



HAL
open science

Etude de l'activité transpositionnelle en condition de stress chez le riz, *Oryza sativa*

Emilie Debladis

► **To cite this version:**

Emilie Debladis. Etude de l'activité transpositionnelle en condition de stress chez le riz, *Oryza sativa*. Biologie végétale. Université de Perpignan, 2016. Français. NNT : 2016PERP0026 . tel-01418445

HAL Id: tel-01418445

<https://theses.hal.science/tel-01418445>

Submitted on 16 Dec 2016

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

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

THÈSE

Pour obtenir le grade de
Docteur

Délivré par
UNIVERSITE DE PERPIGNAN VIA DOMITIA

Préparée au sein de l'école doctorale :
« Energie Environnement » ED305

Et de l'unité de recherche
Laboratoire Génome et Développement des Plantes
UMR 5096

Spécialité : Biologie

Présentée par
Emilie DEBLADIS

**Etude de l'activité transpositionnelle en
condition de stress chez le riz, *Oryza sativa***

Soutenue le 25 novembre 2016 devant le jury composé de :

Pr. Cristina VIEIRA-HEDDI, PR UCBL, UMR 5558, Lyon, France
Dr. Josep M^a CASACUBERTA, DR, CRAG, Barcelone, Espagne
Dr. Clémentine VITTE, CR1 CNRS, UMR GV, Paris, France
Dr. Thierry LAGRANGE, DR CNRS, UMR 5096, Perpignan, France
Dr. Eric LASSERRE, MCF UPVD, UMR 5096, Perpignan, France
Pr. Olivier PANAUD, PR UPVD, UMR 5096, Perpignan, France

Rapporteur
Rapporteur
Examineur
Examineur
Directeur de thèse
Directeur de thèse

Remerciements

Voici donc venu le temps des remerciements qui clôture cette folle aventure qu'est la thèse.

Tout d'abord je dois remercier mes co-directeurs de thèse Olivier Panaud et Eric Lasserre. Olivier, pour l'opportunité que tu m'as donnée de travailler sur ce sujet, et ainsi accéder à la dernière marche du parcours universitaire. Merci pour les libertés que tu m'as laissées tout au long de ce projet. Même si l'apprentissage n'aura pas été une expérience toujours facile pendant ces 3 années, je ressors grandie de cette expérience à tes côtés, où j'ai sans cesse appris à tester ma détermination et dépasser mes exigences. Eric, pour ta droiture, ton sens critique, ton humour corrosif et surtout ta réactivité et ta disponibilité.

J'aimerais également remercier Jean-Marc Deragon pour m'avoir permis de réaliser cette thèse au sein de son laboratoire dont le mode de fonctionnement m'a permis de réaliser ce travail dans de bonnes conditions.

Je veux également remercier tous les nombreux LGDPiens que j'ai eu le plaisir de côtoyer. Merci pour votre bonne humeur, d'avoir répondu à mes questions, indiqué le lieu de rangement d'un produit ou encore montré le fonctionnement d'une machine.

Merci à Christophe Crackers Belin pour son dévouement envers les étudiants, pour les réservations d'amphi et correction de français, moins pour tous ces moments de frayeur...

Un merci spécial à Edouard pour la qPCR, tous les dépannages machines et son inventivité qui oeuvrent de concert avec son sourire.

Un grand merci également aux personnes avec qui j'ai eu le plaisir de travailler durant ces trois années: Marie-Christine, dite MC, qui pas à pas m'a fait partager son savoir dans le domaine de la bio-informatique et, surtout, transmis un peu de sa passion (un exploit et non des moindres!). Merci pour toutes ces analyses bio-informatiques, ces sauvetages de commandes et pour m'avoir encouragée, écoutée et fait bouger des poings.

Merci à Sophie, dite la petite mais certainement pas par l'esprit, d'avoir été mon binôme de nettoyage d'enceinte, pour les recettes

de tarte et autres joyeusetés au citron meringué, et bien d'autres choses.

Merci Marie, pour ta disponibilité et tes conseils. Tu avais tellement raison lorsqu'en première année de thèse tu me disais que: « 3 années, ça passe très vite ». Les voilà derrière moi maintenant!

Merci à Nathalie pour m'avoir toujours proposé ton aide en cas de besoin.

Merci à Christel pour toute son aide et sa patience pour venir à bout des expériences, notamment pour les ajustements de manip's!

Merci à toutes ces personnes parties depuis ou arrivées il y a peu. Je pense notamment à Richard toujours disponible pour discuter des « points obscurs » et ton aide les deux premières années de ma thèse. Merci à Jérémy, pour son humour grinçant mais tordant de temps en temps (c'est pour la rime, « souvent » serait moins crédible), ses petits pots et sa main verte! Merci à Davy, non seulement pour le côté scientifique mais pour ces quelques mois à partager le même bureau, les pauses « grand-air » et ta décontraction.

Merci à tous ces jeudis soirs passés en très bonne compagnie, pour ces moments de détente et de partage (oui, vous voyez de quoi je parle :)).

Je ne peux évidemment pas oublier la petite dernière arrivée, Laura M. pour les co-voiturages, les co-trainages (??), ta bonne humeur contagieuse. D'ailleurs, j'en profite pour remercier ici Laura C. d'avoir toujours répondu avec bonhomie aux appels qui ne t'étaient pas destinés pendant tes heures de lectures ou de classes. Et maintenant qu'on en parle, vous n'avez pas que le prénom en commun!

Merci aux différentes colocos avec qui j'ai eu le plaisir de partager un bureau. The first: Jana, Thaïs & Danyela that I couldn't forget. Don't give up the scientist vibe from you and Parabéns! para os momentos de alegria por anos para vir.

Merci à Mlle Elodie, la meilleure coloc de bureau, avec qui j'ai appris à maîtriser à la perfection la procrastin'roulette qui nous aura bien botté le t**** (euh... motivées). À mon soutien de bout de couloir, à toutes les significations de mots et figures de style toutes aussi vite apprises qu'oubliées, à cette belle amitié qui est née et perdure, merci.

J'aimerais remercier Gouy-gouy, un de ces rares profs que j'ai eu le plaisir d'avoir. Ce n'est certainement pas grâce aux notes distribuées lors des sessions d'interros surprises en fin d'heure de cours mais surtout pour m'avoir inculqué le sens du mot pugnacité. Les notes ne sont qu'une opinion, une grande leçon de vie. Du coup, j'en profite également pour lancer un merci à M. Desbrosses, et Mme Marquès de fait, qui en dernière année de licence m'ont donné la possibilité de poursuivre mon rêve de l'époque de travailler sur les palmiers, ce qui m'a certainement permis d'être ici aujourd'hui.

Merci à mes premiers mentors avec qui j'ai énormément appris et à tous les moments partagés ensemble qui font que même dans les moments de doutes je sais pourquoi je suis là où je suis. Merci à toi Alain R., pour ta joie de vivre et ton savoir insatiable sur les palmiers, merci de m'avoir permis de rencontrer Estelle J., de m'avoir fait confiance pour la réalisation de tous ces projets menés ensemble. Merci Estelle, pour toutes ces possibilités et occasions d'apprendre, sans oublier ton humour caustique et ta culture inégalable des polars. Merci à Thierry, le grand surfeur blond australien (sic Alain R.), d'avoir toujours été présent pour moi, merci pour ta rigueur et ta bonne humeur.

Je tiens également à remercier Agnès A. pour ce stage de M1 qui a été si passionnant et m'a permis d'allier recherche et passion. Merci de m'avoir permis de rencontrer des personnes formidables et qui m'ont énormément transmis dans le domaine de la microscopie.

D'un point de vue plus personnel, je souhaite remercier tous mes amis qui ont suivi mon cheminement, tout en ayant réussi à me faire partager des moments de détente et de bonheur.

Une pensée à ces amis qui m'ont permis de continuer à célébrer la vie pendant ces 3 années : Merci à Arnaud (Finalement!!) et Nanou pour avoir été comme une seconde famille pour moi cette première année de thèse où tout m'était étranger dans cette nouvelle ville qui n'était pas la mienne.

Merci à Charlotte & Mika (mes fanfarons vigneron) pour les rendez-vous paella inaltérables du 15 août, pour les sessions rituelles ciné/brasseurs, les estivales et tant d'autres.

Merci à Chloé & Lionel pour ces moments nature, culture (ma rencontre avec Francis Hallé et son arbre à racines en fleurs sont

gravés dans ma mémoire et sur le papier maintenant) et pour les soirées déguisées qui entretiennent l'imagination.

Merci à Elise, Lilou, loin du téléphone mais proche du cœur (TMTTC ;)).

Merci Elo (ma joyeuse chic & chieuse catalane) & Fred & Nico (Salvador Dali devenu Jesus VOSTFR: Ressouss). Pas de coloc pour nous mais merci Elo pour ton franc parler et ton amitié sincère et sans langue de bois.

Merci à Andréa, Margaux, Claire, Hélène, Camille, Elio (petit bilan littéraire, je ne suis définitivement pas prête pour un catch), Thierry et son fantastique safran (j'en salive encore), Alison (de ne pas répondre), Sophie A. (pour ta patience et ton inspirant courage), Adrienne (que je n'ai pas souvent vue, mais qui nous fait voyager avec elle), Alice & Clémence (on part où??) & Loris. Merci à PE et Çigdem, pour le plus beau des voyages de ces 3 dernières années, à Léa (ma cousine avant tout), et à ma zazinette voisinettes Elsa.

Un grand merci à Kelly & Romain. Merci pour votre écoute, votre soutien, et pour m'avoir réservé de quoi m'occuper pendant quelques mois quand cette thèse sera terminée, vous aviez peur que je m'ennuie? Et merci pour cette excuse de plus de devoir retourner à Montpellier, ville de mon cœur, le plus souvent que je le puisse.

À Céline (ma merveilleuse petite tête bouclée), la soeur que je n'ai jamais eue. On n'aura certainement pas vécu les meilleurs moments ensemble cette année mais nous les avons partagés ensemble et c'est l'essentiel. Fournissons-nous en d'autres de pur bonheur.

Merci à ma famille qui me comble de fierté et me pousse à donner le meilleur de moi-même en espérant les égarer un jour. Une mention spéciale à Nils et à mes parents pour leur amour inconditionnel et d'avoir aidé à faire de moi ce que je suis aujourd'hui.

Et que ne serait des remerciements sans te remercier, toi, Thibaut, d'avoir su gérer avec sang-froid les biens trop nombreux moments de crise. Merci pour ces dernières semaines, nous avons côtoyé le pire, continuons vers le meilleur...

Merci à toutes ces personnes qui peuplent désormais mon cœur...

Table des matières

Table des matières	1
Table des illustrations	4
Liste des abréviations	7
PARTIE 1. Introduction	10
1.1. Éléments transposables (ETs)	10
1.1.1. Impact sur la structure et l'évolution des génomes	10
De leur découverte à leur classification	10
Contribution à l'évolution des génomes	13
1.1.2. La régulation des ETs	16
La diversité des marques épigénétiques associées aux ETs	17
Le contrôle épigénétique et la régulation des ETs chez les plantes	18
L'ARN polymérase IV et la répression des ETs <i>via</i> les siRNAs	20
1.1.3. Les ETs et la réponse au stress	20
Culture <i>in vitro</i>	21
Stress biotiques	22
Stress abiotiques	22
1.2. Détection des ETs par les techniques de séquençage haut débit (Next-Generation Sequencing, NGS)	23
1.2.1. Les techniques d'analyse des données génomiques	23
Principe du séquençage haut-débit	24
Le séquençage haut-débit : techniques de 2 ^{ème} et 3 ^{ème} générations	24
Principales caractéristiques des séquenceurs : Forces et faiblesses	26
1.2.2. Détection des variations structurales associées aux éléments transposables	27
Le paired-end mapping	27
L'alignement des lectures coupées	28
Le Depth of Coverage	29
1.3. Le modèle de l'étude : Le riz	30
1.3.1. Phénologie	30
1.3.2. Les ETs chez le riz, <i>Oryza sativa</i>	31
1.4. Objectifs de la thèse	33

PARTIE 2. Transposition <i>in planta</i> chez le riz <i>Oryza sativa</i> dans un contexte épigénétique relâché	35
2.1. Matériel végétal	37
2.1.1. Génération du mutant <i>Osnrpd1</i> et sélection des plantes	37
Caractérisation des lignées mutantes obtenues	37
Impact de la construction RNAi sur l'accumulation des sRNAs.....	37
2.1.2. Génération de matériel végétal	38
2.2. Méthodes « omiques »	40
2.2.1. Génome	40
2.2.1. sRNAseq	40
2.2.1. Transcriptome	40
2.2.2. Méthylome	41
2.3. Méthodes utilisées pour la détection de néo-insertions d'ETs	42
2.3.1. Détection par approche <i>a priori</i>	42
2.3.2. Mise au point d'une méthode de détection des néo-insertions	42
Identification de séquences non/mal assemblées dans le génome de référence	43
2.3.3. Détection par analyse des clipped-reads	44
2.3.4. Détection par calcul de la profondeur de lecture	45
2.4. Résultats	47
2.4.1. Profil d'accumulation d' <i>OsNRPD1</i> dans les différentes générations <i>Osnrpd1</i>	47
2.4.2. Néo-insertions d'ETs dans les plantes <i>Osnrpd1</i>	47
Détection par approche AAP	47
Détection par le pipeline d'analyse	48
Détection par analyse des clipped-reads	48
Détection par calcul de la profondeur de lecture	49
2.4.3. Validation des néo-insertions	49
Validation par la méthode RBIP	49
Validation par analyse de Southern-blot	50
Validation <i>in silico</i>	51
2.4.4. Impact de la stratégie RNAi sur l'activation rétrotranspositionnelle.....	51
2.4.5. Accumulation des sRNAs dans les plantes <i>Osnrpd1</i>	52
Les ETs activés dans les plantes <i>Osnrpd1</i> cibles des petits ARNs ?	54
2.4.6. Différentiels d'expression entre les plantes Nipponbare (WT) et les plantes <i>Osnrpd1</i> ..	56
Expression différentielle des ETs	57

Expression différentielle et description des gènes	58
2.4.7. Effet de la mutation d' <i>O_sNRPD1</i> sur le méthylome des plantes <i>Osnrpd1</i>	59
2.4.8. Observation phénotypique des plantes	61
Synthèse des principaux résultats de l'étude du mutant <i>Osnrpd1</i> , en contexte épigénétique relâché...	62
Schéma récapitulatif des ETs activés dans les différentes générations de mutants <i>Osnrpd1</i>	63
2.5. Discussion et perspectives	64
2.5.1. Approche expérimentale	64
2.5.2. Réactivation transpositionnelle suite à la modification du contrôle épigénétique	65
PARTIE 3. Activité rétrotranspositionnelle chez le riz <i>Oryza sativa</i> en condition de stress	70
3.1. Matériel végétal	71
Efficacité du stress thermique	71
3.2. Méthodes de détection de néo-insertions dans les plantes stressées thermiquement.....	72
3.3. Résultats.....	73
3.3.1. Activité transpositionnelle détectée par le pipeline d'analyse	73
3.3.2. Activité transpositionnelle détectée par la méthode AAP	73
3.3.3. Activité transpositionnelle détectée par la méthode du DOC	73
3.4. Discussion	75
PARTIE 4. Étude de la dynamique des génomes par la nouvelle technologie Nanopore	77
DISCUSSION GÉNÉRALE	100
RÉFÉRENCES BIBLIOGRAPHIQUES	103
ANNEXES	118

Table des illustrations

- Fig. 1.1 Le système *Ac/Ds* du maïs
- Fig. 1.2 Classification des éléments transposables
- Fig. 1.3 Mécanisme de transposition d'un rétrotransposon à LTR
- Fig. 1.4 Mécanismes de réparation de l'ADN aux sites d'excision d'un transposon ADN
- Fig. 1.5 Variation de la taille des génomes et de leur composition en éléments transposables entre différentes espèces
- Fig. 1.6 Exemple d'effet de l'insertion d'ETs sur la diversité phénotypique
- Fig. 1.7 Caractéristiques des différents états chromatinien chez *A. thaliana*
- Fig. 1.8 Régulation de l'état de compaction de la chromatine par la méthylation
- Fig. 1.9 Mécanisme de répression génétique post-transcriptionnelle (PTGS)
- Fig. 1.10 RNA-directed DNA Methylation
- Fig. 1.11 Déméthylation active
- Fig. 1.12 Les polymérase 4 d'*A. thaliana* & d'*O. sativa*
- Fig. 1.13 Accumulation du rétrotransposon *Tos17* pendant la culture cellulaire
- Fig. 1.14 Fréquence d'insertion de l'élément *Karma*
- Fig. 1.15 Activation du rétrotransposon *Osen* en réponse au stress thermique dans des mutants déficients dans la biogenèse de siRNAs d'*A. thaliana*
- Fig. 1.16 Illustration de la méthode Sanger
- Fig. 1.17 Grands principes des nouvelles techniques de séquençage NGS
- Fig. 1.18 Le Paired-End Mapping (PEM)
- Fig. 1.19 La méthode des lectures coupées
- Fig. 1.20 Méthode de la profondeur de lecture (DOC)
- Fig. 1.21 Initiation des talles pendant l'étape de tallage
- Fig. 1.22 Distribution du nombre de copies de rétrotransposons à LTR (LTR-RT) par famille chez le riz (*O. sativa*)
- Fig. 1.23 Répartition des éléments transposables sur les chromosomes
- Tab. 2.1 Tableau récapitulatif des données de séquençage obtenues pour les plantes *Osnrpd1* T0 et T1
- Fig. 2.1 Représentation schématique du vecteur utilisé pour le RNAi contre *O_sNRPD1*
- Fig. 2.2 Profil d'accumulation des transcrits de *O_sNRPD1*
- Fig. 2.3 Distribution sRNAs sur *O_sNRPD1*

Fig. 2.4 Génotypage des plantes *Osnrpd1* de génération T1

Fig. 2.5 Profil d'accumulation de *OsNRPD1* dans les plantes T1 23.1

Fig. 2.6 Principe de détection d'insertions d'Est

Fig. 2.7 Schéma du matériel Nipponbare sauvage séquencé.

Tab. 2.2 Tableau des positions des séquences mal assemblées dans le génome de référence retrouvées dans les 9 premiers génomes séquencés

Tab. 2.3 Nombre de néo-insertions total et de *Tos17* détectées dans les plantes *Osnrpd1* en fonction du nombre de paires discordantes

Tab. 2.4 Tableaux comparatifs des méthodes utilisées pour la détection des néo-insertions de *Tos17*, *Lullaby* et *Karma*

Fig. 2.8 Profil d'accumulation d'*OsNRPD1* dans 3 générations de plantes *Osnrpd1*

Tab. 2.5 Néo-insertions de *Tos17* et *Lullaby* dans les plantes *Osnrpd1* de génération T0 et T1

Tab. 2.6 Néo-insertions détectées dans les plantes *Osnrpd1* par le pipeline d'analyse

Fig. 2.9 Dot-plot des séquences flanquant la néo-insertion, de l'élément inséré (*Tos17*) et des lectures discordantes et des lectures coupées au point d'insertion sur elles-même

Tab. 2.7 Néo-insertions de *Tos17* et *Lullaby* détectées dans les plantes *Osnrpd1* de génération T0 et T1 par le pipeline d'analyse

Tab. 2.8 Tableau récapitulatif des néo-insertions de *Karma* détectées par analyse des lectures coupées

Tab. 2.9 Tableau des néo-insertions détectées par la méthode du DOC

Fig. 2.10 Validation des néo-insertions par la technique du RBIP (Retrotransposon-Based Insertion polymorphism)

Tab. 2.10 Tableau récapitulatif des néo-insertions confirmées par PCR de *Tos17* et *Lullaby* détectées par analyse à priori, le point exact de l'insertion ainsi que le sens dans lequel est inséré l'élément et l'état de zygotie dans chaque plante *Osnrpd1*

Fig. 2.11 Copies de *Karma* dans la plante Nipponbare et les plantes *Osnrpd1*

Tab. 2.11 Validation *in silico*

Tab. 2.12 Tableau du nombre de copies d'éléments par famille détectées par l'analyse du DOC

Tab. 2.13 Tableau des néo-insertions de *Tos17* et de *Karma* par la méthode du DOC

Fig. 2.12 Distribution des sRNAs dans les plantes *OsNRPD1*

Fig. 2.13 Distribution chromosomique des sRNAs

Fig. 2.14 Accumulation des sRNAs sur les ETs réactivés

Fig. 2.15 Distribution des sRNAs le long d' ETs

Fig. 2.16 Distribution des sRNAs 24-nt le long de *Tos17*

- Fig. 2.17 Représentation globale des ETs différenciellement accumulés
- Fig. 2.18 Représentation des ETs actifs transpositionnellement exprimés dans la plante 23.1-15
- Fig. 2.19 Représentation globale des gènes différenciellement accumulés
- Fig. 2.20 Heatmap des gènes différenciellement exprimés dans la *plante* *Osnrpd1* 23.1-15 versus la plante contrôle WT
- Fig. 2.21 Gene Ontology des gènes différenciellement exprimés
- Fig. 2.22 Cytosines méthylées dans les différents contextes
- Fig. 2.23 Méthylation globale dans les différents génomes
- Fig. 2.24 Niveau de méthylation des plantes *Osnrpd1* comparées au WT
- Fig. 2.25 Distribution de la méthylation de différentes classes d'ETs
- Fig. 2.26 Distribution des longueurs des feuilles drapeaux entre plantes sauvages et *Osnrpd1* de la générations T3
- Fig. 2.27 Distribution des graines fertilisées entre plantes sauvages et transformées sur différentes générations
- Fig. 2.28 Phénotype torsadé de plantes *Osnrpd1*
- Fig. 3.1 Profil d'expression de deux gènes en réponse à la chaleur
- Tab. 3.1 Tableau des insertions détectées pour les plantes stressées thermiquement
- Fig. 3.2 Dot-plot des séquences flanquant le point d'insertion de l'élément 1553, de l'élément MULE inséré et des lectures pairées et des lectures coupées obtenues par séquençage alignées sur elles-mêmes
- Fig. 3.3 Profil de migration des produits PCR des insertions 1553 et 1803 sur gel d'agarose
- Tab. 3.2 Néo-insertions détectées par la méthode du DOC

Liste des abréviations

AAP	Analyse <i>a priori</i>
Ac	Activator
ADN	Acide Désoxyribonucléique
ADN-T/T-DNA	ADN de transfert
AGO	ARGONAUTE
AP	Aspartic Protease
ARN	Acide Ribonucléique
ARNm	Acide Ribonucléique messenger
BLAST	Basic Local Alignment Search Tool
CIGAR	Compact Idiosyncratic Gapped Alignment Report
<i>CIV</i>	Culture <i>in vitro</i>
CMT3	CHROMOMETHYLASE 3
DCL3	DICER LIKE 3
DDM1	DECREASE IN DNA METHYLATION 1
DE	Différentiellement Exprimés
DME	DEMETER
DML	DEMETER-LIKE
DOC	Depth Of Coverage
DRM2	DOMAINS REARRANGED METHYLTRANSFERASE 2
Ds	Dissociator
<i>EgDEF1</i>	<i>Elaeis guineensis DEFICIENS 1</i>
epiRIL	Epigenetic Recombinant Inbred Line
ET	Elément Transposable
GA	Gibbérellines
Gag	Group-specific antigen
GAMYB	Gibberellin- and Abscisic acid-regulated MYB
GFP	Green Fluorescent Protein
GO	Gene Ontology
<i>Gret1</i>	<i>Grapevine Retrotransposon 1</i>
GyDB	Gypsy Database
INT	Intégrase
IRGSP	International Rice Genome Sequencing Project

IS	Insertion Sequences
KYP	KRYPTONITE
LARD	LArge Retrotranposons Derivative
LINE	Long Interspersed Nuclear Element
LTR-RT	Long Terminal Repeat-Retrotransposon
MET1	METHYLTRANSFERASE 1
MITE	Miniature Inverted Transposable Element
<i>mPing</i>	<i>miniature Ping</i>
ncRNA	ARN non codant
NGS	next-generation sequencing
NRPD1a	NUCLEAR RNA POLYMERASE D1a
ONT	Oxford Nanopore Technologies
ORF	Open Reading Frame
PANTHER	Protein ANalysis THrough Evolutionary Relationships
PBS	Primer Binding Site
PCR	Polymerase Chain Reaction
PEM	Paired-End Mapping
POLIV	ARN POLYMERASE IV
PR5	Pathogenesis-related protein 5
PTGS	Post-Transcriptional Gene Silencing
RAPDB	Rice Annotation Project DataBase
RBIP	Retrotransposon-Based Insertion Polymorphism
RdDM	RNA-directed DNA Methylation
RDR2	ARN POLYMERASE ARN DEPENDANTE 2
RH	Ribonuclease H
RISC	RNA-Induced Silencing Complex
RiTE-db	Rice TE Database
RNAi	Ribonucleic acid interference
ROS1	REPRESSOR OF SILENCING 1
RPKM	Reads Per Kilobase of transcript per Million mapped reads
RPM	Reads Per Million
RT	Reverse Transcriptase
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
RTN	Retrotransposon

<i>sde4</i>	<i>silencing defective 4</i>
SINE	Short Interspersed Nuclear Element
sRNA	Small RNA
siRNA	small interfering Ribonucleic acid
SMRT	Single Molecule Real Time
SOLiD	Sequencing by Oligonucleotide Ligation and Detection
SV	Structural variation
TIR	Tandem Inverted Repeat
TGS	Transcriptional Gene Silencing
TRIM	Terminal-repeat Retrotransposons In Miniature
TSD	Target Site Duplication
UDP	Uridine diphosphate
USB	Universal Serial Bus
UTR	UnTranslated Region
VIM1	VARIANT IN METHYLATION 1
VLP	Virus-Like Particle
VvMYBA1	<i>Vitis vinifera</i> MYeloBlastosis Anthocyanin 1
vs.	versus
WT	Wild Type
ZMW	Zero-Mode Waveguide

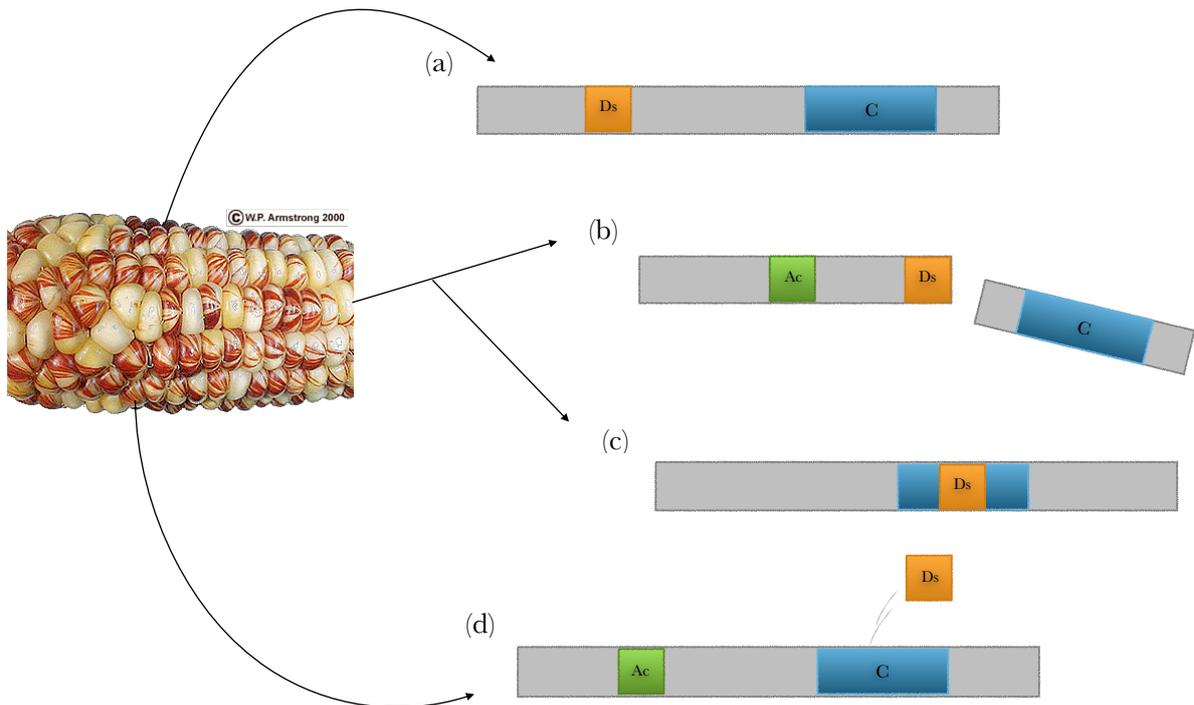
Unités

Gb	Gigabase
kb	Kilobase
M	Million
Mb	Megabase
nm	Nanomètre
nt	Nucléotides
pb	Paire de bases

Fig. 1.1 Le système *Ac/Ds* du maïs

En présence de l'élément activateur (*Ac*), l'élément dissociateur (*Ds*) peut causer des cassures de chromosomes. Le gène de la coloration (ici, *C*) existe sous la forme de 2 allèles, l'allèle dominant conduit au phénotype coloration, l'allèle récessif à une absence de coloration.

- (a) En absence de l'élément *Ac*, *Ds* ne transpose pas et aucune cassure de chromosome ne se produit. Le gène *C* s'exprime sous la version de l'allèle dominant, le grain est coloré.
- (b) En présence de l'élément *Ac*, *Ds* peut transposer dans le génome et causer des cassures de chromosome au point de l'insertion. Si cette cassure provoque la perte de l'allèle dominant du gène *C*, le grain sera non coloré. Cette interruption est irréversible.
- (c) Plus rarement, en présence de l'élément *Ac*, *Ds* peut s'insérer dans l'allèle dominant du gène *C* entraînant un phénotype absence de coloration.
- (d) L'évènement de transposition qui a lieu en (c) crée une interruption transitoire de l'expression du gène *C*. Si l'élément *Ds* s'excise pour transposer ailleurs dans le génome, la fonction de l'allèle dominant est restaurée. On observe alors un patron d'expression de taches de coloration.



PARTIE 1. Introduction

1.1. Éléments transposables (ETs)

1.1.1. Impact sur la structure et l'évolution des génomes

De leur découverte à leur classification :

L'ADN, molécule de base de tout être vivant et support de l'hérédité, était considéré jusque dans les années 1950 comme une entité stable. L'ensemble des molécules d'ADN forme le génome d'un individu. À chaque cycle de division cellulaire, l'ADN est recopié et transmis aux différentes cellules. Des erreurs aléatoires peuvent alors se produire et engendrer des mutations, entraînant ainsi une altération du patrimoine génétique. L'instabilité génétique liée aux ETs a été mise en évidence par Barbara McClintock pour expliquer les différents patrons de coloration retrouvés chez le maïs sud-américain. À partir de ses études sur la structure des chromosomes, elle a introduit ce concept majeur de séquences mobiles dans les génomes. En étudiant les points de cassure au niveau du bras court du chromosome 9 du maïs, elle découvrit que ces cassures apparaissaient fréquemment et se produisaient en même temps que les grains de maïs présentaient des phénotypes de coloration différents. Elle établit que la cassure chromosomique au niveau de ce locus était liée à la présence d'un facteur qu'elle nomma *Ds* pour *Dissociator*. Par le biais de différents croisements, elle démontra que l'élément *Ds* ne pouvait pas induire à lui seul une cassure et que la présence d'un autre élément *Ac*, pour *Activator*, était nécessaire. Ce système impliquant 2 éléments de contrôle fut nommé système *Ac/Ds* (Fig. 1.1). Malgré la rigueur de ses travaux, cette notion d'éléments pouvant se déplacer au sein des génomes était bien trop loin des idées dominantes de l'époque pour en apprécier la pertinence. Il faudra attendre les années 1960 et l'identification d'un système de régulation semblable au système *Ac/Ds*, celui de l'opéron lactose chez la bactérie *Escherichia coli* par Jacob et Monod en 1961 ou encore la découverte des séquences d'ADN IS (Insertion Sequences) (Fiandt *et al.*, 1972; Hirsch *et al.*, 1972).

Depuis, les ETs, composants ubiquitaires des génomes eucaryotes, sont le sujet de nombreuses études menées sur des modèles animaux et végétaux. Les connaissances sur la structure et la composition des génomes ont progressé de manière concomitante avec le séquençage des génomes de nombreuses espèces. Les ETs, retrouvés dans toutes les espèces étudiées jusqu'à présent, nécessitaient une nomenclature uniformisée afin de faciliter les études comparatives entre espèces. Une première classification regroupe les ETs en 2 grandes classes selon leur mode de transposition (Finnegan, 1989). La première se compose des ETs capables de transposer par un mécanisme de

Fig 1.2 Classification des éléments transposables (Wicker *et al.*, 2007)

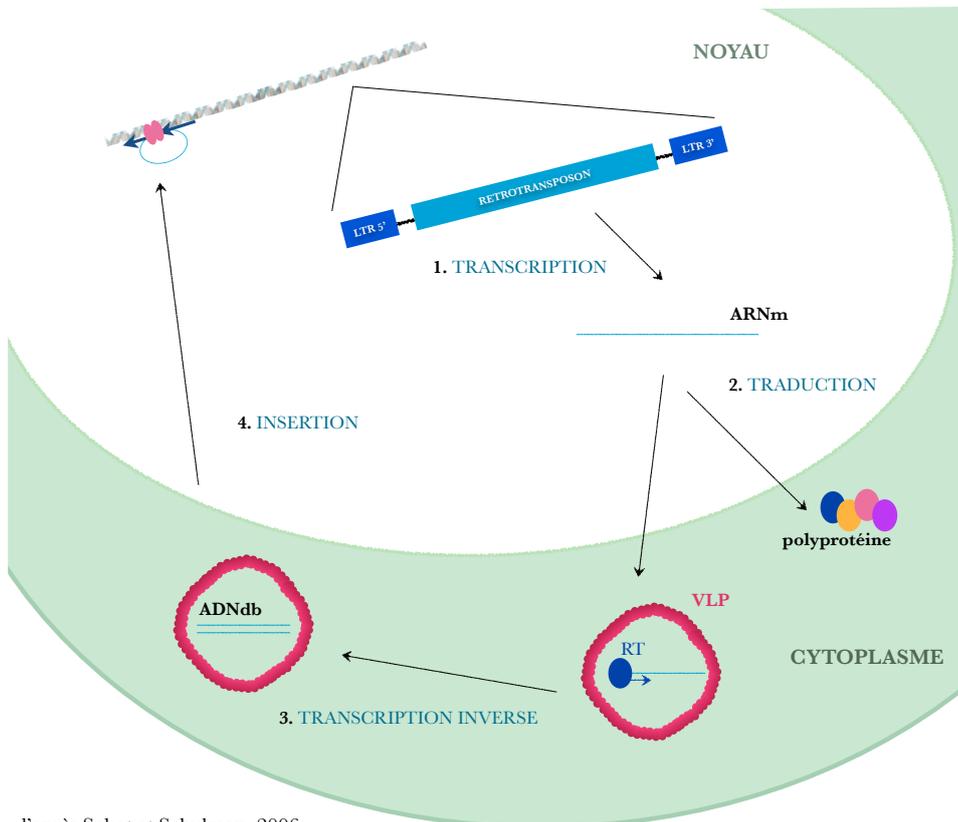
Classification	Structure	TSD	Code	Occurrence	
Order Superfamily					
Class I (retrotransposons)					
LTR	Copia		4-6	RLC	P, M, F, O
	Gypsy		4-6	RLG	P, M, F, O
	Bel-Pao		4-6	RLB	M
	Retrovirus		4-6	RLR	M
	ERV		4-6	RLE	M
DIRS	DIRS		0	RYD	P, M, F, O
	Ngaro		0	RYN	M, F
	VIPER		0	RYV	O
PLE	Penelope		Variable	RPP	P, M, F, O
LINE	R2		Variable	RIR	M
	RTE		Variable	RIT	M
	Jockey		Variable	RIJ	M
	L1		Variable	RIL	P, M, F, O
	I		Variable	RII	P, M, F
SINE	tRNA		Variable	RST	P, M, F
	7SL		Variable	RSL	P, M, F
	5S		Variable	RSS	M, O
Class II (DNA transposons) - Subclass 1					
TIR	Tc1-Mariner		TA	DTT	P, M, F, O
	hAT		8	DTA	P, M, F, O
	Mutator		9-11	DTM	P, M, F, O
	Marlin		8-9	DTE	M, O
	Transib		5	DTR	M, F
	P		8	DTP	P, M
	PiggyBac		TTAA	DTB	M, O
	PIF-Harbinger		3	DTH	P, M, F, O
	CACTA		2-3	DTC	P, M, F
	Crypton		0	DYC	F
Class II (DNA transposons) - Subclass 2					
Helitron		0	DHH	P, M, F	
Maverick		6	DMM	M, F, O	

Structural features	
	Long terminal repeats
	Terminal inverted repeats
	Coding region
	Non-coding region
	Diagnostic feature in non-coding region
	Region that can contain one or more additional ORFs

Protein coding domains	
AP, Aspartic proteinase	APE, Apurinic endonuclease
ENV, Envelope protein	GAG, Capsid protein
POL B, DNA polymerase B	RH, RNase H
Tase, Transposase (* with DDE motif)	ATP, Packaging ATPase
	HEL, Helicase
	RPA, Replication protein A (found only in plants)
	YR, Tyrosine recombinase
	C-INT, C-integrase
	INT, Integrase
	CYP, Cysteine protease
	EN, Endonuclease
	ORF, Open reading frame of unknown function
	RT, Reverse transcriptase
	Y2, YR with YY motif

Species groups	
P, Plants	M, Metazoans
F, Fungi	O, Others

Fig. 1.3 Mécanisme de transposition d'un rétrotransposon à LTR



(1) Transcription du rétrotransposon en ARNm, à partir de la région 5' vers la région 3'. (2) Traduction et synthèse des protéines codées par les domaines gag et pol. Polymérisation des protéines GAG pour former la Virus-like particle (VLP) dans laquelle l'ARNm sert de matrice pour la synthèse du brin complémentaire par la rétrotranscriptase (RT) (3). (4) L'intégrase (cercles roses) assure l'intégration de la nouvelle copie synthétisée à une nouvelle localisation génomique.

« copier/coller » faisant intervenir un intermédiaire ARN. La seconde regroupe les éléments transposant par un mode de « couper/coller » et utilisant un intermédiaire ADN. Cette classification a ensuite été revue pour y intégrer la notion d'ordre incluant des critères structuraux, de la similarité des séquences nucléotidiques ou encore la longueur de la séquence cible dupliquée (TSD, Target Site Duplication) lors de la transposition (Wicker *et al.*, 2007) (Fig. 1.2).

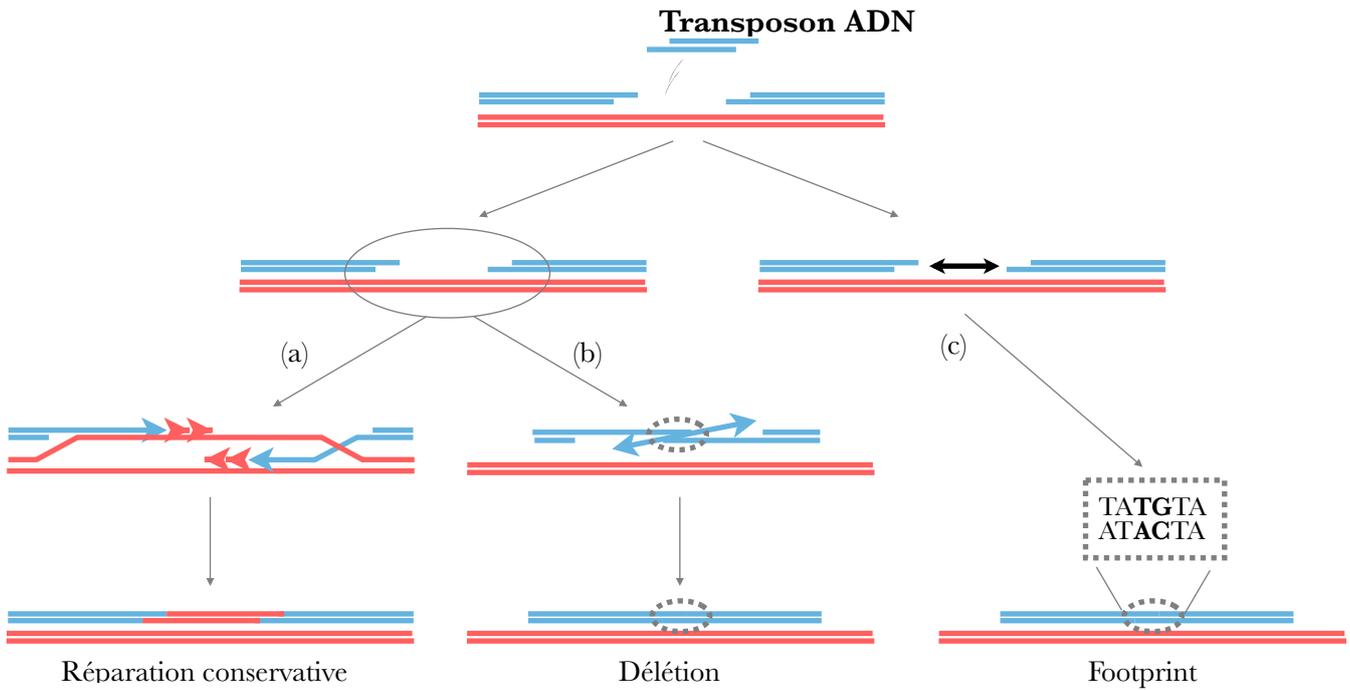
Dans la classe I des ETs, les rétrotransposons (RTNs), on retrouve 2 catégories d'éléments : les RTNs à LTR (Long Terminal Repeat-Retrotransposons, LTR-RTs) qui possèdent 2 séquences terminales répétées et ceux sans LTR qui diffèrent également des premiers par leur mode de transposition.

Les LTR-RTs constituent la classe d'ETs la plus abondante chez les plantes (Vitte et Panaud, 2005; Hawkins *et al.*, 2006). Les LTRs contiennent les régions promotrices des LTR-RTs et leur taille peut atteindre plusieurs milliers de nucléotides. Dans la région interne des RTNs se trouvent 2 ORFs (Open Reading Frames) : l'une, la Gag (Group-specific antigen), codant pour les protéines de la capsid qui se polymériseront pour former la Virus-Like Particle (VLP) dans laquelle se passera la transcription inverse; et l'autre, la Pol, codant une polyprotéine intervenant dans les différentes étapes de la rétrotransposition (Fig. 1.3). Les différentes activités enzymatiques codées par la Pol (et le nom des protéines associées) sont une activité protéase (Aspartic Protease : AP), une activité rétrotranscriptase (Reverse Transcriptase : RT), une activité de dégradation de l'ARN (RNase H : RH) et une activité endonucléase (Intégrase : INT). Le mécanisme de rétrotransposition est initié par la fixation de l'ARN de transfert au niveau du site de liaison PBS (Primer Binding Site). L'ADN ainsi rétrotranscrit retourne dans le noyau afin d'être réinséré dans le génome. Cette dernière étape entre la rétrotranscription et l'insertion n'est pas encore totalement comprise ni détaillée, cependant une forme circulaire a été détectée pour des ETs actifs et pourrait être une forme intermédiaire avant la réinsertion d'un élément dans le génome (Lanciano *et al.*, submitted). Les LTR-RTs sont composés de 2 superfamilles, *Copia* et *Gypsy*, différenciées par la position du domaine INT au sein du domaine Pol (Havecker *et al.*, 2004; Sabot et Schulman, 2006).

Quant aux RTNs dépourvus de LTR, ils regroupent principalement les LINEs (Long Interspersed Nuclear Elements) et les SINEs (Short Interspersed Nuclear Elements) et sont très répétés chez les mammifères. La région promotrice des LINEs est contenue dans la région 5'-UTR. Les SINEs, en revanche, ne codent pas les protéines nécessaires à leur transposition. Ils font partie des éléments qualifiés de non-autonomes et utilisent les LINEs pour leur transposition (Kajikawa et Okada, 2002).

Fig. 1.4 Mécanismes de réparation de l'ADN aux sites d'excision d'un transposon ADN

L'excision d'un transposon ADN génère une cassure double brin du chromosome (bleu). Dans cet exemple, l'insertion du transposon était homozygote, l'élément est aussi inséré sur la chromatide soeur (rouge). La réparation de l'ADN au point d'excision est prise en charge par la machinerie de réparation de l'ADN de différentes manières. Si la chromatide soeur est utilisée comme matrice, alors le processus de réparation régénère une copie intacte du transposon excisé (a). La réparation peut se faire entre les séquences homologues du même chromosome si elles présentent des homologies de part et d'autre de l'excision, générant des délétions variables (b). Alternativement, la réparation peut se faire à partir des extrémités qui sont jointes grâce à un mécanisme semi-conservatif, le non-homologous end joining (NHEJ). Dans ce cas, la réparation sera imprécise, créant des footprints (c).



adapté de J-L Bessereau, ENS, Paris

Les analyses de génomes, de plus en plus nombreuses, ont permis de montrer la diversité structurale des ETs et notamment de distinguer les éléments autonomes des non autonomes, en fonction de la présence ou non d'ORFs fonctionnels codant pour les protéines nécessaires à la transposition. Les ETs non autonomes ont perdu leur capacité à transposer, en accumulant progressivement des mutations, substitutions, insertions et délétions les rendant incapables de produire eux-mêmes la machinerie protéique nécessaire à leur transposition. Cependant, ils peuvent continuer à transposer, en recrutant la machinerie transpositionnelle de leurs partenaires autonomes. Ainsi les LARDs (LArge Retrotransposons Derivatives) sont les éléments non autonomes dérivant des LTR-RTs dont la taille dépasse les 4 000 bases (Kalendar *et al.*, 2004), les plus petits sont qualifiés de TRIMs (Terminal-repeat Retrotransposons In Miniature) (Witte *et al.*, 2001).

Enfin, les éléments de classe II, ou transposons à ADN, regroupent les éléments qui se déplacent en s'excisant d'un site donneur vers un site accepteur par un mécanisme de « couper/coller ». L'excision est assurée par la transposase de l'élément et l'intégration par une activité endonucléase/ligase. Ce mode de transposition, contrairement à celui des RTNs, n'induit pas une multiplication de l'élément à chaque événement de transposition. Pourtant, les transposons sont également présents en milliers de copies dans les génomes. Lorsqu'un transposon s'excise, s'il y a réparation du site donneur par le mécanisme de « non-homologous end joining », alors la copie du site donneur est perdue. Une alternative est possible : lorsqu'un transposon s'excise, la réparation du site donneur peut se faire par recombinaison homologue grâce à la séquence de l'élément sur la chromatide soeur ou du chromosome homologue, la copie du site donneur sera alors conservée (Fig. 1.4).

Dans cette seconde classe d'ETs, les MITEs (Miniature Inverted Transposable Elements) constituent l'un des types de transposon les plus abondants chez les plantes (Wessler *et al.*, 1995; Casacuberta et Santiago, 2003; Yang *et al.*, 2009). Pourtant, ces éléments sont non-autonomes, composés seulement de TIRs (Tandem Inverted Repeats) à leurs extrémités et entre elles de séquences non fonctionnelles (Bureau et Wessler, 1994). L'origine de leur amplification reste encore à élucider, néanmoins pour leur transposition même s'ils ne possèdent pas de séquences codantes pour la transposase, ils peuvent utiliser la transposase d'autres transposons à ADN autonomes (Feschotte *et al.*, 2002; Jiang *et al.*, 2003). Leur présence en grand nombre de copies dans les génomes et leur association fréquente à proximité ou dans des gènes (Bureau et Wessler, 1992; Casacuberta et Santiago, 2003) suggèrent soit que ces éléments se sont insérés au niveau de sites préférentiels, soit qu'ils ont été sélectionnés pendant l'évolution des génomes. En effet, les MITEs sont de petites séquences dont l'insertion dans les gènes fournit de nouveaux sites de liaison pour des facteurs de

transcription et ainsi de nouveaux réseaux de régulation transcriptionnelle (Feschotte, 2008; Naito *et al.*, 2009; Hénaff *et al.*, 2014).

Contribution à l'évolution des génomes :

Selon la théorie Darwinienne que la variabilité est à la base de l'évolution, un réel intérêt investit les biologistes suite à la découverte que le contenu en ADN était constant entre les différentes cellules d'un organisme et également au sein des différents individus d'une espèce donnée (Vendrely et Vendrely, 1948). La taille d'un génome fut alors désignée par le terme de la valeur C, pour valeur constante (Swift, 1950).

Bien que cette valeur ait été considérée dans un premier temps liée à l'état de complexité d'un organisme, cette information est apparue moins appropriée après la découverte que des espèces primitives, les amphibiens et les poissons, contenaient plus d'ADN dans le noyau de leurs cellules que l'Homme. Ces variations de la taille des génomes ont toujours intrigué les biologistes, tant la complexité d'un génome, évaluée en terme de nombre de types cellulaires ne semble pas coïncider avec la taille d'un génome : c'est ce qu'on appelle le paradoxe de la valeur C (Thomas, 1971).

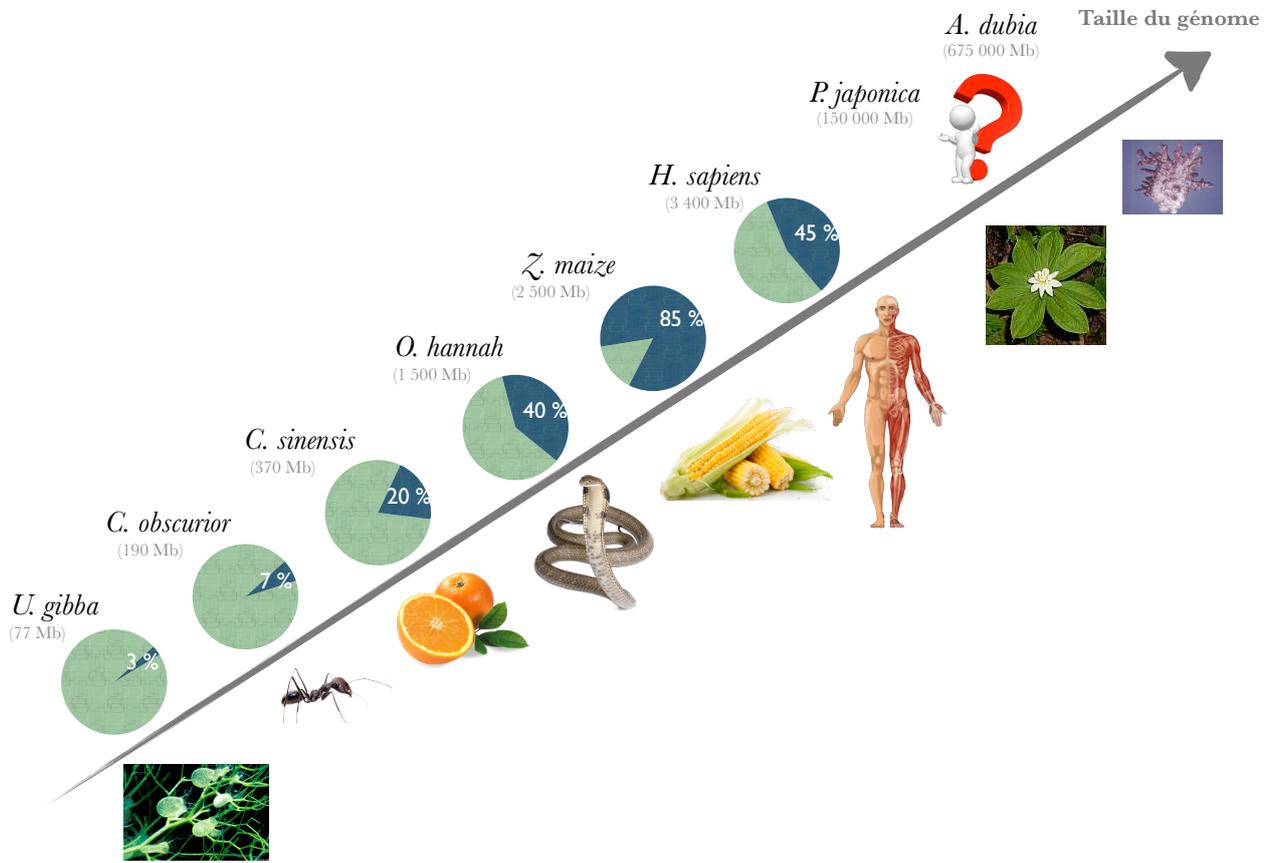
L'apparition de la notion du gène poubelle (Ohno, 1972), qui serait de l'ADN non codant, permettrait d'expliquer qu'un organisme moins complexe que l'Homme mais présentant un plus gros génome aurait en réalité une grande fraction d'ADN qui ne porterait pas d'informations importantes. Après l'introduction de la notion d'ADN non codant dans le paradoxe, un autre indicateur pertinent pour expliquer l'état de complexité d'un génome semblait être le nombre gènes trouvé dans un génome, la valeur G (Bird, 1995). Mais le résultat de cette nouvelle hypothèse mène à un autre complexe, celui de la valeur G (Hahn et Wray, 2002). Le nombre de gènes ne semble pas non plus corrélér avec la complexité apparente des organismes. En effet, chez l'humain on estime à environ 20 000 le nombre de gènes codants pour des protéines (Goodstadt et Ponting, 2006), alors que chez la paramécie, *Paramecium tetraurelia*, le nombre de gènes codants est évalué à 40 000 (Aury *et al.*, 2006).

Finalement, il ne semble pas y avoir de lien simple entre la complexité d'un organisme et la taille de son génome. Cependant, l'augmentation de la taille des génomes serait la conséquence de deux mécanismes, la polyploïdisation (Soltis et Soltis, 1999; Soltis *et al.*, 2009; Paterson *et al.*, 2012) et l'amplification de l'ADN répété non codant (Hawkins *et al.*, 2006; Piegu *et al.*, 2006, Fedoroff, 2012).

La polyploïdisation est un phénomène courant chez les plantes, notamment pour un grand nombre d'espèces cultivées (Wendel, 2000; Comai, 2005; Leitch et Leitch, 2008; Grover et

Fig. 1.5 Variation de la taille des génomes et de leur composition en éléments transposables entre différentes espèces

La part des ETs dans les génomes est représentée en bleu dans les diagrammes. La taille du génome est indiquée entre parenthèses.



Wendel, 2010). Il existe deux catégories de polyploïdes, les autopolyploïdes et les allopolyploïdes, selon le mécanisme d'hybridation entre individus d'une même espèce ou non. Chez les angiospermes, les évènements de duplication du génome sont courants (Cui *et al.*, 2006).

Outre la polyploïdisation, une corrélation positive a été montrée entre la taille des génomes et le nombre d'ETs (Bennetzen et Kellogg, 1997; Gregory, 2005; Vitte et Panaud, 2005) (Fig. 1.5). Particulièrement, les LTR-RTs sont les ETs qui occupent un rôle majeur dans l'augmentation de la taille des génomes chez les plantes (SanMiguel *et al.*, 1998; Kumar et Bennetzen, 1999; Kalendar *et al.*, 2000; Vitte et Panaud, 2005; Piegu *et al.*, 2006; Vitte *et al.*, 2007; Devos, 2010). La capacité des ETs à se multiplier et à s'insérer ailleurs dans les génomes contribue à la dynamique structurale du génome. Ils peuvent générer rapidement un grand nombre de copies d'eux-mêmes et ils constituent la majeure partie des génomes eucaryotes (Bennetzen, 2000; Tenaillon *et al.*, 2010). Chez certaines espèces végétales (comme le blé) plus de 90 % du génome nucléaire est composé d'ETs (Messing et Bennetzen, 2008).

Mais alors, entre ces deux facteurs inflationnistes que sont la polyploïdisation et l'amplification par les éléments transposables, les génomes sont-ils condamnés à grossir inexorablement? (Bennetzen et Kellogg, 1997). Il semble que même si l'évolution des génomes est dominée par une expansion sans doute rapide de la taille des génomes, de nombreux exemples montrent que celle-ci s'accompagnerait également d'une phase réductive plus lente dans le temps (Wolf et Koonin, 2013).

D'une part, il n'est pas rare d'observer après des évènements de polyploïdisation des pertes de gènes dupliqués (Leitch et Bennett, 2004). En prenant les exemples du maïs et du riz, la part de gènes ayant été perdue depuis le dernier évènement de polyploïdisation est estimée à 50 % (Messing *et al.*, 2004; Wang *et al.*, 2005). Il a été montré que les mécanismes de recombinaison sont efficaces pour induire une diminution de la taille des génomes même s'ils ne peuvent pas permettre de ramener la taille d'un génome à ce qu'elle était avant son expansion (Ma *et al.*, 2004). Deux mécanismes principaux ont été identifiés : la recombinaison illégitime et la recombinaison homologue.

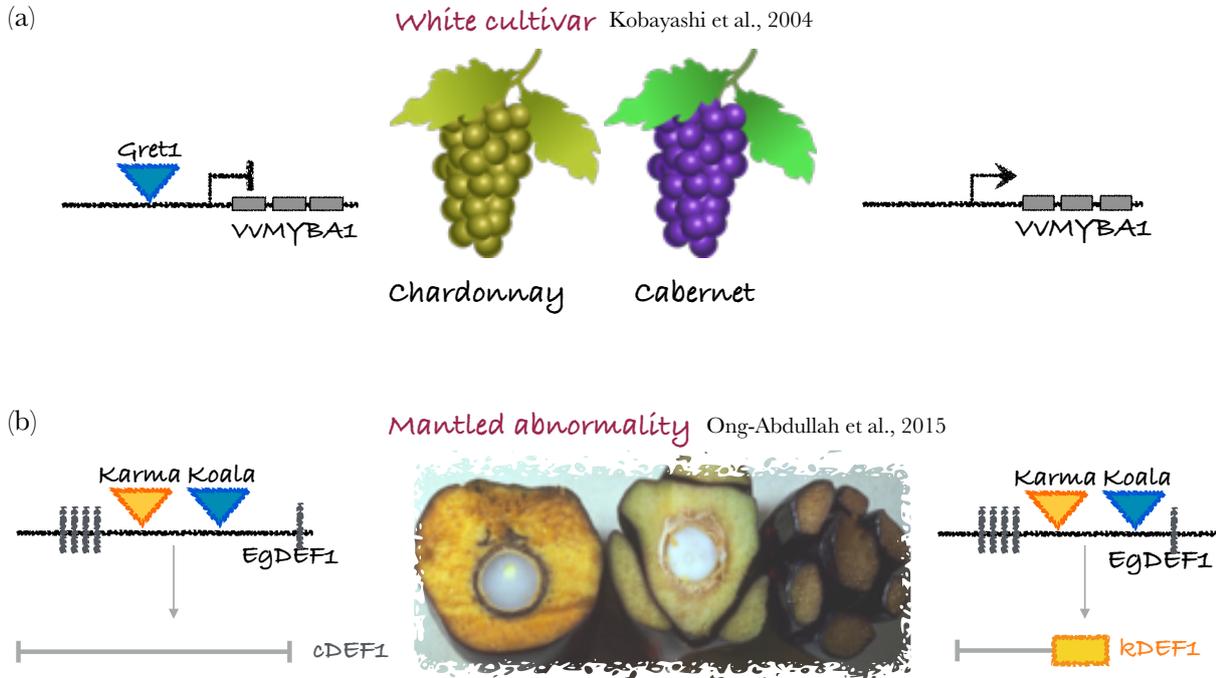
Le mécanisme de recombinaison illégitime implique des petites délétions lors de la reconnaissance de séquences homologues courtes (Ma *et al.*, 2004; Bennetzen *et al.*, 2005). Cette reconnaissance peut être consécutive à des erreurs lors de la réplication de l'ADN ou au moment de la réparation de cassures double-brin (Devos *et al.*, 2002; Vitte et Panaud, 2005).

Le mécanisme de recombinaison homologue inégale implique des délétions plus importantes (Ma *et al.*, 2004; Bennetzen *et al.*, 2005). Pour les LTR-RTNs, un tel évènement est essentiellement reconnu par la présence de solo-LTR, une partie de la région interne de l'élément et d'un de ses

Fig. 1.6 Exemple d'effet de l'insertion d'ETs sur la diversité phénotypique

(a) La couleur blanche des baies de raisin de la variété Chardonnay de la vigne (*Vitis vinifera*) est causée par la présence du rétrotransposon à LTR *Gret1* dans la région promotrice du gène *VvMYBA1*. Cette insertion cause la perte de fonction de ce gène codant pour un facteur de transcription de la synthèse des pigments colorés et entraîne la perte de pigmentation des baies.

(b) Pour le palmier à huile, le transposon *Karma* est inséré dans le gène *EgDEF1*, impliqué dans la mise en place des organes mâles. Une perte de la méthylation en aval de l'insertion de ce gène est associée au phénotype « mantled ».



LTRs ayant été éliminée (Devos *et al.*, 2002).

Ces deux mécanismes n'ont pas la même efficacité selon les espèces considérées (Vitte et Bennetzen, 2006; Hawkins *et al.*, 2008). Chez *A. thaliana*, la recombinaison illégitime serait plus efficace (Devos *et al.*, 2002), alors que chez d'autres espèces, comme pour le riz, ce serait la recombinaison homologue (Bennetzen *et al.*, 2005; Vitte et Panaud, 2005; Vitte *et al.*, 2007; Du *et al.*, 2010). De plus, l'efficacité de ces mécanismes d'élimination couplée à la capacité de certains génomes à être plus « permissifs » que d'autres à l'amplification des ETs pourrait être une des réponses pour expliquer la différence de la taille des génomes (El baidouri et Panaud, 2013).

