



**HAL**  
open science

# Contribution des polymorphismes d'insertions à la stérilité des hybrides chez *Paramecium tetraurelia*

Guillaume Pellerin

► **To cite this version:**

Guillaume Pellerin. Contribution des polymorphismes d'insertions à la stérilité des hybrides chez *Paramecium tetraurelia*. Génétique. Université Pierre et Marie Curie - Paris VI, 2017. Français. NNT : 2017PA066078 . tel-01613046

**HAL Id: tel-01613046**

**<https://theses.hal.science/tel-01613046>**

Submitted on 9 Oct 2017

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

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

Université Pierre et Marie Curie

Ecole doctorale 515 - Complexité du Vivant

**Contribution des polymorphismes  
d'insertions à la stérilité des hybrides  
chez *Paramecium tetraurelia***

Par Guillaume Pellerin

Thèse de doctorat de Biologie

Dirigée par Eric Meyer

**Institut de Biologie de l'Ecole Normale Supérieure**

Présentée et soutenue publiquement le 31/03/2017

Devant un jury composé de :

M <sup>me</sup> BETERMIER Mireille	DR - I2BC, Gif-sur-Yvette	Rapporteur
M. PELISSON Alain	DR émérite - IGH, Montpellier	Rapporteur
M. COLOT Vincent	DR - IBENS, Paris	Examineur
M. HIGUET Dominique	PU - IBPS, Paris	Examineur
M. MEYER Eric	DR - IBENS, Paris	Directeur de thèse



# Remerciements

En premier lieu, je dois remercier l'homme sans qui cette thèse n'aurait jamais vu le jour, Eric Meyer. C'est toi, le scientifique chevronné, qui a su attiser ma curiosité d'étudiant en me vendant le modèle du *genome scanning* chez la paramécie. Tu m'as ensuite offert l'opportunité de travailler dans ton équipe à l'IBENS, pour mon stage puis ma thèse. Toujours disponible pour parler de science, je reste fasciné par ta capacité à raisonner et à expliquer en peu de mots, par oral ou par écrit, des idées souvent complexes. Tu resteras un modèle d'éloquence pour moi.

Ensuite, je souhaite remercier toutes les personnes qui ont contribué à ma formation à la paillasse. Deepankar P. Singh m'a appris toutes les bases de la culture de la paramécie. Je lui dois l'apprentissage des techniques avec toutes les astuces qui permettent de gagner du temps. C'est aussi lui qui m'a appris à anticiper tous les aléas d'une expérience afin d'être sûr d'aller jusqu'au bout. Je remercie Quentin Carradec pour m'avoir appris à injecter et à utiliser R. Merci à Simran Bhullar pour tous ses conseils et ses astuces qui m'ont permis d'avancer. Je suis également reconnaissant envers Véronique Tanty pour l'organisation du labo, les chouquettes et les cookies, et pour son aide quotidienne dans mes expériences. J'espère qu'elle ne m'enverra pas de facture pour tous les services qu'elle m'a rendus. Enfin, je remercie Sophie Malinsky, Ayush Srivastava et toutes les autres personnes qui ont travaillé dans l'équipe pendant ma thèse. Travailler et discuter avec eux a enrichi ma culture scientifique et générale.

Pour continuer, je tiens à remercier toutes les personnes des équipes de Linda Sperling, Mireille Bétermier et Sandra Duharcourt. Nos réunions « paramécie » m'ont beaucoup aidé à assimiler la complexité de notre organisme préféré. Merci pour ces moments de réflexions et pour vos retours sur mon travail.

Merci également à tous les membres de la section Génomique Fonctionnelle, en particulier ceux que je croise régulièrement dans les couloirs ou à côté de la machine à café. J'y inclus les services de gestion et de laverie qui nous apportent une aide remarquable.

Enfin, ma thèse n'aurait pas été si plaisante je n'avais pas pu m'en plaindre auprès de mes amis et de ma femme. Merci à eux de m'avoir écouté et soutenu dans les moments pénibles.

# Table des matières

<b>Remerciements</b>	<b>iii</b>
<b>Table des matières</b>	<b>iv</b>
<b>Abréviations et conventions d'écriture</b>	<b>viii</b>
<b>INTRODUCTION</b> .....	<b>2</b>
<b>I – Les transposons et leur neutralisation</b> .....	<b>3</b>
<b>A. Les transposons</b> .....	<b>3</b>
1) Découverte.....	3
2) Les classes de transposons.....	3
3) Diversité et abondance.....	5
4) Les transposons parasitent le génome. ....	6
5) D'un génome à l'autre .....	7
<b>B. L'ARN interférence</b> .....	<b>7</b>
1) Principe.....	7
2) Biogenèse des siARN .....	9
3) Les protéines Argonautes .....	9
4) Le complexe RISC .....	10
5) La répression des transposons.....	10
<b>C. Quelques systèmes de défense des génomes chez les unicellulaires</b> .	<b>11</b>
1) Comment distinguer les transposons étrangers de son propre ADN ? .....	11
2) Les systèmes procaryotes.....	12
3) Les systèmes chez <i>Neurospora crassa</i> (MSUD, quelling, RIP).....	13
<b>D. Les piARN des métazoaires</b> .....	<b>17</b>
1) Caractéristiques des piARN .....	17
2) La voie piARN de <i>Drosophila melanogaster</i> .....	18
3) La voie piARN de <i>Caenorhabditis elegans</i> .....	20
4) Hérité transgénérationnelle des piARN .....	22
<b>II – La dysgenèse hybride</b> .....	<b>24</b>
<b>A. Le système P-M</b> .....	<b>24</b>
1) Etudes génétiques historiques .....	24
2) Les bases moléculaires de la répression maternelle .....	27
3) Adaptation des F1 dysgéniques à l'élément P.....	29
<b>B. L'isolement reproductif</b> .....	<b>30</b>
1) Généralités.....	30
2) Croisements intraspécifiques .....	31
3) Croisements interspécifiques.....	34
<b>III – Les ciliés et leurs réarrangements génomiques</b> .....	<b>40</b>
<b>A. Généralités sur les ciliés</b> .....	<b>40</b>

1) Position phylogénétique .....	40
2) Dimorphisme nucléaire et cycle sexuel .....	42
3) Les génomes des ciliés .....	43
<b>B. Méthodes d'études chez <i>Paramecium tetraurelia</i> .....</b>	<b>44</b>
1) Cultures .....	44
2) Génétique classique .....	45
3) Génétique inverse .....	47
<b>C. Les réarrangements du génome .....</b>	<b>48</b>
1) La fragmentation des chromosomes .....	49
2) L'élimination interne d'ADN .....	50
3) <i>DNA unscrambling</i> .....	52
4) Les transposases .....	52
<b>D. Le contrôle épigénétique des réarrangements chez <i>Paramecium</i> et <i>Tetrahymena</i> .....</b>	<b>53</b>
1) Approche historique .....	53
2) Principe du <i>genome scanning</i> .....	54
3) Les petits ARN impliqués dans le modèle .....	55
4) la sélection MIC-spécifique dans le MAC maternel .....	58
5) L'élimination guidée par les scnARN des IES .....	59
6) Le rôle possible de la méthylation des adénines .....	61
<b>E. Les apports du modèle <i>Oxytricha trifallax</i> .....</b>	<b>61</b>
1) Otiwi et ses piARN .....	61
2) La méthylation des cytosines .....	62
<b>IV – La spéciation du complexe <i>P. aurelia</i> .....</b>	<b>62</b>
<b>A. Duplications du génome et spéciation .....</b>	<b>62</b>
1) L'histoire du génome des <i>P. aurelia</i> .....	62
2) Devenir des gènes dupliqués .....	63
3) L'hypothèse de Bateson - Dobzansky - Muller .....	64
<b>B. IES et dysgenèse hybride .....</b>	<b>65</b>
1) Histoire évolutive des IES .....	65
2) Prédiction des problèmes d'excision en F1 .....	66
3) Létalité en F2 .....	69
4) Données préliminaires .....	70
5) Projet de thèse .....	73
<b>RESULTATS .....</b>	<b>74</b>
<b>V - Effet des polymorphismes d'insertion d'IES avec l'exemple de mtB .....</b>	<b>75</b>
<b>A. Les polymorphismes d'insertion .....</b>	<b>75</b>
<b>B. L'IES de mtB est contrôlée maternellement .....</b>	<b>75</b>
<b>C. L'IES de mtB est partiellement retenu en F1 .....</b>	<b>76</b>
1) Analyse des taux de rétention en F1 .....	76
2) Comparaison des deux séries .....	77

3) Test d'un biais PCR .....	79
4) Corrélation caryonidale .....	80
5) Taux de rétention et phénotype O* .....	81
<b>D. Les scnARN paternels peuvent exciser l'IES de mtB chez le partenaire naïf. ....</b>	<b>82</b>
1) Croisement avec déplétion de Ptiwi01&09 .....	82
2) Croisement avec la souche mtB+ .....	83
<b>E. L'IES 101 est davantage retenu chez le F1 dérivant du parent naïf... 84</b>	<b>84</b>
<b>VI – Effet des divergences d'IES avec l'exemple de G4404 .... 86</b>	<b>86</b>
<b>A. Les IES 51G4404 et 32G4404 sont contrôlées maternellement. .... 86</b>	<b>86</b>
<b>B. les scnARN ciblant l'IES G4404 sont spécifiques à chaque allèle. ... 87</b>	<b>87</b>
<b>C. les allèles divergents de G4404 sont convenablement excisés chez les F1 51 x 32. .... 89</b>	<b>89</b>
<b>D. Les scnARN paternels assurent l'excision quasi parfaite des IES... 89</b>	<b>89</b>
1) Croisement avec déplétion de Ptiwi01&09 .....	89
2) Croisement 51-G4404+ x 32 .....	91
<b>VII – Mode d'action des scnARN paternels.....93</b>	<b>93</b>
<b>A. Un nouveau modèle .....</b>	<b>93</b>
1) Deux hypothèses.....	93
2) Le paradoxe apparent du type sexuel .....	94
<b>B. Expériences d'injection simultanée des 2 allèles G4404 .....</b>	<b>95</b>
<b>C. Etude de la localisation des protéines Ptiwi .....</b>	<b>97</b>
1) Etude de la localisation de la protéine Ptiwi03 .....	97
2) Etude de la localisation de la protéine Ptiwi01 .....	99
3) Blocage de Ptiwi01 par Nowa1 .....	103
<b>D. Lien avec la méthylation de l'ADN .....</b>	<b>105</b>
1) Méthylation des cytosines .....	105
2) Méthylation des adénines .....	106
<b>E. Excision des allèles identiques dans le croisement 51 x 32 avec extinction des gènes Dcl2/3. .... 109</b>	<b>109</b>
1) Hypothèse .....	109
2) Expérience.....	110
<b>VIII – Avancées sur le modèle du scanning du génome .....</b>	<b>112</b>
<b>A. Existe-t-il une boucle d'amplification ?..... 112</b>	<b>112</b>
1) Réflexions .....	112
2) Etude de la rétention de l'IES de mtB dans des caryonides individuels.....	113
3) Etude de la voie Dcl5.....	114
<b>B. Etude cinétique d'une conjugaison hautement synchronisée..... 117</b>	<b>117</b>
1) Pourquoi et comment synchroniser ? .....	117
2) Tentative de synchronisation avec la WGA.....	117
3) Synchronisation avec des billes magnétiques .....	118
4) Résultats.....	122

<b>MATERIEL &amp; METHODES.....</b>	<b>124</b>
<b>A. Techniques de biologie cellulaire .....</b>	<b>125</b>
<b>B. Techniques de biologie moléculaire .....</b>	<b>127</b>
<b>C. Analyse informatique .....</b>	<b>127</b>
<b>DISCUSSION .....</b>	<b>130</b>
<b>A. Etendue des défauts de réarrangements.....</b>	<b>131</b>
1) F1 51/32 hémizygotes pour des IES .....	131
2) F1 d48/d12 hémizygote pour le gène A .....	131
3) F1 51/32 hétérozygotes pour des IES divergentes.....	132
4) Etude à l'échelle du génome. ....	132
<b>B. Mécanisme des défauts de réarrangements .....</b>	<b>133</b>
1) Une quantité limitée de scnARN ? .....	133
2) Détection de l'état hémizygote par appariement des chromosomes .....	133
3) Détection de l'état hémizygote par des ARN .....	134
4) Mécanisme d'inhibition de l'excision .....	134
5) Implication défaut de scnARN maternels .....	135
<b>C. Un modèle avec des scnARN paternels actifs .....</b>	<b>135</b>
1) Un échange de cytoplasme ? .....	135
2) Une empreinte gamétique ?.....	136
<b>D. Que reste-t-il de la dysgenèse hybride ?.....</b>	<b>136</b>
1) Défauts d'excision des IES et létalité .....	136
2) Et les transposons ? .....	137
3) Dysgenèse suite à une invasion de transposon chez Paramecium.....	137
<b>ANNEXES.....</b>	<b>138</b>
<b>BIBLIOGRAPHIE .....</b>	<b>154</b>

# Abréviations et conventions d'écriture

Écriture des gènes : *DCL2, PTIWI, PGM, TWI1, DCL1...*

Écriture des protéines : Dcl2, Ptiwi, Pgm ou Twi1p, Dcl1p... (avec un p à la fin chez *Tetrahymena*)

Écriture d'une langue étrangère : *Dicer-like, genome scanning, de novo, anlage*

Notation des petits ARN en fonction de leur taille et de la nature du 1er nucléotide: 21U, 22G...

ADN : Acide désoxyribonucléique

Ago : Argonaute

ARN : Acide ribonucléique

ARNi : ARN interférence

ARNm : ARN messenger

BET : Bromure d'éthidium

*cphg* : Copies par génome haploïde

CRISPR : *Clustered Regularly Interspaced Short Palindromic Repeats*

CSR-1 : *Chromosome Segregation and RNAi déficient 1*

DCL : *Dicer-like*

ET : Élément Transposable

EZL : *Enhancer of Zeste-Like*

GFP : *Green Fluorescent Protein*

HP1 : *Heterochromatin Protein 1*

HRDE-1 : *Heritable RNAi Deficient 1*

IES : *Internal Eliminated Sequence*

iesARN : *IES-ARN*

kb : Kilobase(s)

LTR : *Long Terminal Repeat*

MAC : Macronoyau (nom) ou Macronucléaire (adjectif) selon le contexte

MIC : Micronoyau (nom) ou Micronucléaire (adjectif) selon le contexte

MSUD : *Meiotic Silencing of Unpaired DNA*

ND : *Non Discharge*

nt : nucléotide(s)

OTIWI : *Oxytricha trifallax* PIWI

PAF : Paraformaldéhyde

PAZ : *Piwi Argonaut and Zwillle*

pb : Paires de bases

PCR : *Polymerase Chain Reaction*

PGM : *PiggyMac*

PIWI : *P-element Induced Wimpy Testis*

piARN : PIWI Interacting ARN

PRG-1 : *PIWI-related gene 1*

PTIWI : *Paramecium tetraurelia* PIWI

PolII : ARN Polymérase II

RIP : *Repeat-Induced Point mutation*

RISC : *RNA-Induced Silencing Complex*

scnARN : *scan ARN*

siARN : *small interfering ARN*

TIR : *Terminal Inverted Repeats*

TWI : *Tetrahymena thermophila* PIWI

X-TAS : *X-Telomere Associated Sequence*

WAGO : *Worm Argonaute*

WGA : *Wheat Germ Agglutinin*

WGD : *Whole Genome Duplication*

WGP : *Wheat Grass Powder*



# **INTRODUCTION**

# I – Les transposons et leur neutralisation

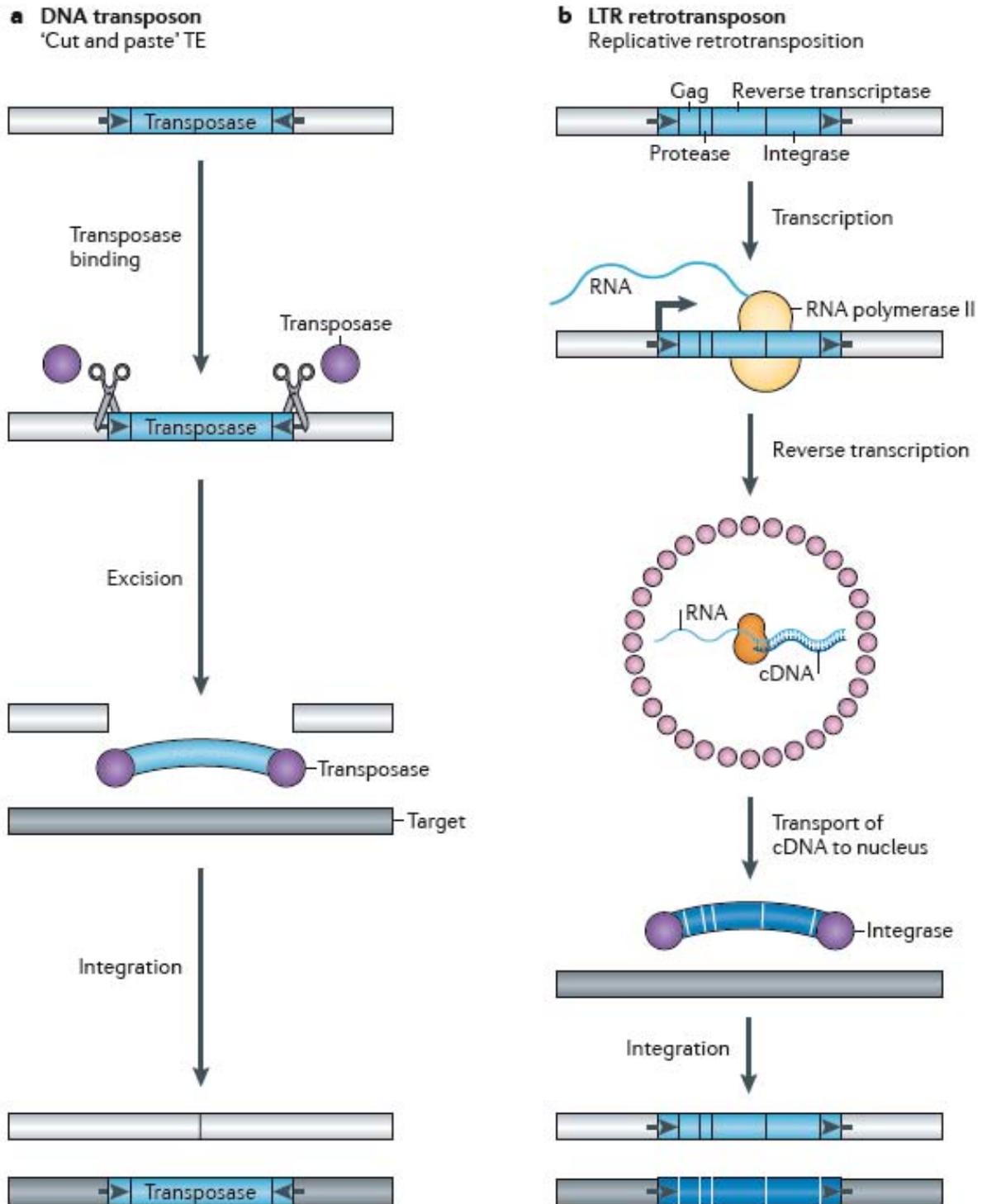
## A. Les transposons

### 1) Découverte

Les éléments transposables (ET) sont des séquences d'ADN capables de se déplacer d'une position à une autre dans un génome. De fait, ils génèrent des mutations par insertion dont certaines se traduisent par un caractère observable. C'est par ce biais que Barbara McClintock a découvert des phénotypes instables chez le maïs ainsi que des réarrangements chromosomiques (McClintock, 1950). Ses travaux l'ont amenés à postuler l'existence du premier élément génétique mobile, le locus *Dissociator* (*Ds*). La transposition qu'elle observait n'était possible qu'en présence d'un autre locus, qu'elle appela *Activator* (*Ac*). Plus tard, il fut établi que le locus *Ds* est un élément dérivé du locus *Ac* ayant perdu sa capacité propre de transposition (Fedoroff et al., 1983). Ce n'est qu'après la caractérisation de transposons dans les organismes modèles (*E. coli*, levure, drosophile, homme) (Behnke, 1990 pour une revue) qu'elle fut récompensée par un prix Nobel en 1983.

### 2) Les classes de transposons

Grâce au séquençage, la nature précise des transposons est aujourd'hui connue et on les répartit en deux classes. La première désigne les rétrotransposons qui utilisent un intermédiaire ARN pour se dupliquer puis s'insérer dans un autre locus. Les éléments de classe II sont les ET à ADN dont la majorité se déplace par un processus de « couper-coller ». Dans ce cas, ils quittent leur position initiale avant de s'intégrer dans un nouveau locus. Ils peuvent néanmoins augmenter leur nombre de copies, par exemple si un élément qui vient d'être répliqué par la machinerie cellulaire transpose vers un locus en aval d'une fourche de réplication. Un mécanisme de mobilisation est présenté pour chaque classe en **figure I.1**.



**Figure I.1 : Mécanismes de mobilisation des transposons.**

(a) La plupart des transposons à ADN possèdent des répétitions terminales inversées (TIR : flèches noires), codent un transposase (disque violet) et se déplacent par « couper-coller ». La transposase se lie aux TIR, excise l'élément de son locus initial (gris clair) puis cible un nouveau locus (gris foncé) dans lequel elle intègre l'élément. La cassure double-brin du site donneur est réparée par recombinaison homologue ou par la voie NHEJ (*Non-Homologous End Joining*). Le site ciblé par l'intégration est typiquement dupliqué sur 4 à 8 pb au cours du processus. (b) Les rétrotransposons se mobilisent par un mécanisme répliatif impliquant un

intermédiaire à ARN. Les rétrotransposons à LTR portent deux longues répétitions terminales (LTR, flèches noires) et codent des protéines ayant des activités de Gag, protéase, reverse transcriptase et intégrase qui sont toutes requises pour la rétrotransposition. Le LTR en 5' contient un promoteur recrutant l'ARN polymérase II (PolII) qui produit l'ARNm de l'ET. Les protéines Gag (disques roses) forment des particules de type viral contenant l'ARNm (bleu clair), la transcriptase inverse (orange) et l'intégrase (disque violet). La transcriptase convertit l'ARNm en ADN complémentaire (ADNc, bleu foncé) qui est pris en charge par l'intégrase afin de l'intégrer dans un nouveau locus. Le processus duplique également le site d'insertion. (Levin and Moran, 2011)

### 3) Diversité et abondance

Afin de mieux appréhender la diversité des transposons, des classifications sont proposées et s'appuient sur différents critères. Dans la classification de Wicker et al. (**Figure I.2**) (Wicker et al., 2007), les rétrotransposons sont divisés en ordres qui partagent une stratégie de réplique commune. Le mécanisme de transposition des éléments à LTR est détaillé dans la figure I.1b. Les membres du groupe DIRS (*Dictyostelium Intermediate Repeat Sequence*) utilisent une tyrosine recombinase au lieu d'une intégrase alors que les PLE (Penelope-like Elements) ont une étape de rétrotranscription proche d'une activité télomérase. Enfin, les SINE (*Short Interspersed Nuclear Element*) rassemblent des ET non autonomes qui dépendent du dernier ordre, les LINE (*Long Interspersed Nuclear Element*) pour transposer. Les ET à ADN sont d'abord divisés en sous-classes en fonction du mécanisme d'excision. S'il implique les deux brins, comme dans le mécanisme présenté avec les TIR, alors le transposon tombera dans la sous-classe 1. En revanche, si un seul brin est clivé, ce sera la sous-classe 2 (Hélitrons et Mavericks). A l'intérieur des ordres, les ET sont subdivisés en superfamilles en fonction des gènes présents et de la taille du site d'insertion. Le niveau inférieur (familles) repose sur la séquence d'ADN en elle-même.

Les ET sont présents dans tous les règnes du vivant. Ils sont mieux connus et très abondants chez les eucaryotes où ils représentent plus de 50 % du génome chez les mammifères et plus de 70 % chez certaines plantes. Par ailleurs, on en trouve chez les bactéries (Casjens, 2003), les archées (Makarova et al., 2014) et même chez les virus géants (Sun et al., 2015). Dans un génome donné, on trouve généralement de nombreuses copies d'un même élément. La plupart des copies sont inactives, c'est-à-dire qu'elles ne sont plus capables de catalyser leur propre transposition. En revanche, elles peuvent rester mobilisables tant qu'elles sont reconnues par la machinerie de transposition d'autres éléments du génome. En pratique, il est difficile de savoir si une copie est active à partir de la simple séquence car on ne sait pas quelles sont les conséquences des substitutions par rapport aux éléments actifs connus.

Classification	Structure	TSD	Code	Occurrence	
Order	Superfamily				
<b>Class I (retrotransposons)</b>					
LTR	Copia	→ GAG AP INT RT RH →	4-6	RLC	P, M, F, O
	Gypsy	→ GAG AP RT RH INT →	4-6	RLG	P, M, F, O
	Bel-Pao	→ GAG AP RT RH INT →	4-6	RLB	M
	Retrovirus	→ GAG AP RT RH INT ENV →	4-6	RLR	M
	ERV	→ GAG AP RT RH INT ENV →	4-6	RLE	M
DIRS	DIRS	→ GAG AP RT RH YR →	0	RYD	P, M, F, O
	Nguro	→ GAG AP RT RH YR → → →	0	RYN	M, F
	VIPER	→ GAG AP RT RH YR → → →	0	RYV	O
PLE	Penelope	← RT EN →	Variable	RPP	P, M, F, O
LINE	R2	— RT EN —	Variable	RIR	M
	RTE	— APE RT —	Variable	RIT	M
	Jockey	— ORF1 — APE RT —	Variable	RIJ	M
	L1	— ORF1 — APE RT —	Variable	RIL	P, M, F, O
	I	— ORF1 — APE RT RH —	Variable	RII	P, M, F
SINE	tRNA	— — —	Variable	RST	P, M, F
	7SL	— — —	Variable	RSL	P, M, F
	5S	— — —	Variable	RSS	M, O
<b>Class II (DNA transposons) - Subclass 1</b>					
TIR	Tc1-Mariner	→ Tase* →	TA	DTT	P, M, F, O
	hAT	→ Tase* →	8	DTA	P, M, F, O
	Mutator	→ Tase* →	9-11	DTM	P, M, F, O
	Merlin	→ Tase* →	8-9	DTE	M, O
	Transib	→ Tase* →	5	DTR	M, F
	P	→ Tase →	8	DTP	P, M
	PiggyBac	→ Tase →	TTAA	DTB	M, O
	PIF-Harbinger	→ Tase* — ORF2 →	3	DTH	P, M, F, O
	CACTA	→ Tase — ORF2 →	2-3	DTC	P, M, F
Crypton	Crypton	→ YR →	0	DYC	F
<b>Class II (DNA transposons) - Subclass 2</b>					
Helitron	Helitron	→ RPA — Y2 HEL →	0	DHH	P, M, F
Maverick	Maverick	→ C-INT — ATP — CYP — POL B →	6	DMM	M, F, O

Structural features			
→ → →	Long terminal repeats	← ← ←	Terminal inverted repeats
— — —	Diagnostic feature in non-coding region	— — —	Non-coding region
— — —	Region that can contain one or more additional ORFs	— — —	
Protein coding domains			
AP, Aspartic proteinase	APE, Apurinic endonuclease	ATP, Packaging ATPase	C-INT, C-integrase
ENV, Envelope protein	GAG, Capsid protein	HEL, Helicase	INT, Integrase
POL B, DNA polymerase B	RH, RNase H	RPA, Replication protein A (found only in plants)	ORF, Open reading frame of unknown function
Tase, Transposase (* with DDE motif)		YR, Tyrosine recombinase	RT, Reverse transcriptase
			Y2, YR with YY motif
Species groups			
P, Plants	M, Metazoans	F, Fungi	O, Others

**Figure I.2 : Classification des transposons eucaryotes.**

La colonne « code » propose un sigle à 3 lettres pour chaque famille du tableau. (Wicker et al., 2007)

#### 4) Les transposons parasitent le génome.

Après leur découverte, les ET ont rapidement été qualifiés d'ADN égoïste puisqu'ils codent des protéines dont la seule fonction est d'assurer leur propre prolifération dans le génome hôte (Doolittle and Sapienza, 1980). Autrement dit, ce sont des parasites moléculaires n'apportant aucun bénéfice au génome ou à la cellule ; ils constituent davantage une menace par leur potentiel mutagène. Toutefois, les transposons peuvent occasionnellement apporter un gain à l'hôte comme cela a été montré chez la drosophile avec le maintien des télomères (Levis et al., 1993) ou chez

les gnathostomes avec la recombinaison VDJ (Agrawal et al., 1998). Au-delà de cet aspect mutagène, la très grande proportion des ET en soi peut constituer un fardeau pour l'organisme. Un unicellulaire qui se diviserait plus lentement qu'un autre parce que son génome contiendrait plus de transposons serait rapidement contre sélectionné par rapport aux autres.

## 5) D'un génome à l'autre

En tant que partie intégrante des génomes, les ET sont nécessairement hérités par transmission verticale d'une génération à l'autre. Franchir la barrière d'espèces est plus compliqué mais pas impossible. Les transferts horizontaux sont fréquents chez les procaryotes, on les attribue à des cas de transformation, de transduction ou de conjugaison entre espèces plus ou moins proches (Ochman et al., 2000). A partir du moment où un gène est porté par un plasmide à large spectre d'hôte, les barrières d'espèces limitent mais n'empêchent pas les transferts horizontaux entre bactéries (Thomas and Nielsen, 2005). Chez les eucaryotes, deux cas de transferts de transposons à TIR sont bien connus : l'élément P chez les drosophiles (Daniels et al., 1990) et les *mariner* chez les insectes (Wallau et al., 2011). Il en existe beaucoup d'autres qui sont listés ici (Schaack et al., 2010) et concernent souvent des rétrotransposons à LTR de la famille *gypsy*. Une explication probable réside dans leur homologie avec des rétrovirus (Lerat and Capy, 1999). La différence entre les deux repose sur la capacité des virus à traverser la membrane cellulaire grâce à la présence d'un gène d'enveloppe. Ainsi l'acquisition d'un tel gène peut transformer un rétrotransposon à LTR en virus (Cotton, 2001) et inversement la mutation de ce gène transformera le virus en simple ET limité au génome. Il est d'ailleurs important de souligner que parmi les rétrotransposons à LTR, la classification proposée englobe la superfamille des rétrovirus.

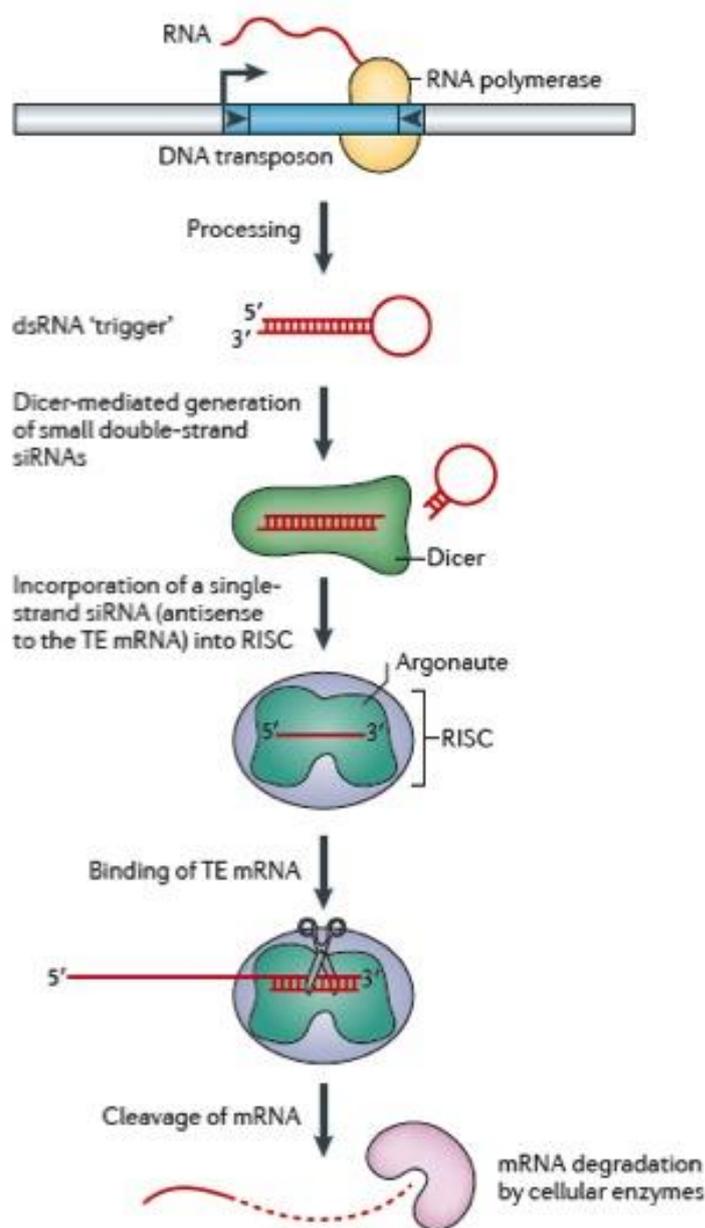
## B. L'ARN interférence

Face à la menace que représentent les transposons actifs, les génomes hôtes ont développé divers mécanismes pour les combattre. Le plus universel d'entre eux est la voie ARN interférence.

### 1) Principe

Après plusieurs observations indépendantes d'un phénomène de répression induit par l'ARN chez plusieurs organismes, la molécule initiatrice fut identifiée comme étant l'ARN double brin (Fire et al., 1998). Malgré la diversification de la voie, le mécanisme reste globalement identique chez toutes les espèces qui le possède (**Figure I.3**). Un long ARNdb est clivé en petits duplex (21-24 nt) qui sont pris en charge par une protéine Argonaute (Ago). Un des deux brins est sélectionné et guide l'activité endonucléasique d'Ago sur un ARNm complémentaire. La probabilité d'obtenir aléatoirement une séquence donnée de 20 nt est de  $10^{-12}$  donc un petit ARN de cette taille est déjà suffisant pour cibler une séquence unique avec 100% d'identité dans un génome d'une centaine de gigabases. L'ARN interférence est donc un outil puissant pour réprimer les transposons dont l'activité génère des ARNdb. Parmi eux, les transposons possédant des séquences complémentaires comme les TIR sont

concernés. Plus généralement, les séquences promotrices des transposons génèrent souvent des transcrits aberrants qui sont détectés par la cellule et transformés en ARNdb par l'action de RdRP (*RNA-dependant RNA Polymerase*). Enfin, la transcription indépendante de plusieurs transposons, insérés les uns dans les autres et dans tous les sens, génère aussi des ARNdb pouvant initier la voie interférence. On parle de voie endo-siARN lorsque l'ARNdb initiateur est produit par la cellule mais le système fonctionne également avec des ARNdb exogènes tels que ceux produits par certains virus.

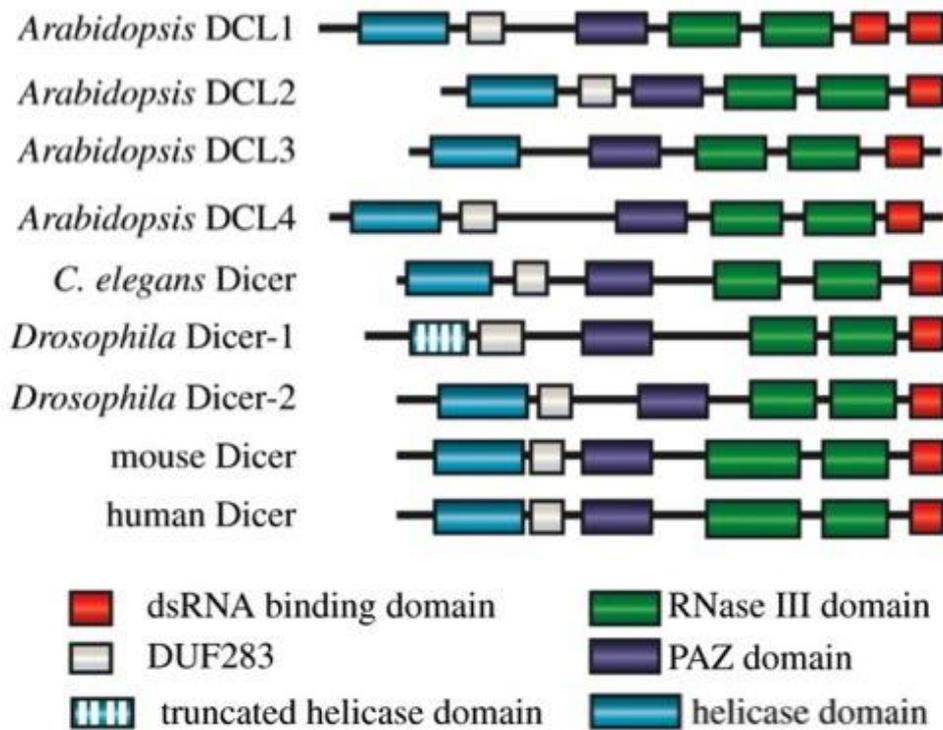


**Figure I.3 : La dégradation des ARNm des transposons par *ARNi*.**

Le transposon à ADN produit un ARNm pouvant former une épingle à cheveux (rouge) par appariement des TIR complémentaires. Celui-ci est clivé par une protéine Dicer en siARN de 21-24 nt. Une protéine Argonaute (Ago) charge le duplex et clive le brin passager (très souvent le brin sens) formant ainsi un complexe RISC (*RNA-Induced Silencing Complex*). Le brin restant cible l'ARNm du transposon et le clive, entraînant ainsi sa dégradation (Levin and Moran, 2011).

## 2) Biogenèse des siARN

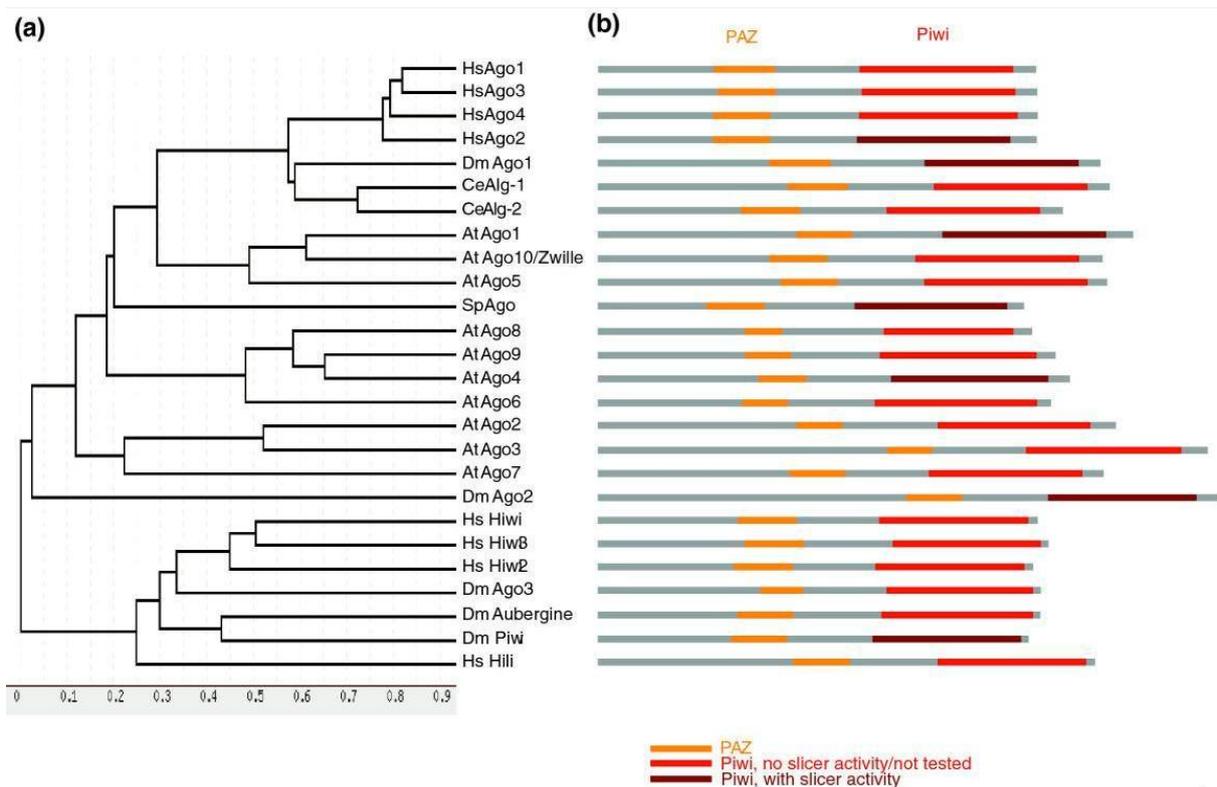
Quelle que soit l'origine de l'ARNdb, la première étape sera un clivage par une protéine de la famille Dicer ou Dicer-like. L'activité endonucléasique est portée par deux domaines RNase-III capables de reconnaître et cliver des ARNdb en courts duplex (21 -28 pb) avec 2 nt sortant en 3'. Un domaine hélicase et un domaine PAZ (*Piwi Argonaut and Zwiile*) complètent la structure typique de ces protéines (**Figure I.4**).



**Figure I.4 : Structure des protéines *Dicer* des eucaryotes.** (Johanson et al., 2013)

## 3) Les protéines Argonautes

Les protéines Argonautes (Ago) chargent le duplex généré par Dicer mais ne gardent que le brin appelé guide au détriment du brin passager (Matranga et al., 2005). Ce choix dépend de l'identité du nucléotide en 5' et de contraintes structurales et thermodynamiques (Noland and Doudna, 2013). Les protéines de la famille Argonaute possèdent toutes un domaine PAZ et un domaine PIWI (*P-element Induced Wimpy Testis*). Une analyse phylogénétique permet de distinguer deux sous-clades bien distincts : les protéines Ago à proprement parler et les protéines Piwi (**Figure 1.5**). Le groupe Ago contient des protéines dont la séquence est proche des Ago d'*Arabidopsis thaliana*. L'autre sous-famille englobe les protéines proches de la protéine Piwi de *Drosophila melanogaster*. La voie impliquant les protéines Piwi sera spécifiquement abordée dans la partie I-D. D'un point de vue structural, il y a trois domaines importants dans toutes ces protéines. Le domaine PAZ, qui existe aussi dans les protéines *Dicer*, fixe l'extrémité 3' de l'ARN. Le domaine PIWI porte l'activité de *slicing* assurant le clivage l'ARNm dans le duplex siARN-ARNm. Enfin le domaine MID (*middle*) assure l'ancrage de l'extrémité 5'P du brin guide.



**Figure I.5 : Phylogénie de la famille Argonaute dans quelques organismes.** (Höck and Meister, 2008)

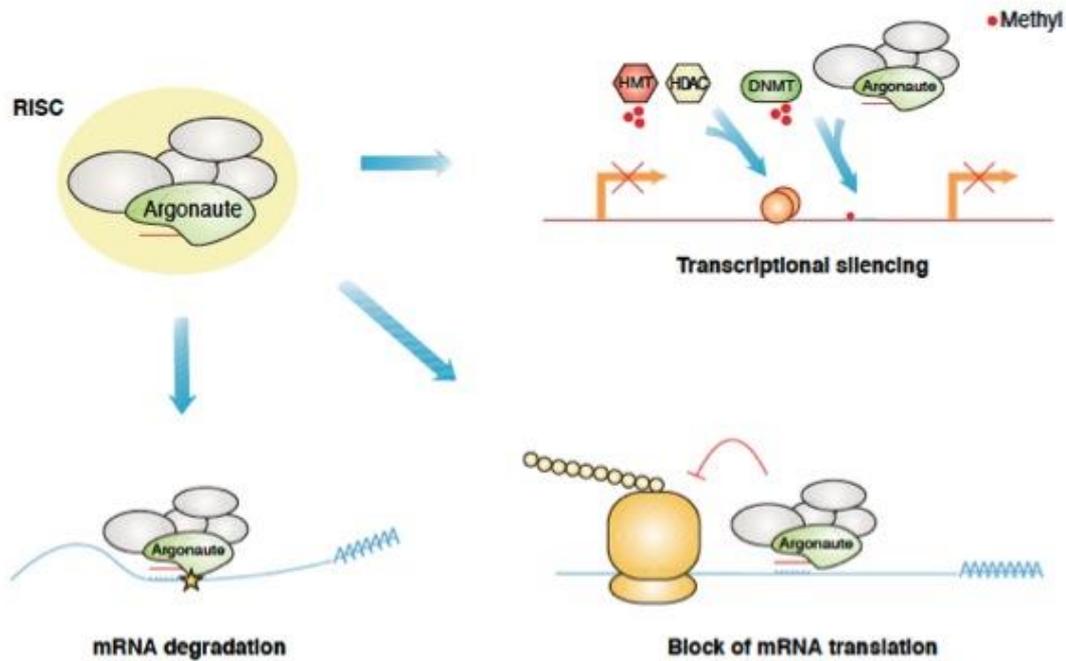
#### 4) Le complexe RISC

C'est l'association du brin guide avec la protéine Ago qui forme la base du complexe RISC (*RNA-Induced Silencing Complex*). Chaque protéine Ago a des spécificités de fixation en termes de taille, de nature du nucléotide en 5' et de tolérance au mésappariement. Par exemple, chez la drosophile, Ago1 se lie à des petits ARN de 22-23 nt avec un U en 5' (notation 22-23U) autorisant deux mésappariements avec la cible alors qu'Ago2 se lie à des petits ARN de type 21C sans mésappariement possible (Okamura et al., 2004). Le chargement du brin guide est un processus ATP-dépendant nécessitant l'intervention de chaperonnes. (Iwasaki et al., 2010).

#### 5) La répression des transposons

Les complexes RISC agissent principalement en clivant les ARNm ciblés par les petits ARN chargés (Figure I.6). Un deuxième mode de répression post-transcriptionnelle consiste à inhiber la traduction des ARNm (Li et al., 2013) au niveau du réticulum endoplasmique. Ce cas de figure concerne plus spécifiquement la voie des microARN qui sont produits à partir de précurseurs endogènes transcrits sous une forme tige-boucle. Les protéines Ago peuvent aussi exercer une répression transcriptionnelle en modifiant la structure chromatinienne des gènes ciblés. Ce phénomène a été initialement observé chez les plantes en réponse à l'insertion d'un transgène ou une

infection virale (Wassenegger et al., 1994). Le mécanisme implique la méthylation des cytosines et peut être renforcé par la méthylation des histones (Borges and Martienssen, 2015 pour une review). Ainsi les nombreux transposons rencontrés dans les génomes des plantes sont maintenus inactifs par l'ensemble de ces modifications de l'ADN et de la chromatine. Chez la levure, il existe un maintien de la répression transcriptionnelle par le complexe RITS (RNA-induced Initiation of Transcriptional gene Silencing) composée notamment de Ago1 et de Chp1, une protéine à chromodomaine assurant le remodelage de la chromatine (Verdel et al., 2004).



**Figure I.6 : Voies de répression induites par les protéines Ago.**

Le ciblage d'un ARNm par un complexe RISC peut entraîner sa dégradation ou bien bloquer sa traduction (répression post-transcriptionnelle). Un appariement avec un transcrite naissant donne accès au locus génomique et permet au complexe RISC d'exercer une répression transcriptionnelle par la méthylation de l'ADN et des méthylations/désacétylations des histones. (Almeida and Allshire, 2005)

## C. Quelques systèmes de défense des génomes chez les unicellulaires

### 1) Comment distinguer les transposons étrangers de son propre ADN ?

La voie ARN interférence repose sur la détection d'ARNdb pour identifier les séquences étrangères provenant de virus ou de transposons. En l'absence de tels intermédiaires, d'autres mécanismes sont nécessaires pour les reconnaître et pour maintenir la répression à tout instant. D'un point de vue théorique, une cellule peut logiquement utiliser deux stratégies pour distinguer un ADN (ou un ARN) étranger

par rapport au sien (Rechavi, 2014). Premièrement, elle peut marquer l'ensemble de son génome afin que toute séquence non marquée soit considérée comme étrangère. En d'autres termes, elle établit une *guest-list* de son génome. Alternativement, elle peut marquer toutes les séquences qu'elle a préalablement identifiées comme étant étrangères. Ce système de *black-list* nécessite un premier contact puis une mise en mémoire de la séquence parasite. Cet aspect a déjà été abordé avec la répression transcriptionnelle induite par Ago car la méthylation de l'ADN ou des histones peut être maintenue au cours de la réplication et transmise au cours de la mitose/méiose. L'ARN interférence, tout comme le système de *guest-list*, peut faire face à toute intrusion sans aucun contact préalable, on peut donc parler d'immunité innée du génome. A l'opposé, un système de *black-list*, capable de garder en mémoire les séquences parasites rencontrées afin de les neutraliser spécifiquement, s'apparente à un système d'immunité acquise. Les organismes utilisent ces deux types d'immunité dont certaines sont détaillées dans les paragraphes suivants.

## 2) Les systèmes procaryotes

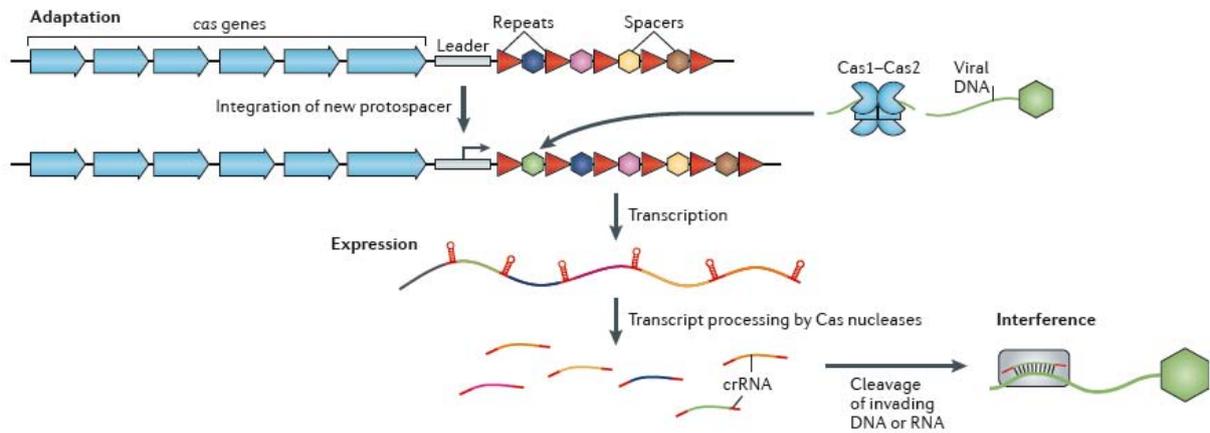
### a. Les systèmes de modification-restriction

Les bactéries possèdent généralement un système de modification-restriction. Sous sa forme la plus commune et la plus simple (le type II), il est composé d'une méthyltransférase et d'une endonucléase. Les deux enzymes reconnaissent un même site palindromique, par exemple GATC, et l'endonucléase ne peut couper que les sites non méthylés. Alors que le génome bactérien sera constamment protégé par la méthylase, une séquence étrangère possédant le site sera immédiatement clivée et dégradée. Ce système nécessite un bon équilibre entre l'activité des deux enzymes qui sont en compétition sur le même site. Typiquement, l'enzyme de restriction est plus active et sera en mesure de cliver rapidement une séquence étrangère alors que le génome hôte sera protégé par l'efficacité de la méthylase à traiter les sites hémiméthylés créés au cours de la réplication (Labrie et al., 2010). La simplicité de ce système ouvre la voie à de nombreux mécanismes d'échappement caractéristiques d'une relation hôte-parasite. Par exemple, le phage K de *Staphylococcus* ne possède pas de site Sau3A dans son génome et le bactériophage Tuc2009 produit une méthylase pour protéger son génome (McGrath et al., 1999).

### b. Les systèmes CRISPR

Les systèmes CRISPR (*Clustered Regularly Interspaced Short Palindromic Repeat*) sont présents chez 45 % des bactéries et 85 % des archées. Le nom reflète la structure de ces régions génomiques initialement identifiées dans plus d'une quarantaine de procaryotes en 2002 (Jansen et al., 2002). Elles sont composées de séquences répétées (21-48 pb) entrecoupées de séquences variables appelées *spacers* (26-72 pb) homologues à des segments viraux ou dérivées de plasmides. A cette organisation s'ajoutent plusieurs gènes *CAS* (*CRISPR-associated*) dont la nature varie en fonction du type de système (Gasiunas et al., 2014). Les protéines Cas1 et Cas2 sont communes à tous les types et assurent l'intégration de nouveaux *spacers*. Le locus CRISPR est transcrit à partir d'une séquence leader puis maturé en petits ARN appelés crRNA (*CRISPR RNA*) par différents gènes Cas. Ces crRNA sont utilisés

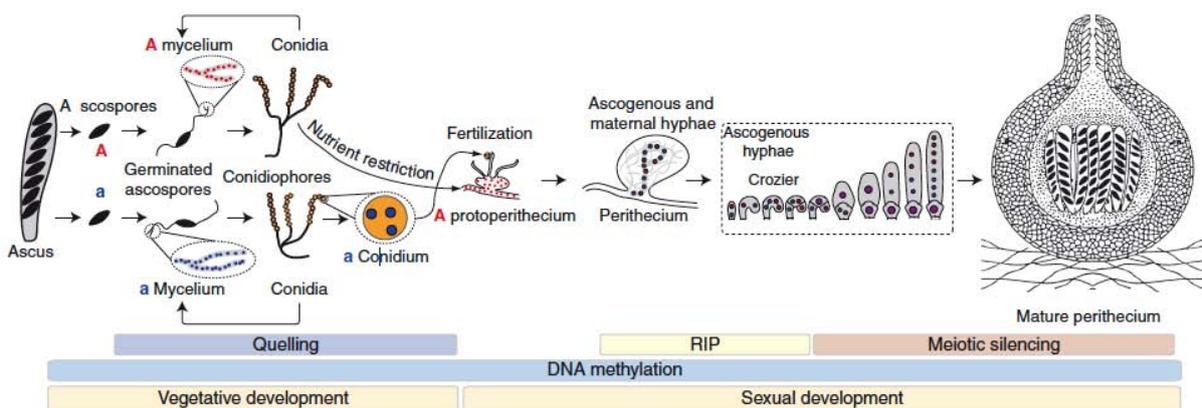
comme guide pour cliver les ADN ou ARN homologues par un processus d'interférence similaire au complexe RISC (**Figure I.8**). En résumé ces systèmes sont adaptatifs et spécifiques de l'élément envahisseur. L'intégration d'un nouvel élément s'apparente à une vaccination qui est d'autant plus efficace qu'elle est transmissible verticalement.



**Figure I.7 : Le système immunitaire CRISPR des procaryotes.** (Koonin and Krupovic, 2015)

### 3) Les systèmes chez *Neurospora crassa* (MSUD, quelling, RIP)

*Neurospora crassa* est un champignon filamenteux utilisé en laboratoire depuis le début du XXème siècle. Son étude a révélé une grande richesse de mécanismes de répressions épigénétiques de type ARNi pouvant contrer les transposons. Les trois mécanismes qui vont être discutés opèrent à différents moments du cycle de vie (**Figure I.8**).



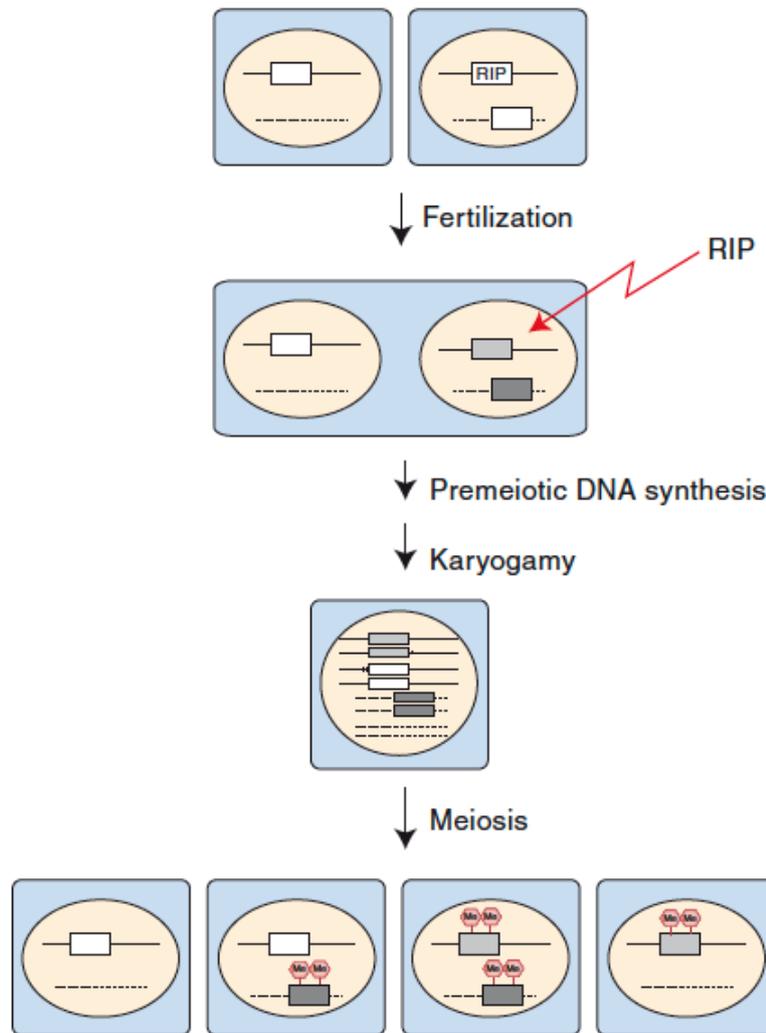
**Figure I.8 : Cycle de vie de *Neurospora crassa*.**

La moitié des ascospores sexuels sont de type sexuel A (rouge) et l'autre moitié de type a (bleu). Ces ascospores forment des mycéliums desquels émergent des fructifications asexuelles (conidiophores). Ces dernières produisent des spores végétatives (conidies)

multinucléées qui peuvent former d'autres mycéliums ou bien féconder une structure femelle (périthèce) de type sexuel opposé. De telles structures sont produites par les mycéliums sexuels en réponse à une privation d'azote. Après la fécondation, les noyaux mâles et femelles coexistent dans le même cytoplasme et forment un hyphé dicaryotique. La caryogamie intervient dans des structures en forme de crochet puis est immédiatement suivie par une méiose et une mitose générant un asque à 8 spores. Les étapes dans lesquelles opèrent les processus épigénétiques décrits sont indiquées sous le schéma du cycle. (Aramayo and Selker, 2013).

#### **a. Le RIP (Repeat-Induced Point mutation)**

Le premier système identifié fut le RIP dans les années 90 (Selker, 1990). Des séquences dupliquées, qu'elles soient endogènes ou non, génétiquement liées ou non, subissent des mutations de C vers T dans les noyaux haploïdes post-fécondation. Il s'agit d'un processus de désamination affectant majoritairement le dinucléotide CpA (Cambareri 1990). En un passage à travers le cycle, 30 % des cytosines des séquences dupliquées sont mutées. Les mutations s'accompagnent souvent de méthylations des cytosines sur le même locus (**Figure I.9**). L'analyse du génome de *N. crassa* a d'ailleurs révélé que la grande majorité des régions méthylées sont des reliques de transposons inactivés par RIP (Selker et al., 2003). Le système de détection des séquences dupliquées n'est pas connu mais il implique vraisemblablement des ARN. Puisque ce système de défense détruit toute séquence qui tenterait de se dupliquer, il doit limiter en contrepartie la possibilité d'évolution du génome par duplication de gènes. Les familles de gènes identifiées montrent que tous les paralogues sont suffisamment divergents pour échapper au RIP (Galagan and Selker, 2004).

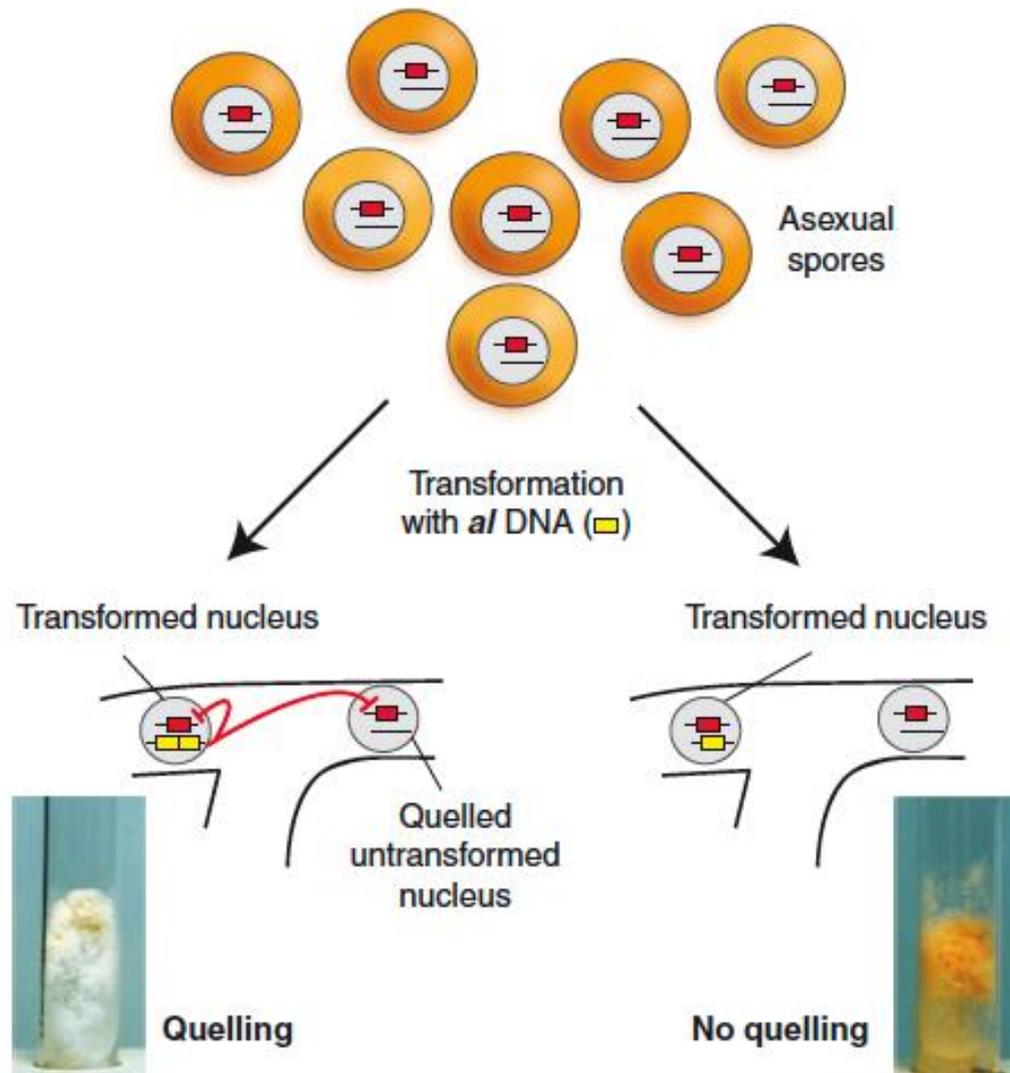


**Figure I.9 : Le RIP (*Repeat-Induced Point mutation*).**

Les traits noirs représentent des chromosomes et les rectangles représentent un gène ou un segment de chromosome qui a été dupliqué dans une souche (en haut à droite). Les duplications sont soumises au RIP (rectangles grisés) entre la fécondation et la caryogamie. Le génotype des 4 ascospores possibles est représenté en bas. Les étiquettes roses 'Me' désignent la méthylation de l'ADN associée au RIP. (Aramayo and Selker, 2013).

### **b. Le quelling**

Lorsque les transformations chez *Neurospora* ont débuté, les chercheurs ont observé que 30 % des individus transformés voyaient leur transgène réprimé. De plus, le gène endogène correspondant était aussi réprimé et on a donné le nom de *quelling* à ce phénomène (Pickford et al., 2002) (**Figure I.10**). Il implique les gènes *QDE-1*, *QDE-2* et *QDE-3* codant respectivement une ARN polymérase ADN/ARN-dépendante, une protéine *Ago-like* et une ADN hélicase. La molécule initiatrice serait un ARN provenant de la transcription du transgène mais sa nature n'est pas clairement identifiée (ARN aberrant ou ARN antisens ou ARN chimérique). Les petits ARN effecteurs sont des 20-21U appelés qiARN (Lee et al., 2009). D'après l'effet sur les transgènes, ce système devrait réprimer les nouvelles insertions d'ET ainsi que l'ET initiateur de la transposition.

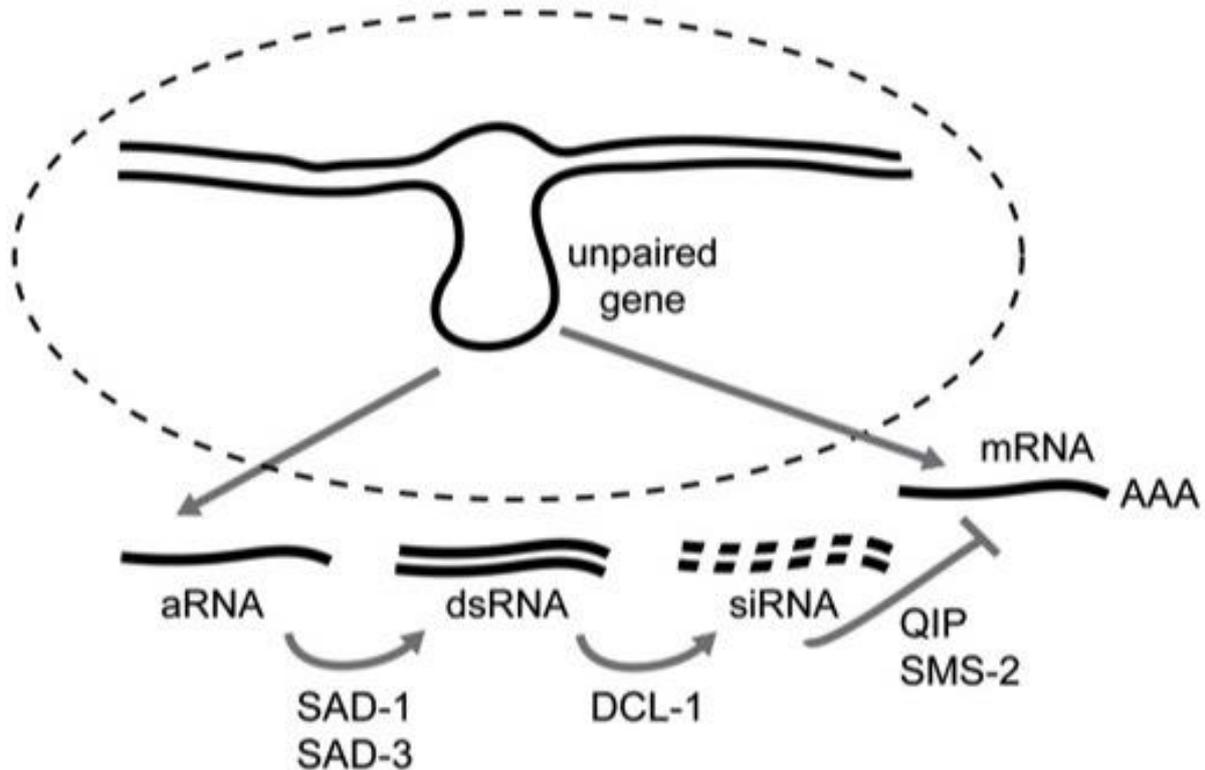


**Figure I.10 : Le quelling.**

Les traits noirs représentent des chromosomes. Des spores asexuelles portant le gène albino sauvage (rectangle orange) sont transformées par une version ectopique (rectangle jaune). 30% des cellules multinucléées transformées montrent une répression du gène endogène (trait rouge) et ont un phénotype [blanc] alors que les autres 70% ne sont pas affectés et sont [orange] (Aramayo and Selker, 2013).

**c. Le MSUD (Meiotic Silencing of Unpaired DNA)**

Ce dernier mécanisme de répression opère spécifiquement pendant la méiose et cible des régions non appariées entre chromosomes homologues (Aramayo and Metzenberg, 1996; Shiu et al., 2001). C'est le cas lorsque qu'une nouvelle séquence (transgène, transposon ou autre) s'intègre dans un allèle, on parle de situation hémizygote. Le mécanisme implique d'abord la comparaison des loci alléliques entre chromosomes homologues par un processus appelé *trans-sensing* puis dans un second temps la répression des séquences ectopiques par une voie de type ARNi (**Figure I.11**). Toutes les séquences homologues à la région détectée sont alors réprimées. L'efficacité du mécanisme augmente avec la taille de la boucle et ne requiert pas la présence d'un promoteur (Lee et al., 2004).



**Figure I.11 : Un modèle du MSUD (Meiotic Silencing of Unpaired DNA).**

Un mécanisme de *trans-sensing* identifie les régions ectopiques des chromosomes homologues et génère un ARN aberrant (aRNA). Les principaux composants du MSUD, nommés en majuscules, sont associés à une structure périnucléaire. Ils convertissent l'ARN aberrant en ARNdb puis en siARN et s'en servent comme guide pour cliver spécifiquement l'ARNm de la région détectée. (Hammond et al., 2011)

Le *trans-sensing* est sensible à la présence d'une séquence ectopique mais également à l'absence d'une homologie suffisante entre deux allèles. Le degré de mésappariement qui peut être détecté en méiose dans un asque hétérozygote a été évalué par différents croisements (Pratt et al., 2004). 6 % de divergence suffisent à induire une répression alors que 3 % ne suffisent pas. Le seuil est abaissé en présence de méthylation induite par le RIP. Il est intéressant de noter qu'un mutant déficient dans la voie MSUD est capable de donner une descendance partiellement viable lorsqu'il est croisé avec des espèces proches de *N. crassa* (Shiu et al., 2001). Ce système de défense permettant de neutraliser l'intégration d'ADN étranger contribue conjointement à l'isolement reproductif de l'espèce.

## D. Les piARN des métazoaires

### 1) Caractéristiques des piARN

Les protéines Piwi furent initialement identifiées chez la mouche du vinaigre comme des facteurs favorisant l'auto-renouvellement des cellules souches germinales. On a ensuite trouvé des orthologues chez d'autres organismes eucaryotes (Cox et al., 1998).

Il s'agit du deuxième clade des protéines de la famille des Argonautes. En 2006, une nouvelle classe de petits ARN associés à ces protéines fut découverte : les piARN (*PIWI Interacting RNA*) (Aravin et al., 2006; Girard et al., 2006). Les piARN sont abondants dans les tissus germinaux et sont caractérisés par un uracile en 5' et une 2'-O-méthylation du sucre en 3'. L'extinction des gènes PIWI résulte généralement en un problème de fertilité. Par ailleurs, les piARN s'alignent principalement sur des séquences de transposons et c'est ainsi que leur rôle primordial dans le contrôle des ET de la lignée germinale fut établi. Malgré la relative conservation des protéines PIWI, les voies identifiées peuvent être assez différentes entre organismes. J'ai choisi de développer deux des systèmes les mieux connus pour illustrer le fonctionnement des piARN.

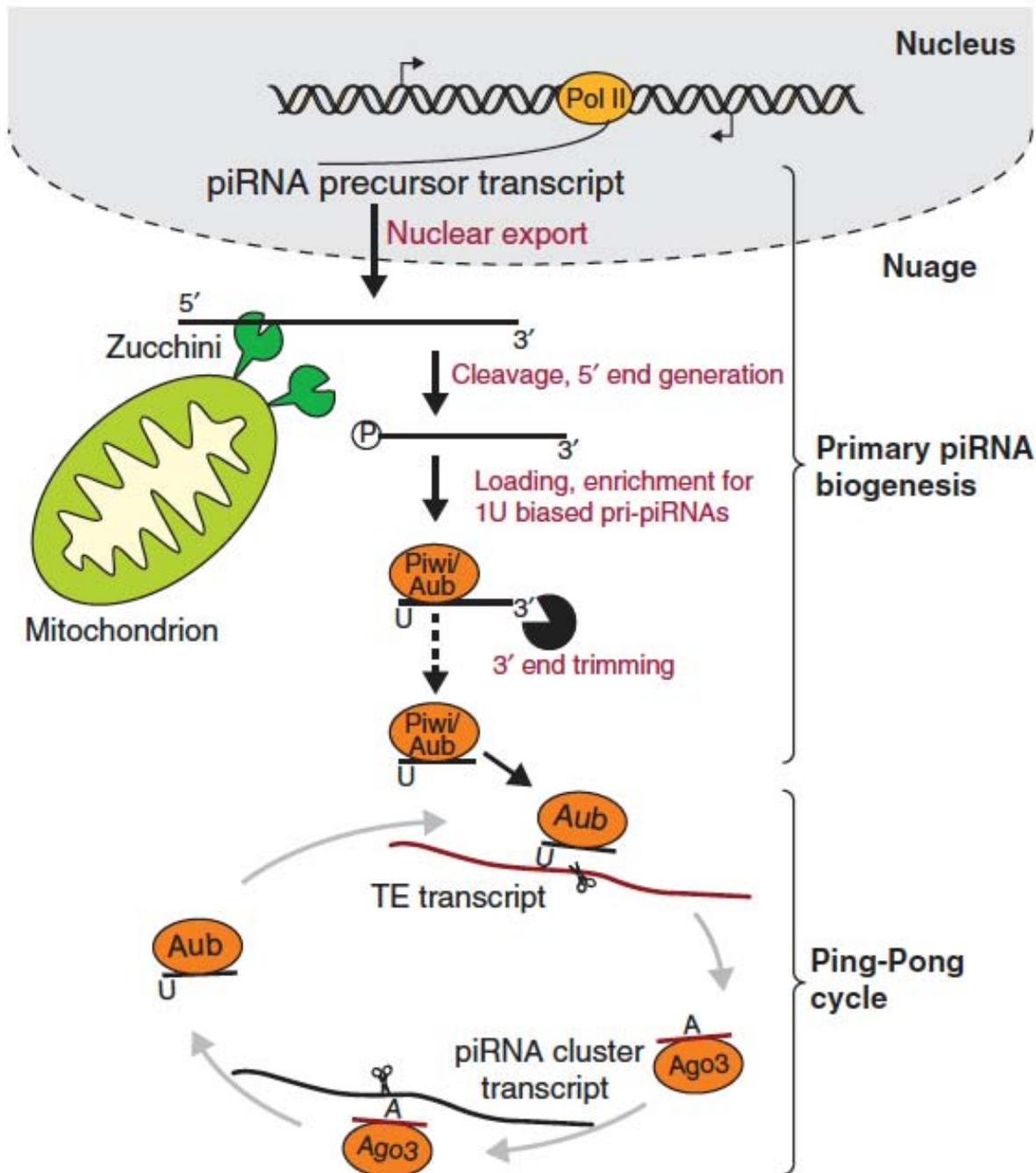
## 2) La voie piARN de *Drosophila melanogaster*

### a. Biogenèse des piARN primaires

Presque tous les piARN de la drosophile dérivent de régions péri-centromériques et sub-télomériques riches en répétitions. Ces clusters sont pauvres en séquences codantes et contiennent principalement des fossiles de transposons (Brennecke et al., 2007). Ils peuvent être transcrits dans un sens ou dans les deux sens mais aucun promoteur commun n'est visible. En revanche, la protéine Rhino, un homologue de HP1 (*Heterochromatin Protein 1*), est spécifiquement associée aux clusters de piARN (Klattenhoff et al., 2009). Les longs transcrits subissent une maturation en 5' par une nucléase (probablement Zucchini (Ipsaro et al., 2012)) et l'ajout d'un 2'-O-méthyl en 3' par Pimet (Saito et al., 2007). Les piARN matures sont pris en charge par les protéines Piwi et Aub (*Aubergine*, une autre protéine PIWI), ils sont antisens relativement aux ARNm des transposons (**Figure I.12**).

### b. Cycle d'amplification ping-pong

La troisième protéine PIWI existant chez la drosophile est Ago3. Elle se lie à des piARN sens dont les 10 premiers nucléotides peuvent être complémentaires au piARN portés par Aub. En particulier, un A complémentaire du 5'U apparaît clairement en 10<sup>ème</sup> position, c'est l'une des principales observations qui a conduit à l'émergence du modèle ping-pong (Brennecke et al., 2007; Gunawardane et al., 2007) (**Figure I.12**). Brièvement, le complexe Aub/piARN reconnaît et clive son ARN cible (typiquement un ARNm d'ET), produisant l'extrémité 5' d'un nouveau piARN qui est pris en charge par Ago3. Le complexe piARN/Ago3 reconnaît et clive un transcrit complémentaire issu d'un cluster, générant une nouvelle extrémité 5' d'un piARN qui sera associée à Aub et ainsi de suite. Tout comme la maturation des piARN, ce cycle se déroule dans une structure périnucléaire dense appelée *nuage* (Brennecke et al., 2007). Le cycle ping-pong permet l'amplification des piARN qui trouvent une cible parmi les ARNm et génère alors une population de piARN secondaires spécifiques des transposons actifs qui auraient échappé à la répression transcriptionnelle.

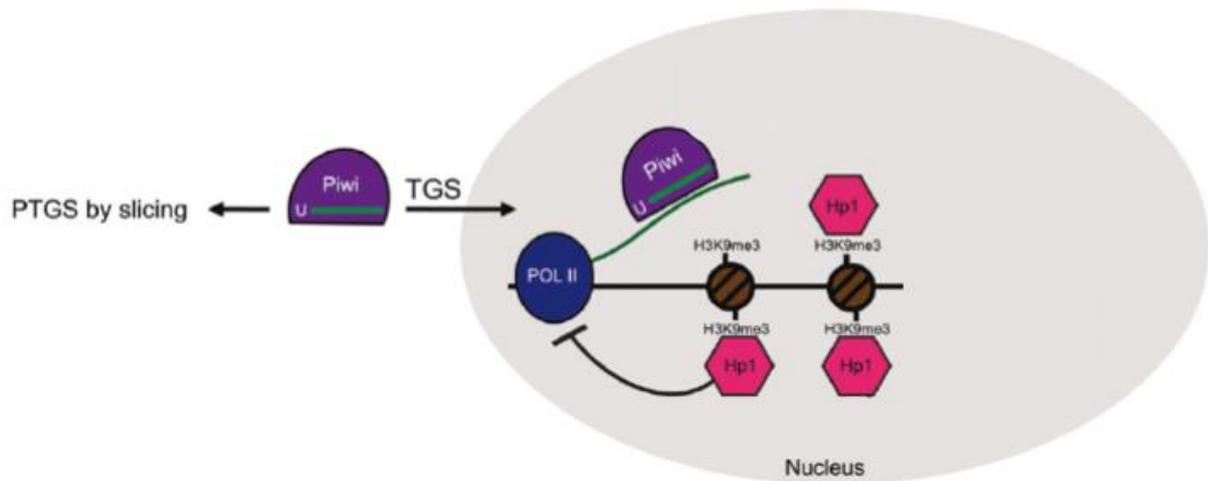


**Figure I.12 : Biogenèse des piARN chez *D. melanogaster*.** (Le Thomas et al., 2014)

### c. Répression

La répression post-transcriptionnelle est bien comprise, elle se fait par un clivage des ARNm dépendant de l'activité catalytique des protéines Piwi. En revanche, la répression transcriptionnelle est moins claire, elle ne requiert pas l'activité catalytique mais dépend de la localisation nucléaire de Piwi (Klenov et al., 2011). La déplétion de Piwi entraîne une perte de la marque H3K9 et une augmentation de l'occupation de la polymérase PolII au niveau des séquences des ET (Le Thomas et al.,

2013). De plus, la protéine HP1 est déposée quand on dirige des piARN sur un système rapporteur. Ainsi, Piwi serait capable d'entrer dans le noyau et de se lier à des transcrits naissants homologues au piARN qu'elle porte afin d'induire la formation d'hétérochromatine (**Figure I.13**).



