

**VETAGRO SUP
CAMPUS VETERINAIRE DE LYON**

Année 2018 - Thèse n°089

***ETUDE DE L'EVOLUTION D'UN ELEMENT TRANSPOSABLE
DU GENOME DE LA CHAUVE-SOURIS PTEROPUS
GIGANTEUS, ESPECE RESERVOIR DE ZOONOSES VIRALES***

THESE

Présentée à l'UNIVERSITE CLAUDE-BERNARD - LYON I
(Médecine - Pharmacie)
et soutenue publiquement le 23 novembre 2018
pour obtenir le grade de Docteur Vétérinaire

par

COLLET Mathilde



**VETAGRO SUP
CAMPUS VETERINAIRE DE LYON**

Année 2018 - Thèse n°089

***ETUDE DE L'EVOLUTION D'UN ELEMENT TRANSPOSABLE
DU GENOME DE LA CHAUVE-SOURIS PTEROPUS
GIGANTEUS, ESPECE RESERVOIR DE ZOONOSES VIRALES***

THESE

Présentée à l'UNIVERSITE CLAUDE-BERNARD - LYON I
(Médecine - Pharmacie)
et soutenue publiquement le 23 novembre 2018
pour obtenir le grade de Docteur Vétérinaire

par

COLLET Mathilde



Liste des Enseignants du Campus Vétérinaire de Lyon (1er mars 2018)

Nom	Prénom	Département	Grade
ABITBOL	Marie	DEPT-BASIC-SCIENCES	Maître de conférences
ALVES-DE-OLIVEIRA	Laurent	DEPT-BASIC-SCIENCES	Maître de conférences
ARCANGIOLI	Marie-Anne	DEPT-ELEVAGE-SPV	Professeur
AYRAL	Florence	DEPT-ELEVAGE-SPV	Maître de conférences
BECKER	Claire	DEPT-ELEVAGE-SPV	Maître de conférences
BELLUCO	Sara	DEPT-AC-LOISIR-SPORT	Maître de conférences
BENAMOU-SMITH	Agnès	DEPT-AC-LOISIR-SPORT	Maître de conférences
BENOIT	Etienne	DEPT-BASIC-SCIENCES	Professeur
BERNY	Philippe	DEPT-BASIC-SCIENCES	Professeur
BONNET-GARIN	Jeanne-Marie	DEPT-BASIC-SCIENCES	Professeur
BOULOCHER	Caroline	DEPT-BASIC-SCIENCES	Maître de conférences
BOURDOISEAU	Gilles	DEPT-ELEVAGE-SPV	Professeur
BOURGOIN	Gilles	DEPT-ELEVAGE-SPV	Maître de conférences
BRUYERE	Pierre	DEPT-BASIC-SCIENCES	Maître de conférences
BUFF	Samuel	DEPT-BASIC-SCIENCES	Maître de conférences
BURONFOSSE	Thierry	DEPT-BASIC-SCIENCES	Professeur
CACHON	Thibaut	DEPT-AC-LOISIR-SPORT	Maître de conférences
CADORÉ	Jean-Luc	DEPT-AC-LOISIR-SPORT	Professeur
CALLAIT-CARDINAL	Marie-Pierre	DEPT-ELEVAGE-SPV	Maître de conférences
CAROZZO	Claude	DEPT-AC-LOISIR-SPORT	Maître de conférences
CHABANNE	Luc	DEPT-AC-LOISIR-SPORT	Professeur
CHALVET-MONFRAY	Karine	DEPT-BASIC-SCIENCES	Professeur
DE BOYER DES ROCHES	Alice	DEPT-ELEVAGE-SPV	Maître de conférences
DELIGNETTE-MULLER	Marie-Laure	DEPT-BASIC-SCIENCES	Professeur
DEMONT	Pierre	DEPT-ELEVAGE-SPV	Professeur
DJELOUADJI	Zorée	DEPT-ELEVAGE-SPV	Maître de conférences
ESCRIOU	Catherine	DEPT-AC-LOISIR-SPORT	Maître de conférences
FRIKHA	Mohamed-Ridha	DEPT-ELEVAGE-SPV	Maître de conférences
GALIA	Wessam	DEPT-ELEVAGE-SPV	Maître de conférences Stagiaire
GILOT-FROMONT	Emmanuelle	DEPT-ELEVAGE-SPV	Professeur
GONTHIER	Alain	DEPT-ELEVAGE-SPV	Maître de conférences
GRANCHER	Denis	DEPT-BASIC-SCIENCES	Maître de conférences
GREZEL	Delphine	DEPT-BASIC-SCIENCES	Maître de conférences
HUGONNARD	Marine	DEPT-AC-LOISIR-SPORT	Maître de conférences
JANKOWIAK	Bernard	DEPT-ELEVAGE-SPV	Maître de conférences Contractuel
JAUSSAUD	Philippe	DEPT-BASIC-SCIENCES	Professeur
JEANNIN	Anne	DEPT-ELEVAGE-SPV	Inspecteur en santé publique vétérinaire (ISPV)
JOSSON-SCHRAMME	Anne	DEPT-BASIC-SCIENCES	Maître de conférences Contractuel
JUNOT	Stéphane	DEPT-AC-LOISIR-SPORT	Maître de conférences
KODJO	Angeli	DEPT-ELEVAGE-SPV	Professeur
KRAFFT	Emilie	DEPT-AC-LOISIR-SPORT	Maître de conférences
LAABERKI	Maria-Halima	DEPT-ELEVAGE-SPV	Maître de conférences
LAMBERT	Véronique	DEPT-BASIC-SCIENCES	Maître de conférences
LE GRAND	Dominique	DEPT-ELEVAGE-SPV	Professeur
LEBLOND	Agnès	DEPT-AC-LOISIR-SPORT	Professeur
LEDOUX	Dorothee	DEPT-ELEVAGE-SPV	Maître de conférences Stagiaire
LEFEBVRE	Sébastien	DEPT-BASIC-SCIENCES	Maître de conférences Stagiaire
LEFRANC-POHL	Anne-Cécile	DEPT-BASIC-SCIENCES	Maître de conférences
LEPAGE	Olivier	DEPT-AC-LOISIR-SPORT	Professeur
LOUZIER	Vanessa	DEPT-BASIC-SCIENCES	Professeur
MARCHAL	Thierry	DEPT-AC-LOISIR-SPORT	Professeur
MATEOS	Stevana	DEPT-AC-LOISIR-SPORT	Maître de conférences Contractuel
MOISSONNIER	Pierre	DEPT-AC-LOISIR-SPORT	Professeur
MOUNIER	Luc	DEPT-ELEVAGE-SPV	Professeur
PEPIN	Michel	DEPT-BASIC-SCIENCES	Professeur
PIN	Didier	DEPT-AC-LOISIR-SPORT	Professeur
PONCE	Frédérique	DEPT-AC-LOISIR-SPORT	Professeur
PORTIER	Karine	DEPT-AC-LOISIR-SPORT	Professeur
POUZOT-NEVORET	Céline	DEPT-AC-LOISIR-SPORT	Maître de conférences
PROUILLAC	Caroline	DEPT-BASIC-SCIENCES	Maître de conférences
REMY	Denise	DEPT-AC-LOISIR-SPORT	Professeur
RENE MARTELLET	Magalie	DEPT-ELEVAGE-SPV	Maître de conférences
RIVES	Germain	DEPT-ELEVAGE-SPV	Maître de conférences Contractuel
ROGER	Thierry	DEPT-BASIC-SCIENCES	Professeur
SABATIER	Philippe	DEPT-ELEVAGE-SPV	Professeur
SAWAYA	Serge	DEPT-BASIC-SCIENCES	Maître de conférences
SCHRAMME	Michael	DEPT-AC-LOISIR-SPORT	Professeur
SERGENTET	Delphine	DEPT-ELEVAGE-SPV	Professeur
THIEBAULT	Jean-Jacques	DEPT-BASIC-SCIENCES	Maître de conférences
THOMAS-CANCIAN	Auréli	DEPT-AC-LOISIR-SPORT	Maître de conférences Contractuel
TORTEREAU	Antonin	DEPT-AC-LOISIR-SPORT	Maître de conférences
VIGUIER	Eric	DEPT-AC-LOISIR-SPORT	Professeur
VIRIEUX-WATRELOT	Dorothee	DEPT-AC-LOISIR-SPORT	Maître de conférences Contractuel
ZENNER	Lionel	DEPT-ELEVAGE-SPV	Professeur

REMERCIEMENTS

À Monsieur le Professeur Damien SANLAVILLE,

De la faculté de médecine de Lyon,

Pour m'avoir fait l'honneur d'accepter de présider mon jury de thèse,
Mes hommages respectueux.

À Madame la Professeure Maria-Halima LAABERKI,

De VetAgro Sup, Campus Vétérinaire de Lyon,

Pour m'avoir fait l'honneur d'accepter d'encadrer ce travail,
Pour votre disponibilité, votre gentillesse et vos précieux conseils,
Soyez assurée de ma reconnaissance.

À Madame la Professeure Véronique LAMBERT,

De VetAgro Sup, Campus Vétérinaire de Lyon,

Pour m'avoir fait l'honneur d'accepter la participation à ce jury,
Pour le soutien que vous m'avez apporté tout au long de mes années d'étude,
Recevez le témoignage de ma gratitude.

À Monsieur le Docteur Frédéric BRUNET,

De l'Institut de Génomique Fonctionnelle de Lyon,

Sans qui ce travail n'aurait jamais vu le jour,
Pour votre participation à la réalisation de ce projet, vos explications et les réponses
apportées à chacune de mes questions,
Pour m'avoir accueillie en stage,
Sincère reconnaissance.

TABLE DES MATIERES

TABLE DES FIGURES.....	13
TABLE DES TABLEAUX.....	15
LISTE DES ABREVIATIONS ET DES SIGLES	17
Introduction.....	21
PARTIE 1. ETUDE BIBLIOGRAPHIQUE DES ET, ROLE DANS L'EVOLUTON	23
I. PRESENTATION DES ET	23
A. Place des ET parmi les différentes sortes d'ADN.....	23
B. Découverte des ET par Barbara McClintock	24
C. Annotation des ET et séquençage	26
II. TAXONOMIE DES ET.....	27
A. Première proposition : Finnegan en 1989	27
B. Propositions de classifications universelles par Jurka (2005) et Wicker (2007).....	28
C. Proposition d'une classification tenant compte de la phylogénie : Piégu en 2015	30
III. DESCRIPTION DES DIFFERENTS ET, MECANISMES DE TRANSPOSITION ET TRANSMISSION	33
A. ET de classe I = rétrotransposons	33
1) Ordre LTR.....	33
2) Ordre DIRS.....	35
3) Ordre PLE.....	36
4) Ordre LINE	37
5) Ordre SINE	38
B. Mécanisme de transposition des rétrotransposons.....	39
1) Mécanisme de transposition des rétrotransposons à LTR.....	39
2) Mécanisme de transposition des rétrotransposons appartenant à l'ordre DIRS	41
3) Mécanisme de transposition des rétrotransposons sans LTR	43
4) Tableau comparatif des transposons à LTR et sans LTR	45
C. ET de classe II = transposons à ADN	45
1) Les transposons à ADN TIR.....	46
2) Les Cryptons	48
3) Les Helitrons ou « rolling-circle transposons ».....	49
4) Les Polintons	50
5) Les MITE	51
D. Mécanisme de transposition des transposons à ADN.....	51

1) Transposases et mécanisme de transposition des transposons à ADN TIR.....	51
2) Mécanisme de transposition des Cryptons.....	52
3) Mécanisme de transposition des Helitrons	53
4) Mécanisme de transposition des Polintons	54
E. Tableau récapitulatif des différents ET	55
F. Transmission des ET par transfert horizontal	57
1) Critères utilisés pour détecter les HGT d'ET	57
2) Eléments impliqués	58
IV. Rôle des ET dans l'évolution des organismes.....	60
A. Taille et composition des génomes	60
1) Le paradoxe de la valeur C	60
2) Richesse relative en familles d'ET et activité	61
3) Distribution non aléatoire des ET au sein des génomes.....	63
B. Interactions entre ET et génome	64
1) Insertions et réarrangements chromosomiques	64
2) Transposition, pseudogènes et brassage d'exons	65
3) Exemples de maladies humaines liées aux ET	66
4) Régulation de l'activité des ET	69
5) Domestication des ET.....	71
6) ET impliqués dans le maintien de la structure des génomes	71
C. ET et évolution des virus.....	72
1) Rétrotransposons et rétrovirus.....	72
2) Polintons et virus.....	72
PARTIE II. ETUDE DES ET DU GENOME DE <i>P. GIGANTEUS</i>	73
I. Présentation de l'espèce <i>P. giganteus</i>	73
A. Description, taxonomie et distribution géographique	73
1) Description	73
2) Taxonomie.....	74
3) Distribution géographique	74
B. Aptitude au vol et immunité.....	75
C. <i>P. giganteus</i> et zoonoses virales	76
1) Zoonoses à virus Nipah et Hendra	76
a) Virus Nipah	76
b) Virus Hendra.....	77

c) Caractéristiques des Henipavirus	78
d) Synthèse comparative des virus Nipah et Hendra	78
2) Système immunitaire des chauves-souris et interaction avec les virus zoonotiques	80
II. Particularités des ET chez les chauves-souris	81
A. Transposons à ADN	81
B. ERV	82
C. Transfert horizontal	83
III. Etude d'un ET de <i>Pteropus giganteus</i>	84
A. Objectifs de l'étude.....	84
B. Matériel et méthode.....	84
1) Matériel.....	84
a) Acteurs	84
b) Etapes et enjeux.....	85
c) Séquençage du génome de <i>P. giganteus</i> (méthodes et algorithmes utilisés).....	85
2) Méthode.....	88
a) Séquences ADN disponibles pour l'étude	88
b) Logiciels et banques de données utilisés	88
c) Genèse des résultats	90
C. Résultats.....	93
1) Reconstruction d'ETBat.....	93
a) 1 ^{ère} partie : reconstruction d'une séquence présentant des homologies avec un Tigger1 chez <i>P. giganteus</i>	93
b) Partie 2 : séquences caractéristiques et ORF d'ETBat	94
c) Partie 3 : étude de la phylogénie d'ETBat par BLASTN dans NCBI	95
D. Discussion	96
1) ETBat de <i>P. giganteus</i> et Tigger1 d' <i>H. sapiens</i>	96
2) Etude de la phylogénie d'ETBat : transfert horizontal	97
3) Etude de la phylogénie d'ETBat : pertinence des bases de données	98
4) Etude de la phylogénie d'ETBat : différents épisodes de domestication	99
Conclusion.....	101
Bibliographie	103

TABLE DES FIGURES

Figure 1 : Profils de renaturation du génome de la bactérie <i>Escherichia coli</i> (<i>E. coli</i>) et du génome de la souris (d'après Cech et Hearst, 1976).	23
Figure 2 : Expérience de Barbara McClintock sur les grains de maïs (d'après Nanjundiah, 1996 et Griffiths et al., 2002).	25
Figure 3 : « Corn and and microscope », photographie des épis de maïs et du microscope de Barbara McClintock, mis à l'exposition au National Museum of Natural History (photo appartenant au domaine public).....	26
Figure 4 : Mode de transposition des ET de classes I et II (Bouchet, 2009).....	28
Figure 5 : Comparaison et contenu des deux propositions pour la classification et l'annotation des ET eucaryotes (Piégu et al., 2015).....	29
Figure 6 : Classification des ET selon les différentes enzymes et voies de transposition utilisées (Curcio et Derbyshire, 2003 ; Piégu et al., 2015).	30
Figure 7 : Classification des ET proposée par Piégu et ses collaborateurs (Piégu et al., 2015 ; Arensburger et al., 2016).	32
Figure 8 : Classification synthétique des ET (d'après Wicker et al., 2007).	32
Figure 9 : Représentation schématique des rétrotransposons à LTR.	33
Figure 10 : Représentation schématique de la structure d'un rétrovirus (Rodrigues et al., 2011).....	34
Figure 11 : Représentation schématique des rétrotransposons DIRS.	36
Figure 12 : Représentation schématique des rétrotransposons Penelope.	36
Figure 13 : Représentation schématique des rétrotransposons LINE.	37
Figure 14 : Représentation schématique des rétrotransposons SINE.	39
Figure 15 : Mécanisme de transposition des rétrotransposons à LTR (d'après Eickbush et al., 2002 et Finnegan, 2012).....	40
Figure 16 : Mécanisme de transposition des éléments apparentés à DIRS1 (d'après Poulter et Goodwin, 2005).....	41
Figure 17 : Mécanisme de transposition des éléments SDR (d'après Poulter et Goodwin, 2005).....	42
Figure 18 : Mécanisme de transposition des rétrotransposons sans LTR (d'après Eickbush et al., 2002 et Finnegan, 2012).....	44
Figure 19 : Représentation schématique des transposons TIR.....	46
Figure 20 : Diagramme en ruban du domaine catalytique DDE de l'intégrase du VIH-1 (Hickman et al., 2010).	48
Figure 21 : Représentation schématique des transposons Crypton.	49
Figure 22 : Représentation schématique des transposons Helitron.....	50

Figure 23 : Représentation schématique des transposons Polinton.....	50
Figure 24 : Mécanisme de transposition des transposons TIR.	52
Figure 25 : Mécanisme de transposition des Cryptons (d'après Goodwin et al., 2003).....	53
Figure 26 : Mécanisme de transposition d'un Helitron (Grabundzija et al., 2016).	54
Figure 27 : Mécanisme de transposition des Polintons (d'après Kapitonov et Jurka, 2006)...	55
Figure 28 : Nombre de transferts horizontaux d'ET de vertébrés par superfamille et par classe d'ET (Gilbert et Feschotte, 2018).....	59
Figure 29 : Relation entre la taille du génome et la proportion de séquences répétées au sein du génome (a) et relation entre la taille du génome et la proportion d'ET (b) chez 502 espèces (148 espèces animales, 81 plantes, 202 champignons et 70 protistes) (Elliott TA et Gregory TR, 2015).....	61
Figure 30 : Part relative en ET de différentes familles des génomes de différentes espèces (Chalopin et al., 2015).	62
Figure 31 : Structure schématique d'un élément Alu (d'après Kim et al., 2016).....	66
Figure 32 : Impact de l'insertion d'un élément Alu dans une région génique sur la transcription (Kim et al., 2016).....	67
Figure 33 : Mécanisme d'interférence par ARN (Slotkin et Martienssen, 2007).	69
Figure 34 : Mécanisme de modifications de la chromatine basé sur l'interférence par ARN (Slotkin et Martienssen, 2007).	70
Figure 35 : Photo de <i>P. giganteus</i> (photographe Tonge Simon J., le 18 octobre 2017 au Parc zoologique Tierpark Berlin, Allemagne).	73
Figure 36 : Distribution géographique de <i>P. giganteus</i> (d'après Molur et al., 2008).	75
Figure 37 : Invasion des génomes des différentes chauves-souris par les transposons à ADN (Platt et al., 2016).....	82
Figure 38 : Vue d'ensemble du pipeline dnaPipeTE (d'après Goubert et al., 2015).	87
Figure 39 : Composition du génome de <i>P. giganteus</i> en ET (d'après Julien Fouret, Viroscan3D).	88
Figure 40 : Identification d'un ET de <i>P. giganteus</i> présentant un intérêt particulier	90
Figure 41 : Représentation schématique de l'alignement des séquences de <i>P. giganteus</i> et de l'ETX grâce à SeaView.....	91
Figure 42 : Identification des ORF d'ETBat.....	92
Figure 43 : Représentation schématique du TIGGER1 <i>Homo sapiens</i> et de son alignement avec la partie commune des deux séquences de <i>P. giganteus</i> qui lui sont quasiment identiques.....	94
Figure 44 : ETBat de <i>P. giganteus</i> , un élément homologue à Tigger1 d' <i>H. sapiens</i> , et ses séquences caractéristiques.	95
Figure 45 : Arbre phylogénétique obtenu par BLASTN de la séquence nucléotidique d'ETBat	96

TABLE DES TABLEAUX

Tableau I: Les types d'ADN eucaryote (d'après Singh-Cundy et Shin, 2017).	24
Tableau II: Tableau comparatif de la transposition des transposons à LTR et sans LTR.....	45
Tableau III : Tableau récapitulatif de la structure des différents ET.	56
Tableau IV: Tableau comparatif concernant les virus Nipah et Hendra (d'après les sites internet de l'Organisation Mondiale de la Santé (OMS), de l'Organisme mondiale de la santé animale (OIE) et de l'Australian Veterinary Association (AVA)).....	78

LISTE DES ABREVIATIONS ET DES SIGLES

A	Adénine
aa	Acides aminés
A3	Famille des protéines APOBEC3
ADN	Acide DésoxyriboNucléique
AP	Protéase aspartique
APE	Endonucléase apurinique
ARN	Acide RiboNucléique
ARNm	Acide RiboNucléique messenger
ARN Pol III	ARN polymérase III
ARNr	Acide RiboNucléique ribosomique
ARNt	Acide RiboNucléique de transfert
ATP	Adénosine TriPhosphate
AVA	Australian Veterinary Association
BLAST	Basic Local Aligement Search Tool
C	Cytosine
dsRNA	Acide RiboNucléique double brin
ELISA	Enzyme-Linked ImmunoSorbent Assay
En	Endonucléase
Env	Protéine d'enveloppe
ERV	Rétrovirus endogène
ET	Elément Transposable
E-value	Expect value
G	Guanine
Gag	Group-specific AntiGen
HeV	Virus Hendra
HGT	Horizontal Gene Transfer

ITR	Inverted Terminal Repeat
ICR	Internal Complementary Region
IHC	ImmunoHistoChimie
INT	INTégrase
kb	Kilobase
LINE	Long INterspersed Element
LTR	Long Terminal Repeat
MITE	Miniature Inverted-repeat Transposable Element
MT	Méthyltransférase
NCBI	National Center for Biotechnology Information
NGS	Next Generation Sequencing
NiV	Virus Nipah
OH	Groupe hydroxyle
OIE	Organisation mondiale de la santé animale
OMS	Organisation Mondiale de la Santé
ORF	Open Reading Frame
pb	Paire de bases
PCR	Polymerase Chain Reaction
Ph. D.	Philosophiæ Doctor
POLB	ADN polymerase B
PRO	PROtéase
RC	Cercle Roulant
RdRP	ARN polymérase dépendant de l'ARN
RepHel	Protéine possédant les domaines Rep et Hel
RH	Ribonucléase H
RNAi	Interférence par ARN
RSS	Recombination Signal Sequence
RT	Transcriptase inverse

SINE	Short INterspersed Element
siRNA	Petits ARN interférents
SDR	Split Direct Repeat
T	Thymine
Tase	Transposase
TERT	TElomerase Reverse Transcriptase
TIR	Terminal Inverted Repeat
Topo II α	Topoisomérase II α
TPRT	Target-Primed RetroTransposition
TR	Sub-terminal repeats
TSD	Target Site Duplication
YR	Tyrosine Recombinase

Introduction

Les éléments transposables (ET) sont des séquences d'ADN capables de se déplacer au sein du génome d'une cellule. Elles sont parsemées aussi bien au sein des génomes procaryotes qu'eucaryotes. Les ET ont été découverts par Barbara McClintock dans les années 1950. La part du génome composée d'ET est très variable d'un organisme à l'autre. Ces éléments représentent environ 45% du génome humain et peuvent représenter plus de 90% du génome chez certaines plantes.

Ces éléments, capables de proliférer et de se déplacer au sein des génomes, ont tout d'abord été considérés comme des parasites, égoïstes et indésirables. En effet, ils sont responsables de mutations et de réarrangements chromosomiques pouvant induire un large panel de maladies génétiques.

Cependant, les mutations induites par les ET participent à la diversité génétique d'un organisme et peuvent s'avérer bénéfiques dans certaines circonstances. Ainsi, certains ET ont été domestiqués et agissent maintenant comme des gènes ou comme des séquences régulatrices essentiels pour leur hôte.

Ces séquences répétées de nombreuses fois dans les génomes posent problème lors de l'assemblage de ceux-ci. En effet, les courtes séquences obtenues par séquençage forment les pièces d'un puzzle qu'il faut par la suite assembler. Les multiples copies d'un ET, dispersées au sein du génome, constituent des pièces qui peuvent parfois s'ajuster parfaitement à plusieurs sites distincts, compliquant l'assemblage. Pour cette raison ainsi que pour leur intérêt évolutif, l'identification et l'annotation des ET figurant au sein d'un génome présente un intérêt majeur.

En raison de son rôle de réservoir d'agents pathogènes zoonotiques émergents, le génome de la chauve-souris frugivore *Pteropus giganteus* a été séquencé. Chez cette dernière, l'infection virale par les virus Nipah et Hendra est asymptomatique. Afin d'identifier les éventuels éléments de résistance à l'infection virale et d'en déduire des stratégies de lutte chez l'humain, le génome et le transcriptome de *P. giganteus* sont en cours d'analyse.

L'étude des ET de *P. giganteus*, présents au sein de la version préliminaire de son génome, permettra de faciliter son assemblage, d'appréhender l'histoire évolutive de ces éléments, voire de découvrir d'éventuelles particularités responsables de sa résistance à l'infection par les virus zoonotiques.

Ce rapport est divisé en deux grandes parties. Une première partie, issue d'une étude bibliographique, présente de façon détaillée les ET, leur découverte, leur place au sein des génomes, leurs mécanismes de transposition ainsi que leur rôle dans l'évolution des êtres vivants.

Une seconde partie débute par la présentation de l'espèce *P. giganteus*, des particularités des ET ainsi que du système immunitaire des chauves-souris. Puis se focalisera sur notre analyse bioinformatique d'un ET spécifique du génome de *P. giganteus*, de sa reconstruction aux hypothèses évolutives qu'implique sa phylogénie.

PARTIE 1. ETUDE BIBLIOGRAPHIQUE DES ET, ROLE DANS L'EVOLUTON

I. PRESENTATION DES ET

A. Place des ET parmi les différentes sortes d'ADN

Historiquement, le génome des organismes pluricellulaires a été divisé en trois catégories de séquences. Ces séquences se distinguent par leur cinétique de renaturation en solution après dénaturation de l'ADN à haute température. Les séquences qui se renaturent le plus rapidement sont répétées des centaines de milliers de fois alors que celles qui se renaturent le plus lentement sont uniques (fig 1). En effet, la réassociation d'une paire de séquences complémentaires résulte de leur collision ; la vitesse de réassociation augmente ainsi avec la concentration car la probabilité de rencontrer une séquence semblable est alors augmentée (Britten et Kohne, 1968). Par exemple, le génome humain est constitué de 10 à 15% de séquences hautement répétées, de 20 à 40% de séquences moyennement répétées et d'environ 50% de séquences uniques. **L'ADN non répété** (ou ADN unique) correspond principalement aux gènes codant pour les protéines. **L'ADN moyennement répété** est majoritairement non codant et comprend les gènes codant pour les ARN de transfert (ARNt), les ARN ribosomiques (ARNr), mais également les ET (transposons, séquences SINE, LINE et rétrovirus endogènes (ERV)). Enfin, **l'ADN hautement répété** est non codant et comprend des séquences répétées disposées en tandem, les microsatellites, les centromères et les télomères (Krahn, 2009).

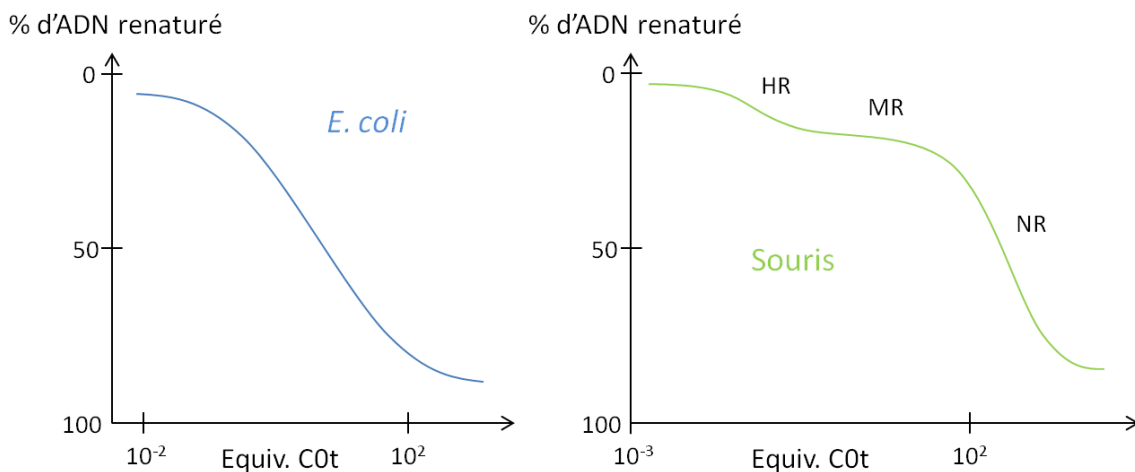


Figure 1 : Profils de renaturation du génome de la bactérie *Escherichia coli* (*E. coli*) et du génome de la souris (d'après Cech et Hearst, 1976).

L'axe Equiv. C0t représente le produit du temps (en secondes) et de la concentration en ADN (en moles de nucléotides par litres) corrigé par les conditions de renaturation. Le génome d'*E. coli*, organisme procaryote, se renature suivant une courbe sigmoïde simple, traduisant la présence d'un seul type d'ADN. Le génome de la souris, organisme eucaryote, se renature selon une courbe constituée de différents paliers, traduisant la présence de types d'ADN différents. Le premier palier (HR) correspond à l'ADN hautement répété, le second (MR) à l'ADN moyennement répété et le troisième (NR) à l'ADN non répété.

Les ET sont « des segments discrets d'ADN **capables de se déplacer dans un génome hôte d'un chromosome ou d'un site plasmidique à un autre** et qui n'encodent pas une machinerie moléculaire spécifique leur permettant d'infecter le génome de nouveaux hôtes par transfert latéral » (Piégu et al., 2015). Les ET ont été **découverts dans tous les génomes procaryotes et eucaryotes** au sein desquels ils ont été recherchés (Lerat, 2001). Ils sont retrouvés dans les régions intergéniques (tab I).

Tableau I: Les types d'ADN eucaryote (d'après Singh-Cundy et Shin, 2017).

Gènes et séquences liées aux gènes				Régions intergéniques			
Exons	Introns	ADN intercalaire	ADN régulateur	ET	Satellites	ADN structural (ex : centromères)	Autre

De nos jours, des génomes entiers ont été analysés et de nombreuses données sont disponibles concernant des séquences caractérisées précisément. En ce qui concerne les ET des génomes eucaryotes, ils sont référencés dans la **base de données Repbase** qui est bien documentée, régulièrement mise à jour (par ajout de séquences consensus, affinage ou extension de séquences déjà présentes, reclassification...) et utilisée pour la quasi-totalité des analyses de séquences de génomes eucaryotes (Bao et al., 2015).

B. Découverte des ET par Barbara McClintock

[Nanjundiah, 1996 ; Kass et Chomet, 2009 ; Ravindran, 2012]

Les ET ont été découverts par **Barbara McClintock** au cours de ses études concernant le maïs.

En 1927, Barbara McClintock est diplômée d'un Ph.D. de botanique à l'université de Cornell pour ses recherches cytogénétiques sur le maïs. Dès lors, l'étude de cet organisme devient son sujet de prédilection. En 1929, Barbara McClintock est capable de distinguer les dix chromosomes du maïs, permettant ainsi de relier les données génétiques au comportement des chromosomes. Elle aide ainsi à mettre en évidence les gènes liés, gènes qui sont hérités conjointement à cause de leur proximité sur un même chromosome, et elle met également en évidence les crossing-overs à l'échelle chromosomique en démontrant que l'échange de segments de chromosomes est à l'origine de recombinaisons génétiques. En 1944, Barbara McClintock est élue à l'**Académie des Sciences** et en 1945 elle est la première femme à être élue présidente de la Genetics Society of America.

Après avoir mis en évidence que les gènes étaient associés aux chromosomes et, par conséquence stables au sein du génome, Barbara McClintock développe l'idée que les gènes pourraient être en réalité instables. En effet, elle a découvert un locus sur le chromosome 9 du maïs au niveau duquel une cassure peut se produire (**locus** nommé **Ds** pour « dissociation ») et qui peut changer de position au sein des chromosomes (fig 2). Elle met également en évidence un second locus nécessaire à la cassure du locus Ds (qu'elle nomme

Ac pour « Activator ») qui est également capable d'initier sa propre **transposition**. Elle décrit ses découvertes dans le Yearbook 1947-1948 de la Carnegie Institution de Washington, une fondation de recherche scientifique. L'instabilité génomique causée par les mutations est déjà connue mais Barbara McClintock introduit l'idée nouvelle que certains « gènes sauteurs » peuvent se déplacer au sein du génome à **une fréquence des centaines de fois supérieure aux mutations**.

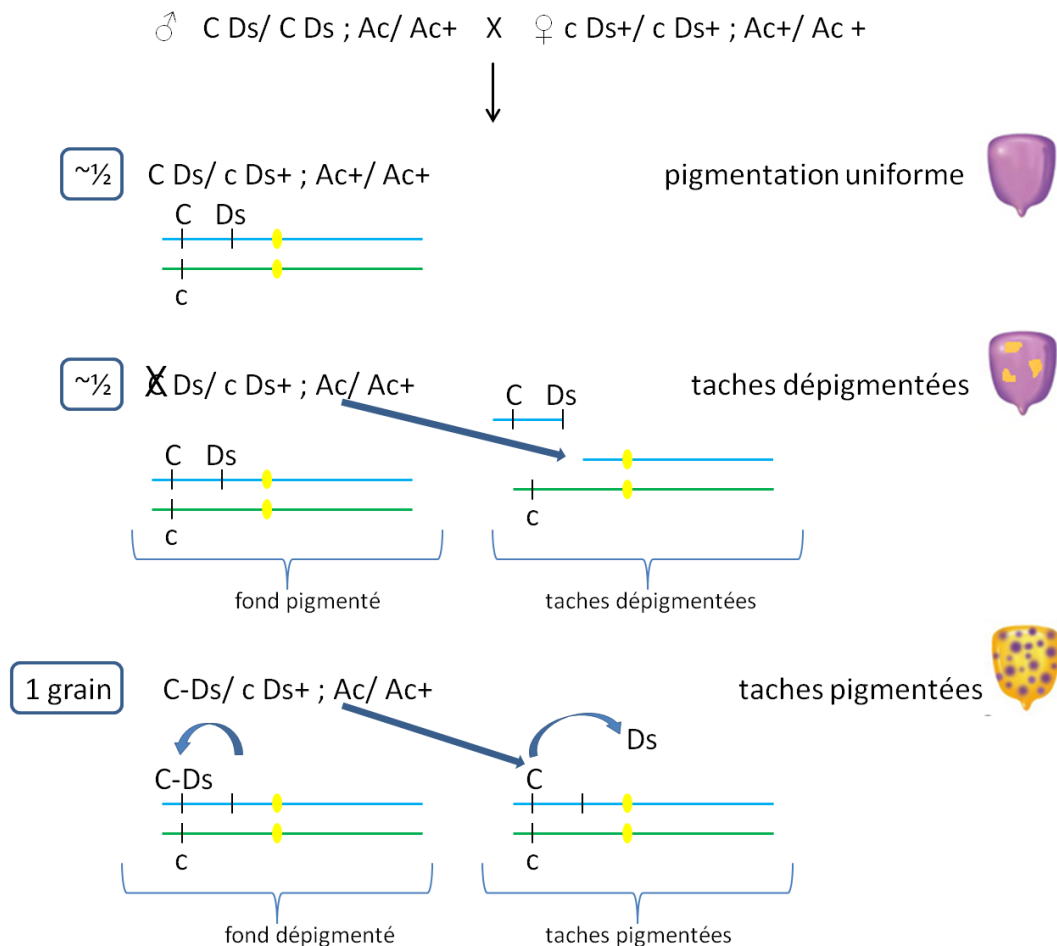


Figure 2 : Expérience de Barbara McClintock sur les grains de maïs (d'après Nanjundiah, 1996 et Griffiths et al., 2002).

L'élément Ds (*Dissociator*) est responsable d'une fréquence élevée de cassures chromosomiques à l'endroit où il est inséré. L'activité de Ds dépend de la présence de l'élément non lié Ac (*Activator*). Les + indiquent l'absence de ces éléments. Le gène C détermine la couleur du grain de maïs : « C » est l'allèle dominant et donne un grain de maïs coloré, « c » est l'allèle récessif et donne un grain de maïs incolore. Lorsqu'un croisement est réalisé entre un mâle homozygote C, homozygote Ds et possédant Ac et une femelle homozygote c, et ne possédant ni Ds, ni Ac, les proportions attendues sont 100% de grains de maïs pigmentés à la première génération car C sera hérité du mâle et dominant. Cependant seulement 50% environ des descendants du croisement sont pigmentés (en l'absence de Ac, Ds est stable et aucune cassure ne se produit au niveau de Ds). Dans environ 50% des cas, lorsqu'Ac est présent, Ds est actif et entraîne la perte de « C » lors de la cassure chromosomique. Le fond du grain est ainsi coloré car C est présent au début du développement, puis en fonction du moment du développement où a lieu la cassure, les taches dépigmentées sont de différentes tailles. Exceptionnellement, Ds s'insère dans C et donne un fond dépigmenté. Celui-ci est instable car lorsque Ds est mobilisé par Ac, C retrouve sa fonction et entraîne ainsi l'apparition de taches pigmentées sur un grain de fond clair.

Ses observations sont tout d'abord incomprises par la communauté scientifique. En effet, les données de cartographie génétique avaient jusque-là montré que les gènes étaient arrangés de manière linéaire et fixe sur les chromosomes. Il était alors difficile pour les scientifiques d'accepter l'idée que certains gènes pouvaient être mobiles au sein du génome.

Au milieu des années 1960, le code génétique est décrypté. Grâce à la révolution de la biologie moléculaire, les gènes ne sont plus un concept abstrait mais des entités moléculaires. Dans les années 1980, les transposons Ds et Ac sont isolés. Les ET sont également découverts au sein des virus phages, des bactéries et des drosophiles. La communauté scientifique reconnaît ainsi petit à petit que les transposons ne sont pas uniquement présents chez le maïs mais probablement répandus au sein des espèces.

C'est seulement en **1983** que le travail de Barbara McClintock est récompensé par le **Prix Nobel de Physiologie ou Médecine** pour sa découverte des ET chez le maïs (fig 3), 35 ans après la publication de son premier rapport sur la transposition.



Figure 3 : « Corn and microscope », photographie des épis de maïs et du microscope de Barbara McClintock, mis à l'exposition au National Museum of Natural History (photo appartenant au domaine public).

Epis de maïs observés par Barbara McClintock démontrant divers phénotypes liés à l'activité des ET.

C. Annotation des ET et séquençage

Lors du séquençage d'un génome, les séquences obtenues (« reads ») sont de courts fragments d'ADN qu'il faut par la suite assembler afin de reconstruire les différents chromosomes. La richesse des génomes eucaryotes en ET pose problème lors de l'assemblage du génome. En effet, tant que les reads sont uniques, ils sont assemblés de

proche en proche. Par contre, lorsque les fragments sont répétés de nombreuses fois au sein du génome de manière identique, le risque est alors d'assembler des fragments non adjacents au sein du génome (Tang, 2006). De plus, la présence de régions codantes chez certains ET peut être source de confusion lors de l'annotation des gènes (Lerat, 2010). Ainsi, la caractérisation des ET d'un génome permet de se focaliser sur les séquences situées en 5' et en 3' afin de repérer les régions dans lesquelles ils sont insérés, facilitant l'assemblage.

Les ET sont impliqués dans l'évolution des génomes à deux échelles de temps. A très longue échelle, ils ont façonné la structure actuelle des génomes par leur accumulation quantitative et par les modifications chromosomiques (réarrangements, délétions, duplications) qu'ils ont entraînés par des mécanismes de recombinaisons ectopiques. A courte échelle, ils sont responsables de mutations pouvant conduire à l'expression de maladies génétiques (Kazazian, 2004).

Ainsi, à la fois pour des raisons techniques et parce qu'ils jouent un rôle majeur dans l'évolution des génomes, il est aujourd'hui particulièrement intéressant d'annoter et de classer les ET d'un génome avant d'effectuer son assemblage (Lerat, 2010).

Lors de leur annotation, les ET sont rattachés à différentes catégories, comme le sont par exemple les différentes espèces ou les différents virus. La classification des ET fera l'objet de la partie suivante.

II. TAXONOMIE DES ET

La classification des ET n'est pas consensuelle au sein de la communauté scientifique. Un appel à la création d'une organisation nommée « International Committee for the Taxonomy of Transposable Element (ICTTE) » a été lancé en 2015 afin de débattre sur ce sujet et de promouvoir l'adoption de standards concernant l'annotation des ET (Arensburger et al., 2016).

A. Première proposition : Finnegan en 1989

La première classification proposée divise les ET en deux classes selon leur mécanisme de transposition (fig 4). La classe I regroupe les éléments transposant par transcription inverse à partir d'un intermédiaire ARN : mécanisme ADN-ARN-ADN. La classe II regroupe les éléments transposant directement de l'ADN à l'ADN (Finnegan, 1989).

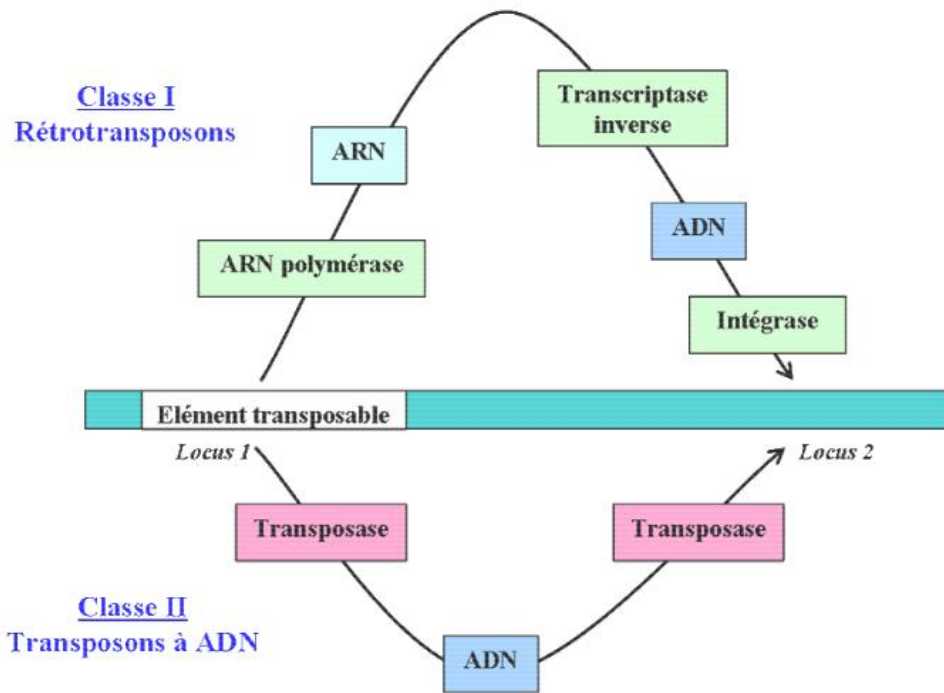


Figure 4 : Mode de transposition des ET de classes I et II (Bouchet, 2009).

Par la suite, les éléments de classe I ont été divisés en deux types selon leurs caractéristiques structurales. Le premier type, communément appelé rétrotransposons à LTR, est caractérisé par de longues répétitions terminales (LTR), trois cadres de lecture ouverts (ORF) codant pour la protéine Gag (« group specific antigen »), une transcriptase inverse (RT) et une protéine d'enveloppe rétrovirale (Env). Le second type, communément appelé rétrotransposons non-LTR, regroupe des ET avec de courtes répétitions terminales, deux ORF codant pour des protéines similaires à Gag et RT et possédant une queue poly-A à leurs extrémités 3' (Piégu et al., 2015). Les éléments appartenant à la classe II ont été caractérisés structurellement comme contenant des répétitions inversées terminales (TIR) et possédant un gène qui code pour une transposase (enzyme permettant la transposition de l'ET) (Finnegan, 1992).