Les ETs ont été considérés pendant longtemps comme des séquences inutiles voire parasites du génome, qui ne faisaient que s'amplifier sans pour autant avoir d'impact sur le développement et le fonctionnement des organismes. Bien que la plupart de ces éléments soient inactifs, les quelques éléments actifs sont susceptibles d'être réactivés et de se multiplier dans les génomes. Les ETs peuvent s'insérer au niveau de régions régulatrices de gènes ou au niveau de séquences codantes ou non-codantes. Ces insertions peuvent être responsables de mutations et comme pour toute mutation cela peut entraîner des modifications de l'information génétique de l'organisme et créer de la diversité (pour revue Wei et Cao, 2016).

En effet, plusieurs exemples montrent que la mobilité des ETs au sein des génomes, et notamment, leur insertion dans les gènes ou au niveau de régions flanquantes de gènes, peut avoir un impact fonctionnel sur ces gènes. Le premier de ces exemples est évidemment celui du maïs sud-américain et de ses grains présentant différents patrons de pigmentation dus à l'insertion du transposon *Ds* dans le gène *BRONZE* (McClintock, 1950). Autre exemple notable, chez les plantes, celui du raisin et de ses cultivars (Kobayashi *et al.*, 2004), l'insertion du LTR-RT *Gret1* (*Grapevine Retrotransposon 1*) dans la région promotrice du gène *VvMYB1* (*Vitis vinifera MYeloBlastosis Anthocyanin 1*) cause la perte de fonction de ce gène codant pour un facteur de transcription de la synthèse des pigments colorés et ainsi la perte de pigmentation des baies chez les cultivars blancs, tel que le Chardonnay (Fig. 1.6 (a)). Il existe d'autres exemples de phénotypes induits par l'insertion d'ETs et qui ne touchent pas les gènes responsables de la coloration. Le caractère pois ridé de Mendel est en réalité le résultat de l'insertion d'un transposon de la famille *Ac/Ds* au niveau du locus *r* (*rugosus*) ce qui a comme conséquence la réduction de la biosynthèse de l'amidon (Bhattacharyya *et al.*, 1990).

Dans le règne animal, on compte également bon nombre d'exemples pour lesquels les réarrangements de gènes, induits par l'insertion d'ETs, créent de la diversité. Les ETs ont joué un rôle essentiel dans l'évolution des espèces, on peut citer notamment la création du placenta chez les mammifères en intégrant leur génome (Emera et Wagner, 2012). Le phénotype « merle »

chez le chien résulte de l'insertion d'un SINE à la jonction d'un intron et d'un exon (Clark *et al.*, 2006). Un exemple du rôle important que semblent jouer les ETs sur l'évolution des espèces est celui découvert chez deux populations de fourmis de l'espèce invasive *Cardiocondyla obscurior*. L'accumulation rapide d'ETs confèrerait à ces populations invasives une source de mutations contribuant à leur diversification génétique. Cette diversité augmenterait alors leur capacité d'adaptation à de nouveaux environnements (Schrader *et al.*, 2014). Chez la souris, Muotri et ses associés (2005) ont observé que la mobilité d'éléments de type LINE était responsable de changements dans les interactions entre neurones. Sachant que chez l'humain, les ETs comptent pour approximativement 45 % du génome et le plus actif de ces ETs étant l'élément *Line1 (L1)*, l'impact sur les fonctions neuronales des ETs est un sujet d'étude très intéressant pour comprendre la diversité entre individus chez l'Homme (Singer *et al.*, 2010). Récemment, il a été montré que le mélanisme industriel chez les phalènes du bouleau (*Biston betularia*), souvent pris comme exemple pour expliquer la sélection naturelle d'individus face à un nouvel environnement, est dû à l'insertion d'un transposon dans un gène impliqué dans le développement des ailes (Van't Hof *et al.*, 2016).

Malgré les avantages adaptatifs permis par les éléments transposables que nous venons de voir, la présence d'ETs peut aussi avoir des effets délétères sur les organismes si l'insertion modifie la structure et la régulation de gènes essentiels pour le développement. Chez l'Homme notamment, à ce jour, l'implication des ETs dans plus d'une centaine de cancers a été démontré (pour revue Hancks et Kazazian, 2016). Chez les plantes, il a récemment été découvert que l'anomalie mantelé (« mantled » en anglais) du palmier à huile était la conséquence de l'insertion d'un transposon « Karma » dans l'intron du gène homéotique *EgDEF1* (*Elaeis guineensis DEFICIENS 1*, Ong-Abdullah *et al.*, 2015) (Fig. 1.6 (b)).

1.1.2. La régulation des ETs

La capacité des ETs à se déplacer au sein des génomes leur confère un pouvoir hautement mutagène. Bien que leur impact puisse être positif sur les génomes, nous venons de voir qu'ils pouvaient également représenter une menace pour l'intégrité des génomes. C'est pourquoi les aptitudes des ETs à la transcription et à la transposition sont par défaut réprimées par l'action combinée de plusieurs mécanismes épigénétiques. Les régulations épigénétiques ne modifient pas la séquence de l'ADN mais jouent sur la méthylation de l'ADN et les modifications post-traductionnelles d'histones ce qui affecte la conformation de la chromatine.

Fig. 1.7 Caractéristiques des différents états chromatinien chez *A. thaliana*

Il existe 4 états chromatinien chez *A. thaliana* définis par les termes CS1, CS2, CS3 et CS4.

(a) Distribution des marques épigénétiques (11 marques histone et 5-méthylcytosine) dans les différents états chromatinien. L'état chromatinien CS1 est associé aux gènes fortement exprimés, ces domaines sont enrichis en marques permissives H3K4me2 et H3K36me3. L'état chromatinien CS2 concerne majoritairement les gènes activés au cours du développement et est enrichi en marque répressive H3K27me3. L'état CS3 est associé à l'hétérochromatine et est enrichi en marques répressives H3K9me2, H3K27me1, H4K20me1 et 5mC. Enfin, l'état chromatinien CS4 ne présente pas d'enrichissement particulier dans aucune des 12 marques épigénétiques, concerne généralement les gènes faiblement exprimés.

(b) Distribution des fractions d'éléments transposables (TEs), des gènes et des séquences intergéniques pour les 4 états chromatinien.

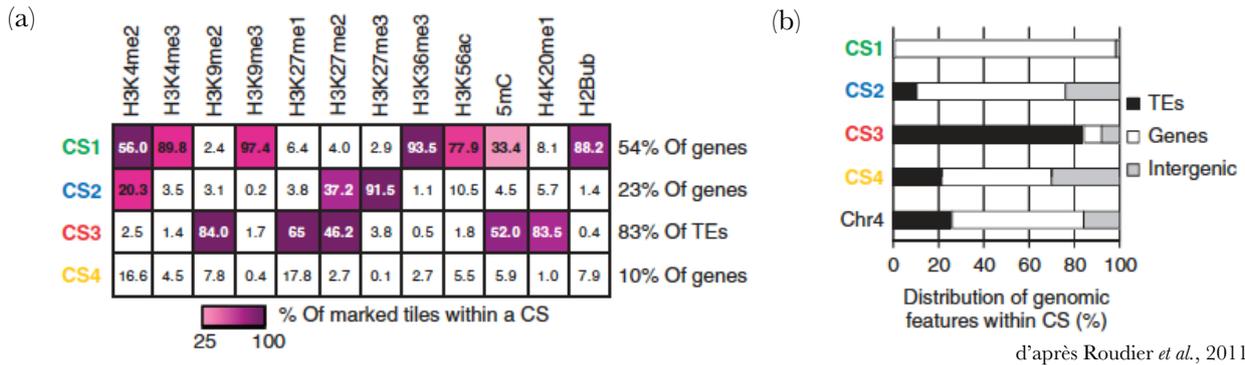
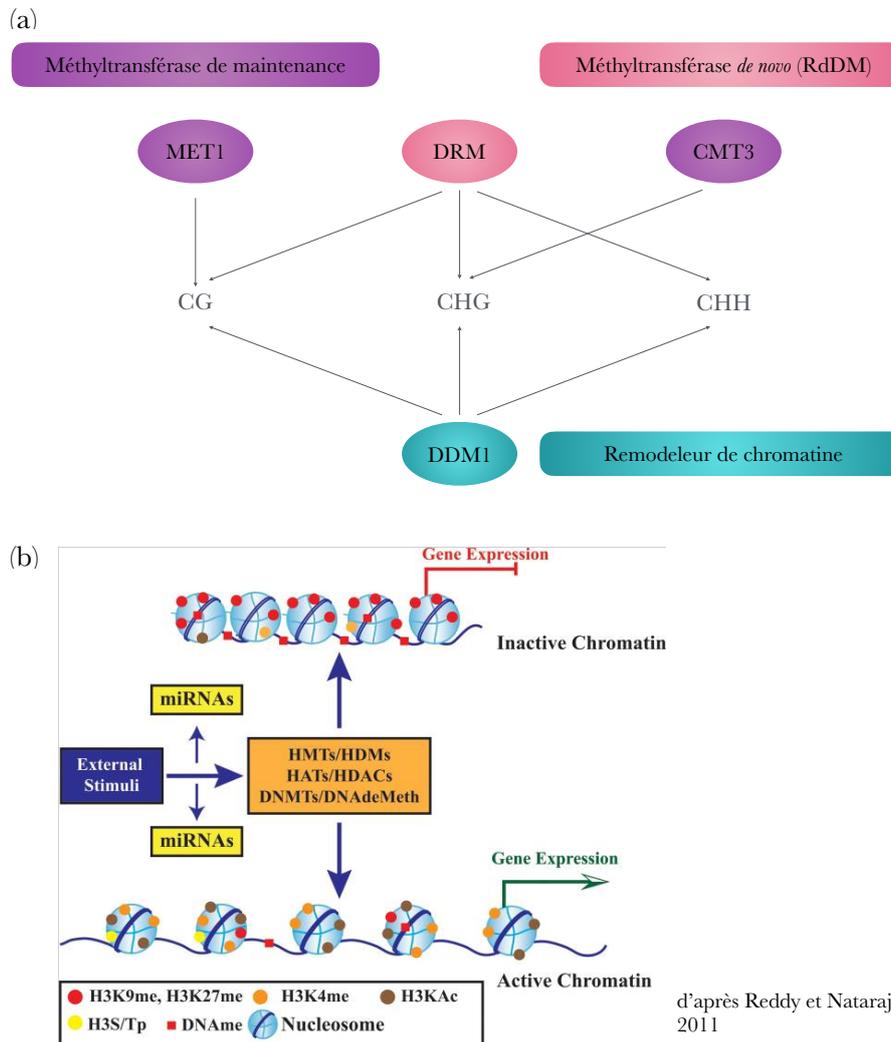


Fig. 1.8 Régulation de l'état de compaction de la chromatine par la méthylation



(a) La méthylation des cytosines dans les différents contextes. En contexte CG, le maintien est assuré par la MÉTHYLTRANSFÉRASE 1 (MET1). En contexte CHG, le maintien se fait par la CHROMOMÉTHYLASE 3 (CMT3). En contexte CHH, le maintien est réalisé via la voie RNA-dépendant DNA Methylation (RdDM). Enfin, la RdDM intervient dans la méthylation de novo via la DNA MÉTHYLTRANSFÉRASE (DRM2). Le rôle supposé de Deficient in DNA Methylation1 (DDM1) est d'aider au recrutement des méthyltransférases dans les 3 contextes.

(b) La méthylation H3Kme est régulée par des histones méthyltransférases/déméthylases (HMTs/HDMs) et H3KAc est régulée par histones acétyltransférases/déacétylases (HATs/HDACs). La méthylation des histones couplée à la méthylation des cytosines par les DNA-méthyltransférases/déméthylases détermine l'état actif ou inactif de la chromatine conduisant à l'expression du gène ou à sa répression.

La diversité des marques épigénétiques associées aux ETs :

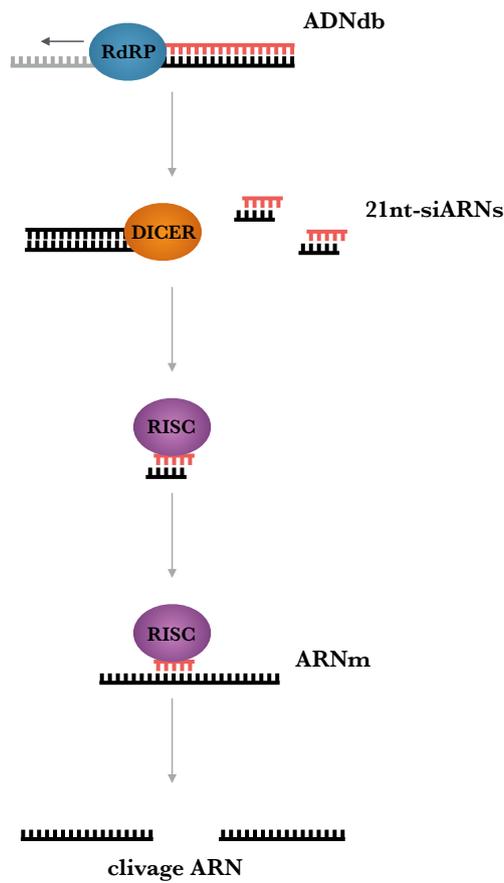
La chromatine, mise en évidence par Flemming (1882), est une structure nucléo-protéique contenant l'ADN. L'unité de base de la chromatine, le nucléosome, est composé d'un octamère d'histones (2 copies de chaque histone H2A, H2B, H3 et H4) autour duquel s'enroule l'ADN. L'histone H1 est quant à elle localisée à l'extérieur du nucléosome au niveau de l'ADN « linker » reliant les nucléosomes entre eux. La chromatine peut subir différentes modifications qui auront pour conséquence de modifier le degré de compaction de l'ADN qui joue un rôle important dans la régulation génique. Cet état de compaction de l'ADN est défini par différents états chromatiniens influencés par la méthylation de l'ADN et des histones (Roudier *et al.*, 2011) (Fig. 1.7).

Parmi les différentes marques associées aux éléments répétés, la mieux caractérisée est la méthylation de l'ADN. Cette méthylation consiste en une modification chimique qui intervient généralement au niveau des cytosines en position 5 du cycle aromatique par ajout d'un groupement méthyl. Cette méthylation intervient surtout pour réguler la transcription : en conduisant à l'hétérochromatinisation, l'accès à la machinerie transcriptionnelle est rendu inaccessible. Contrairement aux animaux où la méthylation des cytosines est restreinte aux dinucléotides CG, chez les plantes la méthylation cible les cytosines en contexte symétrique (i.e. CG et CHG; H=A, T ou C) ou asymétrique (i.e. CHH). Chez *A. thaliana*, la méthylation est presque exclusivement en CG dans les régions géniques et restreinte au corps des gènes (Zhang *et al.*, 2006; Zilberman *et al.*, 2007; Saze et Kakutani, 2011). Cependant, la méthylation est plus abondante au niveau des séquences répétées (Lippman *et al.*, 2004; Lister *et al.*, 2008; Cokus *et al.*, 2008; Zhang *et al.*, 2006; Roudier *et al.*, 2009) ce qui témoigne d'un ciblage plus spécifique de certaines régions.

D'autre part, les modifications des histones sont post-traductionnelles, réversibles et sont majoritairement l'acétylation, la méthylation, la phosphorylation et l'ubiquitinylation. La combinaison de ces modifications constituant le « code des histones » définit des zones actives ou répressives de l'activité transcriptionnelle (Fig. 1.8 (b)).

La plupart des connaissances sur la mise en place de ces marques épigénétiques ciblant les ETs des plantes ont principalement été décrites et découvertes par l'étude de mutants d'*A. thaliana*.

Fig. 1.9 Mécanisme de répression génétique post-transcriptionnelle (PTGS)



Les ARNs double brin synthétisés par une RNA-dépendant RNA polymérase (RdRP) sont reconnus par une ribonucléase DICER qui les dégrade en petits ARNs interférants de 21 nucléotides (21nt-siARNs). Les siARNs sont chargés dans le complexe RNA-Induced Silencing Complex (RISC) pour le guider au niveau de la séquence homologue au siARN pour induire un clivage à cet endroit.

Le contrôle épigénétique et la régulation des ETs chez les plantes :

Les marques épigénétiques que nous venons de voir permettent d'inactiver transcriptionnellement certaines régions pour contenir la prolifération des ETs (Lisch, 2009; Grover et Wendel, 2010; Tenaillon *et al.*, 2010).

La répression des ETs est réalisée par un système de régulation complexe se jouant à 2 niveaux : au niveau transcriptionnel *via* le TGS (Transcriptional Gene Silencing) dans le noyau et au niveau post-transcriptionnel *via* le PTGS (Post-Transcriptional Gene Silencing) dans le cytoplasme. La voie du TGS inhibe la transcription de l'ADN *via* la méthylation de l'ADN et des histones alors que la voie du PTGS cible la dégradation des ARN messagers (ARNm) *via* la reconnaissance de ces transcrits par des petits ARNs (Fig. 1.9).

La méthylation de l'ADN des ETs assure d'une part le maintien de la méthylation et ainsi la transmission d'une mémoire épigénétique et, d'autre part, permet une méthylation *de novo* des séquences non méthylées, comme un ET nouvellement inséré.

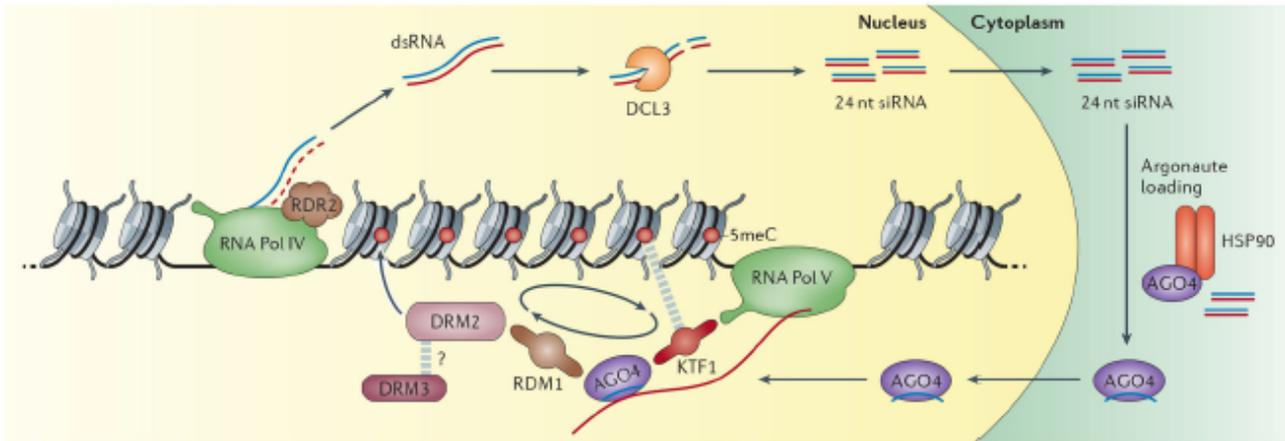
La maintenance de la méthylation est essentielle pour conserver les ETs dans un état inactif. Cette méthylation est assurée par l'action de différentes méthyltransférases spécifiques du contexte nucléotidique dans lequel se trouvent les cytosines (Fig. 1.8 (a)). Ainsi, la méthylation des cytosines en contexte CG est maintenue par la METHYLTRANSFERASE1 (MET1) (Finnegan et Dennis, 1993). D'autre part, la protéine VARIANT IN METHYLATION 1 (VIM1) se lie aux sites CG hémiméthylés (Woo *et al.*, 2007) et favoriserait le recrutement de MET1 (Shook et Richards, 2014). Le maintien de la méthylation au niveau des cytosines en contexte CHG se fait par l'intermédiaire de la CHROMOMETHYLASE3 (CMT3), une chromométhylase spécifique des plantes (Henikoff et Comai, 1998, Lindroth *et al.*, 2001; Cokus *et al.*, 2008). L'action de cette protéine requiert l'action d'une histone méthyltransférase KRYPTONITE (KYP) (Jackson *et al.*, 2002; Du *et al.*, 2012). Dans le cas de la méthylation des cytosines en contexte CHH, le maintien de la méthylation est effectué principalement *via* la voie RdDM (RNA-directed DNA Methylation) par l'action de la méthyltransférase DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) qui assure le rétablissement de la méthylation après chaque cycle de réplication (Cao et Jacobsen, 2002).

En plus des acteurs précédemment cités, il existe d'autres protéines impliquées dans la méthylation de l'ADN. L'action de la protéine de remodelage de la chromatine DDM1 (DECREASE IN DNA METHYLATION 1) a été mise en évidence chez des mutants d'*A. thaliana* où une hypométhylation de l'ADN au niveau des séquences répétées a été observée (Singer *et al.*, 2001; Lippmann *et al.*, 2004) et permet de réprimer certaines familles d'ETs (Tsukahara *et al.*, 2009).

L'action de la machinerie du RdDM est également d'assurer la méthylation *de novo* des cytosines dans les différents contextes nucléotidiques au niveau des transgènes et des ETs (Law et

Fig. 1.10 RNA-directed DNA Methylation

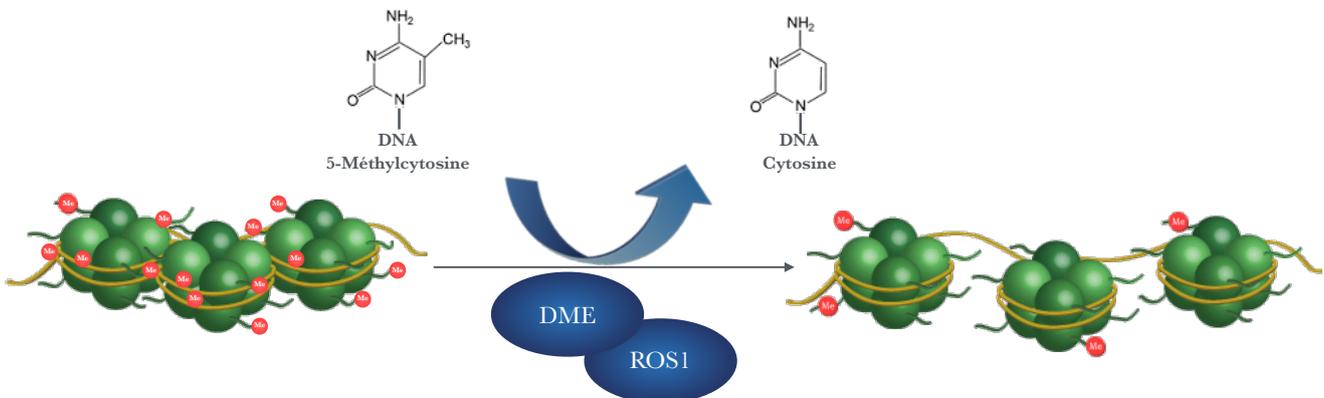
L'ARN polymérase 4 (RNA PolIV) synthétise un ARN simple brin (ligne bleue) à partir de régions hétérochromatiques. La RNA-DEPENDENT POLYMERASE 2 (RDR2) associée à la RNA PolIV synthétise l'ARN complémentaire à partir de la matrice ARN simple brin pour former un ARN double brin (dsRNA). Le dsRNA est clivé en petits ARNs interférents de 24 nucléotides (24 nt siRNA) par la DICER-LIKE 3 (DCL3). Les siRNAs sont ensuite exportés dans le cytoplasme où leur prise en charge est assurée par l'ARGONAUTE 4 (AGO4), qui est facilitée par une HEAT-SHOCK PROTEIN 90 (HSP90), puis retournent dans le noyau. Dans le noyau, les siRNAs ciblent leur séquence complémentaire sur un brin en cours de synthèse par l'ARN polymérase 5 (RNA PolV) (ligne rouge). L'ARN PolV associée au domaine GW/WG de KOW DOMAIN-CONTAINING TRANSCRIPTION FACTOR (KTF1) joue le rôle d'organisateur en interagissant avec AGO4 et une cytosine méthylée (boule rouge). De la même manière, AGO4 interagit avec la DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2) par l'intermédiaire de la RNA-directed DNA METHYLATION1 (RDM1) pour renforcer la méthylation de maintenance. DRM3 est un paralogue de DRM2 requis pour la RdDM mais dont le rôle n'est pas encore complètement connu.



d'après Castel et Martienssen, 2013

Fig. 1.11 Déméthylation active

DEMETER (DME) et REPRESSOR OF SILENCING1 (ROS1) excisent les cytosines méthylées dans les 3 contextes CG, CHG et CHH (ou H = A, T ou G).



Jacobsen, 2010; Haag et Pikaard, 2011) (Fig. 1.10). Ce mécanisme débute par la transcription d'un long ARN simple brin par l'ARN POLYMÉRASE IV (POLIV), spécifique des plantes (Lahmy *et al.*, 2010). Le transcrit formé va ensuite être synthétisé en ARN double brin par l'ARN POLYMÉRASE ARN DÉPENDANTE 2 (RDR2). Ce long transcrit double brin est ensuite clivé par une DICER LIKE 3 (DCL3) en petits ARNs interférents (small interfering RNAs, siRNAs) de 24 nucléotides (24-nt). Les siRNAs sont ensuite pris en charge par la protéine ARGONAUTE4 (AGO4) ou un de ses homologues (AGO6 et AGO9) et forme le complexe RISC (RNA-Induced Silencing Complex). Les siRNAs ainsi chargés vont ensuite interagir au locus cible avec un transcrit complémentaire synthétisé par la POLYMERASE V (POLV), également spécifique des plantes (Wierzbicki *et al.*, 2008; Wierzbicki *et al.*, 2009). Une autre version du modèle est que la POLV sert de plateforme pour recruter directement les siRNAs dans le complexe RISC au niveau de la séquence d'ADN complémentaire (Wierzbicki *et al.*, 2008). La dernière étape est le recrutement de DRM2 qui induit la méthylation *de novo* dans les différents contextes nucléotidiques (Cao et Jacobsen, 2002, Cao *et al.*, 2003).

Ces modifications du profil de méthylation de l'ADN seront perdues si elles ne sont pas maintenues au cours des différents cycles de réplication par les voies de méthylation de maintenance et *de novo*. L'absence de l'activité de ces deux dernières résulte d'une déméthylation dite passive. Il existe cependant une déméthylation active faisant intervenir des déméthylases (pour revue Roldán-Arjona et Ariza, 2009). Chez *A. thaliana*, les protéines DME (DEMETER; Choi *et al.*, 2002) et ROS1 (REPRESSOR OF SILENCING 1; Gong *et al.*, 2002) initient la déméthylation par une action glycosylase/lyase sur le groupement méthyle de la cytosine (Fig. 1.11). En plus de ces 2 protéines, il existe dans le génome d'*A. thaliana* deux autres protéines DML2 et DML3 (DEMETER-LIKE 2/3; Choi *et al.*, 2002).

Même si la plupart des connaissances acquises sur les mécanismes épigénétiques l'ont été chez *A. thaliana*, l'étude de la perturbation du contrôle épigénétique sur les ETs chez le riz suscite un intérêt grandissant. Ainsi, Wei et associés ont trouvé en 2014 que des mutants knock-down *Osdcl3*, présentaient des phénotypes altérés (moins de talles secondaires, angle d'inclinaison de la feuille drapeau plus grand et longueur de la panicule plus courte). DCL3 est un acteur de la voie de répression exercée sur les ETs, cette enzyme intervient dans la coupure des petits ARNs (siRNAs) en 24-nt. Dans cette étude, la levée de la répression par les siRNAs 24-nt dans ces mutants s'accompagne d'une hypométhylation, d'un enrichissement de transposons MITEs et de l'activation de l'expression de gènes contrôlant des traits agronomiques importants.

Récemment, le reséquençage du génome d'un mutant épigénétique du riz, *Oscmt3* a permis d'observer la réactivation et l'insertion de différents éléments transposables (Cheng *et al.*, 2015).

Fig. 1.12 Les polymérases 4 d'*A.thaliana* & d'*O. sativa*

Analyse comparative des séquences d'acides aminés des protéines NRPD1 (NUCLEAR RNA POLYMERASE D 1A) d'*A. thaliana* (At-NRPD1) et d'*Oryza sativa* (Os-NRPD1). La séquence protéique de NRPD1 d'*A. thaliana* a été alignée sur le génome du riz. Deux séquences présentant des homologies avec NRPD1 ont été retrouvées chez *O. sativa* sur le chromosome 4 (Os04g0572600) et sur le chromosome 9 (Os09g0555200). Les acides aminés homologues sont indiqués en turquoise, les non-homologues ne sont pas surlignés.

>Os04g0572600

At-NRPD1-LKRKQLSAESVSSLNEQYKSRNRELKLDIVDLDIQNTNHCSDDQAMKDDNVCITVTVVEASKHSVLELDAI
Os-NRPD1-MKIKRLRLEFIVREIIDQYNTLRKQLNNAIPSVSISNSK-CSVGNQCVKNQTCVMTM-VVQVEINSMSQLDVI

At-NRPD1-RLVLIPLFLLDSPVKGQDQGIKKVNILWTDPRKAPKRNGNHLAGELYLKVTMYGDRGKRNCWTALLETCLPIMDM
Os-NRPD1-KERVIPSILATLLKGFLEFKNVKV-----QCQEDN-----ELVLKVMSEHCKSGKFWATLQNAICIPIMEL

At-NRPD1-IDWGRSHPDNIRQ-CCSVYGIDAGRSIFVANLES AVSDTGKEILREHLLL VADSLSVTGEFVALNAKGWSKQR
Os-NRPD1-IDWERSRPERVYDNFCS-YGIDSAWKFFVESVRSTTDAIGRNIHRQHLLVVADCLSVSGQFHGLSSQGLKQQR

At-NRPD1-QVESTPAPFTQACFSSPSQCFLKAAKEGVRDDLQGSIDALAWGKVPFGTGDQFEIIISP KVHGFTTPVDVYD
Os-NRPD1-TWLSISSPFSEACFSRPAHSFINAAKRDSVDNLSGTLDAIAWGKEPCAGSSGPFKILYSKGSHETKQNEHIYD

At-NRPD1-LLSSTKTMRRNTSAPKSDKATVQPFGLLHSAFLKDIKV-LDGKGI PMS---LLRTIFTWKNI ELLSQSLKRIL
Os-NRPD1-FLHNPEVQALEKNVMDTYRKRTEKTSKRRSALNSEGNATINGGAI SFNQKFLNAKVGIWENIIDMRTSLQNML

At-NRPD1-HSYEINELLNERDEGLVKMVLQLHPNSVEKIGPGVKGIRVAKSK-HGDSCCFEVVRIDGTFEDFSYHKCVLGA
Os-NRPD1-REYTLNEVVTEQDKSCLMEALKFHPRGYDKIGVGIREIKIGVNP GHPSSRCFIVLRNDDTTADFSYNKCVLGA

At-NRPD1-TKI IAPKKMNFYKSK
Os-NRPD1-ANSISPELGSYIENR

>Os09g0555200

At-NRPD1-MIDWGRSHPDNIRQCCSVYGIDAGRSIFVANLES AVSDTGKEILREHLLL VADSLSVTGEFVALNAKGWSKQR
Os-NRPD1-LIDWERSRPERVYDIFCSYGIDSAWKYFVESLRSTTDAIGRNIHRQHLLVVADCLSI SGQFHGLSSQGLKQQR

At-NRPD1-QVESTPAPFTQACFSSPSQCFLKAAKEGVRDDLQGSIDALAWGKVPFGTGDQFEIIISP KVHGFTTPVDVYD
Os-NRPD1-AWLSISSPFSEACFSRPAYSFINAAKRDSVDNLSGALDAIAWGKEPCAGTSGPFKVLYSKGSQKTKQNKNIYD

At-NRPD1-LLSSTKTMRRNTSAPKSDKATVQPFGLLHSAFL-KDIKVL DGKGI PMS---LLRTIFTWKNI ELLSQSLKRIL
Os-NRPD1-FLHNPEVQALEKNFMDTYKQRTEKPSKQRSASFSSKGNATINGGTISVNQKFLDSKVGIWENIIDMRTCLQNML

At-NRPD1-HSYEINELLNERDEGLVKMVLQLHPNSVEKIGPGVKGIRVAKSK-HGDSCCFEVVRIDGTFEDFSYHKCVLGA
Os-NRPD1-REYTLNEVVTEQDKSCLIEALKFHPRGYDKIGVGIREIKIGVNP GHPNSRCFIVQRSDDTSADFSYNKCVLGA

At-NRPD1-TKI IAPKKMNFYKSKYLKN
Os-NRPD1-ANSISPE-LGSYIEKILSN

CMT3 est un acteur majoritaire de la maintenance de la méthylation au niveau des sites CHG d'éléments transposables (Henikoff et Comai, 1998; Cokus *et al.*, 2008).

L'ARN polymérase IV et la répression des ETs via les siRNAs :

Nous avons vu précédemment que POLIV est impliquée dans la voie de répression des ETs. Historiquement, cette enzyme a été identifiée à partir du mutant d'*A. thaliana* *sde4* (*silencing defective 4*) où une perte transitoire de la méthylation est observée dans des lignées surexprimant la GFP (Green Fluorescent Protein, Dalmay *et al.*, 2000). La mutation *sde4* induit une diminution de la méthylation mais ne touche pas la méthylation de maintenance en contexte CG (Chan *et al.*, 2004). SDE4 a ensuite été renommée NRPD1a (NUCLEAR RNA POLYMERASE D1a) et correspond à une des grandes sous-unités de la POLIV. La POLIV est spécifique aux plantes, elle est retrouvée dans le noyau et existe sous 2 formes distinctes, POLIV (POLIVa) et POLV (POLIVb) (Herr *et al.*, 2005; Onodera *et al.*, 2005; Pontier *et al.*, 2005; Lahmy *et al.*, 2009). Ces deux formes diffèrent par leur grande sous-unité NRPD1 (NRPD1a) ou NRPE1 (NRPD1b) formant respectivement avec NRPD2 les POLIVa et POLIVb (Pontier *et al.*, 2005).

Il a été montré dans des mutants *nripd1* d'*A. thaliana*, que l'expression ectopique du gène *FWA* provoque un retard de floraison, et que cette expression est induite par la déméthylation d'ETs de la famille des SINEs en amont du gène *FWA* qui sont réactivés dans ce contexte mutant (Pontier *et al.*, 2005).

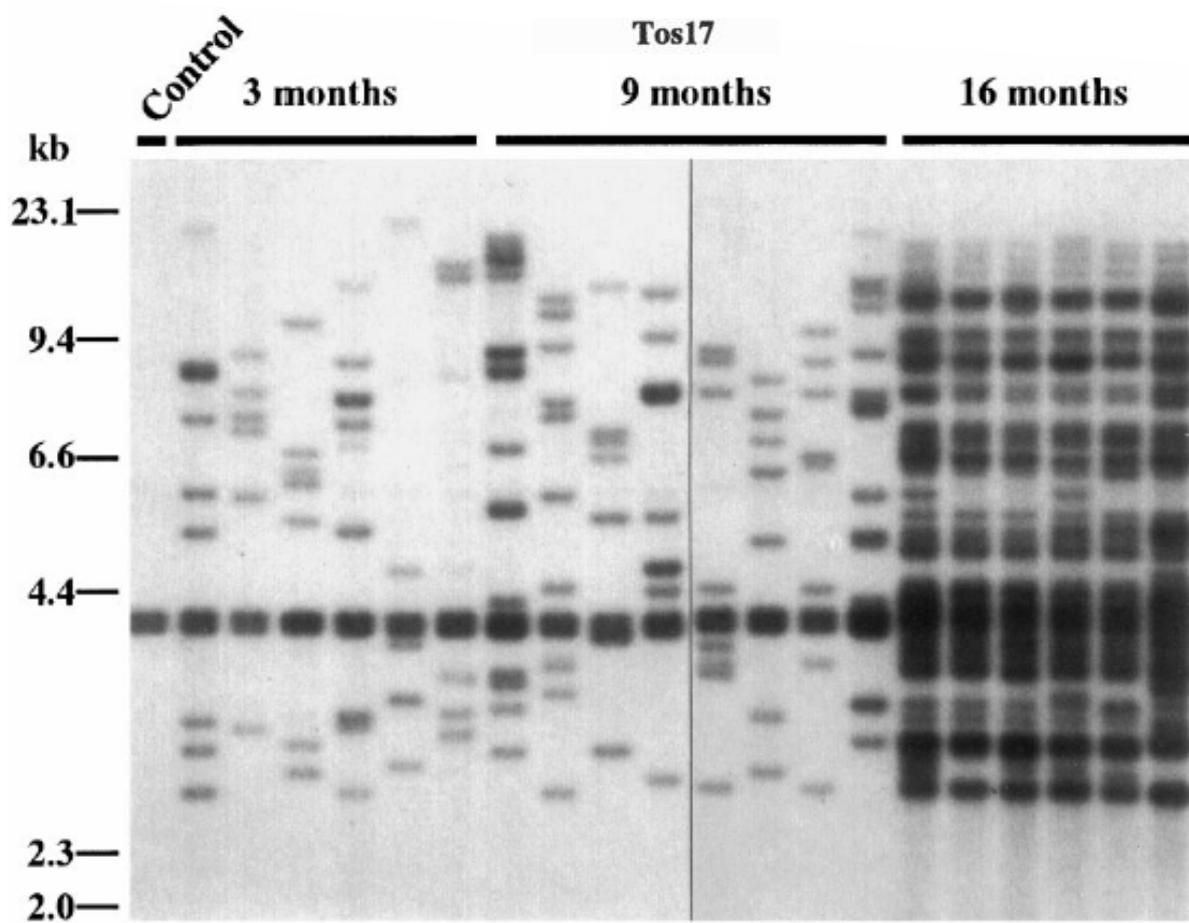
Bien que le nombre d'études portant sur la POLIV chez *A. thaliana* et les mutants affectés dans la voie RdDM chez cette espèce modèle n'aient cessé de susciter un fort intérêt pour la communauté scientifique (Herr *et al.*, 2005; Onodera *et al.*, 2005; Lahmy *et al.*, 2009; Erhard *et al.*, 2013; Yang *et al.*, 2016), nous ne disposons pas du même recul en ce qui concerne le riz, *Oryza sativa*. Une recherche d'homologie de la séquence protéique de NRPD1 a été réalisée chez le riz. *Oryza sativa* spp. *Japonica* contient 2 copies de NRPD1 (Fig. 1.12).

1.1.3. Les ETs et la réponse au stress

La répression des ETs, bien que finement régulés par tous les mécanismes que nous avons vus précédemment, est susceptible d'être levée partiellement ou totalement en cas de stress influant sur la régulation épigénétique du génome. En 1984, B. McClintock émettait cette hypothèse que l'activité des ETs pouvait être induite en réponse à des stress. Une telle réponse au stress, sachant l'impact potentiel des ETs sur l'expression des gènes a été discuté comme pouvant être un moyen d'adaptation aux changements environnementaux pour les ETs (Wessler, 1996).

Fig. 1.13 Accumulation du rétrotransposon *Tos17* pendant la culture cellulaire (Hirochika *et al.*, 1997).

Analyse Southern-blot de *Tos17* dans des plantes Nipponbare régénérées à partir de cals cultivés in vitro pendant 3, 9 et 16 mois. Dans la première colonne, les 2 copies natives de *Tos17* sont détectées dans l'échantillon contrôle.



Bien que la question de savoir comment le stress induit la réactivation des ETs ne soit pas encore complètement comprise (Capy *et al.*, 2000), de nombreux travaux menés chez les plantes ont mis en évidence l'implication de différents types de stress sur la réactivation des ETs.

Culture *in vitro* :

Les études sur l'activation des ETs ont montré que ce n'était pas un phénomène continu, mais qu'elle répond à des stimuli particuliers. La culture *in vitro* (CIV), qui par les ruptures de corrélations entre cellules et les changements environnementaux drastiques qu'elle impose aux plantes, est une source de stress multiples s'exerçant sur les mécanismes préservant l'intégrité du génome.

La culture de cals induit donc une réactivation transcriptionnelle et transpositionnelle d'ETs, le premier exemple décrit chez les plantes étant celui du RTN *Tnt1* dans les cultures de protoplastes du tabac (Pouteau *et al.*, 1991; Grandbastien *et al.*, 1989)

Chez le riz, le premier LTR-RT actif transpositionnellement à avoir été décrit est *Tos17* (Hirochika *et al.*, 1996). Depuis, cet élément fait l'objet de nombreuses études. Il est présent en 2 copies dans le génome du riz *O. sativa* spp. *Japonica* et seule la copie du chromosome 7 est active transpositionnellement. La copie du chromosome 10 est inactive, elle contient des codons stop et des délétions dans ses régions codantes, mais peut cependant être transcriptionnellement réactivée dans des mutants hypométhylés (Hirochika, 2001). Du fait de la capacité de *Tos17* à s'accumuler de façon exponentielle dans les cellules lorsque l'on allonge le temps de culture des cals (Fig. 1.13), ainsi que sa préférence à l'insertion dans les gènes (Miyao *et al.*, 2003), l'induction de son activité a été largement utilisée pour générer des banques de mutants obtenus par mutagenèse insertionnelle (Hirochika *et al.*, 2004; Miyao *et al.*, 2007). Récemment, une étude a montré que *Tos17*, est en réalité un LTR-RT non autonome ne possédant pas de GAG (Sabot, 2014). Les éléments non autonomes utilisent les protéines qu'ils ne peuvent pas coder que synthétisent d'autres éléments. Ainsi, il semble que *Tos17* utiliserait la GAG de son cousin *Lullaby* (Sabot, 2014). *Lullaby* est également activé transpositionnellement par la culture cellulaire (Picault *et al.*, 2009) mais est toutefois moins actif.

La CIV permet également de réactiver d'autres types d'ETs. La rétrotransposition du MITE *miniature Ping* (*mPing*) a été observé dans des cultures cellulaires (Jiang *et al.*, 2003; Kikuchi *et al.*, 2003; Nakazaki *et al.*, 2003). Cet élément est non autonome et utilise la transposase nécessaire à sa transposition codée par les éléments *Ping* et *Pong*. Tout comme pour *Tos17*, bien qu'il soit non autonome, une amplification massive de cet élément a été rapportée pour différentes variétés de riz (Naito *et al.*, 2006) et *mPing* aurait une préférence à l'insertion à proximité de gènes (Naito *et al.*, 2009). De plus, les auteurs ont rapporté que ces insertions à proximité de certains gènes induisaient

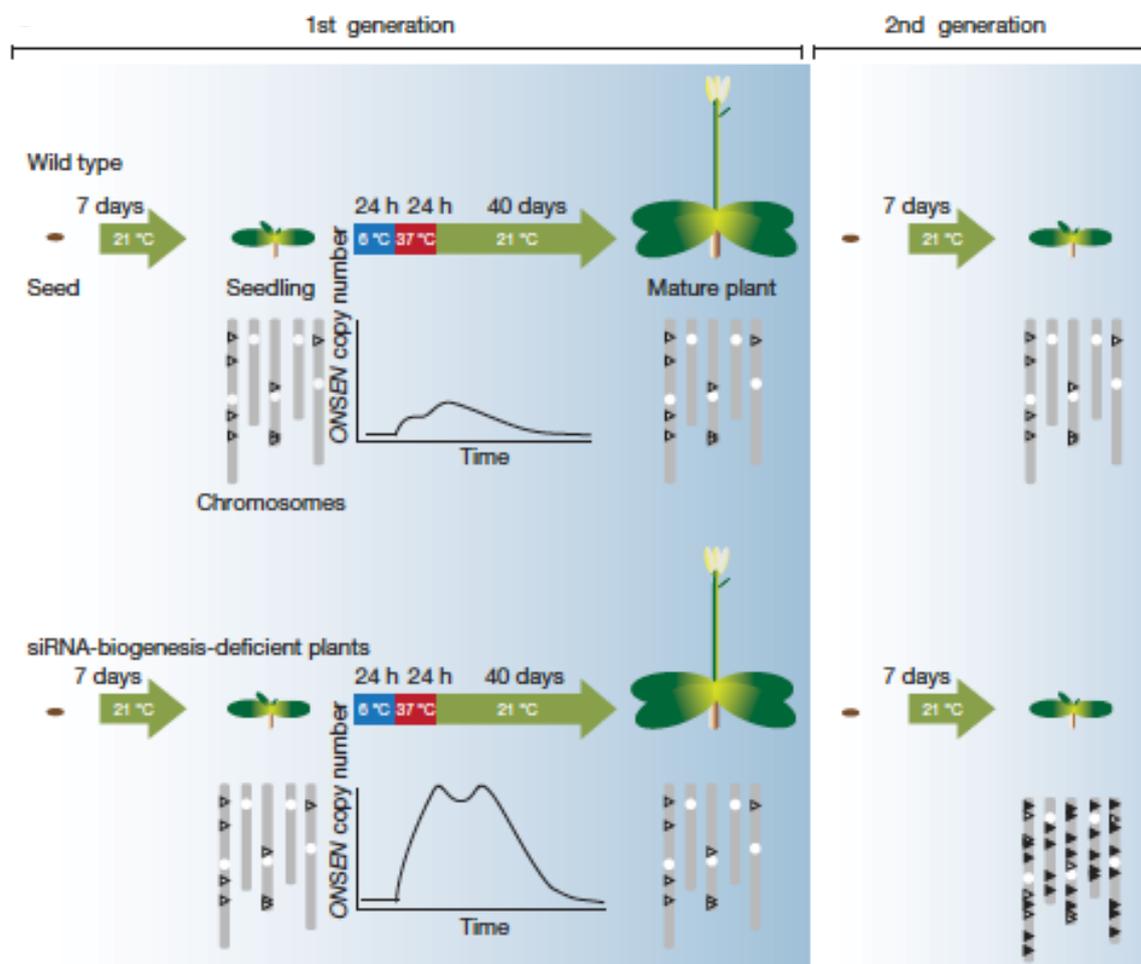
Fig. 1.14 Fréquence d'insertion de l'élément *Karma* (Komatsu *et al.*, 2003)

Generation	Number of Plants Tested ^a	Number of New Insertions	New Insertions per Plant
R0 (callus)	54	0	0
R0 (plants)	13	0	0
R1	15	4	0.27
R2	16	5	0.31
R3	15	10	0.67
R4	23	18	0.78
R5	5	2	0.40
R6	9	2	0.22

^aSum of plants from different lineages.

Fig. 1.15 Activation du rétrotransposon *Onsen* en réponse au stress thermique dans des mutants déficients dans la biogenèse de siRNAs d'*A. thaliana*

La partie haute de la figure représente l'activité transgénérationnelle de *Onsen* dans une plante contrôle soumise à un traitement à la chaleur et la partie basse son activité dans des plantes déficientes en siRNAs. Le graphe illustre la cinétique d'accumulation de *Onsen* dans la plante au cours du stress thermique. Les triangles blancs sur les chromosomes désignent les 8 copies natives de *Onsen* de *A. thaliana* accession Colombia. Les triangles noirs désignent les nouvelles copies de *Onsen* acquises après le traitement à la chaleur. Les cercles blancs montrent les positions des centromères.



d'après Ito *et al.*, 2011

leur expression en condition de stress, ce qui en a fait un nouvel outil génétique pour étudier le changement d'expression de certains gènes en réponse à des stress (Yasuda *et al.*, 2013).

Un autre élément est réactivé par la culture cellulaire, le LINE *Karma* (Komatsu *et al.*, 2003). Les auteurs ont montré que l'activation transcriptionnelle de cet ET était initiée lors de la culture de cals hypométhylés de riz. Cependant, cette activation n'est pas suffisante pour activer transpositionnellement *Karma*, il faut attendre la deuxième génération issue des plantes régénérées pour pouvoir observer des événements de transposition (Fig. 1.14).

Le maintien de la répression des ETs est donc relâché par le passage en *CIV* qui altère le paysage épigénétique du génome (Stroud *et al.*, 2013).

Stress biotiques :

L'implication d'autres facteurs de stress ont également été rapportés comme les stress biotiques qui peuvent être de différentes origines pathogéniques, fongiques ou bactériennes et, dans ces conditions certains ETs sont également réactivés. C'est le cas de la réactivation de l'élément *Tnt1*, chez le Tabac, dont la transcription est induite en présence d'éliciteurs bactériens et fongiques (Grandbastien *et al.*, 1998; 2005). Il est intéressant de noter pour cet exemple, que les motifs régulateurs présents au niveau du LTR de *Tnt1* sont très proches de motifs régulateurs retrouvés dans différents gènes de défenses contre les pathogènes. Ceci serait un exemple de l'adaptation d'un organisme à un changement environnemental (l'attaque fongique) comme suggéré par Wessler en 1996.

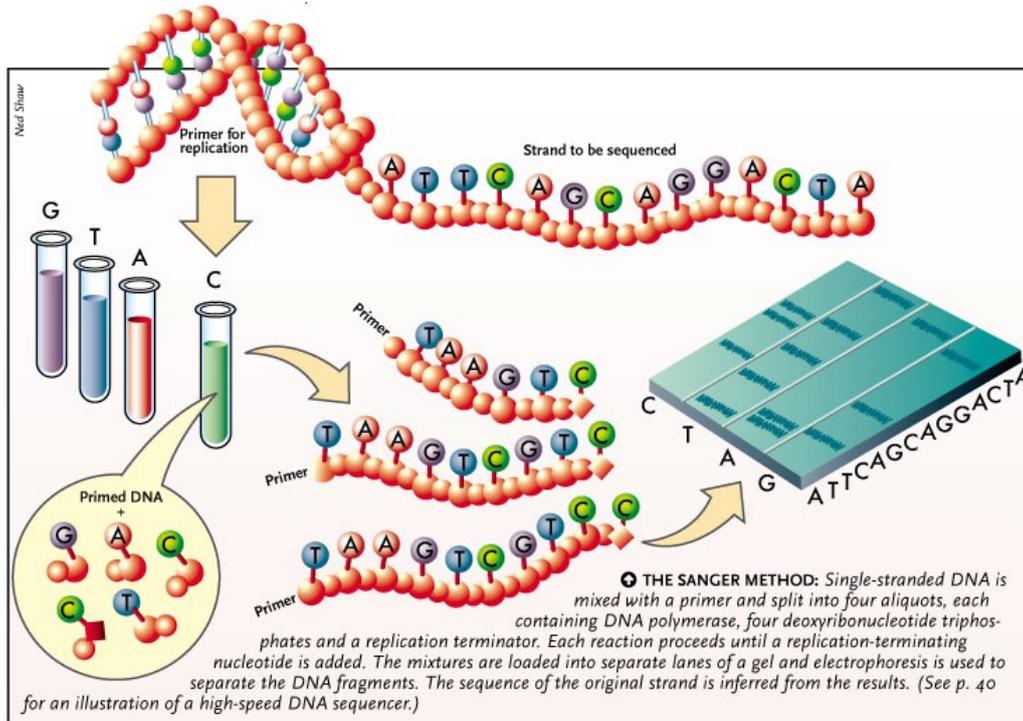
Stress abiotiques :

Enfin, l'implication des stress abiotiques est également importante dans l'activation de certaines familles d'ETs. L'impact du stress thermique a été mis en cause dans la réactivation transcriptionnelle de nombreux ETs chez *A. thaliana* (Pecinka *et al.*, 2010; Tittel-Elmer *et al.*, 2010). Cependant, cette réactivation ne semble pas suffisante pour qu'il y ait transposition. En effet, l'activité transpositionnelle des ETs chez *A. thaliana* a pu être montrée uniquement dans les mutants affectés dans la voie du RdDM, c'est le cas du RTN *Olsen* (Ito *et al.*, 2011; Matsunaga *et al.*, 2012) (Fig. 1.15).

D'autres exemples de réactivation d'ETs induits par des stress abiotiques ont été rapportés : *Tto1* en réponse à l'acide jasmonique (Takeda *et al.*, 1999); *mPing* en culture cellulaire soumis à un stress salin ou froid (Naito *et al.*, 2009; Yasuda *et al.*, 2013); *Tnt1* en réponse à une blessure et d'autres facteurs abiotiques (Mhiri *et al.*, 1997).

Fig. 1.16 Illustration de la méthode Sanger

La méthode de séquençage Sanger est basée sur la fragmentation de l'ADN génomique. Chaque fragment est mis en réaction avec des désoxyribonucléotides (bulle jaune) qui vont permettre la polymérisation. Puis, l'incorporation aléatoires de didésoxyribonucléotides (ddNTP avec le carré dans la bulle jaune) vont induire l'arrêt de la polymérisation et ainsi la formation de fragment d'ADN de longueurs différentes. Les fragments d'ADN vont ensuite migrer par électrophorèse capillaire. Un signal fluorescent correspondant au ddNTP incorporé sera émis qui permettra de déduire la séquence d'ADN.



1.2. Détection des ETs par les techniques de séquençage haut débit (Next-Generation Sequencing, NGS)

Les ETs occupent donc une place importante dans les génomes et peuvent, parfois, être réactivés. L'activité des ETs ne conduisant pas toujours à un phénotype visible, l'identification systématique des événements de transposition est nécessaire pour étudier leur dynamique au sein des génomes. Pour cela, différentes techniques ont été mises au point pour les identifier.

1.2.1. Les techniques d'analyse des données génomiques

Le séquençage de l'ADN est apparu vers la fin des années 1970. Une des méthodes proposées et qui fut la plus utilisée est celle développée par Frederick Sanger (Sanger *et al.*, 1992). Cette technique a révolutionné en son temps le domaine de la génomique et permis d'initier dans les années 1990 le séquençage de grands génomes. Cette technique se base sur l'initiation de la polymérisation de l'ADN à partir d'une amorce complémentaire du fragment d'ADN à séquencer (Fig. 1.16). Puis, une étape d'élongation est réalisée par une ADN polymérase qui utilise comme substrat les 4 désoxyribonucléotides (dATP, dTTP, dCTP et dGTP). Des didésoxyribonucléotides, dont le groupement OH secondaire en 3' est réduit, sont également ajoutés à la réaction. Ces derniers agissent comme terminateurs de chaîne lorsqu'ils sont incorporés, la synthèse ne peut plus se poursuivre au delà et permet d'obtenir un ensemble de fragments d'ADN de tailles différentes. Ces terminateurs de chaînes sont marqués par différentes molécules fluorescentes pour chaque dinucléotide (4 séries de traitement sont donc nécessaires, une pour chaque nucléotide). Les fragments sont ensuite récupérés par électrophorèse sur gel polyacrylamide pour les séparer en fonction de leur longueur. On lit ensuite la séquence en parallèle, en fonction de la taille des fragments, des nucléotides successifs.

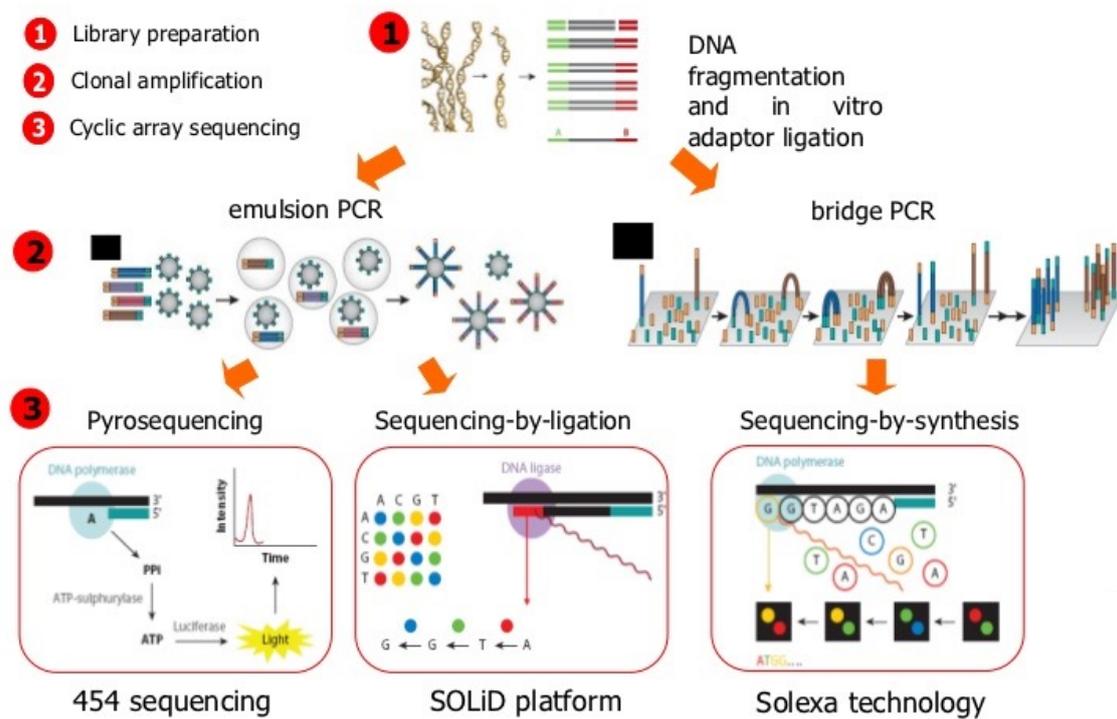
Le séquençage des génomes par cette technique aura nécessité un effort de plus de 10 années de travail. Ce n'est qu'en 2000 que la séquence du génome entier d'*A. thaliana* fut disponible (The Arabidopsis Genome Initiative 2000) et il aura fallu attendre 2004 pour celle de l'humain (International Human Genome Sequencing Consortium).

Puis dans les années 2005, un ensemble de méthodes pouvant produire des millions de séquences en un seul run sont apparues, ce qui leur a conféré leur nom de techniques de séquençage haut-débit. Ces nouvelles techniques sont dites de deuxième génération, relayant ainsi la méthode Sanger au rang de technique de séquençage de première génération.

Fig. 1.17 Grands principes des nouvelles techniques de séquençage NGS

Les 3 techniques de séquençage nouvelle génération utilisent pour le séquençage 3 grandes étapes: (1) la préparation d'une librairie d'ADN marqué avec des adaptateurs dont le principe est commun aux 3, (2) une amplification clonale et (3) le séquençage en parallèle des fragments d'ADN. Les techniques 454 et SOLiD réalisent l'amplification clonale par mise en émulsion des fragments ADN avec les réactifs de PCR alors que la technique Solexa (Illumina) utilise la fixation de l'ADN sur une puce où l'amplification est réalisée au niveau d'un pont d'amplification. Enfin, la méthode de lecture des séquences d'ADN pour la technique 454 utilise le pyroséquençage (lecture de la fluorescence émise après libération du groupement phosphate). La technologie SOLiD lit la fluorescence émise à chaque incorporation de nucléotides. Et la technique Illumina/Solexa détecte un signal lumineux émis à une certaine longueur d'onde en fonction du nucléotides après excitation laser.

Next-generation DNA sequencing



Principe du séquençage haut-débit :

L'essor de ces techniques de séquençage aura entraîné une explosion de la quantité des données générées. Les différentes plateformes commerciales de ces séquenceurs de 2^{ème} génération sont : Roche 454 Genome Sequencer, Illumina Genome Analyzer et Life Technologies SOLiD System. L'amélioration de ces techniques au fil des années ont permis de séquencer plusieurs millions de séquences en un minimum de temps et pour un coût bien plus faible que ce qu'il n'était possible avec la méthode Sanger. Il y a douze ans, la séquence du génome humain était publiée après une dizaine d'année de travail et près de 3 milliards de dollars dépensés, alors qu'aujourd'hui il ne suffirait que d'environ une journée pour séquencer un génome humain pour 1000 dollars (Miller *et al.*, 2015; <http://www.genome.gov>).

Les techniques de séquençage ne cessant d'évoluer, nous sommes maintenant à l'ère des séquenceurs de 3^e génération qui permettent de séquencer directement une molécule d'ADN en temps réel (Greenleaf et Block, 2006; Eid *et al.*, 2009; Munroe et Harris, 2010). Cependant, cette méthode est toujours en cours d'amélioration et j'aborderai ce type de séquenceurs dans la partie 4 de ce manuscrit et discuterai de l'avenir prometteur de cette technologie pour la détection de variations structurales dans les génomes sous la forme d'un article récemment soumis.

Le séquençage haut-débit : techniques de 2^{ème} et 3^{ème} générations :

Les différentes techniques de séquençage de seconde génération ont été développées pour réduire le temps et le coût du séquençage. Ces améliorations ont été rendues possible par le séquençage de milliers de séquences à la fois. Les technologies proposées, Illumina, Roche/454 (Margulies *et al.*, 2005), SOLiD (Valouev *et al.*, 2008), etc.. se basent sur les mêmes grandes étapes (Fig. 1.17), à savoir : (1) La préparation d'une banque d'ADN qui consiste en la fragmentation aléatoire de l'ADN, dont les fragments seront liés à des adaptateurs à chacune des 2 extrémités grâce à une ligase. Les adaptateurs permettront de fixer les fragments à un support : microbilles pour Roche/454 et SOLiD, puce pour Illumina. Les adaptateurs ainsi fixés serviront de site d'amorçage pour l'amplification et le séquençage des fragments d'ADN. (2) une étape d'amplification par PCR de chaque fragment d'ADN réalisée soit en émulsion pour Roche/454 et SOLiD soit par une amplification en ponts pour Illumina, et (3) le séquençage en temps réel de chacun des fragments ADN grâce à la détection d'un signal fluorescent émis par chaque nucléotide ajouté soit par pyroséquençage pour Roche/454, soit par ligation pour SOLiD, soit par incorporation de terminateur réversible pour Illumina.