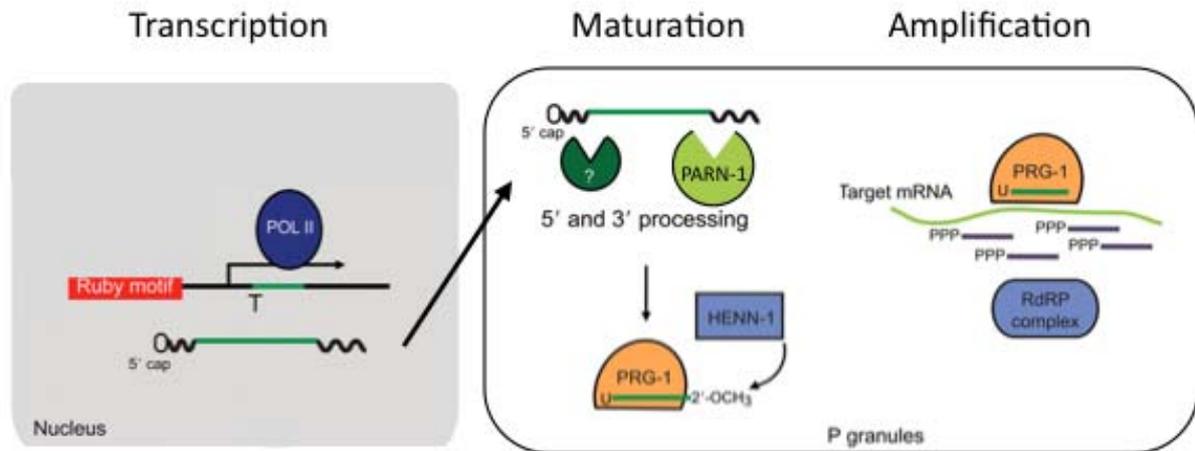
**Figure I.13 : La répression des transposons induite par Piwi.**

Chez *D. melanogaster*, Piwi possède une activité post-transcriptionnelle grâce à son activité catalytique. Elle peut aussi initier le dépôt de la marque répressive H3K9 triméthylée dans le noyau. HP1 est alors recruté et permet le blocage de l'ARN polymérase II. Adapté de (Weick and Miska, 2014).

**3) La voie piARN de Caenorhabditis elegans**

**a. Biogenèse des piARN primaires**

Les piARN de *C. elegans* sont des 21U portés par Prg-1 (*PIWI-related gene 1*), la seule protéine PIWI fonctionnelle du génome (Batista et al., 2008). Deux clusters sur le chromosome IV concentrent les milliers de loci à partir desquels les précurseurs des piARN sont produits. Ils ont en commun un motif spécifique en amont de la séquence transcrite (Ruby et al., 2006) (**Figure I.14**). L'extrémité 3' des précurseurs est réduite jusqu'à 21 nt par l'exonucléase PARN-1 (Tang et al., 2016) puis méthylée par Henn-1 (Montgomery et al., 2012). Étonnamment, les 21U ne s'alignent précisément sur aucun ARNm, que ce soit des gènes cellulaires ou des gènes de transposons. À l'inverse, si l'on autorise 2-3 mésappariements alors la diversité des 16 000 piARN pourrait cibler n'importe quel ARNm existant. Dans les faits, Prg-1 est requise pour la répression des transposons Tc3 (Das et al., 2008) et pour induire la répression d'un transgène (Bagijn et al., 2012)



**Figure I.14 : Biogenèse des piARN chez *C. elegans*.** Adapté de (Weick and Miska, 2014)

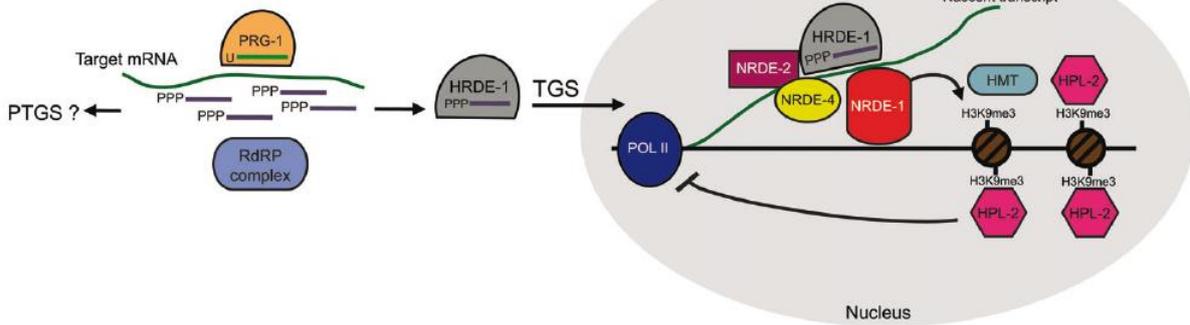
### b. Les siARN secondaires

Lorsqu'on observe une répression dépendante de Prg-1, elle est accompagnée par la génération d'une seconde classe de petits ARN qui s'hybrident parfaitement sur les ARNm cibles. Ils sont produits par l'action de RdRP sur les ARNm ciblés par Prg-1 (Gu et al., 2009). Ce mécanisme d'amplification est l'alternative du nématode au cycle ping-pong de la mouche et des mammifères. Ces ARN secondaires sont des 22G associés à des protéines WAGO (*Worm Argonaute*) qui appartiennent à la famille des Argonautes mais forment une 3ème branche récemment découverte qui n'existe que chez les nématodes (Lee et al., 2012). La répression *de novo* d'un transgène est initiée par Prg-1 puis est maintenue par des 22G associés à des protéines WAGO (Shirayama et al., 2012).

### c. Répression

L'activité catalytique de Prg-1 n'est pas requise pour la production des 22G en aval (Lee et al., 2012). La répression post-transcriptionnelle par Prg-1 n'a pas été clairement étudiée donc on ne sait pas si elle existe. La répression transcriptionnelle est initiée par la translocation nucléaire de Wago-9 qui induit la déposition de H3K9me3 et le blocage de PolII (**Figure I.15**). Wago-9 est également appelé Hrde-1 pour *Heritable RNAi Deficient 1*. D'autres facteurs chromatinienens dont HPL-2 (*HP-Like-2*) et plusieurs histone-méthyltransférases (HMT) sont requis. Enfin les protéines Nrde (*Nuclear RNAi Defective*), dont le rôle est connu dans la répression transcriptionnelle de la voie ARNi classique chez *C. elegans*, sont aussi impliquées dans la voie piARN.

C. *C. elegans*



**Figure I.15 : La répression des transposons induite par Prg-1.** (Weick and Miska, 2014)

#### **d. La protection par Csr-1**

Puisque les piARN peuvent potentiellement cibler n'importe quelle séquence, ils peuvent être dangereux pour les gènes cellulaires exprimés dans les tissus germinaux. Ce problème semble être résolu depuis la découverte de la protéine Argonaute Csr-1 (*Chromosome Segregation and RNAi déficient 1*) qui s'associe à des 22G antisens de ces gènes et les protège des machineries de répression des voies ARNi (Seth et al., 2013; Wedeles et al., 2013). Ainsi on peut distinguer deux classes de petits ARN de type 22G : les premiers sont liés à Hrde-1 et répriment durablement les séquences cibles alors que les deuxièmes sont liés à Csr-1 et contrebalancent l'effet précédent en activant la transcription. Un modèle global d'immunité innée se dessine dans lequel Prg-1 scanne tous les ARNm avec l'ensemble de son répertoire de piARN 21U. Toute séquence non protégée par Csr-1 sera considérée comme étrangère et une répression dépendante de Hrde-1/22G sera mise en place.

#### **4) Hérité transgénérationnelle des piARN**

Les complexes HRDE-1/22G peuvent induire une répression épigénétique transgénérationnelle stable et indépendante de Prg-1 (Ashe et al., 2012). L'épiallèle ainsi généré est associé à la marque H3K9 et possède un effet dominant sur d'autres allèles qui étaient exprimés à la génération précédente (Shirayama et al., 2012), on parle alors de paramutation. La dominance est légèrement plus forte si l'épiallèle est porté par le génome maternel. On pense actuellement que le signal répressif est transmis par les 22G eux-mêmes. En effet, les piARN 21U ne semblent pas nécessaires puisque Prg-1 n'est pas requis et la marque H3K9 n'est pas très stable en l'absence de stimulus (Gu et al., 2012). Parallèlement les complexes Csr-1/22G sont eux aussi capables de transcender les générations en protégeant certains transgènes de la répression piARN-dépendante. Lors d'un croisement entre un ver portant un transgène exprimé (ciblé par Csr-1) et un autre portant un transgène homologue réprimé (ciblé par Hrde-1), c'est l'allèle exprimé qui est paramutateur (Shirayama et al., 2012). De plus, l'allèle converti peut devenir stablement exprimé en l'absence de

l'allèle dominant, c'est-à-dire qu'il peut être « adopté » et considéré comme un gène cellulaire (Seth et al., 2013). Finalement, le système des piARN chez *C. elegans* se transmet à travers les générations avec une *guest-list* portée par les 22G associés à Csr-1 et une *black-list* portée par les 22G associés à Hrde-1.

Chez la drosophile, les clusters de piARN s'apparentent à un piège dans lequel la transposition d'une copie d'un ET initiera la réponse. On peut donc parler de *black-list* qui est transmise génétiquement dans les clusters. Avoir la liste ne suffit pas car la répression des transposons dans la génération suivante nécessite l'initiation du cycle ping-pong via la déposition de piARN dans l'ovocyte (Brennecke et al., 2008; Chambeyron et al., 2008). Cela pose d'ailleurs un problème de dysgenèse hybride lorsque le répertoire des piARN maternels ne contient pas tous les transposons du génome paternel.

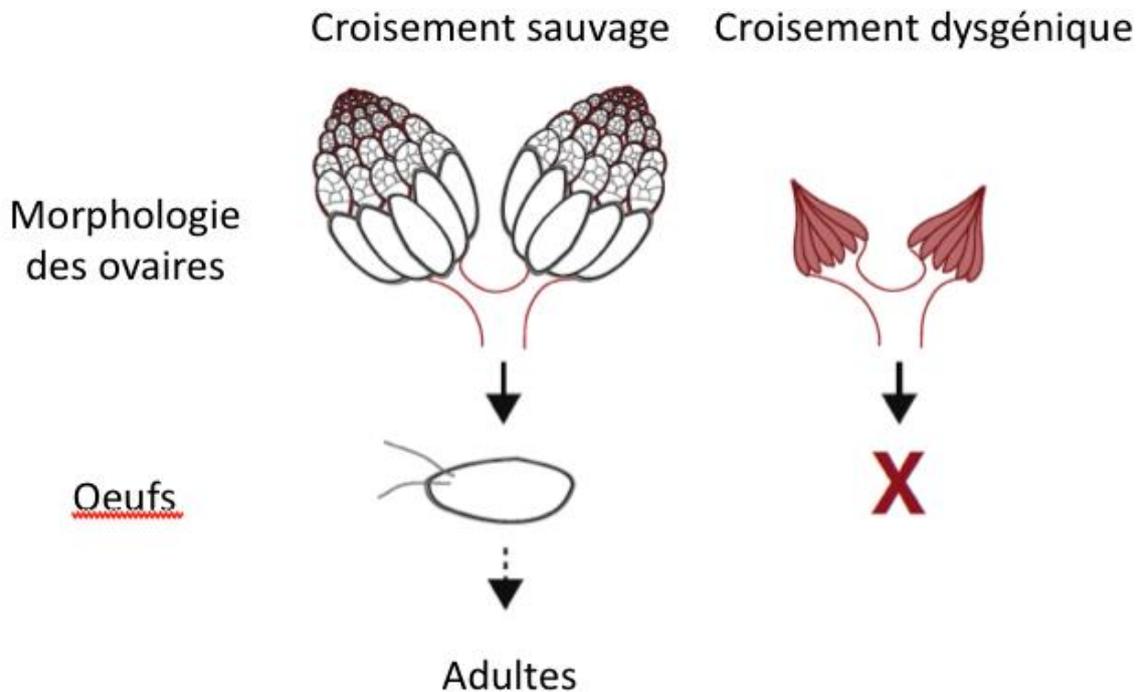
## II – La dysgenèse hybride

### A. Le système P-M

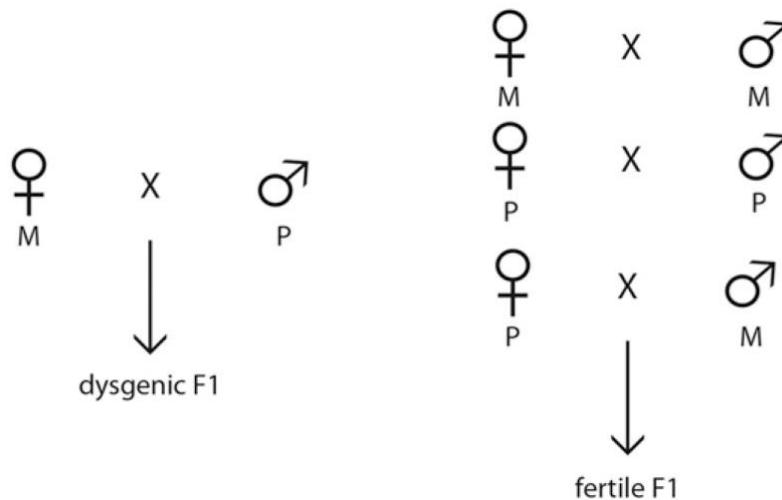
#### 1) Etudes génétiques historiques

##### 1. Découverte du système P-M

Au début des années 60, les généticiens observèrent des effets délétères dans la descendance de croisements entre des souches de *Drosophila melanogaster* sauvages et des souches de laboratoires. Un certain nombre de phénotypes caractéristiques et spécifiques des tissus germinaux furent décrits par la suite : mutations, aberrations chromosomiques, recombinaison chez les mâles, atrophie des ovaires (**Figure II.1**), stérilité, non disjonction, etc. On donna le nom de « dysgenèse hybride » à ce phénomène (Kidwell et al., 1977). Traditionnellement, les souches sont classées en deux types M (*maternal*) et P (*paternal*). Quand un mâle d'une souche P est croisé avec une femelle d'une souche M alors la descendance F1 est dysgénique. En revanche, le croisement réciproque et les croisements M-M ou P-P donnent une progéniture totalement fertile (**Figure II.2**).



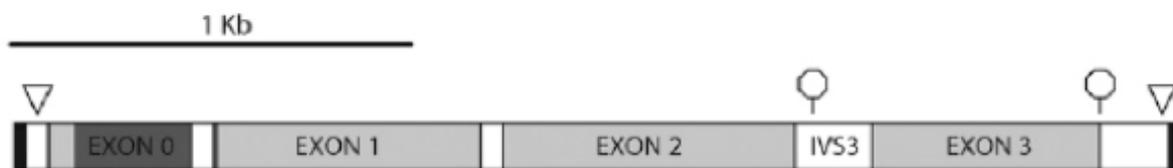
**Figure II.1 : Atrophie ovarienne chez les femelles dysgéniques.** Adapté de (Malone et al., 2015)



**Figure II.2 : Croisement dysgéniques et non dysgéniques.**

## 2. Attribution à l'élément P

C'est l'étude des mutations dans la descendance qui a permis de comprendre l'origine de la dysgénèse. Des études de séquençage ont montré qu'elles sont induites par l'insertion d'éléments génétiques mobiles qui se ressemblent fortement entre eux (Bingham et al., 1982). Puisqu'ils sont typiquement présents en 30-50 copies dans les souches P et quasiment absents des souches M, ces transposons furent appelés éléments P. Les autres familles de transposons étudiées par FISH ne montrent pas de mobilisation donc il s'agit bien de l'élément P qui est responsable de la dysgénèse hybride (Eggleston et al., 1988), un transposon à ADN portant des TIR (**Figure II.3**). Il est apparu dans les populations sauvages de *Drosophila melanogaster* au début du 20ème siècle, vraisemblablement par transfert horizontal depuis *Drosophila wilstoni*. En effet, les éléments P de ces deux espèces ne diffèrent que d'une substitution en moyenne alors que les espèces elles-mêmes divergent depuis plus de 40 millions d'années (Daniels et al., 1990).

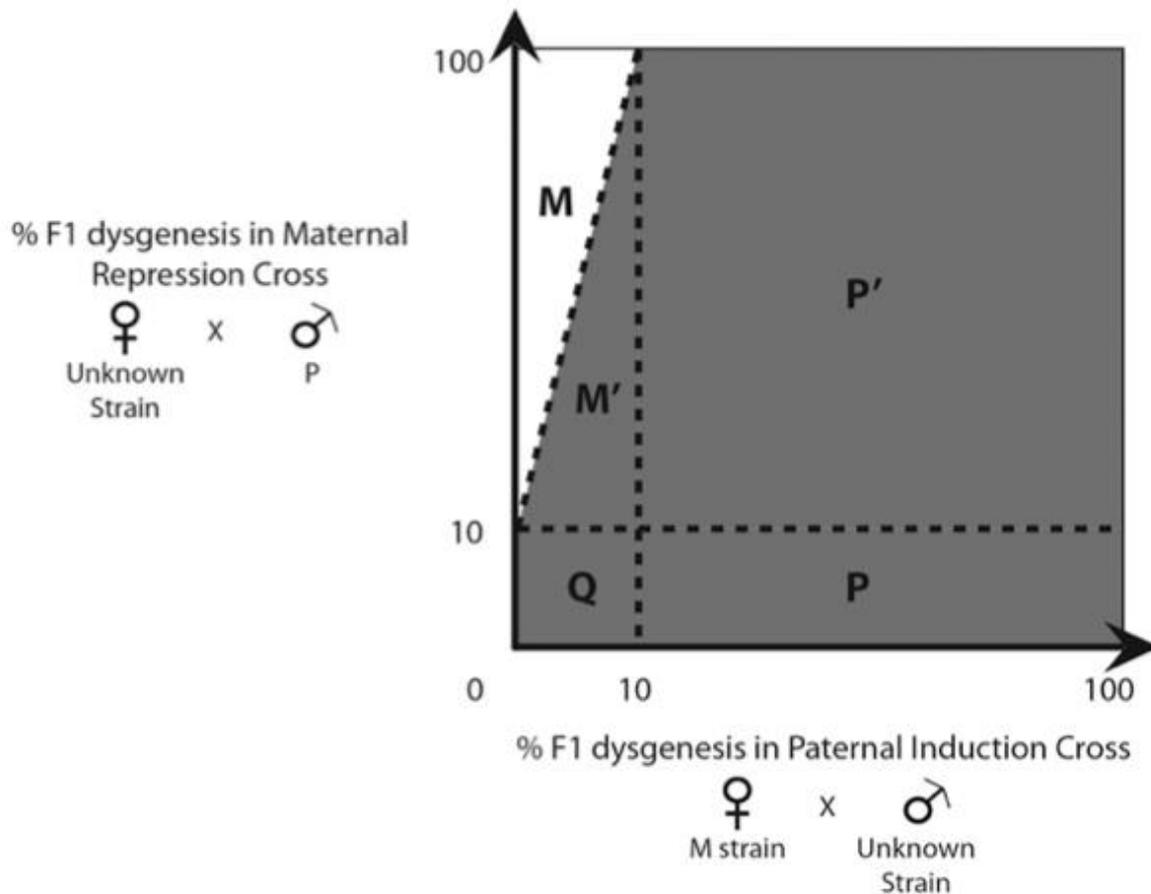


**Figure II.3 : Structure de l'élément P complet.**

L'élément P contient des TIR (noir) et 4 exons (gris) qui codent la transposase P dont le domaine de liaison à l'ADN est indiqué en gris foncé. Les triangles indiquent la position du consensus ATACACTTAA reconnu par la transposase et les octogones indiquent les codons stop. Le troisième intron porte un codon stop et est appelé IVS3 (*Intervening Sequence 3*) (Kelleher, 2016).

### 3. Notion de cytotype

La dysgenèse hybride de référence est le croisement d'un mâle Harwich (souche P de référence) avec une femelle Canton S (souche M de référence). Elle repose sur deux propriétés distinctes : la capacité à induire paternellement la dysgenèse et la capacité à la réprimer maternellement. Une souche P possède ces deux propriétés alors qu'une souche M n'en a aucune. Cependant, de nombreuses souches ne rentrent pas dans ces catégories (**Figure II.4**). C'est le cas des souches Q neutres car elles ne peuvent pas induire la dysgenèse paternellement mais peuvent la réprimer. La dissociation de ces deux facultés montre qu'elles ont deux explications moléculaires distinctes. S'il est assez clair que l'induction nécessite la présence d'éléments P fonctionnels, les bases de la répression maternelle sont quant à elles plus difficiles à expliquer. Des études génétiques ont montré que la répression dépendait à la fois de facteurs chromosomiques et extra-chromosomiques (Ronsseray et al., 1993). Engels proposa alors le nom de « cytotype » pour qualifier cette capacité cellulaire : le cytotype P peut réprimer maternellement l'activité des éléments P alors que le cytotype M n'en est pas capable (Engels, 1979).



**Figure II.4 : Classification des souches au regard de la dysgenèse hybride.**

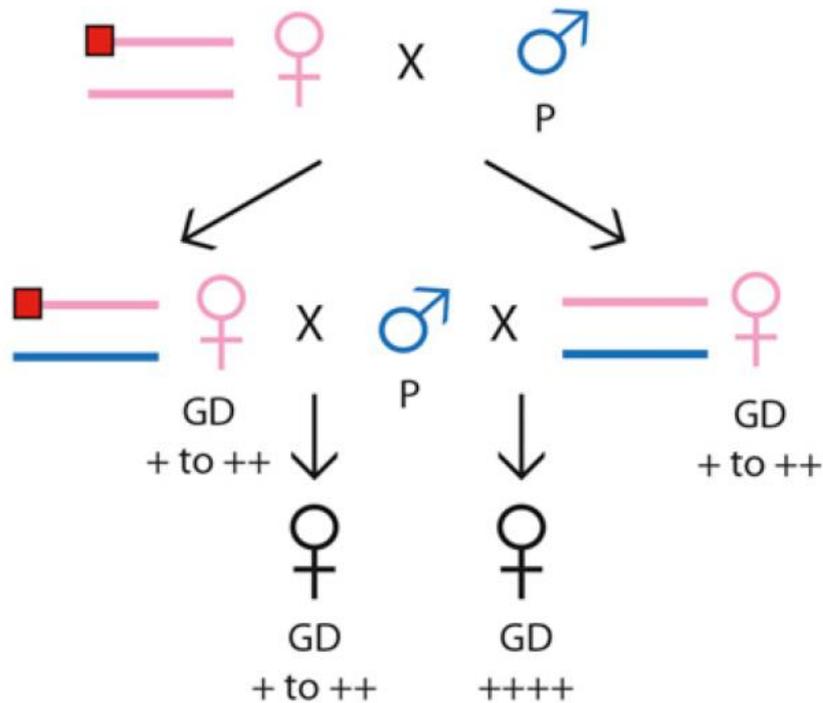
Les souches sont classées selon leur double capacité à induire une dysgenèse dans une souche M de référence (axe des abscisses) et à réprimer la dysgenèse induite une souche P de référence (axe des ordonnées). La fréquence de 10% d'individus dysgéniques en F1 est utilisée comme seuil pour séparer les souches (Kelleher, 2016).

## 2) Les bases moléculaires de la répression maternelle

Deux modèles non mutuellement exclusifs ont été développés pour expliquer le cytotype P. Le premier est celui des protéines répressives codées par les éléments P qui seraient déposées dans les ovocytes (Gloor et al., 1993; Roche et al., 1995). Les répresseurs de type I incluent la transposase tronquée produite lorsque l'ARNm retient le 3ème intron (**Figure II.3**, IVS3). Ce répresseur régule négativement l'activité de la transposase et stimule la rétention de l'intron (boucle positive). Le deuxième modèle est apparu beaucoup plus tard puisqu'il repose sur la voie piARN. Il rend beaucoup mieux compte des effets observés dans le cadre de la dysgenèse hybride et donc c'est ce modèle que je vais développer.

### 1. Déterminants chromosomiques : les clusters de piARN

Des études génétiques ont montré que le cytotype P était associé à la transmission de régions chromosomiques précises contenant des éléments P autonomes (Ronsseray et al., 1991) ou dégénérés (Nitasaka et al., 1987). De plus, si des éléments P conférant l'effet répresseur sont introduits par transgénèse à d'autres positions du génome, ils perdent leur effet (Jensen et al., 2008). Ce n'est donc pas une séquence particulière d'un élément P qui est importante mais plutôt sa localisation dans le génome. Plusieurs études ont montré que la présence d'éléments P dans une région subtélomérique du chromosome X (X-TAS) est suffisante pour expliquer le cytotype P (Ronsseray et al., 1991; Stuart et al., 2002). Lorsqu'on croise une femelle hétérozygote pour ce locus avec un mâle P alors la dysgenèse est limitée dans les F1, indépendamment de leur génotype. En revanche, seules les F1 femelles ayant hérité du locus peuvent à leur tour protéger leur descendance (**Figure II.5**). Le génotype des femelles, à savoir la possession d'éléments P dans la région X-TAS, est donc essentiel pour pouvoir réprimer la dysgenèse en F1. Or cette région est le 4ème cluster de piARN le plus prolifique (Brennecke et al., 2007), ce qui colle parfaitement avec une explication de la dysgenèse par la voie piARN. Enfin, une étude a montré que le cluster X-TAS n'est pas le seul capable de réprimer l'activité de l'élément P dans la lignée germinale. Un transgène intégré dans le bras 2R peut avoir cette capacité et plus encore, il peut convertir d'autres transgènes par un phénomène de paramutation (de Vanssay et al., 2012). La capacité d'un locus à produire des piARN peut donc dépendre de facteurs épigénétiques.



**Figure II.5 : Protection contre la dysgénèse par les éléments P du locus X-TAS.**

La dysgénèse hybride est évaluée par la présence de la dysgénèse gonadique des femelles (GD) et l'activité relative de l'élément P (signes +). Le carré rouge désigne le locus X-TAS comportant des éléments P. Seules les femelles portant ce locus peuvent protéger partiellement leur descendance après un croisement avec un mâle inducteur P. (Kelleher, 2016).

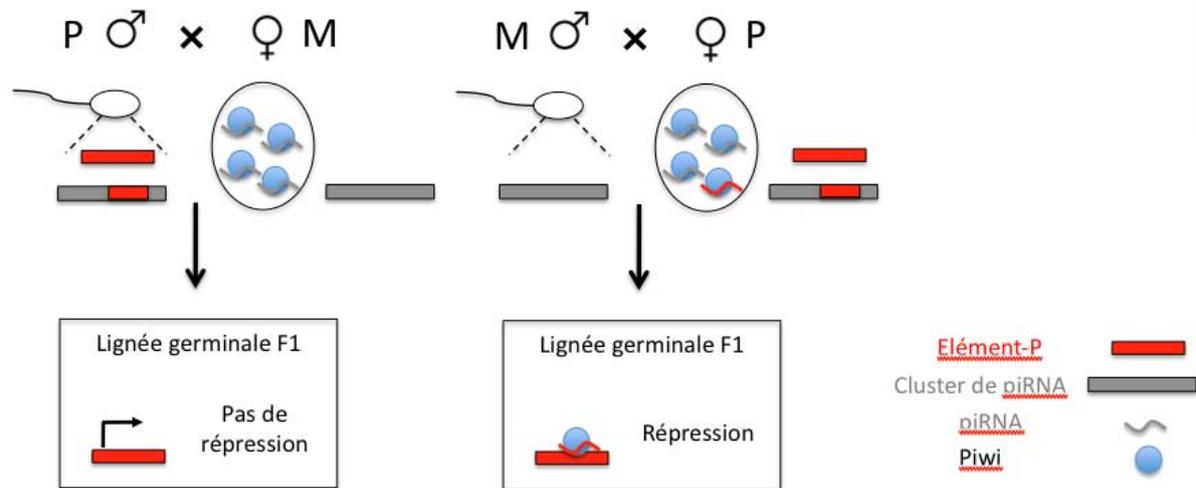
## **2. Déterminants cytoplasmiques : les piARN maternels**

D'après le fonctionnement de la voie chez la drosophile, des piARN maternels doivent être déposés dans l'ovocyte afin de réprimer les transposons dans la future lignée germinale. En effet, des études montrent que la présence d'éléments P dans des clusters est associée à une forte production de piARN dans les ovaires des souches de cytotype P (Brennecke et al., 2008; Khurana et al., 2011). Ces piARN se retrouvent directement dans les embryons des individus F1 puis permettraient aux futurs ovaires de produire à leur tour des piARN via le cycle ping-pong. Leur absence corrèle avec la dysgénèse des F1. Pour renforcer l'implication de la voie piARN dans le phénomène, on peut s'appuyer sur l'observation de différents mutants. Des femelles de cytotype P portant une mutation perte de fonction dans un de ses allèles Piwi ou Aub n'est plus capable de réprimer aussi efficacement l'activité de l'élément P dans sa descendance (Simmons et al., 2010).

## **3. Modèle**

Le modèle ci-dessous (**Figure II.6**) est une interprétation de la dysgénèse hybride à travers la voie des piARN. Une souche P contient des dizaines de copies de l'élément P, complètes ou dégénérées, dont certaines sont intégrées dans des clusters de piARN.

Une souche M n'en contient aucune et ne peut donc pas produire de piARN dirigés contre l'élément P. Ainsi une femelle M ne pourra pas déposer de piARN maternels dans ses ovocytes ce qui permet aux éléments P paternels de transposer dans la lignée germinale. Dans le croisement réciproque, les éléments P sont réprimés par les piARN maternels déposés dans l'ovocyte.



**Figure II.6: Modèle de la dysgénèse hybride expliquée par la voie piARN.**

### 3) Adaptation des F1 dysgéniques à l'élément P

#### 1. La stérilité initiale

D'une part, l'atrophie des ovaires est très clairement visible chez les femelles dysgéniques, indiquant un arrêt du développement de ces tissus. D'autre part, nous avons vu que la dysgénèse est attribuée à la mobilisation de l'élément P dans ces tissus. Entre les deux, de nombreux mécanismes restent à découvrir pour comprendre comment la mobilisation d'un transposon peut conduire à un tel phénotype. Une récente étude a apporté beaucoup d'éléments de réponse à cette question et je vais en faire un résumé (Khurana et al., 2011). Premièrement, les événements de transpositions génèrent des cassures double-brin qui sont détectables et entraînent probablement un arrêt du cycle cellulaire. Cela peut suffire à expliquer l'arrêt du développement de la lignée germinale et donc la stérilité. Parallèlement, les acteurs principaux de la voie piARN (Ago3, Aub et Piwi) présentent une localisation subcellulaire anormale : il n'y a pas de *nuage*. Un autre composant essentiel de la voie piARN, l'hélicase Vasa, voit son expression réduite dans des ovaires dysgéniques. Les auteurs postulent que la voie des dommages à l'ADN est responsable de la phosphorylation et de la dégradation ubiquitine-dépendante de Vasa, ce qui entraînerait la désorganisation du *nuage*. En conséquence, d'autres familles de transposons sont mobilisés et amplifient les cassures dans le génome.

## **2. Evolution de la situation avec le vieillissement**

Toujours dans la même étude, les auteurs rappellent que la fertilité des femelles dysgéniques s'améliore en vieillissant, avec environ un tiers des ovaires produisant des œufs à l'âge de 21 jours. Contrairement aux jeunes drosophiles (2-4 jours), les adultes possèdent un *nuage* fonctionnel et un niveau normal de répression de tous les transposons. Concernant l'élément P, il existe une production *de novo* de piARN à partir de transcrits paternels dérivant probablement de clusters. Très peu de nouvelles insertions dans des clusters ont été observées et aucune n'est retrouvée à la génération suivante. Ce constat s'oppose au modèle des clusters agissant comme des pièges à transposons. En revanche, parmi les familles résidentes mobilisées, une centaine de nouvelles insertions ont été transmises à la génération suivante dont 6 dans des clusters connus de piARN. Elles permettent aux ovaires des F1 dysgéniques de produire *de novo* des piARN qui doivent participer à la réduction de l'activité des transposons dans la génération F2.

## **B. L'isolement reproductif**

### **1) Généralités**

La découverte de la dysgenèse hybride chez la drosophile a rapidement poussé les scientifiques à postuler que les transposons puissent être sources de spéciation (Rose and Doolittle, 1983). La stérilité des F1, même partielle, est un cas d'isolement reproductif post-zygotique qui défavorise les accouplements entre les populations qui les génèrent. Dans ce modèle, les individus se reproduisant davantage avec des partenaires donnant une descendance 100 % fertile ont un meilleur succès reproductif. Il peut s'agir d'un biais comportemental, géographique ou simplement mécanique. La sélection des allèles portés par ces individus va amplifier ce biais jusqu'à ce que les deux populations forment deux nouvelles espèces possédant des barrières secondaires pré-zygotiques. Dans le cas de la dysgenèse évoquée jusqu'à présent, c'est la transposition non contrôlée dans la lignée germinale qui est à l'origine de la stérilité. On attribue à la mobilité des transposons une autre cause de stérilité. Leur dissémination dans les génomes pourrait poser problème à l'appariement des chromosomes homologues en méiose et donc à la formation des gamètes (Rose and Doolittle, 1983).

Dans la thématique autour des transposons et de la spéciation, j'ai choisi de distinguer deux questions. Premièrement, est-ce que les transposons peuvent initier une spéciation ? Autrement dit, l'invasion d'un transposon dans une population peut-elle conduire à isoler cette population au point d'en faire une nouvelle espèce ? C'est cette question qui était posée suite à l'arrivée de l'élément P dans les souches de drosophiles sauvages. Répondre à cette question n'est pas facile car on peut difficilement suivre un tel événement sur une échelle de temps assez longue. Les études qui sont les plus appropriées traitent de croisements intraspécifiques qui ne diffèrent que par des contenus en ET variables. Si on utilise des croisements interspécifiques, on peut certes observer le comportement des ET dans les hybrides mais on peut difficilement le relier à la stérilité/viabilité et encore moins à l'origine de la spéciation. Beaucoup trop d'autres facteurs liés à la divergence des espèces peuvent

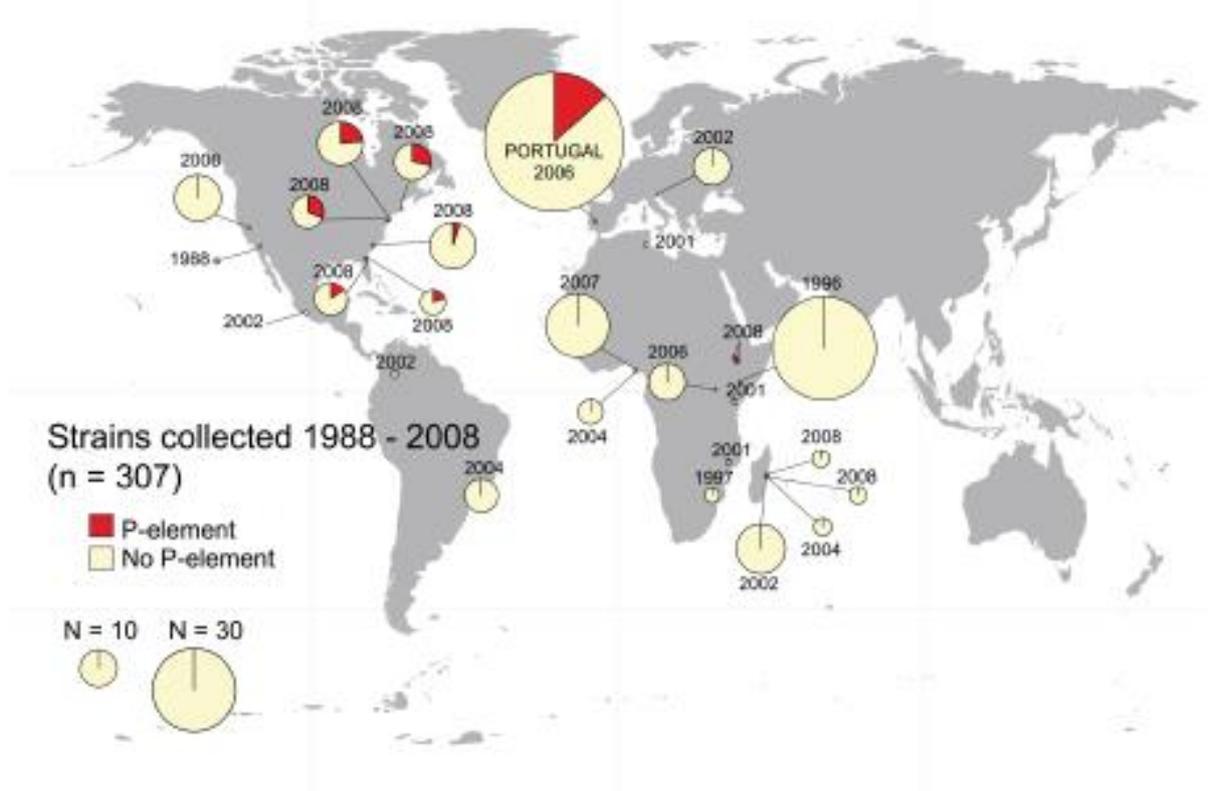
être impliqués. Ce deuxième type d'étude traite plutôt d'une autre question. Les transposons contribuent-ils à l'isolement reproductif post-zygotique entre espèces ?

## 2) Croisements intraspécifiques

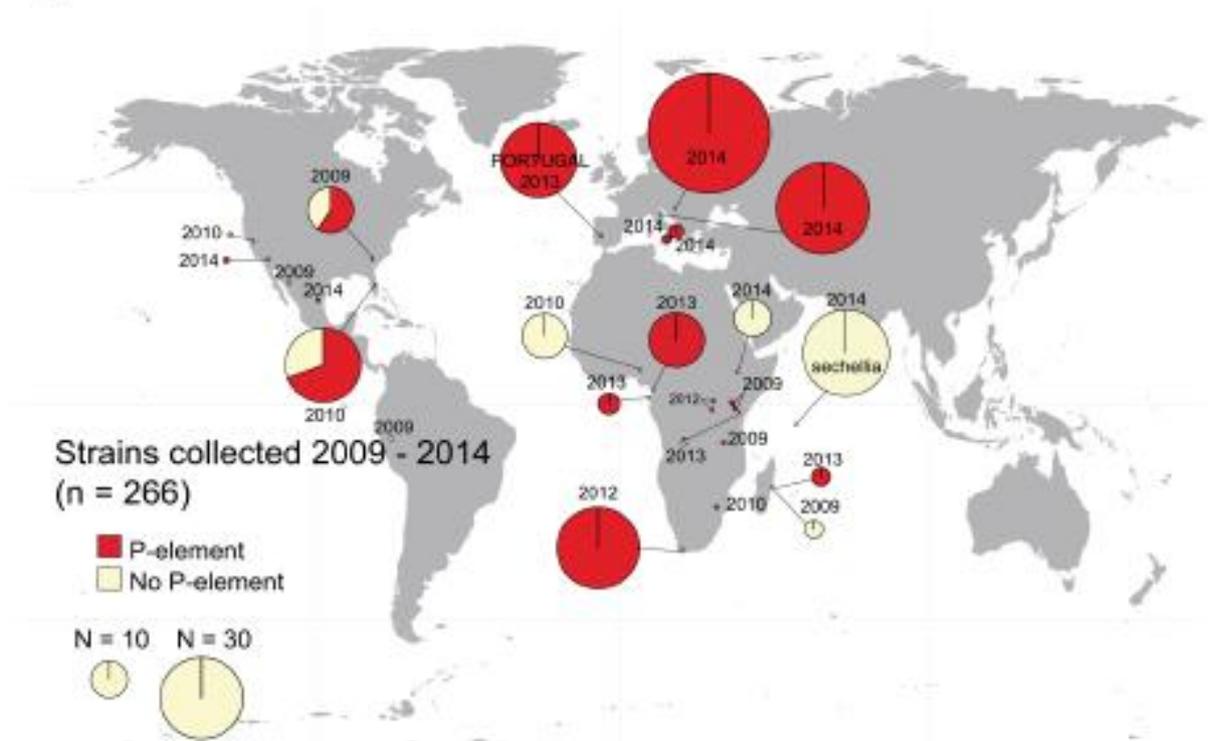
### 1. L'élément P chez *Drosophila melanogaster* et *simulans*

Nous avons vu que le système P-M induit une dysgenèse transitoire et que les femelles semblent rapidement contrer la transposition de l'élément P et ses conséquences (Khurana et al., 2011). En outre, depuis son arrivée au début du 20<sup>ème</sup> siècle, l'élément P semble avoir envahi la totalité de l'espèce à l'échelle du globe (Anxolabéhère et al., 1985), n'épargnant que les souches de laboratoires isolées physiquement. En tout cas, il est difficile d'imaginer que l'invasion de l'élément P pourrait maintenant séparer *D. melanogaster* en 2 espèces. A titre de comparaison, l'élément P a sauté vers l'espèce *D. simulans* en 2006 (Hill et al., 2016). La dysgenèse gonadique est moins marquée chez cette espèce et les auteurs pensent que cela explique pourquoi le transposon a réussi à envahir quasiment toute la population en si peu de temps (**Figure II.7**). Si une forte dysgenèse peut ralentir la propagation d'un ET chez la drosophile, une dysgenèse encore plus forte pourrait-elle générer un événement de spéciation ? Aucune information ne permet d'étayer cette hypothèse pour le moment.

A



B



**Figure II.7 : Présence de l'élément P chez *D. simulans* à l'échelle du globe.**

Les camemberts indiquent la proportion de souches portant l'élément P. Leur taille et localisation indiquent respectivement la taille et le lieu de l'échantillonnage (Hill et al., 2016).

## **2. Les autres dysgénèses hybrides liées à un seul transposon**

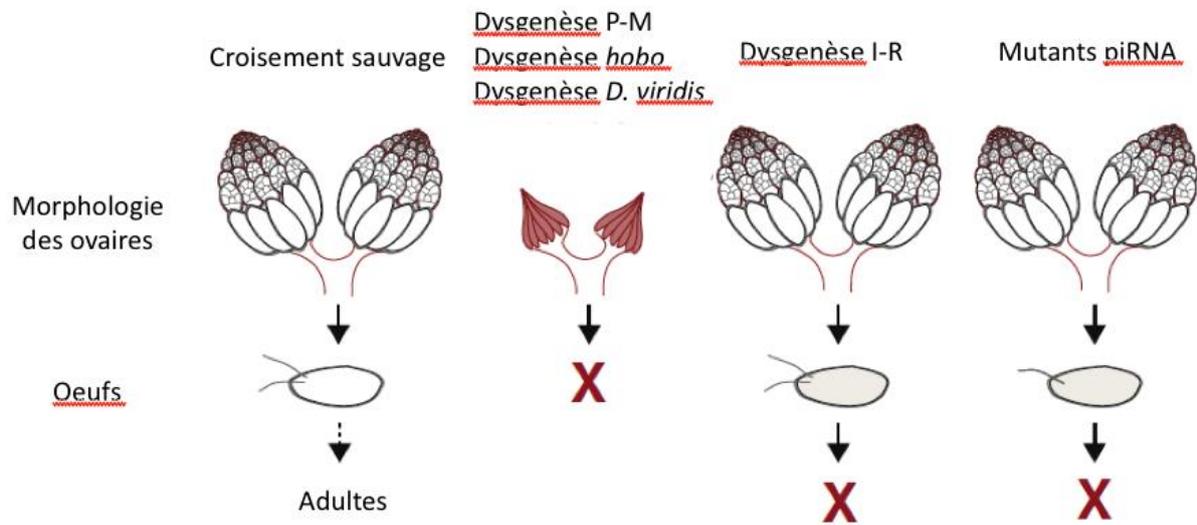
Peu de temps après la découverte du système P-M, le système I-R a été découvert. Dans ce dernier, une souche I (Inducer) porte des copies actives d'un rétrotransposon de la superfamille LINE et induit une dysgénèse lorsqu'un mâle s'accouple avec une femelle R (Reactive) (Bucheton, 1990). Similairement au système P-M, la dysgénèse s'atténue avec l'âge (Bucheton, 1979) et dépend de la production de piARN maternels (Chambeyron et al., 2008). En revanche, les mécanismes aboutissant à la stérilité semblent différents. Les ovaires se développent normalement chez les femelles dysgéniques et la dérégulation des transposons résidents est quasi nulle (Kelleher et al., 2012). De nombreuses figures de méiose sont aberrantes dans les ovocytes, résultant en des *pronuclei* femelles anormaux et donc des embryons peu viables (Orsi et al., 2010). Contrairement au système P-M, il n'y a pas d'enclenchement de la réponse aux dommages de l'ADN, possiblement grâce à la mobilisation de l'élément I par copier-coller. Toujours chez *D. melanogaster*, une troisième dysgénèse hybride a été découverte. Elle est liée à des transposons à TIR de la famille *Hobo* et ressemble phénotypiquement au système P-M mais elle n'a pas été étudiée en détail (Blackman et al., 1987).

## **3. Une dysgénèse liée à plusieurs familles de transposons chez *D. virilis*.**

Une nouvelle dysgénèse, persistante chez l'adulte, est connue lorsqu'un mâle 160 est croisé avec une femelle B9 chez *D. virilis* (Lozovskaya et al., 1990) (**Figure II.8**). Elle est principalement causée par la dérégulation de l'élément *Penelope* existant dans la souche L160 mais pas dans la souche B9 (Evgen'ev et al., 1997). D'autres éléments (*Helena*, *Paris* et *Polyphemus*) sont aussi beaucoup plus abondants dans la souche 160 et contribuent probablement à la dysgénèse (Blumenstiel, 2014; Petrov et al., 1995). Étrangement, il existe des souches neutres pouvant réprimer la dysgénèse malgré l'absence de piARN dirigés contre l'élément *Penelope* (Rozhkov et al., 2013). De plus, la déposition de piARN maternels n'est pas clairement associée à la répression de cet élément (Rozhkov et al., 2010). Ainsi, il a été proposé que la source de cette dysgénèse ne soit pas l'élément *Penelope* en lui-même mais plutôt un problème global de l'activité des clusters conduisant à la dérégulation de nombreuses familles de transposons.

Pour tester cette seconde hypothèse, une étude assez complète à l'échelle du génome a été réalisée (Erwin et al., 2015). De nombreux ET sont dérégulés dans les F1 dysgéniques, dont 10 des 11 familles qui sont plus abondantes dans la souche 160 par rapport à la souche B9. L'abondance relative de ces familles chez les parents et la production de piARN explique partiellement les observations. Dans sa globalité la voie piARN n'est pas affectée chez les F1 mais les auteurs observent un enrichissement en piARN s'alignant sur des gènes cellulaires. L'expression globale du génome est modifiée, sans doute en partie à cause des piARN géniques. Enfin, une analyse de QTL (*Quantitative Trait Locus*) montre que la capacité à induire la dysgénèse est répartie sur l'ensemble du génome. En revanche, la capacité à la réprimer est plus localisée et dépend surtout de trois régions dont deux contiennent des clusters de piARN. Au final, le système décrit chez *D. virilis* est complexe, il ne peut pas se résumer à la simple dérégulation d'un ou plusieurs transposons présents en excès dans la souche 160 par rapport à la souche B9. La question de savoir si cette forme de dysgénèse pourrait davantage contribuer à un isolement reproductif reste

en suspens. Il n'existe pas d'autres dysgénèses de ce type décrites dans la littérature. Pour aller plus loin, il est nécessaire de s'intéresser à des hybrides interspécifiques.



**Figure II.8 : Bilan des dysgénèses connues chez la drosophile.** Adapté de (Malone et al., 2015)

### 3) Croisements interspécifiques

#### 1. Etudes non conclusives chez la drosophile.

Puisque la dysgénèse fut découverte chez la drosophile, c'est au sein du genre que les premières études furent menées. Elles cherchaient à savoir si la mobilisation des transposons observée chez les hybrides (P-M, I-R) était également retrouvée chez les hybrides interspécifiques. Dans ce cas la dérégulation des ET pourrait être un acteur majeur de l'isolement reproductif post-zygotique, et pourquoi pas une source de spéciation. Afin d'étudier le phénomène, le plus simple consistait à regarder l'apparition de mutations ou l'augmentation du taux de recombinaison. Plusieurs publications sont résumées dans la **table 1**. Les premières études étaient plutôt encourageantes mais la théorie prit un coup d'arrêt dans les années 80. D'autres publications attribuaient plutôt la stérilité des hybrides à des loci particuliers, ce qui était difficilement conciliable avec le modèle de mobilisation globale. Pour s'intéresser à la transposition d'éléments particuliers, des expériences de FISH furent menées sur les chromosomes polyténiques des drosophiles. Par exemple une étude a montré la mobilisation de l'élément *Oswaldo* (Labrador et al., 1999) alors qu'une autre n'a trouvé aucune évidence de transposition pour les éléments *copia* et *mdg1* (Guerreiro, 1996).

Espèces	Mesure	Dysgénique / Non dysgénique	Référence
<i>D. pseudoobscura ssp</i>	Mutations visibles	Taux plus élevé chez les hybrides	(Sturtevant, 1939)
<i>Drosophila subgroup affinis</i>	Mutations visibles	Taux plus élevé chez les hybrides	(Miller, 1950)
<i>D. pseudoobscura x D. pseudoobscura bogotana</i> <i>D. simulans x D. sechellia</i> <i>D. virilis vs. D. lummei</i>	Recombinaison chez le mâle	Pas de différence	(Coyne, 1986)
<i>Drosophila subgroup affinis</i>	Mutations visibles	Pas de différence	(Hey, 1988)
<i>D. simulans x D. mauritiana</i>	Mutations visibles liées à l'X	Pas de différence	(Coyne, Jerry A, 1989)
<i>D. serido x D. buzzatii</i>	Réarrangements chromosomiques	Plus élevés chez les hybrides	(Naveira and Fontdevila, 1985)

**Table 1 : Etudes d'hybrides interspécifiques chez la drosophile.**

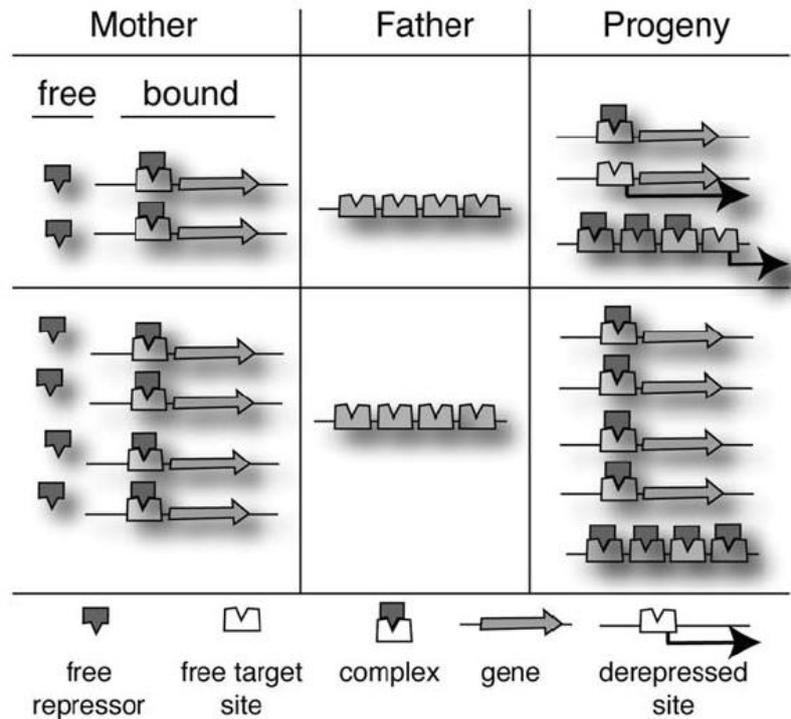
## **2. Mobilisation des ET et phénotypes associés.**

En parallèle de la drosophile, la question préoccupait beaucoup les généticiens des plantes chez qui l'hybridation est un phénomène courant. McClintock avait déjà observé que la transposition était induite par des cassures de chromosomes, elle avait alors proposé que c'était une réponse du génome face à un 'choc génomique' (McClintock, 1984). Ces chocs peuvent être de différentes natures telles qu'un accident de mitose, une infection virale ou encore une caryogamie entre deux gamètes d'espèces différentes. Puisque les hybridations sont fréquentes chez les plantes et s'accompagnent typiquement d'un doublement du matériel génétique (allopolyploïdisation), de nombreuses études se sont intéressées au génome d'hybrides naturels ou artificiels (croisements forcés). De manière générale, il a été observé une instabilité génomique caractérisée par la mobilisation de transposons, un défaut de méthylation de l'ADN, des modifications de l'expression des gènes et des réarrangements chromosomiques (Comai et al., 2003). En réalité, peu d'études montrent des *bursts* de transpositions, il s'agit plutôt d'événements d'insertions des ET les plus actifs qui ne sont pas contre-sélectionnées du fait de la ploïdie (Parisod et al., 2010). Quoiqu'il en soit, la fréquence élevée d'hybridation chez les plantes témoigne de la faiblesse des ET en tant que barrière post-zygotique. Les scientifiques du domaine considèrent davantage le rôle des ET dans la spéciation par leur capacité à réorganiser le caryotype des hybrides par recombinaison entre des courts loci homologues dérivés d'un même ET ancestral. Les hybrides deviennent alors incapables de se reproduire avec leurs deux espèces parentes, d'où la création d'une nouvelle espèce hybride (Templeton, 1981). Cette observation ne se limite pas aux plantes puisque des réarrangements induits par transposition semblent avoir largement contribué à la spéciation de certains poissons (Symonová et al., 2013). En outre, une autre étude sur des poissons montre qu'une hybridation entraîne la

surexpression de nombreux ET et la sous-expression de l'ADN méthyltransférase qui est censée les réprimer (Dion-Côté et al., 2014). Du côté des marsupiaux, des hybrides présentent une réactivation d'éléments rétroviraux endogènes, des réarrangements chromosomiques et une sous-méthylation globale de leur génome (O'Neill et al., 1998). Au final, tous les hybrides interspécifiques présentent des caractéristiques similaires, quel que soit le taxon eucaryote considéré. La mobilisation des ET pourrait être la cause initiale des autres phénotypes (réarrangements chromosomiques, dérégulations de gènes, etc.) mais l'inverse est tout aussi plausible.

### **3. Les modèles de défauts de répresseurs maternels chez les plantes.**

Les plantes ne possèdent pas de protéines Piwi, elles utilisent des complexes [Ago / siARN endogènes] pour cibler les transposons dans les gamétophytes (pollen et sac embryonnaire). Pour expliquer la dérèpression des ET, on retrouve souvent dans la littérature l'hypothèse selon laquelle un défaut de répresseurs maternels (siARN) peut conduire à la transposition d'éléments paternels dans le zygote (Parisod and Senerchia, 2012). A titre d'exemple, la transmission paternelle de l'élément actif EVADE entraîne un *burst* de transposition dans l'embryon alors que sa transmission maternelle le maintient sous contrôle (Reinders et al., 2013). Etrangement, la voie siARN du sac embryonnaire n'est pas impliquée dans le processus, suggérant que l'état réprimé est hérité de la macrospore. La protection maternelle ne dépendrait donc pas du pool maternel de siARN. Chez des hybrides *A. thaliana* x *A. arenosa*, la dysgenèse est associée à la dérèpression de l'élément paternel *Athila* et l'activation de plusieurs gènes normalement réprimés par le complexe Polycomb (Josefsson et al., 2006) ; les siARN ne sont pas étudiés. La mortalité des hybrides est principalement attribuable à l'activation d'un seul gène et elle est curieusement réduite si le génome maternel est en excès (parent tétraploïde). La publication propose alors que la dysgenèse résulterait d'un défaut maternel de complexes répresseurs de la chromatine qui doivent couvrir la nudité de l'ADN paternel après décondensation (**Figure II.9**).

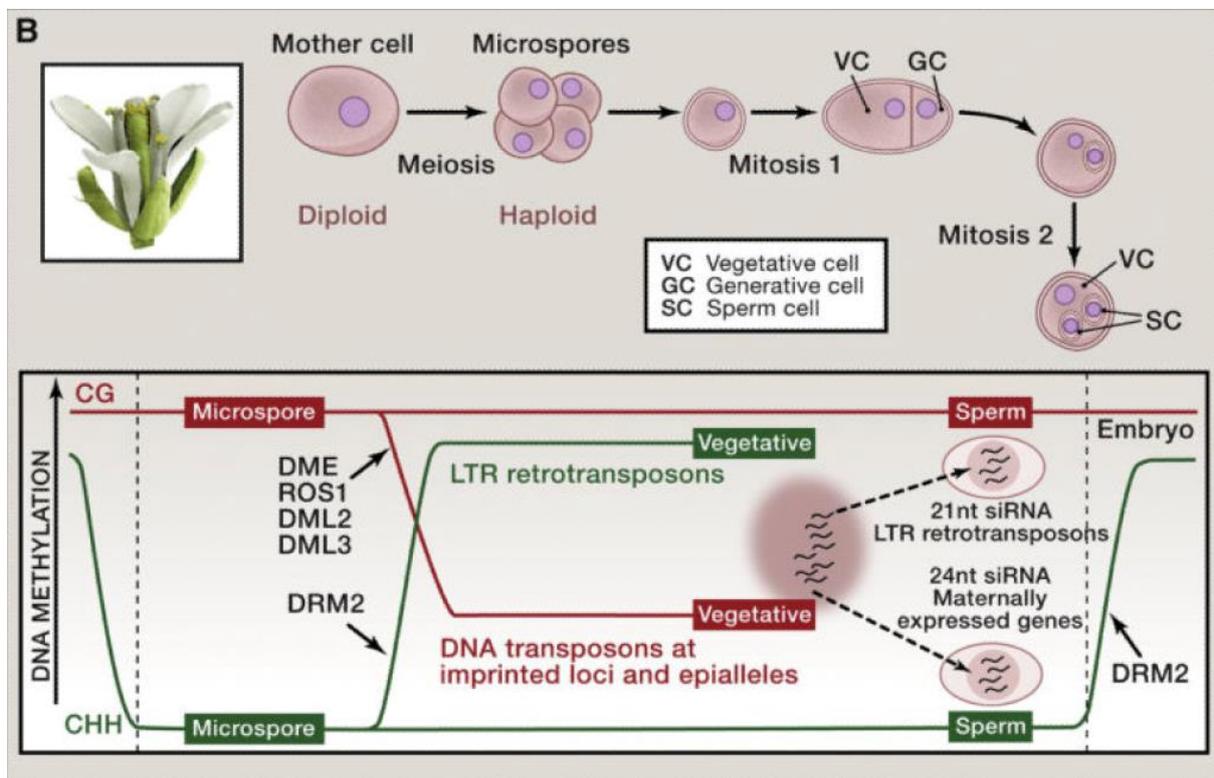


**Figure II.9 : Modèle d'activation des gènes réprimés dans un croisement hybride.**

Le sac embryonnaire porte un nombre limité de régulateurs chromatinien (complexes Polycomb, siARN). Le gamète mâle apporte de nombreux sites qui doivent être ciblés par ces facteurs après la décondensation consécutive à la fécondation. Dans l'exemple, Les répresseurs de la chromatine sont en défaut dans le croisement 'mère 2n x père 2n' et suffisants dans le croisement 'mère 4n x père 2n'. Dans le premier cas, la redistribution des facteurs entre les deux génomes peut conduire à la dérèpression de gènes paternels et maternels (Josefsson et al., 2006).

Pour appuyer ce modèle, les études du niveau de méthylation de l'ADN chez *Arabidopsis* sont très précieuses. Il a été montré que la méthylation CG (et CHG) est maintenue dans la lignée germinale paternelle (noyaux spermatiques) sauf pour quelques loci incluant les régions exprimées maternellement (Calarco et al., 2012). En revanche la méthylation CHH, qui assure la répression de centaines de rétrotransposons, est perdue, mais elle est compensée par la présence de siARN de 21 nt. Ces siARN proviennent du noyau végétatif du grain de pollen et leur activité répressive a été vérifiée pour l'élément *Athila* (Slotkin et al., 2009). De même, les régions ayant perdu leur méthylation CG sont ciblées par des siARN de 24 nt dans les noyaux spermatiques (Calarco et al., 2012). Après la fécondation, le génome paternel connaît une restauration de sa méthylation CHH, or la grande majorité des siARN de 24 nt nécessaires à ce processus sont produits par le génome maternel (Mosher et al., 2009). Ce modèle (**Figure II.10**) explique très bien comment des éléments paternels peuvent s'activer dans l'embryon s'ils ne sont pas connus ou s'ils sont sous-représentés dans le génome maternel, comme c'est le cas pour les hybrides *A. thaliana* x *A. arenosa* étudiés par Josefsson. Pour autant, il n'y a pas, à ma connaissance, d'études à l'échelle du génome montrant une dérégulation des

éléments paternels dans des hybrides corrélée avec des niveaux de siARN. Par exemple, il a été observé que des hybrides du blé présentent une chute globale des siARN ciblant les ET mais aucune corrélation claire entre la quantité de siARN détectée et l'activité des ET n'a été trouvée (Kenan-Eichler et al., 2011).

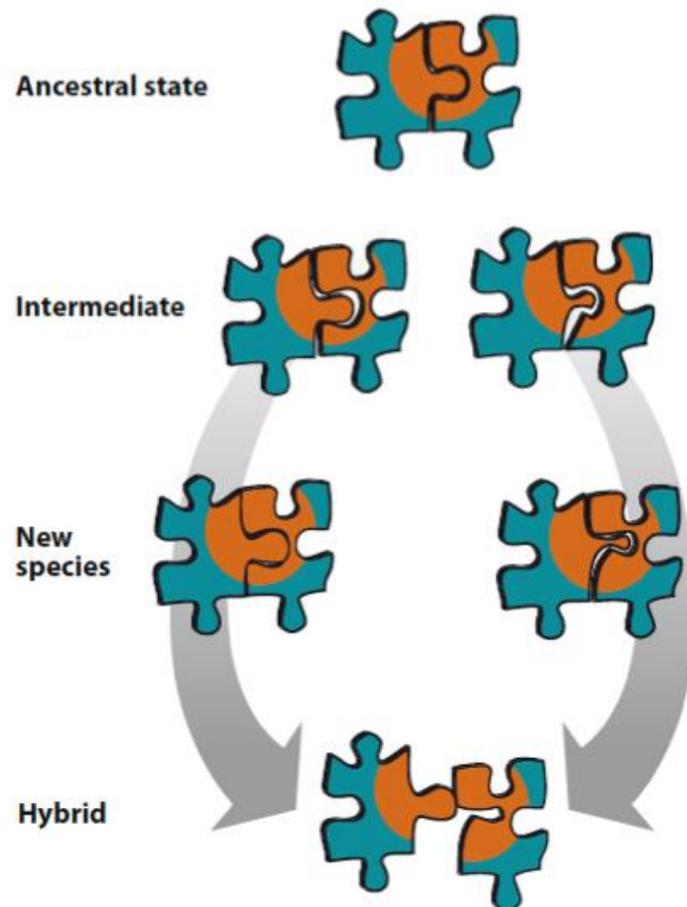


**Figure II.10 : Méthylation de lignée germinale paternelle au cours de la reproduction chez les plantes.** (Heard and Martienssen, 2014)

#### 4. Les modèles de divergence des protéines

Une autre explication pour la dérégulation des transposons chez les hybrides interspécifiques est parfaitement illustrée dans l'étude des hybrides de *D. melanogaster* x *D. simulans* (Kelleher et al., 2012). On y apprend que de nombreuses familles de transposons sont dérégulées dans les ovaires des F1, indépendamment des pools de piARN présents. La voie des piARN est globalement altérée chez l'hybride : défaut de production des piARN, réduction de la signature ping-pong, disruption partielle du *nuage* et activité aberrante des clusters. Ces observations sont caractéristiques des souches mutées dans un gène Piwi, suggérant un problème dans l'activité de ces protéines. Enfin, une de leurs expériences montre qu'un mutant *D. melanogaster aub* ne peut pas être totalement complété par un transgène *aub* de *D. simulans*. Cela signifie que la protéine de cette espèce n'est pas entièrement fonctionnelle dans le fond génétique de l'autre espèce. Tous ces éléments tendent à montrer que la dérégulation des ET observée chez ces hybrides interspécifiques n'est pas liée au contenu relatif en ET entre les deux espèces parentales. Elle découlerait plutôt de la divergence entre les protéines effectrices de la voie piARN qui ne

fonctionneraient pas efficacement ensemble. Cette hypothèse rejoint en fait un modèle classique de spéciation induite par la divergence des allèles créant une incompatibilité chez l'hybride : le modèle BDM (cf. IV-A-3) (**Figure II.11**).



**Figure II.11 : Modèle de spéciation par incompatibilité entre allèles.**

Dans l'espèce ancestrale, deux loci codent des protéines interagissant ensemble. Des mutations s'accumulent dans deux sous-populations et induisent une coévolution spécifique entre les deux partenaires. Chez l'hybride, les allèles sont incapables d'interagir, ce qui provoque la perte d'une fonction essentielle. Adapté de (Maheshwari and Barbash, 2011)

Si c'est la divergence des protéines de la voie piARN qui explique principalement la dérégulation des ET, cela ne veut pas dire que les ET ne sont pas acteurs de la spéciation. En effet, au-delà des possibilités de transferts horizontaux, l'activité des ET peut connaître un regain sur une échelle de temps très courte (Bowen and McDonald, 2001). Les deux cas de figure nécessitent une adaptation rapide du système de défense de l'hôte qui accélère l'évolution des gènes impliqués et donc leur divergence. On sait par exemple que la variabilité de Vasa, un composant du *nuage*, se traduit par une variabilité de répression de l'élément *copia* (Vu and Nuzhdin, 2011). L'adaptation du génome face aux ET peut aussi impliquer des protéines associées à l'hétérochromatine dont la divergence de certaines a été identifiée comme source de stérilité (Bayes and Malik, 2009; Brideau et al., 2006).

# III – Les ciliés et leurs réarrangements génomiques

## A. Généralités sur les ciliés

### 1) Position phylogénétique

La dernière classification des eucaryotes propose cinq groupes majeurs (Adl et al., 2012) (**Figure III.1**), en se basant principalement sur la séquence très conservée de l'ARN 18S de la petite sous-unité ribosomale. Les ciliés appartiennent au groupe des alvéolés, au même titre que les dinoflagellés et les apicomplexans (*Plasmodium falciparum* par exemple). Ils sont rassemblés aujourd'hui avec les straménopiles (diatomées, oomycètes,) et les rhizaires (foraminifères, cercozoaires) au sein du clade « SAR ». Ils doivent leur nom à la présence de cils vibratiles à leur surface qui leur servent à nager et à capturer leurs proies.

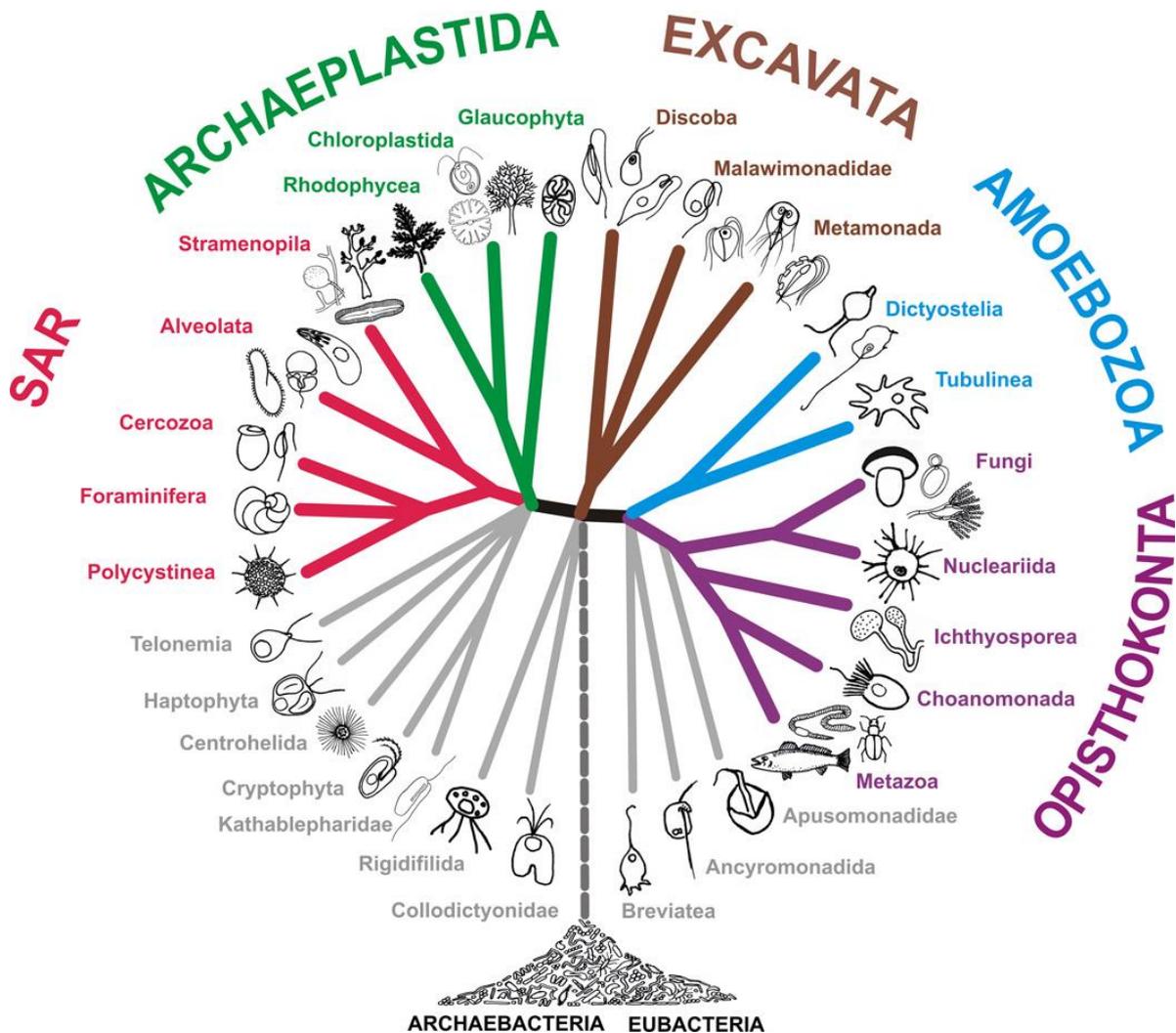
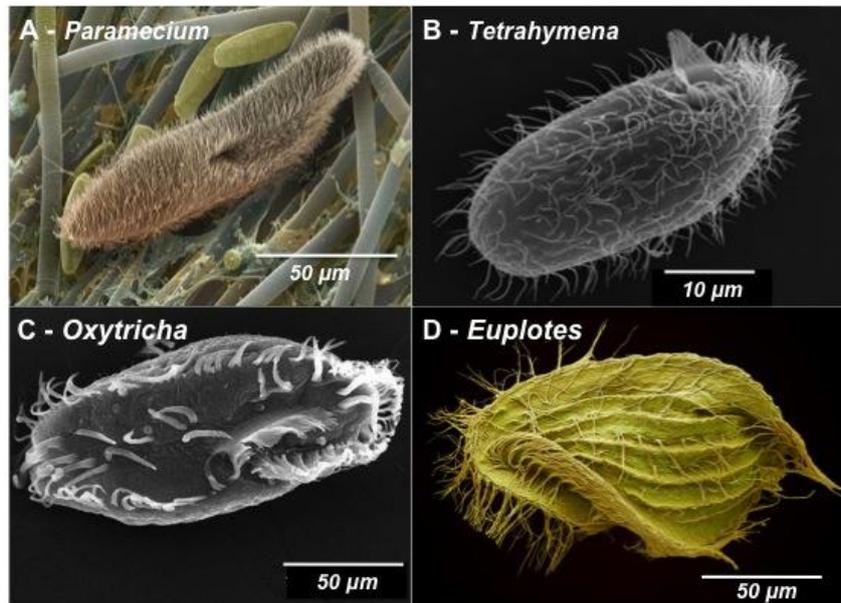
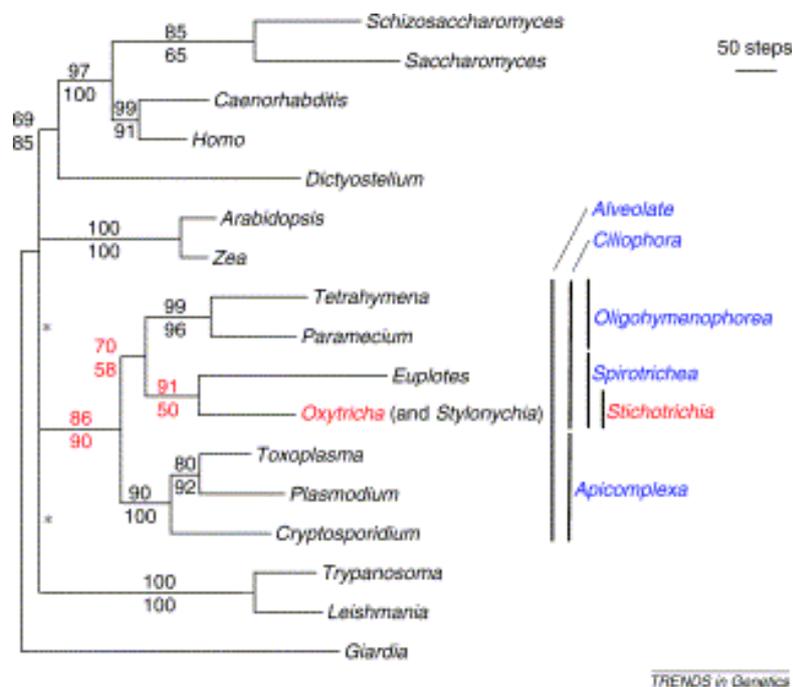


Figure III.1 : Arbre phylogénétique des eucaryotes. (Adl et al., 2012)

Les ciliés les plus étudiés, pour des raisons historiques sont *Paramecium tetraurelia* et *Tetrahymena thermophila*. Ces deux espèces peuvent sembler proches morphologiquement (**Figure III.2**) et phylogénétiquement par rapport à d'autres ciliés. On estime pourtant qu'une centaine de millions d'années les séparent, soit beaucoup plus qu'entre l'homme et le nématode par exemple (**Figure III.3**). D'autres ciliés sont également beaucoup étudiés : *Oxytricha trifallax*, *Stylonychia lemnae* ainsi que le genre *Euplotes*.



**Figure III.2:** Images de microscopie électronique de quelques ciliés.



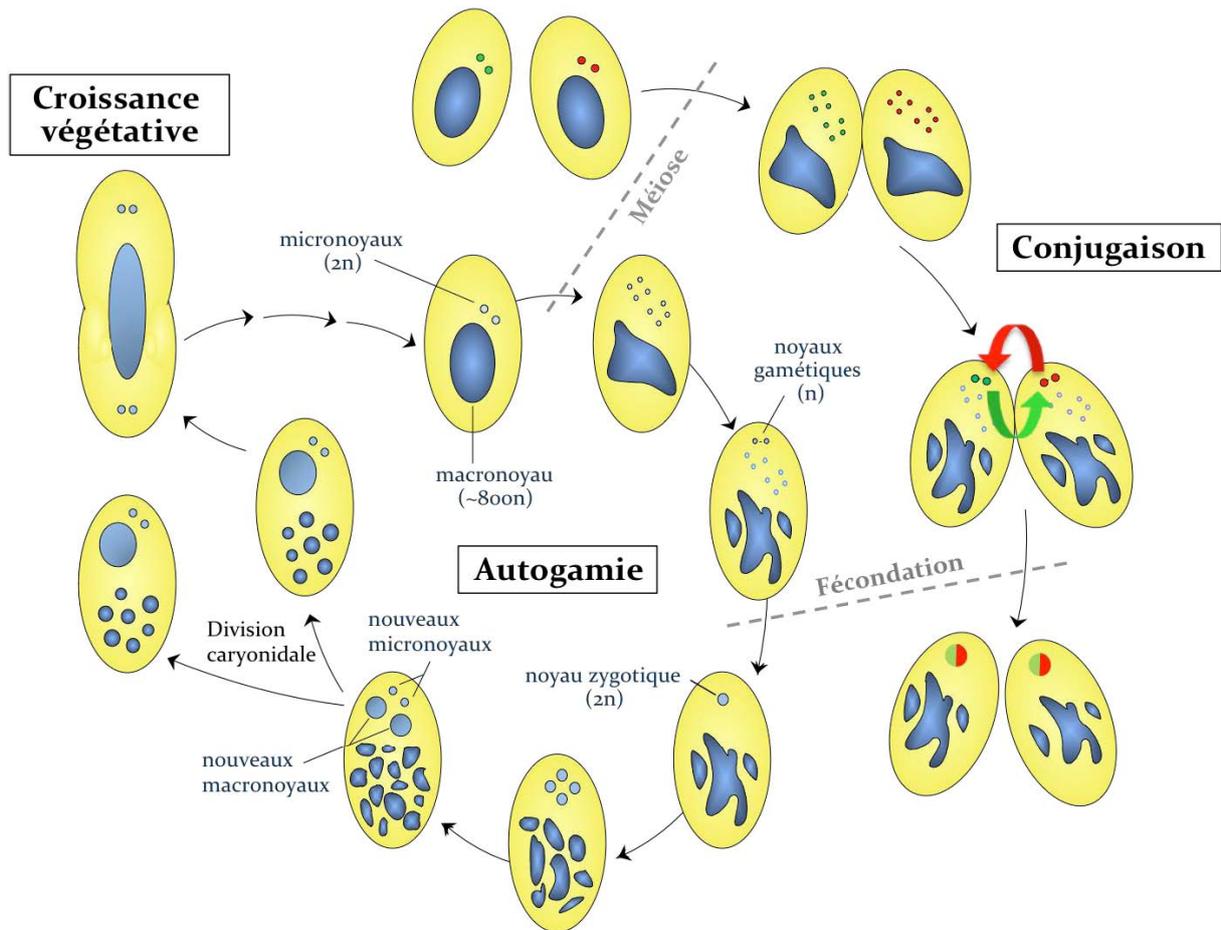
**Figure III.3:** Phylogénie de quelques ciliés. (Doak et al., 2003)

## 2) Dimorphisme nucléaire et cycle sexuel

Les ciliés partagent une autre synapomorphie tout à fait singulière : deux types de noyaux coexistent dans le même cytoplasme. Le macronoyau (MAC) a une fonction somatique alors que le micronoyau (MIC) a une fonction germinale. Le MAC assure la transcription de tous les gènes et exprime ainsi le phénotype de la cellule. Lors d'une division végétative, il double son ADN avant de se répartir plus ou moins équitablement entre les deux cellules filles par un mécanisme amitotique. Sa ploïdie est élevée, allant de 45 n pour *Tetrahymena thermophila* à 800 n pour *Paramecium tetraurelia*. Le MIC en revanche n'est pas nécessaire à la croissance végétative des ciliés, il n'est d'ailleurs jamais transcrit pendant cette phase. Son rôle est d'assurer la transmission du patrimoine génétique de l'organisme à la génération sexuelle suivante.