B. Propositions de classifications universelles par Jurka (2005) et Wicker (2007)

Souhaitant proposer un système de classification universel basé sur une série de critères de classification plus précis que ceux utilisés précédemment, deux mises à jour du système précédemment explicité ont été proposées (Jurka et al., 2005 ; Wicker et al., 2007). Ces deux propositions sont illustrées dans la figure 5. La proposition de Jurka et ses collaborateurs est utilisée pour classer les ET dans Repbase (Jurka et al., 2005, Kapitonov et Jurka, 2008). Ces deux propositions sont basées sur des caractéristiques structurales, et conservent la dichotomie entre rétrotransposons et transposons à ADN proposée par Finnegan (« type » selon Repbase, « classe » selon Wicker). La division suivante se fait en « classe » selon Repbase et en « ordre » selon Wicker, mais ces subdivisions sont très similaires et contiennent les mêmes superfamilles.

Il faut cependant noter que cette classification **ne tient pas compte de l'histoire évolutive des ET**. Elle peut ainsi être à l'origine de malentendus lorsqu'elle est utilisée en présupposant qu'elle reflète la phylogénie et qu'elle regroupe les ET possédant un ancêtre commun, d'où l'importance d'établir une classification tenant compte des principes évolutifs (Piégu et al., 2015).

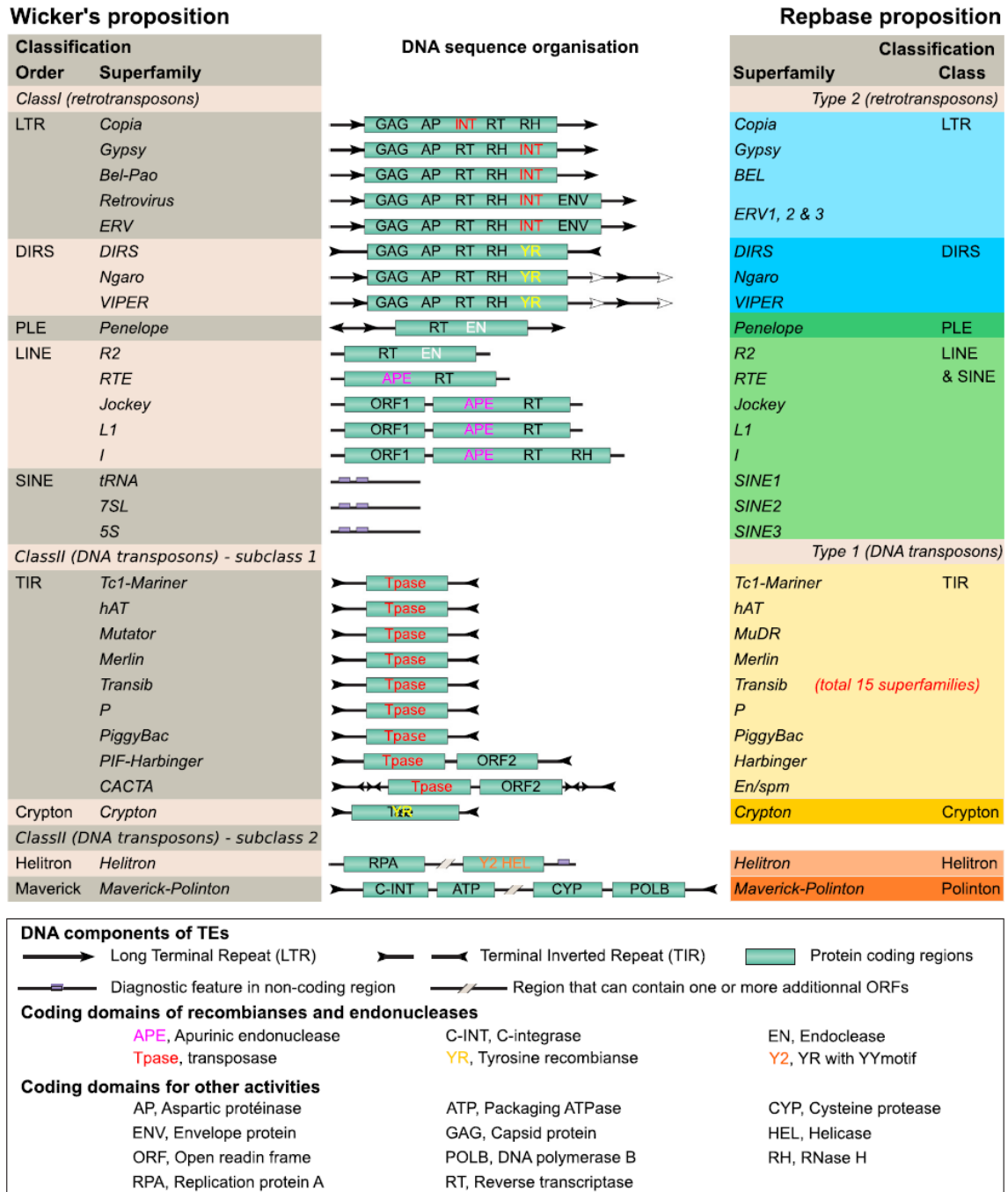


Figure 5 : Comparaison et contenu des deux propositions pour la classification et l'annotation des ET eucaryotes (Piégu et al., 2015).

La proposition utilisée pour classer les ET dans Rebase est présentée à droite (Jurka et al., 2005, Kapitonov et Jurka, 2008) et la proposition de Wicker et ses collaborateurs à gauche (Wicker et al., 2007). La représentation schématique de l'organisation de la séquence d'ADN de chaque superfamille est présentée au centre de la figure. La légende des symboles utilisés est fournie au bas de la figure.

C. Proposition d'une classification tenant compte de la phylogénie : Piégu en 2015

Afin de tenir compte de l'aspect évolutif, les ET pourraient être catégorisés selon la machinerie enzymatique utilisée pour se déplacer d'un locus à un autre. Cinq catégories seraient ainsi définies selon la famille de transposase / intégrase utilisée : DDE-transposases, Y2-transposases ou cercle roulant (RC pour rolling-circle en anglais), tyrosine (Y1)-transposases, sérine (S)-transposases et la combinaison d'activités transcriptase inverse et endonucléase (RT/En) (fig 6). Ces cinq classes seraient par la suite divisées selon les caractéristiques particulières des mécanismes de transposition. Cette classification met en avant l'impact des convergences évolutives dans le contexte de la transposition (phénotypes de transposition similaires mais ayant des origines différentes), et le fait que, comme les virus, les ET ne proviennent pas d'un même ancêtre commun mais de nombreuses classes d'origines différentes (Curcio et Derbyshire, 2003 ; Piégu et al., 2015).

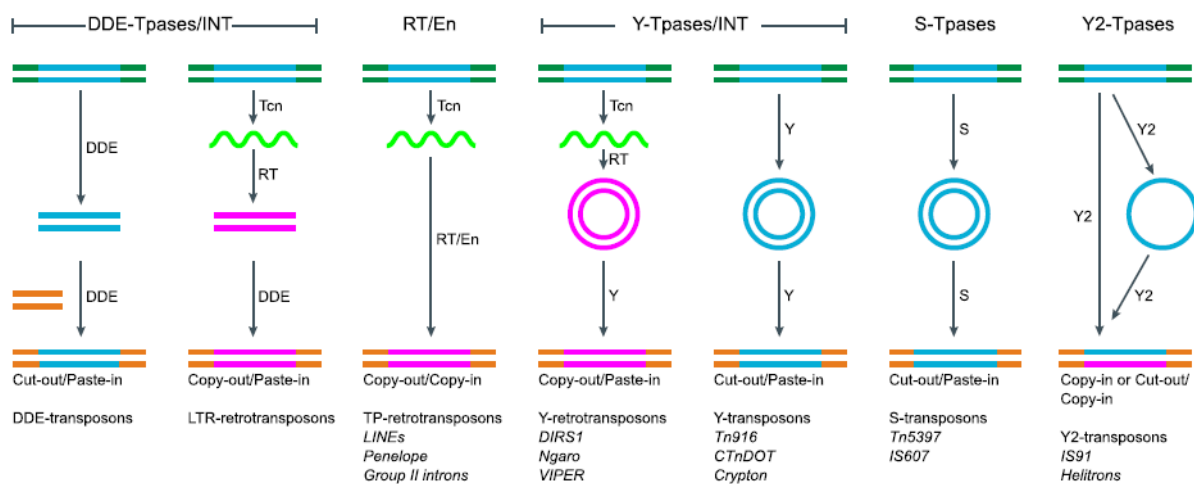


Figure 6 : Classification des ET selon les différentes enzymes et voies de transposition utilisées (Curcio et Derbyshire, 2003 ; Piégu et al., 2015).

En 2015, ayant été constaté que les classifications précédentes sont insuffisantes, les bases d'une nouvelle classification regroupant les ET procaryotes et eucaryotes sont proposées, tenant compte de l'aspect évolutif ainsi que des ET découverts récemment (fig 7). Cette proposition a pour objectif de motiver le rassemblement des différentes communautés scientifiques étudiant les ET afin d'établir une classification universelle et de mettre en commun les ressources et outils utilisés à travers la mise en place d'un comité international (ICTTE) (Piégu et al., 2015 ; Arensburger et al., 2016).

TE Classes with some members having a DNA transposon phenotype

Class	Order	Superfamilies
Nuclease/Recombinase DDE-transposons	<i>Transposition mechanism</i>	<i>Phylogenetic relationships between Nuclease/Recombinase</i>
	<i>DDE transposons with no DNA-transposition intermediate (Copy-in)</i>	<i>Mu</i> <i>Tn3</i>
	<i>DDE/D transposons with a linear dsDNA transposition intermediate (Cut-out/Paste in)</i>	<i>IS1, IS3, IS4, IS701, ISH3, IS1634, IS1182, IS6, IS21, IS30, IS66, IS110, IS630, IS982, IS1380, ISAs1, ISL3</i>
	<i>DDE/D transposons with a linear dsDNA transposition intermediate and using a heteromeric transposase (Cut-out/Paste in)</i>	<i>IS630/Tc1/mariner (ITm)/Zator</i> <i>IS1595-Merlin,</i> <i>IS5/PIF/Harbinger,</i> <i>IS256/MuDR/Mutator/Rehavirus</i> <i>IS1380/PiggyBac,</i> <i>Academ, CACTA/Mirage/Chapaev (CMC), Dada, Hobo/Ac/Tam (hAT), Kolobok, P(?),Sola, Transib/ProtoRag</i> <i>Tn7</i>
<i>DDE transposons with a circular dsDNA transposition intermediate (Copy-out /Paste-in)</i>	<i>IS3</i>	
Y1-transposons	<i>LTR retrotransposons (Copy-out/Paste-in)</i>	<i>Copia</i> <i>Gypsy</i> <i>BEL</i> <i>ERV1</i> <i>ERV2</i> <i>ERV3</i>
	<i>Y1 transposons with a circular dsDNA transposition (Cut-out/Paste in)</i>	<i>IS200/IS605</i> <i>Tn916</i> <i>CTnDOT</i> <i>Crypton</i> <i>DIRS</i> <i>Ngaro</i> <i>VIPER</i>
Y2-transposons	<i>Y2 transposons with a circular ssDNA transposition (Copy-in or -out/ Copy-in)</i>	<i>IS91</i> <i>Helitrons</i>
S-transposons	<i>S transposons with a circular dsDNA transposition (Cut-out/paste-in)</i>	<i>IS607</i> <i>Tn5397</i>
Casposons ^{26,27}	<i>Casposase with a DNA intermediate in a configuration that remains to define (Copy-in or -out/ Copy-in or paste-in)</i>	<i>Casposons</i>
TEs pending classification	<i>?</i>	<i>ISAs1</i>
	<i>?</i>	<i>Fanzor</i>
	<i>Polintons/Mavericks</i>	<i>Mavirus (?)</i>
	<i>DDE integrase (Copy-in or -out/ Copy-in)</i>	<i>Polintons/Mavericks</i> <i>Tlr1</i>
	<i>Transposase putatively related to integrases of LTR retrotransposons</i>	<i>Ginger1</i> <i>Ginger2</i> <i>P(?)</i>
	<i>DDE-transposons with a DDE-transposase having another origin</i>	
	<i>Zisupton (Unknown transposition depending on a "Zisuptase")</i>	<i>Zisupton</i>

TE Classes for TEs with a non-LTR retrotransposon phenotype

Class	Order	Superfamilies
non-LTR retrotransposons Retroposons	<i>Endonuclease (En)</i>	<i>Phylogenetic relationships between endonuclease, then RT</i>
	<i>LINES</i>	<i>LINES with an AP EN</i> <i>LINES with a PD-(D/E)XK EN</i> <i>LINES with both AP and PD-(D/E)XK EN</i>
	<i>Penelope-like elements (PLE)</i>	<i>Athena, no GIY-YIG domain</i> <i>Coprina, no GIY-YIG domain</i> <i>Neptune, GIY-YIG domain</i> <i>Penelope, GIY-YIG domain</i>
	<i>Group II introns</i>	<i>Group II introns</i> <i>Mobile lariat introns</i> <i>Introner-like elements</i>

TE Classes for TEs with an SSE phenotype		
Class	Order	Superfamilies
<i>Machinery for excision of host genes</i>	<i>Transposition mechanism</i>	<i>Phylogenetic relationships between HEN, and-or site into host genes</i>
Intein	LAGLIDADG inteins (HEN dependent HR) HNNH inteins (HEN dependent HR)	Host genes in which each intein specifically inserted could be used, as proposed in InBase (http://tools.neb.com/inbase/)
Group I intron (G1i)	LAGLIDADG G1i (HEN dependent HR) HNNH G1i (HEN dependent HR) His-Cys G1i (HEN dependent HR) GIY-YIG G1i (HEN dependent HR) PD-(D/E)XK G1i (HEN dependent HR) Vsr G1i (?) (HEN dependent HR)	Host sites in which each group I intron specifically inserted could also be used

TE Classes for rare prokaryotic TEs with a retroposon phenotype		
Class	Order	Superfamilies
RT features Retron/msRNA	Transposition mechanism Retron/msRNA (retrotransposition)	Phylogenetic relationships between RT msRNA

Figure 7 : Classification des ET proposée par Piégu et ses collaborateurs (Piégu et al., 2015 ; Arensburger et al., 2016).

Aucun consensus n'existant concernant la classification des ET, la classification proposée par Wicker et ses collaborateurs sera utilisée par la suite pour présenter les différents ET, en gardant à l'esprit les limites de cette classification (fig 8).

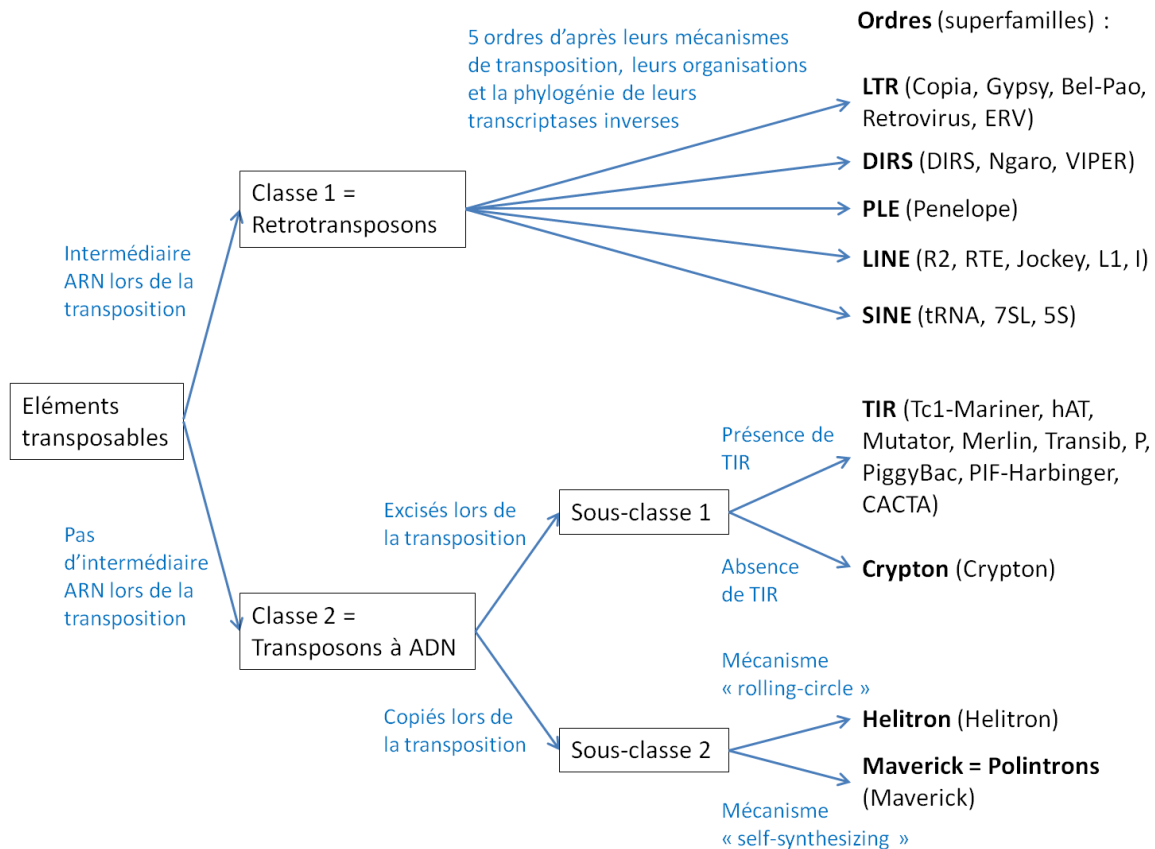


Figure 8 : Classification synthétique des ET (d'après Wicker et al., 2007).

III. DESCRIPTION DES DIFFERENTS ET, MECANISMES DE TRANSPOSITION ET TRANSMISSION

A. ET de classe I = rétrotransposons

Les éléments de classe I transposent via l'intermédiaire d'un ARN. Ainsi la copie originale reste à son locus : la transposition est de type « copier-coller ». Celle-ci est permise par l'enzyme **transcriptase inverse**. Ces ET sont également nommés « **rétrotransposons** ». (Ray et al., 2015, Sotero-Caoi et al., 2017). La classe I est généralement divisée en deux : les **rétrotransposons à LTR** (longues régions terminales répétées) et les **rétrotransposons sans LTR** (Sotero-Caoi et al., 2017). La subdivision en cinq ordres citée précédemment sera utilisée afin de décrire les différents rétrotransposons, puis leurs mécanismes de transposition seront détaillés.

1) Ordre LTR

[Cook et al., 2000 ; Bannert et Kurth, 2006 ; Wicker et al., 2007 ; Antonin Demange, 2013 ; Nefedova et Kim, 2017]

Les rétrotransposons à LTR sont appelés ainsi car ils possèdent à leurs extrémités de longues régions terminales répétées (LTR « long terminal repeat ») (fig 9). Ces LTR mesurent quelques centaines à quelques milliers de paires de bases. Ils commencent par 5'-TG-3' et se terminent par 5'-CA-3'. Les rétrotransposons à LTR contiennent habituellement deux ORF. Le premier ORF correspond au gène *gag* des rétrovirus et code pour les protéines de la capsid. Le second ORF (*pol*) code pour une **protéase aspartique** (AP : enzyme protéolytique de la famille des pepsines qui fonctionne habituellement en solution acide), une **transcriptase inverse** (RT), une **ribonucléase H** (RH : enzyme qui dégrade la partie ARN des hybrides ADN-ARN) et une **intégrase** (INT) (Hausen et Stein, 1970 ; Tang et Wong, 1987). Parfois, un troisième ORF (*env*) code pour un produit responsable de la reconnaissance par les récepteurs cellulaires et de la pénétration du virus dans la cellule. Lors de leur transposition, les rétrotransposons à LTR génèrent une **duplication du site cible** (TSD, pour target site duplication en anglais) de 4 à 6 paires de bases.

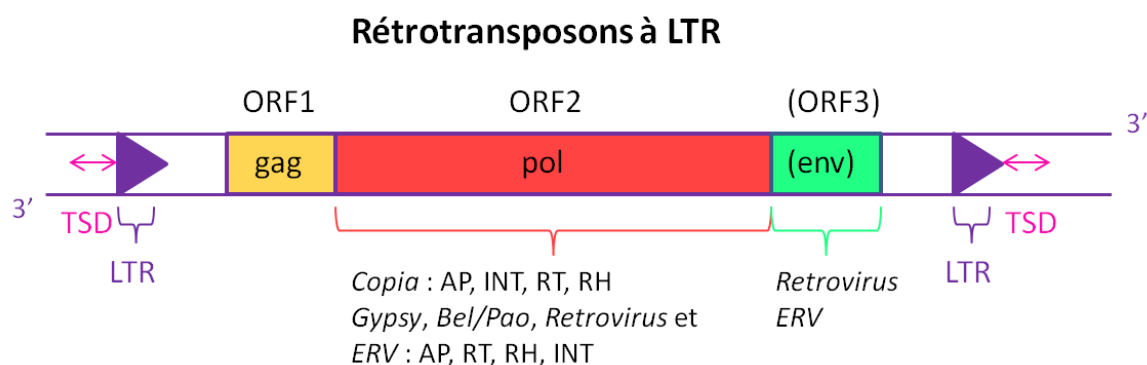


Figure 9 : Représentation schématique des rétrotransposons à LTR.

Les éléments de la superfamille **Copia** se distinguent des autres superfamilles par l'ordre des domaines portés par le gène *pol* (AP-INT-RT-RH chez Copia, AP-RT-RH-INT chez les autres superfamilles). Les éléments des superfamilles **Gypsy** et **Bel/Pao** ont la même structure mais les analyses phylogénétiques indiquent qu'ils forment des superfamilles différentes.

Les rétrotransposons LTR possédant les trois ORF ont une structure identique aux rétrovirus sous forme ADN intégrés dans les génomes de vertébrés. Au regard de l'évolution, les rétrotransposons à LTR et les rétrovirus sont étroitement liés (voir paragraphe Rétrotransposons et rétrovirus). Les rétrovirus peuvent ainsi être classés parmi les rétrotransposons et leurs caractéristiques seront développées dans ce paragraphe. Cependant il faut noter qu'ils sont habituellement classés parmi les virus, dans la famille des *Retroviridae*.

Les **rétrovirus** sont des virus possédant deux copies d'ARN monocaténaire à polarité positive (leur ARN n'est pas directement codant, il n'agit pas comme un ARNm) enfermées dans une capsid ; ils sont enveloppés sous leur forme extracellulaire (fig 10). Tous les rétrovirus possèdent 4 gènes principaux dans l'ordre 5'-gag-pro-pol-env-3' (certains de ces gènes pouvant parfois se chevaucher). *Gag* code pour les protéines de la matrice, de la nucléocapsid et de la capsid, *pro* pour la protéase, *pol* pour la transcriptase inverse et l'intégrase, et *env* pour les protéines transmembranaires et de surface de l'enveloppe. D'autres rétrovirus sont plus complexes et possèdent des gènes de protéines accessoires et régulatrices, ces gènes s'intercalent entre les gènes essentiels et peuvent les chevaucher. Les rétrovirus réalisent la transcription inverse de leur génome grâce à la RT, formant un intermédiaire ADN bi-caténaire aux extrémités duquel se trouvent des LTR qui agissent principalement sur la régulation de la transcription des gènes.

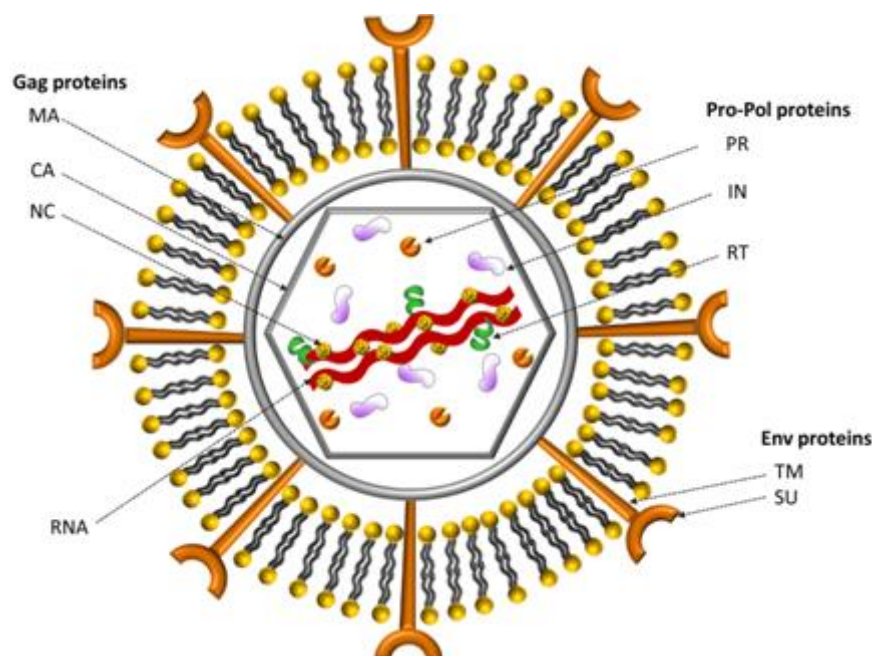


Figure 10 : Représentation schématique de la structure d'un rétrovirus (Rodrigues et al., 2011).

MA : protéine de matrice, CA : protéine de capsid, NC : protéine de la nucléocapsid, RNA : ARN monocaténaire à polarité positive, PR : protéase, IN : intégrase, RT : transcriptase inverse, TM : sous unité transmembranaire des protéines d'enveloppe, SU : protéines de surface de l'enveloppe.

Lorsqu'un rétrovirus infecte une cellule de la lignée germinale, les descendants issus de celle-ci possèdent le provirus dans chacune des cellules nucléées de leur organisme. Ce provirus sera transmis de nouveau à leurs descendants. Le rétrovirus appartiendra alors au génome de l'hôte et sera nommé **rétrovirus endogène** (ERV : « endogenous retrovirus »). Il sera donc transmis verticalement à la descendance, par opposition à la transmission horizontale qui permet la propagation de la forme exogène. Un ERV peut garder la possibilité de se transmettre verticalement et horizontalement pendant un certain temps. Une fois intégré au génome les ERV peuvent proliférer et donner de nombreuses copies à différentes localisations chromosomiques. Les ERV issus d'un même rétrovirus « fondateur » forment une lignée monophylétique qui peut comprendre des provirus fixés et non fixés en fonction de l'activité de réplication, de la dynamique de la population de l'hôte et des pressions sélectives.

Les ERV peuvent proliférer par réinfection lorsqu'ils ont gardé leur capacité de réplication ainsi qu'un gène fonctionnel d'enveloppe. Ils se réinsèrent alors dans la lignée germinale de l'hôte. Grâce à ce mécanisme, certains virus parviennent à se maintenir malgré la sélection purificatrice à laquelle ils sont soumis. Les quelques copies ayant acquis des mutations qui les ont inactivées et qui ont ainsi éliminé les effets néfastes de leur réplication sur l'hôte voient augmenter leur chance d'être fixées. Les ERV peuvent également proliférer par rétrotransposition, en utilisant leurs propres protéines. Les ERV non autonomes (dont certaines protéines fonctionnelles sont manquantes) peuvent proliférer grâce aux protéines d'un autre rétrovirus endogène ou exogène lors d'une co-infection, ou grâce à un rétro-élément éloigné physiquement dans le génome par le mécanisme nommé complémentation en *trans*.

2) *Ordre DIRS*

Les éléments DIRS codent pour une **RT** (Capello et al., 1985) ainsi qu'une **tyrosine recombinase** (YR) (Goodwin et Poulter, 2004) (fig 11). L'action de la YR n'entraîne **pas de TSD** (Goodwin et Poulter, 2001). Cette YR est proche de celle portée par les transposons à ADN appartenant au groupe Crypton (Goodwin et al., 2003). Ils codent également pour une **RH**, une **AP** et une protéine semblable à une **protéine gag** (Goodwin et Poulter, 2004 ; Wicker et al., 2007). Les éléments DIRS-1 possèdent des extrémités répétées inversées (**ITR** pour « Inverted Terminal Repeats ») associées à une région interne qui leur est respectivement inversement complémentaire (**ICR** : « internal complementary region ») alors que les éléments Ngaro, VIPER et PAT présentent des extrémités répétées en sens direct (**SDR** pour « Split Direct Repeats ») (de Chastonay et al., 1992 ; Poulter et Goodwin, 2004 et 2005). Contrairement aux éléments DIRS, les éléments Ngaro et les élément VIPER ne possèdent pas de domaine méthyltransférase (MT) (Poulter et Goodwin, 2005 ; Lorenzi et al., 2006).

Rétrotransposons DIRS

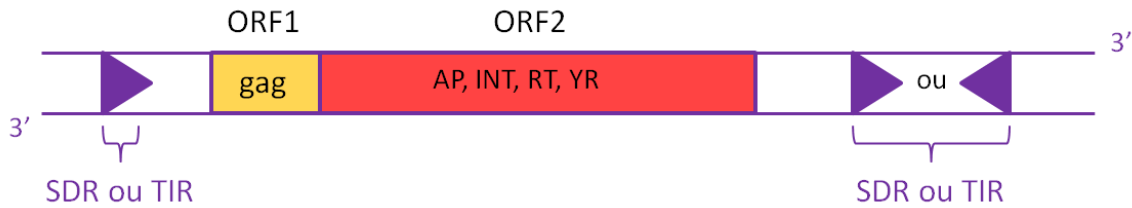


Figure 11 : Représentation schématique des rétrotransposons DIRS.

Le premier élément DIRS a été découvert dans le génome de l'amibe *Dictyostelium discoideum* et a été nommée DIRS-1 pour « Dictyostelium intermediate repeat sequence 1 » (Chung et al., 1983). Deux autres éléments appartenant au groupe DIRS-1 ont été mis en évidence : les éléments PAT chez le nématode *Panagrellus redivivus* (de Chastonay et al. 1992), et les éléments Prt1 chez le champignon filamenteux *Phycomyces blakesleeanus* (Ruiz-Perez et al., 1996). Les éléments Ngaro ont été découverts dans le génome de *Danio rerio*, le poisson-zèbre, et nommés ainsi d'après le nom signifiant « mouche » en Maori. Les éléments VIPER (pour « vestigial interposed retroelement ») ont été identifiés dans le génome de *Trypanosoma cruzi* (Vázquez et al., 2000).

3) Ordre PLE

Les éléments Penelope ont été caractérisés dans le génome d'une drosophile (*Drosophila virilis*). Un élément Penelope est constitué d'une région centrale codant pour une RT ainsi qu'une INT, toutes deux semblables à celles retrouvées chez les rétrovirus (fig 12). Les éléments Penelope sont entourés d'extrémités répétées qui peuvent être dans la même orientation ou inversées (TR : « terminal repeats ») (Evgen'ev et al., 1997). L'insertion d'un élément Penelope génère de courtes TSD (Arkhipova et al., 2013).

Rétrotransposons PLE

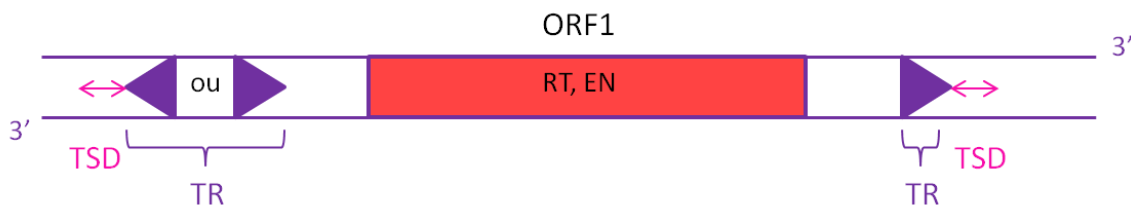


Figure 12 : Représentation schématique des rétrotransposons Penelope.

Certaines copies d'éléments Penelope contiennent un intron mais ce n'est pas un caractère nécessaire. Les éléments Penelope appartiennent à un groupe de rétrotransposons bien distinct des rétrotransposons à LTR et des rétrotransposons sans LTR. Leur RT est plus proche de celle des télomérases que des autres RT (Arkhipova et al., 2003). L'INT des éléments Penelope est en fait une **endonucléase** (EN) appartenant à la famille URI et qui fonctionnerait comme une intégrase (Lyozin et al., 2001).

L'EN Uri n'a été retrouvée chez aucun autre rétroélément. Une première hypothèse évolutive serait qu'elle a été acquise par l'élément Penelope par déplacement de gène non orthologue, par exemple par recombinaison avec un virus codant une EN de la famille Uri tel que chez le baculovirus (Pyatkov et al., 2004). Le déplacement de gènes non orthologues est le remplacement fonctionnel d'une protéine donnée chez une espèce par une protéine non orthologue possédant une activité enzymatique équivalente (Bouvrette, 2010). Une seconde hypothèse serait que ce domaine ait été acquis par fusion indépendante avec la RT ancestrale après sa divergence de la lignée TERT (Telomerase Reverse Transcriptase). Enfin une troisième hypothèse moins probable serait une descendance directe de Penelope d'une lignée provenant d'un ancêtre porteur à la fois des domaines RT et EN Uri. Les subdivisions ultérieures de la lignée TERT auraient alors été accompagnées de la perte du domaine EN. La RT ainsi que l'EN Uri des éléments Penelope sont bien distincts des autres RT ainsi que des autres EN Uri, suggérant ainsi une origine ancienne et monophylétique de cette lignée de rétroéléments Uri (Pyatkov et al., 2004).

4) Ordre LINE

La découverte d'une séquence d'ADN répétée exceptionnellement longue au sein du génome humain a permis de caractériser une nouvelle famille d'ET composée de longues séquences répétées qui sera par la suite nommée LINE pour « Long Interspersed Nucleotide Elements » (Adams et al., 1980 ; Singer, 1982 ; Luan et al., 1993) (fig 13). Aucune répétition de séquence interne ou flanquante n'a été mise en évidence chez ces éléments (Singer, 1982). De nombreuses copies de ces éléments possèdent une **queue polyA** à leur extrémité 3' suggérant qu'ils sont issus d'une transcription inverse à partir d'un ARN (Eickbush et Malik, 2002). Les différents membres d'une même famille de LINE peuvent être de taille variable car la plupart sont **tronqués à leur extrémité 5'** au cours du mécanisme de transposition (Singer et Skowronski, 1985). Les LINE codent tous pour une **RT**. Ils sont divisés en cinq groupes basés sur les relations phylogéniques des transcriptases inverses ainsi que sur la nature et l'enchaînement de leurs domaines protéiques. Ces cinq groupes ont été nommés d'après l'élément caractérisé à l'origine dans chacun d'eux : R2, L1, RTE, I et Jockey (Eickbush et Malik, 2002).

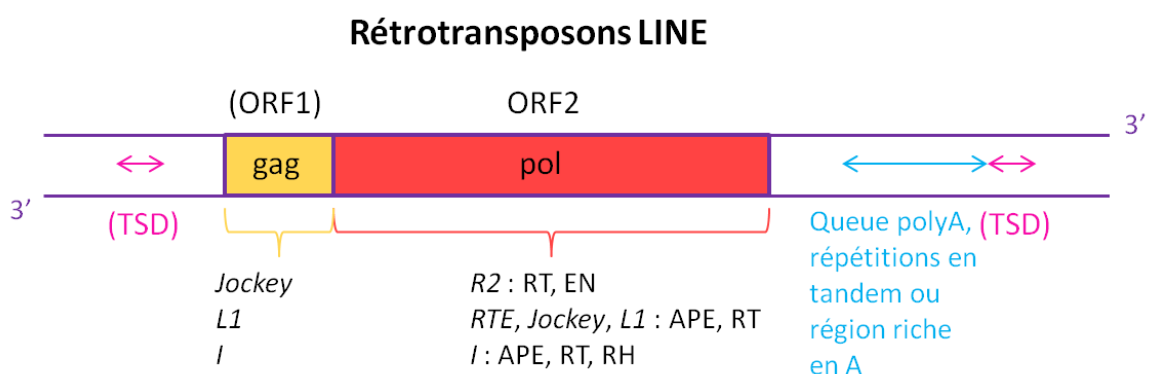


Figure 13 : Représentation schématique des rétrotransposons LINE.

Les éléments de la superfamille **R2** codent pour une **EN** en plus de la RT. La coupure réalisée par cette EN est une coupure décalée (Luan et al., 1993). Grâce aux expériences réalisées par Luan et ses collaborateurs en 1993 sur les éléments R2, le mécanisme de transposition utilisé par les rétrotransposons sans LTR est mis en évidence. Ils suggèrent le terme de « RNA-mediated retrotransposition » pour désigner cette rétrotransposition lors de laquelle l'ARN subit la transcription inverse directement au niveau du site cible de l'intégration (cf paragraphe sur le mécanisme de transposition des rétrotransposons sans LTR) (Luan et al., 1993). Les éléments de la superfamille **RTE** possèdent un seul ORF qui code pour une **endonucléase apurinique** (APE) en plus de la RT (Wicker et al., 2007). Les éléments de la superfamille **Jockey** sont des LINE typiques : ils entraînent une TSD, n'ont pas d'extrémités répétées, sont terminés par une séquence polyA de longueur variable. Ils possèdent deux ORF ; le premier est homologue au gène **gag** des rétrovirus, le second possède des domaines caractéristiques de la RT (Priimägi et al., 1988). Ils codent également pour une **APE** dont la séquence est située en amont de la séquence de la RT (Malik et al., 1999). Les éléments de la superfamille **L1** possèdent également un domaine **APE** situé en amont du domaine de la RT (Malik et al., 1999) ainsi qu'un ORF1 semblable au gène **gag** (Martin, 2006). Enfin, les éléments de la superfamille **I** possèdent, comme les éléments Jockey et L1, une ORF semblable au gène **gag** (Martin, 2006) ainsi qu'une ORF codant une **APE** et une RT, mais cette seconde ORF code également pour une **RH** (Malik et al., 1999).

5) Ordre SINE

Les SINE font entre 80 et 550pb et sont généralement entourés de courtes **TSD** (fig 14). Ils sont formés de trois parties : 1. une partie 5' (« **tête** ») généralement **issue d'un ARNt** (mais pouvant également être issue d'un **ARN 7SL** ou d'un **ARNr 5S**) et qui contient le **promoteur pour l'ARN polymérase III** (ARN Pol III) ; 2. le « **corps central** » qui n'est pas lié à un ARN et dont l'extrémité 3' est semblable au LINE associé, et 3. la « **queue** » en 3' qui est une région riche en A / AT / T, une queue polyT ou une courte répétition en tandem. Le corps et la queue peuvent être tronqués voire absents et le corps peut être formé de deux ou trois unités SINE (dimérisation, trimérisation d'unités identiques ou non) (Kramerov et Vassetzky, 2005).

Rétrotransposons SINE

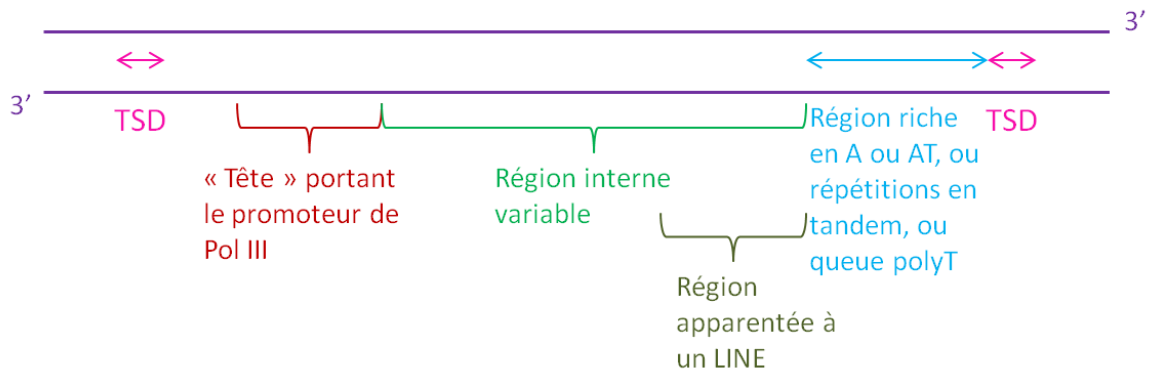


Figure 14 : Représentation schématique des rétrotransposons SINE.

Les SINE (« short interspersed elements ») sont des ET **non autonomes**. Le promoteur interne permet leur transcription par l'ARN Pol III. Les SINE ne sont pas issus de délétions internes des éléments autonomes correspondants contrairement aux autres éléments non autonomes (issus des transposons à ADN, des rétrotransposons LTR et des LINE). Leur partie 5' est issue d'un ARN et leur extrémité 3' est souvent semblable à l'extrémité 3' des LINE appartenant au génome de la même espèce (Kramerov et Vassetzky, 2005). L'**extrémité 3' d'un SINE est reconnue en trans par la RT d'un LINE** possédant la même extrémité 3' et permet ainsi la transposition du SINE par le mécanisme de « target-primed reverse transcription » (TPRT) (Kajikawa et Okada, 2002 ; Kramerov et Vassetzky, 2005). Lorsqu'aucune région apparentée à un LINE n'est présente, la queue riche en A semble s'y substituer fonctionnellement lors de la transposition (Kramerov et Vassetzky, 2005).

B. Mécanisme de transposition des rétrotransposons

Les rétrotransposons peuvent être séparés en trois catégories en fonction du mécanisme utilisé pour intégrer les nouvelles copies au génome de l'hôte. Les **rétrotransposons à LTR** utilisent une **INT de type DDE**, les **rétrotransposons de l'ordre des DIRS** utilisent une **YR** et les **rétrotransposons sans LTR** (LINE et SINE associés) s'intègrent en même temps qu'ils se répliquent grâce à un mécanisme nommé **TPRT** (« target-primed retrotransposition »). Le mécanisme d'intégration des éléments de l'ordre des PLE est encore incertain (Poulter et Butler, 2015).

1) Mécanisme de transposition des rétrotransposons à LTR

Un **intermédiaire à ARN** est d'abord synthétisé, d'un LTR à l'autre, il se termine par une polyadénylation (fig 15). L'étape suivante est la **rétrotransposition** de cet intermédiaire ARN en ADN double brin, qui a lieu dans le cytoplasme. Elle est amorcée par un ARNt, un hétéroduplex ARN-ADN est formé grâce à la RT, puis l'intermédiaire à ARN est éliminé par la RH et le second brin d'ADN est synthétisé par la RT (activité ADN polymérase ADN dépendante de la RT). Il en résulte un **intermédiaire ADN bicaténaire** linéaire contenant des séquences LTR complètes à ses deux extrémités. La dernière étape a lieu dans le noyau, où lorsque celui-ci est rompu lors de la division cellulaire. L'**INT** réalise une coupure à bouts

cohésifs au niveau du site cible du chromosome et lie de manière covalente à cet endroit les extrémités de l'intermédiaire à ADN. La réparation de l'ADN au niveau de cette coupure décalée produit une **TSD** qui est de la même longueur pour toutes les copies d'un même élément (Eickbush et al., 2002 et Finnegan, 2012).