La technologie utilisée dans cette thèse est la technique développée par Illumina (Bentley *et al.*, 2008) qui utilise des terminateurs de chaîne réversibles. Cette méthode utilise la fixation du

premier adaptateur lié à l'extrémité du fragment d'ADN sur un support solide (puce). Pour chaque fragment d'ADN sera ensuite généré un pont d'amplification par fixation du second adaptateur lié à l'autre extrémité du fragment d'ADN à séquencer à une séquence complémentaire fixée sur la puce (Fedurco *et al.*, 2006). Ces adaptateurs vont alors servir d'amorce pour l'amplification clonale de chaque fragment par PCR. Chaque nucléotide, marqué par un fluorophore, émettra un signal lumineux d'une longueur d'onde spécifique lorsqu'il sera ajouté et sera détecté après excitation laser. La dernière étape sera la déprotection par photoclivage du blocage de la terminaison en 3' qui restaurera le groupement fonctionnel 3'OH.

Grâce aux avancées dans les nanotechnologies, la troisième génération de séquenceurs est en train d'émerger. La principale avancée technique des nouveaux séquenceurs est leur capacité à séquencer directement une molécule d'ADN, et cela sans aucune amplification au préalable. Actuellement l'une des principales technologies de 3^{ème} génération utilisée et disponible sur le marché est celle proposée par Pacific Bioscience.

Pacific Bioscience a développé le premier instrument capable de séquencer en temps réel une molécule d'ADN. La méthode « PacBio », ou encore SMRT (Single Molecule Real Time) utilise des nanostructures appelées ZMW (Zero-Mode Waveguide) de 100 nm de diamètre. Ce diamètre, plus petit que les longueurs d'onde utilisées pour l'excitation des fluorochromes, permet d'éviter toute perturbation photonique et assure le confinement du signal fluorescent pour une lecture optimale du nucléotide. Au fond des ZMW, se trouve une molécule d'ADN polymérase immobilisée qui incorpore successivement des nucléotides liés à un fluorophore selon l'ordre de la séquence d'ADN. L'incorporation du nucléotide entraîne l'excitation du fluorophore dont le signal fluorescent est lu par un détecteur de fluorescence ultra sensible pour permettre leur identification. Une fois l'incorporation du nucléotide achevée, le fluorophore est libéré hors du ZMW, ce qui met fin au signal lumineux.

Une autre technologie de 3^{ème} génération a été proposée il y a maintenant 3 décennies (Deamer *et al.*, 2016) et est développée par Oxford Nanopore Technologies (ONT). Cette technique utilise un système de séquençage de la taille d'une clé USB, le MinION. La stratégie ONT est basée sur la démocratisation du séquençage, dont le faible coût et la portabilité de ce séquenceur permettra son utilisation à grande échelle. Le principe est simple : un complexe de protéine est assemblé autour d'un nanopore situé sur une membrane. Un ADN double brin va s'amarrer au nanopore grâce à une enzyme à laquelle il est lié. Une hélicase située au niveau du nanopore permet de séparer la double hélice d'ADN permettant ainsi le passage d'un seul brin à l'intérieur du nanopore. Un courant électrique est appliqué et chaque nucléotide du brin d'ADN est déterminé lorsqu'il passe au travers du nanopore, chacun des 4 types de bases possédant sa propre signature

électronique. De plus, l'autre brin de la molécule est également séquencé, ce qui crée une information redondante qui permet de limiter les erreurs de séquençage.

Principales caractéristiques des séquenceurs : Forces et faiblesses :

La méthode Sanger avait l'avantage de pouvoir produire des séquences assez longues pouvant atteindre le millier de nucléotides. Cependant, cette technique souffre de nombreuses limitations, la plus grande étant le rendement, ne permettant de séquencer qu'une centaine de séquences à la fois seulement. Les autres limitations sont dues au principe de la technique en lui-même, l'ADN doit être fractionné, amplifié, séquencé (étapes très coûteuses en argent et en temps), puis chaque échantillon sera lu sur une machine qui ne peut lire qu'un à la fois.

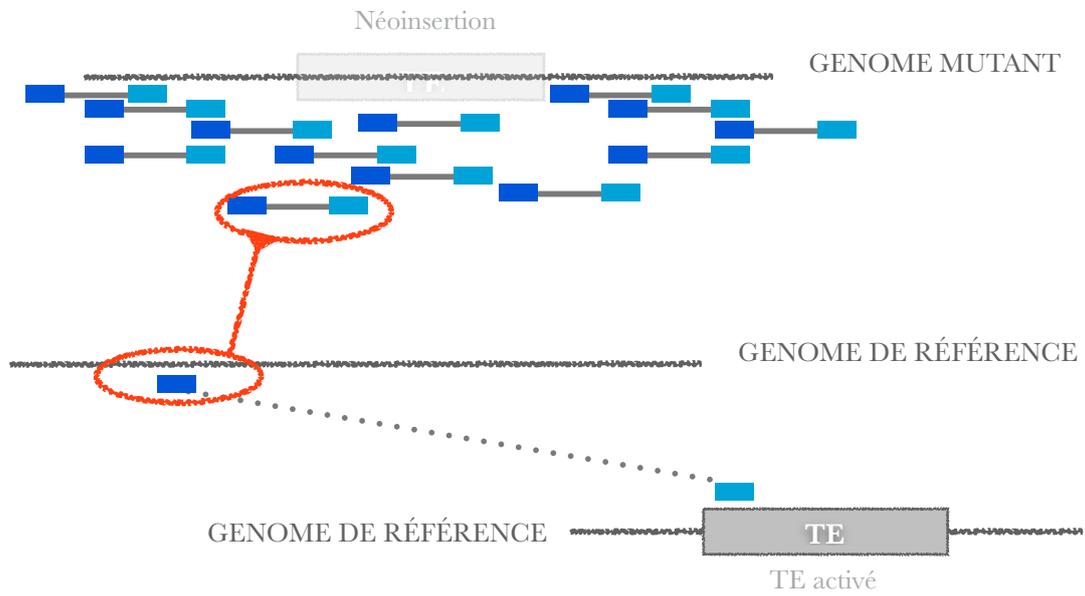
L'intérêt premier des séquenceurs haut-débit réside donc évidemment dans le fait qu'ils ont permis la parallélisation du séquençage de milliers de séquences à la fois. Ces séquenceurs permettent donc d'obtenir plus de séquences, rapidement et le prix du séquençage ramené au gain de temps et au nombre de séquences obtenues en a fait un moyen de séquençage peu onéreux. De plus, là où la méthode Sanger demandait une étape de clonage de l'ADN suivie d'une électrophorèse, ces nouvelles techniques reconstituent et permettent de lire directement la séquence d'ADN par détermination des nucléotides intégrés en temps réel. Néanmoins, ces nouvelles technologies présentent une faiblesse par rapport à la méthode Sanger, elles ne permettent d'obtenir que peu de nucléotides à la fois.

La technique de pyroséquençage 454 permet tout de même de produire des lectures plus longues qu'Illumina ou Solid, d'environ 500 nt. Néanmoins, cette méthode entraîne souvent des erreurs dans les séquences contenant des répétitions de mononucléotides (Rothberg et Leamon, 2008). La technologie proposée par Life Technologies, SOLiD permet d'obtenir des taux d'erreur de séquençage plus faible grâce à la combinaison de 2 fluorophores pour la détection de chaque nucléotide. Cependant, les lectures générées sont très courtes d'environ 50 nt mais avec un taux d'erreur inférieur à 1 %. Du fait de ces différentes caractéristiques, ces techniques sont désormais considérées comme obsolètes et on leur préfère la technique Illumina.

L'avantage de la technique Illumina est sa capacité à générer une très grande quantité de données qui se compte en Gigabases. En revanche, à ses débuts, la longueur des séquences obtenues étaient assez courtes, de l'ordre de quelques dizaines de nucléotides (50 à 100 nt). Les améliorations apportées à cette technologie permettent d'obtenir aujourd'hui des lectures plus longues avec les nouvelles versions de séquenceurs (jusqu'à 350 nt) et avec un taux d'erreur dans le séquençage inférieur à 2 %.

Fig. 1.18 Le Paired-End Mapping (PEM)

Le séquençage de lectures pairées au niveau d'une nouvelle insertion d'élément transposable (ET) donnera des lectures comme indiquées sur le schéma (lectures gauche et droite en bleus). Au niveau de l'insertion, après alignement sur le génome de référence, les paires de lectures sont à cheval entre la séquence du génome de référence et l'ET inséré.



Malgré l'obtention de séquences plutôt courtes, ces techniques ont permis de révolutionner le séquençage de génome entier. Alors qu'il fallait encore au début du siècle déboursier 10 000 \$ pour séquencer 1 million de nucléotides, il n'en faut maintenant que 0,1 \$ (source : genome.gov). Cette évolution du séquençage est telle que les biologistes ne se contentent plus de la séquence d'un seul génome. Maintenant des projets de séquençage de milliers de génomes sont mis en oeuvre, par exemple : le projet 1000 génomes de l'humain (Birney et Soranzo, 2015), celui d'*A. thaliana* (The 1001 Genomes Consortium, 2016).

Désormais avec les nouvelles technologies de 3^{ème} génération et le séquençage de molécule unique il est possible d'obtenir des séquences plus longues pouvant atteindre plusieurs kb. Néanmoins ces technologies sont encore en cours d'amélioration car elles entraînent un fort taux d'erreurs (10 à 15%) alors que les techniques de 2^{ème} génération moins de 1%. Cependant le couplage de données issues des technologies PacBio ou Nanopore avec des données obtenues avec les séquenceurs Illumina, ou l'augmentation de la couverture de séquençage permettraient de corriger les erreurs de séquençage.

1.2.2. Détection des variations structurales associées aux éléments transposables

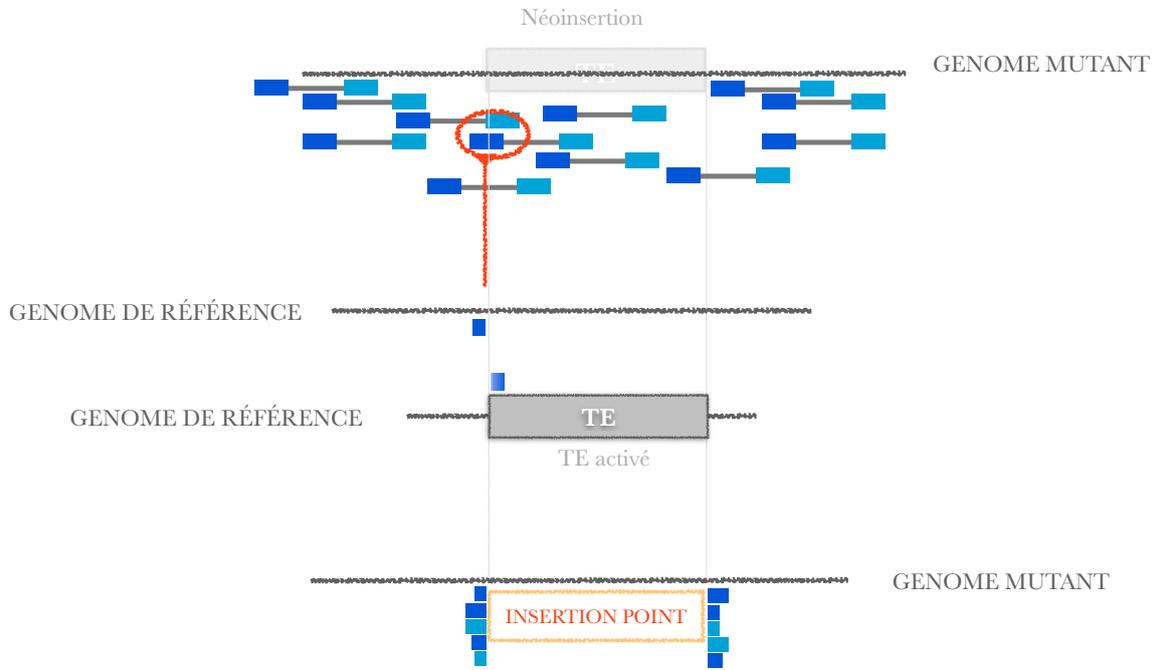
L'identification des ETs dans un génome est essentielle pour étudier la dynamique structurale et fonctionnelle des génomes. Ces dernières années les programmes de détection des variants structuraux sont allés *crescendo* en même temps que les améliorations apportées aux technologies NGS. Il existe, à ce jour, 3 approches différentes pour détecter un évènement de transposition : la méthode dite de Paired-End Mapping (PEM), la méthode de détection des lectures coupées ou chimériques (split-reads) et la méthode dite du Depth Of Coverage (DOC).

Le paired-end mapping :

Cette méthode de détection a été rendue possible par certaines technologies NGS, comme Illumina, qui permet de séquencer les lectures en paired-end, i.e. des 2 extrémités d'un même fragment d'ADN. Les 2 séquences sont séparées par une distance connue qui correspond à la taille des fragments générés lors de la préparation de la banque d'ADN. Ces séquences sont alignées sur le génome de référence et toutes les paires de lectures qui ne s'alignent pas avec la distance attendue, ou qui s'alignent sur 2 chromosomes différents, sont récupérées (Fig. 1.18). Ces dernières sont désignées par le terme « discordantes » et indiquent la présence d'une variation structurale.

Fig. 1.19 La méthode des lectures coupées

La détection de l'insertion d'un élément transposable (ET) se fait par l'identification de lecture coupée où une partie de la lecture s'aligne sur le génome de référence et l'autre partie sur un ET. Le point exact de l'insertion est déterminé à l'endroit de la coupure (trait fin gris).



L'alignement des lectures de ces paires sur une base de données d'ETs permet ensuite d'identifier si la paire discordante est le résultat de l'insertion d'un ET.

Cette technique de séquençage a été utilisée par Korb et ses associés en 2007, puis après un alignement des lectures sur le génome de référence ils ont pu identifier que 90 % des variations structurales étaient dues à des insertions d'ET de la famille des LINE1. Chez les plantes, cette méthode de détection a permis d'identifier un élément actif, appelé *Evadé*, qui est réactivé transpositionnellement dans des lignées epiRILs (epigenetic Recombinant Inbred Lines) hypométhylées d'*A. thaliana* (Mirouze *et al.*, 2009). Chez le riz, la méthode PEM a été utilisée pour détecter l'activité de différentes familles d'ETs dans des lignées *Tos17* (Sabot *et al.*, 2011).

Cette technique a donc montré son efficacité dans la détection de mouvement d'ETs. Elle a l'avantage d'augmenter la résolution de la détection des variations structurales et permet la validation par PCR des sites identifiés. Cependant, elle présente quelques limitations. La première étant évidemment la nécessité de disposer d'un génome de référence. De plus, les séquences obtenues par séquençage étant assez courtes, il est nécessaire que les génomes comparés soient le plus proche possible phylogénétiquement pour pouvoir faire une cartographie propre des lectures. En ce qui concerne la détection des ETs par cette technique, le premier problème réside dans la nature même des ETs. Ce sont des éléments répétés et leur identification par la technique de PEM exige qu'une des lectures de la paire soit ancrée au niveau d'une région unique du génome. Dans le cas d'insertion d'ETs dans des régions complexes du génome, leur identification n'est pas possible et entraîne donc des faux négatifs.

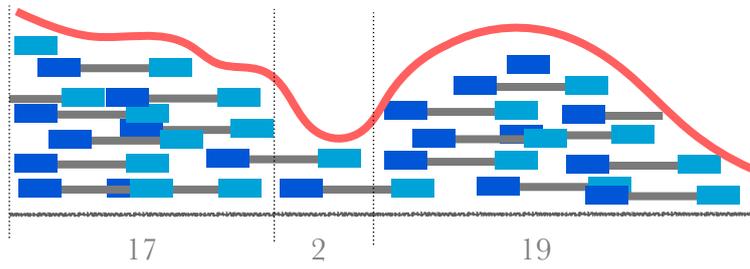
L'alignement des lectures coupées :

Une autre méthode d'identification des variants structuraux est possible depuis que les longueurs des lectures générées par séquençage NGS sont plus grandes (Medvedev *et al.*, 2009; Ye *et al.*, 2009). L'identification d'une lecture coupée se fait par l'analyse des lectures interrompues après alignement des lectures sur le génome de référence (Fig. 1.19). Une partie de la lecture s'aligne à un endroit du génome, l'autre partie ailleurs. Cette méthode permet donc définir la position exacte de l'évènement d'insertion d'un ET. Cependant, cette approche dépend fortement de la taille des lectures qui doivent avoir une longueur suffisante et ne peut se faire qu'au niveau de régions présentes dans le génome de référence (Zhao *et al.*, 2013).

Cette méthode peut être utilisée conjointement avec le PEM pour renforcer la détection d'évènements de transposition (Jiang *et al.*, 2012; Hénaff *et al.*, 2015).

Fig. 1.20 Méthode de la profondeur de lecture (DOC)

Après alignement des lectures séquencées sur la séquence du génome de référence, la profondeur de lecture peut être calculée à chaque position du génome.



GENOME DE
RÉFÉRENCE

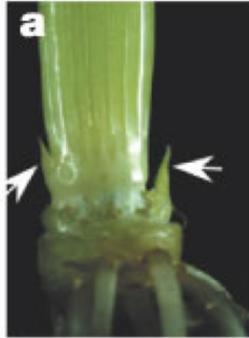
Le Depth of Coverage :

La troisième méthode, celle du DOC, consiste en une évaluation de la profondeur de couverture de différentes régions du génome dans chaque échantillon (Fig. 1.20). Cette valeur de profondeur correspond au nombre de fois qu'a été lu un nucléotide d'une région au cours du séquençage. Cette valeur est ensuite corrélée avec le nombre de copies de la région grâce à une normalisation avec le nombre de lecture correspondant à une région unique. Cette technique s'est avérée efficace dans la détection de variations structurales (Chiang *et al.*, 2009; Yoon *et al.*, 2009; Glusman *et al.*, 2015). Malgré son avantage évident, par rapport aux 2 autres techniques, de pouvoir détecter des variations structurales dans des régions complexes du génome, celles-ci ne peuvent pas être cartographiées sur le génome et ne peuvent de ce fait être validées en laboratoire.

Ces 3 méthodes utilisées pour la détection des variations structurales dans les génomes se sont avérées très utiles pour l'étude de l'activité transpositionnelle et sont devenues incontournables dans le contexte actuel où le séquençage de génomes individuels est devenu routinier.

Fig. 1.21 Initiation des talles pendant l'étape de tallage

Étude microscopique de la formation des bourgeons qui deviendront des talles chez le riz. Les flèches indiquent l'initiation des bourgeons au stade plantule 4 feuilles.



Li *et al.*, 2003

1.3. Le modèle de l'étude : Le riz

Le riz (*Oryza* spp., Linné 1753) est une plante monocotylédone de la famille des Poaceae. Cette plante est une herbacée annuelle qui présente un cycle biologique court, de l'ordre de 3-4 mois (Delseny *et al.*, 2000). Le riz est autogame, cultivé sous différents climats et adapté à différents habitats, irrigué, inondé ou pluvial. Cette capacité d'adaptation à différents milieux en fait l'une des premières céréales cultivées dans le monde avec le blé pour l'alimentation humaine.

Il existe 2 espèces cultivées *O. sativa* (riz asiatique) et *O. glaberrima* (riz ouest africain) qui se différencient par différents traits morphologiques, tallage, taille des grains, etc. A l'instar d'*A. thaliana* chez les dicotylédones, le riz représente une ressource clé et un modèle pour étudier les plantes monocotylédones, et les graminées en particulier, dû à la taille de son génome. La taille du génome du riz est estimé à 450 Mb, l'un des plus petits chez les céréales (cf. 2,7 Gb pour le maïs et 18 pour le blé). Une séquence de haute qualité du génome du riz, *Oryza sativa* spp. *Nipponbare*, est disponible depuis 2004 (International Rice Genome Sequencing Project, IRGSP, 2005;).

Cette espèce est devenue au cours de ces vingt dernières années une cible prioritaire en matière de recherche agronomique au niveau international et de manière concomitante la seconde espèce modèle en biologie végétale après *A. thaliana*.

1.3.1. Phénologie

Le cycle de développement du riz du stade graine à graine dure 3 mois et peut être divisé en 3 phases.

(1) La phase végétative, qui commence à la germination et se divise en 3 stades. Le stade plantule qui correspond à la période comprise entre l'émergence de la première feuille et l'apparition de la cinquième feuille. C'est un stade critique au cours duquel la plante est très fragile (Efsue et Derera, 2012; Ham *et al.*, 2013). Ensuite vient le stade du tallage avec l'apparition de la première talle de manière concomitante à l'apparition de la 5^{ème} feuille (Fig. 1.21). Le tallage se produit à partir de nœuds empilés à la base de la tige, dite « court-nouée », car les nœuds sont séparés par de très courts entre-nœuds. À partir de chaque nœud se développe une talle primaire. Le dernier stade de la phase végétative est le stade de l'élongation des entre-nœuds qui se traduit par une augmentation de la taille de la plante.

(2) La phase reproductive, période pendant laquelle le métabolisme de la plante est le plus important, ses besoins en éléments minéraux sont au maximum. Elle commence par l'initiation paniculaire de la feuille « drapeau » qui émerge à la base du dernier nœud. À partir de ce stade, le

nombre de grains par panicule est déjà déterminé. La panicule se développe à l'intérieur de la talle, c'est le stade de montaison, qui se termine par l'épiaison où la panicule émerge à la base de la feuille drapeau. La panicule est composée de branches primaires, secondaires et parfois tertiaires qui forment un groupe d'épillets qui prend naissance sur le dernier nœud de la tige.

Cette phase est très sensible aux changements de température (Jagadish *et al.*, 2013), au déficit hydrique (Jin *et al.*, 2013) et à la salinité (Kumar et Khare, 2016) qui peuvent entraîner dans certains cas la stérilité des organes reproducteurs qui se traduit par des grains vides .

(3) La phase de maturation commence à partir de la floraison qui se produit du haut vers le bas de la panicule. La fleur s'ouvre et la pollinisation s'ensuit. Après la fécondation, les grains vont se remplir progressivement, et passer par différents stades laiteux, pâteux jusqu'à maturité où le grain atteint sa taille définitive et perd toute coloration verte.

1.3.2. Les ETs chez le riz, *Oryza sativa*

En dépit de sa relative petite taille (450 Mb) pour un génome de graminées (cf. le Maïs avec 2,7 Gb; le blé avec 18 000 Mb), le riz présente une proportion significative d'ETs (plus de 34 %).

Les premiers ETs détectés chez le riz ont été trouvés soit par hasard soit par la détection de régions conservées dans les RTNs. C'est ainsi qu'Hirochika et ses associés ont, en 1992, isolé et estimé à environ 1 000 copies le nombre de RTNs dans le génome du riz en se basant sur la séquence du PBS, très conservée parmi les RTNs des plantes (Wicker et Keller, 2007). Plus tard, dans les années 2000, le programme de séquençage du génome du riz IRGSP a été initié. Une banque de clones BAC a été construite et une première étape fut de séquencer les extrémités des insertions de ces clones. Cela constituait une première opportunité d'étudier la distribution des ETs chez le riz, qui a été saisie par Mao et ses collaborateurs (2000). Les auteurs ont identifié environ 10 000 ETs dont 4 000 RTNs de la famille des *Gypsy* et 1 500 de la famille des *Copia* ainsi qu'une autre famille très représentée, celle des MITEs avec 2 700 éléments.

Finalement, en 2005, le séquençage du génome du riz (IRGSP, 2005) a permis d'établir un catalogue plus complet des ETs qu'il contient. Chez le riz, bien que les MITEs soient les plus abondants (Jiang *et al.*, 2004), ce sont les LTR-RTs qui occupent la part la plus importante du génome (IRGSP, 2005). Avec environ 300 familles regroupant presque 3 700 LTR-RTs (El baidouri et Panaud, 2013), ces éléments occupent à eux-seuls presque 21 % du génome et les autres éléments de classe I, les LINEs, les SINEs ainsi que les éléments de classe 2 ne représentent qu'un peu plus de 10 % du génome.

Fig. 1.22 Distribution du nombre de copies de rétrotransposons à LTR (LTR-RT) par famille chez le riz (*O. sativa*) (El baidouri et Panaud, 2013).

Les nombres de copies des 3 familles les plus représentées sont donnés dans le diagramme circulaire et seules 40 familles de LTR-RT sont représentées.

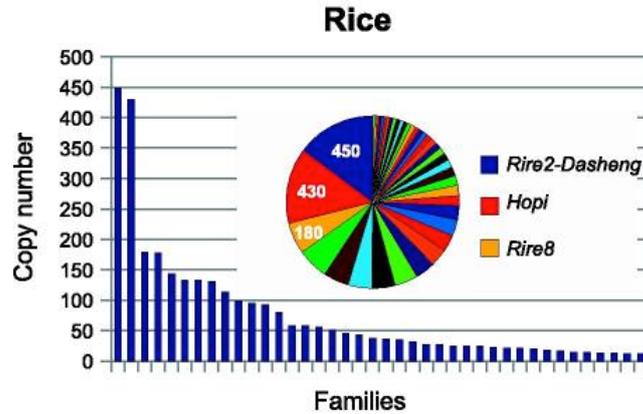
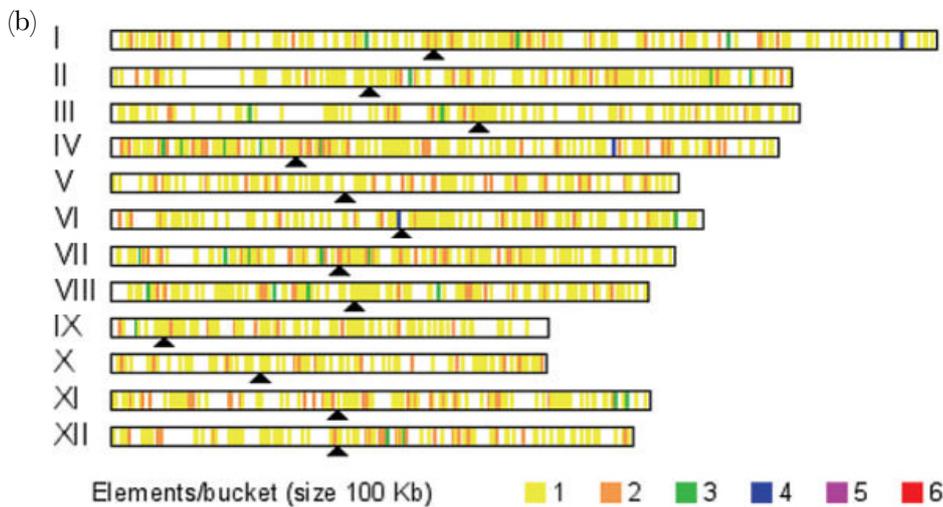
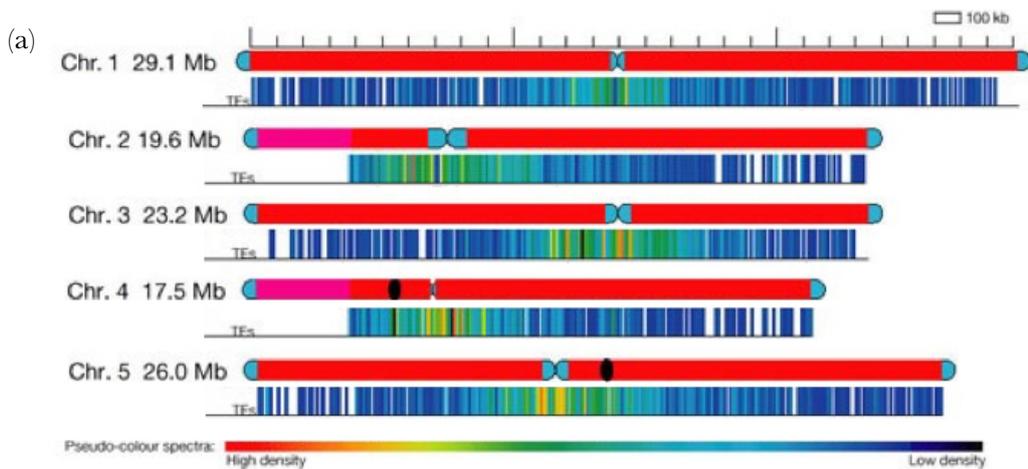


Fig. 1.23 Répartition des éléments transposables sur les chromosomes

(a) D'*Arabidopsis thaliana* (adapté de The Arabidopsis Genome Initiative, 2000).

(b) D'*Oryza sativa* (Chaparro *et al.*, 2007)



La majorité des LTR-RTs du riz ne se répartissent finalement que dans très peu de familles (Baucom *et al.*, 2009), et principalement dans 3 familles que sont *RIRE2-Dasheng*, *Hopi* et *Rire8* (El baidouri et Panaud, 2013) (Fig. 1.22).

Différentes études sur les populations d'ETs du riz ont cherché à estimer l'âge relatif et les périodes auxquelles ils ont envahi le génome afin de mieux comprendre l'impact des ETs sur l'évolution du génome. Pour cela, les LTR-RTs identifiés comme responsables de l'augmentation de la taille des génomes (Vitte et Panaud, 2005; Piegu *et al.*, 2006) représentent des éléments de choix pour la datation des événements de transposition. La datation des LTR-RTs repose sur l'étude du taux de substitutions entre les LTRs des LTR-RTs d'une même famille, qui, lorsqu'ils se copient pour transposer présentent un pourcentage d'identité entre les 2 LTRs de 100%. L'intérêt de ce système de datation a été montré pour évaluer les périodes de burst de transposition (Vitte et Panaud, 2005; El baidouri et Panaud, 2013). D'autre part, chez le riz, la comparaison des ETs à des positions orthologues dans différents génomes d'une même espèce cumulée à la datation des LTR-RTs par le taux de substitutions s'est avéré utile pour la datation d'évènements de domestication (Vitte *et al.*, 2004).

Contrairement aux ETs d'*A. thaliana* distribués majoritairement au niveau des régions péricentromériques (The Arabidopsis Genome Initiative, 2000), les ETs chez le riz semblent répartis de façon plus aléatoires dans le génome (Chaparro *et al.*, 2007) (Fig. 1.23). Cependant, certaines familles de LTR-RT présentent des distributions bien plus distinctes avec une préférence pour les régions hétérochromatiques, comme les régions centromériques (Gao *et al.*, 2015; IRGSP, 2005). Les raisons de cette répartition non aléatoire de certaines familles dans des régions composées presque exclusivement d'ADN non codant, hétérochromatiques et où le taux de recombinaison est faible restent encore inexplicées. Pour expliquer les variations de densité des ETs dans les génomes si l'on considère qu'un ET est un élément égoïste qui parasite les génomes et transpose pour survivre, on peut supposer qu'il y ait :

- Soit une contre sélection des ETs qui s'insèreraient dans ou à proximité des gènes essentiels pour le développement de l'organisme,
- Soit une sélection des ETs insérés dans des régions peu denses en gènes ou où le taux de recombinaison est faible, car les chances d'exercer un effet délétère sur l'hôte y sont plus faibles.

1.4. Objectifs de la thèse

Nous venons de voir dans l'introduction que l'étude des ETs revêt un grand intérêt pour comprendre leur impact sur les génomes. D'une manière générale, bien que la plupart des ETs soient inactifs, ils occupent une grande partie des génomes végétaux. Les conditions d'une telle amplification dans les génomes restent encore assez incompréhensibles malgré la mise en évidence ces dernières années de mécanismes efficaces pour contrôler leur activité et le relâchement de ces mécanismes sous l'impact de stress. Cependant, les pics de transposition qui ont eu lieu au cours de l'évolution se sont produits en conditions naturelles.

C'est dans ce contexte que s'inscrit le sujet de ma thèse dont les interrogations sont :

➤ Quelle est la réponse génomique du riz au relâchement de la voie de répression des ETs *via* la voie des siRNAs?

➤ Et quel est l'impact du stress thermique sur l'activité rétrotranspositionnelle?

Chez le riz, outre le fait que cette plante soit d'un grand intérêt agronomique, nous disposons d'une séquence de référence de bonne qualité, qui nous a permis de répondre à ces questions par une approche génomique.

Ce travail de thèse s'articule autour de 3 axes principaux constituant 3 parties de ce manuscrit :

(1) l'analyse de l'activation rétrotranspositionnelle dans un contexte épigénétique relâché dans le but de répondre à l'hypothèse selon laquelle, dans un contexte où le contrôle épigénétique sur les ETs est relâché, les contrôles d'inhibition des ETs seraient déverrouillés et permettraient la réactivation de certains ETs. Ces plantes ont été caractérisées par une étude génomique, transcriptomique et méthylomique.

La génération du matériel végétal ayant représenté une part non négligeable de ce travail, les données obtenues des différents séquençages n'ont pas encore été exploitées à leur maximum. En accord avec mes directeurs de thèse, je ne présenterai dans ce manuscrit que les résultats analysés. L'objectif étant de soumettre les résultats de cette étude avant la fin de l'année 2016 en vue d'une publication en 2017.

(2) l'analyse de l'activation rétrotranspositionnelle en condition de stress thermique chez le riz *Oryza sativa*. Pour cela, des plantes sauvages et en contexte épigénétique relâché ont été étudiées afin d'observer si le stress thermique est suffisant pour entraîner l'activation de certains ETs *in natura*.

(3) Un nouvel outil pour la détection de variations structurales comme l'insertion d'ETs causée par l'activation transpositionnelle dans les génomes. Les résultats de cette dernière partie font l'objet d'une publication soumise dans Mobile DNA.

Une dernière partie sera consacrée à une discussion générale sur les résultats de ces 3 parties ainsi qu'aux perspectives offertes à l'issue de ce travail.

Tab. 2.1 Tableau récapitulatif des données de séquençage obtenues pour les plantes *Osnrpd1* T0 et T1

	DNA-seq	RNA-seq	siRNA-seq	Méthyl-seq
NB (WT)	✓	✓ x3	✓	✓
23.1	✓		✓	✓
23.1-15	✓	✓ x3	✓	✓
23.1-17	✓		✓	
23.1-23	✓		✓	✓
23.1-31	✓			

Partie 2. Transposition *in planta* chez le riz *Oryza sativa* dans un contexte épigénétique relâché

L'activation des ETs a été principalement observée dans des plantes soumises à différents types de stress et/ou lors de cultures cellulaires. Les méristèmes ont également été identifiés comme des tissus somatiques permissifs à l'activation des ETs (Ohtsu *et al.*, 2007; Vicient *et al.*, 2010). La multiplication cellulaire localisée au sein des méristèmes suivie de la différenciation des cellules caractérisent la croissance des végétaux. Les mutations somatiques ayant lieu dans les méristèmes peuvent donner naissance à des lignées de cellules identiques portant la même mutation. Chez le riz, les cellules méristématiques restent plus ou moins indifférenciées depuis l'embryogenèse précoce jusqu'à la détermination florale. Au début du développement végétatif, elles donnent les tiges, desquelles émerge une panicule portant les grains de riz, et les feuilles. Dans le cas où une mutation somatique serait propagée par division mitotique à toute la tige, celle-ci serait retrouvée dans les cellules formant les organes floraux et l'individu pourra être porteur de cette mutation. La mutation sera alors transmise à la descendance de la plante.

L'un des objectifs de ce travail est de détecter des ETs qui se seraient insérés dans des cellules somatiques du méristème. Nous souhaitons également savoir si les événements de transposition qui se produiraient seront transmis transgénérationnellement. Pour cela, nous avons étudié 3 générations successives de plantes sauvages et mutantes. L'observation d'événements de transposition dans les plantes pourra nous permettre de comprendre où et quand ces événements se sont produits.

De plus, les ETs sont maintenus à l'état silencieux par différents mécanismes épigénétiques. En parallèle de l'étude des plantes sauvages, des plantes mutantes affectées dans une voie épigénétique ont également été étudiées. Cette étude permet de tester l'hypothèse selon laquelle, dans un contexte de contrôle épigénétique relâché, les contrôles d'inhibition de l'activité des ETs sont déverrouillés et permettent la réactivation de certains ETs. Pour cela, nous avons entrepris l'analyse d'un mutant RNAi *Osnrpd1*, affecté dans la voie RdDM, encore jamais décrit chez le riz.

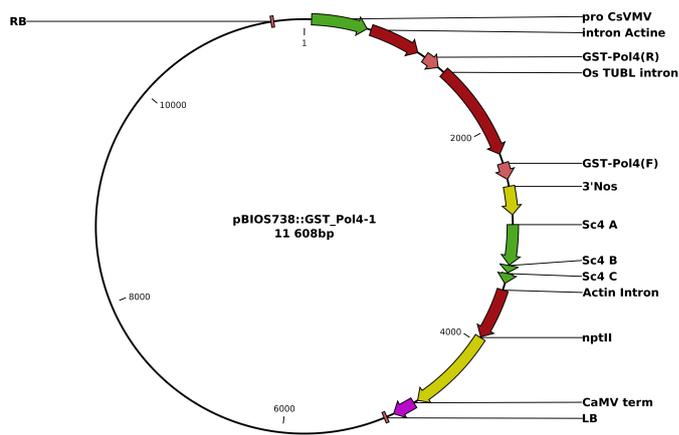
En sus de l'identification génomique de variations structurales causées par l'insertion d'ETs, une combinaison de différentes approches « omiques » a été utilisée afin d'obtenir une vue d'ensemble des mécanismes régissant le contrôle des ETs dans les plantes *Osnrpd1* de génération T0 et T1 (Tab. 2.1), à savoir :

> un séquençage des sRNAs, afin de vérifier l'efficacité de la stratégie RNAi à affecter la biogenèse des siRNAs de 24-nt dans le mutant *Osnrpd1*. Les informations de ce séquençage permettront également de vérifier l'implication des sRNAs dans le contrôle des ETs.

> un séquençage du méthylome. Les siRNAs jouent un rôle dans la méthylation de l'ADN des séquences qu'ils ciblent, et, dans la répression de l'activité des ETs. Ces données permettront d'identifier si le niveau de méthylation joue un rôle dans le relâchement de la répression des ETs dans le mutant *Osnrpd1*.

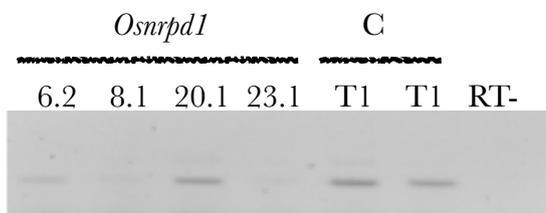
> un séquençage du transcriptome afin de voir : (1) si l'activité transcriptionnelle des ETs réactivés est liée à leur capacité transpositionnelle. (2) si leur insertion à proximité ou dans des gènes pourrait avoir un impact sur leur expression.

Fig. 2.1 Représentation schématique du vecteur utilisé pour le RNAi contre *OsNRPD1*



La séquence de 150 pb ciblant *OsNRPD1* a été clonée. Les deux séquences sont séparées par l'intron du gène tubuline (*TUBL*) et sous contrôle du promoteur du gène du virus de la mosaïque du manioc (*CsVMV*). Des étiquettes de 150pb GST-Pol4 (Gene Specific Tag- Polymerase4) sont insérées de part et d'autre de l'intron *TUBL*. Elles permettront la formation d'un long ARN en épingle à cheveux qui sera clivé en petits ARNs de 21-nt. Le gène *NptII* permet la sélection des plantes. Les séquences RB (Right Border) et LB (Left Border) délimitent l'ADN-T qui sera intégré au génome de la plante.

Fig. 2.2 Profil d'accumulation des transcrits de *OsNRPD1*



6.2, 8.1, 20.1, 23.1 désignent les différentes lignées *Osnrpd1*, T1 la plante contrôle WT. RT- le contrôle négatif de la PCR. Les amorces utilisées sont pol4-F6 et pol4-R6.

2.1. Matériel végétal

2.1.1. Génération du mutant *Osnrpd1* et sélection des plantes

Afin d'identifier les événements de transposition ayant eu lieu dans un mutant affecté dans la voie du RdDM, nous avons, en partenariat avec le CIRAD de Montpellier, généré des lignées RNAi contenant un construit dirigé contre *OsnRDP1*, homologue chez le riz de *NRPD1* identifié chez *A. thaliana* (Herr *et al.*, 2005). Avant mon arrivée au laboratoire, la transformation génétique du riz *Oryza sativa* avait été initiée au CIRAD par l'utilisation d'un ADN de transfert (ADN-T) introduit dans un plasmide porté par la bactérie *Agrobacterium tumefaciens*. L'agroinfection de tissus embryonnaires de riz a permis la génération de cals qui ont été cultivés *in vitro*. La structure du plasmide utilisé pour mener la stratégie RNAi est représentée dans la Figure 2.1.

Une séquence de 150 nt d'*OsnRDP1* a été insérée en répétition inversée. Cette construction génère un ARN en épingle à cheveu (hairpin), générant un ARN double brin qui sera clivé en petits ARNs de 21-nt, ce qui permettra de diriger la dégradation et/ou le blocage de la traduction des ARNm cibles, complémentaires à ces petits ARNs. Le but de cette stratégie RNAi est d'une part d'inactiver post-transcriptionnellement un gène en amont dans la voie RdDM afin de lever les contrôles d'inhibition exercés sur les ETs. La construction introduite cible le gène codant pour *OsnRDP1* qui est la grande sous-unité de la POLIV, elle-même à l'origine de la synthèse des petits ARNs interférents (siRNAs) chez les plantes. D'autre part, si la perte totale de l'expression d'*OsnRDP1* était létale, la méthode RNAi permet d'obtenir une inactivation partielle d'*OsnRDP1* (knock-down) suffisante pour diminuer la production de siRNAs. C'est pourquoi la génération d'un mutant knock-down a été préférée à celle d'un knock-out.

Caractérisation des lignées mutantes obtenues :

Quatre lignées mutantes, 6.2, 8.1, 20.1 et 23.1 ont été régénérées à partir de cals cultivés *in vitro* pendant 6-7 semaines. La lignée 23.1 a été choisie pour être la plante mère (T0) afin de générer le matériel végétal nécessaire à cette étude car cette plante présentait le taux d'expression d'*OsnRDP1* le plus faible mesuré par RT-PCR (Fig. 2.2).

Impact de la construction RNAi sur l'accumulation des siRNAs :

La construction RNAi utilisée pour générer le matériel végétal cible le gène *OsnRDP1*. Ce gène code pour la grande sous-unité de la polymérase 4 qui est impliquée dans la biogenèse des siRNAs de 24-nt à partir de régions hétérochromatiques. Les siRNAs dérivés des ETs jouent un rôle

Fig. 2.3 Distribution sRNAs sur *OsNRPD1*

Représentation des sRNAs de 21-nt (haut) et 24-nt (bas) sur la longueur du gène *OsNRPD1* pour la plante WT Nipponbare (en vert), la plante T0 23.1 (en rouge). Les valeurs sont représentées en ordonnées en RPM (Reads Per Million) sur la longueur d'*OsNRPD1* en abscisse.

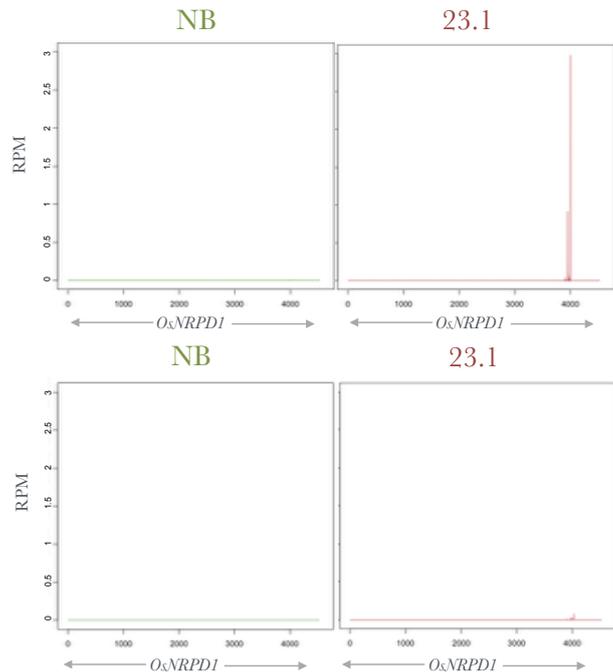
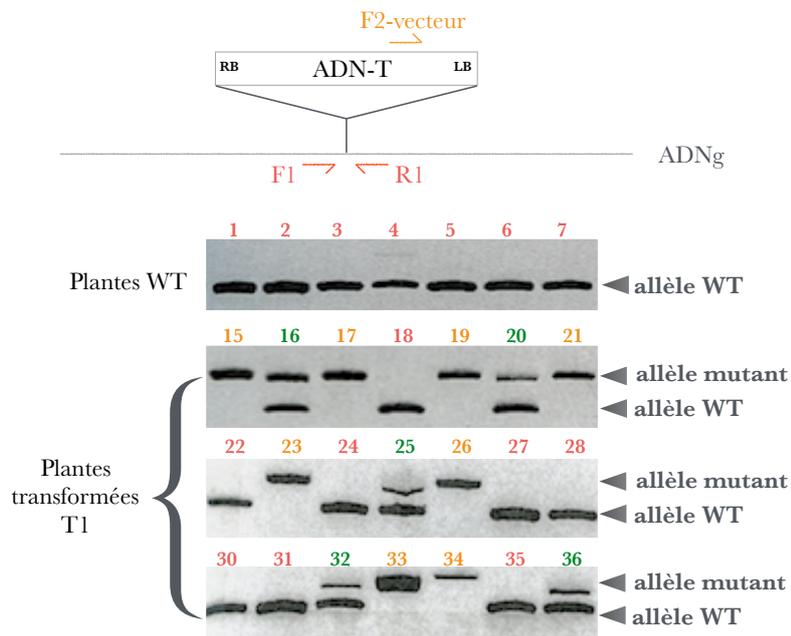


Fig. 2.4 Génotypage des plantes *Osnrpd1* de génération T1



(a) Les amorces utilisées sont 23.1-F1, 23.1-R1 et pol4vect-F1. F1 + R1 permettent l'amplification de l'allèle sauvage, i.e. pas d'insertion de l'ADN-T. F2 + R1 permettent de mettre en évidence la présence de l'ADN-T. Si il est présent à l'état homozygote, le produit de l'amplification par F1 + R1 est trop large pour être amplifié.

(b) Analyse par PCR effectuée sur l'ADN génomiques de feuilles récoltées sur des plantes issues d'une plante RNAi régénérée. Les fragments produits sont issus de l'allèle sauvage (+) ou de l'allèle mutant (m). Les plantes homozygotes sauvages (+/+) sont notées en rouge, les homozygotes mutantes (^m/_m) en orange et les hétérozygotes (+/_m) en vert.

Fig. 2.5 Profil d'accumulation de *OsNRPD1* dans les plantes T1 23.1



RT-PCR réalisées à partir d'extraits de feuilles de plante contrôlé (4) et plantes *Osnrpd1*. L'actine est utilisée comme contrôle de charge. Les plantes en rouge sont homozygotes sauvages (+/+), « 4 » étant une plante Nipponbare contrôlé sauvage. Les plantes en orange sont homozygotes mutantes (^m/_m).

dans leur répression en guidant la méthylation au niveau de leur séquence complémentaire. Nous avons donc réalisé un séquençage des sRNAs afin de caractériser les populations de petits ARNs dans les feuilles des plantes *Osnrpd1*, plantes pour lesquelles nous souhaitons induire un relâchement de la répression des ETs.

Un séquençage a été réalisé sur des échantillons de feuilles Nipponbare WT, et *Osnrpd1* T0 en partenariat avec le laboratoire de Blake Meyers (Université du Missouri).

Le nombre de lectures obtenues pour chacune des librairies de ces échantillons est répertorié dans l'Annexe 6. Le séquençage Illumina des librairies de sRNAs a permis d'obtenir de $2,67.10^7$ à $4,97.10^7$ séquences. Ces séquences de sRNAs débarrassées des séquences des adaptateurs sont alignées sur le génome de référence du riz Nipponbare IRGSP-1.0. Plusieurs lectures sont identiques et le nombre de ces séquences qui s'alignent sur le génome de référence est répertorié dans la colonne « Nombre de séquences différentes ». Le « nombre de lectures totales alignées » sur le génome de référence prend en compte le nombre de lectures de séquence identique, le taux d'alignement des lectures sur le génome varie de 48,1 % à 76,9 %.

La construction introduite dans les mutants *Osnrpd1*, induisant la production de sRNAs de 21-nt, ciblent *OsNRPD1*. L'abondance des sRNAs est évaluée au niveau de la séquence du gène *OsNRPD1* (Figure 2.3). Un pic de sRNAs de 21-nt de 1,6 à 3 RPM (Reads Per Million) est observé au niveau de la séquence cible du gène *OsNRPD1* dans les plantes transformées. Ces résultats suggèrent que la stratégie RNAi pour cibler le gène *OsNRPD1* fonctionne. Une accumulation des sRNAs 24-nt est également observée au niveau de la séquence ciblée (0,1 RPM). Cette accumulation est faible mais existante alors qu'il n'y en a pas dans la plante WT. Ces abondances de sRNAs au niveau la séquence cible seront prise en compte dans les analyses globales des sRNAs.

2.1.2. Génération de matériel végétal

Dans un premier temps, vingt-et-une graines de la plante T0 23.1 ont été semées. De chacune des plantes alors produites, une feuille a été récupérée afin d'en extraire l'ADN pour déterminer l'état de zygote de ces plantes pour l'ADN-T. Cette étape est nécessaire afin d'identifier les plantes homozygotes pour l'insertion de l'ADN-T, et ainsi, les plantes pour lesquelles la mutation knock-down est fixée.

Les résultats de ce génotypage sont présentés dans la Figure 2.4, ils montrent que 8 des plantes cultivées sont homozygotes pour l'insertion (T-DNA^{m/m}), 5 sont hétérozygotes (T-DNA^{+/m}) et 8 sont homozygotes WT (T-DNA^{+/+}) c'est-à-dire qu'elles ont perdu l'insertion de l'ADN-T. De

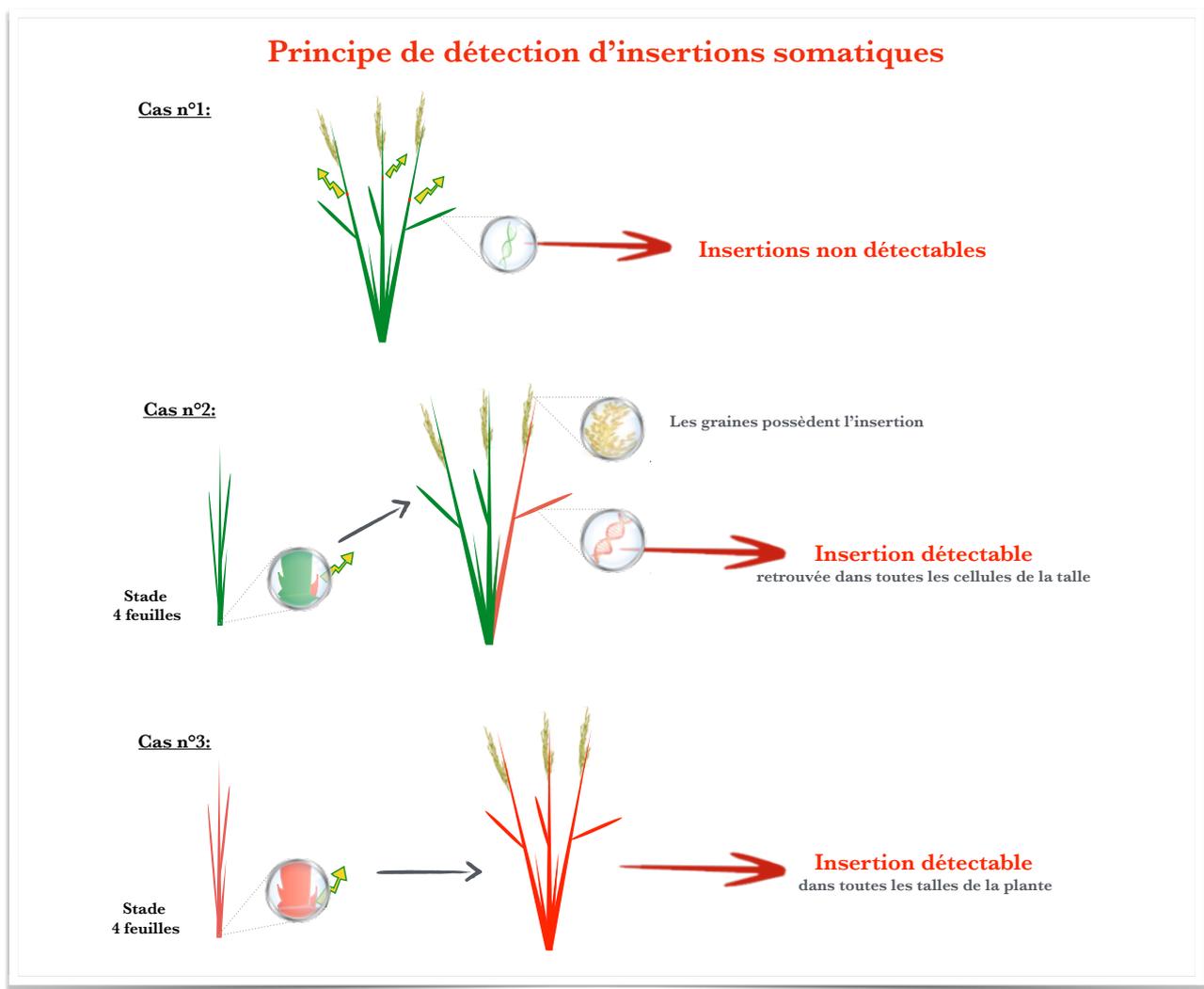


Fig. 2.6 Principe de détection d'insertions d'ETs

Les flèches en forme d'éclair jaune indiquent un évènement de transposition. En rouge sont représentés les organes formés de cellules portant l'évènement de transposition.

Dans le **cas n°1**, un ET transpose dans une cellule d'une talle développée. L'ADN contenu dans la feuille séquencée ne contiendra pas la nouvelle insertion. Cependant, les cellules reproductrices étant produites tardivement à partir des cellules somatiques, l'insertion pourra être retrouvée dans la génération suivante. **N.B.:** si l'ET transpose plus tôt, pendant le développement de la feuille, l'insertion ne sera pas détectable car le pool de cellules avec l'insertion trop faible.

Dans le **cas n°2**, un ET transpose au niveau du méristème, dans une cellule d'où s'initie la formation d'un bourgeon axillaire à partir duquel sera formée la talle. Toutes les cellules de la talle qui se répliquent par division porteront l'insertion. Cette dernière sera détectable dans l'ADN de feuille séquencé.

Dans le **cas n°3**, l'insertion d'un nouvel ET a eu lieu dans le méristème, toutes les talles qui se développent à partir de ce méristème possèdent l'insertion.

plus, le niveau d'expression d'*O_sNRPD1* a été évalué par RT-PCR semi-quantitative pour ces plantes (Fig. 2.5).

Trois plantes T-DNA ^{m/m}, 23.1-15, 23.1-17, 23.1-23 présentant le plus faible taux d'expression d'*O_sNRPD1* ainsi qu'une plante T-DNA ^{+/+}, 23.1-31 ont été choisies pour constituer les plantes de génération T1 pour cette étude. Elles ont été cultivées avec des plantes sauvages d'*Oryza sativa* spp. japonica cv. Nipponbare sur 3 générations successives en conditions normales (Annexe 1) afin de construire le pédigrée de l'étude. De par sa phénologie, le riz est un modèle intéressant pour identifier, grâce à différentes générations, le lieu et le moment des évènements de transposition qui peuvent se produire *in planta* (Fig. 2.6).

Pour chaque plante, des graines issues de 3 talles différentes ont été récoltées et une auto-fécondation a été faite afin de générer 3 plantes de seconde génération (T2). Ce schéma de récolte/semis de graines issues de 3 talles différentes d'une même plante auto-fécondée a été suivi à chaque temps de génération. Les plantes sont toutes nommées en fonction de la talle dont elles descendent. Ainsi, par exemple pour la plante Ta1.1, il s'agit d'une plante de génération T2 provenant de graines récoltées sur la talle 1 de la plante de première génération Ta1.

La figure 2.7 représente le matériel végétal des plantes Nipponbare sauvages. Les différentes plantes dont le génome a été séquencé figurent en bleu sur la figure.

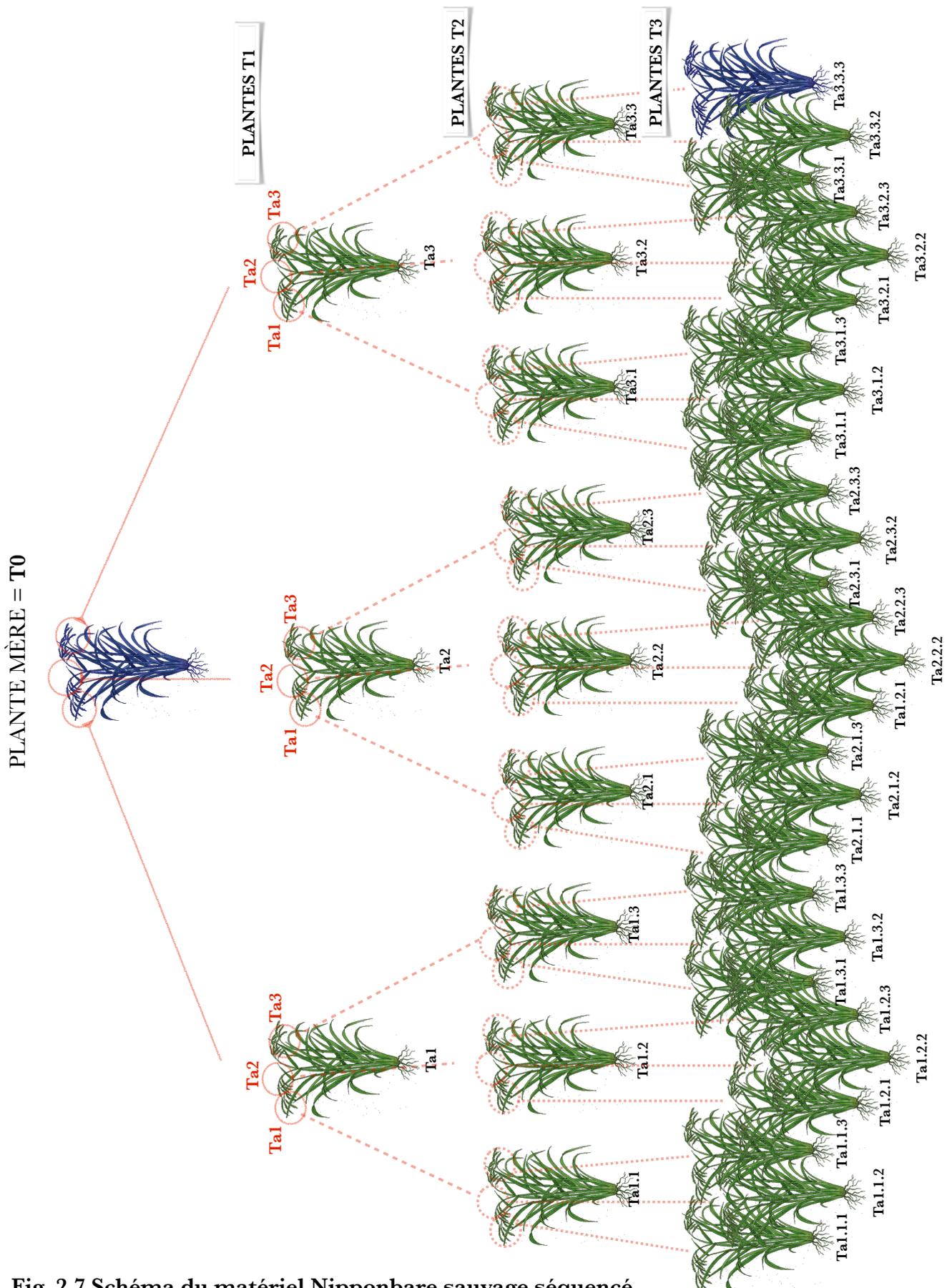


Fig. 2.7 Schéma du matériel Nipponbare sauvage séquencé.

Les différentes générations successives sont désignées par les temps T0, T1, T2 et T3. La plante mère est la plante à partir de laquelle a été produit tout le matériel végétal. Les talles dont les graines ont été utilisées pour produire les plantes de la génération suivante par auto-fécondation sont désignées par Ta1, Ta2 ou Ta3. Les plantes en bleu désignent les plantes dont le génome a été séquencé et celles sur fond rouge ont été séquencées après avoir subi le stress thermique.

2.2. Méthodes « omiques »

2.2.1. Génome

La technologie Hiseq2500 (San Diego, CA, USA) a été choisie pour effectuer un séquençage génomique en paired-end. Le « paired-end » permet de séquencer les deux extrémités d'un fragment d'ADN court (200-400 bases) afin d'obtenir une paire de deux séquences. Les données de ce séquençage ont été utilisées pour détecter les néo-insertions d'ETs dans les plantes mutantes (cf. partie 2.3).

2.2.2. sRNAseq

Le séquençage des sARNs a été réalisé pour la plante WT, la plante *Osnrpd1* : T0 23.1 et trois plantes T1 23.1-15, 23.1-17 et 23.1-23 par le laboratoire de Blake Meyers (Newark, DE, USA). Les alignements des sRNAs sur les séquences des ETs ont été réalisés sur la plateforme du laboratoire de Blake Meyers (<https://mpss.danforthcenter.org>). L'analyse globale des sRNAs a été faite par alignement des séquences sRNAs sur le génome du riz IRGSP-1.0 à l'aide du logiciel Bowtie2 (Langmead et Salzberg, 2012) version 2.2.4. Puis, le nombre de lectures alignées sur le génome est calculé par fenêtre de 100kb. Enfin, le nombre de lecture est normalisé par RPM (reads per million) pour chacune des librairies.