Pour illustrer la dynamique de ces noyaux, nous allons utiliser le cycle sexuel de l'espèce *Paramecium tetraurelia* (**Figure III.4**). Le cycle sexuel est induit par une privation de nourriture. Les MIC, au nombre de deux chez cette espèce, entrent en méiose et produisent huit noyaux haploïdes dont un seul subsiste puis se divise, ce qui forme deux noyaux gamétiques génétiquement identiques. *P. tetraurelia* est capable de s'autoféconder formant donc par caryogamie un noyau zygotique entièrement homozygote, on parle d'autogamie. Deux mitoses successives généreront quatre copies dont les deux les plus antérieures formeront les nouveaux MIC alors que les autres formeront les nouveaux MAC (Grandchamp and Beisson, 1981). La première division cellulaire suivante donnera deux cellules appelées caryonides héritant chacune d'un nouveau MAC. Les deux caryonides frères ont des MIC identiques issus de mitoses alors que leurs MAC peuvent différer en fonction des réarrangements réalisés pendant leur développement. En parallèle, le MAC maternel est fragmenté tout au long de ce processus jusqu'à ce qu'il n'en reste 32 morceaux qui seront ensuite dilués au cours des divisions ultérieures.

La sexualité de *P. tetraurelia* ne se limite pas à l'autogamie (résultant de l'autofécondation), cette espèce est également capable de conjugaison. Cela est possible lorsque des cellules privées de nourriture sont en présence de partenaires compatibles. Les paramécies sont caractérisées par un type sexuel de manière à ce qu'une cellule de type O ne puisse s'accoupler qu'avec une cellule de type E (Sonneborn, 1937). La conjugaison se passe initialement comme l'autogamie dans les deux cellules accouplées (**Figure III.4**). Au stade « deux noyaux gamétiques », un échange réciproque de noyaux, suivi de la caryogamie, forme deux noyaux zygotiques identiques entre eux, un dans chaque partenaire. Le couple se détache ensuite et chaque cellule termine le processus comme à l'autogamie. Lorsqu'une cellule vient de terminer une autogamie ou une conjugaison, elle est incapable de réaliser immédiatement une nouvelle autogamie. Environ 15 à 20 divisions sont nécessaires alors qu'il n'existe aucune restriction pour la conjugaison.



**Figure III.4 : Cycle de vie de *Paramecium tetraurelia*.**

La paramécie se reproduit par scissiparité en croissance végétative. Elle entre en cycle sexuel (autogamie ou conjugaison) en cas de privation de nourriture. La méiose des deux MIC forme huit noyaux haploïdes dont un seul subsiste puis se divise, ce qui forme deux noyaux gamétiques génétiquement identiques. Dans le cas de l'autogamie, l'autofécondation génère un noyau zygotique entièrement homozygote. Dans le cas de la conjugaison, les deux partenaires se fécondent réciproquement et forment chacun le même noyau hétérozygote. Le cycle se poursuit par deux mitoses du noyau zygotique qui formeront les deux nouveaux MIC et les deux nouveaux MAC. La première division qui suit le cycle produit deux caryonides frères héritant chacun d'un des deux MAC. En parallèle, le MAC maternel est fragmenté tout au long du cycle.

### 3) Les génomes des ciliés

#### a. Composition des génomes MAC

D'après ce que nous venons de voir, chaque cilié possède, en plus du génome mitochondrial, deux génomes contenus dans deux types de noyaux. Le génome MAC, par sa forte ploïdie, est le plus accessible. En effet, si l'on prend l'exemple de *Paramecium tetraurelia* qui contient deux MIC diploïdes et un MAC 800 n, le ratio MAC/MIC d'une cellule est d'environ 200. Les efforts de la communauté '*Paramecium*' d'une part et '*Tetrahymena*' d'autre part, ont permis d'obtenir la séquence de ces deux génomes MAC en 2006 d'une taille respective de 72 et 104 Mb (Aury et al., 2006; Eisen et al., 2006). La présence de séquences télomériques à l'extrémité des scaffolds assemblés permet d'estimer que ces génomes sont constitués

d'environ 100 et 225 chromosomes. Ces génomes sont très riches en AT (72 et 80 %) et relativement denses en séquences codantes (78 et 48 %) avec des introns de petite taille (25 et 165 pb en moyenne). Ils contiennent très peu de séquences répétées par comparaison aux autres génomes connus. Le code génétique des oligohyménozoaires est également particulier puisque seul le codon stop UGA est utilisé alors que les codons UAA et UAG sont traduits en glutamine (Caron and Meyer, 1985; Preer et al., 1985).

### **b. Composition des génomes MIC**

Alors que les 5 chromosomes micronucléaires de *Tetrahymena* sont visibles en méiose/mitose, le nombre de chromosomes de *Paramecium tetraurelia* n'est pas connu car ils sont vraisemblablement trop petits. Une analyse cytologique sur 3 autres espèces de *P. aurelia* a montré qu'elles en contiennent entre 30 et 60 paires (Jones, WK, n.d.). Les connaissances initiales sur le génome MIC ont été établies par purification des MIC puis clonage dans des phages afin d'obtenir des banques (Preer et al., 1992; Yao, 1982). Ainsi les génomes MIC et MAC ont pu être comparés pour un certain nombre de loci et de nombreuses séquences MIC-spécifiques ont été trouvées (voir III-C). Avec l'aide de techniques modernes, le génome de *Tetrahymena* a pu être révélé l'année dernière (Hamilton et al., 2016). Concernant *P. tetraurelia*, F. Guerin a mis au point une technique de purification par FACS des micronoyaux et le séquençage a révélé que le génome MIC contient environ 100 Mb (non publié). Ainsi 20 % du génome MIC est absent dans le génome MAC.

## **B. Méthodes d'études chez *Paramecium tetraurelia***

Pour bien comprendre la nature des expériences menées chez cet organisme, il convient de se pencher sur sa biologie. C'est elle qui a fait de lui un modèle au siècle dernier mais c'est aussi elle qui résiste à un certain nombre de techniques pourtant classiques chez d'autres organismes.

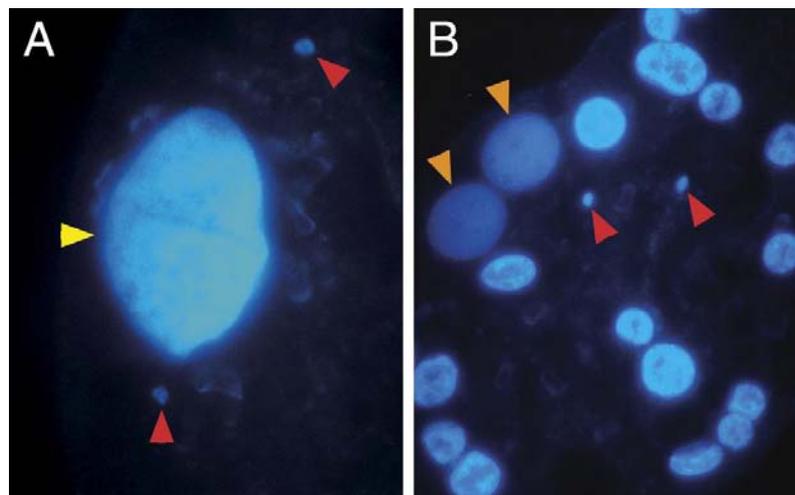
### **1) Cultures**

Les techniques de cultures utilisées aujourd'hui sont héritées de Sonneborn (Sonneborn, 1950) et ont été réécrites au goût du jour ici (Janine Beisson et al., 2010). *P. tetraurelia* se nourrit de bactéries et croît de façon optimale à 27 °C avec un rythme de 4 à 5 divisions par jour. Le milieu de culture standard est constitué d'herbe de blé infusé additionné d'un tampon etensemencé la veille de son utilisation par une souche de *Klebsiella pneumoniae*. La maîtrise de la culture requiert le contrôle du cycle sexuel : tant que les cellules sont nourries, elles vont rester en phase de croissance végétative. Il existe cependant une limite car elles entrent en sénescence au-delà de 200 divisions. Avant d'atteindre ce stade, les cellules ont besoin d'un nouveau MAC en réalisant un événement sexuel (autogamie ou conjugaison).

Lorsqu'on travaille avec des cellules individuelles, on utilise des lames à dépressions qui sont maintenues dans une atmosphère saturée en vapeur d'eau dans des cristallisoirs. La croissance végétative est assurée par le repiquage quotidien, sous loupe binoculaire, d'une cellule dans une nouvelle dépression. En général, on commence un cycle de culture le lundi afin que les cellules repiquées le vendredi

soient suffisamment âgées (en termes de divisions) pour pouvoir tomber autogames dès que la privation de nourriture se fait sentir pendant le week-end (**Figure S1**). L'autogamie est vérifiée par coloration des noyaux au carmin ou au DAPI. La fragmentation du macronoyau témoigne du processus d'autogamie en cours (**Figure III.5**).

Il est également possible de cultiver les paramécies en masse afin d'en extraire des quantités d'ADN, d'ARN ou de protéines exploitables. Pour cela, on utilise un milieu plus riche que pour la culture en dépressions qui permet d'atteindre une densité comprise entre 3 000 et 5 000 cellules par mL. Les cellules sont récupérées par centrifugation dans des récipients en forme de poire permettant de bloquer les cellules dans le fond étroit avant qu'elles ne nagent vers la surface.



**Figure III.5 : Images fluorescentes de cellules marquées au DAPI.**

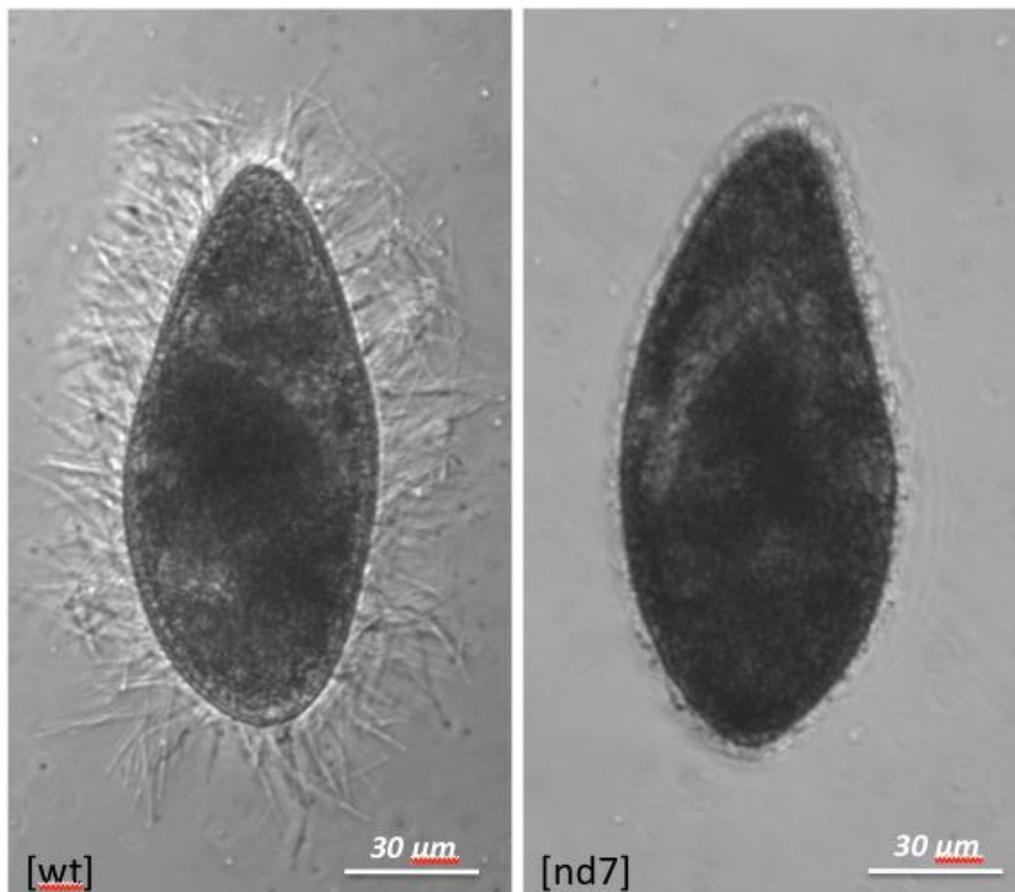
(A) Cellule végétative caractérisée par un MAC imposant (flèche jaune). (B) Cellules post-autogames dans lesquelles le MAC maternel est fragmenté alors que deux nouveaux MAC se forment (flèches orange). Les MIC sont indiqués par les flèches rouges (Janine Beisson et al., 2010).

## 2) Génétique classique

Dans une approche de génétique classique, on réalise une mutagenèse avec des ultraviolets afin de générer des mutations aléatoires. Les mutations qui sont intéressantes sont celles des micronoyaux. Un des intérêts majeurs de la paramécie en tant qu'organisme modèle est sa capacité à réaliser une autogamie qui va permettre, en une génération, d'obtenir une population entièrement homozygote. Si une mutation récessive intéressante est présente dans un allèle d'un des deux MIC, alors la génétique mendélienne prédit qu'un descendant sur quatre sera porteur homozygote de cette mutation. Un crible phénotypique est ensuite réalisé afin d'identifier des mutants déficients dans une voie particulière. Bien que la génétique classique soit moins courante de nos jours, voici deux articles illustratifs : (Marker et al., 2014), (Bhullar et al. en préparation).

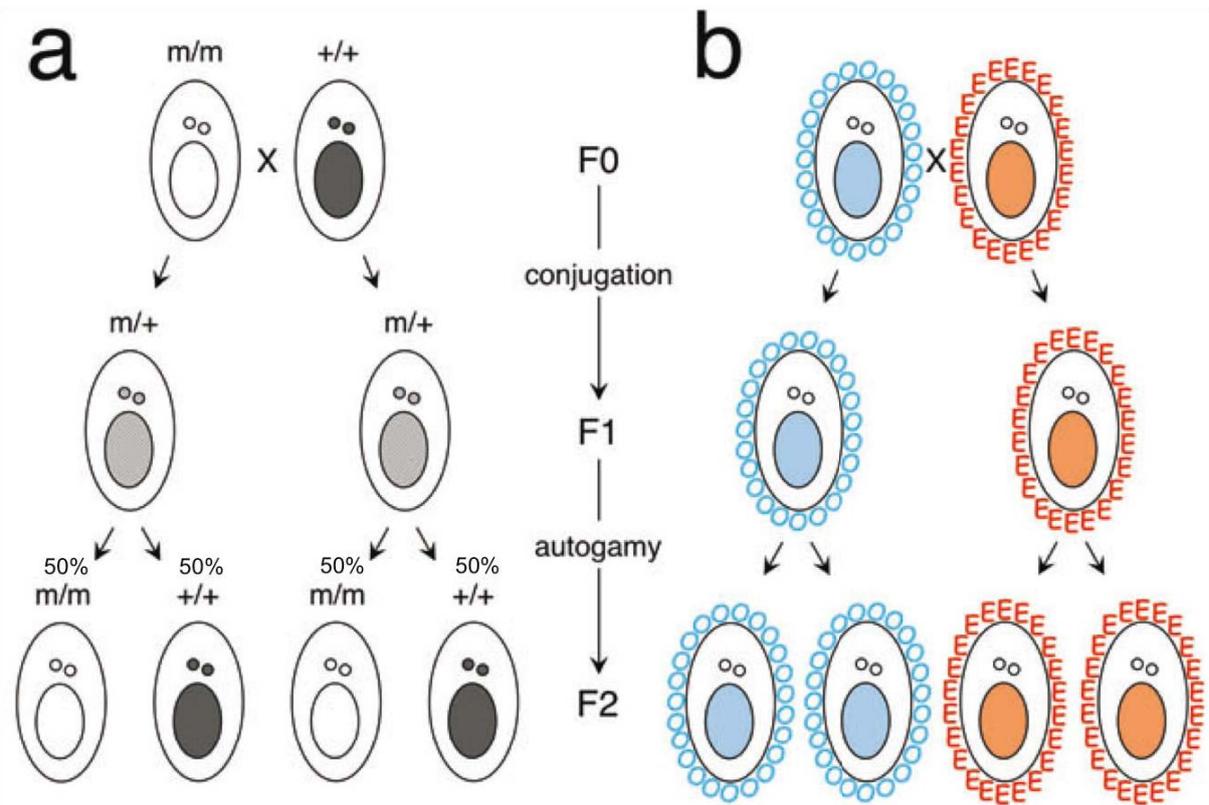
Les souches peuvent être facilement croisées par le biais de la conjugaison. Pour cela, les deux souches doivent être de types sexuels opposés. La réactivité sexuelle est induite par privation de nourriture dans des populations jeunes sinon c'est

l'autogamie qui sera prépondérante. Après mélange des deux populations réactives, des agglutinats se forment puis des couples apparaissent. Ces derniers peuvent être isolés, puis les deux ex-conjugants et enfin les deux paires de caryonides qui en dérivent. Au sein d'un couple, la réussite de l'échange réciproque des noyaux gamétiques n'est pas assurée, elle doit être vérifiée. On se sert le plus souvent d'un marqueur tel qu'une mutation dans le gène *ND7* qui conduit à un phénotype *Non Discharge*. En effet, toute cellule sauvage possède des filaments protéiques, appelés trichocystes, stockés dans des vésicules sous-membranaires. En cas de stress, artificiellement induit avec de l'acide picrique, la paramécie décharge ses trichocystes en une fraction de seconde (**Figure III.6**) (Galvani and Sperling, 2000). La mutation *nd7* est récessive, donc le croisement d'un parent sauvage avec un parent muté donnera deux F1 de phénotype sauvage. Pour l'illustrer, un schéma montrant la ségrégation mendélienne est présenté en **Figure III.7**. Il est parfois utile de connaître l'origine maternelle des caryonides, ce qu'on détermine le plus souvent par un test de type sexuel. En effet, chez *P. tetraurelia* le type sexuel est hérité maternellement, ainsi chaque ex-conjugant (et les 2 caryonides qu'il produira) exprime le même type sexuel que la cellule parentale (**Figure III.7**). Le génotypage de la mitochondrie est une autre méthode permettant de déterminer l'origine maternelle d'un individu.



**Figure III.6 : Test de décharge des trichocystes.**

Observation au microscope de paramécies sauvages (wt) ou mutantes (*nd7*) après contact avec de l'acide picrique.



**Figure III.7 : Hérité mendélienne vs maternelle.**

Chaque cellule représente un clone caryonidal. Dans les générations F1 et F2, les caryonides frères sont identiques entre eux, c'est pourquoi une seule cellule est dessinée. (a) Hérité mendélienne dans un croisement classique. Un seul locus est représenté avec l'allèle sauvage + ou l'allèle muté m. Dans la génération F2, il y a deux génotypes possibles avec une probabilité de 50 % chacun. Les caryonides frères formés par l'autogamie d'un hétérozygote F1 ont nécessairement le même génotype. (b) Hérité maternelle (cytoplasmique) du type sexuel. Des O et des E colorés indiquent le type sexuel et le MAC est également coloré car c'est lui détermine le phénotype.

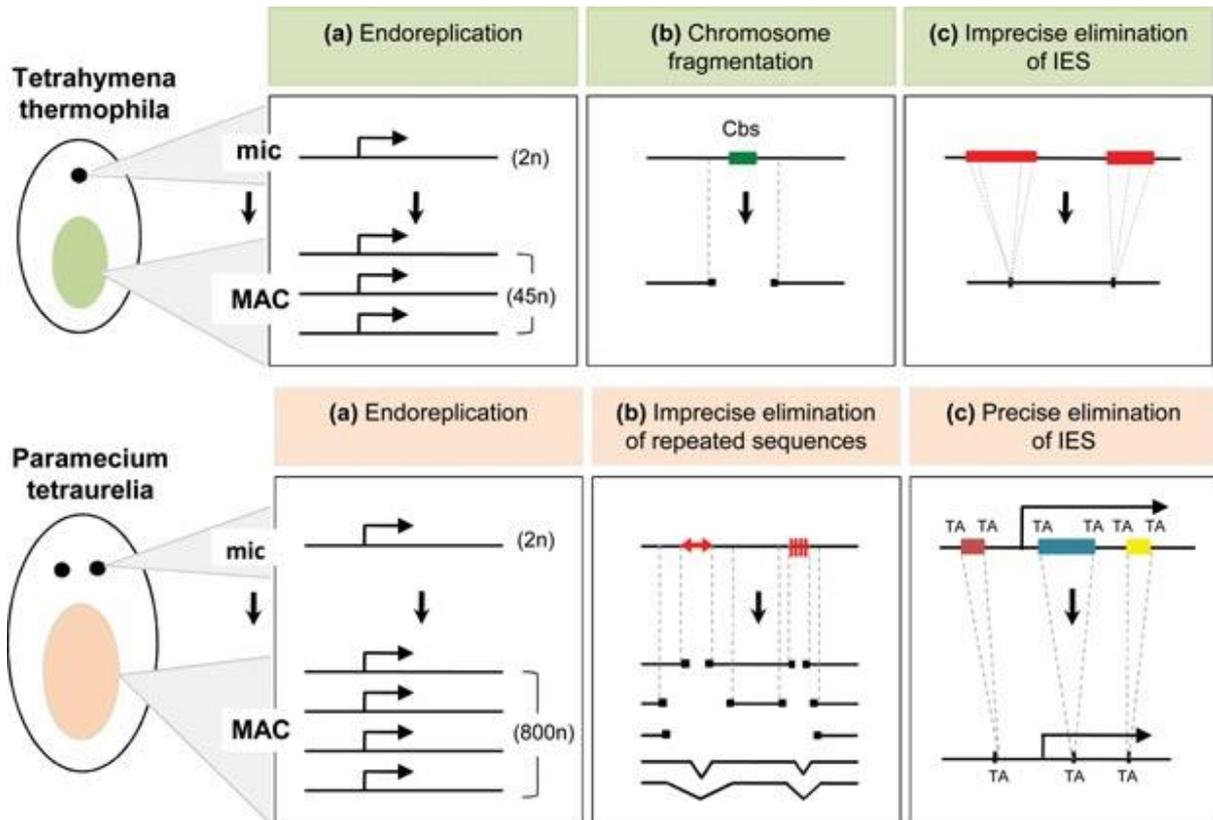
### 3) Génétique inverse

Depuis que son génome a été séquencé, la génétique inverse a pris un essor considérable en paramécologie. Un ensemble de protéines communes à tous les eucaryotes ont ainsi pu être étudiées spécifiquement chez la paramécie en fonction des thématiques de recherche des différents laboratoire : les protéines Piwi (Bouhouche et al., 2011), les protéines de réparation de l'ADN (Kapusta et al., 2011), les protéines BBS impliquées dans le trafic ciliaire (Valentine et al., 2012), etc. La deuxième source de gènes d'intérêts provient d'études transcriptomiques. La connaissance du génome a permis la mise au point de puces à ADN permettant d'étudier le profil d'expression de tous les gènes prédits (Arnaiz et al., 2010). De tels profils peuvent être utilisés afin de sélectionner des gènes candidats impliqués par exemple dans le processus d'autogamie (Nowak et al., 2011)(Arambasic et al., 2014).

Lorsqu'un nouveau gène d'intérêt se présente, le premier réflexe d'un paramécien est de l'inactiver. Contrairement aux organismes modèles classiques, il n'est pas possible d'inactiver un gène de manière permanente chez *P. tetraurelia*. Cela nécessiterait de transformer le MIC, ce qui n'a jamais été réussi jusqu'à présent. Chez *Tetrahymena*, la transformation du MIC est possible par biolistique (Cassidy-Hanley et al., 1997) en réalisant le bombardement à un temps précis de la conjugaison. Le MAC peut être transformé par biolistique, électroporation (Gaertig and Gorovsky, 1992)(Sweet and Allis, 2006) ou directement par micro-injection (Kahn et al., 1993). La seule méthode de transformation qui fonctionne chez *P. tetraurelia* est la micro-injection dans le MAC cependant elle n'aboutit pas à une recombinaison homologe, les fragments d'ADN sont pris en charge par la cellule comme des chromosomes MAC supplémentaires. En revanche, l'injection d'un grand nombre de copies d'un ADN homologe à la séquence codante d'un gène entraîne sa répression. C'est ce que l'on appelle le *transgene-induced silencing* (Ruiz et al., 1998). Toujours basé sur l'ARN interférence, Galvani et Sperling ont développé un outil plus simple en nourrissant les paramécies avec des bactéries produisant des double brins d'ARN homologues à un fragment du gène cible (Galvani and Sperling, 2002). C'est la méthode la plus couramment utilisée aujourd'hui. Enfin, quelle que soit la voie étudiée, une étude de la localisation subcellulaire de la protéine est réalisée par injection du gène additionné d'une étiquette GFP (J. Beisson et al., 2010).

## C. Les réarrangements du génome

Dans le MAC en développement (*anlage*) des ciliés, il y a des cycles d'endoréplication qui permettent d'augmenter la ploïdie. Cependant, la comparaison des séquences MIC et MAC témoigne d'autres événements majeurs de maturation que l'on regroupe sous le terme de réarrangements (**Figure III.8**).



**Figure III.8 : Réarrangements génomiques chez *T. thermophila* et *P. tetraurelia*.**  
Lire le texte pour avoir les détails. (Coyne et al., 2012)

### 1) La fragmentation des chromosomes

*Tetrahymena* possède seulement 5 chromosomes micronucléaires et pourtant les données de séquençage estiment qu'il y a 225 chromosomes MAC (Eisen et al., 2006). Dans une moindre mesure, il y aurait environ 3 fois plus de chromosomes MAC que MIC chez *P. tetraurelia*. En effet, l'analyse des séquences télomériques sur les 188 scaffolds obtenus lors du séquençage MAC indique qu'ils sont majoritairement des chromosomes entiers. Enfin, *Oxytricha trifallax* représente un cas extrême de fragmentation puisque les 120 chromosomes MIC laissent place à 16 000 nanochromosomes dont 90 % ne contiennent qu'un seul gène. Ainsi, la fragmentation des chromosomes semble être une caractéristique commune des ciliés, pourtant les mécanismes en jeu sont très différents. La fragmentation des chromosomes de *Tetrahymena* se fait spécifiquement au niveau de séquences « Cbs » de 15 pb qui sont nécessaires et suffisantes (Yao et al., 1990). Il n'y a pas de mécanismes précis pour les deux autres espèces mentionnées, la fragmentation résulte de l'élimination des séquences MIC-spécifiques. Dans tous les cas, il y a une addition *de novo* de télomères permettant le maintien des chromosomes dans le macronoyau.

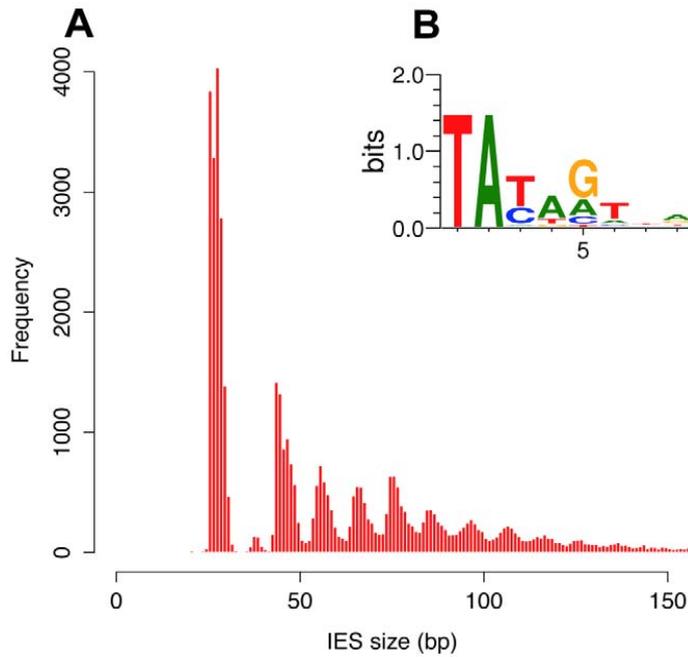
## 2) L'élimination interne d'ADN

### a. Une élimination imprécise

L'élimination de séquences d'ADN chez *Tetrahymena* est connue depuis très longtemps. Des expériences d'hybridation in situ (Yokoyama and Yao, 1982) ou bien des Southern blots (Yao, 1982) ont pu mettre en évidence ces phénomènes. On estime aujourd'hui que 6 000 fragments représentant 15 % de la diversité du MIC sont absents du MAC. Les séquences qui sont éliminées sont appelées IES (*Internal Eliminated Sequences*) et celles qui sont retenues sont appelées MDS (*Macronuclear Destined Sequences*). L'excision des IES chez *Tetrahymena* est imprécise, cela veut dire que les sites de coupure ne sont pas clairement définis par une séquence. Les deux extrémités des MDS sont ensuite raboutées. Chez *Paramecium*, les séquences répétées (transposons et satellites) sont éliminées selon le même principe, puis les extrémités retenues sont soit jointes, soit additionnées de télomères contribuant ainsi à la fragmentation des chromosomes (Le Mouel et al., 2003). 20 % du génome MIC est ainsi éliminé par ce processus au cours de la différenciation macronucléaire. Ces séquences ne sont pas appelées IES chez *Paramecium* car ce nom est réservé à d'autres séquences bien particulières.

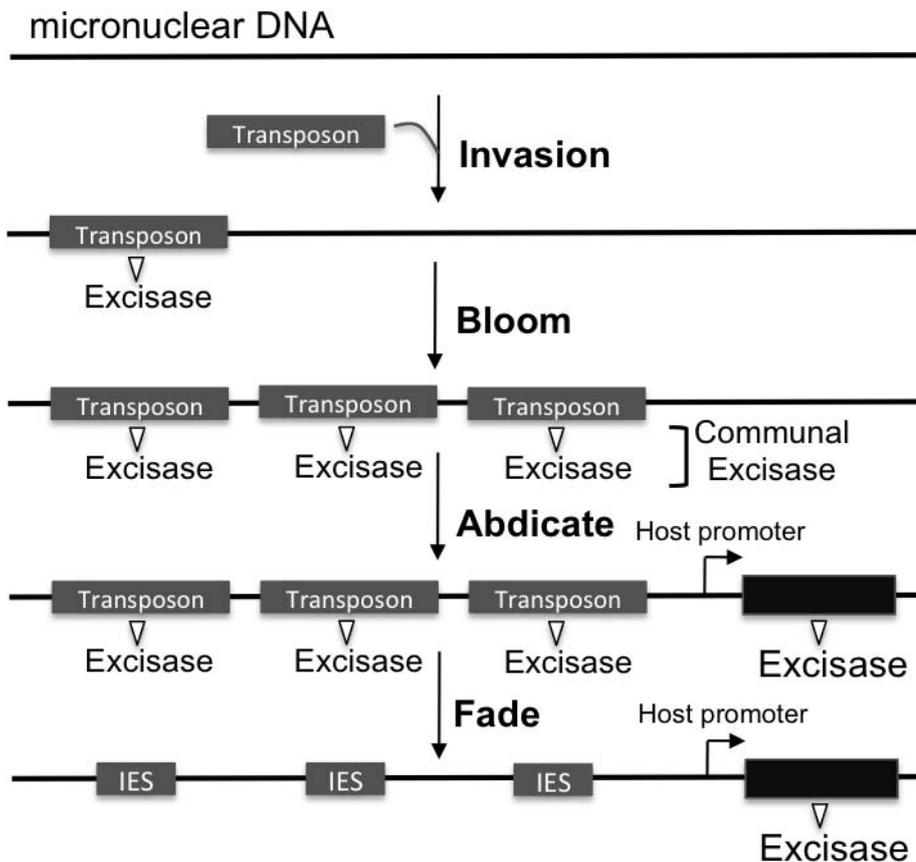
### b. Une élimination précise des IES chez *P. tetraurelia*

Les IES sont de courtes séquences (26 pb à 5 kb) invariablement bordées par des TA. Leur excision est précise puisqu'un seul TA subsiste après l'opération. 45 000 IES ont été dénombrées, aussi bien dans les régions géniques qu'intergéniques. Ainsi, la moitié des gènes de la paramécie sont interrompus par au moins une IES d'où la nécessité d'une excision précise et efficace. Le consensus dégénéré des bornes des IES suggère qu'il s'agit de vestiges de transposons de la famille *Tcl/mariner* (**Figure III.9**) (Arnaiz et al., 2012). Certaines IES, parmi les plus longues, portent d'ailleurs un motif TIR caractéristique. Pour expliquer ces observations, l'hypothèse IBAF (*Invade, Bloom, Abdicate and Fade*) fut proposée dans les années 90 (Klobutcher and Herrick, 1997) (**Figure III.10**). Dans ce modèle, un transposon envahirait le génome germinale et se dupliquerait grâce à l'enzyme qu'il code. La multiplication des copies ne serait pas limitée par le fardeau des mutations insertionnelles car elles sont éliminées pendant le développement du noyau somatique. L'excision de ces séquences reposerait initialement sur le pool de transposases codées par les éléments. Lorsque la cellule parviendrait à domestiquer la fonction d'excision, les transposons seraient voués à dégénérer en courtes séquences dont seuls les TA essentiels à leur reconnaissance seraient maintenus. Il y aurait également une certaine sélection concernant la taille des IES puisqu'elles ont une distribution tout à fait singulière avec une périodicité de 10,2 pb soit la taille du pas de la double hélice d'ADN (**Figure III.9**). Cela suggère que le mécanisme d'excision est contraint spatialement, indépendamment de la séquence d'ADN.



**Figure III.9 : Propriété des IES.**

(A) Histogramme de la taille des IES inférieures à 150 pb. (B) Logo des bornes des IES. (Arnaiz et al., 2012).

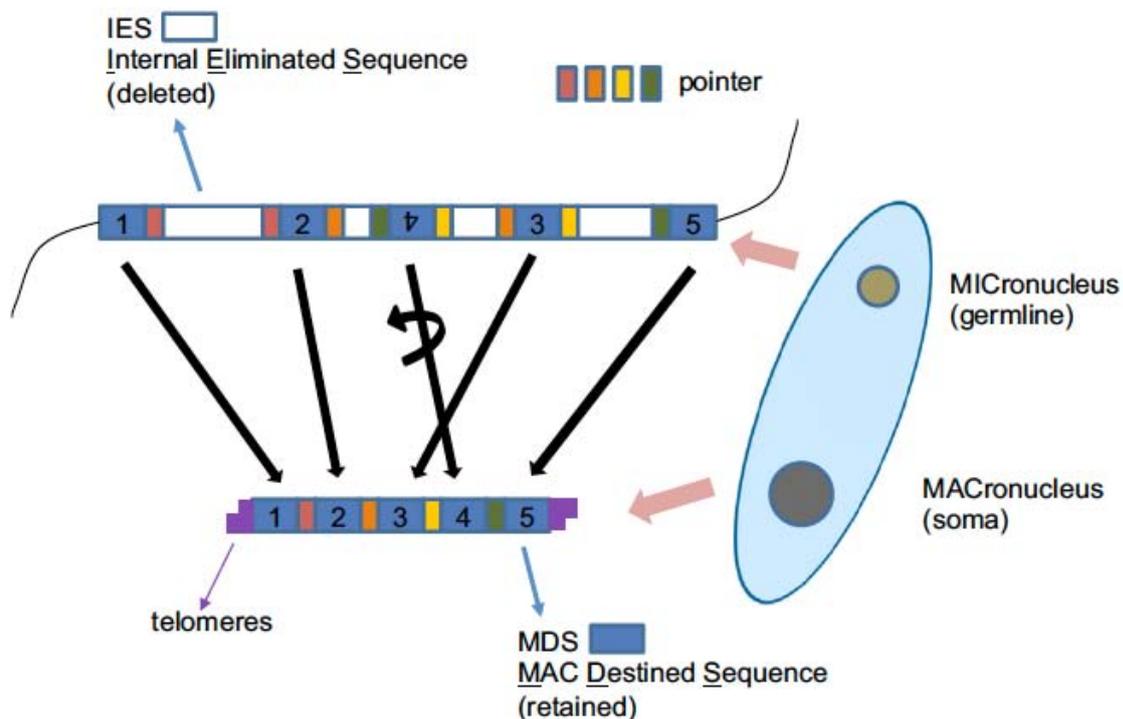


**Figure III.10 : Modèle IBAF (Klobutcher and Herrick, 1997).**

L'ADN micronucléaire est symbolisé par la ligne noire, les transposons et les IES sont représentés par les rectangles gris. Dans la phase d'abdication, une partie de la fonction d'excision devient contrôlée par un promoteur de l'hôte.

### 3) DNA unscrambling

Les réarrangements du génome d'*Oxytricha trifallax* sont encore plus complexes. Non seulement cet organisme élimine 90 % de séquences MIC-spécifiques, mais il doit également réordonner certaines MDS afin de reconstituer un gène macronucléaire fonctionnel (**Figure III.11**) (Greslin et al., 1989). Les MDS peuvent être dans des orientations opposées, provenir de loci différents ou bien encore être utilisées pour construire plusieurs gènes (Chen et al., 2014). L'assemblage exact des gènes qui nécessitent plusieurs MDS requiert des séquences flanquantes de microhomologie appelées pointeurs. Ils sont organisés de telle sorte que le pointeur du MDS  $n$  soit identique au pointeur du MDS  $n+1$  autorisant ainsi une recombinaison exacte.



**Figure III.11 : DNA unscrambling chez *Oxytricha trifallax*.** (Chen et al., 2014)

### 4) Les transposases

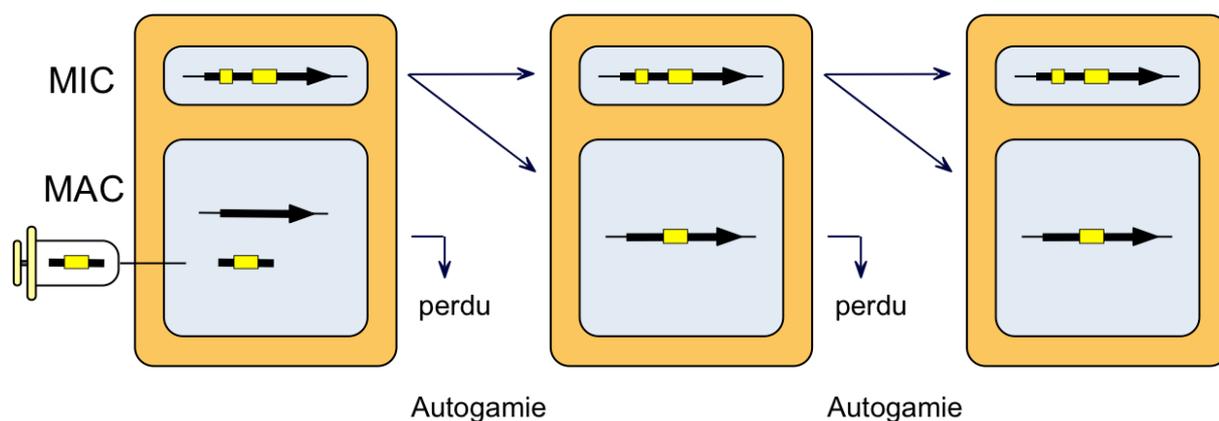
C'est chez *P. tetraurelia* que la première endonucléase responsable des éliminations massives d'ADN fut identifiée. Bien que les IES semblent provenir de transposons de la famille *Tcl/mariner*, la géométrie de leur clivage est différente. C'est une transposase de type Piggybac, qui a été domestiquée (Baudry et al., 2009). Elle s'appelle Piggymac et est responsable de l'excision des IES ainsi que des éléments répétés. Cette découverte a permis d'obtenir indirectement de l'ADN micronucléaire en extrayant des *anlagen* non réarrangés suite à l'inactivation du gène par ARN interférence. Peu de temps après, son orthologue Tpb2p (*Tetrahymena PiggyBac-like*) fut identifié chez *Tetrahymena*. En revanche, il n'existe pas une endonucléase responsable de l'élimination de l'ADN chez *Oxytricha* mais des milliers, portés par les transposons eux-mêmes (Nowacki et al., 2009). Il s'agit peut-être d'une étape évolutive précoce avant la domestication par le génome de l'hôte.

## D. Le contrôle épigénétique des réarrangements chez *Paramecium* et *Tetrahymena*

### 1) Approche historique

Comprendre comment les ciliés distinguent les séquences éliminées des séquences retenues au cours de la différenciation du macronoyau était un challenge majeur qui n'est pas encore totalement résolu aujourd'hui. Des études chez *Tetrahymena* ont montré l'importance des séquences flanquantes des IES pour spécifier les bornes de l'excision. Pour l'IES M en particulier, des séries de polypurines sont nécessaires et suffisantes à l'excision des bornes (Godiska et al., 1993; Godiska and Yao, 1990). Cependant, ces séquences ne sont pas consensuelles d'une IES à l'autre pour celles qui ont été étudiées (Chalker et al., 1999; Fillingham et al., 2001). Par ailleurs, les IES elles-mêmes, tout comme chez *Paramecium*, ne contiennent pas de séquence caractéristique qui permettrait à un facteur protéique de les reconnaître.

Le premier élément de réponse provient de l'étude de la souche d48 de *P. tetraurelia* qui est incapable d'exprimer l'antigène de surface A. La cause en est une délétion de l'extrémité du chromosome MAC qui interrompt la séquence codante du gène en 5' (Forney and Blackburn, 1988). Cette délétion est héréditaire alors que le génome germinal est sauvage, ce qui signifie qu'elle est reproduite à chaque maturation d'un nouveau MAC. La micro-injection dans le MAC d'un plasmide contenant le gène sauvage permet de bloquer la délétion lors de l'autogamie, la génération suivante possède alors la version sauvage du gène dans son MAC (Koizumi and Kobayashi, 1989). L'induction expérimentale de délétions a permis de généraliser ces observations (Garnier et al., 2004)). Ainsi, c'est la présence de la séquence dans le MAC qui détermine si elle sera éliminée ou retenue à la génération suivante. Ce contrôle maternel a également été montré pour une IES insérée dans le gène G codant un autre antigène de surface (Duharcourt et al., 1995) ainsi que pour des IES de *Tetrahymena* (Chalker and Yao, 1996). D'autres IES ont ensuite été testées afin de généraliser le phénomène mais seulement 5 sur 13 ont montré une inhibition de leur excision en réponse à l'injection d'une séquence homologue dans le MAC (Duharcourt et al., 1998). Ces IES sont dites contrôlées maternellement selon le principe évoqué et schématisé dans la **figure III.12**. La rétention, initialement induite par micro-injection, est ensuite maintenue de génération en génération tout comme la délétion dans la souche d48. Il faut souligner que l'inhibition est dépendante du nombre de copies injectées et plus efficace si le fragment injecté contient les séquences flanquantes des IES. De plus, il existe une variabilité importante de l'efficacité du processus en fonction de l'IES étudiée.

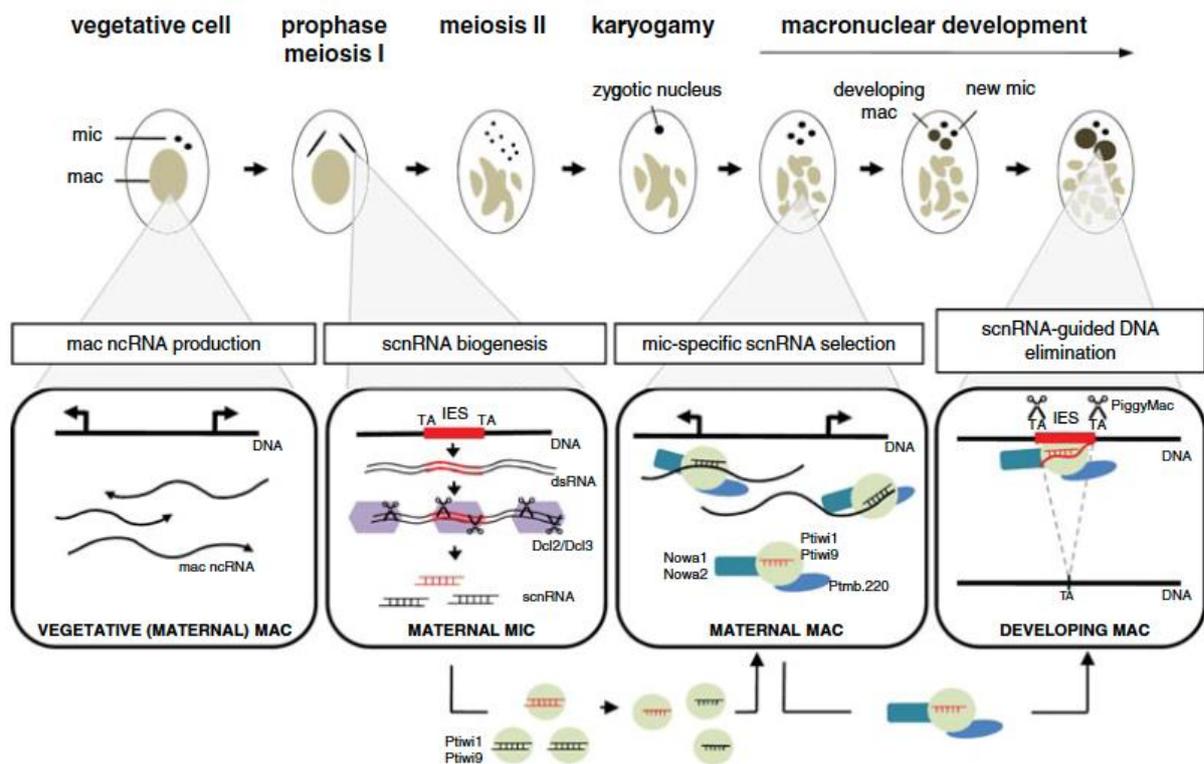


**Figure III.12 : Contrôle maternel des IES.**

La flèche noire représente un gène dans lequel deux IES (rectangles jaunes) sont insérées. L'injection d'une séquence homologue à une IES dans le MAC entraîne spécifiquement la rétention de cette IES à la génération suivante lors du développement de l'*anlage*. Cette rétention est maintenue à l'autogamie suivante.

## 2) Principe du *genome scanning*

Les expériences décrites précédemment ont mis en évidence le fait qu'il existe une communication entre le MAC maternel et le nouveau MAC. En réalité, c'est le génome MIC qui est comparé au génome MAC afin d'identifier les séquences MIC-spécifiques. L'information est portée par des petits ARN puis transmise aux nouveaux MAC afin de guider les réarrangements. Dans un but pédagogique, j'ai choisi de présenter le modèle du *genome scanning* dans sa globalité avant de détailler les différentes étapes dans les parties suivantes (**Figure III.13**). Vous pourrez ainsi vous y référer pour ne pas perdre de vue le schéma général. Bien qu'il soit présenté pour les IES contrôlées maternellement de *P. tetraurelia*, il est aussi valable dans sa globalité pour les séquences répétées, les gènes cellulaires et pour la distinction IES/MDS de *Tetrahymena*.



**Figure III.13 : Modèle du *genome scanning*.**

Les scnARN sont produits au début de la méiose à partir de l'ensemble du génome MIC puis filtrés par appariement avec de longs transcrits non codants dans le MAC maternel. Seuls les scnARN ne trouvant pas d'homologie (en rouge) vont guider l'excision des séquences correspondantes dans le MAC en développement. (Coyne et al., 2012)

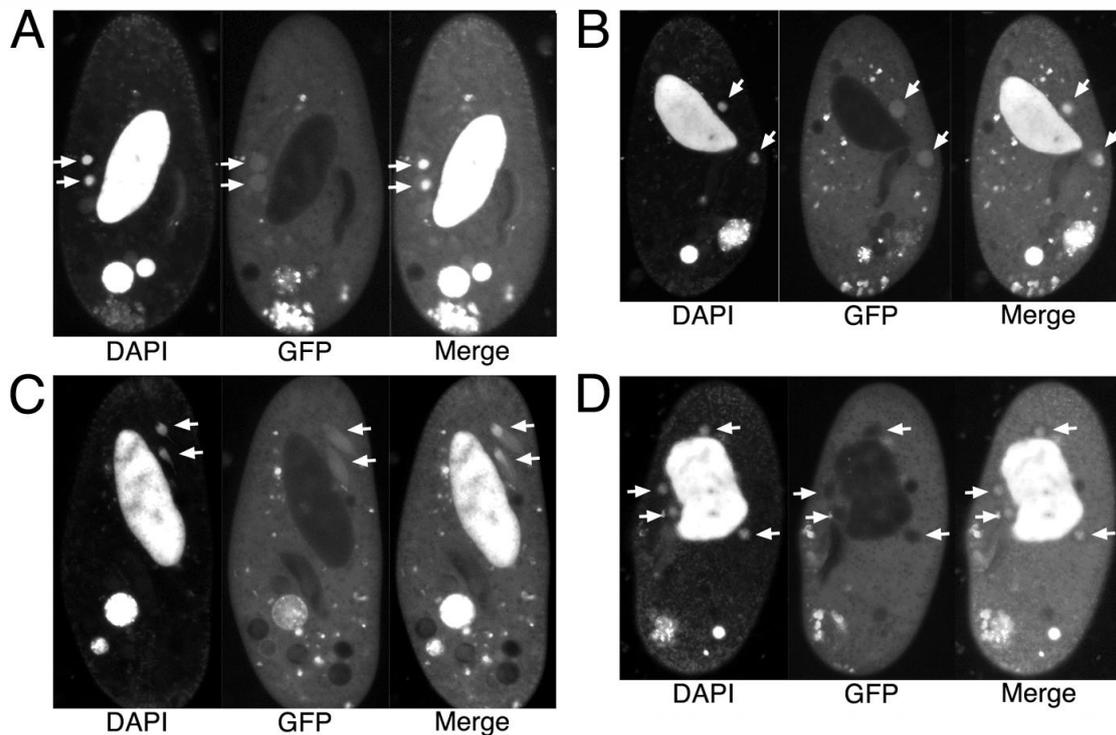
Le système du *genome scanning* apparaît comme un mécanisme de défense inné contre les transposons. Alors qu'une insertion dans le MAC ne pourrait jamais être transmise à la descendance sexuelle, une insertion dans le MIC serait immédiatement détectée. En effet, le locus MIC produirait des scnARN qui ne trouveraient pas d'homologie dans le MAC et donc guideraient l'excision du transposon. Ce système original permet aux ciliés de garder en mémoire les séquences de leur propre ADN sous la forme du MAC somatique. L'efficacité du système explique peut-être pourquoi le génome MIC contient autant d'IES. En effet, les éléments transposables ne devraient pas constituer un fardeau puisqu'ils sont précisément excisés dans le MAC somatique. En l'absence de sélection positive, ils dégénèreraient rapidement en IES.

### 3) Les petits ARN impliqués dans le modèle

#### 1. Biogenèse des scnARN

De petits ARN de 28 nt en moyenne et spécifiques à la conjugaison ont été identifiées pour la première fois chez *Tetrahymena* (Mochizuki et al., 2002). Ils proviennent de transcrits sens et antisens du MIC pendant la méiose (Chalker and Yao, 2001) qui sont clivés par une protéine Dcl1p (*Dicer-like 1 protein*) (Malone et al., 2005). Bien que tout le génome MIC soit globalement couvert par ces ARN, il y a un très fort biais en faveur des IES (Schoeberl et al., 2012). Suite à la proposition du *genome scanning*,

ces petits ARN ont pris le nom de *scan RNA* ou scnARN (Mochizuki and Gorovsky, 2004). Dans la cellule, il a été montré par immunoprécipitation qu'ils sont associés avec la protéine Piwi Twiip (Mochizuki et al., 2002). De fait, ils peuvent être appelés piARN et leur fonction est d'ailleurs identique à celle décrite chez les eucaryotes multicellulaires : réprimer les transposons pendant le développement. Chez *Paramecium*, Les protéines Dcl2/3 génèrent des duplex d'ARN de 25 nt avec une signature 5'-UNG (Lepere et al., 2009) à partir d'une transcription micronucléaire non biaisée vers les IES. Les protéines Ptiwi01/9 chargent le duplex dans le cytoplasme et n'en gardent qu'un seul brin. L'inactivation des protéines Dcl ou Piwi entraîne la disparition des scnARN et la mort de la descendance. Puisque Dcl2-GFP n'est visible dans les MIC qu'en prophase de méiose I, les scnARN ne peuvent être produits que pendant cette courte période (**Figure III.14**) (Lepere et al., 2009).



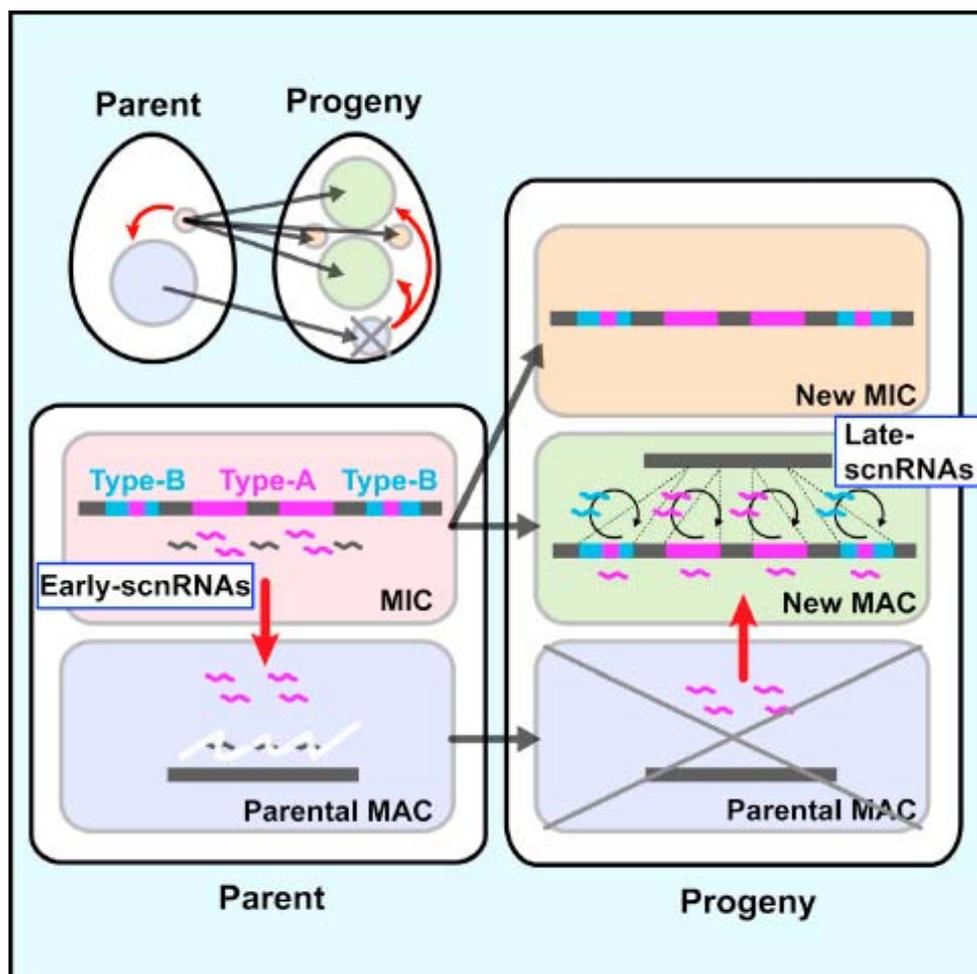
**Figure III.14 : Clichés de microscopie à fluorescence réalisés sur des cellules injectées avec une construction Dcl2-GFP.**

Les MIC sont indiqués par les flèches blanches. (A) Cellule végétative. (B, C) Grossissement des MIC puis stade « croissant » en prophase de méiose I. (D) Fin de méiose I. (Lepere et al., 2009).

## 2. Une deuxième vague de petits ARN

Des études récentes montrent qu'il existe une deuxième vague de petits ARN impliqués dans les réarrangements. Chez *Paramecium*, des iesARN sont produits par une protéine Dicer (Dcl5) dans les *anlagen* (Sandoval et al., 2014). Puisqu'ils ne contiennent que des séquences d'IES, les auteurs proposent une boucle d'amplification selon laquelle les IES excisées seraient transcrites puis converties en

iesARN afin de propager l'excision de cette même IES à l'ensemble des 800 copies du MAC. Cette hypothèse est controversée, notamment parce que cela implique que les IES soient transcrites après leur excision, ce dont on n'a aucune évidence. Chez *Tetrahymena*, des scnARN identiques à ceux produits pendant la méiose ont été identifiés en fin de conjugaison. Ils sont produits par Dcl1p, majoritairement à partir d'IES, puis chargés par les nouvelles protéines Twi1p et Twi11p issues de l'expression du noyau zygotique (Noto et al., 2015). Les auteurs ont défini deux classes d'IES en fonction du ratio de scnARN tardifs/précoces qui s'alignent sur elles. Les IES de type A (40 %) ont un ratio inférieur à 10 et les IES de type B (60 %) ont un ratio supérieur. Dans leur modèle, les IES de type A produisent les scnARN précoces (classiques) qui vont non seulement guider leur excision dans les *anlagen* mais aussi cibler des séquences répétées communes aux deux types d'IES (répétitions A). Les scnARN tardifs sont alors produits par une propagation en cis à partir des répétitions A dans les IES de type B et sont nécessaires à l'excision de la plupart des IES. Un schéma est présenté en **figure III.15**. Puisque les répétitions A peuvent être communes entre plusieurs IES, il y a en réalité tout un réseau de reconnaissance des IES en trans et on peut supposer qu'une IES donnée est rarement dépendante de ses propres scnARN.



**Figure III.15 : Modèle du *genome scanning* proposé chez *Tetrahymena* avec des scnARN précoces et tardifs.**

Lire le texte ci-dessus pour les détails (Noto et al., 2015).

### **3. scnARN et contrôle maternel chez *Paramecium***

Nous avons vu que sur 13 IES testées expérimentalement, 5 présentaient un contrôle maternel avec une sensibilité variable. L'inactivation de la protéine Nowa1, qui n'existe que chez *Paramecium*, affecte spécifiquement l'excision des transposons et des IES contrôlées maternellement (Nowacki et al., 2005). En outre, plus une IES est facilement retenue après injection de sa séquence dans le MAC, plus elle est sensible à la déplétion de Nowa1. Tout comme Ptiwio9, Nowa1 est exprimée à l'autogamie et est d'abord localisée dans le MAC puis dans les *anlagen*. Son rôle n'est pas précisément connu, elle pourrait aider au transport des scnARN et/ou favoriser les appariements entre les scnARN et leurs cibles. Quoiqu'il en soit, des données de séquençage ont montré que son inactivation provoquait la rétention de 30 % des IES (Nowacki Lab), d'où l'idée que ces 30 % forment l'ensemble théorique des IES contrôlées maternellement. Le modèle du genome scanning fut proposé pour expliquer l'hérédité maternelle des réarrangements. En toute logique, les IES contrôlées maternellement devraient dépendre des scnARN pour leur excision. Pourtant, l'inactivation de Dcl2/3 ou de Ptiwio1/9, ayant pour effet la disparition des scnARN, affecte seulement 7,2 % des IES, avec des taux de rétention variables (Sandoval et al., 2014, Nowacki Lab). A première vue, l'ensemble des IES scnARN-dépendantes serait plus restreint que celui des IES contrôlées maternellement mais il s'agit peut-être d'un problème d'efficacité de la double déplétion de Dcl2 et Dcl3. En outre, il semble que les IES retenues après la triple inactivation de Dcl2, 3 et 5 soient les mêmes que pour Nowa1 (Laboratoire Nowacki), alors que l'inactivation de Dcl5, entraînant l'absence d'iesARN, induit une rétention très partielle qui affecte 6,8 % des IES. L'effet synergique observé pourrait finalement nous orienter vers un modèle dans lequel la voie Dcl5 viendrait en secours de la voie Dcl2/3 lorsque les scnARN sont en quantité limitante. Autrement dit, le vrai ensemble des IES scnARN-dépendantes (sensibles à Dcl2/3) ne serait révélé qu'avec la déplétion supplémentaire de Dcl5. Et sous cette condition, cet ensemble d'IES serait bien celui des IES contrôlées maternellement.

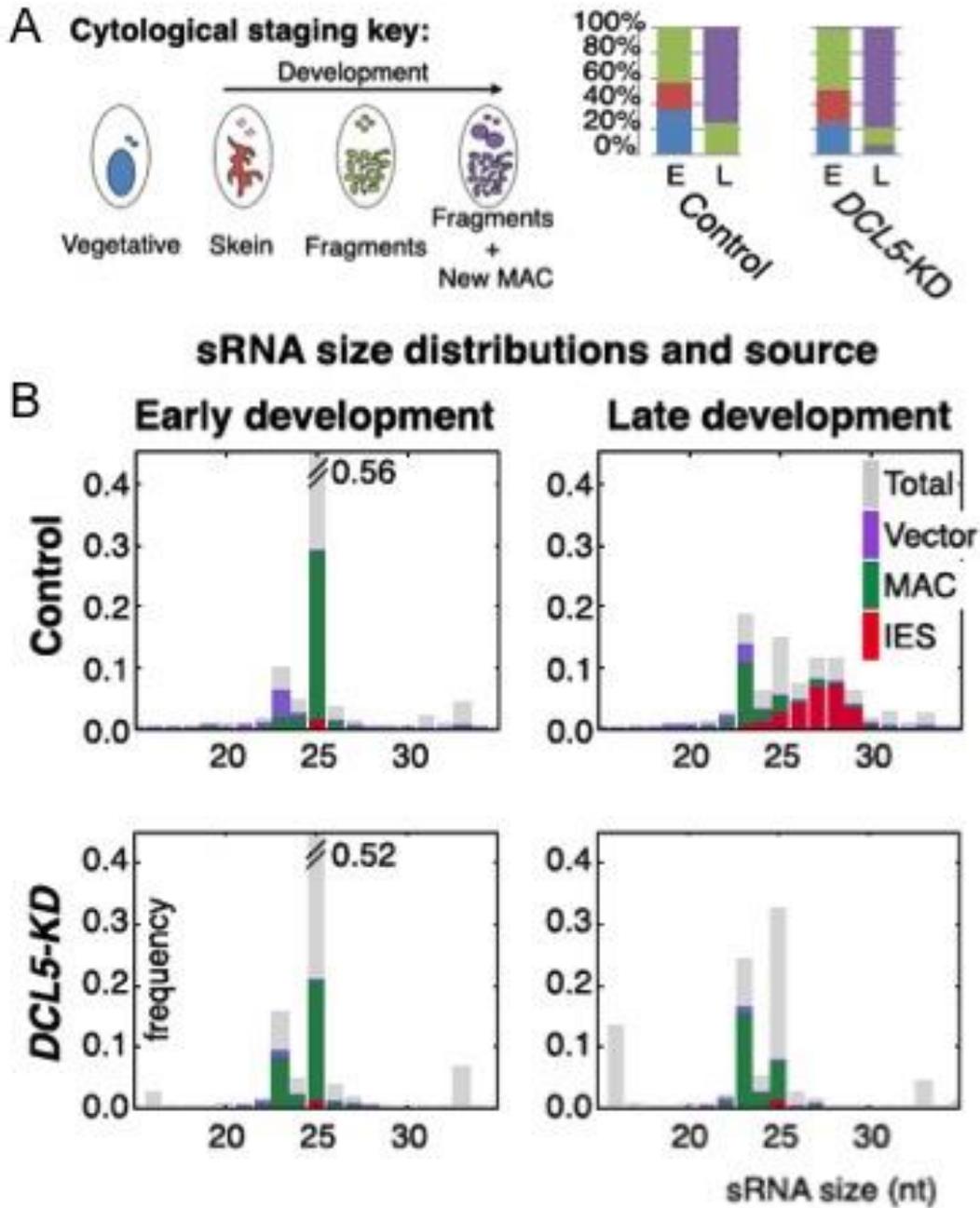
#### **4) la sélection MIC-spécifique dans le MAC maternel**

Afin d'identifier les scnARN MIC-spécifiques, l'ensemble des scnARN doivent être comparés au génome MAC. Des données montrant l'importance de transcrits maternels non codants ont été documentées chez *Tetrahymena* et *Paramecium* (Aronica et al., 2008; Lepere et al., 2008). Ce sont eux, et non directement la séquence d'ADN, qui vont s'apparier avec les scnARN par homologie de séquence. Seuls les scnARN spécifiques du MIC ne trouveront pas de séquence complémentaire et pourront diriger l'excision dans les *anlagen*. Dans une lignée IES+ (rétention héréditaire), l'introduction d'un duplex portant une signature de scnARN et homologue à une portion de l'IES peut déclencher l'excision de l'IES (Lepere et al., 2008). Le processus de sélection qui va marquer différemment les scnARN MIC-spécifiques n'est pas connu. Sont-ils modifiés ? Sont-ils simplement libres de diffuser en attendant de trouver un appariement dans les *anlagen*? Que deviennent les scnARN MAC ? Sont-ils dégradés, séquestrés, ou bien ont-ils un rôle protecteur à jouer ultérieurement dans les *anlagen*? Pour toutes ces questions, on sait seulement qu'il y a un enrichissement en scnARN MIC-spécifiques par rapport aux autres. Chez *Tetrahymena*, des *northern blots* ont montré qu'il y avait initialement 3 fois plus de

scnARN MIC que de scnARN MAC et que ce ratio montait jusqu'à 20 au cours de la conjugaison (Mochizuki and Gorovsky, 2004). Chez *Paramecium*, des données de séquençage en masse montrent également un enrichissement relatif en scnARN MIC (Sandoval et al., 2014) (**Figure III.16**). Toutefois la présence des iesARN de 25 nt, dont la signature est identique à celles des scnARN, vient brouiller cette observation. Lorsqu'ils sont enlevés de l'analyse par l'inactivation de Dcl5, l'enrichissement en scnARN MIC-spécifiques est beaucoup plus faible et les scnARN MAC sont encore quatre fois plus abondants au temps tardif. Il est donc plausible que ces scnARN MAC ne soient pas dégradés et que l'enrichissement relatif observé soit en partie dû à l'apparition des iesARN.

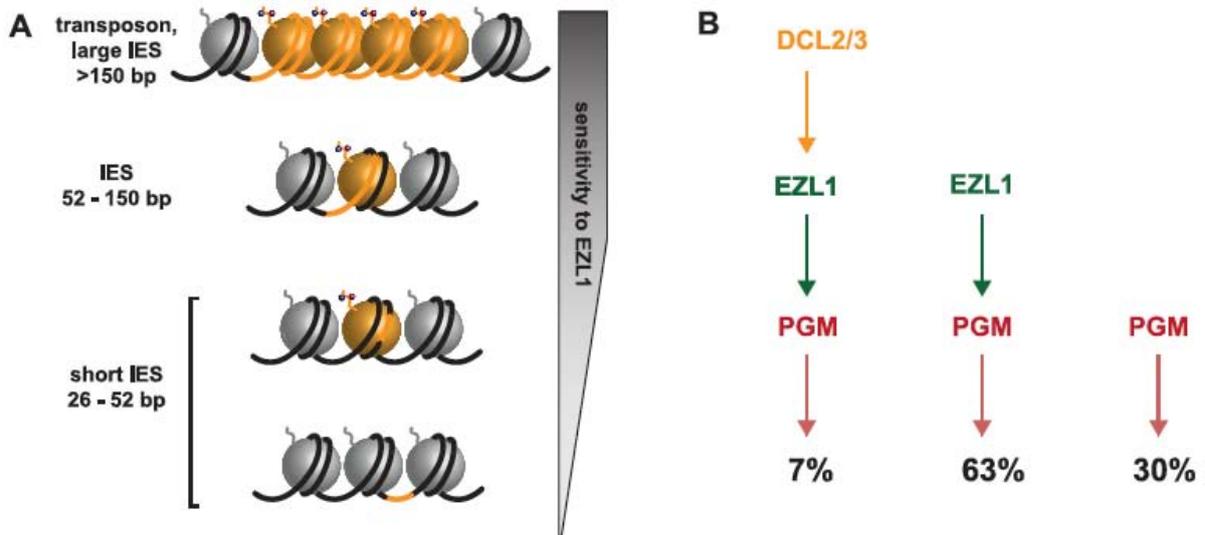
## 5) L'élimination guidée par les scnARN des IES

Après leur sélection, les scnARN MIC-spécifiques sont transportés dans les *anlagen* afin de recruter la machinerie d'excision. Ce transport est étayé par les analyses des protéines de fusion Twi1p, Ptiwi9 et Nowa1 (Bouhouche et al., 2011; Mochizuki et al., 2002; Nowacki et al., 2005). Comme dans beaucoup d'autres organismes, les petits ARN sont capables de guider la formation d'hétérochromatine, au niveau des IES dans le cas présent. La première protéine établissant ce lien est Pdd1p (*Programmed DNA Degradation 1*) chez *Tetrahymena* (Madireddi et al., 1996), dont la séquence montre un chromodomaine homologue à la protéine humaine HP1 (*heterochromatin protein 1*) (Eissenberg and Elgin, 2000). En outre, les méthylations d'histones H3K9 et H3K27 apparaissent spécifiquement pendant la conjugaison et sont dépendantes des scnARN (Liu et al., 2007, 2004; Taverna et al., 2002). Elles sont déposées par l'histone méthyltransférase Ezl1p (*Enhancer of Zeste-Like*) dont un orthologue occupe la même fonction chez *Paramecium* (Lhuillier-Akakpo et al., 2014). Tous ces facteurs hétérochromatiniens sont initialement diffus dans les *anlagen* puis ils se regroupent en foci appelées '*DNA elimination bodies*'. Malgré l'accumulation des connaissances et des facteurs impliqués, aucune expérience *in vitro* n'a permis de montrer l'activité endonucléase de Tbp2p ou Pgm. Un marquage chromatinien des IES paraît satisfaisant comme modèle chez *Tetrahymena* car elles ont une taille moyenne de 2 kb et des bornes peu précises (Fass et al., 2011). De plus, toutes les IES semblent concernées par le mécanisme puisque l'inactivation de Twi1p induit la rétention totale de toutes les IES (Noto et al., 2014). En revanche chez *Paramecium* un marquage chromatinien ne peut pas être suffisant pour guider l'excision précise des petites IES. D'ailleurs 30 % des IES ne dépendent pas d'Ezl1 pour leur excision et parmi les autres, seules 7 % dépendent de Dcl2/3 (**Figure III.17**). Cette sensibilité aux facteurs est assez bien corrélée avec la taille des IES. Sachant que les IES les plus petites sont les plus anciennes (Arnaiz et al., 2012), il semblerait que les IES évoluent progressivement vers une indépendance vis à vis des scnARN puis du marquage chromatinien déposé par Ezl-1. Un autre système de reconnaissance encore inconnu doit nécessairement se mettre en place pour assurer leur excision.



**Figure III.16 : Nature et Distribution de taille des petits ARN au cours de l'autogamie de *P. tetraurelia*.**

Les temps précoce et tardif correspondent à des populations de cellules au cours de l'autogamie dont la nature est représentée en A. Les petits ARN extraits sont un mélange de siARN (23 nt), scnARN (25 nt) et iesARN (21-30 nt). L'inactivation de Dcl5 entraîne la disparition des iesARN. (Sandoval et al. 2014).



**Figure III.17 : Modèle de la programmation des IES pour l'excision rendant compte de leur sensibilité aux différents acteurs de la voie du *genome scanning*.**

(A) Les séquences MIC-spécifiques sont indiquées en orange. Les deux disques colorés représentent la méthylation des histones. La sensibilité à Ezl1 d'une IES dépend de sa taille et de sa position par rapport aux nucléosomes. (B) Représentation des 3 grandes voies impliquées dans le *genome scanning* avec le pourcentage d'IES affectées par leur inactivation (Lhuillier-Akakpo et al., 2014).

## 6) Le rôle possible de la méthylation des adénines

Une piste sérieuse qui permettrait d'expliquer le ciblage précis des IES chez *Paramecium* est la méthylation de l'ADN. Or, il semblerait que la seule base méthylée soit l'adénine (Cummings et al., 1974) avec un pourcentage de 2,4 % dans le MAC et 3,3 % dans le MIC (probable contamination MAC). On imagine par exemple que les TA des bornes des IES pourraient être méthylées ou non méthylées afin de les distinguer des autres TA. Les études chez *Tetrahymena* ont montré que 0,8 % des adénines du MAC sont méthylées, principalement dans des motifs AT (Bromberg et al., 1982). Il est intéressant de noter que ce dinucléotide est palindromique car la méthylation peut alors être conservée au cours de la réplication, et possiblement de génération en génération. Si une telle méthylation pouvait discriminer les IES, alors on comprendrait que la plupart d'entre elles n'aient pas besoin du *genome scanning* pour guider leur excision.

## E. Les apports du modèle *Oxytricha trifallax*.

### 1) Otiwi et ses piARN.

Chez *Oxytricha*, la protéine Piwi (Otiwi1) porte des piARN long de 27 nt avec un U caractéristique en 5' (Fang et al., 2012; Zahler et al., 2012). Ils couvrent le génome MAC, incluant les jonctions entre des MDS adjacents, et ne s'alignent pas sur les MIC (Fang et al., 2012). Contrairement aux oligohyménophores, ce spirotriche produit donc des petits ARN spécifiques du MAC ayant pour rôle le maintien des séquences.

L'opposition de ces deux systèmes peut se justifier par une optimisation économique puisque les MDS d'*Oxytricha* représentent seulement 5 % du génome germinale. Chez *Tetrahymena*, les scnARN sont produits majoritairement à partir des IES donc à partir de 35 % du génome MIC. Ainsi, dans ces 2 espèces, le système produit des petits ARN couvrant la fraction la plus petite du génome, que ce soit pour guider son excision ou sa rétention.

## 2) La méthylation des cytosines

Les IES et les MDS d'*Oxytricha* sont souvent plus petites qu'un tour de nucléosome d'où la nécessité d'une discrimination fine (Chen et al., 2014). Récemment, il a été reporté que cette espèce utilise la méthylation des cytosines *de novo* pour marquer l'élimination de l'ADN (Bracht et al., 2012). En plus de l'étiquetage des IES, la méthylation des cytosines serait aussi utilisée pour conduire la dégradation du MAC maternel entier et pour éliminer les assemblages aberrants qui peuvent se former dans les *anlagen*. La même étude a également identifié un motif de 20 pb comportant de nombreux CC comme étant un site privilégié de méthylation. Toutefois, l'enzyme responsable de cette méthylation demeure inconnue.

# IV – La spéciation du complexe *P. aurelia*

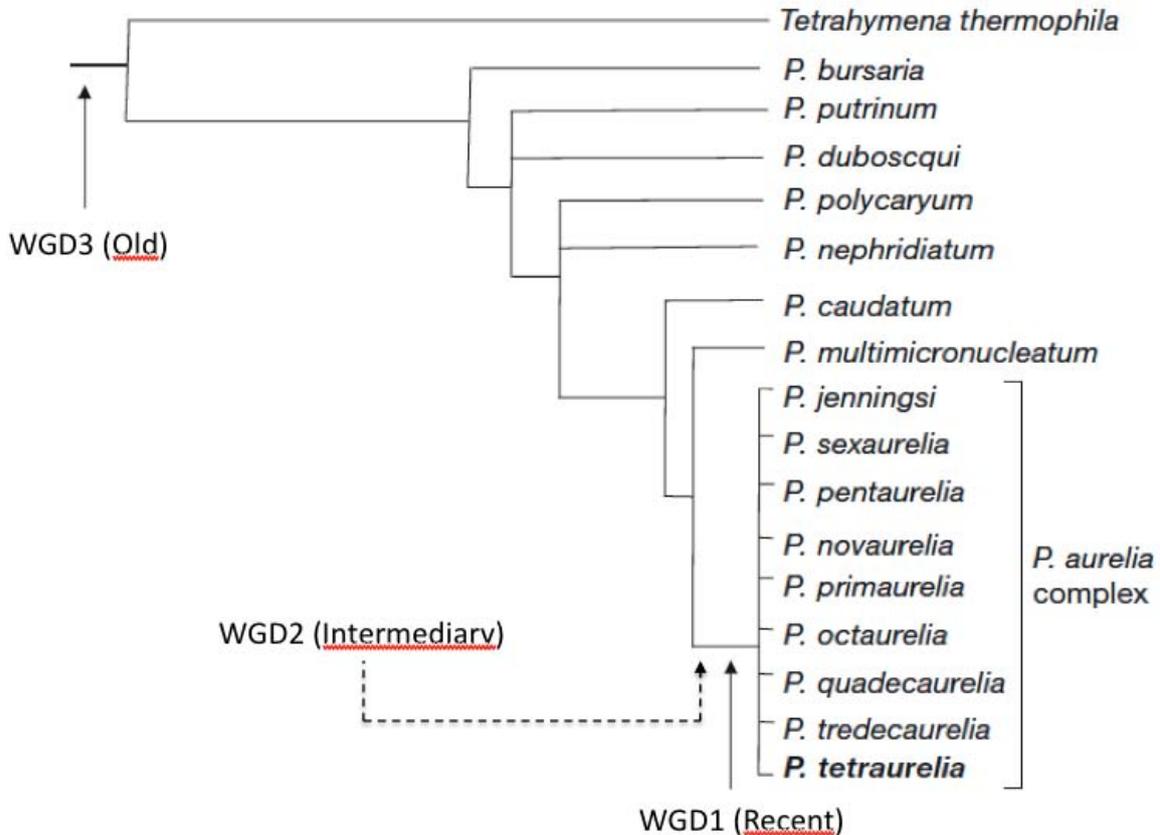
## A. Duplications du génome et spéciation

### 1) L'histoire du génome des *P. aurelia*

L'étude de *Paramecium aurelia* a débuté il y a 80 ans avec les travaux fondateurs de Tracy Sonneborn. Un ensemble de critères morphologiques lui a permis d'accumuler un grand nombre de souches qu'il attribuait à l'espèce « *P. aurelia* ». L'étude du comportement sexuel de ces souches lui ont permis de les classer en 14 « syngènes ». Pour chacun d'entre eux, il avait 2 types sexuels O et E qui réagissaient entre eux et généraient une descendance viable. Entre différents syngènes de types sexuels différents, il observait parfois une réaction d'agglutination et plus rarement une génération F1 partiellement viable mais jamais de F2 viables. Ainsi, suite à l'avènement de la définition biologique de l'espèce, ces syngènes sont devenus des espèces : *primaurelia*, *biaurelia*... *quadecaurelia*. L'espèce *P. sonnebornii* s'est par la suite ajoutée au complexe d'espèces jumelles *Paramecium aurelia*. Ces espèces sont aujourd'hui clairement identifiées par un génotype (Coleman, 2005).