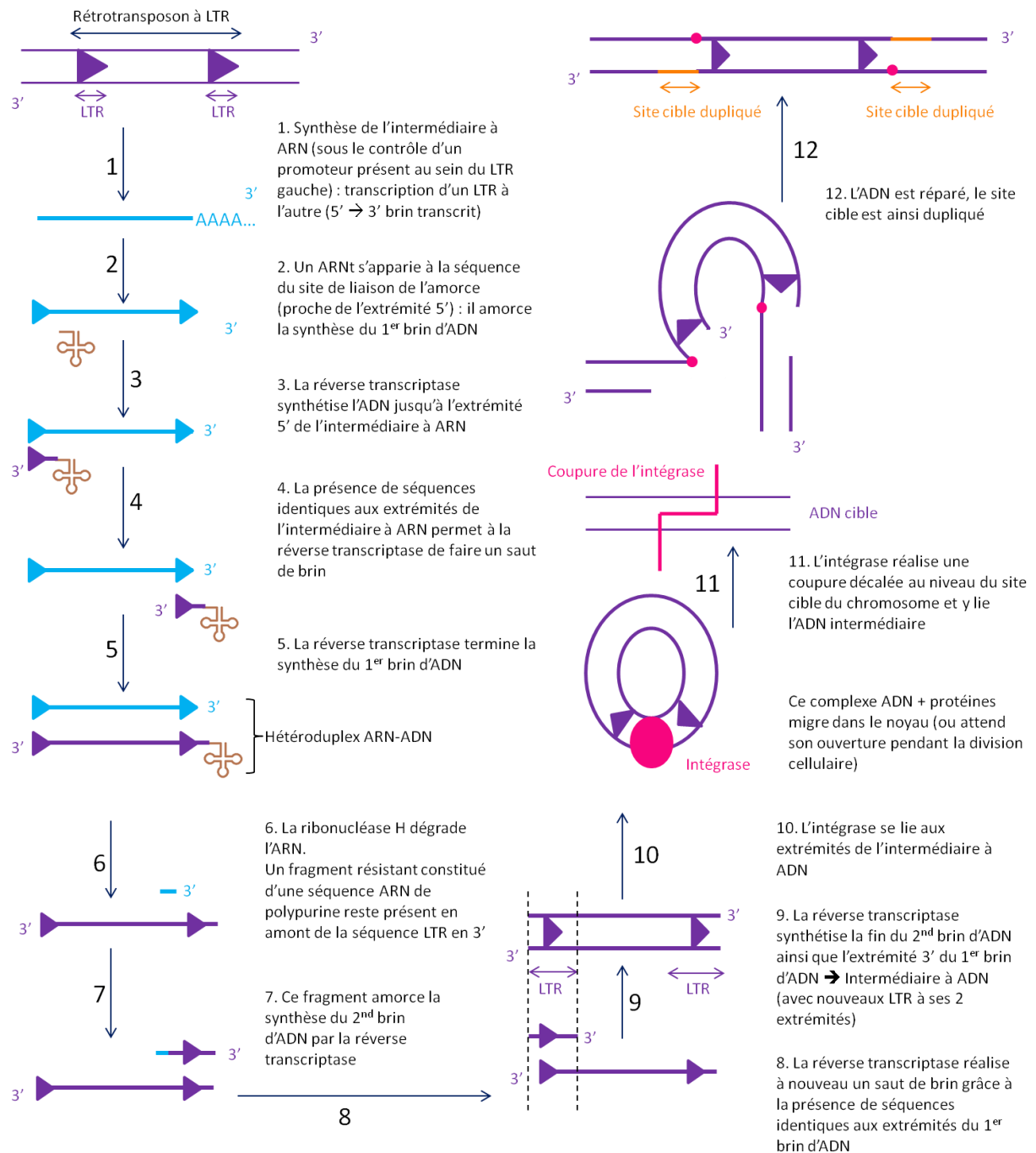


Figure 15 : Mécanisme de transposition des rétrotransposons à LTR (d'après Eickbush et al., 2002 et Finnegan, 2012).

2) Mécanisme de transposition des rétrotransposons appartenant à l'ordre DIRS

Les éléments appartenant à l'ordre DIRS ne codent ni pour une EN (comme les LINE et PLE) ni pour une INT (comme les LTR) mais codent pour une **YR** qui **réalise l'étape d'intégration** de ces éléments. Le mécanisme de transposition proposé est différent selon que les éléments possèdent des ITR ou des SDR (Poulter et Butler, 2015).

Les éléments apparentés à DIRS1 possèdent des **ITR** (ITR gauche long de 316pb et ITR droit long de 352pb) et une **ICR** (les 33pb à gauche de l'ICR sont complémentaires aux premières 33pb de l'ITR de gauche et les 58pb à droite de l'ICR sont complémentaires aux dernières 58pb de l'ITR de droite, 3pb sont communes car l'ICR fait 88pb). A cause de cette structure singulière, un mécanisme de transposition particulier a été proposé pour ces éléments (fig 16). L'élément est transcrit en **ARNm**, puis copié en **ADN simple brin** par la RT. Ensuite, l'ITR gauche partiel s'associe à l'ITR droit permettant de restaurer sa séquence en utilisant l'ITR droit comme modèle. Puis **l'ITR gauche s'associe avec la séquence complémentaire de l'ICR gauche** et **l'ITR droit avec la séquence complémentaire de l'ICR droit**. Ensuite, l'ITR gauche est utilisé comme amorce 3', l'ICR en temps que modèle et la séquence est ligaturée pour obtenir un **ADN circulaire simple brin**. Cette séquence est ensuite convertie en **ADN circulaire double brin**. Cet ADN circulaire double brin sera ensuite intégré à un nouveau site grâce à l'YR, sans entrainer de TSD (Poulter et Goodwin, 2005).

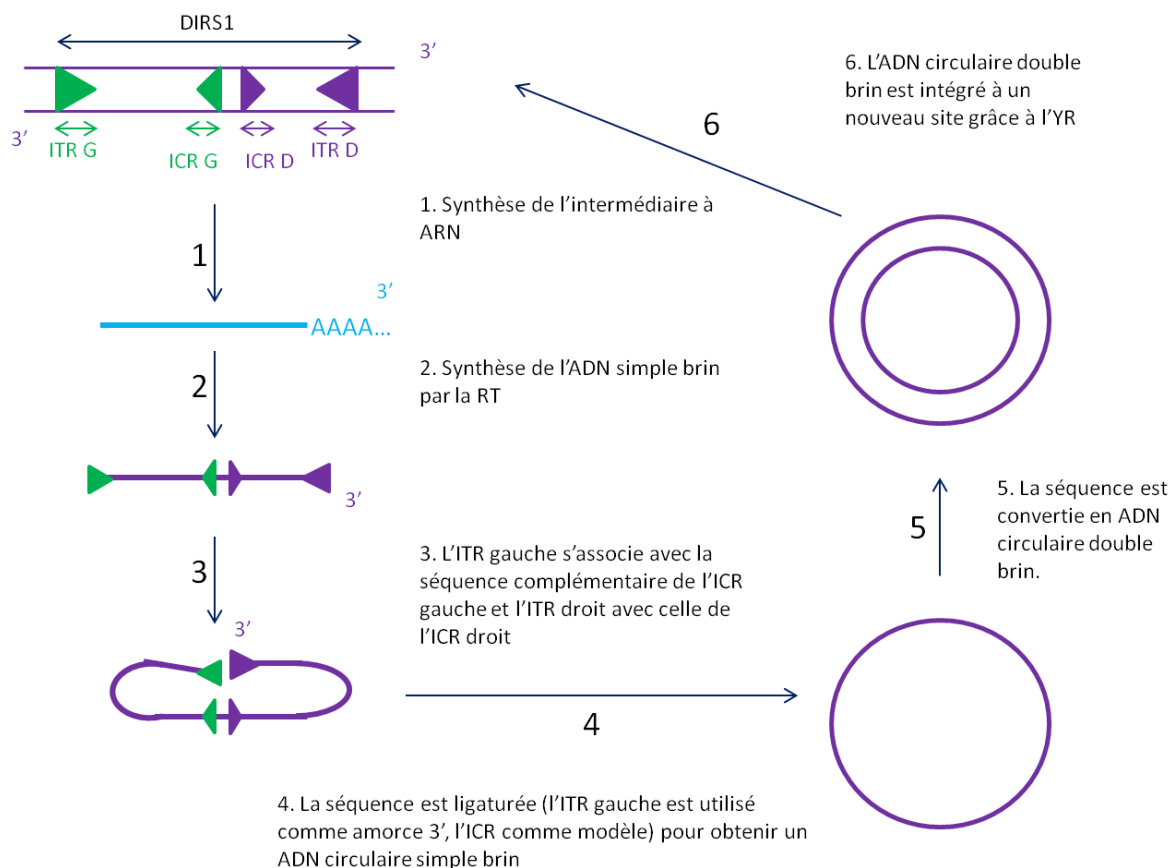


Figure 16 : Mécanisme de transposition des éléments apparentés à DIRS1 (d'après Poulter et Goodwin, 2005).

Les éléments PAT possèdent les mêmes régions codantes que DIRS1 mais leurs séquences répétées sont arrangées différemment (**SDR** dans le cas des PAT), un mécanisme de transposition différent est donc proposé pour les éléments possédant des SDR (fig 17). Afin d'explicitier le mécanisme, les SDR présents dans l'élément sont nommés A1, B1, A2 et B2, de la gauche vers la droite. L'élément est transcrit en **ARNm**, puis copié en **ADN simple brin** par la RT. Ensuite, la **partie A1 de l'ADN simple brin s'associe à la partie complémentaire A2 de l'ARNm** et la **partie B2 de l'ADN simple brin s'associe à la partie complémentaire B1 de l'ARNm**. La RT utilise alors l'extrémité 3' de la séquence A1 comme amorce et la séquence entre B1 et A2 de l'ARNm comme modèle aboutissant à la ligature de l'ADN afin de former un **ADN circulaire simple brin**. Comme pour DIRS1 cette séquence sera ensuite convertie en **ADN circulaire double brin** et intégrée à un nouveau site grâce à l'YR (Poulter et Goodwin, 2005).

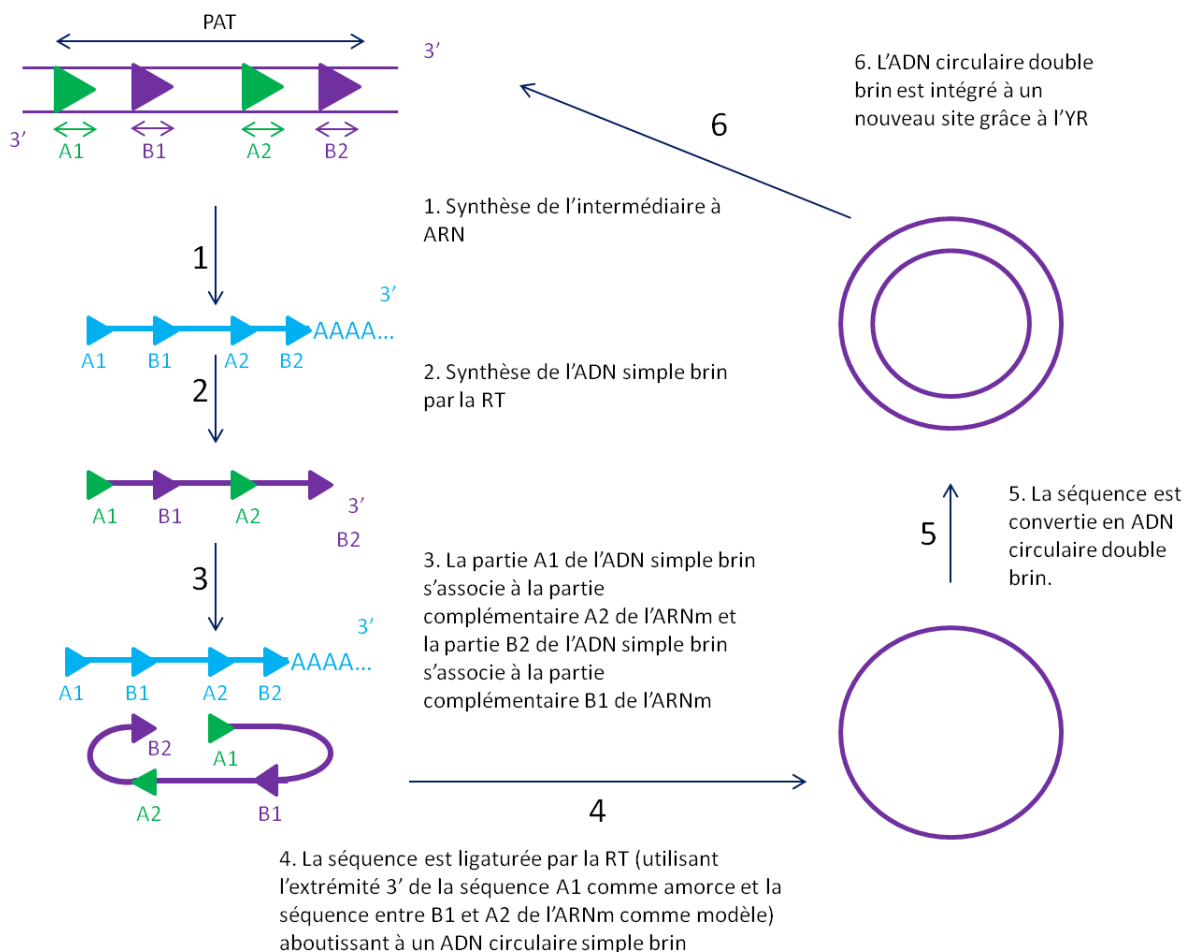


Figure 17 : Mécanisme de transposition des éléments SDR (d'après Poulter et Goodwin, 2005).

3) Mécanisme de transposition des rétrotransposons sans LTR

Contrairement aux rétrotransposons à LTR, ces éléments ne codent pas d'INT, ils ont la même extrémité 3' mais l'extrémité 5' est souvent tronquée et les TSD générés par un même élément n'ont pas tous la même longueur (Eickbush et al., 2002).

Comme pour les rétrotransposons à LTR, un **intermédiaire ARN** est tout d'abord synthétisé, probablement sous le contrôle d'un promoteur interne ou co-transcrit avec les gènes de l'hôte (fig 18). Les rétrotransposons sans LTR codent pour leur propre **EN** ; celle-ci provoque une cassure initiale dans l'ADN cible (et non une rupture du double brin). Le groupe 3' hydroxyle libéré par cette cassure sert à amorcer la transcription inverse de l'intermédiaire ARN. Cette réaction se nomme : « **target-primed reverse transcription** » (TPRT). Ainsi, dans le cas des rétrotransposons sans LTR, la **rétrotransposition** de l'ARN en ADN a lieu **directement au site d'insertion** de la nouvelle copie et non dans le cytoplasme comme pour les rétrotransposons à LTR. Parfois les tentatives de rétrotranscription au site cible de l'ADN sont avortées et plusieurs tentatives sont nécessaires avant d'engager la rétrotranscription. Il en résulte une addition de nucléotides au site cible générant des séquences simples répétées à l'extrémité 3' de l'élément nouvellement intégré. Il en découle des séquences riches en A ou des copies en tandem de répétitions simples (par exemple le motif TAA). Lorsque la RT se sépare de l'ARN avant d'avoir atteint l'extrémité 5' de celui-ci, l'élément inséré est **tronqué à son extrémité 5'**. L'intermédiaire à ARN est dégradé grâce à la RH lorsque celle-ci est codée par le rétrotransposon sans LTR et probablement grâce à l'activité enzymatique cellulaire dans les autres cas. Le mécanisme de synthèse du second brin d'ADN n'est pas élucidé ; il pourrait être réalisé par le rétrotransposon ou par les systèmes cellulaires de réparation de l'ADN. Un modèle alternatif suggère que la cassure du second brin d'ADN génère une amorce pour que la RT réalise la synthèse ADN dépendante du second brin d'ADN (Eickbush et al., 2002 ; Finnegan 2012).

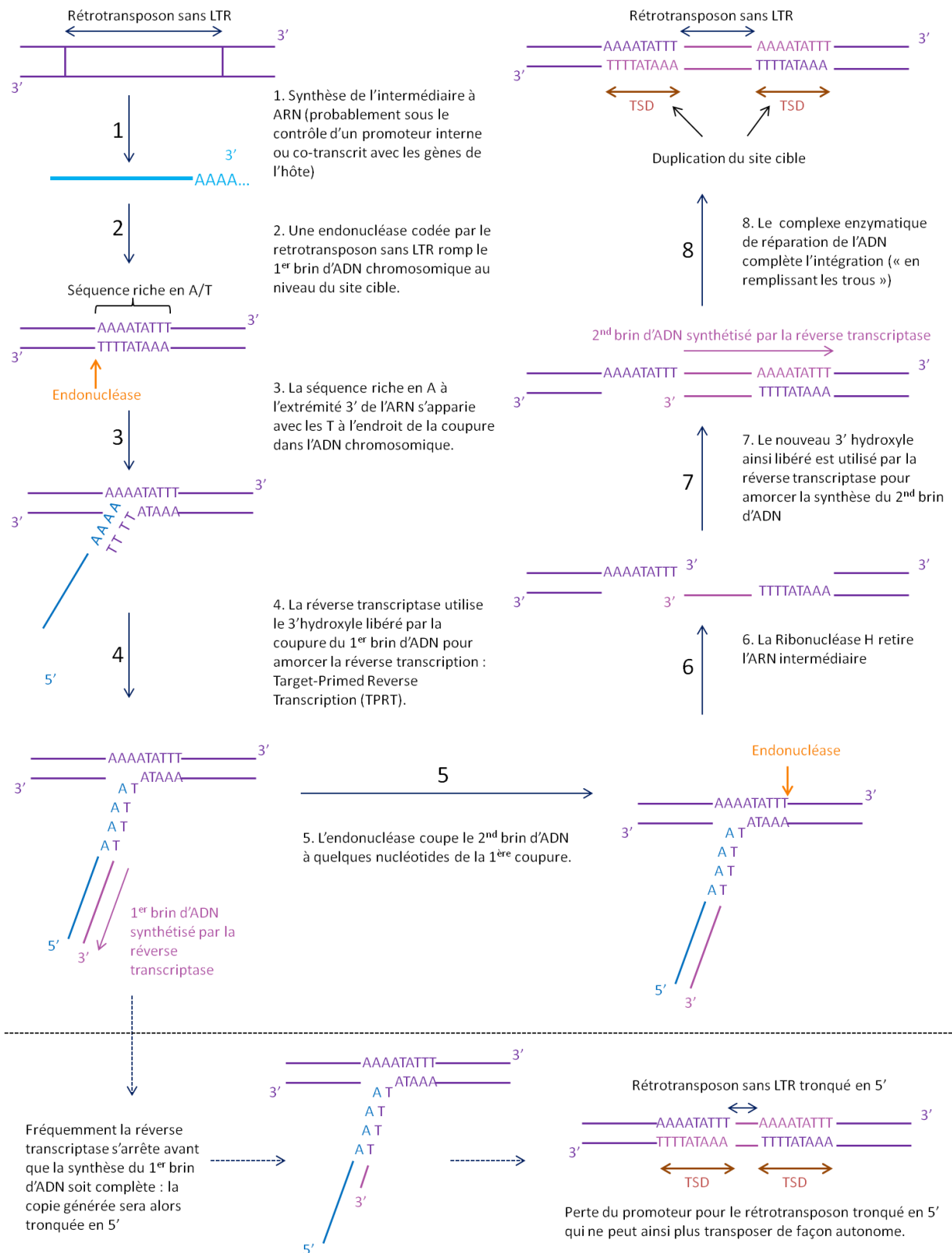


Figure 18 : Mécanisme de transposition des rétrotransposons sans LTR (d'après Eickbush et al., 2002 et Finnegan, 2012).

4) Tableau comparatif des transposons à LTR et sans LTR

Les principales différences concernant les étapes du mécanisme de transposition des rétrotransposons à LTR et sans LTR sont présentées dans le tableau II.

Tableau II: Tableau comparatif de la transposition des transposons à LTR et sans LTR.

	Molécule qui synthétise le nouvel élément, lieu de cette synthèse	Amorce de la synthèse		TSD	Chronologie des étapes de transcription inverse et d'intégration du nouvel élément
		Du 1 ^{er} brin d'ADN	Du 2 nd brin d'ADN		
Transposons à LTR	RT, dans le cytoplasme	ARNt apparié à la séquence du site de liaison de l'amorce (proche de l'extrémité 5' de l'ARN)	Fragment d'une séquence ARN polypurine présente en amont de la séquence LTR en 3'	Oui	Transcription inverse dans le cytoplasme puis migration dans le noyau pour l'intégration
Transposons sans LTR	RT, au niveau du site cible	3' OH libéré par la coupure du 1 ^{er} brin d'ADN	3' OH libéré par la coupure du 2 nd brin d'ADN	Oui	Intégration et transcription inverse simultanées

C. ET de classe II = transposons à ADN

Les éléments regroupés dans la classe II (transposons à ADN) étaient à l'origine ceux qui transposaient grâce à un mécanisme « couper-coller », sans intermédiaire ARN contrairement aux éléments de la classe I. La découverte de nouveaux éléments ainsi que de nouveaux mécanismes de transposition a incité à subdiviser cette classe en fonction des mécanismes de transposition utilisés. Ainsi, les éléments de la classe II peuvent être divisés en deux sous-classes en fonction du nombre de brins d'ADN coupés au cours de la transposition : la sous-classe I regroupe les éléments pour lesquels l'ADN double brin est coupé lors de la transposition alors que les éléments de la sous-classe II sont répliqués sans coupure double brin de l'ADN. La **sous-classe I** comprend les éléments de l'ordre des **TIR** qui utilisent un mécanisme « **couper-coller** » classique lors de leur transposition et les éléments **Crypton**. La **sous-classe 2** comprend les éléments **Helitron** (ou « **Rolling-circle** transposons ») et **Polintons** (ou **Maverick** ou « **Self synthesizing** DNA transposons ») (Wicker et al., 2007 ; Sotero-Caoi et al., 2017).

1) Les transposons à ADN TIR

Les éléments appartenant à l'ordre des TIR possèdent des extrémités terminales inverses (**TIR** pour « terminal inverted repeats ») de taille variable (fig 19). Ils codent pour une **transposase** (Tase) et provoquent une **TSD** lors de leur transposition (Wicker et al., 2007). Neuf superfamilles peuvent être définies selon la séquence des TIR et la taille des TSD : Tc1-Mariner, hAT, Mutator, Merlin, Transib, P, PiggyBac, PIF-Harbinger, CACTA (Wicker et al., 2007).

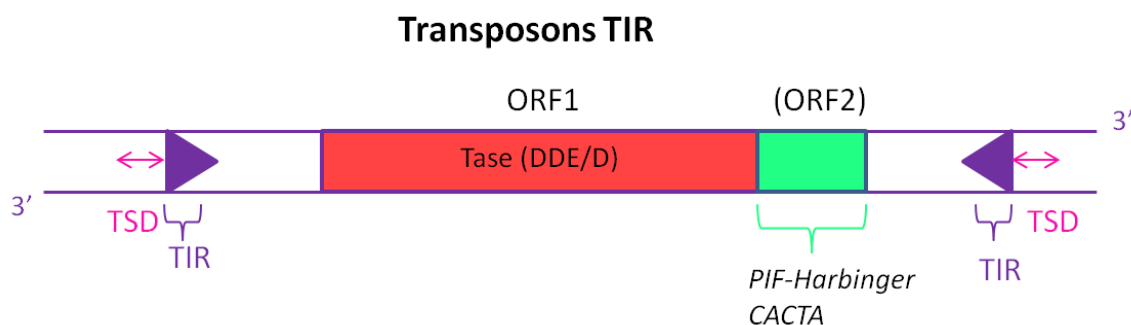


Figure 19 : Représentation schématique des transposons TIR.

Les éléments de la superfamille **Tc1-Mariner** génèrent une **TSD du dinucléotide TA**. Ils sont omniprésents chez les eucaryotes (retrouvés chez les protozoaires, les champignons, les nématodes, les arthropodes, et les chordés). Chez les humains, ils sont considérés comme les fossiles d'une ancienne invasion par un élément actif, alors que leur transposition est la cause principale des mutations spontanées chez le nématode *Caenorhabditis elegans* (Plasterk et Van Luenen, 2002 ; Wicker et al., 2007).

Les éléments de la superfamille **hAT** possèdent des **TSD de 8pb** et des **TIR de 5 à 27pb**. La superfamille a été nommée hAT d'après les premières lettres des trois transposons à l'origine de cette famille : hobo découvert chez la mouche *Drosophila melanogaster*, Activator (Ac) découvert par Barbara McClintock chez le maïs (*Zea mays*) et Tam3 découvert chez la plante nommée muflier (*Antirrhinum majus*). L'élément Dissociation (Ds) découvert par Barbara McClintock en même temps que l'élément Ac est un élément non autonome de cette famille. Les éléments de cette famille ont été identifiés dans la plupart des lignées eucaryotes (champignons, plantes, animaux dont l'Homme) (Kempken et Windhofer, 2001).

Les éléments de la superfamille **Mutator** possèdent des **TSD de 7 à 12pb** et peuvent avoir des TIR relativement longs (environ 100 à 800pb) ou courts voire indétectables. Ces éléments ont été décrits chez les protozoaires (amibes *Entamoeba dispar*, *E. histolytica*, *E. invadens*, *E. moshkovskii*), les champignons (*Fusarium oxysporum*) les plantes (maïs, *Arabidopsis thaliana*, riz, orge, sorgho, lotus...), les animaux (planaire *Schmidtea mediterranea*, sangsue *Helobdella robusta*, nématodes, insectes tels que *Tribolium castaneum*, oursin *Strongylocentrotus purpuratus*, ciones *Ciona intestinalis* et *C. savignyi*, vertébrés tels que cheval, chien, humain) et les virus d'insectes (*Chelonus inanitus bracovirus*, *Glypta fumiferanae ichnovirus*) (Lisch, 2002 ; Chalvet et al., 2003 ; Pritham et al., 2005 ; Wicker et al., 2007 ; Marquez et Pritham, 2010).

Les éléments de la superfamille **Merlin** possèdent des **TSD de 8 à 9pb** et des TIR. Ils ont été décrits dans les génomes des animaux (nématode *Trichuris muris*, trématode *Schistosoma japonicum*, cionnes *C. intestinalis* et *C. savignyi*, vertébrés tels que poisson zèbre *D. rerio*, Xénope lisse *Xenopus laevis*, humain, et insecte *Anopheles gambiae*) ainsi que chez des champignons (*Nosema bombycis*), mais pas chez les plantes (Feschotte, 2004).

Le premier élément intact et actif de la superfamille **Transib**, nommé Hztransib, a été caractérisé chez le ver de l'épi du maïs (*Helicoverpa zea*). Les éléments de cette superfamille codent pour une transposase partageant une similarité significative avec la protéine RAG1 du **système de recombinaison V(D)J** (« Variable, Diversity, Joining ») sur lequel repose la diversité des immunoglobulines des vertébrés. Ces éléments possèdent des TIR et des **TSD de 5pb riches en GC**. Ils ont été identifiés dans le génome des insectes (Bombyx du mûrier *Bombyx mori*, diverses drosophiles, petite guêpe parasitoïde *Nasonia vitripennis* et petit ver de farine *T. castaneum*). Des séquences Transib-like ont été identifiées dans les génomes d'autres animaux (*Ancylostoma caninum*, *S. mediterranea*, *Hydra magnipapillata*), chez le champignon responsable de la rouille du soja (*Phakopsora pachyrhizi*), chez la pomme de terre (*Solanum tuberosum*) ainsi que chez d'autres espèces d'insectes (Chen et Li, 2008).

Les éléments de la superfamille **P** ont été découverts chez *Drosophila melanogaster* chez laquelle ils sont responsables du phénomène d'« hybrid dysgenesis » (développement anormal de la lignée germinale observé lors du croisement entre mâles porteurs des éléments P et femelles non porteuses : fertilité réduite, induction de mutations, réarrangements chromosomiques, lésions de l'ADN, mort des cellules souches germinales). Ces éléments possèdent des **TIR de 31pb** ainsi que des **IIR** (« internal inverted repeats ») **de 11bp** et génèrent **TSD de 8pb**. Des éléments P-like ont été découverts dans de nombreux génomes eucaryotes : chez les vertébrés (humains et autres primates, poisson zèbre, xénope), chez le parasite *Trichomonas vaginalis*, chez la cione *C. intestinalis*, chez l'oursin *S. purpuratus* et chez l'hydre *H. magnipapillata* (Rio, 2002 ; Majumdar et Rio, 2015).

Les éléments de la superfamille **PiggyBac** ont été découverts chez l'insecte « fausse-arpenteuse du chou » (*Trichoplusia ni*). Ces éléments génèrent une **TSD TTAA**. Ils ont été identifiés chez les insectes (*Harmonia axyridis*), les levures (*Saccharomyces cerevisiae*), les protozoaires (*Plasmodium falciparum*), le poisson zèbre, le coq doré (*Gallus gallus*), les mammifères (*Mus musculus*, *H. sapiens*) ainsi que chez le riz (*Oryza sativa*) (Yusa, 2015).

Le premier élément **Harbinger** a été découvert dans le génome d'*A. thaliana* et le premier élément **PIF** (« P instability factors ») dans celui du maïs. La superfamille PIF-Harbinger regroupe les éléments précédemment classés dans la superfamille Harbinger/IS5 et dans la famille PIF. Ils possèdent une **TSD de 3pb** et des **TIR de 14 à 45pb**. Les éléments autonomes possèdent deux ORF, le premier codant pour la transposase et le second pour un domaine de liaison à l'ADN (Jurka et Kapitonov, 2001 ; Nouroz et al., 2016). Ils ont été identifiés chez les plantes, les animaux (moustique *A. gambiae*, poisson-zèbre mais pas chez les mammifères) ainsi que chez le champignon *Moniliophthora perniciosa* (Kapitonov et Jurka, 2004 ; Pereira et al., 2013).

La superfamille **CACTA** est nommée ainsi car ses **TIR de 10 à 28pb** se terminent par un **motif CACTA**. Ses éléments entraînent une **TSD de 3pb**. En plus du motif CACTA, ces éléments possèdent des **répétitions subterminales** (« sub-terminal repeats » : TR) constituées d'un motif de 10 à 20pb répété de façon directe et inversée. Un deuxième ORF

peut également être présent et code pour une protéine dont la fonction est inconnue. Ils ont été décrits dans les génomes des plantes (*Arabidopsis*, pétunia *Petunia hybrida*, carotte *Daucus carota*). Exceptionnellement le **motif CACTG** est présent à la place du motif CACTA comme par exemple chez le riz asiatique (*O. sativa*). Des éléments appartenant à cette superfamille ont également été identifiés chez les animaux avec le **motif CCC** remplaçant le motif CACTA et des **TSD de 2pb** (*T. castaneum* et moustique *Aedes aegypti*) (Wicker et al., 2003 ; DeMarco et al., 2006 ; Nouroz et al., 2017).

Les transposase des éléments de **toutes ces superfamilles** possèdent un **domaine catalytique « DDE/D »**, ce qui suggère une **origine évolutive commune** du mécanisme de transposition « couper-coller » (Yuan et Wessler, 2011). Ce domaine est composé de deux résidus Aspartate (D) et d'un résidu Glutamate (E) ou d'un troisième résidu Aspartate. Cette triade permet à la transposase de catalyser le clivage nucléophile d'une liaison phosphodiester unique. Ce domaine catalytique est impliqué dans la transposition et dans l'intégration rétrovirale. Le cœur catalytique du domaine est constitué d'un mélange d'hélices α et de feuilletts β (fig 20) (Hickman et al., 2010).

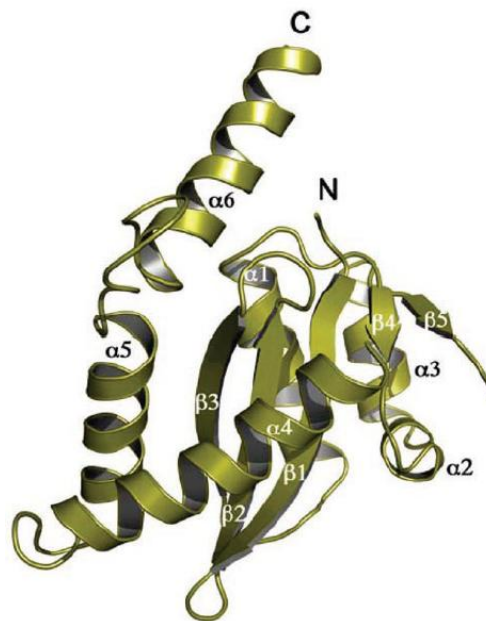


Figure 20 : Diagramme en ruban du domaine catalytique DDE de l'intégrase du VIH-1 (Hickman et al., 2010).

2) Les Cryptons

Les **Cryptons** codent pour une **YR**, comme les rétrotransposons DIRS et Ngaro (fig 21). Cependant, contrairement à ces éléments, ils ne possèdent pas de gène codant pour une RT ou une RH. Ils n'appartiennent donc pas aux rétrotransposons. Les Cryptons sont présents en un faible nombre de copies. Ils ne possèdent pas de TIR ni aucune autre grande répétition, de façon directe ou inversée, que ce soit à leurs extrémités ou à tout autre endroit de leur séquence. Ils semblent générer une **TSD de 4 à 6pb** car de courtes répétitions directes sont observées à leurs extrémités et sont retrouvées au niveau des sites ne

contenant pas le Crypton. Les analyses phylogénétiques de l'YR des Cryptons la regroupent avec celle des éléments DIRS et Ngaro. De plus, leurs extrémités ainsi que leurs sites d'insertion présentent des similitudes suggérant que les Cryptons sont liés de façon évolutive aux groupes DIRS et Ngaro de rétrotransposons à LTR (Goodwin et al., 2003).

Transposons Crypton

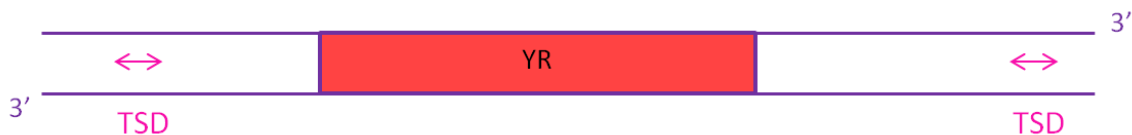


Figure 21 : Représentation schématique des transposons Crypton.

Les Cryptons ont été initialement décrits chez des **champignons pathogènes** (*Coccidioides posadasii*, *Histoplasma capsulatum* et *Cryptococcus neoformans*). Ils ont également été découverts dans les génomes de la diatomée *Phaeodactylum tricornutum*, ainsi que des animaux (moustique *Culex quinquefasciatus*, mollusques *Lottia gigantea*, poisson Médaka *Oryzias latipes*). Des gènes (par exemple *WOC*) dérivés d'éléments cryptons domestiqués sont présents dans le génome humain. Ces gènes contiennent le « domaine 3504 », de fonction inconnue, présent chez les eucaryotes, dérivé de l'YR des Cryptons. La domestication d'éléments semblables aux Cryptons semble avoir été relativement ordinaire dans un passé lointain chez les eucaryotes (la domestication d'un Crypton ayant généré *WOC* semble être le troisième événement le plus ancien de domestication des transposons, il y a plus de 910 millions d'années chez l'ancêtre commun des bilatériens). Les Cryptons sont des **éléments très anciens** qui ont probablement été très **actifs dans un passé lointain** et qui sont actuellement très diversifiés et rarement retrouvés dans les génomes (Kojima et Jurka, 2011).

3) Les Helitrons ou « rolling-circle transposons »

Un **Helitron** (ou transposons en cercle roulant) autonome code pour une protéine constituée des domaines **Rep** (l'initiateur de la réplication en cercle roulant) et **Hel** (une ADN-Hélicase) (fig 22). Le domaine Rep est supposé être l'enzyme impliquée dans les **réactions catalytiques de clivage endonucléolitique**, de **transfert d'ADN** et de **ligature**. Il possède trois motifs conservés au sein des Helitrons et également conservés entre des protéines hautement divergentes catalysant la réplication en cercle roulant de plasmides, de virus à ADN simple brin et de transposons bactériens. Chez certaines espèces, il code également pour une ou plusieurs protéines similaires à la protéine de réplication A, une protéine de liaison à l'ADN simple brin impliquée dans la réplication et la réparation de l'ADN. Chez les espèces animales, la protéine RepHel contient souvent deux autres domaines : l'**APE** habituellement présente au niveau de la portion C terminale de RepHel, et

la **cystéine protéase** précédant le domaine Rep et appartenant à la superfamille des tumeurs ovariennes (OTU) (Kapitonov et Jurka, 2007). Les Helitrons sont les seuls transposons à ADN dont la transposition ne génère pas de duplication du site cible. De plus, le fait qu'ils possèdent un domaine Rep conservé suggère qu'ils transposent par un **mécanisme en cercle roulant** (Kapitonov et Jurka, 2001).

Transposons Helitron



Figure 22 : Représentation schématique des transposons Helitron.

Les Helitrons ont été identifiés chez les plantes (*A. thaliana*, riz asiatique *O. sativa*), les animaux (nématode *C. elegans*, mammifères tels que la chauve-souris *M. lucifugus*) et les champignons (*Aspergillus nidulans*) (Kapitonov et Jurka, 2007; Pritham et Feschotte, 2007). Ils sont présents en faible quantité et limités à de vieux éléments chez la plupart des mammifères excepté les chauves-souris chez lesquelles de nombreux jeunes éléments sont retrouvés (Pritham et Feschotte, 2007 ; Kapitonov et Jurka, 2007).

4) Les Polintons

Les **Polintons** (également appelés **Mavericks** ou « **Self synthesizing DNA transposons** ») sont les transposons les plus complexes présents chez les eucaryotes. Ils peuvent coder jusqu'à dix protéines différentes dont l'ordre est variable. Les protéines spécifiques des Polintons sont une **ADN polymérase B (POLB)**, une **INT**, une **protéase (PRO)** à cystéine et une **ATPase** (fig 23). Les autres protéines ne sont pas classifiées. Ils possèdent également des **TIR de plusieurs centaines de pb** et génèrent une **TSD de 6pb** (Kapitonov et Jurka, 2006).

Transposons Polinton

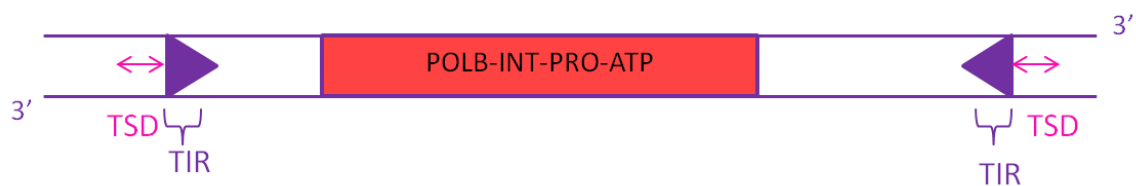


Figure 23 : Représentation schématique des transposons Polinton.

Ils sont présents chez les champignons (*P. pachyrhizi*) et les animaux (poisson-zèbre, crapaud griffu à peau lisse *X. tropicalis*, cione *C. intestinalis*, oursin *S. purpuratus*, drosophile *D. yakuba* et petit vers de farine *T. castaneum*) (Kapitonov et Jurka, 2006).

5) Les MITE

Les éléments **MITE** (pour «Miniature Inverted-repeat Transposable Element» en anglais) sont des transposons à ADN tronqués, **non autonomes**. Ils possèdent des TIR et TSD mais aucun ORF. Ils transposent grâce aux transposases codées par les transposons à ADN autonomes (Lu et al., 2012).

D. Mécanisme de transposition des transposons à ADN

1) Transposases et mécanisme de transposition des transposons à ADN TIR

Les **transposons à ADN TIR** autonomes transposent grâce à une **transposase**, enzyme capable de déplacer des segments d'ADN discrets d'un site donneur vers une nouvelle localisation dans le génome sans utiliser d'intermédiaire ARN. Certains éléments non autonomes ne codent plus pour cette transposase mais possèdent des extrémités pouvant être reconnues par une transposase codée par un autre élément et peuvent ainsi également être déplacés au sein du génome. A l'origine le terme « transposase » était réservé aux enzymes ne nécessitant pas de régions d'homologie significative entre une partie du transposon et le site cible, mais elles peuvent présenter des exigences d'homologie variable ayant conduit à les nommer dans certains cas « résolvas » ou « recombinaisons ». Les transposases peuvent présenter quatre types de domaines catalytiques différents permettant la transposition : domaine catalytique « DDE/D » (voir présentation des transposons à ADN TIR), le domaine catalytique HUH (agissant sur l'ADN simple brin), la sérine transposase et la tyrosine transposase. Ces domaines sont retrouvés également dans d'autres enzymes responsables de réarrangements de l'ADN telles que les intégrases rétrovirales, les invertases, les résolvas, les recombinaisons site-spécifique ainsi que la recombinaison RAG-1 impliquée dans la recombinaison V(D)J (Hickman et Dyda, 2015).

Les transposases des éléments classés dans l'ordre des TIR possèdent toutes un **domaine catalytique DDE/D** (Yuan et Wessler, 2011). Les enzymes DDE sont capables de catalyser deux réactions chimiques : l'**hydrolyse de la liaison phosphodiester** présente à chaque extrémité du transposon (générant ainsi des extrémités 3'OH libres), puis la **trans-estérification** joignant les extrémités 3'OH libres à l'ADN cible (fig 24). Ces étapes se déroulent dans un complexe nucléoprotéique appelé le **transpososome** et permettant l'insertion combinée des deux extrémités du transposon dans l'ADN cible. L'**attaque nucléophile** de l'ADN cible par les extrémités 3'OH se fait de façon **décalée** au niveau de l'ADN cible (généralement décalage de 2 à 9pb) aboutissant à une **TSD** lors de la réparation de l'ADN (dont la longueur est caractéristique pour chaque transposon) (Curcio et Derbyshire, 2003).

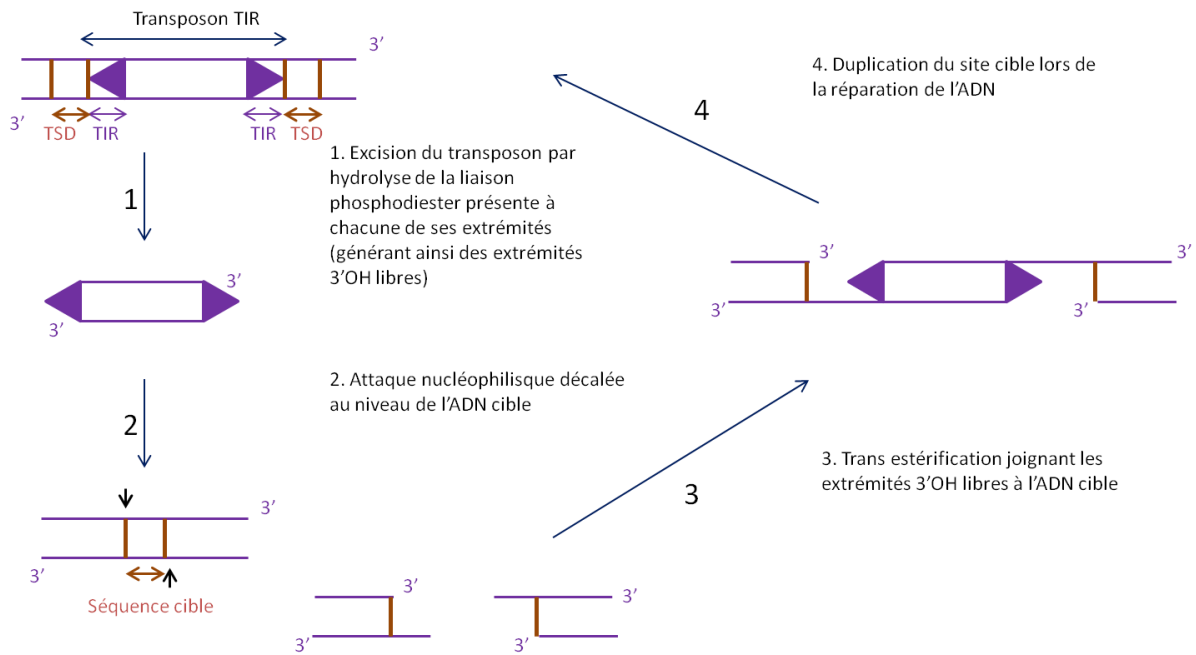


Figure 24 : Mécanisme de transposition des transposons TIR.

2) Mécanisme de transposition des Cryptons

Le mécanisme de transposition de Cryptons est proposé par analogie à celui des transposons procaryotiques codant pour une YR. Tout d'abord, une copie d'un élément Crypton est probablement **excisée du génome hôte par recombinaison** grâce à l'YR au niveau des courtes répétitions directes (TSD) qui entourent cet élément (fig 25). Il en résulte un **ADN circulaire double brin** extrachromosomique ne contenant qu'une copie de séquence directe répétée appelée « séquence donneuse ». La séquence d'ADN circulaire extrachromosomique migre alors vers un nouveau site du génome de l'hôte qui contient une séquence identique ou très similaire à la séquence donneuse, appelée « séquence cible ». La dernière étape consiste en une **recombinaison entre la séquence cible et la séquence donneuse** aboutissant à l'intégration du Crypton au site cible. Ce Crypton sera alors entouré de répétitions directes (correspondant aux TSD) qui seront parfaites si la séquence donneuse et la séquence receveuse sont identiques, ou imparfaites si elles ne sont que très similaires (Goodwin et al., 2003).