2.2.3. Transcriptome

Les transcriptomes d'une plante Nipponbare et d'une plante mutante *Osnrpd1* (23.1-15) cultivées en conditions normales ont été réalisés en « paired-end » avec le séquenceur Hiseq d'Illumina (San Diego, CA, USA).

Dans un premier temps, la qualité des données reçues du séquençage sous forme de fichier FASTQ est vérifiée grâce au programme FastQC. Puis les lectures correspondant à des transcrits d'organelles sont filtrées en alignant les lectures sur les génomes mitochondriaux et chloroplastiques avec le logiciel Bowtie2 (<http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>). Les lectures non alignées sont récupérées et constituent les données d'entrée pour la suite de l'analyse. Ces lectures sont ensuite alignées contre le génome de référence IRGSP-1.0 avec le logiciel TopHat (<http://tophat.cbcb.umb.edu/manual.shtml>).

Pour des meilleurs alignements et détections de transcrits, un fichier d'annotation des éléments transposables du génome au format .gtf est fourni au programme TopHat qui l'aidera à repérer les jonctions exons-exons et ainsi reconstruire les isoformes possibles à partir de ces données.

Afin de faire le comptage du nombre de lectures par locus du génome, la suite de programmes Cufflinks est utilisée (<http://cole-trapnell-lab.github.io/cufflinks>) (Trapnell *et al.*, 2012). Pour cette étape, on part de l'hypothèse que le nombre de lectures venant d'un certain gène est proportionnel à l'abondance de son ARN dans la cellule. Pour pouvoir comparer les transcriptomes des différentes plantes séquencées, les données sont normalisées en utilisant le rpkm (read per kilobase per million), ce qui permet de prendre en compte la taille du gène ou du locus ainsi que la profondeur du séquençage.

Enfin, pour l'analyse différentielle, l'outil Cuffdiff est utilisé pour générer des tableaux de comptage comprenant l'annotation du gène et le nombre de lecture pour chacun des échantillons. Un locus est donné comme différentiellement exprimé avec une FDR (False of Discovery Rate) à 5%, c'est-à-dire que la différence entre la valeur attendue et la valeur observée (en valeur absolue) dépasse ce seuil.

2.2.4. Méthylome

Les analyses du séquençage au bisulfite ont été réalisées en collaboration avec le laboratoire de Blake Meyers (Newark, DE, USA) et P-Y. Chen (Academia sinica, Taiwan) à partir d'une plante WT, la plante *Osnrpd1* T0 23.1 et deux plantes *Osnrpd1* T1 23.1-15Ta2 et 23.1-23Ta3.

2.3. Méthodes utilisées pour la détection de néo-insertions d'ETs

2.3.1. Détection par approche *a priori*

Afin d'établir si des événements de transposition ont eu lieu dans les plantes analysées, nous avons entrepris une approche *a priori* (AAP) des ETs *Tos17* et *Lullaby* connus pour être actif chez le riz après la culture de cals *in vitro* (Hirochika *et al.*, 1996; Picault *et al.*, 2009). Pour cela, des lectures séquencées en paired-end ont été alignées sur les séquences de chacun des 2 ETs. Les paires discordantes dont l'une des lectures s'aligne sur l'un de ces ETs sont récupérées. Il suffit ensuite de voir où s'aligne la deuxième lecture de la paire par BLAST sur la séquence du génome de référence, avec comme condition que cet alignement soit au niveau d'une région unique (hit unique). Les néo-insertions détectées ont été validées par PCR par la technique RBIP (Retrotransposon-Based Insertion Polymorphism)

2.3.2. Mise au point d'une méthode de détection des néo-insertions

En génomique, l'analyse bioinformatique des données est toute aussi importante que leur production. Si les données de séquençage NGS contiennent un taux d'erreur inférieur à 1%, les différentes méthodes et pipelines existants pour la détection des insertions d'ETs n'ont quant à eux pas encore réussi à éliminer le problème des faux positifs et négatifs.

L'objectif de notre équipe lorsque j'ai débuté ma thèse était de créer un pipeline de détection « sur mesure » adapté aux espèces disposant d'une séquence de référence de bonne qualité, comme c'est le cas du riz. Nous avons entrepris l'identification des néo-insertions par l'analyse des données de séquençage obtenues en PEM. Pour cela, deux analyses ont été menées en parallèle afin d'identifier les nouveaux événements de transposition qui se sont produits dans nos plantes (cf. Annexe 4). La première de ces analyses repose sur l'étude des séquences obtenues avec le PEM. Cette méthode de séquençage permet l'identification de paires discordantes en se basant sur l'alignement des lectures (reads) issues du séquençage sur le génome de référence alors que l'autre analyse, complémentaire à la première, se base sur la détection des lectures coupées (split-reads), observées en cas d'alignements incomplets des lectures sur le génome de référence.

Après avoir filtré les séquences correspondant au génome mitochondrial et chloroplastique, un alignement sur le génome de référence est réalisé avec Bowtie2 (Langmead and Salzberg, 2012) version 2.2.4 pour chacune des lectures obtenues pour chaque génome. Les paires discordantes, correspondant à des alignements trop éloignés (de plus de 1 kb) pour appartenir à la même paire,

Tab. 2.2 Tableau des positions des séquences mal assemblées dans le génome de référence retrouvées dans les 9 premiers génomes séquencés.

Les trois premières colonnes du tableau désignent les fenêtres de 10 kb sur un chromosome où se situe une séquence mal assemblée. Les neuf colonnes suivantes représentent les 9 génomes séquences pour former une matrice dans laquelle 1 fait référence à la détection de cette séquence et 0 sa non détection.

Positions		GM	GT3	HST 233	HST 332	23.1	23.1 15	23.1 17	23.1 23	23.1 31
chr2	8 900 000	8 910 000	1	1	1	1	1	1	1	1
chr2	19 400 000	19 410 000	1	1	1	1	1	1	1	1
chr2	19 700 000	19 710 000	1	1	1	1	1	1	1	1
chr3	22 760 000	22 770 000	1	1	1	1	0	1	1	1
chr3	28 700 000	28 710 000	1	1	1	1	1	1	1	1
chr4	910 000	920 000	0	1	1	0	0	1	1	1
chr4	940 000	950 000	1	1	0	1	1	1	0	1
chr4	1 440 000	1 450 000	1	0	1	0	1	1	1	0
chr4	2 300 000	2 310 000	1	1	1	1	0	0	1	1
chr8	3 220 000	3 230 000	1	0	1	1	0	1	0	0
chr8	3 230 000	3 240 000	1	1	1	0	1	0	1	0
chr8	3 240 000	3 240 000	1	0	1	0	0	1	0	0
chr8	28 310 000	28 320 000	1	1	1	1	0	1	1	0
chr10	7 330 000	7 340 000	0	1	1	0	0	1	1	0
chr10	9 270 000	9 280 000	0	1	0	1	0	0	1	0
chr11	23 440 000	23 450 000	0	1	0	1	0	0	1	0

comme les lectures d'une même paire alignées sur des chromosomes différents, ont été récupérées. Deux alignements successifs des lectures ont été effectués pour définir les paires discordantes :

- le premier, en utilisant le paramètre `-k 1` de bowtie (1 alignement), pour optimiser le temps de calcul en réalisant un premier tri des paires discordantes.
- le second, en utilisant le paramètre `-k 20`, afin d'éliminer des faux positifs qui auraient été générés en ne récupérant pas le meilleur alignement lors du premier alignement.

Les paires discordantes sont identifiées par leur flag, un nombre binaire indicatif de la qualité de l'alignement des lectures. Les lectures de ces paires ont ensuite été alignées sur le génome de référence IRGSP-1.0, version 4 (<http://rgp.dna.affrc.go.jp/E/IRGSP/Build4/build4.html>). Les seules paires de lectures récupérées, représentant un événement de transposition (breakpoint), sont celles présentant une des lectures de la paire alignée avec un hit unique sur le génome et l'autre lecture de la paire alignée à au moins 2 endroits différents du génome. Les insertions nichées dans les séquences répétées ne sont pas détectées.

Un alignement des lectures alignées à plusieurs endroits du génome a été réalisé contre la base de données d'ETs, la RITEDb (RIce TE database, <http://www.genome.arizona.edu>), afin d'identifier l'élément qui se serait inséré. Lorsque la lecture s'aligne sur un ET, une région de 2 kb autour de la région du point d'insertion a été extraite. L'alignement de cette région d'insertion avec l'élément potentiellement inséré permet de vérifier qu'il n'y a pas de cousin de l'ET dans cette région, et d'éliminer ce type de faux positif.

En parallèle de cette première partie d'analyse, une deuxième méthode utilisant l'outil d'alignement Segemehl version 0.2.0 (<http://www.bioinf.uni-leipzig.de>) a été utilisée afin d'identifier des splits-reads (ou lectures chimériques), c'est-à-dire que 2 parties de la séquence s'alignent à 2 endroits distincts dans le génome. En sortie d'analyse, les lectures alignées de façon incomplète sur le génome de référence ont été récupérées et comptées par fenêtre de 200 paires de base (pb) sur le génome.

Les régions pour lesquelles un breakpoint a été identifié ont été comparées aux régions contenant des split-reads. Un événement de transposition est établi si dans une même région se trouvent au minimum deux paires discordantes et au minimum 1 split-read (choix des filtres expliqué ci-après).

Identification de séquences non/mal assemblées dans le génome de référence :

Cette analyse a permis d'identifier des différences génomiques existant entre la plante de référence IRGSP-1.0 (MSU Rice Genome Annotation Project) et la plante mère qui est la plante contrôle de notre analyse. Les données obtenues sont énumérées dans le Tableau 2.2. Ce tableau a été construit en fonction du nombre de paires discordantes et de split-reads détectés au sein d'une

Tab. 2.3 Nombre de néo-insertions total et de *Tos17* détectées dans les plantes *Osmrpd1* en fonction du nombre de paires discordantes

	23.1		23.1-15		23.1-17		23.1-23		23.1-31	
filtre	total	<i>Tos17</i>	total	<i>Tos17</i>	total	<i>Tos17</i>	total	<i>Tos17</i>	total	<i>Tos17</i>
>0	159	12	237	7	300	2	883	3	850	7
>1	82	12	132	7	136	2	288	3	301	6
>2	57	9	80	7	75	2	114	3	108	6
>3	52	9	69	7	62	2	92	3	72	6
>4	48	8	60	6	52	2	81	3	57	4
>5	36	7	51	6	47	2	67	3	53	4
>7	27	5	42	5	36	1	49	3	45	4
>10	15	1	33	3	29	1	32	3	28	3
>15	7	0	17	1	13	1	18	3	10	2
>20	3	0	11	0	6	1	10	1	5	1

fenêtre de 10 kb sur le génome, où 1 désigne la présence d'au moins 2 paires discordantes et 1 split-read pour soutenir une néo-insertion dans une fenêtre et 0 moins de 1 paire discordante. Les différences génomiques ont été identifiées comme des variations structurales à 16 positions différentes dans les plantes séquencées. Ces régions ont, de ce fait, été éliminées pour les analyses ultérieures.

Ensuite, plusieurs points de néo-insertion hypothétiques (jusqu'à 883) ont été identifiés dans les plantes *Osnrpd1*. Afin d'affiner la méthode de détection, et diminuer le nombre de faux positifs que nous détectons, différents filtres sur le nombre de paires discordantes ont été testés pour évaluer le seuil de détection des néo-insertions de *Tos17* (Tableau 2.3). Si une paire présente une lecture qui s'aligne en une position définie du génome et l'autre à différentes positions dans le génome alors cette paire discordante est récupérée car potentiellement signe de néo-insertion d'ET. Les insertions d'ETs sont donc détectées seulement s'il s'agit d'une insertion dans une région unique. Ensuite, un filtre sur le nombre de paires discordantes a été réalisé. Toutes les néo-insertions de *Tos17* sont retrouvées dans les plantes si une paire discordante est considérée comme suffisante pour détecter un point de néo-insertion (filtre >0). En revanche, dès que l'on augmente ce filtre à au minimum 2 paires discordantes (filtre > 1), une néo-insertion de *Tos17* n'est plus détectée dans la plante 23.1-31. Le nombre total de néo-insertions détectées avec un filtre strictement supérieur à 1 a lui été divisé de 2 à 3 par rapport au filtre strictement supérieur à 0. Sachant que la profondeur de séquençage minimum pour les plantes *Osnrpd1* est de 14X (Annexe 7), un filtre d'au minimum 2 paires discordantes pour détecter un point de néo-insertion ne devrait pas être trop stringent et a été choisi. Une fois ce filtre réalisé, l'étape suivante dans le pipeline a été de récupérer la lecture de la paire discordante qui ne s'aligne pas à une position unique du génome et de l'aligner sur la base de données d'ETs dont nous disposons pour le riz. Parallèlement à cette méthode d'alignement des lectures paires, un outil de détection Segemehl a été utilisé pour identifier les split-reads, dites chimériques. Un filtre a été réalisé afin de récupérer au moins 2 lectures chimériques par position.

2.3.3. Détection par analyse des clipped-reads

Une autre méthode de détection des néo-insertions utilisant les informations contenues dans les clipped-reads (ou lecture rognée), a été utilisée pour détecter l'activité rétrotranspositionnelle. Ce type de lecture coupée diffère du « split-read », où chacune des deux parties de la lecture doit être alignée sur le génome à des localisations différentes, alors que pour les « clipped-read » seule la partie de la lecture alignée sur le génome est récupérée. Pour cette analyse nous n'avons pas utilisé l'outil Segemehl qui permet de détecter différents types de lectures chimériques, celles présentant

des délétions, insertions mais également des inversions de séquences (etc.). Dans le cas présent, nous souhaitons strictement récupérer dans les plantes analysées, les lectures qui ne s'alignent pas complètement sur le génome, indiquant des rupture d'alignements caractéristiques de néo-insertions d'ETs. Pour cela, un filtre est réalisé pour identifier toutes les régions divergentes entre le génome de référence et le génome Nipponbare contrôle de notre analyse (NB) afin d'éliminer des variations entre génome individuel. Un second filtre après alignement des séquences des mutants sur le génome de référence est réalisé sur le CIGAR (Compact Idiosyncratic Gapped Alignment Report, représentatif de l'alignement) donné dans le fichier SAM (fichier texte) généré après alignement des lectures obtenues par le séquençage sur le génome de référence avec le logiciel Bowtie2. Les paires discordantes dont une lecture s'aligne d'un côté ou de l'autre des points de rupture d'alignement identifiés par les clipped-reads sont utilisés pour déterminer l'ET inséré.

2.3.4. Détection par calcul de la profondeur de couverture

Les 2 premières méthodes de détection utilisées précédemment présentent, en plus d'une production importante de faux positifs pour la première, une contrainte pour la détection d'un point d'insertion. Pour la première, la lecture de la paire discordante qui ne s'aligne pas sur l'ET actif doit s'aligner dans une région unique du génome afin de déterminer le point d'insertion. Pour la seconde, la partie de la lecture alignée doit comme dans la première méthode, être alignée dans une région unique. Dans le but de faire un inventaire exhaustif de toutes les néo-insertions présentes dans les plantes étudiées, la méthode du *Depth of coverage* (DOC) a été utilisée. Cette méthode repose sur le dénombrement des lectures de séquençage s'alignant sur une région précise du génome, correspondant à un rétrotransposon. Elle identifie donc les ETs transposés sur la base de leur augmentation du nombre de copies, sans tenir compte du lieu d'insertion. Ainsi, elle permet d'identifier aussi les ETs insérés dans des régions répétées. L'inconvénient est qu'elle ne permet pas de déterminer le site d'insertion.

Dans un premier temps, l'analyse de la profondeur de lecture a été réalisée sur un jeu de données aléatoire de 92 gènes uniques, afin d'établir la profondeur de lecture pour chacun des génomes. Ainsi, il est possible de normaliser les nombres de lectures alignées à chaque position par la moyenne du nombre de lectures pour une région unique afin de déduire le nombre de copie de l'élément étudié avec un intervalle de confiance à 95% (2σ). Les valeurs obtenues pour chaque génome de la moyenne et des écarts-types sont répertoriées dans l'Annexe 8. Le calcul de la profondeur de lecture a ensuite été testé avec les éléments *Tos17*, *Lullaby* et *Karma* précédemment identifiés afin de confirmer la validité de la méthode. La validation de la méthode s'est faite par

Tab. 2.4 Tableaux comparatifs des méthodes utilisées pour la détection des néo-insertions de *Tos17*, *Lullaby* et *Karma*

Le nombre de néo-insertions détecté dans différentes plantes par l'Analyse À Priori (AAP) est comparé à celui déterminé par la méthode du Depth Of Coverage (DOC). La plante GM désigne la plante Nipponbare contrôle. Les plantes 23.1 sont les plantes mutantes *Osnrpd1*.

(a)

<i>Tos17</i>	AAP vs. DOC	
GM	2	2,8 +/- 0,50
23.1	12	7,4 +/- 2,11
23.1-15	7	6,61 +/- 1,7
23.1-17	3	4,17 +/- 0,9
23.1-23	5	6,33 +/- 1,22
23.1-31	7	6,8 +/- 1,29

(b)

<i>Karma</i>	AAP vs. DOC	
GM	2	2,8 +/- 0,50
23.1	12	7,4 +/- 2,11
23.1-15Taa2.2.2	7	6,61 +/- 1,7
23.1-23Ta3.3.3	3	4,17 +/- 0,9
23.1-31Ta1.2.1	5	6,33 +/- 1,22

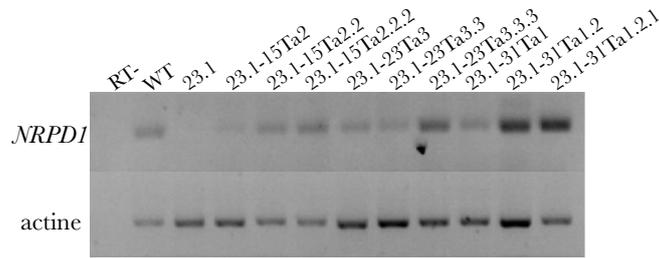
(c)

<i>Lullaby</i>	AAP vs. DOC	
GM	3	2,4 +/- 0,40
23.1	4	2,6 +/- 0,70
23.1-15	5	3,2 +/- 0,81
23.1-17	5	2,8 +/- 0,9
23.1-23	5	3,7 +/- 0,66
23.1-31	5	2,9 +/- 0,56

comparaison avec les résultats du nombre de copies des éléments déterminés avec l'AAP (Tableaux 2.4 a, b et c).

Fig. 2.8 Profil d'accumulation d'*OsNRPD1* dans 3 générations de plantes *Osnrpd1*

Evaluation de l'expression d'*OsNRPD1* par RT-PCR semi-quantitative pour une plante WT, la plante *Osnrpd1* T0 23.1, des plantes *Osnrpd1* T1 (23.1-15Ta2, 23.1-23Ta3 et 23.1-31Ta1), des plantes *Osnrpd1* T2 (23.1-15Ta2.2, 23.1-23Ta3.3 et 23.1-31Ta1.2) et des plantes *Osnrpd1* T3 (23.1-15Ta2.2.2, 23.1-23Ta3.3.3 et 23.1-31Ta1.2.1). RT- est le contrôle négatif de la PCR.



Tab. 2.5 Néo-insertions de *Tos17* et *Lullaby* dans les plantes *Osnrpd1* de génération T0 et T1

Nom de l'élément	Chromosome d'arrivée	Position d'arrivée	Nom de l'insertion	23.1	23.1-15	23.1-17	23.1-23	23.1-31
<i>Lullaby</i>	chr04	30 363 377	Neo1	✓	✓	✓	✓	✓
<i>Tos17</i>	chr01	2 729 566	Neo1	✓	✓			✓
<i>Tos17</i>	chr01	5 815 622	Neo13				✓	✓
<i>Tos17</i>	chr01	7 120 102	Neo8	✓				✓
<i>Tos17</i>	chr01	40 103 794	Neo9	✓				✓
<i>Tos17</i>	chr02	32 884 895	Neo2	✓	✓			
<i>Tos17</i>	chr03	5 572 693	Neo3	✓	✓	✓		✓
<i>Tos17</i>	chr03	79 894 556	Neo4	✓	✓	✓	✓	
<i>Tos17</i>	chr03	34 643 056	Neo5	✓	✓		✓	✓
<i>Tos17</i>	chr07	856 873	Neo10	✓			✓	✓
<i>Tos17</i>	chr09	19 674 689	Neo11	✓			✓	
<i>Tos17</i>	chr09	21 062 775	Neo6	✓	✓			
<i>Tos17</i>	chr11	474 936	Neo12	✓				
<i>Tos17</i>	chr11	25 235 985	Neo7	✓	✓			

Tab. 2.6 Néo-insertions détectées dans les plantes *Osnrpd1* par le pipeline d'analyse

	23.1		23.1-15		23.1-17		23.1-23		23.1-31	
	total	<i>Tos17</i> & <i>Lullaby</i>	total	<i>Tos17</i> & <i>Lullaby</i>	total	<i>Tos17</i> & <i>Lullaby</i>	total	<i>Tos17</i> & <i>Lullaby</i>	total	<i>Tos17</i> & <i>Lullaby</i>
Points de néo-insertions	12	12	11	8	6	3	24	4	17	6

2.4. Résultats

2.4.1. Profil d'accumulation d'*O_sNRPD1* dans les différentes générations *Osnrpd1*

Le but de la transformation génétique RNAi est de reproduire le mécanisme de l'interférence de l'ARN. Cette stratégie a été utilisée contre le gène *O_sNRPD1*, lui-même impliqué dans la voie de l'interférence ARN *in natura*. Afin de vérifier l'efficacité de la construction knock-down d'*O_sNRPD1*, l'abondance des sRNAs au niveau du gène *O_sNRPD1* a été évaluée pour 3 plantes *Osnrpd1* (cf. Annexe 17) et le niveau d'expression de ce gène a été quantifié par RT-PCR. Des plantes de différentes générations ont été testées pour vérifier que l'effet de la construction est maintenu au cours des différentes générations (Fig. 2.8). L'échantillon NB désigne les transcrits de Nipponbare sauvage (WT); 15, 17 et 23 sont les transcrits issus de plantes mutantes *Osnrpd1* homozygotes (T-DNA^{m/m}) et 31 d'une plante homozygote WT (T-DNA^{+/+}).

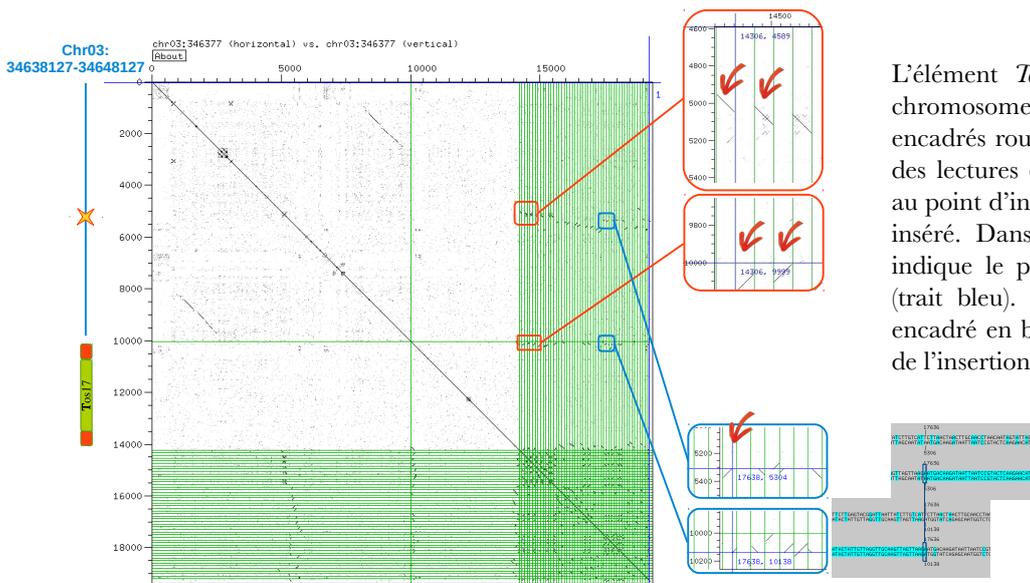
Les résultats de cette expérience montrent que le profil d'expression des transformants en T0 et T1 est plus faible que celui d'une plante Nipponbare WT, ce qui confirme l'efficacité du knock-down dans ces plantes. En revanche, les niveaux d'expression d'*O_sNRPD1* dans les mutants de génération T3 semblent revenir à un niveau similaire par rapport à celui de la plante contrôle voire plus important. Cette dernière constatation suggère que la construction RNAi n'est plus efficace dans les plantes de génération T3.

2.4.2. Néo-insertions d'ETs dans les plantes *Osnrpd1*

Détection par approche AAP :

L'AAP utilisée pour les ETs *Tos17* et *Lullaby* a permis de détecter de nouvelles insertions de ces ETs dans les plantes *Osnrpd1* (Tableau 2.5). Toutes les néo-insertions, sauf une (Néo13) sur le chromosome 1, ont été détectées dans la plante T0 23.1. Cela suggère que toutes ces néo-insertions ont eu lieu dans la cellule du cal à partir de laquelle a été régénéré la plante 23.1. Les différentes plantes T1, ne présentent pas toutes les mêmes insertions, il y a eu une ségrégation. En revanche, la néo-insertion qui n'est pas retrouvée chez la plante 23.1 est retrouvée à la même position dans 2 plantes T1 : 23.1-23 et 23.1-31. Ce résultat suggère que cet événement transpositionnel a eu lieu *in planta* dans la plante régénérée 23.1. Mais cette insertion n'a pas été identifiée dans le pool de cellules séquencées (cf. cas n°1 dans la Fig. 2.6).

Fig. 2.9 Dot-plot des séquences flanquant la néo-insertion, de l'élément inséré (*Tos17*) et des lectures discordantes et des lectures coupées au point d'insertion sur elles-même



L'élément *Tos17* est inséré au niveau du chromosome 3 (étoile jaune). Dans les encadrés rouges se trouvent les alignements des lectures discordantes d'une paire, l'une au point d'insertion et l'autre dans l'élément inséré. Dans les encadrés bleus, la flèche indique le point de coupure de la lecture (trait bleu). Dans les rectangles grisés est encadré en bleu au nucléotide près le point de l'insertion.

Tab. 2.7 Néo-insertions de *Tos17* et *Lullaby* détectées dans les plantes *Osnrpd1* de génération T0 et T1 par le pipeline d'analyse

			23.1	23.1-15	23.1-17	23.1-23	23.1-31	Élément
chr01	2 720 000	2 730 000	1	1	0	0	1	<i>Tos17</i>
chr01	5 810 000	5 820 000	0	0	0	1	1	<i>Tos17</i>
chr01	7 110 000	7 120 000	1	0	0	0	1	<i>Tos17</i>
chr01	7 120 000	7 130 000		0	0	0		
chr01	40 100 000	40 110 000	1	0	0	0	0	<i>Tos17</i>
chr02	32 880 000	32 890 000	1	1	0	0	0	<i>Tos17</i>
chr03	5 570 000	5 580 000	1	1	1	0	0	<i>Tos17</i>
chr03	7 980 000	7 990 000	1	1	0	1	1	<i>Tos17</i>
chr03	34 640 000	34 650 000	1	1	0	1	1	<i>Tos17</i>
chr04	30 360 000	30 370 000	1	1	1	1	1	<i>Lullaby</i>
chr09	19 670 000	19 680 000	1	0	1	1	0	<i>Tos17</i>
chr09	21 060 000	21 070 000	1	1	0	0	0	<i>Tos17</i>
chr11	470 000	480 000	1	0	0	0	0	<i>Tos17</i>
chr11	25 230 000	25 240 000	1	1	0	0	0	<i>Tos17</i>

Tab. 2.8 Tableau récapitulatif des néoinsertions de *Karma* détectées par analyse des lectures coupées.

Élément	Chromosome d'arrivée	Position d'arrivée	T0	T1	T2	T3	T1	T2	T3	T1	T2	T3
			23.1	23.1 15 Ta2	23.1 15 Ta2.2	23.1 15 Ta2.2.2	23.1 23 Ta3	23.1 23 Ta3.3	23.1 23 Ta3.3.3	23.1 31 Ta1	23.1 31 Ta1.2	23.1 31 Ta1.2.1
<i>Karma</i>	chr01	38 120 216									✓	
<i>Karma</i>	chr01	40 172 239									✓	✓
<i>Karma</i>	chr03	3 075 010			✓							
<i>Karma</i>	chr03	27 312 040				✓						
<i>Karma</i>	chr03	27 575 352						✓				
<i>Karma</i>	chr04	26 110 697										✓
<i>Karma</i>	chr04	31 112 482										✓
<i>Karma</i>	chr05	21 726 526								✓		
<i>Karma</i>	chr05	21 745 519								✓		
<i>Karma</i>	chr06	26 100 705										✓
<i>Karma</i>	chr07	3 785 856								✓		
<i>Karma</i>	chr08	20 223 751						✓				
<i>Karma</i>	chr08	27 759 929						✓				
<i>Karma</i>	chr09	17 223 474								✓		
<i>Karma</i>	chr09	17 796 062										✓
<i>Karma</i>	chr10	18 397 299										✓
<i>Karma</i>	chr11	5 218 764								✓		
<i>Karma</i>	chr11	5 779 285										✓
<i>Karma</i>	chr11	21 037 135										✓
<i>Karma</i>	chr11	25 590 961						✓				
<i>Karma</i>	chr12	888 805								✓		

Détection par le pipeline d'analyse :

Une intersection des régions obtenues avec la méthode des paires discordantes et des lectures chimériques a permis de déterminer des points potentiels de néo-insertions. Un résumé des néo-insertions détectées par cette méthode est présenté dans le tableau 2.6. L'étape finale avant la validation en laboratoire est la validation *in silico* par analyse de dotter. Les profils de ces dotters ne permettent pas de valider toutes ces néo-insertions, soit parce que nous n'observons pas de vraies lectures coupées au point d'insertion, soit parce qu'une des lectures de la paire discordante s'aligne au niveau d'une région complexe, comme de courtes répétitions en tandem. En revanche, l'analyse des néo-insertions des éléments *Tos17* et *Lullaby* montre des profils attendus pour la détection d'un vrai positif, un exemple est présenté dans la Figure 2.9. Ainsi pour la plante *Osnrpd1* T0, 11 néo-insertions de *Tos17* ont été détectées avec le pipeline d'analyse, 7 dans la plante T1 23.1-15, 2 dans la plante T1 23.1-17, 4 dans la plante T1 23.1-23 et enfin 5 dans la plante T1 23.1-31 (Tab. 2.7).

Une insertion de *Tos17* n'est pas détectée dans la plante T1 23.1-31 (chr03-5 572 693) avec le pipeline d'analyse des points de néo-insertions. Il s'agit d'un faux négatif non détecté à cause du filtre sur les paires discordantes. Une insertion de *Lullaby* a été détectée dans toutes les plantes *Osnrpd1* T0 et T1 étudiées.

Dans les générations suivantes, les plantes *Osnrpd1* ne présentent pas de nouvelles insertions pour ces éléments, en revanche il y a ségrégation des copies de *Tos17* dans les plantes (données non présentées). Ces résultats suggèrent que la culture cellulaire permet la réactivation rétrotranspositionnelle des éléments *Tos17* et *Lullaby*. En revanche, cette capacité est perdue après la régénération.

Détection par analyse des clipped-reads :

L'analyse des lectures rognées par analyse des données du CIGAR après alignement des lectures du séquençage sur le génome de référence a permis de détecter des néo-insertions de l'élément LINE *Karma*. Les positions de ces insertions ont été établies (Tab. 2.8). Pour les plantes 15 et 23 de la lignée 23.1 d'*Osnrpd1* il faut attendre la 3^{ème} génération d'auto-fécondation avant d'observer des néo-insertions de *Karma*. En revanche, pour la plante 31, qui est la plante ayant perdu la construction RNAi (T-DNA ^{+/+}), on observe des néo-insertions de *Karma* dès la 2^{ème} génération dans la plante 23.1-31Ta1.2.

Tab. 2.9 Tableau des néo-insertions détectées par la méthode du DOC

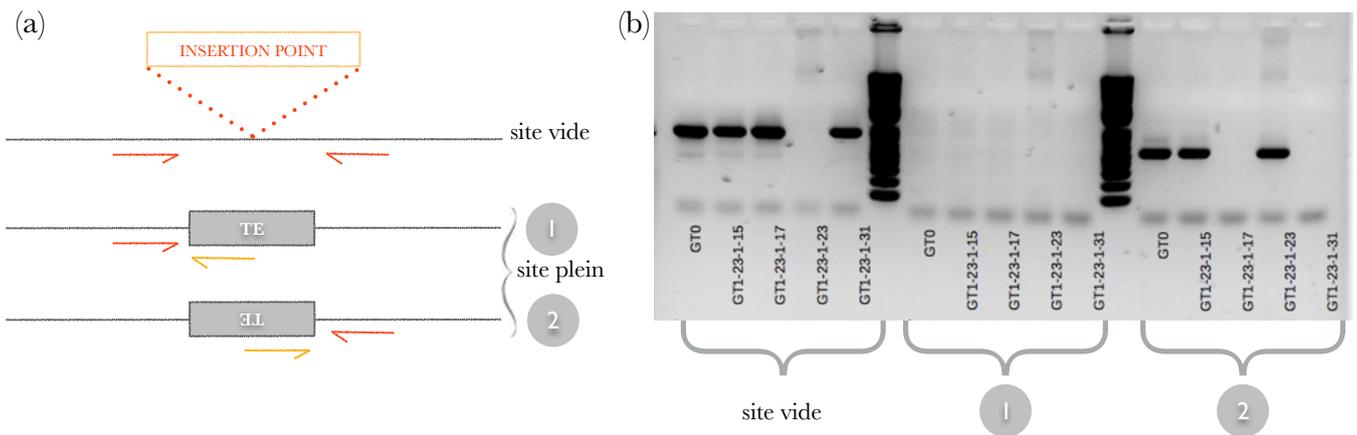
Ce tableau présente le nombre moyen de néo-insertions établi dans différentes plantes *Osmorhiza* cultivées en conditions normales par l'analyse DOC. Ces valeurs moyennes sont encadrées par un intervalle de confiance au seuil 95% pour les écarts-types.

Élément	GM	23.1	23.1-15Ta2	23.1-15Ta2.2	23.1-15Ta2.2.2	23.1-23Ta2	23.1-23Ta2.2	23.1-23Ta3	23.1-23Ta3.3	23.1-23Ta3.3.3	23.1-31Ta1	23.1-31Ta1.2	23.1-31Ta1.2.1
<i>LTR_fam79</i>	44 ^{+/-} .8	46 ^{+/-} .13	199 ^{+/-} .51	39 ^{+/-} .7	211 ^{+/-} .40	95 ^{+/-} .21	172 ^{+/-} .33	122 ^{+/-} .22	215 ^{+/-} .50	106 ^{+/-} .22	133 ^{+/-} .25	109 ^{+/-} .21	128 ^{+/-} .25
<i>LTR_fam80</i>	37 ^{+/-} .7	38 ^{+/-} .11	132 ^{+/-} .34	32 ^{+/-} .5	140 ^{+/-} .27	69 ^{+/-} .15	116 ^{+/-} .22	85 ^{+/-} .15	141 ^{+/-} .33	76 ^{+/-} .15	92 ^{+/-} .17	77 ^{+/-} .15	88 ^{+/-} .17
<i>LTR_fam106</i>	13 ^{+/-} .2	14 ^{+/-} .4	64 ^{+/-} .16	43 ^{+/-} .7	68 ^{+/-} .13	29 ^{+/-} .6	55 ^{+/-} .11	38 ^{+/-} .7	69 ^{+/-} .16	33 ^{+/-} .7	42 ^{+/-} .8	34 ^{+/-} .7	40 ^{+/-} .8
<i>LTR_fam108</i>	52 ^{+/-} .9	51 ^{+/-} .15	151 ^{+/-} .39	113 ^{+/-} .19	157 ^{+/-} .30	88 ^{+/-} .19	134 ^{+/-} .26	104 ^{+/-} .19	159 ^{+/-} .37	96 ^{+/-} .19	111 ^{+/-} .21	96 ^{+/-} .19	106 ^{+/-} .21
<i>LTR_fam124</i>	6 ^{+/-} .1	6 ^{+/-} .2	15 ^{+/-} .4	12 ^{+/-} .2	16 ^{+/-} .3	9 ^{+/-} .2	13 ^{+/-} .3	10 ^{+/-} .2	16 ^{+/-} .4	10 ^{+/-} .2	11 ^{+/-} .2	10 ^{+/-} .2	11 ^{+/-} .2
<i>Lullaby</i>	2 ^{+/-} .0	3 ^{+/-} .1	3 ^{+/-} .1	3 ^{+/-} .1	3 ^{+/-} .1	4 ^{+/-} .1	4 ^{+/-} .1	4 ^{+/-} .1	4 ^{+/-} .1	4 ^{+/-} .0	3 ^{+/-} .1	2 ^{+/-} .0	4 ^{+/-} .1
<i>Tos17</i>	3 ^{+/-} .1	7 ^{+/-} .2	7 ^{+/-} .2	4 ^{+/-} .1	7 ^{+/-} .1	7 ^{+/-} .1	6 ^{+/-} .1	7 ^{+/-} .1	6 ^{+/-} .1	6 ^{+/-} .1	7 ^{+/-} .1	7 ^{+/-} .1	7 ^{+/-} .1

Fig. 2.10 Validation des néo-insertions par la technique du RBIP (Retrotransposon-Based Insertion polymorphism)

(a) Représentation schématique de la technique RBIP. Au niveau du point de néo-insertion, des amorces sont dessinées de part et d'autre de ce point afin d'amplifier le « site vide ». On utilise ensuite ces amorces couplées à une amorce dessinée dans la partie interne de l'élément (en orange) pour amplifier le « site plein ». Le couple d'amorce « site plein » qui permet une amplification indique dans quelle orientation se trouve inséré l'élément transposable.

(b) Gel d'électrophorèse et migration des produits de l'amplification d'une néoinsertion. Il n'y a pas de néo-insertion dans les plantes 23.1-17 et 23.1-31. L'insertion dans la plante 23.1-23 est homozygote et celle dans les plantes 23.1(GT0) et 23.1-15 est hétérozygote. L'élément *Tos17* s'est inséré en anti-sens dans les plantes 23.1, 23.1-15 et 23.1-23.



Détection par calcul de la profondeur de couverture :

Pour chacun des éléments testés le nombre de néo-insertions établi par la méthode du DOC est en accord avec les résultats obtenus avec l'AAP. Néanmoins, pour l'élément *Tos17* de la plante *Osnrpd1* 23.1, la méthode du DOC ne permet de détecter que 60% des copies de *Tos17* (cf. Tab. 2.4).

Par la suite, la méthode du DOC a été appliquée sur une base de données disponible dans l'équipe Génome et Évolution du LGDP qui se compose de 295 séquences de rétrotransposons. Les résultats présentés dans le Tableau 2.9 montrent les rétrotransposons pour lesquels le nombre de copies déduit avec la méthode du DOC diffère de celui de la plante WT.

On peut remarquer, pour ces éléments, que la différence du nombre de copies pour chaque élément n'est pas significativement différente entre la plante contrôle WT et la plante T0 23.1. En revanche, pour toutes les plantes de génération T1 : 23.1-15Ta2, 23.1-23Ta2, 23.1-23Ta3 et 23.1-31Ta1, le nombre de copies de ces éléments est significativement différent. Le nombre de copies pour les éléments *LTR_fam79*, *LTR_fam80*, *LTR_fam106*, *LTR_fam108* et *LTR_fam124* continue d'augmenter significativement dans les plantes de 2^{ème} génération pour les plantes 23.1-23Ta3.2 et 23.1-Ta3.3. Dans les générations successives, le nombre de copies se stabilise ou diminue. On peut aussi noter, pour la plante 23.1-15Ta2.2, que le nombre de copies des éléments à *LTR* est équivalent à ceux des plantes WT et 23.1, en revanche en génération 3, le nombre de copies est équivalent à ceux de la génération 1 pour cette plante.

Ces résultats suggèrent que ces éléments ne sont pas réactivés lors de la culture cellulaire, ni pendant la régénération, mais *in planta* dans les plantes *Osnrpd1*.

2.4.3. Validation des néo-insertions

Validation par la méthode RBIP :

Dans le cadre d'un stage de Master 1, Davy Pierron a réalisé la partie validation des néo-insertions de *Tos17* et *Lullaby*. Puisque le point de néo-insertion est connu, la technique Retrotransposon-Based Insertion Polymorphism (RBIP) a été utilisée. Cette stratégie se base sur l'utilisation d'une simple PCR entre une amorce externe flanquante et une amorce interne du rétrotransposon (approche « externe/interne »). La présence de l'ET sera traduite par une amplification « site plein » et son absence par une non-amplification « site vide ». Comme le sens de l'insertion du rétrotransposon n'est pas connu, le couples d'amorces « externe/interne » en déterminera le sens, en fonction de l'obtention d'un produit d'amplification soit avec l'amorce externe à gauche du point d'insertion soit avec l'amorce externe à droite. Une PCR complémentaire a été réalisée en utilisant des amorces externes de part et d'autre de l'insertion de l'élément

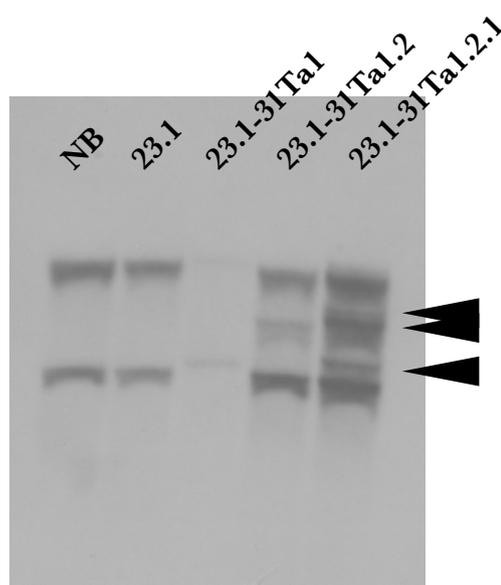
Tab. 2.10 Tableau récapitulatif des néoinsertions confirmées par PCR de *Tos17* et *Lullaby* détectées par analyse à priori, le point exact de l'insertion ainsi que le sens dans lequel est inséré l'élément et l'état de zygote dans chaque plante *Osnrpd1*

Dans ce tableau, He indique un état hémizygote de l'insertion de l'élément et Ho à l'état homozygote. Le sens de l'insertion est indiqué par « + » si elle est dans le même sens que celui de la copie native et par « - » si elle est en sens opposé.

Élément	Néo-insertion	Point exact d'insertion		23.1	23.1.15	23.1.17	23.1.23	23.1.31	Sens
		chr	Coordonnées						
<i>Lullaby</i>	Neo1	chr4	30 363 355	He	He	He	Ho	He	+
<i>Tos17</i>	Neo1	chr1	2 729 575	He	He			Ho	+
<i>Tos17</i>	Neo13	chr1	5 815 622				Ho	He	-
<i>Tos17</i>	Neo8	chr1	7 120 092	He				He	-
<i>Tos17</i>	Neo9	chr1	40 103 765	He				He	-
<i>Tos17</i>	Neo2	chr2	32 884 817	He	He				+
<i>Tos17</i>	Neo3	chr3	5 572 699	He	He	Ho			-
<i>Tos17</i>	Neo4	chr3	7 989 544	He	He		Ho		-
<i>Tos17</i>	Neo5	chr3	34 643 053	He	He		He	Ho	+
<i>Tos17</i>	Neo10	chr7	856 884	He			He	He	+
<i>Tos17</i>	Neo11	chr9	19 674 711	He		He	Ho		+
<i>Tos17</i>	Neo6	chr9	21 062 779	He	Ho				-
<i>Tos17</i>	Neo12	chr11	474 941	He					-
<i>Tos17</i>	Neo7	chr11	25 235 933	He	He				-

Fig. 2.11 Copies de *Karma* dans la plante *Nipponbare* et les plantes *Osnrpd1*

L'ADN génomique a été digéré avec l'enzyme EcoRV. NB est la plante WT, 23.1 la plante T0 *Osnrpd1*. Les plantes 23.1-31Ta1, 23.1-31Ta1.2 et 23.1-31Ta1.2.1 sont respectivement des plantes de 1^{ère}, 2^{ème} et 3^{ème} générations. Les flèches indiquent des bandes surnuméraires dans les plantes *Osnrpd1* 23.1-31Ta1.2 et 23.1-31Ta1.2.1.



(approche « externe/externe ») (Annexe 5). Comme les rétrotransposons peuvent présenter une longue séquence de plusieurs kb, une amplification complète de cet élément entre 2 amorces flanquantes ne sera pas souvent possible. L'utilité de l'approche « externe/externe » est avant tout d'établir l'état de zygote de l'insertion de l'élément. En effet, un individu homozygote pour l'insertion ne présentera pas d'amplification « site vide », alors que ce sera le cas pour un individu hémizyote pour l'insertion. Le schéma de la conception des amorces est présenté dans la Figure 2.10 (a). Un exemple de validation par RBIP d'une insertion de *Tos17* est présenté dans la Figure 2.10 (b).

De plus, les produits des amplifications « sites pleins » ont été séquencés et cela a permis de déterminer les points exacts de chacune des insertions après alignement de séquence sur la base de données pour l'annotation du génome du riz RAPDB (Rice Annotation Project DataBase).

Toutes les insertions détectées par AAP ont été validées, le tableau 2.10 présente l'état de zygote des néo-insertions, ainsi que le sens de l'insertion de l'élément et le point exact de l'insertion. Toutes les insertions de *Tos17* dans la plante T0 *Osnrpd1* 23.1 sont hémizygotiques et cette plante présente le plus de néo-insertions de *Tos17*. Entre les plantes T1 et T0, seul un nouvel événement de transposition a été détecté dans les plantes T1 23.1-23 et T1 23.1-31 à la même position dans le génome sur le chromosome 1 (Néo13). Toutes les autres insertions sont présentes dans la plante *Osnrpd1* T0 dont sont issues les T1.

Validation par analyse de Southern-blot :

Un des désavantages de l'approche du DOC est que, contrairement à l'AAP ou à la détection des lectures coupées, il nous est impossible de savoir où se sont insérées les nouvelles copies d'ETs. Nous ne pouvons pas non plus connaître quels ETs de la famille ont transposé. J'ai donc entrepris de valider les néo-insertions d'ETs par l'analyse de Southern-blot. L'avantage de cette technique est qu'il suffit d'une sonde dessinée dans l'élément à valider pour identifier tous les fragments d'ADN contenant cette sonde. De plus, l'hybridation de l'ADN à cette sonde peut être plus ou moins stringente, ce qui peut nous permettre d'identifier des ETs d'une même famille qui ne seraient pas totalement identiques.

Cette technique a permis de confirmer la réactivation transpositionnelle d'éléments pour les éléments actifs et réinsérés dans des régions uniques, par exemple pour l'élément *karma* (Fig. 2.11). Malheureusement, malgré différents tests avec différentes enzymes de restriction, aucune des insertions d'ETs actifs identifiées par l'analyse DOC qui se seraient insérées au niveau de régions répétées n'a pu être confirmée.

Tab. 2.11 Validation *in silico*

Éléments	GM	23.1-23-Ta3.3.3
<i>LTR_fam124</i>	2010	60118

Tab. 2.12 Tableau du nombre de copies d'éléments par famille détectées par l'analyse du DOC

- (a) Nombre moyen de copies d'éléments par famille dans les mutants Arawak et Altg10 comparés à ceux des plantes WT (GM) et *Osnrpd1* T0 (23.1). Ces valeurs moyenne sont encadrées par un intervalle de confiance au seuil 95% pour les écarts-types.
- (b) Nombre moyen de copies d'éléments par famille dans 3 jeux de données de la plante WT (GM) de couverture 10X.

(a)

Élément	GM	23.1	Arawak	Altg10
<i>LTR_fam79</i>	44 +/- .8	46 +/- .13	19 +/- .12	23 +/- .9
<i>LTR_fam80</i>	37 +/- .7	38 +/- .11	12 +/- .7	15 +/- .6
<i>LTR_fam106</i>	13 +/- .2	14 +/- .4	15 +/- .1	14 +/- .5
<i>LTR_fam108</i>	52 +/- .9	51 +/- .15	52 +/- .33	46 +/- .18
<i>LTR_fam124</i>	6 +/- .1	6 +/- .2	6 +/- .4	5 +/- .2
<i>Lullaby</i>	2 +/- .0	3 +/- .1	2 +/- .2	2 +/- .1
<i>Tos17</i>	3 +/- .1	7 +/- .2	4 +/- .2	3 +/- .1

(b)

Élément	GM-split1	GM-split2	GM-split3
<i>LTR_fam79</i>	43 +/- .9	45 +/- .11	36 +/- .7
<i>LTR_fam80</i>	36 +/- .7	37 +/- .9	31 +/- .6
<i>LTR_fam106</i>	13 +/- .3	13 +/- .3	10 +/- .2
<i>LTR_fam108</i>	51 +/- .10	53 +/- .13	47 +/- .10
<i>LTR_fam124</i>	6 +/- .1	6 +/- .2	5 +/- .1
<i>Lullaby</i>	2 +/- .1	2 +/- .1	3 +/- .1

Validation *in silico* :

Comme la validation par l'analyse de Southern-blot n'a pas permis de valider les néo-insertions d'éléments dans les régions répétées, une validation *in silico* a été effectuée sur l'élément LTR_fam124 qui est l'élément le moins répété des 5 ETs activés. Pour cela, toutes les lectures des paires discordantes qui ne s'alignent pas sur l'ET ont été récupérées puis toutes ces lectures sont alignées sur le génome de référence IRGSP-1.0. Le nombre de régions où ces lectures s'alignent a été récupéré (Tab. 2.11). Pour les plantes *Osnrpd1*, le nombre de régions où les ETs se sont réinsérés est plus important que celui de la plante contrôle (GM).

Néanmoins, la plante contrôle présente des régions d'insertion spécifiques, alors qu'on s'attend à ce que les régions d'insertion chez cette plante soient toutes communes avec les plantes mutantes. Ce résultat suggère, comme pour l'identification des polymorphismes entre les génomes de nos plantes et celui de référence, qu'il existe des variations de séquences entre génomes individuels.

2.4.4. Impact de la mutation RNAi sur l'activation rétrotranspositionnelle

Afin d'évaluer l'impact qu'a pu avoir la construction RNAi sur les plantes transformées *Osnrpd1*, et vérifier que les éléments réactivés le sont à cause de la mutation d'*OsNRPD1* et non pas à cause de la *CIV*, le nombre de copies des différents éléments a été évalué dans 2 mutants disponibles dans l'équipe. Ces deux mutants proviennent d'une collection de mutant ADN-T de riz développée dans le cadre du programme Géoscope au CIRAD de Montpellier (Larmande *et al.*, 2008). Ils ont été transformés et régénérés dans les mêmes conditions que le mutant *Osnrpd1*, permettant ainsi d'évaluer l'impact de la culture des cals sur la transposition. Ces mutants, que l'on nommera *Arawak* et *Altg10* ont été sélectionnés par M. El baidouri (thèse soutenue en 2012) car présentant des phénotypes d'intérêt agronomique. *Arawak* présente un phénotype de feuille tombante et *Altg10* présente un phénotype de nécroses spontanées et le mutant est plus résistant que la plante sauvage en présence du champignon phytopathogène du riz *Magnaporthe grisea*. Un séquençage en paired-end Illumina a été réalisé pour ces 2 mutants. Les profondeurs de couverture de ces séquençages ainsi que les lectures obtenues sont plus faibles que celles utilisées pour l'étude menée dans cette thèse (Annexe 7).

Après avoir déterminé la moyenne du nombre de lectures alignées sur une région unique ainsi que la dispersion de ces valeurs, l'étude du DOC a été réalisée avec les LTR-RTs et l'élément *Karma* du riz présentant plus de copies dans les plantes *Osnrpd1* (Tab. 2.12(a)). On remarque dans cette étude que pour les éléments *LTR_fam79* et *LTR_fam80*, le nombre de copies déduit par l'analyse DOC pour ces éléments est plus petit chez *Arawak* et *Altg10* par rapport à la plante sauvage

Tab. 2.13 Tableau des néo-insertions de *Tos17* et de *Karma* par la méthode du DOC

Nombre moyen de copies de *Tos17* et *Karma* dans les mutants Arawak et Altg10 comparés à ceux des plantes WT (GM) et *Osnrpd1* T0 (23.1). Ces valeurs moyenne sont encadrées par un intervalle de confiance au seuil 95% pour les écarts-types.

(a)

Élément	GM	23.1	Arawak	Altg10
<i>Tos17</i>	3 +/- 1	7 +/- 2	4 +/- 1	3 +/- 1
<i>Karma</i>	2 +/- 0	2 +/- 1	6 +/- 1	7 +/- 3

(b)

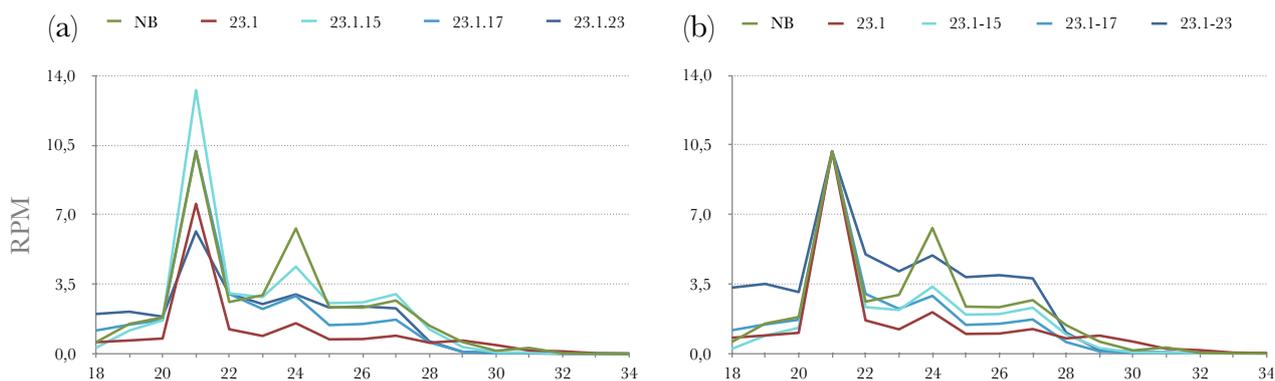
AAP	GM	23.1	Arawak	Altg10
<i>Tos17</i>	3	14	6	3
<i>karma</i>	3	3	13	6

Fig. 2.12 Distribution des sRNAs dans les plantes *OsNRPD1*

Distribution des tailles de sRNAs alignés sur le génomes de référence dans la plante Nipponbare WT et les transformants *Osnrpd1*

(a) corrigée par le nombre de sRNAs de 21-nt s'alignant au niveau de NRPD1, cible de la construction

(b) normalisation par rapport à la production des sRNAs de 21-nt.



(GM). Les profondeurs de séquençage entre ces plantes GM, *Altg10* et *Arawak* étant très différents : 24, 11 et 12 respectivement; nous avons d'abord pensé à un effet de couverture. Pour le vérifier les données de séquençage de la plante GM ont été aléatoirement divisées en 3 jeux de données représentant des couvertures de 10X du génome de riz. L'analyse DOC a ensuite été reconduite sur les éléments réactivés transpositionnellement (Tab. 2.12(b)). Cette dernière analyse montre que le nombre de copies déterminé par DOC varie selon le jeu de données utilisé.

En ce qui concerne les éléments connus pour être réactivés par la *CIV*, *Tos17* et *Karma*, des différences significatives du nombre de copies ont été observées dans ces mutants par l'analyse DOC et l'AAP (Tab. 2.13 (a) et (b)). *Altg10* et *Arawak* sont respectivement des plantes de génération T3 et T4 issues de plantes régénérées à partir de culture cellulaire de 6-7 semaines, comme c'est le cas des transformants *Osnrpd1*. Il est donc probable pour ces 2 mutants que le nombre de copies de *Tos17* ait été plus important dans les plantes T0, mais nous ne disposons pas de ces informations.

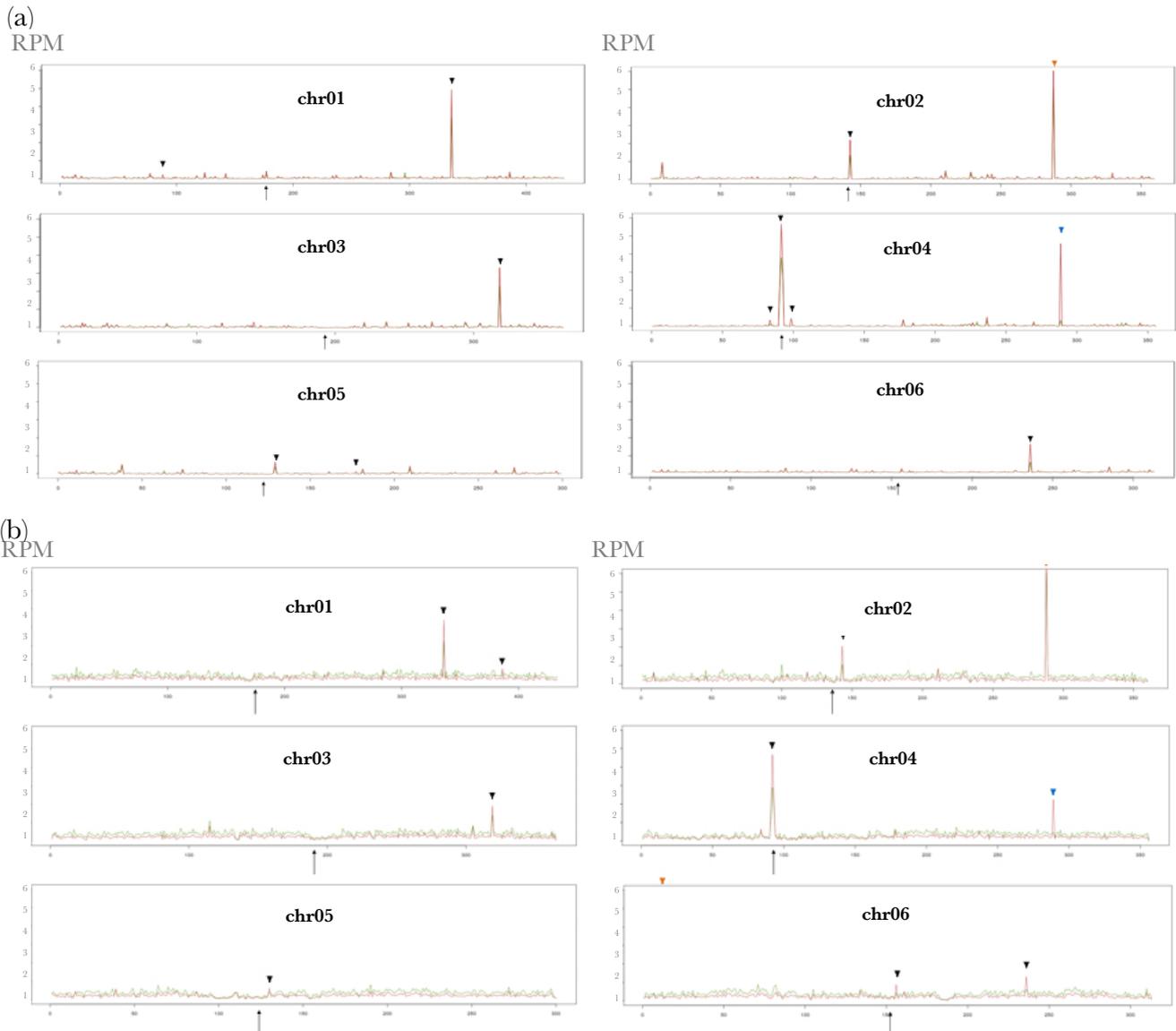
Ces résultats permettent néanmoins de confirmer que la réactivation rétrotranspositionnelle des éléments insérés dans les régions répétées n'est pas la conséquence de la culture cellulaire. Cette réactivation serait due au relâchement de la répression des ETs de la voie RdDM dans les mutants knock-down *Osnrpd1*.

2.4.5. Accumulation des sRNAs dans les plantes *Osnrpd1*

Pour chacune des bibliothèques, les longueurs des séquences de sRNAs alignées sur le génome (distribuées entre 18 et 34 nt) et l'accumulation des sRNAs sont représentées dans la Figure 2.12(a). Pour chaque catégorie de sRNAs une normalisation par rapport au nombre total de sRNAs de chacune des bibliothèques est effectuée. Étant donné que la construction introduite dans les plantes transformées induit un pic de production de sRNAs 21-nt au niveau de la séquence ciblée, la normalisation de chaque bibliothèque est corrigée par le nombre de sRNAs qui s'alignent sur *OsNRPD1*. La plante Nipponbare WT présente 3 pics à 21-nt (28,4%) 24-nt (17,6 %) et 27-nt (7,5 %). Pour les plantes *Osnrpd1*, après avoir enlevé les sRNAs de 21-nt s'alignant sur *OsNRPD1*, on remarque que le taux de production de 21-nt est inférieur à celui du WT pour les plantes 23.1 et 23.1-23, similaire pour la 23.1-17 et supérieur pour la plante 23.1-15. Toutes les plantes transformées présentent une accumulation moins importante de sRNAs de 24-nt (8,7% à 12 %) en comparaison avec la plante contrôle Nipponbare avec la plus faible production pour la plante T0 23.1 (8,7 %). Ces résultats vont dans le sens que la construction RNAi dirigée contre la grande sous-unité OsNRPD1 de la POLIV a eu un impact sur la biogenèse des sRNAs.

