



HAL
open science

L'importance du nucléole et des gènes d'ARN ribosomique 45S dans l'organisation 3D et la stabilité du génome chez *Arabidopsis thaliana*.

Ariadna Picart Picolo

► To cite this version:

Ariadna Picart Picolo. L'importance du nucléole et des gènes d'ARN ribosomique 45S dans l'organisation 3D et la stabilité du génome chez *Arabidopsis thaliana*.. Génomique, Transcriptomique et Protéomique [q-bio.GN]. Université de Perpignan, 2019. Français. NNT : 2019PERP0025 . tel-02457101

HAL Id: tel-02457101

<https://theses.hal.science/tel-02457101>

Submitted on 27 Jan 2020

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
UNIVERSITÉ DE PERPIGNAN VIA DOMITIA

Préparée au sein de l'école doctorale
« **Énergie Environnement** » ED 305 et
de l'unité de recherche LGDP-UMR 5096
CNRS-UPVD

Spécialité : Biologie

Présentée par Ariadna PICART PICOLO

**L'importance du nucléole et des gènes
d'ARN ribosomique 45S dans l'organisation
3D et la stabilité du génome chez
*Arabidopsis thaliana***

Soutenue le 5 novembre 2019 devant le jury composé de

Julio SÁEZ-VÁSQUEZ, Directeur de recherche LGDP, Université de Perpignan Via Domitia	Président de jury
Valérie GAUDIN, Directeur de recherche INRA de Versailles	Rapportrice
Aline PROBST, Directeur de recherche GReD de Clermont-Ferrand	Rapportrice
Nathalie GONZALEZ, Chargée de recherche INRA de Bordeaux	Examinatrice
Stefan GROB, Chargé de recherche Université de Zurich, Suisse	Examineur
Frédéric PONTVIANNE, Chargé de recherche LGDP, Université de Perpignan Via Domitia	Directeur de thèse

Remerciements

Tout d'abord, j'aimerais remercier **Valérie GAUDIN**, **Aline PROBST**, **Nathalie GONZALEZ**, **Stefan GROB** et **Julio SÁEZ VÁSQUEZ** pour avoir accepté de faire partie de mon jury de thèse et évaluer mon travail.

Merci **JEAN MARC DERAGON**, pour m'avoir accueillie au sein du laboratoire LGPD et pour tes cours passionnants de biologie moléculaire pendant mes années de licence.

Un énorme merci pour tout **FRED**, sans toi cette aventure n'aurait pas été possible, tu m'as fait confiance depuis le M2, alors que je me souviens encore de ta tête quand j'ai cassé mon gant en essayant de le mettre la première fois qu'on a manipé ensembles ... Concernant la paillasse, tu m'as tout appris ! J'ai acquis énormément de connaissances au cours de ces années et j'ai grandi humainement aussi, je t'en remercie. Merci également pour avoir été Fred et non pas "mon superviseur de thèse".

Merci à la **superteam MEAC**, pour vos conseils et bonne humeur au labo. Merci **NATHALIE**, pour tes cours sur les TEs, les corrections de ce manuscrit, les moments partagés autour du FACS (JAJA !) et ton soutien le long de ma thèse. Merci **GUILLAUME**, pour tes sages conseils. Merci **JULIE** et **MÉLODY** pour les moments et rires partagés.

Un plus que merci **Marie MIROUZE**, pour tes conseils, ton aide précieuse et ta bonne humeur.

Sans toi, **Martin ROSALIE**, ce manuscrit n'existerait tout simplement pas... en format \LaTeX ! Un grand merci pour ton aide, vraiment !

Merci **SYLVIE**, **DOM**, **ANNE**, **VIVIANE**, **CHRISTEL** (ou la pro du Nanopore) et **MICHELÈLE**, pour votre aide au labo et vos conseils de paillasse ! Un grand merci **ED**, pour les qPCR et pour tout ! C'était vraiment chouette de travailler avec vous tous ! Aussi, je te remercie **CLAIRE**, pour le temps que tu as passé à analyser des noyaux au microscope à cause de moi...

Merci **JEAN RENÉ**, **MYRIAM** et **ÉLISABETH** pour votre patience en ce qui concerne ma mauvaise relation avec les histoires de paperasse.

Un grand merci **Christoph GRUNAU** et **David ROQUIS**, pour votre encadrement pendant mon stage de M1, ma première et très enrichissante expérience dans un labo.

Merci **BELIN**, je pense que je ne pourrais jamais oublier tes cours de phy. veg. (même si je ne sais toujours pas si c'est un bon ou mauvais souvenir); **JEAN-JACQUES**, du PPP aux TP jusqu'à la thèse, merci pour tout; merci **OLIVIER** pour tes cours sans diapos juste énormes; merci également **VALY**, **NATACHA** et **ERIC**. C'était un plaisir de pouvoir travailler avec vous tous, en plus de vous avoir eu comme professeurs.

Une spéciale-dédicace à **MC**, quoi te dire, un merci n'est pas suffisant... donc je t'en fais plein : merci pour m'avoir appris la bioinfo et m'avoir aidé pendant ces trois années de thèse ; merci pour ton écoute, conseils et rires pendant nos réunions dans ton bureau ; merci pour tous les moments passés ensemble en dehors du labo ; merci pour ton soutien... Bref, merci d'être devenue mon amie.

Gracias a ti también **ALHELI**, por tu amistad, buena onda y por ser quien eres !

Merci aux **PILIERS DE LA BIÈRE**, fondements essentiels le long de ma thèse. Thank you **CELSE**, for your advises and our discussions, we keep in touch parce ! Gracias **ABRAHAM** por la risas en el labo. Merci pour tout **CHACHA**, faire ta connaissance et partager le bureau avec toi c'était plus que super !

Merci **ALICIA**, **PANPAN**, **MARC-ALEXANDRE** et **ÉMILIE**, pour m'accompagner dans mes pauses du cerveau, pour vos encouragements et bonne humeur.

Un merci spécial pour **ANTOINE**, **CLÉMENT**, **CAMILLE** et **MANON**, la super-équipe des Doctorants 2016-2019 ! Nous y voilà, ça y est c'est la fin, 3 ans envolés. Je n'arrive pas à y croire... Un plaisir d'avoir pu traverser cette étape avec vous !

Merci à mes parents de Perpignan, **MAX** et **CLAIRE**, pour m'avoir accueilli chez vous dans un moment clé de mon parcours académique... et surtout pour votre amitié. Un énorme merci également à mes super colocs' **FAB** et **MARTIN**, pour plein de choses, mais surtout pour m'avoir nourri et supporté pendant la période de rédaction de ce manuscrit !

Un énorme merci à mes super ex-colocs' **QUENTIN A.K.A MALO** et **CASSOU**, pour avoir fait que le chez moi du début de ma thèse soit le bonheur incarné.

J'aimerais remercier toutes les personnes auxquelles je suis entrain de penser en ce moment même (oui, feed-back nostalgique de ces trois années...), mais souci de temps. En tout cas, des millions de mercis à ma famille choisie !

Des millions de mercis aussi à ma famille non choisie, mais également extraordinaire, avec une mention spéciale pour toi, **MAMA**.

La paraula gràcies es queda curta, tot i així, gràcies, especialment pel dia que em vas dir : «*Jo si fos tu, aniria a la uni, aprofita que t'has tret el batxillerat científic i prova, i si no t'agrada, fas el grau superior...*» Tenies raó, ho havia de provar...

Table des matières

Introduction générale	1
1 Le Nucléole	3
1.1 La synthèse des ribosomes entraîne la formation du nucléole	3
1.1.1 La structure du nucléole	6
1.2 Le nucléole, un organelle multifonctionnel	7
1.2.1 Le nucléole et le métabolisme d'ARNs non ribosomiques	8
1.2.2 Le nucléole et la maturation de complexes snRNP et snoRNP	8
1.2.3 Le nucléole en tant que senseur de stress et dans la réponse au stress	9
1.2.4 Le nucléole dans la régulation du cycle cellulaire	9
1.3 Le Nucléole et l'architecture de la chromatine	10
1.3.1 Différents niveaux d'organisation chromatinienne	10
1.3.1.1 Nucléosome, modifications épigénétiques et la compaction de l'ADN	10
1.3.1.2 L'architecture tridimensionnelle de la chromatine	11
Les interactions intra- et inter- chromosomiques	11
Les Compartiments A et B du Génome et Les Domaines Chromatiniens	12
1.3.1.3 La localisation spatiale	13
1.3.2 Les Domaines chromatiniens associés au nucléole	14
1.3.2.1 La périphérie nucléaire et nucléolaire comme lieu d'ancrage de l'hétérochromatine	15
2 L'ADN ribosomique	17
2.1 Structure et localisation de l'ADNr	17
2.2 La variabilité du nombre de copies de gènes d'ADNr	22
2.3 L'instabilité de l'ADNr	24
2.4 L'ADNr et la stabilité génomique	31
2.4.1 Le nombre de copies de gènes d'ARN ribosomique n'est pas lié au besoin transcriptionnel	31
2.4.2 Importance de l'ADN ribosomique hétérochromatique dans la stabilité du génome	32
Résultats	36
3 La chromatine associée au nucléole d'<i>A. thaliana</i>	39
3.1 Contexte et objectifs	39
3.2 L'étude des NADs chez <i>Arabidopsis thaliana</i>	42

3.2.1	FANS, FANoS et analyse bioinformatique pour l'identification des NADs	42
3.2.2	Comparaison des NADs entre deux échantillons	60
3.2.3	Développement de nouvelles lignées FIB2 ::YFP pour l'étude des NADs	60
3.3	Les NADs en condition de choc thermique	62
3.4	Identification des NADs suite à la transition florale	67
3.5	Rôle des gènes d'ARNr 45S dans la composition des NADs	71
3.6	Discussion	80
4	L'ADN ribosomique et la stabilité du génome	83
	Discussion générale	137
	Matériel et Méthodes	149
6.1	Conditions de culture des plantes	151
6.2	Création de lignées transgéniques FIB2 ::YFP	152
6.3	Identification des NADs chez <i>Arabidopsis thaliana</i>	155
6.4	Visualisation de noyaux et immunolocalisation de protéines nucléaires au microscope confocal	155
6.5	Analyse de l'expression des gènes	157
6.6	Identification des régions différentiellement méthylées (DMRs)	159
6.7	Séquençage Nanopore	160
	Bibliographie	161
	Annexes	i

Table des figures

1.1	La biogenèse des ribosomes chez les eucaryotes	5
1.2	La structure du nucléole	7
2.1	L'ADNr chez les procaryotes	18
2.2	La variabilité interspécifique de l'IGS	21
2.3	Visualisation au microscope de <i>foci</i> de réparation de l'ADN	26
2.4	Réparation des DSBs	27
2.5	Recombinaison Homologue Non Allélique (NAHR)	28
2.6	Modèle d'amplification génique de l'ADNr chez la levure <i>Saccharomyces cerevisiae</i>	30
3.1	Les NADs d' <i>Arabidopsis thaliana</i>	40
3.2	Le <i>knob</i> hk4S d' <i>Arabidopsis thaliana</i> Col-0	41
3.3	Fonctionnement d'un cytomètre en flux trieur de cellules	44
3.4	Système optique du FACS BioRad S3 TM	45
3.5	La fenêtre de tri de noyaux ou nucléoles avec le cytomètre FACS BioRad S3 TM	46
3.6	Création de nouvelles lignées FIB2 ::YFP pour l'étude des NADs	61
3.7	Développement de la lignée 20%ADNr - pEG ::FIB2 ::YFP	63
3.8	Structure du nucléole d' <i>Arabidopsis thaliana</i> suite à un choc thermique	64
3.9	Les NADs d' <i>Arabidopsis thaliana</i> en condition de choc thermique	66
3.10	Décondensation des chromocentres des noyaux de cellules de feuille d' <i>A. thaliana</i> pendant la transition florale	68
3.11	Les NADs d' <i>Arabidopsis thaliana</i> suite à la transition florale	70
3.12	Comparaison des NAD-gènes entre différents échantillons	71
3.13	Comparaison entre les NADs et les PLADs	79

Abréviations

RNP	complexes ribonucléoprotéiques	KEE	KNOT Engaged Element
ARNr	ARN ribosomique	TADs	Topologically Associated Domains
PR	Protéine Ribosomique	CTCF	CCCTC binding Factor
TC	Territoire Chromosomique	CC	Chromocentre
ARNm	ARN messenger	NADs	Nucleolus-Associated Domains
NOR	Nucleolus Organizer Region	LADs	Lamina Associated Domains
IGS	Intergenic Spacer	CRWN	CROWDED NUCLEI
ITS	Internal Transcribed Spacer	NUP1	NUCLEOPORIN1
ETS	External Transcribed Spacer	PLADs	Plant-like Lamina Associated Domains
ADNr	ADN ribosomique	FISH	Fluorescent <i>In situ</i> Hybridization
FC	Fold Change	NCs	Nombre de Copies
ARN Pol I	ARN Polymérase I	IHI	Îlots Hétérochromatiques Interactifs
ARN Pol III	ARN Polymérase III	γ H2A _x	Phosphorylation ser139 H2Ax
snoRNPs	petits complexes RNP nucléolaires	HR	Homologous Recombination
snoARNs	petits complexes RNP nucléolaires	NHEJ	Non Homologous End Joining
GC	composé granulaire	NAHR	Non Allelic Homologous Recombination
DFC	composé fibrillaire dense	RBF	Replication Fork Barrier
FC	centre fibrillaire	SIR2	SIRTOUINE2
ARNt	ARN de transfert	UAF	Upstream Activating Factor
EJC	Exon Junction Complex	FLC	FLOWERING LOCUS
NoDS	pour Nucleolar Detention Signal	HDA6	Histone Déacetylase 6
HSP70	Heat Shock Protein 70	FIB2	FIBRILLARIN2
CDK	cyclin dependent Kinases	NUC1	NUCLEOLIN1
YFP	Yellow Fluorescent Protein	MMS	Méthanesulfonate de Méthyle
siRNA	short-interfering RNA	UV	lumière ultraviolette
snRNP	small nuclear RNP	CT	Choc Thermique
snARN	small nuclear ARN	TF	Transition Florale
TE	Transposable Element	FANS	Fluorecent Activated Nuclear Sorting
H2A	histone 2A	FANoS	Fluorecent Activated Nucleolar Sorting
H2B	histone 2B	CMV	CytoMegalo Virus
H3	histone 3	FACS	Fluorescent Activated Cell Sorting
H4	histone 4	SSC	Side Scatter
kb	kilobases	FSC	Forward Scatter
pb	paires de bases	EG	Early Gate
H3K4	lysine 4 de l'histone H3	Cvi-O	Cape Verde Islands
H3K36	lysine 36 de l'histone H3	Ler	Lansberg
H3K79	lysine 79 de l'histone H3	FAS1	FASCIATA1
H3K9	lysine 9 de l'histone H3	FAS2	FASCIATA2
H3K27	lysine 27 de l'histone H3	MSI1	Suppresseur Multicopie de Ira1
H4K20	lysine 20 de l'histone H3	CAF1	Chromatin Assembly Factor 1
me	méthylation	HIT4	HEAT INTOLERANT 4
ac	acétylation	HX	Index Hétérochromatique
HMT	Histone Méthyl Transférases	RHF	Relative Heterochromatin Fraction
HP1	Heterochromatin Protein 1	DAPI	6'4'-diamidino-2-phénylindole
PRC2	Complex Répressif Polycomb 2		
BRM	domaine BROMO		

Introduction générale

Chapitre 1

Le Nucléole

1.1 La synthèse des ribosomes entraîne la formation du nucléole

Les ribosomes sont des complexes ribonucléoprotéiques (RNP), constitués d'ARN ribosomiques (ARNr) et de protéines ribosomiques (PRs). Deux sous-unités (une grande et une petite) forment ce complexe, dont le rôle est de traduire les ARN messagers (ARNm) en enchaînement d'acides aminés dans les cellules de tous les organismes vivants. La traduction étant un processus vital pour la cellule, les séquences des ARNr comme des PRs ont été extrêmement conservées au cours de l'évolution.

Chez les procaryotes, la grande sous-unité 50S est composée d'ARNr 23S, ARNr 5S et de 34 PRs et la petite sous-unité 30S est constituée d'ARNr 16S et de 21 PRs. Chez les eucaryotes, la petite sous-unité 40S est constituée d'un ARNr 18S associé à 33 protéines ribosomiques. La grande 60S elle est constituée des ARNr 5S, 5.8S et 25S (chez la levure et les végétaux) ou 28S (chez les métazoaires), associés à environ 47 PRs (WEIS *et al.* 2015). La petite sous-unité (30S ou 40S) contient les sites où les ARN de transfert se fixent sur l'ARNm. La grande sous-unité contient le centre peptidyl-transférase, qui catalyse la synthèse de la liaison peptidique entre les acides aminés consécutifs de la protéine (WEIS *et al.* 2015).

Les ARNr constituent le squelette des ribosomes et sont également responsables de leur activité catalytique. En effet, ces ribozymes se replient dans l'espace pour former les sites actifs responsables de la synthèse de protéines, comme le centre peptidyl-transférase (ANDERSON *et al.* 2007). Les organismes procaryotes possèdent un ou plusieurs opérons codant pour les ARN ribosomiques 16S, 23S et 5S (KLAPPENBACH, 2001). Chez les eucaryotes, les gènes codant pour le précurseur d'ARNr 35S (chez la levure), 45S (chez les végétaux et une grande partie des métazoaires) ou 47S (chez les mammifères), sont répétés en tandem (des centaines voire des milliers de fois en fonction de

l'organisme) dans une ou plusieurs régions spécifiques du génome (GARCIA *et al.* 2014; SOCHOROVÁ, 2018). Ces régions s'appellent NORs pour «Nucleolus Organizer Regions» (McCLINTOCK, 1934).

Au sein des NORs, les gènes d'ARNr 35S/45S/47S sont séparés entre eux par un espaceur intergénique non transcrit (IGS) et possèdent la séquence codant pour les ARNr 18S, 5,8S et 25S/28S (Figure 1.1). Ces ARNr sont transcrits par l'ARN Polymérase I (ARN Pol I) en un seul précurseur polycistronique dans le nucléole (SAEZ-VASQUEZ et DELSENY, 2019; WEIS *et al.* 2015).

Au microscope électronique, les gènes d'ARNr 35S/45S/47S en cours de transcription peuvent être visualisés grâce à la présence de structures en «arbre de Noël» (MILLER et BEATTY, 1969). Le tronc de chaque arbre représente l'ADNr et les branches désignent les pré-ARNr en cours de transcription. Les sphères localisées à l'extrémité de chaque branche correspondent aux complexes de maturation et d'assemblage des sous unités ribosomiques. Les gènes d'ARNr 5S sont également organisés en séquences répétées et leur localisation au sein du génome est variable en fonction de l'espèce (GARCIA *et al.* 2014; SOCHOROVÁ, 2018). Ils sont transcrits par l'ARN polymérase III (ARN Pol III). Chez les eucaryotes, les PRs sont produites dans le cytoplasme puis relocalisées dans le nucléole (WEIS *et al.* 2015).

Les précurseurs d'ARNr 35S/45S/47S et 5S subissent des événements de clivage, des repliements et des modifications épitranscriptomiques (modifications chimiques post-transcriptionnelles des riboses, comme la méthylation ou la pseudourydilation). Ces modifications sont réalisées au niveau de sites spécifiques, très conservés au cours de l'évolution, par des petits complexes ribonucléoprotéiques nucléolaires spécialisés (snoRNPs pour small nucleolar RNPs) (revu dans : SAEZ-VASQUEZ et DELSENY, 2019; WATKINS et BOHNSACK, 2012). La spécificité des modifications épitranscriptomiques est rendue possible par l'action de petits ARNs nucléolaires (snoARNs) qui s'hybrident au niveau des sites à modifier. Grâce à cette fixation, ils guident le complexe enzymatique sur sa cible (BACHELLERIE et CAVAILLÉ, 1997; FILIPOWICZ *et al.* 1999). Par exemple, la Fibrillarine (FIB) possède l'activité catalytique responsable de la 2'-o-méthylation des riboses au sein du complexe snoRNP de type C/D du même nom (GALARDI *et al.* 2002).

Ces différentes étapes de maturation se déroulent dans un ordre précis, en parallèle avec les événements de transcription et d'assemblage avec les PRs (GRANNEMAN et BASERGA, 2005; TUROWSKI et TOLLERVEY, 2015; WEIS *et al.* 2015). Parmi les protéines importantes impliquées dans la biogenèse des ribosomes se trouve également la Nucléoline, une des protéines les plus abondantes du nucléole (ANDERSEN *et al.* 2005). Cette protéine joue un rôle majeur dans les processus de transcription et de maturation des précurseurs d'ARNr 35/45/47S, ainsi que dans l'assemblage des ARNr et des PRs en pré-sous unités

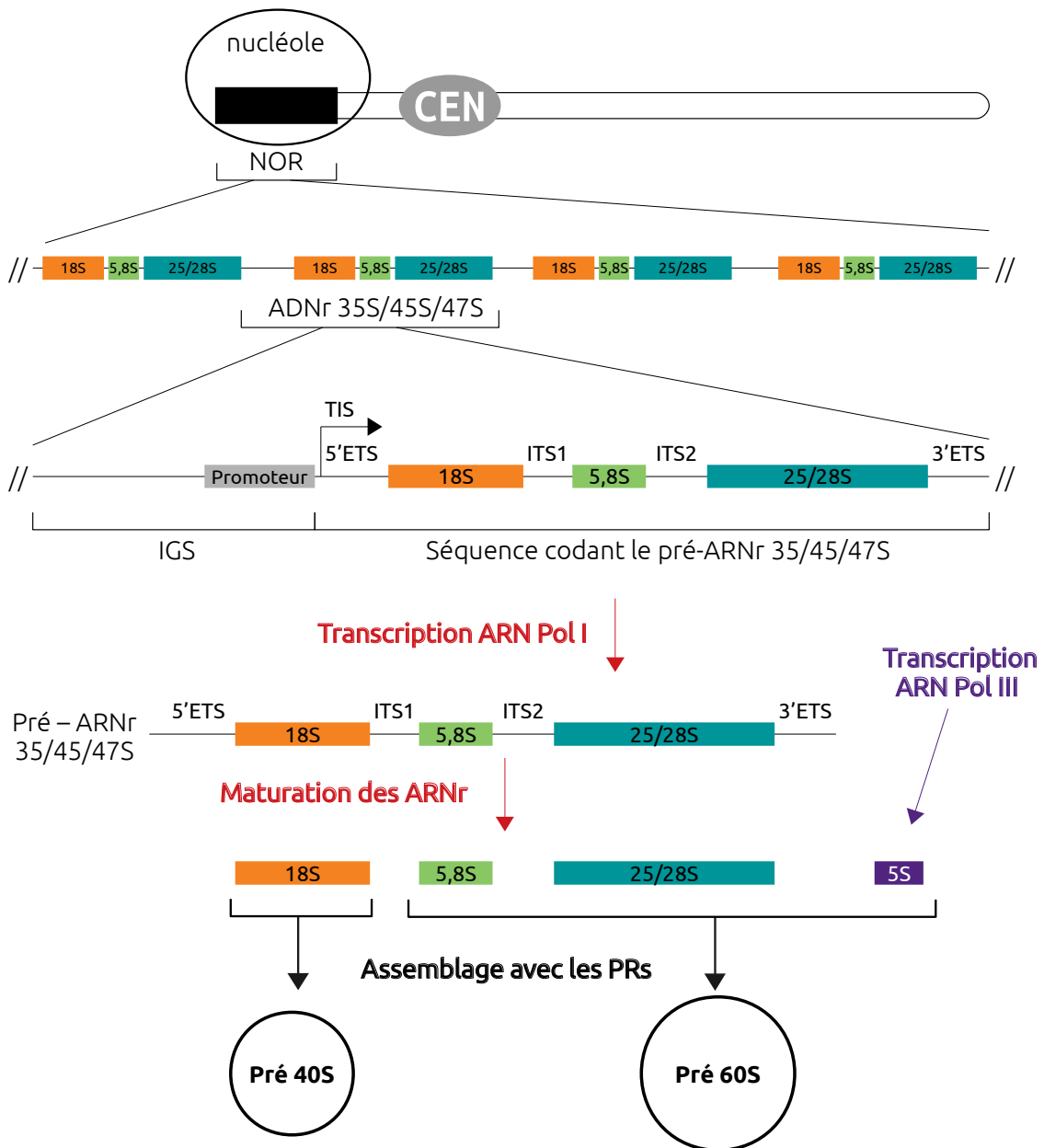


Figure 1.1 – La biogenèse des ribosomes chez les eucaryotes. Les gènes codant pour le précurseur d'ARNr 35S (chez la levure), 45S (chez les végétaux et une grande partie des métazoaires) ou 47S (chez les mammifères), sont répétés en tandem dans une ou plusieurs régions du génome appelées NORs (Nucleolus Organizer Regions). Au sein des NORs, chaque unité de répétition est constituée d'un espaceur intergénique non transcrit (l'IGS, contenant le promoteur de l'ARN Pol I) et de la séquence codant pour le pré-ARNr 35/45/47S. Le pré ARNr 35/45/47S est transcrit dans le nucléole par l'ARN Pol I et contient les ARNr 18S, 5,8S et 25S/28S. Les ARNr 5S sont transcrits par l'ARN pol III dans le noyau. Ces ARNr subissent plusieurs étapes de maturation en parallèle avec l'assemblage des protéines ribosomiques (PRs), pour former les pré sous-unités ribosomiques 40S (avec l'ARN 18S) et 60S (avec les ARNr 5S, 5,8S et 25S/28S).

ribosomiques (GINISTY *et al.* 1999). Finalement, les pré sous-unités ribosomiques (40S et 60S) sont exportées vers le cytoplasme où, après avoir subi d'ultimes modifications, elles s'assembleront autour d'un ARNm pour former un ribosome fonctionnel 80S et éventuellement initier la traduction.

La biogenèse des ribosomes est un processus très énergivore, conservé au cours de l'évolution et nécessaire à la croissance, à la prolifération et à l'homéostasie des cellules. Ce processus requiert l'intervention de plusieurs centaines de facteurs protéiques et snoRNPs, qui œuvrent ensemble autour des gènes codant pour les pré-ARNr 35/45/47S activement transcrits (MÉLÈSE et XUE, 1995). Le résultat de cette agglomération est la formation du plus grand compartiment du noyau, le nucléole.