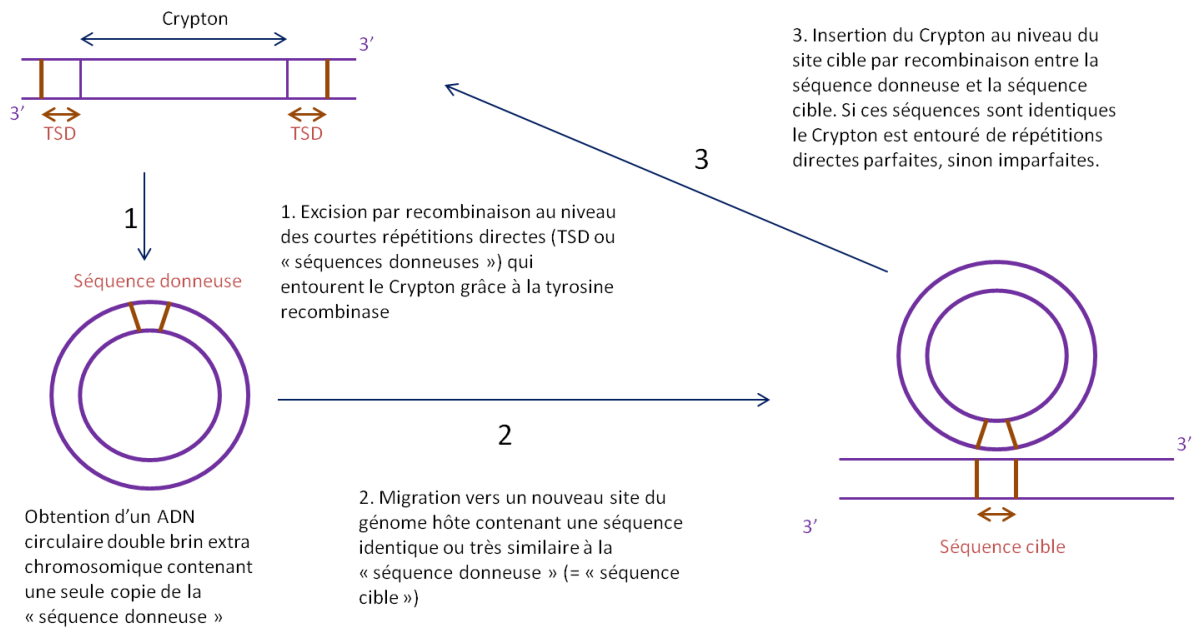


Figure 25 : Mécanisme de transposition des Cryptons (d'après Goodwin et al., 2003).

3) Mécanisme de transposition des Helitrons

Le modèle de transposition proposé pour les Helitrons est nommé « **rolling-circle replication** » à partir d'un intermédiaire ADN simple brin. Tout d'abord, l'extrémité N de la protéine Rep est supposée se lier à l'extrémité 5' d'un Helitron. Les deux premiers résidus de tyrosine de la protéine Rep clivent respectivement l'extrémité 5' du brin donneur de l'Helitron et l'ADN cible et se lient de façon covalente aux extrémités 5' des brins respectifs. Un **transfert de brin** est alors permis par l'attaque de la liaison entre l'ADN de l'Helitron et le premier résidu de tyrosine par le 3'OH libre au niveau de l'ADN cible. La protéine Hélicase permet le déroulement de l'ADN dans le sens 5' vers 3', et l'extrémité 3'OH libre au niveau du brin donneur sert d'amorce à l'ADN polymérase de l'hôte pour synthétiser l'ADN au niveau du brin donneur. L'ADN de l'Helitron lié au site cible reste simple brin grâce à des protéines se liant à lui et le protégeant. Lorsque Rep atteint le terminateur RC à l'extrémité 3' de l'Helitron, le premier résidu de tyrosine clive l'ADN à cet endroit. Un **nouveau transfert de brin** est permis par l'attaque de la liaison entre la deuxième tyrosine et l'ADN cible par l'extrémité 3'OH de l'Helitron générée. Une copie de l'Helitron sous forme d'ADN simple brin est ainsi intégrée au niveau du site cible et peut être répliquée passivement par l'hôte (Kapitonov et Jurka, 2007 ; Thomas et Pritham, 2015).

Un modèle différent est proposé dans le cas de l'Helitron Helraiser (fig 26). Dans ce modèle, l'attaque du brin donneur et la liaison au site cible ne sont pas simultanées. La zone de reconnaissance permettant la coupure de l'Helitron au niveau de son extrémité 3' peut être lue sans entraîner de coupure, alors une région flanquante du génome de l'hôte est mobilisée avec l'Helitron, générant une nouvelle extrémité 3' (Grabundzija et al., 2016).

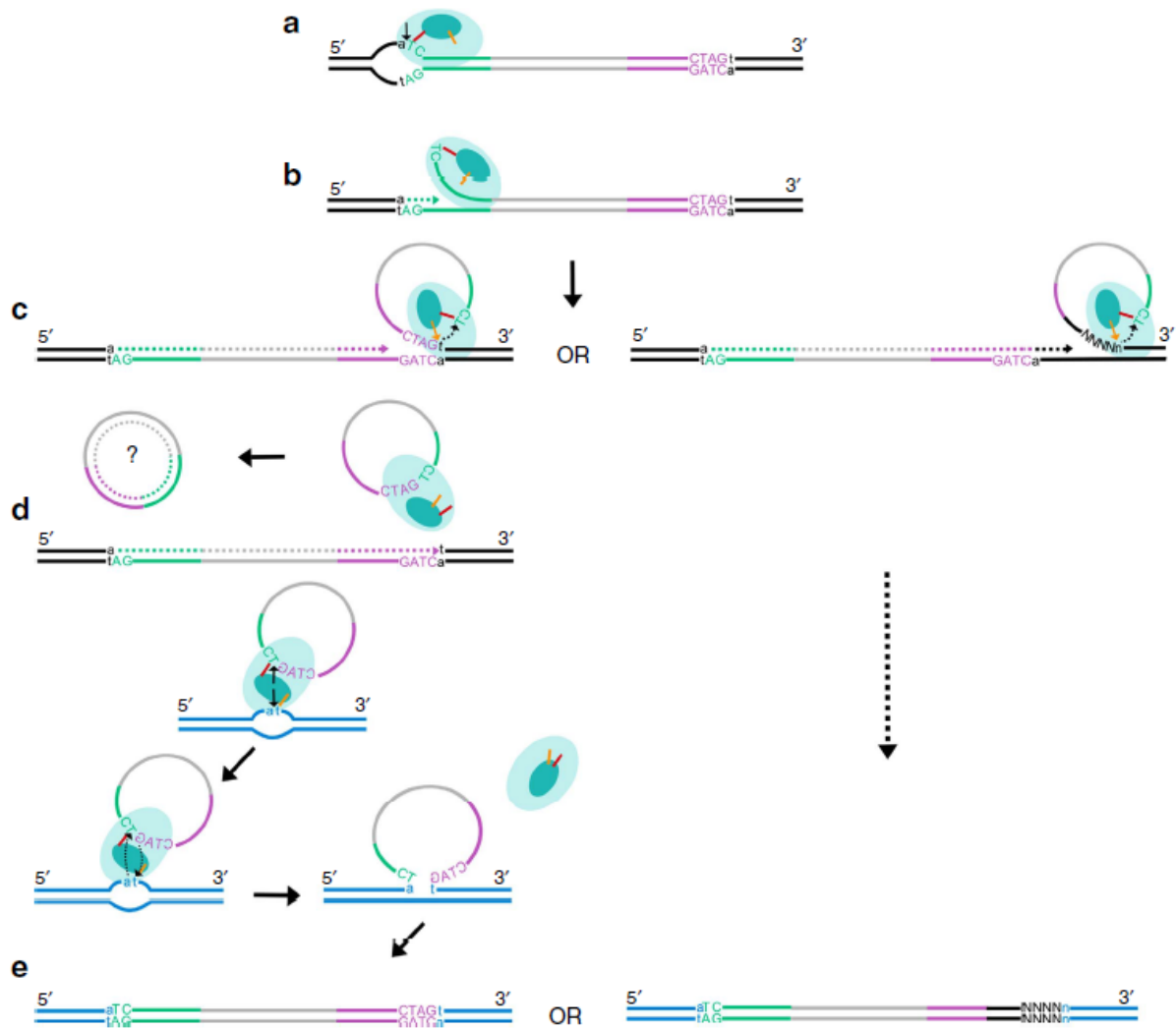


Figure 26 : Mécanisme de transposition d'un Helitron (Grabundzija et al., 2016).

4) Mécanisme de transposition des Polintons

Les Polintons (ou Self synthesizing DNA transposons) semblent utiliser un modèle de transposition complètement différent de ceux décrits précédemment chez les transposons (Kapitonov et Jurka, 2006). Tout d'abord, l'INT excise le Polinton de l'ADN pendant la phase de réplication du génome de l'hôte (fig 27). Le Polinton extra chromosomique simple brin ainsi obtenu a une structure en tige et boucle. La réplication du Polinton est ensuite réalisée par l'ADN polymérase B (POLB) et aboutit à un Polinton double brin. Cette réplication est initiée par une protéine terminale (TP, probablement issue du domaine N terminal de POLB) qui se lie à l'extrémité 5' libre du Polinton. L'INT se lie aux extrémités du Polinton et permet son intégration dans le génome de l'hôte (Kapitonov et Jurka, 2006).

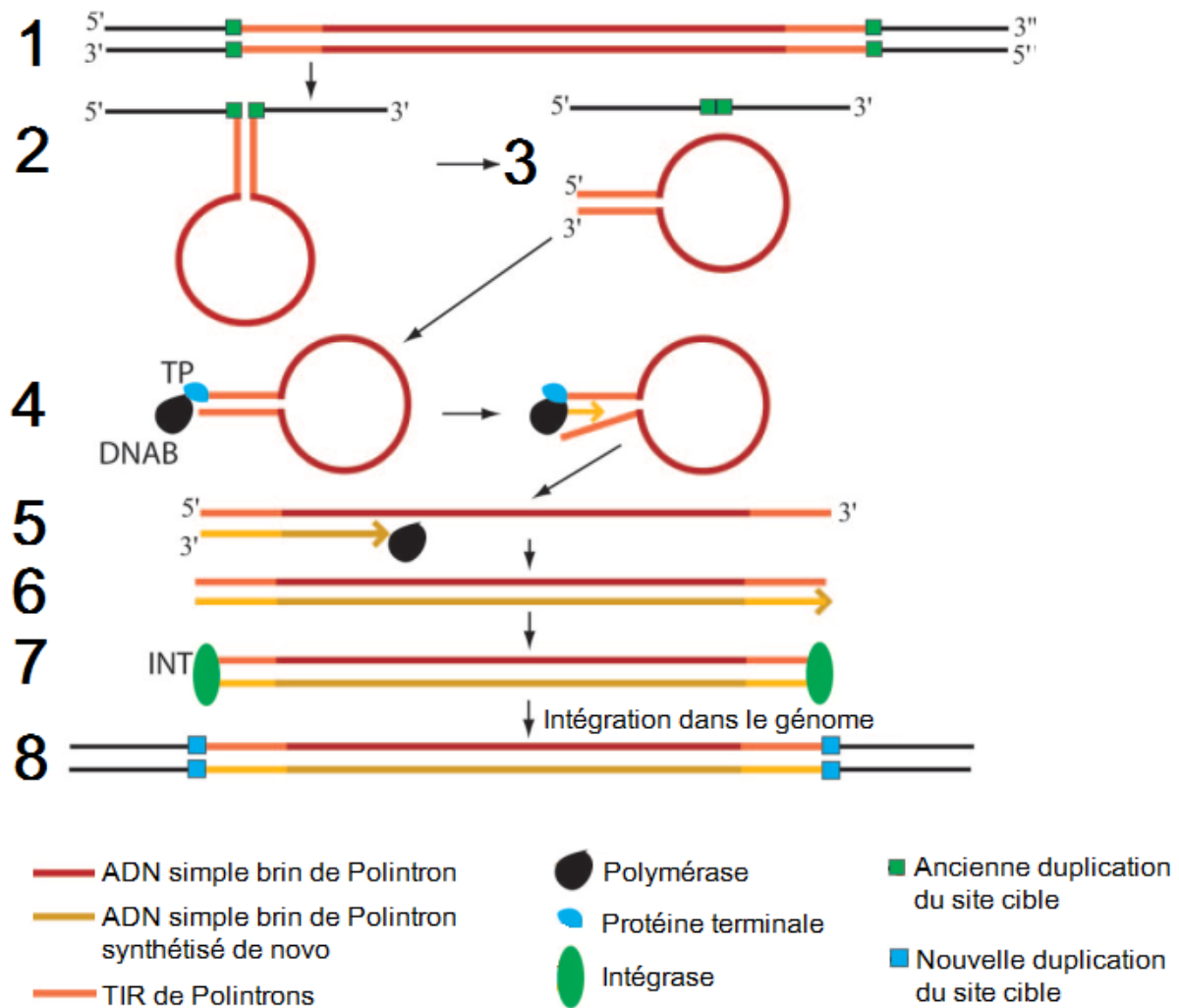


Figure 27 : Mécanisme de transposition des Polintons (d'après Kapitonov et Jurka, 2006).

Un mécanisme alternatif aurait pu être un couper-coller réalisé par l'INT des Polintons, mais Kapitonov et Jurka avancent des arguments suggérant que les Polintons suivent plutôt le mécanisme d'**autosynthèse amorcée par une protéine** qu'ils ont proposé. Parmi ces arguments figurent la conservation des domaines conférant les activités de polymérisation d'ADN dépendant de l'ADN ainsi que de relecture (correction des erreurs) chez des POLB très divergentes indiquant ainsi qu'elles sont nécessaires à la transposition des Polintons (Kapitonov et Jurka, 2006).

E. Tableau récapitulatif des différents ET

La structure des différents ET est présent dans le tableau comparatif suivant (tab III).

Tableau III : Tableau récapitulatif de la structure des différents ET.

Ordre	Superfamilles	Groupes	Protéines codées	Autre particularité	Nature des extrémités	TSD	
LTR	Copia		GAG AP- INT-RT -RH		LTR	Oui	
	Gypsy Bel-Pao		GAG AP-RT – RH-INT				
	Retrovirus ERV		GAG AP-RT – RH-INT ENV				
DIRS	DIRS	DIRS-1	GAG-AP- RT –RH-YR	Domaine MT	ITR (associées à ICR)	Non	
		PAT		Absence de MT	SDR		
	Ngaro						
	VIPER						
PLE	Penelope		RT-EN		TR	Oui	
LINE	R2		RT-EN		5' : parfois troncature ; 3' : polyA / répétition en tandem / région riche en A	Oui en général	
	RTE						
	Jockey		Gag-like				
	L1		EN-APE				
	I		Gag-like EN-APE- RH				
SINE	tRNA		Aucune	ARN Pol III promoteur, extrémité 3' du corps semblable au LINE associé	3' riche en A / AT / T, queue polyT ou courte répétition en tandem	Oui	
	7SL						
	5S						
TIR	Tc1-Mariner		Tase (DDE/D)		TIR	TA	
	hAT				TIR 5-27pb	8pb	
	Mutator				TIR jusqu'à plusieurs 100 ^{aines} pb	7-12pb	
	Merlin				TIR	7-9pb	
	Transib					5pb, riches en GC	
	P				TIR 31 pb	8pb	
	PiggyBac				TIR	TTAA	
	PIF-Harbinger				ORF2	TIR 14-45pb	3pb
	CACTA					TIR 10-28pb	2-3pb
	Cryptons				YR		Pas de TIR
Helitron			Rep-Hel			Non	
Polinton (Maverick)			POLB-INT- PRO-ATP		TIR plusieurs 100 ^{aines} pb	6pb	

F. Transmission des ET par transfert horizontal

L'information génétique peut être transmise d'une espèce hôte à une autre de deux façons : par héritage vertical du parent à la progéniture (transfert vertical), ou par **transfert horizontal** (transmission autre que par la reproduction) (Wallau et al., 2018). Cet échange peut avoir lieu entre des espèces très éloignées phylogénétiquement (Syvanen, 1994). En plus de leur capacité à proliférer au sein d'un génome, les ET sont également capables de passer occasionnellement d'un génome à l'autre par un mécanisme de transfert horizontal de gène (HGT pour Horizontal Gene Transfer).

Pour qu'on puisse observer un HGT d'ET entre deux espèces, il faut tout d'abord que l'ADN de l'ET soit physiquement transféré entre le donneur et le receveur, puis que la fréquence de l'ET augmente chez ce receveur afin d'aboutir à l'invasion de cette nouvelle espèce par l'ET. Ainsi l'ET doit être transmis physiquement au receveur (intervention d'un vecteur), intégré à sa lignée germinale et exprimé en échappant aux mécanismes de répression de la transposition du receveur (Silva et al., 2004).

1) Critères utilisés pour détecter les HGT d'ET

Lorsque les ET sont hérités verticalement, leur phylogénie est similaire à celle de leur hôte. Lorsque ce n'est pas le cas, les écarts par rapport à la phylogénie attendue permettent d'argumenter qu'un HGT a eu lieu. Trois différentes « anomalies phylogénétiques » sont ainsi utilisées pour détecter les HGT : la similitude élevée entre séquences d'ET appartenant à des espèces très éloignées, l'incohérence entre les phylogénies de l'hôte et de l'ET et la distribution éparse d'un ET dans un groupe d'espèces (Loreto et al., 2008 ; Schaack et al., 2010).

Le critère le plus fréquemment utilisé est un **degré de similitude élevée** entre des séquences présentes chez des hôtes appartenant à des **taxons éloignés**. Dans ce cas, la longueur de la branche portant l'ET est beaucoup plus courte qu'attendu compte tenu du temps écoulé depuis la divergence des espèces hôtes (la divergence entre les séquences des ET est inférieure à la divergence entre les séquences des gènes nucléaires non mobiles des hôtes grâce à laquelle leur phylogénie a été construite). La **distribution éparse** d'une famille d'ET parmi des taxons étroitement apparentés signe également un transfert horizontal. On appelle distribution éparse le fait qu'un ET soit présent chez une lignée mais absent chez sa lignée sœur. L'ET a ainsi été acquis par un transfert horizontal qui n'a pas impliqué la lignée sœur. Cependant, cette preuve seule n'est pas suffisante car une autre explication possible serait la perte de l'ET par la lignée sœur (Silva et al., 2004). La localisation de **positions orthologues vides** chez d'autres espèces permet de vérifier que l'ET n'a pas été hérité verticalement (dans ce cas, il serait retrouvé à cette même position génomique chez les autres espèces) (Schaack et al., 2010).

La similitude entre les ET doit être minutieusement analysée car de nombreuses **copies paralogues** peuvent proliférer dans une lignée. La comparaison de copies différentes paralogues conduira à un arbre phylogénétique de l'ET ne correspondant pas à celui de l'espèce et un HGT sera suspecté à tort. Par exemple, les génomes de plusieurs espèces

contiennent plus d'une sous-famille d'ET tels que les mariner, les gypsy et les éléments P (Loreto et al., 2008).

L'identification d'un HGT est complexe car ces différentes « anomalies phylogénétiques » peuvent également être issues d'autres processus évolutifs tels que la perte stochastique et/ou la fixation différentielle du polymorphisme ancestral, la préservation des ET pour la fonction cellulaire grâce à la sélection purificatrice ou bien des taux et modes d'évolution variables des ET. Il faut donc s'appuyer sur plusieurs sources de données pour argumenter en faveur d'un HGT et réfuter les hypothèses alternatives. L'**identification du vecteur ou du mécanisme** ayant permis le HGT constitue une preuve ultime mais est rarement possible (Schaack et al., 2010).

Un vecteur d'ET possible doit pouvoir assurer le transfert physique de l'ADN d'un donneur à la lignée germinale du receveur. Ce vecteur doit donc avoir accès à l'environnement intracellulaire ou aux cellules sans les détruire. Différents vecteurs ont ainsi été proposés. Les **acariens parasites**, par exemple *Proctolaelaps regalis* a été proposé comme vecteur pour le HGT des éléments P entre différentes espèces de Drosophiles (Houck et al., 1991). Les **virus** possèdent des caractéristiques les désignant comme vecteurs potentiels : ils sont capables d'infecter des espèces variées, certains possèdent une étape d'intégration au génome de l'hôte, un stade de leur cycle de vie est intracellulaire et certains ont un tropisme pour les cellules gamétiques (Wallau et al., 2018). Les poxvirus ont été proposés comme vecteurs pour le HGT des rétrotransposons des reptiles aux mammifères (Piskurek et Okada, 2007) et les virus à ADN ont été proposés comme vecteur pour le HGT des SPIN (SPACE INVADERS, transposons à ADN appartenant à la superfamille des hAT) entre différents tétrapodes (*Rodentia* : rongeurs ; *Galago* : Galago ; *M. lucifugus* : Petite chauve-souris brune ; *Tenrecidae* : Tenrec ; *Didelphimorphia* : Opossums ; *A. carolinensis* : Anoles verts et *X. laevis* : Xénopes lisses) (Pace et al., 2008). Les **guêpes parasitoïdes** et les **parasites intracellulaires** (par exemple la bactérie *Wolbachia*) ont également été proposés comme vecteurs possibles d'HGT (Houck et al., 1991).

La **comparaison des taux de mutations synonymes** (Ks) des ET par rapport à ceux des gènes orthologues peut être utilisée pour identifier des HGT. Dans ce cas, l'ET issu de HGT aura été acquis plus récemment et aura accumulé moins de mutations synonymes que les gènes de l'hôte. L'avantage de cette méthode est qu'elle peut être appliquée à des espèces étroitement apparentées (Schaack et al., 2010).

La **datation de l'amplification d'un ET** permet d'estimer la date à laquelle le HGT a pu avoir lieu. En effet, l'amplification d'un ET a lieu immédiatement après l'introduction d'une copie active comme le montrent les données empiriques ainsi que les simulations (Schaack et al., 2010).

2) *Éléments impliqués*

La capacité des ET à réaliser des HGT constitue un **avantage sélectif** car elle augmente leur probabilité de survie au sein d'une gamme taxonomique élargie ainsi que sur de longues périodes d'évolution (Kidwell, 1992).

Les HGT concernent aussi bien les transposons à ADN que les rétrotransposons. Par exemple, chez les drosophiles, 218 cas de transferts horizontaux ont été observés (111 concernent des rétrotransposons et 107 concernent des transposons à ADN) (Schaack et al., 2010).

Certains ET semblent plus propices aux HGT en raison de leurs caractéristiques (fig 28). Par exemple, les ET de classe II passent par un intermédiaire extra chromosomique ADN qui est plus stable que l'intermédiaire extra chromosomique ARN des éléments de classe I. Lorsque les rétrotransposons à LTR transposent, un intermédiaire ADN est produit et peut être intégré dans le génome de la même façon que pour les éléments de classe II, alors que lorsque les rétrotransposons sans LTR transposent, l'intermédiaire ARN subit la transcription inverse directement au niveau du site cible chromosomique. Cette différence pourrait expliquer la rareté des HGT impliquant des rétrotransposons sans LTR. Un moyen pour ces éléments de réaliser un HGT pourrait cependant être d'utiliser un virus comme vecteur, la transcription inverse ayant alors lieu directement dans l'ADN viral ce qui représenterait un évènement très rare. Ainsi, les évènements de HGT semblent **plus répandus chez les éléments de classe II** que chez les éléments de classe I, et parmi les éléments de classe I, plus courants **chez les rétrotransposons à LTR** que chez les rétrotransposons sans LTR (Silva et al., 2004).

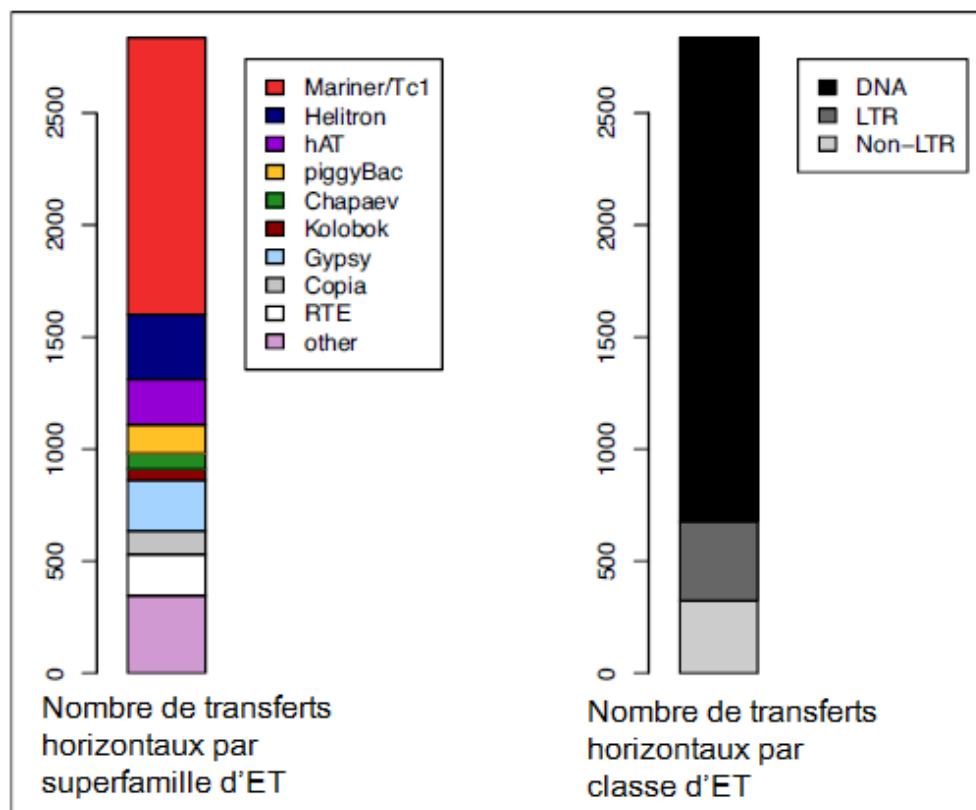


Figure 28 : Nombre de transferts horizontaux d'ET de vertébrés par superfamille et par classe d'ET (Gilbert et Feschotte, 2018).

De plus, l'autorégulation des ET de classe II semble les conduire à l'extinction en les empêchant de maintenir un taux de transposition compatible avec leur survie à long terme dans un même génome. Le HGT de ces éléments pourrait alors être l'évènement permettant à certaines familles d'ET de classe II de survivre (Robertson et Lampe, 1995 ; Silva, 2004).

IV. Rôle des ET dans l'évolution des organismes

Les ET jouent un rôle quantitatif dans la taille des génomes, mais également qualitatif dans leur capacité à perturber l'expression des gènes voire à en créer de nouveaux.

A. Taille et composition des génomes

Avant que l'influence majeure des ET dans la taille des génomes ne soit identifiée, la communauté scientifique faisait face à un paradoxe nommé « le paradoxe de la valeur C ».

1) *Le paradoxe de la valeur C*

La valeur C est la taille d'un génome haploïde exprimée en nombre de paires de bases. Cette valeur est constante pour une espèce donnée mais varie largement d'une espèce à l'autre. La taille du génome, contrairement à ce qui était attendu, n'est pas corrélée significativement à la complexité de l'organisme associé. C'est le **paradoxe de la valeur C** (Thomas, 1971). Cette valeur n'est également pas corrélée au nombre de gènes codant pour des protéines (Pagel et Johnstone, 1992). La quantité d'**ADN non codant** au sein d'un génome peut expliquer une grande partie de ce paradoxe. (Orgel et Crick, 1980 ; Pagel et Johnstone, 1992). Dans cet ADN non codant figure l'ADN répétitif et en particulier les ET. La **taille du génome** est corrélée positivement à la proportion de séquences répétitives au sein du génome et est également **corrélée positivement avec la proportion d'ET** au sein du génome (fig 29). Ainsi les grands génomes contiennent proportionnellement plus d'ET que les petits (Elliott et Gregory, 2015).

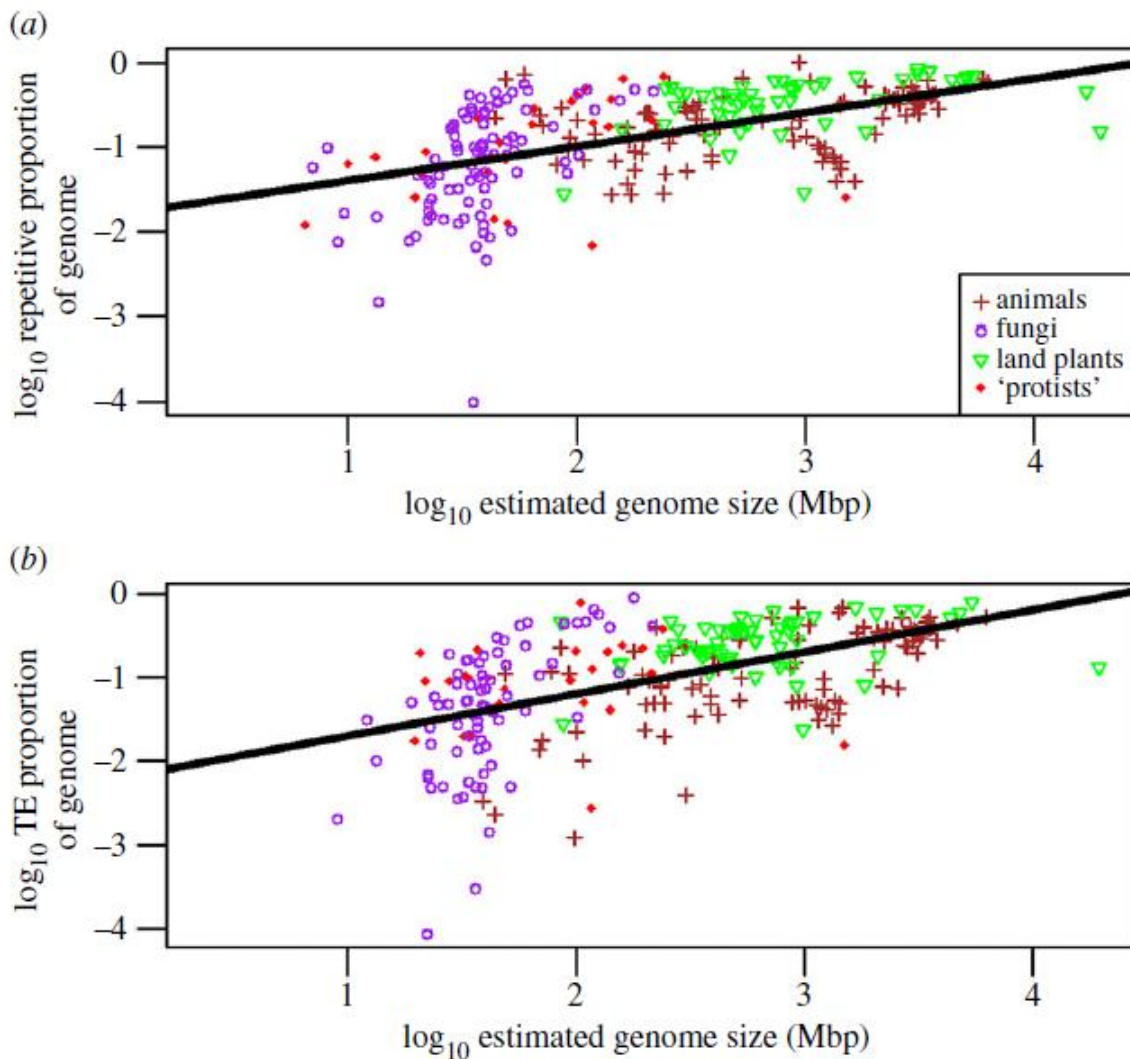


Figure 29 : Relation entre la taille du génome et la proportion de séquences répétées au sein du génome (a) et relation entre la taille du génome et la proportion d'ET (b) chez 502 espèces (148 espèces animales, 81 plantes, 202 champignons et 70 protistes) (Elliott TA et Gregory TR, 2015).

2) Richesse relative en familles d'ET et activité

La taille des génomes est donc corrélée à leur proportion en ET. Cependant, la contribution relative des différentes familles d'ET à la taille d'un génome varie en fonction de l'espèce considérée. Dans certains génomes, les transposons à ADN sont prédominants (par exemple chez le poisson-zèbre *D. rerio*), alors que chez d'autres ce sont les rétrotransposons qui sont prédominants (par exemple chez les mammifères) (fig 30) (Chalopin et al., 2015).

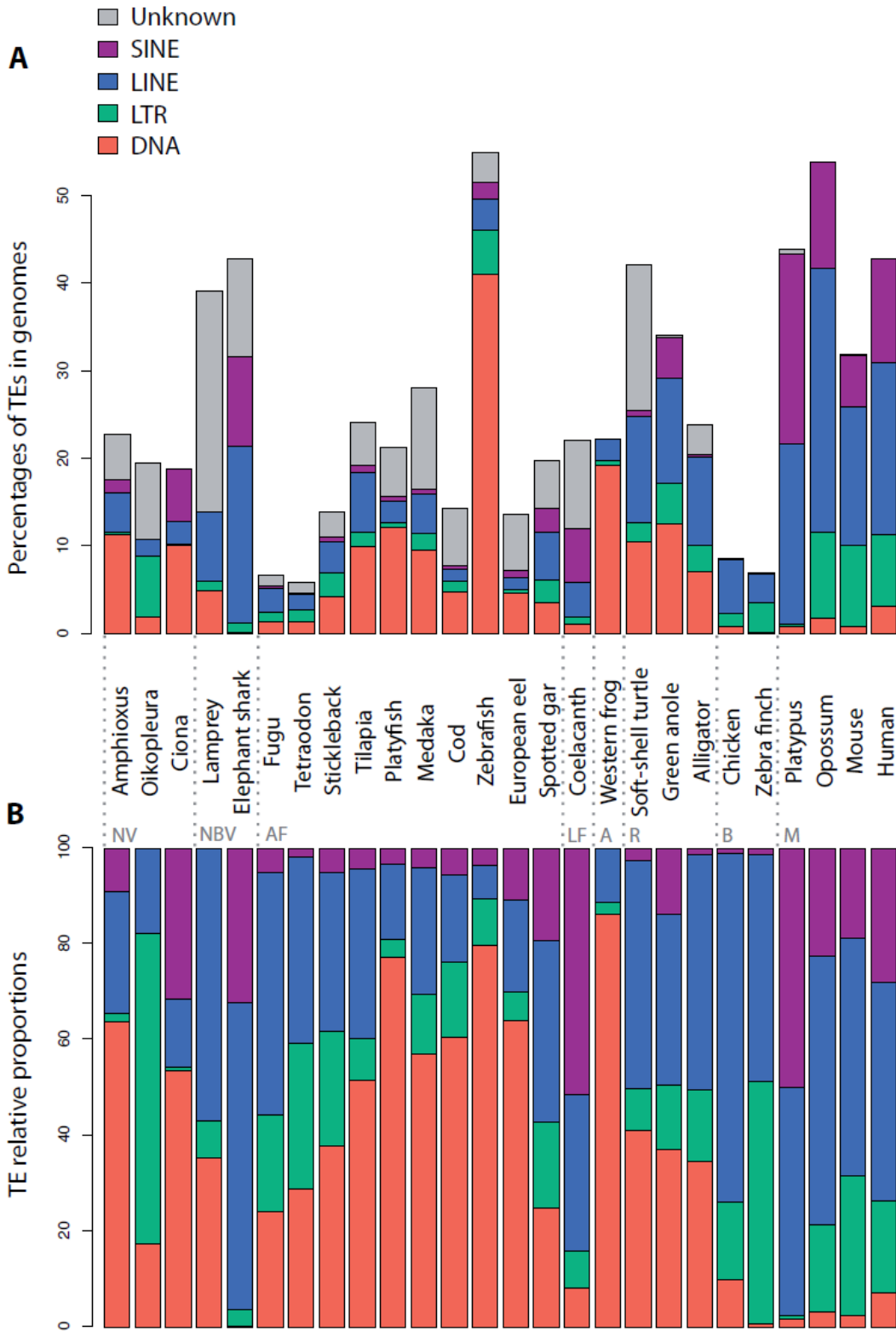


Figure 30 : Part relative en ET de différentes familles des génomes de différentes espèces (Chalopin et al., 2015).

L'activité des différents ET varie également d'un génome à l'autre. Certains génomes contiennent de nombreux ET qui sont pour la plupart inactifs (par exemple l'Homme) alors que d'autres ne contiennent qu'un faible nombre d'ET mais qui restent encore actifs (par exemple le poisson globe) (Elliott et Gregory, 2015).

3) Distribution non aléatoire des ET au sein des génomes

Une grande diversité peut être observée concernant les modèles de distribution des ET dans les génomes. Deux grands types de mécanismes peuvent être à l'origine de ces distributions : une préférence de site cible d'intégration et la sélection naturelle. Il est difficile de distinguer ces deux mécanismes en pratique. Les ET peuvent par exemple être contre-sélectionnés lorsqu'ils s'insèrent dans les régions codantes des gènes ou lorsqu'ils entraînent des recombinaisons. La théorie prévoit ainsi que les ET doivent être retrouvés plus fréquemment dans les régions les moins sujettes à la recombinaison (hétérochromatine des centromères et télomères). Il a été suggéré que la préférence de site cible d'intégration avait évolué car elle favorisait à la fois la propagation de l'ET et l'optimisation de la relation entre l'ET et l'hôte. Différents ET peuvent ainsi occuper différentes niches écologiques au sein du génome hôte (Kidwell et Lisch, 2002). Certains sont essentiellement présents dans des régions pauvres en gènes telles que les télomères ou l'hétérochromatine, alors que d'autres le sont dans des régions riches en gènes. Enfin, d'autres ET ne présente aucun modèle de distribution non aléatoire au sein du génome.

Les rétrotransposons sans LTR **HeT-A**, **TART** et **TAHRE** transposent spécifiquement à l'extrémité des chromosomes des Drosophiles (**télomères**). Ils maintiennent ainsi l'intégrité des télomères en formant de longs réseaux qui prolongent le chromosome et compensent la perte terminale. Ils remplacent ainsi les télomérases qui sont absentes chez les Drosophiles (Pardue et DeBaryshe, 2003 ; George et al., 2010).

Chez *Drosophila melanogaster* les ET sont principalement regroupés au niveau de l'hétérochromatine (segments de chromatine qui restent condensés durant l'interphase, contrairement à l'euchromatine qui se déroule durant l'interphase) (Babu et Verma, 1987 ; Pimpinelli et al., 1995). Cette accumulation pourrait être due à la sélection contre les effets délétères de l'instabilité créée par leur insertion dans l'euchromatine, région plus riche en gène. Cependant elle semble plutôt refléter une interaction évolutive dans laquelle les ET pourraient réparer les nombreuses cassures de l'hétérochromatine de la même façon que certains ET maintiennent la stabilité des télomérases (Dimitri et Junakovic, 1999). Le rétrotransposon à LTR **Ty5** de la levure *Saccharomyces cerevisiae* cible l'**hétérochromatine** en se liant directement, au niveau d'un « targeting domain » situé proche de l'extrémité C de l'INT, à un composant structural de l'hétérochromatine (Sir4) (Xie et al., 2001).

Les éléments **R2** (rétrotransposons sans LTR) des arthropodes s'insèrent tous spécifiquement dans le **gène de l'ARNr 28S**. Cette insertion est responsable d'un arrêt de la production de cet ARNr 28S ; l'hôte peut cependant survivre car son génome possède des centaines de copies de ce gène (Eickbush, 2002).

Contrairement à ces ET qui sont concentrés au niveau de régions spécifiques du génome, d'autres n'ont apparemment aucune spécificité au niveau de leur site d'insertion. Par exemple, les L1 ainsi que les éléments non autonomes qu'ils mobilisent sont dispersés dans tout le génome ce qui est probablement lié à la faible spécificité de site de leur EN (Levin et Moran, 2011).

B. Interactions entre ET et génome

Les ET sont de **puissants agents mutagènes**. Leur puissance égale voire surpasse celle des forts agents mutagènes chimiques ainsi que celle des radiations. Ils sont capables d'induire des changements dans le génome de leur hôte allant de petites modifications silencieuses détectables seulement au niveau de la séquence ADN jusqu'à des modifications ayant des répercussions sur les traits phénotypiques. Les mutations spontanées causées par les ET sont la source d'un **large éventail de variations génétiques** au sein de la population naturelle de quasiment toutes les espèces étudiées. Ces mutations peuvent être considérées comme les conséquences du **comportement égoïste** des ET (pouvant être à l'origine de maladies), mais peuvent également donner naissance à de nouvelles séquences régulatrices ou à de nouveaux gènes utiles à l'hôte et ainsi être la manifestation du **bénéfice évolutif** de la flexibilité génomique que confère les ET (Kidwell et Lisch, 2002).

Les ET sont responsables de mutations via leurs **insertions** au sein du génome hôte lors de leur (rétro)transposition mais également via les **réarrangements chromosomiques** dont ils sont à l'origine par recombinaison ectopique (Kidwell et Lisch, 2002). Les recombinaisons ectopiques sont des processus de réarrangements génomiques atypiques qui se produisent entre des segments d'ADN similaires dans les chromosomes mais non alléliques (Bannert et Kurth, 2006).

1) Insertions et réarrangements chromosomiques

Les insertions d'ET peuvent avoir lieu dans les **régions codantes d'un gène**, elles peuvent alors :

- **interrompre sa fonction** : en entraînant la production d'une protéine non fonctionnelle ou en empêchant la production d'une protéine à la place de la protéine normalement codée par le gène dans lequel s'est inséré l'ET
- être une **mutation silencieuse** : lorsque l'ET inséré dans un exon est exclu de l'ARN mature par l'épissage
- conduire à la formation d'un **nouveau gène** : lorsqu'un ET porteur de sites cryptiques donneurs et/ou accepteurs d'épissage conduit à un épissage différent (Kidwell et Lisch, 2002).

Les insertions d'ET peuvent également avoir lieu dans les **régions non codantes** d'un gène, elles peuvent alors :

- affecter la **régulation** du gène : lorsque l'ET porteurs de séquences régulatrices (par exemple des promoteurs ou amplificateurs) s'insère à proximité d'un gène, lorsqu'un ET s'insère dans la séquence régulatrice d'un gène ou bien lorsqu'une mutation survient au niveau d'un ET déjà fixé le rendant l'ET capable d'affecter la régulation d'un gène
- **évoluer avec le génome de l'hôte** et devenir co-optées pour l'usage de l'hôte.

Ces insertions dans les régions non codantes pourraient également entraîner des **changements à grande échelle** dans la séquence ainsi que la structure de l'ADN (via des réarrangements chromosomiques) affectant l'expression des gènes (Kidwell et Lisch, 2002).

Les ET peuvent également être à l'origine de variations dans le génome lors de leur **excision**. En effet, celle-ci peut aussi bien conduire à l'**addition** de nouvelles séquences qu'à la **délétion** de séquences de l'hôte autour du site d'insertion de l'ET. La **réparation** qui suivra l'excision d'un ET dépend du mécanisme de réparation de l'hôte ; elle peut être complète ou incomplète (Kidwell et Lisch, 2002).

Les ET ont également la capacité d'entraîner des **réarrangements chromosomiques** tels que des **délétions**, des **duplications**, des **inversions** et des **translocations réciproques** (Kidwell et Lisch, 2002). La **recombinaison homologue** entre différentes copies d'un ET peut être à l'origine de ces réarrangements. Un second mécanisme, nommé **transposition alternative**, peut également aboutir à des réarrangements chromosomiques liés aux ET de classe II. La transposition alternative met en jeu des ET ayant une structure similaire mais qui ne sont pas nécessairement identiques. Les extrémités des ET impliquées dans la transposition alternative viennent d'ET différents (appartenant au même chromosome ou à des chromosomes différents) à la différence de la transposition classique qui concerne les extrémités d'un même ET. Un élément hybride est ainsi formé, excisé et réintégré ailleurs dans le génome via le même mécanisme de couper-coller que lors de la transposition classique (Gray, 2000).

2) *Transposition, pseudogènes et brassage d'exons*

En plus de leur capacité à affecter l'expression d'un gène ou sa régulation, les ET peuvent également influencer la partie codante des génomes via la création de pseudogènes ainsi que via le brassage d'exons.

Les ET peuvent être responsables de l'insertion de gènes aboutissant à la formation de **pseudogènes**. Par exemple, chez l'Homme, la machinerie enzymatique des LINE peut également servir en *trans* et conduire à l'insertion d'un transcrit ADN qui n'est pas associé à une séquence LINE. Il en résulte la « rétroposition » du gène transcrit formant un nouveau gène qui possède les caractéristiques d'un pseudogène formé par transcription inverse à partir d'un ARN (« processed pseudogenes » en anglais) telles que la perte d'intron, de promoteur, l'acquisition d'une queue polyA et la présence de TSD de longueur variable (Esnault et al., 2000 ; Kidwell et Lisch, 2002).

