGTAGGA
MT2.1.6	AACCTAAAAC	TTTTTATATG	TATATATATC	TTATTTTAT	ATACGGCTCC	ACCCTTNGAGT	GCACGGACTT	AACTGGCGCA	GCCGGTAGGA
MT2.1.7	AAITTTAAAAT	CCTTCATACG	TATACATATT	TTATCITTCAC	ATACGGTTC	ACCCTTNGAGT	GCACGGACTT	AACTGGCGCA	GCCGGTAGGA
MT2.1.8	AACCTAAAAC	TTTTTATATG	TATATATATT	TTATCITTCAC	ATACGGCTTC	ATCTOGAGT	GCACGGACTT	AATTGGTGT	GTGGTAGGA
MT2.1.9	AACCTAAAAC	CCTTCATATG	TATACACATC	TTATCITTCAC	ATACGGCTCC	ACCCTOGAGT	GCACGGACTT	AACTGGCGCA	GCCGGTAGGA
MT2.1.10	AACTTTAAAAT	TTTTTATACG	TATATACATT	TTATTTTAC	ATACGGCTCC	ATCTOGAGT	GCACGGACTT	AACTGGCGCA	GTGGTAGGA
MT2.1.11	AACTTTAAAAT	TTTTTATACG	TATATACATT	TTATTTTAC	ATACGGCTCC	ATCTOGAGT	GCACGGACTT	AACTGGCGCA	GTGGTAGGA
MT2.1.12	AAITTTAAAAT	CCTTTATATG	TATATATATC	TTATCITTCAC	ATACGGTTC	ACTCTOGAGT	GCACGGACTT	AACTGGCGCA	GTGGTAGGA
MT2.1.13	AACCTAAAAC	CCTTCATACG	TATATATATT	TTATCITTCAC	ATACGGCTCC	ACCCTOGAGT	GCACGGACTT	AACTGGCGCA	GCCGGTAGGA
MT2.2.2	AAITTTGAGT	TTTG-----	--TATACATT	ATATTTTAT	AAAAGGTTTT	ATTTTTGATT	--ATGGATT	AATTGGTGT	GTGGTAGGA
MT2.2.3	AAITTTAAAAT	TTTT-ATATA	TTTATATATT	ATATTTTAT	AAAAGGTTTT	ATTTTTGATT	--ATGGATT	AATTGGTGT	GTGGTAGGA
MT2.2.4	AAITTTAAAAT	TTTTTATATG	TATATATATT	TTATTTTAT	ATATGGTTTT	ATTTTTGATT	GCATGGATT	AATTGGTGT	GTGGTAGGA
MT2.2.5	AAITTTAAAAT	TTTTTATATG	TATATATATT	TTATCITTCAC	ATACGGCTCC	ACCCTOGAGT	GCATGGACTT	AACTGGCGCA	GCCGGTAGGA
MT2.2.6	AACCTAAAAC	CCTTCATACG	TATACACATC	TTATCITTCAC	ATACGGCTCC	ACCCTOGAGT	GCACGGACTT	AACTGGCGCA	GCCGGTAGGA
MT2.2.7	AACCTAAAAT	CCTTCATATG	TATATATATC	TTATCITTCAC	ATACGGCTCC	ACTTTNGAGT	GTATGGACTT	AACTGGCGCA	GCCGGTAGGA
MT2.2.8	AACCTAAAAT	TTTTTATATG	TATATATATC	TTATCITTCAC	ATACGGCTCC	ACCCTOGAGT	GCATGGACTT	AACTGGCGCA	GTGGTAGGA
MT2.2.9	AACCTAAAAC	CCTTCATACA	TATACACATC	TTATCITTCAC	ATACGGCTCC	ACCCTOGAGT	GCACGGACTT	AACTGGCGCA	GTCAATAAAA
MT2.2.10	AACCTAAAAC	CCTTCATACA	TATACACATC	TTATCITTCAC	ATACGGCTCC	ACCCTOGAGT	GCACGGACTT	AACTGGCGCA	GCCGGTAGGA
MT2.2.11	AACCTAAAAC	CCTTTATATG	TATATATATC	TTATCITTCAC	ATACGGCTCC	ACCCTOGAGT	GCACGGACTT	AACTGGCGCA	GCCGGTAGGA
MT2.2.12	AACCTAAAAT	CCTTCATATG	TATATATATC	TTATCITTCAC	ATACGGCTCC	ACTTTNGAGT	GTATGGACTT	AACTGGCGCA	GCCGGTAGGA
MT2.2.13	AAITTTAAAAT	TTTTTATACG	TATACACATT	TTATTTTAT	ATATGGCTCC	ACCCTOGAGT	GCACGGACTT	AACTGGCGCA	GCCGGTAGGA
MT2.2.14	AACCTAAAAT	TTTTTATATG	TATATATATC	TTATCITTCAC	ATACGGCTCC	ACCCTOGAGT	GCATGGACTT	AACTGGCGCA	GTGGTAGGA

Alignment: /Users/benjaminhubert/Desktop/thèse/data/DNA_methylation/STRAINS/MAKINDU/TESTES/TIR52/MAKINDU.txt
Seaview [blocks=10 fontsize=8 A4] on Tue Oct 12 17:02:31 2010

```
1
reference CCTCTGGCTG ACAATAAACT TCACTCTGAT TTTGAACTTC AAGAAAAGTCA GTCGTATTCT TTATTTGAAA TCTTTACACT ACAAAATATCT
MT1.1 TTATTTTTTG ATAATAAAAT TTATTTTGAT TTTGAAATTT AAGAAAAGTTA GTTGTATTTT TTATTGGAAA TTTTATATAT ATAAATATTT
MT1.2 TTATTTTTTG ATAATAAAAT TTATTTTGAT TTTGAAATTT AAGAAAAGTTA GTTGTATTTT TTATTGGAAA TTTTATATAT ATAAATATTT
MT1.3 TTATTTTTTG ATAATAAAAT TTATTTTGAT TTTGAAATTT AAGAAAAGTTA GTTGTATTTT CTATTGGAAA TTTTATATAT ATAAATATTT
MT1.4 --TTTTTTTG ATAATAAAAT TTATTTTGAT TTTGAAATTT AAGAAAAGTTA GTTGTATTTT TTATTGGAAA TTTTATATAT ATAAATATTT
MT1.5 TTATTTTTTG ATAATAAAAT TTATTTTGAT TTTGAAATTT AAGAAAAGTTA GTTGTATTTT TTATTGGAAA TTTTATATAT ATAAATATTT
MT2.1 TTATTTTTTG ATAATAAAAT TTATTTTGAT TTTGAAATTT AAGAAAAGTTA GTTGTATTTT TTATTGGAAA TTTTATATAT ATAAATATTT
MT2.2 TTATTTTTTG ATAATAAAAT TTATTTTGAT TTTGAAATTT AAGAAAAGTTA GTTGTATTTT TTATTGGAAA TTTTATATAT ATAAATATTT
MT2.3 --TTTTTTTG ATAATAAAAT TTATTTTGAT TTTGAAATTT AAGAAAAGTTA GTTGTATTTT TTATTGGAAA TTTTATATAT ATAAATATTT
MT2.4 --TTTTTTTG ATAATAAAAT TTATTTTGAT TTTGAAATTT AAGAAAAGTTA GTTGTATTTT TTATTGGAAA TTTTATATAT ATAAATATTT
MT2.5 TTATTTTTTG ATAATAAAAT TTATTTTGAT TTTGAAATTT AAGAAAAGTTA GTTGTATTTT TTATTGGAAA TTTTATATAT ATAAATATTT
MT2.6 TTATTTTTTG ATAATAAAAT TTATTTTGAT TTTGAAATTT AAGAAAAGTTA GTTGTATTTT TTATTGGAAA TTTTATATAT ATAAATATTT
```

```
91
reference GGGGAAACCT AAAACCCCTC ATACGTATAC ACATCTTATC TTCACATACG GCTCCACCCCT CGAGTGCACG GACTTAACTG GCGCAGCCGG
MT1.1 GTTGAAATTT AAAAATTTTT ATATATTTAT ATATTATATT TTTATAAAAAG GTTTTATTTT TGATT--ATG GATTTAATTG GTGTAGTTGG
MT1.2 GGGGAAATTT AAAA--TTTT ATATATTTAT ATATTATATT TTTATAAAAAG GTTTTATTTT TGATT--ATG GATTTAATTG GTGTAGTTGG
MT1.3 GTTGAAATTT AAAAATTTTT ATATATTTAT ATATTATATT TTTATAAAAAG GTTTTATTTT TGATT--ATG GATTTAATTG GTGTAGTTGG
MT1.4 GGGGAAATTT AAAAATTTTT ATAT--TATAT ATATTTTTAT TTTATAAAAAG GTTTTATTTT TGAT---ATG GATTTAATTG GTGTAGTTGG
MT1.5 GTTGAAATTT AAAAATTTTT ATATATTTAT ATATTATATT TTTATAAAAAG GTTTTATTTT TGATT--ATG GATTTAATTG GTGTAGTTGG
MT2.1 GTTGAAATTT AAAAATTTTT ATATATTTAT ATATTATATT TTTATAAAAAG GTTTTATTTT TGATT--ATG GATTTAATTG GTGTAGTTGG
MT2.2 GTTGAAACCT AAAATTTTTT -----TTAT ATATTATATT TTTATAAAAAG GTTTTATTTT TGATT--ATG GATTTAATTG GTGTAGTTGG
MT2.3 GGGGAAATTT AAAAATTTTT ATAT--TATAT ATATTTTTAT TTTATAATATG GTTTTATTTT TGAT---ATG GATTTAATTG GTGTAGTTGG
MT2.4 GGGGAAATTT AAAAATTTTT ATAT--TATAT ATATTTTTAT TTTATAAAAAG GTTTTATTTT TGAT---ATG GATTTAATTG GTGTAGTTGG
MT2.5 GTTGAAATTT AAAATTTTTT -----TTAT ATATTATATT TTTATAAAAAG GTTTTATTTT TGATT--ATG GATTTAATTG GTGTAGTTGG
MT2.6 GTTGAAATTT AAAATTTTTT -----TTAT ATATTATATT TTTATAAAAAG GTTTTATTTT TGATT--ATG GATTTAATTG GTGTAGTTGG
```

```
181
reference TAGGATGTCC TACCTTTTAA TAATCACCTA CCTGTAAGTA ACCATTGTAA GAAACGAAGC AAACATATATG CAAGATGTCC ACTGAAAAGTG
MT1.1 TAGGATGTTT TATTTAATTA TAATTATTTA TTTGTAAGTA A-TA-TGTAA GAAATGAAGT AAATTATATG TAAGATGTTG ATTGAAAAGTG
MT1.2 TAGGATGTTT TATTTAATTA TAATTATTTA TTTGTAAGTA AATA-TGTAA GAAATGAAGT AAATTATATG TAAGATGTTG ATTGAAAAGTG
MT1.3 TAGGATGTTT TATTTAATTA TAATTATTTA TTTGTAAGTA AATA-TGTAA GAAATGAAGT AAATTATATG TAAGATGTTG ATTGAAAAGTG
MT1.4 TAGGATGTTT TATTTTAA TAATTATTTA TTTGTAAGTA ATTA-TGTAA GAAATGAAGT AAATTATATG TAAGATGTTG ATTGAAAAGTG
MT1.5 TAGGATGTTT TATTTAATTA TAATTATTTA TTTGTAAGTA AAT--TGTAA GAAATGAAGC AAATTATATG TAAGATGTTG ATTGAAAAGTG
MT2.1 TAGGATGTTT TATTTAATTA TAATTATTTA TTTGTAAGTA AATA-TGTAA GAAATGAAGT AAATTATATG TAAGATGTTG ATTGAAAAGTG
MT2.2 TAGGATGTTT TATTTAATTA TAATTATTTA TTTGTAAGTA AATA-TGTAA GAAATGAAGT AAATTATATG TAAGATGTTG ATTGAAAAGTG
MT2.3 TAGGATGTTT TATTTTAA TAATTATTTA TTTGTAAGTA ATTA-TGTAA GAAATGAAGT AAATTATATG TAAGATGTTG ATTGAAAAGTG
MT2.4 TAGGATGTTT TATTTTAA TAATTATTTA TTTGTAAGTA ATTA-TGTAA GAAATGAAGT AAATTATATG TAAGATGTTG ATTGAAAAGTG
MT2.5 TAGGATGTTT TATTTAATTA TAATTATTTA TTTGTAAGTA AATA-TGTAA GAAATGAAGT AAATTATATG TAAGATGTTG ATTGAAAAGTG
MT2.6 TAGGATGTTT TATTTAATTA TAATTATTTA TTTGTAAGTA AATA-TGTAA GAAATGAAGT AAATTATATG TAAGATGTTG ATTGAAAAGTG
```

```
271
reference ACTAGGAACA AATTTTTATA AAACAA
MT1.1 ATTAGGAATA AATTTTTATA AAATAA
MT1.2 ATTAGGAATA AATTTTTATA AAATAA
MT1.3 ATTAGGAATA AATTTTTATA AAATAA
MT1.4 AATAAGGAATA AATTTTTAAA AAATAA
MT1.5 ATTAGGAATA AATTTTTATA AAATAA
MT2.1 ATTAGGAATA AATTTTTATA AAGTAA
MT2.2 ATTAGGAATA AATTTTTATA AAATAA
MT2.3 AATAAGGAATA AATTTTTAAA AAATAA
MT2.4 AATAAGGAATA AATTTTTAAA AAATAA
MT2.5 ATTAGGAATA AATTTTTATA AAATAA
MT2.6 ATTAGGAATA AATTTTTATA AAATAA
```

Alignment: /Users/benjaminhubert/Desktop/thèse/data/DNA_methylation/STRAINS/CHICHARO/OVARIES/TIR01/CHICHARO.txt
 Seaview [blocks=10 fontsize=8 A4] on Tue Oct 12 16:57:17 2010

```

1
reference ACACCCGCATC CGGCCGAATG CTGACATTCG GGCCGAATGC TGACATTGCA CAAAAGTCGC ACTGCAACAT TGTCCCCAGC TAGCCAGCCA
col.1 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.2 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.3 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.4 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.5 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.6 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.7 -----ATA CTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.8 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.9 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.10 ATATTGTATT TGGTTGAATG TTGATATTTT GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.1 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.2 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.3 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.4 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGCAATAT TGTTTTTAGT TAGTTAGTTA
col.5 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.6 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.7 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.8 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.9 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.10 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGCA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.11 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.1.2 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.1.3 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.1.4 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.1.5 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.1.7 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.1.8 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.1.9 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGCA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.1.10 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.1.11 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.1.12 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.1.13 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTCAGTTA
col.2.1 ATATTGTATT TGGTTGAATG TTGATATTTT GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.2.2 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.2.3 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.2.4 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.2.5 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.2.6 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.2.7 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.2.8 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.2.9 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.2.10 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.2.11 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.2.12 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.2.13 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.2.14 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTG
col.2.15 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA
col.2.16 -----ATA TTG-TATTT-GGTTGGAATGT TGATATTGTA TAAAAGTTGT ATTGTAATAT TGTTTTTAGT TAGTTAGTTA

```


Alignment: /Users/benjaminhubert/Desktop/thèse/data/DNA_methylation/STRAINS/CHICHARO/OVARIES/TIR52/CHICHARO.txt
Seaview [blocks=10 fontsize=8 A4] on Tue Oct 12 16:57:36 2010

```
1
C01.1 TTATTTTTCG ATAATAAATT TTATTTTGAT TTTGAATTTT AAGAAAGTTA GTTGTATTTT TTATTGGAAA TTTTATATTT ATAAATATTT
C01.2 TTATTTTTCG ATAATAAATT TTATTTTGAT TTTGAATTTT AAGAAAGTTA GTTGTATTTT TTATTGGAAA TTTTATATTT ATAAATATTT
C01.3 TTATTTTTCG ATAATAAATT ATATTTTGAT TTTGAATTTT AAGAAAGTTA GTTGTATTTT TTATTGGAAA TTTTATATTT ATAAATATTT
C01.4 TCATTTTTCG ACAATAAATT ATATTTTGAT TTTGAATTTT AAGAAAGTTA GTTGTATTTT TTATTGGAAA TTTTATATTT ATAAATATTT
C01.5 TTATTTTTCG ATAATAAATT TTATTTTGAT TTTGAATTTT AAGAAAGTTA GTTGTATTTT TTATTGGAAA TTTTATATTT ATAAATATTT
C02.1 TTATTTTTCG ATAATAAATT TTATTTTGAT TTTGAATTTT AAGAAAGTTA GTTGTATTTT TTATTGGAAA TTTTATATTT ATAAATATTT
C02.2 TTATTTTTCG ATAATAAATT ATATTTTGAT TTTGAATTTT AAGAAAGTTA GTTGTATTTT TTATTGGAAA TTTTATATTT ATAAATATTT
C02.3 TTATTTTTCG ATAATAAATT TTATTTTGAT TTTGAATTTT AAGAAAGTTA GTTGTATTTT TTATTGGAAA TTTTATATTT ATAAATATTT
C02.4 TTATTTTTCG ATAATAAATT TTATTTTGAT TTTGAATTTT AAGAAAGTTA GTTGTATTTT TTATTGGAAA TTTTATATTT ATAAATATTT
C02.5 TTATTTTTCG ATAATAAATT TTATTTTGAT TTTGAATTTT AAGAAAGTTA GTTGTATTTT TTATTGGAAA TTTTATATTT ATAAATATTT

91
C01.1 GTTGAATTTT AAAAATTTTT ATATATTTAT ATATTATATT TTTATAAAAAG GTTTTATTTT TGATTATGGA TTTAATTTGGT GTAGTTGGTA
C01.2 GTTGAATTTT AAAAATTTTT ATATATTTAT ATATTATATT TTTATAAAAAG GTTTTATTTT TGATTATGGA TTTAATTTGGT GTAGTTGGTA
C01.3 GTTGAATTTT AAAAATTTTT ATATATTTAT ATATTATATT TTTATAAAAAG GTTTTATTTT TGATTATGGA TTTAATTTGGT GTAGTTGGTA
C01.4 GTTGAATTTT AAAAATTTTT ATATATTTAT ATATTATATT TTTATAAAAAG GTTTTATTTT TGATTATGGA TTTAATTTGGT GTAGTTGGTA
C01.5 GTTGAATTTT AAAAATTTTT ATATATTTAT ATATTATATT TTTATAAAAAG GTTTTATTTT TGATTATGGA TTTAATTTGGT GTAGTTGGTA
C02.1 GTTGAATTTT AAAAATTTTT ATATATTTAT ATATTATATT TTTATAAAAAG GTTTTATTTT TGATTATGGA TTTAATTTGGT GTAGTTGGTA
C02.2 GTTGAATTTT AAAAATTTTT ATATATTTAT ATATTATATT TTTATAAAAAG GTTTTATTTT TGATTATGGA TTTAATTTGGT GTAGTTGGTA
C02.3 GTTGAATTTT AAAAATTTTT ATATATTTAT ATATTATATT TTTATAAAAAG GTTTTATTTT TGATTATGGA TTTAATTTGGT GTAGTTGGTA
C02.4 GTTGAATTTT AAAAATTTTT ATATATTTAT ATATTATATT TTTATAAAAAG GTTTTATTTT TGATTATGGA TTTAATTTGGT GTAGTTGGTA
C02.5 GTTGAATTTT AAAAATTTTT ATATATTTAT ATATTATATT TTTATAAAAAG GTTTTATTTT TGATTATGGA TTTAATTTGGT GTAGTTGGTA

181
C01.1 GGATGTTTTA TTTATTAATA ATCATTTATT TGTAAAGTAAA TATGTAAGAA ATGAAGTAAA TTATATGTAA GATGTTGATT GAAAGTGATT
C01.2 GGATGTTTTA TTTATTAATA ATCATTTATT TGTAAAGTAAA TATGTAAGAA ATGAAGTAAA TTATATGTAA GATGTTGATT GAAAGTGATT
C01.3 GGATGTTTTA TTTATTAATA ATTATTTATT TGTAAAGTAAA TATGTAAGAA ATGAAGTAAA CTATATGTAA GATGTTGATT GAAAGTGATT
C01.4 GGATGTTTTA TTTATTAATA ATTATTTATT TGTAAAGTAAA TATGTAAGAA ATGAAGTAAA TTATATGTAA GATGTTGATT GAAAGTGATT
C01.5 GGATGTTTTA TTTATTAATA ATTATTTATT TGTAAAGTAAA TATGTAAGAA ATGAAGTAAA TTATATGTAA GATGTTGATT GAAAGTGATT
C02.1 GGATGTTTTA TTTATTAATA ATTATTTATT TGTAAAGTAAA TATGTAAGAA ATGAAGTAAA TTATATGTAA GATGTTGATT GAAAGTGATT
C02.2 GGATGTTTTA TTTATTAATA ATTATTTATT TGTAAAGTAAA TATGTAAGAA ATGAAGTAAA TTATATGTAA GATGTTGATT GAAAGTGATT
C02.3 GGATGTTTTA TTTATTAATA ATTATTTATT TGTAAAGTAAA TATGTAAGAA ATGAAGTAAA TTATATGTAA GATGTTGATT GAAAGTGATT
C02.4 GGATGTTTTA TTTATTAATA ATTATTTATT TGTAAAGTAAA TATGTAAGAA ATGAAGTAAA TTATATGTAA GATGTTGATT GAAAGTGATT
C02.5 GGATGTTTTA TTTATTAATA ATTATTTATT TGTAAAGTAAA TATGTAAGAA ATGAAGTAAA TTATATGTAA GATGTTGATT GAAAGTGATT

271
C01.1 AGGAATAAAT TTTTATAAAA TAA
C01.2 AGGAATAAAT TTTTATAAAA TAA
C01.3 AGGAATAAAT TTTTATAAAA TAA
C01.4 AGGAATAAAT TTTTATAAAA TAA
C01.5 AGGAATAAAT TTTTATAAAA TAA
C02.1 AGGAATAAAT TTTTGTAAAA TAA
C02.2 AGGAATAAAT TTTTATAAAA TAA
C02.3 AGGAATAAAT TTTTATAAAA TAA
C02.4 AGGAATAAAT TTTTATAAAA TAA
C02.5 AGGAATAAAT TTTTATAAAA TAA
```


Alignment: /Users/benjaminhubert/Desktop/thèse/data/DNA methylation/STRAINS/CHICHARO/TESTES/TIRO1/CHICHARO.txt
 Seaview [blocks=10 fontsize=8 A4] on Tue Oct 12 16:57:57 2010

```

1
reference ACACCCGCAATC CGGCCGAATG CFCGACATTC GGCAGCAATG TGCATPTGCA CAAAAGTTCG ACTGCAACAT TGTCGCCAGC TAGCCAGCCA
D1CT1.1 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D1CT1.3 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D1CT1.4 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTATP TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D1CT1.5 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D1CT1.7 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGCA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D1CT1.8 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D1CT1.9 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D1CT1.10 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D1CT1.11 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D1CT1.12 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D1CT2.1 ATATPTGTATP TGGPTGAATG TCGATATPTT GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D1CT2.2 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D1CT2.3 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D1CT2.4 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D1CT2.5 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D1CT2.6 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D1CT2.8 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D1CT2.9 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGCAATAT TGTPPTTAGT TAGTTAGTTA
D1CT2.11 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D1CT2.12 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D1CT2.13 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D1CT2.14 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D1CT2.15 -----ATA TFG-TATTT- GGTGGAATGC TGCATPTGTA CAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D2CT1.2 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D2CT1.3 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D2CT1.6 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D2CT1.7 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D2CT1.8 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D2CT1.9 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D2CT1.11 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D2CT1.14 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D2CT1.15 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D2CT1.16 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D2CT1.17 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D2CT1.18 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGCA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D2CT1.19 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D2CT2.2 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D2CT2.10 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATCGTAATAT TGTPPTTAGT TAGTTAGTTA
D2CT2.12 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D2CT2.13 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D2CT2.15 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D2CT2.16 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D2CT2.17 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D2CT2.18 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D2CT2.20 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA
D2CT2.24 -----ATA TFG-TATTT- GGTGGAATGT TGCATPTGTA TAAAAGTTGT ATTGTAATAT TGTPPTTAGT TAGTTAGTTA

```


Alignment: /Users/benjaminhubert/Desktop/thèse/data/DNA_methylation/STRAINS/CHICHARO/TESTES/TIR52/CHICHARO.txt
 Seaview [blocks=10 fontsize=8 A4] on Tue Oct 12 16:58:18 2010

```

1
reference  CCACFTTCTG  ACAATAAAC  ACACCTGTG  TTTGAACCTC  AAGAAAGTCA  GTCGTATCTE  TTATTGGAAA  TCFTCACACT  ACAAATATCT
CT1.1     TTATFTTTTG  ATAATAAAT  TTATTTTGAT  TTTGAATFTT  AAGAAAGTTA  GTTGTATPTE  T-ATTTGGAAA  TTTTATATAT  ATAAATATTT
CT1.2     TTATFTTTTG  ATAATAAAT  ATATTTTGAT  TTTGAATFTT  AAGAAAGTTA  GTTGTATPTE  TTATTGGAAA  TTTTATATAT  ATAAATATTT
CT1.3     TTATFTTTTG  ATAATAAAT  TTATTTTGAT  TTTGAATFTT  AAGAAAGTTA  GTTGTATPTE  T-ATTTGGAAA  TTTTATATAT  ATAAATATTT
CT1.4     TTATFTTTTG  ATAATAAAT  TTATTTTGAT  TTTGAATFTT  AAGAAAGTTA  GTTGTATPTE  TTATTGGAAA  TTTTATATAT  ATAAATATTT
CT2.1     TTATFTTTTG  ATAATAAAT  ATATTTTGAT  TTTGAATFTT  AAGAAAGTTA  GTTGTATPTE  TTATTGGAAA  TTTTATATAT  ATAAATATTT
CT2.2     TTATFTTTTG  ATAATAAAT  ATATTTTGAT  TTTGAATFTT  AAGAAAGTTA  GTTGTATPTE  TTATTGGAAA  TTTTATATAT  ATAAATATTT
CT2.3     TTATFTTTTG  ATAATAAAT  TTATTTTGAT  TTTGAATFTT  AAGAAAGTTA  GTTGTATPTE  TTATTGGAAA  TTTTATATAT  ATAAATATTT
CT2.4     TTATFTTTTG  ATAATAAAT  ATATTTTGAT  TTTGAATFTT  AAGAAAGTTA  GTTGTATPTE  TTATTGGAAA  TTTTATATAT  ATAAATATTT
  
```

```

91
reference  CCTGAAACTT  AAAAACCTTC  ATACATTTAC  ACATCATATC  TTCACAAAAG  GCTCCACCCF  CGATCACGGA  CTTAACTGGC  GCAGCCGGTA
CT1.1     GTTGAAATTT  AAAAATTTT  ATATATTTAT  ATATTATAT  TTTATAAAAG  GTTTTATPTE  TGATTATGGA  TTTAAITGGT  GTAGTTGGTA
CT1.2     GTTGAAATTT  AAAAATTTT  ATATATTTAT  ATATTATAT  TTTATAAAAG  GTTTTATPTE  TGATTATGGA  TTTAGITGGT  GTAGTTAGTA
CT1.3     GTTGAAATTT  AAAAATTTT  ATATATTTAT  ATATTATAT  TTTATAAAAG  GTTTTATPTE  TGATTATGGA  TTTAAITGGT  GTAGTTGGTA
CT1.4     GTTGAAATTT  AAAAATTTT  ATATATTTAT  ATATTATAT  TTTATAAAAG  GTTTTATPTE  TGATTATGGA  TTTAGITGGT  GTAGTTAGTA
CT2.1     GTTGAAATTT  AAAAATTTT  ATATATTTAT  ATATTATAT  TTTATAAAAG  GTTTTATPTE  TGATTATGGA  TTTAAITGGT  GTAGTTGGTA
CT2.2     GTTGAAATTT  AAAAATTTT  ATATATTTAT  ATATTATAT  TTTATAAAAG  GGTTTATPTE  TGATTATGGA  TTTAAITGGT  GTAGTTGGTA
CT2.3     GTTGAAATTT  AAAAATTTT  ATATATTTAT  ATATTATAT  TTTATAAAAG  GGTTTATPTE  TGATTATGGA  TTTAAITGGT  GTAGTTGGTA
CT2.4     GTTGAAATTT  AAAAATTTT  ATATATTTAT  ATATTATAT  TTTATAAAAG  GTTTTATPTE  TGATTATGGA  TTTAGITGGT  GTAGTTAGTA
  
```

```

181
reference  GGATGTCCTA  CCTATTAATA  ATCACCTACC  TGTAAAGTAA  CATGTAAGAA  ACGAAGCAA  CTATATGCAA  GATGTCGACT  GAAAGTGACT
CT1.1     GGATGTTTTA  TTTATTAATA  ATTATTTATT  TGTAAAGTAA  TATGTAAGAA  ATGAAGTAAA  TTATATGTAA  GATGTTGATT  GAAAGTGATT
CT1.2     GGATGTTTTA  TTTATTAATA  ATTATTTATT  TGTAAAGTAA  TATGTAAGAA  ATGAAGTAAA  TTATATGTAA  GATGTTGATT  GAAAGTGATT
CT1.3     GGATGTTTTA  TTTATTAATA  ATTATTTATT  TGTAAAGTAA  TATGTAAGAA  ATGAAGTAAA  TTATATGTAA  GATGTTGATT  GAAAGTGATT
CT1.4     GGATGTTTTA  TTTATTAATA  ATTATTTATT  TGTAAAGTAA  TATGTAAGAA  ATGAAGTAAA  TTATATGTAA  GATGTTGATT  GAAAGTGATT
CT2.1     GGATGTTTTA  TTTATTAATA  ATTATTTATT  TGTAAAGTAA  TATGTAAGAA  ATGAAGTAAA  TTATATGTAA  GATGTTGATT  GAAAGTGATT
CT2.2     GGATGTTTTA  TTTATTAATA  ATTATTTATT  TGTAAAGTAA  TATGTAAGAA  ATGAAGTAAA  TTATATGTAA  GATGTTGATT  GAAAGTGATT
CT2.3     GGATGTTTTA  TTTATTAATA  ATTATTTATT  TGTAAAGTAA  TATGTAAGAA  ATGAAGTAAA  TTATATGTAA  GATGTTGATT  GAAAGTGATT
CT2.4     GGATGTTTTA  TTTATTAATA  ATTATTTATT  TGTAAAGTAA  TATGTAAGAA  ATGAAGTAAA  TTATATGTAA  GATGTTGATT  GAAAGTGATT
  