Le séquençage du génome de la paramécie a révélé une histoire évolutive unique. Les 40 000 gènes identifiés ont émergé suite à 3 duplications successives du génome (Aury et al., 2006) qui sont nommées WGD1, WGD2 et WGD3 de la plus récente à la plus ancienne. Ainsi, la moitié des gènes de *P. tetraurelia* contient au moins un ohnologue WGD1, c'est-à-dire un paralogue issu de la dernière duplication du génome. Bien qu'aucune duplication n'ait initialement été identifiée dans le génome de *Tetrahymena*, une analyse par homologie a montré que la duplication la plus ancienne était partagée. Les deux duplications les plus récentes sont clairement spécifiques du groupe d'espèces jumelles *P. aurelia* et pourraient donc avoir joué un rôle dans la spéciation. La duplication intermédiaire n'est pas facilement

positionnable sur l'arbre puisqu'aucune espèce n'ayant des signes de WGD2 sans WGD1 n'est connue à ce jour. On peut donc supposer que cette duplication précède la plus récente à la base du groupe *P. aurelia* (**Figure IV.1**).



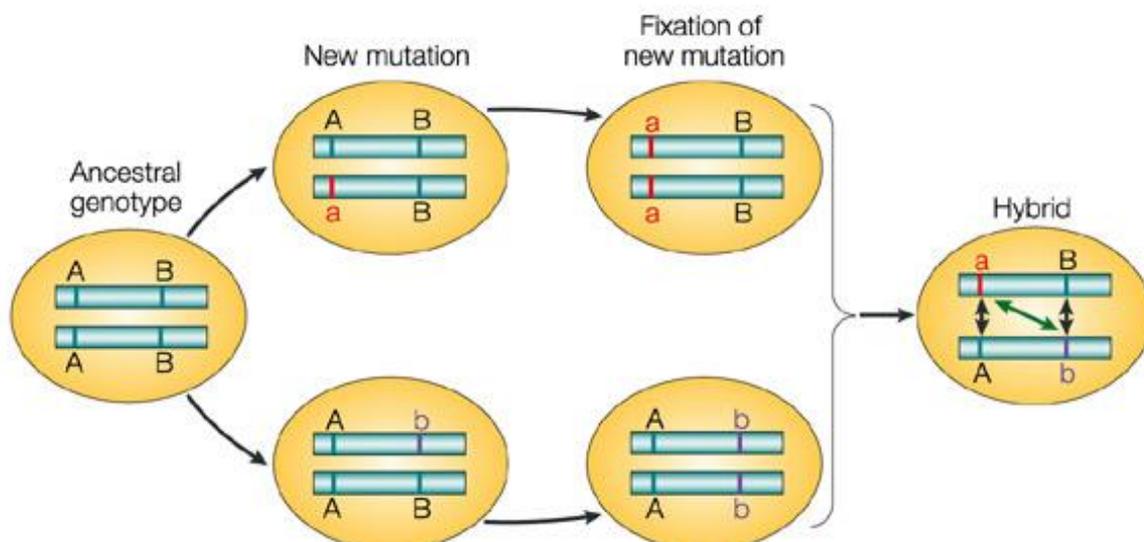
**Figure IV.1 : Les duplications du génome chez les ciliés.** Adapté de (Aury et al., 2006).

## 2) Devenir des gènes dupliqués

Trois duplications successives du génome devraient aboutir à la formation de 8 copies d'un même gène ancestral. En réalité, très peu de familles de gènes sont complètes car les gènes dupliqués peuvent être perdus par dérive génétique. Certaines copies sont encore visibles sous la forme de pseudogènes, c'est-à-dire des copies codant des transcrits aberrants du fait des nombreuses mutations accumulées. En effet, deux gènes issus d'une duplication ont initialement la même fonction et on peut comprendre que la perte d'une de ces deux copies ne soit pas préjudiciable à la cellule. Toutefois, il existe des contraintes qui peuvent favoriser la rétention des ohnologues comme par exemple la stœchiométrie et le fort niveau d'expression du gène (Gout et al., 2009). Il y a également des modèles d'évolution selon lesquels les ohnologues ne gardent pas la même fonction. La copie qui dérive peut occasionnellement accumuler des substitutions qui vont lui conférer une nouvelle fonction : c'est la néo-fonctionnalisation (Ohno, 1970). Il a aussi été proposé que les ohnologues puissent se partager les fonctions du gène ancestral, chacun retenant seulement une partie de la séquence (Force et al., 1999).

### 3) L'hypothèse de Bateson - Dobzansky - Muller

En génétique des populations, on parle d'isolement reproductif lorsque deux populations dérivées d'une même espèce donnent naissance à des hybrides de valeur sélective moindre. Il s'agit d'une étape du processus de spéciation avant que les croisements ne deviennent totalement stériles. Au début du 20<sup>ème</sup> siècle, Bateson, Dobzansky et Muller ont établi un modèle d'incompatibilité hybride entre deux populations accumulant indépendamment l'une de l'autre des mutations neutres ou avantageuses (**Figure IV.2**) (Orr, 1996). Ce modèle nécessite la considération de deux gènes évoluant séparément dans deux populations et dont l'interaction révélée en F1 est négative.

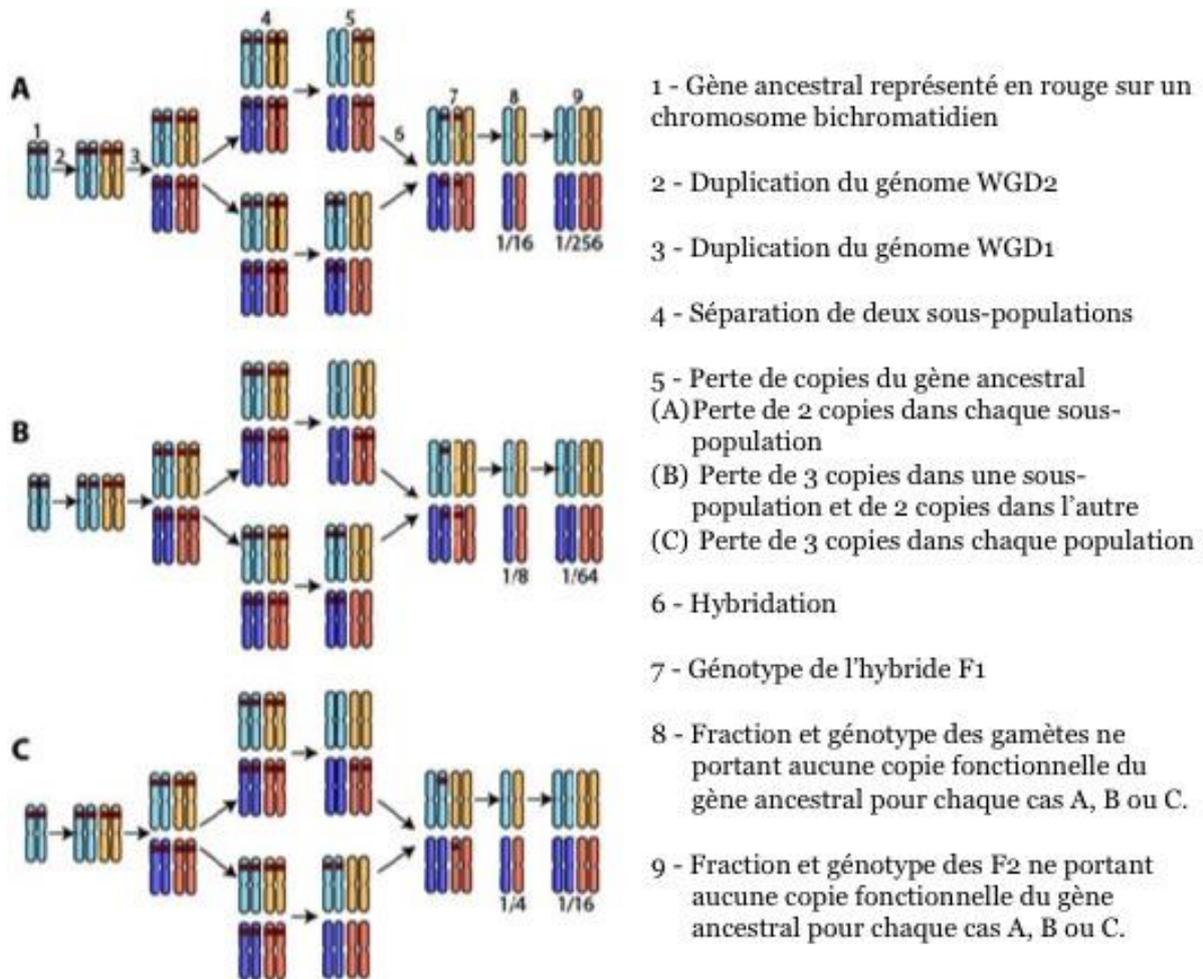


**Figure IV.2 : Modèle d'incompatibilité hybride de Bates-Dobzhansky-Muller.**

Le génotype de la population ancestrale est AA-BB. Après séparation, une sous-population devient aa-BB et l'autre AA-bb. L'incompatibilité apparaît en F1 de la présence simultanée des allèles a et b (double flèche verte).

Le modèle BDM s'applique de manière particulière après une duplication du génome. Si l'on considère les deux gènes du modèle comme étant ohnologues, alors la redondance fonctionnelle va favoriser la perte de l'une des deux copies indépendamment dans les deux populations. On parle de résolution divergente lorsque la copie perdue n'est pas la même et cela aboutira à des hybrides produisant une gamète non fonctionnelle sur 4 (Lynch and Force, 2000). Le groupe *P. aurelia*, avec ses trois duplications successives, est donc une excellente opportunité pour tester ce modèle. Les données de séquençage disponibles n'ont permis de comparer que les génomes de *biaurelia*, *tetraurelia* et *sexaurelia* avec *caudatum* comme groupe externe (Casey L. McGrath et al., 2014; C. L. McGrath et al., 2014). Les deux premiers génomes montrent seulement 113 cas de résolution divergente soit 2 % des

cas étudiés. Cela s'explique en grande partie par les 80 % de cas où un ohnologue a été perdu avant la spéciation. En revanche, 83 % des 5000 résolutions sont divergentes entre *tetraurelia* et *sexaurella*. Ainsi, bien que les résolutions divergentes ne soient pas toujours nombreuses, elles ont pu activement contribuer à la spéciation du groupe *P. aurelia* (**Figure IV.3**).



**Figure IV.3 : Modèle BDM adapté au complexe *P. aurelia*.**

Adapté de (C. L. McGrath et al., 2014).

## B. IES et dysgenèse hybride

### 1) Histoire évolutive des IES

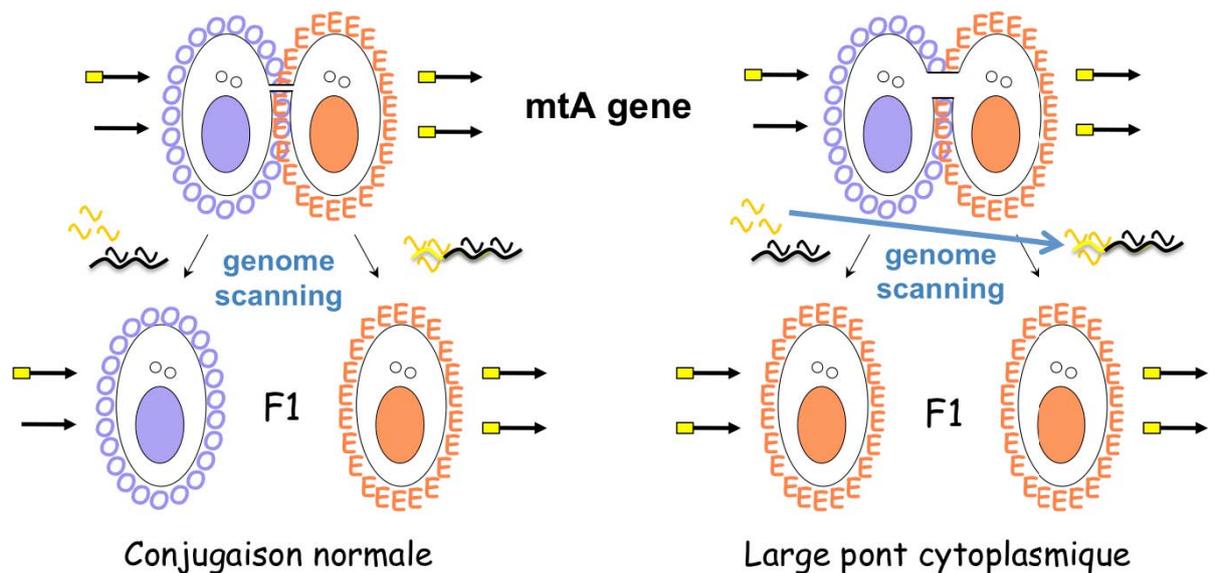
Les duplications successives du génome ont affecté les IES tout autant que les gènes. Ainsi une IES présente avant la duplication WGD3 peut exister en 8 copies si elles n'ont pas été perdues. De même, on peut dire qu'une IES apparue entre les duplications WGD2 et WGD3 existe en 4 copies, et ainsi de suite. En appliquant ce

raisonnement sur des quadruplets de gènes issus de la duplication WGD2 chez *P. tetraurelia*, on a pu estimer que 15 % des IES intragéniques prédataient cette duplication, 69 % sont apparues entre les deux dernières, et enfin 16 % depuis la plus récente (WGD1) (Arnaiz et al., 2012). Cette étude, basée uniquement sur le génome disponible de *P. tetraurelia* n'a pu considérer que 2 000 IES sur les 45 000 identifiées. Le séquençage des autres espèces du groupe, dont l'analyse sera bientôt terminée, permettra d'avoir une vision beaucoup plus large de l'évolution des IES.

## 2) Prédiction des problèmes d'excision en F1

### 1. Indépendance des réarrangements pendant la conjugaison

Chez *P. tetraurelia*, il est connu depuis longtemps que le type sexuel est hérité maternellement au cours d'une autogamie ou d'une conjugaison. Or, ce phénotype est le fruit du réarrangement alternatif de la séquence promotrice du gène *mtA* (Singh et al., 2014). Si le MAC retient cette séquence, alors le gène est exprimé et la cellule sera de type sexuel E, sinon elle sera de type O. Le contrôle de ce réarrangement dépend du *genome scanning*, car le promoteur de *mtA* se comporte comme une IES qui est excisée dans une lignée de type O. Autrement dit, les *scnARN* homologues au promoteur de *mtA* sont MIC-spécifiques dans la cellule O alors qu'ils trouvent une homologie dans les MAC E (*scnARN* MAC). Puisque le système régule non pas une IES dérivée d'un transposon mais bel et bien le fragment d'un gène cellulaire, on parle d'exaptation. Quoiqu'il en soit, l'hérédité de ce phénotype suggère que deux conjugants réarrangent indépendamment leurs *anlagen* au cours de la conjugaison. En revanche, l'induction artificielle de larges ponts cytoplasmiques (Sonneborn, 1947) pendant cet événement résulte en une descendance entièrement de type E. Cela signifie que dans une conjugaison normale, il y a suffisamment peu de cytoplasme échangé pour que les réarrangements soient indépendants. On suppose que l'échange de cytoplasme va permettre la titration des *scnARN* MIC-spécifiques de la cellule O par les transcrits provenant du MAC de la cellule E. Ainsi, ces *scnARN* seront considérés comme des *scnARN* MAC et ne pourront pas cibler l'excision du promoteur de *mtA* dans aucun des MAC zygotiques (**Figure IV.4**).

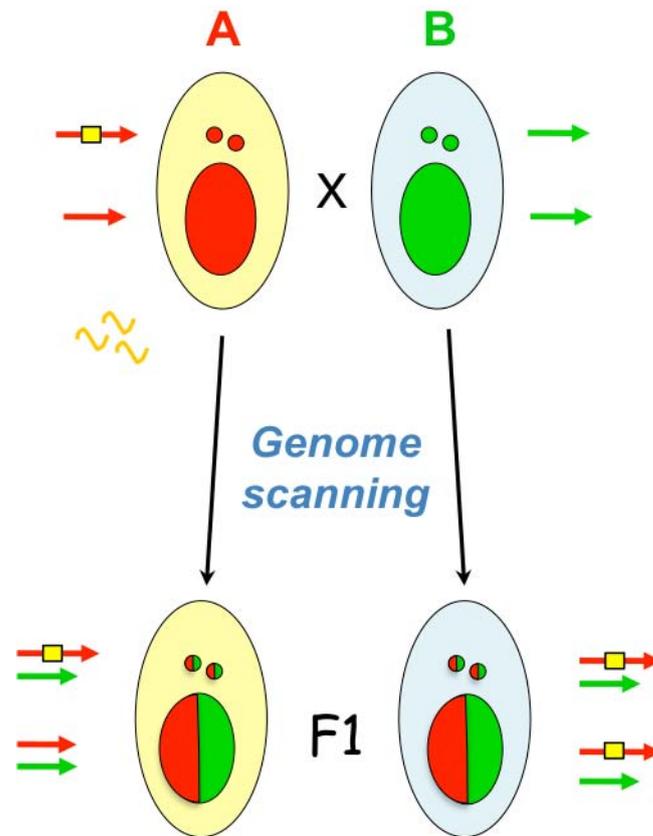


**Figure IV.4 : Hérité maternelle du type sexuel au cours de la conjugaison avec ou sans induction d'un large pont cytoplasmique.**

Le promoteur de *mtA* est représenté par le carré jaune. Le *genome scanning* est illustré par des petits ARN s'hybridant ou non sur des transcrits maternels. Tous les génomes sont homozygotes et schématisés par un seul génotype. Un seul caryonide F1 est représenté par ex-conjugant.

## **2. Les polymorphismes d'IES (absence/présence).**

La dysgenèse hybride décrite chez la drosophile résulte d'un défaut de répression des transposons paternels en l'absence de piARN maternels homologues. On peut prédire un phénomène similaire chez *Paramecium* : le défaut d'excision des IES paternelles en l'absence de scnARN maternels homologues. Un tel cas est prévisible lors d'un croisement entre une souche possédant une IES avec une autre ne la possédant pas. Le partenaire « naïf » n'ayant pas l'IES dans son génome micronucléaire ne peut produire de scnARN homologues lors de la méiose et ne sera donc pas en mesure de guider l'excision de l'IES présente dans l'allèle paternel, pendant le développement du macronoyau hétérozygote. Si l'IES est intragénique, alors l'allèle paternel du gène sera non fonctionnel (**Figure IV.5**).

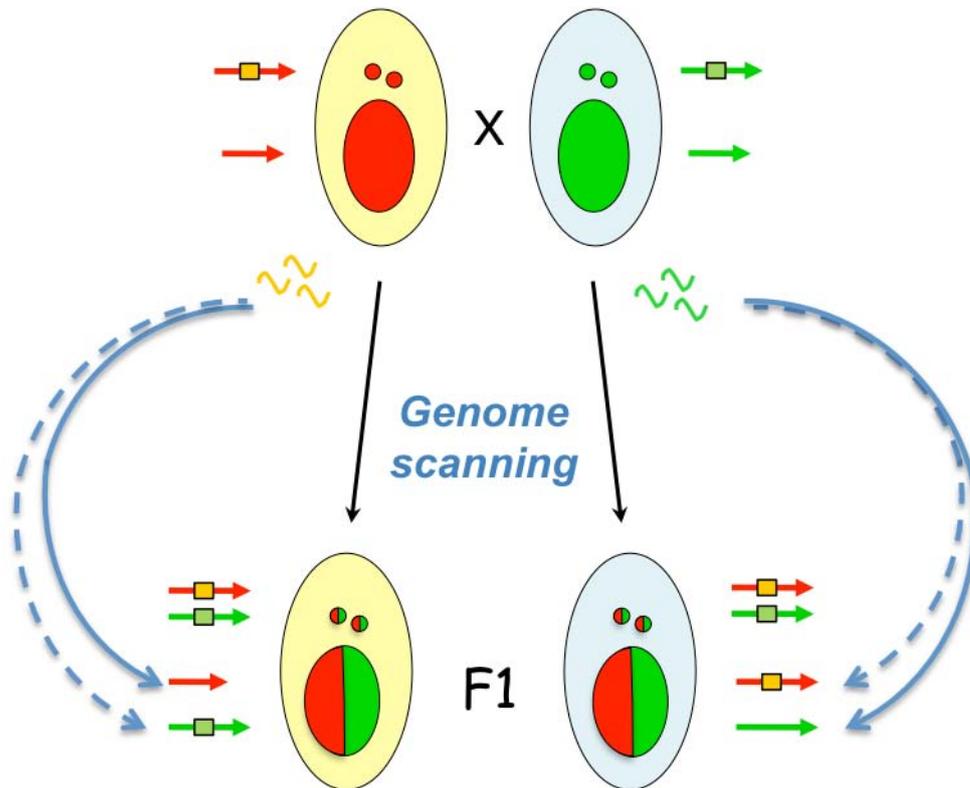


**Figure IV.5: Défaut d'excision prédit en F1 en cas de polymorphisme d'insertion d'une IES intragénique.**

Le parent A (génomme rouge) possède une IES (carré jaune) et produit des scnARN permettant sont excision au cours du développement du nouveau MAC hétérozygote. Le parent B (génomme vert) ne la possède pas, ne produit donc pas de scnARN homologues et retiendra l'IES héritée de son partenaire. Les ex-conjugants sont des hétérozygotes génétiquement identiques mais leurs génomes macronucléaires sont différents. Un seul caryonide F1 est représenté par ex-conjugant.

### **3. Les IES divergentes**

Le raisonnement établi précédemment s'applique aussi lorsque les IES orthologues de deux souches sont trop divergentes au point que les scnARN produits par l'allèle maternel ne peuvent reconnaître l'IES de l'allèle paternel (Figure IV.6). Il n'existe aucune donnée concernant le degré d'homologie nécessaire au ciblage des séquences par les scnARN.



**Figure IV.6 : Défaut d'excision prédit en F1 en cas de divergence allélique des IES.**

Le code couleur est le même que pour la figure IV.5. Chaque F1 peut exciser son allèle maternel mais pas l'allèle paternel hérité de son partenaire.

### 3) Létalité en F2

Si les défauts d'excisions prédits concernent des IES localisées dans des gènes essentiels, alors on pourra observer une létalité en F2. La génération F1, puisqu'elle est hétérozygote, contient au moins l'allèle maternel fonctionnel. En revanche, la moitié des F2 qui hérite de l'allèle paternel non réarrangé ne sera pas viable. Le pourcentage de F2 survivants sera alors de 25 % si l'on considère 2 IES et sera inférieur à 0,1 % à partir de 10 IES. D'ailleurs, il est courant d'observer une létalité en F2 lorsqu'on croise deux souches appartenant à la même espèce. Le croisement des souches 51 et 32 de *P. tetraurelia* donne 25 % de survie en F2 des deux côtés du croisement. Une partie de cette létalité s'explique peut-être par des défauts d'excision en F1. En résumé, puisque 16 % des IES ont été acquises depuis la divergence de *P. tetraurelia* par rapport aux autres espèces de *P. aurelia*, la forme de dysgenèse hybride décrite ici pourrait avoir contribué à la spéciation du groupe. En d'autres termes, la reconnaissance scnARN-dépendante des nouvelles insertions pourrait avoir contribué à l'émergence des espèces jumelles de *P. aurelia* suite à une invasion de transposons dans la population ancestrale. C'est dans le cadre de cette hypothèse que s'inscrit mon projet de thèse. Il est important de noter que ce modèle n'est valable que pour les IES qui dépendent des scnARN pour leur excision c'est-à-dire les IES contrôlées maternellement. Bien que cette fraction soit restreinte, rappelons que

tous les transposons sont reconnus par le genome scanning et que toutes les IES étaient probablement des transposons au moment de leur apparition. Mon projet n'a pas pu porter sur la dysgenèse directement liée aux transposons car ces derniers ne sont pas encore connus (assemblage du génome MIC non publié).

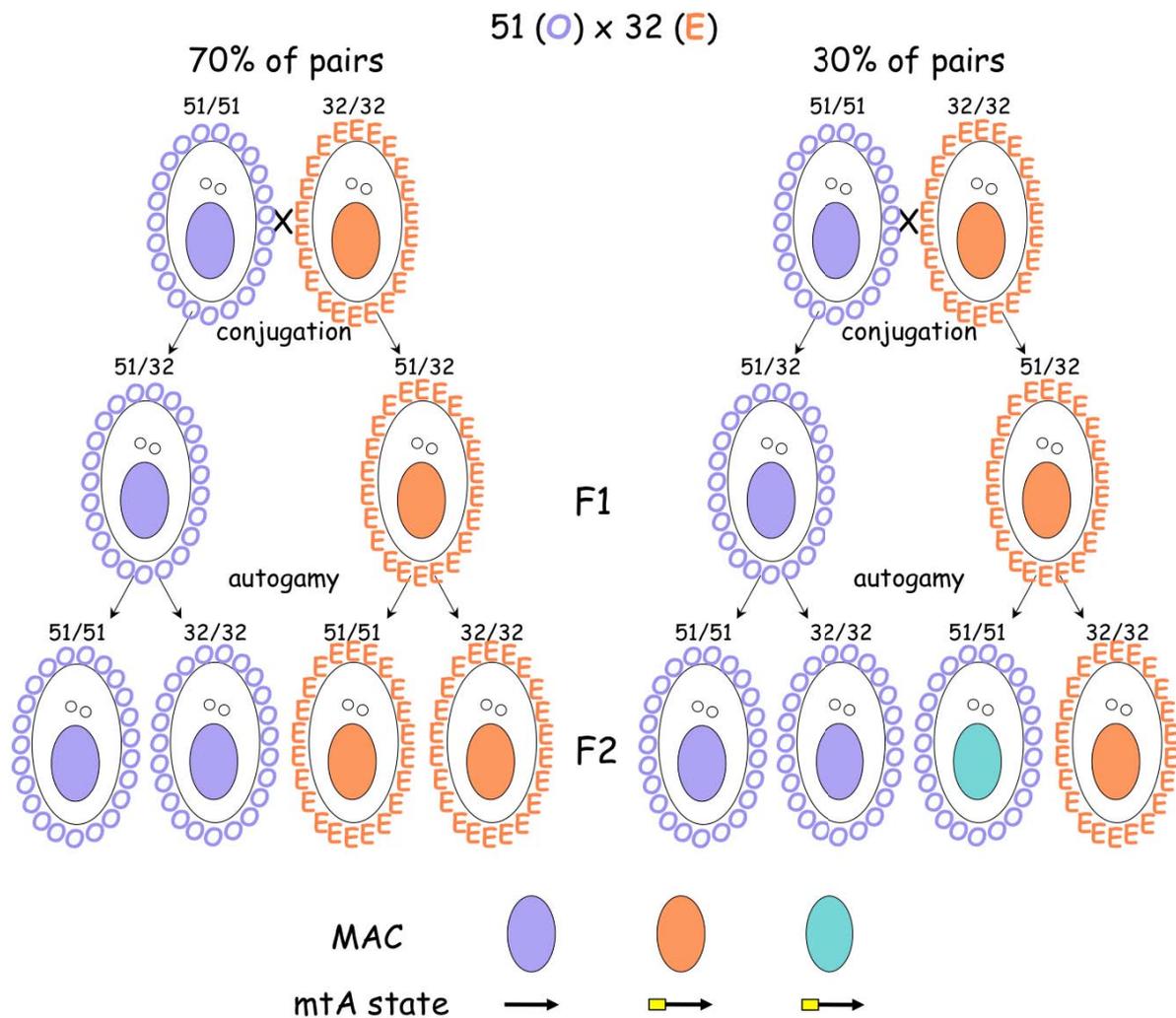
#### 4) Données préliminaires

##### 1. Connaissance des génomes

Afin de tester cette hypothèse, il est nécessaire de bien connaître les génomes de différentes souches et d'identifier les polymorphismes d'IES. Le génome de la souche 51 est bien connu puisqu'il s'agit du génome de référence de l'espèce. La répression de Pgm a permis d'avoir accès aux IES et la récente purification des MIC a permis de générer des données encore plus robustes (non publié). Le génome MAC de la souche 32 a également été séquencé mais il n'y a pas eu d'assemblage des lectures. Ces dernières ont simplement été alignées sur le génome 51 et seules les substitutions présentes dans des séquences alignables ont été prises en compte. Le génome 32 ainsi généré présente 0,3 % de substitutions par rapport au génome 51, c'est-à-dire 3 pour 1 kb. Les régions trop divergentes ou présentes seulement dans la souche 32 ne figurent pas dans ce génome provisoire. Inversement, les régions présentes dans la souche 51 mais pas dans la souche 32 y sont artificiellement maintenues. Enfin, les IES de la souche 32 ont été détectées à partir de lectures d'un séquençage « pgm » en utilisant la même technique que pour la souche 51. Au cours de ma thèse, le génome 32 a été réassemblé *de novo* et les génomes de nombreuses autres souches appartenant à d'autres espèces du groupe *P. aurelia* ont été séquencés dans le cadre d'un projet avec le Génoscope d'Evry.

##### 2. Le défaut d'excision de l'IES de mtB dans le croisement 51x32

Pour comprendre l'émergence de cette hypothèse, il convient de s'intéresser aux croisements réalisés par Brygoo en 1977 avec les souches 51 et 32 (Brygoo, 1977a). Alors qu'un croisement 51 x 51 ou 32 x 32 montre une hérédité maternelle du type sexuel, il a observé un résultat étrange dans le croisement 51-O x 32-E. 30 % des F1 dérivant du parent 32 génèrent des F2 qui deviennent O alors qu'il devraient être E (**Figure IV.7**). Brygoo a réussi à réaliser 11 cycles de *backcross* de la souche 32 vers la souche 51 en réisolant les F2 capables de reproduire le phénotype. Il a ainsi montré qu'un seul locus, appelé mtD, était impliqué et que tous les F2 révertant de E vers O portaient l'allèle mtD51. En outre, il a noté que ce type sexuel O n'avait pas les mêmes propriétés que le type O classique d'où l'appellation O\*. Les cellules O\* produites par le croisement révertent spontanément vers E à l'autogamie avec une fréquence de 1/30 alors que les O classiques ont une fréquence inférieure à 1/50 000. De plus, lorsqu'ils sont recroisés avec le parent 32-E, tous les F1 sont E. Cette observation montre que les cellules O\* sont déterminées pour exprimer le type sexuel E mais ne l'expriment pas car l'allèle mtD est non fonctionnel. Pour le confirmer, notre laboratoire a montré que le promoteur de mtA était normalement retenu dans les F2 O\* issus du croisement 51 x 32 (D. P. Singh).

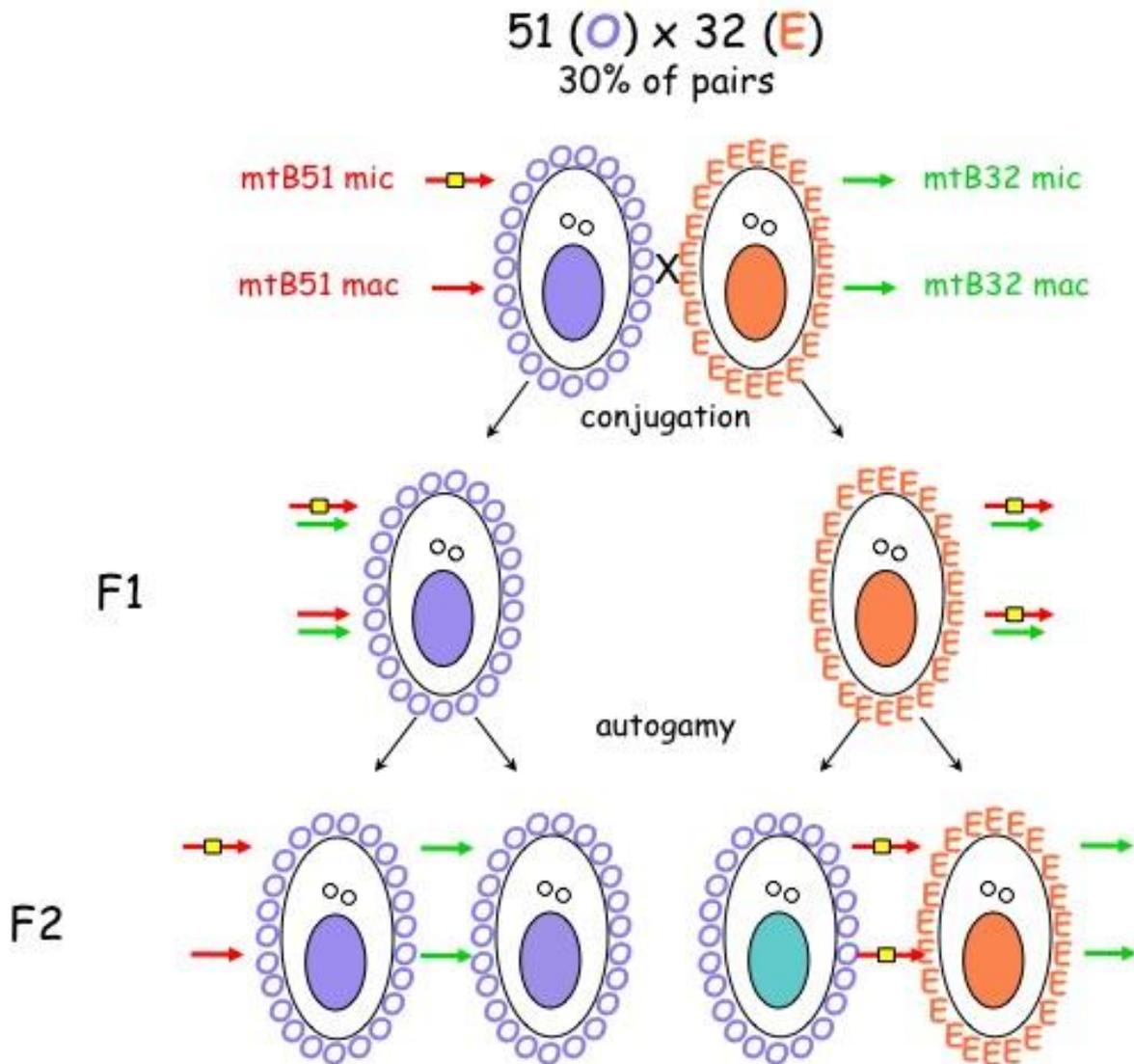


**Figure IV.7 : Phénomène mtD observé par Brygoo en 1977.**

Lors d'un croisement de la souche 51 avec la souche 32, 70 % des paires présentent une hérédité normale du type sexuel. 30 % des couples donnent une descendance F2 dans laquelle le type sexuel O (O violets) est exprimé à la place du type E (E oranges) attendu. Ces F2 furent qualifiés de O\*. Bien qu'ils soient phénotypiquement O, ils diffèrent des O classiques car ils retiennent le promoteur de mtA et sont déterminés pour O par un autre mécanisme, lié au locus mtD. Le génotype du locus mtD est indiqué pour chaque cellule (51/51, 51/32 ou 32/32).

Par ses tests génétiques, Brygoo avait montré que le locus mtD ne faisait pas partie des gènes déjà connus pour empêcher l'expression du type sexuel E : mtA, mtB ou mtC (Byrne, 1973). En séquençant la souche isogénisée de Brygoo dans laquelle l'allèle mtD32 est introduit dans le fond génétique 51, notre laboratoire a identifié le locus mtD comme étant mtB ! L'erreur de Brygoo, qui avait écarté mtB sur la base d'un individu, provient sans doute d'une recombinaison intragénique dans le gène de mtB. Le séquençage du gène a révélé que l'allèle 51 contient une IES surnuméraire par rapport à l'allèle 32. Des PCR sur la région autour de l'IES ont montré que les 30 % de couples générant le phénotype O\* étaient ceux qui présentaient un défaut d'excision en F1 (**Figure IV.8**) (A. Potheikin). Autrement dit, cette observation est la preuve qu'un polymorphisme d'insertion d'IES peut empêcher l'excision correcte de

l'IES dans le F1 dérivant du parent naïf. Une fois que l'IES est retenue dans le MAC F1, alors elle sera retenue en F2 comme toute IES contrôlée maternellement.



**Figure IV.8: Explication moléculaire du phénotype O\* par le statut d'excision de l'IES de mtB.**

Les flèches représentent le gène mtB. L'allèle 51 (rouge) contient une IES (carré jaune) normalement excisée dans le MAC. L'allèle 32 (vert) ne possède pas cette IES. Le croisement 51 x 32 génère 30 % de couples caractérisés par un défaut d'excision conforme au modèle de dysgenèse hybride. Ce défaut est responsable du phénotype O\* pour les F2 héritant de l'allèle 51 non fonctionnel. Le promoteur de mtA, déterminant le type sexuel E, est bien retenu dans le parent 32-E et tous ses descendants. Dans les F2 O\*, le type E ne peut pas s'exprimer car le gène mtB n'est pas fonctionnel.

## 5) Projet de thèse

La redécouverte moléculaire des observations de Brygoo a mis en évidence le fait qu'un polymorphisme d'insertion pouvait générer un défaut de réarrangement dans les F1 hétérozygotes 51/32. Ainsi le défaut d'excision de l'IES de mtB dans les F1 dérivant du parent 32 semble conforter la nécessité de scnARN maternels. Le premier objectif de ma thèse consistait à documenter ce fait en essayant de l'étendre à d'autres polymorphismes d'insertion et à des cas de divergence. A terme, la généralisation de ces observations dans des croisements intersouches devait nous permettre d'évaluer l'impact de ce phénomène dans la spéciation du groupe *P. aurelia*. Très vite, mes résultats ont montré que les défauts de réarrangements étaient limités par le rôle des scnARN paternels. Mon deuxième objectif visait alors à comprendre ce mécanisme totalement nouveau. J'ai travaillé sur deux hypothèses, celle d'une transmission de scnARN entre les partenaires sexuels et celle d'un marquage épigénétique des noyaux gamétiques. Enfin, mon troisième objectif, rendu nécessaire par la découverte de l'action des scnARN paternels, était d'apporter des éléments nouveaux au modèle du *genome scanning*. En particulier, j'ai essayé de développer une méthode pour purifier des cellules en conjugaison afin de réaliser des études cinétiques sur une population hautement synchrone.

# **RESULTATS**

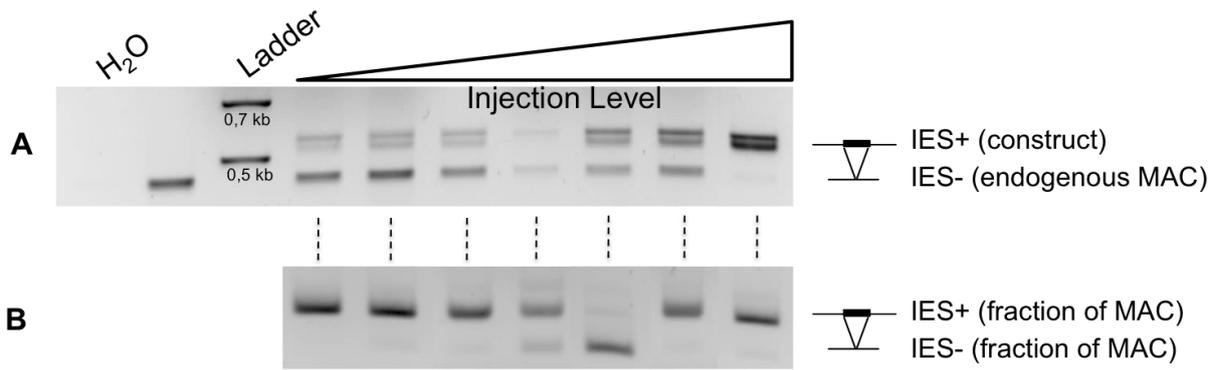
# V - Effet des polymorphismes d'insertion d'IES avec l'exemple de mtB

## A. Les polymorphismes d'insertion

La comparaison des IES identifiées dans les génomes 51 et 32 a permis d'identifier 7 polymorphismes d'insertions. Ce chiffre est sans doute sous-estimé puisque les régions trop divergentes ne sont pas présentes dans notre génome 32 provisoire (voir IV-B-4). L'IES de mtB, qui comporte 134 pb, est absente de la souche 32. Les deux allèles du gène diffèrent de 4 substitutions dont 2 à moins de 25 nt d'une des bornes de l'IES (**Figure S1**). Une autre IES, relativement longue (99 pb), est exclusivement présente sur le scaffold 101 du génome 32. Elle est analysée dans le paragraphe V-E. Les 5 autres cas de polymorphismes trouvés concernent des IES de petite taille dont les tests préliminaires de rétention chez des F1 hétérozygotes 51/32 n'ont pas montré de rétention (Potekhin Lab). Ces IES n'ont pas été étudiées plus en détail par la suite.

## B. L'IES de mtB est contrôlée maternellement

Le défaut d'excision prédit par notre modèle ne concerne que les IES contrôlées maternellement. Puisque l'IES de mtB n'est pas correctement excisée dans 30 % des couples (cf. IV-B-4), alors elle doit probablement en faire partie si notre modèle est correct. Le premier élément suggérant cette propriété est le fait qu'une cellule homozygote retenant l'IES (cellule O\*) garde ce phénotype après autogamie dans 29 cas sur 30 (Brygoo, 1977b). Cela signifie moléculairement que la rétention de l'IES de mtB, lorsqu'elle est établie dans le MAC, induit la rétention de l'IES dans la génération suivante. Pour renforcer cette observation, j'ai injecté un fragment génomique de 3,6 kb contenant l'IES de mtB dans une cellule 51 sauvage. Le niveau d'injection peut être évalué en réalisant une PCR avec des amorces de part et d'autre de l'IES (**Figure V-1 A**). La forme contenant l'IES ne peut provenir que de la construction injectée alors que la forme excisée provient du MAC. L'autogamie des clones transformés entraîne la dégradation de l'ancien MAC et du transgène, ainsi que le réarrangement d'un nouveau MAC. La même PCR peut alors être utilisée pour évaluer le pourcentage des copies MAC qui retiennent l'IES dans les populations post-autogames (**Figure V-1 B**). Parmi les descendances des 7 clones testés, 5 retiennent complètement l'IES, une la retient fortement, et une autre a excisé totalement l'IES. Ce dernier cas montre que la rétention de l'IES n'est pas systématiquement induite alors même qu'un plus faible nombre de fragments injectés peut être suffisant. Les souches retenant l'IES de mtB dans leur MAC sont des épimutants puisque leur MIC est sauvage, elles peuvent être maintenues en stock et seront appelées mtB+ par la suite. Pour conclure, on peut affirmer que l'IES de mtB est contrôlée maternellement.



**Figure V.1 : L'injection de l'IES de mtB entraîne sa rétention à la génération suivante.**

Un fragment génomique de 3,6 kb contenant l'IES de mtB est injecté dans des cellules. (A) Le niveau de rétention est estimé avec une PCR utilisant des amorces autour de l'IES, les produits amplifiés font respectivement 438 et 572 pb pour les formes IES- et IES+. (B) Pour chaque clone, l'ADN d'un pool de 15 cellules post-autogames a été extrait et la même PCR est utilisée pour estimer la fraction du MAC retenant l'IES.

## C. L'IES de mtB est partiellement retenu en F1

### 1) Analyse des taux de rétention en F1

Suite aux résultats préliminaires montrant la rétention de l'IES de mtB dans les F1 51/32, nous avons répété le croisement 51 x 32 plusieurs fois et extrait l'ADN des quatre caryonides F1. On a vérifié qu'ils étaient hétérozygotes et déterminé leur origine maternelle (test du type sexuel et/ou PCR mitochondriale). La quantification des bandes obtenues après PCR sur la région de l'IES a permis de calculer un taux de rétention de l'IES de mtB. Nous avons retenu une seule valeur moyenne de rétention par paire de caryonides frères, on considère qu'elle est représentative de l'ex-conjugal qui les a formés (**Figure V.2 A**). Dans les cas où un seul des deux caryonides a survécu, nous avons utilisé directement la valeur pour représenter l'ex-conjugal. Seul l'allèle 51, contenant une IES dans sa version MIC, peut présenter un taux de rétention. Puisque la PCR ne permet pas de le distinguer de l'allèle 32, le taux de rétention est estimé en supposant que le MAC (800 n) des hétérozygotes possède autant d'allèles 32 que d'allèles 51. La PCR génère souvent une minorité d'hétéroduplex IES-/IES+ qui sont prises en compte dans le calcul. Deux séries totalement indépendantes ont été réalisées : 51-O x 32-E (Laboratoire Meyer) et 51-E x 32-O (Laboratoire Potekhin) (**Figure V.2 B à E**). Un croisement 51-O x 51-E a été réalisé en contrôle de la première série (**Figure S3**) et montre une excision totale de l'IES de mtB dans les F1 51/51.

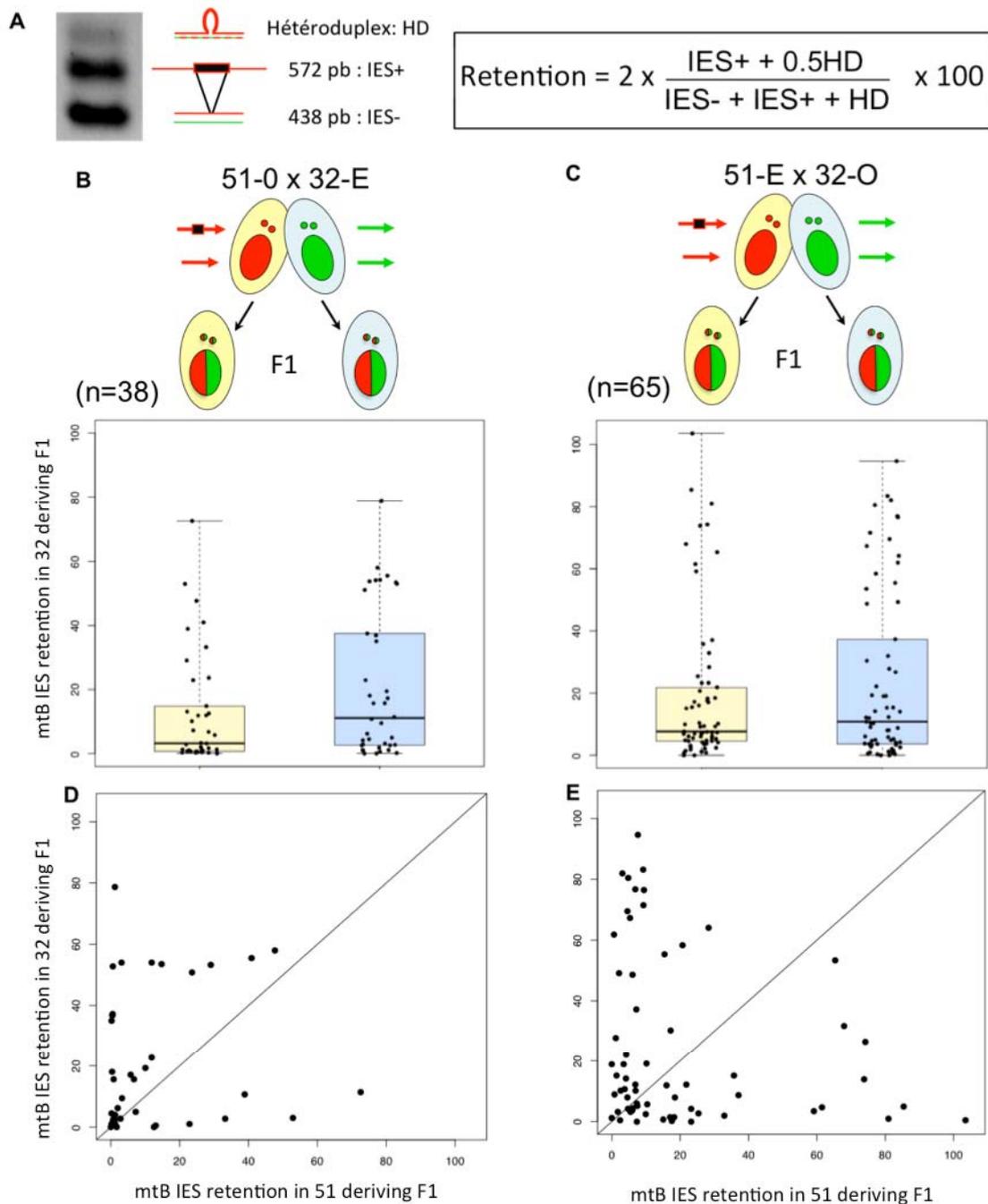
Les résultats montrent une étonnante variabilité dans la rétention de l'IES. Quelle que soit leur origine maternelle, la majorité des F1 affiche un niveau de rétention proche de 0 témoignant de l'excision correcte de l'IES de mtB. En revanche, une fraction non négligeable des F1 retient l'IES à des taux variables, allant parfois jusqu'à 100 %. Malgré le continuum des valeurs de rétention, il semble qu'un groupe de points au-dessus de 40 % se dégage dans les deux séries. Ce groupe, qui contient 25 % des F1, rappelle les 30 % de cas où Brygoo observait l'apparition du phénotype

O\* en F2 (Brygoo 1977). En effet, un taux de rétention de 40 % est probablement suffisant pour induire une rétention totale à la génération suivante par effet maternel (voir V-C-5). Par ailleurs, ce défaut d'excision de l'IES porté par l'allèle 51 dans les ex-conjugants dérivant du parent 32 correspond à la prédiction du modèle. On pourrait l'attribuer à un défaut de scnARN maternels homologues à l'IES cependant l'observation de l'autre côté du croisement ne nous y incite pas. En effet, il semble que les ex-conjugants dérivant du parent 51, qui produisent pourtant les scnARN contre l'IES, éprouvent également des difficultés pour l'exciser. Ce résultat s'oppose à la fois à notre prédiction et aux résultats de Brygoo qui décrivait déjà ce phénomène mais seulement dans 5 % des couples. Ici, il concerne davantage de F1 au point qu'aucune différence significative (test des signes de Wilcoxon) n'existe entre les deux côtés du croisement. Pour aller plus loin, les résultats peuvent être représentés sous la forme d'un nuage de points afin d'observer la corrélation entre la rétention dans les F1 dérivant du parent 51 et celle dans les F1 dérivant du parent 32 (**Figure V.2 D et E**). Les faibles corrélations négatives calculées ne sont pas significatives malgré le manque apparent de points sur la droite  $y = x$  (Test de Spearman).

En résumé, les résultats de ces croisements montrent qu'il existe un défaut d'excision en F1, partiel, et affectant les F1 hétérozygotes indépendamment de leur origine maternelle. On peut parler d'effet hémizygote puisque c'est la présence dans le noyau zygotique d'un allèle portant une IES avec un allèle n'en portant pas qui semble être à l'origine du problème. La deuxième leçon à tirer de ce croisement est le fait que 70 % des F1 dérivant du parent 32 sont capables d'exciser correctement l'IES de l'allèle paternel 51 malgré l'absence de scnARN maternels homologues. Cela signifie que notre modèle actuel du *genome scanning* ne peut pas rendre compte de nos résultats dans le cas présent.

## 2) Comparaison des deux séries

Les résultats obtenus pour les deux séries (**Figure V.2 B et C**) sont relativement similaires mais pas identiques. Un test de Kolmogorov-Smirnov montre une différence significative pour la rétention dans les F1 dérivant du parent 51. Ces différences peuvent avoir une origine technique ou biologique. On peut imaginer que la PCR (polymérase, protocole) ou la quantification des gels puisse avoir un léger effet. Attribuer cette différence aux souches ou au sens du croisement semble plus difficile à concevoir. Notamment, les valeurs de la série V-2B sont globalement plus faibles, dépassant rarement 80 %. Il est possible que les conditions PCR utilisées génèrent un biais en faveur du produit PCR le plus court, sans l'IES, d'où une sous-estimation du taux de rétention réel.

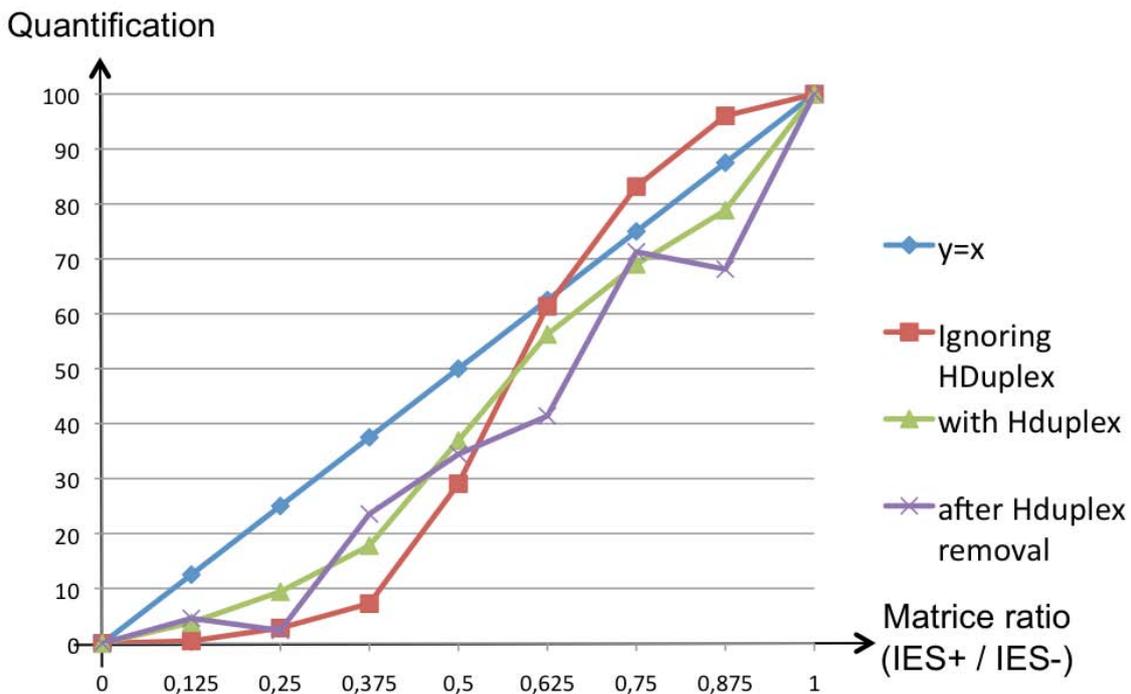


**Figure V.2 : Rétention de l'IES de mtB dans les F1 hétérozygotes 51/32.**

(A) Le calcul du taux de rétention de l'IES de mtB se fait selon la formule indiquée. Seul l'allèle 51 (rouge), possédant une IES dans sa version MIC, peut retenir l'IES dans sa version MAC. Par opposition, l'allèle 32 (vert) n'existe que sous la forme IES-. La forme hétéroduplex HD est considérée dans la formule en ajoutant la moitié du signal à la forme IES+ et l'autre moitié à la forme IES-. (B à E) Deux séries de croisements 51 x 32 indépendants et utilisant des types sexuels inversés ont généré respectivement 38 et 65 couples dont l'ADN des 4 caryonides hétérozygotes a été extrait. Le taux de rétention moyen est calculé pour chaque ex-conjugal à partir des taux des deux caryonides frères. Les valeurs sont représentées par ex-conjugal sous la forme de boxplots avec affichage de tous les points (B et C) et par couple sous la forme de graphique 2D (D et E).

### 3) Test d'un biais PCR

Pour le tester, j'ai mélangé différents ratios molaires des formes IES+ et IES- obtenues à partir de bandes purifiées sur gel. J'ai ensuite amplifié ces mélanges avec la PCR puis quantifié le résultat (**Figure V.3**). Le calcul est fait en prenant en compte ou non la forme hétéroduplex, ou après une étape supplémentaire de polymérisation qui enlève l'hétéroduplex. Les 3 courbes sont en dessous de la droite  $y = x$ , ce qui indique une sous-estimation de la quantité de la forme IES+ initiale. Un ratio équimolaire, équivalent à 50 % de rétention, donne typiquement un résultat autour de 35 %. Cela peut expliquer pourquoi une rétention totale de l'IES dans les hétérozygotes de ma série V-2B ne pourrait pas dépasser un taux de 70-80 %. La prise en compte de la forme hétéroduplex est critique lorsqu'une des deux formes est très minoritaire car la fraction de cette forme piégée dans l'hétéroduplex sera non négligeable. Lorsqu'une PCR de reconditionnement est réalisée pour enlever l'hétéroduplex (Thompson et al., 2002), la quantification est proche de celle faisant le calcul incluant l'hétéroduplex. Cela montre bien qu'il existe un biais intrinsèque, indépendant de cette forme, qui favorise la forme IES-. Au final, enlever l'hétéroduplex ne permet pas d'améliorer la quantification.

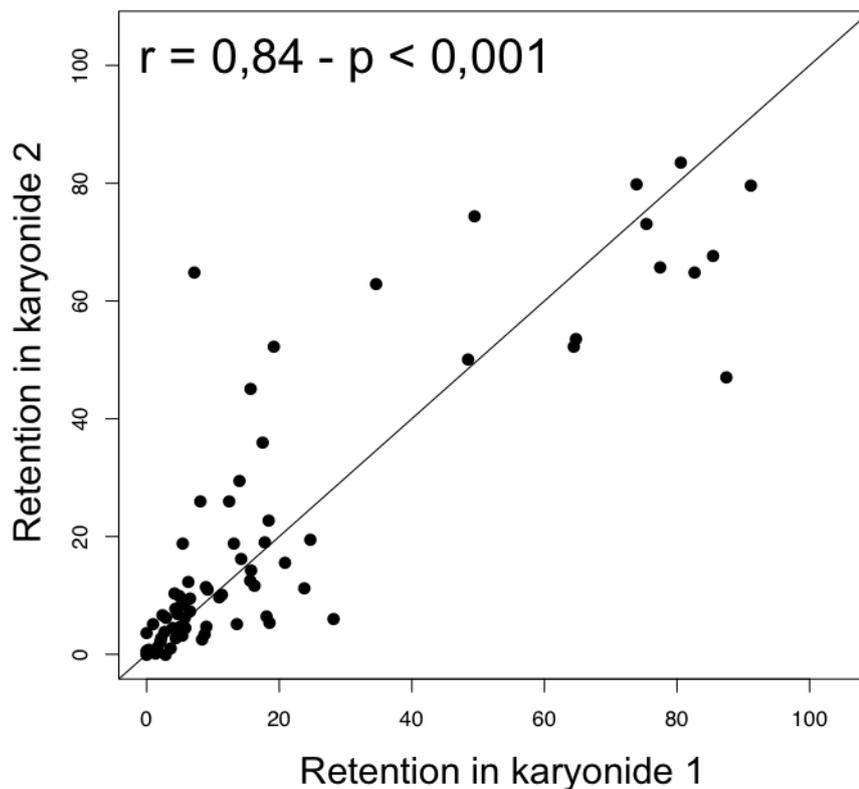


**Figure V.3 : Test du biais PCR dans le calcul de la rétention de l'IES de mtB.**

Différents ratio molaires de produits PCR IES+ et IES- ont été utilisés comme matrice. Après amplification par PCR, les bandes IES+ et IES- sont quantifiées afin de les comparer avec le ratio initial. Le calcul est fait en prenant en compte ou non la forme hétéroduplex, ou après une étape supplémentaire de polymérisation qui enlève l'hétéroduplex.

#### 4) Corrélation caryonidale

Chaque ex-conjugal produit deux nouveaux MAC qui sont séparés à la première division dite caryonidale. Les deux caryonides frères donnent naissance à deux clones possédant chacun un des deux MAC formés indépendamment dans le même cytoplasme. Dans les croisements de la figure V-2C, chaque échantillon d'ADN provient d'une population caryonidale pure, avec souvent deux échantillons par ex-conjugal F1, un pour chaque caryonide. Cela a permis d'évaluer la corrélation pour la rétention de l'IES de mtB dans 73 paires de caryonides frères, origines maternelles confondues (Figure V.4). La forte corrélation positive observée montre que l'excision de l'IES de mtB est guidée par un processus commun aux deux *anlagen*. C'est pourquoi nous avons décidé de garder une seule valeur moyenne de rétention par F1 quand nous disposons des valeurs des deux caryonides frères. Dans les croisements suivants, un seul échantillon d'ADN est extrait pour chaque ex-conjugal, à partir d'une population clonale dérivant d'un seul des deux caryonides.

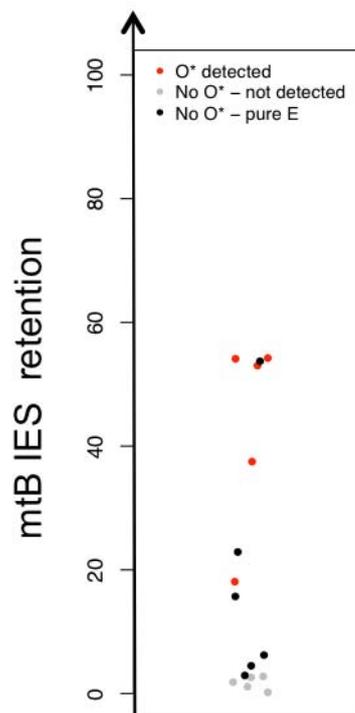


**Figure V.4 : Corrélation caryonidale pour la rétention de l'IES de mtB.**

Pour chacune des 74 paires de caryonides frères, la rétention de l'IES de mtB d'un caryonide arbitraire est représentée en fonction de la rétention dans l'autre caryonide. La méthode de Spearman est utilisée pour calculer le coefficient de corrélation et la p-value associée.

## 5) Taux de rétention et phénotype O\*

La rétention de l'IES de mtB en F1, puisqu'elle est variable, n'induit pas nécessairement une rétention totale à la génération suivante dans les F2 héritant l'allèle 51. Or seuls les allèles non fonctionnels en F2 peuvent causer un phénotype, une réversion du type sexuel de E vers O (phénotype O\*) dans notre cas. Faire le lien entre le taux de rétention en F1 et le phénotype en F2 permet d'évaluer l'impact d'une possible dysgenèse hybride dans le cas où une IES serait présente dans un gène essentiel. Pour le tester, des cultures F1 dérivant du parent 32-E ont été menées à l'autogamie puis renourries afin de déclencher la réactivité sexuelle. L'observation d'agglutination indique qu'une partie des F2 a changé de type sexuel, vraisemblablement parce que le gène mtB est non fonctionnel (**Figure V.5**). A partir de 35 % de rétention en F1, le phénotype O\* est observé dans 4 populations F2 sur 5 alors qu'en deçà de 25 %, il y a une seule observation sur 11. Le seuil n'est donc pas clairement défini mais cette estimation semble cohérente puisqu'elle permet de retrouver les résultats de Brygoo dans mes croisements pour le côté 32 (**Figure V.2 B**). Enfin, nous devons garder à l'esprit que ce seuil concerne l'IES de mtB et n'est probablement pas applicable aux autres IES puisque le degré de contrôle maternel est propre à chaque IES.



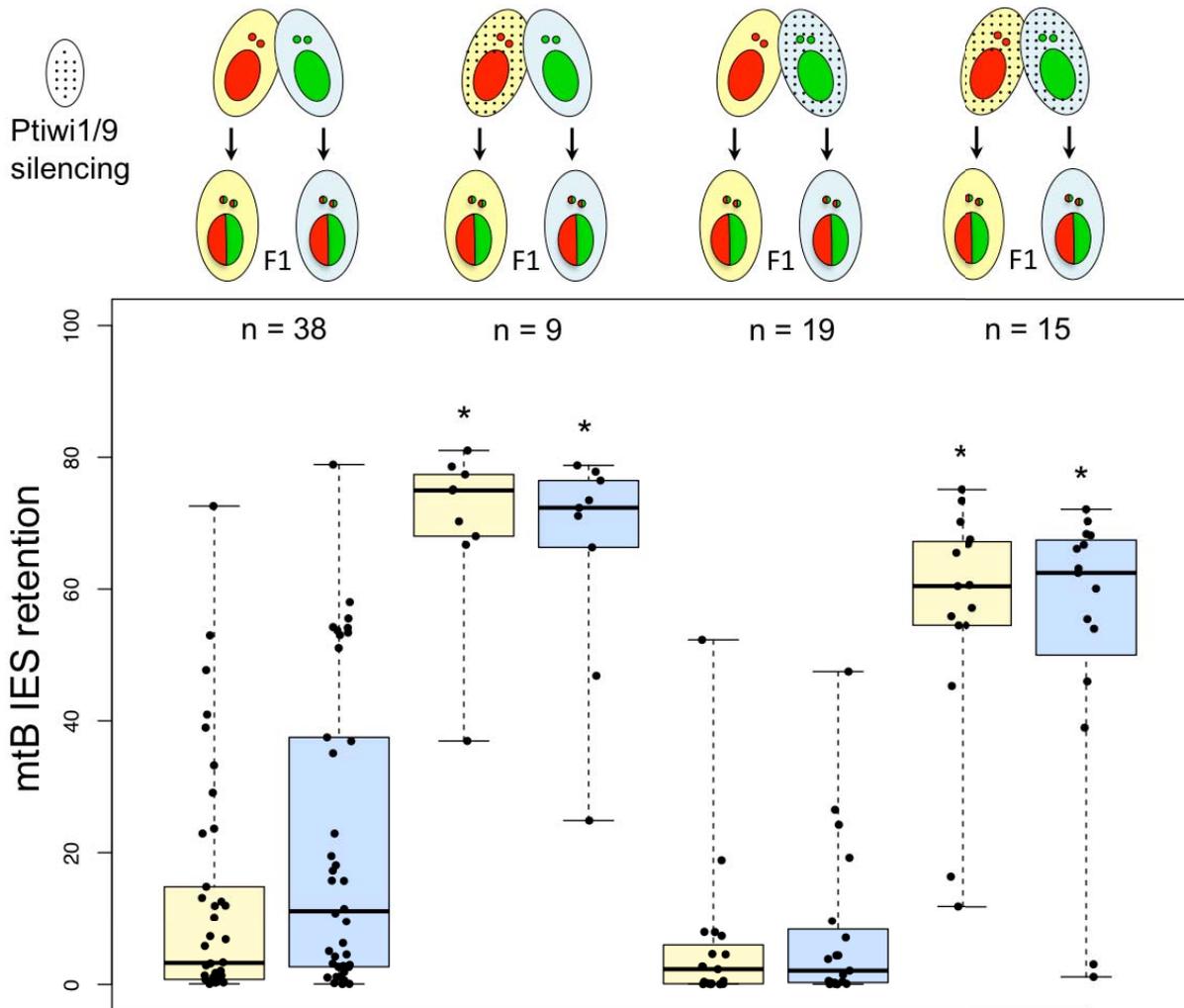
**Figure V.5 : Rétention de l'IES de mtB en F1 et phénotype o\* en F2.**

19 F1 dérivant du parent 32 et de type sexuel E ont été cultivés jusqu'à l'autogamie. Les 19 populations F2 obtenues ont été renourries puis observées afin d'identifier une réaction d'agglutination révélatrice du phénotype O\*. Lorsqu'aucune réaction n'a été observée, il a été possible de confirmer pour la moitié des populations qu'elles étaient entièrement E (points noirs).

## D. Les scnARN paternels peuvent exciser l'IES de mtB chez le partenaire naïf.

### 1) Croisement avec déplétion de Ptiwio1&09

Pour mieux comprendre comment 70 % des F1 dérivant du partenaire naïf parviennent à exciser l'IES de mtB, j'ai interféré avec la voie des scnARN en réprimant les gènes *PTIWIO1* et *PTIWIO9*. J'ai effectué cette répression soit dans le parent 51, soit dans le parent 32, soit dans les deux parents (**Figure V.6**). Tous les F1 sont viables alors que la double répression est létale à l'autogamie (Bouhouche et al., 2011). Il est possible que la méthode ne soit pas aussi efficace en conjugaison car la déplétion se fait par ajout de milieu ARNi la veille du croisement pour déclencher la réactivité. De plus, l'efficacité d'un milieu ARNi peut être variable. En utilisant mon milieu à l'autogamie, j'ai observé seulement un retard de croissance pour deux tiers des individus post-autogames. On peut conclure que la déplétion de Ptiwio1 et Ptiwio9 n'est pas totale dans mes croisements. Néanmoins, elle est suffisante dans le parent 51 pour générer un fort taux de rétention de l'IES dans les F1, indépendamment de leur origine maternelle. Cet effet est visible que le parent 32 soit déplété ou non. En revanche, la répression de la voie dans le seul parent 32 n'affecte pas le taux de rétention par rapport au croisement contrôle. Les protéines Ptiwio1&09 du parent 51 sont donc essentielles à l'excision correcte de l'IES de mtB dans la majorité des F1 hétérozygotes. Cela suggère que les scnARN homologues à l'IES de mtB et produits par le parent 51 sont responsables de l'excision de l'IES dans les F1 dérivant de ce parent mais aussi dans les F1 dérivant du parent naïf. Autrement dit, le F1 dérivant du parent 32 peut utiliser les scnARN paternels (produits par le parent 51) pour exciser l'allèle paternel.

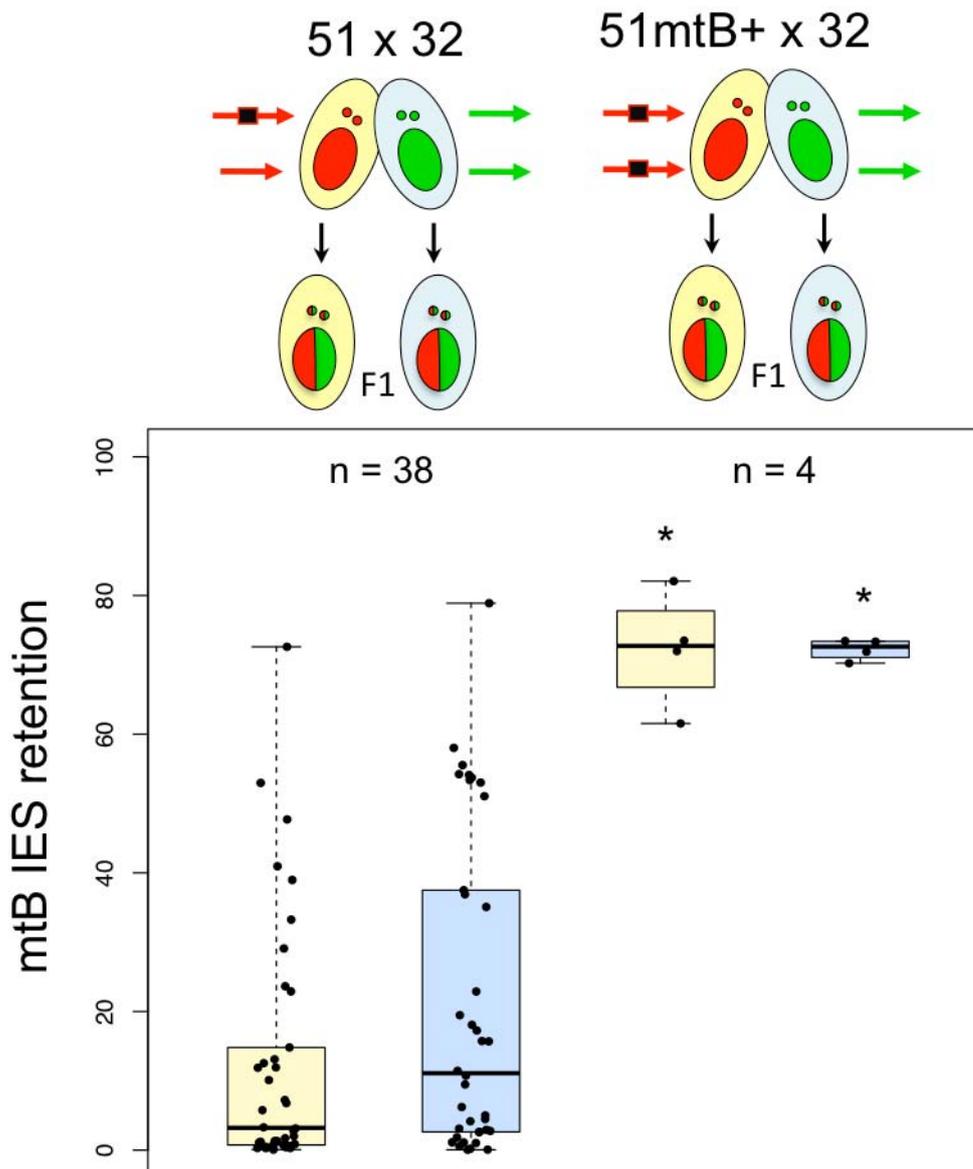


**Figure V.6 : Rétention de l'IES de mtB en F1 en fonction de la répression parentale de *Ptiwi1/9*.**

Les croisements impliquent toujours la souche 51 avec la souche 32 et le contrôle, à gauche, est celui présenté en figure V-2B. Une extinction génique de *PTIW101* et *PTIW109* par ARN interférence a été réalisée dans le parent 51, le parent 32 ou les deux. Elle est matérialisée par le motif à points. Le taux de rétention calculé est présenté par ex-conjuguat sous la forme de boxplots avec affichage de tous les points. Les astérisques indiquent une différence significative (p-value < 0,001, Kolmogorov-Smirnov) par rapport au contrôle.

## 2) Croisement avec la souche mtB+

Afin de confirmer le rôle des scnARN paternels dans l'excision de l'IES de mtB, j'ai utilisé la souche épimutante 51 mtB+ (voir V-B) dans laquelle l'IES de mtB est présente dans le MAC. De cette façon, les scnARN homologues à cette IES ne sont pas sélectionnés pour diriger l'excision de l'IES. Toute la descendance F1 du croisement de 51mtB+ par 32 présente un fort taux de rétention (**Figure V.7**). Une fois de plus, cela montre que les 70 % des F1 capables d'exciser l'IES de mtB dans le croisement contrôle utilisent normalement les scnARN actifs du parent 51 pour y parvenir.



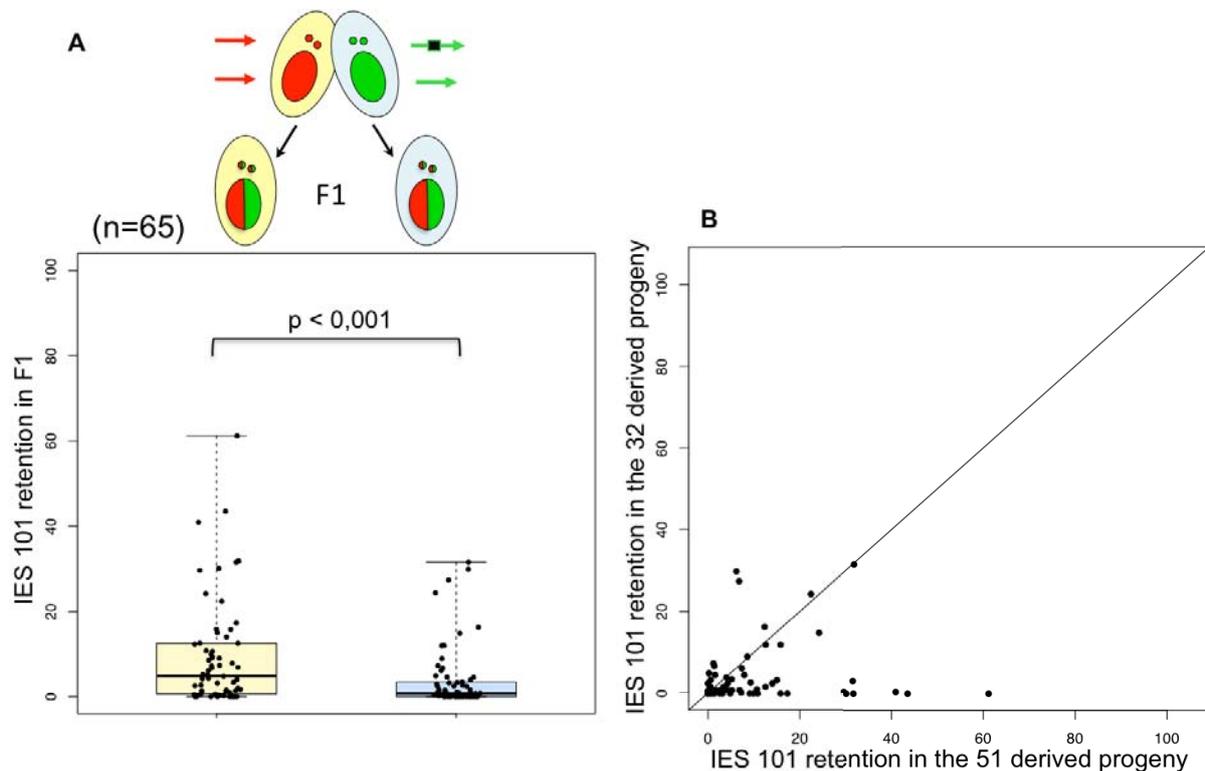
**Figure V.7 : Rétention de l'IES de mtB en F1 en fonction de son statut épigénétique dans le parent 51.**

Une souche épimutante retenant l'IES de mtB dans son MAC est croisée avec la souche 32 (à droite). Le contrôle, à gauche, est celui présenté en figure V-2B. Le taux de rétention calculé est présenté par ex-conjuguat sous la forme de boxplots avec affichage de tous les points. Les astérisques indiquent une différence significative ( $p$ -value < 0,001, Kolmogorov-Smirnov) par rapport au contrôle.

## E. L'IES 101 est davantage retenu chez le F1 dérivant du parent naïf

L'IES 101, longue de 99 pb, est présente dans la souche 32 mais dans la souche 51. De fait, elle est pratiquement symétrique à l'IES de mtB (134 pb) présente dans la souche 51 mais pas dans la souche 32. L'étude de cette IES nous donne un bon élément de comparaison pour extrapoler les résultats précédents. En revanche, le contrôle

maternel de l'IES 101 n'a pas été établi par injection de l'IES dans le MAC et il n'existe pas de jeu de données Dcl2/3 pour la souche 32 qui montrerait au moins la dépendance à la voie des scnARN. Seule sa taille peut suggérer cette dépendance puisque les IES les plus longues sont davantage sensibles à la répression de la voie des scnARN. A partir des échantillons d'ADN du croisement 51 x 32 (V-2C), le laboratoire Potekhin a amplifié la région contenant l'IES afin d'évaluer sa rétention. La méthode utilisée pour quantifier les gels est la même que pour l'IES de mtB. Les résultats montrent un taux de rétention variable mais généralement plus faible que pour l'IES de mtB (**Figure V.8**). Un certain pourcentage des F1 présente un taux de rétention supérieur à 20 %, pouvant potentiellement induire une rétention plus forte à la génération suivante si le contrôle maternel est important. De plus, ces taux élevés sont davantage retrouvés chez la descendance issue du parent 51 naïf, ce qui va dans le sens de notre hypothèse initiale d'un défaut de scnARN maternels mais contredit les observations pour l'IES de mtB. Enfin, il n'y a pas de corrélation entre la rétention d'un côté de la descendance et la rétention de l'autre côté. Il n'y a pas non plus de corrélation entre la rétention de l'IES 101 avec la rétention de l'IES de mtB.