Les ET peuvent également être responsables du **brassage d'exons** (« exon shuffling » en anglais) via l'insertion de séquences géniques dans d'autres gènes lors de leur transposition. Parfois, chez l'humain, lors de la transposition d'un LINE, la machinerie enzymatique outrepasser le signal de polyadénylation en utilisant un second signal de polyadénylation situé en aval. La transposition de ce LINE s'accompagne alors du transfert de la séquence située en 3'. Cette séquence déplacée avec le LINE vers une nouvelle localisation génomique pourra entraîner la production d'une protéine chimère. Dans le génome humain, environ 10% des 129 éléments L1 complets et actifs possèdent une extrémité 3' supposée avoir été déplacée avec le L1 (Esnault et al., 2000 ; Kidwell et Lisch, 2002).

3) Exemples de maladies humaines liées aux ET

Les ET sont une **source majeure d'instabilité génomique**. Ils peuvent être responsables de délétions, d'insertions ou de recombinaisons à l'intérieur d'un génome. Les maladies humaines associées aux ET sont particulièrement documentées dans le cas du rétrotransposon Alu (Kim et al., 2016).

Cependant, les ET peuvent au contraire être associés à la réparation de l'ADN dans le cas de cassures double-brin et ainsi servir au **maintien de la stabilité génomique** (Srikanta et al., 2009).

* Maladies génétiques associées à l'élément Alu

L'élément Alu représente environ 10% du génome humain et 0,1% des désordres génétiques humains lui sont associés. Cet élément est apparu chez les primates il y a 65 millions d'années et s'est propagé dans leur génome avec beaucoup de succès. Il appartient aux SINEs et transpose en utilisant les enzymes de L1 (EN et RT). Il est estimé à 1,1 million de copies dans le génome humain et est toujours actif. Il mesure 300pb et a une structure homodimérique dérivée de l'ARN 7SL (fig 31). Il possède un promoteur à l'ARN pol III en région 5' et une queue polyA en région 3'. Lors de son insertion (mécanisme TPRT), il forme des TSD de 7 à 20pb. Il peut affecter l'expression des gènes via différents mécanismes : insertion à l'intérieur d'un exon, insertion causant un épissage alternatif d'un gène, délétion associée à son insertion, délétion associée à une recombinaison, duplication associée à une recombinaison homologue et conversion associée à une recombinaison homologue (Kim et al., 2016). Lors d'une conversion, une portion de séquence ADN est copiée sur l'emplacement d'une séquence hautement semblable (Hsu et al., 2009).

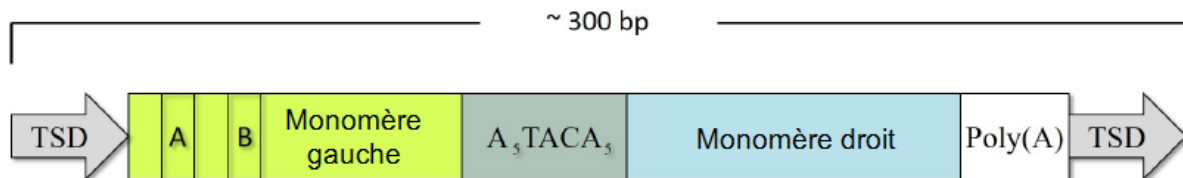


Figure 31 : Structure schématique d'un élément Alu (d'après Kim et al., 2016).

Les boîtes A et B du monomère gauche contiennent les sites de liaison à l'ARN Pol III.

L'**insertion** d'un élément Alu peut être à l'origine d'un **saut d'exon** (« exon skipping ») lorsqu'il s'insère à l'intérieur de celui-ci, d'un **signal de polyadénylation** entraînant la terminaison prématurée de la transcription lorsqu'il s'insère dans une région génique, ainsi que de l'apport d'un **site cryptique d'épissage** conduisant à un transcrit altéré (fig 32) (Kim et al., 2016). Un site cryptique d'épissage est un site d'épissage qui est normalement inactif ou utilisé à un très faible niveau mais qui devient actif lors de mutations du site d'épissage normal (Kapustin et al., 2011). Les éléments Alu sont porteurs de nombreux sites donneurs ou accepteurs d'épissage (Kim et al., 2016).

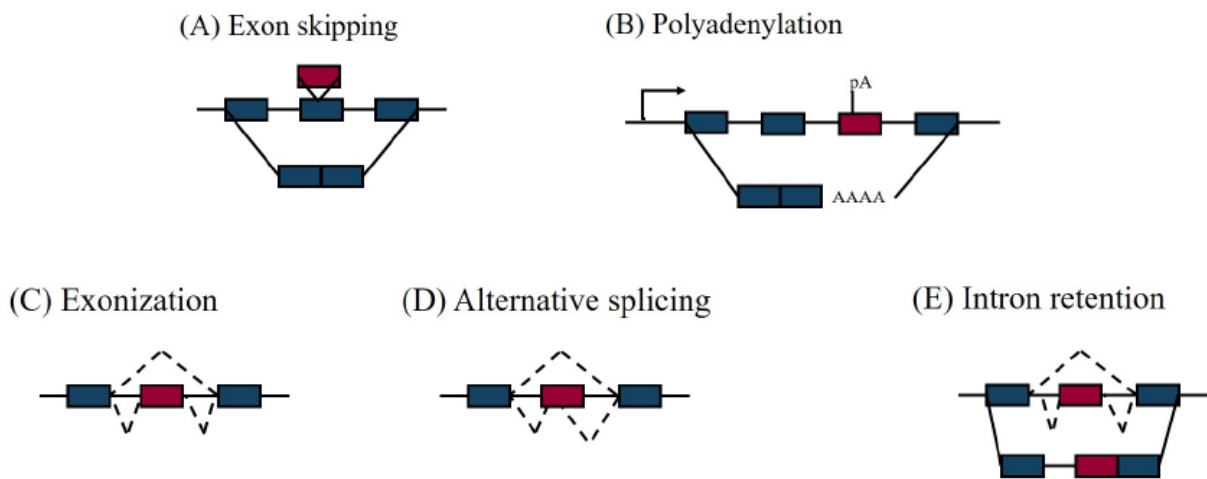


Figure 32 : Impact de l'insertion d'un élément Alu dans une région génique sur la transcription (Kim et al., 2016).

(A) Insertion d'un Alu dans un exon responsable d'un saut d'exon (« exon skipping »), (B) Insertion d'un Alu dans une région génique fournissant un signal de polyadénylation, (C-E) Insertion d'un élément Alu fournissant un site cryptique d'épissage alternatif aboutissant à des transcrits altérés. Les cases bleues représentent les exons, les cases rouges représentent les éléments Alu. La flèche représente le promoteur. Les lignes en tiret indiquent une forme d'épissage alternative.

Différents exemples de maladies humaines liées à l'élément Alu peuvent illustrer l'impact de l'insertion d'éléments Alu dans les régions géniques.

Par exemple, l'**hémophilie A** (un trouble de la coagulation lié au chromosome X causée par un déficit en facteur VIII) peut être liée à l'insertion d'un élément Alu dans un intron du gène codant pour le facteur VIII. Cette insertion est alors à l'origine d'un **saut d'exon** entraînant l'élimination de 39 acides aminés critiques du facteur VIII (Ganguly et al., 2003). L'**hypercalcémie hypocalciurique familiale** (une maladie génétique de transmission autosomique dominante caractérisée par une hypercalcémie modérée, une hypocalciurie relative et des taux de parathormone normaux) et l'**hyperparathyroïdie sévère néonatale** (une maladie caractérisée par une hypercalcémie, une augmentation du taux sérique de parathormone, une déminéralisation squelettique ainsi qu'une hyperplasie cellulaire parathyroïdienne) peuvent être causées par l'insertion d'un élément Alu, porteur d'un site de **polyadénylation**, dans le gène codant pour le récepteur extracellulaire sensible au calcium (CaSR). Il en résulte la production d'une protéine tronquée, affectant ainsi le fonctionnement du CaSR (Bai et al., 1997). L'**atrophie gyrée chorio-rétinienne** (une dégénérescence héréditaire de la rétine et de la choroïde associée à une hyperornithinémie, dont la transmission est autosomique récessive) peut être liée à l'insertion d'un élément Alu dans l'intron 3 du gène de l'ornithine aminotransférase (OAT : enzyme mitochondriale qui catalyse la transamination de l'ornithine en glutamate). Lorsque cet élément Alu subit une transversion cytosine → guanine, il génère un site donneur et active un **site cryptique d'épissage** en amont aboutissant à un épissage erroné responsable de la production d'une protéine anormale (Mitchell et al., 1991 ; Valle et al., 1977).

En plus des mutations induites par leur insertion, les éléments Alu peuvent également être source de **recombinaisons** de par leur abondance ainsi que leur haut niveau

de similitude au sein du génome humain. Ces recombinaisons peuvent entraîner des délétions mais peuvent également réparer des cassures doubles brins dans l'ADN (Kim et al., 2016).

Un exemple de maladie liée à ces recombinaisons est l'**hypertension artérielle pulmonaire**. Des délétions exoniques du gène du récepteur II de la protéine morphogénique osseuse (BMP2) peuvent avoir lieu par **recombinaison homologue non allélique** (délétion des exons 1 à 3 du gène encodant le BMP2) ou par **recombinaison non homologue** (délétion de l'exon 10). Ces mutations ont été mises en évidence chez de nombreux patients atteints d'hypertension artérielle pulmonaire familiale et d'hypertension artérielle pulmonaire idiopathique (Kataoka et al., 2013).

** ET et cancer*

Les recombinaisons liées à des éléments Alu peuvent également avoir lieu dans les cellules somatiques et ainsi être source de cancers comme dans le cas de la **leucémie myéloïde aiguë (LMA)**. Une **recombinaison homologue** entre les séquences **Alu** présents dans les introns 1 et 6 du gène *ALL1* entraîne une duplication partielle en tandem de ce gène responsable de la LMA (Strout et al., 1998).

La méthylation du promoteur de l'ARN pol III permet d'inhiber l'expression des LINE-1. Dans de nombreuses tumeurs malignes humaines, le génome est hypométhylé. L'hypométhylation du promoteur des LINE-1 est corrélée à sa rétrotransposition et est associée à une instabilité génomique ainsi qu'à un mauvais pronostic dans de nombreux cancers (le cancer du poumon non à petites cellules (CPNPC), le cancer du côlon, le carcinome épidermoïde de l'œsophage, le cancer du sein, le carcinome hépatocellulaire et le cancer de l'ovaire). Par exemple, dans le cancer colorectal, **l'insertion d'un élément LINE-1 dans un exon** du gène *APC* (« adenomatous polyposis coli ») compromet fondamentalement son activité anticancéreuse dans le côlon et cause la **polypose adénomateuse familiale (PAF)**. Des mutations supplémentaires peuvent alors entraîner des tumeurs colorectales sporadiques. Ainsi, l'insertion d'éléments LINE-1 peut être **à l'origine ou bien être un facteur favorisant le développement des cancers** lorsqu'elle a lieu dans un codon essentiel d'un gène crucial suppresseur de tumeur (Burns, 2017).

Les ET peuvent également opérer dans la **prolifération de cellules cancéreuses** à travers leur **domestication**. C'est le cas de **SETMAR** aussi appelé **Metnase**, dont le domaine transposase de son gène est dérivé de la transposase d'un élément Mariner. En plus de promouvoir la liaison non homologue des cassures doubles brin de l'ADN, Metnase interagit avec la topoisomérase II α (Topo II α) et améliore son activité de décaténation des chromosomes. Lorsque les cellules sont atteintes de leucémie myéloïde aiguë (LMA), elles ne parviennent pas à stopper leur mitose lorsque la Topo II α est inhibée. Metnase permet aux cellules atteintes de LMA de résister aux effets cytotoxiques des inhibiteurs de la Topo II α (tels que VP-16 qui est largement utilisé dans le traitement du cancer) et de continuer à proliférer (Wray et al., 2009).

4) Régulation de l'activité des ET

Différents mécanismes régulant l'activité des ET existent, tels que l'interférence par ARN et la méthylation de l'ADN.

Le mécanisme de l'**interférence par ARN** (RNAi) permet de réguler l'expression post-transcriptionnelle des ET en les dégradant (fig 33). L'ARN double brin (dsRNA) est clivé par une protéine de la famille dicer afin de former de petits ARN interférents (siRNA) qui sont ensuite incorporés dans un complexe appelé RISC contenant des protéines de la famille Argonaute. L'ensemble clive alors les transcrits dont la séquence est complémentaire à l'ARN interférent. L'ARN double brin peut être formé par le chevauchement antiparallèle de transcrits ou par ARN polymérase dépendant de l'ARN (RdRP) à partir des ARN clivés. Le mécanisme grâce auquel la RNAi cible les ET n'est pas élucidé ; des caractéristiques structurales telles que des épingles à cheveux formées par le repliement des TIR des ET pourraient permettre de les distinguer des transcrits de l'hôte (Slotkin et Martienssen, 2007).

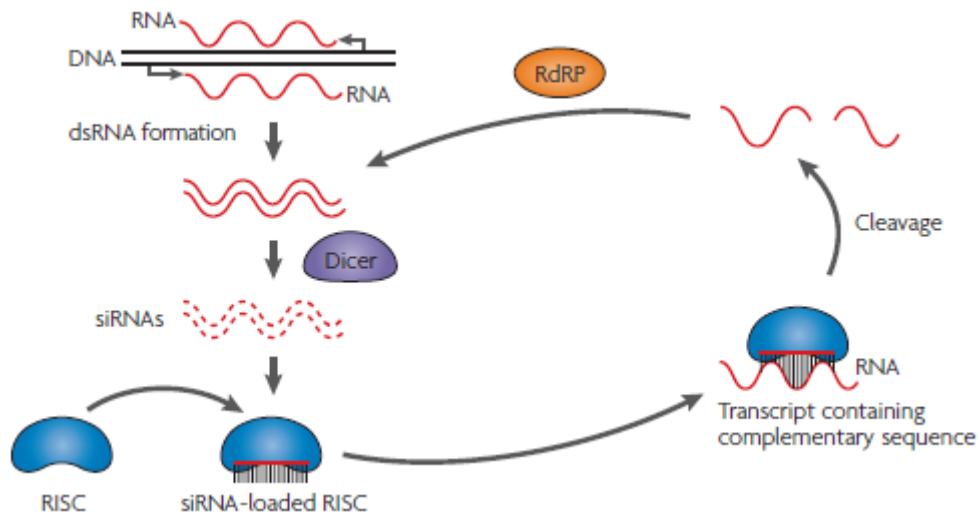


Figure 33 : Mécanisme d'interférence par ARN (Slotkin et Martienssen, 2007).

Des modifications de la chromatine aboutissant à sa condensation (**hétérochromatine**) et donc la rendant **inaccessible à la transcription** peuvent également être **basées sur le mécanisme de RNAi** (fig 34). Les petits ARN interférents peuvent entrer dans la constitution d'un complexe différent contenant une autre protéine de la famille Argonaute (appelée RITS chez *Schizosaccharomyces pombe*). Ce complexe entraîne le clivage des jeunes transcrits qui sont encore liés à l'ARN polymérase II ainsi qu'à l'ADN et cible cette région de la chromatine pour la modifier (méthylation des histones H3K9, et éventuellement méthylation des cytosines dans les organismes dont le génome est méthylé) grâce au recrutement de protéines comprenant une méthyltransférase H3 lysine 9 (H3K9) (Slotkin et Martienssen, 2007).

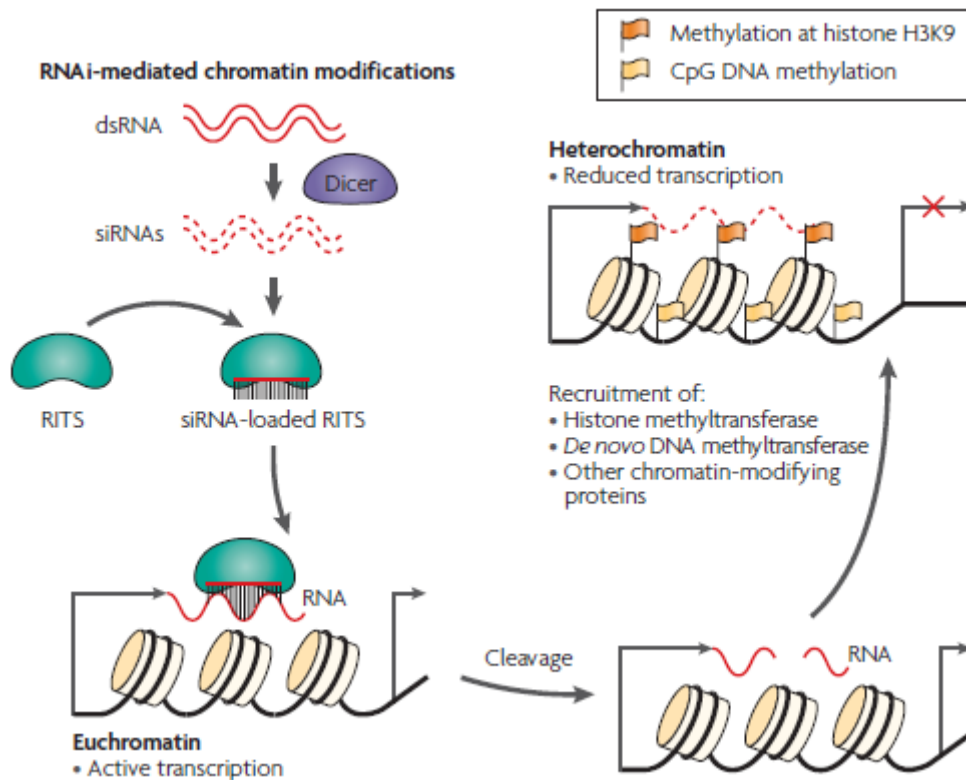


Figure 34 : Mécanisme de modifications de la chromatine basé sur l'interférence par ARN (Slotkin et Martienssen, 2007).

La **méthylation des cytosines** joue également un rôle essentiel dans la défense de l'organisme contre les transposons. Cette méthylation est réalisée par des enzymes appelées ADN méthyltransférases, au niveau du dinucléotide CpG, chez les plantes et les mammifères. Elle permet de réprimer la transcription des ET. Elle est reproduite sur le nouveau brin d'ADN lors de sa réplication et est héritable. On ne connaît pas le mécanisme ciblant les ET dans la méthylation *de novo*, elle se déroule dans les cellules germinales et dans le développement embryonnaire précoce. Le mécanisme de méthylation de l'ADN permet de contrôler l'expression des gènes en entravant, voire en empêchant la plupart des facteurs de transcription des mammifères (nécessitant des sites de liaison riches en GC, voire possédant des CpG dans leurs éléments de reconnaissance de l'ADN) de se lier à l'ADN, réprimant ainsi l'expression de la protéine associée (Bird et Wolffe, 1999 ; Bestor et Bourc'his, 2004 ; Slotkin et Martienssen, 2007).

Cependant, lorsque les cellules sont soumises à un **stress** (changement environnemental qui réduit drastiquement la fitness d'un organisme), l'**activité** de certains ET **augmente**. Il est possible qu'elle soit liée à la libération d'activateurs transcriptionnels nécessaires aux gènes de défense de l'hôte, qui activeraient également les ET grâce à la présence de promoteurs ou de sites de fixation pour ces activateurs dans les régions non traduites des ET. Ce phénomène augmente la variabilité génétique liée aux ET, cette variabilité pouvant s'avérer utile dans des conditions de stress. L'origine de ces sites de fixation n'est pas élucidée, ils pourraient avoir été acquis par les ET à partir de séquences de l'hôte ou, à l'inverse, par les gènes de défense de l'hôte *via* la domestication de séquences d'ET (comme dans le cas du système immunitaire des vertébrés et de RAG1, cf paragraphe suivant) (Capy et al., 2000).

5) Domestication des ET

Certains gènes codant pour des protéines chez les eucaryotes sont en fait dérivés d'ET. On appelle cet événement la **domestication moléculaire**. Elle correspond au passage pour un élément transposable du statut de parasite génomique au statut de gène stationnaire bénéfique à son hôte (Miller et al., 1992). On en trouve chez les plantes, les champignons et les animaux. Certains de ces gènes sont impliqués dans des fonctions essentielles pour leur hôte (Volff, 2006).

Par exemple, la syncytine, une protéine codée par le gène d'enveloppe d'un rétrovirus endogène humain défectueux (HERV-W), est impliquée dans la **morphogenèse placentaire** humaine. Ce gène viral a été domestiqué et a maintenant un rôle physiologique important (Mi et al., 2000).

Un second exemple illustrant le rôle essentiel que peut avoir un gène domestiqué est le **système de recombinaison V(D)J** du système immunitaire des Gnathostomes (vertébrés à mâchoire). Ce système permet de reconnaître un nombre illimité d'antigènes grâce à leur liaison à des récepteurs spécifiques présents sur les cellules B et T. Comme ces récepteurs ne peuvent pas tous être directement codés dans le génome, un système de recombinaison entre trois portions différentes (V pour « variable », D pour « diversity » et J pour « joining »), présentant chacune un peu moins d'une centaine de copies, permet de coder la partie variable des récepteurs. Ce système, nommé « recombinaison V(D)J », joint un segment V, un segment D et un segment J en un seul exon. En plus de ce mécanisme, de petites insertions et délétions sont générées à la jonction des différents segments, augmentant encore la diversité des récepteurs produits. Deux protéines (RAG1 et RAG2) sont responsables de la coupure de l'ADN dans la recombinaison V(D)J. RAG1 est dérivée de la transposase d'un transposon à ADN de la superfamille Transib. De plus, les RSS (pour « recombination signal sequence ») permettant au complexe RAG1/2 de se lier à l'ADN afin de le couper entre les segments V, D et J, semblent dérivés des TIR d'un Transib (Kapitonov et Jurka, 2005).

Les ET ont donc un rôle évolutif dans la formation de nouvelles fonctions cellulaires ainsi que dans leur régulation. Les gènes domestiqués à partir d'ET peuvent être issus de différents gènes des ET : les gènes codant pour les transposase, les INT, les RT, ainsi que les protéines Gag et Env (Volff, 2006).

6) ET impliqués dans le maintien de la structure des génomes

Les ET peuvent également être impliqués dans le maintien de la structure des génomes, comme par exemple dans le cas des **télomères** cités précédemment (rétrotransposons sans LTR **HeT-A**, **TART** et **TAHRE**).

Les LINE-1 semblent pouvoir s'intégrer dans les lésions de l'ADN et les réparer lors de leur rétrotransposition (le mécanisme utilisé est alors différent de la TPRT qui est « endonuclease-dépendant », c'est l'« endonuclease-independent retrotransposition ») dans les cellules de mammifères (Morrish et al., 2002).

Un mécanisme analogue existe pour les éléments Alu, il a été nommé NCAI (« non-classical Alu insertion ») et a été mis en évidence dans les génomes de l'humain, du chimpanzé, du gorille et de l'orang-outan. Ce mécanisme pourrait jouer un rôle dans la réparation des cassures double-brins de l'ADN (Srikanta et al., 2009).

Ainsi, l'insertion des Alu et L1 est généralement réalisée grâce au mécanisme TPRT ; elle est alors associée à une perturbation de la fonction des gènes et est sujette à des recombinaisons ectopiques. Alors que la rétrotransposition non classique de ces éléments semble participer à la réparation de l'ADN. Les familles **L1** et **Alu** contribuent ainsi occasionnellement au **maintien de la stabilité du génome** (Srikanta et al., 2009).

C. ET et évolution des virus

1) Rétrotransposons et rétrovirus

Les **rétrotransposons** et les **rétrovirus** sont très **semblables** au niveau de leur **cycle de vie intercellulaire** ainsi que pour les **protéines majeures** qu'ils codent. Cependant, la présence d'un gène d'enveloppe chez les rétrovirus des vertébrés est ce qui les différencie principalement des rétrotransposons. Ce gène code pour une protéine transmembranaire et pour une protéine de liaison au récepteur de l'hôte. Ensemble, elles sont responsables du caractère infectieux et de la transmission des virus. Leur fonction antigénique entraîne la réponse immunitaire de l'hôte, elles sont ainsi soumises à une forte pression sélective. Leur divergence rapide et leur acquisition ancienne ne permettent pas de savoir si l'acquisition des gènes Env est issue d'un événement unique ou d'événements multiples (Malik et al., 2000).

L'analyse phylogénétique des séquences des RT suggère que l'ancêtre le plus probable des rétrovirus serait un ET possédant les gènes « gag-like » et « pol-like ». Après acquisition de LTR, cet élément aurait donné lieu à des groupes distincts de rétrotransposons à LTR ainsi qu'à trois groupes de virus (les rétrovirus, les hépadnavirus et les caulimovirus) (Xiong et Eickbush, 1990).

2) Polintons et virus

Les **Polintons** sont largement répandus chez divers protistes et animaux, témoignant de leur **origine ancienne** ainsi que de leur succès évolutif. Ils codent pour deux protéines homologues aux protéines de capsides virales. Ainsi, ils **possèdent à la fois des caractéristiques d'ET et des caractéristiques de virus**. Il semble que les Polintons aient évolué directement à partir de bactériophages (Tectiviridae) et soient devenus la source de l'évolution de divers groupes de virus (Adenoviridae, Megavirales, Bidnaviruses) retrouvés chez les eucaryotes (Krupovic et Koonin, 2015).

Après avoir présenté la diversité des ET ainsi que leur intérêt évolutif majeur, la partie suivante se consacre à la présentation de la chauve-souris *P. giganteus* dont le génome a récemment été séquencé, à son rôle de réservoir de virus zoonotiques et à l'étude d'un ET de son génome.

PARTIE II. ETUDE DES ET DU GENOME DE *P. GIGANTEUS*

I. Présentation de l'espèce *P. giganteus*

A. Description, taxonomie et distribution géographique

1) Description

P. giganteus (fig 35) est une chauve-souris possédant de longues oreilles noires ainsi qu'un long museau avec des narines bien développées. Son pelage est de couleur marron et est plus sombre autour des yeux. Elle pèse de 400 à 600 grammes, son corps mesure de 15,5 à 22 centimètres, et elle mesure 0,93 à 1,04 mètres d'envergure (Sivaperuman et Venkataraman, 2018).



Figure 35 : Photo de *P. giganteus* (photographe Tonge Simon J., le 18 octobre 2017 au Parc zoologique Tierpark Berlin, Allemagne).

Les *P. giganteus* vivent en **grandes colonies** regroupant des centaines voire des milliers d'individus. Les colonies ont habituellement un **perchoir** permanent et un ou deux perchoirs temporaires en fonction de la saison. Ces perchoirs sont constitués de grands arbres (tels que le banian, le manguier, le tamarin), aussi bien en zone rurale qu'en zone urbaine (près des champs agricoles et au bord des routes) (Molur et al., 2008 ; Chaturvedi et Singh, 2018).

P. giganteus est **frugivore** et **nectarivore**. Elle participe ainsi à la **pollinisation** et à la **dispersion des graines** (Chaturvedi et Singh, 2018). Elle peut parcourir jusqu'à 150 km la nuit à la recherche de nourriture (Molur et al., 2008).

La population de *P. giganteus* est localement menacée à cause de l'abattage des arbres qui lui servent de perchoir et à cause de la chasse. Elle est chassée pour sa viande (consommée comme « viande de brousse ») ainsi que pour sa graisse corporelle qui est utilisée en potions et comme remède contre les rhumatismes par les médecins locaux. Elle est également considérée comme « vermine » et est détruite par ceux qui la soupçonnent de provoquer des dommages aux cultures fruitières (Molur et al., 2008 ; Mickleburgh et al., 2009 ; Mahmood-UI-Hassan et al., 2010).

2) Taxonomie

Les chauves-souris sont des mammifères appartenant au superordre des Laurasiatheria et au taxon Perissodactyla, qui comprend également les chevaux. Elles sont divisées en deux groupes : le sous-ordre Megachiroptera, comprenant des individus essentiellement frugivores, et le sous-ordre Microchiroptera, comprenant de petits individus insectivores. *P. giganteus* appartient au sous-ordre Megachiroptera et à la famille Pteropodidae (Mahmood-UI-Hassan et al., 2010 ; Hayward et al., 2013).

La **chauve-souris géante d'Inde** (« Indian flying fox » en anglais) a été décrite en 1782 par Brünnich qui lui avait alors donné le nom de *Vespertilio gigantea*. Selon certains auteurs, cette espèce avait déjà été décrite par Linnaeus en 1758 sous le nom de *Vespertilio vampyrus*, et le plus ancien nom disponible pour cette espèce, donc le plus approprié, serait *Pteropus medius*, utilisé par Temminck en 1825. Certains auteurs ne faisant pas la différence entre cette espèce et le Renard volant de Madagascar l'ont appelée *Pteropus edwardsi*. Aujourd'hui, cette espèce est nommée *P. giganteus* par la plupart des auteurs. Trois sous-espèces sont reconnues : *P. giganteus ariel* (utilisé par Allen en 1908), *P. giganteus giganteus* (utilisé par Brünnich en 1782) et *P. giganteus leucocephalus* (utilisé par Hodgson en 1835). D'autres synonymes de *P. giganteus* peuvent également être utilisés : *P. assamensis* (utilisé par McClelland en 1839), *P. kelaarti* (utilisé par Gray en 1871) et *P. ruvicollis* (utilisé par Ogilby en 1840) (Molur et al., 2008 ; Mlíkovský, 2012 ; Integrated Taxonomic Information System).

3) Distribution géographique

P. giganteus est présente en **Asie du Sud** (Bangladesh, Bhoutan, Inde, Maldives, Népal, Pakistan, Sri Lanka) en **Asie du Sud-Est** (Myanmar) et en **Chine** (Qinghai) (fig 36) (Molur et al., 2008).



Figure 36 : Distribution géographique de *P. giganteus* (d'après Molur et al., 2008).

B. Aptitude au vol et immunité

La caractéristique principale différenciant les chauves-souris des autres mammifères est leur **aptitude au vol**. Le vol prolongé entraîne une augmentation du métabolisme oxydatif et donc de la production des sous-produits de ce métabolisme (dérivés réactifs de l'oxygène) responsables de dommages au niveau de l'ADN. De nombreux gènes impliqués dans les voies de contrôle des dommages faits à l'ADN ont été sectionnés de manière positive chez ces mammifères. Il a ainsi été suggéré que l'acquisition du vol au cours de l'évolution a été accompagnée par une **sélection positive des mécanismes limitant les dommages collatéraux causés à l'ADN** (Zhang et al., 2013).

Le gène c-REL fait partie des gènes sélectionnés positivement chez les chauves-souris. Ce gène joue un rôle dans la réponse aux dommages faits à l'ADN mais également dans l'immunité innée et adaptative. Ainsi, il est possible que la sélection des mécanismes de contrôle des dommages faits à l'ADN ait influencé également le **système immunitaire** des chauves-souris (Zhang et al., 2013).

C. *P. giganteus* et zoonoses virales

Les chauves-souris présentent de nombreuses caractéristiques leur permettant d'être un **réservoir et un vecteur idéal pour les agents pathogènes zoonotiques émergents**. Elles sont capables de voler et peuvent ainsi disséminer des virus sur de grandes distances, elles ont une espérance de vie exceptionnellement longue pour leur taille (25 à 35 ans pour certaines) et elles vivent dans de grandes colonies comportant beaucoup de jeunes individus naïfs permettant aux virus de persister. Le risque zoonotique est également augmenté par les activités humaines qui augmentent les interactions entre les chauves-souris, les humains et le bétail (Wang et al., 2011 ; Wynne et Wang, 2013).

1) Zoonoses à virus Nipah et Hendra

P. giganteus est porteuse de zoonoses virales liées aux henipavirus émergents, Nipah et Hendra.

a) Virus Nipah

Le virus Nipah (NiV) a émergé pour la première fois en **1998 en Malaisie**. Les **porcs** infectés présentaient des **symptômes neurologiques et respiratoires**. La plupart des **cas humains** était constituée d'hommes travaillant en **contact avec des porcs**. Ils présentaient de la fièvre, des maux de tête suivis de somnolence et de désorientation pouvant évoluer vers le coma, certains présentaient également des symptômes respiratoires. Des cas similaires ont également été observés à Singapour en 1999 (Centers for Disease Control and Prevention, 1999). Le **réservoir naturel présumé** du NiV était représenté par les chauves-souris frugivores, en particulier du **genre *Pteropus***, et les hommes se contaminaient via des hôtes intermédiaires tels que les porcs (Harcourt et al., 2005). Par la suite, le NiV a émergé au **Bangladesh en 2001** (Hahn et al., 2014) et en **Inde** (Deka et Morshed, 2018). Aucun hôte intermédiaire n'ayant été identifié lors de l'émergence au Bangladesh, une **transmission directe de la chauve-souris à l'homme** a été supposée. Une **transmission interhumaine** a également été mise en évidence (Harcourt et al., 2005).

Le réservoir du NiV au Bangladesh est *P. giganteus*. La densité de population humaine ainsi que les caractéristiques du paysage sont corrélées au risque d'émergence du NiV (Hahn et al., 2014).

Les principaux risques de contamination en Inde et au Bangladesh résident dans la consommation de **sève de palmiers dattiers contaminés** par l'urine ou par la salive des *Pteropus*, dans le **contact avec des humains infectés** et dans la **consommation de viande sauvage** de chauve-souris (Deka et Morshed, 2018).

Le NiV appartient à la famille des *Paramyxoviridae*, à la sous-famille des *Paramyxovirinae* et est étroitement lié au virus Hendra. L'un comme l'autre ont la capacité d'infecter un certain nombre d'espèces (porcs, chats, chiens, chevaux, chauves-souris, humains) et d'être responsables d'une maladie potentiellement mortelle, ce qui est inhabituel au sein de cette famille (Chua et al., 2000).

En 2015, des chauves-souris *P. giganteus* porteuses du NiV (RT-PCR) ont été capturées dans l'est de l'état du Bengale au nord de l'Inde (districts de Dhubri, Assam et Cooch Behar). De plus, de grandes colonies de *P. giganteus* ont été observées à proximité des habitations humaines dans cette région. Ainsi, le risque de contamination de l'espèce humaine est préoccupant (Yadav et al., 2018).

L'encéphalite à NiV est une **maladie à déclaration obligatoire** auprès de l'OIE (OIE Code sanitaire pour les animaux terrestres, 2010) et en France un **danger sanitaire de première catégorie** (Legifrance, 2013).

b) Virus Hendra

Le virus Hendra (HeV) a été décrit pour la première fois en **Australie en septembre 1994**. Responsable d'une fièvre et d'une atteinte respiratoire aiguë mortelle chez les **chevaux** et les **humains**, il a d'abord été considéré comme un virus du genre *Morbillivirus*, famille des *Paramyxoviridae* (Murray et al., 1995).

En octobre 1995, deux chevaux décédés d'une encéphalite en **août 1994** sont rétrospectivement diagnostiqués atteints du HeV. Lors des enquêtes entreprises suite à ces émergences virales, aucun anticorps anti-HeV n'a pu être détecté chez les animaux investigués (plus de 5000 animaux comprenant des chevaux, des bovins, des chiens, des chats et des volailles) et aucun lien épidémiologique n'a été mis en évidence entre les deux foyers. L'infection HeV ne semble pas très contagieuse chez les chevaux. Elle nécessiterait un contact direct ou une transmission mécanique de fluides corporels infectieux (Field et al., 2001).

En **janvier 1999**, une troisième émergence du HeV est signalée en **Australie** chez un cheval atteint d'une pneumonie. La présence d'arbres où les roussettes nichent préférentiellement dans les enclos des chevaux atteints du HeV et le fait que ces cas soient déclarés lors de la saison des naissances des roussettes sont en faveur de l'hypothèse selon laquelle les fluides ou les membranes fœtales seraient une voie d'excrétion du virus chez les roussettes et que l'ingestion d'herbe ou d'aliments contaminés serait le mode de transmission du virus au cheval (Field et al., 2000).

L'isolement du HeV dans le liquide utérin ainsi que dans des poumons et des foies de fœtus de **chauves-souris de la famille Pteropodidae** corrobore les précédents arguments (anticorps capables de neutraliser le HeV et épidémiologie) en faveur du fait que ces chauves-souris soient le **réservoir naturel** de ce virus (Halpin et al., 2000).

Les virus **Nipah et Hendra** sont des membres étroitement apparentés de la famille des Paramyxoviridae. Leurs différences avec les autres membres de la famille ont conduit à les classer dans un **nouveau genre, Henipavirus**, dans la sous-famille Paramyxovirinae. Le virus Hendra et le virus Nipah sont des **agents de biosécurité de niveau 4** (OIE Terrestrial Manual, 2015).

Les chauves-souris de la **famille Pteropodidae** sont sensibles à l'infection par les *Henipavirus* mais celle-ci (expérimentalement et sur le terrain) est généralement non pathogène. Ces caractéristiques indiquent que ces chauves-souris sont des espèces hôtes et font office de **réservoir des Henipavirus** (Halpin et al., 2011).

c) Caractéristiques des Henipavirus

Les virus Hendra et Nipah sont classés dans la famille Paramyxoviridae, la sous-famille Paramyxovirinae et le genre *Henipavirus*. Ils sont pléomorphes (c'est-à-dire qu'ils peuvent revêtir des formes différentes) et les virons mesurent de 180 à 1 900nm de diamètre. Leur génome est constitué d'un ARN négatif monocaténaire non segmenté d'environ 18,2kb. Il possède les gènes codant pour six protéines structurales majeures : une protéine de nucléocapside (N), une phosphoprotéine (P), une protéine de matrice (M), une protéine de fusion (F), une glycoprotéine (G) et une grande polymérase (L) (Lo et Rota, 2012).

Le genre *Henipavirus* contient également des virus récemment identifiés comme le virus Cedar isolé dans l'urine des chauves-souris *Pteropus* en Australie. Sa capacité à se propager et à entraîner l'apparition de signes cliniques est encore inconnue. D'autres *Henipavirus* ont été détectés dans la faune par PCR puis séquençage mais n'ont pas pu être isolés (OIE Terrestrial Manual, 2015).

d) Synthèse comparative des virus Nipah et Hendra

[OIE Terrestrial Manual, 2015 ; OMS Virus Nipah, OMS Infection à virus Hendra, OIE Virus Nipah et AVA The Hendra vaccine]

Tableau IV: Tableau comparatif concernant les virus Nipah et Hendra (d'après les sites internet de l'Organisation Mondiale de la Santé (OMS), de l'Organisation mondiale de la santé animale (OIE) et de l'Australian Veterinary Association (AVA))

Virus	Nipah	Hendra
Taxonomie	Famille : <i>Paramyxoviridae</i> Sous-famille : <i>Paramyxovirinae</i> Genre : <i>Henipavirus</i>	
Statut	Maladie à déclaration obligatoire auprès de l'OIE	N'est pas une maladie à déclaration obligatoire auprès de l'OIE
Flambées épidémiques	Malaisie 1998 Singapour 1999 Bangladesh presque chaque année depuis 2001 Inde orientale périodiquement	Australie (Brisbane) 1994 : 21 chevaux et 2 personnes Côte Est de l'Australie : 11 flambées depuis juillet 2008
Létalité	40 à 75%	
Transmission	- Contact direct avec l'hôte amplificateur (porcs malades, tissus contaminés) (Malaisie) - A partir de l'hôte réservoir naturel : consommation de fruits ou de produits dérivés (jus brut de palmier-dattier) contaminés par urine / salive de chauves-souris infectées (Bangladesh) - Transmission interhumaine par contact direct (sécrétions ou excréctions de personnes infectées, nosocomiale) (Bangladesh)	- Contact direct avec l'hôte amplificateur (contact rapproché pendant les soins ou lors d'autopsie de chevaux)

Symptômes et signes cliniques chez l'humain	<p><u>Clinique</u> : asymptomatique / infection respiratoire aiguë / voire encéphalite mortelle.</p> <p><u>Symptômes</u> :</p> <ul style="list-style-type: none"> - Fièvre, céphalées, myalgies, vomissements, maux de gorge - Vertiges, somnolence, altération de l'état de conscience, signes neurologiques - Pneumonie atypique, problèmes respiratoires sévères, insuffisance respiratoire aiguë - Encéphalite, convulsions, coma 	<p><u>Clinique</u> : syndrome grippal bénin / atteinte respiratoire / voire atteinte neurologique mortelle</p>
Période d'incubation	Généralement 4 à 14 jours (parfois jusqu'à 45 jours)	
Séquelles et rechutes	<ul style="list-style-type: none"> - Guérison complète possible - 20% des patients ont des séquelles neurologiques (troubles convulsifs, altérations de la personnalité) - Quelques rechutes ou encéphalites d'apparition tardive 	
Identification du virus	<ul style="list-style-type: none"> - <u>Identification du virus</u> : après isolement par immunocoloration / neutralisation avec des antisérums spécifiques / caractérisation moléculaire (tissus frais d'organes infectés, de gorge, urine, sécrétions nasales) ou par RT-PCR - <u>Détection Ag</u> : immunohistochimie (IHC) (tissus fixés dans du formol) (à partir de liquides biologiques) - <u>Tests sérologiques</u> : ELISA (dépistage) et tests de neutralisation du virus (VNT : procédure de référence, test de confirmation) 	
Traitement	<ul style="list-style-type: none"> - Aucun traitement ni vaccin spécifique (recherche en cours sur le développement de vaccins en Australie et en France) - Soins intensifs de soutien 	<ul style="list-style-type: none"> - Aucun vaccin - Traitement symptomatique intensif
Réservoir naturel	<i>Pteropus spp</i> (pas de maladie apparente)	
Espèces atteintes	Malaisie en 1999 : porcs, chevaux, chèvres, moutons, chats et chiens	Chevaux (expérimentalement chats atteints mais chiens sans signes cliniques)
Infection de l'hôte amplificateur	Chez les porcs : très contagieux, faible mortalité (sauf porcelet), forte morbidité	Chez les chevaux : peu contagieux, taux de létalité de 75% environ
Symptômes chez l'hôte amplificateur	Porcs : fièvre, atteinte respiratoire et neurologique, toux aboyante, avortement, mort	Chevaux : forte fièvre, tuméfaction faciale, difficultés respiratoires sévères, écoulement nasal mousseux abondant, signes neurologiques (ataxie, myoclonies), coliques

Prévention	<p><u>Porc :</u></p> <ul style="list-style-type: none"> - Aucun vaccin - Abattage des infectés et en contact puis désinfection des sites d'enfouissement à la chaux chlorée, quarantaine, nettoyage et désinfection (hypochlorite de sodium), prévention du contact avec les chauves-souris, surveillance sérologique des élevages à haut risque, interdiction transport porcs dans pays touchés, interdiction temporaire production porcine dans régions affectées <p><u>Homme :</u></p> <ul style="list-style-type: none"> - Réduire risque de transmission de la chauve-souris à l'homme (diminuer l'accès des chauves-souris aux denrées, faire bouillir la sève de palmier-dattier, laver et peler les fruits, jeter les fruits en parti consommés par les chauves-souris) - Réduire le risque de transmission de l'animal à l'homme (équipements de protection lors de la manipulation d'animaux malades, de leurs tissus ou lors de l'abattage, éviter le contact avec animaux infectés, protéger les élevages des chauves-souris) - Réduire le risque de transmission interhumaine (éviter le contact avec les personnes infectées et hygiène lors des soins) 	<p><u>Chevaux :</u></p> <ul style="list-style-type: none"> - Vaccin disponible en Australie depuis le 01/11/2012 (Equivac® HeV, accessible seulement aux vétérinaires). Parmi les chevaux vaccinés en prévention aucun n'a contracté le virus Hendra. Il en est de même pour les chevaux vaccinés après avoir été en contact avec des chevaux infectés par le virus Hendra.
------------	--	---

2) Système immunitaire des chauves-souris et interaction avec les virus zoonotiques

Les chauves-souris de la famille des Pteropodidae, réservoir naturel des *Henipavirus*, ne semblent pas présenter de symptômes lors de leurs infections par NiV et HeV contrairement aux autres mammifères (humain, cheval, porc, chèvre, mouton, chat et chien). Les causes de cette absence de signe clinique ont été étudiées dans le but de trouver des solutions thérapeutiques chez l'humain pour contrôler ces maladies émergentes hautement létales.

HeV et NiV infectent les mammifères en se liant aux **récepteurs de l'éphrine EFNB2 et EFNB3** via leur glycoprotéine d'attachement G. Ces récepteurs sont très conservés au sein des vertébrés, ce qui explique le vaste tropisme d'espèce de ces virus. Il ne semble pas y avoir de différence dans la fonction ou l'utilisation de ces récepteurs entre les différents espèces qui pourrait expliquer leur infection différentielle (Bossart et al., 2008).