Le nucléole est un organelle dépourvu de membrane mais qui possède des propriétés physiques et chimiques qui le séparent du reste du nucléoplasme. En effet, le nucléole fait partie des organelles cellulaires sans membrane qui sont «phase-séparés» (BRANGWYNNE *et al.* 2011), c'est à dire qui se comportent *in vivo* comme un fluide très dense et non miscible avec le reste du milieu (MITREA et KRIWACKI, 2016). La séparation de phase liquide-liquide des organelles sans membrane dépend des caractéristiques physico-chimiques des ARNs et des protéines présentes (la composition en acides aminés et le type de domaines de ces dernières). Ainsi que de la concentration totale de l'ensemble de ces facteurs (MITREA et KRIWACKI, 2016).

Dans le cas du nucléole, la machinerie de transcription des ARNr et les facteurs associés sont les composants plus importants pour son assemblage (BRANGWYNNE *et al.* 2011). En effet, la formation du nucléole est initiée par l'activation transcriptionnelle de l'ARN Pol I et l'arrêt de la transcription dissout le nucléole (BRANGWYNNE *et al.* 2011 ; MITREA et KRIWACKI, 2016). La Fibrillarine et la Nucléoline sont aussi importantes pour cette séparation de phase liquide-liquide (MITREA et KRIWACKI, 2016). Ainsi, l'intégrité structurelle du nucléole dépend des interactions entre l'ensemble de protéines et d'acides nucléiques impliqués dans la biogenèse des ribosomes et de leurs concentrations. Par conséquent, la taille du ou des nucléoles dans le noyau est étroitement liée aux besoins cellulaires en traduction (MITREA et KRIWACKI, 2016).

1.1.1 L'ultrastucture du nucléole reflète les étapes de biogenèse des ribosomes

Au sein du nucléole, les facteurs impliqués dans les différentes étapes de la biogenèse des ribosomes sont organisés dans l'espace en fonction de leur rôle. Cette organisation induit la formation de trois compartiments distincts : le composé granulaire (GC), le composé fibrillaire dense (DFC) et le centre fibrillaire (FC) (Figure 1.2 et STĘPIŃSKI,

2014). Cette structure interne du nucléole est due à des différences de miscibilité entre les protéines des différents compartiments. La nature tripartite du nucléole est donc également maintenue par séparation de phase liquide-liquide (FERIC *et al.* 2016).

Cette configuration tripartite reflète les grandes étapes de la biogenèse des ribosomes.

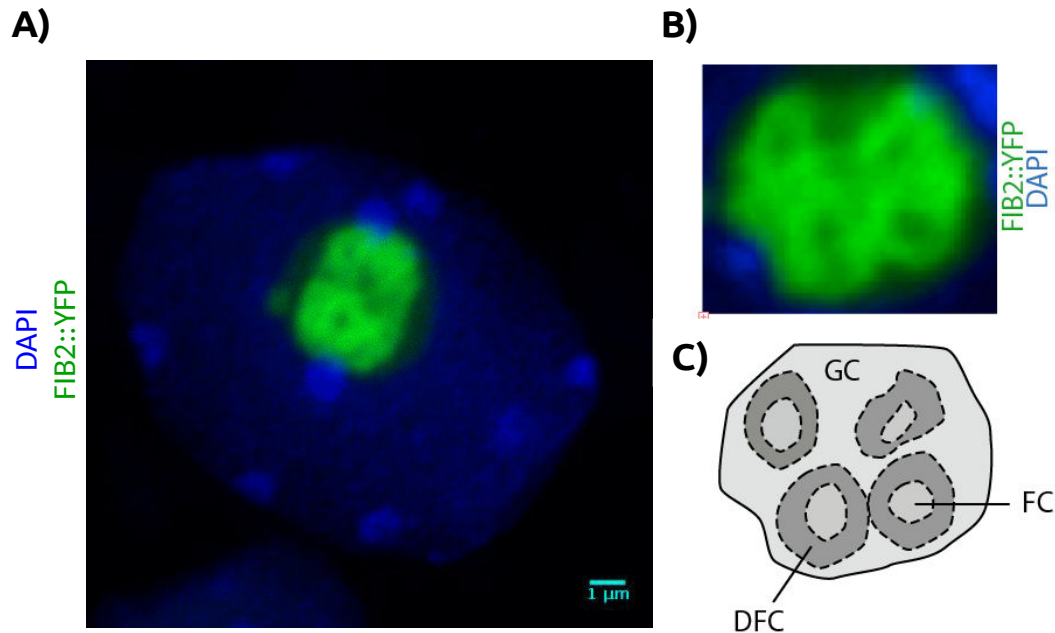


Figure 1.2 – La structure tripartite du nucléole. Observation d'un A) noyau coloré au DAPI et d'un B) nucléole au microscope confocal à épifluorescence, d'une lignée transgénétique d'*A. thaliana* qui exprime le construit 35S ::FIBRILLARINE ::YFP. La FIBRILLARINE est une protéine strictement nucléolaire. Fusionnée à la Yellow Fluorescent Protein (YFP), elle permet de marquer par fluorescence le nucléole et de distinguer les trois compartiments de ce dernier. C) Schéma de la structure tripartite du nucléole observée en B, trois compartiments sont visibles : le composé granulaire (GC), le composé fibrillaire dense (DFC) et le centre fibrillaire (FC).

L'ARN Pol I transcrit les gènes d'ARNr 35S/45S/47S entre la périphérie du FC et la zone interne du DFC. Des événements de maturation précoces des ARNr et le début de l'assemblage avec les PRs ont lieu dans le DFC. D'autres événements plus tardifs de maturation des ARNr ainsi que d'assemblage des pré sous-unités ribosomiques ont lieu dans le GC, où des structures granuleuses peuvent être observées et correspondent aux pré-ribosomes (STĘPIŃSKI, 2014 ; WEIS *et al.* 2015). Chez de nombreuses espèces, surtout végétales, un quatrième composant nommé cavité nucléolaire, dont le rôle reste méconnu, est parfois visible selon le type cellulaire observé (STĘPIŃSKI, 2014 ; WEIS *et al.* 2015).

1.2 Le nucléole, un organelle multifonctionnel

Cependant, la description du nucléole comme lieu uniquement dédié à la biogenèse des ribosomes a évolué ces dernières années. Une multitude d'études ont révélé l'implication du nucléole dans d'autres processus, comme la réponse au stress, la régulation

du cycle cellulaire ou encore le métabolisme d'ARN non ribosomiques. Cette partie va traiter quelques uns des nouveaux rôles attribués à cet organelle.

1.2.1 Le nucléole et le métabolisme d'ARNs non ribosomiques

Il a été montré qu'un grand nombre d'ARN transitent dans le nucléole avant d'être exportés vers le cytoplasme (BOISVERT *et al.* 2007; BROWN et SHAW, 2008; PEDERSON, 1998; PEDERSON et POLITZ, 2000). Parmi ceux-ci, des ARN transcrits par l'ARN Pol III, tels que les ARN de transfert (ARNt), ont été identifiés. Chez la levure, le nucléole est le lieu de transcription et d'étapes de maturation précoces des ARNt (PHIZICKY et HOPPER, 2010).

Le nucléole est également impliqué dans le métabolisme des ARNm à travers plusieurs étapes de leurs vie, *via* la maturation des complexes snRNP de l'épissage ou lors de leur dégradation (BROWN et SHAW, 2008; PENDLE *et al.* 2005). Aussi, chez *Arabidopsis thaliana* (Heynh, 1842), des protéines appartenant au complexe protéique Exon Junction Complex (EJC), ont été identifiées dans le nucléole (PENDLE *et al.* 2005). Ce complexe se forme sur la jonction entre deux exons d'un ARNm en cours d'épissage et est exporté vers le cytosol avec ce dernier. Là, il joue un rôle primordial dans la régulation post-transcriptionnelle de l'ARNm. Il est probable que cette localisation permette l'assemblage correct de l'EJC ou bien qu'elle participe à sa régulation fonctionnelle.

1.2.2 Le nucléole et la maturation de complexes snRNP et snoRNP

Le nucléole participe aussi à l'assemblage et/ou maturation de différents complexes ribonucléoprotéiques nucléaires (snRNP pour small nuclear RNP) et nucléolaires (snoRNP). Tels que le splicéosome, un snRNP intervenant dans l'épissage des ARNm, ou encore le snRNP TR/TRET ou Télomérase. La Télomérase catalyse l'addition de la séquence télomérique à l'extrémité des chromosomes, en utilisant son petit ARN nucléaire (snARN) comme matrice pour la synthèse (BOISVERT *et al.* 2007; BROWN et SHAW, 2008). De plus, chez *Arabidopsis thaliana*, les télomères sont regroupés autour du nucléole (FRANZ *et al.* 2002). Lorsque la fonction (et structure) de ce dernier est perturbée, on observe une perte d'association avec le nucléole accompagnée d'un raccourcissement des télomères (PONTVIANNE *et al.* 2016b). Ce lien pourrait donc être important dans la synthèse ou réplication des télomères chez cette espèce.

1.2.3 Le nucléole en tant que senseur de stress et dans la réponse au stress

Mis à part l'adaptation de la synthèse des ribosomes aux situations de stress cellulaire, comme son inhibition en situation de carences nutritives (BOULON *et al.* 2010), le nucléole participe aussi à la réponse à différents stress à travers d'autres mécanismes. Un mécanisme de «réretention» de certaines protéines a été identifié chez les mammifères suite à un stress thermique ou en conditions d'hypoxie. Ces stress entraînent l'accumulation de transcrits provenant des IGS des gènes d'ARNr 47S au sein du nucléole (AUDAS *et al.* 2012). Ces transcrits, d'environ 400 nucléotides, immobilisent et séquestrent dans le composé granulaire certaines protéines nucléoplasmiques possédant une séquence de détention nucléolaire appelé NoDS (pour Nucleolar Detention Signal) (AUDAS *et al.* 2012).

Ainsi, près de 200 protéines endommagées (mais de manière réversible) à cause du stress, se retrouvent au GC (AUDAS *et al.* 2012; FROTTIN *et al.* 2019). C'est le cas par exemple de l'ADN méthyltransférase DNMT1 et de protéines chaperonnes, comme notamment la HSP70 (Heat Shock Protein 70) (AUDAS *et al.* 2012; FROTTIN *et al.* 2019). Ces protéines, dont le repliement 3D est modifié du fait du stress, s'associent avec les chaperonnes, ce qui empêche leur agrégation et endommagement irréversible. Une fois le stress terminé, les protéines séquestrées reprennent leur état fonctionnel et quittent alors le nucléole (FROTTIN *et al.* 2019).

Plus récemment, des données préliminaires montrent que lors d'un choc thermique il y a également une accumulation de protéines, dont des protéines chaperonnes, au sein du nucléole chez *A. thaliana*, mais il s'agit encore de résultats très préliminaires (MONTACIE, 2019). En effet, l'étude a identifié 3189 protéines nucléolaires en condition de choc thermique, soit 5 fois plus que dans le protéome nucléolaire des plantes contrôle. Également chez les plantes, lors d'une infection par l'Alfa Mosaic Virus, une protéine virale de ce dernier est contenue dans le nucléole pour éviter le développement et la multiplication du virus dans la cellule (KALININA *et al.* 2018).

1.2.4 Le nucléole dans la régulation du cycle cellulaire

La séquestration de protéines par le nucléole est un mécanisme clé pour réguler le cycle cellulaire. Par exemple, chez la levure, la phosphatase Cdc14 inactive, est séquestrée dans le nucléole pendant l'interphase. Elle est activée lors de l'anaphase, étape durant laquelle elle quitte la région nucléolaire. Elle active alors par déphosphorylation, la protéine qui va induire la dégradation des CDK (cyclin dependent Kinases). Les CDK sont les enzymes responsables de l'entrée et du maintien de la mitose. L'activité de la phosphatase Cdc14 permet donc la sortie de mitose, et la bonne ségrégation des chromosomes

(BOISVERT *et al.* 2007).

1.3 Le Nucléole et l'architecture de la chromatine

Chez les organismes eucaryotes, l'ADN est hautement compacté sous forme de chromatine dans le noyau. La chromatine se compose d'ADN génomique nucléaire enroulé de manière cyclique à des octamères d'histones (2 H2A, 2 H2B, 2 H3 et 2 H4), cette unité de base est nommée nucléosome (LUGER et RICHMOND, 1998). La régulation du niveau de compaction et de l'organisation de la chromatine au sein du noyau est essentielle pour la stabilité du génome. En effet, la chromatine doit être agencée de façon à permettre le bon déroulement des processus liés à l'ADN au cours du cycle cellulaire, comme la réplication de l'ADN ou la transcription.

Des nombreux facteurs sont impliqués dans cette régulation qui modifie l'accessibilité aux régions régulatrices des gènes par la machinerie de transcription et ainsi leur capacité à être transcrits. Mais, l'organisation de la chromatine participe également à la régulation de la réplication, la réparation, etc. (KOUZARIDES, 2007 ; ROSA et SHAW, 2013). Cette organisation est très dynamique et peut évoluer selon le type cellulaire étudié, l'environnement ou en relation avec le rythme circadien (YEUNG et NAEF, 2018).

1.3.1 Différents niveaux d'organisation chromatinienne

1.3.1.1 Nucléosome, modifications épigénétiques et la compaction de l'ADN

Le mot épigénétique désigne les modifications chimiques de la chromatine, héréditaires au cours des cycles cellulaires (mais réversibles), qui modulent le niveau de compaction de cette dernière, et l'expression de gènes, sans modifier la séquence d'ADN. Les modifications épigénétiques régulent la condensation de la chromatine par le biais de deux mécanismes. D'une part, en induisant le rapprochement ou l'éloignement des nucléosomes entre eux. D'autre part, en recrutant des protéines non histoniques (des protéines structurales ou des enzymes) qui modifient d'avantage la chromatine.

Une des modifications épigénétiques les plus étudiées est la méthylation de l'ADN, ou plus précisément l'ajout d'un groupement méthyl à une cytosine (5-méthylcytosine ou 5-mC). Chez les mammifères, la méthylation se réalise uniquement (ou presque) dans le contexte CpG, tandis que toutes les cytosines de plantes peuvent être méthylées (CG, CHG et CHH, où H peut être un A un C ou un T) (LAW et JACOBSEN, 2010).

La méthylation de l'ADN induit le plus souvent une répression transcriptionnelle et la condensation du site méthylé. Cette marque est surtout retrouvée au niveau de séquences répétées, comme les ET ou les répétitions centromériques (SCHÜBELER, 2015).

Cependant, dans certains cas la méthylation participe à des changements d'expression de gènes impliqués dans le développement ou encore la répression transcriptionnelle tissu-spécifique de certains gènes (LAW et JACOBSEN, 2010 ; ROSA et SHAW, 2013).

Les modifications post-traductionnelles des histones sont également essentielles pour la régulation du niveau de condensation de la chromatine et la régulation transcriptionnelle chez tous les eucaryotes. Ces marques induisent des rapprochements ou des éloignements entre les nucléosomes par le biais du recrutement de protéines structurales ou d'enzymes. Ces enzymes vont soit «remodeler» les nucléosomes, soit induire des autres modifications épigénétiques.

Parmi ces modifications, la méthylation des histones est une des modifications les plus importantes biologiquement, car elle est impliquée dans nombreux processus essentiels, comme la mitose, et elle y joue des rôles majeurs (ROSA et SHAW, 2013).

La méthylation des lysines H3K4, H3K36 et H3K79, induit une décondensation de la chromatine, contrairement à la méthylation des lysines H3K9, H3K27 et H4K20 ; qui sont des sites de méthylation corrélés à la répression transcriptionnelle. Dans tous les cas, ces lysines peuvent être mono-, di- ou triméthylés, en fonction de quoi les effets sur la régulation de la chromatine seront différentes (KOUZARIDES, 2007). Les histone méthyltransférases (HMT) sont les enzymes responsables de ce transfert de groupement/s méthyle sur les lysines et également sur les arginines (PONTVIANNE *et al.* 2010).

Par exemple, la di et tri-méthylation de l'H3 de la lysine 9 (H3K9me_{2/3}), induisent la formation d'hétérochromatine en servant de site de fixation pour le recrutement de la HP1 (Heterochromatin Protein 1). Lorsque la HP1 se lie sur des H3K9me_{2/3}, elle induit la propagation et le maintien de l'état hétérochromatique (KOUZARIDES, 2007). Un autre exemple est la triméthylation de l'H3K27, réalisée par le complexe répressif Polycomb 2 (PRC2). Cette marque va induire la répression des gènes concernés qui vont par la suite être regroupés et former les domaines Polycomb (KOUZARIDES, 2007).

Chez les plantes, les promoteurs des gènes actifs sont associés à une acétylation plus élevée des lysines des histones H3 et H4, et ces modifications sont reconnues par des remodelateurs de la chromatine contenant le domaine Bromo (BRM). Cela conduit au remodelage des nucléosomes au niveau du promoteur et facilite l'accès aux machineries de la transcription (ROSA et SHAW, 2013).

1.3.1.2 L'architecture tridimensionnelle de la chromatine

Les interactions intra- et inter- chromosomiques La majorité d'interactions chromosomiques se font au sein du même chromosome. Et, dans de nombreux cas, ces interactions intrachromosomiques rapprochent une séquence régulatrice éloignée dans le chro-

mosome (enhancer) avec le promoteur d'un gène, pour activer ainsi son expression. Chez *A. thaliana* il semblerait que le locus FLOWERING LOCUS C (FLC) forme une boucle qui a un impact sur son expression (LIU *et al.* 2016).

Mais il y a également des interactions interchromosomiques. Un des exemples les plus connus, est l'interaction d'îlots hétérochromatiques interactifs (IHI) (FENG *et al.* 2014), qui forment une structure également appelée KNOT Engaged Element (KEE) (GROB *et al.* 2014).

Les Compartiments A et B du Génome et Les Domaines Chromatiniens Globalement, l'ensemble des marques épigénétiques, ainsi que les interactions inter- et intrachromosomiques, puis les protéines non histoniques (structurales ou enzymes, *e.g.* : chaperonnes, HP1, condensines, cohésines) et des ARNs non codants (NOZAWA et GILBERT, 2019) qui se lient à la chromatine, participent à la formation de domaines chromatiniens au sein d'un chromosome. Chaque domaine présente un certain niveau de condensation par rapport aux domaines adjacents. La transcription elle-même joue aussi un rôle primordial dans la formation de domaines euchromatiques (STEENSEL et FURLONG, 2019).

À une échelle plus grande, le génome est divisé en plusieurs compartiments appartenant à deux groupes, les compartiments A où la chromatine est plutôt permissive à la transcription (*i.e.* : euchromatine), et B, où la chromatine est condensée et non favorable à la transcription (*i.e.* : hétérochromatine) (DOĞAN et LIU, 2018). Par conséquent, les compartiments A et B, diffèrent grandement au niveau des leurs marques épigénétiques et niveau de condensation de la chromatine (LIEBERMAN-AIDEN *et al.* 2009).

Contrairement aux chromosomes de mammifères qui sont divisés en plusieurs compartiments A et B, les chromosomes végétaux adoptent souvent une structure simple avec deux compartiments A, correspondant aux deux bras euchromatiques et un compartiment B, au centre, correspondant au centromère, aux régions péri-centromériques et aux NORs (DONG *et al.* 2017 ; FENG *et al.* 2014). C'est le cas notamment d'*Arabidopsis thaliana*.

Les compartiments A et B sont des régions comprenant plusieurs mégabases (Mb). A une échelle plus petite, les chromosomes sont divisés en domaines topologiquement associés (TADs) ou en compartiments A et B locaux, d'environ quelques dizaines à quelques centaines de kb (DOĞAN et LIU, 2018 ; DONG *et al.* 2017).

Chez les animaux, les TADs sont des unités chromosomiques espacées par des bordures enrichies en protéines architecturales telles que la cohésine et le facteur de liaison au CCCTC (CTCF), ainsi qu'en gènes fortement exprimés (BONEV et CAVALLI, 2016). Les TADs actifs présentent moins d'interactions chromosomiques que les TADs inactifs qui sont plus condensés, ainsi que des boucles créées par le contact entre l'activateur et le promoteur de certains gènes. Une mauvaise conformation des TADs peut induire des maladies chez l'homme (LUPIÁÑEZ *et al.* 2016, 2015).

Chez les plantes, on peut noter l'absence de protéines frontalières comme la CTCF (WANG *et al.* 2015a). Cependant, des compartiments A et B locaux ont été définis chez le riz, le millet, la tomate, le maïs, le blé et l'orge (DOĞAN et LIU, 2018; DONG *et al.* 2017, 2019; LIU *et al.* 2017). Ces compartiments ne sont pas présents chez *Arabidopsis thaliana*, en effet, son petit génome ne comprend pas de régions intergéniques hétérochromatiques qui pourraient être à l'origine de la formation de compartiments B locaux au niveau des bras euchromatiques.

1.3.1.3 La localisation spatiale

En plus de l'organisation de la chromatine en domaines et compartiments présentant différents niveaux de condensation, les chromosomes sont positionnés de façon précise dans le noyau.

Chaque chromosome occupe un territoire chromosomique (TC) délimité qui ne se mélange pas avec les territoires des chromosomes voisins (CREMER *et al.* 2001; PECINKA *et al.* 2004; SZCZEPIŃSKA *et al.* 2019). Chez les animaux, la position relative des TC est beaucoup plus homogène que chez *A. thaliana*, où la localisation au sein du noyau d'un TC semble être aléatoire.

Au sein de chaque TC, le chromosome est structuré d'une façon particulière. Les chromosomes d'*A. thaliana*, adoptent une configuration globale en «rosette» (FRANZ *et al.* 2002). Dans cette configuration, le centromère de chaque chromosome et son hétérochromatine péricentromérique adjacente, sont fortement condensés et forment un chromocentre (CC). Les bras euchromatiques du chromosome forment des boucles à partir du CC et les télomères sont regroupés autour du nucléole (ARMSTRONG *et al.* 2001; FRANZ *et al.* 2002).

Les CCs sont positionnés à la périphérie du noyau, sauf celui du chromosome qui possède le NOR activement transcrit. En effet, les NORs participent à la formation du CC (FENG *et al.* 2014; FRANZ *et al.* 2002). Si le NOR est transcriptionnellement actif, le CC s'associe avec la périphérie nucléolaire, s'il est inactif le CC est positionné à la périphérie nucléaire.

Les espèces qui ont un plus grand génome, telles que, le blé, le seigle, l'orge et l'avoine, présentent souvent une configuration des chromosomes en «Rabl» (RABL, 1885), dans laquelle les centromères et les télomères sont situés aux pôles opposés des noyaux (DOĞAN et LIU, 2018; ROSA et SHAW, 2013). Ainsi, les bras chromosomiques forment des TC longitudinaux. Dans cette configuration, plus de 70% du génome est localisé dans l'hémisphère centromérique du noyau et est fortement condensé, et l'autre moitié du noyau où se trouvent les télomères, contient une chromatine plus décondensée (DOĞAN et LIU, 2018; ROSA et SHAW, 2013).

Cependant, le noyau est composé non seulement de la chromatine, mais également de nombreux corps nucléaires, qui ne possèdent pas de membrane les séparant physiquement du nucléoplasme. Ces compartiments influencent également l'organisation spatiale du génome.

Par exemple, chez l'humain et la souris, des interactions inter-chromosomiques entre 3 ou plus chromosomes (appelés *hubs* inter-chromosomiques), comprenant des gènes activement transcrits, sont localisés à la périphérie des *speckles* (QUINODOZ *et al.* 2018). Les *speckles* nucléaires (ou granules interchromatiniens) sont des sites où se concentrent de nombreux facteurs d'épissage de l'ARNm (GALGANSKI *et al.* 2017). La localisation de ces *hubs* autour des *speckles* permet sûrement d'optimiser la production de transcrits.

Parmi ces compartiments, le nucléole est le plus proéminent et possède des caractéristiques qui le rendent non miscible avec le nucléoplasme (séparation de phase liquide-liquide). Par conséquent sa présence a un impact direct dans l'organisation spatiale des chromosomes.

1.3.2 Les Domaines chromatiniens associés au nucléole

L'ADN ribosomique n'est pas le seul à être en contact avec le nucléole. En 2010, deux groupes indépendants ont montré que les 23 chromosomes humains possèdent tous au moins une région associée au nucléole. Ces régions ont été appelées NADs pour Nucleolus-Associated chromatin Domains (KONINGSBRUGGEN *et al.* 2010; NÉMETH *et al.* 2010). Plus tard, les NADs ont été identifiés chez la souris (GUETG et SANTORO, 2012a; VERTII *et al.* 2019).

Les NADs présentent un grand enrichissement en marques épigénétiques répressives de la chromatine et en séquences répétées, notamment en séquences péricentromériques. Par contre, ces domaines sont dépourvus de gènes et ceux présents, ont un faible niveau d'expression. En outre, des *hubs* inter-chromosomiques, comprenant cette fois-ci des gènes avec un faible taux de transcription, sont localisés à la périphérie du nucléole (QUINODOZ *et al.* 2018).

Le nucléole semble donc être un lieu d'ancrage pour l'hétérochromatine. Chez l'humain, une partie non négligeable des compartiments B sont associés au nucléole (DILLINGER *et al.* 2016).

Parmi les NADs identifiés chez l'humain, se trouvent des gènes codant pour des récepteurs olfactifs (NÉMETH *et al.* 2010). Les gènes codant pour les récepteurs olfactifs appartiennent à une grande famille de gènes, sont exprimés de façon tissu-spécifique et présentent également une expression monoallélique (CHESS *et al.* 1994). Comme le nu-

cléole est un environnement peu propice à la transcription de gènes par l'ARN Pol II, leur association avec le nucléole pourrait en partie participer à ces différences d'expression. Ainsi, dans les tissus où ils ne seraient pas exprimés ils s'associeraient avec le nucléole et vice-versa. Dans le cas de l'expression monoallélique, uniquement un des deux allèles serait associé.

Un grand enrichissement en gènes codant pour des ARNt a également été mis en évidence (NÉMETH *et al.* 2010). Cette localisation périnucléolaire de gènes codant pour des ARNt est cohérente avec des études précédentes qui ont montré une activité accrue de l'ARN pol III autour de la périphérie nucléolaire (THOMPSON *et al.* 2003).

1.3.2.1 La périphérie nucléaire et nucléolaire comme lieu d'ancrage de l'hétérochromatine

Il est avéré chez les métazoaires, que de larges portions du génome sont maintenues à la périphérie du noyau, par un réseau de fibres composé en grande partie de lamines nucléaires. Ces domaines ont été appelés LADs, pour Lamina-Associated chromatin Domains.