Fig. 2.13 Distribution chromosomique des sRNAs

Pour chaque chromosome, les distributions des sRNAs (a) 21-nt et (b) 24-nt sont tracées pour les génomes d'une plante Nipponbare (WT) en vert et d'une plante *Osnrpd1* de génération T0 (23.1) en rouge. L'axe des abscisses représente chacun des chromosomes par fenêtre de 100kb. L'axe des ordonnées représente le nombre de sRNAs en Reads Per Million (RPM). Les flèches indiquent la position des centromères. Les triangles bleus indiquent les positions des 2 gènes *O_sNRPD1* du riz sur les chromosomes chr04 et chr09. Les triangles oranges désignent les régions d'ADN ribosomiques et enfin les triangles noirs montrent des positions où l'enrichissement en sRNAs est plus important dans la plante transformée par rapport au WT.



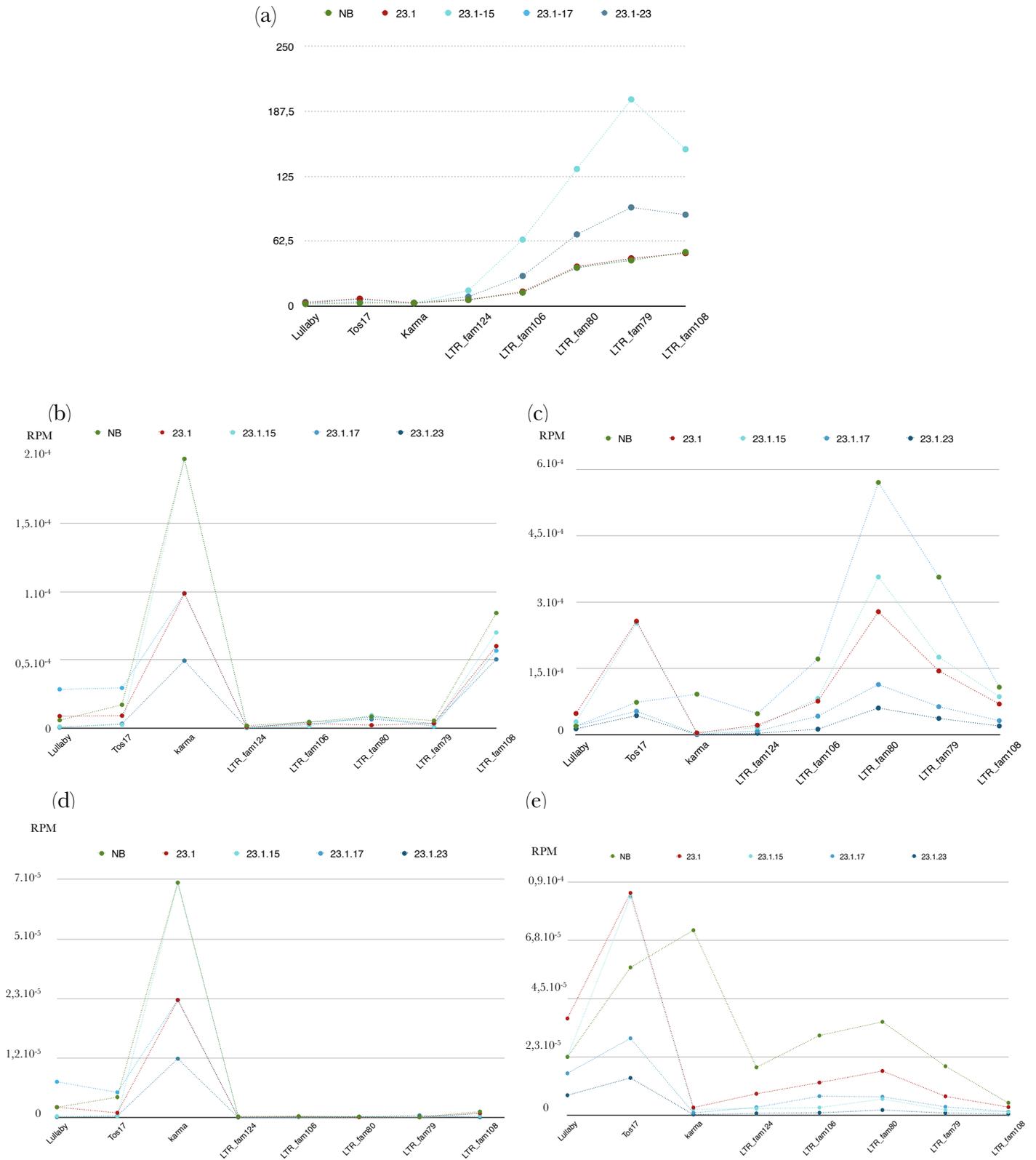
Enfin, pour évaluer l'effet du knock-down sur l'accumulation des sRNAs 24-nt dans chaque plante, les sRNAs ont été évalués en fonction de la quantité des sRNAs 21-nt (Figure 2.12(b)). L'accumulation des sRNAs 24-nt est plus faible dans les plantes transformées par rapport au WT, en revanche cette accumulation est plus importante dans les plantes T1 par rapport à la plante de génération T0. On peut également remarquer que la production des sRNAs 27-nt est plus importante dans la plante 23.1-23 par rapport aux autres plantes.

Afin de visualiser les dispersions globales des sRNAs de 21-nt et de 24-nt sur le génome de la plante mutante *Osnrpd1* 23.1, un alignement de ces sRNAs est réalisé sur les différents chromosomes du riz. Le nombre de sRNAs a été calculé par fenêtre de 100 kb sur le génome puis normalisé en RPM. L'alignement des sRNAs a ensuite été tracé pour chacun des 12 chromosomes du riz *O. sativa* cv. japonica (IRGSP-1.0) dans la figure 2.13 (a) (et Annexe 11) pour les sRNAs de 21-nt et (b) (et Annexe 12) pour les sRNAs de 24-nt. Premièrement, on peut observer que les régions de l'ADN portant de l'ADN ribosomique (indiquées par des triangles oranges) présentent l'accumulation la plus importante de sRNAs de 21-nt et de 24-nt dans la plante contrôle et la plante mutante (23.1). Ensuite, les régions d'*OxNRPD1* sur les chromosomes 4 et 9 (indiquées par des triangles bleus) présentent également une accumulation importante de sRNAs de 21-nt mais seulement pour la plante 23.1. Une accumulation de sRNAs de 24-nt est aussi observée au niveau de ces régions où se trouvent *OxNRPD1* dans la plante 23.1. Les triangles noirs indiquent des régions où l'accumulation de sRNAs est plus importante dans la plante 23.1 par rapport à la plante WT. On remarque que pour les chromosomes 2, 4, 5 et 11, certains de ces triangles sont localisés au niveau des centromères. De plus, la plupart des régions présentant des pics d'accumulation de 21-nt sont des régions de l'ADN codant pour des protéines d'organites. Les régions ciblées par les principaux pics de sRNAs de 21-nt le sont également par les sRNAs de 24-nt. Globalement, l'accumulation des sRNAs de 24-nt est plus faible dans la plante 23.1 en comparaison à la plante WT.

Cette analyse globale de la répartition des sRNAs montre que malgré une diminution de la quantité des sRNAs de 24-nt dans la plante *Osnrpd1*, certaines régions sont toujours préférentiellement ciblées. Cependant, cette représentation globale représente l'accumulation de sRNAs par fenêtre de 100kb du génome, ce qui ne permet pas de voir l'impact local de la mutation. C'est pourquoi l'analyse de régions d'intérêts, comme celles des ETs activés dans les plantes *Osnrpd1*, a été réalisée.

Fig. 2.14 Accumulation des sRNAs sur les ETs réactivés

- (a) Nombre de copies d'ET par famille active dans la plante WT et les plantes *Osmrpd1* déterminé par AAP et/ou DOC.
 (b) et (c) Nombre de sRNAs respectivement de 21-nt et 24-nt par famille d'ET.
 (d) et (e) Nombre de sRNAs respectivement de 21-nt et 24-nt par famille d'ET normalisé par le nombre de copies d'ET de cette famille.



Les ETs activés dans les plantes *Osnrpd1* cibles des petits ARNs ? :

Dans le but de comprendre les mécanismes impliqués dans la mobilisation des ETs réactivés transpositionnellement dans les plantes *Osnrpd1*, une étude a été réalisée sur l'accumulation des petits ARNs de 21-nt et de 24-nt, ciblant les ETs dans les voies du PTGS et du RdDM respectivement. Un récapitulatif du nombre de copies pour chaque famille d'ET, pour lesquelles des néo-insertions ont été détectées par les différentes analyses effectuées précédemment, est donné pour chaque plante dans la Figure 2.14(a). Puis la Figure 2.14(b) et (c) représente les nombres de sRNAs, respectivement de 21-nt et de 24-nt, accumulés au niveau de chaque famille d'ET. L'élément *Lullaby*, réactivé dans la plante T0 pendant la culture cellulaire, est plus ciblé par les sRNAs de 21-nt dans les plantes 23.1 et 23.1-17 par rapport à la plante sauvage NB (WT) et pour l'élément *Tos17*, seule l'accumulation des sRNAs de 21-nt dans la plante 23.1-17 est plus importante que pour la plante WT. Pour les ETs, *Karma*, *LTR_fam124*, *LTR_fam106*, *LTR_fam80*, *LTR_fam108* et *LTR_fam79*, la plante 23.1-15 présente une accumulation plus importante de sRNAs 21-nt que les autres plantes *Osnrpd1*, mais l'accumulation des sRNAs est plus importante dans le WT. Dans les plantes *Osnrpd1*, où la construction RNAi entraîne une production de sRNAs de 24-nt plus faible que dans un contexte normal, les sRNAs s'accumulent aussi au niveau des ETs. En ce qui concerne les éléments *Lullaby* et *Tos17*, ils sont ciblés par une plus grande quantité de ces sRNAs de 24-nt dans les plantes 23.1 et 23.1-15 par rapport au WT. Ces plantes sont également celles qui présentent le plus grand nombre de copies de *Tos17*. Pour les autres éléments actifs, la quantité des sRNAs de 24-nt est plus importante dans la plante WT. On remarque également que les accumulations des sRNAs de 21-nt et de 24-nt suivent des tendances opposées au niveau des éléments activés au cours de la première génération (*LTR_fam124*, *LTR_fam106*, *LTR_fam80*, *LTR_fam108* et *LTR_fam79*). D'autre part, nous avons observé que l'accumulation des ETs réactivés au cours de la génération T1 continuait dans certaines plantes de générations T2 (cf. Tab. 2.9). De ce fait, en faisant l'hypothèse que tous les ETs d'une même famille seraient tous potentiellement actifs et, donc la cible des sRNAs, une normalisation par le nombre de copies (déterminé par AAP et DOC) a été faite (Fig. 2.14 (d) et (e)).

Après la normalisation, on observe pour chacun des éléments les mêmes différences d'accumulation des sRNAs entre les différentes plantes. Néanmoins, dans les plantes *Osnrpd1*, les LTR-RTs activés au cours de la première génération étant plus répétés, la normalisation par le nombre de copies entraîne une diminution forte du nombre de sRNAs de 24-nt.

Afin de voir si la répartition des sRNAs sur les ETs est uniforme ou si certaines régions sont préférentiellement ciblées en fonction de leur activité transpositionnelle, les sRNAs ont été alignés sur des séquences d'ETs. Dans la figure 2.15, les quantités des sRNAs de 24-nt (a) et (b) et de 21-nt (c) et (d) ont été distribuées sur les éléments *Tos17* et *Karma* (a) et (c), pour lesquels des néo-insertions

Fig. 2.15 Distribution des sRNAs le long d' ETs

(a) Répartition des sRNAs de 24-nt sur les ETs *Tos17* et *Karma* actifs transpositionnellement dans les plantes *Osnrpd1*.
 (b) Répartition des sRNAs de 24-nt sur des ETs non réactivés transpositionnellement dans les plantes *Osnrpd1*.
 (c) Répartition des sRNAs de 21-nt sur les ETs *Tos17* et *Karma* actifs transpositionnellement dans les plantes *Osnrpd1*.
 (d) Répartition des sRNAs de 21-nt sur des ETs non réactivés transpositionnellement dans les plantes *Osnrpd1*.
 Les traits rouge indiquent les LTRs (Long Terminal Repeats) de chaque rétrotransposon. Les traits violets les UTRs (UnTranslated Regions).

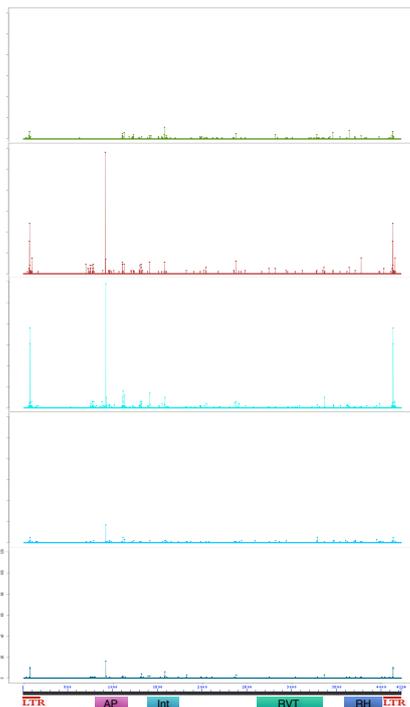
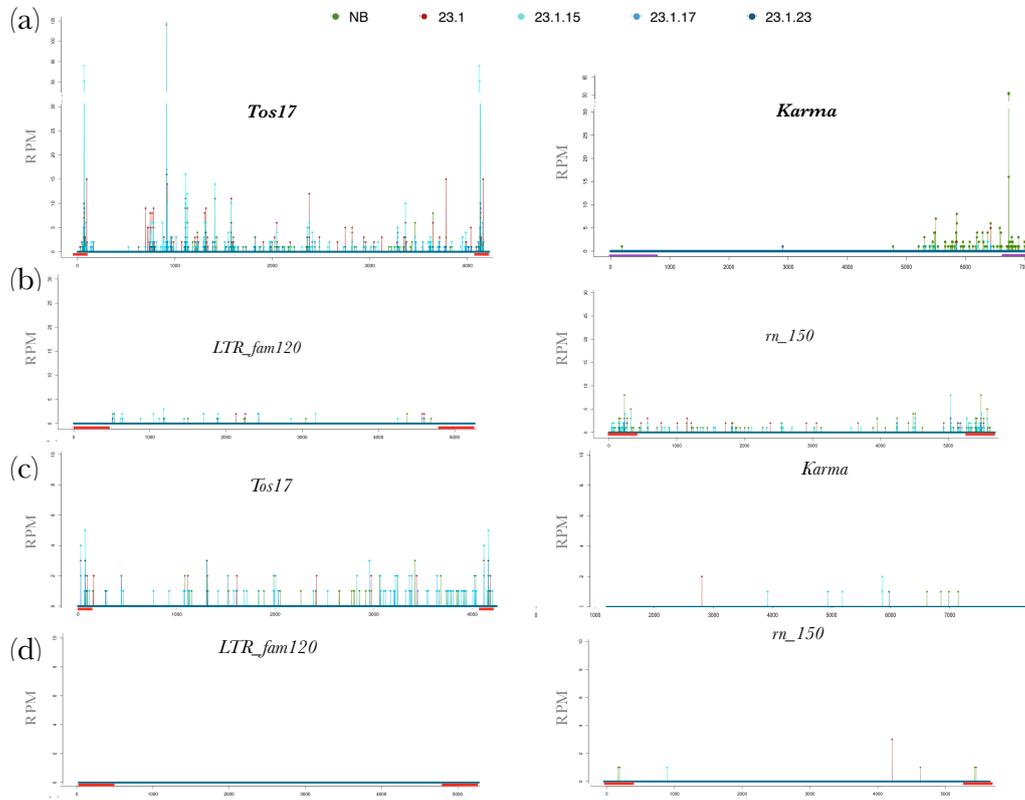


Fig. 2.16 Distribution des sRNAs 24-nt le long de *Tos17*

Représentation de la répartition des sRNAs et des domaines conservés de *Tos17*. Les LTRs (Long Terminal Repeats) sont indiquées en rouge, la gag en violet, Int (Intégrase) en cyan, RVT (ReVerse Transcriptase) en vert, RH (Rnase H) en bleu.

ont été observées dans les plantes mutantes, et pour les éléments *LTR_fam120* et *rn_150* (b) et (d) qui ne sont pas réactivés transpositionnellement dans les plantes mutantes. Sur la même figure sont représentées les distributions des sRNAs pour la plante contrôle NB, WT (en vert), la plante de génération T0, 23.1 (en rouge), et les plantes de génération T1 (en bleu). Les distributions des sRNAs 21-nt et 24-nt pour d'autres éléments actifs et inactifs dans les plantes *Osnrpd1* sont présentées en Annexes 13 à 16.

L'élément *Tos17* est d'une manière générale plus ciblé par les sRNAs de 24-nt et de 21-nt par rapport aux ETs inactifs dans les plantes *Osnrpd1*. On remarque la présence de 2 pics de sRNAs au niveau des LTRs de *Tos17* de 21-nt et de 24-nt. Dans la figure 2.16, la distribution des sRNAs le long de l'élément *Tos17* est présentée pour chacune des plantes utilisées pour tracer les distributions présentées dans la figure 2.14. On peut ainsi distinguer que l'élément *Tos17* est ciblé par une plus grande quantité de sRNAs 24-nt dans les plantes *Osnrpd1* 23.1 et 23.1-15 (présentant plus de copies de *Tos17*, cf Tab. 2.10) par rapport aux autres plantes. On observe également sur cette figure que l'accumulation de sRNAs est plus importante au niveau des LTRs du rétrotransposon, ce que l'on observe également pour les autres éléments dans les plantes où ils se sont néo-insérés (Annexe 13). Ce résultat montre qu'après son activation rétrotranspositionnelle dans les cals, *Tos17* est la cible des sRNAs de 24-nt et notamment ses LTRs. Dans les plantes T1 23.1-23 et 23.1-31, où *Tos17* n'est plus actif, les profils d'accumulation des sRNAs de 24-nt sont similaires au WT où *Tos17* est inactif.

Le profil d'alignement des sRNAs sur la séquence de *Tos17* présente 2 pics au niveau des LTRs mais également un troisième pic important. Dans le but de voir sur quel type de séquence s'aligne ces sRNAs, une recherche des séquences codantes de l'élément *Tos17* a été réalisée avec la base de données GyDB (Gypsy Database 2.0; Lloréns *et al.*, 2008; <http://www.gydb.org>) par comparaison BLASTX, les résultats sont présentés au bas de la figure 2.16. Le troisième pic de sRNAs de 24-nt est localisé au niveau du domaine codant pour la protéase aspartique (AP, Aspartic Protease) de l'élément *Tos17*. Le domaine AP sert à cliver la polyprotéine des LTR-RTs. Ce résultat suggère que la méthylation des ETs est dirigée spécifiquement au niveau de domaines clés pour l'activité des ETs.

En ce qui concerne l'élément *Karma* qui ne présente des néo-insertions qu'à partir de la génération T3 de ces plantes *Osnrpd1* (cf. Tab. 2.8), on remarque que l'accumulation des sRNAs de 24-nt est plus importante dans la plante NB, WT en vert. Le fait que cet élément ne soit pas ciblé par les sRNAs de 24-nt dans les plantes *Osnrpd1* de générations T0 et T1 dans lesquelles nous n'avons pas retrouvé d'événements de transposition suggère que la sous-accumulation de sRNAs de 24-nt n'est pas suffisante pour la rétrotransposition de *Karma* dans ces plantes.

Pour les éléments *LTR_fam120* et *m_150*, qui ne rétrotransposent pas, nous observons des profils de distribution de sRNAs 24-nt différents. L'élément *m_150* est ciblé par les sRNAs de 24-nt avec une plus forte accumulation de ces sRNAs au niveau des LTRs, comme pour l'élément *Tos17* qui a rétrotransposé dans les cals mais qui ne l'est plus dans ces plantes.

Enfin, lorsque nous regardons les distributions des sRNAs de 21-nt sur les ETs (Fig. 2.15 (c) et (d) et Annexes 14 et 16), on remarque que les ETs qui sont actifs rétrotranspositionnellement (*Tos17*, *LTR_fam80*, *LTR_fam108*) sont ciblés par des sRNAs de 21-nt, et particulièrement dans les plantes *Osnrpd1*. Alors que, si l'on regarde les éléments qui ne sont pas réinsérés, l'accumulation des sRNAs de 21-nt est presque inexistante. Ceci suggère que dans un contexte où la synthèse de sRNAs de 24-nt est affectée, la voie induisant la synthèse de sRNAs de 21-nt peut venir renforcer la répression des ETs actifs.

2.4.6. Différentiels d'expression entre les plantes Nipponbare (WT) et les plantes *Osnrpd1*

Notre objectif était principalement d'observer la réactivation de rétrotransposons utilisant un intermédiaire ARN pour leur transposition. Différents verrous s'exercent sur les ETs à différentes étapes de leur cycle de rétrotransposition, la première marque de la mobilisation d'un rétrotransposon est sa capacité à contourner le premier verrou afin de produire des ARNm. Ainsi, même si ils ne sont pas capables de se réinsérer par la suite dans un génome, il est possible d'identifier leur activité transcriptionnelle dans le génome.

Lorsque les ETs se réinsèrent dans le génome, il est possible que l'insertion se passe dans ou à proximité de gènes et, ainsi, influencer l'expression de ces gènes. De plus, la stratégie RNAi dirigée contre *OsnRPD1* entraîne une diminution des siRNAs de 24-nt de façon globale dont la synthèse permet également de guider la méthylation de certains gènes.

Pour évaluer l'impact de la mutation *Osnrpd1* sur l'activité transcriptionnelle, des séquençages des ARNm d'une plante Nipponbare WT et d'une plante *Osnrpd1* de génération T1 (23.1-15) ont été réalisés. Cette expérience comprend, pour chaque génotype, des triplicats techniques provenant de 3 extractions d'ARN totaux de feuilles de riz. La plateforme Illumina a été utilisée pour séquencer les ARNm (séquençage polyA+).

Fig. 2.17 Représentation globale des ETs différemment accumulés

Représentation des proportions des transcrits différemment accumulés dans la plante *Osnrpd1* T1 23.1-15Ta2 par rapport à la plante sauvage Nipponbare. Stables: 61%, Classe I sur-accumulés: 2%, Classe II sur-accumulés: 14%, Classe I sous-accumulés: 4%, classe II sous-accumulés: 19%.

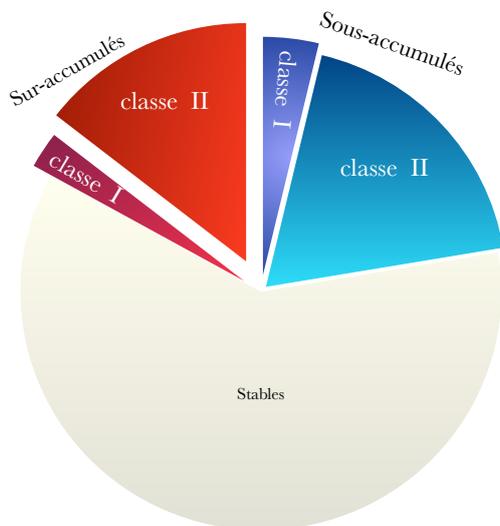


Fig. 2.18 Représentation des ETs actifs transpositionnellement exprimés dans la plante 23.1-15

Représentation du nombre de locus correspondant aux éléments transpositionnellement actifs dans les plantes *Osnrpd1*. Pour chaque élément, le nombre de locus présentant une proportion de transcrits >1 rpkm est indiqué sous le graphique circulaire correspondant à l'élément. Un locus est représenté en blanc si le niveau d'accumulation des transcrits est stable, en rouge si les transcrits sont différemment sur-accumulés et en bleu si ils sont différemment sous-accumulés entre *Osnrpd1* par rapport au WT.

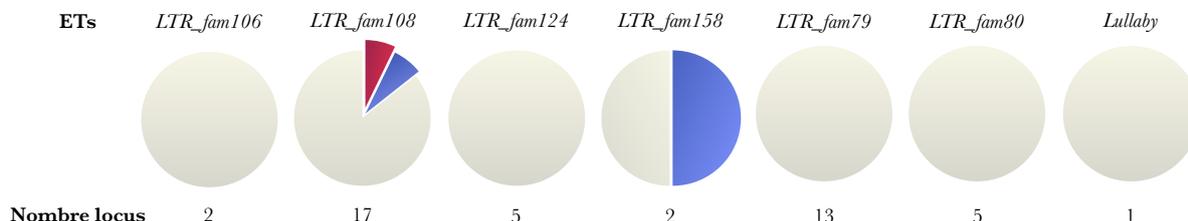
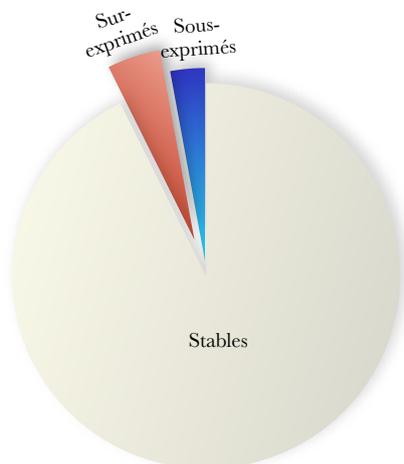


Fig. 2.19 Représentation globale des gènes différemment accumulés

Représentation des proportions des transcrits différemment accumulés dans *Osnrpd1* par rapport à Nipponbare. Stables: 93%, sur-accumulés: 4%, sous-accumulés: 3%.



Expression différentielle des ETs :

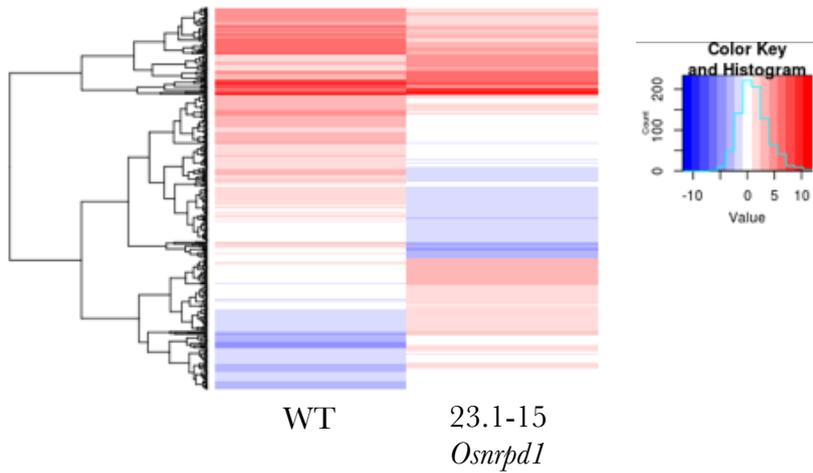
Pour évaluer l'activité transcriptionnelle des ETs, une analyse différentielle des transcrits d'une plante contrôle et d'une plante mutée a été réalisée *in silico* à partir de données de séquençage Illumina. Les séquences des ARNm de cette analyse ont été alignées sur la séquence du génome de référence IRGSP-1.0 et le comptage des séquences alignées a été normalisé en lectures par kilobase et par million (Reads Per Kilobase of transcript per Million mapped reads, RPKM). Les valeurs exprimées en RPKM sont utilisées pour normaliser le comptage des séquences en fonction de la longueur du contig et la profondeur du séquençage. On considère qu'un élément est exprimé lorsqu'au moins une des valeurs pour la plante WT ou la plante *Osnrpd1* est supérieure ou égale à 1. Dans notre étude, 6043 loci présentent un niveau d'accumulation de transcrits suffisant pour considérer qu'un ET est exprimé. L'analyse différentielle compare les valeurs de la plante *Osnrpd1* par rapport au WT, et une expression différentielle est considérée comme significative lorsque le Log2 (Fold Change) est compris entre -2 et 2, avec le Fold change = *Osnrpd1*/WT. Les résultats ont montré que 16% (1025 loci) des ETs étaient sur-exprimés et 23% (1352 loci) des ETs étaient sous-exprimés dans la plante *Osnrpd1* 23.1-15 par rapport à la plante WT. Les ETs les plus affectés par la mutation RNAi semblent être les éléments de classe II (Fig. 2.17). Cependant l'étude différentielle nous montre que globalement, le niveau d'accumulation des transcrits d'ETs issus de la plante *Osnrpd1* serait plus faible par rapport à celui de la plante WT.

Dans la plante 23.1-15, des néo-insertions d'ETs ont été trouvées (cf. Tab. 2.9) par la méthode du DOC. Les éléments *Tos17* et *Lullaby* ne se réinsèrent pas dans les générations suivant la régénération. Dans l'étude de l'expression différentielle, les néo-insertions des loci correspondant à ces ETs ont donc été analysées (Fig. 2.18). Les résultats montrent que pour presque tous les ETs, il n'y a pas de différentiel d'expression entre la plante 23.1-15 et la plante WT. Ces résultats suggèrent qu'au stade auquel les ARNs ont été extraits de la feuille, les ETs ne sont pas actifs transcriptionnellement.

En revanche, pour l'élément *LTR_fam108*, un locus, représentant un élément de cette famille, est différentiellement sous-exprimé (sur le chromosome 02) et un autre est sur-exprimé (correspondant à un locus sur le chromosome 01) dans la plante 23.1-15. Ce résultat suggère que l'activité de chaque membre d'une famille d'ETs n'est pas régulée de la même manière.

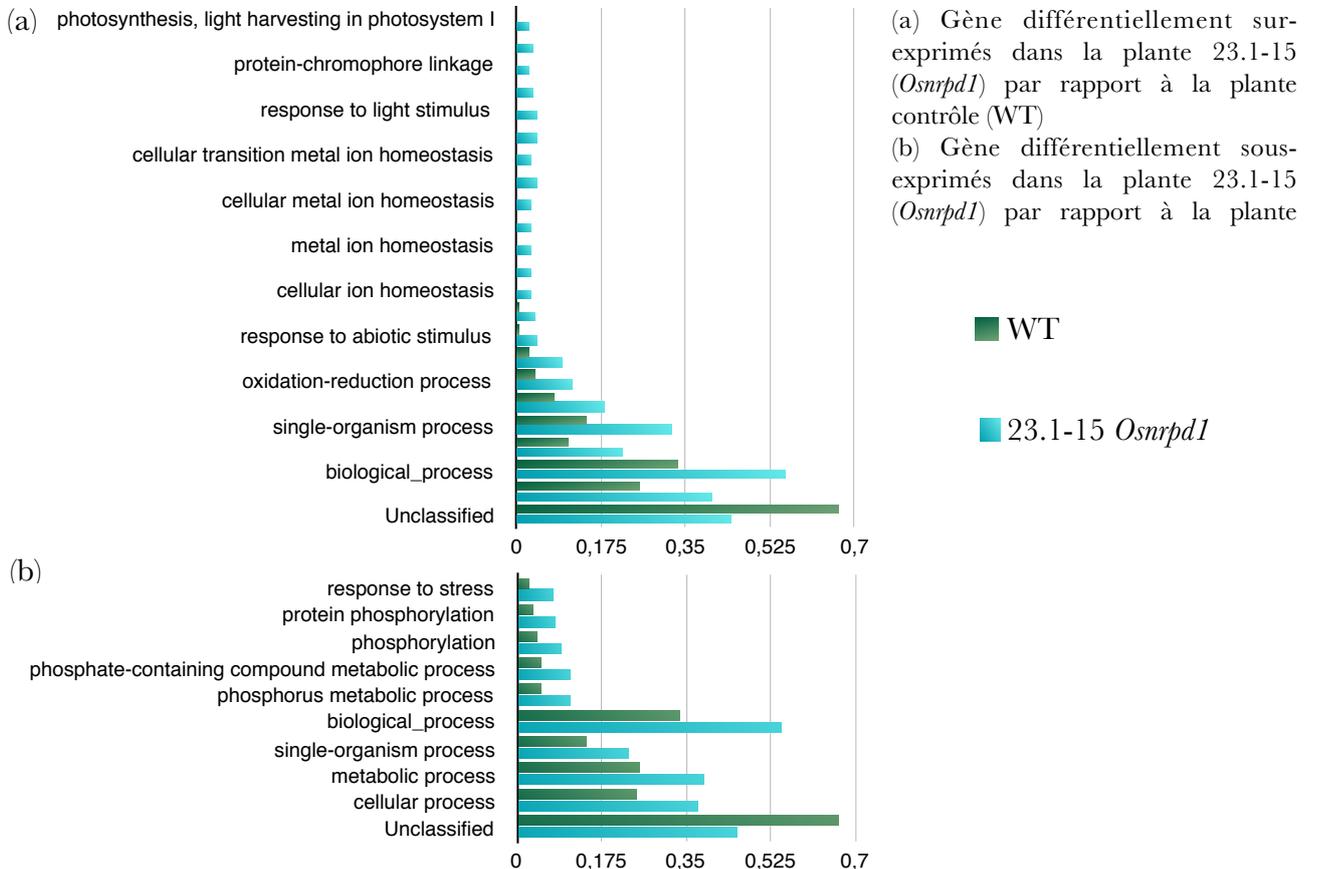
L'élément *Tos17*, présent en 2 copies dans le génome montre un locus sous-exprimé dans la plante 23.1-15 et un qui ne présente pas de différentiel d'expression avec la plante WT. Ces résultats suggèrent que l'activité de *Tos17* est réprimée dans la plante 23.1-15 où il n'est plus actif transpositionnellement.

Fig. 2.20 Heatmap des gènes différentiellement exprimés dans la plante *Osnrpd1* 23.1-15 versus la plante contrôle WT



La représentation heatmap montre la répartition des transcrits différentiellement exprimés (DE) en fonction des valeurs normalisées en RPKM. L'échelle de coloration représente les valeurs données en logarithme décimal.

Fig. 2.21 Gene Ontology des gènes différentiellement exprimés



Expression différentielle et description des gènes :

D'autre part, une étude a également été réalisée pour évaluer le différentiel d'expression de gènes dans la plante *Osnrpd1* 23.1-15 par rapport à la plante WT. Les mêmes filtres que ceux utilisés pour déterminer les différentiels d'expressions des ETs ont été utilisés pour les gènes. Dans un premier temps, une analyse globale des gènes exprimés (9701) a été réalisée et les résultats sont présentés sous forme d'un diagramme circulaire (Fig. 2.19). Les résultats montrent que 93% (8993) des gènes ne présentent pas de différentiel d'expression avec la plante WT, 4% (429) sont significativement sur-exprimés et 3% (279) sont sous-exprimés. Afin de mieux visualiser les 708 gènes différentiellement exprimés (DE) entre la plante *Osnrpd1* 23.1-15 et la plante WT, une heatmap a été construite à partir des valeurs de l'accumulation des transcrits de gènes DE (Fig. 2.20). On remarque que la clusterisation, basée sur les similarités de comportements entre les gènes, permet de différencier 3 principaux clusters. Le premier cluster est représenté par les gènes les plus exprimés dans les 2 plantes, un second par les gènes faiblement exprimés dans la plante *Osnrpd1* et le troisième où les gènes sont faiblement exprimés dans la plante contrôle. Ceci suggère que la mutation d'*OsNRPD1*, par la diminution de synthèse de siRNAs, a un impact sur la régulation de l'expression des gènes.

Afin de comprendre l'impact qu'a pu avoir la mutation d'*OsNRPD1* sur la plante 23.1-15, la classification des gènes DE a été analysée en utilisant les catégories GO (Gene Ontology) de la base de données PANTHER (Protein ANalysis THrough Evolutionary Relationships) pour identifier les processus biologiques affectés (Fig. 2.21). Cette analyse révèle que les gènes différentiellement sur-exprimés dans la plante 23.1-15 sont principalement impliqués dans des processus de réponse à la lumière, le maintien de l'homéostasie d'ions et la réponse à des stimuli abiotiques. Six gènes sont exprimés dans la plante 23.1-15 et ne le sont pas dans la plante WT. Trois d'entre eux sont hypothétiques, les 3 autres codent chacun pour une protéine de la famille des UDP-glucuronosyl/UDP-glucosyltransferase, pour une protéine oxygénase des zéaxanthines (zeaxanthin cleavage oxygenase) et pour une protéine similaire à CsAtPR5, un précurseur de gènes PR5 (Pathogenesis-related protein 5). Les gènes différentiellement sous-exprimés sont en partie induits en réponse au stress et à des processus de phosphorylation. Trois de ces gènes ne sont pas exprimés dans la plante 23.1-15. Il s'agit d'un gène hypothétique, d'un gène codant pour une peroxydase et d'un gène codant pour une protéine de liaison aux GAMYB (Gibberellin- and Abscisic acid- regulated MYB) situé sur le chromosome 10, des séquences spécifiques des régions promotrices de gènes de réponse aux GA (Gibbérélines). Ces résultats montrent que la mutation *Osnrpd1* entraîne de manière directe ou indirecte une modulation de l'expression des gènes induits en réponse au stress. L'insertion d'ETs à

Fig. 2.22 Cytosines méthylées dans les différents contextes

Graphique présentant le nombre de cytosines méthylées dans les différents contextes de chaque génome étudié, une plante contrôle (WT), une plante *Osnrpd1* T0 (23.1) et 2 plantes *Osnrpd1* de génération T1 (23.1-15 et 23.1-23). En contexte CG, CHG et CHH le nombre de cytosines méthylées (en %) est respectivement pour la plante WT: 82.35, 82.94 et 76.76; pour la plante 23.1: 57.81, 64.87 et 66.63; pour la plante 23.1-15: 80.93, 83.65 et 81.02; pour la plante 23.1-23: 85.58, 87.63 et 86.00.

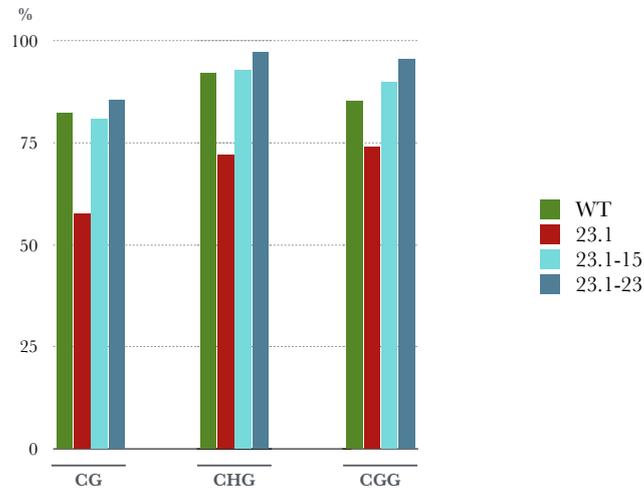
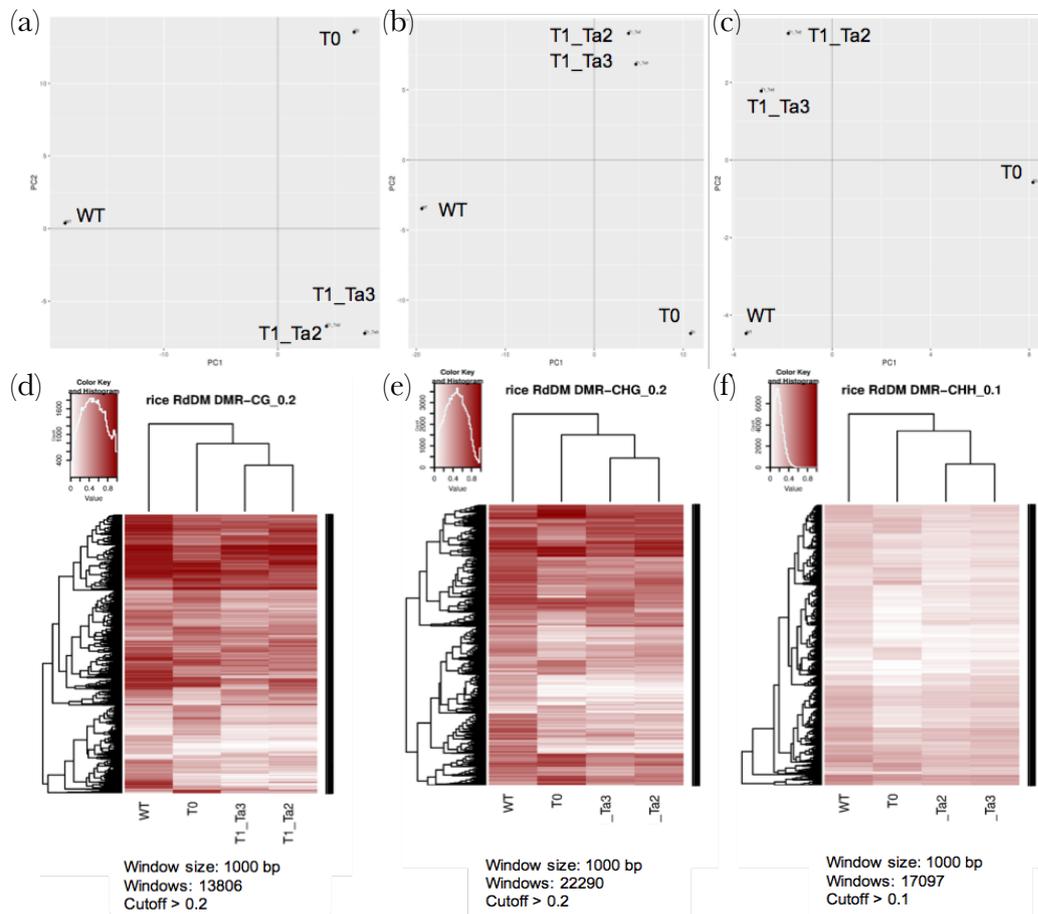


Fig. 2.23 Méthylation globale dans les différents génomes

(a),(b) et (c) Représentations en composante principale des données de méthylation en contexte CG, CHG et CHH respectivement.

(d), (e) et (f) Heatmaps des profils de méthylation dans les différents génomes et dans les contextes CG, CHG et CHH respectivement.



proximité ou dans des gènes peut avoir un impact sur la fonction de gènes. C'est pourquoi la détection d'ETs dans l'environnement local des gènes différentiellement exprimés serait intéressante.

2.4.7. Effet de la mutation d'*OsNRPD1* sur le méthylome des plantes *Osnrpd1*

Afin d'identifier les changements de méthylation de l'ADN induits dans les mutants *Osnrpd1*, l'étude du méthylome d'une plante WT, et 3 plantes *Osnrpd1* : 23.1 (T0), 23.1-15Ta2(T1) et 23.1-23Ta3(T1) a été réalisée par séquençage au bisulfite en collaboration avec le laboratoire de Blake Meyers (Université du Missouri, USA).

Les résultats présentés dans cette partie proviennent d'une analyse préliminaire effectuée par nos collaborateurs. Les profils de méthylation au niveau du génome entier ont été réalisés en déterminant le niveau de méthylation pour chaque cytosine du génome. Les séquences ont été alignées sur le génome de référence du riz Nipponbare (IRGSP-1.0) avec le logiciel BS Seeker2 (Guo *et al.*, 2013). Le nombre de lectures s'alignant sur le génome varie entre 144 M et 179 M, ce qui représente des couvertures variant de 16,85 à 21,92 X du génome (Annexe 9). La conversion au bisulfite permet de convertir les cytosines non méthylées en thymines, l'efficacité de la conversion au bisulfite a été estimée pour chaque cytosine dans les différents contextes (CG, CHG et CHH). L'efficacité de la conversion au bisulfite a été évaluée en considérant les cytosines couvertes par au minimum 4 lectures, le taux de conversion atteint au minimum 96% pour chacune des bibliothèques (Annexe 10).

Ensuite, le nombre de cytosines méthylées dans chaque génome a été évalué (Fig. 2.22). Les résultats montrent que le nombre de cytosines méthylées dans les différents contextes est plus faible dans la plante T0 23.1 par rapport à la plante WT. En revanche, les plantes *Osnrpd1* de génération T1 présentent des taux plus élevés de cytosines méthylées, par rapport à la plante WT, dans tous les contextes pour la plante 23.1-23 et dans les contextes CHG et CHH dans la plante 23.1-15. Une autre étude a été réalisée afin d'examiner les différences de méthylation entre différents génomes en utilisant la méthode d'analyse en composante principale (Fig. 2.23 (a), (b) et (c)). Les profils de méthylation dans les plantes *Osnrpd1* sont différents de ceux de la plante WT mais ils sont également différents entre les plantes *Osnrpd1* T0 et T1. En revanche, ils sont très proches entre les plantes 23.1-15 et 23.1-23 en contexte CG et CHG, moins en contexte CHH. Ceci indique que la mutation RNAi d'*OsNRPD1* entraîne une hypométhylation mais qu'au fil des générations l'efficacité de la mutation diminue et le niveau de méthylation est plus élevé.

Puisque les profils de méthylation dans les différents contextes sont différents entre les différentes plantes étudiées, les changements de méthylation des cytosines dans les différents

Fig. 2.24 Niveau de méthylation des plantes *Osnrpd1* comparées au WT

Répartition de la méthylation de chaque plante *Osnrpd1* sur le génome en comparaison à la méthylation de la plante WT. La méthylation en contexte CG est représentée en vert, la méthylation CHG en bleu et la méthylation CHH en rouge.

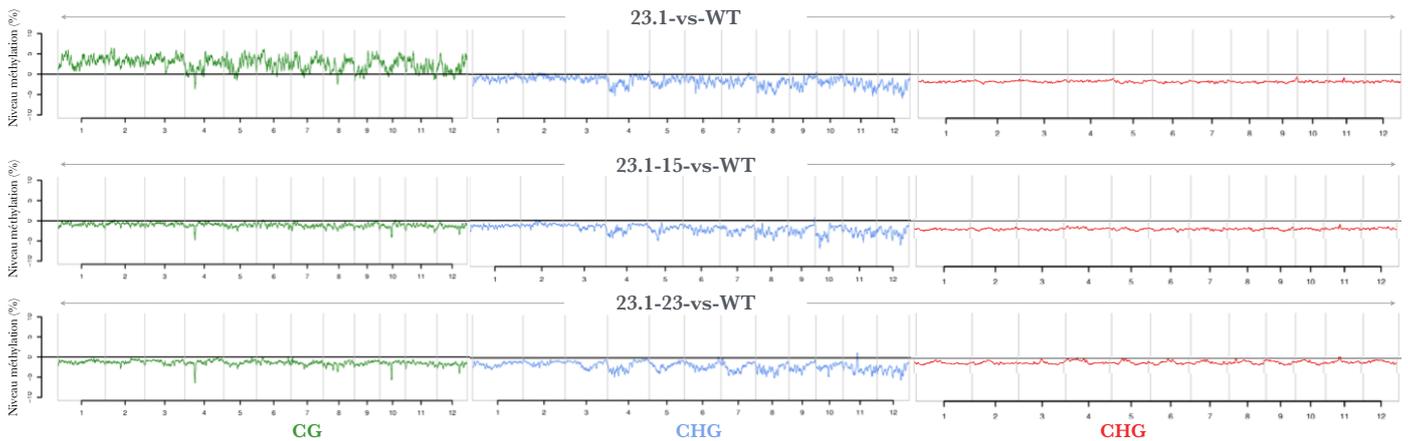
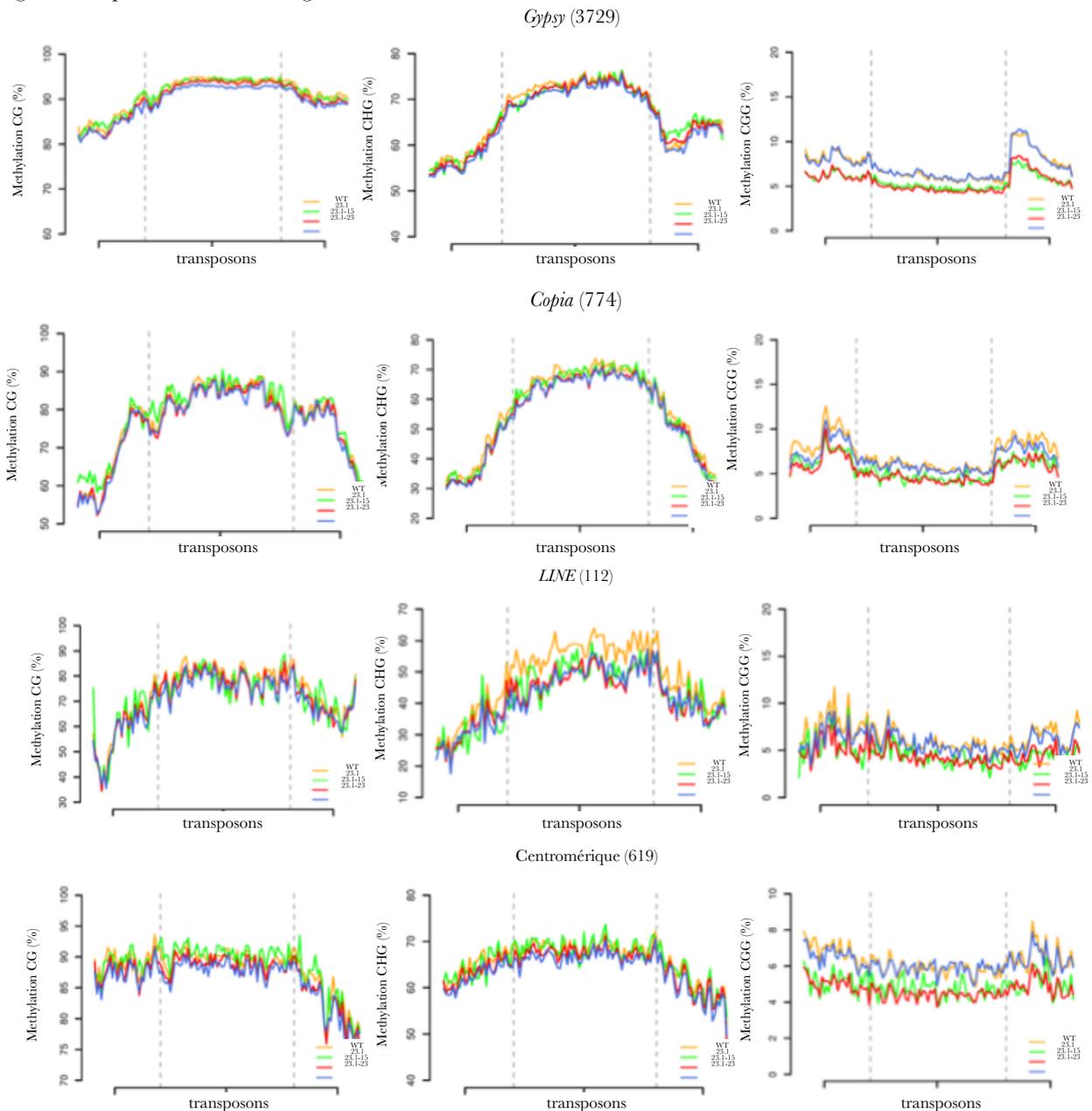


Figure 2.25 Distribution de la méthylation de différentes classes d'ETs

Distribution des niveaux de méthylation des cytosines dans les 3 contextes le long de différentes classes d'ETs et de leurs régions flanquantes à droite et à gauche.



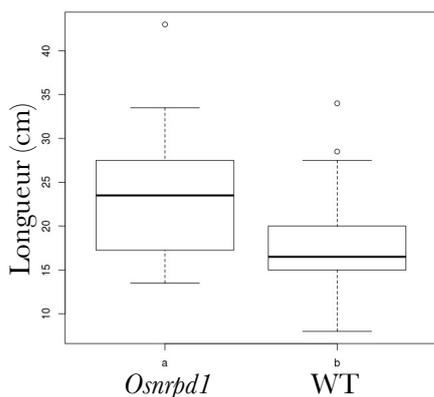
génomomes ont été distribués pour les contextes CG, CHG et CHH (Fig. 2.23 (d), (e) et (f) respectivement). On observe ici aussi que la plante WT présente le profil de méthylation le plus différent en comparaison aux autres plantes *Osnrpd1*. Une hypométhylation dans les plantes *Osnrpd1* était attendue dans des plantes où la stratégie RNAi cible *OsNRPD1*, la grande sous-unité de la polymérase IV.

Afin de voir si la distribution de la méthylation varie en fonction des régions du génome, une analyse comparative a été réalisée pour les méthylations dans les 3 contextes et en comparant les génomes des plantes *Osnrpd1* à celui de la plante WT (Fig. 2.24). Les résultats montrent que la méthylation des cytosines en contexte CG est plus importante dans la plante 23.1 par rapport à la plante contrôle, WT. Les autres méthylations dans les différents contextes sont plus faibles dans les plantes *Osnrpd1* en comparaison au WT. On peut également remarquer que la répartition de la méthylation n'est pas uniforme sur le génome, notamment au niveau de 2 régions fortement hypométhylées sur les chromosomes 4 et 10 en contexte CG.

Enfin, une analyse de la méthylation des cytosines dans les différents contextes a été réalisée sur différentes classes d'ETs afin de voir si la répartition de la méthylation est différente en fonction des génomes étudiés. Dans la Figure 2.25 se trouvent les distributions de la méthylation sur différentes classes d'ETs, les LTR-RTs de la famille des *Gypsy*, des *Copia*, des éléments de la classe des LINEs et des éléments spécifiques de régions centromériques. Les résultats montrent que dans le contexte CHH, la distribution de la méthylation de la plante 23.1-23 est similaire à celle du WT. En revanche, pour les plante 23.1 et 23.1-15, la méthylation en CHH est plus faible. Globalement, on ne remarque pas de différence dans la distribution de la méthylation dans les autres contextes, sauf dans le cas des éléments LINEs où la méthylation en contexte CHG est plus faible dans les plantes *Osnrpd1* par rapport à la plante WT.

Les données obtenues jusqu'à maintenant ne permettent pas encore d'étudier la méthylation des ETs locus par locus. Cependant, ces premiers résultats permettent d'identifier des classes d'ETs pour lesquels les profils de méthylation diffèrent de ceux du WT.

Fig. 2.26 Distribution des longueurs des feuilles drapeaux entre plantes sauvages et *Osnrpd1* de la générations T3



Pour chaque boîte à moustaches, le trait épais représente la médiane de l'échantillon, les limites inférieure et supérieure de la boîte sont respectivement le premier et le troisième quartile, les traits horizontaux reliés à la boîte par des lignes verticales représentent les premiers et neuvième déciles. Les valeurs extrêmes sont représentées par un cercle vide. *Osnrpd1* sont les plantes transformées et les plantes WT sont les plantes sauvages.

Fig. 2.27 Distribution des graines fertilisées entre plantes sauvages et transformées sur différentes générations

Pour chaque boîte à moustaches, le trait épais représente la médiane de l'échantillon, les limites inférieure et supérieure de la boîte sont respectivement le premier et le troisième quartile, les traits horizontaux reliés à la boîte par des lignes verticales représentent les premiers et neuvième déciles. Les valeurs extrêmes sont représentées par un cercle vide. *Osnrpd1* sont les plantes transformées et les plantes WT sont les plantes sauvages.

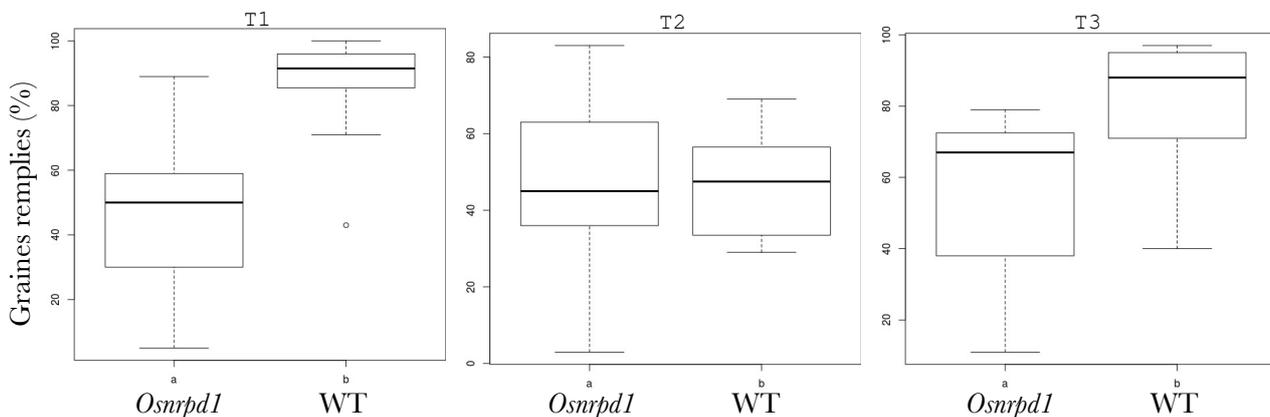
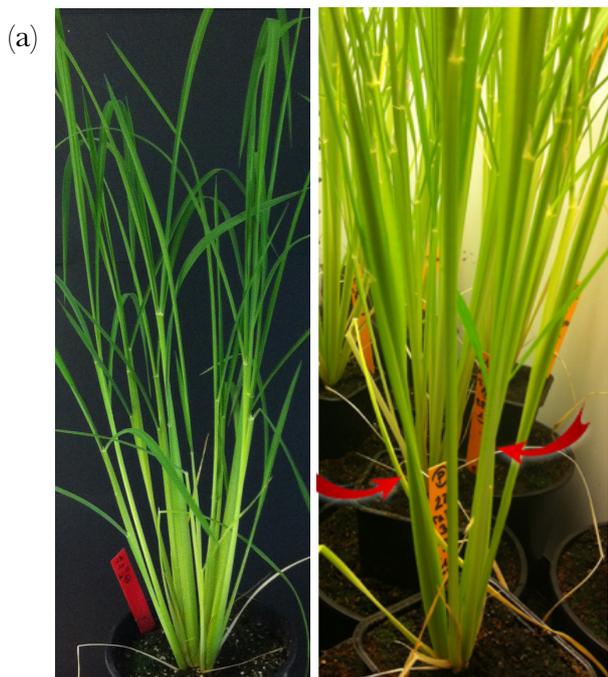


Fig. 2.28 Phénotype torsadé de plantes *Osnrpd1*



(a) Plant de riz *O. sativa* cv. Nipponbare (b) Talles torsadées d'une plante *Osnrpd1* de génération T5. Les flèches rouges pointent les endroits où la talle se tord.

2.4.8. Observation phénotypique des plantes

À l'issue de chaque génération de plantes, différents critères phénologiques et morphologiques ont été étudiés pour chacune des plantes, en particulier le moment de la floraison, le nombre de talles, la longueur de la feuille drapeau ainsi que le taux de remplissage des grains par épillets. Au cours de la deuxième génération, un retard de floraison de plus de 10 jours pour la talle Ta3 de la plante 23.1-23 a été observé par rapport aux plantes WT (donnée non présentée). La génération T3 issue de cette talle a été générée et présente également un retard de floraison, cependant ce retard n'est plus observé dans les plantes de génération T4. En revanche, au cours de la génération T4, c'est une autre plante (23.1-15 Ta2.2.2) qui présente un retard de floraison.

Le nombre de talles présenté par les différentes plantes Nipponbare sauvage et *Osnrpd1* ne montre pas de différences significatives.

Pour la longueur de la feuille drapeau, on observe une tendance à présenter des feuilles drapeaux plus longues chez les plantes *Osnrpd1* de la génération T3 par rapport aux WT (Fig. 2.26).

Après avoir récolté et dénombré le nombre de graines remplies de chaque épillet de chaque plante, une étude en box plot (Annexe 2) a permis de déterminer un défaut de remplissage des graines dans les plantes mutantes par rapport aux sauvages (Fig. 2.27). Les graines remplies correspondant aux graines fertiles, la mesure de ce trait permet d'évaluer l'impact de la mutation sur la capacité de la plante à se reproduire. Pour cette étude, le nombre de graines remplies a été comptabilisé sur 39, 54, 54 épillets de plantes mutantes (T-DNA ^m/_m) à celui de 15, 9, 9 épillets de plantes sauvages respectivement pour les générations T1, T2, et T3. Graphiquement, on peut observer une tendance au remplissage de toutes les graines chez les plantes Nipponbare cultivées en conditions normales pour les générations T1 et T3. En revanche, chez les plantes mutées homozygotes *Osnrpd1*, on observe une dispersion des valeurs du nombre des graines remplies par épillet plus importante avec plus ou moins 50 % des épillets présentant un taux de remplissage inférieur à 50 %. On observe pour les 3 générations de plantes que la valeur de la médiane est plus faible pour les plantes *Osnrpd1*.

Enfin, une autre génération (T5) de plantes *Osnrpd1* a mis en évidence un phénotype particulier des talles de certaines plantes qui présentent une torsion au niveau des talles (Fig. 2.28). Des graines issues de ces plantes ont été semées afin de valider le phénotype observé sur ces plantes.

Synthèse des principaux résultats de l'étude du mutant *Osnrpd1*, en contexte épigénétique relâché

Nomenclature du tableau : \times = Pas de données. \curvearrowright = Hypométhylation par rapport à la plante contrôle. \curvearrowleft = ETs activés transpositionnellement. \downarrow = ETs ne transposent plus, ségrégation. ? = habilité à la transposition indéterminée. LTR_famX = Les familles d'ETs actives transpositionnellement identifiées par DOC.