```

```

271
reference  AGGAACAAAT  TTTTATAAAA  CAA
CT1.1     AGGAATAAAT  TTTTATAAAA  TAA
CT1.2     AGGAATAAAT  TTTTATAAAA  TAA
CT1.3     AGGAATAAAT  TTTTATAAAA  TAA
CT1.4     AGGAATAAAT  TTTTATAAAA  TAA
CT2.1     AGGAATAAAT  TTTTATAAAA  TAA
CT2.2     AGGAATAAAT  TTTTATAAAA  TAA
CT2.3     AGGAATAAAT  TTTTATAAAA  TAA
CT2.4     AGGAATAAAT  TTTTATAAAA  TAA
  
```


Reference	ACTAGGAACA	AATTTTATA	AAACAA
Z01.1	ATTAGGAATA	AATTTTATA	AAATAA
Z01.2	ATTAGGAATA	AATTTTATA	AAATAA
Z01.3	ATTAGGAATA	AATTTTATA	AAATAA
Z01.4	ATTAGGAATA	AATTTTATA	AAATAA
Z01.5	ATTAGGAATA	AATTTTATA	AAATAA
Z01.6	ATTAGGAATA	AATTTTATA	AAATAA
Z01.7	ATTAGGAATA	AATTTTATA	AAATAA
Z01.8	ATTAGGAATA	AATTTTATA	AAATAA
Z01.9	ATTAGGAATA	AATTTTATA	AAATAA
Z01.10	ATTAGGAATA	AATTTTATA	AAATAA
Z01.11	ATTAGGAATA	AATTTTATA	AAACAA
Z01.12	ATTAGGAATA	AATTTTATA	AAATAA
Z02.1	ATTAGGAATA	AATTTTATA	AAATAA
Z02.2	ATTAGGAATA	AATTTTATA	AAATAA
Z02.3	ATTAGGAATA	AATTTTATA	AAATAA
Z02.4	ATTAGGAATA	AATTTTATA	AAATAA
Z02.5	ATTAGGAATA	AATTTTATA	AAATAA
Z02.6	ATTAGGAATA	AATTTTATA	AAATAA
Z02.7	ATTAGGAATA	AATTTTATA	AAATAA
Z02.8	ATTAGGAATA	AATTTTATA	AAATAA
Z02.9	ATTAGGAATA	AATTTTATA	AAATAA
Z02.10	ATAAGAATA	AATTTTAAA	AAATAA
Z02.11	ATAAGAATA	AATTTTAAA	AAATAA

Alignment: /Users/benjaminhubert/Desktop/thèse/data/DNA_methylation/STRAINS/ZIMBABWE/TESTES/TIR52/ZIMBABWE.txt
 Seaview [blocks=10 fontsize=8 A4] on Tue Oct 12 17:06:13 2010

1
 Reference CCTCTGGCTG ACAATAAACT TCACTCTGAT TTTGAACTTC AAGAAAAGTCA GTCGTATTC TTAATTTGAAA TCPTTACACT ACAAAATATCT
 ZT1.1 TTAATTTTGG ATAATAAAAT TTAATTTTGG TTTGAAATTT AAGAAAAGTTA GTTGTATTTT TTAATTTGAAA TTTTATATAT ATAAATATTT
 ZT1.2 TTAATTTTGG ATAATAAAAT TTAATTTTGG TTTGAAATTT AAGAAAAGTTA GTTGTATTTT TTAATTTGAAA TTTTATATAT ATAAATATTT
 ZT1.3 TTAATTTTGG ATAATAAAAT TTAATTTTGG TTTGAAATTT AAGAAAAGTTA GTTGTATTTT TTAATTTGAAA TTTTATATAT ATAAATATTT
 ZT1.4 TTAATTTTGG ATAATAAAAT TTAATTTTGG TTTGAAATTT AAGAAAAGTTA GTTGTATTTT TTAATTTGAAA TTTTATATAT ATAAATATTT
 ZT1.5 TTAATTTTGG ATAATAAAAT TTAATTTTGG TTTGAAATTT AAGAAAAGTTA GTTGTATTTT TTAATTTGAAA TTTTATATAT ATAAATATTT
 ZT1.6 TTAATTTTGG ATAATAAAAT TTAATTTTGG TTTGAAATTT AAGAAAAGTTA GTTGTATTTT TTAATTTGAAA TTTTATATAT ATAAATATTT
 ZT2.1 TTAATTTTGG ATAATAAAAT TTAATTTTGG TTTGAAATTT AAGAAAAGTTA GTTGTATTTT TTAATTTGAAA TTTTATATAT ATAAATATTT
 ZT2.2 TTAATTTTGG ATAATAAAAT TTAATTTTGG TTTGAAATTT AAGAAAAGTTA GTTGTATTTT TTAATTTGAAA TTTTATATAT ATAAATATTT
 ZT2.3 TTAATTTTGG ATAATAAAAT TTAATTTTGG TTTGAAATTT AAGAAAAGTTA GTTGTATTTT TTAATTTGAAA TTTTATATAT ATAAATATTT
 ZT2.4 TTAATTTTGG ATAATAAAAT TTAATTTTGG TTTGAAATTT AAGAAAAGTTA GTTGTATTTT TTAATTTGAAA TTTTATATAT ATAAATATTT
 ZT2.5 TTAATTTTGG ATAATAAAAT TTAATTTTGG TTTGAAATTT AAGAAAAGTTA GTTGTATTTT TTAATTTGAAA TTTTATATAT ATAAATATTT

91
 Reference GGGGAACCTT AAAACCCCTT ACACGTATAC ACATCTTATC TTCACATACG GCTCCACCCCT CGAGTGCACG GACTTAACTG GCGCAGCCGG
 ZT1.1 GTTGAATTTT -----TTAT ATATTATAT TTTATAAAAG GTTTTATTTT TGATT--ATG GATTTAATFG GTGTAGTTGG
 ZT1.2 GTTGAATTTT -----TTAT ATATTATAT TTTATAAAAG GTTTTATTTT TGATT--ATG GATTTAATFG GTGTAGTTGG
 ZT1.3 GTTGAACCTT -----TTAT ATATTATAT TTTATAAAAG GTTTTATTTT TGATT--ATG GATTTAATFG GTGTAGTTGG
 ZT1.4 GTTGAATTTT -----TTAT ATATTATAT TTTATAAAAG GTTTTATTTT TGATT--ATG GATTTAATFG GTGTAGTTGG
 ZT1.5 GTTGAATTTT -----TTAT ATATTATAT TTTATAAAAG GTTTTATTTT TGATT--ATG GATTTAATFG GTGTAGTTGG
 ZT1.6 GTTGAATTTT AAAAATTTT ATATTATAT ATATTATAT TTTATAAAAG GTTTTATTTT TGATT--ATG GATTTAATFG GTGTAGTTGG
 ZT2.1 GTTGAATTTT AAAAATTTT ATATTATAT ATATTATAT TTTATAAAAG GTTTTATTTT TGATT--ATG GATTTAATFG GTGTAGTTGG
 ZT2.2 GTTGAATTTT -----TTAT ATATTATAT TTTATAAAAG GTTTTATTTT TGATT--ATG GATTTAATFG GTGTAGTTGG
 ZT2.3 GTTGAATTTT -----TTAT ATATTATAT TTTATAAAAG GTTTTATTTT TGATT--ATG GATTTAATFG GTGTAGTTGG
 ZT2.4 GTTGAATTTT AAAAATTTT ATATTATAT ATATTATAT TTTATAAAAG GTTTTATTTT TGATT--ATG GATTTAATFG GTGTAGTTGG
 ZT2.5 GTTGAATTTT -----TTAT ATATTATAT TTTATAAAAG GTTTTATTTT TGATT--ATG GATTTAATFG GTGTAGTTGG

181
 Reference TAGGATGTCC TACCTTTTAA TAATCACCTA CCTGTAAGTA ACCATTGTAA GAAACGAAGC AAATATATG CAAGATGTCG ACTGAAAGTG
 ZT1.1 TAGGATGTTT TATTTATTA TAATTATTTA TTTGTAAGTA AATAT-GTAA GAAATGAAGT AAATTATATG TAAGATGTTG ATTGAAAGTG
 ZT1.2 TAGGATGTTT TATTTATTA TAATTATTTA TTTGTAAGTA AATAT-GTAA GAAATGAAGT AAATTATATG TAAGATGTTG ATTGAAAGTG
 ZT1.3 TAGGATGTTT TATTTATTA TAATTATTTA TTTGTAAGTA AATAT-GTAA GAAATGAAGT AAATTATATG TAAGATGTTG ATTGAAAGTG
 ZT1.4 TAGGATGTTT TATTTATTA TAATTATTTA TTTGTAAGTA AATAT-GTAA GAAATGAAGT AAATTATATG TAAGATGTTG ATTGAAAGTG
 ZT1.5 TAGGATGTTT TATTTATTA TAATTATTTA TTTGTAAGTA AATAT-GTAA GAAATGAAGT AAATTATATG TAAGATGTTG ATTGAAAGTG
 ZT1.6 TAGGATGTTT TATTTATTA TAATTATTTA TTTGTAAGTA AATAT-GTAA GAAATGAAGT AAATTATATG TAAGATGTTG ATTGAAAGTG
 ZT2.1 TAGGATGTTT TATTTATTA TAATTATTTA TTTGTAAGTA AATAT-GTAA GAAATGAAGT AAATTATATG TAAGATGTTG ATTGAAAGTG
 ZT2.2 TAGGATGTTT TATTTATTA TAATTATTTA TTTGTAAGTA AATAT-GTAA GAAATGAAGT AAATTATATG TAAGATGTTG ATTGAAAGTG
 ZT2.3 TAGGATGTTT TATTTATTA TAATTATTTA TTTGTAAGTA AATAT-GTAA GAAATGAAGT AAATTATATG TAAGATGTTG ATTGAAAGTG
 ZT2.4 TAGGATGTTT TATTTATTA TAATTATTTA TTTGTAAGTA AATAT-GTAA GAAATGAAGT AAATTATATG TAAGATGTTG ATTGAAAGTG
 ZT2.5 TAGGATGTTT TATTTATTA TAATTATTTA TTTGTAAGTA AATAT-GTAA GAAATGAAGT AAATTATATG TAAGATGTTG ATTGAAAGTG

271
 Reference ACTAGGAACA AATTTTTATA AAACAA
 ZT1.1 ATTAGGAATA AATTTTTATA AAATAA
 ZT1.2 ATTAGGAATA AATTTTTATA AAATAA
 ZT1.3 ATTAGGAATA AATTTTTATA AAATAA
 ZT1.4 ATTAGGAATA AATTTTTATA AAATAA
 ZT1.5 ATTAGGAATA AATTTTTATA AAATAA
 ZT1.6 ATTAGGAATA AATTTTTATA AAATAA
 ZT2.1 ATTAGGAATA AATTTTTATA AAATAA
 ZT2.2 ATTAGGAATA AATTTTTATA AAATAA
 ZT2.3 ATTAGGAATA AATTTTTATA AAATAA
 ZT2.4 ATTAGGAATA AATTTTTATA AAATAA
 ZT2.5 ATTAGGAATA AATTTTTATA AAATAA

reference	AGGAACAAAT	TTTTATAAAA	CAA
S01.1	AGGAATAAGT	TTTTATAAAA	TAA
S01.2	AGGAATAAAT	TTTTATAAAA	TAA
S01.3	AGGAATAAAT	TTTTATAAAA	TAA
S01.4	AGGAATAAAT	TTTTATAAAA	TAA
S01.5	AGGAATAAAT	TTTTATAAAA	TAA
S01.6	AGGAATAAAT	TTTTATAAAA	TAA
S01.7	AGGAATAAAT	TTTTATAAAA	TAA
S01.8	AGGAATAAAT	TTTTATAAAA	TAA
S01.9	AGGAATAAAT	TTTTATAAAA	TAA
S01.10	AGGAATAAAT	TTTTATAAAA	TAA
S01.11	AGGAATAAAT	TTTTATAAAA	TAA
S01.12	AGGAATAAAT	TTTTATAAAA	TAA
S01.13	AGGAATAAAT	TTTTATAAAA	TAA
S01.14	AGGAATAAAT	TTTTATAAAA	TAA
S01.15	AGGAATAAAT	TTTTATAAAA	TAA
S02.1	AGGAATAAAT	TTTTATAAAA	TAA
S02.2	AGGAATAAAT	TTTTATAAAA	TAA
S02.3	AGGAATAAAT	TTTTATAAAA	TAA
S02.4	AGGAATAAAT	TTTTATAAAA	TAA
S02.5	AGGAATAAAT	TTTTATAAAA	TAA
S02.6	AGGAATAAAT	TTTTATAAAA	TAA
S02.7	AGGAATAAAT	TTTTATAAAA	TAA
S02.8	AGGAATAAAT	TTTTATAAAA	TAA
S02.9	AGGAATAAAT	TTTTATAAAA	TAA
S02.10	AGGAATAAAT	TTTTATAAAA	TAA
S02.11	AGGAATAAAT	TTTTATAAAA	TAA

Alignment: /Users/benjaminhubert/Desktop/thèse/data/DNA_methylation/STRAINS/SENEGAL/TESTES/TIR52/SENEGAL.txt
Seaview [blocks=10 fontsize=8 A4] on Tue Oct 12 17:04:32 2010

1
reference CCACCTCTCG ACAATAAACT TCACACTGAT TTTGAACTTC AAGAAAAGTCA GTCGTATTCT TTATTGGAAA TCTTCACACT ACAACTATCT
ST1.1 TTATTTTTG ATAATAAAT TTAGATTGAT TTTGAATTT AAGAAAAGTCA GTTGTATTTT TTATTGGAAA TTTTATATAT ATAATTATTT
ST1.2 TTATTTTTG ATAATAAAT TTAGATTGAT TTTGAATTT AAGAAAAGTCA GTTGTATTTT TTATTGGAAA TTTTATATAT ATAATTATTT
ST1.3 TTATTTTTG ATAATAAAT TTAGATTGAT TTTGAATTT AAGAAAAGTCA GTTGTATTTT TTATTGGAAA TTTTATATAT ATAATTATTT
ST1.4 TTATTTTTG ATAATAAAT ATATATTGAT TTTGAATTT AAGAAAAGTCA GTTGTATTTT TTATTGGAAA TTTTATATAT ATAATTATTT
ST1.5 TTATTTTTG ATAATAAAT TTATATTGAT TTTGAATTT AAGAAAAGTCA GTTGTATTTT TTATTGGAAA TTTTATATAT ATAATTATTT
ST2.1 TTATTTTTG ATAATAAAT TTATATTGAT TTTGAATTT AAGAAAAGTCA GTTGTATTTT TTATTGGAAA TTTTATATAT ATAATTATTT
ST2.2 TTATTTTTG ATAATAAAT TTATATTGAT TTTGAATTT AAGAAAAGTCA GTTGTATTTT TTATTGGAAA TTTTATATAT ATAATTATTT
ST2.3 TTATTTTTG ATAATAAAT TTATATTGAT TTTGAATTT AAGAAAAGTCA GTTGTATTTT TTATTGGAAA TTTTATATAT ATAATTATTT
ST2.4 TTATTTTTG ATAATAAAT TTATATTGAT TTTGAATTT AAGAAAAGTCA GTTGTATTTT TTATTGGAAA TTTTATATAT ATAATTATTT
ST2.5 TTATTTTTG ATAATAAAT TTATATTGAT TTTGAATTT AAGAAAAGTCA GTTGTATTTT TTATTGGAAA TTTTATATAT ATAATTATTT

91
reference GCTGAAACTT AAAAACCTTC ATACATTTAC ACATCATATC TTCACAAAAG GCTCCACCCT CGATCACGGA CTTAACTGGC GCAGCCGGTA
ST1.1 GTTGAATTT AAAAATTTT ATATATTAT ATATTATAT TTTATAAAG GTTTTATTTT TGATTATGGA TTTAATTTGGT GTAGTTGGTA
ST1.2 GTTGAATTT AAAAATTTT ATATATTAT ATATTATAT TTTATAAAG GTTTTATTTT TGATTATGGA TTTAATTTGGT GTAGTTGGTA
ST1.3 GTTGAATTT AAAAATTTT ATATATTAT ATATTATAT TTTATAAAG GTTTTATTTT TGATTATGGA TTTAATTTGGT GTAGTTGGTA
ST1.4 GTTGAATTT AAAAATTTT ATATATTAT ATATTATAT TTTATAAAG GTTTTATTTT TGATTATGGA TTTAATTTGGT GTAGTTGGTA
ST1.5 GTTGAATTT AAAAATTTT ATATATTAT ATATTATAT TTTATAAAG GTTTTATTTT TGATTATGGA TTTAATTTGGT GTAGTTGGTA
ST2.1 GTTGAATTT AAAAATTTT ATATATTAT ATATTATAT TTTATAAAG GTTTTATTTT TGATTATGGA TTTAATTTGGT GTAGTTGGTA
ST2.2 GTTGAATTT AAAAATTTT ATATATTAT ATATTATAT TTTATAAAG GTTTTATTTT TGATTATGGA TTTAATTTGGT GTAGTTGGTA
ST2.3 GTTGAATTT AAAAATTTT ATATATTAT ATATTATAT TTTATAAAG GTTTTATTTT TGATTATGGA TTTAATTTGGT GTAGTTGGTA
ST2.4 GTTGAATTT AAAAATTTT ATATATTAT ATATTATAT TTTATAAAG GTTTTATTTT TGATTATGGA TTTAATTTGGT GTAGTTGGTA
ST2.5 GTTGAATTT AAAAATTTT ATATATTAT ATATTATAT TT-ATAAAG GTTTTATTTT TGATTATGGA TTTAATTTGGT GTAGTTGGTA

181
reference GGATGCTCTA CCTATTAATA ATTACCTACC TGTAAGTAAA CATGTAAGAA ACGAAAACAA CTATATGCAA GATGTCGACT GAAAGTGACT
ST1.1 GGATGTTTTA TTTATTAATA ATTATTTATT TGTAAGTAAA TATGTAAGAA ATGAAATAAA TTATATGTAA GATGTTGATT GAAAGTGATT
ST1.2 GGATGTTTTA TTTATTAATA ATTATTTATT TGTAAGTAAA TATGTAAGAA ATGAAATAAA TTATATGTAA GATGTTGATT GAAAGTGATT
ST1.3 GGATGTTTTA TTTATTAATA ATTATTTATT TGTAAGTAAA TATGTAAGAA ATGAAATAAA TTATATGTAA GATGTTGATT GAAAGTGATT
ST1.4 GGATGTTTTA TTTATTAATA ATTATTTATT TGTAAGTAAA TATGTAAGAA ATGAAATAAA TTATATGTAA GATGTTGATT GAAAGTGATT
ST1.5 GGATGTTTTA TTTATTAATA ATTATTTATT TGTAAGTAAA TATGTAAGAA ATGAAATAAA TTATATGTAA GATGTTGATT GAAAGTGATT
ST2.1 GGATGTTTTA TTTATTAATA ATTATTTATT TGTAAGTAAA TATGTAAGAA ATGAAATAAA TTATATGTAA GATGTTGATT GAAAGTGATT
ST2.2 GGATGTTTTA TTTATTAATA ATTATTTATT TGTAAGTAAA TATGTAAGAA ATGAAATAAA TTATATGTAA GATGTTGATT GAAAGTGATT
ST2.3 GGATGTTTTA TTTATTAATA ATTATTTATT TGTAAGTAAA TATGTAAGAA ATGAAATAAA TTATATGTAA GATGTTGATT GAAAGTGATT
ST2.4 GGATGTTTTA TTTATTAATA ATTATTTATT TGTAAGTAAA TATGTAAGAA ATGAAATAAA TTATATGTAA GATGTTGATT GAAAGTGATT
ST2.5 GGATGTTTTA TTTATTAATA ATTATTTATT TGTAAGTAAA TATGTAAGAA ATGAAATAAA TTATATGTAA GATGTTGATT GAAAGTGATT

271
reference AGGAACAAAT TTTTATAAAA CAA
ST1.1 AGGAATAAAT TTTTATAAAA TAA
ST1.2 AGGAATAAAT TTTTATAAAA TAA
ST1.3 AGGAATAAAT TTTTATAAAA TAA
ST1.4 AGGAATAAAT TTTTATAAAA TAA
ST1.5 AGGAATAAAT TTTTATAAAA TAA
ST2.1 AGGAATAAAT TTTTATAAAA TAA
ST2.2 AGGAATAAAT TTTTATAAAA TAA
ST2.3 AGGAATAAAT TTTTATAAAA TAA
ST2.4 AGGAATAAAT TTTTATAAAA TAA
ST2.5 AGGAATAAAT TTTTATAAAA TAA

Alignment: /Users/benjaminhubert/Desktop/thèse/data/DNA_methylation/ACTINE/MAKINDU.txt
Seaview [blocks=10 fontsize=8 A4] on Tue Oct 12 16:09:39 2010

```
1
Act5C-PB  CTCCCCCGCG CGTCCTCCCA TCGATTGTGG GACGTCGCCG TCACCAGGGT GTGATGGTGG GCATGGGCCA GAAGGACTCG TACGTGGGGT
7.1      TTTTFTGTGT TGTTTTTTTA TTGATTGTGG GATGTTTTTG TTATTAGGGT GTGATGGTGG CATGGGPTA GAAGGATTTG TATGTGGGGT
7.2      TTTTFTGTGT TGTTTTTTTA TTGATTGTGG GATGTTTTTG TTATTAGGGT GTGATGGTGG CATGGGPTA GAAGGATTTG TATGTGGGGT
7.5      TTTTFTGTGT TGTTTTTTTA TTGATTGTGG GATGTTTTTG TTATTAGGGT GTGATGGTGG CATGGGPTA GAAGGATTTG TATGTGGGGT
7.6      TTTTFTGTGT TGTTTTTTTA TTGATTGTGG GATGTTTTTG TTATTAGGGT GTGATGGTGG CATGGGPTA GAAGGATTTG TATGTGGGGT
7.7      TTTTFTGTGT TGTTTTTTTA TTGATTGTGG GATGTTTTTG TTATTAGGGT GTGATGGTGG CATGGGPTA GAAGGATTTG TATGTGGGGT
7.8      TTTTFTGTGT TGTTTTTTTA TTGATTGTGG GATGTTTTTG TTATTAGGGT GTGATGGTGG CATGGGPTA GAAGGATTTG TATGTGGGGT
7.9      TTTTFTGTGT TGTTTTTTTA TTGATTGTGG GATGTTTTTG TTATTAGGGT GTGATGGTGG CATGGGPTA GAAGGATTTG TATGTGGGGT
7.11     TTTTFTGTGT TGTTTTTTTA TTGATTGTGG GATGTTTTTG TTATTAGGGT GTGATGGTGG CATGGGPTA GAAGGATTTG TATGTGGGGT
7.14     TTTTFTGTGT TGTTTTTTTA TTGATTGTGG GATGTTTTTG TTATTAGGGT GTGATGGTGG CATGGGPTA GAAGGATTTG TATGTGGGGT
7.15     TTTTFTGTGT TGTTTTTTTA TTGATTGTGG GATGTTTTTG TTATTAGGGT GTGATGGTGG CATGGGPTA GAAGGATTTG TATGTGGGGT
7.16     TTTTFTGTGT TGTTTTTTTA TTGATTGTGG GATGTTTTTG TTATTAGGGT GTGATGGTGG CATGGGPTA GAAGGATTTG TATGTGGGGT
8.3      TTTTFTGTGT TGTTTTTTTA TTGATTGTGG GATGTTTTTG TTATTAGGGT GTGATGGTGG CATGGGPTA GAAGGATTTG TATGTGGGGT
8.5      TTTTFTGTGT TGTTTTTTTA TTGATTGTGG GATGTTTTTG TTATTAGGGT GTGATGGTGG CATGGGPTA GAAGGATTTG TATGTGGGGT
8.6      TTTTFTGTGT TGTTTTTTTA TTGATTGTGG GATGTTTTTG TTATTAGGGT GTGATGGTGG CATGGGPTA GAAGGATTTG TATGTGGGGT
8.9      TTTTFTGTGT CGTTTTTTTA TTGATTGTGG GATGTTTTTG TTATTAGGGT GTGATGGTGG CATGGGPTA GAAGGATTTG TATGTGGGGT
8.12     TTTTFTGTGT TGTTTTTTTA TTGATTGTGG GATGTTTTTG TTATTAGGGT GTGATGGTGG CATGGGPTA GAAGGATTTG TATGTGGGGT
8.13     TTTTFTGTGT TGTTTTTTTA TTGATTGTGG GATGTTTTTG TTATTAGGGT GTGATGGTGG CATGGGPTA GAAGGATTTG TATGTGGGGT
8.15     TTTTFTGTGT TGTTTTTTTA TTGATTGTGG GATGTTTTTG TTATTAGGGT GTGATGGTGG CATGGGPTA GAAGGATTTG TATGTGGGGT
8.16     TTTTFTGTGT TGTTTTTTTA TTGATTGTGG GATGTTTTTG TTATTAGGGT GTGATGGTGG CATGGGPTA GAAGGATTTG TATGTGGGGT
```

```
91
Act5C-PB  ATGAGGCGCA GAGCAAGCGT GGTATCCTCA CCCTGAAGTA CCCCATYAG CACGGTATCG TGACCAACTG
7.1      ATGAGGTGTA GAGTAAGTGT GGTATTTTTA TTTTGAAGTA TTTTATTGAG TATGGTATTC TGATTAATTC
7.2      ATGAGGTGTA GAGTAAGTGT GGTATTTTTA TTTTGAAGTA TTTTATTGAG TATGGTATTC TGATTAATTC
7.5      ATGAGGTGTA GAGTAAGTGT GGTATTTTTA TTTTGAAGTA TTTTATTGAG TATGGTATTC TGATTAATTC
7.6      ATGAGGTGTA GAGTAAGTGT GGTATTTTTA TTTTGAAGTA TTTTATTGAG TATGGTATTC TGATTAATTC
7.7      ATGAGGTGTA GAGTAAGTGT GGTATTTTTA TTTTGAAGTA TTTTATTGAG TATGGTATTC TGATTAATTC
7.8      ATGAGGTGTA GAGTAAGTGT GGTATTTTTA TTTTGAAGTA TTTTATTGAG TATGGTATTC TGATTAATTC
7.9      ATGAGGTGTA GAGTAAGTGT GGTATTTTTA TTTTGAAGTA TTTTATTGAG TATGGTATTC TGATTAATTC
7.11     ATGAGGTGTA GAGTAAGTGT GGTATTTTTA TTTTGAAGTA TTTTATTGAG TATGGTATTC TGATTAATTC
7.14     ATGAGGCGTA GAGTAAGTGT GGTATTTTTA TTTTGAAGTA TTTTATTGAG TATGGTATTC TGATTAATTC
7.15     ATGAGGTGTA GAGTAAGTGT GGTATTTTTA TTTTGAAGTA TTTTATTGAG TATGGTATTC TGATTAATTC
7.16     ATGAGGTGTA GAGTAAGTGT GGTATTTTTA TTTTGAAGTA TTTTATTGAG TATGGTATTC TGATTAATTC
8.3      ATGAGGTGTA GAGTAAGTGT GGTATTTTTA TTTTGAAGTA TTTTATTGAG TATGGTATTC TGATTAATTC
8.5      ATGAGGTGTA GAGTAAGTGT GGTATTTTTA TTTTGAAGTA TTTTATTGAG TATGGTATTC TGATTAATTC
8.6      ATGAGGTGTA GAGTAAGTGT GGTATTTTTA TTTTGAAGTA TTTTATTGAG TATGGTATTC TGATTAATTC
8.9      ATGAGGTGTA GAGTAAGTGT GGTATTTTTA TTTTGAAGTA TTTTATTGAG TATGGTATTC TGATTAATTC
8.12     ATGAGGTGTA GAGTAAGTGT GGTATTTTTA TTTTGAAGTA TTTTATTGAG TATGGTATTC TGATTAATTC
8.13     ATGAGGTGTA GAGTAAGTGT GGTATTTTTA TTTTGAAGTA TTTTATTGAG TATGGTATTC TGATTAATTC
8.15     ATGAGGTGTA GAGTAAGTGT GGTATTTTTA TTTTGAAGTA TTTTATTGAG TATGGTATTC TGATTAATTC
8.16     ATGAGGTGTA GAGTAAGTGT GGTATTTTTA TTTTGAAGTA TTTTATTGAG TATGGTATTC TGATTAATTC
```