D'autres caractéristiques concernant le rôle des protéines non structurales (V, W et C) de NiV et HeV dans la réponse immunitaire ont été mis en évidence. Cependant, elles sont également présentes chez les autres paramyxovirus et elles ne suffisent pas à expliquer le fort pouvoir pathogène observé chez de nombreux mammifères ainsi que la coexistence asymptomatique observée chez les chauves-souris (Enchéry et Horvat, 2017).

Contrairement aux autres mammifères, les chauves-souris ne possèdent **aucun gène de la famille PYHIN**. Cette famille joue le rôle important de capteur immunitaire de l'ADN intracellulaire activant les voies de l'inflammasome et/ou de l'interféron. Cette perte des capteurs d'ADN intracellulaire pourrait expliquer la coexistence avec les pathogènes habituellement détectés par cette voie (Ahn et al., 2016).

De plus, l'expression des gènes de l'**IFN- α** (appartenant à la première ligne de défense contre l'infection virale chez les vertébrés) est constitutive chez les chauves-souris et n'est pas stimulée lors d'une infection virale. Cette expression inhabituelle de l'IFN- α stimule un sous-ensemble de gènes associé à une activité antivirale et à une résistance aux dommages causés à l'ADN qui pourrait jouer un rôle important dans la capacité des chauves-souris à coexister avec les virus (Zhou et al., 2016).

Les chauves-souris de la famille des Pteropodidae produisent des anticorps neutralisants contre NiV et HeV lors de leur infection, mais en quantité inférieure à celle observée lors de l'infection des autres espèces de mammifères. Ainsi, il a été suggéré que cette réduction de l'immunité pouvait être le résultat d'une relation hôte-pathogène plus mature (Enchéry et Horvat, 2017).

Les membres de la famille des **protéines APOBEC3 (A3)** font partie des facteurs de restriction antiviraux les plus étudiés. Ces protéines sont des ADN cytosine désaminases capables d'entraîner une hypermutation dans les génomes viraux et d'entraver ainsi la synthèse d'ADN. Elles sont capables de limiter la réplication des rétrovirus, des hépadnavirus, des parvovirus et des rétrotransposons. Les espèces du genre *Pteropus* possèdent la gamme la plus grande et la plus diversifiée de gènes APOBEC3 observée au sein des mammifères. Les produits des gènes de la famille A3 sont connus pour causer des dommages à l'ADN, il est possible que la tolérance accrue de ces dommages chez les chauves-souris ait permis l'expansion de cette famille. De plus, la datation de l'expansion du sous-type A3Z1 coïncide avec l'extinction des LINE-1 chez les chauves-souris Pteropodidae il y a environ 24 millions d'années (Hayward et al., 2018).

II. Particularités des ET chez les chauves-souris

Avant d'étudier un ET de l'espèce *P. giganteus*, une présentation de quelques particularités concernant les ET chez les chauves-souris abordera les transposons à ADN, les ERV et les transferts horizontaux.

A. Transposons à ADN

Alors que l'amplification des transposons à ADN semble avoir cessé dans le génome des mammifères il y a 40 millions d'années, de **multiples vagues d'amplification de transposons à ADN** (hAT, Helitron, mariner, Tc2, piggyBac) durant les 40 derniers millions d'années ont été mises en évidence dans le génome de la chauve-souris *Myotis lucifugus*. Une lignée (appartenant aux piggyBac) semble toujours active. L'hypothèse la plus probable mise en avant pour expliquer cette distribution est une possible introduction d'éléments fondateurs autonomes par de multiples épisodes de transfert horizontal à partir d'une

source inconnue. Cette hypothèse est en accord avec la tendance et l'apparente nécessité des transposons à ADN à réaliser des transferts horizontaux (Ray et al., 2008).

Les transposons à ADN peuvent être considérés, de la même façon que les virus, comme des parasites génomiques. Il a ainsi été suggéré que la capacité des chauves-souris à abriter sans dommage des virus pourrait être liée à la même forme de « **tolérance génétique** » que celle dont elle fait preuve vis-à-vis des transposons à ADN. Ou bien, à l'inverse, c'est la capacité des chauves-souris à tolérer l'infection virale qui aurait pu permettre à ces virus de servir de vecteur et d'introduire des transposons à ADN dans leur génome (Ray et al., 2008).

Contrairement à *M. lucifugus*, les transposons à ADN ne représentent qu'une faible proportion du génome de *Miniopterus natalensis*. De plus, aucun Helitron n'a pu être mis en évidence chez cette espèce. Un faible nombre de TcMariner et de piggyBac ont envahi le génome de l'ancêtre commun à ces deux espèces et seules les chauves-souris appartenant à la famille des Vespertilionidae possèdent une accumulation, unique chez les mammifères, de transposons à ADN au sein de leur génome (fig 37) (Platt et al., 2016).

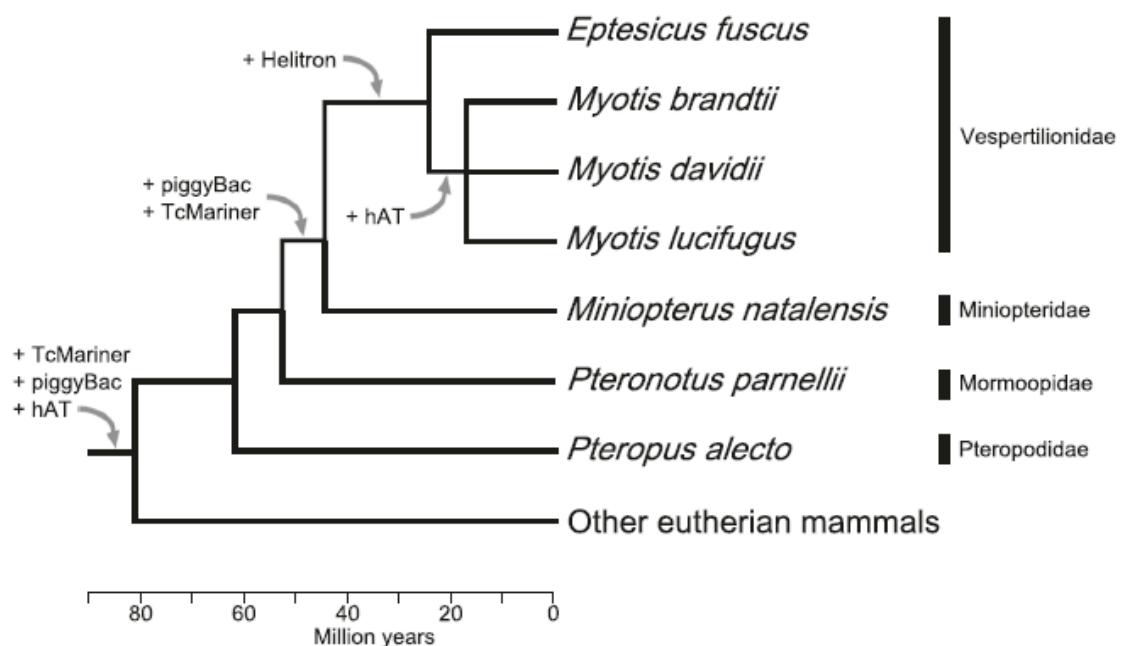


Figure 37 : Invasion des génomes des différentes chauves-souris par les transposons à ADN (Platt et al., 2016)

B. ERV

Le génome des chauves-souris comprend et **exprime des beta-rétrovirus endogènes** (β ERV). Il semble que les beta-rétrovirus aient circulé dans les populations de chauves-souris durant la majeure partie de leur histoire évolutive et qu'ils circulent encore actuellement. Les β ERV sont impliqués dans des maladies telles que le cancer du poumon chez le mouton, ainsi que dans des cancers et des maladies auto-immunes chez l'homme. De plus, il semble

que la distribution des β ERV ne soit pas limitée par la barrière d'espèce. Ainsi, la présence de β ERV ayant la capacité d'être transcrits en éléments viraux actifs chez les chauves-souris, les principaux animaux réservoirs de pathogènes viraux, fait craindre une possible transmission rétrovirale à l'espèce humaine (Hayward et al., 2013).

La course à l'armement menée entre les virus et leur hôte peut prendre des raccourcis, comme dans le cas de l'exaptation d'un rétrovirus aboutissant à une voie de stimulation de la réponse immunitaire par l'ERV correspondant. Par exemple, l'ERV MER41.AIM2 présent chez l'homme régule l'expression des gènes de l'immunité grâce à son domaine de liaison à STAT1. STAT1 est un facteur de transcription, stimulé par l'IFN γ , qui entraîne la production de la protéine AIM2 qui stimule les voies de l'inflammation (Katzourakis et Aswad, 2016). Les motifs de liaison à STAT1 présent chez l'ERV MER41 de l'Homme sont également retrouvés chez les MER41-like des chauves-souris Vespertilionidae, des lémuriformes, des carnivores et des artiodactyles. Ceci suggère qu'ils pourraient également posséder des amplificateurs activés par l'IFN γ (Chuong et al., 2016).

C. Transfert horizontal

Les chauves-souris ont **continuellement acquis de nouveaux transposons à ADN par transfert horizontal**. Parfois, cet échange peut avoir lieu entre des espèces éloignées. Par exemple, deux transposons à ADN ont été transférés horizontalement entre des chauves-souris appartenant à la famille Vespertilionidae et des insectes : l'élément Buster2 retrouvé chez la chauve-souris vespérale (*Nycticeius humeralis*), l'insecte *Rhodnius prolixus* et l'insecte parasite *Mengenilla moldrzyki* et l'élément Kenshin retrouvé chez la chauve-souris *Myotis davidii*, l'insecte *Megachile rotundata* (abeille découpeuse) et l'insecte parasite *M. moldrzyki*. Etant donné que la chauve-souris vespérale se nourrit d'une large gamme d'insectes, il a été suggéré que la prédation sur une grande quantité d'insectes pourrait augmenter l'exposition des chauves-souris aux transferts horizontaux. De plus, l'identification de transposons à ADN quasiment identiques chez des insectes et chez l'insecte parasite *M. moldrzyki* pose la question de savoir si un échange de matériel génétique pourrait être facilité lors de l'interaction hôte-parasite (Tang et al., 2015).

De multiples transferts horizontaux de transposons à ADN nommés SPIN (SPACE INVADERS, appartenant à la superfamille des hAT) ont été mis en évidence chez 7 lignées tétrapodes incluant la Petite chauve-souris brune (*Myotis lucifugus*). Il semble que ces différents tétrapodes aient été infectés dans un intervalle de période évolutive proche, principalement via le même élément actif et à partir de la même espèce source ancestrale, mais ces derniers n'ont pas pu être mis en évidence. Un intermédiaire potentiel à ces transferts horizontaux entre des espèces très divergentes pourrait être les virus à ADN (Pace et al., 2008).

III. Etude d'un ET de *Pteropus giganteus*

A. Objectifs de l'étude

Le génome de *P. giganteus* a récemment été séquencé à partir d'une biopsie réalisée au niveau de la membrane de l'aile d'un spécimen de *P. giganteus* (voir paragraphe suivant *Matériel et méthode*). L'abondance ainsi que la similitude des ET complique l'assemblage des génomes. L'analyse de ces ET **facilitera l'assemblage du génome** de *P. giganteus*.

Cette espèce est le **réservoir de nombreuses zoonoses virales**. Le portage de ces virus est asymptomatique. L'étude des ET de *P. giganteus* pourrait permettre de repérer un ET particulier ayant un lien avec cette résistance au développement de signes cliniques. Cela a par exemple été mis en évidence chez *D. melanogaster* ; l'insertion du rétrotransposon Doc suivie de sa duplication ont permis d'augmenter sa résistance à l'infection virale par le sigma virus (Magwire et al., 2011). De plus, les ET peuvent jouer un rôle dans la régulation de la réponse immunitaire innée de leur hôte lors de leur exaptation, comme dans le cas de l'ERV MER41.AIM2 cité précédemment, qui a permis l'introduction et l'amplification d'activateurs inductibles par l'IFN dans de nombreux génomes de mammifères (Chuong et al., 2016). Notre étude ne se focalisera pas sur la recherche de tels éléments mais nous garderons à l'esprit la possibilité que la résistance particulière de *P. giganteus* au développement de signes cliniques soient issue d'un processus évolutif mettant en jeu des ET.

B. Matériel et méthode

Afin d'étudier les ET de *P. giganteus*, nous disposons des éléments annotés comme ET lors du séquençage préliminaire de son génome ainsi que de logiciels et de banques de données permettant leur analyse bioinformatique.

1) Matériel

Les séquences ADN étudiées ont été obtenues grâce au séquençage du génome de *P. giganteus* dans le cadre du projet BAT-NIPATH. Ce projet ainsi que les différents acteurs et enjeux sont donc présentés dans le paragraphe suivant afin de situer le contexte de l'étude d'un ET de *P. giganteus*.

a) Acteurs

[ECOFECT Projet BAT-NIPATH ; ViroScan3D Projet BAT-NIPATH et la Société Française de BioInformatique, Scaffolding du génome *P. giganteus*, réservoir de nombreux virus hautement pathogènes, 2017]

Le **projet BAT-NIPATH** étudie l'interaction moléculaire entre les chauves-souris frugivores et le virus Nipah hautement pathogène. Il est réalisé en partenariat avec l'équipe « Immunobiologie des infections virales » du Dr Branka Horvat (spécialiste des *Henipavirus*

au sein du Centre International de Recherche en Infectiologie : **CIRI**, l'équipe « Bioinformatique, phylogénie et génomique évolutive » du Dr Marc Bailly-Béchet (bio-informaticien au Laboratoire de Biométrie et Biologie Evolutive : **LBBE**), du Dr Catherine Legras-Lachuer (dirigeant la société **ViroScan 3D**) et de Hervé Raoul (dirigeant le **Laboratoire P4 Jean Mérieux-Inserm**). Ce projet est financé entre autre par le **Laboratoire d'Excellence (LabEx) ECOFECT**. Ce laboratoire a pour objectif d'approfondir les connaissances de la dynamique des maladies infectieuses afin de promouvoir des stratégies de lutte innovantes en replaçant le pathogène dans la complexité de son environnement. Il regroupe pour cela des chercheurs de différents champs disciplinaires — virologie, immunologie, bactériologie, biologie moléculaire, épidémiologie, génétique et génomique des populations, évolution moléculaire, écologie, médecine clinique — et méthodologiques — (bio)mathématique et (bio)informatique.

b) Etapes et enjeux

La première étape du projet BAT-NIPATH est le **séquençage du génome de la chauve-souris *P. giganteus***, hôte principal du virus Nipah ne présentant aucun signe clinique sévère lors de leur infection. Une biopsie a été réalisée au niveau de la membrane de l'aile d'un spécimen de *P. giganteus* mâle détenu en captivité par le zoo de Vienne (Autriche) et âgé de 6 ans. La société Viroscan3D a réalisé le **séquençage de novo** du génome de *P. giganteus* à partir de l'ADN extrait de cette biopsie.

Les étapes suivantes du projet BAT-NIPAH visent à comparer la signature transcriptomique des cellules de *P. giganteus* lors de leur infection par le virus Nipah à celle des cellules humaines afin d'identifier des facteurs spécifiques aux chauves-souris qui seraient responsables du contrôle de l'infection, le but étant de comprendre comment cette chauve-souris parvient à contrôler l'infection par le virus Nipah.

c) Séquençage du génome de *P. giganteus* (méthodes et algorithmes utilisés)

La société **Viroscan3D**, spécialisée en séquençage NGS (« Next Generation Sequencing » en anglais), a réalisé le **séquençage du génome de *P. giganteus*** grâce à la **technologie Illumina** (réalisé par Julien Fouret sous la direction de Catherine Legras-Lachuer).

Les **technologies de séquençage de nouvelle génération (NGS)** sont nommées ainsi par opposition à la méthode de Sanger, considérée comme une technologie de première génération. Parmi ces technologies de séquençage NGS figurent le pyroséquençage grâce à la technologie 454 (2005), la technologie Illumina/Solexa (2006), la technologie SOLiD (Sequencing by Oligo Ligation Detection) (2007) et la technologie Ion Torrent (2010). Ces technologies ont permis de réduire significativement le coût et le temps nécessaire au séquençage. En 10 ans, le coût du séquençage d'un génome humain a été divisé par 10 000 (passant de 10 000 000\$ en 2004 à 1 000\$ en 2014). Les séquences obtenues sont cependant plus courtes qu'avec la méthode Sanger. Le séquençage Nanopore est, quant à lui, considéré comme une technologie de troisième génération car il permet le séquençage de molécules uniques en temps réel (van Dijk et al., 2014).

Le **séquençage par la technologie Illumina/Solexa** est réalisé grâce à la préparation d'une banque d'ADN dénaturé lié par une extrémité à une surface solide recouverte d'oligonucléotides « adaptateurs ». L'extrémité libre de chaque fragment d'ADN s'hybride à un adaptateur complémentaire sur la surface, ce qui initie la synthèse de leur brin complémentaire. De multiples cycles sont ainsi réalisés (amplification en phase solide suivie d'une dénaturation) permettant la création de groupes d'environ 1000 copies de molécules d'ADN monocaténaïres. Le séquençage est ensuite permis par l'ajout des réactifs de synthèse : les amorces, l'ADN polymérase et les quatre nucléotides terminateurs réversibles marqués. Le premier nucléotide incorporé est identifié grâce à sa couleur (émission de lumière à une fréquence particulière), puis sont retirés le terminateur 3' présent sur la base incorporée ainsi que le fluorophore. Le cycle est à nouveau répété afin de déterminer la couleur de la seconde base qui sera ajoutée, et ainsi de suite (Dijk et al., 2014).

Les NGS n'ont pas une précision de 100% ; des erreurs liées aux biais technologiques sont ainsi présentes dans les reads. Par exemple, la technologie Illumina présente un taux d'erreurs d'environ 0,1%, principalement lié aux substitutions du fait de son très haut débit (Peterlongo, 2016). Ainsi, lors de l'assemblage du génome de *P. giganteus*, il faut tout d'abord corriger les erreurs présentes dans les reads. Pour cela, le premier algorithme utilisé est *Lighter* (LIGHTweight ERror corrector).

Lighter est un outil qui réalise trois passages au sein des reads. Le premier permet d'échantillonner les k-mers présents dans les reads de manière aléatoire, cet échantillon est stocké dans un filtre Bloom A (un filtre Bloom est une structure compacte de données probabilistes). Le second utilise le filtre Bloom A pour identifier les k-mers solides ; ils sont stockés dans le filtre Bloom B. Le dernier utilise le filtre Bloom B afin de corriger les erreurs dans les données reads (Song et al., 2014).

Le second algorithme utilisé lors de l'assemblage est *Abyss-pe*, il est basé sur les graphes de De Bruijn et permet l'assemblage en « contigs » (Peng et al., 2010).

Il est souvent difficile d'annoter et d'assembler l'ADN répété lors de l'analyse d'un génome. Le **pipeline bioinformatique dnaPipeTE** (De Novo Assembly and Annotation Pipeline for Transposable Elements) permet d'**assembler**, d'**annoter** et de **quantifier** les séquences répétées à partir de petits jeux de données NGS non assemblées (reads). Pour cela, dnaPipeTE utilise la méthode **Trinity** (Grabherr et al., 2011) afin d'assembler les répétitions figurant au sein de jeux de données obtenus par faible couverture génomique. Il peut ainsi générer des séquences répétées complètes et permettre la récupération d'autres consensus au sein d'une famille d'ET. DnaPipeTE utilise également **RepeatMasker** et la base de données **Repbase** (Bao et al., 2015) afin d'annoter automatiquement les répétitions et de les quantifier. Enfin, dnaPipeTE calcule la distribution d'âge des ET pour les copies les plus récentes, en utilisant la divergence entre les reads et les contigs (Goubert et al., 2015). La figure suivante (fig 38) illustre le fonctionnement du pipeline dnaPipeTE.

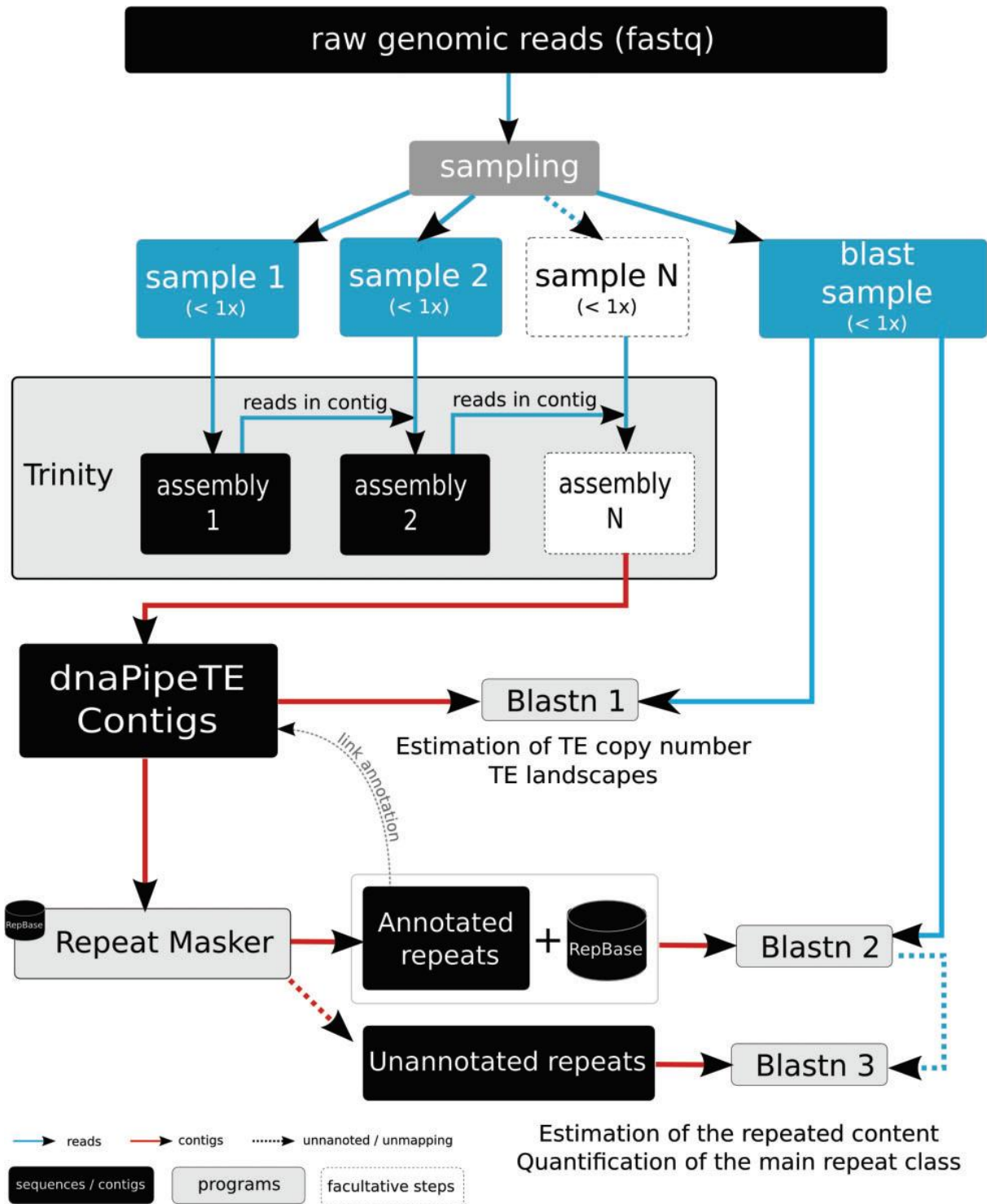


Figure 38 : Vue d'ensemble du pipeline dnaPipeTE (d'après Goubert et al., 2015).

Premièrement, les reads sont échantillonnés au format FASTQ. Ensuite, l'assemblage des répétitions est effectué en utilisant deux ou plusieurs itérations de *Trinity*. Pour chaque itération, les reads précédemment assemblés sont ajoutés à l'échantillon suivant afin d'améliorer l'assemblage. Dans l'étape suivante, les contigs assemblés sont annotés à l'aide de RepeatMasker. Finalement, les reads de "l'échantillon BLAST" sont comparés par BLASTN à tous les contigs afin d'estimer l'abondance relative de chaque ET assemblé et ainsi évaluer la composition en ET. Dans un second BLASTN, le même échantillon est comparé successivement aux contigs annotés joints à la bibliothèque Repbase, puis aux contigs non annotés afin de récupérer des copies qui n'auraient pas été assemblées et d'obtenir une estimation plus globale du contenu en éléments répétés.

Une estimation de la proportion des différents ET est obtenue par séquençage du génome de *P. giganteus* avec une couverture de 0,5x (fig 39) (Julien Fouret, Viroscan3D). La couverture du séquençage est le nombre moyen de fois que chaque nucléotide est séquencé (Peterlongo, 2016).

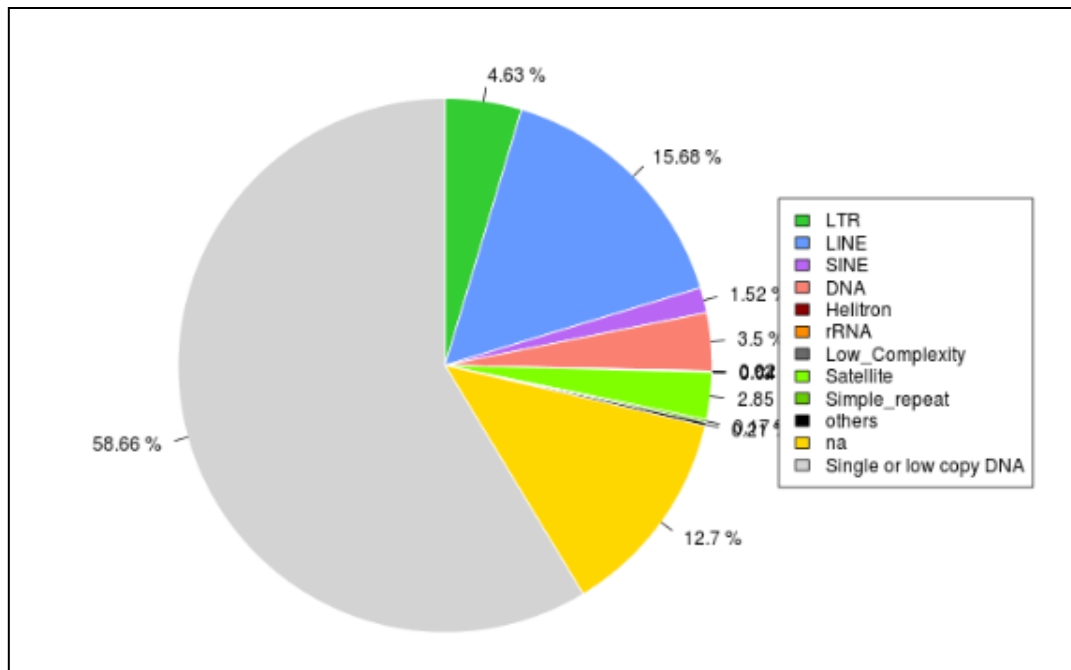


Figure 39 : Composition du génome de *P. giganteus* en ET (d'après Julien Fouret, Viroscan3D).

2) Méthode

a) Séquences ADN disponibles pour l'étude

Les séquences utilisées dans l'étude sont les séquences ADN annotées comme étant des ET lors du séquençage préliminaire du génome de *P. giganteus*. Elles ont été classées dans 5 catégories différentes (cf figure 39) : **DNA** (34 823 éléments), **Helitron** (343 éléments), **LINE** (100 555 éléments), **LTR** (40 746 éléments) et **SINE** (14 796 éléments).

b) Logiciels et banques de données utilisés

Les différents logiciels et banques de données utilisés pour traiter ces séquences ADN sont : Censor de GIRI, BLAST de NCBI, ExpASY Translate Tool, Ensembl et Genomicus qui sont sur des serveurs en ligne, et SeaView un logiciel installé en local.

Censor de GIRI (Genetic Information Research Institute)

L'institut de recherche GIRI est consacré à l'étude et à la caractérisation d'éléments ADN répétitifs ou mobiles chez les espèces eucaryotes. Il maintient et développe la base de données de référence nommée **Repbase**, permettant d'annoter l'ADN répétitif présent dans les génomes eucaryotes. GIRI fournit l'outil **Censor** (<https://www.girinst.org/censor/>) (Kohany et al., 2006) qui permet de comparer des séquences soumises à la base de données de référence afin d'en identifier les portions homologues. Les répétitions identifiées sont

ensuite classées dans un rapport et les portions homologues sont « masquées » dans la séquence soumise avec des symboles « X » (**Repeat Masking**).

BLAST (Basic Local Alignment Search Tool) de NCBI (National Center for Biotechnology Information)

Le programme **BLAST** de NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) (Altschul et al., 1990) permet de trouver des régions de similitude entre les séquences biologiques. Il compare des séquences nucléotidiques ou protéiques à des bases de données de séquences et calcule la signification statistique de la similitude trouvée. Plusieurs comparaisons sont possibles : comparer une séquence nucléotidique à des bases de données de séquences nucléotidiques (**Nucleotide BLAST/BLASTN**), comparer une séquence protéique à des bases de données de séquences protéiques (**Protein BLAST/BLASTP**), comparer une séquence nucléotidique traduite en une séquence protéique à des bases de données de séquences protéiques (**BLASTX**) ou bien comparer une séquence protéique traduite en une séquence nucléotidique à des bases de données de séquences nucléotidiques (**TBLASTN**).

Lorsqu'une région de similitude est trouvée entre deux séquences (« hit »), BLAST fournit comme information un score d'alignement ainsi qu'une valeur E.

Le **score** d'alignement évalue la similitude entre les séquences. Il est calculé grâce à une matrice de similitude qui attribue des scores positifs aux identités ainsi qu'aux remplacements conservateurs et des scores négatifs aux discordances (Altschul et al., 1990).

La **E-value** (Expect value) est un paramètre qui décrit le nombre de hits que l'on peut s'attendre à avoir par hasard lors de la recherche dans une base de données d'une taille particulière. Cette valeur est proportionnelle au volume de la base de données interrogée (par exemple toute la base, seulement les métazoaires, ou seulement les vertébrés ou seulement un génome donné) ; plus celui-ci sera élevé et plus la E-value sera haute pour un hit donné. La E-value diminue exponentiellement à mesure que le score augmente. Plus elle est faible (plus elle est proche de zéro) et plus la correspondance est significative (voir les FAQs BLAST NCBI).

SeaView

Le logiciel **SeaView** (<http://doua.prabi.fr/software/seaview>) permet de visualiser des alignements de séquences (ADN, ARN et protéiques). Sont implémentés dans ce logiciel les programmes d'alignements ClustalO version 2 et Muscle, ainsi que des outils de reconstructions phylogénétiques à partir des séquences alignées.

ExpASy Translate Tool

ExpASy Translate Tool (<https://web.expasy.org/translate/>) est un outil permettant la traduction d'une séquence nucléotidique (ADN ou ARN) en une séquence protéique dans les 6 cadres de lecture possibles (3 possibilités pour débiter la lecture d'un codon sur le brin sens ainsi que sur le brin anti sens).

Ensembl

Ensembl (<http://www.ensembl.org/index.html>) (Zerbino et al., 2018) permet de naviguer dans les génomes de vertébrés. Il permet par exemple de comparer des gènes entre différentes espèces.

Genomicus

Genomicus (<http://www.genomicus.biologie.ens.fr/genomicus-84.01/cgi-bin/search.pl>) (Louis et al., 2015) permet également de naviguer dans les génomes, et cela dans plusieurs dimensions (linéairement le long des axes chromosomiques, transversalement entre différentes espèces et chronologiquement au cours de l'évolution). Lorsqu'un gène est recherché, ses copies orthologues et paralogues sont affichées dans tous les autres génomes métazoaires séquencés et également dans le génome ancestral prédit chez les espèces ancestrales. Genomicus utilise des données provenant d'Ensembl.

c) Genèse des résultats

La **première étape** de l'étude consiste à repérer un élément qui présenterait un intérêt particulier (en raison d'une forte similitude avec un autre ET par exemple) au sein des éléments annotés comme ET chez *P. giganteus* (fig 40).

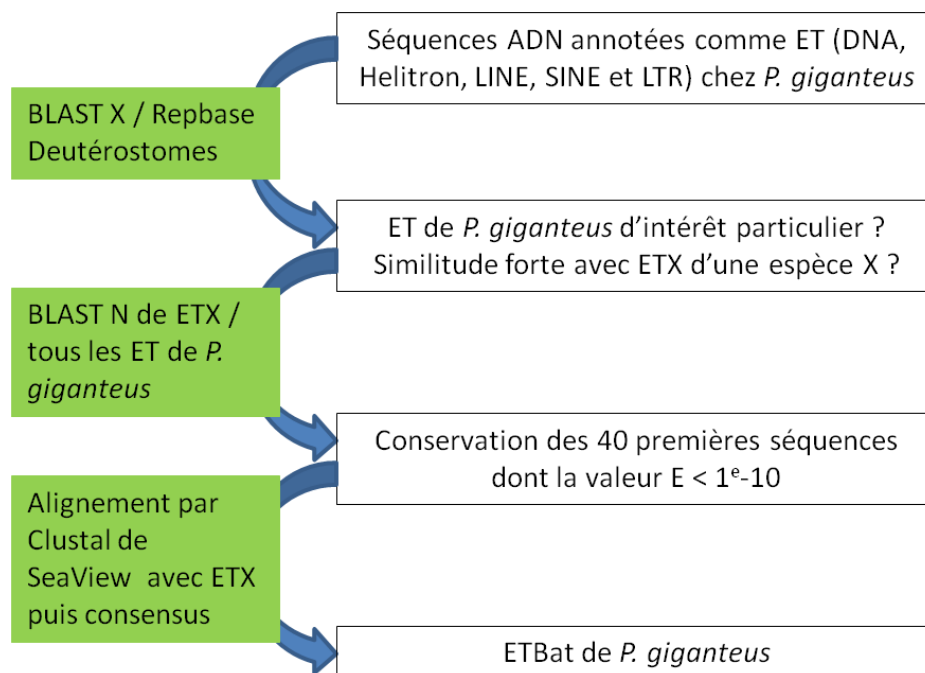


Figure 40 : Identification d'un ET de *P. giganteus* présentant un intérêt particulier

Dans ce but, une large comparaison (par BLASTX) des éléments annotés comme ET chez *P. giganteus* (catégories DNA, Helitron, LINE, LTR et SINE) est réalisée par rapport à tous les éléments annotés comme ET chez les Deutérostomes dans la base de donnée Repbase. Seule cette catégorie de Repbase a été sélectionnée afin de ne pas considérer dans un

premier temps les éléments appartenant à des espèces phylogénétiquement trop distantes de *P. giganteus*. Le taxon Deuterostomia regroupe les espèces appartenant aux Chaetognatha (les Chétognates, qui sont des vers en forme de flèche, appelés aussi vers sagittaires), aux Echinodermata (échinodermes) aux Hemichordata (hémichordés) et aux Chordata (chordés). A l'issue de ce BLASTX, un élément annoté comme ET de *P. giganteus* semblant présenter un intérêt particulier en raison d'une forte similitude avec l'ETX d'une autre espèce X sera retenu et étudié plus en détail.

Les séquences nucléotidiques annotées comme ET chez *P. giganteus* sont de courts reads. Ainsi, la séquence retenue ne représentera donc probablement qu'une partie d'un ET (ou d'un autre élément proche d'un ET) du génome de *P. giganteus*. Afin de reconstruire cet élément dans son intégralité, un BLASTN de l'ETX est réalisé par rapport à l'ensemble des séquences annotées comme ET chez *P. giganteus*. Cette méthode est susceptible d'entraîner un biais dans la reconstruction de l'élément chez *P. giganteus*. En effet, le consensus reconstruit chez *P. giganteus*, à partir de séquences sélectionnées en raison de leur similitude avec le consensus d'un ET de l'espèce X, sera « artificiellement » proche de ce dernier. Les 40 meilleurs hits présentant une valeur $E < 1e^{-10}$ sont conservés, alignés avec l'ETX (fig 41), puis un consensus est obtenu grâce à Seaview (nommé ETBat).

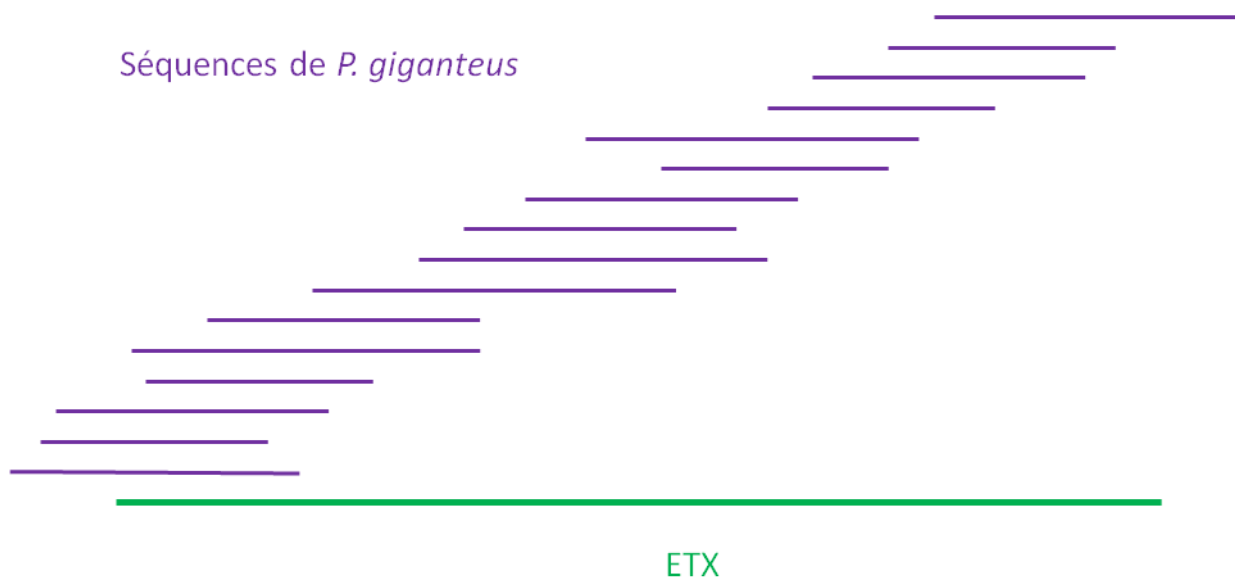


Figure 41 : Représentation schématique de l'alignement des séquences de *P. giganteus* et de l'ETX grâce à SeaView

La **seconde partie** de l'étude vise à repérer les éléments caractéristiques d'ETBat (TIR, LTR ou autre selon le type d'élément identifié) ainsi que les éventuels ORF présents au sein de cette séquence (fig 42).

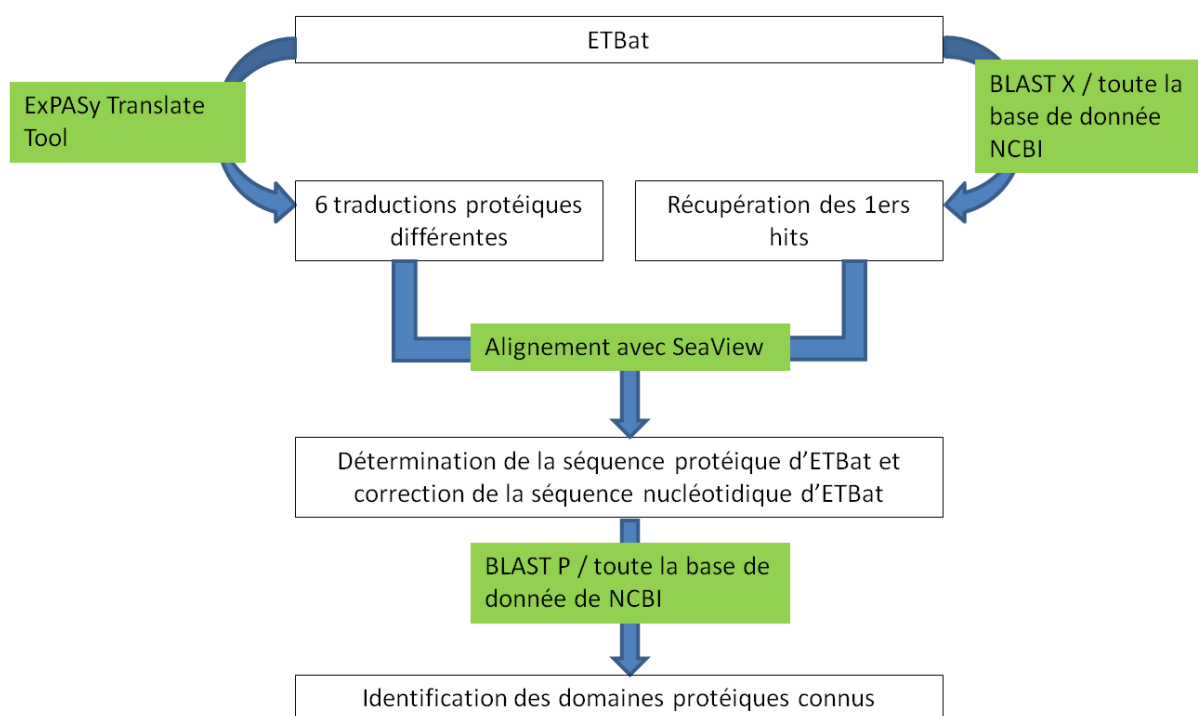


Figure 42 : Identification des ORF d'ETBat

Pour repérer le/les ORF d'ETBat, sa séquence nucléotidique est traduite en protéine grâce à ExPASy Translate Tool. Cet outil donne comme résultat 6 traductions protéiques différentes (les trois phases sur le brin sens et les trois autres sur le brin anti-sens). Pour retrouver les différentes parties de la/les protéine/s codée/s par ETBat, un BLASTX de sa séquence est réalisé (paramètres standards sur NCBI) et les séquences protéiques des premiers hits sont récupérées. Elles correspondent aux protéines de la base de données NCBI présentant une homologie avec une partie de la séquence nucléotidique d'ETBat traduite en protéine dans les 6 cadres possibles. Elles sont ensuite alignées dans SeaView avec les fragments protéiques trouvés grâce à ExPASy Translate Tool afin de reconstruire la/les protéine/s codée/s par ETBat. La détermination de la séquence protéique d'ETBat permet de corriger la séquence ADN obtenue lors de la première partie de l'étude afin d'obtenir le meilleur consensus possible. A la fin de cette étape, les deux séquences obtenues sont donc le consensus de la séquence ADN de ETBat corrigé et la protéine correspondante. Il est alors possible de comparer la séquence protéique d'ETBat aux séquences protéiques semblables appartenant à la base de données NCBI afin d'identifier d'éventuels domaines protéiques connus.

La **troisième partie** de l'étude vise à étudier la phylogénie d'ETBat. Un BLASTN de la séquence nucléotidique d'ETBat est réalisé par rapport à l'ensemble de la base de données NCBI. Les paramètres classiques sont gardés et les 100 premiers hits générés sont conservés. Ils sont classés par E-value croissantes et par max score décroissants pour une E-value identique.

C. Résultats

1) Reconstruction d'ETBat

a) 1^{ère} partie : reconstruction d'une séquence présentant des homologies avec un Tigger1 chez *P. giganteus*

Lors du BLASTX des éléments annotés comme ET dans le génome de *P. giganteus* par rapport à Repbase, deux séquences attirent notre attention car elles possèdent une **E-value de 0,0 avec l'élément nommé « TIGGER1 Homo sapiens »**, c'est-à-dire qu'ils sont **certainement homologues**.

Cet élément a été découvert en 1996 et nommé ainsi en lien avec le personnage Tigrou (« Tigger » en anglais) introduit dans le second volume de Winnie l'ourson (« The House at Pooh Corner » écrit par Milne en 1928) qui utilise sa queue comme un bâton sauteur (« pogo-stick ») ; car Tigger1 représente un élément transposable de mammifère ressemblant fortement au transposon à ADN pogo décrit chez la drosophile (Smit et Riggs, 1996).