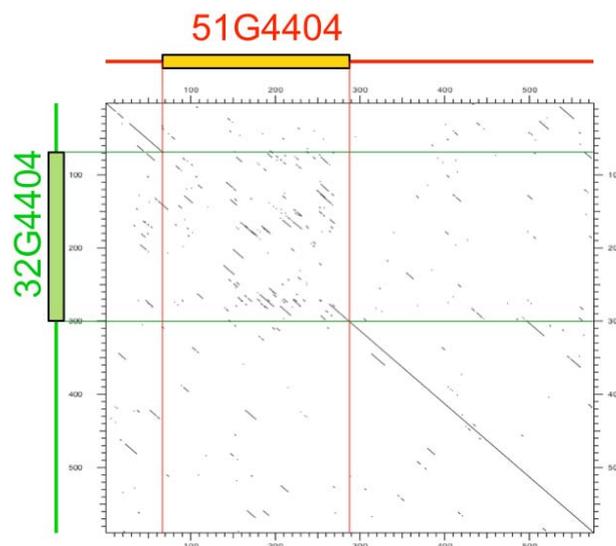
**Figure V.8 : Rétention de l'IES 101 dans les F1 hétérozygotes 51/32.**

A partir des échantillons d'ADN du croisement 51 x 32 (Figure V.2C), une PCR utilisant des amorces encadrant la région de l'IES 101 a été réalisée. Le calcul du taux de rétention est effectué de la même manière que pour l'IES de mtB (Figure V.2A). Dans le cas présent, seul l'allèle 32 (vert) possède l'IES dans sa version MIC et peut donc la retenir dans sa version MAC alors que l'allèle 51 (rouge) n'existe que sous la forme IES-. Le taux de rétention calculé est présenté par ex-conjugué sous la forme de boxplots avec affichage de tous les points (A) et par couple sous la forme de graphique 2D (B). 'p' indique la p-value issue du test de comparaison de la rétention de l'IES entre les deux côtés du croisement (Test des signes de Wilcoxon).

# VI – Effet des divergences d’IES avec l’exemple de G4404

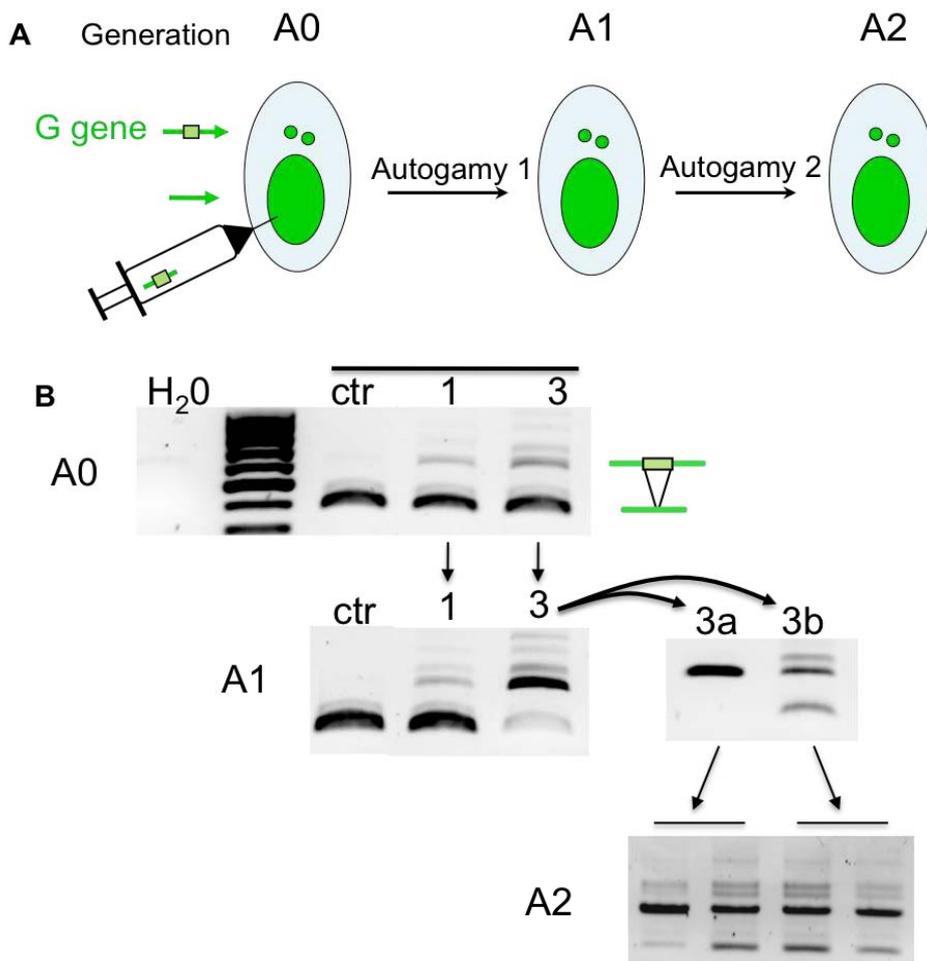
## A. Les IES 51G4404 et 32G4404 sont contrôlées maternellement.

L’IES G4404 fut l’une des premières IES contrôlées maternellement identifiées chez la souche de référence 51 (Duharcourt et al., 1995). Elle est localisée dans la région centrale du gène G codant un antigène de surface. Cette région est intrinsèquement variable, sa version 32 n’était pas présente dans notre génome. Le caractère historique du gène G nous a poussés à en faire séquencer l’allèle 32. Malgré la divergence, les allèles 51 et 32 du gène G peuvent être alignés à l’exception de l’IES G4404 (**Figure VI.1**). Pour autant il s’agit bien d’allèles dérivant d’une IES ancestrale puisqu’ils sont tous les deux insérés au niveau du même TA. Cette situation nous a donné l’opportunité de travailler sur un modèle de dysgenèse hybride liée à la divergence des séquences (cf. IV-B-3). Si le contrôle maternel de l’allèle 51 est bien connu, aucun élément ne peut garantir qu’il existe pour l’allèle 32. Pour le vérifier, j’ai injecté un plasmide linéarisé portant le fragment représenté en figure VI.1 (plasmide p32) dans la souche 32 puis observé la rétention de l’IES après autogamie (**Figure VI.2**). L’analyse du clone injecté n°3 montre que moins d’une copie injectée par génome haploïde (ratio IES+/IES- < 1) suffit à induire une forte rétention de l’IES lors de l’autogamie suivante. Cette rétention, bien que non totale et variable entre individus post-autogames, est ensuite propagée à la génération suivante en l’absence du fragment initial injecté. Finalement, on peut dire que l’IES 32G4404 est bien contrôlée maternellement. En revanche, il est possible que le degré de sensibilité à ce contrôle soit différent pour les deux allèles.



**Figure VI.1 : Matrice d’alignement des allèles 51 et 32 d’une portion du gène G.**

L’alignement des allèles d’une fraction du gène G contenant l’IES G4404 est présenté sous la forme d’un dot plot. Un point est dessiné à chaque fois qu’une fenêtre glissante de 21 nt présente au moins 13 nt identiques entre les deux allèles. L’alignement utilisé est fourni en annexe (**Figure S4**)



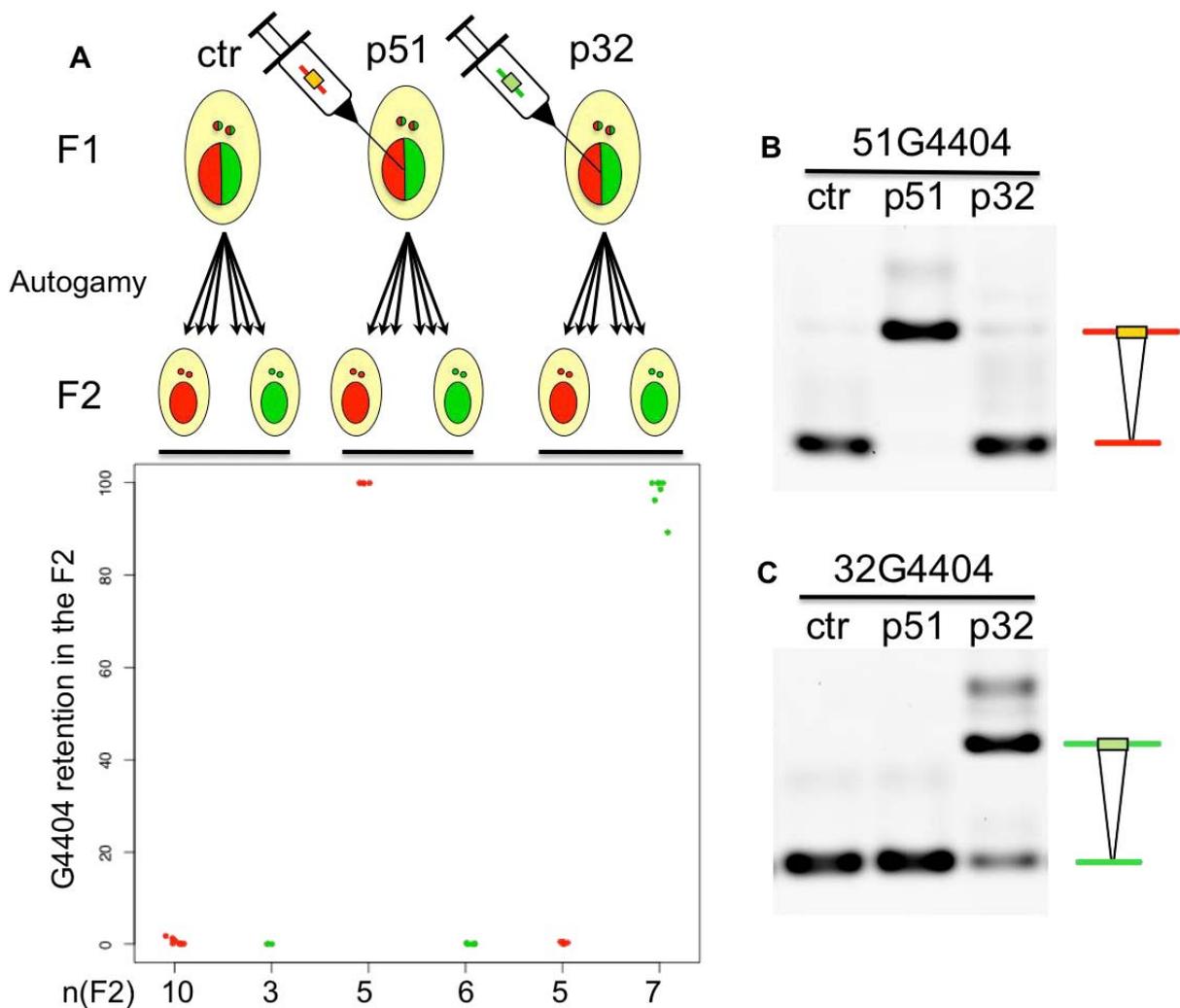
**Figure VI.2 : Contrôle maternel de l'allèle 32G4404**

(A) Chez la souche 32, un fragment du gène G contenant l'IES G4404 est injecté dans le macronoyau (génération A0). Le devenir de l'IES est suivi dans deux clones injectés 1 et 3 ainsi que dans un contrôle non injecté. Après autogamie, 20 cellules sont cultivées ensemble et forment la population A1 dont l'ADN est extrait. Parmi la population post-autogame dérivant du clone 3, deux individus 3a et 3b sont cultivés jusqu'à l'autogamie suivante. Deux duplicats de 20 cellules post-autogames représentent la population A2. (B) Une PCR utilisant des amorces encadrant l'IES permet de révéler un ratio entre la forme IES+ et la forme IES-. Dans la génération A0, la forme IES+ représente le fragment injecté alors que la forme IES- représente le MAC endogène. Dans les générations A1 et A2, les deux formes proviennent du nouveau MAC car le fragment injecté a été éliminé au cours de la première autogamie.

## B. les scnARN ciblant l'IES G4404 sont spécifiques à chaque allèle.

Le modèle de dysgenèse hybride par divergence de séquence suppose que les scnARN ciblant une IES chez une espèce ne puissent pas cibler l'orthologue de l'autre espèce avec laquelle elle est croisée. Dans notre étude intraspécifique, j'ai cherché à montrer que les scnARN ciblant l'allèle 51G4404 sont incapables de reconnaître l'allèle 32G4404, et *vice versa*. Pour cela, j'ai utilisé un clone hétérozygote 51/32 dans lequel les deux allèles sont parfaitement excisés. J'ai injecté un fragment génomique contenant l'IES 51G4404 ou 32G4404, portés respectivement dans les plasmides p51

et p32, dans le MAC de ce clone puis j'ai observé la descendance post-autogame (**Figure VI.3**). Les F2 homozygotes générés par le clone non injecté excisent totalement l'IES G4404, qu'ils héritent de l'allèle 51 ou de l'allèle 32. En revanche, les F2 héritant de l'allèle 51 du parent hétérozygote injecté avec le plasmide 51 ne sont pas capables d'exciser cet allèle, alors que ceux héritant de l'allèle 32 de ce même parent l'excisent totalement. Ainsi, le blocage maternel de l'excision de l'allèle 51 n'affecte pas l'allèle 32, ce qui signifie que les scnARN ciblant l'IES 32G4404 ne peuvent pas être inactivés par appariement avec la séquence 51G4404 injectée dans le MAC parental. Ces scnARN sont alors capables de diriger l'excision de l'allèle 32 dans le MAC F2 mais pas de l'allèle 51 (totale rétention de 51G4404). Ce constat est réciproque puisque qu'on observe un résultat symétrique après l'injection du plasmide p32 dans le clone hétérozygote.



**Figure VI.3 : Spécificité allélique du contrôle maternel de l'IES G4404**

Un hétérozygote F1 51/32, dont les deux allèles macronucléaires de G4404 sont parfaitement excisés, est injecté ou non (ctr) avec le plasmide p51 ou p32. Ces plasmides portent l'IES G4404 de la souche 51 et 32 respectivement. Le niveau d'injection peut être estimé par une PCR amplifiant spécifiquement la région encadrant l'IES : la forme IES+ provient du plasmide alors que la forme IES- provient du MAC endogène (B et C). Après autogamie, l'ADN de 11 à 13 F2 a été extrait pour chaque condition et le niveau de rétention de l'IES de l'allèle hérité est calculé. Le nombre de F2 testés est indiqué en bas du graphique.

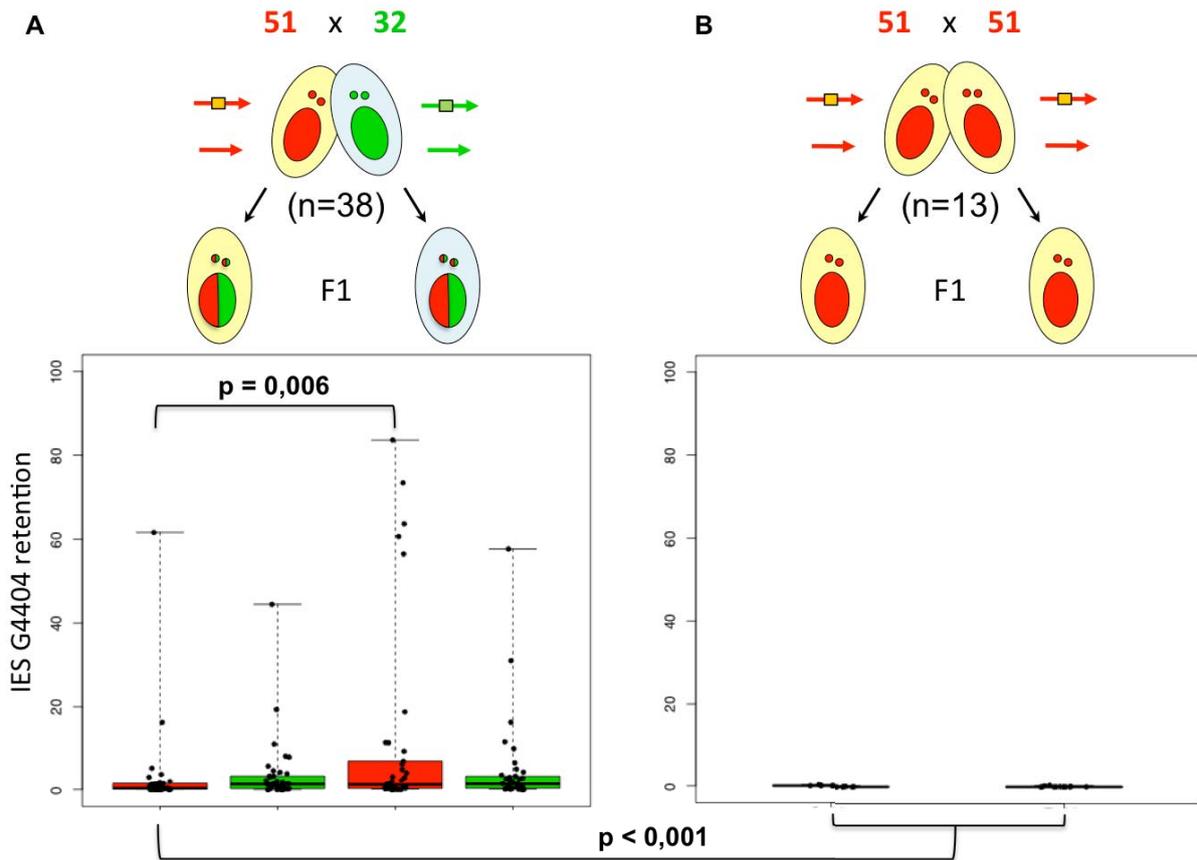
## C. les allèles divergents de G4404 sont convenablement excisés chez les F1 51 x 32.

Puisque nous avons vérifié que les deux allèles de l'IES G4404 sont conformes à notre situation théorique initiale (divergence et contrôle maternel), nous pouvons analyser le résultat du croisement 51 x 32 (**Figure VI.4**). Contrairement à notre prédiction, l'allèle 51 est bien excisé dans les F1 dérivant du parent 32 et l'allèle 32 est bien excisé dans les F1 dérivant du parent 51. Ainsi le défaut de scnARN maternels homologues ne semble pas être un problème pour l'excision de ces allèles. Toutefois, il faut apporter des nuances à cette conclusion. Tout d'abord, cette excision n'est pas complète comme en témoigne le croisement contrôle 51 x 51. De nombreux F1 présentent une rétention inférieure à 10 % mais non égale à 0 tandis que quelques rares cas retiennent l'IES au-delà de 40 %. De plus, ces anomalies de rétention concernent les deux allèles et ne dépendent pas de l'origine maternelle du F1. Il s'agit donc d'un phénomène qui ne s'explique pas par notre modèle. Seul le fait qu'il y ait davantage de cas de forte rétention de l'allèle 51G4404 parmi les F1 dérivant du parent 32 que du parent 51 s'inscrit dans notre modèle.

## D. Les scnARN paternels assurent l'excision quasi parfaite des IES.

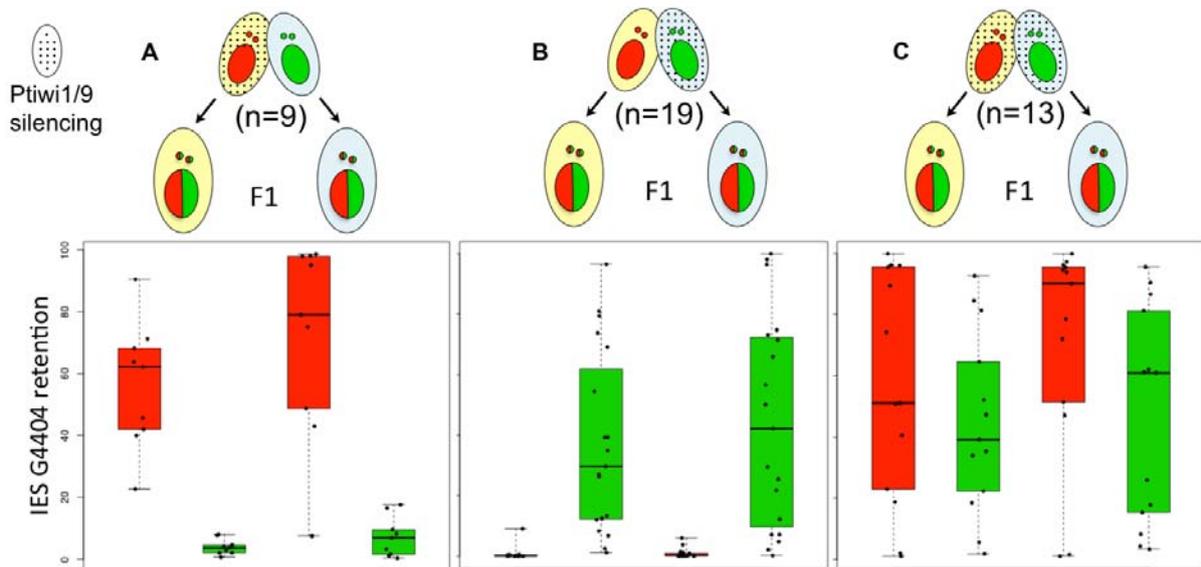
### 1) Croisement avec déplétion de *Ptiwio1&09*

Pour comprendre comment les F1 peuvent exciser des allèles pour lesquels leur parent cytoplasmique ne peut produire de scnARN homologues, j'ai utilisé la même méthode que pour l'étude de l'IES de mtB. J'ai réprimé les gènes *PTIWIO1* et *PTIWIO9* soit dans le parent 51, soit dans le parent 32, soit dans les deux parents (**Figure VI.5**). Lorsque l'extinction génique est réalisée sur le parent 51, le F1 dérivant du parent 51 ne parvient pas à exciser correctement l'allèle 51 (rétention à 60% environ) mais il parvient à exciser l'allèle 32. De l'autre côté, le F1 dérivant du parent 32 donne un résultat tout à fait similaire. On peut donc conclure qu'en l'absence de production de scnARN ciblant l'allèle 51, les deux F1 ne peuvent pas l'exciser. En analysant les deux autres croisements, il apparaît clairement que la répression des gènes *PTIWIO1&9* chez un parent entraîne une forte rétention de l'allèle de ce parent dans les deux F1. Cela veut dire que dans un croisement normal, l'excision d'un allèle en F1 est assurée par les scnARN du parent donnant cet allèle, qu'ils soient d'origine maternelle ou paternelle.



**Figure VI.4 : Rétention de l'IES G4404 dans les F1 hétérozygotes 51/32.**

Les flèches rouges et vertes représentent respectivement les allèles 51 et 32 du gène G. Les versions MIC contiennent l'IES G4404 (rectangle orange ou vert) alors que les versions MAC en sont dépourvues. (A) A partir des échantillons d'ADN du croisement 51 x 32 (Figure V.2B), une PCR allèle-spécifique utilisant des amorces encadrant l'IES G4404 permet de calculer indépendamment le taux de rétention pour chaque allèle (51G4404 en rouge et 32G4404 en vert). Les taux de rétention calculés sont présentés sous la forme de boxplots avec affichage de tous les points. Les taux sont comparés 2 à 2 par F1 et par allèle avec un test des signes de Wilcoxon. La seule p-value significative indique que l'allèle 51 est davantage retenu dans les F1 dérivant du parent 32 que dans ceux dérivant du parent 51. (B) Un croisement contrôle 51 x 51 montre que l'IES 51G4404 est complètement excisée. La rétention de l'IES 51G4404 dans ces F1 homozygotes dérivant de parents 51 est comparée à la rétention de cette même IES dans le F1 51/32 avec un test de Kolmogorov-Smirnov.

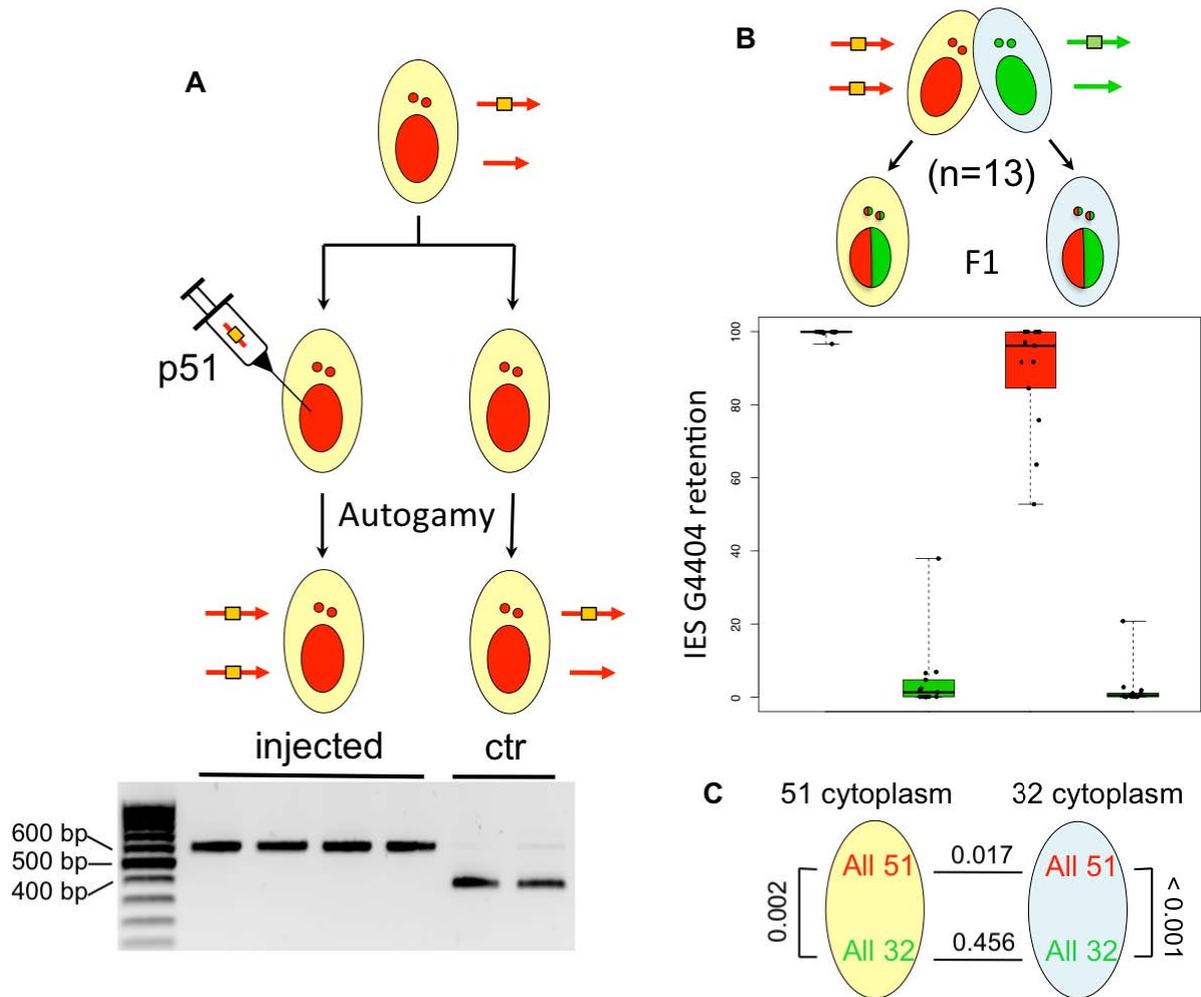


**Figure VI.5 : Rétention de l'IES G4404 en F1 en fonction de la répression parentale de *Ptiwi1/9*.**

Les croisements impliquent toujours la souche 51 avec la souche 32. Il s'agit des mêmes croisements qu'en figure V.6. Une extinction génique de *PTIWI01* et *PTIWI09* par ARN interférence a été réalisée dans le parent 51 (A), le parent 32 (B) ou les deux (C). Elle est matérialisée par le motif à points. Le taux de rétention est toujours calculé en utilisant une PCR encadrant l'IES. Le taux de rétention calculé est présenté par ex-conjugué sous la forme de boxplots avec affichage de tous les points.

## 2) Croisement 51-G4404+ x 32

Pour aller plus loin, j'ai testé l'effet de la titration des scnARN homologues à l'allèle 51G4404 en utilisant une souche 51 épimutante retenant l'IES. Pour la construire, j'ai injecté le plasmide p51 utilisé en VI-B et sélectionné un clone qui retenait totalement l'IES après autogamie (**Figure VI.6 A**). Le croisement de ce clone avec la souche 32 donne des individus F1 qui retiennent totalement l'allèle, indépendamment de leur origine maternelle (**Figure VI.6 B**). L'allèle 32 de l'IES G4404 n'est pas affecté, il est bien excisé dans tous les F1. Le défaut de scnARN MIC-spécifiques ciblant l'allèle 51 est donc la cause de la rétention observée. Dans un croisement sauvage, le F1 dérivant du parent 32 est capable d'exciser l'allèle 51 grâce aux scnARN paternels produits par la cellule 51. Toutefois, on observe ici que les F1 dérivant du parent épimutant semblent retenir davantage l'allèle que les F1 dérivant du parent 32 (100 % contre 50 à 100 % ;  $p = 0,017$ ). Ce résultat suggère que l'action des scnARN est plus efficace dans la cellule qui les a produits que dans la cellule partenaire. Pour émettre des hypothèses, il convient de s'intéresser au mode d'action possible des scnARN, et plus particulièrement à la transmission de l'information qu'ils portent entre les deux partenaires de conjugaison.



**Figure VI.6 : Rétention de l'IES G4404 dans le croisement 51-G4404+ x 32.**

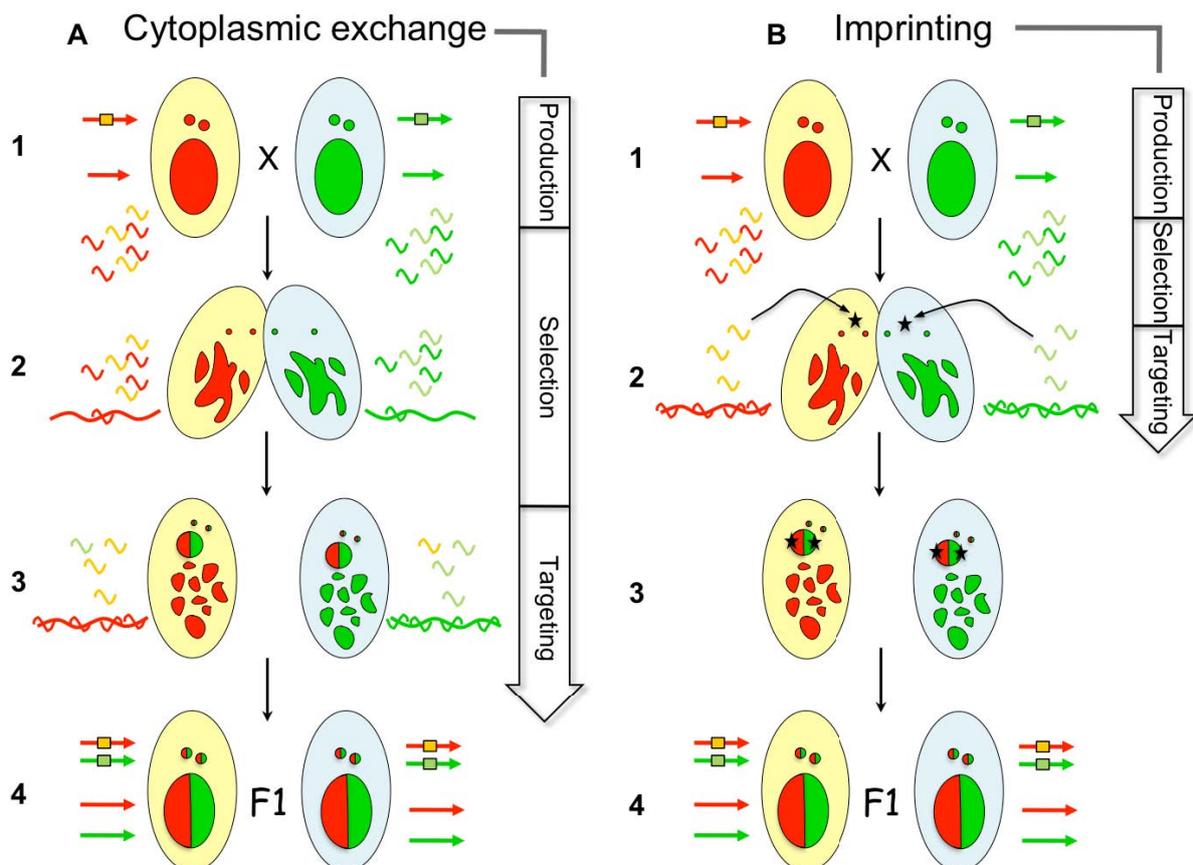
Les flèches rouges et vertes représentent respectivement les allèles 51 et 32 du gène G. Les versions MIC contiennent l'IES G4404 (rectangle orange ou vert) alors que les versions MAC en sont normalement dépourvues. (A) Le plasmide p51 (cf. VI-B) est injecté dans une cellule 51 afin d'induire la rétention macronucléaire de l'allèle 51G4404. Après autogamie, le plasmide est perdu et une PCR encadrant l'IES permet d'évaluer sa rétention dans le nouveau macronoyau (Figure VI.3). (B) Un des 4 clones retenant l'IES dans son MAC est croisé avec la souche 32 et la descendance est analysée. Une PCR utilisant des amorces encadrant l'IES G4404 permet de calculer spécifiquement le taux de rétention pour chaque allèle (51G4404 en rouge et 32G4404 en vert). Le taux de rétention calculé est présenté par ex-conjuguat sous la forme de boxplots avec affichage de tous les points. (C) Les médianes des taux de rétention sont comparés 2 à 2 par F1 et par allèle avec un test des signes de Wilcoxon. Les valeurs indiquées correspondent aux p-value.

# VII – Mode d'action des scnARN paternels

## A. Un nouveau modèle

### 1) Deux hypothèses

Que ce soit un polymorphisme d'insertion ou une divergence allélique, j'ai montré que les IES paternelles sont excisées grâce au scnARN paternels. Ce processus requiert les protéines Ptiwi01&09 du père ainsi que l'absence de l'IES dans le macronoyau paternel. L'hypothèse la plus simple consiste à dire que les scnARN peuvent être échangés au cours de la conjugaison et ainsi diriger l'excision des IES non reconnus par des scnARN maternels (**Figure VII.1 A**). Elle est néanmoins en contradiction avec l'hérédité maternelle du type sexuel (cf. IV-B-2) qui nous a conduit au modèle de dysgenèse hybride. Une deuxième hypothèse propose que les scnARN déposent une empreinte sur les IES dans le noyau gamétique, sous la forme de modifications épigénétiques de la chromatine (histones et/ou ADN) (**Figure VII.1 B**). Cette empreinte pourrait alors être lue par la cellule receveuse afin de les exciser. Les expériences qui vont suivre apportent des éléments permettant de discuter ces deux hypothèses qui sont non exclusives.



**Figure VII.1 : Modèles d'excision des allèles divergents par la voie du genome scanning avec deux hypothèses d'action des scnARN paternels.**

L'étape 1 correspond à l'état initial dans lequel une souche (génome rouge et cytoplasme beige) est croisée avec une autre (génome vert et cytoplasme bleu). Les flèches colorées représentent un gène possédant deux allèles divergents (rouge et vert) et interrompu par une

IES (carré orange ou vert clair) dans sa version MIC. Un seul allèle est représenté pour les parents homozygotes alors que deux allèles sont présents dans les F1 hétérozygotes (étape 4). L'étape 1 est immédiatement suivie de la production de scnARN couvrant l'ensemble du MIC, incluant les IES, et dont la couleur montre l'homologie à l'allèle. L'**étape 2** montre les cellules en conjugaison juste avant l'échange réciproque de noyau gamétiques. Chaque conjugant possède un noyau gamétique femelle (résident) et un noyau gamétique mâle qui migrera dans l'autre cellule. A ce stade, les scnARN ont entamé leur processus de titration par homologie de séquence avec de long ARN non codants maternels (traits ondulés rouges ou verts). Cette sélection doit même être terminée dans l'hypothèse de la programmation (B) afin que les scnARN puissent déposer une empreinte (étoile noire) sur les IES dans les noyaux gamétiques. Dans l'étape 3, les ex-conjugants sont séparés et possède de nouveaux MIC et deux *anlagen* (MAC en développement). On considère que les deux *anlagen* subissent le même développement donc un seul est représenté pour simplifier le schéma. Dans l'hypothèse d'échange cytoplasmique (A), la titration est complète et une fraction des scnARN sélectionnés ont pu migrer dans l'autre cellule. L'ensemble des scnARN maternels et paternels peut alors diriger l'excision de toutes les IES dans l'*anlage* hétérozygote. Dans l'autre hypothèse (B), c'est l'empreinte qui permet d'initier l'excision des IES. Enfin, l'**étape 4** est la même dans les deux hypothèses : le nouveau MAC est parfaitement réarrangé.

## 2) Le paradoxe apparent du type sexuel

L'hérédité maternelle du type sexuel nous avait initialement suggéré que les réarrangements étaient indépendants dans les deux partenaires (voir IV-B-2). Dans le nouveau modèle que nous proposons, ils ne le sont plus car les scnARN produits par une cellule peuvent guider les réarrangements des *anlagen* de son partenaire. Dans ce cadre, les scnARN dirigés contre le promoteur de mtA qui sont produits par le parent O devraient donc induire l'excision de la séquence aussi bien dans le F1 dérivant du parent O que dans le F1 dérivant du parent E (au moins pour l'allèle provenant du parent O). Or nous savons que le F1 dérivant du parent E retient totalement le promoteur de mtA sur toutes ses copies, donc sur les deux allèles (Singh et al., 2014). Pour expliquer ce paradoxe, il faut analyser de près cette situation inédite dans laquelle les deux partenaires ont le même allèle MIC mais pas la même version MAC (version excisée ou non). Les scnARN produits par la cellule O portent une information d'excision qui se retrouve contredite dans la cellule E par la présence du promoteur dans le MAC. En conséquence, ces scnARN ne sont pas en mesure d'exciser la séquence homologue dans les F1 dérivant du parent E. On peut expliquer ce résultat dans l'hypothèse de l'échange cytoplasmique en invoquant la titration tardive des scnARN de la cellule O dans le MAC E. Autrement dit, les scnARN homologues au promoteur de mtA sont bien MIC-spécifiques dans la cellule O mais ceux qui migrent dans la cellule E ne le sont plus. Suivant l'hypothèse de l'empreinte génomique, l'allèle paternel du noyau zygotique de la cellule E serait déjà programmé pour être excisé. Pourtant il ne le sera pas, donc il faut imaginer que l'empreinte est réversible. La présence de la séquence dans le MAC de la cellule E protège cet allèle de l'excision. Il pourrait y avoir un rôle protecteur des scnARN MAC.

## B. Expériences d'injection simultanée des 2 allèles G4404

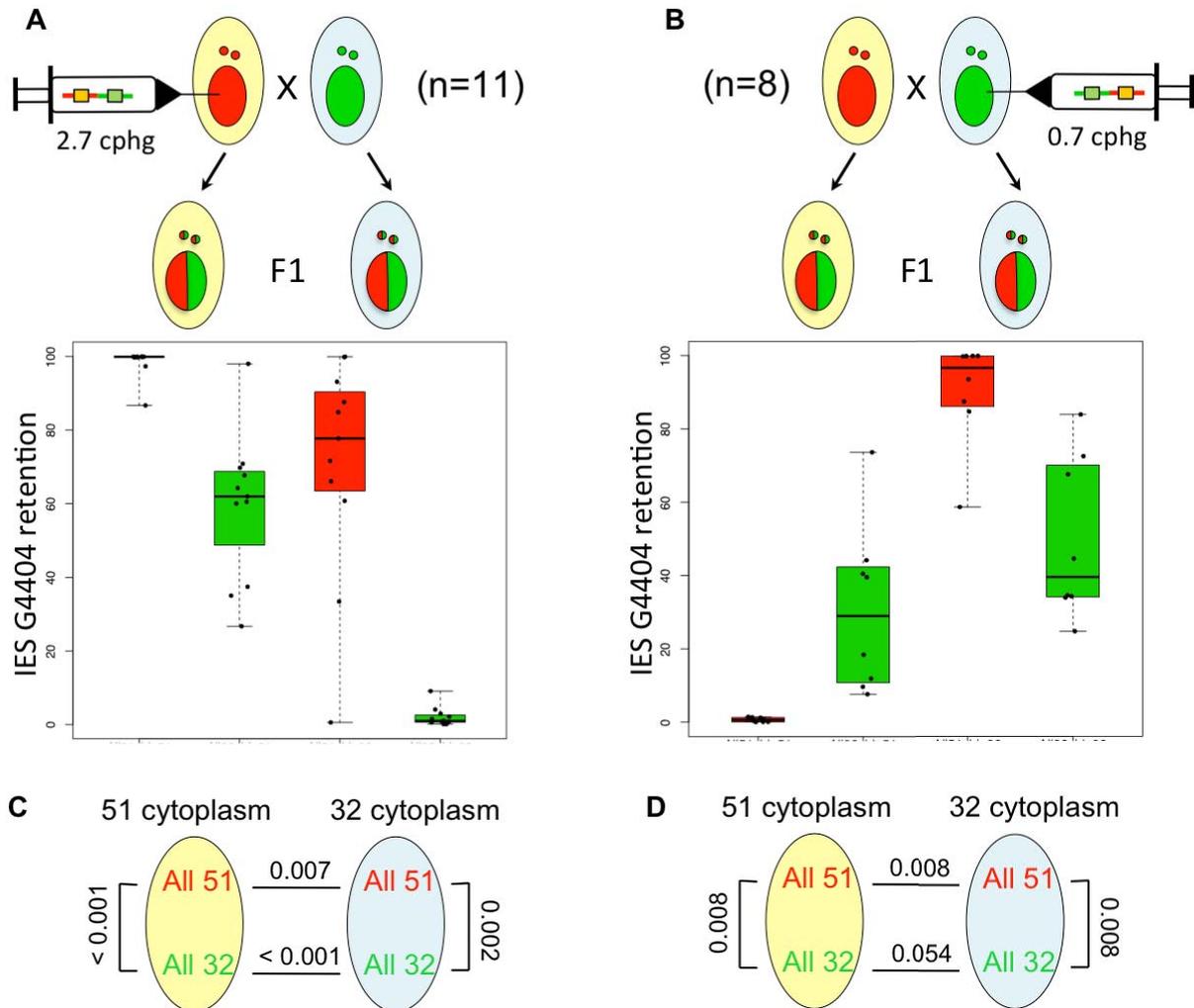
Afin de mettre à l'épreuve notre réinterprétation moléculaire de l'hérédité du type sexuel, nous avons décidé de créer une situation similaire en injectant l'allèle 32G4404 dans la souche parentale 51. La présence de cet allèle dans le MAC de la cellule 51 devrait empêcher l'excision de la séquence homologue qui sera reçu de la cellule 32. En tant que contrôle interne, le plasmide portant l'allèle 32G4004 contient également l'allèle 51G4404. Cela nous permettra de vérifier qu'il y a bien inhibition de l'excision au moins pour l'allèle 51.

Les résultats de ce croisement montrent tout d'abord que la présence de l'allèle 51G4404 dans le MAC du parent 51 induit la rétention de cet allèle à la génération suivante. Cette rétention est totale dans le F1 dérivant du parent 51 et très forte dans l'autre F1 (80 % en moyenne) (**Figure VII.8 A et C, allèle 51 rouge**). Cette différence pourrait suggérer que les scnARN MAC ne soient pas totalement titrés au moment de la fécondation réciproque. Ainsi, certains scnARN homologues à l'IES 51G4404 porteraient une information d'excision dans les nouveaux MAC de la cellule 32. En revanche, l'opération de titration pourrait continuer dans la cellule 51 au-delà de la fécondation jusqu'à ce que tous les scnARN homologues à l'IES soient titrés, d'où une rétention totale dans la cellule 51. Ce mécanisme s'inscrirait très bien dans le cadre de l'hypothèse d'échange de scnARN entre les partenaires. Dans l'autre hypothèse, il faudrait imaginer qu'une empreinte programmant la rétention soit partiellement déposée par les scnARN dans les noyaux gamétiques. Dans la cellule 32 cette empreinte générerait seulement 80 % de rétention alors que dans la cellule 51 l'empreinte pourrait se renforcer car les scnARN MAC sont toujours présents. Au vu des multiples spéculations nécessaires pour maintenir à flot l'hypothèse de l'empreinte génomique, cette analyse favorise plutôt le modèle d'échange cytoplasmique.

Si l'on s'intéresse maintenant à l'allèle 32G4404 (**Figure VII.8 A et C, allèle 32 vert**), on peut noter qu'il est totalement excisé dans le F1 dérivant du parent 32, conformément à un croisement typique. En revanche il est retenu à hauteur de 60 % dans l'autre F1, ce qui montre que la présence de cet allèle sur la construction injectée dans le MAC possède un effet inhibiteur sur l'excision. Cette rétention est celle qu'on attendait d'après notre réinterprétation de l'hérédité du type sexuel. Pourquoi seulement 60 % ? On peut supposer que le contrôle maternel de l'allèle 32G4404 ne permet pas d'induire une rétention totale dans notre expérience.

Afin de confirmer ces résultats, j'ai croisé les souches 51 et 32 en injectant la même construction dans le parent 32. L'analyse allèle par allèle donne un résultat similaire sur lequel on peut avoir le même raisonnement que dans les paragraphes précédents (**Figure VII.8 B et D**). Les valeurs absolues des taux de rétention sont difficilement comparables entre les deux expériences puisque le niveau d'injection (en copies par génome haploïde) n'est pas le même. Néanmoins on peut déduire une information intéressante en comparant les taux de rétention des deux allèles dans les F1 dérivant du parent injecté. Ces F1 retiennent 51G4404 à 100 % et 32G4404 à 60 % dans le 1er croisement, et ils les retiennent à 95 % et 40 % dans le deuxième. Autrement dit, quel

que soit le parent injecté par la construction portant les deux allèles, l'allèle 51 est toujours davantage retenu. On peut en déduire que l'allèle 51 est plus sensible au contrôle maternel que l'allèle 32. C'est bien le fait que les deux allèles soient présents simultanément en un nombre identique de copies qui permet de tirer cette conclusion.



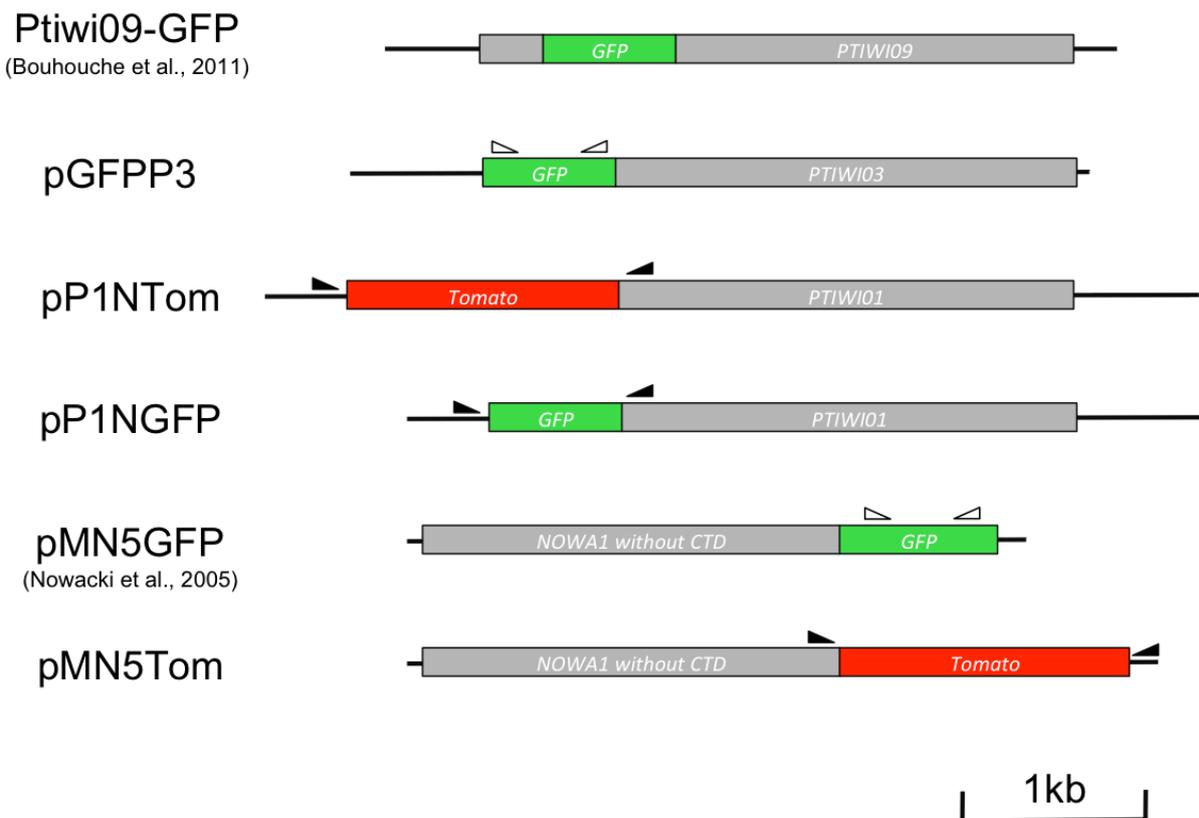
**Figure VII.8 : Croisements 51 x 32 après injection des deux allèles de G4404 dans un parent.**

La souche 51 (A) ou 32 (B) est injectée avec un plasmide contenant les deux allèles de G4404 (à l'intérieur de la seringue). Le nombre de copies du plasmide est estimé en copies par génome haploïde (cphg) à partir d'une PCR. A partir des échantillons d'ADN des F1, une PCR utilisant des amorces encadrant l'IES G4404 permet de calculer spécifiquement le taux de rétention pour chaque allèle (51G4404 en rouge et 32G4404 en vert). Le taux de rétention calculé est présenté par ex-conjuguant sous la forme de boxplots avec affichage de tous les points. (C, D) Les taux de rétention sont comparés 2 à 2 par F1 et par allèle avec un test des signes de Wikcoxon dont les p-value sont indiquées.

## C. Etude de la localisation des protéines Ptiwi

### 1) Etude de la localisation de la protéine Ptiwi03

Puisque les protéines Ptiwi portent les scnARN, l'étude de leurs localisations devrait nous renseigner sur les ARN eux-mêmes. S'ils déposent une empreinte dans les MIC alors on devrait les voir y aller, et s'ils sont échangés au cours de la conjugaison, on devrait également le voir. Lors de l'étude initiale des protéines Ptiwi, un tag GFP avait été inséré après le codon 116 de Ptiwi09 (**Figure VII.2**) et cette construction émet une fluorescence dans le cytoplasme avant de s'accumuler dans le MAC maternel puis dans les nouveaux MAC (Bouhouche et al., 2011). Elle est clairement absente des MIC en méiose I et n'y est pas non plus visible dans les stades suivants.

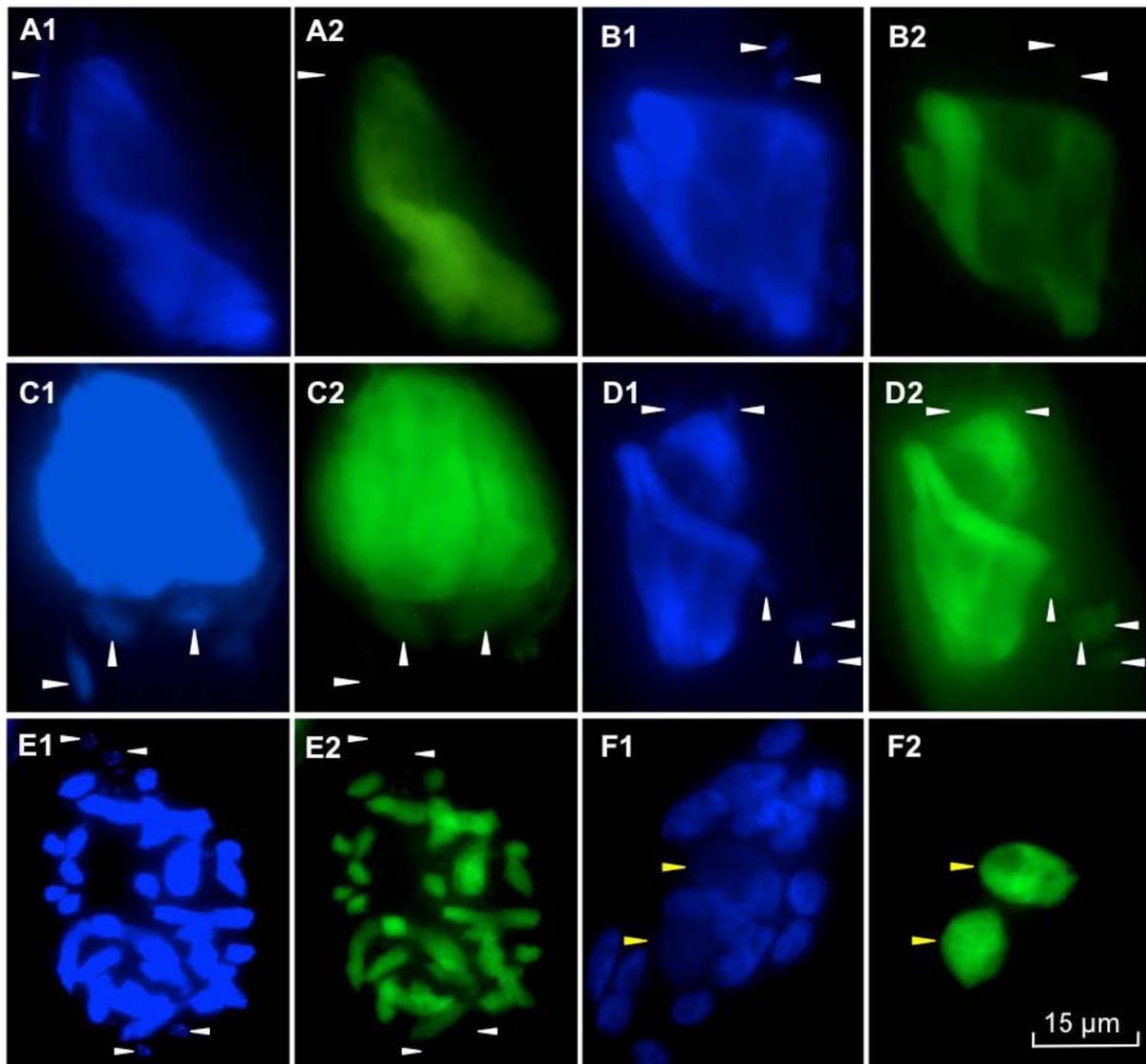


**Figure VII.2 : Schémas des constructions utilisées.**

Chaque schéma représente la séquence, orientée de 5' vers 3', insérée dans le plasmide pUC18. Les noms indiqués sont ceux des plasmides entiers. Les rectangles correspondent aux ORF des gènes de la paramécie (gris), de la GFP (vert) ou de la *tomato* (rouge), ils sont orientés de Nter vers Cter. Les lignes aux extrémités sont les séquences intergéniques, elles contiennent le promoteur en 5' et potentiellement d'autres séquences régulatrices. Les triangles blancs représentent des amorces utilisées pour tester la présence de la construction par PCR. Les triangles noirs représentent des amorces qui amplifient aussi bien la construction que le gène endogène, elles permettent d'estimer le nombre de copies du plasmide par rapport au génome haploïde.

Nous nous sommes donc penchés sur la protéine Ptiwi03, exprimée au même moment, mais dont le rôle ne semble pas essentiel à l'autogamie (Bouhouche et al., 2011). J'ai ajouté une étiquette GFP par trans-PCR en position N-terminale de cette protéine afin de l'étudier. Les clichés de microscopie (**Figure VII.3**) montrent une

localisation macronucléaire identique à Ptiwi09. En revanche, on ne voit pas de fluorescence dans le cytoplasme au début de l'autogamie ni dans les corps basaux. On détecte un signal GFP dans et/ou autour des micronoyaux sous la forme d'une zone de diamètre plus large que celle révélée par le DAPI aux stades 4 et 8 MIC. Il semble donc que Ptiwi03 n'aie pas exactement la même localisation que Ptiwi09. En particulier, elle pourrait emmener des scnARN dans les MIC avant la fécondation. Les stades « noyaux gamétiques » et « noyau zygotique » sont très difficiles à identifier car ils sont rares (événements rapides) et les MIC sont peu visibles de manière générale. Quoiqu'il en soit, après les deux divisions du noyau zygotique, on ne voit plus de signal GFP dans les MIC.



**Figure VII.3 : Dynamique de localisation d'une construction Ptiwi03-GFP au cours d'une autogamie.**

Les photos numérotées 1 et 2 montrent respectivement le marquage au DAPI et la fluorescence GFP. Les lettres correspondent à différents stades de l'autogamie. (A) croissant de méiose I. (B) Méiose I. (C) 4 MIC (3 visibles). (D) 8 MIC (6 visibles). (E) 4 copies du noyau zygotique. (F) *Anlagen*.

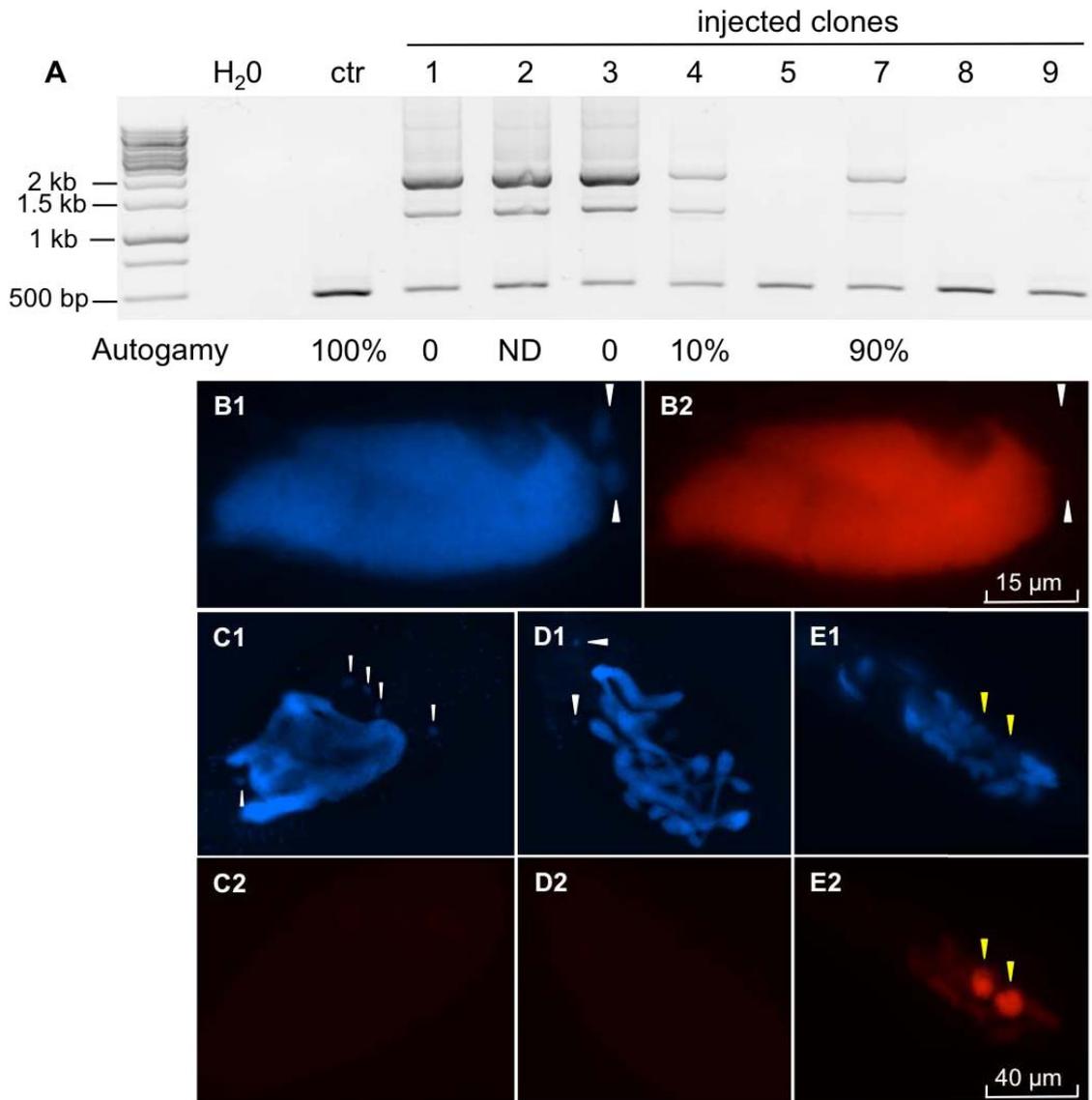
Bien que la présence de Ptiwi03 dans les MIC soit fugace, nous avons décidé de tester l'effet de l'extinction de ce gène dans le croisement 51 x 32 (**Figures S5**). Le résultat pour l'IES de mtB est peu différent du croisement contrôle et celui pour l'IES G4404 n'a rien montré dans les quelques échantillons testés (non montré). Nous n'avons donc pas poursuivi l'étude de cette protéine.

## 2) Etude de la localisation de la protéine Ptiwi01

Toujours à la recherche d'une protéine Piwi capable d'aller dans les MIC, nous avons décidé de construire Ptiwi01-*Tomato*. Ptiwi01 est l'orthologue WGD1 de Ptiwi09 avec une identité de séquence protéique de 99 %. Si l'on peut logiquement supposer qu'elle ait un rôle identique à Ptiwi09, il nous paraît important de le vérifier. De plus, cela nous permet de comparer son profil avec celui de Ptiwi03-GFP, notamment au niveau des MIC. J'ai donc ajouté l'étiquette *tomato*, un dimère de RFP, en position N-terminale par trans-PCR. L'injection de cette construction a montré deux phénotypes bien distincts (**Figure VII.4**). Les clones les plus injectés se sont révélés incapables de réaliser une autogamie normale. La privation de nourriture entraîne un état dans lequel les MIC sont ronds et volumineux, comme en tout début de méiose I. Ils ne sont pas fluorescents, seul le MAC semble héberger la construction. Cet état persiste et ne conduit pas à la mort puisque les paramécies reprennent un cycle végétatif normal si on les nourrit. Les clones les moins injectés réalisent une autogamie qui semble normale. En revanche, le profil d'expression de Ptiwi01-*tomato* est inattendu puisque la fluorescence rouge n'apparaît que tardivement lorsque les *anlagen* sont visibles. Ces derniers concentrent la majorité de la fluorescence, les fragments de l'ancien MAC en émettent un peu et on ne peut rien dire sur les MIC qui ne sont pas visibles, même en DAPI. Le problème d'autogamie rencontré pour les clones les plus injectés ainsi que le profil très différent obtenu par rapport à Ptiwi09-GFP nous a conduit à remettre en cause la construction.

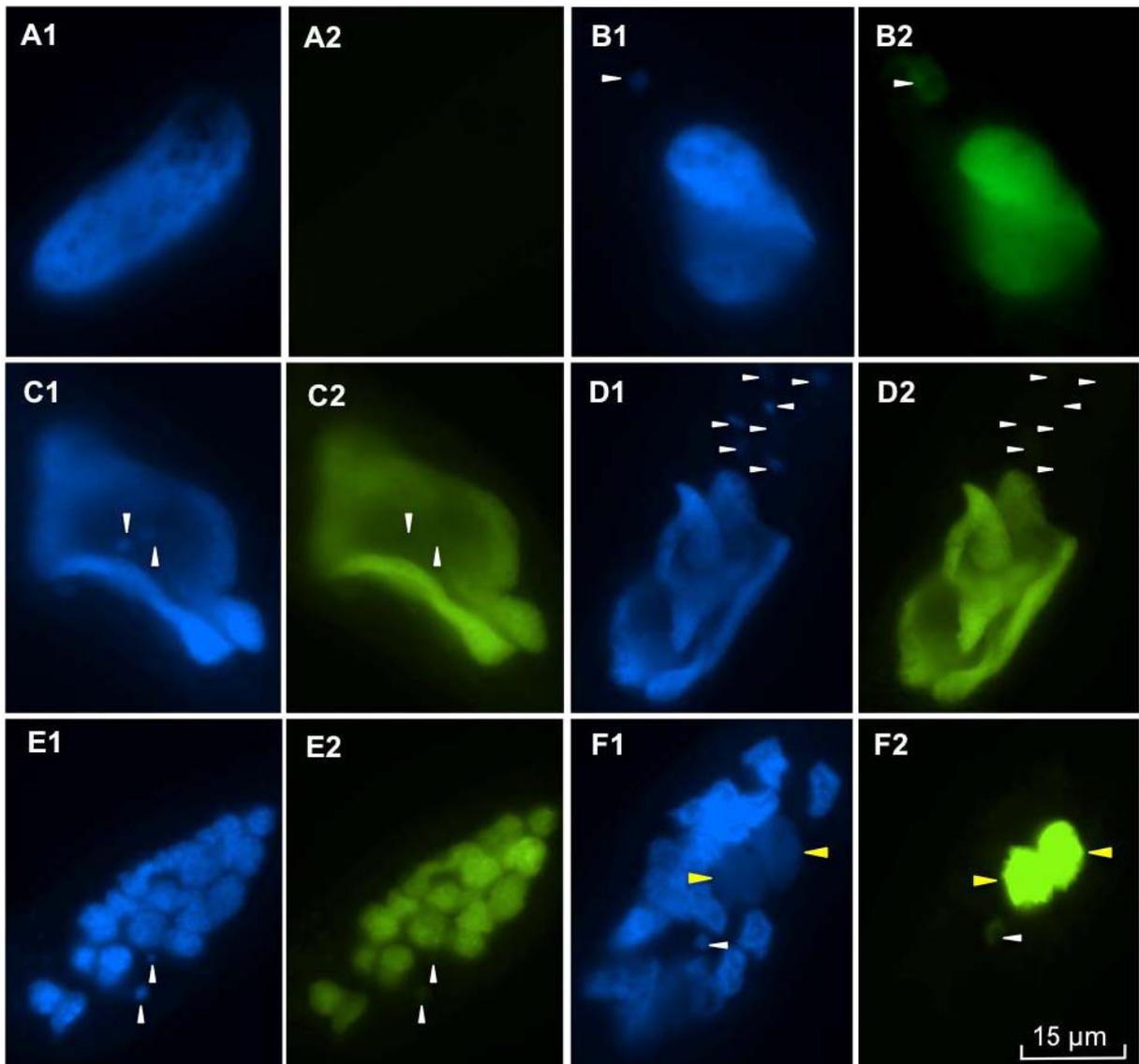
Nous avons alors décidé de remplacer l'étiquette *tomato* par une étiquette GFP pour voir si on observait les mêmes effets. Cette substitution s'est faite par digestion puis clonage puisque nous avons pris soin de placer des sites de restrictions BamHI de part et d'autre de l'étiquette *tomato*. Au début de l'autogamie, la fluorescence apparaît dans le MAC et ne le quitte que pour aller dans les nouveaux MAC lorsque ceux-ci sont bien développés (**Figure VII.5**). Ce profil macronucléaire est donc semblable à celui décrit pour Ptiwi09-GFP, et très différent de celui vu précédemment avec la construction pP1NTom. Dans cette dernière, l'étiquette *tomato*, deux fois plus grosse que la GFP, devait altérer la localisation naturelle de la protéine Ptiwi01. Curieusement, on retrouve ici un signal GFP dans les MIC à deux stades distincts de l'autogamie : le début de méiose I puis plus tardivement dans les nouveaux MIC. Le premier stade, caractérisé par des MIC volumineux et arrondis, montre une fluorescence verte périphérique intense. La structure ainsi dévoilée rappelle la caryolympe décrite dans les noyaux vésiculaires des *P. aurelia* (Fokin, 1997). Dans ces noyaux, la chromatine est concentrée au milieu et serait révélée ici au DAPI, une couche de caryolympe sépare l'ADN de l'enveloppe nucléaire. Cette comparaison reste hypothétique puisque nous n'avons pas poussé l'analyse morphologique et que le stade observé possède potentiellement une architecture

différente du MIC végétatif. La GFP semble ensuite disparaître des MIC avant de revenir marquer les nouveaux MIC en fin d'autogamie, une fois de plus en périphérie du marquage DAPI. Cette observation est totalement nouvelle et pourrait signifier que cette protéine Piwi a un rôle à jouer après que les réarrangements soient finis dans le MAC. On peut imaginer qu'il existe un processus d'empreinte préparant les nouveaux MIC aux réarrangements qui auront lieu à la génération suivante.



**Figure VII.4 : Injection d'une construction *Ptiwio1-tomato* et observation des cellules au cours d'une autogamie.**

(A) Une PCR sur cellules vivantes encadrant le site d'insertion de l'étiquette *tomato* (voir Figure VII.1) permet d'amplifier le gène endogène (484 pb) ou la construction (1918 pb). Une bande intermédiaire est visible et correspond probablement à un produit de trans-PCR puisque l'étiquette *tomato* est un dimère. 'H<sub>2</sub>O' est un contrôle négatif de la PCR et 'ctr' désigne une cellule non injectée. Le pourcentage de cellules capable d'atteindre un stade d'autogamie avec fragmentation du MAC est indiqué dessous du gel. (B à E). Les photos numérotées 1 et 2 montrent respectivement le marquage au DAPI et la fluorescence rouge. (B) Stade unique observé pour les clones 1 à 3. (C à E) Différents stades d'autogamie observés pour le clone 7: '5 MIC' (C), skein (D) et *Anlagen* (E). Les flèches blanches et jaunes indiquent les MIC et les *anlagen*.

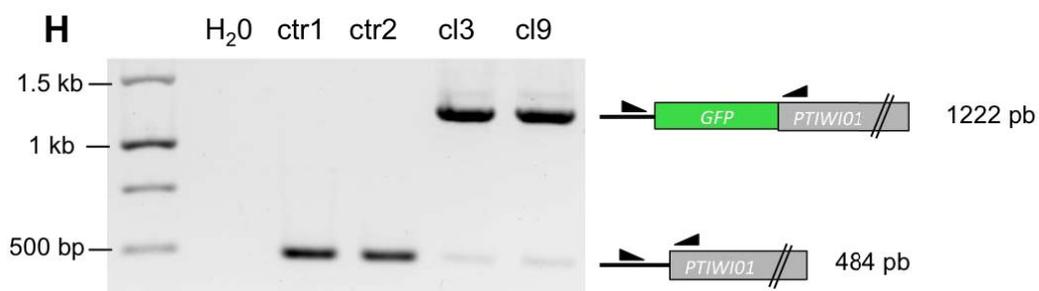
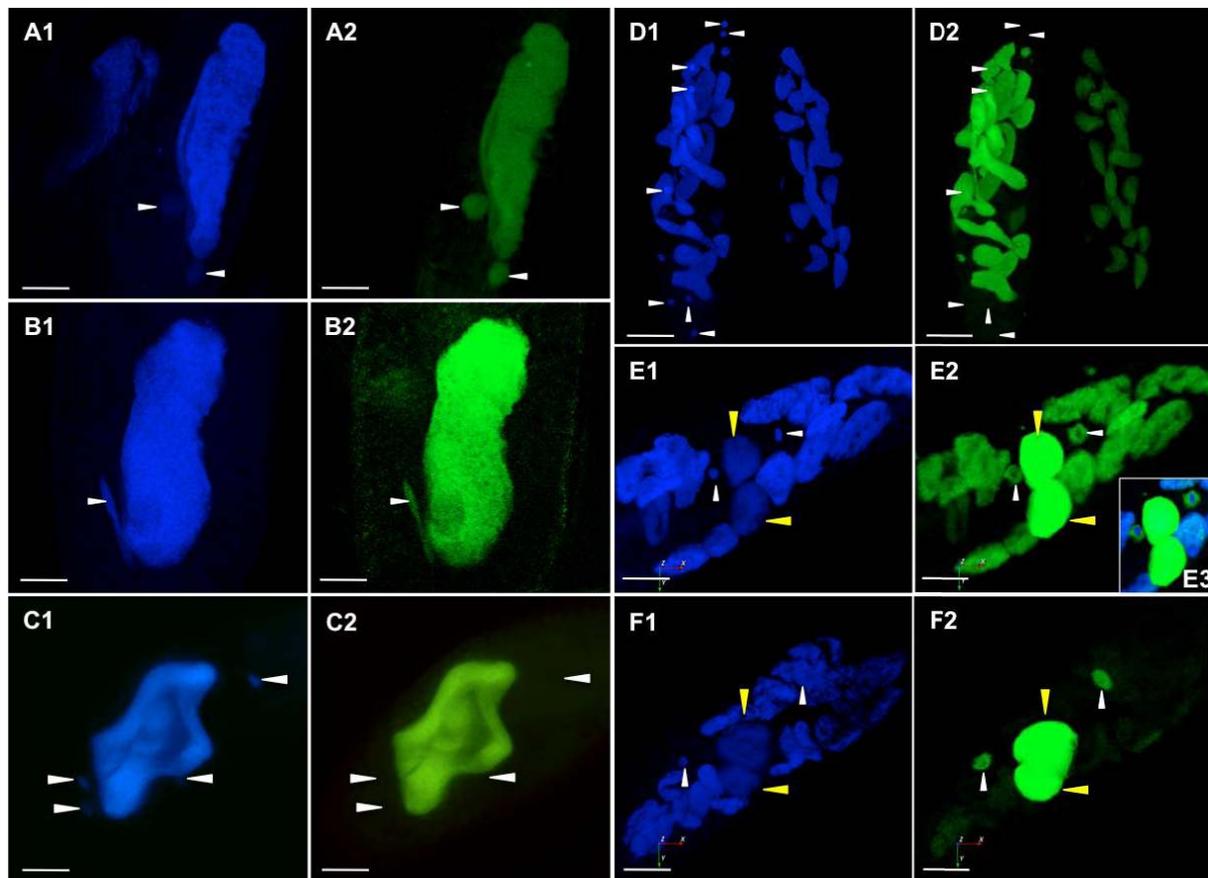


**Figure VII.5 : Observation au microscope en épifluorescence de la dynamique de localisation d'une construction *Ptiwi1-GFP* au cours d'une autogamie.**

Les photos numérotées 1 et 2 montrent respectivement le marquage au DAPI et la fluorescence GFP. Les lettres correspondent à différents stades du cycle sexuel. (A) Stade végétatif. (B) Début de méiose I. (C) Méiose I. (D) 8 MIC (7 visibles). (E) Fragmentation du MAC maternel. (F) *Anlagen*. Les flèches blanches et jaunes indiquent les MIC et les *anlagen*.

Nous avons voulu confirmer les deux observations de fluorescence dans les MIC en répétant l'injection de la construction pP1NGFP. Lors de cette deuxième expérience, le clone injecté choisi a été croisé avec une cellule sauvage afin d'observer la localisation subcellulaire en contexte de conjugaison. Par ailleurs, nous avons eu recours à la microscopie confocale afin d'améliorer la résolution des structures mises en jeu (**Figure VII.6**). Ainsi, les observations précédentes sont confirmées : présence de GFP en début de méiose et dans les nouveaux MIC. Toutefois, il y a des nuances qui méritent d'être détaillées. Tout d'abord, on peut préciser que le signal apparaît en début de méiose et s'attarde au stade 'croissants' avant de disparaître à la fin de la méiose I (4 MIC formés). Il est difficile de prédire si le signal GFP réapparaît très vite dans les nouveaux MIC mais il est bien présent quand les *anlagen* sont

visibles, autour du signal DAPI (**Figure VII.6 E3**). Contrairement à l'expérience précédente, le signal GFP en début de méiose se superpose au signal DAPI plutôt que de l'entourer. Au final, nous ne pouvons pas trancher sur la localisation précise de la protéine à ce stade. Au-delà de l'information de localisation, cette étude au cours de la conjugaison nous apporte un élément capital : la protéine de fusion Ptiwi01-GFP produite par la cellule injectée est capable d'entrer dans le partenaire (**Figure VII.6 D2**). En effet, en l'absence d'une telle translocation au sein d'un couple de cellules, on devrait toujours voir une cellule GFP+ (injectée) et une cellule GFP- (non injectée). C'est bien le cas au début de la conjugaison mais à partir du stade 8 MIC, soit environ 5 h après le mélange des cellules de types sexuels opposés, les deux cellules sont 'vertes'. L'une des deux cellules, probablement dérivant du clone injecté, possède une fluorescence supérieure. Bien qu'on ne puisse pas conclure sur l'échange des protéines Piwi (et de leur scnARN) dans un contexte naturel, c'est-à-dire en l'absence d'étiquette GFP et de surexpression, cela montre que cet échange est physiquement possible.

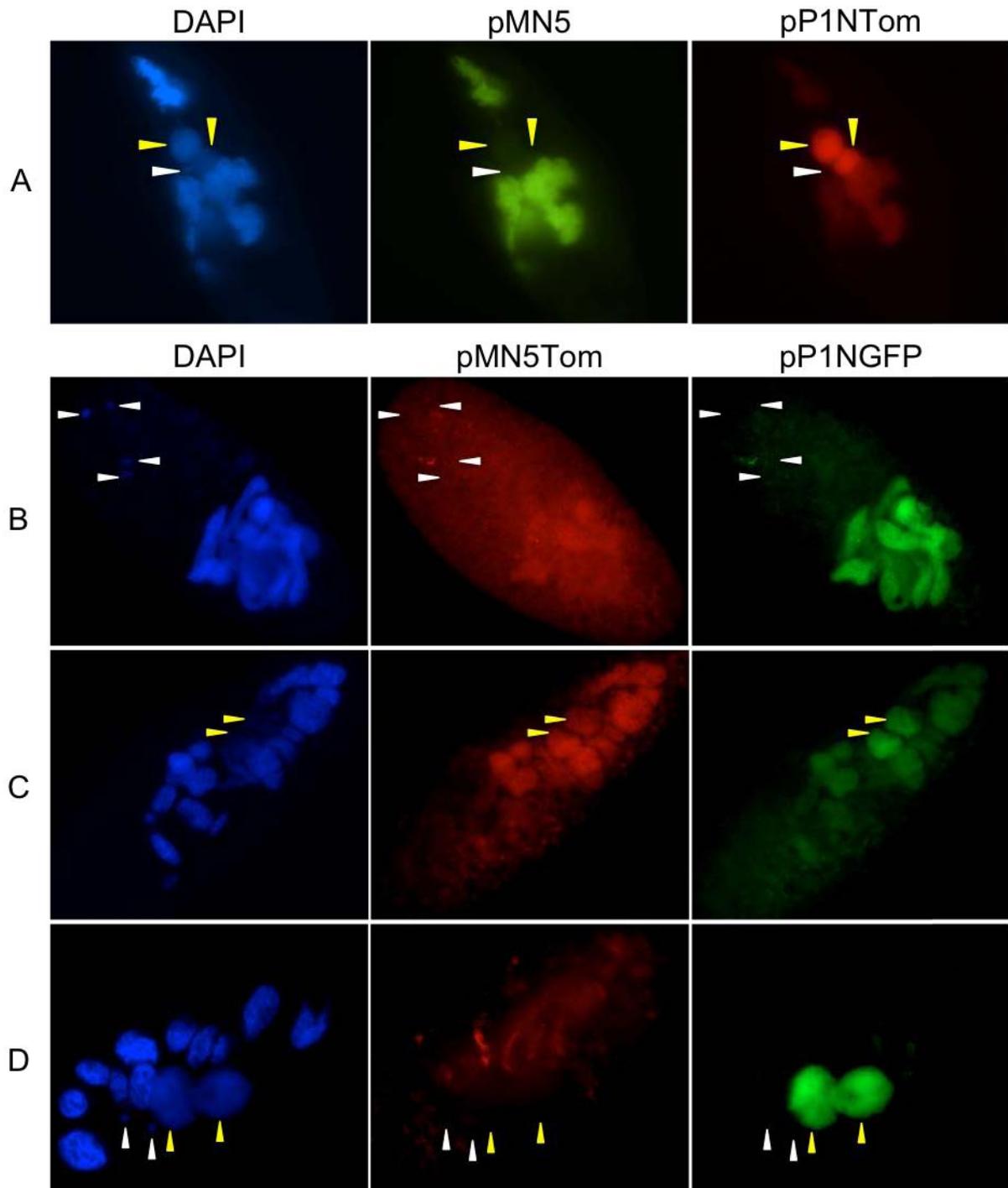


**Figure VII.6 : Observation au microscope confocal de la dynamique de localisation d'une construction *Ptiwi01-GFP* au cours d'une conjugaison.**

Les photos numérotées 1 et 2 montrent respectivement le marquage au DAPI et la fluorescence GFP. La barre représente 20  $\mu\text{m}$ . Les lettres correspondent à différents stades de conjugaison. (A) Début de méiose I. (B) Croissants de méiose I. (C) 4 MIC. (D) 8 MIC. (E et F) *Anlagen*. (E3) est un zoom sur les nouveaux MIC et MAC avec superposition des signaux GFP et DAPI. Les flèches blanches et jaunes indiquent les MIC et les *anlagen*. (H) Une PCR utilisant des amorces (triangles noirs) encadrant le site d'insertion de l'étiquette *GFP* permet d'amplifier le gène endogène (484 pb) ou la construction (1222 pb). 'H2O' est un contrôle négatif de la PCR et 'ctr1 et ctr2' désignent des cellules non injectées. Les photos B sont réalisées sur le clone 9 (cl9) et toutes les autres proviennent du clone 3 (cl3).