Alignment: /Users/benjaminhubert/Desktop/thèse/data/DNA_methylation/PURIFIED_DNA_FROM_MAKINDU/TIR01.txt
Seaview [blocks=10 fontsize=8 A4] on Tue Oct 12 16:31:22 2010

```
1
reference AACACCGCAT CCGGCCGAAT GCTGACATTC CGGCCGAATG CTGACATTGC ACTAAAAGTC GCACTGCAAC ATTGCCCCA GCTAGCCAGC
MO1 ----- -AAT ATTG-TATTT -GGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
MO2 ----- -AAT GTTG-TATTT -GGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
MO3 ----- -AAT ATTG-TATTT -GGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
MO4 ----- -AAT ATTG-TATTT -GGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
MO5 AATATTGTAT TTGGTTGAAT GTTGATATTT TGGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
MO6 ----- -AAT ATTG-TATTT -GGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
MO7 ----- -AAT ATTG-TATTT -GGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
MO8 ----- -AAT ATTG-TATTT -GGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
MO9 ----- -AAT ATTG-TATTT -GGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
MO10 ----- -AAT ATTG-TATTT -GGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
MO11 ----- -AAT ATTG-TATTT -GGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
MO12 ----- -AAT ATCG-TATTT -GGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
MO13 ----- -AAT ATTG-TATTT -GGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
MO14 ----- -AAT ATTG-TATTT -GGTTGAATG TTGATATTGT AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
```

```
91
reference CACATGCCGA GTCGCCATGT TCATTATGTT TACAACCTAAG AACCTATGTA CTTATGTEATA AGATGGAAAC GGAGTATTC AGTAGCCCTCT
MO1 TATATGTFGA GTTGGTATGT TTATTATGTT TATAATTAAG AATTTATGTA TTTATGTEATA AGATGGAAAT GGAGTATTTG AGTAGTTATT
MO2 TATATGTFGA GTTGGTATGT TTATTATGTT TATAATTAAG AATTTATGTA TTTATGTEATA AGATGGAAAT GGAGTATTTG AGTAGTTATT
MO3 TATATGTFGA GTTGGTATGT TTATTATGTT TATAATTAAG AATTTATGTA TTTATGTEATA AGATGGAAAT GGAGTATTTG AGTAGTTATT
MO4 TATATGTFGA GTTGGTATGT TTATTATGTT TATAATTAAG AATTTATGTA TTTATGTEATA AGATGGAAAT GGAGTATTTG AGTAGTTATT
MO5 TATATGTFGA GTTGGTATGT TTATTATGTT TATAATTAAG AATTTATGTA TTTATGTEATA AGATGGAAAT GGAGTATTTG AGTAGTTATT
MO6 TATATGTFGA GTTGGTATGT TTATTATGTT TATAATTAAG AATTTATGTA TTTATGTEATA AGATGGAAAT GGAGTATTTG AGTAGTTATT
MO7 TATATGTFGA GTTGGTATGT TTATTATGTT TATAATTAAG AATTTATGTA TTTATGTEATA AGATGGAAAT GGAGTATTTG AGTAGTTATT
MO8 TATATGTFGA GTTGGTATGT TTATTATGTT TATAATTAAG AATTTATGTA TTTATGTEATA AGATGGAAAT GGAGTATTTG AGTAGTTATT
MO9 TATATGTFGA GTTGGTATGT TTATTATGTT TATAATTAAG AATTTATGTA TTTATGTEATA AGATGGAAAT GGAGTATTTG AGTAGTTATT
MO10 TATATGTFGA GTTGGTATGT TTATTATGTT TATAATTAAG AATTTATGTA TTTATGTEATA AGATGGAAAT GGAGTATTTG AGTAGTTATT
MO11 TATATGTFGA GTTGGTATGT TTATTATGTT TATAATTAAG AATTTATGTA TTTATGTEATA AGATGGAAAT GGAGTATTTG AGTAGTTATT
MO12 TATATGTFGA GTTGGTATGT TTATTATGTT TATAATTAAG AATTTATGTA TTTATGTEATA AGATGGAAAT GGAGTATTTG AGTAGTTATT
MO13 TATATGTFGA GTTGGTATGT TTATTATGTT TATAATTAAG AATTTATGTA TTTATGTEATA AGATGGAAAT GGAGTATTTG AGTAGTTATT
MO14 TATATGTFGA GTTGGTATGT TTATTATGTT TATAATTAAG AATTTATGTA TTTATGTEATA AGATGGAAAT GGAGTATTTG AGTAGTTATT
```

```
181
reference GGCTGACAAT AAACCTCCTC CTGATTTTGA ACTTCAAGAA AGTCAGTCTGT ATTCCTTATT TGAAACTCTT ACACCTACAA TATCTGGGGA
MO1 TTTTGATAAT AAATTTTATT TTGATTTTGA ATTTTAAGAA AGTTAGTCTGT ATTTTCTATT GGAATTTTTC ATATTATAAA TATTTGTTGA
MO2 TTTTGATAAT AAATTTTATT TTGATTTTGA ATTTTAAGAA AGTTAGTCTGT ATTTTCTATT GGAATTTTTC ATATTATAAA TATTTGTTGA
MO3 TTTTGATAAT AAATTTTATT TTGATTTTGA ATTTTAAGAA AGTTAGTCTGT ATTTTCTATT GGAATTTTTC ATATTATAAA TATTTGTTGA
MO4 TTTTGATAAT AAATTTTATT TTGATTTTGA ATTTTAAGAA AGTTAGTCTGT ATTTTCTATT GGAATTTTTC ATATTATAAA TATTTGTTGA
MO5 TTTTGATAAT AAATTTTATT TTGATTTTGA ATTTTAAGAA AGTTAGTCTGT ATTTTCTATT GGAATTTTTC ATATTATAAA TATTTGTTGA
MO6 TTTTGATAAT AAATTTTATT TTGATTTTGA ATTTTAAGAA AGTTAGTCTGT ATTTTCTATT GGAATTTTTC ATATTATAAA TATTTGTTGA
MO7 TTTTGATAAT AAATTTTATT TTGATTTTGA ATTTTAAGAA AGTTAGTCTGT ATTTTCTATT GGAATTTTTC ATATTATAAA TATTTGTTGA
MO8 TTTTGATAAT AAATTTTATT TTGATTTTGA ATTTTAAGAA AGTTAGTCTGT ATTTTCTATT GGAATTTTTC ATATTATAAA TATTTGTTGA
MO9 TTTTGATAAT AAATTTTATT TTGATTTTGA ATTTTAAGAA AGTTAGTCTGT ATTTTCTATT GGAATTTTTC ATATTATAAA TATTTGTTGA
MO10 TTTTGATAAT AAATTTTATT TTGATTTTGA ATTTTAAGAA AGTTAGTCTGT ATTTTCTATT GGAATTTTTC ATATTATAAA TATTTGTTGA
MO11 TTTTGATAAT AAATTTTATT TTGATTTTGA ATTTTAAGAA AGTTAGTCTGT ATTTTCTATT GGAATTTTTC ATATTATAAA TATTTGTTGA
MO12 TTTTGATAAT AAATTTTATT TTGATTTTGA ATTTTAAGAA AGTTAGTCTGT ATTTTCTATT GGAATTTTTC ATATTATAAA TATTTGTTGA
MO13 TTTTGATAAT AAATTTTATT TTGATTTTGA ATTTTAAGAA AGTTAGTCTGT ATTTTCTATT GGAATTTTTC ATATTATAAA TATTTGTTGA
MO14 TTTTGATAAT AAATTTTATT TTGATTTTGA ATTTTAAGAA AGTTAGTCTGT ATTTTCTATT GGAATTTTTC ATATTATAAA TATTTGTTGA
```

```
271
reference AACCTAAAAC CCTTCATACG TATACACATC TTATCTTAC ATACGGCTCC ACCCTCGAGT GCACGGACTT AACTGGCGCA GCCGGTAGGA
MO1 AATTTAGAAT TTTT----- -ATATAT ATATTTTAT AAAAGGTTTT ATTTTGTATT --ATGGATT AATTGGTCTA GTTGGTAGGA
MO2 AATTTAGAAT TTTT----- -ATATAT ATATTTTAT AAAAGGTTTT ATTTTGTATT --ATGGATT AATTGGTCTA GTTGGTAGGA
MO3 AATTTAGAAT TTTT----- -ATATAT ATATTTTAT AAAAGGTTTT ATTTTGTATT --ATGGATT AATTGGTCTA GTTGGTAGGA
MO4 AATTTAGAAT TTTT----- -ATATAT ATATTTTAT AAAAGGTTTT ATTTTGTATT --ATGGATT AATTGGTCTA GTTGGTAGGA
MO5 AATTTAGAAT TTTT----- -ATATAT ATATTTTAT AAAAGGTTTT ATTTTGTATT --ATGGATT AATTGGTCTA GTTGGTAGGA
MO6 AATTTAGAAT TTTT----- -ATATAT ATATTTTAT AAAAGGTTTT ATTTTGTATT --ATGGATT AATTGGTCTA GTTGGTAGGA
MO7 AATTTAGAAT TTTT----- -ATATAT ATATTTTAT AAAAGGTTTT ATTTTGTATT --ATGGATT AATTGGTCTA GTTGGTAGGA
MO8 AATTTAGAAT TTTT----- -ATATAT ATATTTTAT AAAAGGTTTT ATTTTGTATT --ATGGATT AATTGGTCTA GTTGGTAGGA
MO9 AATTTAGAAT TTTT----- -ATATAT ATATTTTAT AAAAGGTTTT ATTTTGTATT --ATGGATT AATTGGTCTA GTTGGTAGGA
MO10 AATTTAGAAT TTTT----- -ATATAT ATATTTTAT AAAAGGTTTT ATTTTGTATT --ATGGATT AATTGGTCTA GTTGGTAGGA
MO11 AATTTAGAAT TTTT----- -ATATAT ATATTTTAT AAAAGGTTTT ATTTTGTATT --ATGGATT AATTGGTCTA GTTGGTAGGA
MO12 AATTTAAAAA TTTTATATA TTTATATAT ATATTTTAT AAAAGGTTTT ATTTTGTATT --ATGGATT AATTGGTCTA GTTGGTAGGA
MO13 AATTTAAAAA TTTTATATA TTTATATAT ATATTTTAT AAAAGGTTTT ATTTTGTATT --ATGGATT AATTGGTCTA GTTGGTAGGA
MO14 AATTTAAAAA TTTTATATA TTTATATAT ATATTTTAT AAAAGGTTTT ATTTTGTATT --ATGGATT AATTGGTCTA GTTGGTAGGA
```

```
361
reference TGTCCCTACCT TTTAATAATC ACCTACCTGT AAGTAACCAT TGTAAAGAAC GAAGCAAAT ATAT
MO1 TGTPTTATTT ATTAATAATT ATTTATPTGT AAGTAATAT -GTAAGAAAT GAAGTAAAT ATAT
MO2 TGTPTTATTT ATTAATAATT ATTTATPTGT AAGTAATAT -GTAAGAAAT GAAGTAAAT ATAT
MO3 TGTPTTATTT ATTAATAATT ATTTATPTGT AAGTAATAT -GTAAGAAAT GAAGTAAAT ATAT
MO4 TGTPTTATTT ATTAATAATT ATTTATPTGT AAGTAATAT -GTAAGAAAT GAAGTAAAT ATAT
MO5 TGTPTTATTT ATTAATAATT ATTTATPTGT AAGTAATAT -GTAAGAAAT GAAGTAAAT ATAT
MO6 TGTPTTATTT ATTAATAATT ATTTATPTGT AAGTAATAT -GTAAGAAAT GAAGTAAAT ATAT
MO7 TGTPTTATTT ATTAATAATT ATTTATPTGT AAGTAATAT -GTAAGAAAT GAAGTAAAT ATAT
MO8 TGTPTTATTT ATTAATAATT ATTTATPTGT AAGTAATAT -GTAAGAAAT GAAGTAAAT ATAT
MO9 TGTPTTATTT ATTAATAATT ATTTATPTGT AAGTAATAT -GTAAGAAAT GAAGTAAAT ATAT
MO10 TGTPTTATTT ATTAATAATT ATTTATPTGT AAGTAATAT -GTAAGAAAT GAAGTAAAT ATAT
MO11 TGTPTTATTT ATTAATAATT ATTTATPTGT AAGTAATAT -GTAAGAAAT GAAGTAAAT ATAT
MO12 TGTPTTATTT ATTAATAATT ATTTATPTGT AAGTAATAT -GTAAGAAAT GAAGTAAAT ATAT
MO13 TGTPTTATTT ATTAATAATT ATTTATPTGT AAGTAATAT -GTAAGAAAT GAAGTAAAT ATAT
MO14 TGTPTTATTT ATTAATAATT ATTTATPTGT AAGTAATAT -GTAAGAAAT GAAGTAAAT ATAT
```

Alignment: /Users/benjaminhubert/Desktop/thèse/data/DNA methylation/PURIFIED_DNA_FROM_MAKINDU/TIR52.txt
Seaview [blocks=10 fontsize=8 A4] on Tue Oct 12 16:31:46 2010

```
1
tirant_complet  CCACTCTCTG ACAATAAACT TGATACTGAT TTTGAAC TTC AAGAAAGTCA GTGCTATTCT TTATTGGAAA TCTTCACACT ACAACTATCT
MO1             TTATTTTTTG ATAATAAAAT TTATTTTGAT TTTGAATTTT AAGAAAAGTTA GTTGTATTTT TTATTGGAAA TTTTATATAT ATAAATATTT
MO2             TTATTTTTTG ATAATAAAAT TTATTTTGAT TTTGAATTTT AAGAAAAGTTA GTTGTATTTT TTATTGGAAA TTTTATATAT ATAAATATTT
MO3             TTATTTTTTG ATAATAAAAT TTATTTTGAT TTTGAATTTT AAGAAAAGTTA GTTGTATTTT TTATTGGAAA TTTTATATAT ATAAATATTT
MO4             -TTTTGGTTG ATAATAAAAT TTATTTTGAT TTTGAATTTT AAGAAAAGTTA GTTGTATTTT TTATTGGAAA TTTTATATAT ATAAATATTT
MO5             -TTTTGGTTG ATAATAAAAT TTATTTTGAT TTTGAATTTT AAGAAAAGTTA GTTGTATTTT TTATTGGAAA TTTTATATAT ATAAATATTT
MO6             ---TGGTTG ATAATAAAAT TTATTTTGAT TTTGAATTTT AAGAAAAGTTA GTTGTATTTT TTATTGGAAA TTTTATATAT ATAAATATTT
MO7             -TTTTGGTTG ATAATAAAAT TTATTTTGAT TTTGAATTTT AAGAAAAGTTA GTTGTATTTT TTATTGGAAA TTTTATATAT ATAAATATTT
MO8             -TTTTGGTTG ATAATAAAAT TTATTTTGAT TTTGAATTTT AAGAAAAGTTA GTTGTATTTT TTATTGGAAA TTTTATATAT ATAAATATTT
MO9             -TTTTGGTTG ATAATAAAAT TTATTTTGAT TTTGAATTTT AAGAAAAGTTA GTTGTATTTT TTATTGGAAA TCTTATATAT ATAAATATTT
MO10            -TTTTGGTTG ATAATAAAAT TTATTTTGAT TTTGAATTTT AAGAAAAGTTA GTTGTATTTT TTATTGGAAA TTTTATATAT ATAAATATTT
MO11            TTATTTTTTG ATAATAAAAT TTATTTTGAT TTTGAATTTT AAGAAAAGTCA GTTGTATTTT TTATTGGAAA TTTTATATAT ATAAATATTT

91
tirant_complet  GCTGAAACTT AAAAACCTTC ATACATTTAC ACATCATATC TTCACAAAAG GCTCCACCCT CGATCACGGA CTTAACTGGC GCAGCCGGTA
MO1             GTTGAAA--- -AGAAATTTG TGGTGGATAT ATATTATATT TTTATAAAAAG GTTTTATTTT TGATTATGGA TTTAACTGGT GTAGTTGGTA
MO2             GTTGAAA--- -AGAAATTTG TGGTGGTAT ATATTATATT TTTATAAAAAG GTTTTATTTT TGATTATGGA TTTAACTGGT GTAGTTGGTA
MO3             GTTGAAA--- -AGAAATTTG --GTGGTAT ATATTATATT TTTATAAAAAG GTTTTATTTT TGATTATGGA TTTAACTGGT GTAGTTGGTA
MO4             GGGGAAATTT -AAAAATTTT ATATGTATAT ATATTATATT TTTATATATG GTTTTATTTT TGATTATGGA TTTAACTGGT GTAGTTGGTA
MO5             GGGGAAATTT -AAAAATTTT ATATGTATAT ATATTATATT TTTATATATG GTTTTATTTT TGATTATGGA TTTAACTGGT GTAGTTGGTA
MO6             GGGGAAATTT -AAAAATTTT ATATGTATAT ATATTATATT TTTATATATG GTTTTATTTT TGATTATGGA TTTAACTGGT GTAGTTGGTA
MO7             GGGGAAATTT -AAAAATTTT ATATGTATAT ATATTATATT TTTATATATG GTTTTATTTT TGATTATGGA TTTAACTGGT GTAGTTGGTA
MO8             GGGGAAA--- -AAAAATTTT ATATGTATAT ATATTATATT TTTATAAATG GTTTTATTTT TGATTATGGA TTTAACTGGT GTAGTTGGTA
MO9             GGGGAAATTT -AAAAATTTT ATATGTATAT ATATTATATT TTTATATATG GTTTTATTTT TGATTATGGA TTTAACTGGT GTAGTTGGTA
MO10            GGGGAAA--- -AAAAATTTT ATATGTATAT ATATTATATT TTTATAAAAAG GTTTTATTTT TGATTATGGA TTTAACTGGT GTAGTTGGTA
MO11            GTTGAAA--- -AAAAATTTT ATATAATTTAT ATATTATATT TTTATAAAAAG GTTTTATTTT TGATTATGGA TTTAACTGGT GTAGTTGGTA

181
tirant_complet  GGATGTCCTA CCTATTAATA ATTACCTACC TGTAAAGTAAA CATGTAAGAA ACGAAACAAA CTATATGCAA GATGTCGACT GAAAGTGACT
MO1             GGATGTTTTA TTTATTAATA ATTATTTTAT TGTAAAGTAAA TATGTAAGAA ATGAAGTAAA TTATATGTAA GATGTTGATT GAAAGTGATT
MO2             GGATGTTTTA TTTATTAATA ATTATTTTAT TGTAAAGTAAA TATGTAAGAA ATGAAGTAAA TTATATGTAA GATGTTGATT GAAAGTGATT
MO3             GGATGTTTTA TTTATTAATA ATTATTTTAT TGTAAAGTAAA TATGTAAGAA ATGAAGTAAA TTATATGTAA GATGTTGATT GAAAGTGATT
MO4             AGATGTTTTA TTTTTAATA ATTATTTTAT TGTAAAGTAAAT TATGTAAGAA ATGAAGTAAA TTATATGTAA GATGTTGATT GAAAGTGAAT
MO5             GGATGTTTTA TTTTTAATA ATTATTTTAT TGTAAAGTAAAT TATGTAAGAA ATGAAGTAAA TTATATGTAA GATGTTGATT GAAAGTGAAT
MO6             GGATGTTTTA TTTTTAATA ATTATTTTAT TGTAAAGTAAAT TATGTAAGAA ATGAAGTAAA TTATATGTAA GATGTTGATT GAAAGTGAAT
MO7             GGATGTTTTA TTTTTAATA ATTATTTTAT TGTAAAGTAAAT TATGTAAGAA ATGAAGTAAA TTATATGTAA GATGTTGATT GAAAGTGAAT
MO8             GGATGTTTTA TTTTTAATA ATTATTTTAT TGTAAAGTAAAT TATGTAAGAA ATGAAGTAAA TTATATGTAA GATGTTGATT GAAAGTGAAT
MO9             GGATGTTTTA TTTTTAATA ATTATTTTAT TGTAAAGTAAAT TATGTAAGAA ATGAAGTAAA TTATATGTAA GATGTTGATT GAAAGTGAAT
MO10            GGATGTTTTA TTTTTAATA ATTATTTTAT TGTAAAGTAAAT TATGTAAGAA ATGAAGTAAA TTATATGTAA GATGTTGATT GAAAGTGAAT
MO11            GGATGTTTTA TTTATTAATA ATTATTTTAT TGTAAAGTAAA TATGTAAGAA ATGAAGTAAA TTATATGTAA GATGTTGATT GAAAGTGAAT

271
tirant_complet  AGGAACAAAT TTTTATAAAA CAA
MO1             AGGAATAAAT TTTTATAAAA TAA
MO2             AGGAATAAAT TTTTATAAAA TAA
MO3             AGGAATAAAT TTTTATAAAA TAA
MO4             AAGAATAAAT TTTTAAAAAA TAA
MO5             AAGAATAAAT TTTTAAAAAA TAA
MO6             AAGAATAAAT TTTTAAAAAA TAA
MO7             AAGAATAAAT TTTTAAAAAA TAA
MO8             AAGAATAAAT TTTTAAAAAA TAA
MO9             AAGAATAAAT TTTTAAAAAA TAA
MO10            AAGAATAAAT TTTTAAAAAA TAA
MO11            AGGAATAAAT TTTTATAAAA TAA
```


361

reference	CCTACCTATT	AATAATTACC	TACCTGTAAG	TAAACNTGTA	AGAAACGAAA	CAAACCTATAT
1.1	TTTATTTTATT	AATAATTATT	TATTTGTAAG	TAAATATGTC	AGAAATGAAA	TAAATTATAT
1.2	TTTATTTTATT	AATAATTATT	TATTTGTAAG	TAAATATGTA	AGAAATGAAA	TAAATTATAT
1.3	TTTATTTTATT	AATAATTATT	TATTTGTAAG	TAAATATGTA	AGAAATGAAA	TAAATTATAT
1.4	TTTATTTTATT	AATAATTATT	TATTTGTAAG	TAAATATGTA	AGAAATGAAA	TAAATTATAT
1.5	TTTACTTATT	AATAATTATT	TATTTGTAAG	TAAATATGTA	AGAAATGAAA	TAAATTATAT
1.6	TTTATTTTATT	AATAATTATT	TATTTGTAAG	TAAATATGTA	AGAAATGAAA	TAAATTATAT
1.7	TTTATTTTATT	AATAATTATT	TATTTGTAAG	TAAATATGTA	AGAAATGAAA	TAAATTATAT
1.8	TTTATTTTATT	AATAATTATT	TATTTGTAAG	TAAATATGTA	AGAAATGAAA	TAAATTATAT
1.9	TTTATTTTATT	AATAATTATT	TATTTGTAAG	TAAATATGTC	AGAAATGAAA	TAAATTATAT
2.1	TTTATTTTATT	AATAATTATT	TATTTGTAAG	TAAATATGTA	AGAAATGAAA	TAAATTATAT
2.2	TTTATTTTATT	AATAATTATT	TATTTGTAAG	TAAATATGTA	AGAAATGAAA	TAAATTATAT
2.3	TTTATTTTATT	AATAATTATT	TATTTGTAAG	TAAATATGTA	AGAAATGAAA	TAAATTATAT
2.4	TTTATTTTATT	AATAATTATT	TATTTGTAAG	TAAATATGTA	AGAAATGAAA	TAAATTATAT
2.5	TTTATTTTATT	AATAATTATT	TATTTGTAAG	TAAATATGTA	AGAAATGAAA	TAAATTATAT
2.7	TTTATTTTATT	AATAATTATT	TATTTGTAAG	TAAATATGTA	AGAAACGAAA	TAAATTATAT
2.8	TTTATTTTATT	AATAATTATT	TATTTGTAAG	TAAATATGTA	AGAAATGAAA	TAAATTATAT
2.10	TTTATTTTATT	AATAATTACT	TATTTGTAAG	TAAATATGTA	AGAAATGAAA	TAAATTATAT
2.11	TTTATTTTATT	AATAATTATT	TATTTGTAAG	TAAATATGTA	AGAAATGAAA	TAAATTATAT