Les deux séquences très semblables au Tigger1 d'*H. sapiens* (989pb et 1072pb) appartiennent toutes les deux à la **catégorie DNA** des ET de *P. giganteus*. Lorsqu'elles sont alignées dans SeaView, il apparaît qu'elles sont identiques sur quasiment toute leur longueur (de leur 1^{er} nucléotide au 927^{ème}). La partie commune de ces séquences est soumise à Censor qui montre l'alignement local avec le Tigger1 d'*H. sapiens*. La séquence consensus du Tigger1 d'*H. sapiens* indexée dans Repbase fait 2418pb, elle est composée de TIR de 23pb, d'un site cible TA et de 2 ORF (ORF1 bases 425 à 1789 et ORF2 bases 1811 à 2206). L'ORF1 code pour une transposase possédant un domaine catalytique DDE. Cette transposase est très similaire à la protéine centromérique CENP-B (centromere-associated protein B) des mammifères suggérant que cette dernière est dérivée d'une transposase pogo-like par exaptation (Smit et Riggs, 1996). Une exaptation est une fonctionnalité qui améliore la fitness mais qui n'avait à l'origine pas été sélectionnée pour la fonction qu'elle occupe actuellement, contrairement au cas d'une adaptation (Gould et Vrba, 1982). La protéine CENP-B n'a pas été isolée chez des espèces autres que les espèces appartenant aux mammifères (Casola et al., 2007). La partie commune aux deux séquences retenues chez *P. giganteus* s'aligne avec la partie centrale du Tigger1 d'*H. sapiens* (de la base 465 à 1393) (fig 43).

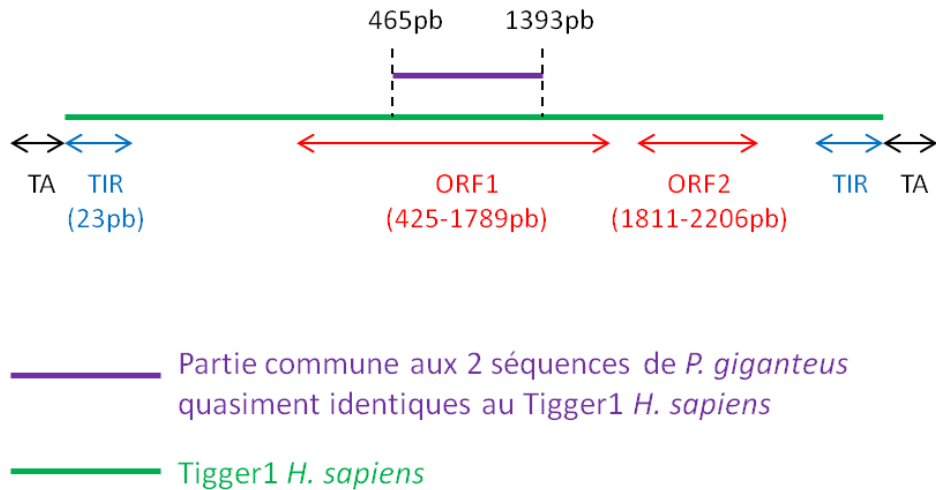


Figure 43 : Représentation schématique du TIGGER1 *Homo sapiens* et de son alignement avec la partie commune des deux séquences de *P. giganteus* qui lui sont quasiment identiques

Ces deux fragments d'ET sont-ils les seuls qui existent dans le génome de *P. giganteus* ou bien est-il possible de reconstruire un élément semblable à un Tigger1 complet ?

Afin de répondre à cette question, une tentative de reconstruction de l'intégralité d'un ET semblable à Tigger1 est entreprise. Pour cela, la séquence du Tigger1 d'*H. sapiens* est utilisée dans un BLASTN en retour par rapport à l'ensemble des séquences annotées comme ET dans le génome de *P. giganteus* afin de récupérer tous les reads homologues à cet élément. Les 40 premières séquences possédant le meilleur score parmi les 1841 ayant une E-value < $1e^{-10}$ sont conservées puis alignées grâce au programme ClustalO. Grâce au programme SeaView, une séquence consensus est déduite de cet alignement. La séquence obtenue n'est pas totalement déterminée avec cette méthode. Le consensus utilisé est le consensus 60, c'est-à-dire constitué des nucléotides qui sont identiques dans plus de 60% des séquences alignées de *P. giganteus*. Pour les sites n'étant pas représentés à plus de 60%, le logiciel propose un "N", et c'est un choix réalisé à l'œil qui est fait au cas par cas si cela est faisable. La séquence nucléotidique obtenue est nommée ETBat.

b) Partie 2 : séquences caractéristiques et ORF d'ETBat

L'alignement de la séquence nucléotidique du Tigger1 d'*H. sapiens* et ETBat (2409pb) dans Seaview permet de repérer qu'ETBat possède également des **TIR de 23pb semblables** à ceux du Tigger1 d'*H. sapiens* et possède le **même site cible TA** (fig 44).

Lorsqu'un BLASTP de la séquence protéique d'ETBat est réalisé, deux domaines fortement similaires à d'autres protéines connues sont identifiés. Le premier domaine est très similaire au **domaine putatif de liaison à l'ADN** des protéines CENP-B, mouse jerky et transposase (42-109 aa, E-value $2,01e^{-10}$). Le second domaine présente également de fortes similitudes avec le **domaine catalytique DDE** des endonucléases de la superfamille DDE ; cette superfamille inclut la protéine CENP-B (175-362 aa, E-value $4,53e^{-55}$).

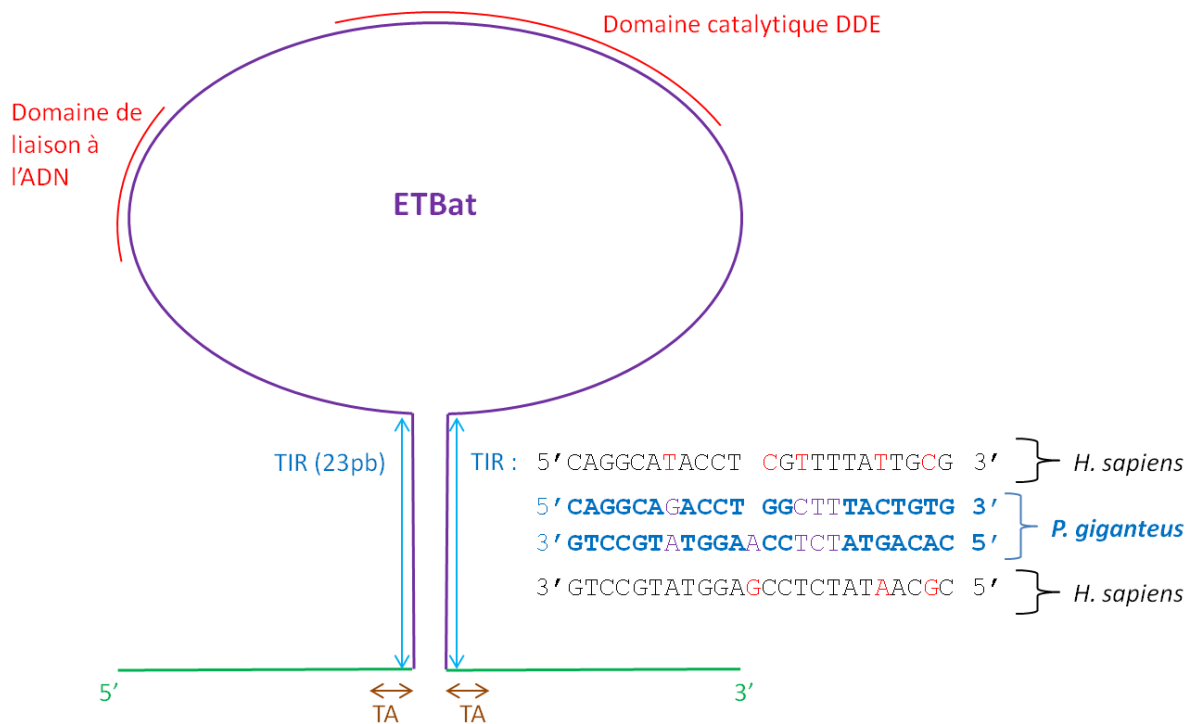


Figure 44 : ETBat de *P. giganteus*, un élément homologue à *Tigger1* d'*H. sapiens*, et ses séquences caractéristiques.

Les bases azotées des TIR de *P. giganteus* figurent en bleu. Lorsque la complémentarité des bases est respectée elles sont également en gras. Les bases azotées des TIR d'*H. sapiens* figurent en noir lorsqu'elles sont identiques à celles des TIR de *P. giganteus*, en rouge sinon.

c) Partie 3 : étude de la phylogénie d'ETBat par BLASTN dans NCBI

Grâce à un BLASTN de la séquence nucléotidique d'ETBat, les 100 séquences (hits) qui lui sont le plus proches, indexées dans la base de données NCBI, sont récupérées. Ces hits sont donnés par ordre croissant de E-value et pour une E-value identique par ordre décroissant de « max score ». Ils sont présentés dans l'arbre suivant (fig 45).



Figure 45 : Arbre phylogénétique obtenu par BLASTN de la séquence nucléotidique d'ETBat

Les groupes de séquences retrouvés dans cet arbre phylogénétique concernent des protéines d'humain, de nombreux scaffolds du kiwi de Mantlell (*Apteryx australis mantelli*), le Tigger1 d'*H. sapiens*, des TIGD1 de primates, des protéines et TIGD1 de carnivores (*Vulpes vulpes*, *Panthera pardus* et *Mustela putorius furo*), des protéines de Cétartiodactyles (*Physeter catodon*, *Bubalus bubalis*, *Bos taurus* et *Bos indicus*), des protéines de Chiroptères (*Rhinolophus sinicus*, *Roussettus aegyptiacus* et *Pteropus alecto*), des protéines d'Archosauriens (*Gavialis gangeticus* et *Dromaius novaehollandiae*), une protéine du rhinocéros blanc du Sud (*Ceratotherium simum simum*) et une protéine du cheval (*Equus caballus*).

D. Discussion

1) ETBat de *P. giganteus* et Tigger1 d'*H. sapiens*

L'élément ETBat reconstitué chez *P. giganteus* possède les caractéristiques d'un Tigger1 (site cible, TIR et ORF homologue à CENPB). Il semble donc que la séquence reconstruite chez *P. giganteus* soit la **séquence consensus de l'élément Tigger1** chez cette espèce. Cependant, la séquence construite pourrait être légèrement différente si d'autres séquences avaient été utilisées en plus des 40 sélectionnées, ou si d'autres choix avaient été faits lors du remplacement manuel des nucléotides non déterminés par le consensus⁶⁰.

Lors du BLASTN d'ETBat, le meilleur hit correspond à la séquence de l'ET **Tigger 1 d'*H. sapiens*** (« max score » 3576, « query cover » 98%, « E value » 0,0 et « ident » 93%), cette même séquence avait été retrouvée lors de la première étape comme présentant la

meilleure homologie avec une des séquences annotées comme ET chez *P. giganteus*. Ce résultat peut être lié au biais induit par la façon de reconstruire ETBat comme il peut également être un signe de **transfert horizontal** entre ces deux espèces. Afin de les distinguer, il faudrait rechercher les différentes copies de Tigger1 dans le nouvel assemblage du génome de *P. giganteus* ainsi que dans celui d'*H. sapiens* et calculer les ratios de substitution nucléotidique (dN/dS) pour estimer le temps de divergence de ces éléments. Des vecteurs pouvant être envisagés comme responsables de ce transfert sont les virus ou les parasites. Les virus envisagés sont ceux qui étaient capables d'infecter à la fois les ancêtres de *P. giganteus* et d'*H. sapiens*. Les parasites envisagés sont ceux qui auraient pu être échangés lors de la consommation de la viande crue d'ancêtres de *P. giganteus* par des ancêtres d'*H. sapiens* avant l'utilisation du feu.

D'autres séquences également très semblables à ETBat sont retrouvées dans le génome humain. Ces séquences correspondent à des **introns** (1^{er} intron du gène KCNC1, 4^{ème} intron du gène GRIN2B, 28^{ème} intron du gène COL4A5 et 1^{er} et 14^{ème} introns du gène RGS6), suggérant que des Tigger1 sont présents dans ces introns. Afin de vérifier s'il s'agit bien d'insertions de ce Tigger1, ces hits pourraient être comparés à la séquence consensus du Tigger1 d'*H. sapiens*. Pour dater ces insertions, il serait intéressant d'étudier les gènes orthologues à KCNC1, GRIN2B, COL4A5 et RGS6 chez les espèces et les taxons voisins afin de déterminer si l'élément Tigger1 est également présent au sein des mêmes introns de ces gènes.

2) Etude de la phylogénie d'ETBat : transfert horizontal

La plus grande quantité de hits concerne des **scaffolds du génome d'un oiseau, le kiwi de Mantell** (*Apteryx australis mantelli*), avec des alignements sur la quasi-totalité de la longueur d'ETBat pour nombre d'entre eux. Au vu de ces résultats, il est possible qu'un élément transposable très semblable à ETBat soit présent dans le génome du kiwi de Mantell, à la fois sous forme complète ou quasi complète, ainsi que sous forme dégénérée. Une si forte similitude avec une espèce aussi éloignée phylogénétiquement (appartenant aux Archosauriens) nous conduit à poser l'**hypothèse d'un transfert horizontal**. Un constat similaire avait été fait lors de l'étude des transposons de la chauve-souris *Miniopterus natalensis* (Platt et al., 2016). Il avait été remarqué que l'élément nommé Tigger1 chez cette chauve-souris partageait plus de 97% de similitude sur plus de 92% de sa longueur avec plus de 20 insertions présentes chez le kiwi de Mantell. Cet élément était également très similaire au Tigger1 de l'humain comme notre élément ETBat. D'autres espèces d'**Archosauriens** sont également retrouvées parmi les hits : l'**émeu d'Australie** (*Dromaius novaehollandiae*) (3 séquences indexées « **tigger transposable element-derived protein1-like** » présentes sur des scaffold non placés) et le **gavial du Gange** (*Gavialis gangeticus*) (séquence indexée « **TIGD1** »). Si l'hypothèse du transfert horizontal pouvait expliquer cette distribution, alors il pourrait avoir eu lieu **chez un ancêtre commun aux Archosauriens**. Afin d'investiguer cette hypothèse de transfert horizontal, il faudrait étudier la répartition de cet ET au sein des différents taxons voisins de ces espèces. Pour cela, il conviendrait de voir si la répartition de cet ET correspond ou non à la phylogénie de ces espèces et de vérifier que la divergence entre leurs ET est bien inférieure à celle attendue étant donné leur temps de divergence (Schaack et al., 2010).

Un des hits concernant cet oiseau s'aligne avec la totalité d'ETBat. L'étude de cette séquence permet de constater qu'elle possède des TIR de seulement 13pb, c'est-à-dire la moitié de la longueur des TIR d'ETBat et de Tigger1 d'*H. sapiens*. Aucune autre séquence quasi complète retrouvée parmi les hits ne s'aligne avec le TIR en 5', mais la plupart s'alignent avec le TIR en 3'. Il serait intéressant de rechercher la séquence complète contenant les deux TIR dans l'ensemble du génome du kiwi de Mantell afin de déterminer si d'autres copies de cette séquence existent (Le Duc et al., 2015).

3) Etude de la phylogénie d'ETBat : pertinence des bases de données

Parmi les séquences proches de l'élément ETBat, un groupe de séquences nommées **TIGD1** (tigger transposable element derived 1) est présent chez des **primates**. Ces éléments sont bien documentés : TIGD1 est un **élément transposable domestiqué** faisant parti d'une paire de gènes en position bidirectionnelle (c'est-à-dire utilisant le même promoteur mais orientées en sens inverse). Sa séquence possède le domaine catalytique DDE/D mais un des TIR est absent. TIGD1 a été détecté dans le génome de toutes les espèces de primates mais est absent dans les autres groupes de mammifères qui ont été étudiés. Sa domestication est donc supposée avoir eu lieu lors de la divergence de la lignée des primates, il y a 67 à 98 millions d'années (Kalitsis et Saffery, 2009). Etant donné la très forte homologie entre ETBat et Tigger1 d'*H. sapiens* et le fait que le gène TIGD1 soit issu d'un Tigger1 domestiqué, il n'est pas surprenant de retrouver ces gènes parmi les séquences présentant une très forte homologie avec ETBat. Cependant, comme d'autres séquences retrouvées dans les hits appartenant à des espèces non primates sont annotées comme TIGD1, quelle est alors la pertinence de cette annotation dans la base de données ? Par exemple, concernant la séquence « TIGD1 » du gavial du Gange, elle est annotée comme étant un gène orthologue au TIGD1 d'*H. sapiens*, présent sur un scaffold, mais les autres gènes sur ce scaffold entourant « TIGD1 » n'ont pas été identifiés dans la base de données NCBI. Il n'y a ainsi aucun indice indiquant que cet élément fasse partie de la même synténie que celle observée chez les primates (TIGD1 et compris entre les gènes CHRNG et EIF4E2 chez les primates). Ce gène contient également un TIR en 3' n'ayant qu'un nucléotide différent par rapport au TIR du TIGD1 d'*H. sapiens*. Il semble donc que le nom de TIGD1 lui ait été attribué en raison de sa forte homologie avec TIGD1 d'*H. sapiens* mais sans prise en compte de l'existence d'une synténie conservée. Cette annotation peut conduire à des erreurs d'interprétation si on suppose que tous les gènes nommés TIGD1 constituent des gènes orthologues (c'est-à-dire descendant d'une même séquence unique présente chez un ancêtre commun). Au vu du nombre d'éléments semblables à ETBat retrouvés chez le kiwi de Mantell, espèce appartenant aux Archosauriens comme le gavial du Gange, il pourrait être intéressant de voir si une synténie commune à ces deux espèces peut être identifiée. Par exemple, en identifiant les gènes présents sur le scaffold sur lequel a été identifié le "TIGD1" du gavial et en les recherchant dans le génome du kiwi. L'espèce la plus proche possédant un élément annoté TIGD1 dans Genomicus est la tortue trionyx de Chine (*Pelodiscus sinensis*). Cet élément se trouve entre les gènes CBLN2 et RTTN. Si une synténie est retrouvée chez le kiwi de Mantell et le gavial du Gange, il serait intéressant de voir si elle correspond à celle de la trionyx de Chine.

4) Etude de la phylogénie d'ETBat : différents épisodes de domestication

Un hit correspond à un élément annoté comme TIGD1 chez le furet (*Mustela putorius furo*) est positionné entre le gène NUP35 et le gène DUSP19 sur un scaffold non placé. Ce hit peut être rapproché du hit concernant plus de la moitié de la longueur du **gène DUSP19** (« dual specificity phosphatase 19 ») du renard roux (*Vulpes vulpes*). De plus, dans Ensembl cette synténie est également conservée chez un autre représentant des Caniformes, le chien (*Canis lupus familiaris*) (chromosome 36). Un hit nommé « tigger transposable element-derived protein 1 » concerne le léopard (*Panthera pardus*), espèce appartenant à l'ordre des Carnivores comme les trois précédentes. Selon Ensembl, le léopard ne possède pas de gène orthologue à TIGD1. Cependant, lorsque ce hit est visualisé avec Genome Browser, il apparaît que cet élément est situé juste à côté d'un gène identifié comme DUSP19. Les séquences appartenant au renard roux et au léopard possèdent un TIR en 3' mais ne semble pas posséder de TIR en 5'. Comme TIGD1 chez les primates, cette séquence a également perdu un TIR (Kalitsis et Saffery, 2009). Il est donc probable qu'un **autre épisode de domestication d'un Tigger1** ait eu lieu chez un ancêtre commun aux **Carnivores**, à proximité du gène DUSP19. Afin de conforter cette hypothèse, il faudrait étudier ce locus dans les génomes de tous les Carnivores qui sont séquencés.

D'autres hits semblent suggérer un **autre épisode de domestication** ayant eu lieu chez un ancêtre commun aux **Cétartiodactyles**. En effet, des hits avec des protéines annotées « tigger transposable element-derived protein 1 » ou « protein1-like » chez le grand cachalot (*Physeter catodon*), le buffle d'eau (*Bubalus bubalis*), le bovin domestique (*Bos taurus*) et le zébu (*Bos indicus*) présentent la même synténie (proximité du gène CTDP1). Un TIR est présent en 3' chez le Buffle d'eau mais aucun TIR n'est repérable chez les autres espèces. Il faudrait de la même façon vérifier si ce locus contient le même type d'élément dans tous les génomes Cétartiodactyles qui ont été séquencés.

Un **épisode de domestication encore différent**, chez un ancêtre commun aux **Chiroptères**, est également suggéré par une autre synténie retrouvée chez des hits présents chez le renard volant noir (*Pteropus alecto*), la roussette d'Égypte (*Rousettus aegyptiacus*) et la chauve-souris chinoise en fer à cheval roux (*Rhinolophus sinicus*) (hit avec gènes supposés homologues à C15orf40 d'*H. sapiens* placés entre les gènes RAMAC et BTBD1 sur des scaffold non placés). Ces gènes supposés homologues ont été annotés chez toutes les classes de vertébrés dans la base de données NCBI. Dans le but d'étudier si l'insertion d'un Tigger1 dans ce locus a eu lieu chez un ancêtre commun aux Vertébrés, il serait intéressant de regarder s'ils possèdent tous une région homologue à ETBat. Dans le cas contraire il faudrait identifier à partir de quel taxon cette insertion est présente pour pouvoir dater ce supposé épisode de domestication (datation relative). De plus, *P. giganteus* appartient également aux Chiroptères ; il serait intéressant de regarder si la synténie est conservée, c'est-à-dire si l'élément ETBat est lui aussi placé dans un gène homologue à C15orf40 d'*H. sapiens* placé entre les gènes RAMAC et BTBD1 dans le génome de *P. giganteus*.

Un hit concerne un « tigger transposable element-derived protein1-like » chez le rhinocéros blanc du Sud (*Ceratotherium simum simum*). Cependant ce gène est porté par un scaffold non placé et les locus l'entourant n'ont pas été identifiés. Il faudrait étudier ces locus afin de savoir si on peut rattacher cet élément à une des synténies observées chez les autres Laurasiathériens (celle des Carnivores, celle des Cétartiodactyles ou celle des

Chiroptères) ou si elle peut évoquer un épisode de domestication encore différent, voire un autre évènement de transfert horizontal.

Un hit concerne le **4^{ème} intron du gène CRISP3** du **Cheval** (*Equus caballus*). Cette région du génome du cheval est très riche en ET (LINE, MIR, ERV et SINE). Ce résultat suggère qu'un ET très semblable à Tigger1 serait également inséré dans cette région.

L'acquisition de la version définitive du séquençage du génome de *P. giganteus* ainsi que des investigations ultérieures permettront de tester les hypothèses avancées dans notre étude.

CONCLUSION

Les ET occupent une grande proportion des génomes eucaryotes. En raison de leur mobilité et de leur abondance, ils sont responsables de la plasticité génomique à deux échelles de temps. A courte échelle, en entraînant des mutations et des remaniements génomiques délétères contre sélectionnés par la mort des individus (ou leur baisse de fitness). Et sur le long terme, lorsque ces remaniements présentent un avantage sélectif (voire sont neutres ou quasi neutres), pouvant aller de la modification de l'expression d'un gène jusqu'à l'exaptation. Les ET sont ainsi une source majeure d'évolution.

En pratique, lorsque ces séquences fortement répétées ne sont que trop peu divergentes, leur présence complique l'assemblage des génomes lors de leur séquençage.

Il est donc important d'annoter ces éléments, d'étudier leur phylogénie, et de ne pas les considérer seulement comme de l'ADN parasite.

La chauve-souris *P. giganteus* porte de façon asymptomatique des zoonoses dues à des *Henipavirus* entraînant des symptômes graves chez l'Homme. Les éléments responsables de cette résistance n'ont pas encore été identifiés.

Des recherches ont été entreprises sur cette espèce et son génome a récemment été séquencé. En plus de l'étude du génome ainsi que du transcriptome, un intérêt particulier a été porté à l'étude des ET de *P. giganteus*. Suite au séquençage préliminaire du génome, l'annotation de ces éléments facilitera son assemblage final. L'étude de ses ET pourrait également permettre de rechercher et d'identifier un ET proche d'un gène associé à la résistance au développement de signes cliniques chez cette espèce réservoir de zoonoses virales. Nos recherches se sont focalisées sur un élément remarquable en raison de sa forte homologie avec l'ET Tigger1 d'humain.

La séquence consensus de cet élément a été reconstruite et nommée ETBat. Elle porte les caractéristiques d'un Tigger1. L'étude de la phylogénie d'ETBat permet de suspecter que différents épisodes de domestication de l'élément Tigger1 aient eu lieu au cours de l'évolution. Elle permet également de suspecter des événements de transferts horizontaux, dont un concernant un Tigger1 ayant possiblement eu lieu entre *P. giganteus* et *H. sapiens* (voire entre des ancêtres de ces deux espèces), éventuellement pas le biais d'un vecteur.

Bibliographie

- ADAMS, J. W., KAUFMAN, R. E., KRETSCHMER, P. J., HARRISON, M. et NIENHUIS, A. W., 1980. A family of long reiterated DNA sequences, one copy of which is next to the human beta globin gene. In : *Nucleic Acids Research*. 20 décembre 1980. Vol. 8, n° 24, p. 6113-6128.
- AHN, M., CUI, J., IRVING, A. T. et WANG, L.-F., 2016. Unique Loss of the PYHIN Gene Family in Bats Amongst Mammals: Implications for Inflammasome Sensing. In : *Scientific Reports*. 24 février 2016. Vol. 6, p. 21722. DOI 10.1038/srep21722.
- ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. et LIPMAN, D. J., 1990. Basic local alignment search tool. In : *Journal of Molecular Biology*. 5 octobre 1990. Vol. 215, n° 3, p. 403-410. DOI 10.1016/S0022-2836(05)80360-2.
- ARENSBURGER, P., PIÉGU, B. et BIGOT, Y., 2016. The future of transposable element annotation and their classification in the light of functional genomics - what we can learn from the fables of Jean de la Fontaine ? In : *Mobile Genetic Elements*. 2016. Vol. 6, n° 6, p. e1256852. DOI 10.1080/2159256X.2016.1256852.
- ARKHIPOVA, I. R., PYATKOV, K. I., MESELSON, M. et EVGEN'EV, M. B., 2003. Retroelements containing introns in diverse invertebrate taxa. In : *Nature Genetics*. Février 2003. Vol. 33, n° 2, p. 123-124. DOI 10.1038/ng1074.
- ARKHIPOVA, I. R., YUSHENOVA, I. A. et RODRIGUEZ, F., 2013. Endonuclease-containing Penelope retrotransposons in the bdelloid rotifer *Adineta vaga* exhibit unusual structural features and play a role in expansion of host gene families. In : *Mobile DNA*. 27 août 2013. Vol. 4, n° 1, p. 19. DOI 10.1186/1759-8753-4-19.
- AVA, Australian Veterinary Association. *The Hendra vaccine* [en ligne]. Disponible à l'adresse : <https://www.ava.com.au/hendra-virus-vaccine> [consulté le 21 juillet 2018].
- BABU, A. et VERMA, R. S., 1987. Chromosome Structure: Euchromatin and Heterochromatin. In : BOURNE, G. H., JEON, K. W. et FRIEDLANDER, M. (éd.), *International Review of Cytology* [en ligne]. S.l. : Academic Press. p. 1-60. [Consulté le 27 septembre 2018]. Disponible à l'adresse : <http://www.sciencedirect.com/science/article/pii/S0074769608614357>.
- BAI, M., JANICIC, N., TRIVEDI, S., QUINN, S. J., COLE, D. E., BROWN, E. M. et HENDY, G. N., 1997. Markedly reduced activity of mutant calcium-sensing receptor with an inserted Alu element from a kindred with familial hypocalciuric hypercalcemia and neonatal severe hyperparathyroidism. In : *The Journal of Clinical Investigation*. 15 avril 1997. Vol. 99, n° 8, p. 1917-1925. DOI 10.1172/JCI119359.
- BANNERT, N. et KURTH, R., 2006. The evolutionary dynamics of human endogenous retroviral families. In : *Annual Review of Genomics and Human Genetics*. 2006. Vol. 7, p. 149-173. DOI 10.1146/annurev.genom.7.080505.115700.

BAO, W., KOJIMA, K. K. et KOHANY, O., 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. In : *Mobile DNA* [en ligne]. 2 juin 2015. Vol. 6. [Consulté le 30 septembre 2018]. DOI 10.1186/s13100-015-0041-9. Disponible à l'adresse : <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4455052/>.

BESTOR, T. H. et BOURC'HIS, D., 2004. Transposon silencing and imprint establishment in mammalian germ cells. In : *Cold Spring Harbor Symposia on Quantitative Biology*. 2004. Vol. 69, p. 381-387. DOI 10.1101/sqb.2004.69.381.

BIRD, A. P. et WOLFFE, A. P., 1999. Methylation-induced repression--belts, braces, and chromatin. In : *Cell*. 24 novembre 1999. Vol. 99, n° 5, p. 451-454.

BOSSART, K. N., TACHEDJIAN, M., MCEACHERN, J. A., CRAMERI, G., ZHU, Z., DIMITROV, D. S., BRODER, C. C. et WANG, L.-F., 2008. Functional studies of host-specific ephrin-B ligands as Henipavirus receptors. In : *Virology*. 15 mars 2008. Vol. 372, n° 2, p. 357-371. DOI 10.1016/j.virol.2007.11.011.

BOUCHET, N., 2009. *Mécanismes de transposition et de régulation de la transposase de l'élément mariner Mos1*. Thèse de doctorat en sciences de la Vie. Tours : Université François – Rabelais, 286p.

BOUVRETTE, L. P. B., 2010. *Analyse de la corrélation conditionnelle dérivée de la coévolution d'un système de trois gènes par un modèle du maximum de vraisemblance*. Master en bio-informatique. Montréal : Université de Montréal, 112p.

BRITTEN, R. J. et KOHNE, D. E., 1968. Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. In : *Science (New York, N.Y.)*. 9 août 1968. Vol. 161, n° 3841, p. 529-540.

BURNS, K. H., 2017. Transposable elements in cancer. In : *Nature Reviews. Cancer*. 2017. Vol. 17, n° 7, p. 415-424. DOI 10.1038/nrc.2017.35.

CAPPELLO, J., HANDELSMAN, K. et LODISH, H. F., 1985. Sequence of *Dictyostelium* DIRS-1: an apparent retrotransposon with inverted terminal repeats and an internal circle junction sequence. In : *Cell*. Novembre 1985. Vol. 43, n° 1, p. 105-115.

CAPY, P., GASPERI, G., BIÉMONT, C. et BAZIN, C., 2000. Stress and transposable elements: co-evolution or useful parasites? In : *Heredity*. Août 2000. Vol. 85 (Pt 2), p. 101-106.

CECH, T. R. et HEARST, J. E., 1976. Organization of highly repeated sequences in mouse main-band DNA. In : *Journal of Molecular Biology*. 25 janvier 1976. Vol. 100, n° 3, p. 227-256.

Centers for Disease Control and Prevention. Morbidity and Mortality Weekly Report. *Outbreak of Hendra-Like Virus -- Malaysia and Singapore, 1998-1999* [en ligne]. Disponible à l'adresse : <https://www.cdc.gov/mmwr/preview/mmwrhtml/00056866.htm> [consulté le 28 juillet 2018].

CHALVET, F., GRIMALDI, C., KAPER, F., LANGIN, T. et DABOUSSI, M.-J., 2003. Hop, an active Mutator-like element in the genome of the fungus *Fusarium oxysporum*. In : *Molecular Biology and Evolution*. Août 2003. Vol. 20, n° 8, p. 1362-1375. DOI 10.1093/molbev/msg155.

CHATURVEDI, S. K., SINGH, S., 2018. Status and distribution of Indian flying fox (*Pteropus giganteus*) in district Chitrakoot, Uttar Pradesh, India [en ligne] *International Journal of Zoology Studies*, Volume 3, pp 126-128. Disponible à l'adresse : www.zoologyjournals.com/download/302/3-2-22-665.pdf [consulté le 7 septembre 2018].

CHEN, S. et LI, X., 2008. Molecular characterization of the first intact Transib transposon from *Helicoverpa zea*. In : *Gene*. 31 janvier 2008. Vol. 408, n° 1-2, p. 51-63. DOI 10.1016/j.gene.2007.10.015.

CHUA, K. B., BELLINI, W. J., ROTA, P. A., HARCOURT, B. H., TAMIN, A., LAM, S. K., KSIAZEK, T. G., ROLLIN, P. E., ZAKI, S. R., SHIEH, W., GOLDSMITH, C. S., GUBLER, D. J., ROHRIG, J. T., EATON, B., GOULD, A. R., OLSON, J., FIELD, H., DANIELS, P., LING, A. E., PETERS, C. J., ANDERSON, L. J. et MAHY, B. W., 2000. Nipah virus: a recently emergent deadly paramyxovirus. In : *Science (New York, N.Y.)*. 26 mai 2000. Vol. 288, n° 5470, p. 1432-1435.

CHUNG, S., ZUKER, C. et LODISH, H. F., 1983. A repetitive and apparently transposable DNA sequence in *Dictyostelium discoideum* associated with developmentally regulated RNAs. In : *Nucleic Acids Research*. 25 juillet 1983. Vol. 11, n° 14, p. 4835-4852.

CHUONG, E. B., ELDE, N. C. et FESCHOTTE, C., 2016. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. In : *Science (New York, N.Y.)*. 4 mars 2016. Vol. 351, n° 6277, p. 1083-1087. DOI 10.1126/science.aad5497.

COOK, J. M., MARTIN, J., LEWIN, A., SINDEN, R. E. et TRISTEM, M., 2000. Systematic screening of *Anopheles* mosquito genomes yields evidence for a major clade of Pao-like retrotransposons. In : *Insect Molecular Biology*. Février 2000. Vol. 9, n° 1, p. 109-117.

CURCIO, M. J. et DERBYSHIRE, K. M., 2003. The outs and ins of transposition: from mu to kangaroo. In : *Nature Reviews. Molecular Cell Biology*. Novembre 2003. Vol. 4, n° 11, p. 865-877. DOI 10.1038/nrm1241.

DE CHASTONAY, Y., FELDER, H., LINK, C., AEBY, P., TOBLER, H. et MÜLLER, F., 1992. Unusual features of the retroid element PAT from the nematode *Panagrellus redivivus*. In : *Nucleic Acids Research*. 11 avril 1992. Vol. 20, n° 7, p. 1623-1628.

DEKA, M., MORSHED, N., DEKA, M. A. et MORSHED, N., 2018. Mapping Disease Transmission Risk of Nipah Virus in South and Southeast Asia. In : *Tropical Medicine and Infectious Disease*. 30 mai 2018. Vol. 3, n° 2, p. 57. DOI 10.3390/tropicalmed3020057.

DEMANGE, A., 2013. *Protéines à motif tripartite (TRIM) chez le porc (Sus scrofa) et répliation du rétrovirus endogène porcin*. Thèse de doctorat en Biologie. Rennes : Université de Rennes 1, 216p.

DEMARCO, R., VENANCIO, T. M. et VERJOVSKI-ALMEIDA, S., 2006. SmTRC1, a novel *Schistosoma mansoni* DNA transposon, discloses new families of animal and fungi transposons belonging to the CACTA superfamily. In : *BMC evolutionary biology*. 7 novembre 2006. Vol. 6, p. 89. DOI 10.1186/1471-2148-6-89.

DIMITRI, P. et JUNAKOVIC, N., 1999. Revising the selfish DNA hypothesis: new evidence on accumulation of transposable elements in heterochromatin. In : *Trends in genetics: TIG*. Avril 1999. Vol. 15, n° 4, p. 123-124.

ECOFECT. *Projet BAT-NIPATH* [en ligne]. Disponible à l'adresse : <http://ecofect.universite-lyon.fr/recherche/projet-bat-nipath-315482.kjsp> [consulté le 2 août 2018].

EICKBUSH, T. H., 2002. R2 and related site-specific non-long terminal repeat retrotransposons. In : CRAIG, N. L., CRAIGIE, R., GELLERT, M., LAMBOWITZ, A. M. (éd.) *Mobile DNA II*. Washington : ASM Press, p. 813-835.

EICKBUSH, T. H., MALIK, H. S., 2002. Origins and evolution of retrotransposons. In : CRAIG, N. L., CRAIGIE, R., GELLERT, M., LAMBOWITZ, A. M. (éd.) *Mobile DNA II*. Washington : ASM Press, p. 1111-1144.

ELANGOVA, V., MATHUR, V., KUMAR, M., PRIYA, Y. S., 2018. Diversity and conservation of chiropteran fauna. In : SIVAPERUMAN, C., VENKATARAMAN, K. (éd.) *Indian Hotspots: Vertebrate Faunal Diversity, Conservation and Management, Volume 1*. Singapore : Springer, p. 57-88.

ELLIOTT, T. A. et GREGORY, T. R., 2015. What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. In : *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*. 26 septembre 2015. Vol. 370, n° 1678, p. 20140331. DOI 10.1098/rstb.2014.0331.

ENCHÉRY, F. et HORVAT, B., 2017. Understanding the interaction between henipaviruses and their natural host, fruit bats: Paving the way toward control of highly lethal infection in humans. In : *International Reviews of Immunology*. Avril 2017. Vol. 36, n° 2, p. 108-121. DOI 10.1080/08830185.2016.1255883.

ESNAULT, C., MAESTRE, J. et HEIDMANN, T., 2000. Human LINE retrotransposons generate processed pseudogenes. In : *Nature Genetics*. Avril 2000. Vol. 24, n° 4, p. 363-367. DOI 10.1038/74184.

EVGEN'EV, M. B., ZELENTOVA, H., SHOSTAK, N., KOZITSINA, M., BARSKYI, V., LANKENAU, D. H. et CORCES, V. G., 1997. Penelope, a new family of transposable elements and its possible role in hybrid dysgenesis in *Drosophila virilis*. In : *Proceedings of the National Academy of Sciences of the United States of America*. 7 janvier 1997. Vol. 94, n° 1, p. 196-201.

FESCHOTTE, C., 2004. Merlin, a new superfamily of DNA transposons identified in diverse animal genomes and related to bacterial IS1016 insertion sequences. In : *Molecular Biology and Evolution*. Septembre 2004. Vol. 21, n° 9, p. 1769-1780. DOI 10.1093/molbev/msh188.

FIELD, H. E., BARRATT, P. C., HUGHES, R. J., SHIELD, J. et SULLIVAN, N. D., 2000. A fatal case of Hendra virus infection in a horse in north Queensland: clinical and epidemiological features. In : *Australian Veterinary Journal*. Avril 2000. Vol. 78, n° 4, p. 279-280.

FIELD, H., YOUNG, P., YOB, J. M., MILLS, J., HALL, L. et MACKENZIE, J., 2001. The natural history of Hendra and Nipah viruses. In : *Microbes and Infection*. Avril 2001. Vol. 3, n° 4, p. 307-314.

FINNEGAN, D. J., 1989. Eukaryotic transposable elements and genome evolution. In : *Trends in genetics: TIG*. Avril 1989. Vol. 5, n° 4, p. 103-107.

FINNEGAN, D. J., 1992. Transposable elements. In : *Current Opinion in Genetics & Development*. Décembre 1992. Vol. 2, n° 6, p. 861-867.

FINNEGAN, D. J., 2012. Retrotransposons. In : *Current biology: CB*. 5 juin 2012. Vol. 22, n° 11, p. R432-437. DOI 10.1016/j.cub.2012.04.025.

GANGULY, A., DUNBAR, T., CHEN, P., GODMILOW, L. et GANGULY, T., 2003. Exon skipping caused by an intronic insertion of a young Alu Yb9 element leads to severe hemophilia A. In : *Human Genetics*. Septembre 2003. Vol. 113, n° 4, p. 348-352. DOI 10.1007/s00439-003-0986-5.

GEORGE, J. A., TRAVERSE, K. L., DEBARYSHE, P. G., KELLEY, K. J. et PARDUE, M.-L., 2010. Evolution of diverse mechanisms for protecting chromosome ends by *Drosophila* TART telomere retrotransposons. In : *Proceedings of the National Academy of Sciences of the United States of America*. 7 décembre 2010. Vol. 107, n° 49, p. 21052-21057. DOI 10.1073/pnas.1015926107.

GILBERT, C. et FESCHOTTE, C., 2018. Horizontal acquisition of transposable elements and viral sequences: patterns and consequences. In : *Current Opinion in Genetics & Development*. Avril 2018. Vol. 49, p. 15-24. DOI 10.1016/j.gde.2018.02.007.

GOODWIN, T. J. et POULTER, R. T., 2001. The DIRS1 group of retrotransposons. In : *Molecular Biology and Evolution*. Novembre 2001. Vol. 18, n° 11, p. 2067-2082. DOI 10.1093/oxfordjournals.molbev.a003748.

GOODWIN, T. J. D., BUTLER, M. I. et POULTER, R. T. M., 2003. Cryptons: a group of tyrosine-recombinase-encoding DNA transposons from pathogenic fungi. In : *Microbiology (Reading, England)*. Novembre 2003. Vol. 149, n° Pt 11, p. 3099-3109. DOI 10.1099/mic.0.26529-0.

GOODWIN, T. J. D. et POULTER, R. T. M., 2004. A new group of tyrosine recombinase-encoding retrotransposons. In : *Molecular Biology and Evolution*. Avril 2004. Vol. 21, n° 4, p. 746-759. DOI 10.1093/molbev/msh072.

GOUBERT, C., MODOLO, L., VIEIRA, C., VALIENTEMORO, C., MAVINGUI, P. et BOULESTEIX, M., 2015. De novo assembly and annotation of the Asian tiger mosquito (*Aedes albopictus*) repeatome with dnaPipeTE from raw genomic reads and comparative analysis with the yellow fever mosquito (*Aedes aegypti*). In : *Genome Biology and Evolution*. 11 mars 2015. Vol. 7, n° 4, p. 1192-1205. DOI 10.1093/gbe/evv050.

GOUY, M., GUINDON, S. et GASCUEL, O., 2010. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. In : *Molecular Biology and Evolution*. Février 2010. Vol. 27, n° 2, p. 221-224. DOI 10.1093/molbev/msp259.

GRABHERR, M. G., HAAS, B. J., YASSOUR, M., LEVIN, J. Z., THOMPSON, D. A., AMIT, I., ADICONIS, X., FAN, L., RAYCHOWDHURY, R., ZENG, Q., CHEN, Z., MAUCELI, E., HACOEN, N., GNIRKE, A., RHIND, N., DI PALMA, F., BIRREN, B. W., NUSBAUM, C., LINDBLAD-TOH, K., FRIEDMAN, N. et REGEV, A., 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. In : *Nature biotechnology*. 15 mai 2011. Vol. 29, n° 7, p. 644-652. DOI 10.1038/nbt.1883.

GRABUNDZIJA, I., MESSING, S. A., THOMAS, J., COSBY, R. L., BILIC, I., MISKEY, C., GOGOL-DÖRING, A., KAPITONOV, V., DIEM, T., DALDA, A., JURKA, J., PRITHAM, E. J., DYDA, F., IZSVÁK, Z. et IVICS, Z., 2016. A Helitron transposon reconstructed from bats reveals a novel mechanism of genome shuffling in eukaryotes. In : *Nature Communications*. 2 mars 2016. Vol. 7, p. 10716. DOI 10.1038/ncomms10716.

GRAY, Y. H., 2000. It takes two transposons to tango: transposable-element-mediated chromosomal rearrangements. In : *Trends in genetics: TIG*. Octobre 2000. Vol. 16, n° 10, p. 461-468.

GRIFFITHS, A., MILLER, J., SUZUKI, D., LEWONTIN, R., GELBART, W. Les éléments génétiques transposables. In : (2002) *Introduction à l'analyse génétique*. Paris : De Boeck, p. 601-622.

HAHN, M. B., GURLEY, E. S., EPSTEIN, J. H., ISLAM, M. S., PATZ, J. A., DASZAK, P. et LUBY, S. P., 2014. The role of landscape composition and configuration on *Pteropus giganteus* roosting ecology and Nipah virus spillover risk in Bangladesh. In : *The American Journal of Tropical Medicine and Hygiene*. Février 2014. Vol. 90, n° 2, p. 247-255. DOI 10.4269/ajtmh.13-0256.

HALPIN, K., YOUNG, P. L., FIELD, H. E. et MACKENZIE, J. S., 2000. Isolation of Hendra virus from pteropid bats: a natural reservoir of Hendra virus. In : *The Journal of General Virology*. Août 2000. Vol. 81, n° Pt 8, p. 1927-1932. DOI 10.1099/0022-1317-81-8-1927.