### 3) Blocage de *Ptiwi01* par *Nowa1*

Puisque nous avons vu que les scnARN produits par une cellule peuvent transmettre une information à la cellule partenaire, nous avons essayé de bloquer physiquement cette transmission. Ce blocage permettrait alors d'étudier l'importance du transfert dans notre croisement 51 x 32. Pour y parvenir, nous avons utilisé la construction pMN5, dérivée de la protéine *Nowa1*, qui s'accumule dans le MAC maternel et n'en sort pas (Nowacki et al., 2005). Puisque cette protéine interagit probablement avec les Piwi via ses répétitions GGWG (type *Ago hook* (El-Shami et al., 2007)), elle pourrait bloquer avec elle les protéines Piwi dans le MAC maternel, les empêchant ainsi de délivrer une quelconque information aux noyaux gamétiques ou au partenaire. Cette expérience nécessite d'injecter à la fois la version tronquée de *Nowa* et *Ptiwi01* pour en suivre la localisation. Ceci peut se faire soit par injection d'un mélange des plasmides, soit par injection séquentielle d'un plasmide puis de l'autre. J'ai utilisé la première méthode en injectant un mélange pMN5:pP1NTom de ratio 10:1 afin de d'avoir pMN5GFP en excès pour piéger toutes les protéines Piwi. Étonnamment, les clones injectés présentaient la fluorescence rouge de pP1NTom mais aucune fluorescence verte malgré la confirmation PCR que la construction pMN5GFP était présente (données non montrées). La deuxième méthode a révélé des profils de localisation identiques pour les protéines fluorescentes par rapport à l'injection indépendante des deux constructions. En particulier, dans les stades tardifs d'autogamie, pMN5GFP reste dans les fragments du MAC maternel tandis que pP1NTom s'accumule dans les nouveaux MAC (**Figure VII.7 A**). Autrement dit, le piégeage initialement imaginé dans cette expérience n'a pas lieu. Étant donné les problèmes de localisation rencontrés avec la construction pP1NTom, nous avons décidé d'inverser les étiquettes fluorescentes. J'ai construit et injecté le plasmide pMN5Tom puis j'ai injecté pP1NGFP. Premièrement, la construction pMN5Tom donne un signal différent de pMN5GFP puisque le MAC maternel se distingue à peine du cytoplasme avant l'apparition des *anlagen* (**Figure VII.7 B à D**). À partir de ce moment, il se concentre dans les fragments macronucléaires. Deuxièmement, les *anlagen* émettent souvent une fluorescence rouge ce qui signifie que la localisation de la protéine pMN5Tom n'est pas restreinte à l'ancien MAC. Enfin, la protéine pP1NGFP s'accumule normalement dans les *anlagen* confirmant ainsi l'échec du système de piégeage. Nous n'avons pas poursuivi ce projet par la suite.



**Figure VII.7: Dynamique de localisation des protéines *Nowa1* tronquée et *Ptiwio1* au cours d'une autogamie.**

Les constructions pMN5 et pP1NTom (A) ou pMN5Tom et pP1NGFP (B à D) ont été injectées séquentiellement dans une cellule. La nature du signal observé pour chaque case est indiquée en haut de la colonne. Les stades représentés sont '4 MIC' (B) et *anlagen* (A, C et D). Les flèches blanches et jaunes indiquent les MIC et les *anlagen*.

## D. Lien avec la méthylation de l'ADN

S'il existe une empreinte génomique dans les noyaux gamétiques, alors on peut réfléchir sur la forme qu'elle pourrait prendre. Les marques épigénétiques les plus courantes sont les modifications des histones et la méthylation de l'ADN. Cette dernière semble plus appropriée à la taille des IES donc nous avons fait des recherches dans ce sens. Elles rejoignent la thématique plus globale du rôle possible de la méthylation dans les réarrangements programmés de l'ADN chez la paramécie. Les études préliminaires de notre laboratoire ont identifié des AT méthylés dans le génome de *Paramecium tetraurelia*, principalement dans le MAC. D'autres sites sont méthylés et demandent une analyse approfondie.

### 1) Méthylation des cytosines

La méthylation des cytosines sur le carbone 5 est une marque de régulation épigénétique bien connue, notamment chez les plantes. Elle est présente chez certains ciliés tels qu'*Oxytricha* (Bracht et al., 2012) mais n'a jamais été détectée chez les oligohyménozoaires. Pourtant, les CpG sont sous-représentés dans le génome de *Paramecium tetraurelia*, avec seulement 42 et 44 % des CpG normalement attendus respectivement dans le MAC et les IES. Chez les organismes qui sont capables de méthyler les CpG, on l'explique par l'hypermutableté de la cytosine méthylée. En effet, sa désamination spontanée les convertit en TpG qui ne sont pas toujours réparés. Une sous-représentation du dinucléotide CpG de l'ordre de 43 % pourrait donc indiquer la présence d'une enzyme qui le méthyle (Schorderet and Gartler, 1992). Afin d'étudier la question, j'ai réalisé une PCR sur un échantillon d'ADN traité au bisulfite selon la méthode suivante (Henderson et al., 2010). La conversion des cytosines non méthylées en uraciles rend les brins non complémentaires. La PCR est alors biaisé en faveur d'un seul brin, celui que l'on souhaite étudier. Le produit d'amplification est ensuite inséré dans un vecteur et le tout est transformé dans des bactéries. Chaque clone possède alors un plasmide dont la séquence varie en fonction du niveau de méthylation de la molécule ayant servi de matrice à l'amplicon. J'ai utilisé cette méthode en travaillant à partir d'un échantillon d'ADN de la souche 32 obtenu après inactivation de Pgm. En effet si les cytosines méthylées servent à discriminer les séquences qui doivent être excisées alors cette discrimination doit être présente au moment où l'enzyme coupe. L'inactivation de l'enzyme permet donc d'optimiser le nombre de séquences marquées. De plus, la région choisie est à cheval sur la jonction IES-MAC de 32G4404 avec 16 cytosines dans l'IES et 19 dans la séquence MAC flanquante. Seul l'ADN des MIC ou des *anlagen* non réarrangés peut donc être amplifié. A l'issue du séquençage de 21 clones, j'ai observé 10 cytosines qui n'ont pas été converties dans un clone, dont 7 dans la région IES (**Figure VII.9**). Par ailleurs, deux cytosines de la région flanquante sont non converties dans 3 clones. Cela fait un total de 16 cas de non conversion parmi 735 possibles soit un taux de 2,2 %. Si ce taux est dû au traitement bisulfite, alors la probabilité de toucher 3 fois la même cytosine est de 1 ‰. On peut en déduire que les deux cytosines non converties dans 3 clones représentent une réalité chimique ou biologique. Si ce n'est pas une propriété de la séquence elle-même alors on peut supposer qu'il existe un taux de méthylation de l'ordre de 14 %. Les deux contextes sont TTACCACG et GTTCACT, d'où un motif potentiel CAC. Ce motif est présent une autre fois dans l'IES et une autre

fois dans la séquence flanquante. Dans ces deux cas, la 1ère cytosine a montré un cas de non conversion et la deuxième aucun. Au final, il existe peut-être un motif de méthylation mais il faudrait davantage de données pour conclure. D'autre part, ce taux semble trop faible (14 %) pour avoir un rôle important. Une piste potentiellement intéressante est celle de la méthylation sur le carbone 4 puisqu'il a été montré que cette dernière pouvait également inhiber la conversion par le bisulfite (Vilkaitis and Klimasauskas, 1999). L'inhibition varie de 30 à 100 % selon la séquence et cela pourrait expliquer les faibles taux observés. En revanche, cette méthylation sur le carbone 4 n'a jamais été détectée chez les eucaryotes. De récentes données de séquençage de molécule unique Pac-Bio devraient nous permettre d'en savoir plus.

GGTattgtaaatttaaaatataatTTTTgaatagTTTTtagataaaaagtagtgatt**C**tattattaaatgtaattgTTT  
aaattgaatttt**C**ttcattttttatTTTcTaaattgTtctaaaa**C**ttaaattagagTtaataatagaatat**C**acaa  
ttagtttg**C**tttgTTTTcttttatgTTTTatt**ca**caagaa**ca**aatatattttaaaaattt**ta**cttaataat**ACTGC**  
TTGTTAAGAAT**CAAAAGCTACTT**G**CAAGGATTAC****CG****ACATAAAACAAATGCTCCAC**  
ATAAAC**CAACTAGTTCACTCTT**TATGTATTTGGGTAGATAATTTT**GTTATATATTTGG**

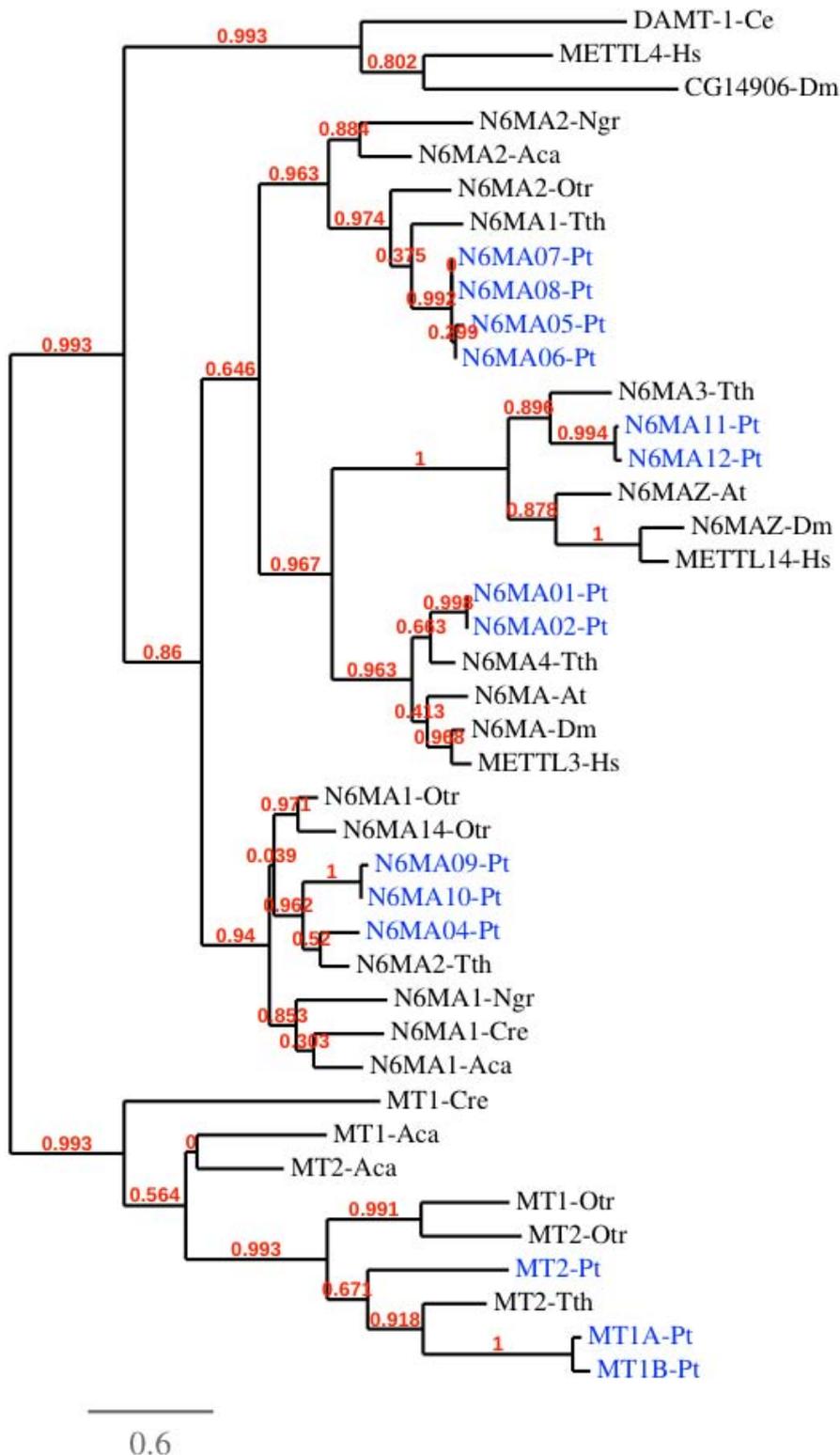
Nb of unconverted C sequenced out of 21 clones  
**C** = 0    **C** = 1    **C** = 2    **C** = 3

**Figure VII.9 : Analyse de la méthylation des cytosines dans une région IES-MAC.**

La séquence est orientée de 5' vers 3', avec l'IES 32G4404 en minuscule. La première séquence soulignée est l'amorce utilisée pour la PCR et la deuxième est le complément inverse de la deuxième amorce. Chaque cytosine de la séquence est colorée en fonction du nombre de clones dans lesquelles elle n'était convertie. Le seul motif CG de la séquence est encadré.

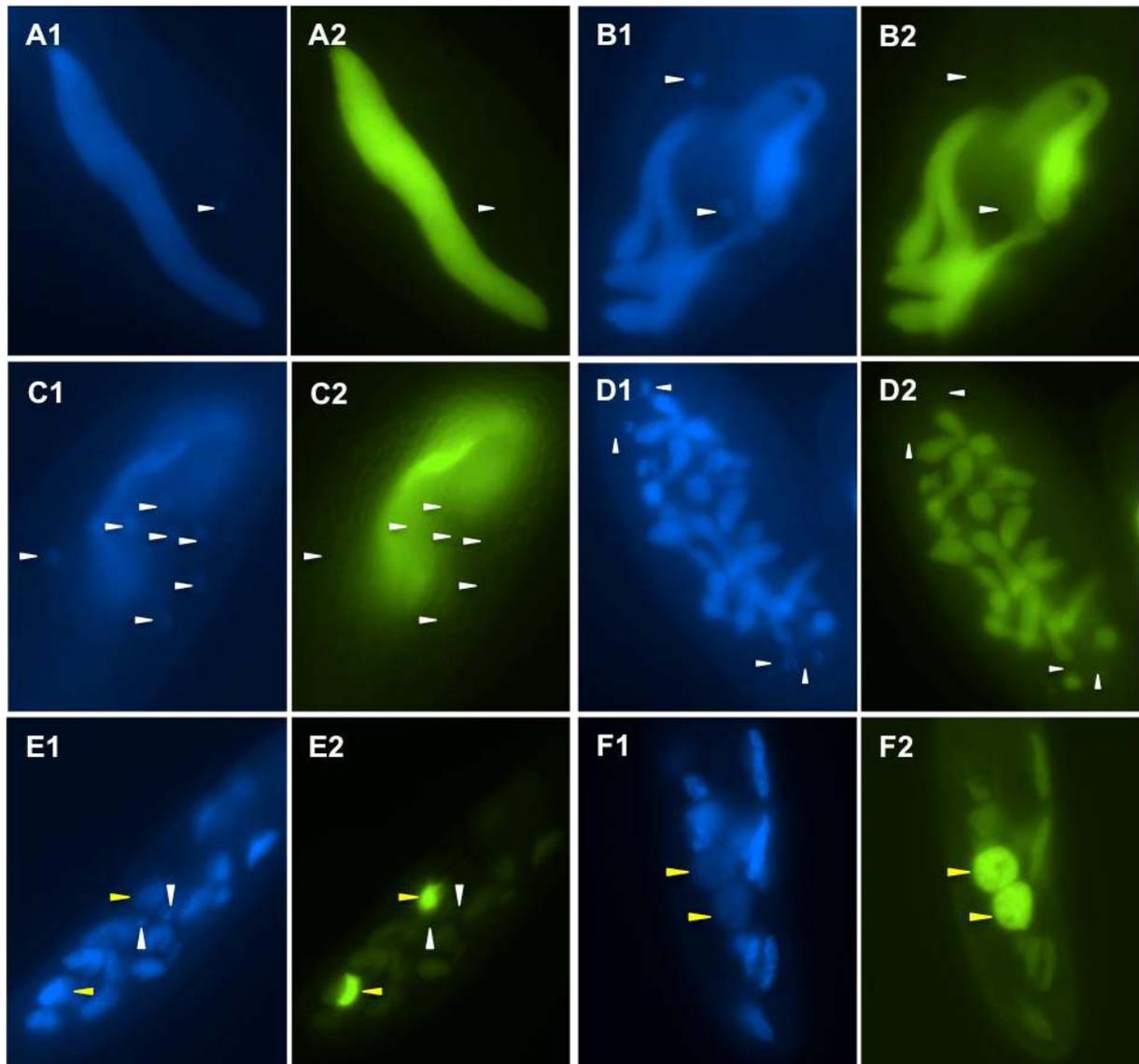
**2) Méthylation des adénines**

Les méthylation des adénines, beaucoup moins connue chez les eucaryotes, est une piste plus prometteuse puisque qu'on sait qu'elle est abondante dans le génome de *Paramecium* (2,4 % des adénines dans le MAC). De récentes publications (Greer et al., 2015; Iyer et al., 2016) ont permis de rechercher par homologie des adénines méthylases chez la paramécie et de nombreuses ont été identifiées (**Figure VII.10**). Deux groupes assez éloignés se distinguent, l'un noté N6MA et l'autre MT. Le 1er groupe se subdivise en 4 clades de 4, 2, 2 et 3 gènes. Les gènes du 1er clade, surexprimés à l'autogamie, sont de bons candidats pour jouer un rôle dans les réarrangements du génome. Ils sont étudiés par le groupe de S. Duharcourt (IJM, Paris, France). Les deux clades suivants sont des orthologues clairs des méthylases connues pour modifier les ARN chez d'autres eucaryotes. Nous avons alors choisi de travailler sur le clade contenant N6MA04, 09 et 10. De plus nous avons étudié les 3 gènes du groupe MT : MT1A, 1B et 2. Les expériences d'extinction génique par ARNi durant la phase végétative ou à l'autogamie n'ont donné aucun phénotype sur les cellules ou leurs descendants post-autogames (données non montrées). J'ai aussi construit les protéines de fusion N6MA04-GFP et N6MA09-GFP afin de voir leur localisation subcellulaire. Elles ont un profil identique avec un signal exclusivement macronucléaire à tous les stades du cycle de vie, ce qui semble exclure un rôle dans la programmation précoce des réarrangements (**Figure VII.11 et VII.12**).



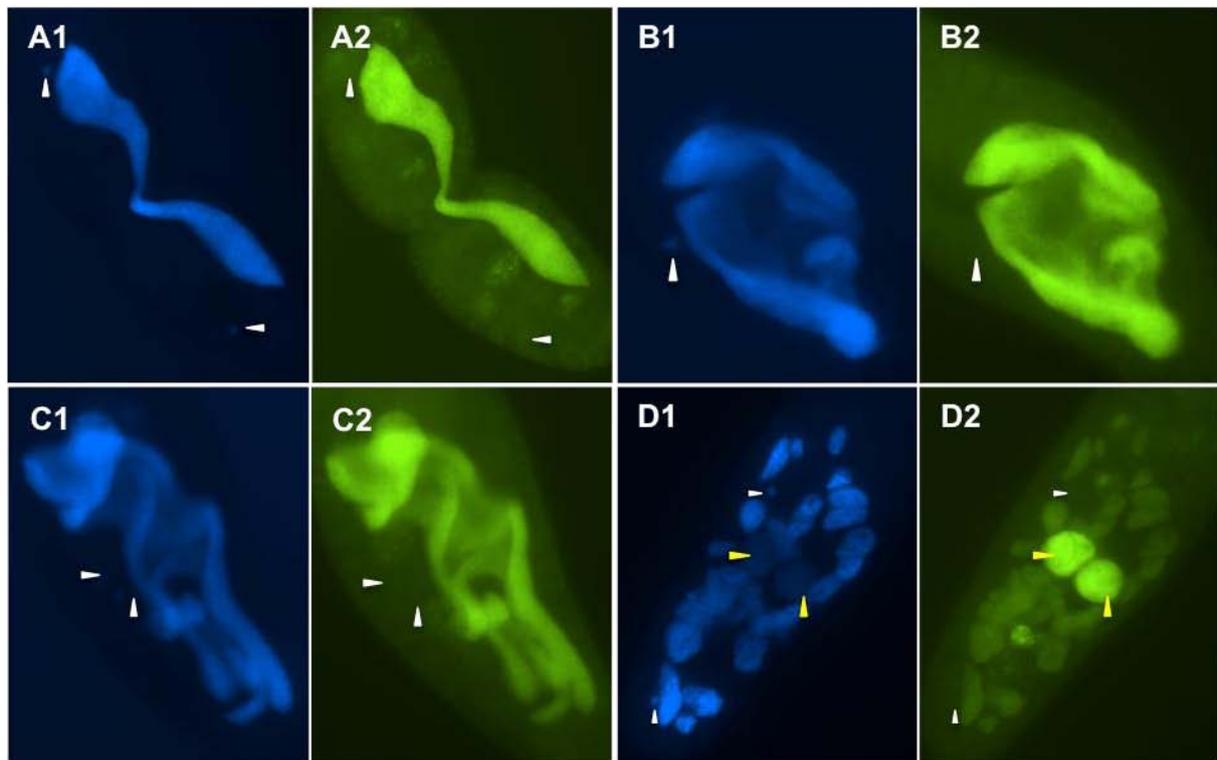
**Figure VII.10 : Arbre phylogénétique de méthyl-transférase**

L'alignement se fait sur le domaine catalytique MTA-70 avec MUSCLE 3.8.31 (paramètres par défaut). L'arbre est calculé par PhyML 3.0 avec la méthode approximative et rapide aLRT puis représenté avec TreeDyn 198.3. Aca *Acanthamoeba castellanii*, Cre *Chlamydomonas reinhardtii*, Ngr *Naegleria gruberi*, Tth *Tetrahymena thermophila*, Pt *Paramecium tetraurelia*, Otr *Oxytricha trifallax*, Hs *Homo sapiens*, Dm *Drosophila melanogaster*, At *Arabidopsis thaliana*, Ce *Caenorhabditis elegans*.



**Figure VII.11 : Dynamique de localisation de N6MA04-GFP au cours d'une autogamie.**

Les photos numérotées 1 et 2 montrent respectivement le marquage au DAPI et la fluorescence GFP. Les lettres correspondent à différents stades du cycle de vie. (A) Cellule en division végétative. (B) Méiose I ou II. (C) 8 MIC. (D) 4 copies du noyau zygotique. (E et F) *Anlagen*. Les flèches blanches et jaunes indiquent les MIC et les *anlagen*.



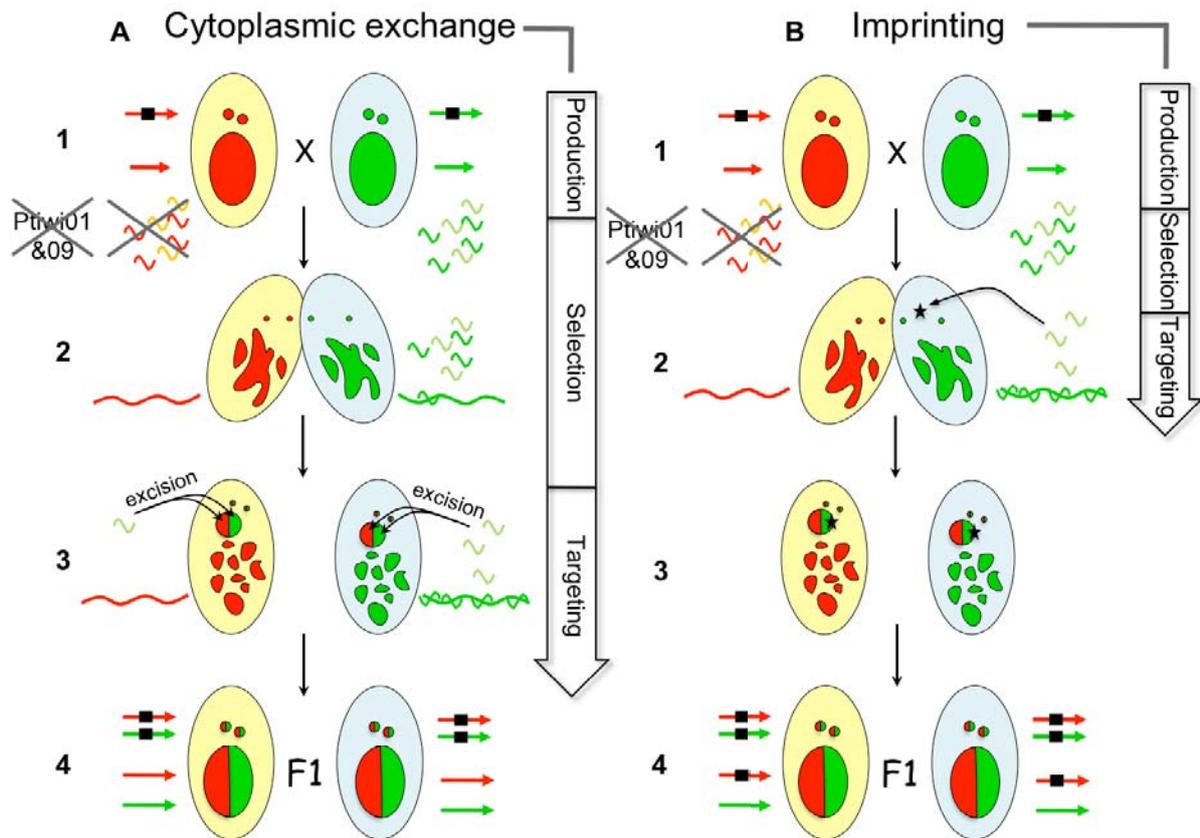
**Figure VII.12 : Dynamique de localisation de N6MA09-GFP au cours d'une autogamie.**

Les photos numérotées 1 et 2 montrent respectivement le marquage au DAPI et la fluorescence GFP. Les lettres correspondent à différents stades du cycle de vie. (A) Cellule en division végétative. (B) Méiose I. (C) Méiose II. (D) *Anlagen*. Les flèches blanches et jaunes indiquent les MIC et les *anlagen*.

## E. Excision des allèles identiques dans le croisement 51 x 32 avec extinction des gènes *Dcl2/3*.

### 1) Hypothèse

L'expérience suivante nous donne probablement l'argument le plus fort en faveur de l'échange cytoplasmique et au détriment de l'hypothèse de l'empreinte. Nous avons vu dans le cas des allèles divergents 51G4404 et 32G4404 que la déplétion de *Ptiwi* dans le parent 51 induit la rétention de l'allèle 51 dans tous les F1. Les scnARN du parent 32 ne pouvaient pas guider l'excision de l'allèle en question par manque d'homologie. Nous nous sommes alors demandé dans quelle mesure ceux-ci pouvaient guider l'excision des allèles 51 pour les IES dont les deux allèles sont identiques ou très proches en séquence. Suivant l'hypothèse d'un échange cytoplasmique des scnARN, on peut logiquement supposer que les scnARN du parent 32 vont assurer l'excision des allèles 51 introduit par le noyau gamétique mâle 32, et que les scnARN d'origine 32 migrant dans la cellule 51 vont également diriger l'excision des allèles 51 locaux (**Figure VII.13 A**). En revanche, si le rôle des scnARN se limite à la déposition d'une empreinte dans les noyaux gamétiques avant l'échange réciproque des noyaux mâles, alors les scnARN d'origine 32 ne pourront pas guider l'excision des allèles 51 car il sera trop tard (**Figure VII.13 B**).



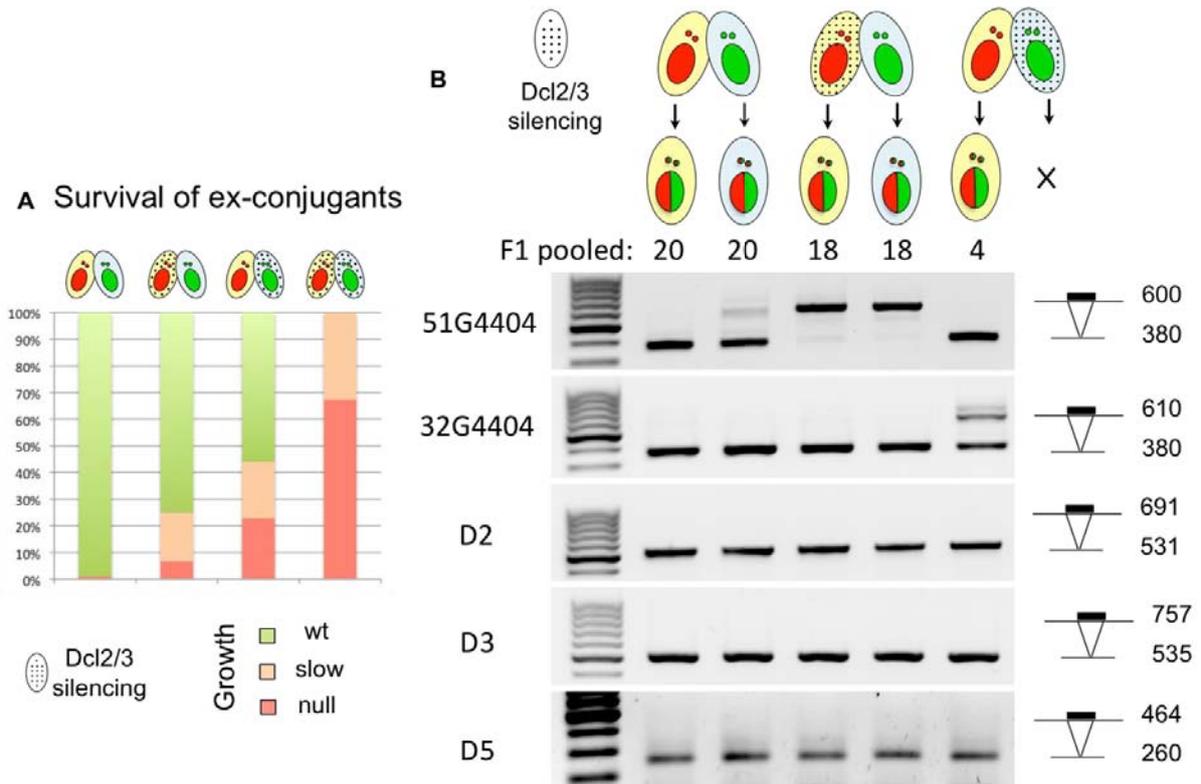
**Figure VII.13 : Modèles d'excision des allèles identiques par la voie du genome scanning avec deux hypothèses d'action des scnARN paternels.**

Ces modèles dérivent de la figure VII.1, dont la légende est bien détaillée. Ici, le parent de gauche (génomme rouge, cytoplasme beige) est soumis à une déplétion des protéines Ptiwi01&09 rendant ses scnARN instables. Les allèles rouge et vert du gène sont identiques, tout comme l'IES qui est présente dans la version MIC (carré noir). (A) Dans l'hypothèse de l'échange cytoplasmique, les scnARN de la cellule 32 sont capables d'exciser les allèles 51 dans les *anlagen* hétérozygotes de l'ex-conjugant 32 et de l'ex-conjugant 51. Ainsi, les F1 ont des génomes correctement réarrangés. (B) Dans l'hypothèse d'une empreinte, seuls les noyaux gamétiques 32 peuvent être programmés (étoile noire), ainsi les allèles 51 ne seront pas excisés dans les F1 hétérozygotes.

## 2) Expérience

L'expérience d'extinction génique de Ptiwi01&09 dans un parent a déjà été réalisée (voir VI-D-1). Nous aurions pu tester notre hypothèse dans ces échantillons d'ADN de F1 mais deux problèmes se posaient. Tout d'abord, les échantillons d'ADN sont limités et ne permettaient pas de tester la rétention de plusieurs nouvelles IES. Ensuite, nous devons choisir des IES qui doivent être fortement retenus en cas de déplétion de Ptiwi01&09. Or l'efficacité de la déplétion réalisée était moyenne dans mon expérience précédente au vu du taux partiel de rétention de G4404 et de l'absence de létalité quand les deux parents étaient déplétés. De plus, il n'existe pas encore de données de séquençage listant les IES sensibles à ces facteurs donc il serait difficile de choisir les IES à tester. En conséquence, nous avons décidé de réaliser un nouveau croisement 51 x 32 en déplétant cette fois les protéines Dcl2&3, responsables de la synthèse des scnARN. La déplétion se fait toujours par ARNi dans le parent 51, le parent 32, aucun parent, ou les deux parents. Dans ce dernier cas, aucun F1 ne

survit donc la déplétion est efficace (**Figure VII.14 A**). Lorsque le parent 51 est déplété, la majorité des F1 survit mais lorsque c'est le parent 32 qui est traité, de nombreux F1 meurent rapidement, surtout ceux dérivant du parent 32. Contrairement aux autres croisements, je n'ai pas extrait l'ADN des F1 individuellement mais j'ai réalisé un pool de plusieurs cultures de F1 avant d'en faire une extraction. Le nombre d'individus F1 réunis dans chaque *pool* est indiqué (**Figure VII.14 B**). Il y en a respectivement 20 et 18 pour le croisement contrôle et celui avec le parent 51 déplété, pour chaque côté du croisement. Lorsque le parent 32 subit la déplétion de Dcl2/3, seuls 4 F1 dérivant du parent 51 ont poussé suffisamment pour que j'en extraie l'ADN. A partir de ces 5 échantillons d'ADN, j'ai d'abord réalisé les PCR sur les allèles 51 et 32 de G4404. Ces allèles ne présentent pas de rétention dans le croisement contrôle à l'exception de l'allèle 51 dans les F1 dérivant du parent 32 (rétention partielle). Puisque nous observons l'image donnée par un pool de 20 F1, cela peut signifier que tous les individus la retiennent partiellement ou que quelques individus la retiennent fortement et les autres pas du tout. Le résultat de la figure VI.4 favorise la seconde option. Lorsque le parent 51 a été soumis au *silencing* de Dcl2/3, l'allèle 51G4404 est retenu à 100 % dans tous les F1 et l'allèle 32G4404 pas du tout. Enfin lorsque le parent 32 est déplété, l'allèle 51 est bien excisé et c'est l'allèle 32 qui est retenu mais pas à 100 %. Nous l'expliquons par le fait que l'allèle 32 est moins soumis au contrôle maternel que l'allèle 51 (voir VII-C, dernier paragraphe) et donc probablement moins sensible à l'absence de Dcl2/3. Avec ces observations, on peut dire que l'ARNi dirigé contre Dcl2/3 a été globalement efficace. Nous avons ensuite sélectionné 3 IES connues pour être sensibles à Dcl2/3 et dont les allèles 51 et 32 sont identiques ou presque (**Table S2**). Les IES sont totalement excisées dans toutes les conditions, donc le parent produisant des scnARN peut compenser l'absence de Dcl2/3 dans l'autre parent. Cette compensation n'est envisageable qu'avec une action post-fécondation des scnARN et une possibilité d'échange cytoplasmique de ces scnARN.



**Figure VII.14 : Croisement 51 x 32 avec déplétion de Dcl2/3 et analyse de l'excision des allèles identiques entre les deux souches.**

Croisement de la souche 51 (noyaux rouges, cytoplasme beige) avec la souche 32 (noyaux verts, cytoplasme bleu) et silencing de Dcl2/3 (motif à points noirs dans le cytoplasme du parent concerné). (A) Pour chacun des 4 croisements, 30 couples donnant 60 ex-conjugants sont isolés et leur viabilité est comptabilisée après 4 jours à 27 °C. (B) Pour chaque origine maternelle de chaque croisement, si la viabilité le permet, des F1 hétérozygotes ont été réunies dans un pool dont l'ADN a été extrait. Les résultats de 5 PCR permettent d'évaluer la rétention des IES dont le nom est noté à gauche des gels. Les bandes de l'échelle de poids moléculaire sont espacées de 100 pb avec la bande la plus intense représentant 500 pb.

## VIII – Avancées sur le modèle du scanning du génome

### A. Existe-t-il une boucle d'amplification ?

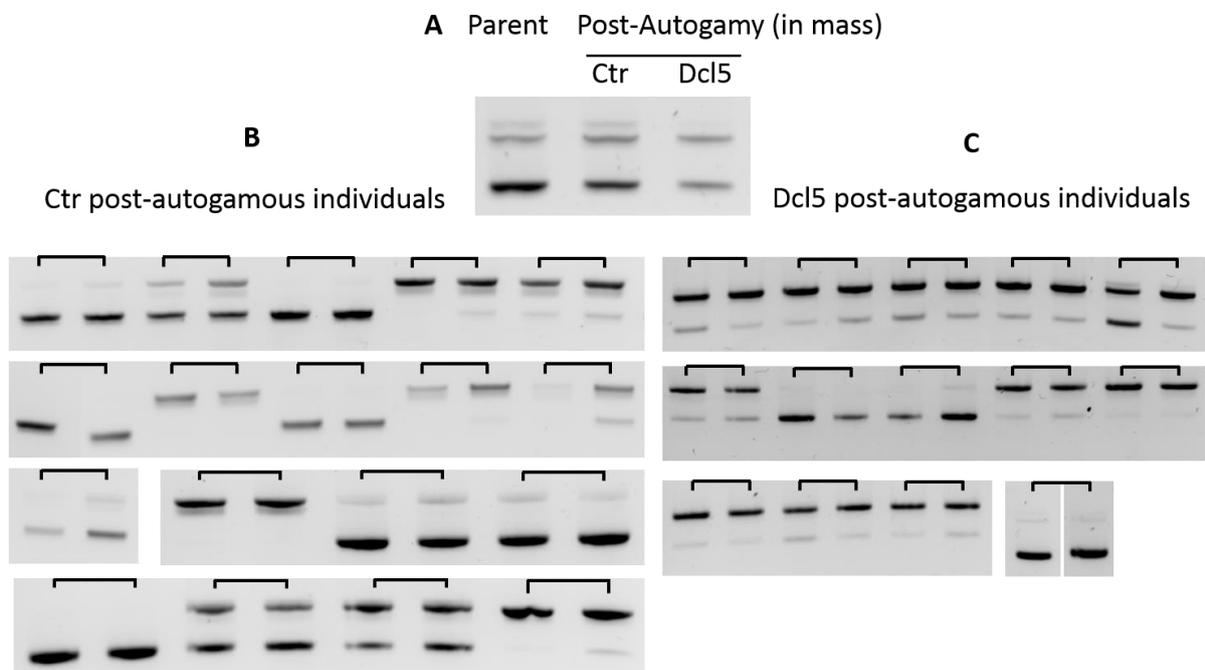
#### 1) Réflexions

Dans le MAC de *Paramecium*, il existe environ 800 copies du génome haploïde. Les réarrangements débutent dans les *anlagen* après 3-4 cycles d'endoréplication (Bétermier et al., 2000) c'est-à-dire lorsqu'il y a entre 16 et 32 copies dans chacun des *anlagen*. Plus l'excision des IES est tardive, plus il y a de copies à traiter. Afin de faciliter le processus, une boucle d'amplification serait tout à fait appropriée, d'autant plus que c'est un mécanisme courant dans les voies des piARN à l'image du cycle ping-pong. Une telle boucle pourrait être facultative dans la majorité des cas en raison de l'abondance des scnARN. En revanche, dans un croisement intersouches

nécessitant l'importation de scnARN paternels, un mécanisme d'amplification pourrait être indispensable à l'excision des allèles paternels à partir d'une poignée de scnARN échangés. Ce raisonnement nous a poussés à mener quelques expériences pour mettre en évidence un phénomène d'amplification.

## 2) Etude de la rétention de l'IES de mtB dans des caryonides individuels

Dans notre modèle du *genome scanning*, une boucle d'amplification devrait assurer l'excision totale d'une IES à partir du moment où une copie est excisée. En conséquence, tout MAC nouvellement réarrangé devrait présenter des IES soit excisées à 100 %, soit retenues à 100 %. Afin de tester ce raisonnement, nous avons utilisé un clone de la souche 51 qui retient l'IES de mtB à 50 % environ après autogamie (**Figure VIII.1 ctr**). Cette image de rétention à 50 % est l'image moyenne représentée par 20 paires de caryonides. Lorsqu'on analyse la rétention par caryonide, on a accès à plus d'informations. Tout d'abord, les taux de rétention de l'IES entre deux caryonides frères sont très proches, indiquant que le déterminisme des réarrangements est le même pour les deux *anlagen*. Il pourrait avoir lieu avant la division du noyau zygotique, et dans ce cas il faut repenser à la notion d'empreinte. Ou alors, le signal déterminant l'excision d'une IES est cytoplasmique et affecte pareillement les deux *anlagen*, ce qui pourrait s'expliquer par une diffusion homogène des scnARN dans le cytoplasme. L'analyse caryonidale de la rétention de l'IES de mtB montre également une grande variabilité entre paires de caryonides frères. Sur les 18 testés, 8 excisent totalement l'IES, 6 la retiennent à 100 % (ou presque) et les quelques autres la retiennent partiellement. Ce résultat montre que l'excision des copies de l'IES est plutôt un système tout ou rien favorisant l'idée d'une boucle d'amplification.



**Figure VIII.1 : Comparaison de l'excision de l'IES de mtB après autogamie, en masse ou par individu.**

Tous les gels montrent le produit de la PCR utilisant des amorces autour de l'IES de mtB, ils permettent d'estimer le taux de rétention de l'IES. Les produits amplifiés sont respectivement

438 et 572 pb pour les formes IES- et IES+. (A) Rétention de l'IES pour le parent initial et pour les deux populations post-autogames, représentées par un pool de 15 individus, après déplétion de Dcl5 (puits Dcl5) ou non (puits ctr). (B et C) Rétention de l'IES de mtB pour des individus post-autogames de la population ctr (B) ou Dcl5 (C). Les crochets relient deux caryonides frères.

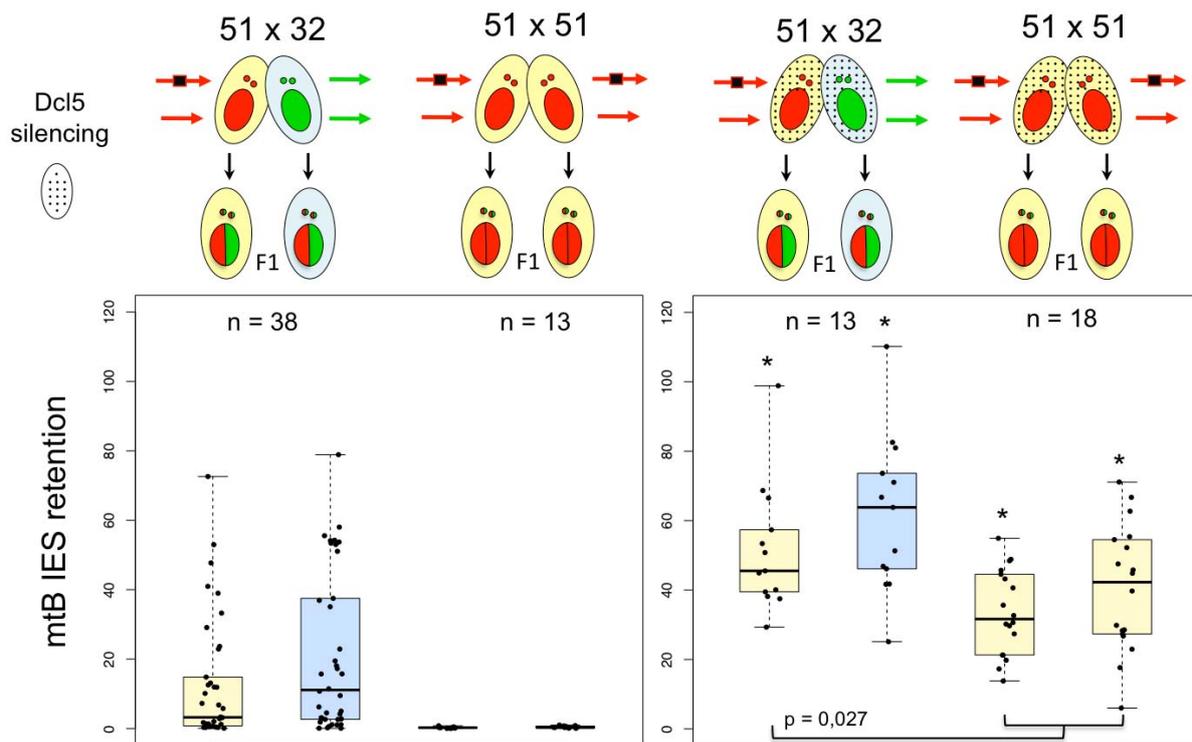
### 3) Etude de la voie Dcl5

#### 1. Présentation de la protéine Dcl5

Dans l'introduction, j'avais déjà évoqué l'hypothèse d'une boucle d'amplification faisant intervenir des iesARN produits par Dcl5 (Sandoval et al., 2014). Cette deuxième vague de petits ARN serait produite tardivement à partir d'IES excisées puis transcrites. Toutefois, l'extinction du gène Dcl5 pendant l'autogamie affecte peu d'IES et pour la plupart de celles-ci, elle n'affecte que modestement leur excision. De plus elle n'entraîne pas de létalité des clones post-autogames. Nous avons tout de même réalisé plusieurs expériences de déplétion de cette protéine, en imaginant que son rôle puisse être critique dans certaines conditions comme par exemple le croisement 51 x 32.

#### 2. Effet de Dcl5 sur l'IES de mtB

J'ai réalisé l'analyse de la rétention de l'IES de mtB après autogamie du clone donnant normalement 50 % de rétention. La déplétion de Dcl5 durant l'autogamie augmente le taux global de rétention (**Figure VIII.1 Dcl5**). Au niveau individuel, la corrélation caryonidale est toujours très forte et les paires de caryonides frères retiennent l'IES à un taux proche de 100 % (11 cas sur 14) ou alors ils l'excisent totalement (3 cas sur 14). Par rapport à la situation contrôle, il semble qu'il y ait davantage de paires retenant fortement l'IES alors que celles qui la retenaient partiellement n'existent plus. Lorsqu'elle est fonctionnelle, la protéine Dcl5 favoriserait l'excision de l'IES de mtB sans pour autant assurer son excision sur toutes les copies. Dans des croisements 51 x 32 et 51 x 51, j'ai ciblé l'expression de Dcl5 par ARNi dans les deux parents (**Figure VIII.2**). Dans un croisement 51 x 51, cela a pour effet d'induire la rétention de l'IES de mtB qui passe de 0 à 40 %. Dcl5 est donc un facteur essentiel à l'excision de l'IES de mtB. Ce résultat confirme la rétention calculée dans les données de séquençage obtenues après déplétion de Dcl5 au cours d'une autogamie (Sandoval et al., 2014) dans lesquelles l'IES est retenue à 35 %. Cette valeur est d'ailleurs sous-estimée puisque les échantillons d'ADN séquencés ne sont pas des pures *anlagen* et contiennent des fragments de l'ancien MAC. Si on regarde maintenant le croisement 51 x 32, l'extinction génique de *DCL5* augmente la rétention de l'IES de mtB par rapport au croisement 51 x 32 non traité. En revanche, cette rétention n'est pas plus élevée que dans le croisement 51 x 51 précédent. Finalement, on peut conclure que l'excision de l'IES de mtB est dépendante de Dcl5, que ce soit à l'autogamie, en conjugaison 51 x 51 ou 51 x 32. En revanche aucun argument ne permet ici de dire que cette protéine agit dans une quelconque voie d'amplification. Par ailleurs, aucun rôle spécifique au croisement 51 x 32 ne semble apparaître.

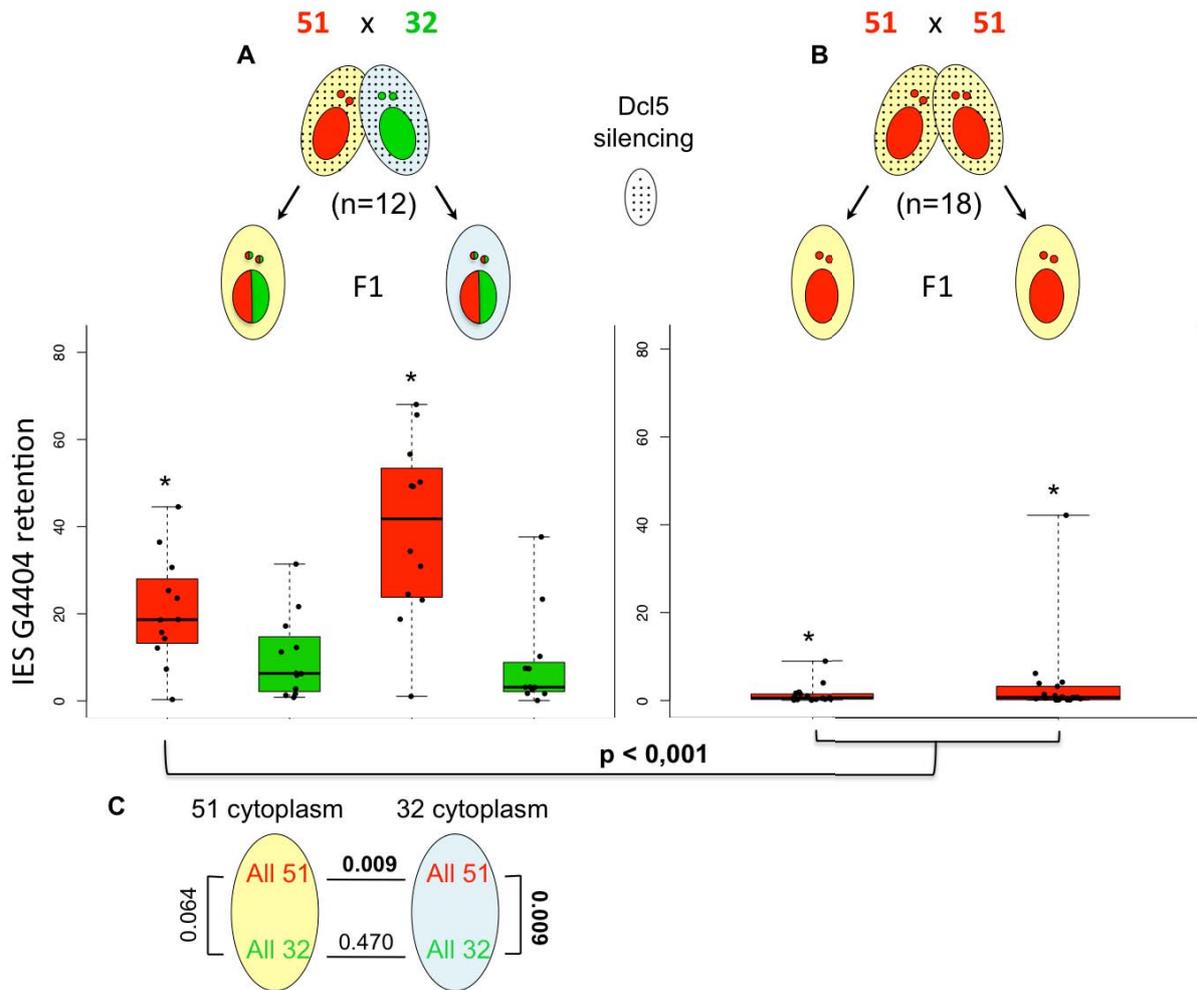


**Figure VIII.2 : Répartition de l'IES de mtB en F1 en fonction de la répression parentale de Dcl5.**

Une extinction génique de *DCL5* par ARN interférence a été réalisée dans les croisements 51 x 32 et 51 x 51. Elle est matérialisée par le motif à points. Les croisements contrôles sans déplétion sont rappelés à gauche. Le taux de rétention calculé est présenté par ex-conjuguat sous la forme de boxplots avec affichage de tous les points. Les astérisques indiquent une différence significative par rapport au contrôle. "p" indique la p-value issue du test de comparaison de la rétention de l'IES dans un F1 dérivant d'un parent 51 entre un hétérozygote 51/32 et un homozygote 51/51. Le test Kolmogorov-Smirnov a été utilisé.

### 3. Effet de Dcl5 sur l'IES G4404

Afin de profiter au maximum des croisements réalisés, j'ai regardé le devenir de l'IES G4404 lorsque Dcl5 est déplété dans les deux parents (**Figure VIII.3**). Dans un croisement 51 x 51, la rétention est très faible mais significativement supérieure au contrôle sans déplétion (Figure VI.4). Ce résultat contraste avec la publication sur Dcl5 dans laquelle aucune rétention n'a été détectée pour 51G4404 après déplétion de Dcl5 pendant l'autogamie (Sandoval et al., 2014). Cela suggère que Dcl5 aurait un rôle spécifique lors de la conjugaison. Alternativement, il peut s'agir d'un biais technique car notre méthode utilisant des individus et un test paramétrique est plus sensible que la leur, d'où la détection de cette faible rétention. Dans le croisement 51 x 32, la rétention de l'allèle 51G4404 est variable mais clairement plus élevée que dans le croisement contrôle. Elle est plus élevée dans les F1 dérivant du parent 32 mais c'est aussi le cas dans le croisement contrôle. L'allèle 32G4404 semble aussi plus retenu, surtout dans les F1 dérivant du parent 51 mais le test statistique ne permet pas de l'affirmer avec un risque inférieur à 1 %. Au final, Dcl5 semble avoir un rôle important dans l'excision de l'IES G4404 mais seulement lors du croisement 51 x 32. Ce résultat nous encourage dans l'idée que Dcl5 puisse être vital dans les croisements intersouches.



**Figure VIII.3 : Rétention de l'IES G4404 en F1 en cas de déplétion de Dcl5.**

Une extinction génique de *DCL5* par ARN interférence a été réalisée dans les croisements 51 x 32 (A) et 51 x 51 (B). Elle est matérialisée par le motif à points. Les croisements contrôles peuvent être visualisés dans la figure VI.4. Une PCR utilisant des amorces encadrant l'IES G4404 permet de calculer spécifiquement le taux de rétention pour chaque allèle (51G4404 en rouge et 32G4404 en vert). Les taux de rétention calculés sont présentés sous la forme de boxplots avec affichage de tous les points. Les astérisques indiquent que la médiane observée est significativement différente de celle des croisements contrôles. (C) Pour le croisement 51 x 32, les taux sont comparés 2 à 2 par F1 et par allèle avec un test des signes de Wilcoxon. Les p-value inférieures à 0,01 sont écrites en gras.

**4. Conclusion sur Dcl5**

Les études menées sur Dcl5 ne permettent pas de comprendre clairement son rôle. Tout d'abord, la sensibilité de l'IES de mtB à ce facteur lors de l'autogamie se rajoute aux effets qu'on aurait pu observer dans les croisements. On peut donc difficilement faire le lien entre l'effet hémizygote et la protéine Dcl5. De plus, elle est exprimée pendant le développement des anlagen, ce qui rend incertaine notre technique d'ARNi. En effet, la réactivité est induite lorsque toutes les bactéries sont digérées et nous ne savons pas dans quelle mesure la voie ARNi perdure. Les protéines Dcl5 pourraient avoir le temps d'être synthétisées pour agir lors du développement des anlagen. Il reste tout de même un effet important de la déplétion de Dcl5 dans le

croisement 51 x 32. L'allèle 51G4404 est beaucoup plus retenu par rapport au contrôle. Notre hypothèse de départ proposait que Dcl5 soit requis pour amplifier le signal d'excision à partir de quelques scnARN paternels importés. Si elle était vraie, alors la déplétion de Dcl5 aurait dû avoir peu d'effet sur l'excision des allèles maternels. C'est partiellement vrai car l'allèle 51G4404 est davantage retenu dans les F1 dérivant du parent 32 mais la différence n'est pas très marquée. Pour résumer, Dcl5 semble être un facteur de la voie des scnARN qui aurait un rôle important lors des conjugaisons intersouches mais on n'en sait pas plus.

## **B. Etude cinétique d'une conjugaison hautement synchronisée**

### **1) Pourquoi et comment synchroniser ?**

La voie du *genome scanning* implique de nombreux acteurs : ADN, scnARN, iesARN, ARNnc et protéines. Les données de cinétiques actuelles nous ont déjà fourni un aperçu du déroulement des réarrangements programmés du génome chez la paramécie. Cependant, ces études sont réalisées par autogamie avec des synchronies (laps de temps entre la première et la dernière cellule de la culture à entrer en méiose) de l'ordre de 8-10 h. En effet, l'autogamie ne peut s'enclencher qu'à un moment précis du cycle cellulaire, même si toutes les cellules reçoivent le signal de privation de nourriture en même temps la synchronie ne peut être meilleure que la durée d'un cycle cellulaire (6 h à 27 °C). Afin d'étudier les réarrangements tout en contournant le problème, il faudrait réaliser une conjugaison massive et cette solution n'est pas miraculeuse. Lorsqu'on mélange deux cultures réactives de types sexuels opposés, seulement la moitié des cellules s'agglutinent et forment des couples. L'autre moitié s'accouple plus tard voire jamais. Il faut donc stopper la formation de nouveaux couples après la première vague afin que toutes les cellules en conjugaison soient synchrones. Idéalement, il faut aussi isoler les couples des autres cellules afin d'avoir une population pure. Cette dernière étape n'est pas indispensable pour l'étude des scnARN puisque les cellules qui ne conjuguent pas n'en produisent pas. Elle est beaucoup plus importante pour l'étude des ARNm ou de l'excision de l'ADN car les autres cellules généreront un bruit de fond gênant.

### **2) Tentative de synchronisation avec la WGA**

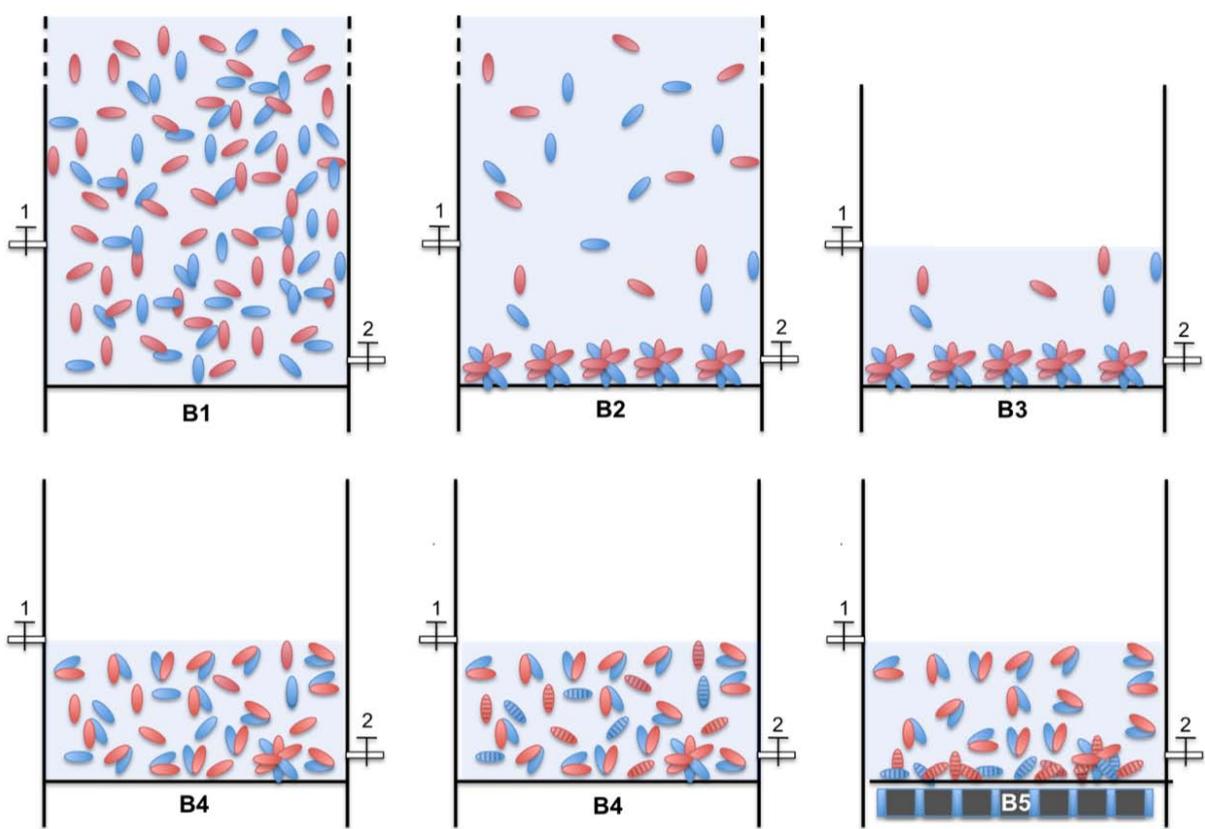
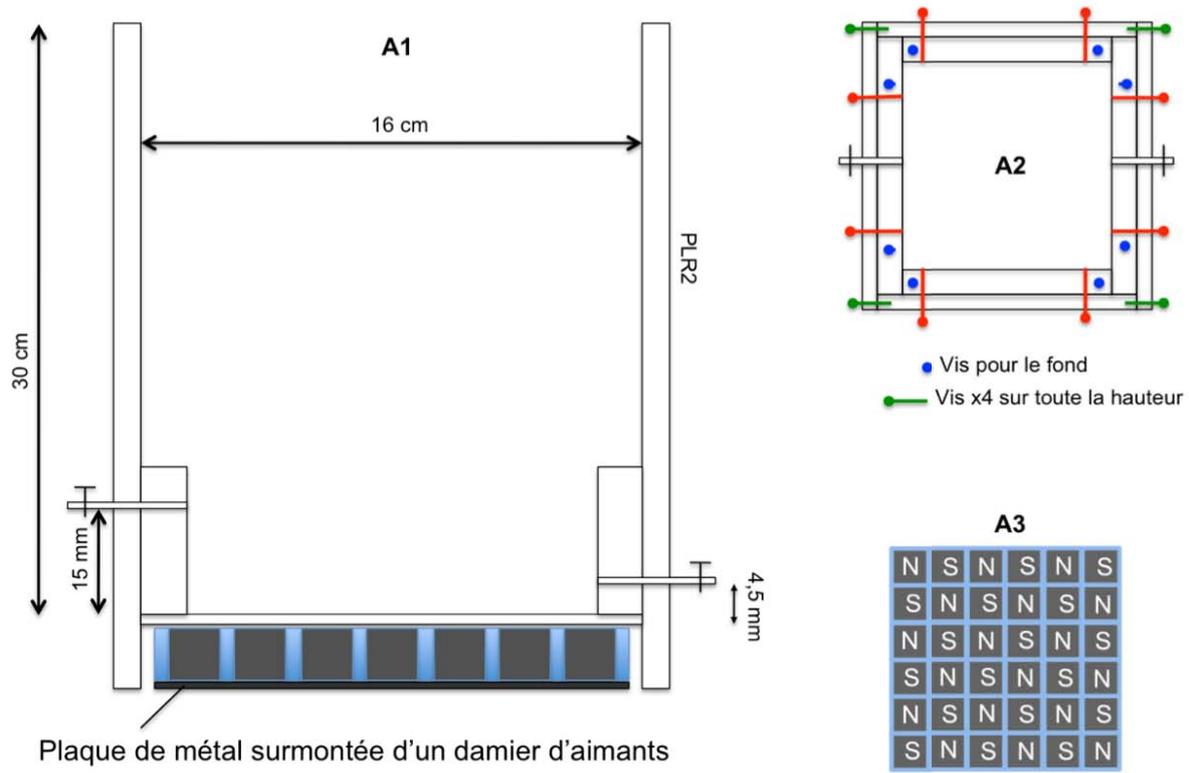
La WGA (*Wheat Germ Agglutinin*) est une lectine couramment utilisée en biologie. Elle est décrite comme inhibitrice de la formation des paires chez *P. primaurelia* lorsque les cellules réactives sont incubées avec cette protéine (Corrado et al., 1997). Nous pensions donc ajouter la WGA après la formation de la majorité des couples (au bout de 2h30) afin de prévenir la formation de nouvelles paires non synchrones. Notre protocole requiert que la WGA bloque efficacement et rapidement la formation de nouvelles paires tout en n'affectant pas les paires déjà formées. Mes tests ont montré que la WGA ne bloque pas la formation des couples mais favorise leur séparation car ils se dissocient plus tôt. En analysant ces ex-conjugants, je me suis rendu compte qu'elle les empêche de réaliser un échange réciproque de noyaux gamétiques. Ce résultat corrobore le fait que la WGA inhibe en réalité la fusion des cytoplasmes (Pape, R., 1988), d'où l'impossibilité d'échanger les noyaux gamétiques.

Par ailleurs, la WGA ne restait pas active suffisamment longtemps pour réaliser une longue cinétique. J'ai donc arrêté les essais de synchronisation avec cette molécule.

### 3) Synchronisation avec des billes magnétiques

#### **a. Principe**

Au cours de la conjugaison, les paramécies perdent leur appareil oral et sont donc incapables d'ingérer des particules de leur milieu. C'est pourquoi le laboratoire de Forney a eu l'idée d'ajouter des particules magnétiques dans le milieu pour discriminer les couples des cellules célibataires (Matsuda and Forney, 2006). Les cellules non engagées ingèrent de nombreuses billes et peuvent être piégées avec un aimant suffisamment puissant. Cette technique n'a pas été utilisée depuis et une première tentative de reproduction de la technique dans un laboratoire parisien n'a pas porté ses fruits. Nous avons pourtant décidé de réessayer dans notre laboratoire, avec l'aide de physiciens. Ils nous ont aidés à réaliser un damier d'aimants pour optimiser l'attraction des cellules gorgées de billes magnétiques et à construire une cuve dans laquelle on pouvait réaliser l'opération (**Figure VIII.4**). Les billes, d'un diamètre de 500 nm, sont composées d'un cœur magnétique encapsulé dans un polymère portant des groupes carboxyles. Les aimants sont des cubes en néodyme de 12 mm de côté ayant une force d'adhérence de 6,3 kg (voir matériel et méthodes pour plus d'informations).



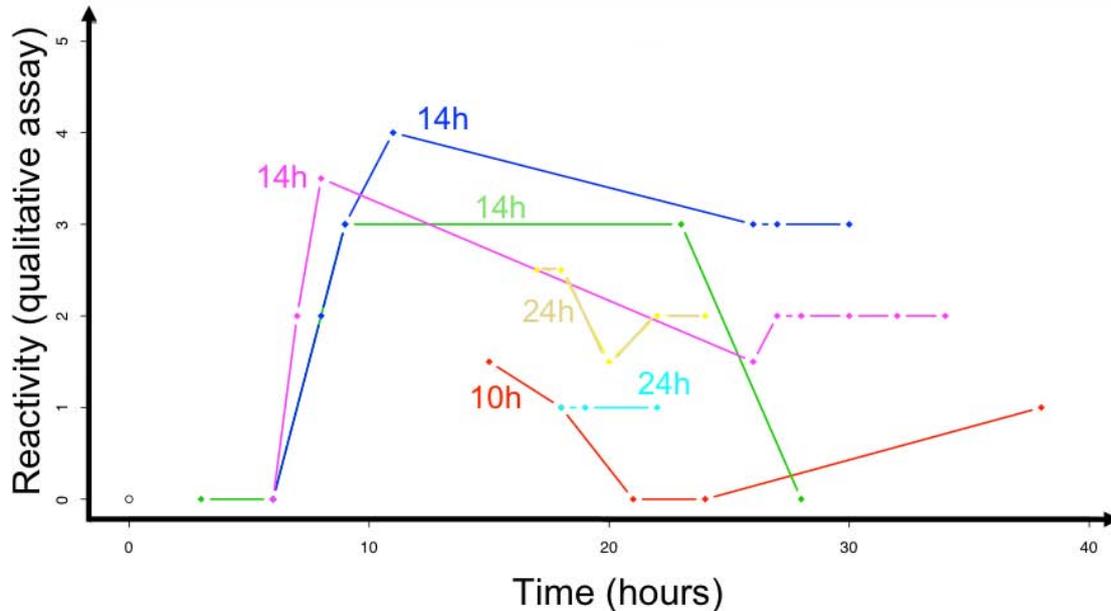
**Figure VIII.4 : Principe de récupération de cellules en conjugaison par utilisation de billes magnétiques dans une cuve adaptée.**

(A) La cuve (A1) peut contenir 5 L de culture. Elle est caractérisée par un fond très fin (2 mm) et deux robinets permettant de récupérer le volume supérieur pour deux hauteurs précises. A sa base, les côtés de la cuve sont additionnés de renforts qui étaient nécessaires pour visser

l'ensemble. Les vis sont schématisées en (A2). Enfin, un damier d'aimants (A3) réalisé en plexiglas avec insertion d'aimants cubiques en alternance de pôles Nord (N) et Sud (S) peut être logé sous la cuve. (B) A partir d'un mélange de cellules réactives de types sexuels opposés (B1), l'ouverture du 1er robinet au bout de 30 minutes permet d'évacuer le volume supérieur contenant des cellules non réactives et de garder principalement les agglutinats (B2 à B3). Environ 2 h après le mélange, les agglutinats ont donné naissance à des couples et on ajoute alors des billes magnétiques à la culture (B4). Les cellules non appariées ingèrent des billes pendant 30 minutes et sont alors piégées avec le damier d'aimant (B4 à B5). Il ne reste qu'à ouvrir le deuxième robinet pour récupérer les couples.

### ***b. Recherche des paramètres optimaux***

Mes premiers tests ont été effectués dans des tubes de 2 mL avec seulement un aimant. Ils m'ont permis de définir un certain nombre de paramètres comme le diamètre et la concentration des billes à utiliser ainsi que le temps nécessaire pour que les paramécies saturer leurs vacuoles digestives en billes. Une étude sur la formation des phagosomes lorsque *Paramecium* est nourri avec des particules solides nous a apportée des éléments importants pour choisir ces paramètres (Fok et al., 1988). J'ai aussi défini le déroulement de toute la procédure. On peut enlever le volume supérieur de cellules non réactives 15 à 60 minutes après le mélange des cultures réactives. Au bout de 2 heures, l'ajout des billes magnétiques entame le processus de discrimination entre les couples formés et les autres cellules. Ce timing est dépendant de la température (27 °C) et requiert une bonne réactivité sexuelle sinon il faut attendre plus longtemps. Enfin, deux minutes en contact avec l'aimant sont suffisantes pour piéger les cellules ayant ingérés des billes. Ces paramètres sont toujours vrais à plus grande échelle, et le damier est suffisant pour retenir les cellules ayant ingéré des billes magnétiques. A plus grande échelle, la cuve ne s'est pas montrée utile. Elle ne supporte pas la pression des 5 L de culture et les débits des robinets sont insuffisants pour mener à bien l'expérience. Des tests dans de petits erlenmeyers indiquent que le damier d'aimants est suffisant pour retenir les cellules ayant ingéré des billes tandis que l'on verse la culture dans un nouveau récipient. C'est donc par ce moyen que sont collectés les couples. Enfin, pour obtenir le maximum de couples, il est impératif que la réactivité sexuelle initiale soit forte. Les cultures doivent être très jeunes c'est-à-dire que les cellules ont fait peu de divisions depuis leur dernier cycle sexuel (autogamie ou conjugaison). La privation de nourriture déclenchera alors la réactivité sexuelle plutôt que l'autogamie, elle peut être réalisée de deux manières. La première consiste simplement à renourrir une culture carencée avec un volume de milieu frais. De cette façon, les cellules pourront réaliser une division avant de ressentir le manque de nourriture. Il est néanmoins difficile de prédire combien de temps nécessitera la reprise de la croissance, la digestion de toute la nourriture et enfin la survenue de la réactivité. La deuxième méthode est la centrifugation brutale d'une culture en croissance et sa resuspension dans du milieu épuisé (sans bactéries). Cela permet de retirer instantanément toute nourriture pour les paramécies. Cette deuxième méthode est plus lourde à réaliser mais beaucoup plus fiable en terme de réactivité obtenue et de reproductibilité (**Figure VIII.5**).

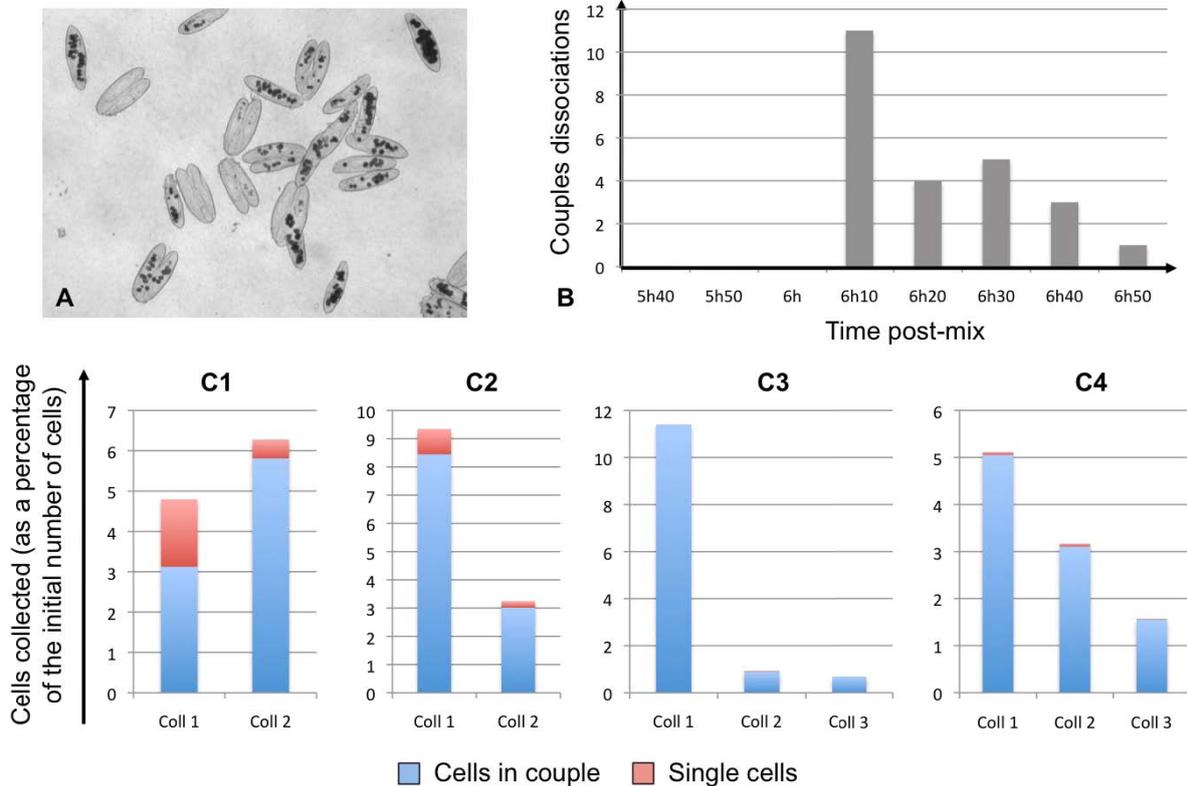


**Figure VIII.5 : Induction de la réactivité.**

Des cultures sont nourries avec du milieu frais puis centrifugées après un 10, 14 ou 24 h afin d'induire une réactivité sexuelle. Chaque courbe est une culture différente, donc d'une couleur différente, et le temps séparant l'alimentation de la centrifugation est indiqué juste à côté. La réactivité est évaluée de manière quantitative sur une échelle de 1 à 5, elle est présentée en fonction du temps dont le zéro est l'étape de centrifugation.

### **c. Efficacité et rendement**

L'ajout de billes magnétiques au bout de deux heures permet en 30 minutes d'obtenir une bonne discrimination entre les couples et les autres cellules (**Figure VIII.6 A**). Si certains couples contiennent des billes magnétiques, c'est probablement qu'ils se sont formés après l'ajout de billes, chaque partenaire ayant eu le temps d'ingérer des billes avant. Ces couples seront probablement piégés par l'aimant donc seuls les couples formés avant 2h seront récupérés après le tri magnétique. En se basant sur l'étape de dissociation, les couples collectés ont une synchronie de 50 minutes environ à 27 °C (**Figure VIII.6 B**). Divers tests effectués avec 50 mL de culture nous donnent un aperçu du rendement et de la pureté de la fraction collectée (**Figure VIII.6 C**). Avec des cellules réactives obtenues par nourrissage-centrifugation, il est possible de collecter (en trois fois) 10 à 15 % des cellules initiales dont plus de 99 % sont en couple. Le rendement reste faible mais ne pourra pas être amélioré car le paramètre limitant est le nombre de couples qui se forment dans la fenêtre de temps désiré. Ce paramètre dépend uniquement de la réactivité, il ne peut pas être modifié à moins d'utiliser un mutant *high reactivity* (Matsuda and Forney, 2006) ou une autre espèce présentant des taux d'accouplement plus élevés.