Alignment: /Users/benjaminhubert/Desktop/thèse/data/DNA_methylation/MAKINDU_SITES/M2.txt
Seaview [blocks=10 fontsize=8 A4] on Tue Oct 12 16:28:44 2010

```
1
reference_hdl27  GATCAGCTGT CAATTTTATG GGGCCCCCTT TCCCAACTC CTTTTCTG GGGGTGTGA CGCACATATT TTAGCCCTT
1  GATTAGTGT TAATTTT-IT GGGGTTTTTT TTTTAAATTT TTTTTTTTG GGGGTGTGA TGTATATATT TTAGTTTTT
2  GATTAGTGT TAATTTTATG GG---TCCTT TTTTAAATTT TTTTTTTTG GGGGTGTGA TGTATATATT TTAGTTTTT
3  GATTAGTGT TAATTTTATG GGGTTTTTTT TTTTAAATTT TTTTTTTTG GGGGTGTGA TGTATATATT TTAGTTTTT
4  GATTAGTGT TAATTTTATG GGGTTTTTTT TTTT-AAATTT TTTTTTTTG GGGGTGTGA TGTATATATT TTAGTTTTT
5  GATTAGTGT TAATTTTATG GGGTTTTTTT TTTTAAATTT TTTTTTTTG GGGGTGTGA TGTATATATT TTAGTTTTT
6  GATTAGTGT TAATTTTATG GGGTTTTTTT TTTTAAATTT TTTTTTTTG GGGGTGTGA TGTATATATT TTAGTTTTT
7  GATTAGTGT TAATTTTATT GGGTTTTTTT TTTTAAATTT TTTTTTTTG GGGGTGTGA TGTATATATT TTAGTTTTT
8  GATTAGTGT TAATTTTATG GGGTTTTTTT TTTT-AAATTT TTTTTTTTG GGGGTGTGA TGTATATATT TTAGTTTTT
9  GATTAGTGT TAATTTTATT GGGTTTTTTT TTTTAAATTT TTTTTTTTG GGGGTGTGA TGTATATATT TTAGTTTTT
10 GATTAGTGT TAATTTTATG GG---TCCTT TTTTAAATTT TTTTTTTTG GGGGTGTGA TGTATATATT TTAGTTTTT
11 GATTAGTGT TAATTTTATG GGGTTTTTTT TTTT-AAATTT TTTTTTTTG GGGGTGTGA TGTATATATT TTAGTTTTT
12 GATTAGTGT TAATTTTATG GGGTTTTTTT TTTTAAATTT TTTTTTTTG GGGGTGTGA TGTATATATT TTAGTTTTT
```

```
81
reference_hdl27  TTGNGTGTGT GCCCAACTTT GTTGTCAATTT CATTGCTTA GATTATTTTG ATATGCATCT GAGCACTTGA TTTGGGCTTA
1  TTGNGTGTGT GTTTAAATTTT GTTGTCAATTT TAATTTGTTA GATTATTTTG ATATGCATCT GAGTATTGA TTTGGGCTTA
2  TTGNGTGTGT GTTTAAATTTT GTTGTCAATTT TAATTTGTTA GATTATTTTG ATATGCATCT GAGTATTGA TTTGGGCTTA
3  TTGNGTGTGT GTTTAAATTTT GTTGTCAATTT TAATTTGTTA GATTATTTTG ATATGCATCT GAGTATTGA TTTGGGCTTA
4  TTGNGTGTGT GTTTAAATTTT GTTGTCAATTT TAATTTGTTA GATTATTTTG ATATGCATCT GAGTATTGA TTTGGGCTTA
5  TTGNGTGTGT GTTTAAATTTT GTTGTCAATTT TAATTTGTTA GATTATTTTG ATATGCATCT GAGTATTGA TTTGGGCTTA
6  TTGNGTGTGT GTTTAAATTTT GTTGTCAATTT TAATTTGTTA GATTATTTTG ATATGCATCT GAGTATTGA TTTGGGCTTA
7  TTGNGTGTGT GTTTAAATTTT GTTGTCAATTT TAATTTGTTA GATTATTTTG ATATGCATCT GAGTATTGA TTTGGGCTTA
8  TTGNGTGTGT GTTTAAATTTT GTTGTCAATTT TAATTTGTTA GATTATTTTG ATATGCATCT GAGTATTGA TTTGGGCTTA
9  TTGNGTGTGT GTTTAAATTTT GTTGTCAATTT TAATTTGTTA GATTATTTTG ATATGCATCT GAGTATTGA TTTGGGCTTA
10 TTGNGTGTGT GTTTAAATTTT GTTGTCAATTT TAATTTGTTA GATTATTTTG ATATGCATCT GAGTATTGA TTTGGGCTTA
11 TTGNGTGTGT --TTAAATTTT GTTGTCAATTT TAATTTGTTA GATTATTTTG ATATGCATCT GAGTATTGA TTTGGGCTTA
12 TTGNGTGTGT GTTTAAATTTT GTTGTCAATTT TAATTTGTTA GATTATTTTG ATATGCATCT GAGTATTGA TTTGGGCTTA
```

```
161
reference_hdl27  GACAGGCAAA TGAATTTAAT GACCGCGCGA GTTACCACCC CACACTGTAA ACCCCACGC CTCTAAACAA ATCATCGGAC
1  GATAGGTAAA TGAATTTAAT GATGTGTGTA GTTATTATTT TATTTTTTAA ATTTTATGT TTTTAAATAA ATTGTTGGAT
2  GATAGGTAAA TGAATTTAAT GATGTGTGTA GTTATTATTT TATTTTTTAA ATTTTATGT TTTTAAATAA ATTGTTGGAT
3  GATAGGTAAA TGAATTTAAT GATGTGTGTA GTTATTATTT TATTTTTTAA ATTTTATGT TTTTAAATAA ATTGTTGGAT
4  GATAGGTAAA TGAATTTAAT GATGTGTGTA GTTATTATTT TATTTTTTAA ATTTTATGT TTTTAAATAA ATTGTTGGAT
5  GATAGGTAAA TGAATTTAAT GATGTGTGTA GTTATTATTT TATTTTTTAA ATTTTATGT TTTTAAATAA ATTGTTGGAT
6  GATAGGTAAA TGAATTTAAT GATGTGTGTA GTTATTATTT TATTTTTTAA ATTTTATGT TTTTAAATAA ATTGTTGGAT
7  GATAGGTAAA TGAATTTAAT GATGTGTGTA GTTATTATTT TATTTTTTAA ATTTTATGT TTTTAAATAA ATTGTTGGAT
8  GATAGGTAAA TGAATTTAAT GATGTGTGTA GTTATTATTT TATTTTTTAA ATTTTATGT TTTTAAATAA ATTGTTGGAT
9  GATAGGTAAA TGAATTTAAT GATGTGTGTA GTTATTATTT TATTTTTTAA ATTTTATGT TTTTAAATAA ATTGTTGGAT
10 GATAGGTAAA TGAATTTAAT GATGTGTGTA GTTATTATTT TATTTTTTAA ATTTTATGT TTTTAAATAA ATTGTTGGAT
11 GATAGGTAAA TGAATTTAAT GATGTGTGTA GTTATTATTT TATTTTTTAA ATTTTATGT TTTTAAATAA ATTGTTGGAT
12 GATAGGTAAA TGAATTTAAT GATGTGTGTA GTTATTATTT TATTTTTTAA ATTTTATGT TTTTAAATAA ATTGTTGGAT
```

```
241
reference_hdl27  ACTCAACCGG GAAGACGGCA ACTGGAACAC CGCATCCGGC CGAATGCTGA CATTCCGGCC GAATGCTGAC ATTACACAAA
1  ATTTAATTTG GAAGATGGTA ATTGGAATAT TGTATTNGGT TGAATGTTGA TATTTNGGTT GAATGTTGAT ATTGTATAAA
2  ATTTAATTTG GAAGATGGTA ATTGGAATAT TGTATTNGGT TGAATGTTGA TATTTNGGTT GAATGTTGAT ATTGTATAAA
3  ATTTAATTTG GAAGATGGTA ATTGGAATAT TGTATTNGGT TGAATGTTGA TATTTNGGTT GAATGTTGAT ATTGTATAAA
4  ATTTAATTTG GAAGATGGTA ATTGGAATAT TGTATTNGGT TGAATGTTGA TATTTNGGTT GAATGTTGAT ATTGTATAAA
5  ATTTAATTTG GAAGATGGTA ATTGGAATAT TGTATTNGGT TGAATGTTGA TATTTNGGTT GAATGTTGAT ATTGTATAAA
6  ATTTAATTTG GAAGATGGTA ATTGGAATAT TGTATTNGGT TGAATGTTGA TATTTNGGTT GAATGTTGAT ATTGTATAAA
7  ATTTAATTTG GAAGATGGTA ATTGGAATAT TGTATTNGGT TGAATGTTGA TATTTNGGTT GAATGTTGAT ATTGTATAAA
8  ATTTAATTTG GAAGATGGTA ATTGGAATAT TGTATTNGGT TGAATGTTGA TATTTNGGTT GAATGTTGAT ATTGTATAAA
9  ATTTAATTTG GAAGATGGTA ATTGGAATAT TGTATTNGGT TGAATGTTGA TATTTNGGTT GAATGTTGAT ATTGTATAAA
10 ATTTAATTTG GAAGATGGTA ATTGGAATAT TGTATTNGGT TGAATGTTGA TATTTNGGTT GAATGTTGAT ATTGTATAAA
11 ATTTAATTTG GAAGATGGTA ATTGGAATAT TGTATTNGGT TGAATGTTGA TATTTNGGTT GAATGTTGAT ATTGTATAAA
12 ATTTAATTTG GAAGATGGTA ATTGGAATAT TGTATTNGGT TGAATGTTGA TATTTNGGTT GAATGTTGAT ATTGTATAAA
```

```
321
reference_hdl27  AGTCGCACAG CAACATTGTC CCCAGCTAGC CAGCCACATG CCGAGTCGGC ATGTTTATTA TGCTTACAAT TAAGAACCTA
1  AGTTGTATTG TAATATTGTT TTTAGTTAGT TAGTTGTATG TTGAGTTGGT ATGTTTATTA TGTTTATAAT TAAGAAATTTA
2  AGTTGTATTG TAATATTGTT TTTAGTTAGT TAGTTGTATG TTGAGTTGGT ATGTTTATTA TGTTTATAAT TAAGAAATTTA
3  AGTTGTATTG TAATATTGTT TTTAGTTAGT TAGTTGTATG TTGAGTTGGT ATGTTTATTA TGTTTATAAT TAAGAAATTTA
4  AGTTGTATTG TAATATTGTT TTTAGTTAGT TAGTTGTATG TTGAGTTGGT ATGTTTATTA TGTTTATAAT TAAGAAATTTA
5  AGTTGTATTG TAATATTGTT TTTAGTTAGT TAGTTGTATG TTGAGTTGGT ATGTTTATTA TGTTTATAAT TAAGAAATTTA
6  AGTTGTATTG TAATATTGTT TTTAGTTAGT TAGTTGTATG TTGAGTTGGT ATGTTTATTA TGTTTATAAT TAAGAAATTTA
7  AGTTGTATTG TAATATTGTT TCTAGTTAGT TA----TATG TTGAGTTGGT ATGTTTATTA TGTTTATAAT TAAGAAATTTA
8  AGTTGTATTG TAATATTGTT TTTAGTTAGT TAGTTGTATG TTGAGTTGGT ATGTTTATTA TGTTTATAAT TAAGAAATTTA
9  AGTTGTATTG TAATATTGTT TTTAGTTAGT TAGTTGTATG TTGAGTTGGT ATGTTTATTA TGTTTATAAT TAAGAAATTTA
10 AGTTGTATTG TAATATTGTT TTTAGTTAGT TAGTTGTATG TTGAGTTGGT ATGTTTATTA TGTTTATAAT TAAGAAATTTA
11 AGTTGTATTG TAATATTGTT TTTAGTTAGT TAGTTGTATG TTGAGTTGGT ATGTTTATTA TGTTTATAAT TAAGAAATTTA
12 AGTTGTATTG TAATATTGTT TTTAGTTAGT TAGTTGTATG TTGAGTTGGT ATGTTTATTA TGTTTATAAT TAAGAAATTTA
```

```
401
reference_hdl27  TGTACTT
1  TGTATTT
2  TGTATTT
3  TGTATTT
4  TGTATTT
5  TGTATTT
6  TGTATTT
7  TGTATTT
8  TGTATTT
9  TGTATTT
10 TGTATTT
11 TGTATTT
12 TGTATTT
```

Alignment: /Users/benjaminhubert/Desktop/these/data/DNA_methylation/MAKINDU_SITES/M3_1.txt
Seaview [blocks=10 fontsize=8 A4] on Tue Oct 12 16:29:07 2010

1
Reference_HDL15 GTAAAGGTAA ATGTGCGGAT TTCTACTGTT TTCPTTTGCA CCTCTTTCGC GCGAGTTACC ACCCCACCCC CTAACCCCC
1 GTAAAGGTAA ATGTGCGGAT TTCTACTGTT TTCPTTTGTA TCTCTTTCGC GCGAGTTACC ATTTTATTTT TTAACCCCC
2 GTAAAGGTAA ATGTGCGGAT TTCTACTGTT TTCPTTTGTA TCTCTTTCGC GCGAGTTACC ATTTTATTTT TTAACCCCC
3 GTAAAGGTAA ATGTGCGGAT TTCTACTGTT TTCPTTTGCA CCTCTTTCGC GCGAGTTACC ACCCCACCCC CTAACCCCC
4 GTAAAGGTAA ATGTGCGGAT TTCTACTGTT TTCPTTTGCA CCTCTTTCGC GCGGGTATC ATCTTACTTT TTAATCTCT
5 GTAAAGGTAA ATGTGCGGAT TTCTACTGTT TTCPTTTGCA CCTCTTTCGC GCGAGTTACC ACCCCATCCC CTAACCCCC
6 GTAAAGGTAA ATGTGCGGAT TTCTACTGTT TTCPTTTGCA CCTCTTTCGC GTGAGTTACC ACCCCACCCC CTAACCCCC
7 GTAAAGGTAA ATGTGCGGAT TTCTACTGTT TTCPTTTGCA CCTCTTTCGC GCGAGTTACC ACCCCACCCC CTAACCCCC
8 GTAAAGGTAA ATGTGCGGAT TTCTACTGTT TTCPTTTGCA CCTCTTTCGC GCGAGTTACC ACCCCACCCC CTAACCCCC
9 GTAAAGGTAA ATGTGCGGAT TTCTACTGTT TTCPTTTGCA CCTCTTTCGC GCGAGTTACC ACCCCACCCC CTAACCCCC
10 GTAAAGGTAA ATGTGCGGAT TTCTACTGTT TTCPTTTGCA CCTCTTTCGC GCGAGTTACC ACCCCACCCC CTAACCCCC
11 GTAAAGGTAA ATGTGCGGAT TTCTACTGTT TTCPTTTGCA CCTCTTTCGC GCGAGTTACC ATCCACCCC CTAACCCCC
12 GTAAAGGTAA ATGTGCGGAT TTCTACTGTT TTCPTTTGCA CCTCTTTCGC GCGAGTTACC ACCCCACCCC CTAACCCCC
13 GTAAAGGTAA ATGTGCGGAT TTTTACTGTT TTTTPTTGTA TTTTPTTGT GTGAGTTATT ATTTTATTTT TTAATTTTT
14 GTAAAGGTAA ATGTGCGGAT TTTTACTGTT TTTTPTTGCA TTTTPTTGT GTGAGTTATC ATTTTATTTT TTAATTTTT
15 GTAAAGGTAA ATGTGCGGAT TTTTACTGTT TTTTPTTGCA TTTTPTTGT GTGAGTTATC ATTTTATTTT TTAATTTTT

81
Reference_HDL15 ACGCCCTAA ACAATCGTC GGACACTCAA CCGGGAAGAC GGCACCTGGA ACACCCGATC CGGCCGAATG CTGACATTCC
1 ACGCCCTAA ACAATCGTC GGACACTCAA CTGGGAAGAC GGCACCTGGA ACACCCGATC CGGCCGAATG CTGACATTCC
2 ACGCCCTAA ACAATCGTC GGACACTCAA CTGGGAAGAC GGCACCTGGA ACACCCGATC CGGCCGAATG CTGACATTCC
3 ACGCCCTAA ACAATCGTC GGACACTCAA CCGGGAAGAC GGCACCTGGA ACACCCGATC CGGCCGAATG CTGACATTCC
4 ACGCCCTAA ACAATCGTC GGACACTCAA CCGGGAAGAC GGCACCTGGA ACACCCGATC CGGCCGAATG CTGACATTCC
5 ACGCCCTAA ACAATCGTC GGACACTCAA CCGGGAAGAC GGCACCTGGA ACACCCGATC CGGCCGAATG CTGACATTCC
6 ACGCCCTAA ACAATCGTC GGACACTCAA CCGGGAAGAC GGCACCTGGA ACACCCGATC CGGCCGAATG CTGACATTCC
7 ACGCCCTAA ACAATCGTC GGACACTCAA CCGGGAAGAC GGCACCTAGA ACACCCGATC CGGCCGAATG CTGACATTCC
8 ACGCCCTAA ACAATCGTC GGACTACTCAA CCGGGAAGAC GGCACCTGGA ACACCCGATC CGGCCGAATG CTGACATTCC
9 ACGCCCTAA ACAATCGTC GGACACTCAA CCGGGAAGAC GGCACCTGGA ACACCCGATC CGGCCGAATG CTGACATTCC
10 ACGCCCTAA ACAATCGTC GGACACTCAA CCGGGAAGAC GGCACCTGGA ACACCCGATC CGGCCGAATG CTGACATTCC
11 ACGCCCTAA ACAATCGTC GGACACTCAA CCGGGAAGAC GGCACCTGGA ACACCCGATC CGGCCGAATG CTGACATTCC
12 ACGCCCTAA ACAATCGTC GGACACTCAA CCGGGAAGAC GGCACCTGGA ACACCCGATC CGGCCGAATG CTGACATTCC
13 ATGTTTTTAA ATAAATTGTT GGATATTAA TTGGGAAGAT GGTAAATTGGA ATATTGTATT TGGTTGAATG TTGATATTTT
14 ATGTTTTTAA ATAAATTGTT GGATATTAA TTGGGAAGAT GGTAAATTGGA ATATTGTATT TGGTTGAATG TTGATATTTT
15 ATGTTTTTAA ATAAATTGTT GGATATTAA TTGGGAAGAT GGTAAATTGGA ATATTGTATT TGGTTGAATG TTGATATTTT

161
Reference_HDL15 GGCCGAATGC TGACATTGCA CTAAGTTCG CACTGCAACA TTGTCCCAG CTAGCCAGCC ACATGCCGAG TCGGCATGTT
1 GGTCGAATGC TGACATTGTA CTAAGTTCG CACTGCAACA TTGTCCCAG CTAGCCAGCC ACATGCCGAG TCGGCATGTT
2 GGTCGAATGC TGACATTGTA CTAAGTTCG CACTGCAACA TTGTCCCAG CTAGCCAGCC ACATGCCGAG TCGGCATGTT
3 GGCCGAATGC TGACATTGCA CTAAGTTCG CACTGCAACA TTGTCTCCAG CTAGCCAGCC ACATGCCGAG TCGGCATGTT
4 GGCCGAATGC TGACATTGCA CTAAGTTCG CACTGCAACA TTGTCTCCAG CTAGCCAGCC ACATGCCGAG TCGGCATGTT
5 GGCCGAATGC TGACATTGCA CTAAGTTCG CACTGCAACA TTGTCCCAG CTAGCCAGCC ACATGCCGAG TCGGCATGTT
6 GGTCGAATGC TGACATTGCA CTAAGTTCG CACTGCAACA TTGTCCCAG CTAGCCAGCC ACATGCCGAG TCGGCATGTT
7 GGCCGAATGC TGACATTGCA CTAAGTTCG CACTGCAACA TTGTCCCAG CTAGCCAGCC ACATGCCGAG TCGGCATGTT
8 GGCCGAATGC TGATATTGCA CTAAGTTCG CACTGCAACA TTGTCCCAG CTAGCCAGCC ACATGCCGAG TCGGCATGTT
9 GGCCGAATGC TGATATTGCA CTAAGTTCG CACTGCAACA TTGTCCCAG CTAGCCAGCC ACATGCCGAG TCGGCATGTT
10 GGCCGAATGC TGACATTGCA CTAAGTTCG CACTGCAACA TTGTCCCAG CTAGCCAGCC ACATGCCGAG TCGGCATGTT
11 GGCCGAATGC TGACATTGTA CTAAGTTCG CACTGCAACA TTGTCCCAG CTAGCCAGCC ACATGCCGAG TCGGCATGTT
12 GGCCGAATGC TGACATTGCA CTAAGTTCG CACTGCAACA TTGTCCCAG CTAGCCAGCC ACATGCCGAG TCGGCATGTT
13 GGTCGAATGC TGATATTGTA TTAAGTTCG TATTGTAATA TTGTTTTAG TTAGTTAGTT ATATGTTGAG TTGATATTTT
14 GGTCGAATGC TGATATTGTA TTAAGTTCG TATTGTAATA TTGTTTTAG TTAGTTAGTT ATATGTTGAG TTGATATTTT
15 GGTCGAATGC TGATATTGTA TTAAGTTCG TATTGTAATA TTGTTTTAG TTAGTTAGTT ATATGTTGAG TTGATATTTT

241
Reference_HDL15 CATTATGTTT ACAACTAAGA ACCATGTATC TTATGTATAA GATGGAAACG GAGTATPCGA GTAGCCTCTG CTTGACAATA
1 TATTATGTTT ACAACTAAGA ATTTATGTAT TTATGTATAA GATGGAAACG GAGTATPCGA GTAGCCTCTG CTTGACAATA
2 TATTATGTTT ACAACTAAGA ATTTATGTAT TTATGTATAA GATGGAAACG GAGTATPCGA GTAGCCTCTG CTTGACAATA
3 CATTATGTTT ACAACTAAGA ACCATGTATC TTATGTATAA GATGGAAACG GAGTATPCGA GTAGCCTCTG CTTGACAATA
4 TATTATGTTT ACAACTAAGA ACCATGTATC TTATGTATAA GATGGAAACG GAGTATPCGA GTAGCCTCTG CTTGACAATA
5 CATTATGTTT ACAACTAAGA ATCTATGTAT TTATGTATAA GATGGAAACG GAGTATPCGA GTAGCCTCTG CTTGACAATA
6 TATTATGTTT ACAACTAAGA ACCATGTATC TTATGTATAA GATGGAAACG GAGTATPCGA GTAGCCTCTG CTTGACAATA
7 CATTATGTTT ACAACTAAGA ACCATGTATC TTATGTATAA GATGGAAACG GAGTATPCGA GTAGCCTCTG CTTGACAATA
8 CATTATGTTT ATAAATAAGA ATTTATGTAT TTATGTATAA GATGGAAACG GAGTATPCGA GTAGCCTCTG CTTGACAATA
9 CATTATGTTT ACAACTAAGA ATCTATGTAT TTATGTATAA GATGGAAACG GAGTATPCGA GTAGCCTCTG CTTGACAATA
10 CATTATGTTT ACAACTAAGA ACCATGTATC TTATGTATAA GATGGAAACG GAGTATPCGA GTAGCCTCTG CTTGACAATA
11 CATTATGTTT ACAACTAAGA ACCATGTATC TTATGTATAA GATGGAAACG GAGTATPCGA GTAGCCTCTG CTTGACAATA
12 CATTATGTTT ACAACTAAGA ACCATGTATC TTATGTATAA GATGGAAACG GAGTATPCGA GTAGCCTCTG CTTGACAATA
13 TATTATGTTT ATAAATAAGA ATTTATGTAT TTATGTATAA GATGGAAATG GAGTATTTGA GTAGTTTTTG GTTGATAATA
14 TATTATGTTT ATAAATAAGA ATTTATGTAT TTATGTATAA GATGGAAATG GAGTATTTGA GTAGTTTTTG GTTGATAATA
15 TATTATGTTT ATAAATAAGA ATTTATGTAT TTATGTATAA GATGGAAATG GAGTATTTGA GTAGTTTTTG GTTGATAATA

321
Reference_HDL15 AACTTCACCTC TGATTTTGAA CTTCAAGAAA GTCAGTCGTA TTCCTTATTT GAAATCTTTA CACTACAAAT ATCTGGGGAA
1 AACTTCACCTC TGATTTTGAA CTTCAAGAAA GTCAGTCGTA TTCCTTATTT GAAATCTTTA CACTACAAAT ATCTGGGGAA
2 AACTTCACCTC TGATTTTGAA CTTCAAGAAA GTCAGTCGTA TTCCTTATTT GAAATCTTTA CACTACAAAT ATCTGGGGAA
3 AACTTCACCTC TGATTTTGAA CTTCAAGAAA GTCAGTCGTA TTCCTTATTT GAAATCTTTA CACTACAAAT ATCTGGGGAA
4 AACTTCACCTC TGATTTTGAA CCTCAAGAAA GTCAGTCGTA TTCCTTATTT GAAATCTTTA CACTACAAAT ATCTGGGGAA
5 AACTTCACCTC TGATTTTGAA CTTCAAGAAA GTCAGTCGTA TTCCTTATTT GAAATCTTTA CACTACAAAT ATCTGGGGAA
6 AACTTCACCTC TGATTTTGAA CTTCAAGAAA GTCAGTCGTA TTCCTTATTT GAAATCTTTA CACTACAAAT ATCTGGGGAA
7 AACTTCACCTC TGATTTTGAA CTTCAAGAAA GTCAGTCGTA TTCCTTATTT GAAATCTTTA CACTACAAAT ATCTGGGGAA
8 AACTTCACCTC TGATTTTGAA CTTCAAGAAA GTCAGTCGTA TTCCTTATTT GAAATCTTTA CACTACAAAT ATCTGGGGAA
9 AACTTCGCTC TGATTTTGAA CTTCAAGAAA GTCAGTCGTA TTCCTTATTT GAAATCTTTA CACTACAAAT ATCTGGGGAA
10 AACTTCACCTC TGATTTTGAA CTTCAAGAAA GTCAGTCGTA TTCCTTATTT GAAATCTTTA CACTACAAAT ATCTGGGGAA
11 AACTTCACCTC TGATTTTGAA CTTCAAGAAA GTTAGTCGTA TTCCTTATTT GAAATCTTTA CACTACAAAT ATCTGGGGAA
12 AACTTCACCTC TGATTTTGAA CTTCAAGAAA GTCAGTCGTA TTCCTTATTT GAAATCTTTA CACTACAAAT ATCTGGGGAA
13 AATTTTATTT TGATTTTGAA TTTTAAGAAA GTTAGTTGTA TTTTATTTT GAAATTTTTA TATTATAAAT ATTTGGGGAA
14 AATTTTATTT TGATTTTGAA TTTTAAGAAA GTTAGTTGTA TTTTATTTT GAAATTTTTA TATTATAAAT ATTTGGGGAA
15 AATTTTATTT TGATTTTGAA TTTTAAGAAA GTTAGTTGTA TTTTATTTT GAAATTTTTA TATTATAAAT ATTTGGGGAA

401
Reference_HDL15 ACCTAAAACC CTTTCATACGT ATACACACCT TATCTTCACA TACGGCTCCA CCCFCGAGTG CACGGACTTA ACTGGCCGAG
1 ACCTAAAATT TTTCATATGT ATATATATCT TATTTTTATA TATGGCTTTA TTCTCGAGTG CACGGACTTA ACTGGCCGAG
2 ACCTAAAATT TTTCATATGT ATATATATCT TATTTTTATA TATGGCTTTA TTCTCGAGTG CACGGACTTA ACTGGCCGAG
3 ACCTAAAACC CTTTCATACGT ATACACACCT TATCTTCACA TACGGCTCCA CCCFCGAGTG CACGGACTTA ACTGGCCGAG
4 ACCTAAAACC CTTTCATACGT ATACACACCT TATCTTCACA TACGGCTCCA CCCFCGAGTG CACGGACTTA ACTGGCCGAG
5 ACTTAAAATT TTTTATATGT ATATACATCT TATCTTCACA TACGGCTCCA CCCFCGAGTG CACGGACTTA ACTGGCCGAG
6 ACCTAAAACC CTTTCATACGT ATACACACCT TATCTTCACA TACGGCTCCA CCCFCGAGTG CACGGACTTA ACTGGCCGAG
7 ACCTAAAACC CCTCATACT ATACACACCT TATCTTCACA TACGGCTCCA CCCFCGAGTG CACGGACTTA ACTGGCCGAG
8 ACCTAAAACC CTTTCATACGT ATACACACCT TATCTTCACA TACGGCTCCA CCCFCGAGTG CACGGACTTA ACTGGCCGAG
9 ACCTAAAACC CTTTCATACGT ATACACACCT TATCTTCACA TACGGCTCCA CCCFCGAGTG CACGGACTTA ACTGGCCGAG
10 ACCTAAAACC CTTTCATACGT ATACACACCT TATCTTCACA TACGGCTCCA CCCFCGAGTG CACGGACTTA ACTGGCCGAG
11 ACCTAAAACC CTTTATATGT ATACACACCT TATCTTCACA TACGGCTCCA TCCFCGAGTG CACGGACTTA ACTGGCCGAG
12 ACCTAAAACC CTTTCATACGT ATACACACCT TATCTTCACA TACGGCTCCA CCCFCGAGTG CACGGACTTA ACTGGCCGAG
13 ATTTAAAATT TTTTATATGT ATATATATTT TATTTTTGTA TATGGTTTTA TTTTFCAGTG TATGGATTTA ATTTGGGTAG
14 ATTTAAAATT TTTTATATGT ATATATATTT TATTTTTGTA TATGGTTTTA TTTTFCAGTG TATGGATTTA ATTTGGGTAG
15 ATTTAAAATT TTTTATATGT ATATATATTT TATTTTTGTA TATGGTTTTA TTTTFCAGTG TATGGATTTA ATTTGGGTAG

481
Reference_HDL15 CCGGTAGGAT GTCCCTACCT TTAATAATCA CCTACCTGTA AGTAACCATT GTAAGAAACG AAGCAAATA TATGCAAGAT
1 CCGGTAGGAT GTCCCTACCT TTAATAATCA TCTACCTGTA AGTAACCATT GTAAGAAACG AAGCAAATA TATGCAAGAT
2 CCGGTAGGAT GTCCCTACCT TTAATAATCA TCTACCTGTA AGTAACCATT GTAAGAAACG AAGCAAATA TATGCAAGAT
3 CCGGTAGGAT GTCCCTACCT TTAATAATCA CCTACCTGTA AGTAACCATT GTAAGAAACG AAGCAAATA TATGCAAGAT
4 CCGGTAGGAT GTCCCTACCT TTAATAATCA CTTATTTGTA AGTAATTATT GTAAGAAACG AAGCAAATA TATGCAAGAT
5 CCGGTAGGAT GTCCCTACCT TTAATAATCA TCTACCTGTA AGTAACCATT GTAAGAAACG AAGCAAATA TATGCAAGAT
6 CCGGTAGGAT GTCCCTACCT TTAATAATCA TCTACCTGTA AGTAACCATT GTAAGAAACG AAGCAAATA TATGCAAGAT
7 CCGGTAGGAT GTCCCTACCT TTAATAATCA TCTACCTGTA AGTAACCATT GTAAGAAACG AAGCAAATA TATGCAAGAT
8 CCGGTAGGAT GTTTTACCT TTAATAATCA CCTACCTGTA AGTAACCATT GTAAGAAACG AAGCAAATA TATGCAAGAT
9 TCGGTAGGAT GTTCTACCT TTAATAATCA CCTACCTGTA AGTAACCATT GTAAGAAACG AAGCAAATA TATGCAAGAT
10 CCGGTAGGAT GTCCCTACCT TTAATAATCA CCTACCTGTA AGTAACCATT GTAAGAAACG AAGCAAATA TATGCAAGAT
11 CCGGTAGGAT GTCCCTACCT TTAATAATCA CCTACCTGTA AGTAACCATT GTAAGAAACG AAGCAAATA TATGCAAGAT
12 CCGGTAGGAT GTCCCTACCT TTAATAATCA CCTACCTGTA AGTAACCATT GTAAGAAACG AAGCAAATA TATGCAAGAT
13 TTGGTAGGAT GTTTTATTT TTAATAATTA TTTATTTGTA AGTAATTATT GTAAGAAATG AAGTAAATTA TATGCAAGAT
14 TTGGTAGGAT GTTTTATTT TTAATAATTA TTTATTTGTA AGTAATTATT GTAAGAAATG AAGTAAATTA TATGCAAGAT
15 TTGGTAGGAT GTTTTATTT TTAATAATTA TTTATTTGTA AGTAATTATT GTAAGAAATG AAGTAAATTA TATGCAAGAT

561
Reference_HDL15 GTCGACTG
1 GTCGACTG
2 GTCGACTG
3 GTCGACTG
4 GTCGACTG
5 GTCGACTG
6 GTCGACTG
7 GTCGACT
8 GTCGACTG
9 GTCGACTG
10 GTCGACTG
11 GTCGACTG
12 GTCGACTG
13 GTCGACTG
14 GTCGACTG
15 GTCGACTG

Alignment: /Users/benjaminhubert/Desktop/thèse/data/DNA methylation/NAKINDU_SITES/M3_2.txt
Seaview [blocks=10 fontsize=8 A4] on Tue Oct 12 16:29:31 2010

1
Reference_HDL15 TAAATGTCGC GATTTCTACT GTTTTCTTTT GCACCTCTTT CGCGCGAGTT ACCACCCAC CCTCTAAACC CCCACGCCCTC
1 TAAATGTCGC GATTTTATT GTTTTFTTTT GTATFTTTT TGTGTGAGTT ATTATTTTAT TTTTAAATTT TTTATGTTTT
2 TAAATGTCGC GATTTTATT GTTTTFTTTT GTATFTTTT TGTGTGAGTT ATTATTTTAT TTTTAAATTT TTTATGTTTT
3 TAAATGTCGC GATTTTATT GTTTTFTTTT GTATFTTTT TGTGTGAGTT ATTATTTTAT TTTTAAATTT TTTATGTTTT
4 TAAATGTCGC GATTTTATT GTTTTFTTTT GTATFTTTT TGTGTGAGTT ATTATTTTAT TTTTAAATTT TTTATGTTTT
5 TAAATGTCGC GATTTTATT GTTTTFTTTT GTATFTTTT TGTGTGAGTT ATCATTTTAT TTTTAAATTT TTTATGTTTT
6 TAAATGTCGC GATTTTATT GTTTTFTTTT GTATFTTTT TGTGTGAGTT ATTATTTTAT TTTTAAATTT TTTATGTTTT
7 TAAATGTCGC GATTTTATT GTTTTFTTTT GTATFTTTT TGTGTGAGTT ATTATTTTAT TTTTAAATTT TTTATGTTTT
8 TAAATGTCGC GATTTTATT GTTTTFTTTT GTATFTTTT TGTGTGAGTT ATTATTTTAT TTTTAAATTT TTTATGTTTT