HALPIN, K., HYATT, A. D., FOGARTY, R., MIDDLETON, D., BINGHAM, J., EPSTEIN, J. H., RAHMAN, S. A., HUGHES, T., SMITH, C., FIELD, H. E., DASZAK, P. et HENIPAVIRUS ECOLOGY RESEARCH GROUP, 2011. Pteropid bats are confirmed as the reservoir hosts of henipaviruses: a comprehensive experimental study of virus transmission. In : *The American Journal of Tropical Medicine and Hygiene*. Novembre 2011. Vol. 85, n° 5, p. 946-951. DOI 10.4269/ajtmh.2011.10-0567.

HARCOURT, B. H., LOWE, L., TAMIN, A., LIU, X., BANKAMP, B., BOWDEN, N., ROLLIN, P. E., COMER, J. A., KSIAZEK, T. G., HOSSAIN, M. J., GURLEY, E. S., BREIMAN, R. F., BELLINI, W. J. et ROTA, P. A., 2005. Genetic characterization of Nipah virus, Bangladesh, 2004. In : *Emerging Infectious Diseases*. Octobre 2005. Vol. 11, n° 10, p. 1594-1597. DOI 10.3201/eid1110.050513.

HAUSEN, P. et STEIN, H., 1970. Ribonuclease H. An enzyme degrading the RNA moiety of DNA-RNA hybrids. In : *European Journal of Biochemistry*. Juin 1970. Vol. 14, n° 2, p. 278-283.

HAYWARD, J. A., TACHEDJIAN, M., CUI, J., CHENG, A. Z., JOHNSON, A., BAKER, M. L., HARRIS, R. S., WANG, L.-F. et TACHEDJIAN, G., 2018. Differential Evolution of Antiretroviral Restriction Factors in Pteropid Bats as Revealed by APOBEC3 Gene Complexity. In : *Molecular Biology and Evolution*. 1 juillet 2018. Vol. 35, n° 7, p. 1626-1637. DOI 10.1093/molbev/msy048.

HAYWARD, J. A., TACHEDJIAN, M., CUI, J., FIELD, H., HOLMES, E. C., WANG, L.-F. et TACHEDJIAN, G., 2013. Identification of diverse full-length endogenous betaretroviruses in megabats and microbats. In : *Retrovirology*. 27 mars 2013. Vol. 10, n° 1, p. 35. DOI 10.1186/1742-4690-10-35.

HICKMAN, A. B. et DYDA, F., 2015. Mechanisms of DNA Transposition. In : *Microbiology Spectrum*. Avril 2015. Vol. 3, n° 2, p. MDNA3-0034-2014. DOI 10.1128/microbiolspec.MDNA3-0034-2014.

HICKMAN, A. B., CHANDLER, M. et DYDA, F., 2010. Integrating prokaryotes and eukaryotes: DNA transposases in light of structure. In : *Critical Reviews in Biochemistry and Molecular Biology*. Février 2010. Vol. 45, n° 1, p. 50-69. DOI 10.3109/10409230903505596.

HOUCK, M. A., CLARK, J. B., PETERSON, K. R. et KIDWELL, M. G., 1991. Possible horizontal transfer of *Drosophila* genes by the mite *Proctolaelaps regalis*. In : *Science (New York, N.Y.)*. 6 septembre 1991. Vol. 253, n° 5024, p. 1125-1128.

HSU, C.-H., ZHANG, Y., HARDISON, R. et MILLER, W., 2009. Whole-Genome Analysis of Gene Conversion Events. In : *Comparative Genomics* [en ligne]. S.l. : Springer, Berlin, Heidelberg. 27 septembre 2009. p. 181-192. [Consulté le 26 mars 2018]. Disponible à l'adresse : https://link.springer.com/chapter/10.1007/978-3-642-04744-2_15.

Integrated Taxonomic Information System. *Pteropus giganteus* (Brünnich, 1782) [en ligne]. Disponible à l'adresse : https://www.itis.gov/servlet/SingleRpt/SingleRpt?search_topic=TSN&search_value=631629#null [consulté le 7 octobre 2018].

JURKA, J. et KAPITONOV, V. V., 2001. PIFs meet Tourists and Harbingers: a superfamily reunion. In : *Proceedings of the National Academy of Sciences of the United States of America*. 23 octobre 2001. Vol. 98, n° 22, p. 12315-12316. DOI 10.1073/pnas.231490598.

JURKA, J., KAPITONOV, V. V., PAVLICEK, A., KLONOWSKI, P., KOHANY, O. et WALICHIEWICZ, J., 2005. Repbase Update, a database of eukaryotic repetitive elements. In : *Cytogenetic and Genome Research*. 2005. Vol. 110, n° 1-4, p. 462-467. DOI 10.1159/000084979.

KAJIKAWA, M. et OKADA, N., 2002. LINEs mobilize SINEs in the eel through a shared 3' sequence. In : *Cell*. 1 novembre 2002. Vol. 111, n° 3, p. 433-444.

KALITSIS, P. et SAFFERY, R., 2009. Inherent promoter bidirectionality facilitates maintenance of sequence integrity and transcription of parasitic DNA in mammalian genomes. In : *BMC genomics*. 27 octobre 2009. Vol. 10, p. 498. DOI 10.1186/1471-2164-10-498.

- KAPITONOV, V. V. et JURKA, J., 2001. Rolling-circle transposons in eukaryotes. In : *Proceedings of the National Academy of Sciences of the United States of America*. 17 juillet 2001. Vol. 98, n° 15, p. 8714-8719. DOI 10.1073/pnas.151269298.
- KAPITONOV, V. V. et JURKA, J., 2004. Harbinger transposons and an ancient HARBI1 gene derived from a transposase. In : *DNA and cell biology*. Mai 2004. Vol. 23, n° 5, p. 311-324. DOI 10.1089/104454904323090949.
- KAPITONOV, V. V. et JURKA, J., 2005. RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. In : *PLoS biology*. Juin 2005. Vol. 3, n° 6, p. e181. DOI 10.1371/journal.pbio.0030181.
- KAPITONOV, V. V. et JURKA, J., 2006. Self-synthesizing DNA transposons in eukaryotes. In : *Proceedings of the National Academy of Sciences of the United States of America*. 21 mars 2006. Vol. 103, n° 12, p. 4540-4545. DOI 10.1073/pnas.0600833103.
- KAPITONOV, V. V. et JURKA, J., 2007. Helitrons on a roll: eukaryotic rolling-circle transposons. In : *Trends in genetics: TIG*. Octobre 2007. Vol. 23, n° 10, p. 521-529. DOI 10.1016/j.tig.2007.08.004.
- KAPITONOV, V. V. et JURKA, J., 2008. A universal classification of eukaryotic transposable elements implemented in Repbase. In : *Nature Reviews. Genetics*. 2008. Vol. 9, n° 5, p. 411-412 ; author reply 414. DOI 10.1038/nrg2165-c1.
- KAPUSTIN, Y., CHAN, E., SARKAR, R., WONG, F., VORECHOVSKY, I., WINSTON, R. M., TATUSOVA, T. et DIBB, N. J., 2011. Cryptic splice sites and split genes. In : *Nucleic Acids Research*. Août 2011. Vol. 39, n° 14, p. 5837-5844. DOI 10.1093/nar/gkr203.
- KASS, L. B. et CHOMET, P., 2009. Barbara McClintock. In : BENNETZEN, Jeffrey L. et HAKE, Sarah (éd.), *Handbook of Maize: Genetics and Genomics* [en ligne]. New York, NY : Springer New York. p. 17-52. [Consulté le 23 septembre 2018]. ISBN 978-0-387-77863-1. Disponible à l'adresse : https://doi.org/10.1007/978-0-387-77863-1_2.
- KATAOKA, M., AIMI, Y., YANAGISAWA, R., ONO, M., OKA, A., FUKUDA, K., YOSHINO, H., SATOH, T. et GAMOU, S., 2013. Alu-mediated nonallelic homologous and nonhomologous recombination in the BMPR2 gene in heritable pulmonary arterial hypertension. In : *Genetics in Medicine: Official Journal of the American College of Medical Genetics*. Décembre 2013. Vol. 15, n° 12, p. 941-947. DOI 10.1038/gim.2013.41.
- KATZOURAKIS, A. et ASWAD, A., 2016. Evolution: Endogenous Viruses Provide Shortcuts in Antiviral Immunity. In : *Current biology: CB*. 23 mai 2016. Vol. 26, n° 10, p. R427-429. DOI 10.1016/j.cub.2016.03.072.
- KAZAZIAN, H. H., 2004. Mobile elements: drivers of genome evolution. In : *Science (New York, N.Y.)*. 12 mars 2004. Vol. 303, n° 5664, p. 1626-1632. DOI 10.1126/science.1089670.
- KEMPKEN, F. et WINDHOFER, F., 2001. The hAT family: a versatile transposon group common to plants, fungi, animals, and man. In : *Chromosoma*. Avril 2001. Vol. 110, n° 1, p. 1-9.

KIDWELL, M. G., 1992. Horizontal transfer of P elements and other short inverted repeat transposons. In : *Genetica*. 1992. Vol. 86, n° 1-3, p. 275-286.

KIDWELL, M. G., LISCH, D. R., 2002. Transposable elements as source of genomic variation. In : CRAIG, N. L., CRAIGIE, R., GELLERT, M., LAMBOWITZ, A. M. (éd.) *Mobile DNA II*. Washington : ASM Press, p. 59-90.

KIM, S., CHO, C.-S., HAN, K. et LEE, J., 2016. Structural Variation of Alu Element and Human Disease. In : *Genomics & Informatics*. Septembre 2016. Vol. 14, n° 3, p. 70-77. DOI 10.5808/GI.2016.14.3.70.

KOHANY, O., GENTLES, A. J., HANKUS, L. et JURKA, J., 2006. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. In : *BMC bioinformatics*. 25 octobre 2006. Vol. 7, p. 474. DOI 10.1186/1471-2105-7-474.

KOJIMA, K. K. et JURKA, J., 2011. Crypton transposons: identification of new diverse families and ancient domestication events. In : *Mobile DNA*. 19 octobre 2011. Vol. 2, n° 1, p. 12. DOI 10.1186/1759-8753-2-12.

KRAHN, M., 2009. Master de pathologie humaine. *Site de la Faculté de Médecine de Marseille* [en ligne], Disponible à l'adresse : http://www.master-pathologie-humaine.org/IMG/pdf/M_KRAHN_Organisation_Genome_et_Analyses_genomiques_2009.pdf [consulté le 23 septembre 2018].

KRAMEROV, D. A. et VASSETZKY, N. S., 2005. Short retrotransposons in eukaryotic genomes. In : *International Review of Cytology*. 2005. Vol. 247, p. 165-221. DOI 10.1016/S0074-7696(05)47004-7.

KRUPOVIC, M. et KOONIN, E. V., 2015. Polintons: a hotbed of eukaryotic virus, transposon and plasmid evolution. In : *Nature Reviews. Microbiology*. 2015. Vol. 13, n° 2, p. 105-115. DOI 10.1038/nrmicro3389.

LE DUC, D., RENAUD, G., KRISHNAN, A., ALMÉN, M. S., HUYNEN, L., PROHASKA, S. J., ONGYERTH, M., BITARELLO, B. D., SCHIÖTH, H. B., HOFREITER, M., STADLER, P. F., PRÜFER, K., LAMBERT, D., KELSO, J. et SCHÖNEBERG, T., 2015. Kiwi genome provides insights into evolution of a nocturnal lifestyle. In : *Genome Biology*. 23 juillet 2015. Vol. 16, p. 147. DOI 10.1186/s13059-015-0711-4.

Legifrance (JORF n°0187 du 13 août 2013, page 13832, texte n° 112). *Arrêté du 29 juillet 2013 relatif à la définition des dangers sanitaires de première et deuxième catégorie pour les espèces animales* [en ligne]. Disponible à l'adresse : <https://www.legifrance.gouv.fr/eli/arrete/2013/7/29/AGRG1320208A/jo/texte> [consulté le 11 octobre 2018].

LERAT, E., 2010. Identifying repeats and transposable elements in sequenced genomes: how to find your way through the dense forest of programs. In : *Heredity*. Juin 2010. Vol. 104, n° 6, p. 520-533. DOI 10.1038/hdy.2009.165.

LERAT, E., 2001. *Comparaison de séquences d'éléments transposables et de gènes d'hôte chez cinq espèces : A. thaliana, C. elegans, D. melanogaster, H. sapiens et S. cerevisiae* [en ligne]. Thèse de doctorat. S.l. : Université Claude Bernard - Lyon I. [Consulté le 20 janvier 2018]. Disponible à l'adresse : <https://tel.archives-ouvertes.fr/tel-00005207/document>.

LEVIN, H. L. et MORAN, J. V., 2011. Dynamic interactions between transposable elements and their hosts. In : *Nature Reviews. Genetics*. 18 août 2011. Vol. 12, n° 9, p. 615-627. DOI 10.1038/nrg3030.

LISCH, D., 2002. Mutator transposons. In : *Trends in Plant Science*. Novembre 2002. Vol. 7, n° 11, p. 498-504.

LORENZI, H. A., ROBLEDO, G. et LEVIN, M. J., 2006. The VIPER elements of trypanosomes constitute a novel group of tyrosine recombinase-encoding retrotransposons. In : *Molecular and Biochemical Parasitology*. Février 2006. Vol. 145, n° 2, p. 184-194. DOI 10.1016/j.molbiopara.2005.10.002.

LORETO, E. L. S., CARARETO, C. M. A. et CAPY, P., 2008. Revisiting horizontal transfer of transposable elements in *Drosophila*. In : *Heredity*. Juin 2008. Vol. 100, n° 6, p. 545-554. DOI 10.1038/sj.hdy.6801094.

LOUIS, A., NGUYEN, N. T. T., MUFFATO, M. et ROEST CROLLIUS, H., 2015. Genomic update 2015: KaryoView and MatrixView provide a genome-wide perspective to multispecies comparative genomics. In : *Nucleic Acids Research*. Janvier 2015. Vol. 43, n° Database issue, p. D682-689. DOI 10.1093/nar/gku1112.

LU, C., CHEN, J., ZHANG, Y., HU, Q., SU, W. et KUANG, H., 2012. Miniature Inverted-Repeat Transposable Elements (MITEs) Have Been Accumulated through Amplification Bursts and Play Important Roles in Gene Expression and Species Diversity in *Oryza sativa*. In : *Molecular Biology and Evolution*. 1 mars 2012. Vol. 29, n° 3, p. 1005-1017. DOI 10.1093/molbev/msr282.

LUAN, D. D., KORMAN, M. H., JAKUBCZAK, J. L. et EICKBUSH, T. H., 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. In : *Cell*. 26 février 1993. Vol. 72, n° 4, p. 595-605.

LYOZIN, G. T., MAKAROVA, K. S., VELIKODVORSKAJA, V. V., ZELENTSOVA, H. S., KHECHUMIAN, R. R., KIDWELL, M. G., KOONIN, E. V. et EVGEN'EV, M. B., 2001. The structure and evolution of Penelope in the virilis species group of *Drosophila*: an ancient lineage of retroelements. In : *Journal of Molecular Evolution*. Mai 2001. Vol. 52, n° 5, p. 445-456. DOI 10.1007/s002390010174.

MAGWIRE, M. M., BAYER, F., WEBSTER, C. L., CAO, C. et JIGGINS, F. M., 2011. Successive increases in the resistance of *Drosophila* to viral infection through a transposon insertion followed by a duplication. In : *PLoS genetics*. Octobre 2011. Vol. 7, n° 10, p. e1002337. DOI 10.1371/journal.pgen.1002337.

- MAHMOOD-UL-HASSAN, M., GULRAIZ, T. L., RANA, S. A. et JAVID, A., 2010. The Diet of Indian Flying-Foxes (*Pteropus giganteus*) in Urban Habitats of Pakistan. In : *Acta Chiropterologica*. 1 décembre 2010. Vol. 12, n° 2, p. 341-347. DOI 10.3161/150811010X537927.
- MAJUMDAR, S. et RIO, D. C., 2015. P Transposable Elements in *Drosophila* and other Eukaryotic Organisms. In : *Microbiology Spectrum*. Avril 2015. Vol. 3, n° 2, p. MDNA3-0004-2014. DOI 10.1128/microbiolspec.MDNA3-0004-2014.
- MALIK, H. S., BURKE, W. D. et EICKBUSH, T. H., 1999. The age and evolution of non-LTR retrotransposable elements. In : *Molecular Biology and Evolution*. Juin 1999. Vol. 16, n° 6, p. 793-805. DOI 10.1093/oxfordjournals.molbev.a026164.
- MALIK, H. S., HENIKOFF, S. et EICKBUSH, T. H., 2000. Poised for contagion: evolutionary origins of the infectious abilities of invertebrate retroviruses. In : *Genome Research*. Septembre 2000. Vol. 10, n° 9, p. 1307-1318.
- MARQUEZ, C. P. et PRITHAM, E. J., 2010. Phantom, a new subclass of Mutator DNA transposons found in insect viruses and widely distributed in animals. In : *Genetics*. Août 2010. Vol. 185, n° 4, p. 1507-1517. DOI 10.1534/genetics.110.116673.
- MARTIN, S. L., 2006. The ORF1 protein encoded by LINE-1: structure and function during L1 retrotransposition. In : *Journal of Biomedicine & Biotechnology*. 2006. Vol. 2006, n° 1, p. 45621. DOI 10.1155/JBB/2006/45621.
- MI, S., LEE, X., LI, X., VELDMAN, G. M., FINNERTY, H., RACIE, L., LAVALLIE, E., TANG, X. Y., EDOUARD, P., HOWES, S., KEITH, J. C. et MCCOY, J. M., 2000. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. In : *Nature*. 17 février 2000. Vol. 403, n° 6771, p. 785-789. DOI 10.1038/35001608.
- MICKLEBURGH, S., WAYLEN, K. et RACEY, P., 2009. Bats as bushmeat: a global review. In : *Oryx*. Avril 2009. Vol. 43, n° 2, p. 217-234. DOI 10.1017/S0030605308000938.
- MILLER, W. J., HAGEMANN, S., REITER, E. et PINSKER, W., 1992. P-element homologous sequences are tandemly repeated in the genome of *Drosophila guanache*. In : *Proceedings of the National Academy of Sciences of the United States of America*. 1 mai 1992. Vol. 89, n° 9, p. 4018-4022.
- MLÍKOVSKÝ, J., 2012. Correct name for the Indian flying fox (Pteropodidae). In : *Vespertilio*. 2012. Vol. 16, p. 203-204.
- MOLUR, S., SRINIVASULU, C., BATES, P. & FRANCIS, C., 2008. *Pteropus giganteus*. *The IUCN Red List of Threatened Species 2008* [en ligne], Disponible à l'adresse : <http://dx.doi.org/10.2305/IUCN.UK.2008.RLTS.T18725A8511108.en> [consulté le 7 septembre 2018].
- MORRISH, T. A., GILBERT, N., MYERS, J. S., VINCENT, B. J., STAMATO, T. D., TACCIOLI, G. E., BATZER, M. A. et MORAN, J. V., 2002. DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. In : *Nature Genetics*. Juin 2002. Vol. 31, n° 2, p. 159-165. DOI 10.1038/ng898.

MURRAY, K., SELLECK, P., HOOPER, P., HYATT, A., GOULD, A., GLEESON, L., WESTBURY, H., HILEY, L., SELVEY, L. et RODWELL, B., 1995. A morbillivirus that caused fatal disease in horses and humans. In : *Science (New York, N.Y.)*. 7 avril 1995. Vol. 268, n° 5207, p. 94-97.

NANJUNDIAH, V., 1996. Barbara McClintock and the discovery of jumping genes. In : *Resonance*. 1 octobre 1996. Vol. 1, n° 10, p. 56-62. DOI 10.1007/BF02839098.

NCBI. *BLAST FAQs* [en ligne]. Disponible à l'adresse : https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=FAQ [consulté le 25 août 2018].

NEFEDOVA, L. et KIM, A., 2017. Mechanisms of LTR-Retroelement Transposition: Lessons from *Drosophila melanogaster*. In : *Viruses*. 16 avril 2017. Vol. 9, n° 4. DOI 10.3390/v9040081.

NOUROZ, F., NOREEN, S. et HARRISON, J. S. H., 2017. Identification and evolutionary dynamics of cacta DNA transposons in brassica. In : *Pakistan Journal of Botany*. 2017. Vol. 49, n° 2, p. 789-798.

NOUROZ, F., NOREEN, S. et HESLOP-HARRISON, J. S. (Pat), 2016. Characterization and Diversity of Novel PIF/Harbinger DNA Transposons in Brassica Genomes. In : [en ligne]. 19 février 2016. [Consulté le 27 septembre 2018]. Disponible à l'adresse : <https://lra.le.ac.uk/handle/2381/37674>.

OIE, Organisation Mondiale de la Santé animale. *Code sanitaire pour les animaux terrestres, Volume 1, Dispositions générales, Dix-neuvième édition, 2010* [en ligne]. Disponible à l'adresse : <https://www.oie.int/doc/ged/D7598.PDF> [consulté le 28 juillet 2018].

OIE, Organisation Mondiale de la Santé animale. *Virus Nipah* [en ligne]. Disponible à l'adresse : <http://www.oie.int/fr/sante-animale-dans-le-monde/maladies-animales/virus-nipah/> [consulté le 21 juillet 2018].

OIE Terrestrial Manual 2015. *Chapter 2.1.14 Nipah and Hendra virus diseases* [en ligne]. Disponible à l'adresse : http://www.oie.int/fileadmin/Home/eng/Health_standards/tahm/2.01.14_NIPAH_HENDRA.pdf [consulté le 29 juillet 2018].

OMS, Organisation Mondiale de la Santé. *Infection à virus Hendra* [en ligne]. Disponible à l'adresse : <http://www.who.int/csr/disease/hendra/fr/> [consulté le 21 juillet 2018].

OMS, Organisation Mondiale de la Santé. *Virus Nipah* [en ligne]. Disponible à l'adresse : <http://www.who.int/fr/news-room/fact-sheets/detail/nipah-virus> [consulté le 21 juillet 2018].

ORGEL, L. E. et CRICK, F. H. C., 1980. Selfish DNA: the ultimate parasite. In : *Nature*. Avril 1980. Vol. 284, n° 5757, p. 604-607. DOI 10.1038/284604a0.

- PACE, J. K., GILBERT, C., CLARK, M. S. et FESCHOTTE, C., 2008. Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. In : *Proceedings of the National Academy of Sciences of the United States of America*. 4 novembre 2008. Vol. 105, n° 44, p. 17023-17028. DOI 10.1073/pnas.0806548105.
- PAGEL, M. et JOHNSTONE, R. A., 1992. Variation across species in the size of the nuclear genome supports the junk-DNA explanation for the C-value paradox. In : *Proceedings. Biological Sciences*. 22 août 1992. Vol. 249, n° 1325, p. 119-124. DOI 10.1098/rspb.1992.0093.
- PARDUE, M.-L. et DEBARYSHE, P. G., 2003. Retrotransposons provide an evolutionarily robust non-telomerase mechanism to maintain telomeres. In : *Annual Review of Genetics*. 2003. Vol. 37, p. 485-511. DOI 10.1146/annurev.genet.38.072902.093115.
- PENG, Y., LEUNG, H. C. M., YIU S. M., CHIN, F. Y. L., 2010. IDBA – A Practical Iterative de Bruijn Graph De Novo Assembler. In : *Proceedings of Research in Computational Molecular Biology, 14th Annual International Conference, RECOMB 2010, April 25-28, 2010, Lisbon, Portugal*. Springer, Berlin, 579p.
- PEREIRA, J. F., ALMEIDA, A. P. M. M., COTA, J., PAMPHILE, J. A., DA SILVA, G. F., DE ARAÚJO, E. F., GRAMACHO, K. P., BROMMONSCHENKEL, S. H., PEREIRA, G. A. G. et DE QUEIROZ, M. V., 2013. Boto, a class II transposon in *Moniliophthora perniciosa*, is the first representative of the PIF/Harbinger superfamily in a phytopathogenic fungus. In : *Microbiology (Reading, England)*. Janvier 2013. Vol. 159, n° Pt 1, p. 112-125. DOI 10.1099/mic.0.062901-0.
- PETERLONGO, P., 2016. *Lire les lectures : analyse de données de séquençage*. Habilitation à Diriger des Recherches. Rennes : Université de Rennes 1, 118p.
- PIÉGU, B., BIRE, S., ARENSBURGER, P. et BIGOT, Y., 2015. A survey of transposable element classification systems--a call for a fundamental update to meet the challenge of their diversity and complexity. In : *Molecular Phylogenetics and Evolution*. Mai 2015. Vol. 86, p. 90-109. DOI 10.1016/j.ympev.2015.03.009.
- PIMPINELLI, S., BERLOCO, M., FANTI, L., DIMITRI, P., BONACCORSI, S., MARCHETTI, E., CAZZI, R., CAGGESE, C. et GATTI, M., 1995. Transposable elements are stable structural components of *Drosophila melanogaster* heterochromatin. In : *Proceedings of the National Academy of Sciences of the United States of America*. 25 avril 1995. Vol. 92, n° 9, p. 3804-3808.
- PISKUREK, O. et OKADA, N., 2007. Poxviruses as possible vectors for horizontal transfer of retrotransposons from reptiles to mammals. In : *Proceedings of the National Academy of Sciences of the United States of America*. 17 juillet 2007. Vol. 104, n° 29, p. 12046-12051. DOI 10.1073/pnas.0700531104.
- PLASTERK, R. H. A., VAN LUENEN, H. G. A. M., 2002. The Tc1/mariner family of transposable elements. In : CRAIG, N. L., CRAIGIE, R., GELLERT, M., LAMBOWITZ, A. M. (éd.) *Mobile DNA II*. Washington : ASM Press, p. 512-532.

- PLATT, R. N., MANGUM, S. F. et RAY, D. A., 2016. Pinpointing the vesper bat transposon revolution using the *Miniopterus natalensis* genome. In : *Mobile DNA*. 2016. Vol. 7, p. 12. DOI 10.1186/s13100-016-0071-y.
- POULTER, R. T. M. et GOODWIN, T. J. D., 2005. DIRS-1 and the other tyrosine recombinase retrotransposons. In : *Cytogenetic and Genome Research*. 2005. Vol. 110, n° 1-4, p. 575-588. DOI 10.1159/000084991.
- POULTER, R. T. M. et BUTLER, M. I., 2015. Tyrosine Recombinase Retrotransposons and Transposons. In : *Microbiology Spectrum*. avril 2015. Vol. 3, n° 2, p. MDNA3- 0036-2014. DOI 10.1128/microbiolspec.MDNA3-0036-2014.
- PRIIMÄGI, A. F., MIZROKHI, L. J. et ILYIN, Y. V., 1988. The *Drosophila* mobile element jockey belongs to LINEs and contains coding sequences homologous to some retroviral proteins. In : *Gene*. 30 octobre 1988. Vol. 70, n° 2, p. 253-262.
- PRITHAM, E. J. et FESCHOTTE, C., 2007. Massive amplification of rolling-circle transposons in the lineage of the bat *Myotis lucifugus*. In : *Proceedings of the National Academy of Sciences of the United States of America*. 6 février 2007. Vol. 104, n° 6, p. 1895-1900. DOI 10.1073/pnas.0609601104.
- PRITHAM, E. J., FESCHOTTE, C. et WESSLER, S. R., 2005. Unexpected diversity and differential success of DNA transposons in four species of entamoeba protozoans. In : *Molecular Biology and Evolution*. Septembre 2005. Vol. 22, n° 9, p. 1751-1763. DOI 10.1093/molbev/msi169.
- PYATKOV, K. I., ARKHIPOVA, I. R., MALKOVA, N. V., FINNEGAN, D. J. et EVGEN'EV, M. B., 2004. Reverse transcriptase and endonuclease activities encoded by Penelope-like retroelements. In : *Proceedings of the National Academy of Sciences of the United States of America*. 12 octobre 2004. Vol. 101, n° 41, p. 14719-14724. DOI 10.1073/pnas.0406281101.
- RAVINDRAN, S., 2012. Barbara McClintock and the discovery of jumping genes. In : *Proceedings of the National Academy of Sciences*. 11 décembre 2012. Vol. 109, n° 50, p. 20198-20199. DOI 10.1073/pnas.1219372109.
- RAY, D. A., FESCHOTTE, C., PAGAN, H. J. T., SMITH, J. D., PRITHAM, E. J., ARENSBURGER, P., ATKINSON, P. W. et CRAIG, N. L., 2008. Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*. In : *Genome Research*. Mai 2008. Vol. 18, n° 5, p. 717-728. DOI 10.1101/gr.071886.107.
- RAY, D. A., PAGAN, H. J., PLATT, R. N., KROLL, A. R., SCHAACK, S. et STEVENS, R. D., 2015. Differential SINE evolution in vesper and non-vesper bats. In : *Mobile DNA*. 2015. Vol. 6, p. 10. DOI 10.1186/s13100-015-0038-4.
- RIO, D. C. P transposable elements in *Drosophila melanogaster*., 2002. In : CRAIG, N. L., CRAIGIE, R., GELLERT, M., LAMBOWITZ, A. M. (éd.) *Mobile DNA II*. Washington : ASM Press, p. 484-518.

- ROBERTSON, H. M. et LAMPE, D. J., 1995. Recent horizontal transfer of a mariner transposable element among and between Diptera and Neuroptera. In : *Molecular Biology and Evolution*. Septembre 1995. Vol. 12, n° 5, p. 850-862. DOI 10.1093/oxfordjournals.molbev.a040262.
- RODRIGUES, A., ALVES, P. M., et COROADINHA, A., 2011. Production of Retroviral and Lentiviral Gene Therapy Vectors: Challenges in the Manufacturing of Lipid Enveloped Virus. In : XU, K. (dir.). *Viral Gene Therapy* [en ligne]. 20 juillet 2011. Chapitre 2. IntechOpen, DOI: 10.5772/18615. Disponible à l'adresse : <https://www.intechopen.com/books/viral-gene-therapy/production-of-retroviral-and-lentiviral-gene-therapy-vectors-challenges-in-the-manufacturing-of-lipi> [consulté le 26 septembre 2018].
- ROTA P. A., LO M. K. Molecular Virology of the Henipaviruses., 2012. In : LEE B., ROTA P. A. (éd.) *Henipavirus. Ecology, Molecular Virology, and Pathogenesis*. Berlin : Springer, pp 41-58.
- RUIZ-PÉREZ, V. L., MURILLO, F. J. et TORRES-MARTÍNEZ, S., 1996. Prt1, an unusual retrotransposon-like sequence in the fungus *Phycomyces blakesleeanus*. In : *Molecular & general genetics: MGG*. 13 décembre 1996. Vol. 253, n° 3, p. 324-333.
- SCHAACK, S., GILBERT, C. et FESCHOTTE, C., 2010. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. In : *Trends in Ecology & Evolution*. Septembre 2010. Vol. 25, n° 9, p. 537-546. DOI 10.1016/j.tree.2010.06.001.
- SILVA, J. C., LORETO, E. L. et CLARK, J. B., 2004. Factors that affect the horizontal transfer of transposable elements. In : *Current Issues in Molecular Biology*. Janvier 2004. Vol. 6, n° 1, p. 57-71.
- SINGER, M. F., 1982. SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes. In : *Cell*. Mars 1982. Vol. 28, n° 3, p. 433-434.
- SINGER, M. F. et SKOWRONSKI, J., 1985. Making sense out of LINES: long interspersed repeat sequences in mammalian genomes. In : *Trends in Biochemical Sciences*. 1 mars 1985. Vol. 10, n° 3, p. 119-122. DOI 10.1016/0968-0004(85)90271-3.
- SINGH-CUNDY, A., SHIN, G., 2017. L'ADN et les gènes. In : *Découvrir la biologie*. Paris : De Boeck Supérieur, p. 240-259.
- SLOTKIN, R. K. et MARTIENSSSEN, R., 2007. Transposable elements and the epigenetic regulation of the genome. In : *Nature Reviews. Genetics*. Avril 2007. Vol. 8, n° 4, p. 272-285. DOI 10.1038/nrg2072.
- SMIT, A. F. et RIGGS, A. D., 1996. Tiggers and DNA transposon fossils in the human genome. In : *Proceedings of the National Academy of Sciences of the United States of America*. 20 février 1996. Vol. 93, n° 4, p. 1443-1448.

Société Française de BioInformatique. *Scaffolding du génome Pteropus giganteus, réservoir de nombreux virus hautement pathogènes* [en ligne]. Disponible à l'adresse : <https://www.sfbi.fr/content/scaffolding-du-g%C3%A9nome-pteropus-giganteus-r%C3%A9servoir-de-nombreux-virus-hautement-pathog%C3%A8nes> [consulté le 2 août 2018].

SONG, L., FLOREA, L. et LANGMEAD, B., 2014. Lighter: fast and memory-efficient sequencing error correction without counting. In : *Genome Biology*. 2014. Vol. 15, n° 11, p. 509. DOI 10.1186/s13059-014-0509-9.

SOTERO-CAIO, C. G., PLATT, R. N., SUH, A. et RAY, D. A., 2017. Evolution and Diversity of Transposable Elements in Vertebrate Genomes. In : *Genome Biology and Evolution*. 1 janvier 2017. Vol. 9, n° 1, p. 161-177. DOI 10.1093/gbe/evw264.

SRIKANTA, D., SEN, S. K., HUANG, C. T., CONLIN, E. M., RHODES, R. M. et BATZER, M. A., 2009. An alternative pathway for Alu retrotransposition suggests a role in DNA double-strand break repair. In : *Genomics*. Mars 2009. Vol. 93, n° 3, p. 205-212. DOI 10.1016/j.ygeno.2008.09.016.

STROUT, M. P., MARCUCCI, G., BLOOMFIELD, C. D. et CALIGIURI, M. A., 1998. The partial tandem duplication of ALL1 (MLL) is consistently generated by Alu-mediated homologous recombination in acute myeloid leukemia. In : *Proceedings of the National Academy of Sciences of the United States of America*. 3 mars 1998. Vol. 95, n° 5, p. 2390-2395.

SYVANEN, M., 1994. Horizontal gene transfer: evidence and possible consequences. In : *Annual Review of Genetics*. 1994. Vol. 28, p. 237-261. DOI 10.1146/annurev.ge.28.120194.001321.

TANG, H., 2007. Genome assembly, rearrangement, and repeats. In : *Chemical Reviews*. Août 2007. Vol. 107, n° 8, p. 3391-3406. DOI 10.1021/cr0683008.

TANG, J. et WONG, R. N., 1987. Evolution in the structure and function of aspartic proteases. In : *Journal of Cellular Biochemistry*. Janvier 1987. Vol. 33, n° 1, p. 53-63. DOI 10.1002/jcb.240330106.

TANG, Z., ZHANG, H.-H., HUANG, K., ZHANG, X.-G., HAN, M.-J. et ZHANG, Z., 2015. Repeated horizontal transfers of four DNA transposons in invertebrates and bats. In : *Mobile DNA*. 2015. Vol. 6, n° 1, p. 3. DOI 10.1186/s13100-014-0033-1.

THOMAS, C. A., 1971. The genetic organization of chromosomes. In : *Annual Review of Genetics*. 1971. Vol. 5, p. 237-256. DOI 10.1146/annurev.ge.05.120171.001321.

THOMAS, J. et PRITHAM, E. J., 2015. Helitrons, the Eukaryotic Rolling-circle Transposable Elements. In : *Microbiology Spectrum*. Août 2015. Vol. 3, n° 4. DOI 10.1128/microbiolspec.MDNA3-0049-2014.

VAN DIJK, E. L., AUGER, H., JASZCZYSZYN, Y. et THERMES, C., 2014. Ten years of next-generation sequencing technology. In : *Trends in genetics: TIG*. Septembre 2014. Vol. 30, n° 9, p. 418-426. DOI 10.1016/j.tig.2014.07.001.

VAZQUEZ, M., BEN-DOV, C., LORENZI, H., MOORE, T., SCHIJMAN, A. et LEVIN, M. J., 2000. The short interspersed repetitive element of *Trypanosoma cruzi*, SIRE, is part of VIPER, an unusual retroelement related to long terminal repeat retrotransposons. In : *Proceedings of the National Academy of Sciences of the United States of America*. 29 février 2000. Vol. 97, n° 5, p. 2128-2133. DOI 10.1073/pnas.050578397.

ViroScan3D. *Projet BAT-NIPATH* [en ligne]. Disponible à l'adresse : <http://www.viroscan3d.com/solutionsrd/projets-de-rd/> [consulté le 2 août 2018].

VOLFF, J.-N., 2006. Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. In : *BioEssays*. 1 septembre 2006. Vol. 28, n° 9, p. 913-922. DOI 10.1002/bies.20452.

WALLAU, G. L., VIEIRA, C. et LORETO, É. L. S., 2018. Genetic exchange in eukaryotes through horizontal transfer: connected by the mobilome. In : *Mobile DNA*. 2018. Vol. 9, p. 6. DOI 10.1186/s13100-018-0112-9.

WANG, L.-F., WALKER, P. J. et POON, L. L. M., 2011. Mass extinctions, biodiversity and mitochondrial function: are bats « special » as reservoirs for emerging viruses? In : *Current Opinion in Virology*. Décembre 2011. Vol. 1, n° 6, p. 649-657. DOI 10.1016/j.coviro.2011.10.013.

WICKER, T., GUYOT, R., YAHIAOUI, N. et KELLER, B., 2003. CACTA transposons in Triticeae. A diverse family of high-copy repetitive elements. In : *Plant Physiology*. Mai 2003. Vol. 132, n° 1, p. 52-63. DOI 10.1104/pp.102.015743.

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. et SCHULMAN, A. H., 2007. A unified classification system for eukaryotic transposable elements. In : *Nature Reviews. Genetics*. Décembre 2007. Vol. 8, n° 12, p. 973-982. DOI 10.1038/nrg2165.

WRAY, J., WILLIAMSON, E. A., SHEEMA, S., LEE, S.-H., LIBBY, E., WILLMAN, C. L., NICKOLOFF, J. A. et HROMAS, R., 2009. Metnase mediates chromosome decatenation in acute leukemia cells. In : *Blood*. 27 août 2009. Vol. 114, n° 9, p. 1852-1858. DOI 10.1182/blood-2008-08-175760.

XIE, W., GAI, X., ZHU, Y., ZAPPULLA, D. C., STERNGLANZ, R. et VOYTAS, D. F., 2001. Targeting of the yeast Ty5 retrotransposon to silent chromatin is mediated by interactions between integrase and Sir4p. In : *Molecular and Cellular Biology*. octobre 2001. Vol. 21, n° 19, p. 6606-6614.

XIONG, Y. et EICKBUSH, T. H., 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. In : *The EMBO journal*. Octobre 1990. Vol. 9, n° 10, p. 3353-3362.

YADAV, P., SUDEEP, A., GOKHALE, M., PAWAR, S., SHETE, A., PATIL, D., KUMAR, V., LAKRA, R., SARKALE, P., NICHOL, S. et MOURYA, D., 2018. Circulation of Nipah virus in Pteropus giganteus bats in northeast region of India, 2015. In : *The Indian Journal of Medical Research*. 2018. Vol. 147, n° 3, p. 318-320. DOI 10.4103/ijmr.IJMR_1488_16.

YUAN, Y.-W. et WESSLER, S. R., 2011. The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies. In : *Proceedings of the National Academy of Sciences of the United States of America*. 10 mai 2011. Vol. 108, n° 19, p. 7884-7889. DOI 10.1073/pnas.1104208108.

YUSA, K., 2015. PiggyBac Transposon. In : *Microbiology Spectrum*. Avril 2015. Vol. 3, n° 2, p. MDNA3-0028-2014. DOI 10.1128/microbiolspec.MDNA3-0028-2014.

ZERBINO, D. R., ACHUTHAN, P., AKANNI, W., AMODE, M. R., BARRELL, D., BHAI, J., BILLIS, K., CUMMINS, C., GALL, A., GIRÓN, C. G., GIL, L., GORDON, L., HAGGERTY, L., HASKELL, E., HOURLIER, T., IZUOGU, O. G., JANACEK, S. H., JUETTEMANN, T., TO, J. K., LAIRD, M. R., LAVIDAS, I., LIU, Z., LOVELAND, J. E., MAUREL, T., MCLAREN, W., MOORE, B., MUDGE, J., MURPHY, D. N., NEWMAN, V., NUHN, M., OGEH, D., ONG, C. K., PARKER, A., PATRICIO, M., RIAT, H. S., SCHUILENBURG, H., SHEPPARD, D., SPARROW, H., TAYLOR, K., THORMANN, A., VULLO, A., WALTS, B., ZADISSA, A., FRANKISH, A., HUNT, S. E., KOSTADIMA, M., LANGRIDGE, N., MARTIN, F. J., MUFFATO, M., PERRY, E., RUFFIER, M., STAINES, D. M., TREVANION, S. J., AKEN, B. L., CUNNINGHAM, F., YATES, A. et FLICEK, P., 2018. Ensembl 2018. In : *Nucleic Acids Research*. 4 janvier 2018. Vol. 46, n° D1, p. D754-D761. DOI 10.1093/nar/gkx1098.

ZHANG, G., COWLED, C., SHI, Z., HUANG, Z., BISHOP-LILLY, K. A., FANG, X., WYNNE, J. W., XIONG, Z., BAKER, M. L., ZHAO, W., TACHEDJIAN, M., ZHU, Y., ZHOU, P., JIANG, X., NG, J., YANG, L., WU, L., XIAO, J., FENG, Y., CHEN, Y., SUN, X., ZHANG, Y., MARSH, G. A., CRAMERI, G., BRODER, C. C., FREY, K. G., WANG, L.-F. et WANG, J., 2013. Comparative analysis of bat genomes provides insight into the evolution of flight and immunity. In : *Science (New York, N.Y.)*. 25 janvier 2013. Vol. 339, n° 6118, p. 456-460. DOI 10.1126/science.1230835.

ZHOU, P., TACHEDJIAN, M., WYNNE, J. W., BOYD, V., CUI, J., SMITH, I., COWLED, C., NG, J. H. J., MOK, L., MICHALSKI, W. P., MENDENHALL, I. H., TACHEDJIAN, G., WANG, L.-F. et BAKER, M. L., 2016. Contraction of the type I IFN locus and unusual constitutive expression of IFN- α in bats. In : *Proceedings of the National Academy of Sciences of the United States of America*. 8 mars 2016. Vol. 113, n° 10, p. 2696-2701. DOI 10.1073/pnas.1518240113.

COLLET Mathilde

ETUDE DE L'EVOLUTION D'UN ELEMENT TRANSPOSABLE DU GENOME DE LA CHAUVESOURIS *PTEROPUS GIGANTEUS*, ESPECE RESERVOIR DE ZOONOSES VIRALES

Thèse d'Etat de Doctorat Vétérinaire : Lyon, le 23 novembre 2018

RESUME :

Le travail présenté est une étude bibliographique des éléments transposables (ET) au sein des génomes eucaryotes illustrée par l'étude d'un ET de la chauve-souris *P. giganteus*, espèce réservoir de zoonoses virales.

La première partie de cette étude porte sur l'étude bibliographique des différents ET. Elle détaille successivement leur classification, leurs caractéristiques, leurs mécanismes de transposition puis leurs rôles dans l'évolution des organismes.

La seconde partie se consacre à l'étude d'un ET du génome de *P. giganteus*, grâce aux données préliminaires de séquençage de son génome. Cette partie débute par la présentation de cette espèce, des zoonoses dont elle est le réservoir ainsi que des particularités concernant l'immunité et les ET des chauves-souris. Puis elle se focalise sur l'analyse bioinformatique d'un élément présentant une forte homologie avec l'ET Tigger1 d'*H. sapiens*.

Cet élément, reconstruit chez *P. giganteus*, présente également les caractéristiques d'un Tigger1. L'étude de la phylogénie de cet ET permet de suspecter que différents épisodes de domestication de l'élément Tigger1 aient eu lieu au cours de l'évolution, ainsi que des événements de transferts horizontaux dont un ayant possiblement eu lieu entre les ancêtres de *P. giganteus* et *H. sapiens*.

La version définitive du séquençage du génome de *P. giganteus* ainsi que des investigations ultérieures permettront de tester les hypothèses avancées dans cette étude.

MOTS CLES :

- | | |
|---------------|----------------|
| - Transposons | - Ptéropodidés |
| - Evolution | - Zoonoses |

JURY :

- | | |
|------------------|---|
| Président : | Monsieur le Professeur Damien SANLAVILLE |
| 1er Assesseur : | Madame la Professeure Maria-Halima LAABERKI |
| 2ème Assesseur : | Madame la Professeure Véronique LAMBERT |

DATE DE SOUTENANCE : 23 novembre 2018