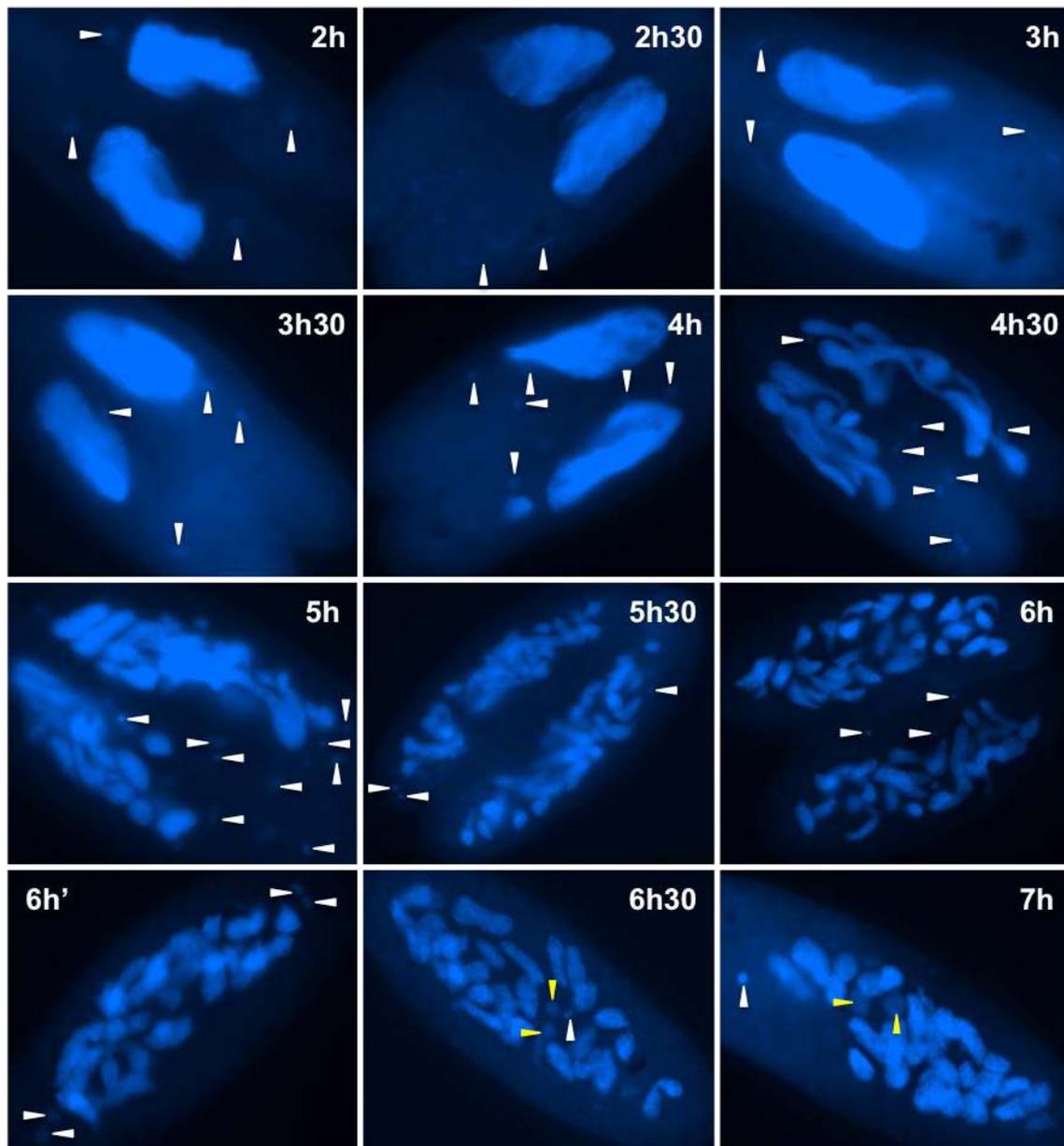


**Figure VIII.6 : Efficacité et rendement du tri magnétique de couples.**

(A) Observation au microscope des cellules 30 min après l'addition de billes magnétiques et 2h30 après le mélange de cellules réactives. (B) 24 des couples récupérés par tri magnétique (ajout des billes à 2h) sont isolés puis observés toutes les 10 minutes à partir de 5h40 afin de noter leur dissociation. Les observations répétées se font à température ambiante (22°C) par rapport au début de la conjugaison qui se passe à 27°C (C) Quantité et taux d'accouplement des cellules collectées par tri magnétique pour quatre expériences de conjugaison avec 50 mL de culture. La réactivité a été obtenue par centrifugation (C2 à C4) ou non (C1). Les billes magnétiques ont été ajoutées au bout de 2h (C2, C3) ou 2h30 (C1, C4). A l'issue d'une première collection (coll 1), une deuxième est réalisée (coll 2) en ajoutant du milieu épuisé puis en appliquant à nouveau l'aimant. Une troisième collection est réalisée pour C3 et C4.

#### 4) Résultats

A petite échelle, l'isolation de couples en quantité abondante m'a permis de réaliser une étude cinétique précise des premiers stades de la conjugaison à 27 °C (**Figure VIII.7**). La synchronie est d'environ 50 min d'après la fenêtre de séparation d'une trentaine de couples. Cette analyse est précieuse pour choisir les points de cinétique les plus pertinents. Par exemple, nous savons que les scnARN sont produits au stade croissant de méiose I donc il faut collecter les petits ARN à la fin de ce stade pour avoir la population initiale de scnARN. Si la technique utilisant les billes magnétiques a pu permettre d'isoler massivement des couples de *P. primaurelia*, je n'ai pas encore pu l'appliquer à grande échelle pour *P. tetraurelia* afin de réaliser l'étude cinétique. Ces données ne sauraient tarder et aideront grandement à déchiffrer la chorégraphie des différents acteurs de la voie du *genome scanning*.



**Figure VIII.7 : Cinétique d'autogamie suivie au DAPI**

Une conjugaison est lancée par mélange de cellules de types sexuels opposés. Au bout de 2 h, une première lame est montée puis des billes magnétiques sont ajoutées dans la culture. A 2h30, une deuxième lame est réalisée puis un tri magnétique permet de récupérer uniquement des couples. Des lames sont alors montées toutes les 30 min jusqu'à 7 h. Les flèches blanches et jaunes indiquent les MIC et les *anlagen*.

**MATERIEL**  
**&**  
**METHODES**

## A. Techniques de biologie cellulaire

### **Paramécies : souches, culture**

Les souches 51 et 32 utilisées sont entièrement homozygotes. Le mutant ND7 dérive du mutant décrit par Cohen et Skouri (Skouri and Cohen, 1997) après plusieurs croisements retour avec la souche 51. Les paramécies sont cultivées à 27 °C dans un milieu composé d'une infusion d'herbe de blé (WGP) tamponnéeensemencée par *Klebsiella pneumoniae* et additionnée extemporanément de  $\beta$ -sitostérol (0,8  $\mu\text{g}/\text{mL}$ ). Le contrôle de la croissance cellulaire se fait en utilisant le protocole d'isolement quotidien établi par Sonneborn en 1937 (**Figure S1**) afin de maîtriser l'âge des cellules, c'est-à-dire le nombre de divisions effectuées depuis le dernier événement de reproduction sexuelle. Après 5 jours de croissance à 27 °C, les paramécies ont effectué environ 20 divisions et peuvent donc toutes entrer en autogamie. Cette dernière est déclenchée par jeûne modéré. Pour vérifier qu'une dépression est 100 % autogame, une trentaine de cellules sont fixés au PAF (1 %) et marquées au DAPI (0,5  $\mu\text{g}/\text{mL}$ ) afin d'observer la fragmentation du macronoyau.

### **Croisements**

La conjugaison est obtenue par mélange de cellules de types sexuels opposés dans un état physiologique de réactivité déclenché par privation de nourriture chez de jeunes cellules post-autogames. Pour tester le type sexuel de cellules, on les rend réactives puis on les mélange avec des cellules réactives de chacun des deux types sexuels. Si on observe une réaction avec des cellules de type E cela veut dire qu'elles sont O et *vice versa*. Lorsqu'on s'intéresse à la descendance du croisement, on isole des couples puis lorsque ces derniers se séparent, on isole les ex-conjugants. La première division donne alors naissance à des caryonides frères, chacun héritant de l'un des deux macronoyaux distincts qui se sont développés dans le même cytoplasme. Pour chaque couple, on peut donc isoler jusqu'à 4 caryonides.

### **Test de létalité**

Lorsqu'une culture en dépression ou en masse présente 100 % de cellules autogames, 30 cellules sont isolées dans 250  $\mu\text{L}$  de milieu WGP frais. Après 4 jours à 27 °C, on observe l'état des cellules pour chaque dépression.

### **Extinction génique par ARNi lors de l'autogamie et de la conjugaison**

Afin d'inactiver un gène cible, des paramécies sont cultivées dans un milieuensemencé par des bactéries *E. coli* HT115 modifiées (Galvani and Sperling, 2002). Elles ont été transformées avec un vecteur plasmidique dans lequel on a inséré un fragment du gène d'intérêt entre deux promoteurs T7 convergents (**Figure S7**). Elles produisent alors des ARNdb qui sont ingérés par les paramécies et à partir desquels ces dernières vont produire des siARN. Le milieu est préparé en diluant une préculture de bactéries HT115 contenant le plasmide d'intérêt dans du milieu WGP. Le lendemain elles sont diluées au quart puis additionnées d'IPTG (0,4 mM) après doublement de la densité optique. Un temps d'incubation minimum de 4 heures à 37°C est respecté.

Pour éteindre les gènes désirés lors la conjugaison, des cellules jeunes sont nourries avec 2 à 3 volumes de milieu ARNi rendant alors les cellules réactives le lendemain. Les couples puis les ex-conjugants sont isolés dans du milieu WGP classique lorsqu'il s'agit de gènes précoces (*PTIWI*, *DCL2/3*) ou dans du milieu ARNi frais pour les gènes tardifs (*DCL5*). Les caryonides sont toujours isolés dans du milieu WGP. Pour éteindre les gènes cibles lors d'une conjugaison, une cinquantaine de cellules âgées est déposée dans une dépression remplie de milieu inhibiteur. Le surlendemain, lorsque qu'elles sont carencées et 100 % autogames, des cellules peuvent être isolées individuellement ou bien 20 cellules sont isolées pour une analyse 'en masse'.

### ***Marqueur ND7 et test de décharge des trichocystes***

Tous les croisements effectués impliquent un parent muté dans le gène *ND7* (*Non Discharge*). Ce dernier est impliqué dans l'exocytose des trichocystes, structures protéiques cristallines libérées en cas de stress. Lors de la conjugaison, l'échange réciproque des micronoyaux gamétiques n'a effectivement lieu que dans un couple sur 2. Pour le vérifier, les caryonides sont soumis au test tric : une trentaine de cellules est prélevée dans une dépression puis deux à trois gouttes d'acide picrique provoquant la décharge des trichocystes sont ajoutées sur le volume prélevé. Le résultat est observé au microscope avec un grossissement x 100. Les cellules sauvages apparaissent avec un manteau de trichocystes déchargés [tric+] alors que les cellules présentant la mutation paraissent « nues » [tric-]. S'il y a eu échange, les caryonides du couple sont hétérozygotes et donc [tric+] car la mutation *nd7* est récessive.

### ***Microinjections***

Les cellules sont injectées selon la procédure détaillée ici (J. Beisson et al., 2010). L'ADN est purifié, débarrassé des poussières par passage dans une colonne Ultrafree-MC et concentré à 3-5 mg/mL. Les cellules sont lavées dans un milieu Volvic-BSA avant l'injection.

### ***Montage et observation des lames***

Les cellules sont fixées 10 min dans du PAF 1 % en dépression. Le PAF est ensuite retirée puis on ajoute un mélange TBST (Tris 10 mM pH 7,4 ; NaCl 0,15 M ; Tween20 1 %, EGTA 0,1M ; MgCl<sub>2</sub> 2 mM) et DAPI (1 µg/mL). Au bout de 10 min, les cellules sont prélevées et mises sur une lame avec une goutte de Citifluor. Les lames sont scellées avec une lamelle et du vernis.

Les lames sont majoritairement observées en épifluorescence avec un microscope Nikon Eclipse Ti-S, une lampe à mercure et le logiciel Nis-Element. Le temps d'exposition choisi dépend de l'intensité de la fluorescence. Pour une cellule donnée, le signal DAPI est toujours observé en dernier puisque l'excitation par les UV peut induire une fluorescence non spécifique dans le vert et le rouge. Si cela est précisé, certaines lames sont observées en confocal avec un microscope *Leica TCS SP5*.

### ***Billes magnétiques et aimants***

Les billes, d'un diamètre de 500 nm, sont composées d'un cœur magnétique encapsulé dans un polymère portant des groupes carboxyles. Elles sont fabriquées par la société Ademtech qui les commercialise sous le nom de Carboxyl-Adembeads.

Elles sont conservées dans une solution aqueuse à 0,09 % d'azoture de sodium (NaN<sub>3</sub>). Avant utilisation, elles sont lavées 3 fois dans de l'eau. La quantité de billes suffisante pour les expériences réalisées est 1 mL de la solution commerciale pour 4 L de culture à 3 000 cellules/mL. Les aimants sont des cubes en néodyme de 12 mm de côté ayant une force d'adhérence de 6,3 kg. Ils sont produits par la société Wecraft et revendus par Supermagnete.

## B. Techniques de biologie moléculaire

### **PCR**

Les PCR sont majoritairement réalisées avec le kit Phusion (Thermo Scientific) et parfois avec le kit Long Range (Roche) ou le kit DynazymeII (Thermo Scientific). Les amorces utilisées sont listées dans la **table S1**.

### **Clonages**

Les produits PCR et le vecteur sont digérés par les enzymes requises en suivant les recommandations des fournisseurs. Le plasmide est ensuite déphosphorylé et les ligations sont réalisées sur la nuit à 18 °C. Les constructions GFP ou *Tomato* sont introduites par électroporation dans des bactéries compétentes *sure* (Agilent Technologies) alors que les constructions utilisées en ARNi sont directement introduites dans HT115.

### **Extractions d'ADN**

Pour extraire l'ADN de cultures en dépression, deux méthodes successives ont été utilisées. Dans la première méthode, 200 à 250 µL d'une culture saturée en dépression (100 à 200 cellules) sont mélangés avec un tampon de lyse (EDTA 0,5M-pH9 ; SDS 1 % ; Sarkosyl 1 % ; Protéinase K 1 mg/mL) préchauffé à 55 °C. Après 10 minutes de lyse, 250 µL de tampon PB du kit de purification PCR QIAquick sont ajoutés. La suite de l'extraction se fait selon les recommandations du kit. Une deuxième méthode utilise le kit GenElute Mammalian Genomic DNA Purification (SigmaAldrich) avec 200 µL de culture et la solution de lyse C.

### **Traitement au bisulfite**

Le traitement est effectué en utilisant le kit EpiTect Bisulfite (QIAGEN).

## C. Analyse informatique

### **Quantification des gels**

Les produits PCR utilisés pour mesurer la rétention d'une IES sont déposés sur un gel d'agarose 1,5 % sans BET en tampon TBE (Tris, Borate, EDTA) 1X. Après une migration suffisante, le gel est incubé 20 min dans un tampon TBE-BET (0,25 µg/µL), puis lavé deux fois 20 min en tampon TBE. Le gel est ensuite scanné avec un Typhoon FLA 7000 pour révéler la fluorescence du BET. Les bandes de l'image sont quantifiées avec le logiciel ImageQuant TL.

### ***Analyses statistiques***

Les différents tests statistiques sont réalisés avec le logiciel R. Pour comparer la rétention d'une IES dans les deux F1 d'un croisement ou la rétention des deux allèles dans le même F1, le test des signes de Wilcoxon est utilisé. Pour comparer la rétention dans un croisement donné par rapport à un autre croisement, le test de Kolmogorov-Smirnov à deux échantillons est utilisé. Enfin, les coefficients de corrélation et les tests associés sont réalisés avec la méthode de Spearman. Dans tous les tests, les deux paramètres comparés sont dit significativement différents lorsque la p-value est inférieure à 0,01 ou encore moins si cela est précisé.

### ***Traitement des images***

Les flèches et la légende de toutes les images sont ajoutées manuellement. Les images de microscopie confocale sont des vues de reconstructions 3D réalisées avec le logiciel Icy à partir des stacks.



# **DISCUSSION**

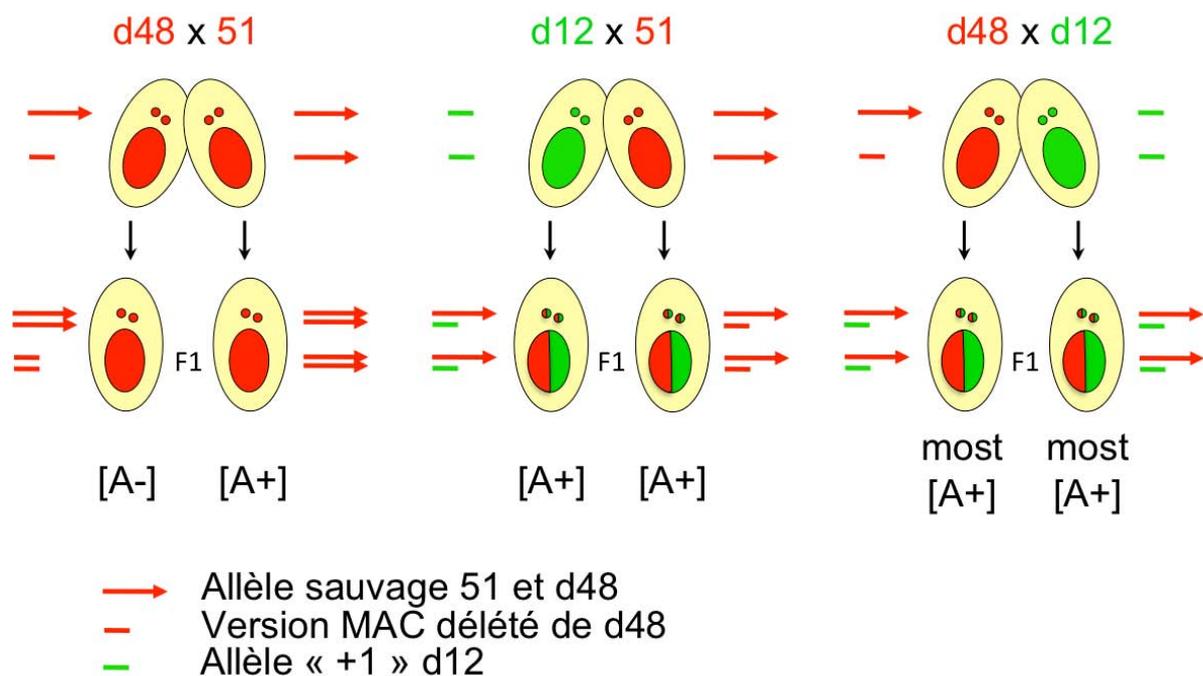
## A. Etendue des défauts de réarrangements

### 1) F1 51/32 hémizygotes pour des IES

L'idée d'une forme de dysgenèse hybride est née de la découverte de la rétention de l'IES de mtB (134 pb) chez 30 % des F1 51/32. Cette rétention était due, pensait-on, au défaut de scnARN maternels produits par la cellule 32. En réalité, j'ai montré que ce n'est pas le cas puisque les scnARN paternels peuvent diriger efficacement l'excision de l'IES dans la cellule 32. Toutefois, il existe bien un défaut de réarrangement qui affecte 30 % des F1 indépendamment de leur origine maternelle. Il semble que l'on ne puisse l'attribuer qu'à la situation hémizygote du noyau zygotique pour le gène mtB. C'est un phénomène totalement nouveau que nous avons appelé « effet hémizygote ». L'étude de l'IES 101 (99 pb, voir V-E), spécifique de la souche 32, montre un effet avec des taux de rétentions plus faibles. Les 5 autres cas de polymorphismes absence/présence concernent de petites IES qui sont parfaitement excisées dans les F1 51/32. Ces données nous suggèrent que la taille pourrait être un facteur important dans le déclenchement de l'effet. A moins que cette tendance ne reflète plutôt le degré de contrôle maternel. En effet, les IES les plus longues sont plus sensibles à la déplétion de dcl2/3, Ezl1 ou encore TFIS4 (Lhuillier-Akakpo et al., 2014; Maliszewska-Olejniczak et al., 2015; Sandoval et al., 2014).

### 2) F1 d48/d12 hémizygote pour le gène A

Un phénomène similaire a été décrit depuis longtemps dans le croisement d48 x d12 (Rudman et al., 1991) (**Figure X.1**). La souche d48 possède une délétion macronucléaire terminale du gène A dans son MAC. Il s'agit d'une délétion héritée maternellement qui suit le modèle du *genome scanning*, on peut parler d'épimutation. La souche d12 porte un allèle avec une délétion du même ordre de taille, il s'agit d'une mutation mendélienne. On se situe à l'échelle du gène et non d'une simple IES comme dans le cas de mtB. Le résultat du croisement d48 x d12 montre que les deux F1 sont souvent de phénotype [A+], ce qui signifie que la délétion macronucléaire est difficilement réalisée dans ce contexte hémizygote. La rétention observée semble plus fréquente que pour l'IES de mtB car elle concerne 6 F1 sur 7 dérivant de d48 et 7 F1 sur 8 dérivant de d12. C'est toutefois difficile à comparer car il s'agit un test d'expression du gène A et on ne sais pas si elle nécessite un grand nombre de copies fonctionnelles. En tout cas, il est probable que l'effet hémizygote affecte toutes les régions et pas seulement les IES. On peut également supposer que l'inhibition de l'excision est d'autant plus forte que la région est longue. Le tester reste difficile puisqu'il n'est pas possible de créer des délétions MIC chez *Paramecium*.



**Figure X.1 : Devenir du gène A dans les croisements impliquant 51, d48 et d12.**

Les trois souches 51, d48, et d12 ont le même fond génétiques mais possèdent des combinaisons de génomes MIC et MAC différentes concernant le gène A (flèche). 51 et d48 possèdent un allèle sauvage mais il est normalement retenu au cours du développement chez 51 alors qu'il est tronqué par délétion héréditaire chez d48. La souche d12 possède un allèle tronqué « +1 ». Sous chaque F1 est indiqué son phénotype, il dépend uniquement du MAC. Interprétation de (Rudman et al., 1991).

### 3) F1 51/32 hétérozygotes pour des IES divergentes

Mon analyse de l'IES G4404 montre qu'il y a très peu de rétention dans les F1 51/32 malgré une différence significative par rapport au croisement 51 x 51 (Figure VI.4). Ces défauts d'excision concernent les deux allèles et tous les F1 quelle que soit leur origine maternelle. Finalement, ce phénomène ressemble à la rétention des IES dans les F1 hémizygotés. On peut donc supposer que l'effet hémizygoté s'applique peut-être dans une moindre mesure pour les allèles divergents. Il est impossible de prédire si une faible divergence entre allèles pourrait générer des défauts d'excision chez des F1 hétérozygotés. Etudier d'autres cas est indispensable pour mieux connaître ce phénomène.

### 4) Etude à l'échelle du génome.

La suite logique de mon travail sur ces quelques IES est de mesurer l'étendue des défauts de réarrangements chez les F1 51/32. Dans ce but, j'ai déjà réalisé deux pools de F1 51/32 dérivant du parent 51 et du parent 32 dont le séquençage est en cours. L'analyse de cette expérience permettra tout d'abord de généraliser l'effet hémizygoté. En plus des 7 polymorphismes d'insertion connus, on espère trouver d'autres cas grâce à la comparaison du génome 51 avec un assemblage récent *de novo* du génome 32. Si ce nombre est suffisant, nous saurons si l'effet hémizygoté concerne uniquement des IES contrôlées maternellement, c'est-à-dire l'ensemble des IES sensibles à la déplétion de Nowa1. Ces données existent pour la souche 51 mais pas

pour la souche 32. Deuxièmement, l'analyse du séquençage nous permettra d'évaluer les conséquences de la divergence des allèles sur la rétention des IES en F1. L'IES G4404 est un cas singulier dans lequel les deux allèles ne s'alignent pas tant la divergence est grande. Nous verrons s'il existe un effet pour des cas plus modérés.

## B. Mécanisme des défauts de réarrangements

### 1) Une quantité limitée de scnARN ?

La découverte de la rétention de l'IES de mtB dans des F1 dérivant du parent 51 a été surprenante. Puisque les scnARN homologues à l'IES sont produits dans ce parent, ils devraient être en abondance et exciser totalement l'IES dans les *anlagen*. Une première hypothèse pour expliquer l'effet hémizygote est un raisonnement quantitatif. Dans le cadre d'un partage total des scnARN entre partenaires, on pourrait dire que les scnARN sont parfois en quantité limitée et ne peuvent pas exciser correctement l'IES dans au moins un des partenaires. Cette hypothèse pourrait expliquer la faible corrélation négative observée pour la rétention de l'IES de mtB dans un F1 par rapport à l'autre (Figure 5.2D). Néanmoins, cette notion de quantité limitée de scnARN est abstraite, elle n'a jamais été proposée dans d'autres cas. Dans le croisement contrôle 51 x 51, la quantité n'est jamais limitante : il y a deux fois plus de scnARN car ils sont produits par les deux parents mais aussi deux fois plus de copies à exciser. Et cela marche dans 100 % des cas.

### 2) Détection de l'état hémizygote par appariement des chromosomes

Dans le MSUD décrit chez *Neurospora* (voir I-C-3), l'appariement des chromosomes homologues en méiose permet de détecter les situations hémizygotiques. L'association intime des chromosomes homologues condensés (*synapsis*) en prophase I semble en effet propice à la détection des régions non appariées. Pourtant, ni cette association, ni la formation de *crossover* n'est requise pour le MSUD d'après l'analyse de mutants (Kelly and Aramayo, 2007). Ces données suggèrent que la détection (*trans-sensing*) a lieu avant, lors de l'alignement instable des chromosomes. La détection de l'homologie entre chromosomes est en réalité peu connue. Une des protéines impliquées dans la voie MSUD est Sad-6, elle est homologue à la protéine Rad54 humaine qui intervient dans la recombinaison. Elle permet l'invasion d'un brin d'ADN dans une région homologue située sur un autre chromosome. Récemment, une étude a montré que la transcription simultanée d'un ARN non codant sur les chromosomes II de *S. Pombe* servait de point d'ancrage à l'association de ces chromosomes homologues en tout début de méiose (Ding et al., 2012) Ce processus est indépendant de la recombinaison homologue. Finalement, le mécanisme de détection initiale de l'homologie en méiose n'est pas très bien connu. D'ailleurs, l'appariement entre régions homologues n'est pas limité à ce stade. Chez les mammifères par exemple, la recombinaison V(D)J, l'inactivation de l'X ou encore les phénomènes de paramutation/empreinte impliquent des interactions homologues (Apte and Meller, 2012; Brandt et al., 2010; Marahrens, 1999). Dans le cas extrême des diptères, tous les chromosomes sont appariés de bout en bout dans les tissus somatiques (Stevens, 1908).

Chez la plupart des ciliés, on observe une prophase avec un stade « croissant » dans lequel les centromères et les télomères se concentrent dans les extrémités opposés des noyaux allongés (Loidl and Scherthan, 2004). Cette particularité faciliterait l'alignement des chromosomes homologues pour compenser l'absence de complexe synaptonémal chez ces organismes. Cela suggère l'existence d'une machinerie importante permettant l'alignement des chromosomes homologues. Cet équipement serait crucial pour permettre la détection des régions hétérozygotes dans notre cas mais il doit pour cela être actif au moment du développement des *anlagen*. En effet, l'inhibition de l'excision que j'ai mise en évidence intervient lorsque le noyau zygotique est hémizygote, c'est-à-dire après la méiose et la fécondation. Pourtant, l'idée que l'état hémizygote puisse être détecté pendant le développement n'est pas aberrante au vu de la diversité des mécanismes évoqués précédemment. D'ailleurs, un système artificiel chez la drosophile a montré que cela était possible (Duan et al., 2013). En outre, il existe un stade polytène dans les *anlagen* des spirotriches qui pourrait favoriser la reconnaissance des régions homologues (Prescott, 1994).

### 3) Détection de l'état hémizygote par des ARN

Chez les ciliés, le modèle du *genome scanning* est une comparaison des génomes MIC et MAC. Ce mécanisme se fait indirectement par recherche d'homologie entre des copies ARN de ces génomes. Si ce système existe, peut-on imaginer un deuxième dialogue d'ARN à l'échelle des *anlagen* pour permettre la détection de l'hémizygotie ? A ce stade, les seuls petits ARN détectés sont les scnARN et les iesARN mais il y a aussi des transcrits naissants qui permettent le ciblage par ces ARN. Si les chromosomes homologues sont proches spatialement, peut être que leurs transcrits naissants pourraient s'apparier de manière complémentaire. Si un des allèles présente une séquence ectopique, alors la boucle formée pourrait être détectée.

### 4) Mécanisme d'inhibition de l'excision

Quel que soit le mode de détection, il reste à comprendre comment il peut générer environ 30 % de forte rétention dans le cas de mtB. Dans le cas du MSUD, la séquence non appariée génère un ARN aberrant qui est converti en ARNdb par une RdRP. L'action de protéines *Dicer* crée des petits ARN pris en charge par une protéine Argonaute-like afin d'induire la répression transcriptionnelle. Il a été proposé que ce mécanisme soit un système de défense face aux insertions de transposons. Bien qu'il soit confiné à la méiose, il est probable que si les séquences non appariées pouvaient être détectées à un autre moment, la voie pourrait également s'enclencher. Pour appuyer cette idée, des chercheurs ont utilisé une RdRP de *C. elegans* capable de réprimer les régions chromosomiques non appariés à la méiose. Ils ont introduit cette protéine ainsi qu'un transgène chez la drosophile. L'appariement des chromosomes dans les mitoses embryonnaires de ce diptère permet alors le déclenchement de la répression du transgène non apparié (Duan et al., 2013). Dans notre cas, s'il y a formation de petits ARN ciblant la séquence ectopique pendant le développement des *anlagen*, alors ils pourraient provoquer la dégradation des transcrits. En conséquence, les scnARN ne parviendraient pas à s'apparier sur ces transcrits pour guider l'élimination de l'ADN. La compétition entre les siARN et les scnARN pourrait être à l'origine de la variabilité de la rétention observée. Finalement, une voie similaire au MSUD est tout à fait envisageable. Les acteurs n'auraient pas été

détectés jusqu'à présent puisque l'étude des réarrangements se fait traditionnellement pendant l'autogamie de souches entièrement homozygotes. Néanmoins, l'intérêt de conserver un tel mécanisme est étrange. La rétention d'une IES est souvent délétère pour la cellule puisqu'elle peut conduire à l'interruption d'un gène. La rétention d'une région plus grande, typiquement suite à l'insertion d'un transposon, est encore plus grave puisque cela pourrait conduire à la prolifération du transposon dans le MAC. Enfin, la perte différentielle des ohnologues dans des populations En conclusion, si on peut envisager un mécanisme similaire au MSUD pour l'effet hémizygote, ses conséquences vis-à-vis des transposons sont totalement opposés.

## 5) Implication défaut de scnARN maternels

Notre prédiction de défaut de scnARN maternel s'est révélée erronée au vu de la rétention de mtB et de G4404 dans les F1 hétérozygotes (Figure V.2 et VI.4). Néanmoins, quelques données subsistent en faveur du modèle initial. Tout d'abord, les observations de Brygoo étaient assez claires (Brygoo, 1977b) : il voyait très peu de rétention de l'IES de mtB (interprétation actuelle du phénotype O\* en F2) dans les F1 dérivant du parent 51. Malgré l'observation indépendante dans deux laboratoires d'une rétention aussi forte chez le parent 51 que le parent 32, le doute subsiste. Deuxièmement, le profil de rétention des allèles de G4404 chez les F1 51/32, montre que l'allèle 51G4404 est davantage retenu dans le F1 dérivant du parent 32, conformément à notre prédiction. L'allèle 32G4404 n'est pas retenu plus fortement dans le F1 dérivant du parent 51 mais il est possible que cet allèle ne soit pas suffisamment soumis au contrôle maternel (cf. VII-C) pour qu'il y ait un effet notable. Enfin, l'analyse de l'IES 101 (Figure V.8) révèle également une rétention plus forte dans le F1 dérivant du parent 51 naïf que dans l'autre. Dans l'ensemble, notre modèle de défaut de scnARN maternels traduit très mal les résultats obtenus mais il colle à certaines observations difficilement explicables autrement.

## C. Un modèle avec des scnARN paternels actifs

### 1) Un échange de cytoplasme ?

Les différents résultats que j'ai obtenus peuvent être assez facilement expliqués par un modèle d'échange des scnARN entre conjugants et le processus de sélection par appariement sur les transcrits macronucléaires. L'observation de la diffusion de fluorescence d'un partenaire à l'autre montre que la protéine Ptiwio1 est capable de migrer durant la conjugaison (Figure VII.6D). Il peut s'agir simplement d'un processus passif de diffusion ou bien un processus actif dont le rôle serait d'assurer le réarrangement des allèles échangés. Rappelons qu'il y a forcément un minimum d'échange cytoplasmique permettant le transfert des noyaux gamétiques mâles (Inaba et al., 1966; Pape, R., 1988), et encore, environ un couple sur deux n'échange pas réciproquement ses noyaux chez *Paramecium tetraurelia*. A l'inverse, l'induction d'un large pont cytoplasmique annule l'hérédité maternelle du type sexuel, les deux ex-conjugants devenant E (Figure IV.4). Il y aurait donc une quantité limitée de cytoplasme qui serait échangé au cours de la conjugaison chez *P. tetraurelia*. Cette quantité serait suffisante pour autoriser l'échange de scnARN mais insuffisante pour

que tous les scnARN produit par une cellule puissent être titrés par les transcrit du MAC de son partenaire.

## 2) Une empreinte gamétique ?

L'hypothèse d'une empreinte déposée par les scnARN dans les gamètes est séduisante. Elle s'affranchit de l'échange cytoplasmique et permet une discrimination allélique claire. En revanche, mes résultats ne permettent pas d'étayer cette hypothèse. Tout d'abord, les résultats des croisements avec l'injection des deux allèles ou plus simplement l'hérédité du type sexuel montrent qu'il ne peut pas y avoir d'empreinte irréversible (voir paradoxe mtA). Et dans ce cas, il faudrait ajouter un nouveau mécanisme pour expliquer la réversion, plus puissant, rendant inutile l'existence d'une empreinte. L'excision des allèles identiques après déplétion uniparentale de Dcl2/3 s'oppose également à l'hypothèse de l'empreinte (voir VII-E). De plus, aucune protéine Piwi n'a pu être localisée dans les noyaux gamétiques ou zygotiques. Et l'hypothèse requiert une sélection très rapide des scnARN afin que l'information puisse être transmise dans ces noyaux.

Quant à la nature de cette empreinte, nous n'avons aucune information. Les approches pour identifier la méthylation de l'ADN n'ont pas permis d'observer une discrimination entre les IES et les séquences purement macronucléaires. Enfin, d'un point de vue théorique, si la marque déposée par les scnARN n'est pas répliquable, alors elle ne sera présente que sur une copie parmi les 8 ou les 16 copies au début de la phase d'excision. Ce fait rendrait indispensable une voie d'amplification. Cela pourrait être le rôle des iesARN produits par Dcl5 mais nous n'avons presque rien trouvé de concluant dans cette direction. Si ce ne sont pas les iesARN, il faudrait alors que les scnARN redéposent la marque à chaque cycle d'endoréplication donc ils devraient accompagner le noyau gamétique, ce qui revient à résumer l'hypothèse à un échange cytoplasmique. Imaginons plutôt que l'empreinte gamétique soit répliquable ou auto-entretenu d'une manière ou d'une autre. Puisque les scnARN sont requis pour l'excision de ces IES, cette empreinte gamétique ne serait pas maintenue, en revanche, pendant la différenciation des nouveaux MIC à partir du noyau zygotique, ou à travers la mitose des MIC pendant la croissance végétative. Finalement, les données ne permettent pas d'étayer l'hypothèse de l'empreinte et sa nature reste entièrement spéculative.

## D. Que reste-t-il de la dysgenèse hybride ?

### 1) Défauts d'excision des IES et létalité

Quelle que soit leur origine, les polymorphismes d'insertion d'IES sont des sources de défauts de réarrangements. Mes résultats montrent qu'il s'agit au mieux de 30 % avec l'exemple de l'IES de mtB. La rétention d'une IES intragénique produira dans la majorité des cas un décalage du cadre de lecture et donc un allèle non fonctionnel. (IES 3n sous représentées?). Pour le moment, aucun lien direct entre un défaut de réarrangement et la létalité en F2 du croisement 51 x 32 (80 %) n'a été montré. Le gène mtB n'est pas essentiel et on ignore si le gène portant l'IES 101 est important. Peut-être qu'on ne pourra jamais attribuer une partie de la létalité à un locus en particulier mais la multiplicité des IES retenues peut avoir un impact considérable. Chez *P. tetraurelia*, un gène sur deux est interrompu par une IES (Arnaiz et al., 2012)

donc le risque est grand. D'un autre côté, la moitié des gènes ont un orthologue WGD1 ce qui diminue la probabilité qu'un gène donné soit essentiel. Chez les hétérozygotes F1, on pense que la rétention d'une IES dans un allèle n'a pas d'effet car il reste l'autre allèle. Pourtant, des individus hétérozygotes pour deux gènes impliqués dans la même voie peuvent révéler un phénotype par additivité. Il est vrai que les défauts de réarrangements en F1 seront plutôt délétères pour les F2 qui en héritent. Ils pourront même être aggravés s'ils concernent effectivement des IES contrôlées maternellement comme nous le pensons pour le moment. A partir d'un certain seuil de rétention en F1, la rétention sera totale en F2. Afin de rechercher une cause de la létalité des F2 51/32 sans a priori sur les mécanismes, nous avons récemment lancé le séquençage d'une quarantaine de F2 survivants. Cela devrait nous permettre d'identifier des *loci* présents sous un seul allèle. L'absence de l'autre allèle pourrait traduire un défaut de réarrangement affectant la viabilité des F2 qui en héritent. Une autre cause possible est l'incompatibilité cytonucléaire qui est une déclinaison du modèle de spéciation Dobzhansky-Muller (Chou and Leu, 2015). L'origine des loci monoalléliques devra donc être vérifiée expérimentalement. Pour aller plus loin et selon les résultats obtenus, nous verrons s'il est possible d'étendre l'étude à des croisements interspécifiques. Ainsi nous pourrions savoir si les phénomènes mis en évidence ont pu contribuer à la spéciation du groupe *P. aurelia*.

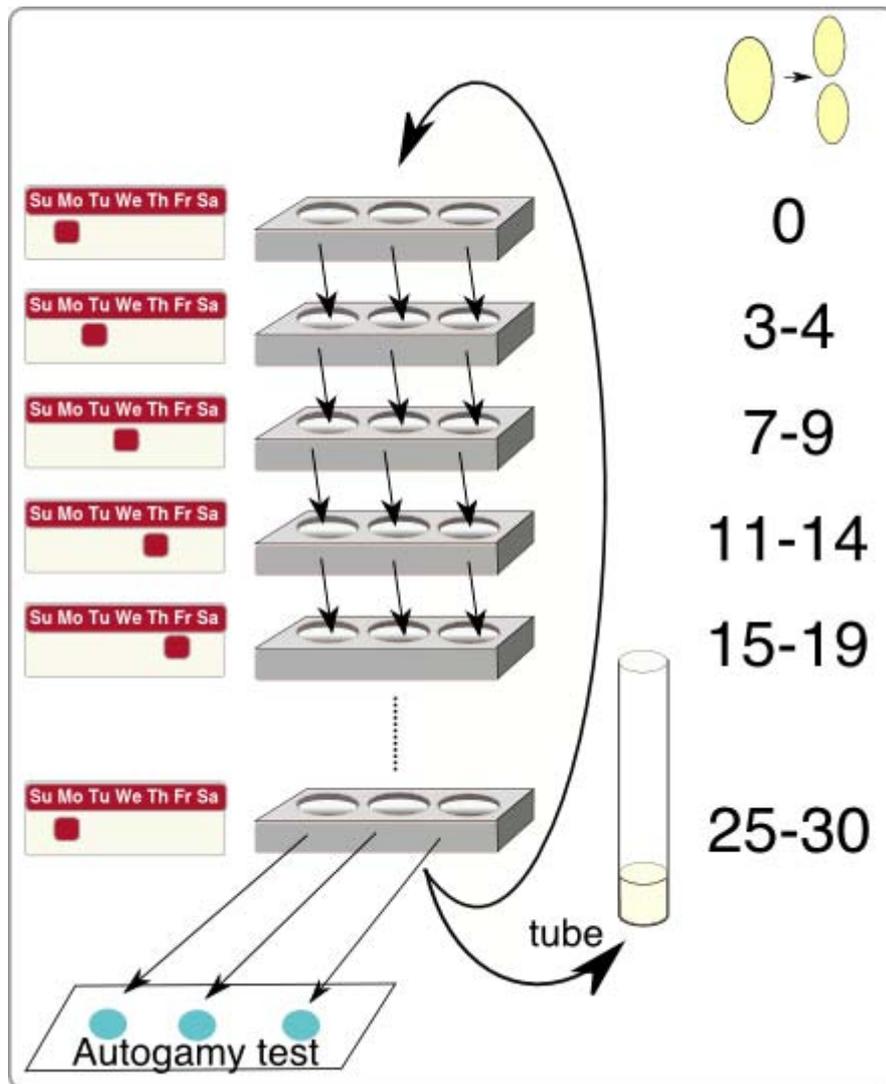
## 2) Et les transposons ?

Le phénomène de dysgenèse hybride chez la drosophile implique des transposons, or nous nous sommes concentrés sur les IES qui ne sont que des vestiges de transposons. En réalité, c'est la découverte de la rétention de l'IES de mtB qui nous a rapprochée de la dysgenèse, d'où l'étude focalisée sur les IES. Dans un deuxième temps, nous pourrions voir si les transposons eux-mêmes sont impliqués dans le système que nous avons décrit chez *Paramecium tetraurelia*. Le catalogue des transposons commence tout juste à être établi grâce à l'assemblage récent du MIC. Il n'existe pas encore de répertoires pour différentes souches ou espèces afin d'étudier leur impact sur des hybrides. De plus, nous ne savons pas s'il existe des transposons fonctionnels dans les MIC. Les premières recherches de transposons par blast n'ont pas permis d'identifier des copies multiples d'un élément pour témoigner d'une activité. Leur répression continue dans les MIC devrait très vite les faire dégénérer d'où l'impossibilité de prédire si leur rétention dans le MAC conduirait à leur prolifération.

## 3) Dysgenèse suite à une invasion de transposon chez *Paramecium*

Mon étude porte sur les polymorphismes absence/présence qui existent dans les souches aujourd'hui. Pour les polymorphismes qui sont apparues par perte des IES, les conséquences observées sur la descendance hybride actuelle sont proches de la situation au moment où les polymorphismes sont apparus. Les cas d'insertion sont très différents puisqu'ils ne peuvent concerner que des éléments mobiles, dont certains codant une transposase fonctionnelle. Dans ce cadre, la rétention d'éléments autonomes dans les hybrides, par l'effet hémizygote, serait beaucoup plus grave car elle pourrait amener à leur expression macronucléaire et à une prolifération sans le MIC de tous les éléments qu'ils reconnaissent. Cette idée suggère que la dysgenèse observée aujourd'hui avec des vestiges de transposons sous-estime largement l'impact d'une invasion de transposons dans une population de paramécies.

# **ANNEXES**



**Figure S1 : Conditions de culture de *Paramecium tetraurelia*.**

Pour mettre en culture une souche à partir d'un stock, 3 cellules sont isolées chacune dans une dépression contenant 250  $\mu\text{L}$  de milieu. Le jour suivant, les cellules sont comptées pour évaluer le nombre de divisions et chaque clone est répliqué par transfert d'une cellule unique dans une nouvelle dépression. On ajoute 750  $\mu\text{L}$  de milieu dans les dépressions initiales ce qui permettra d'atteindre 1000 cellules dans 1 mL. Si le cycle débute un lundi alors les duplicats transférés le vendredi sont laissés dans 1mL tout le week-end. Le lundi suivant, l'autogamie de ces dépressions est évaluée par un marquage au DAPI puis observation au microscope à fluorescence. Dans les dépressions 100 % autogames, les cellules peuvent être utilisées pour commencer un nouveau cycle et le reste est transféré dans un tube de 20 mL avec 4 mL de milieu frais. Ce tube est incubé une nuit à 27 °C puis peut être stocké à basse température (14 °C) quelques semaines avant utilisation. (Janine Beisson et al., 2010)

51MIC-mtB ATGTTTATAAGCACACTAACTTGTTATTAGAACTAAGAAAAGGCACTCTAGAATTAGAC  
32MIC-mtB ATGTTTATAAGCACACTAACTTGTTATTAGAACTAAGAAAAGGCACTCTAGAATTAGAC  
\*\*\*\*\*

51MIC-mtB ATCATAAAATAGCATTTATAATAGCATAGTTTAAATGAATGAAATAGAAATTAATGCATTG  
32MIC-mtB ATCATAAAATAGCATTTATAATAGCATAGTTTAAATGAATGAAATAGAAATTAATGCATTG  
\*\*\*\*\*

51MIC-mtB ATGGACTTAACAAAATAAATAAAAGAATCAATCACTGATGATGAATAGAGGGTTGGGATT  
32MIC-mtB ATGGACTTAACAAAATAAATAAAAGAATCAATCACTGATGATGAATAGAGGGTTGGGATT  
\*\*\*\*\*

51MIC-mtB CTAAATGAAATTGATGCTATTTTATTAAATATATTGTAAGAAAAAGAGAGACCAATTCT  
32MIC-mtB CTAAATGAAATTGATGCTATTTTATTAAATATATTGTAAGAAAAAGAGAGACCAATTCT  
\*\*\*\*\*

51MIC-mtB AATACTGATAACTTTTTAAATTATGTACAATAAAGCAATTAATAATAATATCGGTATAGCA  
32MIC-mtB AATACTGATAACTTTTTAAATTATGTACAATAAAGCAATTAATAATAATATCGGTATAGCA  
\*\*\*\*\*

51MIC-mtB ATACTTAACAATTATGGCCAATTCAGTAAGATTATAACACAATATATCTAGTCTTGACTG  
32MIC-mtB ATACTTAACAATTATGGCCAATTCAGTAAGATTATAACACAATATATCTAGTCTTGACTG  
\*\*\*\*\*

51MIC-mtB ATCAACTTACAAGAAATATTCTTGAATTAATAATATTGAAATTAGAAGCAGTTAACTTTT  
32MIC-mtB ATCAACTTACAAGAAATATTCTTGAATTAATAATATTGAAATTAGAAGCAGTTAACTTTT  
\*\*\*\*\*

51MIC-mtB TTTCATTAATCAGCAAAGTCTCTTAAAATATGTTGAGGGAAAAATTTAATGACTGCTGTC  
32MIC-mtB TTTCATTAATCAGCAAAGTCTCTTAAAATATGTTGAGGGAAAAATTTAATGACTGCTGTC  
\*\*\*\*\*

51MIC-mtB TATTACAGTAGAGTAATGTGAAAGGAGATAAAATAAAATTATTTGTATTTAATCTTCTTCT  
32MIC-mtB TATT-----  
\*\*\*\*

51MIC-mtB GTGGAATGTCTTTTGAATTTCTTCGATTAGCAAATTAACAAATTTAGAACTATTTAAAAA  
32MIC-mtB -----

51MIC-mtB AAGACTGCTTCCTATTGTATATAAAGGGGAATGCAAAGTAAACCTTTTAATTGGCTATGTA  
32MIC-mtB -----ATAAAGGGAAATGCAAAGGAAACCTTTTAATTGGCTATGTA  
\*\*\*\*\*

51MIC-mtB TTCAAAGAGAAATAAAAAAAAAAAGTATTTAGTATTTGAAACAATTTGCTAAAGAAAAAGA  
32MIC-mtB TTCAAAGAGAAATAAAAAAAAAAAGTATTTAGTATTTGAAACAATTTGCTAAAGAAAAAGA  
\*\*\*\*\*

51MIC-mtB AAAGAGGAAATTAATAACTCTTTAGCGAAAAGTAAATTATTATCATTACATCGCAGATA  
32MIC-mtB AAAGAGGAAATTAATAACTCTTTAGCGAAAAGTAAATTATTATCATTACATCGCAGATA  
\*\*\*\*\*

51MIC-mtB CAGATGCAGAACCAATATAATAAGAGCTAATCTTAATGGCAAAGTACTTAAAATCGCTGT  
32MIC-mtB CAGATGCAGAACCAATATAATAAGAGCTAATCTTAATGGCAAAGTACTTAAAATCGCTGT  
\*\*\*\*\*

51MIC-mtB AAGTCACAATATCAAAGACGAATCTATAATTACATGATGATTTTATAGAATCTTTCAGAG  
32MIC-mtB AAGTCACAATATCAAAGACGAATCTATAATTACATGATGATTTTATAGAATCTTTCAGAG

```

*****
51MIC-mtB  ATTCAAATGATATACTACTACACAATCTATCTTCACTAATACCAAATTAATTAAACTAAG
32MIC-mtB  ATTCAAATGATATACTACTACACAATCTATCTTCACTAATACCAAATTAATTAAACTAAG
*****

51MIC-mtB  TGGCTGTCTGTGAGATATATGAATTGGATTAGCATCTAAATTTTCAGTGTTGAAGCCTTGT
32MIC-mtB  TGGCTGTCTGTGAGATATATGAATTGGATTAGCATCTAAATTTTCAGTGTTGAAGCCTTGT
*****

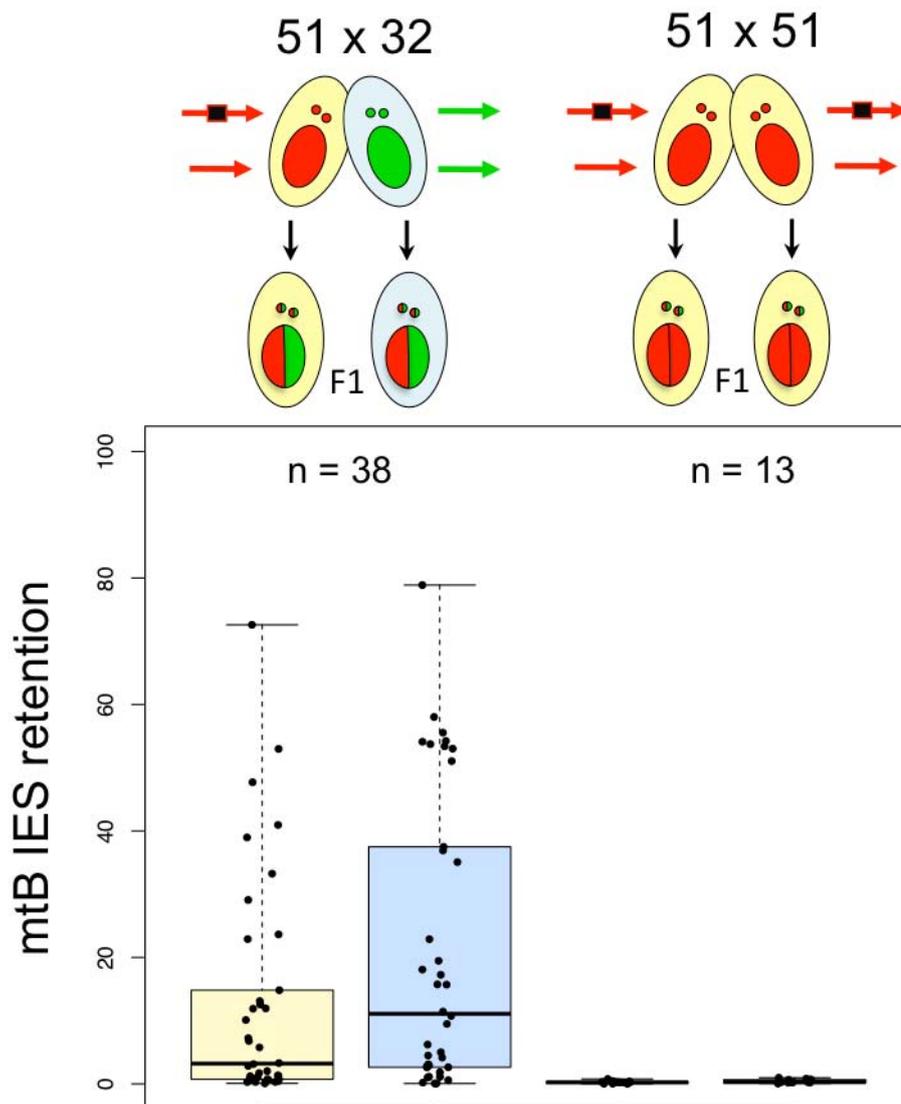
51MIC-mtB  TATCTGATCCTTTCATTAAGAAGAATGAAGGTATTATGTAATATGGATATTTATAGTAAA
32MIC-mtB  TATCTGATCCTTTCATTAAGAAGAATGAAGGTATTATGTAATATGGATATTTATAGTAAA
*****

51MIC-mtB  GTGGAGTAAATTAGTTAGGAAATTGGGTGGTTAGGAGCCGACTTATGAAGACTTTATTAT
32MIC-mtB  GTGGAGTAAATTAGTTAGGAAATTGGGTGGTTAGGAGCCGACTTATGAAGACTTTATTAT
*****

51MIC-mtB  ATGA
32MIC-mtB  ATGA
*****

```

**Figure S2 : Alignement MUSCLE (3.8) des gènes MIC mtB des souches 51 et 32.**  
L'IES de mtB est surlignée en jaune et les substitutions sont annotées d'un rectangle rouge.



**Figure S3 : Excision de l'IES de mtB chez des F1 51/32 et 51/51.**

La souche 51 (génomme rouge) est croisée avec la souche 32 (génomme vert) (A) ou avec elle-même (B). Le calcul du taux de rétention de l'IES de mtB se fait par quantification d'une PCR encadrant l'IES (voir V-2). Seul l'allèle 51, possédant une IES dans sa version MIC, peut retenir l'IES dans sa version MAC. Par opposition, l'allèle 32 (vert) n'existe que sous la forme IES-. Le taux de rétention calculé est présenté par ex-conjuguat sous la forme de boxplots avec affichage de tous les points.

51G4404 1 GGAAGTGGAT**TTGGT**AGCTGCAACAT**GTGCTGC**ATATAAT**GTAGG**TTGTATTAATTT**TAAAA**  
32G4404 1 GGAAGTGGAC**TAAC**AGCTGCAAT**TTGCGCA**AATTAT**CATGCAA**TTGTATTAATTT**CAAAA**

51G4404 61 GATGGTATT**ATAAA****TCGAAT**TAT**TAA---**TT**CTGA**ATAGATTT**CTAAT****TCTTGA**ATA**ACA**  
32G4404 61 GATGGTATT**GTAA****ACTTAA**AAT**ACA**ATTT**TTGA**ATAG**CTTT****TAGATA**AAA**AGTAG**AT**T**

51G4404 118 **CGTATG**ATT**TAATG****CTC**AT**T**CAT**CCAA**ATT**AGCC**TTT**GT**TATA**---**AT**TCAA**TTT**C**AAAA  
32G4404 121 **CTAT**TATT**AAATG--****TAAT****TGTT**AAATT**GAAT**TTT**TCTC**ATTTT**TAT**TTT**C**TAAA

51G4404 175 **G**TAGT**CCAAAA**ACT**TGAA-----**AATAT**TTCAA**AAA**---**ACAAT**ATATA**TT**TACT**GT  
32G4404 179 **TTGT**TCT**AAAA**ACT**TAA**ATT**AGAG**TTAATATA**ATAGAA**TAT**CA**CAAT**-TAG**TT**GC**CT**T**

51G4404 223 **ATA**ATATTTT**CCA**ATATTT**TG**TAATT**ACC**T**AAAA**TTAAT**TCAA**CTTTT**TAAAG---**TTA  
32G4404 238 **GT---**TTTT**CTTT**TAT**GTTT**TATT**CA**CA**AGAA**CAAT**---**ATATTTT**AAAA**ATTTA

51G4404 280 **CTTA**ATAAT**ACCG**CTGTT**AGGA**AG**CCAA**AGCT**GG**TTG**CAA**AGATT**ACTCC**ACATA**AAAA**  
32G4404 291 **CTTA**ATAAT**ACTG**CTT**GTTA**AGAA**TCAA**AGCT**ACT**TG**CAA**AGATT**ACG**ACATA**AAAA**

51G4404 340 AAATGCT**TCG**CATA**ATCA**ACT**GG**TT**CC**TT**CAT**GTATTT**GGG**T**GT**ATA**GC**TCT**TG**TAT  
32G4404 351 AAATGCT**CC**ACATA**AA**CAACT**AG**TT**CA**CT**CT**CATGTATTT**GGG**T**AG**ATA**AT**TCT**TG**CTAT

51G4404 400 ACAC**TTGG****TGAAAA**AAATT**GTAG**T**TCT**ATA**ACAG**GA**ACTGG**ACTT**GT**AT**CATG**CT**CAAT**GT  
32G4404 411 ACAC**TTGG****AGATGT**AAATT**GTAG**T**GTA**ATA**ACTGG**AA**AGG**CTT**GT**AT**CATG**CT**CAAT**GT

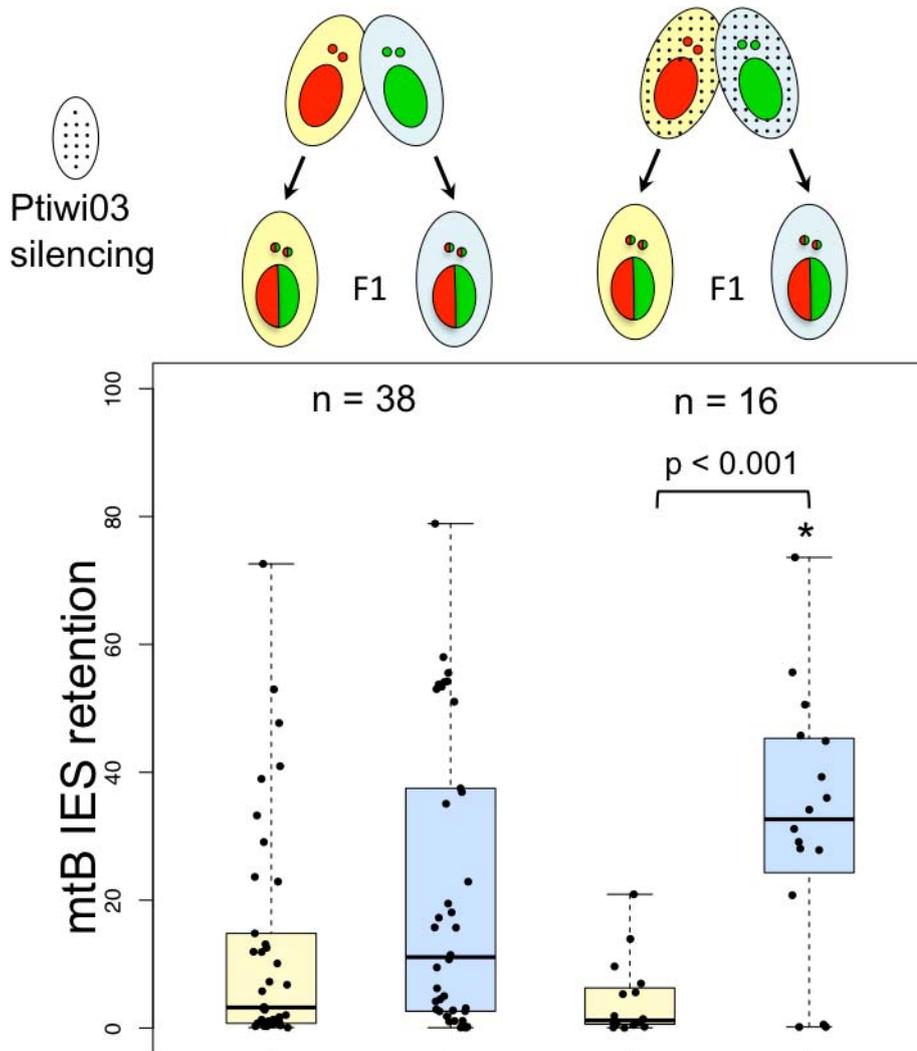
51G4404 460 **TAA**TCTT**ACAG****CG**AGGTT**GTAC**ATCT**ATTACT**GT**GGA**AC**AAA**AT**GT**TAAGACTT**AAAA**  
32G4404 471 **CAAG**CTT**ATAG**CT**TTAG**GTT**GTAC**ATCT**CTTTCG**GAT**GGA**ACC**CAAG**T**GT**TAAGACTT**AAAA**

51G4404 520 **TCAT**CT**TG**CGAATAATAT**GCAG**GT**ACA**AC**CG**CAAGTT**GC**ACAAAA**ACTG**CT**AC**CT**C**AAAA  
32G4404 531 **TC****TAC**AT**GC**GAATAATAT**GC**TGGT**GCAG**CTTT**AAG**TT**GTAC**AAAA**CA**CC**ACA**T**C**AAAG

51G4404 580 **TG**TTATTTAT**AGG**CT  
32G4404 591 **TG**CTATTTATA**AGG**TT

**Figure S4 : Alignement des séquences générant le dot plot de la figure VI.1**

Les allèles 51 et 32 d'une portion du gène G portant l'IES G4404 sont alignés. L'IES est surlignée en jaune, les substitutions sont écrites en rouge.



**Figure S5 : Effet de la déplétion de *Ptiwi03* sur le croisement 51x32.**

Une extinction génique de *PTIWIO3* a été réalisée sur les parents du croisement 51 x 32. Le contrôle, à gauche, est celui présenté en figure V-2B. Le taux de rétention calculé est présenté par ex-conjugué sous la forme de boxplots avec affichage de tous les points. L'astérisque indique une différence significative (p-value < 0,01, Kolmogorov-Smirnov) par rapport au contrôle. "p" indique la p-value issue du test de comparaison de l'IES entre les deux côtés du croisement (Test des signes de Wilcoxon).

```

32ref      GGTATTGTAAATTTAAAATATAATTTTTTTGAATAGTTTTTAGATAAAAAGTAGTGATTCCCT
BS33      GGTATTGTAAATTTAAAATATAATTTTTTTGAATAGTTTTTAGATAAAAAGTAGTGATTTTT
BS30      GGTATTGTAAATTTAAAATATAATTTTTTTGAATAGTTTTTAGATAAAAAGTAGTGATTTTT
BS32      GGTATTGTAAATTTAAAATATAATTTTTTTGAATAGTTTTTAGATAAAAAGTAGTGATTTTT
BS8       GGTATTGTAAATTTAAAATATAATTTTTTTGAATAGTTTTTAGATAAAAAGTAGTGATTTTT
BS7       GGTATTGTAAATTTAAAATATAATTTTTTTGAATAGTTTTTAGATAAAAAGTAGTGATTTTT
BS23      GGTATTGTAAATTTAAAATATAATTTTTTTGAATAGTTTTTAGATAAAAAGTAGTGATTTTT
BS6       GGTATTGTAAATTTAAAATATAATTTTTTTGAATAGTTTTTAGATAAAAAGTAGTGATTTTT
BS16      GGTATTGTAAATTTAAAATATAATTTTTTTGAATAGTTTTTAGATAAAAAGTAGTGATTTTT
BS31      GGTATTGTAAATTTAAAATATAATTTTTTTGAATAGTTTTTAGATAAAAAGTAGTGATTTTT
BS1       GGTATTGTAAATTTAAAATATAATTTTTTTGAATAGTTTTTAGATAAAAAGTAGTGATTCTT
BS19      GGTATTGTAAATTTAAAATATAATTTTTTTGAATAGTTTTTAGATAAAAAGTAGTGATTTTT
BS9       GGTATTGTAAATTTAAAATATAATTTTTTTGAATAGTTTTTAGATAAAAAGTAGTGATTTTT
BS13      GGTATTGTAAATTTAAAATATAATTTTTTTGAATAGTTTTTAGATAAAAAGTAGTGATTTTT
BS3       GGTATTGTAAATTTAAAATATAATTTTTTTGAATAGTTTTTAGATAAAAAGTAGTGATTTTT
BS20      GGTATTGTAAATTTAAAATATAATTTTTTTGAATAGTTTTTAGATAAAAAGTAGTGATTTTT
BS29      GGTATTGTAAATTTAAAATATAATTTTTTTGAATAGTTTTTAGATAAAAAGTAGTGATTTTT
BS11      GGTATTGTAAATTTAAAATATAATTTTTTTGAATAGTTTTTAGATAAAAAGTAGTGATTTTT
BS12      GGTATTGTAAATTTAAAATATAATTTTTTTGAATAGTTTTTAGATAAAAAGTAGTGATTTTT
BS34      GGTATTGTAAATTTAAAATATAATTTTTTTGAATAGTTTTTAGATAAAAAGTAGTGATTTTT
BS4       GGTATTGTAAATTTAAAATATAATTTTTTTGAATAGTTTTTAGATAAAAAGTAGTGATTTTT
BS5       GGTATTGTAAATTTAAAATATAATTTTTTTGAATAGTTTTTAGATAAAAAGTAGTGATTTTT
met       *****10*

```

```

32ref      ATTATTAATGTAATTGTTTAAATTGAATTTTTCTTCATTTTTTTATTCTAAATTGTT
BS33      ATTATTAATGTAATTGTTTAAATTGAATTTTTTTTTATTTTTTTTAAATTGTT
BS30      ATTATTAATGTAATTGTTTAAATTGAATTTTTTTTT-TTTTTTTTATTTTTTAAATTGTT
BS32      ATTATTAATGTAATTGTTTAAATTGAATTTTTTTTTATTTTTTTTAAATTGTT
BS8       ATTATTAATGTAATTGTTTAAATTGAATTTTTTTTTATTTTTTTTAAATTGTT
BS7       ATTATTAATGTAATTGTTTAAATTGAATTTTTTTTTATTTTTTTTAAATTGTT
BS23      ATTATTAATGTAATTGTTTAAATTGAATTTTTTTTTATTTTTTTTAAATTGTT
BS6       ATTATTAATGTAATTGTTTAAATTGAATTTTTTTTTATTTTTTTTAAATTGTT
BS16      ATTATTAATGTAATTGTTTAAATTGAATTTTTTTTCATTTTTTTTAAATTGTT
BS31      ATTATTAATGTAATTGTTTAAATTGAATTTTTTTTTATTTTTTTTAAATTGTT
BS1       ATTATTAATGTAATTGTTTAAATTGAATTTTTTTTTATTTTTTTTAAATTGTT
BS19      ATTATTAATGTAATTGTTTAAATTGAATTTTTTTTTATTTTTTTTAAATTGTT
BS9       ATTATTAATGTAATTGTTTAAATTGAATTTTTTTTTATTTTTTTTAAATTGTT
BS13      ATTATTAATGTAATTGTTTAAATTGAATTTTTTTTTATTTTTTTTAAATTGTT
BS3       ATTATTAATGTAATTGTTTAAATTGAATTTTTTTTTATTTTTTTTAAATTGTT
BS20      ATTATTAATGTAATTGTTTAAATTGAATTTTTTTTTATTTTTTTTAAATTGTT
BS29      ATTATTAATGTAATTGTTTAAATTGAATTTTTTTTTATTTTTTTTAAATTGTT
BS11      ATTATTAATGTAATTGTTTAAATTGAATTTTTTTTTATTTTTTTTAAATTGTT
BS12      ATTATTAATGTAATTGTTTAAATTGAATTTTTTTTTATTTTTTTTAAATTGTT
BS34      ATTATTAATGTAATTGTTTAAATTGAATTTTTTTTTATTTTTTTTAAATTGTT
BS4       ATTATTAATGTAATTGTTTAAATTGAATTTTTTTTTATTTTTTTTAAATTGTT
BS5       ATTATTAATGTAATTGTTTAAATTGAATTTTTTTTTATTTTTTTTAAATTGTT
met       *****0**1*****1*****

```

32ref CTAAAACTTAAATTAGAGTTAATATAATAGAATATCACAATTAGTTTGCCTTTG-TTTT  
BS33 TTAAAAATTTAAATTAGAGTTAATATAATAGAATATCATAATTAGTTTGTTTTTG-TTTT  
BS30 TTAGAAATTTAAATTAGAGTTAATATAATAGAATATTATAATTAGTTTGTTTTTG-TTTT  
BS32 TTAAAAATTTAAATTAGAGTTAATATAATAGAATATTATAATTAGTTTGTTTTTG--TTT  
BS8 TTAAAAATTTAAATTAGAGTTAATATAATAGAATATTATAATTAGTTTGTTTTTG-TTTT  
BS7 TTAAAAATTTAAATTAGAGTTAATATAATAGAATATTATAATTAGTTTGTTTTTG-TTTT  
BS23 TTAAAACTTAAATTAGAGTTAATATAATAGAATATTATAATTAGTTTGTTTTTG-TTTT  
BS6 TTAAAAATTTAAATTAGAGTTAATATAATAGAATATTATAATTAGTTTGTTTTTG-TTTT  
BS16 TTAAAAATTTAAATTAGAGTTAATATAATAGAATATTATAATTAGTTTGTTTTTG-TTTT  
BS31 TTAAAAATTTAAATTAGAGTTAATATAATAGAATATTATAATTAGTTTGTTTTTG-TTTT  
BS1 TTAAAAATTTAAATTAGAGTTAATATAATAGAATATTATAATTAGTTTGTTTTTG-TTTT  
BS19 TTAAAAATTTAAATTAGAGTTAATATAATAGAATATTATAATTAGTTTGTCTTTG-TTTT  
BS9 TTAAAAATTTAAATTAGAGTTAATATAATAGAATATTATAATTAGTTTGTTTTTG-TTTT  
BS13 TTAAAAATTTAAATTAGAGTTAATATAATAGAATATTATAATTAGTTTGTTTTTGTTTTT  
BS3 TTAAAAATTTAAATTAGAGTTAATATAATAGAATATTATAATTAGTTTGTTTTTG-TTTT  
BS20 TTAAAAATTTAAATTAGAGTTAATATAATAGAATATTATAATTAGTTTGTTTTTG-TTTT  
BS29 TTAAAAATTTAAATTAGAGTTAATATAATAGAATATTATAATTAGTTTGTTTTTG-TTTT  
BS11 TTAAAAATTTAAATTAGAGTTAATATAATAGAATATTATAATTAGTTTGTTTTTG--TTT  
BS12 TTAAAAATTTAAATTAGAGTTAATATAATAGAATATTATAATTAGTTTGTTTTTG-TTTT  
BS34 TTAAAAATTTAAATTAGAGTTAATATAATAGAATATTATAATTAGTTTGTTTTTG-TTTT  
BS4 TTAAAAATTTAAATTAGAGTTAATATAATAGAATATTATAATTAGTTTGTTTTTG-TTTT  
BS5 TTAAAAATTTAAATTAGAGTTAATATAATAGAATATTATAATTAGTTTGTTTTTG-TTTT  
met 0\*\* \*\*\*1\*\*\*\*\*1\*0\*\*\*\*\*01\*\*\* \*\*

32ref TCTTTTATGTTTTTATTCAACAAGAACAATATATTTTAAAAATTTTACTTAATAATACT  
BS33 TTTTTTATGTTTTTATTTAATAAGAATAAATATATTTTTGAAAATTTTATTTAATAAATATT  
BS30 TTTTTTATGTTTTTATTTAATAAGAATAAATATATTTTTAAAAATTTTATTTAATAAATATT  
BS32 TTTTTTATGTTTTTATTTAATAAGAATAAATATATTTTTAAAAATTTTATTTAATAAATATT  
BS8 TTTTTTATGTTTTTATTTAATAAGAATAAATATATTTTTAAAAATTTTATTTAATAAATATT  
BS7 TTTTTTATGTTTTTATTTAATAAGAACAATATATTTTTAAAAATTTTATTTAATAAATATT  
BS23 TTTTTTATGTTTTTATTTAATAAGAATAAATATATTTTTAAAAATTTTATTTAATAAATATT  
BS6 TTTTTTATGTTTTTATTTAATAAGAATAAATATATTTTTAAAAATTTTATTTAATAAATATT  
BS16 TTTTTTATGTTTTTATTTAATAAGAATAAATATATTTTTAAAAATTTTATTTAATAAATATT  
BS31 TTTTTTATGTTTTTATTTAATAAGAATAAATATATTTTTAAAAATTTTATTTAATAAATATT  
BS1 TTTTTTATGTTTTTATTTAATAAGAATAAATATATTTTTAAAAATTTTATTTAATAAATATT  
BS19 TTTTTTATGTTTTTATTTAATAAGAATAAATATATTTTTAAAAATTTTATTTAATAAATATT  
BS9 TTTTTTATGTTTTTATTTAATAAGAATAAATATATTTTTAAAAATTTTATTTAATAAATATT  
BS13 TTTTTTATGTTTTTATTTAATAAGAATAAATATATTTTTAAAAATTTTATTTAATAAATATT  
BS3 TTTTTTATGTTTTTATTTAATAAGAATAAATATATTTTTAAAAATTTTATTTAATAAATATT  
BS20 TTTTTTATGTTTTTATTTAATAAGAATAAATATATTTTTAAAAATTTTATTTAATAAATATT  
BS29 TTTTTTATGTTTTTATTTAATAAGAATAAATATATTTTTAAAAATTTTATTTAATAAATATT  
BS11 TTTTTTATGTTTTTATTTAATAAGAATAAATATATTTTTAAAAATTTTATTTAATAAATATT  
BS12 TTTTTTATGTTTTTATTTAATAAGAATAAATATATTTTTAAAAATTTTATTTAATAAATATT  
BS34 TTTTTTATGTTTTTATTTAATAAGAATAAATATATTTTTAAAAATTTTATTTAATAAATATT  
BS4 TTTTTTATG-TTTTATTTAATAAGAATAAATATATTTTTAAAAATTTTATTTAATAAATATT  
BS5 TTTTTTATG-TTTTATTTAATAAGAATAAATATATTTTTAAAAATTTTATTTAATAAATATT  
met \*0\*\*\*\*\* \*\*\*\*\*0\*\*0\*\*\*\*\*1\*\*\*\*\* \*\*\*\*\*0\*\*\*\*\*0\*

```

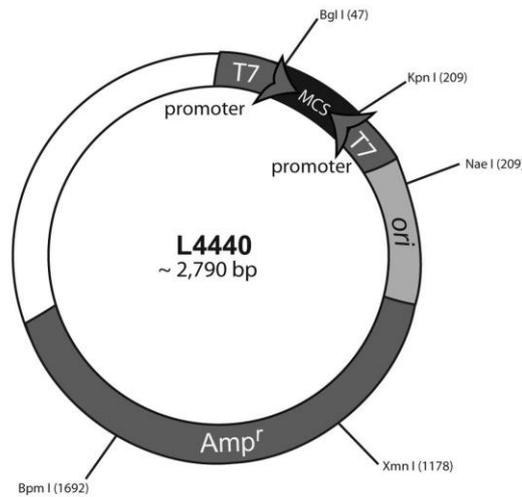
32ref      GCTTGTTAAGAATCAAAAGCTACTTGCAAGGATTACACGACATAAAACAAATGCTCCACA
BS33      GTTTGTTAAGAATTTAAAGTTATTTGTAAGGATTATATGATATAAAAATAAATGTTTCATA
BS30      GTTTGTTAAGAATTTAAAGTTATTTGTAAGGATTATATGATATAAAAATAAATGTCTTATA
BS32      GTTTGTTAAGAATTTAAAGTTATTTGTAAGGATTATATGATATAAAAATAAATGTTTTATA
BS8       GTTTGTTAAGAATTTAAAGTTATTTGTAAGGATTATATGGTATAAAAATAAATGTTTTATA
BS7       GTTTGTTAAGAATTTAAAGTTATTTGTAAGGATTACATGATATAAAAATAAATGTTTTATA
BS23      GTTTGTTAAGAATTTAAAGTTATTTGTAAGGATTACATGATATAAAAATAAATGTTTTATA
BS6       GTTTGTTAAGAATTTAAAGTTATTTGTAAGGATTACATGATATAAAAATAAATGTTTTATA
BS16      GTTTGTTAAGAATTTAAAGTTATTTGTAAGGATTATATGATATAAAAATAAATGTTTTATA
BS31      GTTTGTTAAGAATTTAAAGTTATTTGTAAGGATTATATGATATAAAAATAAATGTTTTATA
BS1       GTTTGTTAAGAATTTAAAGTTATTTGTAAGGATTATATGATATAAAAATAAATGTTTTATA
BS19      GTTTGTTAAGAATTTAAAGTTATTTGTAAGGATTATATGATATAAAAATAAATGTTTTATA
BS9       GTTTGTTAAGAATTTAAAGTTATTTGTAAGGATTATATGATATAAAAATAAATGTTTTATA
BS13      GTTTGTTAAGAATTTAAAGTTATTTGTAAGGATTATATGATATAAAAATAAATGTTTTATA
BS3       GTTTGTTAAGAATTTAAAGTTATTTGTAAGGATTATATGATATAAAAATAAATGTTTTATA
BS20      GTTTGTTAAGAATTTAAAGTTATTTGTAAGGATTATATGATATAAAAATAAATGTTTTATA
BS29      GTTTGTTAAGAATTTAAAGTTATTTGTAAGGATTATATGATATAAAAATAAATGTTTTATA
BS11      GTTTGTTAAGAATTTAAAGTTATTTGTAAGGATTATATGATATAAAAATAAATGTTTTATA
BS12      GTTTGTTAAGAATTTAAAGTTATTTGTAAGGATTATATGATATAAAAATAAATGTTTTATA
BS34      GTTTGTTAAGAATTTAAAGTTATTTGTAAGGATTATATGATATAAAAATAAATGTTTTATA
BS4       GTTTGTTAAGAATTTAAAGTTATTTGTAAGGATTATATGATATAAAAATAAATGTTTTATA
BS5       GTTTGTTAAGAATTTAAAGTTATTTGTAAGGATTATATGATATAAAAATAAATGTTTTATA
met       *0*****0*****0**0***0*****3*0* 0*****0*****0 01*0*

32ref      TAAACAACCTAGTTCACTCTTATGTATTTGGGTAGATAAATTTTGTATATATTTGG
BS33      TAAATAATTAGTTTATTTTTATGTATTTGGGTAGATAAATTTTGTATATATTTGG
BS30      TAAATAATTAGTTTATTTTTATGTATTTGGGTAGATAAATTTTGTATATATTTGG
BS32      TGAATAATTAGTTTATTTTTATGTATTTGGGTAGATAAATTTTGTATATATTTGG
BS8       TAAATAATTAGTTTCATTTTTATGTATTTGGGTAGATAAATTTTGTATATATTTGG
BS7       TAAATAATTAGTTTATTTTTATGTATTTGGGTAGATAAATTTTGTATATATTTGG
BS23      TAAATAATTAGTTTATTTTTATGTATTTGGGTAGATAAATTTTGTATATATTTGG
BS6       TAAATAATTAGTTTATTTTTATGTATTTGGGTAGATAAATTTTGTATATATTTGG
BS16      TAAATAATTAGTTTCATTTTTATGTATTTGGGTAGATAAATTTTGTATATATTTGG
BS31      TAAATAATTAGTTTCATTTTTATGTATTTGGGTAGATAAATTTTGTATATATTTGG
BS1       TAAATAATTAGTTTATTTTTATGTATTTGGGTAGATAAATTTTGTATATATTTGG
BS19      TAAATAATTAGTTTATTTTTATGTATTTGGGTAGATAAATTTTGTATATATTTGG
BS9       TAAACAATTAGTTTATTTTTATGTATTTGGGTAGATAAATTTTGTATATATTTGG
BS13      TAAATAACTAGTTTATTTTTATGTATTTGGGTAGATAAATTTTGTATATATTTGG
BS3       TAAATAATTAGTTTATTTTTATGTATTTGGGTAGATAAATTTTGTATATATTTGG
BS20      TAAATAATTAGTTTATTTTTATGTATTTGGGTAGATAAATTTTGTATATATTTGG
BS29      TAAATAATTAGTTTATTTTTATGTATTTGGGTAGATAAATTTTGTATATATTTGG
BS11      TAAATAATTAGTTTATTTTTATGTATTTGGGTAGATAAATTTTGTATATATTTGG
BS12      TAAATAATTAGTTTATTTTTATGTATTTGGGTAGATAAATTTTGTATATATTTGG
BS34      TAAATAATTAGTTTATTTTTATGTATTTGGGTAGATAAATTTTGTATATATTTGG
BS4       TAAATAATTAGTTTATTTTTATGTATTTGGGTAGATAAATTTTGTATATATTTGG
BS5       TAAATAATTAGTTTATTTTTATGTATTTGGGTAGATAAATTTTGTATATATTTGG
met       * **1**1*****3*0*0*****

```

**Figure S6 : Alignement des séquences des clones utilisé pour déterminer la méthylation des cytosines.**

La première séquence soulignée est l'amorce utilisée pour la PCR et la deuxième est le complément inverse de la deuxième amorce. La séquence surlignée en jaune est le génome de référence (MIC de la souche 32). Pour chaque cytosine de cette séquence, le nombre d'occurrences de cytosines non converties est indiqué dans la dernière ligne.