81
Reference_HDL15 TAAACAAATC GTCGGACACT CAACCCGGAA GACCGCAACT GGAACACCGC ATCCGGCCGA ATGCTGACAT FCCCGCCGAA
1 TAAATAAATT GTTGGATATT TAATTGGGAA GATGGTAATT GGAATATTGT ATTTGGTTGA ATGTTGATAT TTTGGTTGAA
2 TAAATAAATT GTTGGATATT TAATTGGGAA GATGGTAATT GGAATATTGT ATTTGGTTGA ATGTTGATAT TTTGGTTGAA
3 TAAATAAATT GTTGGATATT TAATTGGGAA GATGGTAATT GGAATATTGT ATTTGGTTGA ATGTTGATAT TTTGGTTGAA
4 TAAATAAATT GTTGGATATT TAATTGGGAA GATGGTAATT GGAATATTGT ATTTGGTTGA ATGTTGATAT TTTGGTTGAA
5 TAAATAAATT GTTGGATATT TAATTGGGAA GATGGTAATT GGAATATTGT ATTTGGTTGA ATGTTGATAT TTTGGTTGAA
6 TAAATAAATT GTTGGATATT TAATTGGGAA GATGGTAATT GGAATATTGT ATTTGGTTGA ATGTTGATAT TTTGGTTGAA
7 TAAATAAATT GTTGGATATT TAATTGGGAA GATGGTAATT GGAATATTGT ATTTGGTTGA ATGTTGATAT TTTGGTTGAA
8 TAAATAAATT GTTGGATATT TAATTGGGAA GATGGTAATT GGAATATTGT ATTTGGTTGA ATGTTGATAT TTTGGTTGAA

161
Reference_HDL15 TGCTGACATT GCACTAAAAG TCGCACTGCA ACATTGTCCC CAGCTAGCCA GCCACATGCC GAGTCGGCAT GTTCATTATG
1 TGTTGATATT GTATTAAGA TGTATTGTA ATATTGTTTT TAGTTAGTTA GTTATATGTT GAGTTGGTAT GTTTATTATG
2 TGTTGATATT GTATTAAGA TGTATTGTA ATATTGTTTT TAGTTAGTTA GTTATATGTT GAGTTGGTAT GTTTATTATG
3 TGTTGATATT GTATTAAGA TGTATTGTA ATATTGTTTT TAGTTAGTTA GTTATATGTT GAGTTGGTAT GTTTATTATG
4 TGTTGATATT GTATTAAGA TGTATTGTA ATATTGTTTT TAGTTAGTTA GTTATATGTT GAGTTGGTAT GTTTATTATG
5 TGTTGATATT GTATTAAGA TGTATTGTA ATATTGTTTT TAGTTAGTTA GTTATATGTT GAGTTGGTAT GTTTATTATG
6 TGTTGATATT GTATTAAGA TGTATTGTA ATATTGTTTT TAGTTAGTTA GTTATATGTT GAGTTGGTAT GTTTATTATG
7 TGTTGATATT GTATTAAGA TGTATTGTA ATATTGTTTT TAGTTAGTTA GTTATATGTT GAGTTGGTAT GTTTATTATG
8 TGTTGATATT GTATTAAGA TGTATTGTA ATATTGTTTT TAGTTAGTTA GTTATATGTT GAGTTGGTAT GTTTATTATG

241
Reference_HDL15 TTTACAATA AGAACCTATG TACTTATGTA TAAGATGGAA ACGGAGTATT CGAGTAGCCT CTGGCTGACA ATAAACTTCA
1 TTTATAATTA AGAATTTATG TATTTATGTA TAAGATGGAA ATGGAGTATT TGAGTAGTTT TTGGTTGATA ATAAATTTTA
2 TTTATAATTA AGAATTTATG TATTTATGTA TAAGATGGAA ATGGAGTATT TGAGTAGTTT TTGGTTGATA ATAAATTTTA
3 TTTATAATTA AGAATTTATG TATTTATGTA TAAGATGGAA ATGGAGTATT TGAGTAGTTT TTGGTTGATA ATAAATTTTA
4 TTTATAATTA AGAATTTATG TATTTATGTA TAAGATGGAA ATGGAGTATT TGAGTAGTTT TTGGTTGATA ATAAATTTTA
5 TTTATAATTA AGAATTTATG TATTTATGTA TAAGATGGAA ATGGAGTATT TGAGTAGTTT TTGGTTGATA ATAAATTTTA
6 TTTATAATTA AGAATTTATG TATTTATGTA TAAGATGGAA ATGGAGTATT TGAGTAGTTT TTGGTTGATA ATAAATTTTA
7 TTTATAATTA AGAATTTATG TATTTATGTA TAAGATGGAA ATGGAGTATT TGAGTAGTTT TTGGTTGATA ATAAATTTTA
8 TTTATAATTA AGAATTTATG TACTTATGTA TAAGATGGAA ATGGAGTATT TGAGTAGTTT TTGGTTGATA ATAAATTTTA

321
Reference_HDL15 CTCTGATTTT GAACCTCAA
1 TTTTGATTTT GAATTTTAA
2 TTTTGATTTT GAATTTTAA
3 TTTTGATTTT GAATTTTAA
4 TTTTGATTTT GAATTTTAA
5 TTTTGATTTT GAATTTTAA
6 TTTTGATTTT GAATTTTAA
7 TTTTGATTTT GAATTTTAA
8 TTTTGATTTT GAATTTTAA

Alignment: /Users/benjaminhubert/Desktop/thèse/data/DNA_methylation/ANTISENS/OVARIES.txt
 Seaview [blocks=10 fontsize=8 A4] on Tue Oct 12 16:15:23 2010

```

1
reference  TTGTTTAGTA GCCTGTGAGT TGGCCCTTCT GCTGTTGATT TTGTGGCGTA GGCCGGCTTA CGACTGTAAAG GCCGGCTTAC GACTGTAACG
R115      TTGTTTAGTA GTTTGTGAGT TGGTTTTTTT GTTGTGATT  TTGTGGTGTG GGTGCGTTTA TGATTGTAAG GTTGGTTTAT GATTGTAATG
R11      TTGTTTAGTA GTTTGTGAGT TGGTTTTTTT GTTGTGATT  TTGTGGTGTG GGTGCGTTTA TGATTGTAAG GTTGGTTTAT GATTGTAATG
R110     TTGTTTAGTA GTTTGTGAGT TGGTTTTTTT GTTGTGATT  TTGTGGTGTG GGTGCGTTTA TGATTGTAAG GTTGGTTTAT GATTGTAATG
R111     TTGTTTAGTA GTTTGTGAGT TGGTTTTTTT GTTGTGATT  TTGTGGTGTG GGTGCGTTTA TGATTGTAAG GTTGGTTTAT GATTGTAATG
R113     TTGTTTAGTA GTTTGTGAGT TGGTTTTTTT GTTGTGATT  TTGTGGTGTG GGTGCGTTTA TGATTGTAAG GTTGGTTTAT GATTGTAATG
R12      TTGTTTAGTA GTTTGTGAGT TGGTTTTTTT GTTGTGATT  TTGTGGTGTG GGTGCGTTTA TGATTGTAAG GTTGGTTTAT GATTGTAATG
R13      TTGTTTAGTA GTTTGTGAGT TGGTTTTTTT GTTGTGATT  TTGTGGTGTG GGTGCGTTTA TGATTGTAAG GTTGGTTTAT GATTGTAATG
R14      TTGTTTAGTA GTTTGTGAGT TGGTTTTTTT GTTGTGATT  TTGTGGTGTG GGTGCGTTTA TGATTGTAAG GTTGGTTTAT GATTGTAATG
R15      TTGTTTAGTA GTTTGTGAGT TGGTTTTTTT GTTGTGATT  TTGTGGTGTG GGTGCGTTTA TGATTGTAAG GTTGGTTTAT GATTGTAATG
R410    TTGTTTAGTA GTTTGTGAGT TGGTTTTTTT GTTGTGATT  TTGTGGTGTG GGTGCGTTTA TGATTGTAAG GTTGGTTTAT GATTGTAATG
R17     TTGTTTAGTA GTTTGTGAGT TGGTTTTTTT GTTGTGATT  TTGTGGTGTG GGTGCGTTTA TGATTGTAAG GTTGGTTTAT GATTGTAATG
R19     TTGTTTAGTA GTTTGTGAGT TGGTTTTTTT GTTGTGATT  TTGTGGTGTG GGTGCGTTTA -----T GATTGTAATG
R114    TTGTTTAGTA GTTTGTGAGT TGGTTTTTTT GTTGTGATT  TTGTGGTGTG GGTGCGTTTA -----T GATTGTAATG
R116    TTGTTTAGAA GTTTGTGAGT TGGTTTTTTT GTTGTGATT  TTGTGGTGTG GGTGCGTTTA -----T GATTGTAATG
  
```

```

91
reference  TGATTTTACG CGTGACGTTG TAACAGGGGT CGATCGGTCG GTGTACGGCT CAGCCGTACA AGTAATACAA ATGTTGATTG TTGGATACAT
R115     TGATTTTACG TGTGATGTTG TAATAGGGGT TGATTGGTTG GTGTATGGTT TAGTTGTTAT AGTAATATAA ATGTTGATTG TTGGATATAT
R11      TGATTTT-AG TGTGATGTTG TAATAGGGGT TGATTGGTTG GTGTATGGTT TAGTTGTTAT AGTAATATAA ATGTTGATTG TTGGATATAT
R110    TGATTTTACG TGTGATGTTG TAATAGGGGT TGATTGGTTG GTGTATGGTT TAGTTGTTAT AGTAATATAA ATGTTGATTG TTGGATATAT
R111    TGATTTTACG TGTGATGTTG TAATAGGGGT TGATTGGTTG GTGTATGGTT TAGTTGTTAT AGTAATATAA ATGTTGATTG TTGGATATAT
R113    TGATTTTACG TGTGATGTTG TAATAGGGGT TGATTGGTTG GTGTATGGTT TAGTTGTTAT AGTAATATAA ATGTTGATTG TTGGATATAT
R12     TGATTTTACG TGTGATGTTG TAATAGGGGT TGATTGGTTG GTGTATGGTT TAGTTGTTAT AGTAATATAA ATGTTGATTG TTGGATATAT
R13     TGATTTTACG TGTGATGTTG TAATAGGGGT TGATTGGTTG GTGTATGGTT TAGTTGTTAT AGTAATATAA ATGTTGATTG TTGGATATAT
R14     TGATTTTACG TGTGATGTTG TAATAGGGGT TGATTGGTTG GTGTATGGTT TAGTTGTTAT AGTAATATAA ATGTTGATTG TTGGATATAT
R15     TGATTTTACG TGTGATGTTG TAATAGGGGT TGATTGGTTG GTGTATGGTT TAGTTGTTAT AGTAATATAA ATGTTGATTG TTGGATATAT
R410   TGATTTTACG TGTGATGTTG TAATAGGGGT TGATTGGTTG GTGTATGGTT TAGTTGTTAT AGTAATATAA ATGTTGATTG TTGGATATAT
R17    TGATTTTACG TGTGATGTTG TAATAGGGGT TGATTGGTTG GTGTATGGTT TAGTTGTTAT AGTAATATAA ATGTTGATTG TTGGATATAT
R19    TGTTTT-AG TGTGATGTTG TAATAGGGGT TGATTGGTTG GTGTATGGTT TAGTTGTTAT AGTAATATAA ATGTTGATTG TTGGATATAT
R114   TGTGTTTTAA TGTGATGTTG TAATAGGGGT TGATTGGATT ----TATGCTT TAGTTGTTAT AGTAATATAA ATGTTGATTG TTGGATATAT
R116   TGTGTTTTAA TGTGATGTTG TAATAGGGGT CGATTGGATT ----TATGCTT TAGTTGTTAT AGTAATATAA ATGTTGATTG TTGGATATAT
  
```

```

181
reference  GAATACATAT TCTACCTTTG CCTCATAAGC TCATCGG--A GACCCAGCTG TAPTTGAAGT GAGACTAAAA CTGGAAGTTC TTTCAGTCAG
R115     GAATACATAT TTTATTTTTG TTTTATAAGT TTATPCC--A GATTCATGCT TAPTTGAAGT GAGATTAAAA TTGGAAGTTC TTTTAGTTAG
R11      GAATACATAT TTTATTTTTG TTTTATAAGT TTATPCC--A GAACAGTCTG TAPTTGCTCT GAGATTAAAA TTGGAAGTTC TTTTAGTTAG
R110    GAATACATAT TTTATTTTTG TTTTATAAGT TTATPCC--A GATTCATGCT TAPTTGAAGT GAGATTAAAA TTGGAAGTTC TTTTAGTTAG
R111    GAATACATAT TTTATTTTTG TTTTATAAGT TTATPCC--A GATTCATGCT TAPTTGAAGT GAGATTAAAA TTGGAAGTTC TTTTAGTTAG
R113    GAATACATAT TTTATTTTTG TTTTATAAGT TTATPCC--A GATTCATGCT TAPTTGAAGT GAGATTAAAA TTGGAAGTTC TTTTAGTTAG
R12     GAATACATAT TTTATTTTTG TTTTATAAGT TTATPCC--A GAACAGTCTG TAPTTGCTCT GAGATTAAAA TTGGAAGTTC TTTTAGTTAG
R13     GAATACATAT TTTATTTTTG TTTTATAAGT TTATPCC--A GATTCATGCT TAPTTGAAGT GAGATTAAAA TTGGAAGTTC TTTTAGTTAG
R14     GAATACATAT TTTATTTTTG TTTTATAAGT TTATPCC--A GATTCATGCT TAPTTGAAGT GAGATTAAAA TTGGAAGTTC TTTTAGTTAG
R15     GAATACATAT TTTATTTTTG TTTTATAAGT TTATPCC--A GATTCATGCT TAPTTGAAGT GAGATTAAAA TTGGAAGTTC TTTTAGTTAG
R410   GAATACATAT TTTATTTTTG TTTTATAAGT TTATPCC--A GATTCATGCT TAPTTGAAGT GAGATTAAAA TTGGAAGTTC TTTTAGTTAG
R17    GAATACATAT TTTATTTTTG TTTTATAAGT TTATPCC--A GATTCATGCT TAPTTGAAGT GAGATTAAAA TTGGAAGTTC TTTTAGTTAG
R19    GAATACATAT TTTATTTTTG TTTTATAAGT TTATPCC--A GATTCATGCT TAPTTGAAGT GAGATTAAAA TTGGAAGTTC TTTTAGTTAG
R114   GAATACATAT TTTAATTTTTG GTTTATGATT TTATPCCGAA AAAAGATGCT TAPTTGAGCT GAGATTAAAA TTGGAAGTTC TTTTAGTTAG
R116   GAATACATAT TTTAATTTTTG GTTTATGATT TTATPCCGAA AAAAGATGCT TAPTTGAGCT GAGATTAAAA TTGGAAGTTC TTTTAGTTAG
  
```

```

271
reference  CATAAGAAAT AAATTTTAGA AATGCGATGT TTATAGACCC CTTTGGATTT TGGGAAGTAT GCATATGCTG AGAATAGAAG TGTATGCTGA
R115     CATAAGAAAT AAATTTTAGA AATGCGATGT TTATAGATTT TTTTGGATTT TGGGAAGTAT GCATATGCTG AGAATAGAAG TGTATGCTGA
R11      CATAAGAAAT AAATTTTAGA AATGCGATGT TTATAGATGA TTTTGAATTT TGGGAAGTAT GCATATGCTG AGAATAGAAG TGTATGCTGA
R110    CATAAGAAAT AAATTTTAGA AATGCGATGT TTATAGATTT TTTTGGATTT TGGGAAGTAT GCATATGCTG AGAATAGAAG TGTATGCTGA
R111    CATAAGAAAT AAATTTTAGA AATGCGATGT TTATAGATTT TTTTGGATTT TGGGAAGTAT GCATATGCTG AGAATAGAAG TGTATGCTGA
R113    CATAAGAAAT AAATTTTAGA AATGCGATGT TTATAGATTT TTTTGGATTT TGGGAAGTAT GCATATGCTG AGAATAGAAG TGTATGCTGA
R12     CATAAGAAAT AAATTTTAGA AATGCGATGT TTATAGATGA TTTTGAATTT TGGGAAGTAT GCATATGCTG AGAATAGAAG TGTATGCTGA
R13     CATAAGAAAT AAATTTTAGA AATGCGATGT TTATAGATTT TTTTGGATTT TGGGAAGTAT GCATATGCTG AGAATAGAAG TGTATGCTGA
R14     CATAAGAAAT AAATTTTAGA AATGCGATGT TTATAGATTT TTTTGGATTT TGGGAAGTAT GCATATGCTG AGAATAGAAG TGTATGCTGA
R15     CATAAGAAAT AAATTTTAGA AATGCGATGT TTATAGATTT TTTTGGATTT TGGGAAGTAT GCATATGCTG AGAATAGAAG TGTATGCTGA
R410   CATAAGAAAT AAATTTTAGA AATGCGATGT TTATAGATTT TTTTGGATTT TGGGAAGTAT GCATATGCTG AGAATAGAAG TGTATGCTGA
R17    CATAAGAAAT AAATTTTAGA AATGCGATGT TTATAGATTT TTTTGGATTT TGGGAAGTAT GCATATGCTG AGAATAGAAG TGTATGCTGA
R19    CATAAGAAAT AAATTTTAGA AATGCGATGT TTATAGATGA TTTTGAATTT TGGGAAGTAT GCATATGCTG AGAATAGAAG TGTATGCTGA
R114   TGTAA--AAT AAATTTTAGA AATATGCTGT ATATAGATGA TTTTGGATTT TGGGAAGTAT ATATATATAT GCTATAGGAA TGTATGCTGA
R116   TGTAA--AAT AAATTTTAGA AATATGCTGT ATATAGATGA TTTTGGATTT TGGGAAGTAT ATATATATAT GCTATAGAAA TGTATGCTGA
  
```

```

351
reference  GGTGGGAGCT CACGTCCTTG AATTGACCGC GTCGGCCATC CTCACGGATG GAAAATTATT AGTCGATCGA
R115     GGTGGGAGCT TATGTCPTTG AATTGATGCT GTTGGTTATT TTATAGGATG GATAAATTATT AATCGATCGA
R11      GGTGGGAGCT AGTGT--TTG AATTGATGCT GTTGGTTATT TTATAGGATG GATAAATTATT AATCGATCGA
R110    GGTGGGAGCT TATGTCPTTG AATTGATGCT GTTGGTTATT TTATAGGATG GATAAATTATT AATCGATCGA
R111    GGTGGGAGCT TATGTCPTTG AATTGATGCT GTTGGTTATT TTATAGGATG GATAAATTATT AATCGATCGA
R113    GGTGGGAGCT TATGTCPTTG AATTGATGCT GTTGGTTATT TTATAGGATG GATAAATTATT AATCGATCGA
R12     GGTGGGAGCT AGTGT--TTG AATTGATGCT GTTGGTTATT TTATAGGATG GATAAATTATT AATCGATCGA
R13     GGTGGGAGCT TATGTCPTTG AATTGATGCT GTTGGTTATT TTATAGGATG GATAAATTATT AATCGATCGA
R14     GGTGGGAGCT TATGTCPTTG AATTGATGCT GTTGGTTATT TTATAGGATG GATAAATTATT AATCGATCGA
R15     GGTGGGAGCT TATGTCPTTG AATTGATGCT GTTGGTTATT TTATAGGATG GATAAATTATT AATCGATCGA
R410   GGTGGGAGCT TATGTCPTTG AATTGATGCT GTTGGTTATT TTATAGGATG GATAAATTATT AATCGATCGA
R17    GGTGGGAGCT TATGTCPTTG AATTGATGCT GTTGGTTATT TTATAGGATG GATAAATTATT AATCGATCGA
R19    GGTGGGAGCT AGTGT--TTG AATTGATGCT GTTGGTTATT TTATAGGATG GATAAATTATT AATCGATCGA
R114   GGTGGGAGCT TATGTCPTTG AATTGATGCT GTTGGTTATT TTATAGGATG GATAAATTATT AATCGATCGA
R116   GGTGGGAGCT TATGTCPTTG AATTGATGCT GTTGGTTATT TTATAGGATG GATAAATTATT AATCGATCGA
  
```


Alignment: /Users/benjaminhubert/Desktop/thèse/data/DNA_methylation/ANTISENS/TESTES.txt
Seaview [blocks=10 fontsize=8 A4] on Tue Oct 12 16:15:46 2010

1
reference TGTGTTAGTA GCCTGTGAGT TGGCCCTTCT GCTGTTGATT TTGTGGCGTA GGCCGGCTTA CGACTGTAAG GCCCGCTTAC GACTGTAACG
R410 TTGTTTAGTA GTTGTGAGT TGGTTTTTTT GTTGTGATT TTGTGGTSTA GGTGGGTTTA TGATTGTAAG GTTGGTTTAT GATTGTAATG
R411 TTGTTTAGTA GTTGTGAGT TGGTTTTTTT GTTGTGATT TTGTGGTSTA GGTGGGTTTA TGATTGTAAG GTTGGTTTAT GATTGTAATG
R412 TTGTTTAGTA GTTGTGAGT TGGTTTTTTT GTTGTGATT TTGTGGTSTA GGTGGGTTTA TGATTGTAAG GTTGGTTTAT GATTGTAATG
R414 TTGTTTAGTA GTTGTGAGT TGGTTTTTTT GTTGTGATT TTGTGGTSTA GGTGGGTTTA TGATTGTAAG GTTGGTTTAT GATTGTAATG
R415 TTGTTTAGTA GTTGTGAGT TGGTTTTTTT GTTGTGATT TTGTGGTSTA GGTGGGTTTA TGATTGTAAG GTTGGTTTAT GATTGTAATG
R416 TTGTTTAGTA GTTGTGAGT TGGTTTTTTT GTTGTGATT TTGTGGTSTA GGTGGGTTTA TGATTGTAAG GTTGGTTTAT GATTGTAATG
R42 TTGTTTAGTA GTTGTGAGT TGGTTTTTTT GTTGTGATT TTGTGGTSTA GGTGGGTTTA TGATTGTAAG GTTGGTTTAT GATTGTAATG
R43 TTGTTTAGTA GTTGTGAGT TGGTTTTTTT GTTGTGATT TTGTGGTSTA GGTGGGTTTA TGATTGTAAG GTTGGTTTAT GATTGTAATG
R44 TTGTTTAGTA GTTGTGAGT TGGTTTTTTT GTTGTGATT TTGTGGTSTA GGTGGGTTTA TGATTGTAAG GTTGGTTTAT GATTGTAATG
R45 TTGTTTAGTA GTTGTGAGT TGGTTTTTTT GTTGTGATT TTGTGGTSTA GGTGGGTTTA TGATTGTAAG GTTGGTTTAT GATTGTAATG
R47 TTGTTTAGTA GTTGTGAGT TGGTTTTTTT GTTGTGATT TTGTGGTSTA GGTGGGTTTA TGATTGTAAG GTTGGTTTAT GATTGTAATG
R48 TTGTTTAGTA GTTGTGAGT TGGTTTTTTT GTTGTGATT TTGTGGTSTA GGTGGGTTTA TGATTGTAAG GTTGGTTTAT GATTGTAATG
R49 TTGTTTAGTA GTTGTGAGT TGGTTTTTTT GCTGTTGATT TTGTGGTSTA GGTGGGTTTA TGATTGTAAG GTTGGTTTAT GATTGTAATG

91
reference TGATTTTCAG CGTGACGTTG TAACAGGGGT CGATCGGTGC GTGTACGGCT CAGCCGTACA AGTAATACAA ATGTTGATTC TTGGATACAT
R410 TGATTTTTCAG TGTGATGTTG TAATAGGGGT TGATTTGGTTG GTGTATGGTT TAGTTGTATA AGTAATATAA ATGTTGATTC TTGGATACAT
R411 TGATTTTTCAG TGTGATGTTG TAATAGGGGT TGATTTGGTTG GTGTATGGTT TAGTTGTATA AGTAATATAA ATGTTGATTC TTGGATACAT
R412 TGATTTTTCAG TGTGATGTTG TAATAGGGGT TGATTTGGTTG GTGTATGGTT TAGTTGTATA AGTAATATAA ATGTTGATTC TTGGATACAT
R414 TGATTTTTCAG TGTGATGTTG TAATAGGGGT TGATTTGGTTG GTGTATGGTT TAGTTGTATA AGTAATATAA ATGTTGATTC TTGGATACAT
R415 TGATTTTTCAG TGTGATGTTG TAATAGGGGT TGATTTGGTTG GTGTATGGTT TAGTTGTATA AGTAATATAA ATGTTGATTC TTGGATACAT
R416 TGATTTTTCAG TGTGATGTTG TAATAGGGGT TGATTTGGTTG GTGTATGGTT TAGTTGTATA AGTAATATAA ATGTTGATTC TTGGATACAT
R42 TG-TTTTTCAG TGTGATGTTG TAATAGGGGT TGATTTGGTTG GTGTATGGTT TAGTTGTATA AGTAATATAA ATGTTGATTC TTGGATACAT
R43 TGATTTTTCAG TGTGATGTTG TAATAGGGGT TGATTTGGTTG GTGTATGGTT TAGTTGTATA AGTAATATAA ATGTTGATTC TTGGATACAT
R44 TGATTTTTCAG TGTGATGTTG TAATAGGGGT TGATTTGGTTG GTGTATGGTT TAGTTGTATA AGTAATATAA ATGTTGATTC TTGGATACAT
R45 TGATTTTTCAG TGTGATGTTG TAATAGGGGT TGATTTGGTTG GTGTATGGTT TAGTTGTATA AGTAATATAA ATGTTGATTC TTGGATACAT
R47 TGATTTTTCAG TGTGATGTTG TAATAGGGGT TGATTTGGTTG GTGTATGGTT TAGTTGTATA AGTAATATAA ATGTTGATTC TTGGATACAT
R48 TGATTTTTCAG TGTGATGTTG TAATAGGGGT TGATTTGGTTG GTGTATGGTT TAGTTGTATA AGTAATATAA ATGTTGATTC TTGGATACAT
R49 TGATTTTTCAG TGTGATGTTG TAATAGGGGT TGATTTGGTTG GTGTATGGTT TAGTTGTATA AGTAATATAA ATGTTGATTC TTGGATACAT

181
reference GAATACATAT TCTACCITTC CCTCATAAGC TCATCGGAGA CCGACTGTTA TTGAAGTGA GACTAAAAC TGAAGTTCCT TCAGTCAGCA
R410 GAATATATAT TTTATTTTTG TTTTATAAGT TTATTTGGAGA TTGATTTGTTA TTGAAGTGA GATTAATAAT TGAAGTTTTT TTAGTTAGTA
R411 GAATATATAT TTTATTTTTG TTTTATAAGT TTATTTGGAGA TTGATTTGTTA TTGAAGTGA GATTAATAAT TGAAGTTTTT TTAGTTAGTA
R412 GAATATATAT TTTATTTTTG TTTTATAAGT TTATTTGGAGA TTGATTTGTTA TTGAAGTGA GATTAATAAT TGAAGTTTTT TTAGTTAGTA
R414 GAATATATAT TTTATTTTTG TTTTATAAGT TTATTTGGAGA TTGATTTGTTA TTGAAGTGA GATTAATAAT TGAAGTTTTT TTAGTTAGTA
R415 GAATATATAT TTTATTTTTG TTTTATAAGT TTATTTGGAGA TTGATTTGTTA TTGAAGTGA GATTAATAAT TGAAGTTTTT TTAGTTAGTA
R416 GAATATATAT TTTATTTTTG TTTTATAAGT TTATTTGGAGA TTGATTTGTTA TTGAAGTGA GATTAATAAT TGAAGTTTTT TTAGTTAGTA
R42 GAATACATAT TTTATTTTTG TTTTATAAGT TTATTTGGAGA TTGATTTGTTA TTGAAGTGA GATTAATAAT TGAAGTTTTT TTAGTTAGTA
R43 GAATATATAT TTTATTTTTG TTTTATAAGT TTATTTGGAGA TTGATTTGTTA TTGAAGTGA GATTAATAAT TGAAGTTTTT TTAGTTAGTA
R44 GAATATATAT TTTATTTTTG TTTTATAAGT TTATTTGGAGA TTGATTTGTTA TTGAAGTGA GATTAATAAT TGAAGTTTTT TTAGTTAGTA
R45 GAATATATAT TTTATTTTTG TTTTATAAGT TTATTTGGAGA TTGATTTGTTA TTGAAGTGA GATTAATAAT TGAAGTTTTT TTAGTTAGTA
R47 GAATATATAT TTTATTTTTG TTTTATAAGT TTATTTGGAGA TTGATTTGTTA TTGAAGTGA GATTAATAAT TGAAGTTTTT TTAGTTAGTA
R48 GAATATATAT TTTATTTTTG TTTTATAAGT TTATTTGGAGA TTGATTTGTTA TTGAAGTGA GATTAATAAT TGAAGTTTTT TTAGTTAGTA
R49 GAATATATAT TTTATTTTTG TTTTATAAGT TTATTTGGAGA TTGATTTGTTA TTGAAGTGA GATTAATAAT TGAAGTTTTT TTAGTTAGTA