Les LADs ont été identifiés chez la drosophile (PICKERSGILL *et al.* 2006), la souris (VERTII *et al.* 2019) et l'humain (GUELEN *et al.* 2008; KIND *et al.* 2013). Comme les NADs, ces régions présentent des caractéristiques propres à l'hétérochromatine et comportent un faible nombre de gènes, qui ont tendance à être peu transcrits. En effet, de nombreux compartiments B s'associent avec les lamines (LIEBERMAN-AIDEN *et al.* 2009). Chez l'humain et la souris, les LADs représentent environ 40% du génome et leur taille varie de 40 kb à plus de 30 Mb (PICKERSGILL *et al.* 2006).

Dans les cellules végétales, les gènes codant des protéines homologues à la lamine nucléaire n'ont pas été identifiées. Cependant, certaines protéines partagent des caractéristiques avec les lamines, comme la localisation au niveau de la matrice nucléaire. Parmi ces dernières, les protéines CROWDED NUCLEI (CRWN), jouent un rôle très important dans l'architecture nucléaire. Chez le double mutant *crwn1,2*, on observe une réduction significative de la taille des noyaux et une morphologie nucléaire non différenciée (DITTMER *et al.* 2007; POULET *et al.* 2017). De plus, la structure tripartite des protéines CRWN est similaire à celle des lamines nucléaires.

Chez *A. thaliana*, les domaines associés à la périphérie du noyau ont été identifiés par immunoprécipitation de la chromatine associée aux protéines NUCLEOPORIN1 (NUP1) et CRWN. Les régions génomiques identifiées sont également des domaines hétérochromatiques (BI *et al.* 2017; HU *et al.* 2019).

De manière intéressante, il existe un certain chevauchement entre les LADs et les

NADs chez l'humain et la souris (KIND *et al.* 2013; NÉMETH *et al.* 2010; VERTII *et al.* 2019). Certaines régions hétérochromatiques peuvent s'ancrer soit à la périphérie du noyau, soit à la périphérie du nucléole (KIND *et al.* 2013; POMBO et DILLON, 2015).

Premier axe de la thèse

L'ANR NucleoReg développée par F. Pontvianne et dans laquelle s'insère mon projet de thèse, a pour but de réaliser une cartographie de la dynamique des NADs dans différents contextes, afin d'évaluer l'impact du nucléole dans l'organisation globale de la chromatine et la régulation transcriptionnelle de gènes transcrits par l'ARN pol II chez *A. thaliana*. Avant de commencer ma thèse, l'équipe a pu mettre en évidence chez *A. thaliana*, que des domaines chromatinien provenant des 5 chromosomes, étaient associés au nucléole (PONTVIANNE *et al.* 2016b). Comme chez les métazoaires, ces régions présentent des caractéristiques propres à l'hétérochromatine.

Cependant, 907 gènes font également partie des NADs et présentent un niveau d'expression transcriptionnelle plus faible comparé à l'ensemble des gènes (PONTVIANNE *et al.* 2016b). Afin de mettre en évidence une possible voie de régulation transcriptionnelle par séquestration nucléolaire, j'ai identifié les NADs en condition de choc thermique et suite à la transition florale.

Deuxièmement, pour déterminer s'il existe un chevauchement entre les domaines associés à la périphérie nucléolaire et la périphérie nucléaire chez *A. thaliana*, une analyse comparative entre les NADs et les PLADs (Plant-like Lamina Associated chromatin Domains) a été réalisée.

Chapitre 2

L'ADN ribosomique

2.1 Structure et localisation dans le génome de l'ADN ribosomique (ADNr)

Les gènes d'ARNr sont universels. Chez les procaryotes, les gènes d'ARNr sont organisés sous forme d'opérons (unité d'ADN fonctionnelle regroupant des gènes sous le contrôle d'un même promoteur). Chaque opéron code pour les trois ARNr (16S, 23S et 5S), ainsi que des ARN de transfert (Figure 2.1A et KLAPPENBACH, 2001). En fonction de l'espèce, 1 à 15 opérons sont dispersés dans le génome (KLAPPENBACH, 2001). Par exemple, *Escherichia coli* (T. Escherich, 1885) possède environ 7 opérons proches d'une origine de réplication de l'ADN (Figure 2.1B et ELLWOOD et NOMURA, 1982).

En raison de la grande demande en protéines et donc de ribosomes, le nombre de copies de gènes codant pour le pré-ARNr 35/45/47S et 5S dans les organismes plus complexes a augmenté pour amplifier le nombre de transcrits. A l'inverse, les gènes codant pour les PRs sont dans la plupart du temps en simple copie dans le génome.

Dans la majorité de cas, les gènes d'ARNr 35S/45S/47S sont séparés physiquement des clusters d'ARNr 5S (arrangement type-S). Cependant, dans moins de 5% des cas, ils sont liés dans la même unité de répétition au sein des NORs, c'est l'arrangement type-L (GARCIA *et al.* 2014; SOCHOROVÁ, 2018). Ainsi, chez *Saccharomyces cerevisiae* (E. C. Hansen, 1883), les gènes codant pour le pré ARNr 5S se trouvent en simple copie après chaque gène 35S. Tandis que chez *Schizosaccharomyces pombe* (Lindner, 1893), environ 30 copies d'ADNr 5S sont dispersées dans le génome (MAO *et al.* 1982).

Chez certaines plantes comme les *Asteroideae* (une sous-famille des Angiospermes), on retrouve fréquemment un arrangement type-L de ces gènes (GARCIA *et al.* 2010). Ou encore chez certains *Gnetales*, *Coniferales* (Gymnospermes) et chez le fossile vivant *Ginkgo biloba* (von Linné, 1771) (GARCIA et KOVAŘÍK, 2013). *G. biloba* est un Spermatophyte, seul représentant actuel de la famille des *Ginkgoaceae*, la famille d'arbres plus ancienne

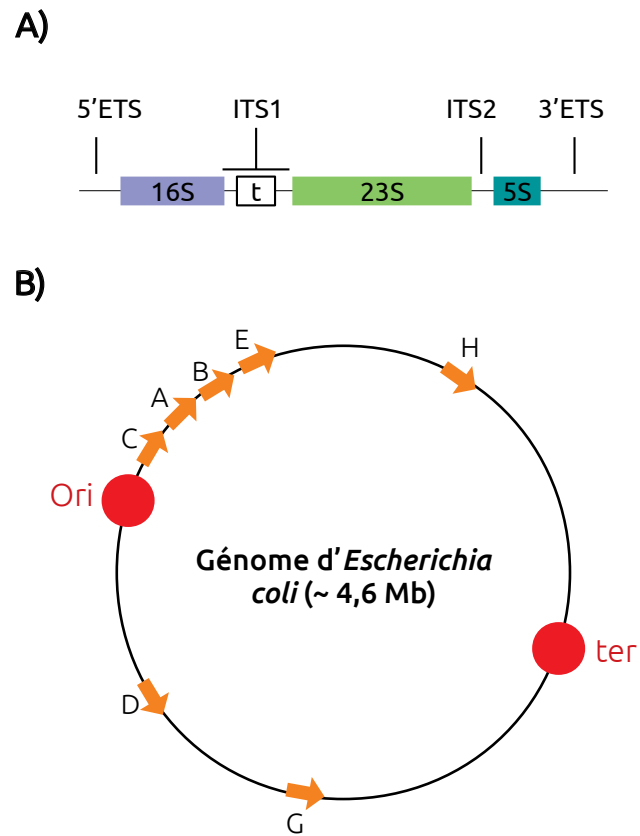


Figure 2.1 – L'ADNr chez les procaryotes, adapté de MAEDA *et al.* 2015. **A)** Chez les procaryotes, un opéron code pour les trois ARNr (16S, 23S et 5S) ainsi que pour un ou plusieurs des ARN de transfert (t). **B)** Distribution dans le génome des 7 opérons d'ARNr d'*Escherichia coli*. Chaque flèche orange représente un opéron d'ARNr. L'origine de réplication (Ori) et le site de terminaison de la réplication (ter) sont représentés en rouge.

connue, apparue il y a plus de 270 millions d'années. Ceci laisse penser qu'à la base les deux gènes étaient liés (arrangement type-L) puis ils se sont séparés au cours de l'évolution (GARCIA et KOVAŘÍK, 2013).

La structure et organisation des NORs est très conservée chez les organismes eucaryotes. Chaque unité de répétition des clusters d'ADNr est divisée en deux parties majeures (Figure 1.1). La région codante et l'espaceur intergénique non transcrit (IGS pour Internal Genomic Spacer). C'est dans la région codante, d'environ 6,9 kb (35S) chez la levure, ~10 kb chez *A. thaliana* (RABANAL *et al.* 2017b) et jusqu'à ~13 kb (47S) chez les mammifères (McSTAY, 2016), que se trouvent les séquences codant les ARNr 18S, 5.8S, et 25S/28S. Ces séquences sont séparées par deux espaceurs internes transcrits (ITS1 et ITS2 pour Internal Transcribed Spacer). Les espaceurs externes transcrits (5'ETS et 3'ETS pour External Transcribed Spacer) entourent les pré-ARNr.

En raison de la grande conservation au cours de l'évolution des séquences codant les ARNr, elles sont utilisées pour reconstruire l'histoire évolutive des espèces et les relations phylogéniques entre ces dernières, et ce chez les procaryotes comme chez les eucaryotes (GARCIA *et al.* 2014; KLAPPENBACH, 2001; SOCHOROVÁ, 2018).

En revanche, contrairement à la partie transcrite, une variabilité beaucoup plus importante en taille et séquence est observée au sein de l'IGS entre les espèces. Il peut varier de ~1 kb chez certaines espèces de plantes à ~30 kb chez les mammifères (McSTAY, 2016; ROGERS et BENDICH, 1987). Par exemple, chez l'homme (*Homo sapiens* [von Linné, 1758]), un IGS fait entre 10 et 30 kb contre 2 à 5 kb chez *A. thaliana*.

De manière générale, l'IGS contient le promoteur proximal du gène d'ARNr 35/45/47S adjacent, des séquences régulatrices de la synthèse des pré-ARNr, des promoteurs espaceurs et des séquences de liaison aux facteurs de terminaison de la transcription (McSTAY et GRUMMT, 2008; SALIM et GERTON, 2019). Chez les métazoaires, l'IGS contient également une origine de répllication de l'ADN (COEN et DOVER, 1982; GONZALEZ et SYLVESTER, 1995; GROZDANOV *et al.* 2003; GRUENDLER *et al.* 1991; KOBAYASHI, 2014; ROGERS et BENDICH, 1987; SUZUKI *et al.* 1996).

Cependant, la présence ou non de ces éléments, leur agencement, leur nombre et séquence varient beaucoup au niveau inter-spécifique (Figure 2.2). Par exemple, chez *A. thaliana*, en plus du promoteur proximal, l'IGS contient trois sites de séquences répétées, qui comportent de nombreux sites de restriction *Sall* et deux promoteurs espaceurs. Chez la levure (*Saccharomyces cerevisiae*), chaque IGS est divisé (en IGS1 et IGS2) par un gène codant pour le précurseur 5S en orientation inverse des gènes 35S (KOBAYASHI, 2014). L'IGS1 contient aussi un promoteur bidirectionnel de l'ARN pol II, qui permet la transcription d'un ARN non-codant (l'E-pro) et aucune séquence répétée n'a été identifiée (KOBAYASHI, 2014). Chez l'homme, l'IGS contient des séquences microsatellites, des

éléments transposables (*e.g.* : des éléments ALU ou LTR), un enhancer, un promoteur de transcription non codante de l'IGS et un promoteur espaceur (AGRAWAL et GANLEY, 2018; GONZALEZ et SYLVESTER, 1995; SALIM et GERTON, 2019).

Depuis 1974, la production de données concernant le nombre et la localisation de ces clusters d'ADNr n'a pas cessé d'augmenter. Ces données ont été regroupées en deux bases de données, la base de données d'ADNr de plantes (première sortie en 2011, puis une troisième version en 2016; <http://www.plantrdnadatabase.com>) et une autre concernant les animaux (première sortie en 2016; <http://www.animalrdnadatabase.com>).

Environ 50% des publications référencées ont été produites durant cette dernière décennie (GARCIA *et al.* 2014; SOCHOROVÁ, 2018). La majorité des études (~95%) concernent des analyses microscopiques de l'ADNr par FISH (Fluorescent *In Situ* Hybridization). Cette technique est très efficace pour l'ADNr en raison de son abondance, sa nature répétée et la conservation de séquence entre les espèces. En effet, il est possible de suivre par microscopie l'ADNr d'une espèce sans connaître son génome avec des sondes spécifiques pour les séquences 18S, 25S-28S, etc., car elles sont très similaires entre les espèces.

Ainsi, la base de données de plantes rassemble plus de 600 études, 2949 caryotypes de 1791 espèces de plantes, la majorité étant des Angiospermes (GARCIA *et al.* 2014). La base de données des animaux recense 1343 espèces de vertébrés et invertébrés (SOCHOROVÁ, 2018).

Au niveau interspécifique, les organismes présentent une grande diversité du nombre de clusters d'ADNr ou NOR, de 1 à plusieurs dizaines de *loci* en fonction de l'espèce (de 2 à 74 pour l'ADNr 5S et de 2 à 54 pour les sites d'ADNr 35S/45S/47S par génome diploïde). Il est à noter qu'aucune corrélation n'existe entre le nombre de *loci* d'ADNr 5S ou 35S/45S/47S et le nombre de chromosomes de l'espèce (GARCIA *et al.* 2014; GARCIA *et al.* 2017; SOCHOROVÁ, 2018). Cependant, un nombre réduit de *loci* est préféré.

En considérant les 1358 caryotypes animaux, le nombre médian de sites 45S/47S et 5S par génome diploïde est de 2 (environ 60% des caryotypes animaux ont un seul *locus* 5S ou 45S/47S par génome haploïde) (SOCHOROVÁ, 2018). Dans la plupart des végétaux, on retrouve en médiane 4 *loci* 45S par génome diploïde et 2 *loci* 5S. Le pourcentage de caryotypes avec un *locus* unique 5S étant de 40,35% tandis que 27,94% ont un seul *locus* d'ADNr 45S (GARCIA *et al.* 2014; GARCIA *et al.* 2017). *Pinus nigra* (R. Legay, 1785) avec 26 *loci* 45S et 4 5S est une des espèces au plus grand nombre de *loci*, ou encore *Alstroemeria aurea* (Graham, 1833), avec 26 *loci* 45S et 14 5S. Le nombre de *loci* d'ADNr 5S et 45S sont positivement corrélés chez les plantes, contrairement aux animaux, où aucune corrélation significative est observée (GARCIA *et al.* 2017; SOCHOROVÁ, 2018).

Les *loci* d'ADNr sont localisés sur différents chromosomes (gonosomes ou auto-

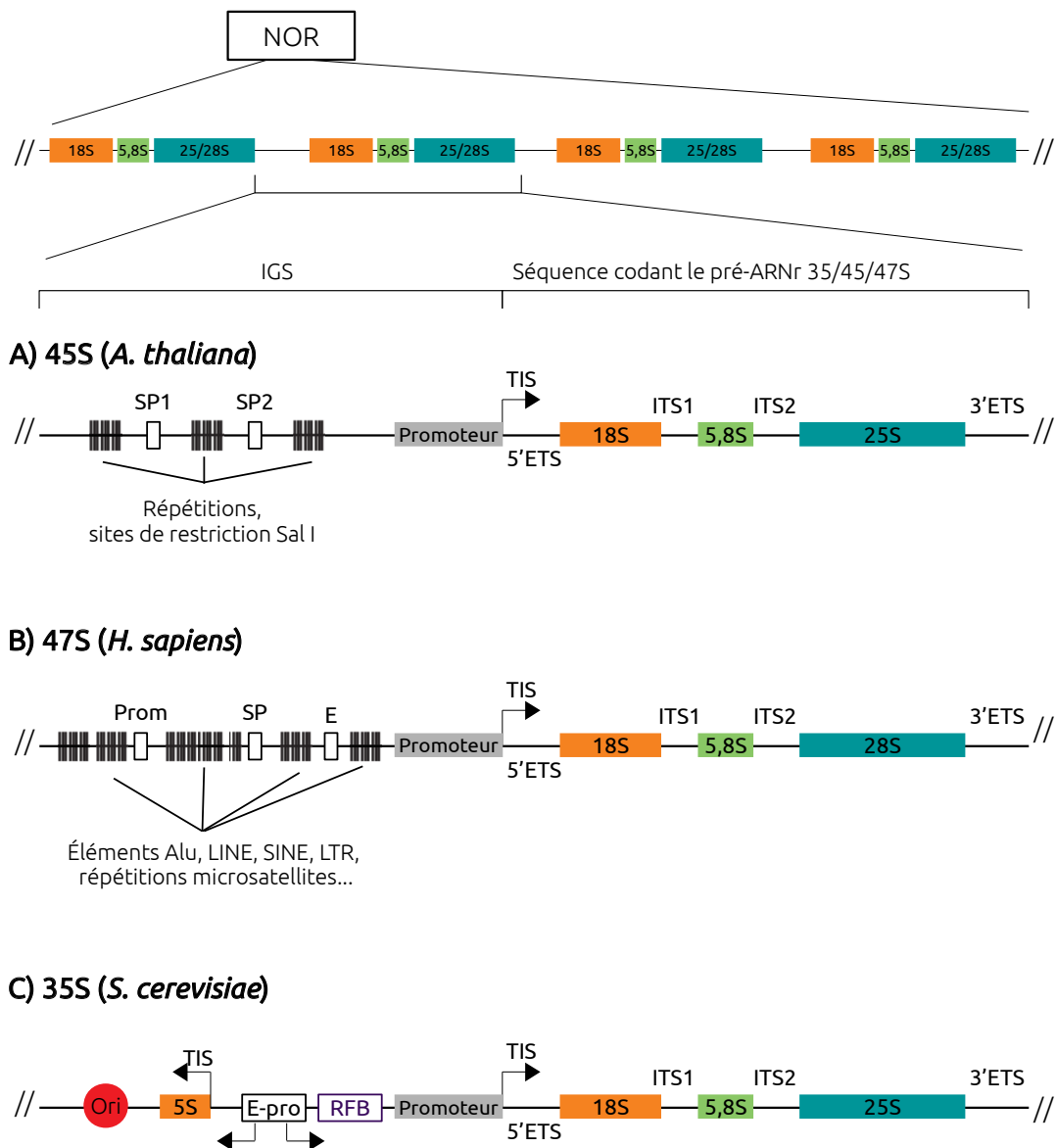


Figure 2.2 – La variabilité interspécifique de l'espace intergénique (IGS). Les unités de répétition de l'ADNr sont divisées en deux parties, la séquence codant pour les gènes d'ARNr 35/45/47S et l'IGS. **A)** L'IGS d'*A. thaliana*, en plus du promoteur proximal (Promoteur), et deux promoteurs espaceurs (SP), l'IGS contient trois sites de séquences répétées, qui comportent de nombreux sites de restriction *SalI*. **B)** Chez l'homme, l'IGS contient de nombreuses séquences répétées comme des séquences microsatellites ou des éléments transposables, un enhancer (E), un promoteur de transcription non codante de l'IGS (Prom) et un promoteur espaceur (PS). **C)** Chez la levure (*Saccharomyces cerevisiae*), chaque IGS contient un gène codant pour le pré-ARNr 5S en orientation inverse des gènes 35S (arrangement type-L). L'IGS contient aussi un promoteur bidirectionnel de l'ARN pol II (E-pro) et aucune séquence répétée n'a été identifiée.

somes) en fonction de l'espèce. Cependant, dans la plupart de cas les *loci* d'ADNr 5S sont en position intra-chromosomique (souvent péricentromérique). Inversement, les *loci* d'ADNr 45S sont en position terminale (proches ou adjacents aux télomères). La position chez l'ancêtre commun du *loci* d'ADNr 35S/45S/47S (contenant aussi l'ARNr 5S) est donc reconstruit comme terminal (GARCIA *et al.* 2017 ; SOCHOROVÁ, 2018).

De plus, chez la grande majorité d'organismes, les *loci* 45/47S se trouvent à l'extrémité du bras court des chromosomes appelés «acrocentriques». Ces chromosomes ont la particularité d'avoir un bras chromosomique plus petit que l'autre. En conséquence, le centromère est plus proche d'un des deux télomères et la quasi totalité du bras court est constitué d'ADNr (McSTAY, 2016).

Chez l'homme par exemple, environ 300 répétitions d'ADNr sont distribuées entre les NORs sur les bras courts de chacun des cinq chromosomes acrocentriques HSA13, HSA14, HSA15, HSA21 et HSA22 (McSTAY, 2016). La mouche *Drosophila melanogaster* (Meigen, 1830) possède jusqu'à 600 répétitions d'ADNr, réparties entre les NORs du bras court du chromosome Y entièrement hétérochromatique et dans l'hétérochromatine péricentromérique du chromosome X (RITOSSA et SPIEGELMAN, 1965). Chez *A. thaliana*, ~700 copies sont distribuées entre deux NORs proches des télomères des bras courts des chromosomes acrocentriques II et IV (COPENHAVER et PIKAARD, 1996a,b).

Mais il existe des exceptions à cette organisation, comme la levure (*Saccharomyces cerevisiae*), qui possède un seul cluster de gènes d'ARN ribosomique 35S, avec environ 150 copies au bras long du chromosome XII. Ou encore la souris *Mus musculus* (von Linné, 1758) où la majorité de chromosomes sont acrocentriques et les clusters d'ADNr 47S se trouvent dans les régions centromériques des chromosomes 12, 15, 16, 18 et 19 (KURIHARA *et al.* 1994).

2.2 La variabilité inter- et intraspécifique du nombre de copies de gènes d'ARN ribosomique

Au niveau évolutif, malgré la grande conservation de la séquence et structure des gènes codant les ARNr, ainsi qu'une tendance à avoir peu de *loci* d'ADNr, il existe une grande fluctuation du nombre de copies de gènes d'ARNr (NC). En effet, le NC peut varier de quelques unités à plusieurs milliers en fonction de l'organisme (LONG et DAWID, 1980 ; ROGERS et BENDICH, 1987).

Très récemment, une étude comprenant une grande gamme phylogénétique du règne des *Fungi*, a révélée une variation de 11 à 251 copies en fonction de l'espèce (LOFGREN *et al.* 2019). Les animaux présentent généralement entre 100 et 1.000 copies de gènes d'ARNr par cellule diploïde, tandis que les plantes contiennent entre 500 et 40.000 copies (LONG

et DAWID, 1980 ; ROGERS et BENDICH, 1987 ; ROSATO *et al.* 2016). Par exemple, chez les Gnétophytes, l'une des quatre lignées de Gymnospermes, le nombre d'unités d'ADNr est compris entre 2.500 à 6.000 (WANG *et al.* 2019). Parmi les grands clades des Bryophytes (*Marchantiophyta*, *Bryophyta* et *Anthocerotophyta*), ce nombre varie de 500 à 2.500 copies (avec une moyenne de 1197 ± 630) par génome diploïde (ROSATO *et al.* 2016).

Aucune association entre la position phylogénétique et le nombre de copies de gènes d'ARNr que présente une espèce a été mis en évidence, à l'exception des *Fungi* où la similitude entre le NCs est inversement corrélé à la distance phylogénétique (LOFGREN *et al.* 2019). Cependant, il existe une corrélation positive entre le NCs et la taille du génome chez les eucaryotes, même si le mécanisme responsable de cette relation reste inconnu (GARCIA *et al.* 2017 ; PROKOPOWICH *et al.* 2003).

Au niveau intra-spécifique, de nombreuses études ont montré une grande variabilité au sein de populations naturelles. C'est le cas par exemple dans des populations de crustacés d'eau douce du genre *Daphnia* (Müller, 1785) (EAGLE et CREASE, 2012), chez *Aspergillus fumigatus* (Fresenius, 1863), un champignon responsable de maladies chez les humains et chez les oiseaux (HERRERA *et al.* 2009) ou encore chez la levure (KOBAYASHI, 2014).

C'est également le cas chez la drosophile, où les populations présentent une variation en quantité d'ADNr pouvant aller jusqu'à un facteur de 6 fois (ALDRICH et MAGGERT, 2014 ; LYCKEGAARD et CLARK, 1989). Ce facteur peut être de jusqu'à 10 fois chez l'humain et chez la souris. En effet, l'humain présente de quelques dizaines à environ 410 copies de gènes 45S et la souris, de 31 à 289 copies (GIBBONS *et al.* 2015).

Le NC peut changer sur une courte période de temps, d'une génération à une autre (après une méiose ou une mitose s'il s'agit d'un organisme unicellulaire) et même au sein d'un individu. Les populations naturelles d'*A. thaliana* présentent aussi une grande variabilité du nombre de copies. De plus, suite à un croisement entre un individu à fort nombre de copies (contenant ~2.500 répétitions par génome haploïde) et un autre à faible nombre de copies (environ 500), la descendance de la génération F1 (qui possède ~1.500 copies) manifeste des modifications successives du nombre de copies (qui diffère du nombre de copies des parents), déjà évidentes dès la génération F2 (RABANAL *et al.* 2017b).

Le génotypage du NC de gènes 45S dans trois généalogies humaines constituées de huit frères et sœurs, a montré des différences de taille entre les clusters d'ADNr de ces derniers, pouvant représenter un écart de jusqu'à deux fois du nombre de copies (GIBBONS *et al.* 2015). Une autre étude chez l'humain qui compare également les clusters d'ADNr au sein de familles, a démontré que ces derniers sont soumis à un arrangement méiotique (de parent à fils) à une fréquence de >10% par *locus* d'ADNr. En d'autres mots, un

locus d'ADNr subit une délétion ou expansion dans 10% des méioses (STULTS *et al.* 2007). Plus encore, cette variabilité a aussi été observée au sein d'individus. En effet, l'étude de la longueur des NORs de globules blancs humains a montré une grande variabilité entre les cellules d'un même individu (STULTS *et al.* 2007).

Finalement, une analyse du NC de gènes 47S d'individus et leurs données d'expression (séquençage ARN à haut débit) correspondantes a pu corrélérer de manière significative l'expression d'environ 200 gènes transcrits par l'ARN Pol II au dosage d'ADNr. Les produits de ces gènes sont enrichis fonctionnellement en facteurs apparentés à la biogenèse des ribosomes et au nucléole. Cependant, ils sont aussi enrichis en facteurs impliqués dans le dépôt de marques épigénétiques de la chromatine (GIBBONS *et al.* 2014). Chez le maïs (*Zea mays*; von Linné, 1753) également, une analyse de 256 individus présentant une variation de 1.061 à 17.347 copies de gènes 45S (3.246 en moyenne), a identifié 32 gènes avec une expression corrélée significativement avec le nombre de copies. Ces gènes sont liés à la biogenèse des ribosomes (LI *et al.* 2018). Mais pourquoi donc existe-t-il une telle variabilité du nombre de copies au sein des clusters d'ADNr ?