**Fragment PTIWIO3 :**

Gtattgataataaaatctactcgctaaaataattgatggaattgaaaaacaagatcaaaccatttgggaactttacaaaa  
tcaggaagaagctgagtaaagtagtcaattaagtttattaatcgattaagtgagaatcatataagcaagaactaaataa  
gctagattaataaagtagtagtgagaaatcaattctaaatggttctataggaaaaactgggagtaaatttttggaa  
aaaaactataaaataaagagaacaacttagaagtataatgtttatagtttagattggccaggagtccaatgcattat  
taaccaagtgccttttaaaattttaagcctaaactaataatcgattgtgcatttaaaattctaagatacagaactgccttagag  
gaattagactcattaaaaccatactgaattgattattggttagattgtgatgacgacctataaaaaatttactaaatt  
ctggttagaaataggaatgaatcccaatcaacatttcaaaattaaaatggcgaagctatcactttgagagttattatagta  
caaagtataataaaaaatcattccaaattaacctttataaaaagcaattgtgactggaaaaaaggatgaaagtc

**Fragment PTIWIO1 :**

Ttctaaattctaattcttaattgctatgctctgcttgacaaaatagtttctatcctttgttaccaactgaaaccatttgcattta  
gcttctgatgacttattgaaaagtcttgagataatgatattaaagtcagcttggttctcaatcttcctactagtttgagagttgc  
tgtatcttgattgtctcatttccactcaacgaacctaaaactaatccattatcagcttcttagataagctttgggaataaatgt  
tttagttcaagctcataaataagatcaattaagcttagagttaaggctctaaatttcttagtgctcactactttctcaggtgag  
gaatctgtaatttctaatttgattgataaactggtctggatgggttcaatctca

Fragment PTIWI09 :

acccatcaagaccagttaccaatacaaaactagaataactgattcctccccagagaaagtccagcgaggcactcaagaat  
tcagacctcaactccaaaactaattgacatctattatgagcttgaactagaacatttaccacaaaactaatccaagaggccg  
ataatggattagtaggctcattgagtggaatgagactaatcaagacactgcaactctaaactagtaggaaagatcga  
aaataaagctgatctgaatatcatctcaagactcttaaataagtgatcagaagtcaattacaaatggtttctgtgggaa  
acaaaggatagaaactcttttgggtccagtcgagcacagtaatttaaggactaaaatttag

Fragment DCL2 :

gaattcaattaatttaactatcaaattcagctcaaaatgagtttacagatttaacgaatttatgtaataaaaaaatccgt  
ttgatcctttcagtatatgatatacaaacagcaatgagaaagcaatatcatattctagaaattagtgaataaaggccat  
ttagaatctggatgaagcattataatattatggagtaactctaaatgagatagaaggaatgaatgttcatgaattgttcac  
tagaaaaactcattctgggaatgtgcaatcctatagataaaattatcagagttccatgaatatatagcctttagcgcata  
ctttacatatcaattaattcctcaggaatgttccaaatattggcatttagcacaatgaatgtgattagtgtactcaggcattcc  
atcagagattgcacatcactcaaacctatccaacctttaaagtaacctcaattaataccatctaataattgtaagaatccc  
tctagattataaaacctatgatattgtttgcattctctttcattcactaatatgctaattcaagattggagtcacagactcagt  
cgaaagcat

Fragment DCL3 :

gtcaacttaagaatttgatcacatcactgaattcgtaatctegaatgtaacaggcaccacatatagcctcataaatatctgct  
tactgtttcttgataattgaaatcatccacatcatttattgtcgtggatcatgattcttactttttgtaatgtatctttgaattc  
attgttcgaatttcagaattgaatacacatcggacaaatgtttgggtgattaatttaattctaaggaagacaatagacttt  
cattattaatcctttatcttcaacaacattgctgatgacagatacctaataactgagctcctcaacatttctaggtttaatagt  
ttcttctaagttagaagatggacatgaaattgcttcttctattcgttctatctatgcgatatttattcattctattatagacattc  
agtttgcgcaaatgcatcatctgacatatcaattatgagaatagctaccaattaaattcacatattcaatccgtacaggata  
agcgaacagttcactctccagccaaaatatttattagcttttagacactagaaatgaacatttagaagtatttaacaate  
tgggcacactaaaataatctaaaatgtgtatcactcaaattttaataatccgcttaatgatgcatttaattatgtgtttcatt  
gttgaataaatagaattactaatCctggtttaattgtgaatttctacgaattgcttagactgagttcattgaattaggcag  
ctctttactatcaaaaaatttatataaagtaatgtctaaattggttacttgataatgctactagatct

Fragment DCL5 :

gatctgaaatcagaatgtaagaaatcaataattggtttattaaaataactttctaataatgaaattcctcattgctttactgeat  
cttataagaatgacaagcctatttatctatactttctactaatattggtttggtgacttcaaggccaatcctttagaattatagg  
ccaataaaaatatacatatagacggtccacagccagaatatacagaagtcttgtttaaaagtgtaaattaagaaattgtatt  
tgattatgtttcaatttagcagagttttgattataatgtgagattacaaatgagttggtatgatctacctaataaatcaag  
gctcatacaataaattccacaactcatcaatgggtgatatagaattatttaattggtacc

**Figure S7 : Vecteur et fragments utilisés pour l'ARNi**

**Table S1 : Liste des oligos utilisés**

Ils sont listés suivant l'ordre de présentation des expériences auxquelles ils se rattachent.

<b>Nom</b>	<b>Séquence</b>	<b>Utilisation</b>
mtBup	tccatatcataactaaatccagc	Amplification du fragment de 3,6 kb contenant l'IES de mtB
mtBdown	ccactcctcagaattatcaaacc	Amplification du fragment de 3,6 kb contenant l'IES de mtB
Bmac4F	atcggtatagcaataacttaacaattatggc	Test de rétention de l'IES de mtB
Bmac4R	cgtctttgatattgtgacttacagcg	Test de rétention de l'IES de mtB
COXIIA	ttgtgaggtgtatttccacagatttg	Déterminer l'origine maternelle d'un F1 51/32
COXIIB	cttaaagcatagaagtagtcatcagc	Déterminer l'origine maternelle d'un F1 51/32
101F	attttagctgcaaatgcctctattacttac	Test de rétention de l'IES 101
101R	gatttaagtagggatcgggatttacc	Test de rétention de l'IES 101
32GA	caggaactggactaacagctgaatt	Test de rétention de l'IES 32G4404
32GB	gaacctataaatagcactttgatgtggttg	Test de rétention de l'IES 32G4404
51GA	ggaactggattgtagctgaaca	Test de rétention de l'IES 51G4404
51GB	agccctataaataacattttgaggtagcag	Test de rétention de l'IES 51G4404
P3F	catgctgcagccctattccatttgatcaatgctcacc	Amplification du gène Ptiwi03
P3R	aattcgagctetaaatctctgtccttaaatccatgttcatttc	Amplification du gène Ptiwi03
P3Af	ctcatattgtttactactctctctgtg	trans-PCR pour construire Ptiwi03-GFP
P3Bf	atcaggaagaagctgagtaaatag	trans-PCR pour construire Ptiwi03-GFP
P3Cr	cacaacaaaagtagccatagagattac	trans-PCR pour construire Ptiwi03-GFP
P3Dr	gctctaccttaacctctgagtaa	trans-PCR pour construire Ptiwi03-GFP
GFP F2	ggccaacactgtcactactttaac	Tester l'injection des constructions GFP
GFP R2	gtctgctaattgaactgatccatcttc	Tester l'injection des constructions GFP
PstI.P1F	catgctgcaggcagaactacattaatatgagttctga	Amplification du gène Ptiwi03
SacI.P1R	aattcgagctcgacatatttaataacaacacaatctgt	Amplification du gène Ptiwi03
P1midF	tcagagcaactgtgaggagcaataag	trans-PCR pour construire Ptiwi01-Tom
PP1TomR	tggatcccattttgtattatatacatatttactttgattg	trans-PCR pour construire Ptiwi01-Tom
PP1TomF	ataatacaaaatgggatccaagggagaagaag	trans-PCR pour construire Ptiwi01-Tom

TomP1R	ggaacatggatcccttgtaaagttcatccattccgtaaag	trans-PCR pour construire Ptiwi01-Tom
TomP1F	ttacaagggatccatgttccaaaatattaacttaagcaaac	trans-PCR pour construire Ptiwi01-Tom
P1CHR	ctgaaacatttgcatttagcttctgatgac	trans-PCR pour construire Ptiwi01-Tom + tester l'injection
ProP1F	caatcataaattgtactaaacatcttactcacact	Séquencer la construction pMN5Tom
TomMidR	tctttgatgacagccatgttgt	Séquencer la construction pMN5Tom
N1endR	aacacacaccattaattcaatttga	Tester l'injection des constructions pMN5
N1AspF	acaattacaacaacaataacaacaacca	Tester l'injection des constructions pMN5
N1endR	aacacacaccattaattcaatttga	Tester l'injection des constructions pMN5
32G44-BS1	ggtattgtaaatttaaataatTTTTgaatagt	Amplification de l'ADN traité au bisulfite
32G44-BS2	ccaaatataatacaaaaattatctacccaatacata	Amplification de l'ADN traité au bisulfite
PstI.N6MA4	aaaacctgcagattaatagctgtttatgaatttaggatc	Amplification du gène N6MA04
SacI.N6MA4	ataatgagctcgectaaacaatagattaaataacttacaatc	Amplification du gène N6MA04
PstI.N6MA9	tgaactgcagaattcttattgtgaaatagaatccttet	Amplification du gène N6MA09
SacI.N6MA9	aagtggagctcaatgaagctatgaaaataaatgggtg	Amplification du gène N6MA09
N6MA4B	cggatcccattaccgectcttttaattattaattatatttttac	trans-PCR pour construire N6MA04- GFP
N6MA4C	gaggcgggtaatgggatccggaggaatg	trans-PCR pour construire N6MA04- GFP
N6MA4D	ctgattgcatggatccaccttgtaaagttcatcc	trans-PCR pour construire N6MA04- GFP
N6MA4E	aggtggatccatgcaatcagaagataagctaaagaagc	trans-PCR pour construire N6MA04- GFP
N6MA4F	ttacagatctcttctcatcagactacattca	trans-PCR pour construire N6MA04- GFP
N6MA9B	cggatcccattcacaattaagatttaaattattattatttgg	trans-PCR pour construire N6MA09- GFP
N6MA9C	taatttgtgaatgggatccggaggaatgctc	trans-PCR pour construire N6MA09- GFP
N6MA9D	ttccttacatggatccaccttgtaaagttcatc	trans-PCR pour construire N6MA09- GFP
N6MA9E	aggtggatccatgtaaggaagacgtagcca	trans-PCR pour construire N6MA09- GFP
N6M49F	agattcaatgctagcctctactttggatagc	trans-PCR pour construire N6MA09-

		GFP
D2F	ggagcaggttcccactact	Test de rétention de l'IES D2
D2R	gctcaaatgactagcgtcgatt	Test de rétention de l'IES D2
D3F	gcagcataactagtagattaattggc	Test de rétention de l'IES D3
D3R	gcactactcataaagaaactgtcg	Test de rétention de l'IES D3
D5F	gctaaattagccaatattttcttgagc	Test de rétention de l'IES D5
D5R	cattttaattatatacacacttttgacatgat	Test de rétention de l'IES D5

**Table S2 : Informations sur les IES D2, D3 et D5.**

Caractéristiques des trois IES D2, D3 et D5 sensibles à la déplétion de Dcl2/Dcl3. Les substitutions entre les souches 51 et 32 sont indiquées pour les séquences flanquantes incluses dans les oligos et pour l'IES en elle-même.

<b>Nom</b>	<b>ID</b>	<b>Taille</b>	<b>Séquences flanquantes</b>	<b>IES</b>
D2	IESPGM.PTET51.1.2.435685	160 pb	1 subst.	1 subst.
D3	IESPGM.PTET51.1.39.393494	222 pb	4 subst.	100% identique
D5	IESPGM.PTET51.1.50.213822	204 pb	100% identique	100% identique

# **BIBLIOGRAPHIE**

- Adl, S.M., Simpson, A.G., Lane, C.E., Lukeš, J., Bass, D., Bowser, S.S., Brown, M., Burki, F., Dunthorn, M., Hampl, V., Heiss, A., Hoppenrath, M., Lara, E., leGall, L., Lynn, D.H., McManus, H., Mitchell, E.A.D., Mozley-Stanridge, S.E., Parfrey, L.W., Pawlowski, J., Rueckert, S., Shadwick, L., Schoch, C., Smirnov, A., Spiegel, F.W., 2012. The revised classification of eukaryotes. *J. Eukaryot. Microbiol.* 59, 429–493. doi:10.1111/j.1550-7408.2012.00644.x
- Agrawal, A., Eastman, Q.M., Schatz, D.G., 1998. Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* 394, 744–751. doi:10.1038/29457
- Almeida, R., Allshire, R.C., 2005. RNA silencing and genome regulation. *Trends Cell Biol.* 15, 251–258. doi:10.1016/j.tcb.2005.03.006
- Anxolabéhère, D., Nouaud, D., Périquet, G., Tchen, P., 1985. P-element distribution in Eurasian populations of *Drosophila melanogaster*: A genetic and molecular analysis. *Proc. Natl. Acad. Sci. U. S. A.* 82, 5418–5422.
- Apte, M.S., Meller, V.H., 2012. Homologue Pairing in Flies and Mammals: Gene Regulation When Two Are Involved. *Genet. Res. Int.* 2012. doi:10.1155/2012/430587
- Aramayo, R., Metzberg, R.L., 1996. Meiotic transvection in fungi. *Cell* 86, 103–113.
- Aramayo, R., Selker, E.U., 2013. *Neurospora crassa*, a Model System for Epigenetics Research. *Cold Spring Harb. Perspect. Biol.* 5. doi:10.1101/cshperspect.a017921
- Arambasic, M., Sandoval, P.Y., Hoehener, C., Singh, A., Swart, E.C., Nowacki, M., 2014. Pdsg1 and Pdsg2, Novel Proteins Involved in Developmental Genome Remodelling in *Paramecium*. *PLoS ONE* 9, e112899. doi:10.1371/journal.pone.0112899
- Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M.J., Kuramochi-Miyagawa, S., Nakano, T., Chien, M., Russo, J.J., Ju, J., Sheridan, R., Sander, C., Zavolan, M., Tuschl, T., 2006. A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* 442, 203–207. doi:10.1038/nature04916
- Arnaiz, O., Goût, J.-F., Bétermier, M., Bouhouche, K., Cohen, J., Duret, L., Kapusta, A., Meyer, E., Sperling, L., 2010. Gene expression in a paleopolyploid: a transcriptome resource for the ciliate *Paramecium tetraurelia*. *BMC Genomics* 11, 547. doi:10.1186/1471-2164-11-547
- Arnaiz, O., Mathy, N., Baudry, C., Malinsky, S., Aury, J.-M., Denby Wilkes, C., Garnier, O., Labadie, K., Lauderdale, B.E., Le Mouël, A., Marmignon, A., Nowacki, M., Poulain, J., Prajer, M., Wincker, P., Meyer, E., Duharcourt, S., Duret, L., Bétermier, M., Sperling, L., 2012. The *Paramecium* Germline Genome Provides a Niche for Intragenic Parasitic DNA: Evolutionary Dynamics of Internal Eliminated Sequences. *PLoS Genet.* 8, e1002984. doi:10.1371/journal.pgen.1002984
- Aronica, L., Bednenko, J., Noto, T., DeSouza, L.V., Siu, K.W.M., Loidl, J., Pearlman, R.E., Gorovsky, M.A., Mochizuki, K., 2008. Study of an RNA helicase implicates small RNA–noncoding RNA interactions in programmed DNA elimination in *Tetrahymena*. *Genes Dev.* 22, 2228–2241. doi:10.1101/gad.481908

Ashe, A., Sapetschnig, A., Weick, E.-M., Mitchell, J., Bagijn, M.P., Cording, A.C., Doebley, A.-L., Goldstein, L.D., Lehrbach, N.J., Le Pen, J., Pintacuda, G., Sakaguchi, A., Sarkies, P., Ahmed, S., Miska, E.A., 2012. piRNAs Can Trigger a Multigenerational Epigenetic Memory in the Germline of *C. elegans*. *Cell* 150, 88–99. doi:10.1016/j.cell.2012.06.018

Aury, J.-M., Jaillon, O., Duret, L., Noel, B., Jubin, C., Porcel, B.M., Ségurens, B., Daubin, V., Anthouard, V., Aiach, N., Arnaiz, O., Billaut, A., Beisson, J., Blanc, I., Bouhouche, K., Câmara, F., Duharcourt, S., Guigo, R., Gogendeau, D., Katinka, M., Keller, A.-M., Kissmehl, R., Klotz, C., Koll, F., Le Mouël, A., Lepère, G., Malinsky, S., Nowacki, M., Nowak, J.K., Plattner, H., Poulain, J., Ruiz, F., Serrano, V., Zagulski, M., Dessen, P., Bétermier, M., Weissenbach, J., Scarpelli, C., Schächter, V., Sperling, L., Meyer, E., Cohen, J., Wincker, P., 2006. Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* 444, 171–178. doi:10.1038/nature05230

Bagijn, M.P., Goldstein, L.D., Sapetschnig, A., Weick, E.-M., Bouasker, S., Lehrbach, N.J., Simard, M.J., Miska, E.A., 2012. Function, Targets, and Evolution of *Caenorhabditis elegans* piRNAs. *Science* 337, 574–578. doi:10.1126/science.1220952

Baldauf, S.L., 2003. The Deep Roots of Eukaryotes. *Science* 300, 1703–1706. doi:10.1126/science.1085544

Batista, P.J., Ruby, J.G., Claycomb, J.M., Chiang, R., Fahlgren, N., Kasschau, K.D., Chaves, D.A., Gu, W., Vasale, J.J., Duan, S., Conte Jr., D., Luo, S., Schroth, G.P., Carrington, J.C., Bartel, D.P., Mello, C.C., 2008. PRG-1 and 21U-RNAs Interact to Form the piRNA Complex Required for Fertility in *C. elegans*. *Mol. Cell* 31, 67–78. doi:10.1016/j.molcel.2008.06.002

Baudry, C., Malinsky, S., Restituto, M., Kapusta, A., Rosa, S., Meyer, E., Bétermier, M., 2009. PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate *Paramecium tetraurelia*. *Genes Dev.* 23, 2478–2483. doi:10.1101/gad.547309

Bayes, J.J., Malik, H.S., 2009. Altered heterochromatin binding by a hybrid sterility protein in *Drosophila* sibling species. *Science* 326, 1538–1541. doi:10.1126/science.1181756

Behnke, D., 1990. DOUGLAS E. BERG and MARTHA M. HOWE (Editors), *Mobile DNA*. XIII + 972 S., 300 + 2 Abb., 36 Tab. Washington, D.C. 1989. American Society for Microbiology. \$ 95.00. ISBN: 1-55581-005-5. *J. Basic Microbiol.* 30, 498–498. doi:10.1002/jobm.3620300707

Beisson, J., Bétermier, M., Bré, M.-H., Cohen, J., Duharcourt, S., Duret, L., Kung, C., Malinsky, S., Meyer, E., Preer, J.R., Sperling, L., 2010. Maintaining Clonal *Paramecium tetraurelia* Cell Lines of Controlled Age through Daily Reisolation. *Cold Spring Harb. Protoc.* 2010, pdb.prot5361. doi:10.1101/pdb.prot5361

Beisson, J., Bétermier, M., Bré, M.-H., Cohen, J., Duharcourt, S., Duret, L., Kung, C., Malinsky, S., Meyer, E., Preer, J.R., Sperling, L., 2010. DNA Microinjection into the Macronucleus of *Paramecium*. *Cold Spring Harb. Protoc.* 2010, pdb.prot5364-pdb.prot5364. doi:10.1101/pdb.prot5364

- Bétermier, M., Duhaucourt, S., Seitz, H., Meyer, E., 2000. Timing of Developmentally Programmed Excision and Circularization of Paramecium Internal Eliminated Sequences. *Mol. Cell. Biol.* 20, 1553–1561. doi:10.1128/MCB.20.5.1553-1561.2000
- Bingham, P.M., Kidwell, M.G., Rubin, G.M., 1982. The molecular basis of P-M hybrid dysgenesis: the role of the P element, a P-strain-specific transposon family. *Cell* 29, 995–1004.
- Blackman, R.K., Grimaila, R., Koehler, M.M., Gelbart, W.M., 1987. Mobilization of hobo elements residing within the decapentaplegic gene complex: suggestion of a new hybrid dysgenesis system in *Drosophila melanogaster*. *Cell* 49, 497–505.
- Blumenstiel, J.P., 2014. Whole genome sequencing in *Drosophila virilis* identifies Polyphemus, a recently activated Tc1-like transposon with a possible role in hybrid dysgenesis. *Mob. DNA* 5, 6. doi:10.1186/1759-8753-5-6
- Borges, F., Martienssen, R.A., 2015. The expanding world of small RNAs in plants. *Nat. Rev. Mol. Cell Biol.* 16, 727–741. doi:10.1038/nrm4085
- Bouhouche, K., Gout, J.-F., Kapusta, A., Betermier, M., Meyer, E., 2011. Functional specialization of Piwi proteins in *Paramecium tetraurelia* from post-transcriptional gene silencing to genome remodelling. *Nucleic Acids Res.* 39, 4249–4264. doi:10.1093/nar/gkq1283
- Bowen, N.J., McDonald, J.F., 2001. *Drosophila* euchromatic LTR retrotransposons are much younger than the host species in which they reside. *Genome Res.* 11, 1527–1540. doi:10.1101/gr.164201
- Bracht, J.R., Perlman, D.H., Landweber, L.F., 2012. Cytosine methylation and hydroxymethylation mark DNA for elimination in *Oxytricha trifallax*. *Genome Biol.* 13, R99. doi:10.1186/gb-2012-13-10-r99
- Brandt, V.L., Hewitt, S.L., Skok, J.A., 2010. It takes two. *Nucleus* 1, 23–29. doi:10.4161/nucl.1.1.10595
- Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., Hannon, G.J., 2007. Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity in *Drosophila*. *Cell* 128, 1089–1103. doi:10.1016/j.cell.2007.01.043
- Brennecke, J., Malone, C.D., Aravin, A.A., Sachidanandam, R., Stark, A., Hannon, G.J., 2008. An Epigenetic Role for Maternally Inherited piRNAs in Transposon Silencing. *Science* 322, 1387–1392. doi:10.1126/science.1165171
- Brideau, N.J., Flores, H.A., Wang, J., Maheshwari, S., Wang, X., Barbash, D.A., 2006. Two Dobzhansky-Muller Genes Interact to Cause Hybrid Lethality in *Drosophila*. *Science* 314, 1292–1295. doi:10.1126/science.1133953
- Bromberg, S., Pratt, K., Hattman, S., 1982. Sequence specificity of DNA adenine methylase in the protozoan *Tetrahymena thermophila*. *J. Bacteriol.* 150, 993–996.
- Brygoo, Y., 1977a. Genetic analysis of mating-type differentiation in *Paramecium tetraurelia*. *Genetics* 87, 633–653.

- Brygoo, Y., 1977b. Genetic analysis of mating-type differentiation in *Paramecium tetraurelia*. *Genetics* 87, 633–653.
- Bucheton, A., 1990. I transposable elements and I-R hybrid dysgenesis in *Drosophila*. *Trends Genet. TIG* 6, 16–21.
- Bucheton, A., 1979. Non-Mendelian Female Sterility in *DROSOPHILA MELANOGASTER*: Influence of Aging and Thermic Treatments. III. Cumulative Effects Induced by These Factors. *Genetics* 93, 131.
- Byrne, B.C., 1973. Mutational Analysis of Mating Type Inheritance in *Syngen 4* of *PARAMECIUM AURELIA*. *Genetics* 74, 63–80.
- Calarco, J.P., Borges, F., Donoghue, M.T.A., Van Ex, F., Jullien, P.E., Lopes, T., Gardner, R., Berger, F., Feijó, J.A., Becker, J.D., Martienssen, R.A., 2012. Reprogramming of DNA Methylation in Pollen Guides Epigenetic Inheritance via Small RNA. *Cell* 151, 194–205. doi:10.1016/j.cell.2012.09.001
- Caron, F., Meyer, E., 1985. Does *Paramecium primaurelia* use a different genetic code in its macronucleus? *Nature* 314, 185–188.
- Casjens, S., 2003. Prophages and bacterial genomics: what have we learned so far? *Mol. Microbiol.* 49, 277–300. doi:10.1046/j.1365-2958.2003.03580.x
- Cassidy-Hanley, D., Bowen, J., Lee, J.H., Cole, E., VerPlank, L.A., Gaertig, J., Gorovsky, M.A., Bruns, P.J., 1997. Germline and Somatic Transformation of Mating *Tetrahymena thermophila* by Particle Bombardment. *Genetics* 146, 135–147.
- Chalker, D.L., La Terza, A., Wilson, A., Kroenke, C.D., Yao, M.C., 1999. Flanking regulatory sequences of the *Tetrahymena R* deletion element determine the boundaries of DNA rearrangement. *Mol. Cell. Biol.* 19, 5631–5641.
- Chalker, D.L., Yao, M.-C., 2001. Nongenic, bidirectional transcription precedes and may promote developmental DNA deletion in *Tetrahymena thermophila*. *Genes Dev.* 15, 1287–1298. doi:10.1101/gad.884601
- Chalker, D.L., Yao, M.C., 1996. Non-Mendelian, heritable blocks to DNA rearrangement are induced by loading the somatic nucleus of *Tetrahymena thermophila* with germ line-limited DNA. *Mol. Cell. Biol.* 16, 3658–3667.
- Chambeyron, S., Popkova, A., Payen-Groschêne, G., Brun, C., Laouini, D., Pelisson, A., Bucheton, A., 2008. piRNA-mediated nuclear accumulation of retrotransposon transcripts in the *Drosophila* female germline. *Proc. Natl. Acad. Sci.* 105, 14964–14969. doi:10.1073/pnas.0805943105
- Chen, X., Bracht, J.R., Goldman, A.D., Dolzhenko, E., Clay, D.M., Swart, E.C., Perlman, D.H., Doak, T.G., Stuart, A., Amemiya, C.T., Sebra, R.P., Landweber, L.F., 2014. The Architecture of a Scrambled Genome Reveals Massive Levels of Genomic Rearrangement during Development. *Cell* 158, 1187–1198. doi:10.1016/j.cell.2014.07.034
- Chou, J.-Y., Leu, J.-Y., 2015. The Red Queen in mitochondria: cyto-nuclear co-evolution, hybrid breakdown and human disease. *Front. Genet.* 6. doi:10.3389/fgene.2015.00187

- Coleman, A.W., 2005. *Paramecium aurelia* Revisited. *J. Eukaryot. Microbiol.* 52, 68–77. doi:10.1111/j.1550-7408.2005.3327r.x
- Comai, L., Madlung, A., Josefsson, C., Tyagi, A., 2003. Do the different parental “heteromes” cause genomic shock in newly formed allopolyploids? *Philos. Trans. R. Soc. B Biol. Sci.* 358, 1149–1155. doi:10.1098/rstb.2003.1305
- Corrado, M.U.D., Locatelli, D., Paleari, L., Bottiroli, G., 1997. Lectin-Binding Sites Involved in *Paramecium primaurelia* Mating Pair Formation. *J. Eukaryot. Microbiol.* 44, 603–608. doi:10.1111/j.1550-7408.1997.tb05966.x
- Cotton, J., 2001. Retroviruses from retrotransposons. *Genome Biol.* 2, 1. doi:10.1186/gb-2001-2-2-reports0006
- Cox, D.N., Chao, A., Baker, J., Chang, L., Qiao, D., Lin, H., 1998. A novel class of evolutionarily conserved genes defined by piwi are essential for stem cell self-renewal. *Genes Dev.* 12, 3715–3727. doi:10.1101/gad.12.23.3715
- Coyne, J.A., 1986. Meiotic Segregation and Male Recombination in Interspecific Hybrids of *Drosophila*. *Genetics* 114, 485–494.
- Coyne, Jerry A, 1989. Heredity - Abstract of article: Mutation rates in hybrids between sibling species of *Drosophila*. *Heredity* 63, 155–162.
- Coyne, R.S., Lhuillier-Akakpo, M., Duharcourt, S., 2012. RNA-guided DNA rearrangements in ciliates: Is the best genome defence a good offence? *Biol. Cell* 104, 309–325. doi:10.1111/boc.201100057
- Cummings, D.J., Tait, A., Goddard, J.M., 1974. Methylated bases in DNA from *Paramecium aurelia*. *Biochim. Biophys. Acta BBA - Nucleic Acids Protein Synth.* 374, 1–11. doi:10.1016/0005-2787(74)90194-4
- Daniels, S.B., Peterson, K.R., Strausbaugh, L.D., Kidwell, M.G., Chovnick, A., 1990. Evidence for horizontal transmission of the P transposable element between *Drosophila* species. *Genetics* 124, 339–355.
- Das, P.P., Bagijn, M.P., Goldstein, L.D., Woolford, J.R., Lehrbach, N.J., Sapetschnig, A., Buhecha, H.R., Gilchrist, M.J., Howe, K.L., Stark, R., Matthews, N., Berezikov, E., Ketting, R.F., Tavaré, S., Miska, E.A., 2008. Piwi and piRNAs Act Upstream of an Endogenous siRNA Pathway to Suppress Tc3 Transposon Mobility in the *Caenorhabditis elegans* Germline. *Mol. Cell* 31, 79–90. doi:10.1016/j.molcel.2008.06.003
- de Vanssay, A., Bougé, A.-L., Boivin, A., Hermant, C., Teyssset, L., Delmarre, V., Antoniewski, C., Ronsseray, S., 2012. Paramutation in *Drosophila* linked to emergence of a piRNA-producing locus. *Nature* 490, 112–115. doi:10.1038/nature11416
- Ding, D.-Q., Okamasa, K., Yamane, M., Tsutsumi, C., Haraguchi, T., Yamamoto, M., Hiraoka, Y., 2012. Meiosis-Specific Noncoding RNA Mediates Robust Pairing of Homologous Chromosomes in Meiosis. *Science* 336, 732–736. doi:10.1126/science.1219518
- Dion-Côté, A.-M., Renaut, S., Normandeau, E., Bernatchez, L., 2014. RNA-seq Reveals Transcriptomic Shock Involving Transposable Elements Reactivation in Hybrids of Young Lake Whitefish Species. *Mol. Biol. Evol.* 31, 1188–1199. doi:10.1093/molbev/msu069

- Doak, T.G., Cavalcanti, A.R.O., Stover, N.A., Dunn, D.M., Weiss, R., Herrick, G., Landweber, L.F., 2003. Sequencing the *Oxytricha trifallax* macronuclear genome: a pilot project. *Trends Genet.* 19, 603–607. doi:10.1016/j.tig.2003.09.013
- Doolittle, W.F., Sapienza, C., 1980. Selfish genes, the phenotype paradigm and genome evolution. *Nature* 284, 601–603. doi:10.1038/284601a0
- Duan, G., Saint, R.B., Helliwell, C.A., Behm, C.A., Wang, M.-B., Waterhouse, P.M., Gordon, K.H.J., 2013. *C. elegans* RNA-dependent RNA polymerases rrf-1 and ego-1 silence *Drosophila* transgenes by differing mechanisms. *Cell. Mol. Life Sci.* 70, 1469–1481. doi:10.1007/s00018-012-1218-8
- Duharcourt, S., Butler, A., Meyer, E., 1995. Epigenetic self-regulation of developmental excision of an internal eliminated sequence on *Paramecium tetraurelia*. *Genes Dev.* 9, 2065–2077. doi:10.1101/gad.9.16.2065
- Duharcourt, S., Keller, A.-M., Meyer, E., 1998. Homology-Dependent Maternal Inhibition of Developmental Excision of Internal Eliminated Sequences in *Paramecium tetraurelia*. *Mol. Cell. Biol.* 18, 7075–7085.
- Eggleston, W.B., Johnson-Schlitz, D.M., Engels, W.R., 1988. P-M hybrid dysgenesis does not mobilize other transposable element families in *D. melanogaster*. *Nature* 331, 368–370. doi:10.1038/331368a0
- Eisen, J.A., Coyne, R.S., Wu, M., Wu, D., Thiagarajan, M., Wortman, J.R., Badger, J.H., Ren, Q., Amedeo, P., Jones, K.M., Tallon, L.J., Delcher, A.L., Salzberg, S.L., Silva, J.C., Haas, B.J., Majoros, W.H., Farzad, M., Carlton, J.M., Smith, R.K., Jr., Garg, J., Pearlman, R.E., Karrer, K.M., Sun, L., Manning, G., Elde, N.C., Turkewitz, A.P., Asai, D.J., Wilkes, D.E., Wang, Y., Cai, H., Collins, K., Stewart, B.A., Lee, S.R., Wilamowska, K., Weinberg, Z., Ruzzo, W.L., Wloga, D., Gaertig, J., Frankel, J., Tsao, C.-C., Gorovsky, M.A., Keeling, P.J., Waller, R.F., Patron, N.J., Cherry, J.M., Stover, N.A., Krieger, C.J., del Toro, C., Ryder, H.F., Williamson, S.C., Barbeau, R.A., Hamilton, E.P., Orias, E., 2006. Macronuclear Genome Sequence of the Ciliate *Tetrahymena thermophila*, a Model Eukaryote. *PLoS Biol* 4, e286. doi:10.1371/journal.pbio.0040286
- Eissenberg, J.C., Elgin, S.C., 2000. The HP1 protein family: getting a grip on chromatin. *Curr. Opin. Genet. Dev.* 10, 204–210.
- El-Shami, M., Pontier, D., Lahmy, S., Braun, L., Picart, C., Vega, D., Hakimi, M.-A., Jacobsen, S.E., Cooke, R., Lagrange, T., 2007. Reiterated WG/GW motifs form functionally and evolutionarily conserved ARGONAUTE-binding platforms in RNAi-related components. *Genes Dev.* 21, 2539–2544. doi:10.1101/gad.451207
- Engels, W.R., 1979. Hybrid dysgenesis in *Drosophila melanogaster*: rules of inheritance of female sterility. *Genet. Res.* 33, 219–236. doi:10.1017/S001667230001836X
- Erwin, A.A., Galdos, M.A., Wickersheim, M.L., Harrison, C.C., Marr, K.D., Colicchio, J.M., Blumenstiel, J.P., 2015. piRNAs Are Associated with Diverse Transgenerational Effects on Gene and Transposon Expression in a Hybrid Dysgenic Syndrome of *D. virilis*. *PLoS Genet.* 11, e1005332. doi:10.1371/journal.pgen.1005332

- Evgen'ev, M.B., Zelentsova, H., Shostak, N., Kozitsina, M., Barskyi, V., Lankenau, D.H., Corces, V.G., 1997. Penelope, a new family of transposable elements and its possible role in hybrid dysgenesis in *Drosophila virilis*. *Proc. Natl. Acad. Sci. U. S. A.* 94, 196–201.
- Fang, W., Wang, X., Bracht, J.R., Nowacki, M., Landweber, L.F., 2012. Piwi-Interacting RNAs Protect DNA against Loss during *Oxytricha* Genome Rearrangement. *Cell* 151, 1243–1255. doi:10.1016/j.cell.2012.10.045
- Fass, J.N., Joshi, N.A., Couvillion, M.T., Bowen, J., Gorovsky, M.A., Hamilton, E.P., Orias, E., Hong, K., Coyne, R.S., Eisen, J.A., Chalker, D.L., Lin, D., Collins, K., 2011. Genome-Scale Analysis of Programmed DNA Elimination Sites in *Tetrahymena thermophila*. *G3 GenesGenomesGenetics* 1, 515–522. doi:10.1534/g3.111.000927
- Fedoroff, N., Wessler, S., Shure, M., 1983. Isolation of the transposable maize controlling elements Ac and Ds. *Cell* 35, 235–242.
- Fillingham, J.S., Bruno, D., Pearlman, R.E., 2001. Cis-acting requirements in flanking DNA for the programmed elimination of *mse2.9*: a common mechanism for deletion of internal eliminated sequences from the developing macronucleus of *Tetrahymena thermophila*. *Nucleic Acids Res.* 29, 488–498.
- Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., Mello, C.C., 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391, 806–811. doi:10.1038/35888
- Fok, A.K., Sison, B.C., Ueno, M.S., Allen, R.D., 1988. Phagosome formation in *Paramecium*: effects of solid particles. *J. Cell Sci.* 90, 517–524.
- Fokin, S., 1997. Morphological Diversity of the Micronuclei in *Paramecium*. *Arch. Für Protistenkd.* 148, 375–387. doi:10.1016/S0003-9365(97)80017-6
- Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., Postlethwait, J., 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151, 1531–1545.
- Forney, J.D., Blackburn, E.H., 1988. Developmentally controlled telomere addition in wild-type and mutant *paramecia*. *Mol. Cell. Biol.* 8, 251–258.
- Gaertig, J., Gorovsky, M.A., 1992. Efficient mass transformation of *Tetrahymena thermophila* by electroporation of conjugants. *Proc. Natl. Acad. Sci. U. S. A.* 89, 9196–9200.
- Galagan, J.E., Selker, E.U., 2004. RIP: the evolutionary cost of genome defense. *Trends Genet.* TIG 20, 417–423. doi:10.1016/j.tig.2004.07.007
- Galvani, A., Sperling, L., 2002. RNA interference by feeding in *Paramecium*. *Trends Genet.* 18, 11–12. doi:10.1016/S0168-9525(01)02548-3
- Galvani, A., Sperling, L., 2000. Regulation of secretory protein gene expression in *Paramecium*. *Eur. J. Biochem.* 267, 3226–3234. doi:10.1046/j.1432-1327.2000.01341.x
- Garnier, O., Serrano, V., Duhaucourt, S., Meyer, E., 2004. RNA-Mediated Programming of Developmental Genome Rearrangements in *Paramecium tetraurelia*. *Mol. Cell. Biol.* 24, 7370–7379. doi:10.1128/MCB.24.17.7370-7379.2004

- Gasiunas, G., Sinkunas, T., Siksnys, V., 2014. Molecular mechanisms of CRISPR-mediated microbial immunity. *Cell. Mol. Life Sci.* 71, 449–465. doi:10.1007/s00018-013-1438-6
- Girard, A., Sachidanandam, R., Hannon, G.J., Carmell, M.A., 2006. A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* 442, 199–202. doi:10.1038/nature04917
- Gloor, G.B., Preston, C.R., Johnson-Schlitz, D.M., Nassif, N.A., Phillis, R.W., Benz, W.K., Robertson, H.M., Engels, W.R., 1993. Type I repressors of P element mobility. *Genetics* 135, 81–95.
- Godiska, R., James, C., Yao, M.C., 1993. A distant 10-bp sequence specifies the boundaries of a programmed DNA deletion in *Tetrahymena*. *Genes Dev.* 7, 2357–2365.
- Godiska, R., Yao, M.-C., 1990. A programmed site-specific DNA rearrangement in *Tetrahymena thermophila* requires flanking polypurine tracts. *Cell* 61, 1237–1246. doi:10.1016/0092-8674(90)90688-B
- Gout, J.-F., Duret, L., Kahn, D., 2009. Differential Retention of Metabolic Genes Following Whole-Genome Duplication. *Mol. Biol. Evol.* 26, 1067–1072. doi:10.1093/molbev/msp026
- Grandchamp, S., Beisson, J., 1981. Positional control of nuclear differentiation in paramecium. *Dev. Biol.* 81, 336–341. doi:10.1016/0012-1606(81)90297-9
- Greer, E.L., Blanco, M.A., Gu, L., Sendinc, E., Liu, J., Aristizábal-Corrales, D., Hsu, C.-H., Aravind, L., He, C., Shi, Y., 2015. DNA Methylation on N6-Adenine in *C. elegans*. *Cell* 161, 868–878. doi:10.1016/j.cell.2015.04.005
- Greslin, A.F., Prescott, D.M., Oka, Y., Loukin, S.H., Chappell, J.C., 1989. Reordering of nine exons is necessary to form a functional actin gene in *Oxytricha nova*. *Proc. Natl. Acad. Sci. U. S. A.* 86, 6264–6268.
- Gu, S.G., Pak, J., Guang, S., Maniar, J.M., Kennedy, S., Fire, A., 2012. Amplification of siRNA in *Caenorhabditis elegans* generates a transgenerational sequence-targeted histone H3 lysine 9 methylation footprint. *Nat. Genet.* 44, 157–164. doi:10.1038/ng.1039
- Gu, W., Shirayama, M., Conte, D., Vasale, J., Batista, P.J., Claycomb, J.M., Moresco, J.J., Youngman, E.M., Keys, J., Stoltz, M.J., Chen, C.-C.G., Chaves, D.A., Duan, S., Kasschau, K.D., Fahlgren, N., Yates, J.R., Mitani, S., Carrington, J.C., Mello, C.C., 2009. Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the *C. elegans* germline. *Mol. Cell* 36, 231–244. doi:10.1016/j.molcel.2009.09.020
- Guerreiro, M.G., 1996. Heredity - Abstract of article: Behaviour of the transposable elements copia and mdg1 in hybrids between the sibling species *Drosophila melanogaster* and *D. simulans*. *Heredity* 77, 40–46.
- Gunawardane, L.S., Saito, K., Nishida, K.M., Miyoshi, K., Kawamura, Y., Nagami, T., Siomi, H., Siomi, M.C., 2007. A slicer-mediated mechanism for repeat-associated siRNA 5' end formation in *Drosophila*. *Science* 315, 1587–1590. doi:10.1126/science.1140494
- Hamilton, E.P., Kapusta, A., Huvos, P.E., Bidwell, S.L., Zafar, N., Tang, H., Hadjithomas, M., Krishnakumar, V., Badger, J.H., Caler, E.V., Russ, C., Zeng, Q., Fan, L., Levin, J.Z., Shea, T., Young, S.K., Hegarty, R., Daza, R., Gujja, S., Wortman, J.R., Birren, B.W., Nusbaum, C.,

- Thomas, J., Carey, C.M., Pritham, E.J., Feschotte, C., Noto, T., Mochizuki, K., Papazyan, R., Taverna, S.D., Dear, P.H., Cassidy-Hanley, D.M., Xiong, J., Miao, W., Orias, E., Coyne, R.S., 2016. Structure of the germline genome of *Tetrahymena thermophila* and relationship to the massively rearranged somatic genome. *eLife* 5, e19090. doi:10.7554/eLife.19090
- Hammond, T.M., Xiao, H., Boone, E.C., Perdue, T.D., Pukkila, P.J., Shiu, P.K.T., 2011. SAD-3, a Putative Helicase Required for Meiotic Silencing by Unpaired DNA, Interacts with Other Components of the Silencing Machinery. *G3 GenesGenomesGenetics* 1, 369–376. doi:10.1534/g3.111.000570
- Heard, E., Martienssen, R.A., 2014. Transgenerational Epigenetic Inheritance: myths and mechanisms. *Cell* 157, 95–109. doi:10.1016/j.cell.2014.02.045
- Henderson, I.R., Chan, S.R., Cao, X., Johnson, L., Jacobsen, S.E., 2010. Accurate sodium bisulfite sequencing in plants. *Epigenetics Off. J. DNA Methylation Soc.* 5, 47–49.
- Hey, J., 1988. Speciation via hybrid dysgenesis: negative evidence from the *Drosophila affinis* subgroup. *Genetica* 78, 97–103. doi:10.1007/BF00058840
- Hill, T., Schlötterer, C., Betancourt, A.J., 2016. Hybrid Dysgenesis in *Drosophila simulans* Associated with a Rapid Invasion of the P-Element. *PLoS Genet.* 12. doi:10.1371/journal.pgen.1005920
- Höck, J., Meister, G., 2008. The Argonaute protein family. *Genome Biol.* 9, 210. doi:10.1186/gb-2008-9-2-210
- Inaba, F., Imamoto, K., Suganuma, Y., 1966. Electron-microscopic Observations on Nuclear Exchange during Conjugation in *Paramecium multimicronucleatum*. *Proc. Jpn. Acad.* 42, 394–398. doi:10.2183/pjab1945.42.394
- Ipsaro, J.J., Haase, A.D., Knott, S.R., Joshua-Tor, L., Hannon, G.J., 2012. The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis. *Nature* 491, 279–283. doi:10.1038/nature11502
- Iwasaki, S., Kobayashi, M., Yoda, M., Sakaguchi, Y., Katsuma, S., Suzuki, T., Tomari, Y., 2010. Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. *Mol. Cell* 39, 292–299. doi:10.1016/j.molcel.2010.05.015
- Iyer, L.M., Zhang, D., Aravind, L., 2016. Adenine methylation in eukaryotes: Apprehending the complex evolutionary history and functional potential of an epigenetic modification. *BioEssays* 38, 27–40. doi:10.1002/bies.201500104
- Jansen, R., Embden, J.D.A. van, Gaastra, W., Schouls, L.M., 2002. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* 43, 1565–1575.
- Jensen, P.A., Stuart, J.R., Goodpaster, M.P., Goodman, J.W., Simmons, M.J., 2008. Cytotype regulation of P transposable elements in *Drosophila melanogaster*: repressor polypeptides or piRNAs? *Genetics* 179, 1785–1793. doi:10.1534/genetics.108.087072
- Johanson, T.M., Lew, A.M., Chong, M.M.W., 2013. MicroRNA-independent roles of the RNase III enzymes Drosha and Dicer. *Open Biol.* 3. doi:10.1098/rsob.130144

- Jones, W.K., n.d. Nuclear differentiation in *Paramecium*., PhD thesis, University of Wales, Aberystwyth, UK.
- Josefsson, C., Dilkes, B., Comai, L., 2006. Parent-Dependent Loss of Gene Silencing during Interspecies Hybridization. *Curr. Biol.* 16, 1322–1328. doi:10.1016/j.cub.2006.05.045
- Kahn, R.W., Andersen, B.H., Brunk, C.F., 1993. Transformation of *Tetrahymena thermophila* by microinjection of a foreign gene. *Proc. Natl. Acad. Sci. U. S. A.* 90, 9295–9299.
- Kapusta, A., Matsuda, A., Marmignon, A., Ku, M., Silve, A., Meyer, E., Forney, J.D., Malinsky, S., Bétermier, M., 2011. Highly Precise and Developmentally Programmed Genome Assembly in *Paramecium* Requires Ligase IV–Dependent End Joining. *PLoS Genet* 7, e1002049. doi:10.1371/journal.pgen.1002049
- Kelleher, E.S., 2016. Reexamining the P-Element Invasion of *Drosophila melanogaster* Through the Lens of piRNA Silencing. *Genetics* 203, 1513–1531. doi:10.1534/genetics.115.184119
- Kelleher, E.S., Edelman, N.B., Barbash, D.A., 2012. *Drosophila* Interspecific Hybrids Phenocopy piRNA-Pathway Mutants. *PLoS Biol.* 10, e1001428. doi:10.1371/journal.pbio.1001428
- Kelly, W.G., Aramayo, R., 2007. Meiotic silencing and the epigenetics of sex. *Chromosome Res. Int. J. Mol. Supramol. Evol. Asp. Chromosome Biol.* 15, 633–651. doi:10.1007/s10577-007-1143-0
- Kenan-Eichler, M., Leshkowitz, D., Tal, L., Noor, E., Melamed-Bessudo, C., Feldman, M., Levy, A.A., 2011. Wheat Hybridization and Polyploidization Results in Deregulation of Small RNAs. *Genetics* 188, 263–272. doi:10.1534/genetics.111.128348
- Khurana, J.S., Wang, J., Xu, J., Koppetsch, B.S., Thomson, T.C., Nowosielska, A., Li, C., Zamore, P.D., Weng, Z., Theurkauf, W.E., 2011. Adaptation to P Element Transposon Invasion in *Drosophila melanogaster*. *Cell* 147, 1551–1563. doi:10.1016/j.cell.2011.11.042
- Kidwell, M.G., Kidwell, J.F., Sved, J.A., 1977. Hybrid Dysgenesis in *Drosophila Melanogaster*: A Syndrome of Aberrant Traits Including Mutation, Sterility and Male Recombination. *Genetics* 86, 813–833.
- Klattenhoff, C., Xi, H., Li, C., Lee, S., Xu, J., Khurana, J.S., Zhang, F., Schultz, N., Koppetsch, B.S., Nowosielska, A., Seitz, H., Zamore, P.D., Weng, Z., Theurkauf, W.E., 2009. The *Drosophila* HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. *Cell* 138, 1137–1149. doi:10.1016/j.cell.2009.07.014
- Klenov, M.S., Sokolova, O.A., Yakushev, E.Y., Stolyarenko, A.D., Mikhaleva, E.A., Lavrov, S.A., Gvozdev, V.A., 2011. Separation of stem cell maintenance and transposon silencing functions of Piwi protein. *Proc. Natl. Acad. Sci.* 108, 18760–18765. doi:10.1073/pnas.1106676108
- Klobutcher, L.A., Herrick, G., 1997. Developmental genome reorganization in ciliated protozoa: the transposon link. *Prog. Nucleic Acid Res. Mol. Biol.* 56, 1–62.

- Koizumi, S., Kobayashi, S., 1989. Microinjection of plasmid DNA encoding the A surface antigen of *Paramecium tetraurelia* restores the ability to regenerate a wild-type macronucleus. *Mol. Cell. Biol.* 9, 4398–4401.
- Koonin, E.V., Krupovic, M., 2015. Evolution of adaptive immunity from transposable elements combined with innate immune systems. *Nat. Rev. Genet.* 16, 184–192. doi:10.1038/nrg3859
- Labrador, M., Farré, M., Utzet, F., Fontdevila, A., 1999. Interspecific hybridization increases transposition rates of *Osvaldo*. *Mol. Biol. Evol.* 16, 931–937.
- Labrie, S.J., Samson, J.E., Moineau, S., 2010. Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.* 8, 317–327. doi:10.1038/nrmicro2315
- Le Mouel, A., Butler, A., Caron, F., Meyer, E., 2003. Developmentally Regulated Chromosome Fragmentation Linked to Imprecise Elimination of Repeated Sequences in *Paramecia*. *Eukaryot. Cell* 2, 1076–1090. doi:10.1128/EC.2.5.1076-1090.2003
- Le Thomas, A., Rogers, A.K., Webster, A., Marinov, G.K., Liao, S.E., Perkins, E.M., Hur, J.K., Aravin, A.A., Tóth, K.F., 2013. Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes Dev.* 27, 390–399. doi:10.1101/gad.209841.112
- Le Thomas, A., Tóth, K.F., Aravin, A.A., 2014. To be or not to be a piRNA: genomic origin and processing of piRNAs. *Genome Biol.* 15, 204. doi:10.1186/gb4154
- Lee, D.W., Seong, K.-Y., Pratt, R.J., Baker, K., Aramayo, R., 2004. Properties of unpaired DNA required for efficient silencing in *Neurospora crassa*. *Genetics* 167, 131–150.
- Lee, H.-C., Chang, S.-S., Choudhary, S., Aalto, A.P., Maiti, M., Bamford, D.H., Liu, Y., 2009. qiRNA is a new type of small interfering RNA induced by DNA damage. *Nature* 459, 274–277. doi:10.1038/nature08041
- Lee, H.-C., Gu, W., Shirayama, M., Youngman, E., Conte, D., Mello, C.C., 2012. *C. elegans* piRNAs Mediate the Genome-wide Surveillance of Germline Transcripts. *Cell* 150, 78–87. doi:10.1016/j.cell.2012.06.016
- Lepere, G., Betermier, M., Meyer, E., Duharcourt, S., 2008. Maternal noncoding transcripts antagonize the targeting of DNA elimination by scanRNAs in *Paramecium tetraurelia*. *Genes Dev.* 22, 1501–1512. doi:10.1101/gad.473008
- Lepere, G., Nowacki, M., Serrano, V., Gout, J.-F., Guglielmi, G., Duharcourt, S., Meyer, E., 2009. Silencing-associated and meiosis-specific small RNA pathways in *Paramecium tetraurelia*. *Nucleic Acids Res.* 37, 903–915. doi:10.1093/nar/gkn1018
- Lerat, E., Capy, P., 1999. Retrotransposons and retroviruses: analysis of the envelope gene. *Mol. Biol. Evol.* 16, 1198–1207.
- Levin, H.L., Moran, J.V., 2011. Dynamic interactions between transposable elements and their hosts. *Nat. Rev. Genet.* 12, 615–627. doi:10.1038/nrg3030
- Levis, R.W., Ganesan, R., Houtchens, K., Tolar, L.A., Sheen, F.M., 1993. Transposons in place of telomeric repeats at a *Drosophila* telomere. *Cell* 75, 1083–1093.

- Lhuillier-Akakpo, M., Frapporti, A., Denby Wilkes, C., Matelot, M., Vervoort, M., Sperling, L., Duharcourt, S., 2014. Local Effect of Enhancer of Zeste-Like Reveals Cooperation of Epigenetic and cis-Acting Determinants for Zygotic Genome Rearrangements. *PLoS Genet.* 10, e1004665. doi:10.1371/journal.pgen.1004665
- Li, S., Liu, L., Zhuang, X., Yu, Y., Liu, X., Cui, X., Ji, L., Pan, Z., Cao, X., Mo, B., Zhang, F., Raikhel, N., Jiang, L., Chen, X., 2013. microRNAs inhibit the translation of target mRNAs on the endoplasmic reticulum in Arabidopsis. *Cell* 153, 562–574. doi:10.1016/j.cell.2013.04.005
- Liu, Y., Mochizuki, K., Gorovsky, M.A., 2004. Histone H3 lysine 9 methylation is required for DNA elimination in developing macronuclei in Tetrahymena. *Proc. Natl. Acad. Sci. U. S. A.* 101, 1679–1684. doi:10.1073/pnas.0305421101
- Liu, Y., Taverna, S.D., Muratore, T.L., Shabanowitz, J., Hunt, D.F., Allis, C.D., 2007. RNAi-dependent H3K27 methylation is required for heterochromatin formation and DNA elimination in Tetrahymena. *Genes Dev.* 21, 1530–1545. doi:10.1101/gad.1544207
- Loidl, J., Scherthan, H., 2004. Organization and pairing of meiotic chromosomes in the ciliate Tetrahymena thermophila. *J. Cell Sci.* 117, 5791–5801. doi:10.1242/jcs.01504
- Lozovskaya, E.R., Scheinker, V.S., Evgen'ev, M.B., 1990. A hybrid dysgenesis syndrome in *Drosophila virilis*. *Genetics* 126, 619–623.
- Lynch, M., Force, A.G., 2000. The Origin of Interspecific Genomic Incompatibility via Gene Duplication. *Am. Nat.* 156, 590–605. doi:10.1086/316992
- Madireddi, M.T., Coyne, R.S., Smothers, J.F., Mickey, K.M., Yao, M.C., Allis, C.D., 1996. Pdd1p, a novel chromodomain-containing protein, links heterochromatin assembly and DNA elimination in Tetrahymena. *Cell* 87, 75–84.
- Maheshwari, S., Barbash, D.A., 2011. The Genetics of Hybrid Incompatibilities. *Annu. Rev. Genet.* 45, 331–355. doi:10.1146/annurev-genet-110410-132514
- Makarova, K.S., Wolf, Y.I., Forterre, P., Prangishvili, D., Krupovic, M., Koonin, E.V., 2014. Dark matter in archaeal genomes: a rich source of novel mobile elements, defense systems and secretory complexes. *Extremophiles* 18, 877–893. doi:10.1007/s00792-014-0672-7
- Maliszewska-Olejniczak, K., Gruchota, J., Gromadka, R., Denby Wilkes, C., Arnaiz, O., Mathy, N., Duharcourt, S., Bétermier, M., Nowak, J.K., 2015. TFIIS-Dependent Non-coding Transcription Regulates Developmental Genome Rearrangements. *PLoS Genet.* 11. doi:10.1371/journal.pgen.1005383
- Malone, C.D., Anderson, A.M., Motl, J.A., Rexer, C.H., Chalker, D.L., 2005. Germ Line Transcripts Are Processed by a Dicer-Like Protein That Is Essential for Developmentally Programmed Genome Rearrangements of Tetrahymena thermophila. *Mol. Cell. Biol.* 25, 9151–9164. doi:10.1128/MCB.25.20.9151-9164.2005
- Malone, C.D., Lehmann, R., Teixeira, F.K., 2015. The cellular basis of hybrid dysgenesis and Stellate regulation in *Drosophila*. *Curr. Opin. Genet. Dev.* 34, 88–94. doi:10.1016/j.gde.2015.09.003
- Marahrens, Y., 1999. X-inactivation by chromosomal pairing events. *Genes Dev.* 13, 2624–2632.

- Marker, S., Carradec, Q., Tanty, V., Arnaiz, O., Meyer, E., 2014. A forward genetic screen reveals essential and non-essential RNAi factors in *Paramecium tetraurelia*. *Nucleic Acids Res.* 42, 7268–7280. doi:10.1093/nar/gku223
- Matranga, C., Tomari, Y., Shin, C., Bartel, D.P., Zamore, P.D., 2005. Passenger-Strand Cleavage Facilitates Assembly of siRNA into Ago2-Containing RNAi Enzyme Complexes. *Cell* 123, 607–620. doi:10.1016/j.cell.2005.08.044
- Matsuda, A., Forney, J.D., 2006. The SUMO Pathway Is Developmentally Regulated and Required for Programmed DNA Elimination in *Paramecium tetraurelia*. *Eukaryot. Cell* 5, 806–815. doi:10.1128/EC.5.5.806-815.2006
- McClintock, B., 1984. The significance of responses of the genome to challenge. *Science* 226, 792–801. doi:10.1126/science.15739260
- McClintock, B., 1950. The Origin and Behavior of Mutable Loci in Maize. *Proc. Natl. Acad. Sci. U. S. A.* 36, 344–355.
- McGrath, C.L., Gout, J.-F., Doak, T.G., Yanagi, A., Lynch, M., 2014. Insights into Three Whole-Genome Duplications Gleaned from the *Paramecium caudatum* Genome Sequence. *Genetics* 197, 1417–1428. doi:10.1534/genetics.114.163287
- McGrath, C.L., Gout, J.-F., Johri, P., Doak, T.G., Lynch, M., 2014. Differential retention and divergent resolution of duplicate genes following whole-genome duplication. *Genome Res.* 24, 1665–1675. doi:10.1101/gr.173740.114
- McGrath, S., Seegers, J.F., Fitzgerald, G.F., van Sinderen, D., 1999. Molecular characterization of a phage-encoded resistance system in *Lactococcus lactis*. *Appl. Environ. Microbiol.* 65, 1891–1899.
- Miller, D.D., 1950. Observations on Two Cases of Interspecific Hybridization with *Drosophila athabasca*. *Am. Nat.* 84, 81–93.
- Mochizuki, K., Fine, N.A., Fujisawa, T., Gorovsky, M.A., 2002. Analysis of a *piwi*-Related Gene Implicates Small RNAs in Genome Rearrangement in *Tetrahymena*. *Cell* 110, 689–699.
- Mochizuki, K., Gorovsky, M.A., 2004. Conjugation-specific small RNAs in *Tetrahymena* have predicted properties of scan (scn) RNAs involved in genome rearrangement. *Genes Dev.* 18, 2068–2073. doi:10.1101/gad.1219904
- Montgomery, T.A., Rim, Y.-S., Zhang, C., Downen, R.H., Phillips, C.M., Fischer, S.E.J., Ruvkun, G., 2012. PIWI Associated siRNAs and piRNAs Specifically Require the *Caenorhabditis elegans* HEN1 Ortholog henn-1. *PLOS Genet* 8, e1002616. doi:10.1371/journal.pgen.1002616
- Mosher, R.A., Melnyk, C.W., Kelly, K.A., Dunn, R.M., Studholme, D.J., Baulcombe, D.C., 2009. Uniparental expression of PolIV-dependent siRNAs in developing endosperm of *Arabidopsis*. *Nature* 460, 283–286. doi:10.1038/nature08084
- Naveira, H., Fontdevila, A., 1985. The evolutionary history of *Drosophila buzzatii*. IX. High frequencies of new chromosome rearrangements induced by introgressive hybridization. *Chromosoma* 91, 87–94.

- Nitasaka, E., Mukai, T., Yamazaki, T., 1987. Repressor of P elements in *Drosophila melanogaster*: Cytotype determination by a defective P element carrying only open reading frames 0 through 2. *Proc. Natl. Acad. Sci. U. S. A.* 84, 7605–7608.
- Noland, C.L., Doudna, J.A., 2013. Multiple sensors ensure guide strand selection in human RNAi pathways. *RNA* 19, 639–648. doi:10.1261/rna.037424.112
- Noto, T., Kataoka, K., Suhren, J.H., Hayashi, A., Woolcock, K.J., Gorovsky, M.A., Mochizuki, K., 2015. Small-RNA-Mediated Genome-wide trans-Recognition Network in *Tetrahymena* DNA Elimination. *Mol. Cell.* doi:10.1016/j.molcel.2015.05.024
- Noto, T., Kurth, H.M., Mochizuki, K., 2014. Analysis of Piwi-loaded small RNAs in *Tetrahymena*. *Methods Mol. Biol. Clifton NJ* 1093, 209–224. doi:10.1007/978-1-62703-694-8\_17
- Nowacki, M., Higgins, B.P., Maquilan, G.M., Swart, E.C., Doak, T.G., Landweber, L.F., 2009. A Functional Role for Transposases in a Large Eukaryotic Genome. *Science* 324, 935–938. doi:10.1126/science.1170023
- Nowacki, M., Zagorski-Ostojka, W., Meyer, E., 2005. Nowa1p and Nowa2p: Novel Putative RNA Binding Proteins Involved in trans-Nuclear Crosstalk in *Paramecium tetraurelia*. *Curr. Biol.* 15, 1616–1628. doi:10.1016/j.cub.2005.07.033
- Nowak, J.K., Gromadka, R., Juszczuk, M., Jerka-Dziadosz, M., Maliszewska, K., Mucchielli, M.-H., Gout, J.-F., Arnaiz, O., Agier, N., Tang, T., Aggerbeck, L.P., Cohen, J., Delacroix, H., Sperling, L., Herbert, C.J., Zagulski, M., Bétermier, M., 2011. Functional Study of Genes Essential for Autogamy and Nuclear Reorganization in *Paramecium*. *Eukaryot. Cell* 10, 363–372. doi:10.1128/EC.00258-10
- Ochman, H., Lawrence, J.G., Groisman, E.A., 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature* 405, 299–304. doi:10.1038/35012500
- Ohno, D.S., 1970. The Creation of a New Gene from a Redundant Duplicate of an Old Gene, in: *Evolution by Gene Duplication*. Springer Berlin Heidelberg, pp. 71–82. doi:10.1007/978-3-642-86659-3\_14
- Okamura, K., Ishizuka, A., Siomi, H., Siomi, M.C., 2004. Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways. *Genes Dev.* 18, 1655–1666. doi:10.1101/gad.1210204
- O'Neill, R.J.W., O'Neill, M.J., Graves, J.A.M., 1998. Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. *Nature* 393, 68–72. doi:10.1038/29985
- Orr, H.A., 1996. Dobzhansky, Bateson, and the Genetics of Speciation. *Genetics* 144, 1331–1335.
- Orsi, G.A., Joyce, E.F., Couble, P., McKim, K.S., Loppin, B., 2010. *Drosophila* I-R hybrid dysgenesis is associated with catastrophic meiosis and abnormal zygote formation. *J Cell Sci* 123, 3515–3524. doi:10.1242/jcs.073890

- Pape, R., Haacke-Bell, B., Lüthe, N., Plattner, H., 1988. Conjugation of *Paramecium tetraurelia* cells: selective wheat germ agglutinin binding, reversible local trichocyst detachment and secretory function repair. *J. Cell Sci.* 90, 37–49.
- Parisod, C., Alix, K., Just, J., Petit, M., Sarilar, V., Mhiri, C., Ainouche, M., Chalhoub, B., Grandbastien, M.-A., 2010. Impact of transposable elements on the organization and function of allopolyploid genomes. *New Phytol.* 186, 37–45. doi:10.1111/j.1469-8137.2009.03096.x
- Parisod, C., Senerchia, N., 2012. Responses of transposable elements to polyploidy, in: *Plant Transposable Elements*. Springer, pp. 147–168.
- Petrov, D.A., Schutzman, J.L., Hartl, D.L., Lozovskaya, E.R., 1995. Diverse transposable elements are mobilized in hybrid dysgenesis in *Drosophila virilis*. *Proc. Natl. Acad. Sci. U. S. A.* 92, 8050–8054.
- Pickford, A.S., Catalanotto, C., Cogoni, C., Macino, G., 2002. Quelling in *Neurospora crassa*. *Adv. Genet.* 46, 277–303.
- Pratt, R.J., Lee, D.W., Aramayo, R., 2004. DNA Methylation Affects Meiotic trans-sensing, Not Meiotic Silencing, in *Neurospora*. *Genetics* 168, 1925–1935. doi:10.1534/genetics.104.031526
- Preer, J.R., Preer, L.B., Rudman, B.M., Barnett, A.J., 1985. Deviation from the universal code shown by the gene for surface protein 51A in *Paramecium*. *Nature* 314, 188–190.
- Preer, L.B., Hamilton, G., Preer, J.R., 1992. Micronuclear DNA from *Paramecium tetraurelia*: serotype 51 A gene has internally eliminated sequences. *J. Protozool.* 39, 678–682.
- Prescott, D.M., 1994. The DNA of ciliated protozoa. *Microbiol. Rev.* 58, 233–267.
- Rechavi, O., 2014. Guest list or black list: heritable small RNAs as immunogenic memories. *Trends Cell Biol.* 24, 212–220. doi:10.1016/j.tcb.2013.10.003
- Reinders, J., Mirouze, M., Nicolet, J., Paszkowski, J., 2013. Parent- of- origin control of transgenerational retrotransposon proliferation in *Arabidopsis*. *EMBO Rep.* 14, 823–828. doi:10.1038/embor.2013.95
- Roche, S.E., Schiff, M., Rio, D.C., 1995. P-element repressor autoregulation involves germline transcriptional repression and reduction of third intron splicing. *Genes Dev.* 9, 1278–1288.
- Ronsseray, S., Lehmann, M., Anxolabéhère, D., 1991. The maternally inherited regulation of P elements in *Drosophila melanogaster* can be elicited by two P copies at cytological site 1A on the X chromosome. *Genetics* 129, 501–512.
- Ronsseray, S., Lemaitre, B., Coen, D., 1993. Maternal inheritance of P cytotype in *Drosophila melanogaster*: a “pre-P cytotype” is strictly extra-chromosomally transmitted. *Mol. Gen. Genet.* MGG 241, 115–123.
- Rose, M.R., Doolittle, W.F., 1983. *Molecular Biological Mechanisms of Speciation*. *Science* 220, 157–162. doi:10.1126/science.220.4593.157

- Rozhkov, N.V., Aravin, A.A., Zelentsova, E.S., Schostak, N.G., Sachidanandam, R., McCombie, W.R., Hannon, G.J., Evgen'ev, M.B., 2010. Small RNA-based silencing strategies for transposons in the process of invading *Drosophila* species. *RNA* 16, 1634–1645. doi:10.1261/rna.2217810
- Rozhkov, N.V., Schostak, N.G., Zelentsova, E.S., Yushenova, I.A., Zatssepina, O.G., Evgen'ev, M.B., 2013. Evolution and Dynamics of Small RNA Response to a Retroelement Invasion in *Drosophila*. *Mol. Biol. Evol.* 30, 397–408. doi:10.1093/molbev/mss241
- Ruby, J.G., Jan, C., Player, C., Axtell, M.J., Lee, W., Nusbaum, C., Ge, H., Bartel, D.P., 2006. Large-Scale Sequencing Reveals 21U-RNAs and Additional MicroRNAs and Endogenous siRNAs in *C. elegans*. *Cell* 127, 1193–1207. doi:10.1016/j.cell.2006.10.040
- Rudman, B., Preer, L.B., Polisky, B., Preer, J.R., 1991. Mutants affecting processing of DNA in macronuclear development in paramecium. *Genetics* 129, 47–56.
- Ruiz, F., Vayssié, L., Klotz, C., Sperling, L., Madeddu, L., 1998. Homology-dependent Gene Silencing in Paramecium. *Mol. Biol. Cell* 9, 931–943. doi:10.1091/mbc.9.4.931
- Saito, K., Sakaguchi, Y., Suzuki, T., Suzuki, T., Siomi, H., Siomi, M.C., 2007. Pimet, the *Drosophila* homolog of HEN1, mediates 2'-O-methylation of Piwi- interacting RNAs at their 3' ends. *Genes Dev.* 21, 1603–1608. doi:10.1101/gad.1563607
- Sandoval, P.Y., Swart, E.C., Arambasic, M., Nowacki, M., 2014. Functional Diversification of Dicer-like Proteins and Small RNAs Required for Genome Sculpting. *Dev. Cell* 28, 174–188. doi:10.1016/j.devcel.2013.12.010
- Schaack, S., Gilbert, C., Feschotte, C., 2010. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends Ecol. Evol.* 25, 537–546. doi:10.1016/j.tree.2010.06.001
- Schoeberl, U.E., Kurth, H.M., Noto, T., Mochizuki, K., 2012. Biased transcription and selective degradation of small RNAs shape the pattern of DNA elimination in *Tetrahymena*. *Genes Dev.* 26, 1729–1742. doi:10.1101/gad.196493.112
- Schorderet, D.F., Gartler, S.M., 1992. Analysis of CpG suppression in methylated and nonmethylated species. *Proc. Natl. Acad. Sci. U. S. A.* 89, 957–961.
- Selker, E.U., 1990. Premeiotic instability of repeated sequences in *Neurospora crassa*. *Annu. Rev. Genet.* 24, 579–613. doi:10.1146/annurev.ge.24.120190.003051
- Selker, E.U., Tountas, N.A., Cross, S.H., Margolin, B.S., Murphy, J.G., Bird, A.P., Freitag, M., 2003. The methylated component of the *Neurospora crassa* genome. *Nature* 422, 893–897. doi:10.1038/nature01564
- Seth, M., Shirayama, M., Gu, W., Ishidate, T., Conte, D., Mello, C.C., 2013. The *C. elegans* CSR-1 Argonaute pathway counteracts epigenetic silencing to promote germline gene expression. *Dev. Cell* 27, 656–663. doi:10.1016/j.devcel.2013.11.014
- Shirayama, M., Seth, M., Lee, H.-C., Gu, W., Ishidate, T., Conte, D., Mello, C.C., 2012. piRNAs Initiate an Epigenetic Memory of Nonself RNA in the *C. elegans* Germline. *Cell* 150, 65–77. doi:10.1016/j.cell.2012.06.015

- Shiu, P.K.T., Raju, N.B., Zickler, D., Metzberg, R.L., 2001. Meiotic Silencing by Unpaired DNA. *Cell* 107, 905–916. doi:10.1016/S0092-8674(01)00609-2
- Simmons, M.J., Thorp, M.W., Buschette, J.T., Peterson, K., Cross, E.W., Bjorklund, E.L., 2010. Maternal impairment of transposon regulation in *Drosophila melanogaster* by mutations in the genes *aubergine*, *piwi* and *Suppressor of variegation 205*. *Genet. Res.* 92, 261–272. doi:10.1017/S0016672310000352
- Singh, D.P., Saudemont, B., Guglielmi, G., Arnaiz, O., Goût, J.-F., Prajer, M., Potekhin, A., Przybòs, E., Aubusson-Fleury, A., Bhullar, S., Bouhouche, K., Lhuillier-Akakpo, M., Tanty, V., Blugeon, C., Alberti, A., Labadie, K., Aury, J.-M., Sperling, L., Duharcourt, S., Meyer, E., 2014. Genome-defence small RNAs exapted for epigenetic mating-type inheritance. *Nature* 509, 447–452. doi:10.1038/nature13318
- Skouri, F., Cohen, J., 1997. Genetic approach to regulated exocytosis using functional complementation in *Paramecium*: identification of the ND7 gene required for membrane fusion. *Mol. Biol. Cell* 8, 1063–1071.
- Slotkin, R.K., Vaughn, M., Borges, F., Tanurdžić, M., Becker, J.D., Feijó, J.A., Martienssen, R.A., 2009. Epigenetic Reprogramming and Small RNA Silencing of Transposable Elements in Pollen. *Cell* 136, 461–472. doi:10.1016/j.cell.2008.12.038
- Sonneborn, T.M., 1950. Methods in the general biology and genetics of *paramecium aurelia*. *J. Exp. Zool.* 113, 87–147. doi:10.1002/jez.1401130106
- Sonneborn, T.M., 1947. Recent advances in the genetics of *Paramecium* and *Euplotes*. *Adv. Genet.* 1, 263–358.
- Sonneborn, T.M., 1937. Sex, sex inheritance and sex determination in *Paramecium aurelia*. *Proc. Natl. Acad. Sci. U. S. A.* 23, 378.
- Stevens, N.M., 1908. A study of the germ cells of certain diptera, with reference to the heterochromosomes and the phenomena of synapsis. *J. Exp. Zool.* 5, 359–374. doi:10.1002/jez.1400050304
- Stuart, J.R., Haley, K.J., Swedzinski, D., Lockner, S., Kocian, P.E., Merriman, P.J., Simmons, M.J., 2002. Telomeric P elements associated with cytotypic regulation of the P transposon family in *Drosophila melanogaster*. *Genetics* 162, 1641–1654.
- Sturtevant, A.H., 1939. High Mutation Frequency Induced by Hybridization. *Proc. Natl. Acad. Sci. U. S. A.* 25, 308.
- Sun, C., Feschotte, C., Wu, Z., Mueller, R.L., 2015. DNA transposons have colonized the genome of the giant virus *Pandoravirus salinus*. *BMC Biol.* 13, 38. doi:10.1186/s12915-015-0145-1
- Sweet, M.T., Allis, C.D., 2006. Transformation of *Tetrahymena thermophila* by Electroporation. *CSH Protoc.* 2006. doi:10.1101/pdb.prot4502
- Symonová, R., Majtánová, Z., Sember, A., Staaks, G.B., Bohlen, J., Freyhof, J., Rábová, M., Ráb, P., 2013. Genome differentiation in a species pair of coregonine fishes: an extremely rapid speciation driven by stress-activated retrotransposons mediating extensive ribosomal DNA multiplications. *BMC Evol. Biol.* 13, 42. doi:10.1186/1471-2148-13-42

- Tang, W., Tu, S., Lee, H.-C., Weng, Z., Mello, C.C., 2016. The RNase PARN-1 Trims piRNA 3' Ends to Promote Transcriptome Surveillance in *C. elegans*. *Cell* 164, 974–984. doi:10.1016/j.cell.2016.02.008
- Taverna, S.D., Coyne, R.S., Allis, C.D., 2002. Methylation of Histone H3 at Lysine 9 Targets Programmed DNA Elimination in *Tetrahymena*. *Cell* 110, 701–711. doi:10.1016/S0092-8674(02)00941-8
- Templeton, A.R., 1981. Mechanisms of Speciation--A Population Genetic Approach. *Annu. Rev. Ecol. Syst.* 12, 23–48.
- Thomas, C.M., Nielsen, K.M., 2005. Mechanisms of, and Barriers to, Horizontal Gene Transfer between Bacteria. *Nat. Rev. Microbiol.* 3, 711–721. doi:10.1038/nrmicro1234
- Thompson, J.R., Marcelino, L.A., Polz, M.F., 2002. Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by “reconditioning PCR.” *Nucleic Acids Res.* 30, 2083–2088. doi:10.1093/nar/30.9.2083
- Valentine, M.S., Rajendran, A., Yano, J., Weeraratne, S.D., Beisson, J., Cohen, J., Koll, F., Van Houten, J., 2012. *Paramecium* BBS genes are key to presence of channels in Cilia. *Cilia* 1, 16. doi:10.1186/2046-2530-1-16
- Verdel, A., Jia, S., Gerber, S., Sugiyama, T., Gygi, S., Grewal, S.I.S., Moazed, D., 2004. RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* 303, 672–676. doi:10.1126/science.1093686
- Vilkaitis, G., Klimasauskas, S., 1999. Bisulfite sequencing protocol displays both 5-methylcytosine and N4-methylcytosine. *Anal. Biochem.* 271, 116–119. doi:10.1006/abio.1999.4116
- Vu, W., Nuzhdin, S., 2011. Genetic variation of copia suppression in *Drosophila melanogaster*. *Heredity* 106, 207–217. doi:10.1038/hdy.2010.41
- Wallau, G.L., Hua-Van, A., Capy, P., Loreto, E.L.S., 2011. The evolutionary history of mariner-like elements in Neotropical drosophilids. *Genetica* 139, 327–338. doi:10.1007/s10709-011-9552-6
- Wassenegger, M., Heimes, S., Riedel, L., Sanger, H.L., 1994. RNA-directed de novo methylation of genomic sequences in plants. *Cell* 76, 567–576.
- Wedeles, C.J., Wu, M.Z., Claycomb, J.M., 2013. Protection of Germline Gene Expression by the *C. elegans* Argonaute CSR-1. *Dev. Cell* 27, 664–671. doi:10.1016/j.devcel.2013.11.016
- Weick, E.-M., Miska, E.A., 2014. piRNAs: from biogenesis to function. *Development* 141, 3458–3471. doi:10.1242/dev.094037
- Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., Schulman, A.H., 2007. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8, 973–982. doi:10.1038/nrg2165
- Yao, M.C., 1982. Elimination of specific DNA sequences from the somatic nucleus of the ciliate *Tetrahymena*. *J. Cell Biol.* 92, 783–789.

Yao, M.-C., Yao, C.-H., Monks, B., 1990. The controlling sequence for site-specific chromosome breakage in tetrahymena. *Cell* 63, 763–772. doi:10.1016/0092-8674(90)90142-2

Yokoyama, R.W., Yao, M.C., 1982. Elimination of DNA sequences during macronuclear differentiation in *Tetrahymena thermophila*, as detected by in situ hybridization. *Chromosoma* 85, 11–22.

Zahler, A.M., Neeb, Z.T., Lin, A., Katzman, S., 2012. Mating of the Stichotrichous Ciliate *Oxytricha trifallax* Induces Production of a Class of 27 nt Small RNAs Derived from the Parental Macronucleus. *PLoS ONE* 7. doi:10.1371/journal.pone.0042371