271
reference TAAGAAATAA ACTTTAGAAA TGTGATGTTT ATAGACCCTT TTGGATTTTG GGAAGTATGC ATATGTTAGT AATAGAAGTG TATGCCGAGG
R410 TAAGAAATAA ATTTTAGAAA TGTGATGTTT ATAGATTTTT TTGGATTTTG GGAAGTATGT ATATGTTAGT AATAGAAGTG TATGTTGAGG
R411 TAAGAAATAA ATTTTAGAAA TGTGATGTTT ATAGATTTTT TTGGATTTTG GGAAGTATGT ATATGTTAGT AATAGAAGTG TATGTTGAGG
R412 TAAGAAATAA ATTTTAGAAA TGTGATGTTT ATAGATTTTT TTGGATTTTG GGAAGTATGT ATATGTTAGT AATAGAAGTG TATGTTGAGG
R414 TAAGAAATAA ATTTTAGAAA TGTGATGTTT ATAGATTTTT TTGGATTTTG GGAAGTATGT ATATGTTAGT AATAGAAGTG TATGTTGAGG
R415 TAAGAAATAA ATTTTAGAAA TGTGATGTTT ATAGATTTTT TTGGATTTTG GGAAGTATGT ATATGTTAGT AATAGAAGTG TATGTTGAGG
R416 TAAGAAATAA ATTTTAGAAA TGTGATGTTT ATAGATTTTT TTGGATTTTG GGAAGTATGT ATATGTTAGT AATAGAAGTG TATGTTGAGG
R42 TAAGAAATAA TTTTAGAAG TGTGATGTTT ATAGATGATT TTGAATTTCT GGAAGTATGT AAATGTTAGT TATAGAAGTG TTTTTGAGG
R43 TAAGAAATAA ATTTTAGAAA TGTGATGTTT ATAGATTTTT TTGGATTTTG GGAAGTATGT ATATGTTAGT AATAGAAGTG TATGTTGAGG
R44 TAAGAAATAA ATTTTAGAAA TGTGATGTTT ATAGATTTTT TTGGATTTTG GGAAGTATGT ATATGTTAGT AATAGAAGTG TATGTTGAGG
R45 TAAGAAATAA ATTTTAGAAA TGTGATGTTT ATAGATTTTT TTGGATTTTG GGAAGTATGT ATATGTTAGT AATAGAAGTG TATGTTGAGG
R47 TAAGAAATAA ATTTTAGAAA TGTGATGTTT ATAGATTTTT TTGGATTTTG GGAAGTATGT ATATGTTAGT AATAGAAGTG TATGTTGAGG
R48 TAAGAAATAA ATTTTAGAAA TGTGATGTTT ATAGATTTTT TTGGATTTTG GGAAGTATGT ATATGTTAGT AATAGAAGTG TATGTTGAGG
R49 TAAGAAATAA ATTTTAGAAA TGTGATGTTT ATAGATTTTT TTGGATTTTG GGAAGTATGT ATATGTTAGT AATAGAAGTG TATGTTGAGG

361
reference TGGGAGCTCA CGTGCCCTGAA TTGACCCCGT CGGCCATCCT ACAGGATGGA AAATATTATG TGGATGGA
R410 TGGGAGTTTA TGTGTTTGAA TTGATTTGTTG TGGTTATTTT ATAGGATGGA AAATATTATG TGGATGGA
R411 TGGGAGTTTA TGTGTTTGAA TTGATTTGTTG TGGTTATTTT ATAGGATGGA AAATATTATG TGGATGGA
R412 TGGGAGTTTA TGTGTTTGAA TTGATTTGTTG TGGTTATTTT ATAGGATGGA AAATATTATG TGGATGGA
R414 TGGGAGTTTA TGTGTTTGAA TTGATTTGTTG TGGTTATTTT ATAGGATGGA AAATATTATG TGGATGGA
R415 TGGGAGTTTA TGTGTTTGAA TTGATTTGTTG TGGTTATTTT ATAGGATGGA AAATATTATG TGGATGGA
R416 TGGGAGTTTA TGTGTTTGAA TTGATTTGTTG TGGTTATTTT ATAGGATGGA AAATATTATG TGGATGGA
R42 TGGGAGTT-A -GTGTTTGAA TTGATTTGTTG TGGTTATTTT ATGGATGGA AAATATTATG TGGATGGA
R43 TGGGAGTTTA TGTGTTTGAA TTGATTTGTTG TGGTTATTTT ATAGGATGGA AAATATTATG TGGATGGA
R44 TGGGAGTTTA TGTGTTTGAA TTGATTTGTTG TGGTTATTTT ATAGGATGGA AAATATTATG TGGATGGA
R45 TGGGAGTTTA TGTGTTTGAA TTGATTTGTTG TGGTTATTTT ATAGGATGGA AAATATTATG TGGATGGA
R47 TGGGAGTTTA TGTGTTTGAA TTGATTTGTTG TGGTTATTTT ATAGGATGGA AAATATTATG TGGATGGA
R48 TGGGAGTTTA TGTGTTTGAA TTGATTTGTTG TGGTTATTTT ATAGGATGGA AAATATTATG TGGATGGA
R49 TGGGAGTTTA TGTGTTTGAA TTGATTTGTTG TGGTTATTTT ATAGGATGGA AAATATTATG TGGATGGA


```
1
reference AACACCCGCAT CCGGCCGGAAT GCTGACATTC CGGCCGAATG CTGACATTCG ACTAAAAAGTC GCACTGCAAC ATNGTCCCCA GCTAGCCAGC
A.1 AATATGTAT TGGTTGAAT GTTGATATTT TGGTTGAATG TTGATATTTG ATAAAAAGTT GTATTGTAAT ATNGTTTTTA GTTAGTTAGT
A.2 -----AAT ATTG-TATTT -GGTTGAATG TTGATATTTG ATAAAA-GTT GTATTGTAAT ATNGTTTTTA GTTAGTTAGT
A.3 -----AAT ATTG-TATTT -GGTTGAATG TTGATATTTG ATAAAA-GTT GTATTGTAAT ATNGTTTTTA GTTAGTTAGT
A.4 AACACCCGCAT CCGGCCGGAAT GCTGACATTC CGGCCGAATG CTGACATTCG ACTAAAAAGTC GCACTGCAAC ATNGTCCCCA GCTAGCCAGC
A.5 -----AAT ATTG-TATTT -GGTTGAATG TTGATATTTG ATAAAA-GTT GTATTGTAAT ATNGTTTTTA GTTAGTTAGT
A.6 AATATGTAT TGGTTGAAT GTTGATATTT TGGTTGAATG TTGATATTTG ATAAAAAGTT GTATTGTAAT ATNGTTTTTA GTTAGTTAGT
A.7 -----AAT ATTG-TATTT -GGTTGAATG TTGATATTTG ATAAAA-GTT GTATTGTAAT ATNGTTTTTA GTTAGTTAGT
A.8 AATATGTAT TGGTTGAAT GTTGATATTT TGGTTGAATG TTGATATTTG ATAAAAAGTT GTATTGTAAT ATNGTTTTTA GTTAGTTAGT
A.9 -----AAT ATTG-TATTT -GGTTGAATG TTGATATTTG ATAAAA-GTT GTATTGTAAT ATNGTTTTTA GTTAGTTAGT
A.10 -----AAT ATTG-TATTT -GGTTGAATG TTGATATTTG ATAAAA-GTT GTATTGTAAT ATNGTTTTTA GTTAGTTAGT
A.11 -----AAT ATTG-TATTT -GGTTGAATG TTGATATTTG ATAAAA-GTT GTATTGTAAT ATNGTTTTTA GTTAGTTAGT
A.12 AACACCCGCAT CCGGCCGGAAT GCTGACATTC TGGCCGAATG CTGACATTCG ACTAAAAAGTT GTATTGTAAT ATNGTCCCCA GCTAGCCAGC
A.13 AACACCCGCAT CCGACCAAT GCTGACATTC CGGCCGAATG CTGACATTCG AAAAAAGTC GCACTGCAAC ATATCCCCA GCTAGCCAGC
A.14 -----AAT ATTG-TATTT -GGTTGAATG TTGATATTTG ATAAAA-GTT GTATTGTAAT ATNGTTTTTA GTTAGTTAGT
A.15 -----AAT ATTG-TATTT -GGTTGAATG TTGATATTTG ATAAAA-GTT GTATTGTAAT ATNGTTTTTA GTTAGTTAGT
A.16 -----AAT ATTG-TATTT -GGTTGAATG TTGATATTTG ATAAAA-GTT GTATTGTAAT ATNGTTTTTA GTTAGTTAGT
A.17 -----AAT ATTG-TATTT -GGTTGAATG TTGATATTTG ATAAAA-GTT GTATTGTAAT ATNGTTTTTA GTTAGTTAGT
A.18 -----AAT ATTG-TATTT -GGTTGAATG TTGATATTTG ATAAAA-GTT GTATTGTAAT ATNGTTTTTA GTTAGTTAGT
```

```
91
reference CACATGCCGA GTCGGCATGT TCATTATGTT TACAACCTAAG AACCTATGTA CTTATGTATA AGATGGAAAC GGAGTATTCG AGTAGCCTCT
A.1 TATATGTTGA GTTGGTATGT TTATTATGTT TATAATTAAG AATTTATGTA TTTATGTATA AGATGGAAAT GGAGTATTTG AGTAGTTTTT
A.2 TATATGTTGA GTTGGTATGT TTATTATGTT TATAATTAAG AATTTATGTA TTTATGTATA AGATGGAAAT GGAGTATTTG AGTAGTTATT
A.3 TATATGTTGA GTTGGTATGT TTATTATGTT TATAATTAAG AATTTATGTA TTTATGTATA AGATGGAAAT GGAGTATTTG AGTAGTTATT
A.4 CACATGCCGA GTCGGCATGT TCATTATGTT TACAACCTAAG AACCTATGTA CTTATGTATA AGATGGAAAC GGAGTATTCG AGTAGCTCT
A.5 TATATGTTGA GTTGGTATGT TTATTATGTT TATAATTAAG AATTTATGTA TTTATGTATA AGATGGAAAT GGAGTATTTG AGTAGTTATT
A.6 TATATGTTGA GTTGGTATGT TTATTATGTT TATAATTAAG AATTTATGTA TTTATGTATA AGATGGAAAT GGAGTATTTG AGTAGTTTTT
A.7 TATATGTTGA GTTGGTATGT TTATTATGTT TATAATTAAG AATTTATGTA TTTATGTATA AGATGGAAAT GGAGTATTTG AGTAGTTATT
A.8 TATATGTTGA GTTGGTATGT TTATTATGTT TATAATTAAG AATTTATGTA TTTATGTATA AGATGGAAAT GGAGTATTTG AGTAGTTATT
A.9 TATATGTTGA GTTGGTATGT TTATTATGTT TATAATTAAG AATTTATGTA TTTATGTATA AGATGGAAAT GGAGTATTTG AGTAGTTATT
A.10 TATATGTTGA GTTGGTATGT TTATTATGTT TATAATTAAG AATTTATGTA TTTATGTATA AGATGGAAAT GGAGTATTTG AGTAGTTATT
A.11 TATATGTTGA GTTGGTATGT TTATTATGTT TATAATTAAG AATTTATGTA TTTATGTATA AGATGGAAAT GGAGTATTTG AGTAGTTATT
A.12 CACATGCCGA GTCGGCATGT TCATTATGTT TACAACCTAAG AACCTATGTA CTTATGTATA AGATGGAAAC GGAGTATTCG AGTAGCCTCT
A.13 CACATGCCGA GTCGGCATGT TCATTATGTT TACAACCTAAG AACCTATGTA CTTATGTATA AGATGGAAAC GGAGTATTCG AGTAGCCTCT
A.14 TATATGTTGA GTTGGTATGT TTATTATGTT TATAATTAAG AATTTATGTA TTTATGTATA AGATGGAAAC GGAGTATTTG AGTAGTTATT
A.15 TATATGTTGA GTTGGTATGT TTATTATGTT TATAATTAAG AATTTATGTA TTTATGTATA AGATGGAAAT GGAGTATTTG AGTAGTTTTT
A.16 TATATGTTGA GTTGGTATGT TTATTATGTT TATAATTAAG AATTTATGTA TTTATGTATA AGATGGAAAT GGAGTATTTG AGTAGTTATT
A.17 TATATGTTGA GTTGGTATGT TTATTATGTT TATAATTAAG AATTTATGTA TTTATGTATA AGATGGAAAT GGAGTATTTG AGTAGTTATT
A.18 TATATGTTGA GTTGGTATGT TTATTATGTT TATAATTAAG AATTTATGTA TTTATGTATA AGATGGAAAT GGAGTATTTG AGTAGTTATT
```

```
181
reference GGCTGACAAT AAACCTTCACT CNGATTTTGA ACTTCAAGAA AGTCAGTCGT ATTCITTTATT TGAATCTTTT ACACCTACAAA TATCTGGGGA
A.1 GGTGATAAT AAATTTTAT TTGATTTTGA ATTTTAAGAA AGTTAGITGT ATTTTTTATT TGAATTTTTT ATATTATAAA TATTTGGGGA
A.2 TTTGATAAT AAATTTTAT TTGATTTTGA ATTTTAAGAA AGTTAGITGT ATTTTTTATT TGAATTTTTT ATATTATAAA TATTTGGGGA
A.3 TTTGATAAT AAATTTTAT TTGATTTTGA ATTTTAAGAA AGTTAGITGT ATTTTTTATT TGAATTTTTT ATATTATAAA TATTTGGGGA
A.4 GGCTGACAAT AAACCTTCACT CNGATTTTGA ACTTCAAGAA AGTCAGTCGT ATTCITTTATT TGAATCTTTT ACACCTACAAA TATCTGGGGA
A.5 TTTGATAAT AAATTTTAT TTGATTTTGA ATTTTAAGAA AGTTAGITGT ATTTTTTATT TGAATTTTTT ATATTATAAA TATTTGGGGA
A.6 GGTGATAAT AAATTTTAT TTGATTTTGA ATTTTAAGAA AGTTAGITGT ATTTTTTATT TGAATTTTTT ATATTATAAA TATTTGGGGA
A.7 TTTGATAAT AAATTTTAT TTGATTTTGA ATTTTAAGAA AGTTAGITGT ATTTTTTATT TGAATTTTTT ATATTATAAA TATTTGGGGA
A.8 TTTGATAAT AAATTTTAT TTGATTTTGA ATTTTAAGAA AGTTAGITGT ATTTTTTATT TGAATTTTTT ATATTATAAA TATTTGGGGA
A.9 TTTGATAAT AAATTTTAT TTGATTTTGA ATTTTAAGAA AGTTAGITGT ATTTTTTATT TGAATTTTTT ATATTATAAA TATTTGGGGA
A.10 TTTGATAAT AAATTTTAT TTGATTTTGA ATTTTAAGAA AGTTAGITGT ATTTTTTATT TGAATTTTTT ATATTATAAA TATTTGGGGA
A.11 TTTGATAAT AAATTTTAT TTGATTTTGA ATTTTAAGAA AGTTAGITGT ATTTTTTATT TGAATTTTTT ATATTATAAA TATTTGGGGA
A.12 GGCTGACAAT AAACCTTCACT CNGATTTTGA ACTTCAAGAA AGTCAGTCGT ATTCITTTATT TGAATCTTTT ACACCTACAAA TATCTGGGGA
A.13 GGCTGACAAT AAACCTTCACT CNGATTTTGA ACTTCAAGAA AGTCAGTCGT ATTCITTTATT TGAATCTTTT ACACCTACAAA TATCTGGGGA
A.14 TTTGATAAT AAATTTTAT TTGATTTTGA ATTTTAAGAA AGTTAGITGT ATTTTTTATT TGAATTTTTT ATATTATAAA TATTTGGGGA
A.15 TTTGATAAT AAATTTTAT TTGATTTTGA ATTTTAAGAA AGTTAGITGT ATTTTTTATT TGAATTTTTT ATATTATAAA TATTTGGGGA
A.16 TTTGATAAT AAATTTTAT TTGATTTTGA ATTTTAAGAA AGTTAGITGT ATTTTTTATT TGAATTTTTT ATATTATAAA TATTTGGGGA
A.17 TTTGATAAT AAATTTTAT TTGATTTTGA ATTTTAAGAA AGTTAGITGT ATTTTTTATT TGAATTTTTT ATATTATAAA TATTTGGGGA
A.18 TTTGATAAT AAATTTTAT TTGATTTTGA ATTTTAAGAA AGTTAGITGT ATTTTTTATT TGAATTTTTT ATATTATAAA TATTTGGGGA
```

```
271
reference AACCT----- --AAAACCTT TCATACG--- -TATACACAT CTTATCTTCA CATACGGCTC CACCCCTCGAG TGCACGGACT TAACTGGCGC
A.1 AATTT----- --AAAACCTT TCATAC--- -TTTACACAT CATATCTTCA CGAARGACTC CACCCCTCGAT C--ACGGACT TAACTGGCGC
A.2 AATTT----- --AGAATTTT TTTTGT---- --TATATAT TATATTTTTA TAAAAGGTTT TATTTTGTAT T--ATGGATT TAACTGGTGT
A.3 AATTT----- --AAAATTTT TTATATG--- -TATATATAT TTTATTTTTA TATATGGTTT TATTTTGTAG TGTATGGATT TAACTGGTGT
A.4 AACCT----- --AAAACCTT TCATACG--- -TATATATAT CTTATCTTCA TATATGGCTC CACCTTCGAG TGCACGGACT TAACTGGCGC
A.5 AATTT----- --AAAATTTT T--ATAT--- -TTTATATAT TATATTTTTA TAAAAGGTTT TATTTTGTAT T--ATGGATT TAACTGGTGT
A.6 AATTT----- --AAAATTTT TTATATG--- -TATATATAT TTTATTTTTA TATATGGTTT TATTTTGTAG TGTATGGATT TAACTGGTGT
A.7 AATTT----- --AAAATTTT TTATATA--- -TTTATATAT TATATTTTTA TAAAAGGTTT TATTTTGTAT T--ATGGATT TAACTGGTGT
A.8 AATTT----- --AGAATTTT TTTTGT---- --TATATAT TATATTTTTA TAAAAGGTTT TATTTTGTAT T--ATGGATT TAACTGGTGT
A.9 AATTT----- --AAAATTTT TTATAT--- -TTTATATAT TATATTTTTA TAAAAGGTTT TATTTTGTAT T--ATGGATT TAACTGGTGT
A.10 AATTT----- --AGAATTTT TTTTGT---- --TATATAT TATATTTTTA TAAAAGGTTT TATTTTGTAT T--ATGGATT TAACTGGTGT
A.11 AATTT----- --AGAATTTT TTTTGT---- --TATATAT TATATTTTTA TAAAAGGTTT TATTTTGTAT T--ATGGATT TAACTGGTGT
A.12 AATCT----- --AAAACCTT TCATATG--- -TATACACAT CTTATCTTCA CATACGGCTC CATCCTAGAG TGCACGGACT TAACTGGCGC
A.13 AACCT----- --AAAACCTT TCATAC--- -TATACACAT CTTATCTTCA CAACAACACT CACCCCTCGAG TGCACGGACT TAACTGGCGC
A.14 AATTT----- --AGAATTTT TTTTGT---- --TATATAT TATATTTTTA TAAAAGGTTT TATTTTGTAT T--ATGGATT TAACTGGTGT
A.15 AATTT----- --AAAATTTT TTATATA--- -TTTATATAT CATATTTTTA TAAAAGGTTT TATTTTGTAT A--ATGGATT TAACTGGTGT
A.16 AATTTGGTGA GTAGAAATTT GTTGGTTGTT GGTATATAT TATATTTTTA TAAAAGGTTT TATTTTGTAT T--ATGGATT TAACTGGTGT
A.17 AACCTTGTGA GTAGAATTTT GTTGGTTGTT GGTATATAT TATATTTTTA TAAAAGGTTT TATTTTGTAT T--ATGGATT TAACTGGTGT
A.18 AATTTTGTGA GTAGAATTTT GTTGGTTGTT GGTATATAT TATATTTTTA TAAAAGGTTT TATTTTGTAT T--ATGGATT TAACTGGTGT
```


reference	AGCCGGTAGG	ATGTCCTACC	TTTTAATAAT	CACCTACCTG	TAAGTAACCA	TTGTAAGAAA	CGAAGCAAAC	TATAT
A.1	AGCCGGTAGG	ATGTCCTACC	TATTAATAAT	CACCTACCTG	TAAGTAAACA	T-GTAAGAAA	CGAAGCAAAC	TATAT
A.2	AGTTGGTAGG	ATGPTTTATT	TATTAATAAT	TATTTATTTG	TAAGTAAATA	T-GTAAGAAA	TGAAGTAAAT	TATAT
A.3	AGTTGGTAGG	ATGPTTTATT	TTTTAATAAT	TATTTATTTG	TAAGTAAATTA	TTGTAAGAAA	TGAAGTAAAT	TATAT
A.4	AGTCGGTAGG	ATGTCCTACC	TTTTAATAAT	CACCTACCTG	TAAGTAACCA	TTGTAAGAAA	CGAAGCAAAC	TATAT
A.5	AGTTGGTAGG	ATGPTTTATT	TATTAATAAT	TATTTATTTG	TAAGTAAATA	T-GTAAGAAA	TGAAGTAAAT	TATAT
A.6	AGTTGGTAGG	ATGPTTTATT	TTTTAATAAT	TATTTATTTG	TAAGTAAATTA	TTGTAAGAAA	TGAAGTAAAT	TATAT
A.7	AGTTGGTAGG	ATGPTTTATT	TATTAATAAT	TATTTATTTG	TAAGTAAATA	T-GTAAGAAA	TGAAGTAAAT	TATAT
A.8	AGTTGGTAGG	ATGPTTTATT	TATTAATAAT	TATTTATTTG	TAAGTAAATA	T-GTAAGAAA	TGAAGTAAAT	TATAT
A.9	AGTTGGTAGG	ATGPTTTATT	TATTAATAAT	TATTTATTTG	TAAGTAAATA	T-GTAAGAAA	TGAAGTAAAT	TATAT
A.10	AGTTGGTAGG	ATGPTTTATT	TATTAATAAT	TATTTATTTG	TAAGTAAATA	T-GTAAGAAA	TGAAGTAAAT	TATAT
A.11	AGTTGGTAGG	ATGPTTTATT	TATTAATAAT	TATTTATTTG	TAAGTAAATA	T-GTAAGAAA	TGAAGTAAAT	TATAT
A.12	AGCCGGTAGG	ATGTCCTACC	TTTTAATAAT	CACCTACCTG	TAAGTAAACA	TTGTAAGAAA	CGAAGCAAAC	TATAT
A.13	AGCCGGTAGG	ATGTCCTACC	TTTTAATAAT	CACCTACCTG	TAAGTAACCA	TTGTAAGAAA	CGAAGCAAAC	TATAT
A.14	AGTTGGTAGG	ATGPTTTATT	TATTAATAAT	TATTTATTTG	TAAGTAAATA	T-GTAAGAAA	TGAAGTAAAT	TATAT
A.15	AGTTGGTAGG	ATGPTTTATT	TATTAATAAT	TATTTATTTG	TAAGTAAATA	T-GTAAGAAA	TGAAGTAAAT	TATAT
A.16	AGTTGGTAGG	ATGPTTTATT	TATTAATAAT	TATTTATTTG	TAAGTAAATA	T-GTAAGAAA	TGAAGTAAAT	TATAT
A.17	AGTTGGTAGG	ATGPTTTATT	TATTAATAAT	TATTTATTTG	TAAGTAAATA	T-GTAAGAAA	TGAAGTAAAT	TATAT
A.18	AGTTGGTAGG	ATGPTTTATT	TATTAATAAT	TATTTATTTG	TAAGTAAATA	T-GTAAGAAA	TGAAGTAAAT	TATAT

Alignment: /Users/benjaminhubert/Desktop/thèse/data/DNA_methylation/EXPERIMENTAL_CONDITIONS/MAKINDU_OVARIES/+PROT_K.t
 Seaview [blocks=10 fontsize=8 A4] on Tue Oct 12 16:21:10 2010

```

1
reference AACACCCGAT CCGGCCGAAT GCTGACATTC CGGCCGAATG CTGACATTCG ACTAAAA-GT CGCACTGCAA CATTGTCCCC AGCTAGCCAG
B.1 -----AAT ATTC-TATTT -GGTTGAATG TTGATATGTT AT-AAAA-GT TGTATTCGTA TATTGTTTTT AGTTAGTTEAG
B.2 -----AAT ATTC-TATTT -GGTTGAATG TTGATATGTT AT-AAAA-GT TGTATTCGTA TATTGTTTTT AGTTAGTTEAG
B.3 -----AAT ATTC-TATTT -GGTTGAATG TTGATATGTT AT-AAAA-GT TGTATTCGTA TATTGTTTTT AGTTAGTTEAG
B.4 AATATTGTAT TCGGTTGGAAT GTTGGTATTT TGGTTGAATG TTGATATGTT AT-AAAAAGT TGTATTCGTA TACTGTTTTT AGTTAGTTEAG
B.5 -----AAT ATTC-TATTT -GGTTGAATG TTGATATGTT AT-AAAA-GT TGTATTCGTA TATTGTTTTT AGTTAGTTEAG
B.6 -----AAT ATTC-TATTT -GGTTGAATG TTGATATGTT AT-AAAA-GT TGTATTCGTA TATTGTTTTT AGTTAGTTEAG
B.7 AACACCCGAT CCGGCCGAAT GCTGACATTC CGGCCGAATG CTGACATTCG ATAAAA-GT CGCACTGCAA CATTGTCCCC AGCTAGCCAG
  
```

```

91
reference CCACATGCCG AGTCGGCATG TTCATTATGT TTACAACATA GAACCTATGT ACTTATGTAT AAGATGGAAA CGGAGTATTC GAGTAGCCTC
B.1 TTATATGTTG AGTTGGTATG TTTATTTATG TTATAAATTA GAATTTATGT ATTTATGTAT AAGGTGGAAA TGGAGTATTT GAGTAGTTEAT
B.2 TTATATGTTG AGTTGGTATG TTTATTTATG TTATAAATTA GAATTTATGT ATTTATGTAT AAGATGGAAA TGGAGTATTT GAGTAGTTEAT
B.3 TTATATGTTG AGTTGGTATG TTTATTTATG TTATAAATTA GAATTTATGT ATTTATGTAT AAGATGGAAA TGGAGTATTT GAGTAGTTEAT
B.4 TTATATGTTG AGTTGGTATG TTTATTTATG TTATAAATTA GAATTTATGT ATTTATGTAT AAGATGGAAA TGGAGTATTT GAGTAGTTEAT
B.5 TTATATGTTG AGTTGGTATG TTTATTTATG TTATAAATTA GAATTTATGT ATTTATGTAT AAGATGGAAA TGGAGTATTT GAGTAGTTEAT
B.6 TTATATGTTG AGTTGGTATG TTTATTTATG TTATAAATTA GAATTTATGT ATTTATGTAT AAGATGGAAA TGGAGTATTT GAGTAGTTEAT
B.7 TTATATGTTG AGTTGGTATG TTTATTTATG TTATAAATTA GAACCTATGT ACTTATGTAT AAGATGGAAA CGGAGTATTC GAGTAGCCTC
  
```

```

181
reference TGGCTGACAA TAAACTTCAC TCTGATTTTG AACTTCAAGA AAGTCAGTCC TATTCCTTAT TTGAAATCTT TACACTACAA ATATCTGGGG
B.1 TTTTGGATAA TAAATTTTAT TTTGATTTTG AATTTAAGA AAGTTAGTTG TATTTTTTAT TGGAAATTTT TATAATTATA ATATTTGTTG
B.2 TTTTGGATAA TAAATTTTAT TTTGATTTTG AATTTAAGA AAGTTAGTTG TATTTTTTAT TGGAAATTTT TATAATTATA ATATTTGTTG
B.3 TTTTGGATAA TAAATTTTAT TTTGATTTTG AATTTAAGA AAGTTAGTTG TATTTTTTAT TGGAAATTTT TATAATTATA ATATTTGTTG
B.4 TTTTGGATAA TAAATTTTAT TTTGATTTTG AATTTAAGA AAGTTAGTTG TATTTTTTAT TGGAAATTTT TATAATTATA ATATTTGTTG
B.5 TTTTGGATAA TAAATTTTAT TTTGATTTTG AATTTAAGA AAGTTAGTTG TATTTTTTAT TGGAAATTTT TATAATTATA ATATTTGTTG
B.6 TTTTGGATAA TAAATTTTAT TTTGATTTTG AATTTAAGA AAGTTAGTTG TATTTTTTAT TGGAAATTTT TATAATTATA ATATTTGTTG
B.7 TGGCTGACAA TAAACTTCAC TCTGATTTTG AACTTCAAGA AAGTCAGTCC TATTCCTTAT TTGAAATCTT TACACTACAA ATATCTGGGG
  
```

```

271
reference AAACCTAAAA CCCTTCATAC GTATACACAT CTTATCTTCA CATACGGCTC CACCCTCGAG TGCACGGACT TAACTGGCGC AGCCGGTAGG
B.1 AAATTT---- --TTATATAT TATATTTTAA TAAAAGGTTT TACTTTTGAT T--ATGGATT TAATTTGGTGT AGTTGGTAGG
B.2 AAATTT---- --TTATATAT TATATTTTAA CAAAAGGTTT TATTTTTGAT T--ATGGATT TAATTTGGTGT AGTTGGTAGG
B.3 AAATTT---- --TTATATAT TATATTTTAA TAAAAGGTTT TATTTTTGAT T--ATGGATT TAATTTGGTGT AGTTGGTAGG
B.4 AAATTTAAAA ATTTTATAT ATTTATATAT TATATTTTAA TATATGGTTT TATTTTTGAG TGTATGGATT TAATTTGGTGT AGTTGGTAGG
B.5 AAATTTAAAA ATTTTATAT ATTTATATAT TATATTTTAA TAAAAGGTTT TATTTTTGAT T--ATGGATT TAATTTGGTGT AGTTGGTAGG
B.6 AAATTTAAAA ATTTTATAT ATTTATATAT TATATTTTAA TAAAAGGTTT TATTTTTGAT T--ATGGATT TAATTTGGTGT AGTTGGTAGG
B.7 AAATCTAAAA CCCTTTATAT GTATACACAT CTTATCTTAA TATATGGTCC CACCCTCGAG TGCACGGACT TAACTGGCGC AGCCGGTAGG
  