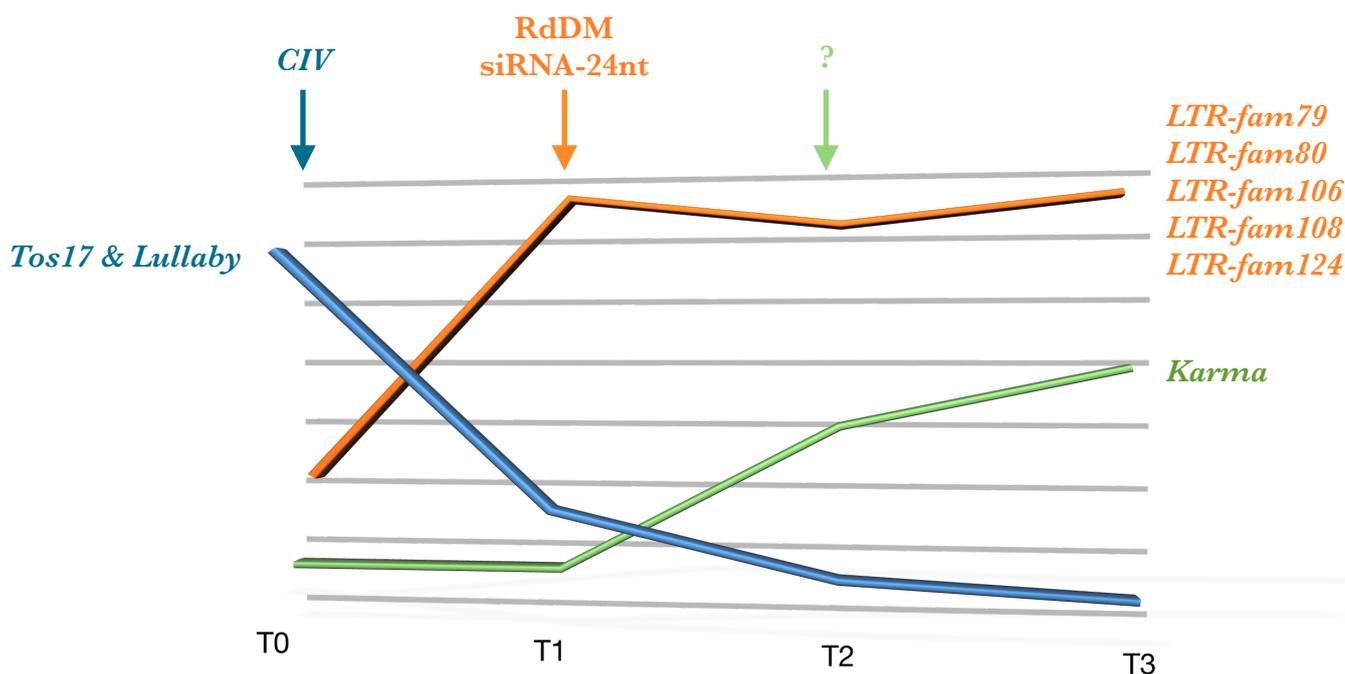
Données		Principaux résultats					
Plantes		WT	Cals	<i>Osnrpd1</i> T0	<i>Osnrpd1</i> T1	<i>Osnrpd1</i> T2	<i>Osnrpd1</i> T3
sRNA-seq 24-nt (RPM)		6.3	\times	2.07	2.3	\times	\times
BS-seq (%)	CG	82.35	\times	57.81	82.5	\times	\times
	CHG	82.94	\times	64.87	85.6	\times	\times
	CHH	76.76	\times	66.63	83.5	\times	\times
	ETs		\times	\curvearrowright	\curvearrowright	\times	\times
RNA-seq		>Différentiel d'expression de gènes entre une plante T1 <i>Osnrpd1</i> et le WT					
DNA-seq		Génération dans laquelle les ETs sont activés transpositionnellement					
Tos17		\times	\curvearrowleft	\curvearrowleft	\downarrow	\downarrow	\downarrow
Karma		\times	\times	\times	\curvearrowleft	\curvearrowleft	\curvearrowleft
LTR_famX		\times	\times	\curvearrowleft	?	?	?

Schéma récapitulatif des ETs activés dans les différentes générations de mutants

Osnrpd1

La mobilité des éléments *Tos17* et *Lullaby* a été montrée pendant la culture cellulaire. Celle de l'élément *Karma* a lieu dans les régénérants de plantes cultivées *in vitro*. Enfin, les éléments *LTR_fam79*, *LTR_fam80*, *LTR_fam106*, *LTR_fam108* et *LTR_fam124* sont réactivés dans les plantes *Osnrpd1* et leur mobilité n'a jamais été observée dans des plantes cultivées *in vitro*.

CIV = Culture *in vitro*; RdDM = RNA-directed DNA methylation; ? = Déclencheur indéterminé.



2.5. Discussion et perspectives

2.5.1. Approche expérimentale

Bien que les mécanismes de mobilisation des ETs aient été le sujet de nombreuses études, le(s) déclencheur(s) de leur réactivation *in planta* en conditions naturelles sont encore inconnus. En revanche, des réactivations d'ETs ont été observées dans des conditions dites plutôt artificielles, dans des plantes cultivées en CIV (Hirochika *et al.*, 1996; Grandbastien *et al.*, 1989; Picault *et al.*, 2009) ou dans des mutants épigénétiques (pour revue : Slotkin et Martienssen, 2007; Cui et Cao, 2014). Pour cela, nous avons entrepris d'étudier la mobilisation des ETs, des LTR-RTs particulièrement, à l'échelle du génome entier et ce pour différentes générations de plantes.

De plus, les ETs sont en partie réprimés par une voie de répression spécifique chez les plantes, la voie du RdDM, dont l'un des acteurs est POLIV. Jusqu'à présent, aucune étude de mutants *Osnrpd1* n'a encore été réalisée chez le riz. L'étude de ces mutants devait permettre de voir si dans un contexte où la répression des ETs est relâchée, l'activation d'ETs était facilitée.

Les différentes méthodes de détection utilisées ont été efficaces pour détecter des néo-insertions et leurs positions, ce qui nous a permis de les valider par PCR. Néanmoins, ces méthodes de détection ne permettaient pas d'identifier des néo-insertions qui auraient eu lieu dans des régions répétées du génome. Or, certains ETs sont connus pour cibler préférentiellement des régions géniques mais également des régions complexes lorsqu'ils se réinsèrent dans les génomes (Miyao *et al.*, 2003; Yang *et al.*, 2007). La méthode du DOC a été utilisée avec succès puisque des néo-insertions d'autres ETs actifs dans les plantes mutantes ont été identifiés alors que ces ETs n'avaient, jusqu'alors, jamais été décrits comme transpositionnellement actifs. Pour sa réalisation, cette méthode nécessite une base de données d'ETs. Hormis le fait que la base de données des LTR-RTs du riz ait été expertisée au sein de notre équipe, nous avons choisi d'étudier ces éléments car ils représentent la plus grande part des ETs dans le génome. Cependant, cette méthode de détection ne permet pas de déterminer avec précision le nombre exact de néo-insertions (Yoon *et al.*, 2009; Glusman *et al.*, 2015) ni leur position dans le génome contrairement aux autres méthodes d'alignements des lectures. De plus, l'état d'hétérozygotie des insertions peut conduire à une sous-estimation du nombre de néo-insertions. Ce cas de figure a été observé avec l'élément *Tos17* dans la plante T0, où toutes les insertions sont à l'état hétérozygote ce qui conduit à une estimation presque diminuée de moitié par rapport au nombre réel d'insertions.

Pour valider en laboratoire ces néo-insertions, une simple PCR n'est donc pas envisageable. La méthode du Southern-blot a donc été testée car elle nécessite seulement une sonde

complémentaire au fragment d'ADN à amplifier. Cette sonde est plus grande que la séquence d'une amorce ce qui permet d'être moins stringent quant à la séquence que nous voulions amplifier, critère intéressant pour amplifier différents ETs d'une même famille. Malheureusement, cette étude n'a pas permis de valider les néo-insertions d'éléments des régions répétées. Une des étapes du Southern-blot est une étape de digestion par des enzymes de restriction. L'efficacité de reconnaissance et de coupure d'un site donné dépend de son contexte et de l'état de complexité dans lequel se trouve l'ADN, ce qui pourrait être une explication de l'inefficacité à valider les néo-insertions dans les régions répétées du génome. Étant donné que les familles d'éléments néo-insérés et non validés regroupent plusieurs éléments, il faut également envisager que la sonde n'ait pas été dessinée dans la région la mieux conservée entre tous les représentants de cette famille. Si un seul élément de la famille est réactivé et qu'il ne contient pas la séquence de la sonde, ou qu'elle est très divergente, alors la validation n'est pas possible. Même si cette explication est envisageable, la probabilité que pour tous les éléments la sonde dessinée ne permet pas de valider au moins une des néo-insertions est très faible.

De ce fait, une validation *in silico* des néo-insertions a été réalisée pour un des éléments activés transpositionnellement. Cette validation permet de confirmer que le nombre de sites d'insertions est plus important pour les néo-insertions des plantes mutantes par rapport au sauvage. Étant donné que les sites d'insertion correspondent à des régions répétées du génome, le nombre de ces sites n'est pas proportionnel au nombre de néo-insertions détectées. Une analyse approfondie de ces régions répétées pourrait néanmoins permettre de valider quelques régions par PCR.

2.5.2. Réactivation transpositionnelle suite à la modification du contrôle épigénétique

Dans les plantes *Osnrpd1*, la stratégie RNAi ciblant le gène codant pour *OsNRPD1* a entraîné un défaut dans la biogenèse des sRNAs de 24-nt. L'analyse de la répartition globale de ces sRNAs sur les chromosomes du riz montre que des régions où des pics de sRNAs de 21-nt et de 24-nt sont observés dans les plantes WT le sont également dans les plantes *Osnrpd1*. Ces régions correspondent à de l'ADN codant pour des protéines adressées à des organites et des séquences d'ADN transcrites en ARN ribosomiques. Ce résultat semble indiquer qu'il s'agit de petits ARNs non codants (ncRNAs) dérivés de ces régions, dont la taille des plus petits peut varier de 20-nt à 24-nt (Lung *et al.*, 2006).

L'analyse globale des sRNAs montre que l'accumulation des sRNAs de 24-nt est moins importante dans les plantes *Osnrpd1*. Le niveau d'accumulation de ces sRNAs est le plus faible dans

la plante T0 23.1 puis remonte dans les plantes de génération T1. Ce résultat coïncide avec l'analyse semi-quantitative de l'expression d'*O_sNRPD1*, dont la sous-expression est transitoire. En effet, le niveau d'expression d'*O_sNRPD1* est plus faible dans la génération T0. L'activité d'*O_sNRPD1* étant diminuée, la biogenèse des siRNAs de 24-nt est affectée et la quantité des sRNAs de 24-nt est la plus faible dans cette plante. Puis, dans les générations T1, le niveau d'expression d'*O_sNRPD1* est plus fort qu'en T0 mais toujours plus faible que dans la plante WT. La stratégie RNAi entraîne une diminution des siRNAs de 24-nt, qui est accompagnée d'une hypométhylation des cytosines dans les 3 contextes, ce qui est en accord avec les études antérieures sur les différents acteurs de la voie RdDM (pour revue : Ito, 2012; Matzke et Mosher, 2014).

Le séquençage du méthylome a d'ailleurs permis de montrer chez différentes classes d'ETs, principalement en contexte CHH, une hypométhylation dans les plantes T0 23.1 et T1 23.1-15. Le niveau de méthylation des ETs est proche de celui de la plante WT pour la plante T1 23.1-23, qui est une plante pour laquelle nous avons identifié une accumulation plus importante des sRNAs de 24-nt. Ce résultat laisse penser que l'impact de la mutation *Osnrpd1* est moins important dans cette plante. Quoiqu'il en soit, ce défaut de méthylation des ETs est observé principalement en contexte CHH dont la méthylation *de novo* est assurée par le RdDM (Law et Jacobsen, 2010).

Cependant, même si une hypométhylation de différentes classes d'ETs a été observée, dont les rétrotransposons, des néo-insertions pour seulement 8 familles de LTR-RTs ont été détectées.. L'hypométhylation d'ETs n'est pas toujours suffisante pour réactiver les ETs, ce qui est d'ailleurs le cas dans le mutant *nrpd1* d'*A. thaliana* (Ito *et al.*, 2011).

La mobilité de LTR-RTs a été identifiée pour un petit nombre de familles dans le mutant *Osnrpd1*. Notamment, l'activation transpositionnelle des éléments *Tos17* et *Lullaby* dans les plantes *Osnrpd1* a eu lieu, à une exception près, au moment de la culture cellulaire, ce qui est en accord avec des études antérieures réalisées dans des cals et plantes régénérées à partir de cals (Hirochika, 2001; Picault *et al.*, 2009). Nos travaux ont pu mettre en évidence un évènement de transposition de l'élément *Tos17* qui a eu lieu *in planta* lors de la régénération de la plante T0, ce qui a déjà été observé par Hirochika et associés (1996). Néanmoins, douze copies de l'élément *Tos17* ont été détectées dans les plantes *Osnrpd1* dont la culture cellulaire a duré 7 semaines alors que le nombre de néo-insertions de *Tos17* observé dans des cals cultivés pendant 12 semaines est de l'ordre de 2 (Hirochika *et al.*, 1996). Sachant que le nombre de copies de *Tos17* augmente avec le temps de culture, il est tentant de vouloir proposer que la mutation RNAi affectant la voie du RdDM puisse avoir favorisé la capacité de *Tos17* à transposer. Deux autres mutants T-DNA (*Arawak* et *Altg10*) générés au sein de l'équipe collaboratrice du CIRAD étaient disponibles dans notre équipe. Le nombre de copies de *Tos17* a été évalué dans ces mutants T-DNA. Ces mutants ont été générés par

culture cellulaire, il n'est donc pas étonnant d'observer plus de copies de *Tos17* que dans un contexte sauvage. En revanche, *Arawak* et *Altg10* sont respectivement des plantes régénérées de génération T3 et T4. Dans les plantes régénérées, *Tos17* n'est plus actif et il y a une ségrégation des insertions à l'état hétérozygote lors de la méiose, ce que nous avons observé dans les plantes *Osnrpd1* et qui laisse à penser que pour les mutants *Arawak* et *Altg10*, le nombre d'insertions était plus grand. Nous ne disposons malheureusement pas des plantes T0 ayant permis d'obtenir ces plantes pour vérifier notre hypothèse. Néanmoins, dans une étude menée par l'équipe collaboratrice ayant générée les mutants *Osnrpd1*, le nombre moyen de néo-insertions de *Tos17* détectées dans 384 lignées T-DNA issues de cals cultivés pendant 7 semaines est de 3,37 (Piffanelli *et al.*, 2007). Mais en l'absence de matériel contrôle cultivé dans les mêmes conditions pour un même temps de culture, nous ne pouvons pas répondre à cette hypothèse.

D'autre part, si la levée du contrôle épigénétique, due à la mutation RNAi du gène *OsNRPD1* était à l'origine des néo-insertions de *Tos17*, alors nous devrions retrouver d'autres néo-insertions de cet élément dans les générations suivantes. Ce n'est pas le cas, mais, nous avons observé que le niveau de siRNAs de 24-nt remonte dans la génération T1. De plus, l'analyse de la distribution des sRNAs sur les ETs activés montre que les sRNAs de 24-nt sont plus abondants dans les plantes *Osnrpd1*, notamment sur *Tos17*. Pourtant, la mutation dirigée contre *OsNRPD1* induit une diminution des siRNAs de 24-nt. Ce résultat suggère que l'activité récente de *Tos17* est réprimée. L'élément *Tos17* ayant une préférence pour les régions à proximité des gènes (Yamazaki *et al.*, 2001; Miyao *et al.*, 2003), il est potentiellement hautement mutagène. Une étude préliminaire des régions où les néo-insertions de *Tos17* ont été détectées montre que 50% de ces insertions sont dans des gènes. Cependant, aucun d'entre eux n'est différentiellement exprimé dans la plante T1 23.1-15 pour laquelle nous avons les données de RNA-seq. De plus, pour l'élément *Tos17*, l'accumulation des sRNAs de 21-nt est plus importante dans les plantes de génération T1 par rapport à la plante WT ou T0. Il est possible qu'une autre voie vienne en renfort de la voie RdDM pour contrôler la répression de *Tos17* par l'intermédiaire des sRNAs de 21-nt comme cela l'a été suggéré pour des ETs d'*A. thaliana* (Nuthikattu *et al.*, 2013).

Les autres ETs activés dans les plantes T1 sont également ciblés par les sRNAs de 24-nt, et l'analyse DOC suggère qu'ils ne transposent plus ensuite.

Ces résultats suggèrent que la voie du RdDM permet de maintenir le contrôle de la répression de ces ETs *via* les sRNAs de 24-nt. De plus, il est intéressant de noter que les sRNAs de 24-nt ciblent préférentiellement les LTRs qui contiennent les régions promotrices des LTR-RTs et codent pour les protéines requises pour leur transposition. Cette observation d'une accumulation de siRNAs de 24-nt dérivés des LTRs a été faite précédemment chez *A. thaliana* (Mari-ordóñez *et al.*,

2013). Les auteurs ont pu faire une corrélation entre méthylation du LTR 5' et réactivation du rétrotransposon *Évadé* (*EVD*). Lorsque les données brutes de méthylation seront disponibles, il sera intéressant de regarder l'état de méthylation des LTR-RTs réactivés dans les plantes *Osnrpd1* où une accumulation de sRNAs de 24-nt est observée au niveau des LTRs.

L'activité rétrotranspositionnelle d'éléments appartenant à 5 familles de LTR-RTs, dont la mobilité n'avait encore jamais été identifiée, a été observée dans les mutants *Osnrpd1*. De plus, parmi ces LTR-RTs, *LTR_fam106*, *LTR_fam108* et *LTR_fam124* ne sont pas réactivés par la culture cellulaire puisque les mutants *Altg10* et *Arawak* ne présentent pas de différences significatives du nombre de copies d'éléments pour chacune de ces familles par rapport aux plantes WT et 23.1. La réactivation de ces ETs est donc la conséquence du relâchement de la répression induite par la voie RdDM. Cependant, nous ne pouvons pas savoir où sont insérés les ETs dans les plantes T1 puisque l'identification de leur mobilité a été faite par l'étude du DOC. Les lectures n'ont pas été alignées sur le génome mais sur la séquence d'un représentant de chaque famille d'ETs pour faire une estimation du nombre de copies de cet élément dans chaque génome. De ce fait, il n'est pas possible de savoir si toutes ces insertions ont eu lieu lors de la régénération de la plante T0 ou dans les plantes T1. Sachant que le nombre de familles de LTR-RTs expertisées par l'équipe chez le riz est de 295, il apparaît que très peu d'éléments sont mobilisés suite à la perturbation du contrôle épigénétique induite par la mutation *Osnrpd1*. Il est assez troublant que les ETs restent immobiles dans des mutants où la biogenèse de siRNAs est affectée alors que leur implication dans la répression des ETs, par l'intermédiaire de la méthylation, est reconnue (Lister *et al.*, 2008). Cependant, même si le nombre d'ETs réactivés apparaît assez faible, dans le mutant *nprpd1* d'*Arabidopsis*, la transposition d'ETs n'a pas été observée en condition de culture normale mais suite à l'application d'un stress thermique (Ito *et al.*, 2011, Matsunaga *et al.*, 2015). De plus, dans le mutant épigénétique du riz *Oscmt3*, seules 8 familles d'ETs sont mobilisées, dont *Tos17* et *Lullaby* (Cheng *et al.*, 2015).

Nous avons également détecté des néo-insertions de l'élément *Karma* dans les régénérants des plantes *Osnrpd1*, ce qui est en accord avec les travaux publiés par Komatsu et associés (2003). De plus, ces auteurs ont montrés que *Karma* est hypométhylé dans les cals et les régénérants. Nous ne pouvons actuellement pas regarder l'état de la méthylation au niveau de cet élément car nous n'avons pas accès aux données brutes du séquençage bisulfite. Néanmoins, nous avons montré que l'accumulation des sRNAs de 24-nt était beaucoup moins importante dans les plantes *Osnrpd1* par rapport au WT, ce qui laisse penser que la méthylation médiée par les siRNAs de 24-nt doit être fortement diminuée. Cependant, la transposition de cet élément n'est observée qu'à partir de la génération T2 des régénérants. Le mécanisme pouvant expliquer cette transposition dans les

régénérants n'est pas connu, cependant Komatsu et associés proposent que la régulation de l'élément *Karma* permet de limiter sa transposition à certaines étapes du développement.

Le fait qu'un élément comme *Karma* puisse être actif transpositionnellement après un laps de temps de différentes générations soulève la question de la régulation d'autres ETs dans les plantes mutées. De plus, après 5 générations, certaines plantes *Osnrpd1* présentent un phénotype touchant le développement des talles qui est confirmé dans la génération T6 de ces plantes. Ce phénotype pourrait être la conséquence d'une insertion d'un ET à proximité d'un gène contrôlant le développement.

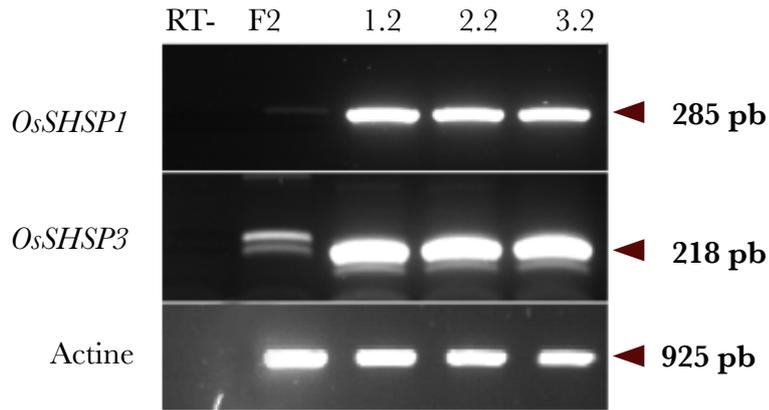
PARTIE 3. Activité rétrotranspositionnelle chez le riz *Oryza sativa* en condition de stress

L'origine des bursts de transposition qui ont eu lieu dans la nature au cours de l'évolution pour expliquer l'invasion des ETs dans les génomes n'a pas encore été identifiée. On sait néanmoins que l'activation transpositionnelle des ETs est observée dans des plantes soumises à différents types de stress. Dans cette partie nous nous sommes intéressé au stress thermique qui s'inscrit dans le contexte actuel de changement climatique. Les futurs changements globaux prédits rendent les recherches sur le riz, comme sur toutes les plantes d'intérêt agronomique, très importantes quant à l'adaptation des plantes. En effet, ces changements vont entraîner différents stress auxquels les plantes seront soumises, comme l'augmentation de la température, des inondations plus marquées causées par des pluies plus puissantes et plus fréquentes dans les régions tropicales, des périodes de sécheresse plus fréquentes dans les régions arides, ou encore des conditions de salinité élevées causées par l'évaporation de l'eau des sols. Il est donc important de comprendre comment le riz va s'adapter face à ces changements, dont l'accélération présente un danger pour la sécurité alimentaire à l'échelle planétaire.

L'étude de l'impact génomique du stress en conditions naturelles peut nous apporter des pistes quant à l'origine des bursts transpositionnels mais également sur l'adaptabilité génomique du riz face à un stress.

Nous supposons que des néo-insertions de rétrotransposons se produisent *in planta* lors de l'application d'un stress thermique important. De plus, l'étude d'un mutant d'*A. thaliana* affecté dans la voie du RdDM, *nripd1*, et soumis à un stress thermique a permis d'observer que l'accumulation de ces stress entraînait la réactivation d'un RTN, *Osen* (Ito *et al.*, 2011). Nous avons donc entrepris de générer différentes populations de riz sauvage et en contexte épigénétique relâché, *Osnripd1*, et de soumettre ces populations à un stress thermique. Nous pourrions ainsi étudier la dynamique transpositionnelle chez le riz en réponse à un stress thermique.

Fig. 3.1 Profil d'expression de deux gènes en réponse à la chaleur.



Des plantules de riz ont été soumises à un stress thermique (42°C, 5h/jr pendant 5 jrs). Le gène de l'actine du riz a été utilisé comme contrôle interne. F2 désignent la plante Nipponbare contrôle (F2) et 1.2, 2.2 et 3.2 sont des plantes stressées thermiquement. Les gènes *OsSHSP1* et *OsSHSP2*, codant pour des small heat shock protein, sont des gènes fortement induits en conditions de stress (Ham *et al.*, 2013).

3.1. Matériel végétal

Le stress joue un rôle de déclencheur pour la réactivation transpositionnelle des ETs. Nous souhaitons observer les évènements de transposition qui pourraient avoir lieu en réponse à un stress thermique. Or, il n'est pas possible de prédire dans quelles cellules ces évènements se produiront. La croissance du riz, comme pour tous les végétaux, commence au niveau d'un méristème dont les cellules indifférenciées vont se diviser et permettre la croissance de la plante. Chez le riz, à partir du méristème, des cellules vont se diviser pour former différentes talles qui vont émerger après la formation de la 4^{ème} feuille. Un stress a donc été appliqué au stade 4 feuilles afin de maximiser les chances qu'un nouvel événement de transposition se retrouve dans toutes les cellules de la plante. Ce stress est réalisé à 45 °C, soit une augmentation d'au moins 17 °C, par rapport aux conditions optimales pour la culture du riz, et a été appliqué pendant 5h (de 10h à 15h) et pendant 5 jours consécutifs. Le même protocole de génération des plantes que celui mis en place (cf partie 2.1.2) a été suivi, et le stress a été appliqué à partir de la génération T1. Deux plantes sauvages stressées thermiquement de génération T3 ainsi que 3 plantes T2 *Osnrpd1* ont été envoyées à séquencer. Les plantes stressées *Osnrpd1* HST-23.1-15Ta2.2, HST-23.1-23Ta2.2 et HST-23.1-23Ta3.3 sont les plantes « soeurs » des plantes 23.1-15Ta2.2, 23.1-23Ta2.2 et 23.1-23Ta3.3 étudiées dans la partie 2.

Efficacité du stress thermique :

L'observation des plantes stressées thermiquement n'a pas permis de mettre en évidence de phénotypes aberrants. Afin d'évaluer l'efficacité du stress thermique exercé sur les plantes, des RT-PCR semi-quantitatives ont été réalisées à partir d'ARNs extraits de feuilles après 2 jours de stress thermique. L'expression de 2 gènes (*OsSHSP1* et *OsSHSP2*) codant pour des protéines HSPs (Heat Shock Proteins) a été testée. Ces gènes sont sur-exprimés en conditions de stress et notamment en conditions de stress thermique (Ham *et al.*, 2013). Les résultats présentés dans la Fig. 3.1 montrent que ces 2 gènes *OsSHSP1* et *OsSHSP2* sont fortement induits dans les feuilles des plantes 1.2, 1.2 et 1.3 stressées thermiquement.

3.2. Méthodes de détection de néo-insertions dans les plantes stressées thermiquement

Les différentes méthodes de détection de mouvements d'ETs utilisées dans la partie 2 ont été utilisées pour la détection des ETs dans les génomes des plantes soumises au stress thermique. Ainsi les régions retrouvées dans toutes les plantes Nipponbare qui diffèrent de la séquence du génome de référence ont été retrouvées dans les plantes stressées thermiquement.

Tab. 3.1 Tableau des insertions détectées pour les plantes stressées thermiquement

Plante	Breakpoint	Localisation	Élément
HST-233	1553	chr03: 13 606 060	MULE
HST-332	1803	chr03: 31 820 943	MITE inséré dans <i>Bajie</i>

Fig. 3.2 Dot-plot des séquences flanquant le point d'insertion de l'élément 1553, de l'élément MULE inséré et des lectures pairées et des lectures coupées obtenues par séquençage alignées sur elles-mêmes

Des agrandissements des encadrés noirs montrent les paires discordantes. Dans les encadrés rouges sont montrés les endroits des lectures coupées.

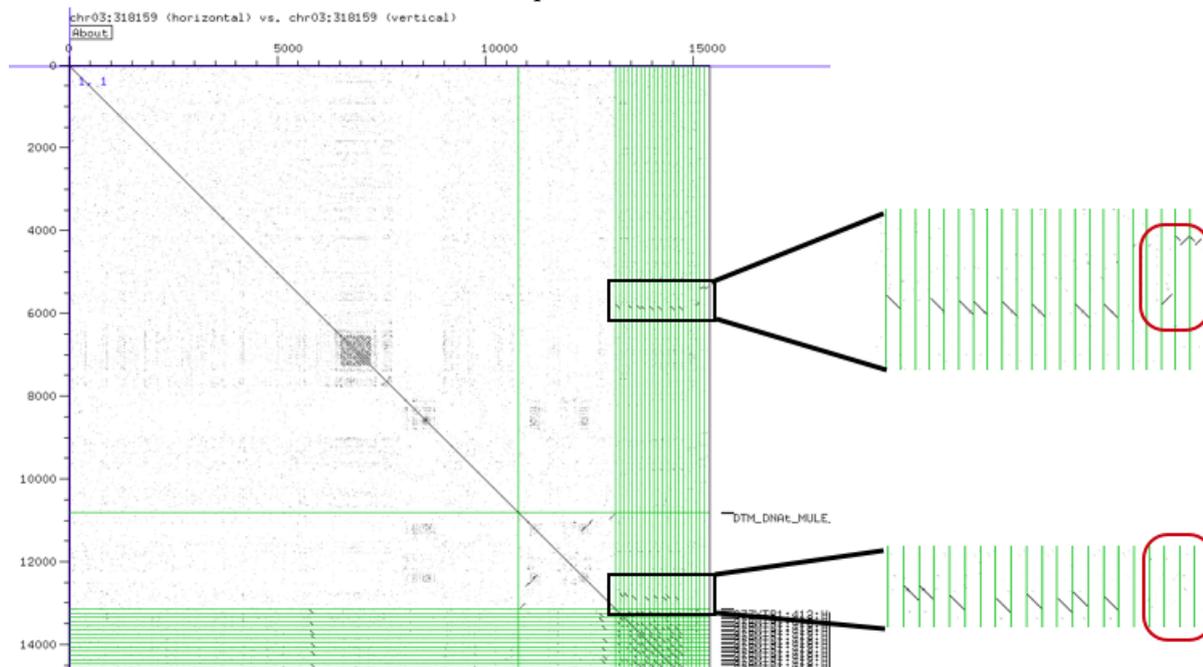
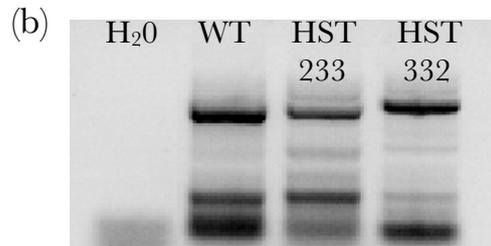
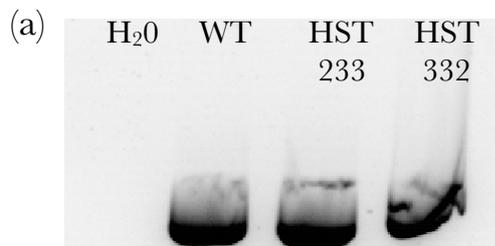


Fig. 3.3 Profil de migration des produits PCR des insertions 1553 et 1803 sur gel d'agarose

(a) test de l'insertion 1553 site plein, (b) test de l'insertion 1803 site plein. H₂O, eau; WT, plante Nipponbare contrôle; HST, plante stressées thermiquement.



3.3. Résultats

3.3.1. Activité transpositionnelle détectée par le pipeline d'analyse

Les néo-insertions détectées dans les plantes sauvages stressées thermiquement par la méthode de PEM couplée à l'identification de lectures coupées sont présentées dans le Tableau 3.1. Dans un premier temps, une validation manuelle de ces breakpoints a été réalisée par analyse de dot-plot. Un exemple de dot-plot est présenté dans la Figure 3.2.

Nous observons que les paires discordantes s'alignent bien de façon à ce qu'une lecture de la paire s'aligne au niveau du point d'insertion identifié et l'autre lecture s'aligne au niveau de la séquence de l'élément identifié comme s'étant inséré. En revanche, nous n'observons pas de vraies lectures coupées, où une partie de la lecture devrait s'aligner au niveau du point d'insertion et l'autre partie de la lecture dans l'élément. Les mêmes caractéristiques de dot-plot ont été observées pour l'élément 1803.

Une validation par PCR a été réalisée pour ces 2 éléments en dessinant des amorces dans des paires obtenues par le séquençage en paired-end (Annexe 5). L'amplification de ces fragments est présentée dans la Fig. 3.3 (a) pour l'élément 1553 et (b) pour l'élément 1803. On remarque que les profils de migration des produits d'amplification des sites de néo-insertions pour les 3 plantes séquencées (WT et HSTs) sont similaires.

Malgré l'identification de paires discordantes, l'amplification PCR de ces éléments ne valide pas l'insertion d'élément dans les plantes stressées thermiquement.

3.3.2. Activité transpositionnelle détectée par la méthode AAP

Les ETs *Tos17* et *Lullaby* ne sont pas activés rétrotranspositionnellement dans les plantes sauvages. En revanche, nous avons retrouvé pour les plantes *Osmrpd1*, les mêmes insertions de ces ETs que celles retrouvées dans leur plante soeur respectives (données non présentées).

3.3.3. Activité transpositionnelle détectée par la méthode du DOC

Dans les plantes stressées thermiquement, la méthode du DOC a permis d'identifier un LTR-RT, *LTR_fam86*, qui présente un nombre de copies significativement plus important dans les plantes HST-23.1-15Ta2.2 et HST-23.1-23Ta3.3 par rapport à la plante WT et 23.1. Dans les plantes non stressées thermiquement 23.1-15Ta2.2 et 23.1-23Ta3.3 le nombre de copies pour cet

Tab. 3.2 Néo-insertions détectées par la méthode du DOC

Élément	GM	23.1	23.1-15Ta2.2	HST 23.1-15Ta2.2	23.1-23Ta2.2	HST 23.1-23Ta2.2	23.1-23Ta3.3	HST 23.1-23Ta3.3
<i>LTR_fam86</i>	51 +/-9	50 +/-14	77 +/-13	82 +/-16	77 +/-15	71 +/-14	72 +/-14	86 +/-20

élément n'est pas significativement différent ni par rapport aux plantes WT et 23.1 ni par rapport aux plantes stressées (Tab. 3.2).

Ces résultats suggèrent que la mutation RNAi a également permis la réactivation transpositionnelle de cet élément dans les plantes *Osnrpd1* et le stress thermique permettrait une mobilisation plus importante de cette famille d'ET.

3.4. Discussion

L'analyse des plantes semble montrer que le stress thermique appliqué favorise la réactivation rétrotranspositionnelle d'une seule famille d'élément *LTR_fam86*.

Cependant, il n'est pas possible d'affirmer qu'aucune autre activité transpositionnelle n'a pas eu lieu dans les plantes en réponse au stress. En effet, la méthode du DOC n'a été utilisée que pour analyser la mobilisation de LTR-RTs. En revanche, le pipeline développé par l'équipe peut détecter les autres ETs disponibles dans la base de données du riz mais des améliorations doivent encore être apportées à cette méthode qui entraîne la détection de faux positifs et négatifs.

Le stress appliqué sur les plantes n'était peut-être pas assez fort pour entraîner une forte activité rétrotranspositionnelle des plantes. En effet, même si l'impact de ce stress sur la sur-expression de 2 gènes codant pour des HSPs a été observé, nous n'avons pas noté de changements drastiques de l'aspect physiologique des plantes, ni de réactivation transpositionnelle. Il n'est pas possible pour le moment de tester un stress à une température plus élevée car les enceintes dans lesquelles le matériel a été cultivé ne sont pas destinées à des températures plus hautes que 45°C.

En 2015, Li et ses collaborateurs ont soumis des riz de la variété Nipponbare à une température de 45°C pendant 45h consécutives et au stade 2 feuilles lorsque les plantes ne sont que des plantules. Les résultats de leur étude montre que ces plantes stressées présentent un aspect morphologique, drastiquement différent de la plante WT, caractérisé par des feuilles sclérosées. J'ai effectué un test préliminaire en soumettant des plantes WT de manière continue à une température de 45°C. L'objectif était d'atteindre un stade subléthal pour la plante, mais même après une dizaine de jours, les plantes soumises à un tel stress résistent, présentent des scléroses qu'aux extrémités des feuilles et semblent s'être adaptées à la différence de température. L'application du stress au stade 2 feuilles est envisageable mais il s'agit d'un stade précoce de développement qui pourrait limiter le taux de survivants fertiles pour générer les différentes générations nécessaires à l'étude. Les études de stress thermique à ces stades de développement étant très limitées, nous ne pouvons pas prévoir les résultats d'un tel stress.

La rétrotransposition d'ETs n'a été établie que pour une seule famille dans nos plantes en réponse au stress thermique appliquée, cependant, une étude très récente montre qu'un stress thermique sévère (39°C) appliqué au stade graine induit une forte induction de l'expression d'ETs (Chen *et al.*, 2016).

Les ETs les plus représentés dans leur analyse sont les éléments Pong, *Ac/Ds* et Mariner appartenant à la classe des transposons. Il serait sans doute intéressant de tester notre stress au stade graines, mais même si l'activation transcriptionnelle d'ETs a souvent été décrite (Finatto *et al.*, 2015; Grandbastien, 2015) elle ne s'accompagne pas toujours d'une activation transpositionnelle des ETs, ces derniers étant réprimés par des contrôles stricts pour prévenir les réarrangements génomiques qu'ils pourraient induire.

Enfin, la réactivation transpositionnelle des ETs a été observée en réponse à des stress divers (Wessler, 1996; Grandbastien, 1998; Capy *et al.*, 2000). Or, nous n'avons étudié que l'impact du stress thermique. Il est possible que ce stress ne soit pas le stress induisant la meilleure réponse transpositionnelle. Tester l'impact du stress salin avait été envisagé au début de cette thèse et le protocole expérimental ainsi que les paramètres de culture des plantes ont été mis au point. Le matériel végétal est à développer et nécessiterait un an de culture pour produire le pédigrée.

PARTIE 4. Étude de la dynamique des génomes par la nouvelle technologie Nanopore

L'utilisation du PEM présente des avantages majeurs pour la détection des variations structurales. D'une part, parce que cette pratique permet d'obtenir plus de lectures que le séquençage en *single-end* pour soutenir la présence d'une variation structurale (SV) et, d'autre part, parce qu'elle permet de couvrir plus largement les régions où des insertions ont pu se produire et ainsi sont plus informatives quant à la séquence insérée. Cette propriété est particulièrement intéressante pour la détection d'ETs et la discrimination de nouvelles insertions.

Dans les parties 2 et 3, la combinaison du PEM et les différentes approches utilisées pour l'étude des néo-insertions nous a permis de dresser un inventaire des LTR-RTs réactivés dans les plantes de l'étude. Néanmoins, chacune de ces méthodes présentent des avantages et des inconvénients. Le pipeline développé au sein de l'équipe nous a permis de mettre en évidence un polymorphisme entre le génome des plantes de l'étude (sauvage et mutantes) et le génome de la plante de référence. Ces différences génomiques entre la plante Nipponbare de référence et sauvage (contrôle de notre étude) pourraient être la conséquence d'un mauvais ou d'un manque d'assemblage de certaines régions dans la séquence de référence. En effet, même si la séquence du génome du riz est de très bonne qualité, l'assemblage de certaines régions complexes du génome est souvent difficile. Il est également envisageable que malgré l'appartenance de ces plantes à la même sous-espèce, *O. sativa* spp. *Japonica*, et au même cultivar (Nipponbare), des mutations ponctuelles et réarrangements de génome individuel aient pu avoir lieu.

Cependant, le nombre de faux positifs est encore important à l'issue de notre analyse. Ces résultats pointent les limites de la détection des mouvements d'ETs avec ces techniques. Même si les améliorations apportées aux méthodes de détection des SVs l'ont été de manière concomitante avec les progrès des techniques de séquençage, la détection de nombreux faux positifs et négatifs limitent encore l'efficacité de ces méthodes.

De plus, la propriété des ETs à être répétés dans les génomes empêchent leur détection dans les régions complexes du génome.

Depuis quelque années, des nouvelles technologies, qualifiées de 3^{ème} génération, apparaissent sur le marché du séquençage. Ces techniques se basent sur la lecture directe d'une seule molécule d'ADN. Cette approche permet de générer des séquences d'ADN de plusieurs kb mais avec un taux d'erreur élevé (15% environ). Nous pensons cependant que, malgré ce taux d'erreurs, la longueur de la séquence obtenue sera suffisante pour discriminer de nouveaux évènements de transposition.

Afin d'évaluer l'intérêt d'utiliser cette technique pour la détection des variations structurales, j'ai participé à la détection, dans les données de séquençage Nanopore, des insertions du RTN *Évadé* dans une lignée de plante d'*A. thaliana* dans laquelle sa mobilisation a été identifiée (Reinders *et al.*, 2009; Mirouze *et al.*, 2009). Ce travail est présenté, ci-après, sous forme d'article récemment soumis dans Mobile DNA.

Transposable element insertion caught in a single-Oxford-Nanopore-read

Emilie Debladis¹, Christel Llauro-Berger², Marie-Christine Carpentier², Marie Mirouze³, Olivier Panaud^{1,*}

Author affiliations

¹ *Université Perpignan Via Domitia, Laboratoire Génome et Développement des Plantes, UMR5096, F-66860, Perpignan, France*

² *CNRS, Laboratoire Génome et Développement des Plantes, UMR5096, F-66860, Perpignan, France*

³ *Institut de Recherche pour le Développement, UMR232 DLADE Diversité Adaptation et Développement des Plantes, Université Montpellier 2, 911 av. Agropolis F-34394 Montpellier, France*

*For correspondance : panaud@univ-perp.fr

Background :

Transposable elements (TEs) contribute to both structural and functional dynamics of most eukaryotic genomes. Because of their propensity to densely populate plant and animal genomes, the precise estimation of the impact of transposition on genomic diversity has been considered as one of the main challenges of today's genomics. The recent development of NGS (next generation sequencing) technologies has opened new perspectives in population genomics, by providing new methods for high throughput detection of Transposable Elements-associated Structural Variants (TEASV). However, these have relied on the Illumina platform that generates short reads (up to 350 nucleotides). This limitation in size of sequence reads can cause high false discovery rate (FDR) and therefore limit the power of detection of TEASVs, especially in the case of large, complex genomes. The newest sequencing technologies, such as Oxford Nanopore Technologies (ONT) can generate long reads (several kilobases in routine) thus representing a promising tool for TEASV detection in plant and animals.

Results :

Here we present the results of a pilot experiment for TEASV detection on the model plant species *Arabidopsis thaliana* and show that ONT can be used efficiently to detect TE movements. We generated a ~2.5x genome coverage of a *met1*-derived epigenetic recombinant inbred line (epiRIL) using a MinION™ device with R7 chemistry. We were able to detect 9 new copies of the LTR-retrotransposon *Evadé* among the 13 estimated new copies that this line harbors.

Conclusions :

Even if the Nanopore technology generates data with an average of 12% sequencing error rate, 70% of estimated *Evadé* neo-insertions were detected using a MinION device with only a 2.5x genome coverage. This new method of single-molecule sequencing is therefore a good alternative to Illumina for the detection of transposition in plants and animals.

Introduction

Transposable elements (TEs) are the most widespread class of repetitive sequences in eukaryotic genomes. Because of their nature and ability to increase in copy number, TEs play a major role in the structure and evolution of the genome in most eukaryotic lineages (Britten and Kohne, 1968; Feschotte and Pritham, 2007). Transposable Elements-associated structural variants (TEASV) are common in natural and domesticated populations thus contributing to a large extent

to genomic diversity, while they also can influence gene expression (Lisch, 2013; Zhao *et al.*, 2016) and lead to phenotypic variations (Kobayashi *et al.*, 2004; Wei and Cao, 2016). Therefore, the estimation of the contribution of TEASVs to biological diversity in plant and animals has become one of the objectives of genomics today.

In the past decade, the emergence of Next Generation Sequencing (NGS) technologies based on Illumina platform has opened new perspectives for the development of new methods to conduct systematic genome-wide surveys of TEASV. Several methods for the detection of TEASVs have been thus proposed (Makalowski *et al.*, 2012; Ewing, 2015).

TEASV detection can be achieved through three distinct methods that can be used alone or in combination : the paired-end mapping (PEM), the split-read mapping (SRM) and the depth of coverage (DOC). PEM is based on the mapping of both ends of the amplicons generated for Illumina sequencing on a reference genome. Structural variants are detected when both ends of the amplicon map at different locations in the genome (at a distance which is significantly larger than the average size of the inserts of the Illumina library). The PEM was developed in order to best exploit the sequencing data generated by an Illumina platform, i.e. short reads that are paired from the same amplicon. In plants, the PEM method allowed the detection of active TEs in rice (Sabot *et al.*, 2011; Cheng *et al.*, 2015) and *Arabidopsis thaliana*. In this latter case, the authors identified an active LTR-retrotransposon *Évadé* (*EVD*) in an hypomethylated mutant plant (Mirouze *et al.* 2009). However, one of the major limitations of the method is that the mapping of short reads in complex genomes is not always reliable and can lead to high FDR (Handsaker *et al.*, 2011; Mills *et al.*, 2011). This is mostly due to their presence in high copy number in genomes, thus challenging their detection in repeated regions. The SRM method is based on the identification of sequencing reads that span the junction between a new TE insertion and its native insertion site. It is conceptually one of the most reliable methods for TEASV detection, but the short read length produced by the Illumina technology has so far impeded its reliable use for exhaustive TEASV detection. The recent improvement of the technology in this matter has however led some authors to successfully use the SRM method (Medvedev *et al.*, 2009; Ye *et al.*, 2009; Quadrana *et al.*, 2016).

However, like in the case of PEM, SRM does not allow any reliable detection of TE insertions in repeated sequences because of short read length. The third method, DOC, consists in estimating the sequencing coverage of known TE families to detect an increase in their copy number. It is therefore restricted to TEs that transpose via a copy and paste process, such as LTR retrotransposons, LINEs or SINEs. In addition, the DOC method is not very precise (the exact number of newly transposed copies cant not be ascertained), but it is very robust and can detect

insertions in repeated regions, unlike the two previous methods described above (Yoon *et al.* 2009; Glusman *et al.*, 2015). The major limitation of the DOC is that it does not provide any information on the position of new insertions and can therefore not be used for further functional characterization of TEASVs.

The examples cited above show that the use of the Illumina technology has made possible the detection of TEASVs in eukaryotic genomes at some extent with however some limitations inherent to the short read length. The recent development of third generation technologies that allow the sequencing of a single molecule could open new perspectives for TEASV detection : the two available platforms, *i.e.* Pacific Biosciences and Nanopore Technologies indeed allow to generate long reads (several kb) that could improve the reliability of TEASVs detection by providing enough sequence information to unambiguously map new TE insertion sites.

The Nanopore technology was conceptualized three decades ago (Deamer *et al.*, 2016) but lead only two years ago to the commercial release of the smallest hand-held sequencers, called MinION™. This new Nanopore sequencer is based on DNA characteristics where the use of a positive electrical potential lead to an ionic current. As DNA molecules are negatively charged, they can translocate through a protein nanopore leading to changes in current that are used to determine the nucleotide base sequence (Feng *et al.*, 2015; Laver *et al.*, 2015). So far, both Pacific Biosciences and Nanopore technologies have suffered from high sequencing error rate, which impedes the direct use of the sequencing data for genome assemblies. However, in the case of TEASV detection, the trade-off between sequencing length and high error rate may be in favor of the former since we anticipate that a read of several kilobases could be unambiguously mapped onto a reference genome even if sequencing is achieved with an error rate of 10-15 %. In this report, we tested this hypothesis by sequencing the genome of a partially hypomethylated line of the plant species *A. thaliana* (Reinders *et al.*, 2009; Mirouze *et al.*, 2009) in which 13 new copies of the LTR retrotransposon *EVD* were identified. We indeed show that the vast majority of the reads generated using the MinION device can be mapped onto the *Arabidopsis thaliana* reference genome sequence and that in addition, new insertions of *EVD* were easily and unambiguously mapped.

Results

We first conducted a quantitative polymerase chain reaction (qPCR) analysis of the epiRIL12 from an hypomethylated *met1* mutant seedlings in order to quantify the neo-insertions of the *EVD* retrotransposon (a member of *ATCOPIA93* family shown to be active in such material).

The relative *EVD* copy number of genomic DNA derived from *A. thaliana* Colombia wild type plant (WT) and the epiRIL12 was quantified with an actin control and then calculation of the epiRIL12-to-WT ratio was estimated. Since the WT harbors two native copies of *EVD*, DNA accumulation was normalized at 2 for the WT. A significant increase of 8 fold was found in the epiRIL12 in comparison to the wild type corresponding from 13 to 26 estimated new *EVD* copies in this line (taking heterozygosity into account) (Fig. 1).

Overview of the sequencing data

3 flow cells were used to sequence genomic DNA library with two MinION devices : one with the first available version of the MinION and the other two with the latest MkI release. The first MinION run (FLO-MAP003) generated template sequence reads for 49,061 different DNA fragments (232 Mb). 98% produced Template reads (i.e. generated by the forward strand sequencing) and 40% produced Complement reads (i.e. generated by the reverse strand sequencing). 35% gave Two Direction reads classified as « All 2D » when template and complement reads can be merged into a consensus sequence.

The second and third runs (FLO-MAP103) produced 9,796 and 4,644 DNA fragments (66 and 38 Mb) respectively; 100% were template reads, and 6,437 and 3,008 complement reads, respectively. Of these, a total of 5,490 and 2,357 were converted into 2D reads and classified as « full 2D ».

Because sequenced reads obtained from the 3 MinION sequencing runs corresponded to the same sample of epiRIL12, all sequence data were concatenated into a final multifasta file. The read length statistics are illustrated in a boxplot which can be found in supplementary files, the median of a read length is 4,600 kb (Additional file 1). Then, Nanopore reads were aligned on the *Arabidopsis* reference genome (Col-0) with allowed gaps, and shared almost 72% of identity with the Col-0 genome (Additional file 2). Altogether, 336 Mb of sequence reads from the epiRIL12 were obtained, corresponding to a genome coverage of 2.5x.

Detection of *EVD* insertion sites by MinION reads analyses

Sixteen nanopore reads ranging from 4.2 to 8.8 kb showed homology with the *EVD* retrotransposon. Four of these reads represented only *EVD* sequence (i.e. 100 % of the reads corresponded to the element). Eleven reads showed partial homology to *EVD*, with one part of the sequence corresponding to the element and the other to the *Arabidopsis* genome, which corresponded

to 9 distinct insertion sites. One among these 9 insertion sites corresponded to the native copy of *EVD* present on the chromosome 5. Finally, one 7kb read spanned the complete *EVD* sequence flanked by both insertion site sequences (Fig. 2). To our knowledge, this is the first example of sequencing of a new LTR-retrotransposon insertion in a single read.

In conclusion, in the analyzed epi12 sample we detected 9 new *EVD* insertion sites and one of the *EVD* native copies (Fig. 3).

Validation of *EVD* neo-insertions sites by Retrotransposon-based insertion polymorphism (RBIP)

The wet-lab validation of *EVD* neoinsertion sites detected above was performed using PCR assay. All the primers set showed single clear amplified band of the expected sizes (Fig. 4A to D). The sizes of the PCR products for full sites were estimated to vary from 1636 to 1961 bp and approximately 900 bp for empty sites. Three neo-insertions of *EVD*, detected by the MinION sequencer, are presented in Fig. 4. Two of *EVD* neoinsertions are heterozygous (neo4 and neo6), and one insertion (neo7) is homozygous.

Conclusion/Discussion

In order to demonstrate the utility of Nanopore technology for TEASV detection, we sequenced an epiRIL from *A. thaliana*. We used 3 runs of MinION that in our hands flow cells produced yields up from 38 to 232 Mb that reached a total of 336 Mb. This difference in performance are mostly related to variations in flow cells quality. The median of sequence reads mapped on the arabidopsis genome was 4600 kb (suppl. data) and even if the sequencing error rate was evaluated at 12%, we were able to detect 70 % of estimated *EVD* insertions with only 2.5 fold depth of coverage. As we anticipated, long reads therefore allowed to circumvent the high FDR caused by the short Illumina reads. Reliable mapping of the reads was achieved using conventional BLAST searches using the pseudomolecules of the reference genome of *A. thaliana*, regardless of the significant sequencing error rate (i.e. 12 % in our three runs). This work was a proof of concept for the use of Nanopore technology for TEASV detection. In this regard, we only focussed on the detection of *EVD* transposition. This was achieved very simply by blasting the sequence of this retrotransposon on the nanopore reads and then further characterizing these sequences to identify the insertion sites of the element. The real power of this new technology could only be demonstrated if it allows to identify exhaustively all transposition events in an individual without

prior knowledge of active TE families. In the case of the *Arabidopsis* epiRIL that we tested, we searched for all reads showing partial homology with the WT and only found those corresponding to *EVD*, suggesting that only this LTR retrotransposon is transpositionally active in this material. Further analyses of other lines or different species may allow to further test the nanopore technology for the detection of TEASVs in larger, more complex genomes.

The sequencing yields that we obtained in our three runs are very low in comparison to that announced by the manufacturer of the MinION device (i.e. 335 Mbp using three flow cells compared to 400Mbp per flow cell). The latest flow cells recently released have supposedly been improved in terms of reliability of sequence yields. These should be further tested. At present, the cost of Nanopore sequencing is still much higher than that of other technologies. If the yield of the latest released flow cells is what the manufacturer announces, then the cost per Mbp is ~2.5€ (compared to ~5c for Illumina platforms). However, the long read length greatly improves the reliability of TEASVs detection, especially in high complexity regions, which makes the latest sequencing technologies such as Nanopore the tools of choice to study transposition at the whole genome scale.

Material and methods

Plant material

Arabidopsis thaliana WT ecotype Columbia-0 and epiRIL12 plant materials previously described in Reinders *et al.* (2009), were used in this study. Plants were grown in soil under a 16h/8h (light/dark) cycle after 2 days at 4°C for stratification.

DNA analysis

Genomic DNA from epiRIL12 was extracted from seedlings ground to a fine powder in liquid N₂. The powder was resuspended in 10ml of CTAB2X extraction buffer (2% CTAB, 100 mM Tris-HCl pH 8, 20 mM EDTA pH 8, 1.4 M NaCl, 5% N-lauryol sarcosine de sodium, 0.2% 2-mercaptoethanol) and incubated for 60 min at 65 °C. Then, an equal volume of chloroform was added and the emulsion was maintained during 10 min before centrifugation at 4 500 rpm for 10 min at room temperature. The nucleic acids were precipitated with isopropanol (0.7 v/v) at -80 °C for 15 min and centrifuged at 4 °C at 4 500 rpm for 45 min. Nucleic acids were further washed with 75% ethanol and centrifuged at 4 °C at 4 500 rpm for 10 min. Finally, the pellet was air dried and DNA was resuspended in 300 µl TE and was treated with 10 µg of RNase A (Qiagen, Hilden, Germany) for 30 min at 65 °C before further analysis.

DNA was quantified using the Quant-iT dsDNA High sensitivity Assay (HS) Kit (Life Technologies) according to the manufacturer's instructions.

For qPCR analysis of *EVD* DNA copies, the Takyon No RoxSYBR MasterMix dTTP Blue Kit (Eurogentec, Liège, Belgium) was used and *ACTIN2* was used to normalize DNA levels. DNA copy number was performed in a final volume of 10 μ l employing a LightCycler® 480 (Roche, Basel, Switzerland). PCR were performed with primers summarize in Additional file 3 with the following conditions : 95 °C for 5 min followed by 45 cycles of amplification composed of 10 s at 95 °C, 10 s at 60 °C and 10 s at 72 °C, then a final soak at 37 °C for 30 s was done. The *EVD* DNA copies were determined by the $2\Delta\Delta CT$ method and diagram represent relative quantity of amplification compared to the WT, which was taken as 2.

MinION library preparation

The MinION™ sequencing library was generated with the sequencing Kit (SQK-MAP-006) (Oxford Nanopore Technologies, Oxford, UK) according to the manufacturer's instructions.

Four micrograms of DNA were fragmented as follows : DNA was loaded into a G-tube (Covaris, Brighton, UK) and spun at 6 000 rpm in an eppendorf 5424 for 1 min before inverting the tube to centrifuge again for 1 min.

One microgram of fragmented DNA was end-repaired and dA-tailed using NEBNext Ultra II end-repair/dA-tailing module (Biolabs, New England, USA, cat. no. E7546S) as per the manufacturer's instructions except for the thermal cycler program performed with 5 min at 20 °C, followed by 5 min at 65 °C and terminated with 5 min at 4 °C.

The DNA was further purified with 1.0 \times vol. Agencourt AMPure XP beads (Beckman Coulter, High Wycombe, UK). After 2 washes with 70 % of ethanol, beads were air dried and the DNA was eluted with 31 μ l of Gibco distilled water. A ligation with biotinylated hairpin adapters was then performed by adding 50 μ l Blunt/TA ligase master MIX (Biolabs, New England, US, cat. no. M0367S) to the A-tailed library.

Library was enriched of molecules fixed with biotinylated hairpin thanks to the use of MyOne™ Streptavidin C1 Dynabeads (ThermoFisher) prepared with bead binding buffer (BBB). Then, library was eluted off beads using 25 μ l of elution buffer (ELB). The DNA concentration in the final library was quantified using a Quant-iT.

Sequencing MinION library

Three Oxford Nanopore flow cells : 1 FLO-MAP003 and 2 FLO-MAP103 were used in this study.

A MinION flow cell was removed out of its packaging and inserted into the MinION. Then, the number of available pores was recorded to evaluate the flow cell's quality with the MinKNOW™ software. The version v0.51.1.62 of MinKNOW™ software was used with the flow cell FLO-MAP003 and version v0.51.3.40 with the 2 other flow cells FLO-MAP103.

Before starting the analyses, the Minknow application was first used to make the platform QC before starting the analyses and a map/NC-48-h sequencing run flow (Map 103) protocol was chosen and initiated. 500 µl of a priming mix (26.6 µl of *Fuel mix (FMX)*, 500 µl of *Running buffer (RNB)*, 473.4 µl of Nuclease-free water) were introduced 2 times in the MinION flow cell to prime the sensory array. Thirty-six ng of the prepared library was diluted into a mix composed of 75 µl of *RNB*, 4 µl of *FMX* and Nuclease-free water was added to reach a final volume of 150 µl. The 150 µl priming mix was loaded into the sample loading port of the flow cell. The sequencing reaction was started. After 24 hours, 150 µl of a priming mix with 36 ng of the library was loaded again. One of the flow-cells FLO-MAP103 was loaded with a priming mix containing 108 ng of the prepared library. Base-calling was performed through data transference using the Metrichor™ agent (v2.39.3) and 2D base-calling workflow (v1.69). During the sequencing run, one additional freshly diluted aliquot of DNA library was loaded after 24 h of initial input.

MinION data analysis

Raw reads were first extracted from the native HDF5 format MinIon reads using poretools (Loman and Quinlan, 2014). To perform alignments, a BLAST multifasta database of all the MinIon reads was created with formatdb (version 2.2.26). Then the *EVD* sequence was aligned against the multifasta database with blastall (version 2.2.26) with the default parameters. For each blast hit, a dotter (version 4.1) with the Nanopore read against the *EVD* sequence was produced. In parallel, the same Nanopore read was aligned against the *Arabidopsis thaliana* genome, using ncbi BLAST with default parameters in order to identify position of the flanking region. The start and end position of the alignment were kept and a sequence with plus or minus 1 kb was extracted of the *Arabidopsis thaliana* genome for each position. The extracted sequence and the *EVD* sequence were concatenated and a dotter with the 2D fasta MinION read and the concatenated sequence was performed.

Transposition of other TE families was tentatively detected from the sequencing data : all reads were blasted against the Col0 reference genome sequence. All reads showing partial homology were filtered out of the dataset.

RBIP

RBIP is a PCR-based marker strategy using a three-primer set derived from the flanking region and the LTR to the LTR-RT TE. Here, RBIP primer sets were designed to amplify the full sites between the internal *EVD* sequence and its flanking sequence.

Primers used to confirm the point of retrotransposon neo-insertion were designed with the Primer3 tool <http://primer3.ut.ee/>, 0.4.0 version (Koressaar and Remm, 2007; Untergasser *et al.*, 2012) (Additional file 2).

PCR was performed in volume of 23 μ l containing 5 μ l of Gotaq® DNA polymerase Buffer (Promega, Madison, WI, USA), 14.8 μ l water, 1 μ l of 240 μ M dNTP , 1 μ l of 10 μ M of each primer, 0.06 U/ μ l of Gotaq® DNA polymerase and 2 μ l of 1 ng/ μ l DNA.

Amplification was conducted under the same parameters except for the elongation step depending of the expected amplicon size. Regions containing the entire *EVD* insertion was amplified with an elongation time of 5 min. All other parameters were identical to the classical PCR approach. The amplification was performed with the program described as follows : Initial hot start (3 min, 94 °C) and then 32 cycles with denaturation (30 s, 94 °C), annealing (30 s, 60 °C), elongation (1-5 min, 72 °C), with a final soak (4 min, 20 °C).

The amplified product was electrophoresed at 100 V on a 1 % agarose gel (Euromedex, Souffelweyersheim, France) in a 1X TAE buffer (40 mM Tris, 20 mM glacial acetic acid, 1 mM EDTA pH 8). The 1 % agarose gel solution was prepared with Gel red™ 10 000X stock reagent (Interchim, Montluçon, France) diluted at 1 :10 000.