2.3 L'ADN ribosomique, un des *loci* les plus instables du génome

Au sein d'un génome, l'apparition de cassures ou d'anomalies structurelles chromosomiques comme des délétions, inversions ou encore insertions n'est pas aléatoire. En effet, il existe des *loci* dits «fragiles» qui sont plus propices à subir des lésions ou des erreurs lors de la réplication ou réparation de l'ADN.

L'ADNr est connu pour être un des *loci* les plus instables du génome à cause de leur nature répétée mais aussi de leur taux de transcription qui est le plus élevé du génome (McSTAY, 2016; SALIM et GERTON, 2019). Ses caractéristiques le rendent propice à une des lésions les plus fréquentes et dangereuses pour l'intégrité génomique, les cassures double brin de l'ADN (DSB pour Double Strand Break en anglais).

Les DSBs peuvent entraîner des pertes ou des gains de copies après des événements de recombinaison délétères au cours de leur réparation (KOBAYASHI, 2014; SCHUBERT *et al.* 2004; TSANG et CARR, 2008). En effet, les cellules de levure, de mammifères et de plantes, présentent en conditions normales, des taux de $\gamma H2AX$ (marqueur des cassures double brin de l'ADN) plus élevés dans les nucléoles que dans tout le nucléoplasme (HUANG *et al.* 2012; KOBAYASHI, 2011; PAREDES *et al.* 2018), et même par rapport à d'autres régions répétées comme les centromères (HUANG *et al.* 2012).

Parmi tous les facteurs qui peuvent endommager l'ADN (l'ADN peut subir jusqu'à 1.000.000 de lésions par cellule et par jour!), les facteurs endogènes sont une des sources majeures d'instabilité génomique. Par exemple, les processus liés à la transmission et expression de l'information génétique qui ont, entre autres, le potentiel d'induire des coupures simple et double brin de l'ADN (GABORIAUD et WU, 2019; KIM et JINKS-ROBERTSON, 2012; MARNEF *et al.* 2017; TSANG et CARR, 2008).

En effet, lors de la réplication de l'ADN, le matériel génétique est susceptible de subir des lésions comme des DSBs. Le réplisome rencontre fréquemment des obstacles sur l'ADN matrice tels que des protéines non-histoniennes liées à l'ADN comme celles de la machinerie de transcription. L'un des obstacles plus problématiques au niveau des fourches de réplication, notamment à des régions fortement transcrites, car souvent aboutit à des collisions entre les deux machineries (qui peuvent entraîner des DSBs). Lorsque les fourches de réplication rencontrent de telles barrières, le réplisome ralentit et se stabilise. Si le problème n'est pas résolu, la fourche de réplication peut se rompre et le réplisome est alors dissocié de l'ADN, qui est alors exposé à des possibles lésions (GABORIAUD et WU, 2019; KIM et JINKS-ROBERTSON, 2012; TSANG et CARR, 2008).

D'un autre côté, la transcription elle-même est aussi une grande source de DSBs à cause de structures secondaires qui peuvent se former au niveau du brin d'ADN activement transcrit (MARNEF *et al.* 2017). Plus encore, une forte activité transcriptionnelle combinée à un état ouvert de la chromatine, augmente la sensibilité de l'ADN à des agents mutagènes endogènes (des agents génotoxiques produits par le métabolisme cellulaire) ou exogènes (comme des rayonnements ultraviolets) (GABORIAUD et WU, 2019; HUANG *et al.* 2012; KIM et JINKS-ROBERTSON, 2012).

Les DSBs déclenchent l'activation de la réparation de l'ADN. La kinase ATM (Ataxia Telangiectasia-Mutated) reconnaît les DSBs et initie la transduction du signal. Une fois activée par son recrutement au niveau de la cassure double brin, elle phosphoryle la sérine 139 du variant d'histone $H2A_X$ ($\gamma H2A_X$) des nucléosomes entourant le site endommagé. Des centaines à des milliers de $H2A_X$ sont phosphorylés (KUO et YANG, 2008). Grâce à cette phosphorylation de l'histone $H2A_X$, il est possible de repérer par immunofluorescence les foci de réparation des DSBs formés, par le biais d'anticorps anti- $\gamma H2A_X$ (KUO et YANG, 2008 et Figure 2.3).

Ensuite, en fonction du cycle cellulaire ou de la position génomique et nucléaire, les DSBs sont réparés soit par recombinaison homologue (RH) soit par jonction d'extrémité non-homologue (NHEJ) (Figure 2.4). Brièvement, lors de la NHEJ, les extrémités cassées sont alignées et liguées ensemble, ce qui ne requiert pas de séquence complémentaire pour réparer la cassure. Ces revues abordent plus en détail cette voie de réparation : HASTINGS *et al.* 2009; HER et BUNTING, 2018; SCHUBERT *et al.* 2004. Ce mécanisme est très mutagène car il conduit régulièrement à des petites délétions de 1 à 4 paires de bases

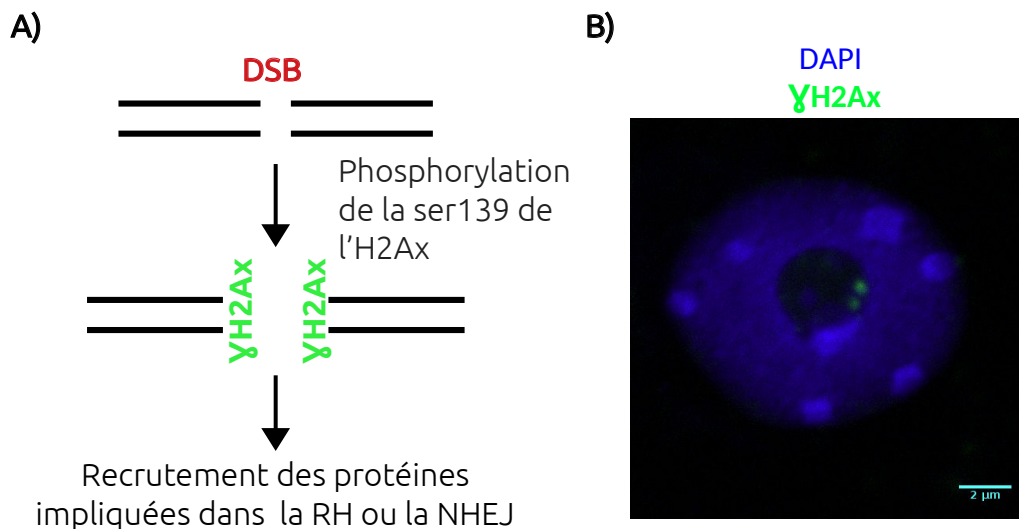


Figure 2.3 – Visualisation au microscope de *foci* de réparation des cassures double brin de l'ADN (DSB). **A)** La kinase ATM (Ataxia Telangiectasia-Mutated) reconnaît les DSBs et phosphoryle la sérine 139 du variant d'histone $H2A_X$ ($\gamma H2A_X$) des nucléosomes entourant le DSB. Cette phosphorylation induit le recrutement des protéines impliquées dans la recombinaison homologue (RH) ou la jonction d'extrémités non-homologue (NHEJ). **B)** Observation au microscope confocal à épifluorescence d'un noyau d'*Arabidopsis thaliana* coloré au DAPI (bleu), présentant deux *foci* de réparation de l'ADN marqués par immunofluorescence avec des anti-corps anti- $H2A_X$ (vert).

(pb) (HASTINGS *et al.* 2009). En revanche, la RH répare l'ADN sans erreur au niveau de la séquence car utilise un ADN complémentaire comme matrice.

La RH commence par la production d'extrémités cohésives au niveau de la cassure double brin, ensuite les extrémités d'un seul brin de l'ADN endommagé vont envahir l'ADN non endommagé et se lier au brin complémentaire. Ce brin complémentaire sert de matrice pour la synthèse d'ADN entre les deux extrémités de la cassure et il s'agit normalement d'une chromatide sœur (Figure 2.4).

Par conséquent, il est considéré que la réparation par RH est limitée aux phases S et G2 du cycle cellulaire et la réparation par NHEJ prédomine dans les phases G0 et G1. Mais de nombreuses études montrent que la réparation de DSBs est aussi accomplie par RH durant l'interphase, en particulier au niveau de séquences répétées en tandem qui sont hautement recombino-gènes. Notamment l'ADNr, où se mécanisme semble être majoritaire (MARNEF *et al.* 2019; McSTAY, 2016; SLUIS et McSTAY, 2015; TSANG et CARR, 2008). En effet, dans ces cas, les répétitions au sein de la même chromatide, du chromosome homologue ou d'un chromosome non-homologue (par exemple un NOR d'un autre chromosome) peuvent être utilisées comme matrice pour la RH.

Si les DSBs ne sont pas réparés, ils peuvent entraîner la mort cellulaire. Si ils sont mal réparés, ils peuvent induire des réarrangements chromosomiques, particulièrement au niveau de séquences répétées en tandem puisqu'elles sont propices à des recombinaisons

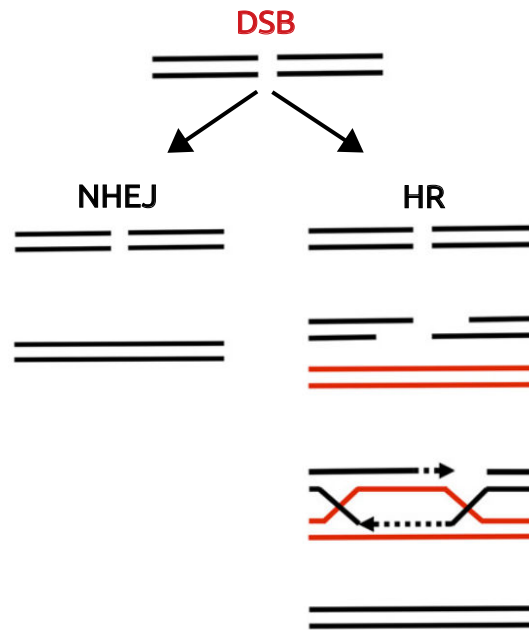
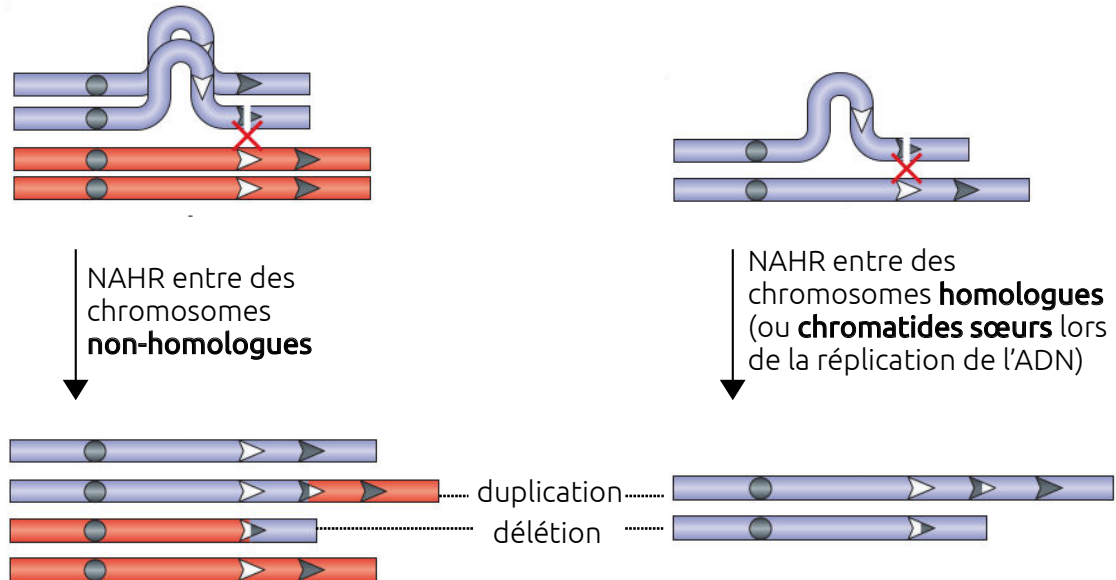


Figure 2.4 – Les deux mécanismes de réparation des cassures double brin de l'ADN (DSB), adapté de HER et BUNTING, 2018. La jonction d'extrémités non-homologue (NHEJ) lie ensemble les extrémités cassées. La recombinaison homologue (RH) répare l'ADN en utilisant un ADN complémentaire comme matrice. La RH commence par la production d'extrémités cohésives au niveau de la cassure double brin, ensuite les extrémités d'un seul brin de l'ADN endommagé vont envahir l'ADN non endommagé et se lier au brin complémentaire. Ce brin complémentaire sert de matrice pour la synthèse d'ADN entre les deux extrémités de la cassure.

homologues non alléliques (NAHR) (KRASILEVA, 2019; SASAKI *et al.* 2010; SCHUBERT *et al.* 2004). La NAHR est une recombinaison homologue entre des séquences d'ADN répétées à des positions non alléliques (Figure 2.5). En effet, l'échange se fait entre deux segments d'ADN très similaires mais qui ne sont pas homologues en termes de longueur et position dans le chromosome. Elle peut avoir lieu entre des chromatides sœurs lors de la réplication de l'ADN ou entre des chromosomes homologues, non-homologues et au sein de la même chromatide si le DSB a lieu durant l'interphase.

Le choix des segments d'ADN non alléliques pour la recombinaison définissent le résultat du croisement, qui conduit à une délétion et duplication réciproque d'unités de répétition entre les deux segments d'ADN ou sur une seule chromatide (Figure 2.4 et HASTINGS *et al.* 2009; KRASILEVA, 2019; PENG et KARPEN, 2008; SASAKI *et al.* 2010; TSANG et CARR, 2008). Par exemple, si la recombinaison se fait entre des répétitions de la même chromatide (recombinaison homologue intrachromosomique), les unités qui interviennent forment une boucle entre le site donneur et l'accepteur (endommagé). Cette boucle est ensuite excisée, formant ainsi un ADN circulaire extrachromosomique et réduisant le nombre de répétitions au sein du chromosome (KOBAYASHI, 2014; KRASILEVA, 2019; PENG et KARPEN, 2008; SASAKI *et al.* 2010; TSANG et CARR, 2008).

A) NAHR interchromosomique



B) NAHR intrachromosomique

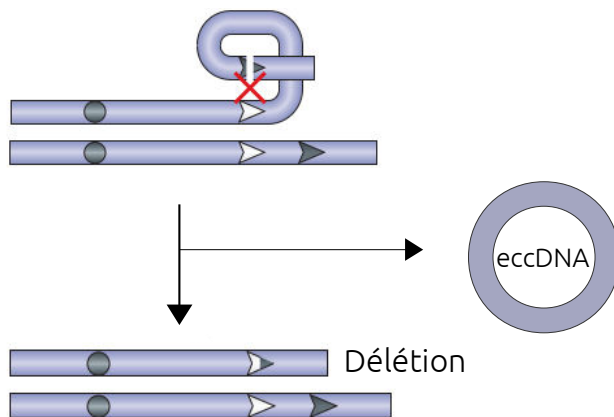


Figure 2.5 – La Recombinaison Homologue Non Allélique (NAHR), adapté de SASAKI *et al.* 2010. La NAHR est une recombinaison homologue entre des séquences d'ADN à des positions non alléliques. A) La NAHR interchromosomique peut avoir lieu entre des chromosomes homologues et non-homologues ou entre des chromatides sœurs lors de la réplication de l'ADN. Cette NAHR conduit à une délétion et duplication réciproque d'un segment d'ADN entre les deux chromatides. B) La NAHR intrachromosomique conduit à une délétion au sein de la chromatide. Les unités qui interviennent pour la réparation du DSB forment une boucle entre le site donneur et l'accepteur. Cette boucle est excisée formant ainsi un ADN circulaire extrachromosomique (eccDNA). Les DSBs sont représentés par une croix rouge.

On vient de voir que l'ADNr présente toutes les caractéristiques favorables à des expansions et contractions dues à des recombinaisons homologues non alléliques entre des unités de répétition lors de la réparation d'un DSB. Par conséquent c'est une des régions plus dynamiques en termes de variation du nombre de copies. Mais malgré cette grande instabilité, au sein d'une espèce on retrouve un équilibre du NC d'ARNr, qui est maintenu dans un intervalle (entre un minimum et un maximum). En effet, la longueur de l'ADNr change toujours de manière cyclique, les événements menant à la réduction de l'ADNr (comme des NAHRs) s'alternent avec des amplifications de ce dernier (KOBAYASHI, 2014). Chez la drosophile par exemple, suite à une suppression artificielle d'unités de répétition de l'ADNr, une augmentation allant jusqu'à 15 copies par génération est observée (PAREDES et MAGGERT, 2009).

Cette maintenance d'un équilibre à long terme du NC suggère que la cellule a mis au point un (ou plusieurs) système permettant de restaurer le nombre optimal de répétitions suite à des altérations au sein de l'ADNr.

Un tel mécanisme a été décrit chez la levure, où un système d'«*amplification génique*» permet la récupération des copies perdues par des événements de recombinaison homologue non allélique lors de l'interphase (KOBAYASHI, 2011, 2014; NELSON *et al.* 2019). Au cours de la phase S du cycle cellulaire, la réplication de l'ADNr commence à partir des origines de réplication bidirectionnels situés à l'IGS. La fourche de réplication s'arrête lorsqu'elle rencontre un site de blocage de la fourche de réplication (RFB) par l'effet de la protéine Fob1, qui s'associe spécifiquement à la séquence de la RFB.

Lorsque la fourche de réplication est arrêtée, un DSB est induit et l'extrémité cassée est réparée par recombinaison homologue avec la chromatide sœur. Si le DSB est réparée avec l'unité de répétition la plus proche de la chromatide sœur (recombinaison égale des chromatides sœurs), le nombre de répétitions ne change pas (Figure 2.6). Par contre, si le DSB est réparée avec une unité plus éloignée de la chromatide sœur, certaines unités sont répliquées deux fois et le nombre de répétitions augmente (recombinaison inégale des chromatides sœurs)(Figure 2.6).

Le choix du type de recombinaison se fait par la transcription du promoteur non codant E-pro qui produit un ARN non-codant. Ainsi, dans des situations où le nombre de copies de l'ADNr est normal, la protéine SIRTOUINE2 (SIR2, une NAD-dépendante histone déacetylase) inhibe la transcription d'E-pro. Ceci permet au complexe de la cohésine (qui relie les chromatides sœurs jusqu'à la division cellulaire) de s'associer à l'IGS. Les DSB sont réparés par recombinaison égale des chromatides sœurs (Figure 2.6).

Dans le cas où le nombre de copies est réduit, la répression par SIR2 est supprimée et E-pro est activé. En effet, le complexe Upstream Activating Factor (UAF) de l'ARN Pol I, se fixe sur les promoteurs proximaux de l'ADNr pour activer la transcription par l'ARN Pol I. Lorsque le nombre de copies est réduit, les sites de fixation de l'UAF sont saturés et

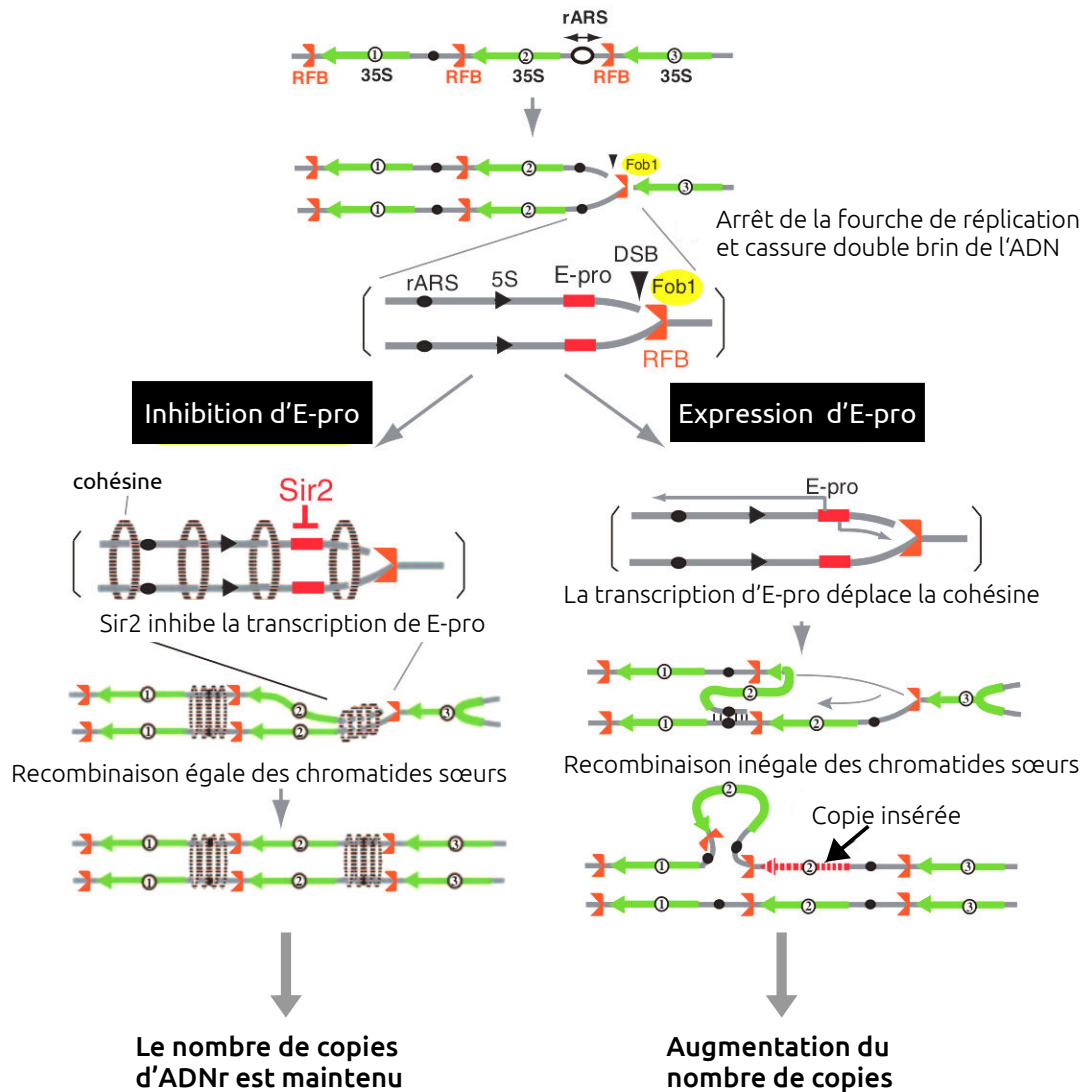


Figure 2.6 – Modèle d'amplification génique de l'ADNr 35S chez la levure *Saccharomyces cerevisiae*, adapté de KOBAYASHI, 2014. La protéine Fob1 s'associe au site de blocage de la fourche de réplication (RFB). Quand la fourche de réplication arrive sur un site où Fob1 est présent, elle est arrêtée et une cassure double brin de l'ADN (DSB) est induite. Le DSB est ensuite réparé par recombinaison homologue avec la chromatide sœur. Dans des situations normales, SIR2 inhibe la transcription d'E-pro et le DSB est réparé par recombinaison égale des chromatides sœurs. Il n'y a pas de perte du nombre de copies des gènes d'ARNr 35S. Dans le cas où le nombre de copies est réduit, E-pro est activé. Sa transcription déplace le complexe de la cohésine et lors de la réparation du DSB, certaines unités sont répliquées deux fois (recombinaison inégale des chromatides sœurs). Par conséquent, le nombre de copies augmente.

les complexes UAF «libres» se fixent sur les promoteurs de la SIR2. Ceci inhibe l'histone déacétylase SIR2 et l'ARN non codant E-pro est transcrit. Cette transcription déplace le complexe de la cohésine de l'IGS et la réparation des DSB se fait avec des répétitions éloignées de la chromatide sœur. Par conséquent, le nombre de copies d'ADNr augmente (Figure 2.6).

2.4 L'ADN ribosomique et la stabilité génomique

2.4.1 Le nombre de copies de gènes d'ARN ribosomique n'est pas lié au besoin transcriptionnel

Malgré la forte demande en transcrits, seul un sous-ensemble de gènes d'ARNr est transcrit à un moment donné dans une cellule eucaryote. Dans la plupart du temps, les autres copies sont associées à des marques répressives de la chromatine et condensées. D'abord, les organismes qui possèdent plusieurs NORs présentent souvent une expression différentielle de ces derniers. En effet, l'ensemble des copies d'un NOR (inactif) ne seront pas transcrites et l'ensemble du NOR va se condenser et créer un domaine hétérochromatique ne se distinguant pas de l'hétérochromatine constitutive environnante (McSTAY et GRUMMT, 2008). Ensuite, au sein d'un NOR actif, une partie des répétitions d'ADNr se trouve dans un état condensé silencieux et l'autre partie est décondensée et activement transcrite, autour de laquelle se forme le nucléole.

Cette inhibition NOR-spécifique des copies d'ADNr est maintenue à travers les cycles cellulaires (McSTAY et GRUMMT, 2008). La morphologie ouverte très particulière des NORs transcriptionnellement actifs correspond à ce qui a été appelé «constriction secondaire» (les constriction primaires étant les centromères) par Barbara McClintock, lors de ses observations cytologiques sur le maïs, sans connaître le lien avec la synthèse des ribosomes (McCLINTOCK, 1934).

Chez l'écotype d'*A. thaliana* Columbia (Col-0), l'ensemble des gènes d'ARNr sont exprimés au début de la germination de la graine. Puis, quatre jours après la germination, les gènes d'ARNr présents sur le chromosome 2, sont réprimés par l'accumulation de modifications épigénétiques de type répressives, et le NOR2 est le plus souvent exclu du nucléole (CHANDRASEKHARA *et al.* 2016; EARLEY *et al.* 2010; PONTVIANNE *et al.* 2012, 2013). Cependant, en fonction de l'écotype, c'est le NOR2 ou les deux qui seront actifs (RABANAL *et al.* 2017a).