```

```

361
reference ATGTCCTACC TTTTAATAAT CACCTACCTG TAAGTAACCA TTGTAAGAAA CGAAGCAAAAC TATAT
B.1 ATGTTTTAT TATTAATAAT TATTTATTTG TAAGTAAATA T-GTAAGAAA TGAAGTAAAT TATAT
B.2 ATGTTTTAT TATTAATAAT TATTTATTTG TAAGTAAATA T-GTAAGAAA TGAAGTAAAT TATAT
B.3 ATGTTTTAT TATTAATAAT TATTTATTTG TAAGTAAATA T-GTAAGAAA TGAAGTAAAT TATAT
B.4 ATGTTTTAT TTTTAATAAT TATTTATTTG TAAGTAAATA TTGTAAGAAA TGAAGTAAAT TATAT
B.5 ATGTTTTAT TATTAATAAT TATTTATTTG TAAGTAAATA T-GTAAGAAA TGAAGTAAAT TATAT
B.6 ATGTTTTAT TATTAATAAT TATTTATTTG TAAGTAAATA T-GTAAGAAA TGAAGTAAAT TATAT
B.7 ATGTCCTACC TTTTAATAAT TATTTATTTG TAAGTAAATA TTGTAAGAAA CGAAGCAAAAC TATAT
  
```

Alignment: /Users/benjaminhubert/Desktop/thèse/data/DNA_methylation/EXPERIMENTAL_CONDITIONS/MAKINDU_OVARIES/+PROT_K_RNA
 Seaview [blocks=10 fontsize=8 A4] on Tue Oct 12 16:20:38 2010

```

1
reference AACACCCGCAT CCGGCCGAAT GCTGACATTC CGGCCGAATG CTGACATTCG ACTAAAAGTC GCACATGCAAC ATTGTCGCCA GCTAGCCAGC
C.1 -----AAAT ATTG-TATTT -GGTTGAATG TTGATATTCG AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
C.2 -----AAAT ATTG-TATTT -GGTTGAATG TTGATATTCG AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
C.3 -----AAAT ATTG-TATTT -GGTTGAATG TTGATATTCG AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
C.4 -----AAAT ATTG-TATTT -GGTTGAATG TTGATATTCG AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
C.5 -----AAAT ATTG-TATTT -GGTTGAATG TTGATATTCG AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
C.6 -----AAAT ATTG-TATTT -GGTTGAATG TTGATATTCG AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
C.7 -----AAAT ATTG-TATTT -GGTTGAATG TTGATATTCG AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
C.8 -----AAAT ATTG-TATTT -GGTTGAATG TTGATATTCG AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
C.9 -----AAAT ATTG-TATTT -GGTTGAATG TTGATATTCG AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
C.10 -----AAAT ATTG-TATTT -GGTTGAATG TTGATATTCG AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
C.11 AATATGTAT TTGGTTGAAT GTTGATATTT TGGTTGAATG TTGATATTCG AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
C.12 -----AAAT ATTG-TATTT -GGTTGAATG TTGATATTCG AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
C.13 -----AAAT ATTG-TATTT -GGTTGAATG TTGATATTCG AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
C.14 AATATGTAT TTGGTTGAAT GTTGATATTT TGGTTGAATG TTGATATTCG AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
C.15 -----AAAT ATTG-TATTT -GGTTGAATG TTGATATTCG AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
C.16 AATATGTAT TTGGTTGAAT GTTGATATTT TGGTTGAATG TTGATATTCG AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT
C.17 AATATGTAT TTGGTTGAAT GTTGATATTT TGGTTGAATG TTGATATTCG AT-AAAAGTT GTATTGTAAT ATTGTTTTTA GTTAGTTAGT

```

```

91
reference CACATGCCGA GTCCGCATGT TCATTATGTT TACAACCTAG AACCTATGTA CTTATGTATA AGATGGAAAC GGAGTATTCG AGTAGCCCTC
C.1 TATATGTTGA GTTGGTATGT TTATTATGTT TATAAATTAAG AATTTATGTA TTTATGTATA AGATGGAAAN GGAGTATTCG AGTAGTTATT
C.2 TATATGTTGA GTTGGTATGT TTATTATGTT TATAAATTAAG AATTTATGTA TTTATGTATA AGATGGAAAT GGAGTATTCG AGTAGTTATT
C.3 TATATGTTGA GTTGGTATGT TTATTATGTT TATAAATTAAG AATTTATGTA TTTATGTATA AGATGGAAAT GGAGTATTCG AGTAGTTATT
C.4 TATATGTTGA GTTGGTATGT TTATTATGTT TATAAATTAAG AATTTATGTA TTTATGTATA AGATGGAAAT GGAGTATTCG AGTAGTTATT
C.5 TATATGTTGA GTTGGTATGT TTATTATGTT TATAAATTAAG AATTTATGTA TTTATGTATA AGATGGAAAT GGAGTATTCG AGTAGTTATT
C.6 TATATGTTGA GTTGGTATGT TTATTATGTT TATAAATTAAG AATTTATGTA TTTATGTATA AGATGGAAAT GGAGTATTCG AGTAGTTATT
C.7 TATATGTTGA GTTGGTATGT TTATTATGTT TATAAATTAAG AATTTATGTA TTTATGTATA AGATGGAAAT GGAGTATTCG AGTAGTTATT
C.8 TATATGTTGA GTTGGTATGT TTATTATGTT TATAAATTAAG AATTTATGTA TTTATGTATA AGATGGAAAT GGAGTATTCG AGTAGTTATT
C.9 TATATGTTGA GTTGGTATGT TTATTATGTT TATAAATTAAG AATTTATGTA TTTATGTATA AGATGGAAAT GGAGTATTCG AGTAGTTATT
C.10 TATATGTTGA GTTGGTATGT TTATTATGTT TATAAATTAAG AATTTATGTA TTTATGTATA AGATGGAAAC GGAGTATTCG AGTAGTTATT
C.11 TATATGTTGA GTTGGTATGT TTATTATGTT TATAAATTAAG AATTTATGTA TTTATGTATA AGATGGAAAT GGAGTATTCG AGTAGTTATT
C.12 TATATGTTGA GTTGGTATGT TTATTATGTT TATAAATTAAG AATTTATGTA TTTATGTATA AGATGGAAAT GGAGTATTCG AGTAGTTATT
C.13 TATATGTTGA GTTGGTATGT TTATTATGTT TATAAATTAAG AATTTATGTA TTTATGTATA AGATGGAAAT GGAGTATTCG AGTAGTTATT
C.14 TATATGTTGA GTTGGTATGT TTATTATGTT TATAAATTAAG AATTTATGTA TTTATGTATA AGATGGAAAT GGAGTATTCG AGTAGTTATT
C.15 TATATGTTGA GTTGGTATGT TTATTATGTT TATAAATTAAG AATTTATGTA TTTATGTATA AGATGGAAAT GGAGTATTCG AGTAGTTATT
C.16 TATATGTTGA GTTGGTATGT TTATTATGTT TATAAATTAAG AATTTATGTA TTTATGTATA AGATGGAAAT GGAGTATTCG AGTAGTTATT
C.17 TATATGTTGA GTTGGTATGT TTATTATGTT TATAAATTAAG AATTTATGTA TTTATGTATA AGATGGAAAT GGAGTATTCG AGTAGTTATT

```

```

181
reference GGCTGACAAT AAACCTCAC TCGATTTTGA ACTTCAAGAA A-GTCAGTCC TATTCCTTAT TTGAAATCTT TACACTACAA ATATCTGGGG
C.1 TTTTGATAAT AAATTTTAT TTTGATTTGA ATTTTAAGAA A-GTTAGTTC TATTTTAT TTGAAATCTT TATATTATAA ATATTTGTTG
C.2 TTTTGATAAT AAATTTTAT TTTGATTTGA ATTTTAAGAA A-GTTAGTTC TATTTTAT TTGAAATCTT TATATTATAA ATATTTGTTG
C.3 TTTTGATAAT AAATTTTAT TTTGATTTGA ATTTTAAGAA A-GTTAGTTC TATTTTAT TTGAAATCTT TATATTATAA ATATTTGTTG
C.4 TTTTGATAAT AAATTTTAT TTTGATTTGA ATTTTAAGAA A-GTTAGTTC TATTTTAT TTGAAATCTT TATATTATAA ATATTTGTTG
C.5 GGTTGATAAT AAATTTTAT TTTGATTTGA ATTTTAAGAA A-GTTAGTTC TATTTTAT TTGAAATCTT TATATTATAA ATATTTGTTG
C.6 TTTTGATAAT AAATTTTAT TTTGATTTGA ATTTTAAGAA A-GTTAGTTC TATTTTAT TTGAAATCTT TATATTATAA ATATTTGTTG
C.7 TTTTGACAAT AAATTTTAT TTTGATTTGA ATTTTAAGAA A-GTTAGTTC TATTTTAT TTGAAATCTT TATATTATAA ATATTTGTTG
C.8 TTTTGATAAT AAATTTTAT TTTGATTTGA ATTTTAAGAA A-GTTAGTTC TATTTTAT TTGAAATCTT TATATTATAA ATATTTGTTG
C.9 TTTTGATAAT AAATTTTAT TTTGATTTGA ATTTTAAGAA A-GTTAGTTC TATTTTAT TTGAAATCTT TATATTATAA ATATTTGTTG
C.10 TTTTGATAAT AAATTTTAT TTTGATTTGA ATTTTAAGAA A-GTTAGTTC TATTTTAT TTGAAATCTT TATATTATAA ATATTTGTTG
C.11 GGTTGATAAT AAATTTTAT TTTGATTTGA ATTTTAAGAA A-GTTAGTTC TATTTTAT TTGAAATCTT TATATTATAA ATATTTGTTG
C.12 GGTTGATAAT AAATTTTAT TTTGATTTGA ATTTTAAGAA A-GTTAGTTC TATTTTAT TTGAAATCTT TATATTATAA ATATTTGTTG
C.13 TTTTGATAAT AAATTTTAT TTTGATTTGA ATTTTAAGAA A-GTTAGTTC TATTTTAT TTGAAATCTT TATATTATAA ATATTTGTTG
C.14 TTTTGATAAT AAATTTTAT TTTGATTTGA ATTTTAAGAA A-GTTAGTTC TATTTTAT TTGAAATCTT TATATTATAA ATATTTGTTG
C.15 TTTTGATAAT AAATTTTAT TTTGATTTGA ATTTTAAGAA A-GTTAGTTC TATTTTAT TTGAAATCTT TATATTATAA ATATTTGTTG
C.16 GGTTGATAAT AAATTTTAT TTTGATTTGA ATTTTAAGAA A-GTTAGTTC TATTTTAT TTGAAATCTT TATATTATAA ATATTTGTTG
C.17 TTTTGATAAT AAATTTTAT TTTGATTTGA ATTTTAAGAA A-GTTAGTTC TATTTTAT TTGAAATCTT TATATTATAA ATATTTGTTG

```

```

271
reference AAACCTAAAA CCCTTC---- -----A TACGTATACA CAPCTTATCT TCACATACGG CTCACCCTC GAGTGCACGG ACTTAACTGG
C.1 AAATTTAGAA TTT----- TTT-----TTT T--GTATATA T--ATATTT TTATAAAGG TTTTATTTTT GATT--ATGG ATTTAACTGG
C.2 AAATTT----- -----TTT----- ---TTATATA T--ATATTT TTATAATGG TTTTATTTTT GAGTGCATGG ATTTAACTGG
C.3 AAATTT----- -----TTT----- ---TTATATA T--ATATTT TTATAAAGG TTTTATTTTT GATT--ATGG ATTTAACTGG
C.4 AAATTT----- -----TTT----- ---TTATATA T--ATATTT TTATAAAGG TTTTATTTTT GATT--ATGG ATTTAACTGG
C.5 AAATTTAAAA TTTTT----- -----A TATGTATATA T--TTATTT TTATAATGG TTTTATTTTT GAGTGCATGG ATTTAACTGG
C.6 AAATTTAAAA TTT-----TTA TATTTATATA T--ATATTT TTATAAAGG TTTTATTTTT GATT--ATGG ATTTAACTGG
C.7 AAATTT----- -----TTT----- ---TTATATA T--ATATTT TTATAAAGG TTTTATTTTT GATT--ATGG ATTTAACTGG
C.8 AAATTT----- -----TTT----- ---TTATATA T--ATATTT TTATAAAGG TTTTATTTTT GATT--ATGG ATTTAACTGG
C.9 AAATTTAAAA TTT-----TTA TATTT-ATATA T--ATATTT TTATAAAGG TTTTATTTTT GATT--ATGG ATTTAACTGG
C.10 AAATTT----- -----TTT----- ---TTATATA T--ATATTT TTATAAAGG TTTTATTTTT GATT--ATGG ATTTAACTGG
C.11 AAATTTAAAA TTTTT----- -----A TATGTATATA T--TTATTT TTATAATGG TTTTATTTTT GAGTGCATGG ATTTAACTGG
C.12 AAATTTAAAA TTTTT----- -----A TATGTATATA TATTTATTT TTATAAATGG TTTTATTTTT GAGTGCATGG ATTTAACTGG
C.13 AAATTT----- -----TTT----- ---TTATATA T--ATATTT TTATAAAGG TTTTATTTTT GATT--ATGG ATTTAACTGG
C.14 AAATTTGCTT AGTAGAATTT TCTTGGTTGT TGGTTATATA TT--ATATTT TTATAAAGG TTTTATTTTT GATT--ATGG ATTTAACTGG
C.15 AAATTT----- -----TTT----- ---TTATATA T--ATATTT TTATAAAGG TTTTATTTTT GATT--ATGG ATTTAACTGG
C.16 AAATTTAAAA TTTTT----- -----A TATGTATATA TATTTATTT TTATAATGG TTTTATTTTT GAGTGCATGG ATTTAACTGG
C.17 AAATTTGCTT AGTAGAATTT TCTTGGTTGT TGGTTATATA TT--ATATTT TTATAAAGG TTTTATTTTT GATT--ATGG ATTTAACTGG

```

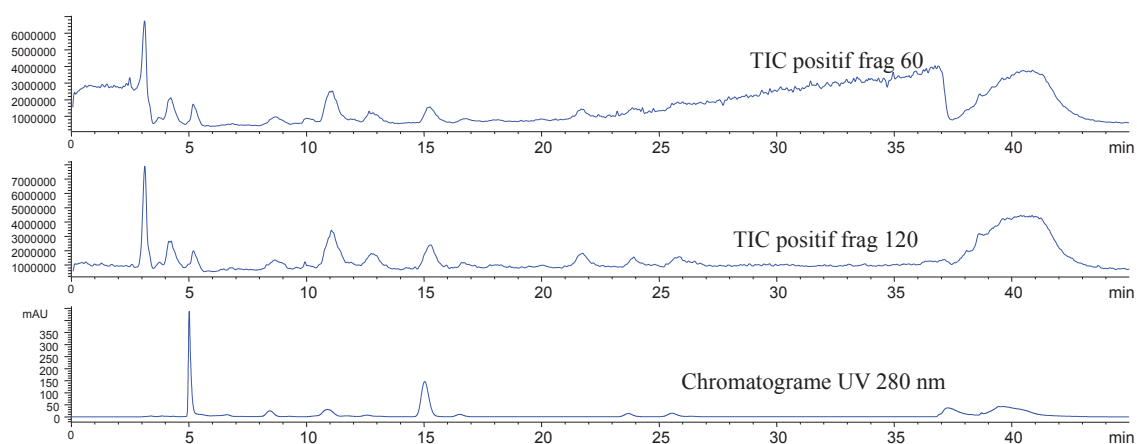
```

reference CGCAGCCGGT AGGATGTCCT ACCTTTTAAAT AATCACCTAC CTGTAAGTAA CCATTGTAAG AAACGAAGCA AACTATAT
C.1       TGTAGTTGGT AGGATGTTTT ATTTATTAAT AATTATTTAT TTGTAAGTAA ATAT-GTAAG AAATGAAGTA AATTATAT
C.2       TGTAGTTGGT AGGATGTTTT ATTTTAAAT AATTATTTAT TTGTAAGTAA TTATTGTAAG AAATGAAGTA AATTATAT
C.3       TGTAGTTGGT AGGATGTTTT ATTTATTAAT AATTATTTAT TTGTAAGTAA ATAT-GTAAG AAATGAAGTA AATTATAT
C.4       TGTAGTTGGT AGGATGTTTT ATTTATTAAT AATTACTTAT TTGTAAGTAA ATAT-GTAAG AAATGAAGTA AATTATAT
C.5       TGTAGTTGGT AGGATGTTTT ATTTTAAAT AATTATTTAT TTGTAAGTAA TTATTGTAAG AAATGAAGTA AATTATAT
C.6       TGTAGTTGGT AGGATGTTTT ATTTATTAAT AATTATTTAT TTGTAAGTAA ATAT-GTAAG AAATGAAGTA AATTATAT
C.7       TGTAGTTGGT AGGATGTTTT ATTTATTAAT AATTATTTAT TTGTAAGTAA ATAT-GTAAG AAATGAAGTA AATTATAT
C.8       TGTAGTTGGT AGGATGTTTT ATTTATTAAT AATTATTTAT TTGTAAGTAA ATAT-GTAAG AAATGAAGTA AATTATAT
C.9       TGTAGTTGGT AGGATGTTTT ATTTATTAAT AATTATTTAT TTGTAAGTAA ATAT-GTAAG AAATGAAGTA AATTATAT
C.10      TGTAGTTGGT AGGATGTTTT ATTTATTAAT AATTATTTAT TTGTAAGTAA ATAT-GTAAG AAATGAAGTA AATTATAT
C.11      TGTAGTTGGT AGGATGTTTT ATTTTAAAT AATTATTTAT TTGTAAGTAA TTATTGTAAG AAATGAAGTA AATTATAT
C.12      TGTAGTTGGT AGGATGTTTT ATTTTAAAT AATTATTTAT TTGTAAGTAA TTATTGTAAG AAATGAAGTA AATTATAT
C.13      TGTAGTTGGT AGGATGTTTT ATTTATTAAT AATTATTTAT TTGTAAGTAA ATAT-GTAAG AAATGAAGTA AATTATAT
C.14      TGTAGTTGGT AGGATGTTTT ATTTATTAAT AATTATTTAT TTGTAAGTAA TTATTGTAAG AAATGAAGTA AATTATAT
C.15      TGTAGTTGGT GGGATGTTTT ATTTTAAAT AATTATTTAT TTGTAAGTAA TTATTGTAAG AAATGAAGTA AATTATAT
C.16      TGTAGTTGGT AGGATGTTTT ATTTTAAAT AATTATTTAT TTGTAAGTAA TTATTGTAAG AAATGAAGTA AATTATAT
C.17      TGTAGTTGGT AGGATGTTTT ATTTATTAAT AATTATTTAT TTGTAAGTAA TTATTGTAAG AAATGAAGTA AATTATAT

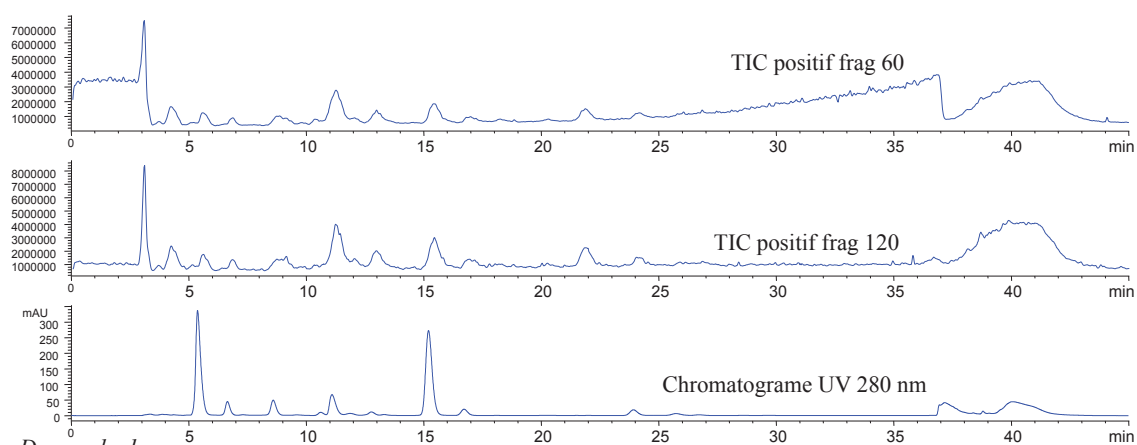
```


2. HPLC et spectrométrie de masse

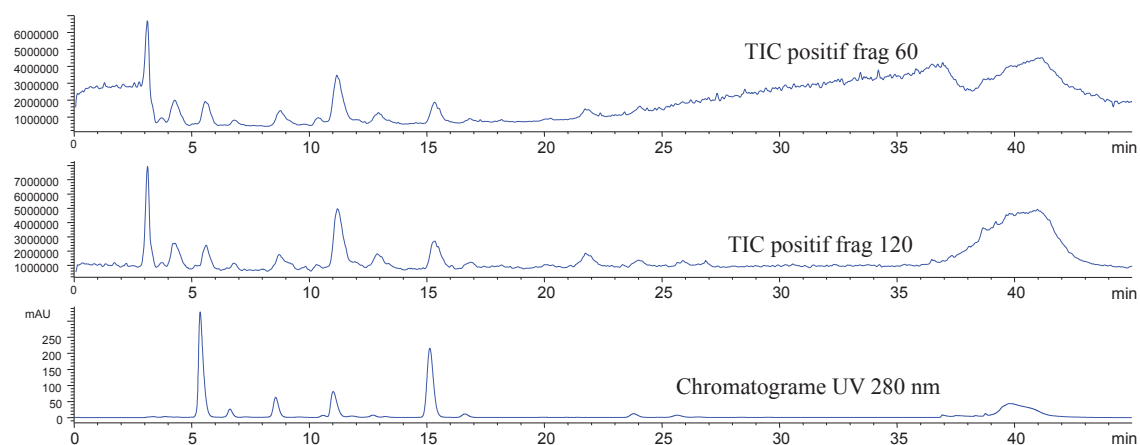
Figure supplémentaire 8. . Chromatogrammes en courant ionique total et UV obtenus par HPLC pour les 12 génomes séquencés de drosophiles.