Sequencing of PCR amplicons

PCR amplicons were purified with the GeneClean® purification kit (MP Biomedicals, Santa Ana, CA, USA) according to the manufacturer's instructions. Sanger sequencing was performed with the same primers used to obtain these PCR amplicons on the ABI3130X1 Genetic Analyser sequencer (Applied Biosystems, Thermo Scientific, Waltham, MA, USA) at the Sequencing platform of the LGDP, Perpignan, France. Analysis were then performed using the Ridom TraceEdit software (Ridom Bioinformatics, Würzburg, Germany).

Acknowledgments

We would like to thank Jerzy Paszkowski and Jon Reinders for providing epi12 seeds.

Fig. 1. Accumulation of *EVD* DNA.

Fig. 2. Characterization of *EVD* neoinsertion 4.

Fig. 3. Representation of *EVD* copies in epiRIL12 sample.

Fig. 4. Validation of *EVD* insertions in epiRIL12 detected by MinION.

Additional file 1. Box-plot of MinION read length.

Additional file 2. Box-plot of identity of MinION read mapped on Col0

Additional file 3. Primers used for quantitative PCR and for real-time PCR

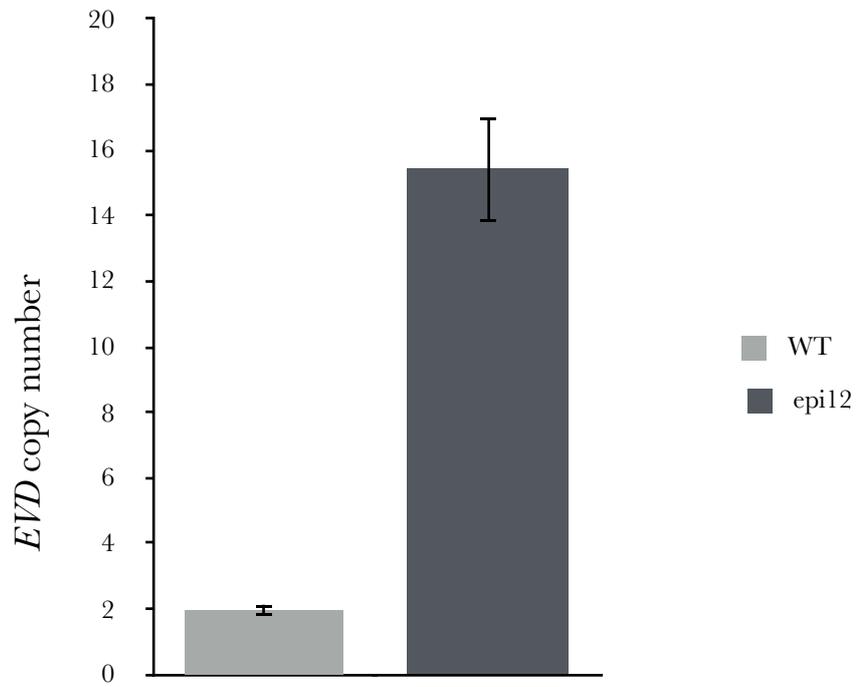


Fig. 1
Accumulation of *EVD* DNA. DNA accumulation of *EVD* for wild type (WT) and epiRIL12 (epi12) measured by qPCR (mean \pm s.e.m., $n = 3$ technical repetitions).

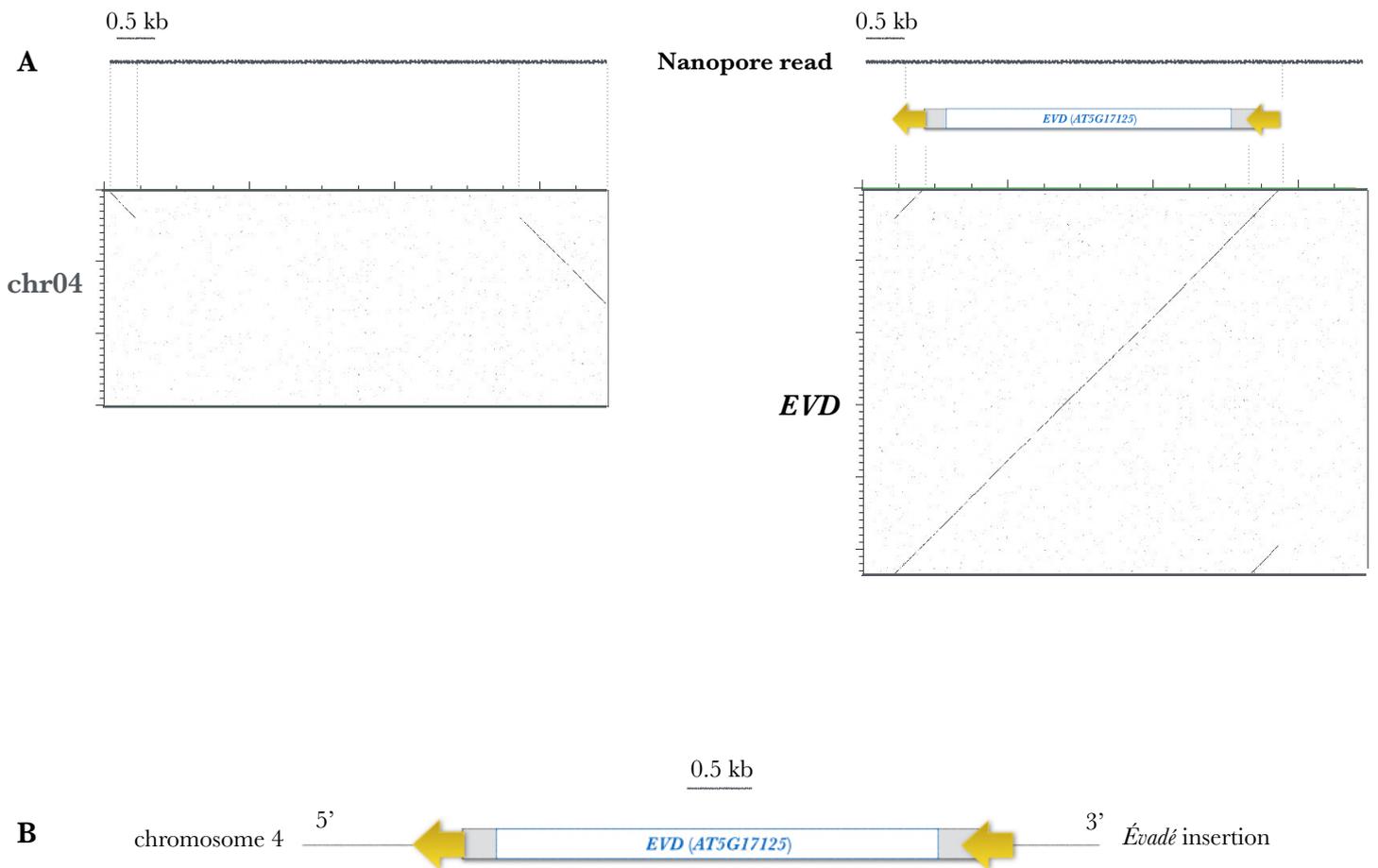


Fig. 2

Characterization of *EVD* neoinsertion 4. (A) Dot-plots comparison of a 7kb Nanopore read with the region of the insertion on the 4th chromosome (chr04) (left) and the sequence of *EVD* (right). Blue box : Open Reading Frame, grey boxes : UTR and yellow arrows : long terminal repeats. (B) Schematic representation of *EVD* insertion in the chromosome 4. The orientation of the *EVD* LTRs are indicated by yellow arrows, the *EVD* insertion is in antisense.

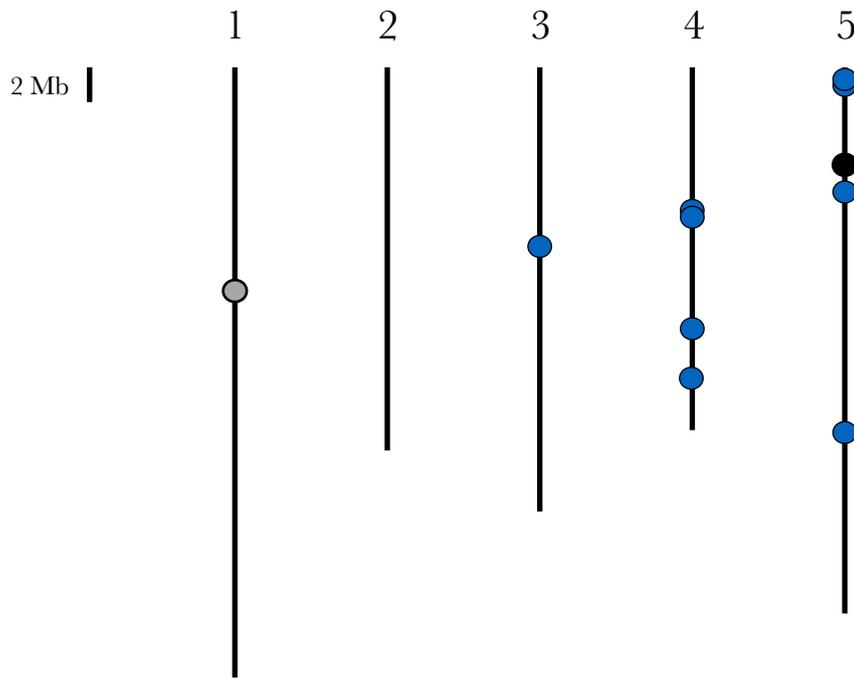


Fig. 3

Representation of *EVD* copies in epiRIL12 sample.

The native *EVD* copy detected by ONT is indicated by black circle and the non detected one by grey circle. The new *EVD* copies detected by ONT are indicated by blue circles. The zygosity state are not shown.

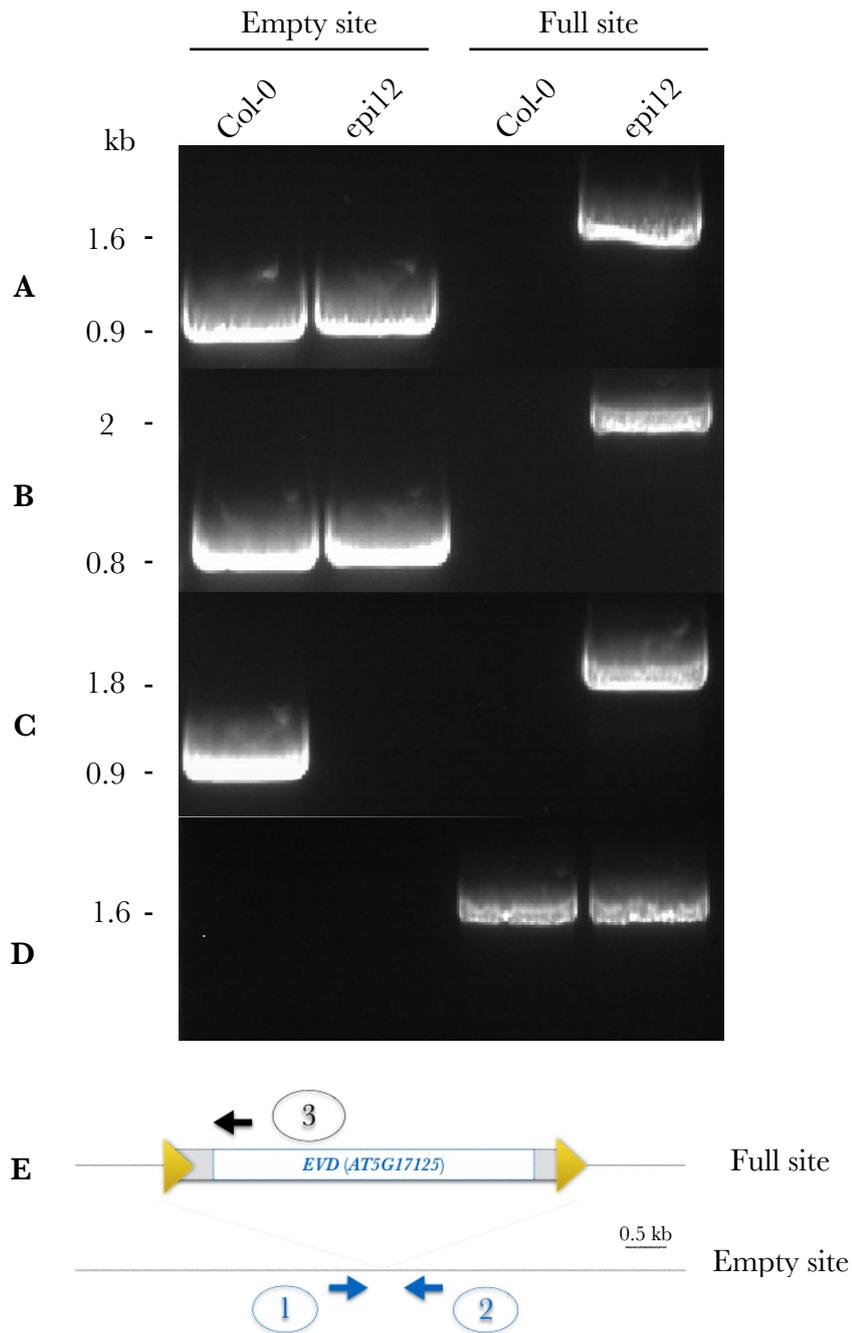


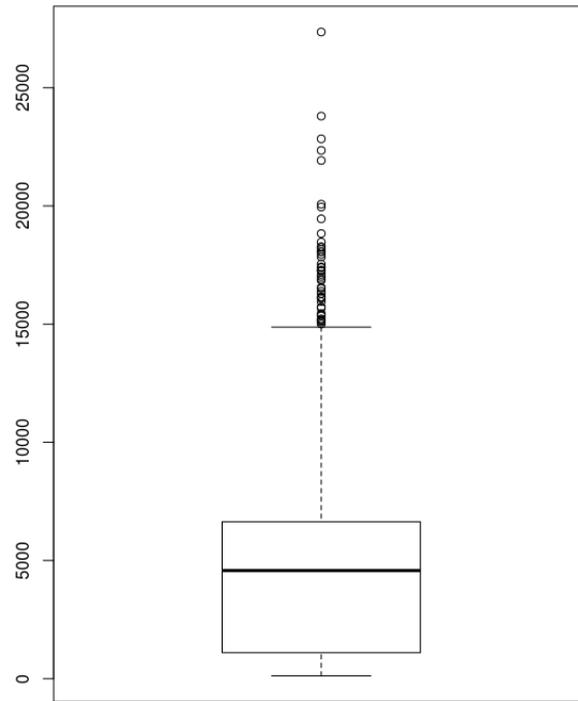
Fig. 4

Validation of *EVD* insertions in epiRIL12 detected by MinION.

PCR detection of *EVD* insertion polymorphism for (A) neo-insertion 4, (B) neo-insertion 6, (C) neo-insertion 7 and (D) a native copy of *EVD* on chromosome 5 as control. (E) Schematic illustration of RBIP primer design. Legend as in Figure 2. PCR with primers 1 and 2 is designed to detect empty sites and PCR with primers 1 and 3 (sense insertion) or 3 and 2 (antisense insertion) are designed to detect full sites.

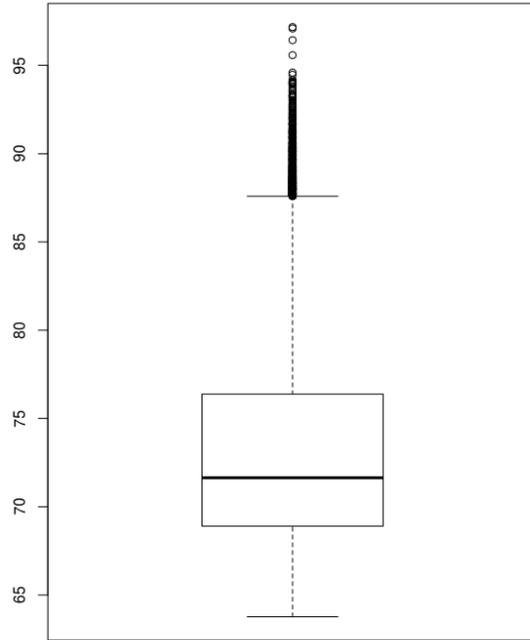
Additional file 1.

Box-plot of MinION read length.



Additional file 2.

Box-plot of identity of MinION read mapped on Col10



Additional file 3.

Primers used for quantitative PCR and for real-time PCR

Primer name	Sequence 5'→3'
qACT2	
Forward	GCCAATCTACGAGGGTTTCT
Reverse	TTACAATTTCCCGCTCTGCT
qEVD29	
Forward	CCGTTGAAGAGGACTCACGAG
Reverse	TTCCTTGGTCTTGGTGATGAGA
neo4	
Forward	TCCTCCACCGCAATTACCGC
Reverse	ACTCCAATGACGCAGCACCA
neo6	
Forward	ACGCGGCAATAGCCAGGAAT
Reverse	GCATCAAGCAGAAGAATGGTGGA
neo7	
Forward	GCACTTGATTTGTTCCGTGGCA
Reverse	AGCCGCTTGTAAGTAAAGAGGA

References

- Britten RJ., Kohne DE. Repeated sequences in DNA. *Science*. 1968; 161 :529-40.
- Cheng C., Tarutani Y., Miyao A., Ito T., Yamazaki M., Sakai H., Fukai E., Hirochika H. Loss of function mutations in the rice chromomethylase OsCMT3a cause a burst of transposition. *Plant J*. 2015; 83 : 1069-81.
- Deamer D. Akeson M., Branton D. Three decades of nano pore sequencing. *Nat. Biotechnol*. 2016; 34 : 518-24.
- Ewing AD. Transposable element detection from whole genome sequence data. *Mobile DNA*. 2015; 6 :24.
- Feng Y., Zhang Y., Ying C., Wang D., Du C. Nanopore-based fourth-generation DNA sequencing technology. *Genomics, Proteomics and Bioinformatics*. 2015; 13 : 4-16.
- Feschotte C., Pritham E.J. DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev.Genet*. 2007; 41 : 331-68.
- Glusman G., Severson A., Dhankani V., Robinson M., Farrah T., Mauldin D.E., Stittrich A.B., Ament S.A., Roach J.C., Brunkow M.E., Bodian D.L., Vockley J.G., Shmulevich I., Niederhuber J.E., Hood L. Identification of copy number variants in whole-genome data using Reference Coverage Profiles. *Front Genet*. 2015; 6 : 6 :45.
- Handsaker R.E., Kom J.M., Nemes J., McCarroll S.A. Discovery and genotyping of genome structural polymorphism by sequencing on a population scale. *Nat. Genet*. 2011; 43 : 269–76.
- Kobayashi S., Goto-Yamamoto N., Hirochika H. Retrotransposon-induced mutations in grape skin color. *Science*. 2004; 304 : 982-82.
- Koressaar T., Remm M. Enhancements and modifications of primer design program Primer3. *Bioinformatics*. 2007; 23 :1289-91.
- Laver T., Harrison J., O'Neill P.A., Moore K., Farbos A., Paszkiewicz K., Studholme D.J. Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomol. Detect. and Quantif*. 2015; 3 : 1-8.
- Lisch D. How important are transposons for plant evolution? *Nat. Rev. Genet*. 2013; 14 : 49-61.
- Loman N.J., Quinlan A.R. Poretools : a toolkit for analyzing nanopore sequence data. *Bioinformatics*. 2014; 30 : 3399-401.
- Makalowski W., Pande A., Gotea V., Makalowska I. Transposable elements and their identification. *Methods in Molecular Biology*. 2012; 855 : 337–59.
- Medvedev P., Stanciu M., Brudno M. Computational methods for discovering structural variation with next-generation sequencing. *Nat. Methods*. 2009; 6 :S13–20.
- Mills R.E., Walker K., Stewart C., Handsaker R.E., Chen K., Alkan C., Abyzov A., Yoon S.C., Ye K., Cheetham R.K., Chinwalla A., Conrad D.F., Fu Y., Grubert F, Hajirasouliha I., Hormozdiari

F., Iakoucheva L.M., Iqbal Z., Kang S., Kidd J.M., Konkel M.K., Korn J., Khurana E., Kural D., Lam H.Y., Leng J., Li R., Li Y., Lin C.Y., Luo R., Mu X.J., Nemesh J., Peckham H.E., Rausch T., Scally A., Shi X., Stromberg M.P., Stütz A.M., Urban A.E., Walker J.A., Wu J., Zhang Y., Zhang Z.D., Batzer M.A., Ding L., Marth G.T., McVean G., Sebat J., Snyder M., Wang J., Ye K., Eichler E.E., Gerstein M.B., Hurles M.E., Lee C., McCarroll S.A., Korbel J.O.. Mapping copy number variation by population-scale genome sequencing. *Nature*. 2011; 470 : 59–65.

Mirouze M., Reinders J., Bucher E., Nishimura T., Schneeberger K., Ossowski S., Cao J., Weigel D., Paszkowski J., Mathieu O. Selective epigenetic control of retrotransposition in *Arabidopsis*. *Nature*. 2009; 427-30.

Quadrana L., Bortolini Silveira A., Mayhew G.F., LeBlanc C., Martienssen R.A., Jeddloh J.A., Colot V. The *Arabidopsis thaliana* mobilome and its impact at the species level. *eLife*. 2016; 5 : e15716.

Reinders J., Wulff B. B., Mirouze M., Mari-Ordonez A., Dapp M., Rozhon W., Bucher E., Theiler G., Paszkowski J. Compromised stability of DNA methylation and transposon immobilization in mosaic *Arabidopsis* epigenomes. *Genes Dev*. 2009; 23 : 939-50.

Sabot F., Picault N., El-Baidouri M., Llauro C., Chaparro C., Piegu B., Roulin A., Guiderdoni E., Delabastide M., McCombie R., Panaud O. Transpositional landscape of the rice genome revealed by paired-end mapping of high-throughput re-sequencing data. *Plant J*. 2011; 66 : 241-6.

Untergasser A., Cutcutache I., Koreasaar T., Ye J., Faircloth BC., Remm M., Rozen SG. Primer3-new capabilities and interfaces. *Nucleic Acids Res*. 2012; 40 : e115.

Wei L., Cao X. The effect of transposable elements on phenotypic variation : insights from plants to humans. *Sri. China Life Sci*. 2016; 59 : 24-37.

Ye K., Schulz M.H., Long Q., Apweiler R., Ning Z. Pindel : a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics*. 2009; 25 : 2865–71.

Yoon S., Xuan Z., Makarov V., Ye K., Sebat J. Sensitive and accurate detection of copy number variants using read depth of coverage. *Genome Res*. 2009; 19 : 1586–92.

Zhao D., Ferguson A.A., Jiang N. What makes up plant genomes : The vanishing line between transposable elements and genes. *Biochim. Biophys. Acta*. 2016; 1859 : 336-80.

Les résultats obtenus par ce travail montrent que l'utilisation des données obtenues avec les séquenceurs de 3^{ème} génération pourra permettre une étude exhaustive des variations structurales (SVs) dans les génomes. Malgré le taux d'erreurs de ces séquenceurs, c'est la première fois qu'un séquençage nous permet de détecter l'insertion complète d'un ET, et, de plus avec ses régions flanquantes, correspondant à l'endroit de son insertion. Ceci devrait permettre de diminuer le nombre de faux positifs détectés avec les méthodes actuelles de détections de SVs. Quant au problème des faux négatifs, il devrait être résolu avec l'augmentation de la couverture du séquençage que nous pourrions obtenir par cette technique.

La plante utilisée lors de ce travail, est issue d'une lignée épigénétique recombinante dérivée du mutant *met1* (epiRIL). Le choix de ce matériel a été fait : d'une part, car la mobilité du rétrotransposon *Évadé* y a été établit (Mirouze *et al.*, 2009), et d'autre part, car le génome d'*A. thaliana* est relativement petit. Ce travail a servi de contrôle pour valider cette nouvelle approche basée sur le séquençage.

A l'issue de ce travail, nous projetons de réaliser le séquençage d'ADN d'un mutant *Osnrpd1* de génération T5 ou T6 pour lesquels un phénotype de talles torsadées a été observé. Ceci afin de tenter d'établir un lien causal entre le phénotype et de potentielles néo-insertions d'ETs.

DISCUSSION GÉNÉRALE

Les ETs occupent une part prépondérante de certains génomes (44% chez l'homme; 34% chez le riz) et peuvent atteindre plus de 85% du génome de certaines espèces végétales comme le maïs (Schnable *et al.*, 2009). Cette répartition dans les génomes suggère qu'au cours de l'évolution des bursts de transposition se sont produits de manière importante ou fréquente (Piegu *et al.*, 2006; Belyayev, 2014). L'origine de ces bursts de transposition est, aujourd'hui encore, inconnue. L'intérêt de l'étude des ETs est encore assez récent et les études sur leur régulation sont de plus en plus nombreuses. Plusieurs analyses génomiques et génétiques chez les plantes ont permis de montrer l'existence de différents mécanismes pour prévenir de nouvelles insertions d'ETs.

Cependant, seulement certaines familles sont présentes en très grand nombre de copies, pourquoi certaines familles sont-elles réactivées alors que d'autres non? Pendant plusieurs années les ETs ont été considérés comme des entités égoïstes. Depuis, leur implication dans l'évolution des génomes a changé cette vision. Néanmoins, l'impact de la réactivation des ETs a plusieurs conséquences sur les génomes allant des effets délétères (Miki *et al.*, 1992; Mazoyer *et al.*, 2005), aux avantages évolutifs qu'il confèrent (Kobayashi *et al.*, 2004; Hazzouri *et al.*, 2015) ou n'entraîne aucune conséquence visible sur le développement normal de l'hôte.

Il est alors évident que les contrôles mis en place pour les maintenir à l'état inactif sont indispensables pour prévenir les éventuelles perturbations qu'ils pourraient entraîner sur le développement de leur hôte. Pourtant, quelques éléments peuvent contourner ces mécanismes de répression sous l'impact de certains stress (Capy, 2000; Grandbastien, 2015). C'est alors que l'on peut se demander pourquoi ces familles sont réactivées. Seroit-ce parce que la nature égoïste de ces éléments induit leur multiplication pour que leur présence dans les génomes soit maintenue? Dans ce cas, l'intérêt d'un ET est de survivre et, ainsi, de cibler des régions complexes du génome non transcrites/essentiels pour le développement de l'organisme et de limiter/contourner leur répression. Dans le cas contraire, ces éléments seraient voués à disparaître par les mécanismes d'élimination de séquences du génome (Devos *et al.*, 2002). Une autre hypothèse résiderait dans la nature même des ETs, plus une famille est grande plus la répression des éléments de cette famille sera difficile, le nombre d'ETs à réprimer étant plus important, le nombre de siRNAs mobilisés pour réprimer chaque élément de cette famille est plus faible (Lisch et Slotkin, 2011). Dans le cas contraire, une famille peu répétée pourrait plus facilement contourner les mécanismes de répression comme c'est le cas pour les éléments *Tos17*, *Lullaby* et *Karma* chez le riz. Ou alors leur réactivation serait-elle due à l'impact de stress particuliers pour activer ou réprimer certains gènes dans le but de moduler leur expression afin que l'organisme s'adapte à un nouvel environnement plus rapidement?

Ces séquences mobiles ne seraient alors plus seulement des séquences égoïstes mais leur présence dans les génomes serait indispensable pour pouvoir répondre aux changements brutaux imposés par des stress occasionnels.

Certes, l'activation des ETs n'a pas été montrée *in natura* puisque la majorité des réactivations d'ETs ont été observées soit après un passage par la *CIV* (Hirochika *et al.*, 1996; ; Kalendar *et al.*, 2000; Komatsu *et al.*, 2003, Picault *et al.*, 2009), soit en réponse à des stress (Grandbastien *et al.*, 1989), soit dans des mutants épigénétiques (Mirouze *et al.*, 2009; Tsukahara *et al.*, 2009). De ce fait, l'origine des mécanismes impliqués dans les bursts transpositionnels qui ont eu lieu au cours de l'évolution dans les populations naturelles ne peut encore qu'être formulée de façon hypothétique. Cependant, l'analyse de l'activité rétrotranspositionnelle dans les plantes de génotype sauvage et en contexte épigénétique nous permet de mieux comprendre l'impact de la régulation de l'activité des ETs chez le riz. Bien que l'étape de la *CIV* ne soit pas retrouvée naturellement dans la nature, elle peut être considérée comme un stress s'il est défini par des conditions physiologiques altérées qui perturbent l'équilibre nécessaire au bon développement de la plante (pour revue : Gaspar *et al.*, 2002). De plus, dans la nature, le matériel génétique de tous les êtres vivants est sans cesse l'objet de modifications et des mutants naturels peuvent apparaître. Ces mutations pourraient résulter en un relâchement des voies de répression des ETs conduisant à leur réactivation et ainsi mimer les résultats obtenus dans l'étude de certains mutants épigénétiques. Les stress environnementaux, comme le stress thermique, qui ont eu lieu au cours de l'évolution, pourraient avoir augmenté la capacité de transposer des ETs dans les génomes de ces mutants naturels à l'image du rétrotransposon *Osen* dans le mutant épigénétique *nrpd1* d'*A. thaliana* (Ito *et al.*, 2011). Dans le cas de notre étude, il ne semble pas que l'accumulation de la perturbation de la biogenèse des siRNAs et du stress thermique appliqué soit suffisante pour expliquer des bursts de transposition. Est-ce le stress thermique qui n'était pas assez important pour entraîner un burst transpositionnel ou n'était-ce pas le stress approprié?

Quoiqu'il en soit, les origines de l'activation des ETs peuvent être diverses et notre étude confirme que les ETs ne sont pas tous régulés de la même manière. L'étude de l'activité transpositionnelle du mutant *Osnrpd1* a permis de mettre en évidence l'activité de 8 familles d'ETs. La mobilité de 3 de ces familles d'ETs, *Tos17*, *Lullaby* et *Karma*, a été provoquée par le passage en *CIV* (Hirochika *et al.*, 1996; Komatsu *et al.*, 2003; Picault *et al.*, 2009). Récemment, Cheng et associés (2015) ont montré que la perte de la méthylation réalisée par l'intermédiaire de CMT3 permettait la réactivation de 8 autres éléments dont *Tos17* et *Lullaby*, précédemment cités, mais également le

transposon *mPing* dont la mobilité a également été montrée dans des cultures cellulaires (Jiang *et al.*, 2003; Nakazaki *et al.*, 2003). Il semble donc que l'altération de certaines voies de répression engendre la réactivation de familles d'ETs bien spécifiques. Notre étude confirme que les ETs sont contrôlés par un « mille-feuille » de répression (Rigal et Mathieu, 2011) dont les différentes couches sont génétiques et épigénétiques mais également environnementales. La contribution de chacune de ces différentes couches de répression est variable selon la famille d'ETs. Ainsi, chez le riz, on peut d'ores-et-déjà définir pour quelques familles les différentes couches impliquées dans leur répression.

Les bursts ayant eu lieu au cours de l'évolution seraient-ils la conséquence de différents facteurs ayant perturbés le paysage épigénétique et ainsi la réactivation de certaines familles? Ou alors n'y aurait-il pas eu un seul burst transpositionnel dont la cause est toujours inconnue, qui serait soit l'accumulation de différents facteurs/stress soit une couche encore non identifiée qui contrôlerait plusieurs voire toutes les familles d'ETs. Ainsi se pose la question de savoir s'il existe une couche de répression qui serait à la base de ce « mille-feuille » et dont la dérégulation aurait entraîné un burst de transposition au cours de l'évolution. Et si l'origine des réactivations d'ETs ne s'expliquait pas par la perturbation d'une seule de ces couches de répression mais que les bursts résultaient d'une accumulation de différents facteurs?

L'activité transpositionnelle et les mécanismes de répression sont encore loin d'être entièrement connus. L'étude des mécanismes contrôlant la mobilité des ETs est encore assez récente, et ces mécanismes sont souvent étudiés lorsque l'insertion d'ET entraîne un phénotype observable (Mirouze *et al.*, 2009; Tsukahara *et al.*, 2009; Miura *et al.*, 2001). Cependant, à l'aide des nouvelles techniques de séquençage de 3^{ème} génération, qui permettront un séquençage rapide et à moindre coût, la prospection des ETs dans des populations entières pourrait permettre de mieux comprendre la dynamique transpositionnelle des ETs au cours de l'évolution.

RÉFÉRENCES BIBLIOGRAPHIQUES

Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*. 2000; 408 : 796-815.

Aury J.-M., Jaillon O., Duret L., Noel B., Jubin C., Porcel B.M., Ségurens B., Daubin V., Anthouard V., Aiach N., *et al.* Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature*. 2006; 171-8.

Baucom R.S., Estill J.C., Leebens-Mack J., Bennetzen J.L. Natural selection on gene function drives the evolution of LTR retrotransposon families in the rice genome. *Nature*. 2009; 19 : 243-54.

Belyayev A. Bursts of transposable elements as an evolutionary driving force. *J. Evol. Biol.* 2014; 27 : 2573-84.

Bennetzen J.L. Transposable element contributes to plant gene and genome evolution. *Plant Mol. Biol.* 2000; 42 : 251-69.

Bennetzen J.L., Ma J., Devos K.M. Mechanisms of recent genome size variation in flowering plants. *Ann. Bot.* 2005; 95 : 127-32.

Bennetzen J.L., Kellogg E.A. Do plants have a one-way ticket to genomic obesity? *Plant Cell*. 1997; 9 : 1509-14.

Bentley D.R., Balasubramanian S., Swerdlow H.P., Smith G.P., Milton J., Brown C.G., Hall K.P., Evers D.J., Barnes C.L., Bignell H.R., *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 2008; 456 : 53-9.

Bhattacharyya M.K., Smith A.M., Noel Ellis T.H., Hedley C., Martin C. The wrinkled-seed character of pea described by Mendel is caused by a transposon-like insertion in a gene encoding starch-branching enzyme. *Cell*. 1990; 60 : 115-22.

Bird A.P. Gene number, noise reduction and biological complexity. *Trends Genet.* 1995; 11 : 94-100.

Birney E., Soranzo N. Human genomics : The end of the start for population sequencing. *Nature*. 2015; 526 : 52-3.

Bureau T.E., Wessler S.R. Tourist : a large family of small inverted repeat elements frequently associated with maize genes. *Plant Cell*. 1992; 4 : 1283-94.

Bureau T.E., Wessler S.R. Stowaway : a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. *Plant Cell*. 1994; 6 : 907-16.

Cao X., Jacobsen S.E. Role of the arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. *Curr. Biol.* 2002; 12 : 1138-44.

Cao X., Aufsatz W., Zilberman D., Mette M.F., Huang M.S., Matzke M., Jacobsen S.E. *Curr. Biol.* 2003; 13 : 2212-7.

- Capy P., Gasperi G., Biéumont C., Bazin C. Stress and transposable elements : co-evolution or useful parasites? *Heredity*. 2000; 101-6.
- Casacuberta J.M., Santiago N. Plant LTR-retrotransposons and MITEs : control of transposition and impact on the evolution of plant genes and genomes. *Gene*. 2003; 311 :1-11.
- Castel S.E., Martienssen R.A. RNA interference in the nucleus : roles for small RNAs in transcription, epigenetic and beyond. *Nat. Rev. Genet.* 2013; 14 : 100-12.
- Chan S.W.L., Zilberman D., Xie Z., Johansen L.K., Carrington J.C., Jacobsen S.E. RNA silencing genes control de novo DNA methylation. *Science*. 2004; 303 : 1336.
- Chaparro C., Guyot R., Zuccolo A., Piégu B., Panaud O. RetrOryza : a database of the rice LTR-retrotransposons. *Nucleic Acids Res.* 2007; 35(Database issue) : D66-70.
- Chen C., Begcy K., Liu K., Folsom J.J., Wang Z., Zhang C., Walia H. Heat stress yields a unique MADS box transcription factor in determining seed size and thermal sensitivity. *Plant Physiol.* 2016; 17 : 606-22.
- Cheng C., Tarutani Y., Miyao A., Ito T., Yamazaki M., Sakai H., Fukai E., Hirochika H. Loss of function mutations in the rice chromomethylase OsCMT3a cause a burst of transposition. *Plant J.* 2015; 83 : 1069-81.
- Chiang DY., Getz G., Jaffe DB., O'Kelly MJ., Zhao X., Carter SL., Russ C., Nusbaum C., Meyerson M., Lander ES. High-resolution mapping of copy-number alterations with massively parallel sequencing. *Nat. Methods.* 2009; 6 :99-103.
- Choi Y., Gehring M., Johnson L., Hannon M., Harada J.J., Goldberg R.B., Jacobsen S.E., Fischer R.L. DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in arabidopsis. *Cell.* 2002; 110 : 33-42.
- Clark L.A., Wahl J.M., Rees C.A., Murphy K.E. Retrotransposon insertion in *SILV* is responsible for merle patterning of the domestic dog. *Proc. Natl. Acad. Sci USA.* 2006; 103 : 1376-81.
- Cokus S.J., Feng S., Zhang X., Chen Z., Merriman B., Haudenschild C.D., Pradhan S., Nelson S.F., Pellegrini M., Jacobsen S.E. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature.* 2008; 452 : 215-9.
- Comai L. The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* 2005; 6 : 836-46.
- Cui L., Wall P.K., Leebens-Mack J.H., Lindsay B.G., Soltis D.E., Doyle J.J., Soltis P.S., Carlson J.E., Arumuganathan K., Barakat A., *et al.* Widespread genome duplications throughout the history of flowering plants. *Genome Res.* 2006; 16 : 738-49.
- Cui X., Cao X. Epigenetic regulation and functional exaptation of transposable elements in higher plants. *Curr. Opin. Plant Biol.* 2014; 21 : 83-8.
- Dalmay T., Hamilton A., Rudd S., Angell S., Baulcombe D.C. An RNA-dependent RNA polymerase gene in Arabidopsis is required for posttranscriptional gene silencing mediated by a transgene but not by a virus. *Cell.* 2000; 101 : 543-53.

- Deamer D., Akeson M., Branton D. Three decades of nanopore sequencing. *Nat. Biotechnol.* 2016; 34 : 518-24.
- Delseny M., Ghesquière A., Lagoda P., Regad F. Le riz, un modèle en génomique. *Biofutur.* 2000; 2000 : 31-5.
- Devos K.M., Brown J.K., Bennetzen J.L. Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Res.* 2002; 12 : 1075-9.
- Devos K.M. Grass genome organization and evolution. *Curr. Opin. Plant Biol.* 2010; 13 : 139-45.
- Du J., Tian Z., Hans C., Laten H., Cannon S., Jackson S.A., Shoemaker R.C., Ma J. Evolutionary conservation, diversity and specificity of LTR-retrotransposon in flowering plants : insights from genome-wide analysis and multi-specific comparison. *Plant J.* 2010; 63 : 584-98.
- Du J., Zhong X., Bernatavichute Y.V., Stroud H., Feng S., Caro E., Vashisht A.A., Terragni J., Chin H.G., Tu A., Hetzel J., *et al.* Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. *Cell.* 2012; 151 : 167-80.
- Efissue A.A., Derera J. Genotypic response in rice during the vegetative phase under water stress and non-stress conditions. *Journal of Crop Improvement.* 2012; 26 : 816-34.
- Eid J., Fehr A., Gray J., *et al.* Real-time DNA sequencing from single polymerase molecules. *Science.* 2009; 323 : 133-8.
- El baidouri M. Les éléments transposables chez les plantes : impact sur la dynamique et l'évolution des génomes. Thèse de doctorat en Sciences de la vie, Biologie, sous la direction d'Olivier Panaud. 2012.
- El baidouri M., Panaud O. Comparative genomic paleontology across plant kingdom reveals the dynamics of TE-driven genome evolution. *Genome Biol. Evol.* 2013; 5 : 594-65.
- Emera D., Wagner G.P. Transposable element recruitments in the mammalian placenta : impacts and mechanisms. *Brief. Funct. Genomics.* 2012; 11 : 267-76.
- Erhard K.F., Parkinson S.E., Gross S.M., Barbour J-E.R., Lim J.P., Hollick J.B. Maize RNA Polymerase IV defines *trans*-generational epigenetic variation. *Plant Cell.* 2013; 25 : 808-19.
- Fedoroff N. Transposable elements, epigenetics, and genome evolution. *Science.* 2012; 338 : 758-67.
- Fedurco M., Romieu A., Williams S., Lawrence I., Turcatti G. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Res.* 2006; 34 : e22.
- Feschotte C. Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* 2008; 9 : 397-405.
- Feschotte C., Jiang N., Wessler S.R. Plant transposable elements : where genetics meets genomics. *Nat. Rev. Genet.* 2002; 3 : 329-41.

- Fiandt M., Szybalski W., Malamy M.H. Polar mutations in lac, gal and phage λ consist of a few IS-DNA sequences inserted with either orientation. *Molec. Gen. Genet.* 1972; 119 : 223-31.
- Finatto T., Costa de Oliveira A., Chaparro C., da Maia L.C., Farias D.R., Woyann L.G., Mistura C.C., Soares-Bresolin A.P., Llauro C., Panaud O., Picault N. Abiotic stress and genome dynamics : specific genes and transposable elements response to iron excess in rice. *Rice.* 2015; 8 : 13.
- Finnegan D.J. Eukaryotic transposable elements and genome evolution. *Trends Genet.* 1989; 5 : 103-107.
- Finnegan E.J., Dennis E.S. Isolation and identification by sequence homology of a putative cytosine methyltransferase from *Arabidopsis thaliana*. *Nucleic Acids Res.* 1993; 21 : 2383-8.
- Flemming L.R. *Zellsubstanz, Kern und Zelltheilung*. Leipzig, F.C.W. Vogel. 1882.
- Gao D., Jiang N., Wing R.A., Jiang J., Jackson S.A. Transposons play an important role in the evolution and diversification of centromeres among closely related species. *Front. Plant Sci.* 2015; 6 : 216.
- Gaspar T., Franck T., Bisbis B., Kevers C., Jouve L., Hausman J.F., Ommes J.D. Concepts in plant stress physiology. Application to plant tissue cultures. *Plant Growth Regulation.* 2002; 37 : 263-85.
- Glusman G., Severson A., Dhankani V., Robinson M., Farrah T., Mauldin D.E., Stittrich A.B., Ament S.A., Roach J.C., Brunkow M.E., *et al.* Identification of copy number variants in whole-genome data using Reference Coverage Profiles. *Front Genet.* 2015; 6 : 45.
- Gong, Z., Morales-Ruiz, T., Ariza, R.R., Roldán-Arjona, T., David, L., and Zhu, J.-K. ROS1, a Repressor of Transcriptional Gene Silencing in Arabidopsis, Encodes a DNA Glycosylase/Lyase. *Cell.* 2002; 111 : 803-14.
- Goodstadt L., Ponting C.P. Phylogenetic reconstruction of orthology, paralogy, and conserved synteny for dog and human. *PLoS Comput. Biol.* 2006; 2 : e133.
- Grandbastien M.A., Spielmann A., Caboche M. Tnt1, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. *Nature.* 1989; 337 : 376-80.
- Grandbastien M.A. Activation of plant retrotransposons under stress conditions. *Trends Plant Sci.* 1998; 3 : 181-87.
- Grandbastien M.A., Audeon C., Bonnivard E., Casacuberta J.M., Chalhoub B., Costa A.P., Le Q.H., Melayah D., Petit M., Poncet C., *et al.* Stress activation and genomic impact of Tnt1 retrotransposons in Solanaceae. *Cytogenet. Genome Res.* 2005; 110 : 229-41.
- Grandbastien M.A. LTR retrotransposons, handy hitchhikers of plant regulation and stress response. *Bioch. Biophys. Acta.* 2015; 1849 : 403-16.
- Greenleaf W.J., Block S.M. Single-molecule, motion-based DNA sequencing using RNA polymerase. *Science.* 2006; 313 : 801.
- Gregory T.R. Synergy between sequence and size in large-scale genomics. *Nat. Rev. Genet.* 2005; 6 : 699-708.

- Grover C.E., Wendel J.F. Recent insights into mechanisms of genome size change in plants. *J. Bot.* 2010; 1-8.
- Guo W., Fiziev P., Yan W., Cokus S., Sun X., Zhang M.Q., Chen P.Y., Pellegrini M. BS-Seeker2 : a versatile aligning pipeline for bisulfite sequencing data. *BMC Genomics.* 2013; 14 : 774.
- Haag J.R., Pikaard C.S. Multisubunit RNA polymerases IV and V : Purveyors of noncoding RNA for plant gene silencing. *Nat. Rev. Mol. Cell Biol.* 2011; 12 : 483-92.
- Hahn M.W., Wray G.A. The g-value paradox. *Evol. Dev.* 2002; 4 : 73-5.
- Ham D.J., Moon J.C., Hwang S.G., Jang C.S. Molecular characterization of two small heat shock protein genes in rice : their expression patterns, localizations, networks, and heterogeneous overexpressions. *Mol. Biol. Rep.* 2013. 40; 6709-20.
- Hancks D.C., Kazazian H.K. Jr. Roles for retrotransposon insertions in human disease. *Mob. DNA.* 2016; 7 : 9.
- Havecker E.R., Cao X., Voytas D.F. The diversity of LTR retrotransposons. *Genome Biol.* 2004; 5 : 225.
- Hawkins J.S., Kim H., Nason J.D., Wing R.A., Wendel J.F. Differential lineage specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. *Genome Res.* 2006; 16 : 1252-61.
- Hawkins J.S., Grover C.E., Wendel J.F. Repeated big bangs and the expanding universe : directionality in plant genome size evolution. *Plant Sci.* 2008; 174 : 557-62.
- Hazzouri K.M., Flowers J.M., Visser H.J., Khierallah H.S.M., Rosas U., Pham G.M., Meyer R.S., Johansen C.K., Fresquez Z.A., Masmoudi K., *et al.* Whole genome re-sequencing of date palms yields insights into diversification of a fruit tree crop. *Nature Comm.* 2015; 6 : 8824.
- Hénaff E., Vives C., Desvoyes B., Chaurasia A., Payet J., Gutierrez C., Casacuberta J.M. Extensive amplification of the E2F transcription factor binding sites by transposons during evolution of Brassica species. *Plant J.* 2014; 77 : 852-62.
- Hénaff E., Zapata L., Casacuberta J. M., Ossowski S. Jitterbug : somatic and germline transposon insertion detection at single-nucleotide resolution. *BMC Genomics.* 2015; 16 : 768.
- Henikoff S., Comai L. A DNA methyltransferase homolog with a chromodomain exists in multiple polymorphic forms in *Arabidopsis*. *Genetics.* 1998; 149 : 307-18.
- Herr A.J., Jensen M.B., Dalmay T., Baulcombe D.C. RNA polymerase IV directs silencing of endogenous DNA. *Science.* 2005; 308 : 118-20.
- Hirochika H. Retrotransposons of rice : their regulation and use for genome analysis. *Plant Mol. Biol.* 1997; 35 : 231-40.
- Hirochika H. Contribution of the Tos17 retrotransposon to rice functional genomics. *Curr. Opin. Plant Biol.* 2001; 4 : 118-22.

- Hirochika H., Fukuchi A., Kikuchi F. Retrotransposon families in rice. *Mol. Gen. Genet.* 1992; 233 : 209-16.
- Hirochika H., Guiderdoni E., An G., Hsing Y.I., Eun M.Y., Han C.D., Upadhyay N., Ramachandran S., Zhang Q., Pereira A., *et al.* Rice mutant resources for gene discovery. *Plant Mol. Biol.* 2004; 54 : 325-34.
- Hirochika H., Sugimoto K., Otsuki Y., Tsugawa H., Kanda M. Retrotransposons of rice involved in mutations induced by tissue culture. *Proc. Natl. Acad. Sci. USA.* 1996; 93 : 7783-88.
- Hirsch H.J., Starlinger P., Brachet P. Two kinds of insertions in bacterial genes. *Mol. Gen. Genet.* 1972; 119 : 191-206.
- International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature.* 2001; 409 : 860-921.
- International Rice Genome Sequencing Project. The map-based sequence of the rice genome. *Nature.* 2005; 436 : 793–800.
- Ito H. Small RNAs and transposon silencing in plants. *Dev. Growth Differ.* 2012; 54 : 100-7.
- Ito H., Gaubert H., Bucher E., Mirouze M., Vaillant I., Paszkowski J. An siRNA pathway prevents transgenerational retrotransposition in plants subjected to stress. *Nature.* 2011; 472 : 115-9.
- Jackson J.P., Lindroth A.M., Cao X., Jacobsen S.E. Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature.* 2002; 416 : 556-60.
- Jacob F., Monod J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* 1961; 3 : 318-56.
- Jagadish S.V.K., Craufurd P.Q., Shi W., Oane R. A phenotypic marker for quantifying heat stress impact during microsporogenesis in rice (*Oryza sativa*). *Funct. Plant Biol.* 2013; 41 : 48–55.
- Jiang N. Bao Z., Zhang X., Hirochika H., Eddy S.R., McCouch S.R., Wessler S.R. An active DNA transposon family in rice. *Nature.* 2003; 421 : 163-7.
- Jiang N., Feschotte C., Zhang X., Wessler S.R. Using rice to understand the origin and amplification of miniature inverted repeat transposable elements (MITEs). *Curr. Opin Plant Biol.* 2004; 7 : 115-9.
- Jiang Y., Wang Y., Brudno M. PRISM : pair-read informed split-read mapping for base-pair level detection of insertion, deletion and structural variants. *Bioinformatics.* 2012; 28 : 2576-83.
- Jin Y., Yang H., Wei Z., Ma H., Ge X. Rice male development under drought stress : phenotypic changes and stage-dependent transcriptomic reprogramming. *Mol. Plant.* 2013; 6 : 1630-45.
- Kajikawa M., Okada N. LINEs mobilize SINEs in the eel through a shared 3' sequence. *Cell.* 2002; 111 : 433-44.
- Kalendar R., Tanskanen J., Immonen S., Nevo E., Schulman A.H., Tanskanen J., Immonen S., Nevo E., Schulman A.H., Immonen S., *et al.* Genome evolution of wild barley (*Hordeum spontaneum*)

- by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. *Proc. Natl. Acad. Sci.* 2000; 97 : 6603-7.
- Kalendar R., Vicient C.M., Peleg O., Anamthawat-Jonsson K., Bolshoy A., Schulman A.H. Large retrotransposon derivatives : Abundant, conserved but nonautonomous retroelements of barley and related genomes. *Genetics*. 2004; 166 :1437-50.
- Kikuchi K., Terauchi K., Wada M., Hirano H.Y. The plant MITE mPing is mobilized in anther culture. *Nature*. 2003; 421 : 167-70.
- Kobayashi S., Goto-Yamamoto N., Hirochika H. Retrotransposon-induced mutations in grape skin color. *Science*. 2004; 304 : 982-82.
- Komatsu M., Shimamoto K., Kyojuka J. Two-Step Regulation and Continuous Retrotransposition of the Rice LINE-Type Retrotransposon *Karma*. *Plant Cell*. 2003; 15 : 1934-44.
- Korbel J.O., Urban A.E., Affourtit J.P., Godwin B., Grubert F., Simons J.F., Kim P.M., Palejev D., Carriero N.J., Du L., *et al.* Paired-end mapping reveals extensive structural variation in the human genome. *Science*. 2007; 318 : 420-6.
- Kumar A., Bennetzen J.L. Plant retrotransposons. *Ann. Rev. Genet.* 1999; 33 : 479-532.
- Kumar V., Khare T. Differential growth and yield responses of salt-tolerant and susceptible rice cultivars to individual (Na^+ and Cl^-) and additive stress effects of NaCl. *Acta Physiol. Plant.* 2016;
- Lahmy S., Bies-Etheve N., Lagrange T. Plant-specific multisubunit RNA polymerase in gene silencing. *Epigenetics*. 2010; 5 : 4-8.
- Lahmy S., Pontier D., Cavel E., Vega D., El-shami M., Kanno T., Lagrange T. PolIV(PolIVb) function in RNA-directed DNA methylation requires the conserved active site and an additional plant-specific subunit. *PNAS*. 2009; 106 : 941-6.
- Lanciano S., Carpentier M-C., Llauro C., Robakowska-Hyzorek D., Lasserre E., Ghesquière A., Panaud O., Mirouze M. Sequencing the extrachromosomal circular mobilome reveals tissue-specific transposable element activity in plants. *PLOS Genet.* 2016. submitted
- Langmead B., Salzberg S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods*. 2012; 9 : 357-9.
- Larmande P., Gay C., Lorieux M., Périn C., Bouniol M., Droc G., Sallaud C., Perez P., Barnola I., Biderre-Petit C., *et al.* Oryza Tag Line, a phenotypic mutant database for the Genoplante rice insertion line library. *Nucleic Acids Res.* 2008; 36(Database issue) : D1022-7.
- Law J.A., Jacobsen S.E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* 2010; 11 : 204-20.
- Leitch I.J., Bennett M.D. Genome downsizing in polyploid plants. *J Linn. Soc.* 2004; 82 : 651-63.
- Leitch A.R., Leitch I.J. Genomic plasticity and the diversity of polyploid plants. *Science*. 2008; 320 : 481-3.

Li X., Qian Q., Fu Z., Wang Y., Xiong G., Zeng D., Wang X., Liu X., Teng S., Hiroshi F., *et al.* Control of tillering in rice. *Nature*. 2003; 422 : 618-21.

Li X-M., Chao D-Y., Wu Y., Huang X., Chen K., Cui L-G., Su L., Ye W-W., Chen H., Chen H-C., *et al.* Natural alleles of a proteasome $\alpha 2$ subunit gene contribute to thermotolerance and adaptation of African rice. *Nature Genet.* 2015; 47 : 827-33.

Lindroth A.M., Cao X., Jackson J.P., Zilberman D., McCallum C.M., Henikoff S., Jacobsen S.E. Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. *Science*. 2001; 292 : 2077-80.

Lippman A., Gendrel A.V., Black M., Vaughn M., Dehla N., McCombie W.R., Lavine K., Mittal V., May B., Kasschau K.D., *et al.* Role of transposable elements in heterochromatin and epigenetic control. *Nature*. 2004; 430 : 471-6.

Lisch D. Epigenetic regulation of transposable elements in plants. *Annu. Rev. Plant Biol.* 2009; 60 : 43-66.

Lisch D., Slotkin R.K. Strategies for silencing and escape : the ancient struggle between transposable elements and their hosts. *Int. Rev. Cell Mol. Biol.* 2011; 292 : 119-52.

Lister R., O'Malley R.C., Tonti-Filippini J., Gregory B.D., Berry C.C., Millar A.H., Ecker J.R. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell*. 2008; 133 : 523-36.

Llorens C., Futami R., Bezemer D., Moya A. The Gypsy Database (GyDB) of mobile genetic elements. *Nucleic Acids Res.* 2008; 36 : 38-46.

Lung B., Zemann A., Madej M.J., Schuelke M., Tschritz S., Ruf S., Bock R., Hüttenhofer A. Identification of small non-coding RNAs from mitochondria and chloroplasts. *Nuc. Acids Res.* 2006; 34 : 3842-52.

Ma J., Devos K.M., Bennetzen J.L. Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genome Res.* 2004; 14 : 860-9.

Mao L., Wood T.C., Yu Y., Budiman M.A., Tomkins J., Woo S., Sasinowski M., Presting G., Frisch D., Goff S., *et al.* Rice Transposable Elements : A Survey of 73,000 Sequence-Tagged-Connectors. *Genome Res.* 2000; 10 : 982-90.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.-J., Chen, Z., *et al.* Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. 2005; 437 : 376–380.

Marí-Ordóñez A., Marchais A., Etcheverry M., Martin A., Colot V., Voinnet O. Reconstructing *de novo* silencing of an active plant retrotransposon. *Nature Genet.* 2013; 45 : 1029-39.

Matsunaga W., Kobayashi A., Kato A., Ito H. The effects of heat induction and the siRNA biogenesis pathway on the transgenerational transposition of ONSEN, a copia-like retrotransposon in Arabidopsis thaliana. *Plant Cell Physiol.* 2012; 53 : 824-33.

- Matsunaga W., Ohama N., Tanabe N., Masuta Y., Masuda S., Mitani N., Yamaguchi-Shinozaki K., Ma J.F., Kato A., Ito H. A small RNA mediated regulation of a stress-activated retrotransposon and the tissue specific transposition during the reproductive period in *Arabidopsis*. *Front. Plant. Sci.* 2015; 6 : 48.
- Matzke M. A., Mosher R. A. RNA-directed DNA methylation : an epigenetic pathway of increasing complexity. *Nature Rev.* 2014; 15 : 394-108.
- Mazoyer S. Genomic rearrangements in the BRCA1 and BRCA2 genes. *Hum Mutat.* 2005; 25 : 415-22.
- McClintock B. The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. USA.* 1950; 36 : 344-55.
- McClintock B. The significance of responses of the genome to challenge. *Science.* 1984; 226 : 792-801.
- Medvedev P., Stanciu M., Brudno M. Computational methods for discovering structural variation with next-generation sequencing. *Nat. Methods.* 2009; 6 :S13–20.
- Messing J., Bharti AK., Karlowski W.M., Gundlach H., Kim H.R., Yu Y., Wei F., Fuks G., Soderlund C.A., Mayer K.F.X., Wing R.A. Sequence composition and genome organization of maize. *Proc. Natl. Acad. Sci. USA.* 2004; 101 : 14349-54.
- Messing J., Bennetzen J.L. Grass genome structure and evolution. *Gen. Dyn.* 2008; 4 : 41-56.
- Mhiri C., Morel J.B., Vernhettes S., Casacuberta J.M., Lucas H., Grandbastien M.A. The promoter of the tobacco Tnt1 retrotransposon is induced by wounding and by abiotic stress. *Plant Mol. Biol.* 1997; 33 : 257-66.
- Miki Y., Nishisho I., Horii A., Miyoshi Y., Utsunomiya J., Kinzler K.W., Vogelstein B., Nakamura Y. Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Res.* 1992; 52 : 643-45.
- Miller N.A., Farrow E.G., Gibson M., Willig L.K., Twist G., Yoo B., Marrs T., Corder S., Krivohlavek L., Walter A. *et al.* A 26-hour system of highly sensitive whole genome sequencing for emergency management of genetic diseases. *Genome Med.* 2015; 7 : 100.
- Mirouze M., Reinders J., Bucher E., Nishimura T., Schneeberger K., Ossowski S., Cao J., Weigel D., Paszkowski J., Mathieu O. Selective epigenetic control of retrotransposition in *Arabidopsis*. *Nature.* 2009; 427-30.
- Miura A., Yonebayashi S., Watanabe K., Toyama T., Shimada H., and Kakutani T. Mobilization of transposons by a mutation abolishing full dna methylation in *arabidopsis*. *Nature.* 2001; 411 : 212-4.
- Miyao A., Tanaka K., Murata K., Sawaki H., Takeda S., Abe K., Shinozuka Y., Onosato K., Hirochika H. Target site specificity of the Tos17 retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. *Plant Cell.* 2003; 15 : 1771-80.

- Miyao A., Iwasaki Y., Kitano H., Itoh J-I, Maekawa M., Murata K., Yatou O., Nagato Y., Hiroshima H. A large-scale collection of phenotypic data describing an insertional mutant population to facilitate functional analysis of rice genes. *Plant Mol. Biol.* 2007; 63 : 625-35.
- Munroe D.J., Harris T.J.R. Third-generation sequencing fireworks at Marco Island. *Nat. Biotech.* 2010; 28 : 426-8.
- Muotri A.R., Chu V.T., Marchetto M.C.N., Deng W., Moran J.V., Gage F.H. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature.* 2005; 435 : 903-10.
- Naito K., Cho E., Yang G., Campbell M.A., Yano K., Okumoto Y., Tanisaka T., Wessler S.R. Dramatic amplification of a rice transposable element during recent domestication. *Proc. Natl. Acad. Sci. USA.* 2006; 103 : 17620-5.
- Naito K., Zhang F., Tsukiyama T., Saito H., Hancock C.N., Richardson A.O., Okumoto Y., Tanisaka T., Wessler S.R. Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature.* 2009; 461 : 1130-4.
- Nakazaki T., Okumoto Y., Horibata A., Yamahira S., Teraishi M., Nishida H., Inoue H., Tanisaka T. Mobilization of a transposon in the rice genome. *Nature.* 2003; 421 : 170-2.
- Nuthikattu S., McCue A.D., Panda K., Fultz D., DeFraia C., Thomas E.N., Slotkin R.K. The initiation of epigenetic silencing of active transposable elements is triggered by RDR6 and 21-22 nucleotides small interfering RNAs. *Plant Physiol.* 2013; 162 : 116-31.
- Ohno S. So much «junk» DNA in our genome. *Brookhaven Symp. Biol.* 1972; 23 : 366-70.
- Ohtsu K., Smith M.B., Emrich S.J., Borsuk L.A., Zhou R., Chen T., Zhang X., Timmermans M.C., Beck J., Buckner B., *et al.* Global gene expression analysis of the shoot apical meristem of maize (*Zea mays* L.). *Plant J.* 2007; 52 : 391-404.
- Onodera Y., Haag J.R., Ream T., Costa Nunes P., Pontes O., Pikaard C.S. Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. *Cell.* 2005; 120 : 613-22.
- Ong-Abdullah M., Ordway J.M., Jiang N., Ooi S.E., Kok S.Y., Sarpan N., Azimi N., Hashim A.T., Ishak Z., Rosli S.K. *et al.* Loss of Karma transposon methylation underlies the mantled somaclonal variant of oil palm. *Nature.* 2015; 525 : 533-7.
- Paterson A.H., Wendel J.F., Gundlach H., Guo H., Jenkins J., Jin D., Llewellyn D., Showmaker K.C., Shu S., Udall J., *et al.* Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature.* 2012; 492 : 423-7.
- Pecinka A., Dinh H.Q., Baubec T., Rosa M., Lettner N., Mittelsten Scheid O. Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in Arabidopsis. *The Plant cell.* 2010; 22 : 3118-29.
- Picault N., Chaparro C., Piegu B., Stenger W., Formey D., Llauro C., Descombin J., Sabot F., Lasserre E., Meynard D., *et al.* Identification of an active LTR retrotransposon in rice. *Plant J.* 2009; 58 : 754-65.