Les gènes d'ARNr sont donc présents en deux états, un état actif (*i.e.* : euchromatique) et un état inactif (*i.e.* : hétérochromatique). Pour la plus grande partie de gènes inactifs, l'état condensé est hérité pendant la division cellulaire. Mais au sein des parties actives

des NORs, se trouvent des copies exprimées et non exprimées intercalées entre elles, et leurs nombres respectifs varient au cours de la vie d'un organisme selon les besoins de la cellule (MCSTAY et GRUMMT, 2008). Deux grands mécanismes sont impliqués, le contrôle de l'état chromatinien des gènes d'ARNr et la régulation du taux de transcription par l'ARN Pol I.

Les modifications épigénétiques modulent le niveau de condensation de la chromatine au niveau des gènes d'ARNr en modifiant ainsi l'accessibilité de la machinerie transcriptionnelle. Ainsi, les gènes actifs sont peu méthylés et on retrouve des marques d'histones liées à un état décondensé de la chromatine (*i.e.* : triméthylation de la lysine 4 de l'histone H3 [H3K4me3] ou encore l'acétylation de la lysine 9 de l'histone H3 [H3K9ac]). Contrairement aux gènes inactifs, enrichis en modifications post-traductionnelles des histones de type repressif (*i.e.* : di- ou triméthylation de la lysine 9 de l'histone H3 [H3K9me2/3] ou encore la méthylation de la lysine 20 de l'histone H4 [H4K20me]) et présentant un taux très élevé de méthylation des cytosines (MCSTAY et GRUMMT, 2008).

L'histone déacétylase 6 (HDA6) joue un rôle majeur dans l'inactivation transcriptionnelle des gènes d'ARN ribosomique chez *A. thaliana*. Cette protéine est impliquée dans la condensation de la chromatine du NOR2 mise en place au cours du développement de la plante et a pour cible plusieurs lysines de l'histone H3. En effet, chez les mutants *hda6*, les deux NORs apparaissent décondensées et sont enrichies en H3K4me3 (EARLEY, 2006; PROBST *et al.* 2004).

Le nombre de copies de gènes d'ARNr au sein d'une espèce est maintenu au delà du nombre de copies nécessaires pour répondre à la demande cellulaire en ARN ribosomique. Par conséquent, la variation interspécifique du nombre de copies n'est pas liée à des différences d'expression des gènes d'ARNr. Chez le maïs par exemple, aucune corrélation n'existe entre le NC d'ARNr 45S et entre l'expression de ces derniers dans les tissus. En d'autres mots, les lignées de maïs avec plus de copies d'ARNr n'ont pas un niveau d'expression global d'ARNr plus élevé (LI *et al.* 2018).

2.4.2 Importance de l'ADN ribosomique hétérochromatique dans la stabilité du génome

Nous avons vu que la réparation de ces DSBs peuvent induire une variabilité du nombre de copies au sein de ces *loci*. Mais en dépit de cette instabilité, une quantité de gènes d'ARNr inactive reste maintenue chez tous les organismes. Il est donc clair que cette hétérochromatine constitutive formée par une partie des copies d'ADNr est sous une pression de sélection positive et a donc été conservée au cours de l'évolution.

La question est de savoir pourquoi? Malgré l'importance de ces gènes et leur universalité au sein des organismes vivants, peu de travaux ont soulevé cette question. Cependant,

quelques données intéressantes ont été découvertes chez la levure, la drosophile ou des cellules de souris. Ces études ont démontré le rôle de l'ADNr hétérochromatique dans la stabilité des gènes d'ARNr et de l'hétérochromatine.

Des drosophiles mutantes avec un nombre réduit de copies de gènes d'ARNr, présentent moins d'hétérochromatine. Ainsi, l'expression de nombreux gènes est affectée mais le niveau de transcription des ARNr reste inchangé (PAREDES et MAGGERT, 2009; PAREDES *et al.* 2011). De manière intéressante, le nombre de gènes différentiellement exprimés était positivement corrélé à la taille de la délétion de l'ADNr. Ces gènes présentent des enrichissements fonctionnels pour des gènes codant pour des protéines localisées dans les mitochondries et impliqués dans le transfert d'électrons, ce qui est en accord avec les résultats obtenus par Gibbons et des collaborateurs (*cf.* I.3.2 et GIBBONS *et al.* 2014).

En outre, ces gènes sensibles à la réduction d'ADNr ne sont pas regroupés physiquement ou proches de régions hétérochromatiques, ils sont répartis dans le génome (PAREDES et MAGGERT, 2009; PAREDES *et al.* 2011). L'ensemble de ces résultats indique que la fraction hétérochromatique de l'ADNr est importante pour la structure tridimensionnelle du génome (les TADs), en particulier dans la mise en place ou la stabilité des compartiments hétérochromatiques au sein du noyau.

D'un autre côté, chez *Saccharomyces cerevisiae*, lorsque des copies d'ADNr sont perdues, les cellules deviennent plus sensibles aux DSBs induits par des agents mutagènes comme le méthanesulfonate de méthyle (MMS) ou la lumière ultraviolette (UV) (IDE *et al.* 2010; SALIM *et al.* 2017). En 2010, Ide et ses collaborateurs ont créé une série de souches de levure contenant un nombre réduit de répétitions d'ADNr (20 à 80 copies). Ils ont ainsi démontré que leur sensibilité au MMS et aux UV augmentait avec la diminution du nombre de répétitions d'ADNr. Cette instabilité génomique avait comme conséquence la sénescence prématurée des cellules (IDE *et al.* 2010).

Dans ces souches, la plupart des unités sont transcrites car la demande en molécules d'ARNr ne permet pas d'en garder une partie inactive (IDE *et al.* 2010) et présentent les structures en «arbre de Noël» lors de leur visualisation par microscopie électronique (KOBAYASHI, 2014). La même étude a démontré que cette sensibilité dépend de l'activité transcriptionnelle de l'ADNr, qui interfère avec la cohésion entre les *loci* de l'ADNr des chromatides sœurs lors de la recombinaison homologe.

Dans des souches de levure normales, les unités inactives fournissent un espace pour que les enzymes de réparation puissent accéder à l'ADN endommagé à cause, entre autres, de la forte activité transcriptionnelle des copies actives. En effet, les unités d'ADNr inactives sont associées à la condensine, un complexe protéique qui condense l'ADNr et connecte les chromatides sœurs pour leur réparation par RH (IDE *et al.* 2010). Dans une

souche avec un faible nombre de répétitions d'ADNr, la condensine ne peut pas accéder à l'ADNr et donc aucun dommage susceptible de se produire ne sera réparé (IDE *et al.* 2010).

En accord avec ces résultats, des cellules souches hématopoïétiques (cellules localisées dans la moelle osseuse, à partir desquelles sont générées les différentes lignées de cellules sanguines du corps) de souris qui ont un nombre réduit de copies de gènes d'ARNr sont également plus sensibles aux agents endommageant de l'ADN tels que la bléomycine, le MMS et les rayons X (XU *et al.* 2017).

Un autre exemple illustrant l'importance de l'état condensé d'une partie des unités de répétition au sein de l'ADNr pour la protection de l'ensemble des copies est une étude de Peng et Karpen de 2007. Ils ont montré chez des drosophiles mutantes pour l'histone méthyltransférase Su(var)3-9, où la formation de l'hétérochromatine constitutive est perturbée, une instabilité accrue de l'ADNr et l'apparition d'ADN circulaires extrachromosomiques (PENG et KARPEN, 2007). En effet, Su(var)3-9 est responsable de la méthylation de la lysine 9 de l'histone H3 (H3K9me3), une des marques impliquées dans la condensation de séquences répétées, comme l'ADNr ou les répétitions centromériques. Plus tard, une réduction du nombre de copies d'ADN ribosomique a été montrée chez ces mêmes mutants (PAREDES et MAGGERT, 2009).

De plus, la suralimentation de mouches qui accroît l'activité de mTOR (une kinase responsable de l'augmentation de la biogenèse des ribosomes lorsque les acides aminés sont abondants), est également accompagnée d'une réduction du nombre de copies d'ADNr (ALDRICH et MAGGERT, 2015). Xu et des collaborateurs ont montré ce même phénotype dans des cellules souches hématopoïétiques de souris dépourvues de PTEN (une phosphatase qui régule négativement l'activité de mTOR), qui présentent également un nombre réduit de copies de gènes d'ARNr (XU *et al.* 2017).

L'ensemble de ces résultats suggèrent que la perte de copies de gènes d'ARNr observée chez la drosophile comme chez la souris, est due à l'augmentation du nombre de copies actives, ce qui induit plus de cassures double brin au niveau de l'ADNr à cause de la forte activité transcriptionnelle. Comme il y a peu de copies inactives disponibles pour la fixation des complexes protéiques pour la recombinaison homologue entre des chromatides sœurs ou homologues, les DSBs sont réparés surtout par de la recombinaison homologue intrachromosomique (non allélique), ce qui a comme conséquence la formation d'ADN circulaires extrachromosomiques et la perte de copies d'ADNr.

Deuxième axe de la thèse

Les études que je viens de présenter, montrent que la fraction inactive des copies d'ADNr influence l'expression des gènes chez la drosophile et est nécessaire pour main-

tenir la stabilité de l'ADNr chez la levure et la souris (IDE *et al.* 2010 ; PAREDES et MAGGERT, 2009 ; PAREDES *et al.* 2011 ; XU *et al.* 2017). Cependant aucune étude n'a été réalisée jusqu'à présent chez les plantes.

Des lignées *A. thaliana* dans lesquelles le nombre de copies à été réduit artificiellement jusqu'à rester que 20% de la totalité des copies (environ 200 copies par génome haploïde), présentent le même niveau de production d'ARNr et leur croissance n'est pas altérée (PAVLIŠTOVÁ *et al.* 2016). Au sein de populations naturelles d'*A. thaliana*, il existe une grande variabilité du nombre de copies, qui fluctue entre ~500 et ~2.500 répétitions par génome haploïde (RABANAL *et al.* 2017b). Cependant, une réduction si importante du nombre de copies n'a pas été observée, alors que 20% des copies suffisent à maintenir la production en ARNr.

Grâce à ces lignées avec 80% d'ADNr en moins, il a été possible d'étudier l'importance des ces copies «en plus» pour l'intégrité du génome d'*A. thaliana*, le deuxième axe de ma thèse.

Résultats

Chapitre 3

Les domaines chromatiniens associés au nucléole d'*Arabidopsis thaliana*

3.1 Contexte et objectifs

Dans l'introduction générale, on a vu que les NADs correspondaient à des régions génomiques associées aux nucléoles et qu'ils appartiennent aux compartiments B hétérochromatiques du génome (*cf.* section). En 2016, l'équipe a mis en évidence que les 5 chromosomes d'*Arabidopsis thaliana* possèdent aussi au moins une région associée au nucléole, ce qui correspond à environ 8 % du génome (PONTVIANNE *et al.* 2016b et Figure 3.1).

Ces régions font en moyenne 11 kilobases et présentent également des caractéristiques propres à l'hétérochromatine. Ainsi, elles sont enrichies en marques épigénétiques de type répressives, sont composées essentiellement de séquences répétées comme les éléments transposables (TEs) et les gènes présents sont peu ou pas transcrits (PONTVIANNE *et al.* 2016b). En effet, parmi les neuf états chromatiniens définis chez cette espèce (SEQUEIRA-MENDES *et al.* 2014), les états 8 et 9 sont particulièrement représentés dans les NADs, caractérisés par un enrichissement en méthylation CG de l'ADN, en diméthylation de la lysine 9 de l'histone H3 (H3K9me2) et monométhylation de la lysine 27 de l'histone H3 (H3K27me).

Par ailleurs, l'ensemble des régions sub-téломériques font aussi partie des NADs. Cette organisation était attendue car chez *A. thaliana* les télomères se regroupent autour du nucléole (ARMSTRONG *et al.* 2001; FRANSZ *et al.* 2002; PONTVIANNE *et al.* 2016b).

Il est à noter que l'activité transcriptionnelle des NORs est déterminante pour la constitution des NADs. En effet, la majeure partie (~70%) des NADs des cellules sont localisées au niveau du bras court du chromosome 4 (PONTVIANNE *et al.* 2016b et Figure 3.1). À ce stade de développement, les copies de gènes d'ARNr 45S du NOR2 sont

inactives et ce sont les gènes d'ARNr issus du NOR4 qui sont activement transcrits. Cette expression différentielle des NORs induit la formation du nucléole autour du NOR4, tandis que le NOR2 reste très condensé et exclu du nucléole (PONTVIANNE *et al.* 2013). La totalité du bras court du chromosome 4, qui inclue le NOR4, le *knob* hk4S et les régions péri-centromériques, est associé au nucléole (Figure 3.1). Le *knob* hk4S est une région hétérochromatique du bras court du chromosome 4, qui résulte d'une inversion paracentrique incluant une région péri-centromérique (Figure 3.2 et FRANSZ *et al.* 1998, 2016). Chez *A. thaliana*, les régions péri-centromériques des chromosomes 2 et 4 interagissent

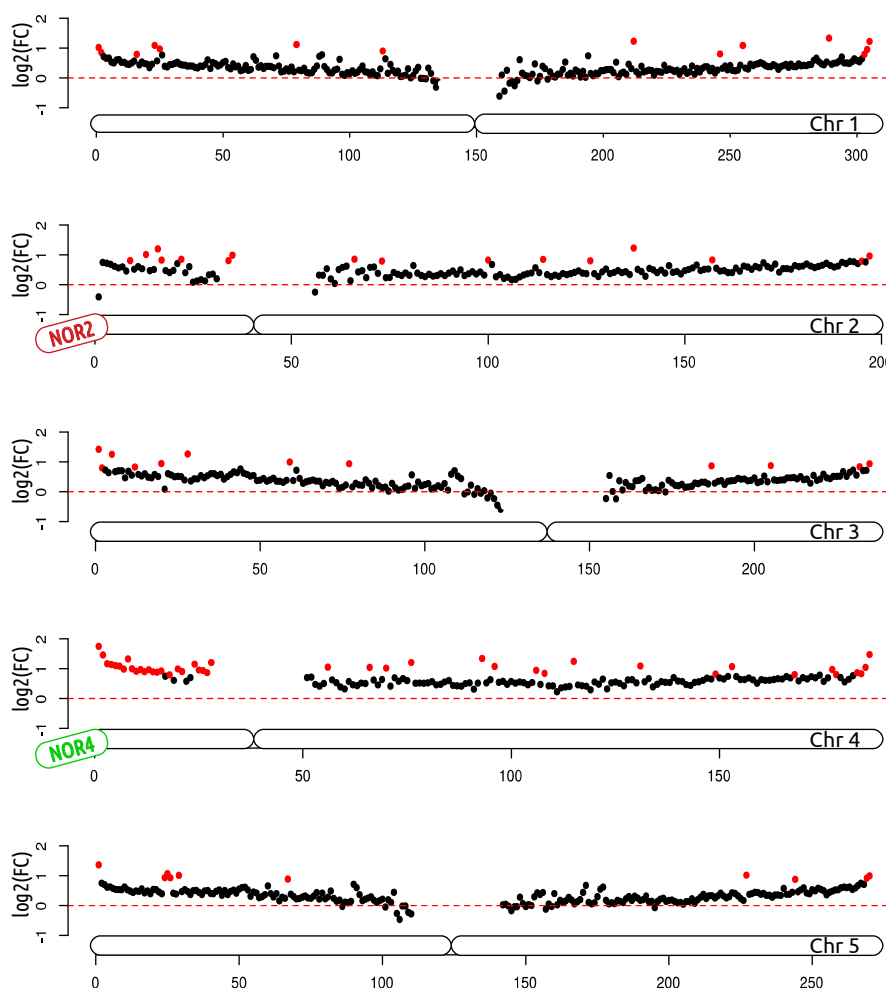


Figure 3.1 – Les NADs d'*Arabidopsis thaliana*. Enrichissement (ou fold change) en ADN nucléaire des fenêtres génomiques de 100 kilobases de chaque chromosome dans les cellules de feuille de l'écotype sauvage Col-0 âgées de 21 jours. Chaque fenêtre est positionnée en fonction de sa localisation le long du chromosome. Le $\log_2(\text{FC})$ est représenté en ordonnée. Les NADs sont représentés en rouge, seuil = 2. Les données des régions péri-centromériques ont été masquées, car la difficulté d'alignement de ces séquences sont source potentielle d'artefacts et peuvent mener à des mauvaises interprétations. Cette figure a été produite à partir de nouvelles données de séquençage ADN générées pendant ma thèse et on retrouve la même distribution des NADs que dans l'article publié en 2016 par l'équipe (PONTVIANNE *et al.* 2016b).

avec leurs NORs respectifs qui, dès lors, font partie du chromocentre (FENG *et al.* 2014),

ainsi que le *knob* hk4S, qui interagit également avec le NOR4 et le centromère 4. De manière intéressante, lorsque le NOR2 est actif, comme dans le mutant *nucléoline 1* (ou *nuc1*), il s'associe avec le nucléole, tout comme une partie du bras court du chromosome 2 et la région centromérique associée (PONTVIANNE *et al.* 2016b).

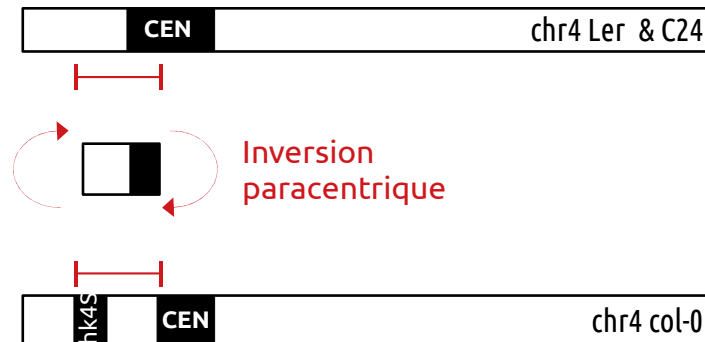


Figure 3.2 – Le *knob* hk4S d'*Arabidopsis thaliana* Col-0. Suite à une inversion paracentrique qui c'est produit au niveau du bras court du chromosome 4, l'écotype col-O possède le *knob* hk4S, contrairement aux écotypes Ler et C24 (FRANSZ *et al.* 1998, 2016).

En excluant les gènes d'ARNr, parmi les NADs se trouvent 907 gènes appelés NAD-gènes, ce qui correspond à 3,1% des 29393 gènes référencés. De plus, les NADs contiennent 3640 éléments transposables (NAD-TEs) correspondant à 11,7% des 31189 TEs annotés dans le génome de référence TAIR10 (annotation Araport11 : www.araport.org/data/araport11; KRISHNAKUMAR *et al.* 2015). Comme chez l'humain, les NAD-gènes sont très enrichis en gènes codant pour des ARNt (NÉMETH *et al.* 2010) et en pseudogènes, l'ensemble des deux représentant ~20% de l'ensemble (PONTVIANNE *et al.* 2016b).

Les pseudogènes comme les TEs sont généralement inactifs transcriptionnellement. Néanmoins, l'intégralité des NAD-gènes provenant des 5 chromosomes ont tendance à être moins exprimés que l'ensemble des gènes. Comme l'ARN Pol II est absente au niveau du nucléole (SCHUBERT et WEISSHART, 2015) et que la chromatine périnucléolaire est condensée, il se peut que cet environnement répressif joue un rôle dans la régulation transcriptionnelle des NAD-gènes.

Un des objectifs de ma thèse a donc été de déterminer s'il existe une voie active de régulation transcriptionnelle par séquestration nucléolaire de gènes transcrits par l'ARN polymérase II, ou si la localisation périnucléolaire de ces gènes est un phénomène passif.

Pour répondre à cette question, nous avons tout d'abord travaillé sur l'optimisation du protocole d'identification des NADs. Cette étape nous a permis de publier un article méthodologique dans *Methods in Molecular Biology* (volume 1675 : Plant chromatin dynamics) en 2018. J'ai ensuite analysé les NADs suite à un choc thermique (CT) et après

la transition florale (TF).

En effet, dans ces deux conditions, un grand remaniement de la chromatine se produit au sein des noyaux de cellules de feuilles d'*A. thaliana*. Cette restructuration pourrait induire des changements au niveau des NADs, liés à la réponse au stress thermique ou au changement de phase développementale. Le CT est une altération physiologique à court terme suite à laquelle la cellule reprend son activité, contrairement à la TF qui induit un changement de développement et donc une reprogrammation génétique maintenue par la suite. L'identification des NADs suite à ces deux événements permettra de déterminer, dans le cas d'une possible régulation génétique par séquestration nucléolaire, s'il s'agit d'un processus qui intervient dans des réponses rapides ou des changements à long terme.

La chromatine périnucléolaire est donc constituée de domaines chromatiniens silencieux et peu enrichis en gènes. Plusieurs études ont montré que la périphérie nucléaire est également un point d'ancrage de l'hétérochromatine chez les métazoaires (POMBO et DILLON, 2015) et chez *Arabidopsis thaliana* (BI *et al.* 2017; HU *et al.* 2019). Chez l'humain, des études ont montré une grande superposition entre les LADs et les NADs. Il est possible que ces régions peu ou pas transcrites par l'ARN polymérase II soient maintenues soit à la périphérie du nucléole soit à la périphérie du noyau.

Un des autres objectifs de mes travaux de thèse a donc été de réaliser une analyse comparative des NADs et des LADs-*like* chez *A. thaliana*. Cette analyse comparative a été publiée en 2019 dans le journal *Nucleus*. Cette publication constitue la dernière partie de ce chapitre de résultats.

3.2 L'étude des NADs chez *Arabidopsis thaliana*

3.2.1 FANS, FANoS et analyse bioinformatique pour l'identification des NADs

Pour identifier les domaines chromatiniens associés au nucléole, il est nécessaire d'isoler les noyaux et les nucléoles. Pour cela, nous utilisons les techniques de «*Fluorescent Activated Nucleus Sorting*» (FANS) et de «*Fluorescent Activated Nucleolus Sorting*» (FANoS), respectivement (PONTVIANNE *et al.* 2013; PONTVIANNE *et al.* 2016a). Ces techniques permettent d'isoler des nucléoles et des noyaux directement à partir d'une plante entière, et non pas à partir de protoplastes. De plus, le FANoS permet d'isoler des nucléoles avec peu de contamination, contrairement aux méthodes classiques d'isolement par fractionnement cellulaire (HACOT *et al.* 2010).

Ce tri de noyaux ou de nucléoles est réalisé par le biais d'un cytomètre de flux, grâce à la fluorescence d'une protéine strictement nucléolaire, la FIBRILLARINE 2 (FIB2), une méthyltransférase impliquée dans la maturation des ARNr dans le DFC, fusionnée au gène rapporteur YFP (Yellow Fluorescent Protein). Ainsi, des lignées transgéniques possédant la séquence codante du gène FIB2 fusionnée à la YFP, sous le contrôle d'un promoteur ubiquitaire comme le 35S CMV (promoteur de CytoMegalovirus) ou spécifique, sont utilisées. En bref, après fixation au formaldéhyde, la lyse des cellules permet la mise en suspension des noyaux et autres composants cellulaires dans du tampon de tri. La moitié du filtrat cellulaire est passée au cytomètre de flux afin de trier les noyaux. L'autre moitié est soniquée afin de casser l'enveloppe nucléaire et permettre le tri de nucléoles par cytométrie en flux. Les étapes de préparation des échantillons pour leurs analyses par cytométrie en flux sont détaillées dans l'article publié dans *Methods in Molecular Biology*.

La cytométrie en flux est une technique qui permet de comptabiliser et caractériser individuellement des particules en suspension dans un liquide, en les faisant traverser des faisceaux laser (Figure 3.3). L'échantillon est acheminé vers la buse qui permet de faire passer individuellement les particules à travers les faisceaux laser. La lumière réémise (par diffusion ou fluorescence) est captée par des détecteurs et permet de définir la taille, la complexité ou encore la fluorescence de chacune de ces particules (cellules, bactéries, noyaux, etc.).

Les différents rayons lumineux émis par la particule sont divisés et dirigés en fonction de leur longueur d'onde vers le détecteur correspondant, par un système de miroirs dichroïques et de filtres optiques (Figure 3.4).

En fonction du détecteur, trois paramètres peuvent être mesurés pour caractériser les particules (Figure 3.4) :

- Premièrement, la mesure de la diffraction de la lumière aux petits angles (entre 1 et 10 degrés) par rapport à l'axe d'émission du rayon laser par le détecteur FSC (Forward Scatter). Ce détecteur se trouve en face du faisceau laser et renseigne sur la taille des particules.
- Deuxièmement, la mesure de la lumière diffusée aux grands angles (à 90 degrés de l'axe d'émission du laser) captée par le détecteur SSC (Side Scatter), qui renseigne sur granularité (ou complexité interne) des particules.
- Finalement, la mesure de l'intensité de la fluorescence émise par la particule.

Les différents détecteurs captent les photons diffusés (FSC et SSC) ou émis (fluorescence) par la particule et convertissent ces signaux optiques en signaux électroniques qui sont par la suite numérisés, ce qui permet de les analyser informatiquement (Figure 3.5).