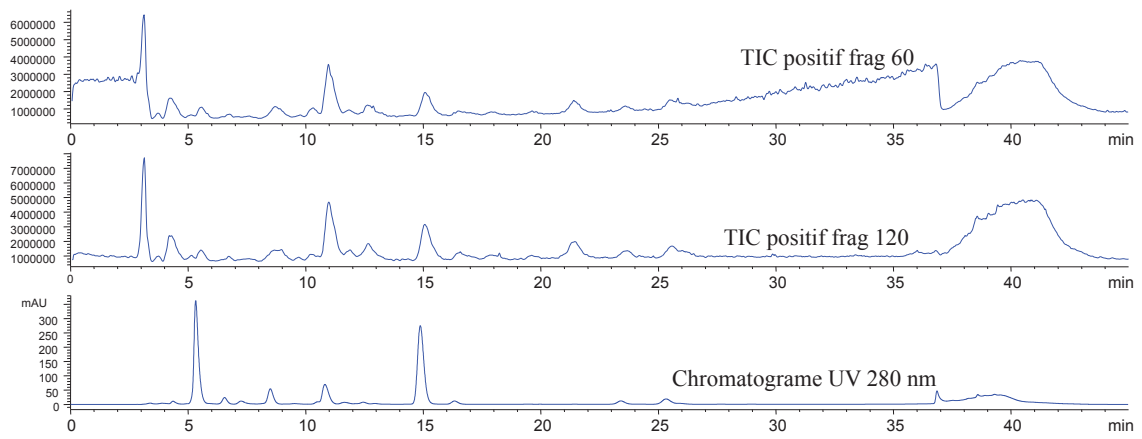
D. willistoni



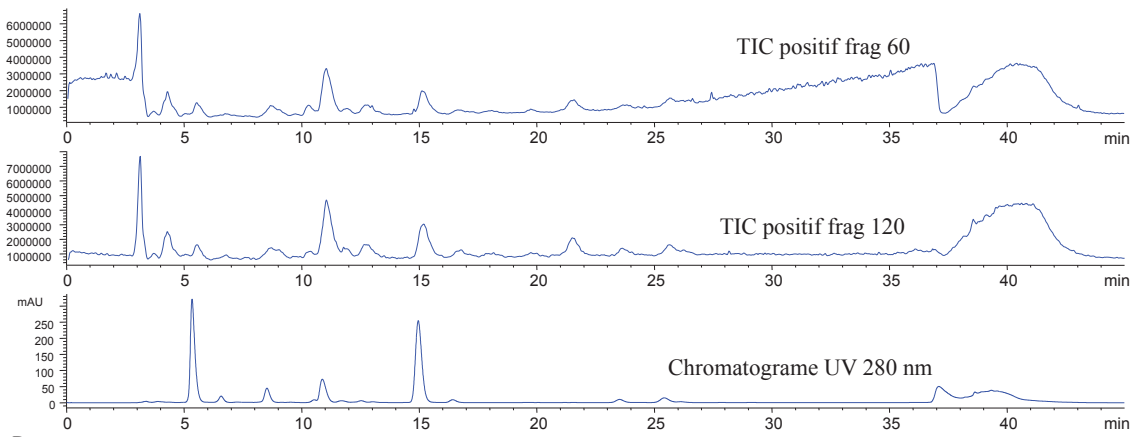
D. pseudoobscura



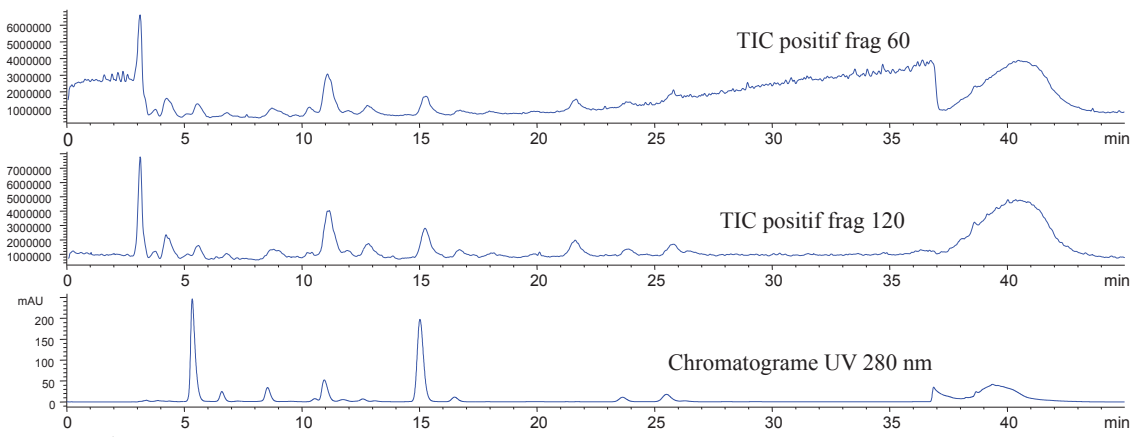
D. melanogaster



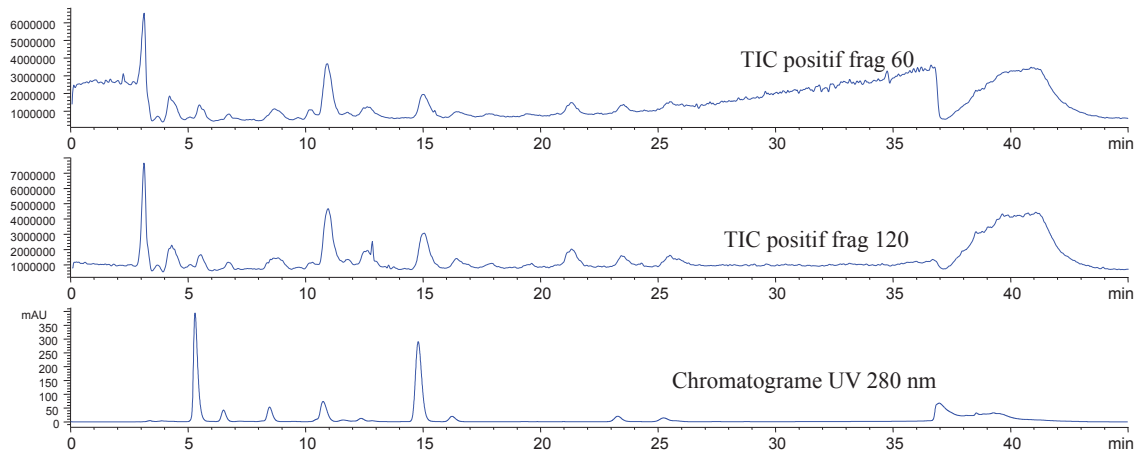
D. virilis



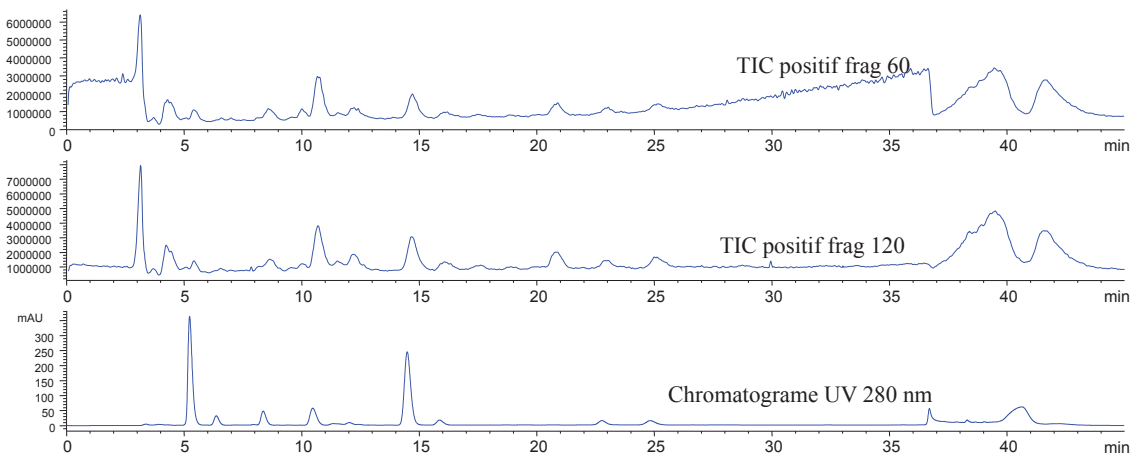
D. erecta



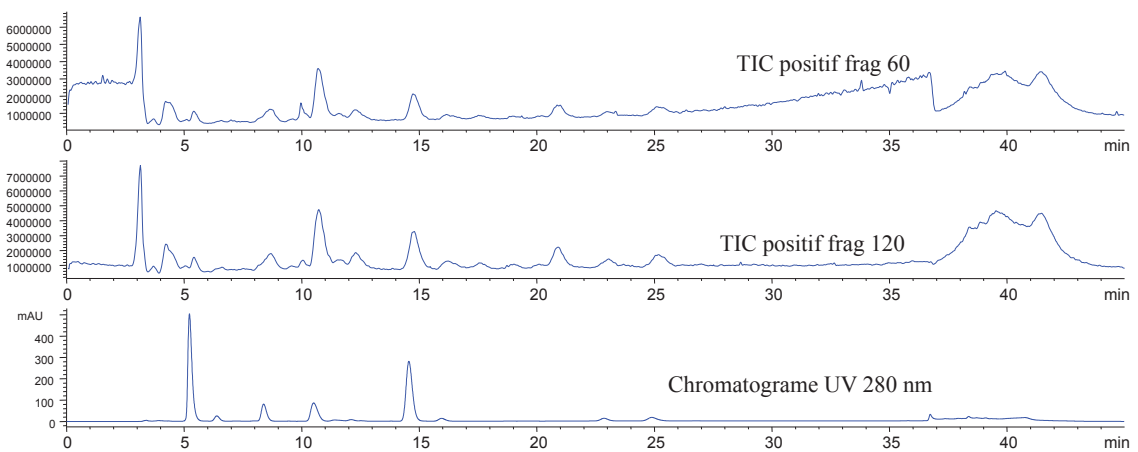
D. persimilis



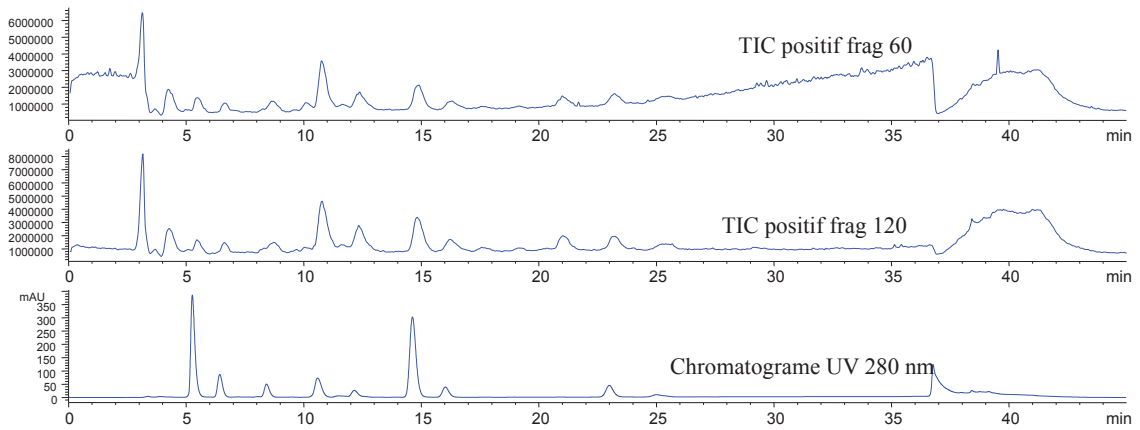
D. yakuba



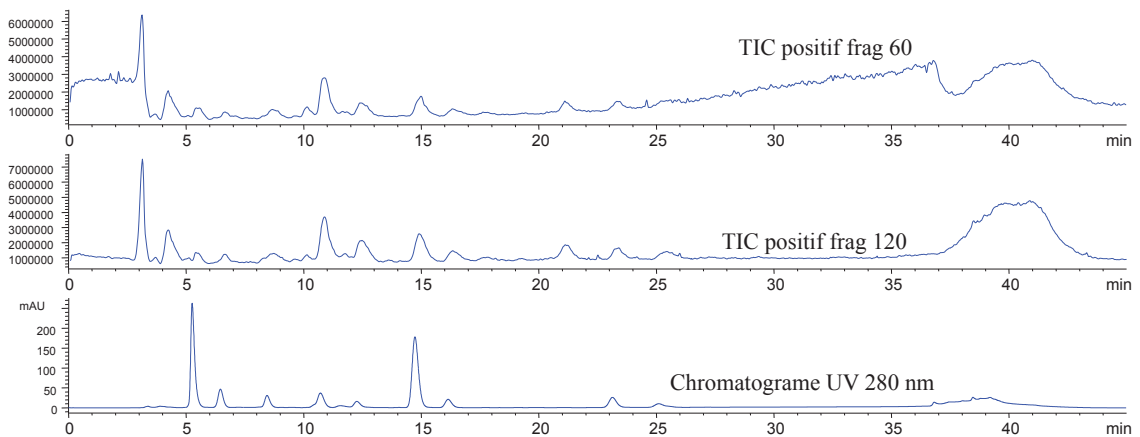
D. ananassae



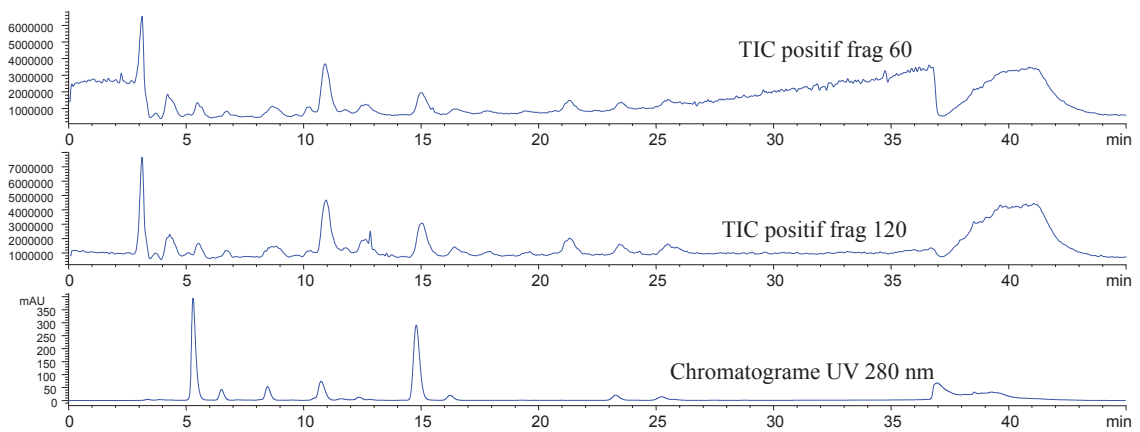
D. simulans



D. mojavensis



D. grimshawi



D. sechellia

BIBLIOGRAPHIE

- Arabidopsis Genome Consortium, 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, 408(6814), p.796-815.
- Aravin, A.A., Hannon, G.J. & Brennecke, J., 2007. The Piwi-piRNA Pathway Provides an Adaptive Defense in the Transposon Arms Race. *Science*, 318(5851), p.761-764.
- Azuara, V. et al., 2006. Chromatin signatures of pluripotent cell lines. *Nature Cell Biology*, 8(5), p.532-538.
- Berdasco, M., Fraga, M.F. & Esteller, M., 2009. Quantification of global DNA methylation by capillary electrophoresis and mass spectrometry. *Methods in Molecular Biology (Clifton, N.J.)*, 507, p.23-34.
- Berger, S.L., 2007. The complex language of chromatin regulation during transcription. *Nature*, 447(7143), p.407-412.
- Bernstein, B.E. et al., 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, 125(2), p.315-326.
- Bessereau, J., 2006. Transposons in *C. elegans*. *WormBook*.
- Bibikova, M. & Fan, J., 2010. Genome-wide DNA methylation profiling. *Wiley Interdisciplinary Reviews. Systems Biology and Medicine*, 2(2), p.210-223.
- Bourc'his, D. & Bestor, T.H., 2004. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature*, 431(7004), p.96-99.
- Brennecke, J. et al., 2008. An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science (New York, N.Y.)*, 322(5906), p.1387-1392.
- Brown, S.E. et al., 2007. Variations in DNA methylation patterns during the cell cycle of HeLa cells. *Epigenetics: Official Journal of the DNA Methylation Society*, 2(1), p.54-65.
- Brummel, T.J. et al., 1994. Characterization and relationship of Dpp receptors encoded by the saxophone and thick veins genes in *Drosophila*. *Cell*, 78(2), p.251-261.
- C. elegans* Sequencing Consortium, 1998. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science (New York, N.Y.)*, 282(5396), p.2012-2018.
- Cañizares, J. et al., 2000. Tirant is a new member of the gypsy family of retrotransposons in *Drosophila melanogaster*. *Genome / National Research Council Canada* 43(1), p.9-14.
- Cassidy, S.B., Dykens, E. & Williams, C.A., 2000. Prader-Willi and Angelman syndromes: sister imprinted disorders. *American Journal of Medical Genetics*, 97(2), p.136-146.
- Chambeyron, S. et al., 2008. piRNA-mediated nuclear accumulation of retrotransposon transcripts in the *Drosophila* female germline. *Proceedings of the National Academy*

- of Sciences of the United States of America*, 105(39), p.14964-14969.
- Chan, S.W., Henderson, I.R. & Jacobsen, S.E., 2005. Gardening The Genome: DNA Methylation in *Arabidopsis thaliana*. *Nat Rev Genet*, 6(7), p.590.
- Clark, A.G. et al., 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature*, 450(7167), p.203-218.
- Cokus, S.J. et al., 2008. Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature*, 452(7184), p.215-219.
- D'Ávila, M.F. et al., 2010. Sex-specific methylation in *Drosophila*: an investigation of the *Sophophora* subgenus. *Genetica*. Available at: <http://www.ncbi.nlm.nih.gov/gate1.inist.fr/pubmed/20640488> [Accédé Août 14, 2010].
- Desset, S. et al., 2008. In *Drosophila melanogaster* the COM locus directs the somatic silencing of two retrotransposons through both Piwi-dependent and -independent pathways. *PLoS One*, 3(2), p.e1526.
- Dowsett, A.P. & Young, M.W., 1982. Differing levels of dispersed repetitive DNA among closely related species of *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America*, 79(15), p.4570-4574.
- Drosophila 12 Genomes Consortium, 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature*, 450(7167), p.203-218.
- Eads, C.A. & Laird, P.W., 2002. Combined bisulfite restriction analysis (COBRA). *Methods in Molecular Biology (Clifton, N.J.)*, 200, p.71-85.
- Edwards, J.R. et al., 2010. Chromatin and sequence features that define the fine and gross structure of genomic methylation patterns. *Genome Research*.
- Elango, N. et al., 2009. DNA methylation is widespread and associated with differential gene expression in castes of the honeybee, *Apis mellifera*. *Proceedings of the National Academy of Sciences of the United States of America*, 106(27), p.11206-11211.
- Fablet, M. et al., 2009. Genomic environment influences the dynamics of the tirant LTR retrotransposon in *Drosophila*. *The FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology*, 23(5), p.1482-1489.
- Fablet, M. et al., 2006. Ongoing loss of the tirant transposable element in natural populations of *Drosophila simulans*. *Gene*, 375, p.54-62.
- Feschotte, C., 2008. Transposable elements and the evolution of regulatory networks. *Nature Reviews. Genetics*, 9(5), p.397-405.
- Field, L.M. et al., 2004. DNA methylation in insects. *Insect Molecular Biology*, 13(2), p.109-115.

- Finnegan, D.J., 1989. Eukaryotic transposable elements and genome evolution. *Trends in Genetics*, 5, p.103-107.
- Fraga, M.F. et al., 2005. Epigenetic differences arise during the lifetime of monozygotic twins. *Proceedings of the National Academy of Sciences of the United States of America*, 102(30), p.10604-10609.
- Fugmann, S.D., 2010. The origins of the Rag genes--from transposition to V(D)J recombination. *Seminars in Immunology*, 22(1), p.10-16.
- Garcia, R.N. et al., 2007. First evidence of methylation in the genome of *Drosophila willistoni*. *Genetica*, 131(1), p.91-105.
- Genereux, D.P. et al., 2008. Errors in the bisulfite conversion of DNA: modulating inappropriate- and failed-conversion frequencies. *Nucl. Acids Res.*, 36(22), p.e150.
- Girton, J.R. & Johansen, K.M., 2008a. Chromatin structure and the regulation of gene expression: the lessons of PEV in *Drosophila*. *Advances in Genetics*, 61, p.1-43.
- Girton, J.R. & Johansen, K.M., 2008. Chapter 1 Chromatin Structure and the Regulation of Gene Expression: The Lessons of PEV in *Drosophila*.
- Goffeau et al.1997. The yeast genome directory. *Nature*, 387(6632 Suppl), p.5.
- Goll, M.G. et al., 2006. Methylation of tRNA^{Asp} by the DNA methyltransferase homolog Dnmt2. *Science (New York, N.Y.)*, 311(5759), p.395-398.
- Goodman, M. et al., 1998. Toward a phylogenetic classification of Primates based on DNA evidence complemented by fossil evidence. *Molecular Phylogenetics and Evolution*, 9(3), p.585-598.
- Gou, D. et al., 2010. SETDB1 Is Involved in Postembryonic DNA Methylation and Gene Silencing in *Drosophila*. *PLoS ONE*, 5(5), p.e10581.
- Gruntman, E. et al., 2008. Kismeth: Analyzer of plant methylation states through bisulfite sequencing. *BMC Bioinformatics*, 9(1), p.371.
- Haniford, D.B., 2006. Transpososome dynamics and regulation in Tn10 transposition. *Critical Reviews in Biochemistry and Molecular Biology*, 41(6), p.407-424.
- Hayatsu, H., 2008. The bisulfite genomic sequencing used in the analysis of epigenetic states, a technique in the emerging environmental genotoxicology research. *Mutation Research*, 659(1-2), p.77-82.
- Henderson, I.R. et al., 2010. Accurate sodium bisulfite sequencing in plants. *Epigenetics : official journal of the DNA Methylation Society*, 5(1), p.47-49.
- Hermann, A., Schmitt, S. & Jeltsch, A., 2003. The human Dnmt2 has residual DNA-(cytosine-C5) methyltransferase activity. *The Journal of Biological Chemistry*, 278(34), p.31717-31721.

- Holliday, R. & Grigg, G., 1993. DNA methylation and mutation. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 285(1), p.61-67.
- Huang, Y. et al., 2010. The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PloS One*, 5(1), p.e8888.
- Jacinto, F.V., Ballestar, E. & Esteller, M., 2008. Methyl-DNA immunoprecipitation (MeDIP): hunting down the DNA methylome. *BioTechniques*, 44(1), p.35, 37, 39 passim.
- Jenuwein, T. & Allis, C.D., 2001. Translating the histone code. *Science (New York, N.Y.)*, 293(5532), p.1074-1080.
- Kapitonov, V.V. & Jurka, J., 2008. A universal classification of eukaryotic transposable elements implemented in Repbase. *Nat Rev Genet*, 9(5), p.411-412.
- Kazazian, H.H., 2004. Mobile elements: drivers of genome evolution. *Science (New York, N.Y.)*, 303(5664), p.1626-1632.
- Kok, R.M. et al., 2007. Global DNA methylation measured by liquid chromatography-tandem mass spectrometry: analytical technique, reference values and determinants in healthy subjects. *Clinical Chemistry and Laboratory Medicine: CCLM / FESCC*, 45(7), p.903-911.
- Kouzminova, E. & Selker, E.U., 2001. dim-2 encodes a DNA methyltransferase responsible for all known cytosine methylation in *Neurospora*. *EMBO J*, 20(15), p.4309-4323.
- Krauss, V., Eisenhardt, C. & Unger, T., 2009. The genome of the stick insect *Medauroidea extradentata* is strongly methylated within genes and repetitive DNA. *PloS One*, 4(9), p.e7223.
- Kriaucionis, S. & Heintz, N., 2009. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science (New York, N.Y.)*, 324(5929), p.929-930.
- Kucharski, R. et al., 2008. Nutritional control of reproductive status in honeybees via DNA methylation. *Science (New York, N.Y.)*, 319(5871), p.1827-1830.
- Kunert, N. et al., 2003. A Dnmt2-like protein mediates DNA methylation in *Drosophila*. *Development (Cambridge, England)*, 130(21), p.5083-5090.
- Kuramochi-Miyagawa, S. et al., 2008. DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes & Development*, 22(7), p.908-917.
- Lachaise, D. & Silvain, J., 2004. How two Afrotropical endemics made two cosmopolitan human commensals: the *Drosophila melanogaster*-*D. simulans* palaeogeographic riddle. *Genetica*, 120(1-3), p.17-39.

- Lander, E.S. et al., 2001. Initial sequencing and analysis of the human genome. *Nature*, 409(6822), p.860-921.
- Larkin, M.A. et al., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics (Oxford, England)*, 23(21), p.2947-2948.
- Le Rouzic, A., Dupas, S. & Capy, P., 2007. Genome ecosystem and transposable elements species. *Gene*, 390(1-2), p.214-220.
- Lee, Y.C.G. & Langley, C.H., 2010. Transposable elements in natural populations of *Drosophila melanogaster*. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 365(1544), p.1219-1228.
- Lesage, P. & Todeschini, A.L., 2005. Happy together: the life and times of Ty retrotransposons and their hosts. *Cytogenetic and Genome Research*, 110(1-4), p.70-90.
- Lin, M. et al., 2005. DNA methyltransferase gene dDnmt2 and longevity of *Drosophila*. *The Journal of Biological Chemistry*, 280(2), p.861-864.
- Lippman, Z. et al., 2004. Role of transposable elements in heterochromatin and epigenetic control. *Nature*, 430(6998), p.471-476.
- Lister, R. et al., 2008. Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell*, 133(3), p.523-536.
- Lyko, F., Ramsahoye, B.H. & Jaenisch, R., 2000. DNA methylation in *Drosophila melanogaster*. *Nature*, 408(6812), p.538-540.
- Lyko, F., Whittaker, A.J. et al., 2000. The putative *Drosophila* methyltransferase gene dDnmt2 is contained in a transposon-like element and is expressed specifically in ovaries. *Mechanisms of Development*, 95(1-2), p.215-217.
- Malone, C.D. et al., 2009. Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell*, 137(3), p.522-535.
- Marhold J.[1] et al., 2004. Conservation of DNA methylation in dipteran insects. *Insect Molecular Biology*, 13, p.117-123.
- Marsano, R.M. et al., 2000. The complete Tirant transposable element in *Drosophila melanogaster* shows a structural relationship with retrovirus-like retrotransposons. *Gene*, 247(1-2), p.87-95.
- Mathieu, O. & Bender, J., 2004. RNA-directed DNA methylation. *Journal of Cell Science*, 117(Pt 21), p.4881-4888.
- Matzke, M. et al., 2007. Targets of RNA-directed DNA methylation. *Current Opinion in Plant Biology*, 10(5), p.512-519.

- Miki, Y. et al., 1996. Mutation analysis in the BRCA2 gene in primary breast cancers. *Nature Genetics*, 13(2), p.245-247.
- Miura, A. et al., 2001. Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis. *Nature*, 411(6834), p.212-214.
- Morgan, H.D. et al., 2005. Epigenetic reprogramming in mammals. *Human Molecular Genetics*, 14 Spec No 1, p.R47-58.
- Olivieri, D. et al., 2010. An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in Drosophila. *The EMBO Journal*, 29(19), p.3301-3317.
- Pardue, M. et al., 2005. Two retrotransposons maintain telomeres in Drosophila. *Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology*, 13(5), p.443-453.
- Parks, A.R. & Peters, J.E., 2009. Tn7 elements: engendering diversity from chromosomes to episomes. *Plasmid*, 61(1), p.1-14.
- Pasyukova, E.G., Nuzhdin, S.V. & Filatov, D.A., 1998. The relationship between the rate of transposition and transposable element copy number for copia and Doc retrotransposons of Drosophila melanogaster. *Genetical Research*, 72(1), p.1-11.
- Pavlopoulou, A. & Kossida, S., 2009. Phylogenetic analysis of the eukaryotic RNA (cytosine-5)-methyltransferases. *Genomics*, 93(4), p.350-357.
- Peng, J.C. & Karpen, G.H., 2009. Heterochromatic genome stability requires regulators of histone H3 K9 methylation. *PLoS Genetics*, 5(3), p.e1000435.
- Peng, J.C. & Karpen, G.H., 2007. H3K9 methylation and RNA interference regulate nucleolar organization and repeated DNA stability. *Nat Cell Biol*, 9(1), p.25-35.
- Phalke, S. et al., 2009. Retrotransposon silencing and telomere integrity in somatic cells of Drosophila depends on the cytosine-5 methyltransferase DNMT2. *Nature Genetics*, 41(6), p.696-702.
- Picot, S. et al., 2008. The mariner transposable element in natural populations of Drosophila simulans. *Heredity*, 101(1), p.53-59.
- Pomraning, K.R., Smith, K.M. & Freitag, M., 2009. Genome-wide high throughput analysis of DNA methylation in eukaryotes. *Methods (San Diego, Calif.)*, 47(3), p.142-150.
- Ponger, L. & Li, W., 2005. Evolutionary diversification of DNA methyltransferases in eukaryotic genomes. *Molecular Biology and Evolution*, 22(4), p.1119-1128.
- Rita Rebollo, Béatrice Horard, Benjamin Hubert, Cristina Vieira, 2010. Jumping genes and epigenetics: Towards new species. *Gene*.

- Reik, W., Dean, W. & Walter, J., 2001. Epigenetic reprogramming in mammalian development. *Science (New York, N.Y.)*, 293(5532), p.1089-1093.
- Rein, T. et al., 1997. Absence of an Unusual “Densely Methylated Island” at the Hamster dhfr ori- β . *Journal of Biological Chemistry*, 272(15), p.10021 -10029.
- Rozhkov, N.V. et al., 2010. Small RNA-based silencing strategies for transposons in the process of invading *Drosophila* species. *RNA (New York, N.Y.)*, 16(8), p.1634-1645.
- Saito, K. et al., 2006. Specific association of Piwi with rasiRNAs derived from retrotransposon and heterochromatic regions in the *Drosophila* genome. *Genes & Development*, 20(16), p.2214-2222.
- Salzberg, A. et al., 2004. Identification of methylated sequences in genomic DNA of adult *Drosophila melanogaster*. *Biochemical and Biophysical Research Communications*, 322(2), p.465-469.
- Sandhu, J. et al., 2009. Determination of 5-methyl-2'-deoxycytidine in genomic DNA using high performance liquid chromatography-ultraviolet detection. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences*, 877(20-21), p.1957-1961.
- Schaefer, M. & Lyko, F., 2010. Solving the Dnmt2 enigma. *Chromosoma*, 119(1), p.35-40.
- Schnable, P.S. et al., 2009. The B73 maize genome: complexity, diversity, and dynamics. *Science (New York, N.Y.)*, 326(5956), p.1112-1115.
- Schotta, G. et al., 2002. Central role of *Drosophila* SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. *The EMBO Journal*, 21(5), p.1121-1131.
- Schotta, G. et al., 2004. A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes & Development*, 18(11), p.1251-1262.
- Schubert, D. et al., 2006. Silencing by plant Polycomb-group genes requires dispersed trimethylation of histone H3 at lysine 27. *The EMBO Journal*, 25(19), p.4638-4649.
- Selker, E.U., Fritz, D.Y. & Singer, M.J., 1993. Dense nonsymmetrical DNA methylation resulting from repeat-induced point mutation in *Neurospora*. *Science (New York, N.Y.)*, 262(5140), p.1724-1728.
- Selker, E.U. et al., 2003. The methylated component of the *Neurospora crassa* genome. *Nature*, 422(6934), p.893-897.
- Siomi, H. & Siomi, M.C., 2009. On the road to reading the RNA-interference code. *Nature*, 457(7228), p.396-404.
- Slotkin, R.K. et al., 2009. Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell*, 136(3), p.461-472.

- Stach, D. et al., 2003. Capillary electrophoretic analysis of genomic DNA methylation levels. *Nucleic Acids Research*, 31(2), p.E2.
- Stoye, J.P., 2009. Proviral protein provides placental function. *Proceedings of the National Academy of Sciences*, 106(29), p.11827-11828.
- Strahl, B.D. & Allis, C.D., 2000. The language of covalent histone modifications. *Nature*, 403(6765), p.41-45.
- Suganuma, T. & Workman, J.L., 2010. WD40 Repeats Arrange Histone Tails for Spreading of Silencing. *Journal of Molecular Cell Biology*, 2(2), p.81 -83.
- Sun, H. et al., 2010. Genome-wide mapping of RNA Pol-II promoter usage in mouse tissues by ChIP-seq. *Nucleic Acids Research*. Available at: <http://www.ncbi.nlm.nih.gov/gate1.inist.fr/pubmed/20843783> [Accédé Octobre 26, 2010].
- Suzuki, M.M. & Bird, A., 2008. DNA methylation landscapes: provocative insights from epigenomics. *Nature Reviews. Genetics*, 9(6), p.465-476.
- Tahiliani, M. et al., 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science (New York, N.Y.)*, 324(5929), p.930-935.
- Tasheva, E.S. & Roufa, D.J., 1995. A densely methylated DNA island is associated with a chromosomal replication origin in the human RPS14 locus. *Somatic Cell and Molecular Genetics*, 21(6), p.369-383.
- Tasheva, E.S. & Roufa, D.J., 1994. Densely methylated DNA islands in mammalian chromosomal replication origins. *Molecular and Cellular Biology*, 14(9), p.5636-5644.
- Terracol, R. & Lengyel, J.A., 1994. The thick veins gene of *Drosophila* is required for dorsoventral polarity of the embryo. *Genetics*, 138(1), p.165-178.
- Vagin, V.V. et al., 2006. A distinct small RNA pathway silences selfish genetic elements in the germline. *Science (New York, N.Y.)*, 313(5785), p.320-324.
- Valinluck, V. et al., 2004. Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). *Nucleic Acids Research*, 32(14), p.4100-4108.
- Vaughn, M.W. et al., 2007. Epigenetic natural variation in *Arabidopsis thaliana*. *PLoS Biology*, 5(7), p.e174.
- Venner, S., Feschotte, C. & Biéumont, C., 2009. Dynamics of transposable elements: towards a community ecology of the genome. *Trends in Genetics: TIG*, 25(7), p.317-323.

- Vieira, C. & Biémont, C., 1997. Transposition rate of the 412 retrotransposable element is independent of copy number in natural populations of *Drosophila simulans*. *Molecular Biology and Evolution*, 14(2), p.185-188.
- Vieira, C. et al., 1999. Wake up of transposable elements following *Drosophila simulans* worldwide colonization. *Molecular Biology and Evolution*, 16(9), p.1251-1255.
- Vieira, C. & Biémont, C., 2004. Transposable element dynamics in two sibling species: *Drosophila melanogaster* and *Drosophila simulans*. *Genetica*, 120(1-3), p.115-123.
- Vincze, T., Posfai, J. & Roberts, R.J., 2003. NEBcutter: a program to cleave DNA with restriction enzymes. *Nucleic Acids Research*, 31(13), p.3688-3691.
- Volff, J., 2006. Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 28(9), p.913-922.
- Wang, M.B. et al., 2001. Replicating satellite RNA induces sequence-specific DNA methylation and truncated transcripts in plants. *RNA*, 7(1), p.16-28.
- Wang, Y. et al., 2006. Functional CpG methylation system in a social insect. *Science (New York, N.Y.)*, 314(5799), p.645-647.
- Warnecke, P.M. et al., 2002. Identification and resolution of artifacts in bisulfite sequencing. *Methods (San Diego, Calif.)*, 27(2), p.101-107.
- Waterston, R.H. et al., 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature*, 420(6915), p.520-562.
- Weiler, K.S. & Wakimoto, B.T., 1995. Heterochromatin and gene expression in *Drosophila*. *Annual Review of Genetics*, 29, p.577-605.
- Werren, J.H. et al., 2010. Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science (New York, N.Y.)*, 327(5963), p.343-348.
- Wicker, T. et al., 2007. A unified classification system for eukaryotic transposable elements. *Nature Reviews. Genetics*, 8(12), p.973-982.
- Wojdacz, T.K., Hansen, L.L. & Dobrovic, A., 2008. A new approach to primer design for the control of PCR bias in methylation studies. *BMC Research Notes*, 1, p.54.
- Xiang, H. et al., 2010. Single base-resolution methylome of the silkworm reveals a sparse epigenomic map. *Nature Biotechnology*, 28(5), p.516-520.
- Zemach, A. et al., 2010. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science (New York, N.Y.)*, 328(5980), p.916-919.
- Zhang, X. et al., 2006. Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell*, 126(6), p.1189-1201.

Zilberman, D. & Henikoff, S., 2007. Genome-wide analysis of DNA methylation patterns. *Development (Cambridge, England)*, 134(22), p.3959-3965.