- Piegu B., Guyot R., Picault N., Roulin A., Saniyal A., Kim H., Collura K., Brar D.S., Jackson S., Wing R.A., *et al.* Doubling genome size without polyploidization : Dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome Res.* 2006; 16 : 1262-9.
- Piffanelli P., Droc G., Mieulet D., Lanau N., Bès M., Bourgeois E., Rouvière C., Gavory F., Cruaud C., Ghesquière A., Guiderdoni E. Large-scale characterization of *Tos17* insertion sites in a rice T-DNA mutant library. *Plant Mol. Biol.* 2007; 65 : 587-601.
- Pontier D., Yahubyan G., Vega D., Bulski A., Saez-Vasquez J., Hakimi M-A., Lerbs-Mache S., Colot V., Lagrange T. Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in *Arabidopsis*. *Genes Dev.* 2005; 19 : 2030-40.
- Pouteau S., Kuttner E., Grandbastien M.A., Caboche M. Specific expression of the *Tnt1* retrotransposon in protoplasts. *EMBO J.* 1991; 10 : 1911-18.
- Reddy M. A., Natarajan R. Epigenetic mechanisms in diabetic vascular complications. *Cardiovasc. Res.* 2011; 90 : 421-9.
- Reinders J., Wulff B. B., Mirouze M., Mari-Ordonez A., Dapp M., Rozhon W., Bucher E., Theiler G., Paszkowski J. Compromised stability of DNA methylation and transposon immobilization in mosaic *Arabidopsis* epigenomes. *Genes Dev.* 2009; 23 : 939-50.
- Rigal M., Mathieu O. A « mille-feuille » of silencing : epigenetic control of transposable elements. *Biochim. Biophys. Acta.* 2011; 1809 : 452-8.
- Roldán-Arjona T., Ariza R.R. Repair and tolerance of oxidative DNA damage in plants. *Mutat. Res.* 2009; 681 : 169-79.
- Rothberg J.M., Leamon J.H. The development and impact of 454 sequencing. *Nat. Biotechnol.* 2008; 26 : 1117-24.
- Roudier F., Teixeira F.K., Colot V. Chromatin indexing in *Arabidopsis* : an epigenomic tale of tails and more. *Trends Genet.* 2009; 25 : 511-7.
- Roudier F., Ahmed I., Berard C., Sarazin A., Mary-Huard T., Cortijo S., Bouyer D., Caillieux E., Duvernois-Berthet E., Al-Shikhley L., *et al.* Integrative epigenomic mapping defines four main chromatin states in *Arabidopsis*. *EMBO J.* 2011; 30 : 1928-38.
- Sabot F., Picault N., El-Baidouri M., Llauro C., Chaparro C., Piegu B., Roulin A., Guiderdoni E., Delabastide M., McCombie R., Panaud O. Transpositional landscape of the rice genome revealed by paired-end mapping of high-throughput re-sequencing data. *Plant J.* 2011; 66 : 241-6.
- Sabot F. *Tos17* rice element : Incomplete but effective. *Mob. DNA.* 2014; 5 : 10.
- Sabot F., Schulman A.H. Parasitism and the retrotransposon life cycle in plants : a hitchhiker's guide to the genome. *Heredity.* 2006; 97 : 381-8.
- Sanger F., Nickten S., Coulson A.R. DNA sequencing with chain-terminating inhibitors. 1977. *Biotech.* 1992; 24 : 104-8.

- SanMiguel P., Gaut B.S., Tikhonov A., Nakajima Y., Bennetzen J.L. The paleontology of intergene retrotransposons of maize. *Nat. Genet.* 1998; 20 : 43-5.
- Saze H., Kakutani T. Differentiation of epigenetic modifications between transposons and genes." *Curr. Opin. Plant Biol.* 2011; 14 : 81-7.
- Schnable P.S., Ware D., Fulton R.S., Stein J.C., Wei F., Pasternak S., Liang C., Zhang J., Fulton L., Graves T.A. *et al.* The B73 Maize Genome : Complexity, Diversity, and Dynamics. *Science.* 2009; 326 : 1112-5.
- Schrader L., Kim J.W., Ence D., Zimin A., Klein A., Wyschetzki K., Weichselgartner T., Kemena C., Stökl J., Schultner E., *et al.* Transposable element islands facilitate adaptation to novel environments in an invasive species. *Nat. Comm.* 2014; 5 : 5495.
- Shook M.S., Richards E.J. VIM proteins regulate transcription exclusively through the MET1 cytosine methylation pathway. *Epigenetics.* 2014; 9 : 980-6.
- Singer T., Yordan C., Martienssen R.A. Robertson's *Mutator* transposons in *A. thaliana* are regulated by the chromatin-remodeling gene *Decrease in DNA Methylation (DDM1)*. *Genes Dev.* R.A. 2001; 15 : 591-602.
- Singer T., McConnell M.J., Marchetto M.C.N., Coufal N.G., Gage F.H. LINE-1 retrotransposons : mediators of somatic variation in neuronal genomes? *Trends in Neurosciences.* 2010; 33 : 345-54.
- Slotkin R.K., Martienssen R. Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* 2007; 8 : 272-85.
- Soltis D.E., Soltis P.S. Polyploidy : recurrent formation and genome evolution. *Trends Ecol. Evol.* 1999; 14 : 348-52.
- Soltis D.E., Albert V.A., Leebens-Mack J., Bell C.D., Paterson A.H., Zheng C., Sankoff D., dePamphilis C.W., Wall P.K., Soltis P.S. Polyploidy and angiosperm diversification. *Am. J. Bot.* 2009; 96 : 336-48.
- Stroud H., Ding B., Simon S.A., Feng S., Bellizzi M., Pellegrini M., Wang G.L., Meyers B.C., Jacobsen S.E. Plants regenerated from tissue culture contain stable epigenome changes in rice. *Elife.* 2013; 2 : e00354.
- Swift H. The constancy of deoxyribose nucleic acid in plant nuclei. *Proc. Natl. Acad. Sci. USA.* 1950; 36 : 643-54.
- Takeda S., Sugimoto K., Otsuki H., Hirochika H. A 13-bp *cis*-regulatory element in the LTR promoter of the tobacco retrotransposon *Tto1* is involved in responsiveness to tissue culture, wounding, methyl jasmonate and fungal elicitors. *Plant J.* 1999; 18 : 383-93.
- Tenaillon M.I., Hollister J.D., Gaut B.S. Plant transposable elements in three dimensions. *Trends Plant Sci.* 2010;15 : 471-8.

- The 1001 Genomes Consortium. 1135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell*. 2016; 166 : 481-91.
- Thomas C.A. The genetic organization of chromosomes. *Ann. Rev. Genet.* 1971; 5 : 237-56.
- Tittel-Elmer M., Bucher E., Broger L., Mathieu O., Paszkowski J., Vaillant I. Stress-induced activation of heterochromatic transcription. *PLoS Genet.* 2010; 6 : e1001175.
- Trapnell C., Roberts A., Goff L., Pertea G., Kim D., Kelley D.R., Pimentel H., Salzberg S.L., Rinn J.L. and Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks). *Nat. Protoc.* 7(3) : 562–578.
- Tsukahara S., Kobayashi A., Kawabe A., Mathieu O., Miura A., Kakutani T. Bursts of retrotransposition reproduced in *Arabidopsis*. *Nature*. 2009; 461 : 423-6.
- Valouev A., Johnson D.S., Sundquist A., Medina C., Anton E., Batzoglou S., Myers R.M., Sidow A. Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. *Nat. Methods*. 2008; 5 : 829-34.
- Van't Hof A.E., Campagne P., Rigden D.J., Yung C.J., Lingley J., Quail M.A., Hall N., Darby A.C., Saccheri I.J. The industrial melanism mutation in British peppered moths is a transposable element. *Nature*. 2016; 534 : 102-5.
- Vendrely R., Vendrely C. La teneur du noyau cellulaire en acide désoxyribonucléique à travers les organes, les individus et les espèces animales. *Experientia*. 1948; 4 : 434-6.
- Vicient C.M. Transcriptional activity of transposable elements in maize. *BMC Genomics*. 2010; 11 : 601.
- Vitte C., Bennetzen J.L. Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution. *Proc. Natl. Acad. Sci. USA*. 2006; 103 : 17638-43.
- Vitte C., Panaud O. LTR retrotransposons and flowering plant genome size : Emergence of the increase/decrease model. *Cytogenet Genome Res.* 2005; 110 : 91–107.
- Vitte C., Ishii T., Lamy F., Brar D., Panaud O. Genomic paleontology provides evidence for two distinct origins of Asian rice (*Oryza sativa* L.). *Mol. Genet. Genom.* 2004; 272 : 504-11.
- Vitte C., Panaud O., Quesneville H. LTR retrotransposons in rice (*Oryza sativa*, L.) : Recent burst amplifications followed by rapid DNA loss. *BMC Genomics*. 2007; 8 : 218.
- Wang X., Shi X., Hao B., Ge S., Luo J. Duplication and DNA segmental loss in the rice genome : implications for diploidization. *New Phytol.* 2005; 165 : 937-46.
- Wei L., Cao X. The effect of transposable elements on phenotypic variation : insights from plants to humans. *Sri. China Life Sci.* 2016; 59 : 24-37.
- Wei L., Gu L., Song X., Cui X., Lu Z., Zhou M., Wang L., Hu F., Zhai J., Meyers B.C., Cao X. Dicer-like 3 produces transposable element-associated 24-nt siRNAs that control agricultural traits in rice. *Proc. Natl. Acad. Sci. USA*. 2014; 111 : 3877-82.

- Wendel J.F. Genome evolution in polyploids. *Plant Mol. Biol.* 2000; 42 : 225-49.
- Wessler S.R., Bureau T.E., White S.E. LTR-retrotransposons and MITEs : important players in the evolution of plant genomes. *Curr. Opin. Genet. Dev.* 1995; 5 : 814-21.
- Wessler S.R. Turned on by stress. Plant retrotransposons. *Curr. Biol.* 1996; 6 : 959-61.
- Wicker T., Sabot F., Hua-Van A., Bennetzen J.L., Capy P., Chalhoub B., Flavell A., Leroy P., Morgante M., Panaud O., Paux E., SanMiguel P., Schulman A.H. A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genetics* 2007; 8 :973-82.
- Wicker T., Keller B. Genome-wide comparative analysis of *cop* retrotransposons in Triticeae, rice, and *Arabidopsis* reveals conserved ancient evolutionary lineages and distinct dynamics of individual *cop* families. *Genome Res.* 2007; 17 : 1072-81.
- Wierzbicki A.T., Haag J.R., Pikaard C.S. Noncoding Transcription by RNA Polymerase PolIVb/Pol V Mediates Transcriptional Silencing of Overlapping and Adjacent Genes. *Cell.* 2008; 135 : 635-48.
- Wierzbicki A.T., Ream T.S., Haag J.R., Pikaard C.S. RNA polymerase V transcription guides ARGONAUTE4 to chromatin. *Nat. Genet.* 2009; 41 : 630-4.
- Witte C.P., Le Q.H., Bureau T., Kumar A. Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes. *Proc. Natl. Acad. Sci. U S A.* 2001; 98 :13778-83.
- Wolf Y., Koonin E.V. Genome reduction as the dominant mode of evolution. *Bioessays.* 2013; 35 : 829-37.
- Woo H.R., Pontes O., Pikaard C.S., Richards E.J. VIM1, a methylcytosine-binding protein required for centromeric heterochromatinization. *Genes Dev.* 2007; 21 : 267-77.
- Yamazaki M., Tsugawa H., Miyao A., Yano M., Wu J., Yamamoto S., Matsumoto T., Sasaki T., Hirochika H. The rice retortansposon Tos17 prefers low-copy-number sequences as integration targets. *Mol. Genet. Genomics.* 2001; 265 : 236-44.
- Yang G., Nagel D.H., Feschotte C.D., Hancock C.N., Wessler S.R. Tuned for transposition : molecular determinants underlying the hyperactivity of a *Stowaway* MITE. *Science.* 2009; 325 : 1391-4.
- Yang G., Zhang F., Hancock C.N., Kessler W.R. Transposition of the rice miniature inverted repeat transposable element *mPing* in *Arabidopsis thaliana*. *PNAS.* 2007; 104 : 10962-67.
- Yang D-L., Zhang G., Tang K., Li J., Yang L., Huang H., Zhang H., Zhu J-K. Dicer-independant RNA-directed DNA methylation in *Arabidopsis*. *Cell Res.* 2016; 26 : 66-82.
- Yasuda K., Ito M., Sugita T., Tsukiyama T., Saito H., Naito K., Teraishi M., Tanisaka T., Okumoto Y. Utilization of transposable element mPing as a novel genetic tool for modification of the stress response in rice. *Mol. Breed.* 2013; 32 : 505-516.

Ye K., Schulz M.H., Long Q., Apweiler R., Ning Z. Pindel : a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics*. 2009; 25 : 2865–71.

Yoon S., Xuan Z., Makarov V., Ye K., Sebat J. Sensitive and accurate detection of copy number variants using read depth of coverage. *Genome Res*. 2009; 19 :1586–92.

Zhang X., Yazaki J., Sundaresan A., Cokus S., Chan S.W., Chen H., Henderson I.R., Shinn P., Pellegrini M., Jacobsen S.E., Ecker J.R. Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell*. 2006; 126 : 1189-201.

Zhao M., Wang Q., Wang Q., Jia P., Zhao Z. Computational tools for copy number variation (CNV) detection using next-generation sequencing data : features and perspectives. *BMC Bioinformatics*. 2013; 14 Suppl 11 :S1.

Zilberman D., Gehring M., Tran R.K., Ballinger T., Henikoff S. Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nat. Genet*. 2007; 39 : 61-9.

ANNEXES

ANNEXE 1 : Conditions de culture

Conditions normales

Les plants de riz ont été cultivés en chambre de culture (Percival, USA). Pour mimer les conditions de culture retrouvées dans la nature ces plantes sont soumises à un cycle 12h/12h de lumière/obscurité. Pendant la période de luminosité, les plantes sont cultivées à 28°C et à 26°C pour la période à l'obscurité. L'humidité relative est de 80% pendant le cycle jour et de 70% pour le cycle nuit. L'intensité lumineuse varie graduellement toutes les 40 minutes de 0% à 100% au début du cycle jour et inversement au début du cycle nuit.

Nous avons obtenu les 3 générations de plantes témoins nécessaires pour l'étude et récolté une feuille de chaque plante à la floraison. Ces feuilles sont conservées à -80°C.

Conditions de stress thermique

Pour cette étude, le protocole de Xiang et associés (2013) a été adapté pour nos plantes. Ces dernières ont été cultivées dans les mêmes conditions que les plantes témoins pendant environ deux semaines jusqu'au stade 4 feuilles. Puis, les plantes ont été soumises à une élévation de température à 42°C pendant 5h (de 10h à 15h) et cela pendant 5 jours.

ANNEXE 2 : Calcul du remplissage des graines par épillet

La dispersion des graines remplies (fertiles) par épillet des différentes plantes étudiées a été représentée par des boîtes à moustaches (boxplot). Dans ce type de représentation, on représente et on interprète les valeurs des médianes et non pas des moyennes. La lecture d'une boîte à moustaches se fait comme suit :

- La valeur du 1^{er} quartile Q_1 (25% des effectifs) correspond au trait inférieur de la boîte,
- La valeur du 2^{ème} quartile Q_2 (50% des effectifs) correspond au trait intérieur de la boîte et désigne la médiane.
- La valeur du 3^{ème} quartile Q_3 (75% des effectifs) correspond au trait supérieur de la boîte,
- Les deux traits supérieurs et inférieurs, les moustaches, représentent les valeurs menant aux valeurs extrêmes respectivement la valeur la plus importante et la plus faible,
- Les valeurs dites extrêmes (outliers) situées au-delà des valeurs adjacentes sont individualisées et sont représentées par des marqueurs (carré, cercle, etc.).

ANNEXE 3 : Techniques de base en biologie moléculaire

Extraction d'ADN

L'ADN a été isolé à partir de feuilles de riz récoltées après un mois de culture. Les feuilles sont broyées dans de l'azote liquide; l'ADN est re-dissout dans le tampon d'extraction CTAB (2% CTAB, 100 mM Tris-HCl pH 8, 20 mM EDTA pH 8, 1,4 M NaCl, 2% PVP 40) et incubé à 65 °C pendant 60 min. L'extraction d'ADN est réalisée au chloroforme/isopropanol et l'élution dans le tampon TE en utilisant les inhibiteurs de RNase (RNase A, Thermoscientific). L'ADN est enfin précipité par addition de 2 volumes d'éthanol 100% et lavé à l'éthanol 70%.

Amplification par PCR et électrophorèse en gel d'agarose

Les régions d'intérêts sont amplifiées par PCR à partir d'ADN génomique à l'aide d'amorces (cf. Annexe 5). L'amplification est réalisée avec l'ADN polymérase Gotaq® de Promega (Madison, WI, USA). 40 ng d'ADN sont ajoutés aux 12,5 µl de réactifs PCR : tampon Gotaq® Green Master Mix 2X, dNTP 240 µM, amorce sens 1 µM, amorce anti-sens 1 µM, polymérase Gotaq® 0,06 unités / µl. Le programme utilisé pour amplifier l'ADN se décompose comme suit : une dénaturation préliminaire de 3 min à 94°C, suivi de 30 cycles de dénaturation (94°C, 45 sec), d'hybridation (45 sec à une température correspondant au (T_m-5°C) des amorces utilisées) et de polymérisation (72°C, 1min30). Enfin une étape d'élongation finale est effectuée pendant 5 min à 72°C.

La qualité de l'ADN est vérifiée sur un gel d'agarose 1% (Euromedex, Souffelweyersheim, France) fait dans un tampon TAE 1X (40 mM Tris ; 20 mM d'acide acétique glacial ; 1mM EDTA). Du Gel red™ 10 000X utilisé au 1/20 000 (Interchim, Montluçon, France) est inclus dans le gel. Le gel est déposé dans l'appareil d'électrophorèse dans du tampon TAE 1X pour migration. Les gels sont ensuite révélés sous U.V.

Purification d'ADN sur gel d'agarose pour séquençage

La bande correspondant au fragment à purifier est découpée du gel sur banc U.V. (300 nm). L'ADN extrait du gel est purifié sur colonne à l'aide du kit de purification GeneClean® de MP Biomedicals (Santa Ana, CA, USA) selon les recommandations du fournisseurs.

Séquençage

Pour séquencer l'ADN, la quantité d'ADN requis dépend de la taille du fragment à séquencer (Exp : 32 ng pour un fragment de 1kb). La réaction se passe en présence de 1µl d'amorce à 10 µM. Le séquençage est ensuite réalisé au sein de la plateforme du LGDP (Perpignan, France) qui utilise le séquenceur ABI3130X1 Genetic Analyzer (Applied Biosystems, ThermoScientific, Waltham, MA, USA). Les graphiques obtenus sont alors visualisés et analysés avec le logiciel Ridom TraceEdit (Ridom Bioinformatics, Würzburg, Germany).

Extraction d'ARN total et séparation par électrophorèse

Des ARNs totaux ont été isolés à partir de feuilles de riz après 1 mois de culture avec la méthode utilisant comme réactif le trizol (Invitrogen, <http://www.invitrogen.com>). Les ARNs ont ensuite subi un traitement à la DNase I (RQ1 RNase-Free DNase, Promega) en suivant les recommandations du fabricant et ont été reverse-transcrits en utilisant le kit Goscript reverse transcriptase (Promega) en utilisant les recommandations du fabricant. Les quantités des échantillons d'ADNc ont été ajustées en estimant l'expression du gène codant pour la protéine de l'actine utilisée comme contrôle interne. Les séquences des amorces sont listées dans le tableau ci-après. Le programme PCR a été réalisé comme suit : 3 min à 94°C, suivi de 30 cycles de 45 sec de dénaturation, 45 sec d'hybridation et 1 min 30 sec pour la polymérisation à 72°C avec une étape d'élongation finale de 5 min à 72°C.

Pour les amorces des gènes Os0316030 (Os3) et Os01g04380 (Os1) les conditions de PCR décrites par Ham et associés (2013) ont été suivies (cf. Annexe 5).

Traitement DNase

Afin d'éliminer toute contamination par de l'ADN génomique, les ARNs totaux extraits subissent un traitement DNase avec RQ1 (Promega, Madison, WI, USA). À 800 ng d'ARN total dans 8 µl, on ajoute 1 µl de tampon RQ1 DNase 10X et 1 µl de RQ1 RNasefree DNaseI (1 U). La digestion est effectuée pendant 30 min à 37°C. La réaction est ensuite arrêtée par addition de 1 µl de solution Stop suivie d'une incubation de 10 min à 65°C pour dénaturer la DNaseI afin d'arrêter la réaction.

Reverse transcription (RT)

La transcription inverse est réalisée avec les ARNs préalablement traités à la DNase en utilisant 1 µl d'oligodT pour un volume final de 12 µl. Le mélange est ensuite incubé pendant 5 min à 70°C puis placés pour 5 min dans de la glace. Les ARNs sont alors rétrotranscrits en ajoutant dans chaque tube 8 µl d'un mix contenant 0,5 µM dNTP, 20 U RNAsin, 100 U Goscript enzyme et 1X de tampon Goscript (Promega, Madison, WI, USA). Le mélange est ensuite incubé dans un thermocycleur pour subir 5 min d'hybridation à 25°C, une transcription inverse à 50°C pendant 60 min et enfin une étape d'incubation à 70°C pendant 15 min afin d'inactiver l'enzyme. Les produits de RT ainsi obtenus serviront pour évaluer l'expression génique.

ANNEXE 4 : Analyse haut-débit

Séquençage de génome et analyses bio-informatiques pour la détection de néo-insertions

La technologie Illumina HiSeq 2500 (San Diego, CA, USA) a été choisie pour effectuer un séquençage en paired-end. Le «paired-end mapping» permet de séquençer les deux extrémités d'un fragment d'ADN court (200-400 bases) afin d'obtenir une paire de deux séquences.

Alignement des reads séquencés et méthodes de détection des néo-insertions par le pipeline d'analyse

Dans un premier temps, la qualité des lectures issues du séquençage a été vérifiée à l'aide du programme FastQC v1.3 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Deux analyses sont menées en parallèle afin d'identifier les nouveaux événements de transposition qui se sont produits dans nos plantes.

La première de ces analyses repose sur l'étude des séquences obtenues avec le Paired-end mapping (PEM). Cette méthode de séquençage permet l'identification de paires discordantes en se basant sur l'alignement des séquences (reads) issues du séquençage sur le génome de référence alors que l'autre analyse, complémentaire à la première, se base sur la détection des lectures splittées ou split-reads, observés en cas d'alignements incomplets des reads sur notre génome de référence. Pour les deux analyses des filtres pour éliminer les séquences correspondants au génome mitochondrial et chloroplastique ont été réalisés. Ensuite pour la première analyse, un alignement avec Bowtie2 (Langmead and Salzberg, 2012) version 2.2.4 est effectué en utilisant le mode d'alignement end-to-end afin d'obtenir un alignement complet des reads et les paramètres -X 1000 (longueur maximale du fragment) et -k 1 (obtenir un seul alignement). En sortie de cette analyse, un fichier d'alignement des reads au format SAM. Seules les séquences identifiées comme des paires discordantes sont extraites du fichier SAM. Les paires discordantes correspondent aux alignements trop proches ou trop éloignés (exp. : reads d'une même paire sur un chromosome différents) pour appartenir à la même paire, elles sont identifiées grâce à un nombre binaire, le flag. Les fichiers SAM sont convertis au format binaire correspondant (fichiers BAM) grâce à l'extension Bamtools. Un autre filtre des organelles, de l'ADN ribosomique, des microsatellites, et de génome chloroplastique et mitochondrial est réalisé afin d'éliminer des séquences qui n'ont pas été identifiées lors du premier filtre. Ensuite, les séquences restantes sont alignées de nouveau sur le génome de référence, mais, cette fois le paramètre -k 20 est utilisé. En effet, dans la première partie de l'analyse, utiliser le paramètre -k 1 permet d'optimiser le temps de calcul puisque le programme recherche le premier alignement correspondant au paramètre. En récupérant les 20 autres séquences qui peuvent s'aligner sur notre génome de référence nous pouvons encore éliminer des faux positifs. Un tri a ensuite été effectué sur les position des reads appartenant à une même paire R1 et R2. Si ces reads sont éloignés de plus de 1000 pb, ils sont gardés car considérés comme une marque d'une néo-insertion. Les reads sont ensuite alignés contre le génome de référence IRGSP1, version 4 (<http://rgp.dna.affrc.go.jp/E/IRGSP/Build4/build4.html>). Seules les paires de reads qui présentent un des read de la paire avec un hit unique sur le génome et l'autre read de la paire avec au moins 2 alignements sur le génome sont récupérées. Pour chacun de ces éléments, correspondant à un breakpoint ou événement de transposition des fichiers multifasta sont créés. Dans ces fichiers, les séquences correspondent à une néo-insertion au sein d'un site unique du génome. En effet, on ne peut pas détecter les insertions qui seraient nichées dans les séquences répétées. On effectue un blast des séquences obtenues contre la base de données d'éléments transposables, la RITEDb (<http://www.genome.arizona.edu>), afin d'identifier l'élément qui se serait inséré en demandant un pourcentage d'identité de 98 %. Lorsqu'un élément a été identifié, on extrait du génome de référence, 2 kb autour de la région de l'hypothétique point d'insertion pour aligner cette séquence sur celle de l'élément transposable inséré (blast2seq). Ainsi, la présence d'un cousin de l'élément transposable identifié dans cette région peut être écartée si il n'y a aucun alignement. En sortie, des

fichiers au format bed (fichiers textes) sont créés. Ces fichiers bed contiennent toutes les coordonnées des breakpoints détectés le long des 12 chromosomes du riz, ainsi que la nature de l'élément transposable qui y serait inséré. En parallèle de cette première partie d'analyse, une deuxième méthode de détection est appliquée après le filtre sur les génomes mitochondriaux et chloroplastiques. Cette deuxième méthode utilise un autre outil d'alignement nommé segemehl version 0.2.0 (<http://www.bioinf.uni-leipzig.de>), avec les paramètres : seed = 20 (longueur de l'indel détecté). En sortie d'analyse, les séquences alignées de façon incomplètes sur le génome de référence (split-reads) sont récupérées. La présence d'une néo-insertion est établie si autour d'une région de 200 pb maximum se trouve au moins 2 split-reads. Enfin, un autre fichier au format bed est créé et il contient les coordonnées des régions où sont présents au moins 2 split-reads. L'étape finale de cette analyse utilise l'outil de Bedtools, intersect, pour croiser les coordonnées contenues dans les 2 fichiers au format bed issues des 2 analyses. Ainsi, nous obtenons à l'arrivée les coordonnées des nouveaux événements de transposition détectés dans une région unique et présentant au moins 2 split-reads.

Validation manuelle des éléments transposables

Tous les résultats obtenus sont ensuite repris de façon individuelle afin de vérifier que le programme de détection des événements de transposition n'ait pas détecté de faux-positif. Pour cela, on lie dans un fichier multifasta : une séquence de ± 2 kb autour du point d'insertion extraite à l'aide de l'outil extractseq, la séquence de l'élément identifié par le programme est récupérée sur la base de donnée RITEdb, le fichier multifasta des reads issus de la première partie du programme d'analyse (PEM) et enfin le fichier multifasta des split-reads obtenus avec segemehl. Un alignement, ou dotter, de ce fichier multifasta sur lui-même est ensuite effectué pour vérifier que nous avons bien un read de la paire dans la séquence flanquante au point d'insertion identifié et l'autre read de cette même paire dans l'élément que nous avons identifié comme ayant subi un événement de transposition (encadrés rouges). Les split-reads sont également observés dans le dotter pour observer le point exact de l'insertion (encadrés bleus), un exemple est présenté dans la figure 5.

Séquençage du transcriptome et traitement des données de séquençage

Les transcriptomes d'une plante Nipponbare cultivée en conditions normales et d'une plante mutante Osnrpd1 (23.1-15) ont été réalisés en « paired-end » avec le séquenceur Hiseq d'Illumina (San Diego, CA, USA).

Dans un premier temps la qualité des données reçues du séquençage sous forme de fichiers fastq sont vérifiées grâce au programme FastQC. Puis les lectures correspondant à des transcrits d'organelles sont filtrées en alignant les lectures sur les génomes mitochondriaux et chloroplastiques avec le logiciel Bowtie2 (<http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>). Les lectures non alignées sont récupérées et constituent les données d'entrée pour la suite de l'analyse. Ces lectures sont ensuite alignées contre le génome de référence IRGSP1.0 avec le logiciel TopHat (<http://tophat.cbcb.umb.edu/manual.shtml>). Pour un meilleur alignement et détection des transcrits, un fichier d'annotation des éléments transposables du génome au format .gtf est fourni au programme TopHat qui l'aidera à repérer les jonctions exons-exons et ainsi reconstruire les isoformes possibles à partir de ces données.

Afin de faire le comptage du nombre de lectures par locus du génome, la suite de programmes Cufflinks est utilisé (<http://cole-trapnell-lab.github.io/cufflinks>) (Trapnell et al., 2012). Pour cette étape, on part de l'hypothèse que le nombre de lectures venant d'un certain gène est proportionnel à l'abondance de son ARN dans la cellule. Pour pouvoir comparer les transcriptomes des différentes plantes séquencées, les données sont normalisées en utilisant le rpkm (read per kilobase per million), ce qui permet de prendre en compte la taille du gène ou du locus ainsi que la profondeur du séquençage. Enfin, pour l'analyse différentielle l'outil Cuffdiff est utilisé pour générer des tableaux de comptage comprenant l'annotation du gène et le nombre de lecture pour chacun des échantillons. Un locus est donné comme différentiellement avec une FDR (False of Discovery Rate)

à 5%, c'est-à-dire que la différence entre la valeur attendue et la valeur observée (en valeur absolue) dépasse ce seuil.

Langmead B., Salzberg S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods.* 2012; 9 : 357-9.

Trapnell C., Roberts A., Goff L., Pertea G., Kim D., Kelley D.R., Pimentel H., Salzberg S.L., Rinn J.L. and Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks). *Nat. Protoc.* 7(3) : 562–578.

Xiang J., Ran J., Zou J., Zhou X., Liu A., Zhang X., Peng Y., Tang N., Luo G., Chen X. 2013. Heat shock factor OsHsfB2b negatively regulates drought and salt tolerance in rice. *Plant Cell Rep.* 32(11) : 1795-1806.

ANNEXE 5 : Liste des amorces

Nom	Séquence 5'-3'	
23.1-F1	TCGTTAGGTGGGTGCAATTT	Génotypage
23.1-R1	GAGGCAGTGATTGTTCGATGA	
pol4vect-F1	AATCCAGATCCCCGAATTA	
pol4-F6	GCAGGATGGTGATCTTGTTT	RT-PCR polIV
pol4-R6	TGCAGATAGGGCAATCAGAG	
actine F	GTCTGCGATAATGGAAGTGG	
actine R	ATCTCCTTGCTCATCCTGTCAG	
Os03g16030 F (Os3)	CGTGCCGGGGCTGAAGAAGG	Efficacité stress thermique
Os03g16030 R (Os3)	TGACGGTGAGCACGCCGTTT	
Os01g04380 F (Os1)	CGGCCACCTCCGACAACGAC	
Os01g04380 R (Os1)	CCATGCCGGCCTTCACCTGG	

ANNEXE 6 : Caractéristiques des bibliothèques des petits ARNs (sRNAs)

Plante	Génotype	Nombre de séquences totales	Nombre de séquences différentes alignées sur IRGSP1.0	Nombre de lectures totales alignées sur IRGSP-1.0
NB	Wild type	42 947 842	3 989 933	33 010 927 (76,9 %)
23.1	T0 hémizygote T-DNA +/-	26 741 611	1 393 131	12 862 000 (48,1 %)
23.1-15	T1 homozygote T-DNA +/+	49 722 396	2 502 923	36 113 929 (72,6 %)
23.1-17	T1 homozygote T-DNA +/+	43 073 286	2 050 974	23 250 353 (53,9 %)
23.1-23	T1 homozygote T-DNA -/-	39 763 958	1 920 288	23 459 949 (59,0 %)

ANNEXE 7 : Nombre de lectures du séquençage d'ADN et alignées sur le génome du riz, *Oryza sativa*

Plantes	Nombre de lectures	Taille des	Couverture
GM	100 487 264	102	24
23.1	57 988 006	102	14
23.1 15	85 512 870	102	20
23.1 17	81 278 626	102	19
23.1 23	77 013 568	102	18
23.1 31	78 393 308	102	19
23.1 15Ta2	81 791 238	151	29
23.1 15Ta2.2	119 204 936	151	42
23.1 15Ta2.2.2	100 470 204	151	35
23.1 23Ta2	106 351 706	151	37
23.1 23Ta2.2	98 997 130	151	35
23.1 23Ta3	114 480 724	151	40
23.1 23Ta3.3	79 788 124	151	28
23.1 23Ta3.3.3	113 876 292	151	40
23.1 31Ta1	80 255 752	151	28
23.1 31Ta1.2	100 530 724	151	35
23.1 31Ta1.2.1	80 402 282	151	28
HST-23.1-15 Ta2.2	107 893 808	151	38
HST-23.1-23 Ta 2.2	130 103 666	151	46
HST-23.1-23 Ta3.3	106 056 128	151	37
Altg10	61 413 058	76	11
Arawak	52 770 510	101	12

ANNEXE 8 : Statistiques d'alignement des lectures sur 92 régions uniques du génome du riz, *Oryza sativa*

Plante	Moyenne	écart-type
NB	264,15	23,76
23.1	163,56	23,49
23.1-15 Ta2	201,22	25,88
23.1-15 Ta2.2	307,26	26,43
23.1-15 Ta2.2.2	254,33	24,27
23.1-23 Ta2	267,14	29,44
23.1-23 Ta2.2	255,81	24,73
23.1-23 Ta3	289,84	26,01
23.1-23 Ta3.3	201,41	23,45
23.1-23 Ta3.3.3	287,03	29,17
23.1-31 Ta1	194,33	18,51
23.1-31 Ta1.2	253,77	24,85
23.1-31 Ta1.2.1	203,12	19,62
HST-23.1-15 Ta2.2	272,03	26,24
HST-23.1-23 Ta2.2	328,18	32,68
HST-23.1-23 Ta3.3	272,00	27,31

ANNEXE 9 : Statistiques des données de séquençage au bisulfite

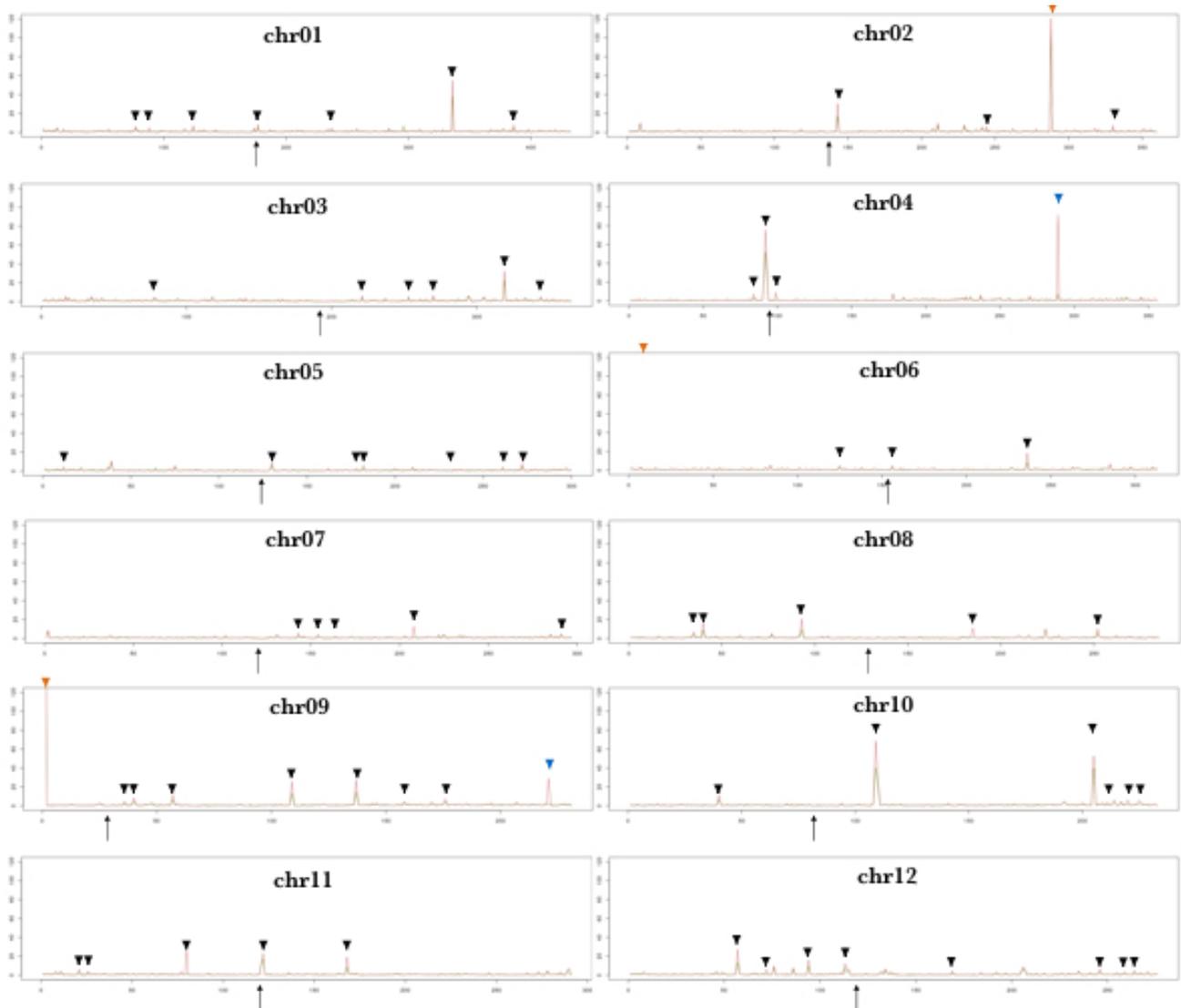
Échantillon	Nombre de lectures totales	Nombres de lectures alignées sur le génome	Alignement (%)	CG (%)	CHG (%)	CHH (%)
NB_WT	204 239 547	156 315 617	76.54	50.84	27.63	7.1
T0 23.1	185 928 414	144 471 165	77.70	54.15	25.95	5.35
T1 23.1-15Ta2	192 961 567	145 470 648	75.39	49.71	25.61	5.04
T1 23.1-23Ta3	231 780 810	179 568 611	77.47	49.61	25.66	6.07

ANNEXE 10 : Efficacité de la conversion bisulfite

Échantillon	Alignement (%)	Couverture	Nombre de Cytosine	Nombre de Cytosine + Thymine	Taux de conversion	Taux de non conversion
NB_WT	2.44	1 650.57	2 027 034	55 487 431	96.35	3.65
T0 23.1	3.61	2 204.44	2 152 163	78 209 949	97.25	2.75
T1 23.1-15Ta2	4.57	2 766.73	1 945 086	99 371 825	98.04	1.96
T1 23.1-23Ta3	3.48	2.675.23	2 471 141	88 434 034	97.21	2.79

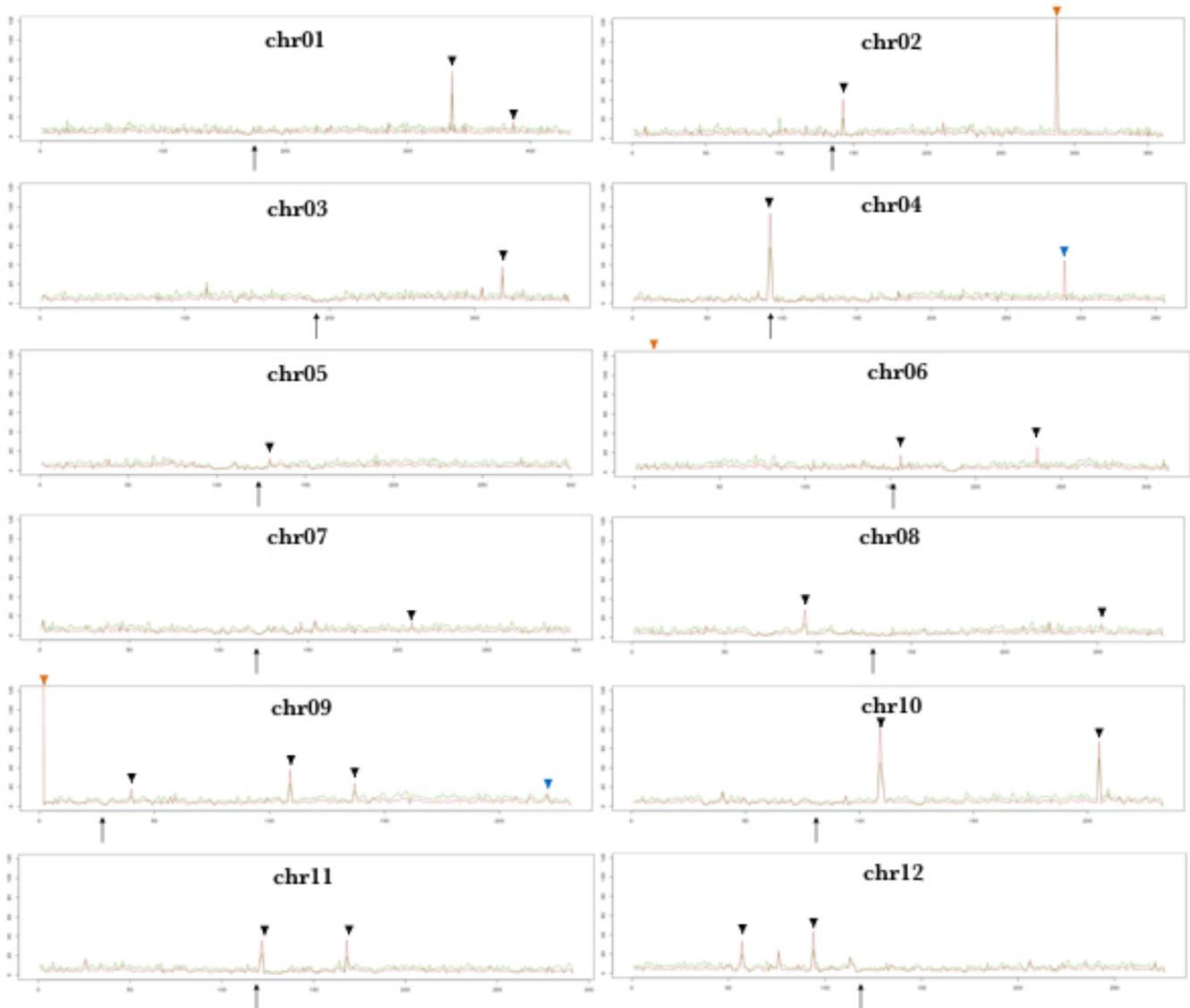
ANNEXE 11 : Distribution chromosomique des sRNAs de 21-nt sur les 12 chromosomes du riz, *Oryza sativa*

Pour chaque chromosome, la distribution des sRNAs 21-nt est tracée pour les génomes d'une plante Nipponbare (WT) en vert et d'une plante *Osmrpd1* de génération T0 (23.1) en rouge. L'axe des abscisses représente chacun des chromosomes par fenêtre de 100kb. L'axe des ordonnées représente le nombre de sRNAs en Reads Per Million (RPM). Les flèches indiquent la position des centromères. Les triangles bleus indiquent les positions des 2 gènes *OxMRPD1* du riz sur les chromosomes chr04 et chr09. Les triangles oranges désignent les régions d'ADN ribosomiques et enfin les triangles noirs montrent des positions où l'enrichissement en sRNAs 21-nt est plus important dans la plante transformée par rapport au WT.



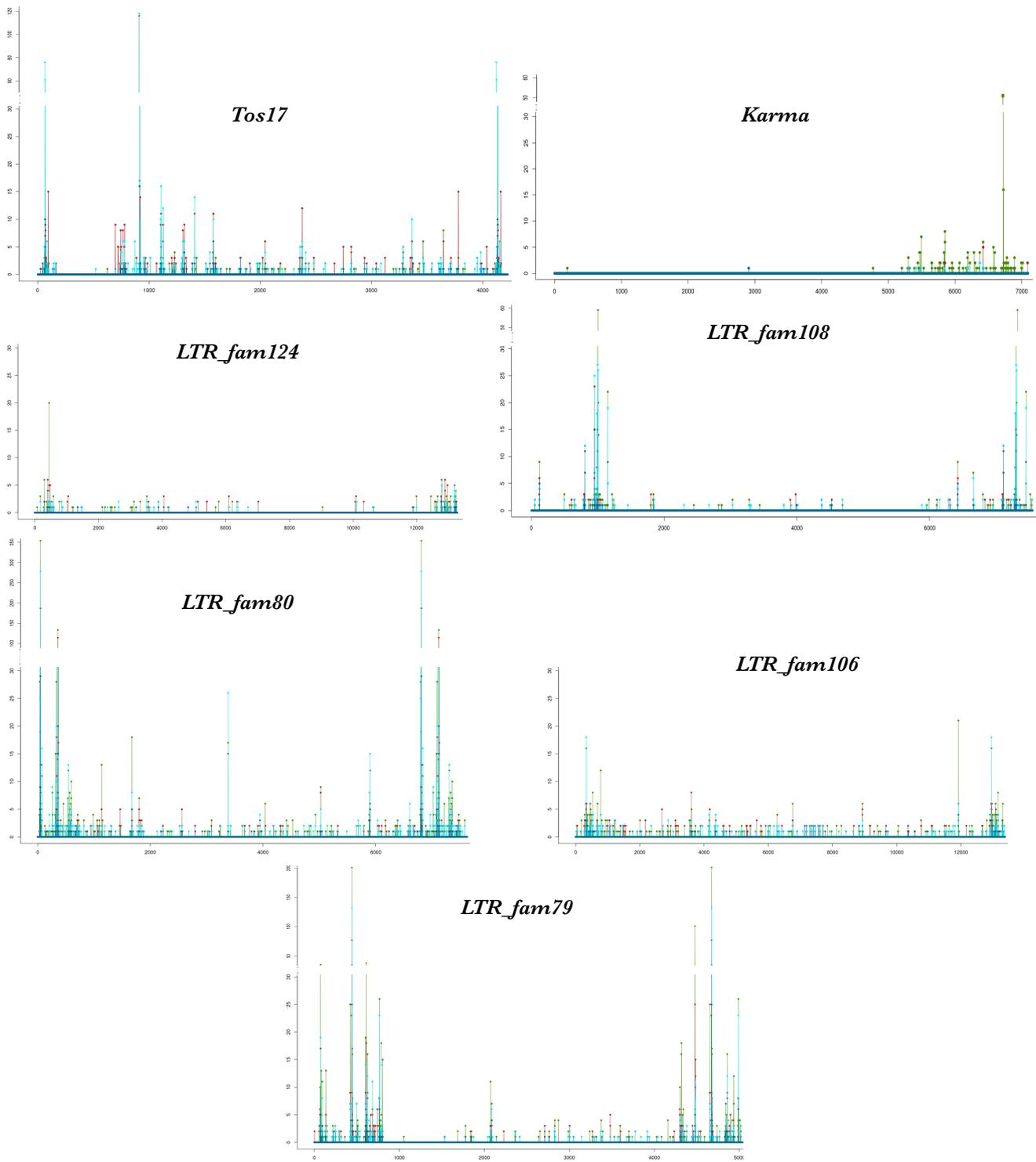
ANNEXE 12 : Distribution chromosomique des sRNAs de 24-nt sur les 12 chromosomes du riz, *Oryza sativa*

Pour chaque chromosome, la distribution des sRNAs 24-nt est tracée pour les génomes d'une plante Nipponbare (WT) en vert et d'une plante *Osmrpd1* de génération T0 (23.1) en rouge. L'axe des abscisses représente chacun des chromosomes par fenêtre de 100kb. L'axe des ordonnées représente le nombre de sRNAs en Reads Per Million (RPM). Les flèches indiquent la position des centromères. Les triangles bleus indiquent les positions des 2 gènes *O_sNRPD1* du riz sur les chromosomes chr04 et chr09. Les triangles oranges désignent les régions d'ADN ribosomiques et enfin les triangles noirs montrent des positions où l'enrichissement en sRNAs 24-nt est plus important dans la plante transformée par rapport au WT.



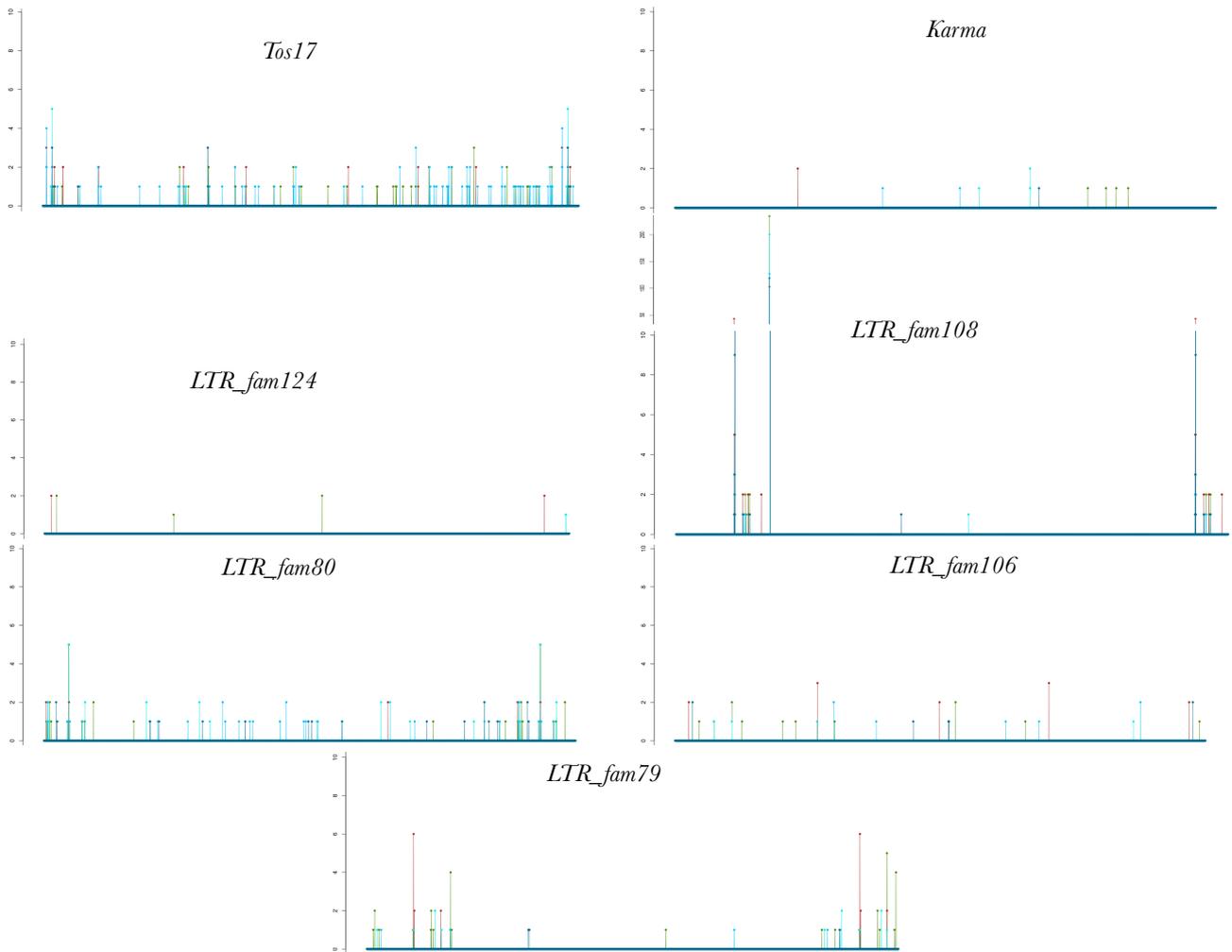
ANNEXE 13 : Distribution des sRNAs de 24-nt le long d'ETs actifs dans les plantes

Osnrpd1



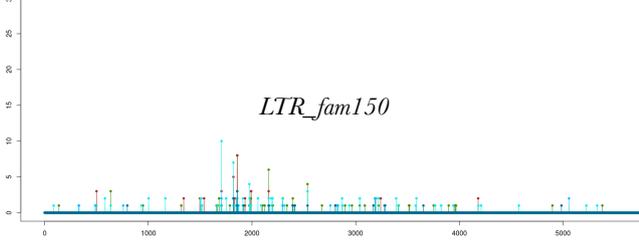
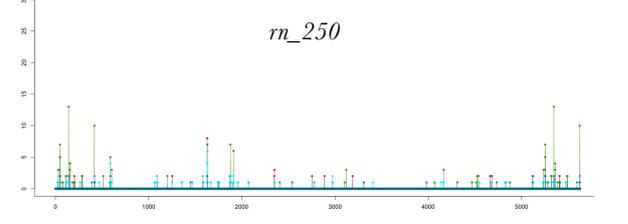
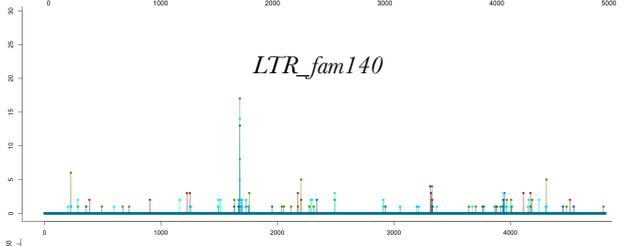
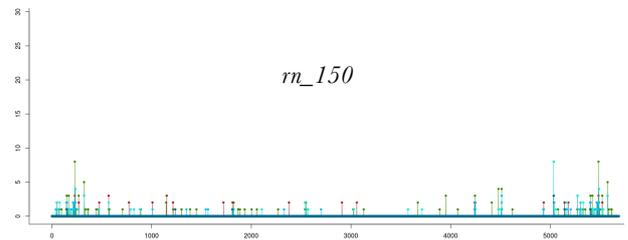
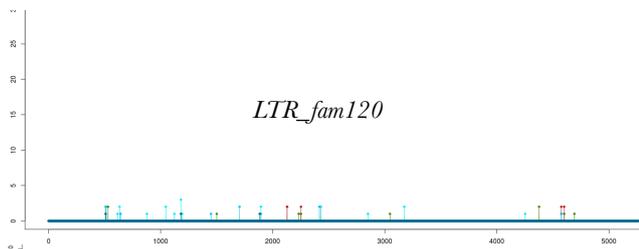
ANNEXE 14 : Distribution des sRNAs de 21-nt le long d'ETs actifs dans les plantes

Osnrpd1

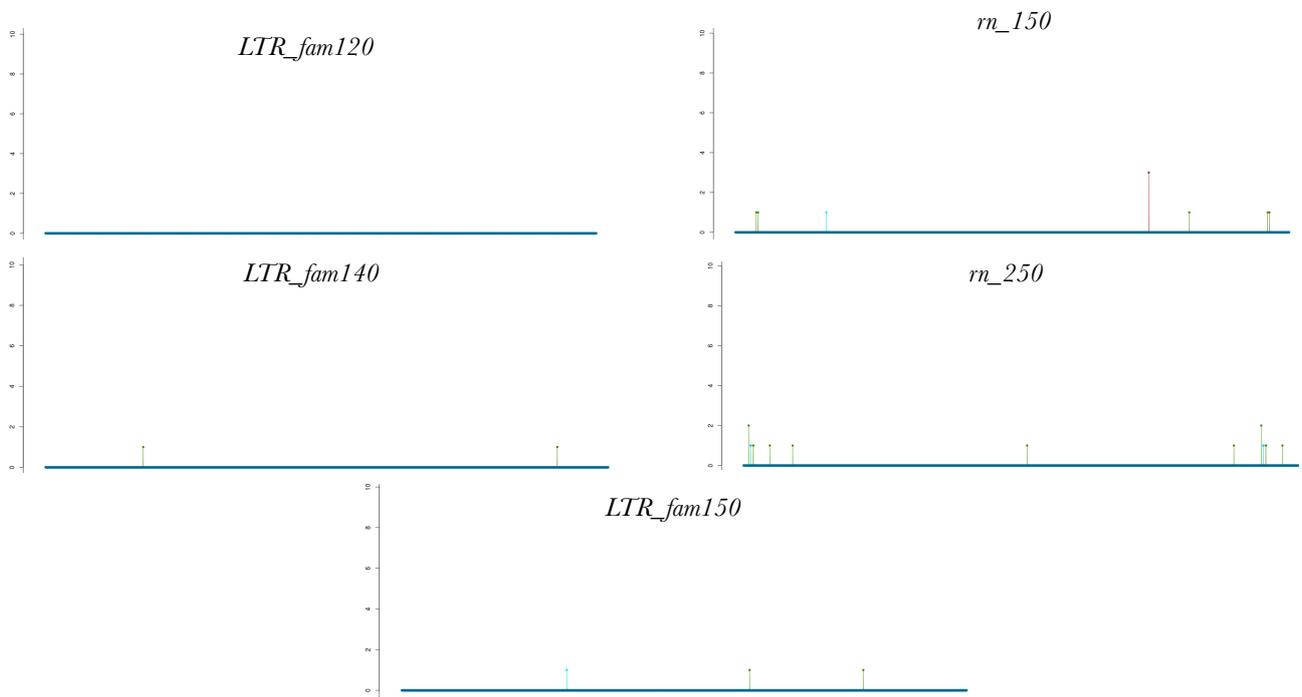


ANNEXE 15 : Distribution des sRNAs de 24-nt le long d'ETs inactifs dans les plantes

Osnrpd1



ANNEXE 16 : Distribution des sRNAs de 21-nt le long d'ETs inactifs dans les plantes *Osnrpd1*



ANNEXE 17 : Distribution des sRNAs sur *OsNRPD1* pour 3 plantes *Osnrpd1* de première génération.



RÉSUMÉ

Les éléments transposables (ETs) sont des composants ubiquitaires des génomes eucaryotes, parfois prépondérants chez les plantes. Ce sont des séquences mobiles, potentiellement mutagènes, reconnus comme des acteurs de l'évolution des génomes. Cependant, la plupart des ETs sont aujourd'hui inactifs car réprimés par des mécanismes épigénétiques très efficaces. Néanmoins, ces derniers peuvent être relâchés par des stress, conduisant à la réactivation d'ETs. De tels stress sont-ils suffisants pour activer la transposition dans les populations naturelles? L'application répétée d'un stress peut-elle expliquer les pics d'activité transpositionnelle qui ont eu lieu en conditions naturelles? De récents travaux chez un mutant d'*Arabidopsis thaliana*, affecté dans une voie de répression d'ETs, le RdDM (RNA-directed DNA Methylation), ont démontré qu'un stress thermique conduisait à la réactivation transpositionnelle d'un ET. Mes travaux de thèse portent sur l'étude de riz sauvage et d'un mutant non décrit, affecté dans le RdDM, cultivés en conditions normales ou de stress thermique sur plusieurs générations. Les objectifs de mes travaux ont été de déterminer (1) l'impact de la mutation sur les différentes étapes d'activation rétrotranspositionnelle et (2) l'activation rétrotranspositionnelle en réponse à un stress thermique. Une part importante de ce travail a été consacrée au développement et à la comparaison de méthodes d'identification des mouvements d'ETs et différentes approches « omiques » ont été utilisées. La réactivation de 5 ETs dans les plantes mutantes, dont la mobilité n'avait pas encore été observée, suggère que la voie RdDM est impliquée dans le contrôle de leur répression. De plus, nos résultats confirment que les ETs ne sont pas tous réprimés par les mêmes voies de régulation.

ABSTRACT

Transposable elements (TEs) are ubiquitous among eukaryotic genomes sometimes overriding in plants. Due to their ability to replicate and transpose, they are potentially mutagenic and recognized as actors of genome evolution. However, the analysis of the transpositional activity of TEs in different plant species have shown that most of them are maintained in a transcriptionally inactive state through powerful and specific epigenetic mechanisms. These silencing processes can nevertheless be alleviated under stress conditions, leading to TE reactivation. Are these stress sufficient to activate transposition in natural populations? Are repeated heat stress able to trigger transposition and therefore lead to bursts of transposition? In recent reports, reactivation of retrotransposons has been shown in *Arabidopsis thaliana* mutants impaired in the RdDM pathway (RNA-directed DNA Methylation) and submitted to heat stress. My PhD works reports the study of of a wild rice and a new rice mutant, affected in the RdDM, cultivated under optimal or heat stress conditions over generations. Here, we propose to determine (1) the impact of the mutation at the different levels leading to the retrotranspositional activation and (2) the retrotranspositional activity in response to heat stress. An important part of this work has been devoted to the development and the comparison of different methods to identify TE movements, and different -omics approaches have been used. The reactivation of 5 new TEs in mutants, suggests that the RdDM pathway is involved in the control of the repression of these TEs. Furthermore, our result confirm that all TEs are not regulated through the same pathways but are under the control of different lock.