En fonction des propriétés optiques, différentes populations de particules sont dé-

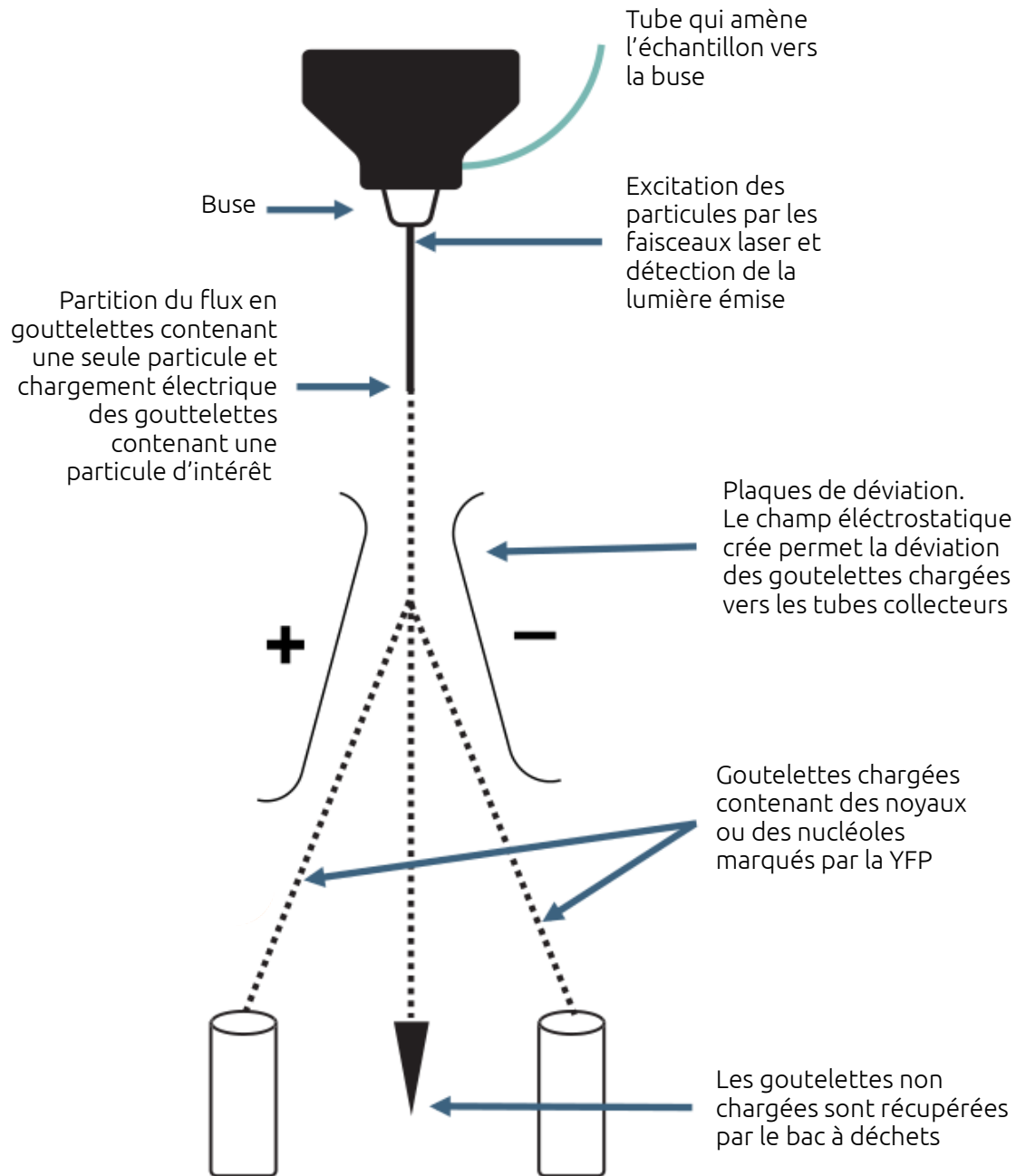


Figure 3.3 – Fonctionnement d'un cytomètre en flux trieur de cellules. L'échantillon est dirigé vers la buse. Le flux de particules traverse les faisceaux laser et la lumière émise par chaque particule est captée par des détecteurs. Ensuite le flux est divisé en gouttelettes contenant chacune une seule particule. Si la particule possède les caractéristiques physiques voulues par l'utilisateur, elle sera chargée électriquement. Lorsque la gouttelette chargée arrive au niveau du champ électrostatique créé par les plaques de déviation, elle est déviée et collectée dans des tubes. Adapté du guide du FACS BioRad S3™.

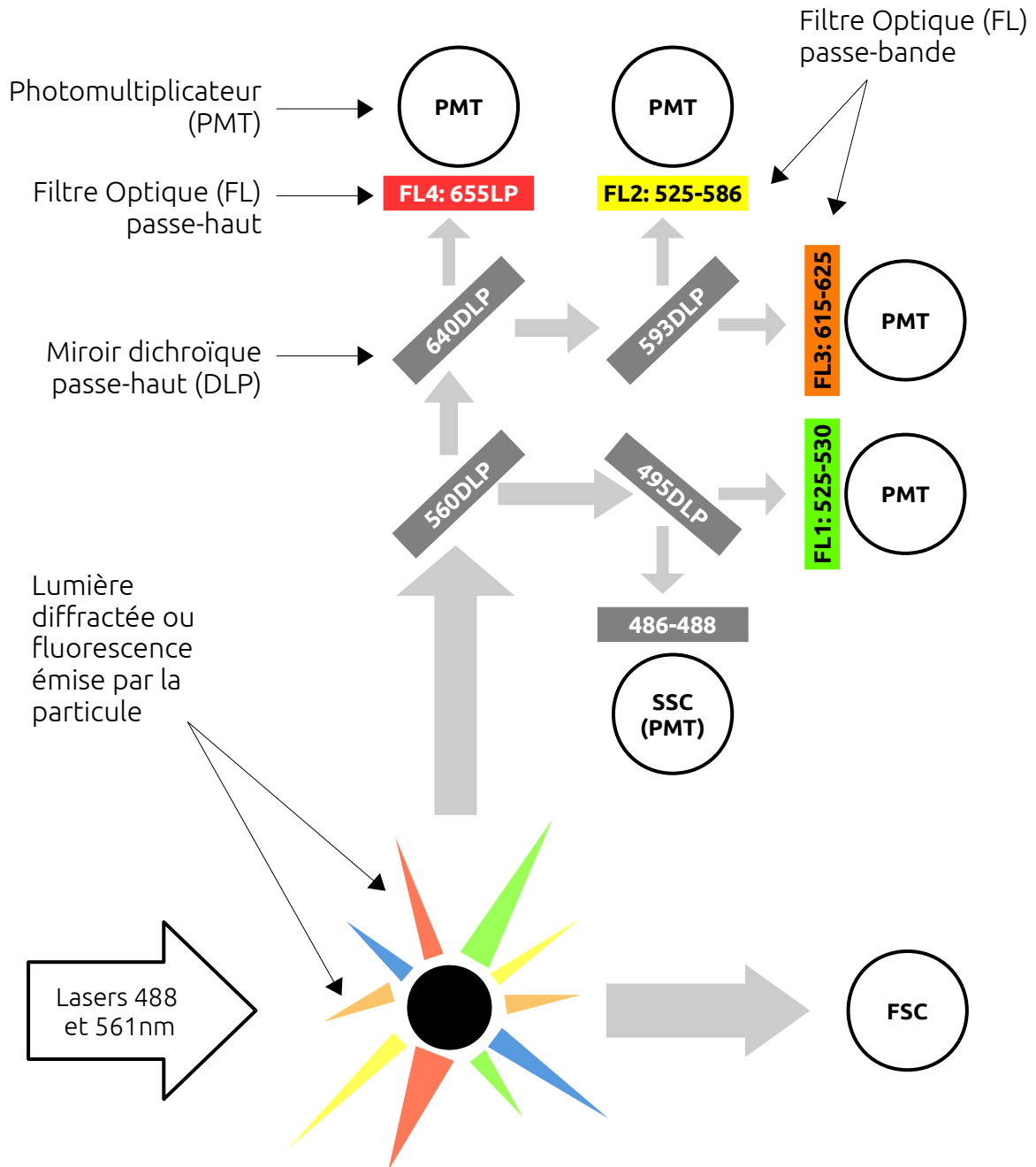


Figure 3.4 – Système optique du FACS BioRad S3TM. Ce cytomètre de flux comporte deux lasers (488 nm et 561 nm). La particule traverse les faisceaux laser, diffracte la lumière et peut également émettre de la fluorescence. La lumière diffractée aux petits angles est captée par le détecteur FSC (Forward Scatter) se trouvant dans l'axe des rayons laser. Les rayons lumineux entre 486 et 488 nm diffractés aux grands angles sont captés par le détecteur SSC (Side Scatter). La lumière émise est divisée et dirigée vers les détecteurs par un système de miroirs dichroïques passe-haut et de filtres optiques. Les miroirs dichroïques passe-haut reflètent les rayons lumineux au-dessous d'une certaine longueur d'onde et laissent passer les autres rayons. Ex/ le miroir 560DLP reflète les rayons avec une longueur d'onde inférieure à 560 nm et laisse passer ceux qui ont une longueur d'onde supérieure à 560 nm. Les filtres optiques passe-bande, absorbent tous les rayons lumineux sauf certaines longueurs d'onde, comme le FL1 qui laisse passer les rayons de 525 à 530 nm et les dirige vers le détecteur (photomultiplicateur) associé. Adapté du guide du FACS BioRad S3TM.

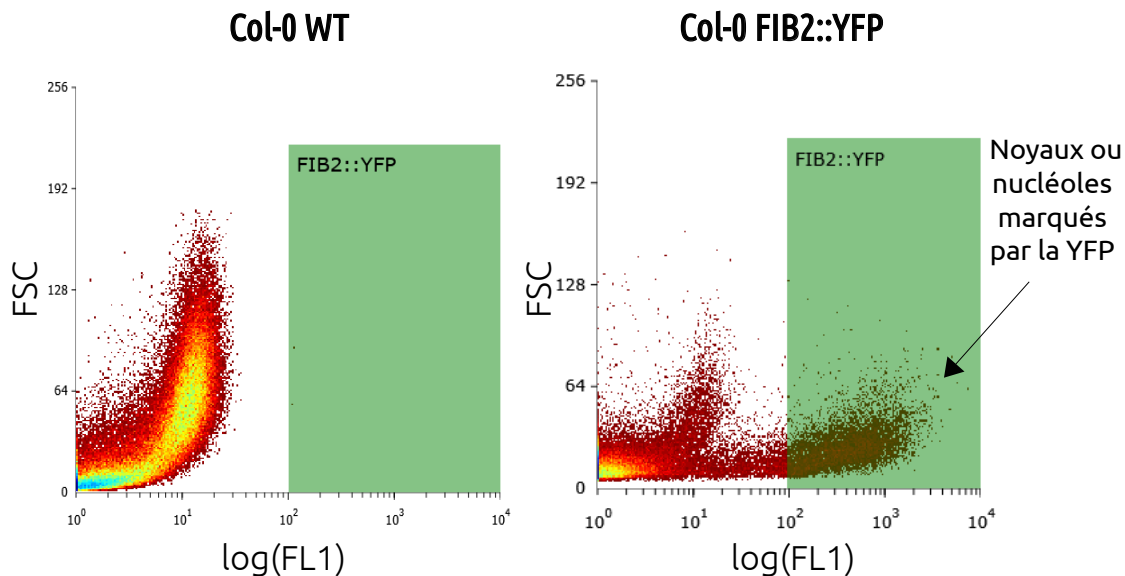


Figure 3.5 – La fenêtre de tri de noyaux ou nucléoles avec le cytomètre FACS BioRad S3TM, par le biais de la fluorescence du construit PROMOTEUR ::YFP ::FIB2. Graphiques représentant la taille de la particule analysée (signal FSC) en fonction de sa fluorescence entre 252 et 530 nm (la protéine de fusion YFP ::FIB2 émet de la fluorescence à une longueur d’onde de 527 nanomètres (nm)). Cette fluorescence est dirigée avec le filtre optique passe-bande 525-530 (FL1) vers le détecteur qui numérise l’intensité du rayon lumineux. La fenêtre de tri est définie avec le logiciel ProSortTM à partir de l’échantillon sauvage (Col-0 WT) car il permet de discerner les signaux dus à de l’auto-fluorescence et ceux provenant de la YFP. Les particules (noyaux ou nucléoles) incluses dans la fenêtre de tri (rectangle vert) seront déviées vers un tube collecteur.

finies. Ensuite, avec la fonction de tri des cytomètres en flux, il est possible de séparer physiquement une de ces populations et la collecter pour des analyses ultérieures (Figure 3.5). Le liquide contenant les particules en suspension est mis en vibration à un point précis pour donner des gouttelettes, chacune contenant une seule particule (Figure 3.3).

Le programme de tri permet la définition d’une zone de tri (Figure 3.5) en fonction des caractéristiques physiques des particules d’intérêt. Les gouttelettes qui contiennent en suspension une particule présentant ces caractéristiques, seront chargées électriquement. Si une goutte est chargée, quand elle arrive au niveau des plaques de déflexion fortement chargées, elle est déviée du côté de la plaque de polarité opposée et peut être ainsi collectée (Figure 3.3).

Les techniques FANS et FANoS reposent sur la fluorescence de la protéine de fusion FIB2 ::YFP pour trier par cytométrie en flux des noyaux ou des nucléoles d’un échantillon. Le réglage du cytomètre en flux trieur de cellules et les paramètres nécessaires pour définir la zone de tri de noyaux ou nucléoles grâce à la fluorescence de la protéine de fusion FIB2 ::YFP sont détaillés dans l’article méthodologique «Fluorescence-Activated Nucleolus Sorting in *Arabidopsis*», PONTVIANNE *et al.* 2016a.

Dans cette publication, comme pour la première identification des NADs chez *A. thaliana*, c'est le FACS AriaTM III (BD biosciences, États Unis) qui a été utilisé par l'équipe. Pour mes travaux de thèse j'ai utilisé le FACS BioRad S3TM avec le logiciel ProSortTM (BioRad, États Unis), en gardant les mêmes paramètres (FSC et fluorescence de la YFP) pour définir la zone de tri (Figure 3.5).

Les nucléoles ou les noyaux contenus dans une micro-goutte d'environ 3 nanolitres (nL) sont récupérés dans des tubes collecteurs. Ainsi, 600.000 nucléoles (quantité nécessaire pour produire un échantillon d'ADN nucléolaire de 10 nanogrammes) représentent un volume d'environ 2 millilitres (mL) s'ils sont triés avec le FACS AriaTM III, ou 3 mL avec le FACS BioRad S3TM. L'ADN nucléaire comme nucléolaire sont donc très dilués et il est impératif de les concentrer en limitant les pertes. Il est à noter que ce type de tri est beaucoup plus lent que le tri de cellules entières, car l'échantillon à trier se trouve dilué à de nombreuses autres particules assez variées. De ce fait, au début de ma thèse, j'ai réalisé plusieurs tests afin d'optimiser la quantité de matériel végétal utilisée, l'efficacité de tri et le rendement de la quantité d'ADN obtenue après purification.

Après purification et dosage, les ADN nucléaires et nucléolaires sont séquencés. Le protocole bioinformatique d'identification des NADs qui a été mis au point par Marie-Christine Carpentier (PONTVIANNE *et al.* 2016b), la bioinformaticienne du laboratoire, est le suivant.

La normalisation du nombre de lectures d'ADN provenant d'une fenêtre génomique au sein d'une même bibliothèque de séquences (ou «*library*»), permet de s'affranchir des différences observées entre des échantillons qui sont dues à la différence de couverture de séquençage entre ces derniers. Dans ce cas, pour évaluer l'enrichissement nucléolaire d'une fenêtre génomique, on ne compare pas le nombre brut de lectures provenant de la fenêtre en question entre les libraries d'ADN nucléolaire ou nucléaire, mais la proportion de lectures provenant de cette partie du génome sur l'ensemble des séquences de la library (*i.e.* : nombre de lectures provenant de la fenêtre génomique divisé par le nombre total de séquences de la bibliothèque).

Ainsi, les différences observées sont une réalité biologique et non une différence de couverture entre les deux échantillons. Ensuite, un domaine chromatinien est défini comme associé au nucléole si l'enrichissement (fold change ou FC) du nombre normalisé de lectures alignées provenant de l'ADN nucléolaire par rapport à l'ADN nucléaire est supérieur ou égal à 2 sur au moins dix fenêtres de suite (1000 pb).

Après l'identification des NADs, les gènes et TEs présents dans ces régions sont identifiés (NAD-gènes et NAD-TEs). Le même pipeline d'analyse mais en découpant le génome de référence en fenêtres de 100 kilobases (kb) est utilisé pour représenter graphi-

quement les NADs et permettre ainsi de les localiser au sein des chromosomes. Après comptage et normalisation du nombre de lectures d'ADN alignées sur des fenêtres génomiques de 100 kb, le logarithme binaire (\log_2) du FC de chaque fenêtre est représenté graphiquement (figure 3.1).

Le protocole complet pour l'identification des NADs chez *A. thaliana* avec les optimisations apportées, a fait l'objet en 2018 d'une publication dans la série de livres *Methods in Molecular Biology*, et est inséré à la suite de ce paragraphe. Dans cet article, le kit ChIP DNA Clean & ConcentratorTM (Zymo Research, États-Unis) est indiqué pour purifier l'ADN et le kit Quant-iTTM PicoGreen (Thermo Fisher Scientific, États-Unis) pour le doser. Il est à noter que la méthode de purification d'ADN au Phénol-Chloroforme (*cf.* Matériel et Méthodes) et le kit QuantiFluor[®] dsDNA System (Promega, États-Unis), selon les recommandations du fabricant, ont également été utilisés durant ma thèse. La méthode de purification d'ADN au Phénol-Chloroforme a été utilisée car elle permet d'obtenir le même rendement d'ADN tout en réduisant le coût de l'expérience. De même, le kit QuantiFluor[®] dsDNA System est une méthode de dosage plus simple et rapide que le kit Quant-iTTM PicoGreen et permet également la quantification de faibles concentrations d'ADN.

Chapter 7

A Method to Identify Nucleolus-Associated Chromatin Domains (NADs)

Marie-Christine Carpentier, Ariadna Picart-Piccolo,
and Frédéric Pontvianne

Abstract

The nuclear context needs to be taken into consideration to better understand the mechanisms shaping the epigenome and its organization, and therefore its impact on gene expression. For example, in *Arabidopsis*, heterochromatin is preferentially localized at the nuclear and the nucleolar periphery. Although chromatin domains associating with the nuclear periphery remain to be identified in plant cells, Nucleolus Associated chromatin Domains (NADs) can be identified thanks to a protocol allowing the isolation of pure nucleoli. We describe here the protocol enabling the identification of NADs in *Arabidopsis*. Providing the transfer of a nucleolus marker as described here in other crop species, this protocol is broadly applicable.

Key words Nucleolus-associated chromatin domains, NADs, Nucleolus, Nucleus, Chromatin, FACS

1 Introduction

Gene accessibility to transcription factors and RNA polymerases is a key step in gene transcriptional regulation. This accessibility depends on local chromatin structure, but also on specific localization within the nucleus (i.e., the nuclear context) [1]. Global genome-wide approaches are routinely used to identify the precise chromatin context for each gene, revealing the existence of several chromatin states that correlate with the expression level of genes that they contain. In *Arabidopsis thaliana* (*Arabidopsis*), up to nine different chromatin states, characterized by specific combination of chromatin modifications, have been defined [2, 3]. However, the position of a gene with respect to a nuclear pore, nuclear lamina, and the nucleolus, as well as interchromosome or intrachromosome interactions, may also affect its transcriptional regulation, so they have to be taken into consideration too [1].

In the nucleoplasm, the largest nuclear body is the nucleolus, which can represent up to half of the nuclear volume in a certain cell

type. The nucleolus is a direct consequence of ribosome biogenesis, but it is also a plurifunctional body implicated in other mechanisms including stress sensing, cell cycle progression, viral replication, and RNP biogenesis [4, 5]. Similar to the nuclear periphery, a thick heterochromatin layer surrounds the nucleolus, constraining chromatin domain mobility within the nucleoplasm [6]. In mammalian cells, nucleolus-associated chromatin domain (NAD) identification showed that all 23 human chromosomes possess at least one region associated with the nucleolus [7, 8]. In *Arabidopsis* leaf cells, all five chromosomes also have regions associating with the nucleolus [9, 10]. In both cases, NADs are composed of genomic regions enriched in silent chromatin modifications associated with transcriptional repression, such as DNA and/or Lysine 9 Histone 3 methylation. The identification of NADs is a complementary approach that allows to directly take into consideration the local environment of a locus. For example, RNA polymerase II tends to be much less abundant within the nucleolus. Thus, whether the association of a gene with the nucleolus influences its transcriptional regulation remains an open question, which requires additional studies, and particularly, the analysis of chromatin states in NADs.

Here we provide a detailed protocol allowing the identification of NADs in plant cells, and more specifically in *Arabidopsis* (Fig. 1). Four major steps are required and presented in this chapter: (1) nuclei and nucleoli isolation by fluorescence-activated cell sorting, (2) nuclear and nucleolar DNA recovery, (3) DNA-seq library preparation and sequencing, and (4) bioinformatic analyses of the sequences to identify NADs, as well as their respective chromatin states. These steps are indicated in the flowchart in Fig. 2.

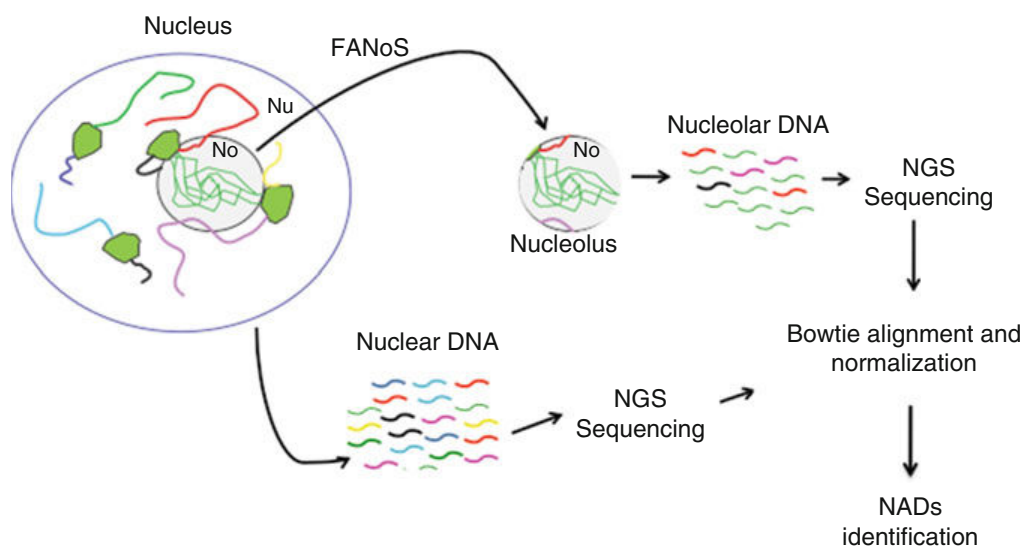


Fig. 1 NADs ID general strategy. Strategy used to purify and sequence total nuclear (Nu) and nucleolar (No) DNA in order to identify NADs by high throughput sequencing

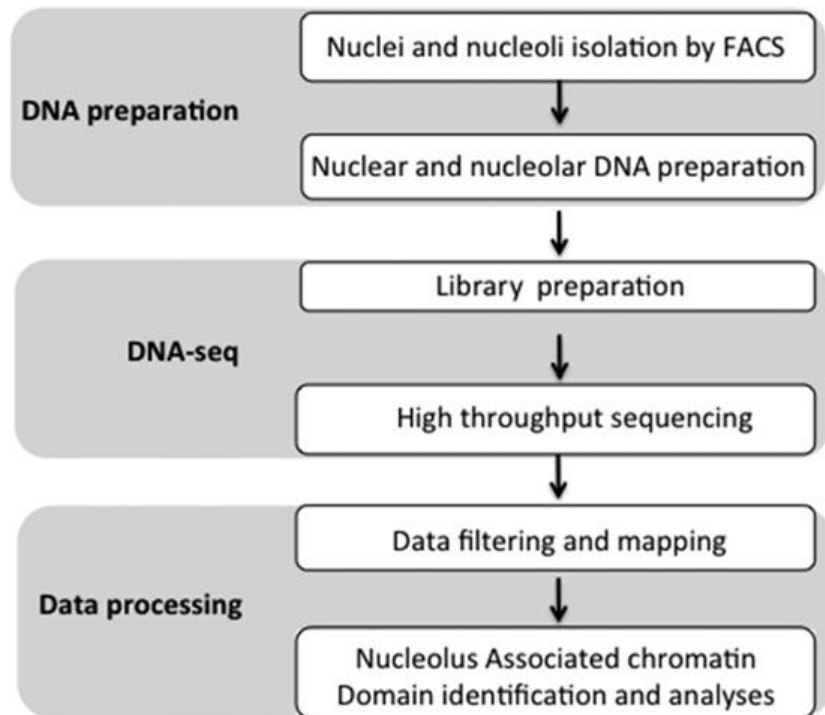


Fig. 2 Flowchart showing a schematic overview of the entire procedure

2 Materials and Reagents

2.1 Materials

1. Plants expressing the chimeric protein YFP fused to FIBRILLARIN 2 to mark the nucleoli (*see Note 1*).
2. Vertically sloped filters with a 30 μm mesh size (*see Note 2*).
3. Gentle sonicator device for DNA shearing (*see Note 3*).
4. Flow cytometer/Cell sorter: Fluorescence-Activated Cell Sorting (FACS) is carried out by flow cytometry with a cell sorter. YFP or GFP is excited with a 488 nm blue laser of 100 mW and signal emission is captured with a 505 long pass filter and a 530/30 (515–545 nm) filter. The sorting procedure is carried out in sterile conditions and the sheath fluid is $1 \times$ PBS.
5. A fluorimetric DNA quantification device distinguishing single stranded and double stranded DNA (e.g., Qubit).
6. Bioanalyzer system to control the size, the quantity and quality of DNA.
7. High-throughput DNA sequencer device.

2.2 Reagents

1. TEN buffer: 10 mM Tris-HCl pH 7.5, 10 mM Na-EDTA pH 8, and 100 mM NaCl. Has to be freshly prepared from stock solution (1 M Tris-HCl pH 8, 500 mM Na-EDTA pH 8, and 5 M NaCl, all kept at room temperature).
2. 36% formaldehyde solution.

3. FACS Solution (or Galbraith's buffer): 45 mM MgCl₂, 20 mM 3-(N-morpholino)propanesulfonic acid (MOPS), 30 mM sodium citrate, 0.1% Triton X-100 adjusted to pH 7 with NaOH. Can be prepared in advance, in 5 mL aliquot and stored at -20 °C for several months.
4. 10× Phosphate buffer saline (PBS): 1370 mM sodium chloride (NaCl), 27 mM potassium chloride (KCl), 100 mM disodium hydrogen phosphate (Na₂HPO₄), and 18 mM potassium dihydrogen phosphate (KH₂PO₄). Can be stored at room temperature for several months.
5. DNase-free RNase A, used at a concentration of 10 µg/mL.
6. DNase-free Proteinase K, used at a concentration of 10 µg/mL.
7. A kit allowing the purification and the concentration of low amount of DNA (*see Note 4*).
8. A kit enabling the preparation of a DNA library compatible for NGS sequencing from a low amount of DNA (*see Note 5*).

2.3 Software and Databases

1. Bowtie2.
2. Bedcoverage.
3. R.
4. Unix system and awk command line.
5. Computing cluster.
6. TAIR10 reference genome sequence database (*see Note 6*): https://www.arabidopsis.org/download_files/Genes/TAIR10_genome_release/TAIR10_chromosome_files/TAIR10_chr_all.fas.

3 Methods

3.1 Nuclei and Nucleoli Isolation by Fluorescence-Activated Cell Sorting

First, nuclear and nucleolar DNA have to be purified from isolated nuclei or nucleoli respectively. Two main protocols are available to isolate nuclei and nucleoli: (1) a protocol based on successive centrifugation steps with a sucrose gradient and variable salt concentrations or (2) a protocol based on the utilization of a fluorescence-activated cell sorter (FACS). However, in (1), protoplasts have to be the starting material [11], while the entire organism and any plant tissue can be used in (2).

Thus, we recommend to purify nuclei or nucleoli by FACS using the approaches named FANS for fluorescence-activated nucleus sorting and FANoS for fluorescence-activated nucleolus sorting [12]. This purification is based on the labeling of nucleoli via a fluorescent marker that delimits the nucleolar position in the cell of interest *in planta*. Hence, plants expressing the nucleolus-

specific protein Fibrillarin 2 (FIB2) coding sequence from *Arabidopsis* (At4G25630), with the Yellow Fluorescent Protein (YFP) at its C-terminal part and under the control of the ubiquitous promoter 35S (YFP:FIB2 plants) are used here (*see Note 1*) [13]. The tissues is prepared for FACS as follows:

1. Fix 1 g tissue in 10 mL of TEN buffer with 4% formaldehyde on a rotor in a cold room for 20 min.
2. Wash twice for 10 min in 10 mL of ice-cold TEN buffer.
3. Mince the tissue with a razor blade in 1 mL FACS buffer (*see Note 7*).
4. Filter the homogenate through a 30 μm mesh diameter nylon membrane into a 1.5 mL tube (*see Note 8*).
5. To sort nuclei by FANS (*see Note 9*), the filtered homogenate can be directly subjected to FACS.
6. To sort nucleoli using FANoS, the filtered homogenate should be sonicated four times for 5 min at medium power using a Bioruptor to liberate the nucleoli, and can then be subjected to FACS.
7. Sorting is carried out with a fluorescent-activated cell sorter (FACS). To facilitate the setting of the sorting window, it is better to first analyse the sample containing the fluorescent marker, then the negative control. Because the nucleolar marker used may not be ubiquitously expressed, we recommend to always perform nuclear sorting using the fluorescent signal of YFP:FIB2.
8. Nucleolar sorting is performed using the YFP:FIB2 signal (*see Note 10*). As for nuclear sorting, the negative control sample (without a nucleolar marker) is used to determine the area where no YFP signal is detected from the particle cloud. For NADs identification by high throughput sequencing, at least 500,000 nuclei and 1 million nucleoli are necessary as starting material.

3.2 Nuclear and Nucleolar DNA Recovery

1. To homogenize the samples, sonicate the sorted nuclei five times for 5 min (with a 30 s ON and 30 s OFF interval) at medium power, using a Bioruptor (*see Note 3*). In parallel, sonicate the sorted nucleoli one time for 5 min using the same settings (*see Note 11*).
2. To disrupt the nuclei or nucleoli, place the samples for 10 min at 95 °C.
3. To deteriorate the RNA present in the samples, incubate them for 30 min at 37 °C with RNase A at a concentration of 10 $\mu\text{g}/\text{mL}$.
4. To degrade the proteins, incubate for 30 min at 45 °C with Proteinase K at a concentration of 10 $\mu\text{g}/\text{mL}$.

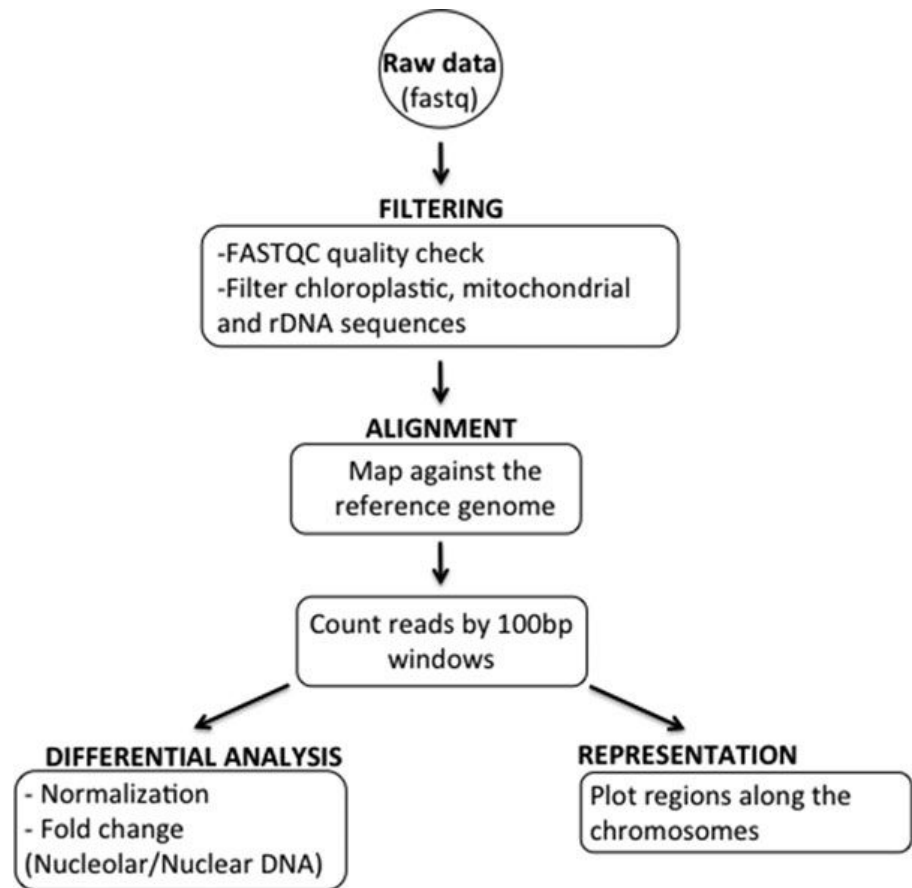


Fig. 3 NADs ID pipeline. Bioinformatic pipeline used to identify NADs from fastQ reads

5. Boil the samples for 10 min at 99 °C to inactivate Proteinase K.
6. For high-throughput DNA sequencing, nuclear and nucleolar DNA is purified and concentrated using the ChIP DNA Clean & Concentrator kit according to the manufacturer's instructions.
7. The Quant-iT PicoGreen dsDNA kit can be used to determine the concentration of the samples, according to the manufacturer's instructions.

3.3 DNA-Seq Library Preparation and Sequencing

The libraries are generated via the Nextera XT DNA library preparation kit or equivalent and dosed with for example the Agilent High Sensitivity DNA kit, according to the manufacturer's instructions. DNA samples are then subjected to high-throughput paired-end sequencing (*see Note 12*).

3.4 DNA-Seq Data Analysis

The pipeline described below is also presented in the Fig. 3.

3.4.1 Processing of the Data

1. Download the TAIR10 genome reference from the TAIR website (*see Note 13*).
2. Separate the organellar (chloroplasts and mitochondria) and ribosomal DNA from the nuclear chromosomes by generating

two different files (For example *chrC_M_rDNA.fa* and *TAIR10_5chr.fasta* respectively).

3. Create an index with chloroplastic and mitochondrial TAIR10 chromosomes and rDNA (*chrC_M_rDNA.fa*) sequences:

```
bowtie2-build index chrC_M_rDNA.fa chrC_M_rDNA
```

4. Filter the nonnuclear sequences from the raw data by mapping the fastq reads against the sequences above, i.e., chrC_M_rDNA, on the cluster, in parallel, with 12 threads per sample:

```
bowtie2 -sensitive -p 12 -un-conc $out-filtered.fastq
-x chrC_M_rDNA -1 $fq1 -2 $fq2 -S $out.sam
```

5. Keep the pairs that did not align with the index (in fastq format).

3.4.2 Dataset Alignment on the Reference Genome

1. Create an index with the five chromosomes of the TAIR10 database.

```
bowtie2-build TAIR10_5chr.fasta TAIR10_5chr
```

2. Map the filtered fastq libraries against the Arabidopsis chromosomes:

```
bowtie2 -sensitive-local -p 12 -x TAIR10_5chr -1
$fq1_A2 -2 $fq2_A2 -S $out.sam
```

3.4.3 Quantification

1. Segment the TAIR10 chromosomes by 100 bp windows:

```
bedtools make windows -g size_chr_TAIR10.txt -w 100 >
TAIR10_100bpwindows.bed
```

2. Convert the sam alignment files in sorted bam files:

```
samtools view -bS $in.sam -o $out.bam and samtools
sort $out.bam $out.sort
```

3. Count for each library, the number of reads that map against each 100 bp window:

```
coveragebed -abam $bam -b TAIR10_100bpwindows.bed >
$out.100pb.bed.coverage
```

4. Reformat the quantification table obtained with bedcoverage for data treatment in R:

```
awk -F '\t' '{print $1 < _ > $2 < _ > $3 < \t > $4}'
$out.100pb.bed.coverage
```


3.4.4 Differential Analysis in R

In the following commands, the initial file names used are *Col0_N* (nuclear DNA) and *Col0_No* (nucleolar DNA).

1. Normalize the counted data by the sequence coverage of each library.

```
tmpRESC$Col0_N_norm ← tmpRESC$Col0_N/sum(tmpRESC$Col0_N)
```

2. Filter the uncovered windows, i.e., keep the 100 bp windows that have a minimum of 10 aligned reads from each DNA sample

```
FRESC ← tmpRESC[which(tmpRESC$mCol0_N > 10), ]
```

3. For each 100 bp window, calculate the mean number of mapped reads between the nuclear (N) or nucleolar (No) DNA replicates of the same condition

```
tmpRESC$mCol0_N ← rowMeans(tmpRESC[, 2:3])
```

4. Determine the fold change (No/N) for each condition.

```
FRESC$FC ← -FRESC$Col0_No_n/FRESC$Col0_N_n
```

5. A region is defined as a NAD when the window is differentially covered, i.e., fold change ≥ 2 or fold change ≤ 0.5 .

```
FRESC$DC ← ifelse(FRESC$FC >= 2 | FRESC$FC <= 0.5, TRUE, FALSE)
```

3.4.5 Plot Regions Along the Chromosomes in R

For each chromosome, plot the log₂ fold change (log₂ No/N) of 100 kb windows according to their chromosome position. 100 kb windows are used here to allow a better visualization of the NADs along the chromosomes. To obtain them, restart the analyses from Subheading 3.4.3, using 100 kb instead of 100 bp windows. Every 100 kb window is represented by a black dot; nucleolus enriched genomic regions above the threshold (NADs) are shown in grey.

```
plot(log2(No/N), ylim=c(-1.5, 3.5), col=ifelse(log2(No/N) > 0.5, « gray », « black »), pch=16)
```

An example of a chromosome plot is presented in Fig. 4.

3.4.6 Define the chromatin states of the NADs

1. Create a bed file for each chromatin state in TAIR10 reference genome starting from the Supplemental Data Set 2 published in [3]:

```
awk -F "\t" '{print "chr"$1"\t"$2"\t"$3"\tstate1"}'
State1_TAIR.csv > reformat-State1_TAIR.txt
cat reformat-State* > reformat-State_all_TAIR.bed
bedtools sort -i reformat-State_all_TAIR.bed > reformat-
State_all_TAIR.sort.bed
```

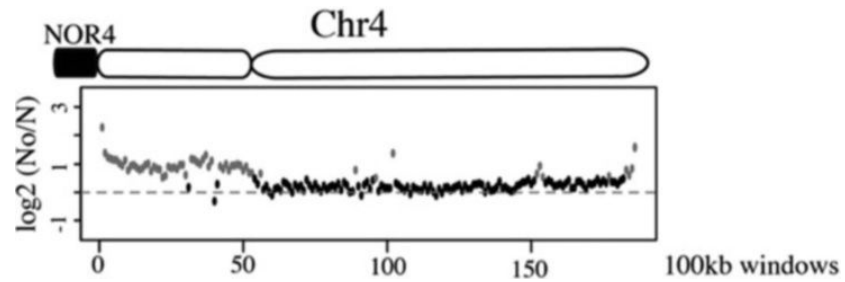


Fig. 4 Visualisation of NADs position along the chromosome. The relative enrichment of a given genomic segment in the nucleolus can be visualized using a chromosome plot. In this example, nucleolar and nuclear DNA from wild-type Col-0 plants were used. The *y axis* displays the fold change ratio No/N. Each dot represents a 100-kb window along chromosome 4. Nucleolus enriched genomic regions above the threshold are shown in *grey*

2. Intersect the NADs with the chromatin state bed files:

```
bedtools intersect -a $TAIR10_states.bed -b NADS_re-
gions.bed -wa -wb
```

3. Sum for each chromatin state the number of NADs from the reference genome:

```
with home-made perl program
perl parse_state.pl > sum_state_TAIR10.tab
```

4. Create a histogram showing the proportion of sequences in each chromatin state:

```
barplot((ref$V1/sum(ref$V1)*100), col="lightblue",
border="blue", names.arg=ref$V2, las=2, ylim=c(0, 30))
```

4 Notes

1. The chimeric protein YFP::FIB2 is expressed under control of the 35S promoter, as described in [9]. However, other promoters including cell-specific promoters can be used to isolate nucleoli from specific cell-types.
2. Depending on the nozzle size in the FACS machine, 50 μ m mesh sizes can also be used.
3. Because the sonication procedure depends on the device used, we describe here the setting used for the Bioruptor standard sonicator device (Diagenode). If an alternative sonication device is used, make sure to adapt the settings properly.
4. We used the Nextera XT DNA library preparation kit (Illumina, USA) at this stage, but alternative kits could also be used here.

5. For instance, the ChIP DNA Clean & Concentrator™ kit (Zymo Research, USA) could be used at this stage.
6. The reference genome from Araport, 11 can also be used here: <https://www.araport.org/data/araport11>
7. Nuclei isolation by FACS can be performed using the fluorescent signal of the YFP:FIB2 marker. Alternatively, FANS can be performed after DNA labeling with a DNA marker like DAPI or propidium iodide.
8. To mince the samples, you can use a plastic weight boat or the interior of a petri dish.
9. It is important to use low retention eppendorf tubes because nucleoli tend to stick on the surface of regular tubes.
10. The size of the nucleoli detected also varies because they tend to aggregate with each other once released from the nucleus.
11. When sorted by FACS, each nucleus or nucleolus is isolated in an approximately 3 nL drop. Therefore, a centrifugation step can be added here: 100 g 15 min. Keep only 100 µL of the supernatant and resuspend the nuclei or nucleoli in this volume.
12. For a better alignment of the reads along the reference genome, paired-end sequencing of at least 100 bp is recommended.
13. Of course, any reference genome could be used here.

Acknowledgments

This work was supported by the CNRS and ANR JCJC NucleoReg (ANR-15-CE12-0013-01) to FP.

References

1. Bickmore WA, van Steensel B (2013) Genome architecture: domain organization of interphase chromosomes. *Cell* 152:1270–1284. doi:[10.1016/j.cell.2013.02.001](https://doi.org/10.1016/j.cell.2013.02.001)
2. Roudier F, Ahmed I, Berard C et al (2011) Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. *EMBO J* 30:1928–1938. doi:[10.1038/emboj.2011.103](https://doi.org/10.1038/emboj.2011.103)
3. Sequeira-Mendes J, Araguez I, Peiro R et al (2014) The functional topography of the Arabidopsis genome is organized in a reduced number of linear motifs of chromatin states. *Plant Cell* 26:2351–2366. doi:[10.1105/tpc.114.124578](https://doi.org/10.1105/tpc.114.124578)
4. Boulon S, Westman BJ, Hutten S et al (2010) The nucleolus under stress. *Mol Cell* 40:216–227. doi:[10.1016/j.molcel.2010.09.024](https://doi.org/10.1016/j.molcel.2010.09.024)
5. Pederson T (2011) The nucleolus. *Cold Spring Harb Perspect Biol*. doi:[10.1101/cshperspect.a000638](https://doi.org/10.1101/cshperspect.a000638)
6. Ferreira J, Paoletta G, Ramos C, Lamond AI (1997) Spatial organization of large-scale chromatin domains in the nucleus: a magnified view of single chromosome territories. *J Cell Biol* 139:1597–1610
7. van Koningsbruggen S, Gierlinski M, Schofield P et al (2010) High-resolution whole-genome sequencing reveals that specific chromatin

- domains from most human chromosomes associate with nucleoli. *Mol Biol Cell* 21:3735–3748. doi:[10.1091/mbc.E10-06-0508](https://doi.org/10.1091/mbc.E10-06-0508)
8. Nemeth A, Conesa A, Santoyo-Lopez J et al (2010) Initial genomics of the human nucleolus. *PLoS Genet* 6:e1000889. doi:[10.1371/journal.pgen.1000889](https://doi.org/10.1371/journal.pgen.1000889)
 9. Pontvianne F, Carpentier M-C, Durut N et al (2016) Identification of nucleolus-associated chromatin domains reveals a role for the nucleolus in 3D organization of the *A. thaliana* genome. *Cell Rep* 16:1574–1587. doi:[10.1016/j.celrep.2016.07.016](https://doi.org/10.1016/j.celrep.2016.07.016)
 10. Picart C, Pontvianne F (2016) Plant nucleolar DNA: green light shed on the role of Nucleolin in genome organization. *Nucleus* Austin Tex 8:11–16. doi:[10.1080/19491034.2016.1236167](https://doi.org/10.1080/19491034.2016.1236167)
 11. Hacot S, Coute Y, Belin S et al (2010) Isolation of nucleoli. *Curr Protoc Cell Biol* Chapter 3: Unit3.36. doi:[10.1002/0471143030.cb0336s47](https://doi.org/10.1002/0471143030.cb0336s47)
 12. Pontvianne F, Boyer-Clavel M, Saez-Vasquez J (2016) Fluorescence-activated nucleolus sorting in Arabidopsis. *Methods Mol Biol* 1455:203–211. doi:[10.1007/978-1-4939-3792-9_15](https://doi.org/10.1007/978-1-4939-3792-9_15)
 13. Pontvianne F, Blevins T, Chandrasekhara C et al (2013) Subnuclear partitioning of rRNA genes between the nucleolus and nucleoplasm reflects alternative epiallelic states. *Genes Dev* 27:1545–1550. doi:[10.1101/gad.221648.113](https://doi.org/10.1101/gad.221648.113)

3.2.2 Comparaison des NADs entre deux échantillons

Pour pouvoir comparer l'enrichissement nucléolaire entre deux échantillons (stressé et contrôle, par exemple), un nouveau FC est calculé à partir des FCs des deux conditions à comparer (FC/FC). Proches de 1, cela signifie qu'il n'y a pas de différences d'association avec le nucléole dans les conditions testées. Si le FC est élevé ($>1,5$) ou proche de 0, on peut considérer un gain ou une perte d'association avec le nucléole, respectivement. Pour visualiser ces différences d'association avec le nucléole, le \log_2 du FC dans la condition étudiée divisé par le FC chez le témoin ($\log_2 \text{FC/FC}$) de chaque fenêtre est représenté (Figure 3.11).

3.2.3 Développement de nouvelles lignées FIB2 ::YFP pour l'étude des NADs

Pour isoler des noyaux ou nucléoles par FANS et FANoS, respectivement, des lignées transgéniques exprimant le marqueur nucléolaire FIB2 ::YFP sont utilisées. J'ai donc produit le plasmide pEarlyGate101 ::FIB2 ::YFP (pEG ::FIB2 ::YFP) avec la méthode de clonage Gateway[®] (Thermo Fisher Scientific, États Unis), en introduisant la séquence codante de la FIB2 dans le plasmide de destination pEarlyGate101 (pEG101), entre le promoteur ubiquitaire 35S CMV et la séquence codante de la YFP (*cf.* Figure 3.6 et partie Matériel et Méthodes).

Le plasmide a été ensuite validé par séquençage Sanger et utilisé pour transformer des plantes *via* insertion d'ADN-T. La bactérie *Agrobacterium tumefaciens* (Beijerinck et van Delden, 1902) est alors utilisée pour transformer les plantes *via* leurs boutons floraux (CLOUGH et BENT, 1998). Après transformation, des lignées de la deuxième génération de transformants (T2) homozygotes pour l'insertion ont été sélectionnées (*cf.* partie Matériel et Méthodes).

Un des objectifs de l'ANR nucleoReg a été de révéler l'existence potentielle d'une variabilité intraspécifique au niveau des NADs. Pour cela, les écotypes Cvi-O (Cape Verde Islands), C24 et Ler (Lansberg) d'*A. thaliana* ont été transformés et sélectionnés. Ils seront utilisés ultérieurement par l'équipe. De manière intéressante, le *knob* hk4S du bras court du chromosome 4 n'est pas présent dans les écotypes Ler et C24 (Figure 3.2 et FRANZ *et al.* 2016). L'identification des NADs dans ces écotypes permettra de déterminer si en absence du *knob* hk4S, la totalité du bras court du chromosome 4 continue d'être associé avec le nucléole.

De plus, dans l'écotype Ler, les deux NORs sont actifs (RABANAL *et al.* 2017a). Par conséquent, l'identification des NADs dans cet écotype permettra d'analyser l'impact de l'activité transcriptionnelle du NOR2 dans l'organisation du génome autour du nucléole dans

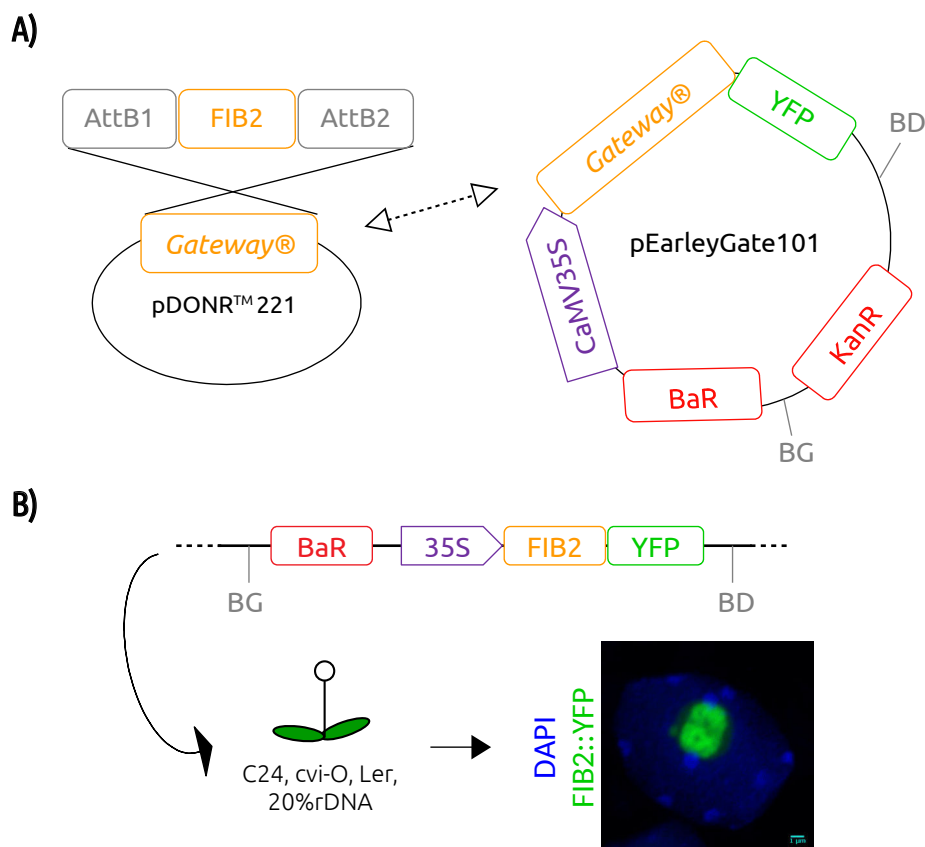


Figure 3.6 – Création de nouvelles lignées FIB2 ::YFP pour l'étude des NADs. **A)** La séquence codante du gène *FIBRILLARINE2* a été amplifiée par PCR (les sites de recombinaison homologue Att B1 et B2, de part et d'autre de la séquence codante, étaient incluses dans les amorces). Ce fragment d'ADN a été inséré dans le plasmide donneur pDONR221 par recombinaison homologue. Le plasmide de destination pEarleyGate101, possède un gène de sélection pour les bactéries (KanR : gène de résistance à la Kanamycine) et entre les bordures gauche (BG) et droite (BD) de l'ADN-T : un gène de sélection pour les plantes (BaR : gène de résistance au BASTA), la séquence codante de l'YFP sous contrôle du promoteur ubiquitaire 35S CMV, et entre ces deux derniers se trouve la cassette Gateway®. Cette cassette permet d'insérer la séquence codant la FIB2, à partir du pDONR™221, en amont de l'YFP. **B)** Des *Agrobacterium tumefaciens* ont été transformées avec le plasmide pEG101 ::YFP ::FIB2, et utilisées pour insérer ce dernier dans différents écotypes et la lignée 20%ADNr d'*A. thaliana* pour l'étude des NADs.

un contexte naturel, ce qui permettra de compléter les données obtenues dans les mutants *nuc1* et *hda6* dans lesquels les NOR2 et NOR4 sont actifs (PONTVIANNE *et al.* 2016b).

Deuxièmement, une lignée Col-0 avec 80% de copies de gènes d'ARNr 45S en moins (20%ADNr) a également été transformée. La lignée 20%ADNr a été créée par complémentation fonctionnelle des mutants *fasciata1* et *fasciata2* (Figure 3.7A et PAVLIŠTOVÁ *et al.* 2016). Les protéines FASCIATA1 (FAS1) et FAS2, ainsi que le SUPPRESSEUR MULTICOPIE DE IRA1 (MSI1) sont les trois sous-unités du facteur d'assemblage de la chromatine 1 (CAF1) chez *A. thaliana*.

CAF1 est une histone chaperonne qui réalise la déposition des histones H3 et H4 sur l'ADN lors de sa réplication. Lorsque *FAS1* ou *FAS2* sont mutés, il y a une réduction de la longueur des télomères des deux extrémités des 5 chromosomes et une perte progressive de copies de gènes d'ARNr 45S. Les autres séquences répétées ne sont pas affectées (Figure 3.7B et MOZGOVÁ *et al.* 2010).

En effet, des unités de répétition sont perdues au niveau des deux NORs (NOR2 et NOR4) et après 4 générations, il ne reste que 20% de la totalité des gènes 45S, le minimum requis *a priori* pour maintenir le niveau de production d'ARNr et permettre la survie de la plante (MOZGOVÁ *et al.* 2010; PONTVIANNE *et al.* 2013).

Le croisement entre la 4^{ème} génération des mutants *fasciata1* et *fasciata2*, complémente les deux mutations et les télomères récupèrent leur longueur normale. Cependant, le nombre de copies de gènes d'ARN 45S n'est pas rétabli. Au contraire, la descendance du croisement entre *fas1* et *fas2* présente une grande hétérogénéité du nombre de copies de gènes d'ARNr 45S, ce qui a permis de sélectionner deux lignées (6 et 9) 20%ADNr, dans lesquelles uniquement 20% de la totalité des copies sont présentes. Cette quantité est maintenue au cours des générations suivantes (PAVLIŠTOVÁ *et al.* 2016). La lignée 6 (L6) 20%ADNr, a été transformée avec le plasmide pEG::FIB2::YFP, et les NADs ont été identifiés (voir le prochain chapitre de résultats).

3.3 Les NADs en condition de choc thermique

Le choc thermique (CT) induit la reprogrammation transcriptionnelle de centaines de gènes, la dénaturation de protéines et l'accumulation d'Espèces Réactives de l'Oxygène (ROS) (WANG *et al.* 2013). Chez *Arabidopsis thaliana*, cette reprogrammation s'accompagne d'une décondensation de la totalité des chromocentres dans un grand nombre de cellules (PECINKA *et al.* 2010), qui est nécessaire à la réponse au stress thermique (WANG *et al.* 2013).

Ainsi, suite à un CT de 24 heures à 37 degrés Celsius (°C), environ 70% des noyaux de feuilles d'*A. thaliana* présentent des chromocentres décondensés. La protéine HIT4 (pour HEAT-INTOLERANT 4), localisée au niveau des chromocentres, est impliquée dans ce

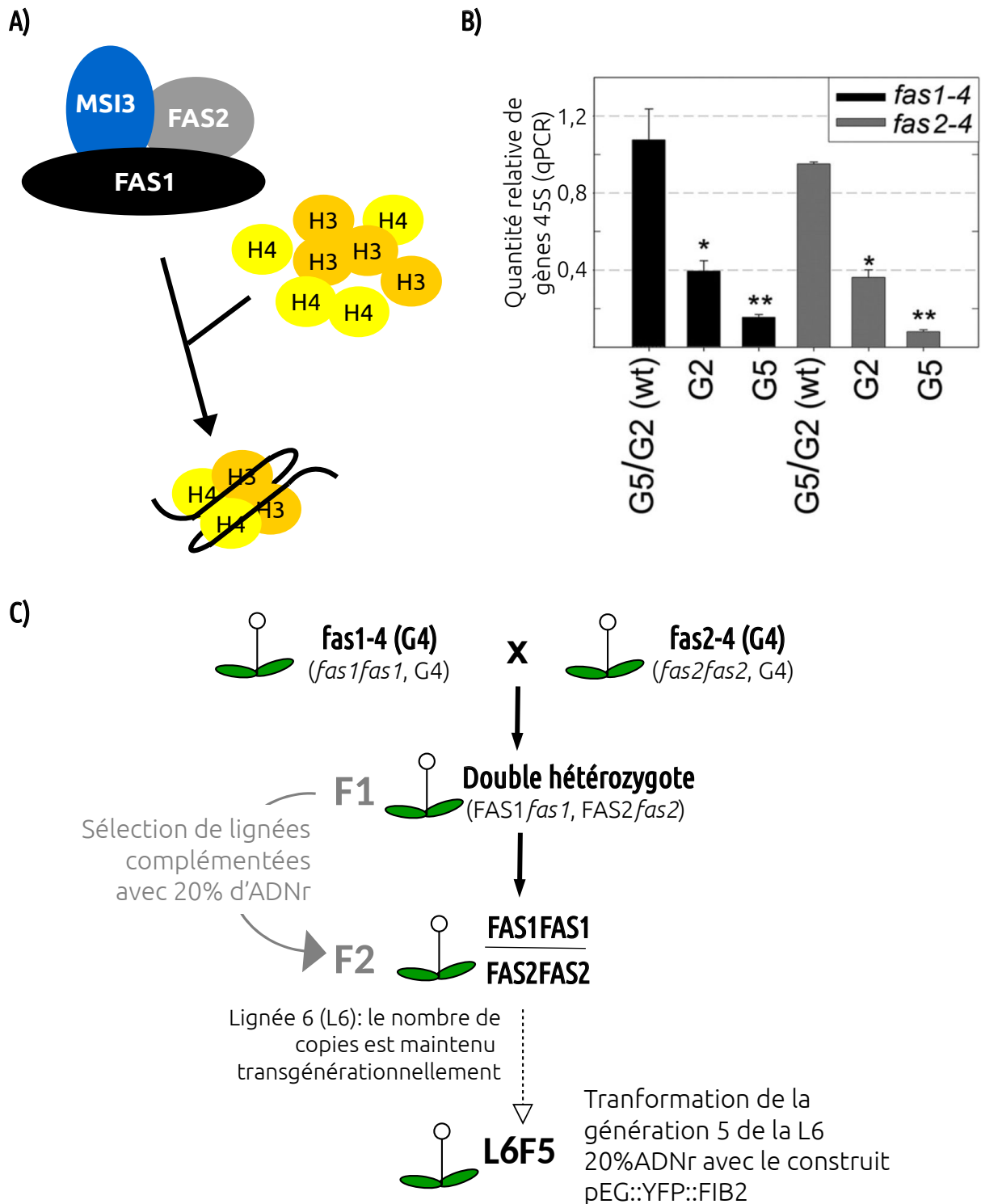


Figure 3.7 – Développement de la lignée 20%ADNr - pEG ::FIB2 ::YFP. **A)** L’histone chaperonne Chromatin Assembly Factor 1 (CAF1) des plantes est composée de trois sous-unités : MSI3, FAS1 et FAS2. Ce complexe dépose les histones H3 et H4 lors de la réplication de l’ADN. **B)** Histogramme représentant la quantité relative de copies de gènes d’ARNr 45S dans les mutants *fasciata1* et *2* en fonction de la génération (adapté de MOZGOVÁ *et al.* 2010). On constate une perte progressive de copies de gènes d’ARNr 45S dans ces mutants, qui après quatre générations ne contiennent que 20% de la totalité des copies de gènes d’ARN ribosomique 45S. **C)** Le croisement entre la génération 4 des mutants *fasciata1* et *2*, permet la complémentation fonctionnelle des deux protéines. La descendance récupère la totalité des télomères et de façon hétérogène le nombre de copies de gènes d’ARNr 45S, sauf les lignées 20%ADNr 6 et 9. Dans les lignées L6 et L9, la quantité relative de copies est maintenue de façon transgénérationnelle (pendant plus de 6 générations) à 20% par rapport à la plante sauvage (PAVLIŠTOVÁ *et al.* 2016). Des plantes de la cinquième génération de la lignée 6 (L6F5) ont été utilisées pour créer la lignée 20%ADNr - pEG ::YFP ::FIB2

relâchement de l'hétérochromatine qui induit l'expression des répétitions centromériques de 180 pb et d'autres séquences répétées comme l'élément transposable ONSEN (MIROUZE *et al.* 2009; WANG *et al.* 2013).

Ce remodelage de la chromatine semble essentiel à la survie de la plante lors d'un CT. En effet, les mutants *hit4*, dans lesquels cette décondensation des chromocentres ne se produit pas, résistent moins bien au stress thermique. Par ailleurs, un CT affecte la fonction nucléolaire et la biogenèse des ribosomes, ce qui entraîne la relocalisation ou dissociation des protéines nucléolaires et potentiellement, des changements structuraux au niveau des trois compartiments du nucléole (FC, DFC et GC) (Figure 3.8 et JACOB *et al.* 2013; PONTVIANNE, 2016).

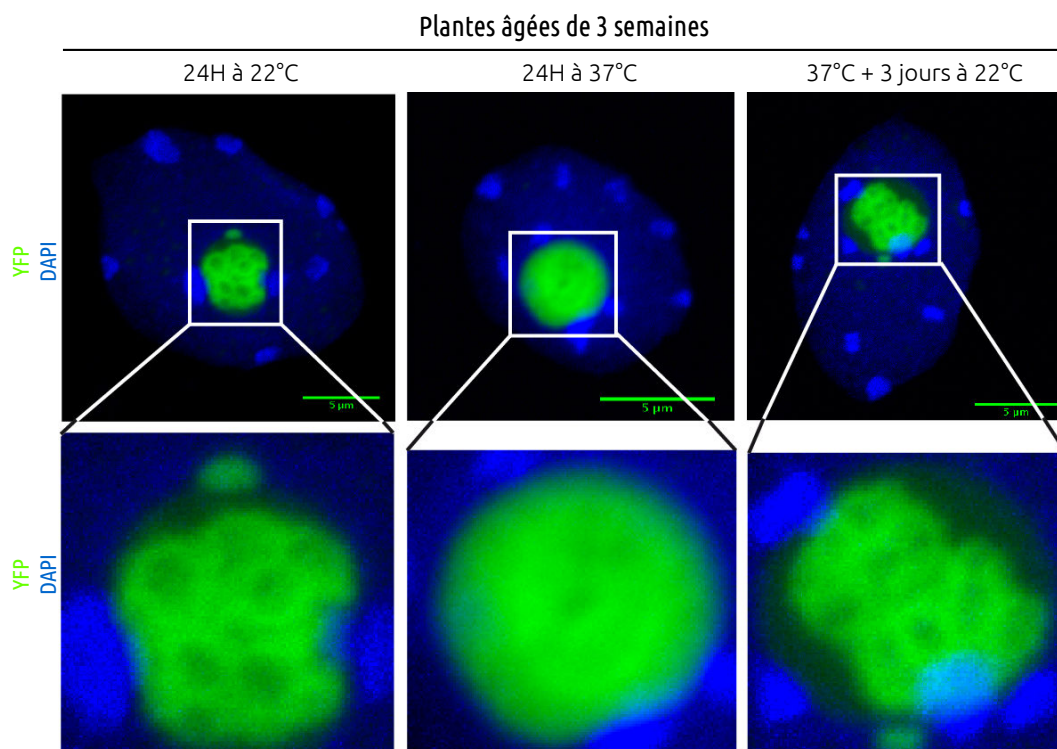


Figure 3.8 – Structure du nucléole d'*Arabidopsis thaliana* suite à un choc thermique. De gauche à droite : noyaux de la lignée Col-0 pEG ::YFP ::FIB2 colorés au DAPI et observés au microscope confocal en conditions contrôles (22°C), suite à un choc thermique de 24 heures à 37°C et trois jours après ce choc. Suite à un choc thermique, on ne distingue plus les trois compartiments du nucléole (PONTVIANNE, 2016).

Afin de déterminer si ces changements structuraux, à la fois au niveau du nucléole et de la chromatine, induisent des changements au niveau des domaines chromatiniens associés au nucléole, j'ai identifié les NADs suite à un CT. Des graines de la lignée transgénique Col-0 pEG101 ::FIB2 ::YFP (PONTVIANNE *et al.* 2013), ont été semées en terre et cultivées en chambre de culture Percival® pendant trois semaines. Ensuite, un stress thermique de 24 heures à 37°C a été réalisé en maintenant la même photopériode, intensité lumineuse et humidité. Les lots témoins ont été maintenus à 22°C durant le stress.

Pour l'identification des NADs, un échantillon d'ADN nucléaire et deux d'ADN nucléaire de plantes de 21 jours stressées et de plantes contrôle ont été séquencés (séquençage illumina® en «paired-end», 2x150 pb) par la plateforme de séquençage Bioenvironnement de l'Université de Perpignan *Via Domitia* (cf. Matériel et Méthodes).

En parallèle, une analyse transcriptomique a également pu être réalisée. En effet, des données d'ARN-seq (séquençage à haut débit du transcriptome) des ARN polyadénylés (pour sélectionner les ARNm stables) de plantes de 21 jours cultivées dans les mêmes conditions contrôle puis de stress, étaient déjà disponibles au laboratoire. J'ai donc pu réaliser l'analyse différentielle de l'expression des gènes lors de ce CT afin de comparer ces différences d'expression avec les différences d'association avec le nucléole. Le détail de cette analyse bioinformatique se trouve dans la partie Matériel et Méthodes de ce manuscrit de thèse. Les caractéristiques de séquençage et d'alignement des différentes bibliothèques de séquences (ADN ou ARN) se trouvent en Annexe 1.

Remarquablement, malgré le remaniement important de l'architecture chromatinienne accompagné d'une décondensation de la totalité des chromocentres (WANG *et al.* 2015b) et une déstructuration du nucléole (PONTVIANNE, 2016) lors d'un CT chez *A. thaliana*, on ne constate pas de différences au niveau des domaines chromatinien associés au nucléole. En effet, on peut voir que la distribution des régions associées au nucléole sont très similaires aux conditions contrôles (Figure 3.9A). À l'échelle chromosomique, on retrouve en effet, l'association du bras court du chromosome 4 et de sa région centromérique, des régions subtélomériques et de quelques petites régions hétérochromatiques au sein des 5 chromosomes.

Néanmoins, à l'échelle génique, l'ensemble des gènes associés sont assez différents. Une grande partie des NAD-gènes identifiés avec les échantillons d'ADN provenant des plantes témoin (580 sur 884 identifiés), présentent un appauvrissement de leur association avec le nucléole, et 716 gènes qui ne sont pas associés au nucléole chez les plantes qui sont restées à 22°C, deviennent des NAD-gènes suite à un CT. Cependant, la majorité de ces derniers proviennent également du bras court du chromosome 4 (Figure 3.9B).

De son côté, l'analyse transcriptomique ne montre pas de relation entre cet enrichissement ou appauvrissement d'association avec le nucléole et des changements d'expression. Effectivement, le choc thermique induit la reprogrammation transcriptionnelle de centaines de gènes, au total 527 gènes sont sur-accumulés et 640 sont inhibés ($p_{adj} < 0.05$ et $FC > 4$). Parmi les gènes activés, seuls 7 se dissocient du nucléole suite à un CT. Dans le même temps, parmi les 640 gènes réprimés, uniquement 23 font partie des NAD-gènes suite à un CT (Figure 3.9C). Il est à noter que ces 7 ou 23 gènes ne sont pas liés à la réponse au CT et ne sont pas impliqués dans un processus biologique particulier (analyse *Gene Ontology* avec le logiciel ThaleMine d'Araport ; [https :](https://)

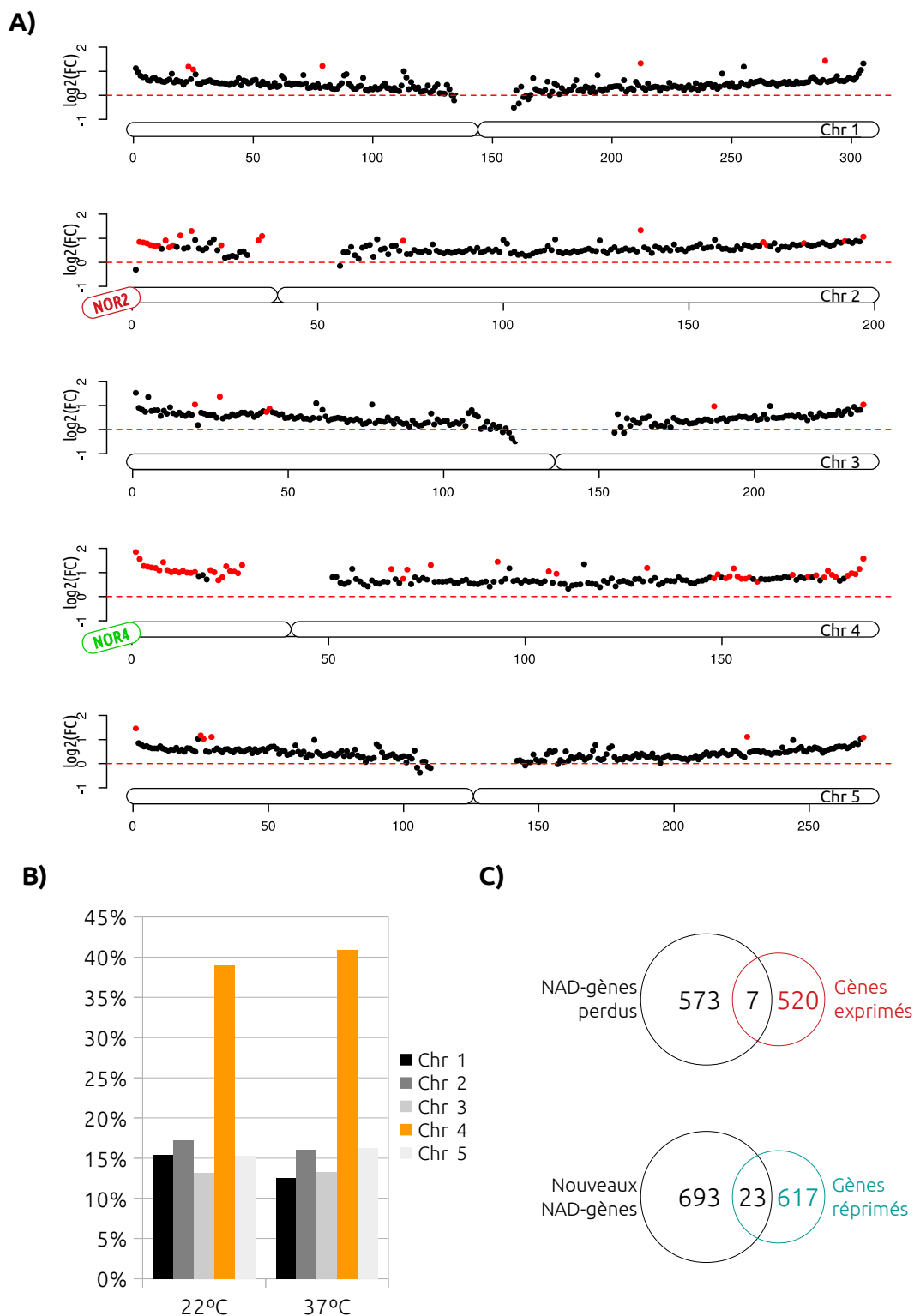


Figure 3.9 – Les NADs d'*Arabidopsis thaliana* en condition de choc thermique. **A)** Enrichissement (ou fold change) nucléolaire de fenêtres génomiques de 100 kilobases chez l'écotype sauvage Col-0 suite à un choc thermique de 24h à 37°C. Chaque fenêtre est positionnée le long du chromosome et en ordonnées est représenté le $\log_2(\text{FC})$. Les NADs sont représentés en rouge, seuil = 2. Les données des régions péri-centromériques ont été masquées, car la difficulté d'alignement de ces séquences sont source potentielle d'artefacts et peuvent mener à des mauvaises interprétations. **B)** Proportion de NAD-gènes par chromosome (en %) en conditions normales (22°C) ou suite à un CT (37°C). **C)** Diagrammes de Venn représentant la superposition entre les gènes activés lors d'un CT et ceux qui perdent leur association avec le nucléole (en haut); et la superposition entre les gènes inhibés et ceux qui s'associent avec le nucléole lors d'un CT (en bas).

//apps.araport.org/thalemine/ et KRISHNAKUMAR *et al.* 2015).

3.4 Identification des NADs suite à la transition florale

Lors de la transition florale (TF), le passage de la phase végétative à la phase de reproduction sexuée, s'accompagne d'une reprogrammation génétique et d'un remaniement de la chromatine (TESSADORI *et al.* 2007). Cette transition se déroule généralement après 4 semaines de croissance chez *A. thaliana*, en fonction des conditions environnantes.

En 2007, Tessadori et des collaborateurs ont mesuré l'hétérogénéité de la fraction d'hétérochromatine relative (RHF) et de l'index hétérochromatique (HX) des noyaux de cellules du mésenchyme foliaire, à différents stades de développement autour de la TF. La RHF est le ratio entre la surface et l'intensité fluorescente de tous les chromocentres par rapport à la surface et l'intensité fluorescente totale du noyau. L'HX est le pourcentage de noyaux présentant des chromocentres bien définis sur l'ensemble analysé.

Ainsi, environ 4 jours avant et jusqu'à 2 jours après l'apparition des boutons floraux (~27 jours après la germination, juste avant que la tige de la hampe florale soit visible), ils ont mis en évidence une réduction des deux indices RHF et HX. Cela suggère une décondensation de l'hétérochromatine du génome liée à la floraison. Pour vérifier cette hypothèse, ils ont réalisé la même étude dans différents écotypes ou mutants où la TF se déroule à d'autres moments du développement. Une décondensation des chromocentres lors de la TF était observable chez toutes les plantes étudiées, indépendamment de l'âge des plantes lors de la transition (TESSADORI *et al.* 2007).

Il est possible que ces remaniements de la structure 3D du génome soient liés à la reprogrammation transcriptionnelle nécessaire au changement de phase développementale. Dans ce cas, la localisation nucléaire de domaines chromatiniens ainsi que de gènes, maintenus inhibés ou activés pendant la phase de reproduction sexuée, peut être modifiée. Il est donc possible que cette transition développementale induise des différences dans la composition des NADs. Pour tester cette hypothèse, j'ai identifié les NADs juste après la TF, lors de la reconstitution des chromocentres.

Le moment de la transition florale dépend de l'espèce mais également des conditions de culture. En effet, la TF dans les individus de l'étude de Tessadori et ces collaborateurs (2007) se déroule plus précocement que chez le même écotype Col-O, dans les conditions de croissance des chambres de culture du laboratoire. Par conséquent, j'ai d'abord réalisé un suivi phénotypique de la lignée transgénique Col-0 pEG::FIB2::YFP, accompagné d'une analyse microscopique de noyaux de cellules de feuilles colorés au DAPI. Cette analyse m'a permis de définir de façon précise le moment de la TF de cette lignée au niveau des feuilles dans les conditions de culture de notre laboratoire. L'isolement

de nucléoles a été réalisé avant (plantes âgées de 21 jours) et après la transition florale (plantes âgées de 29 jours), juste après le remaniement chromatinien associé à cette transition.

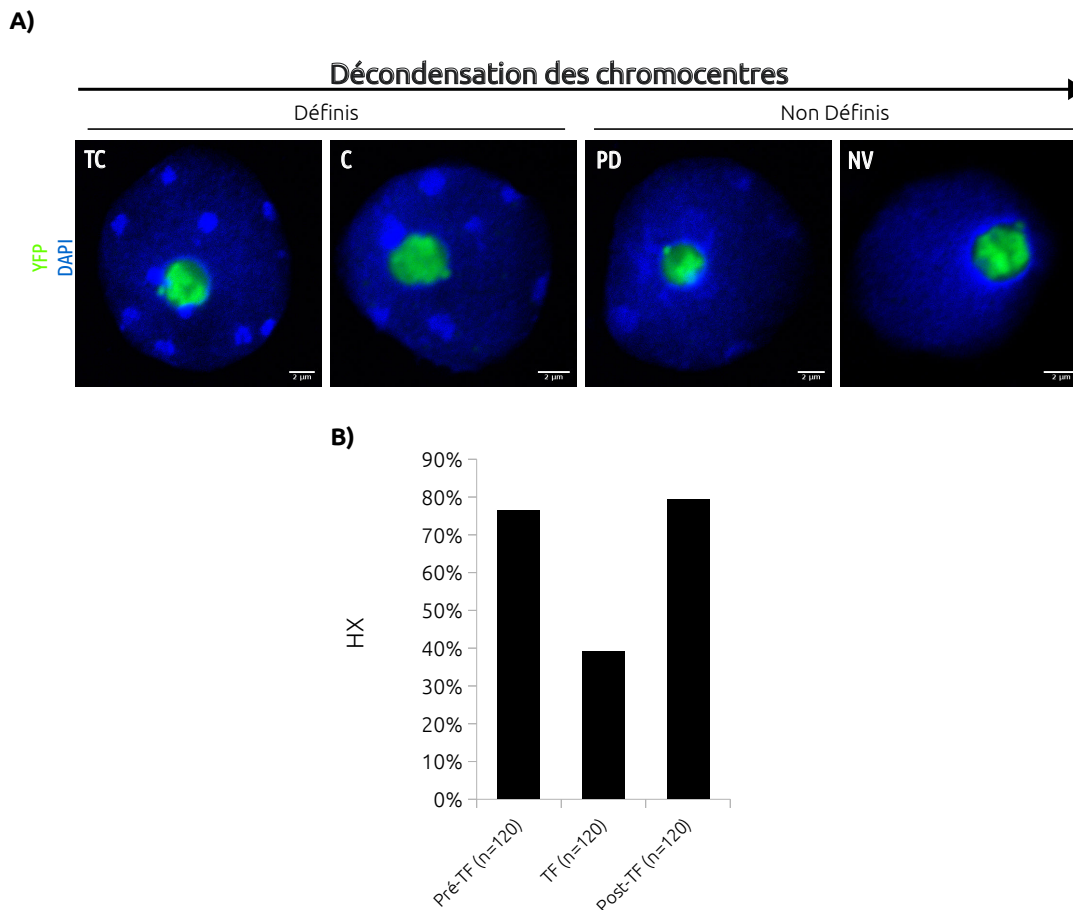


Figure 3.10 – Décondensation des chromocentres des noyaux de cellules de feuille d'*A. thaliana* pendant la transition florale. **A)** Noyaux isolés Col-0 colorés au DAPI et observés au microscope confocal. De gauche à droite : chromocentres très condensés (TC), condensés (C), peu définis (PD) ou non visibles (NV). La distribution de la FIB2 ::YFP correspond au signal vert. **B)** Pourcentage de noyaux présentant des chromocentres définis sur la totalité de noyaux observés (indes hétérochromatique ou HX), avant la transition florale (Pré-TF), lors de la TF et après la transition florale (Post-TF). Lors de la TF, environ 60% (sur 120) des noyaux analysés présentent des chromocentres mal définis ou non détectables.

J'ai débuté le suivi à partir du 22ème jour de croissance jusqu'au trente-unième jour après la germination. Chaque jour, deux extractions de noyaux de feuilles d'un total de 6 individus ont été réalisées, ainsi qu'une coloration au DAPI (4',6-diamidino-2-phénylindole) de ces noyaux pour leur observation au microscope confocal (*cf.* partie Matériel et Méthodes). Les différentes préparations cytologiques ont été regroupées en fonction du phénotype de la plante, suivant les observations de TESSADORI *et al.* 2007 : les noyaux provenant de plantes où la rosette n'est constituée que de feuilles (autour de ~23 jours de croissance), ceux provenant de plantes où des petites feuilles centrales ren-

ferment les boutons floraux (plantes âgées d'environ 25 à ~28 jours) et finalement, ceux issus de plantes où la tige de la hampe florale est visible. Ces trois groupes correspondent aux stades Pré-TF, TF et Post-TF.

En interphase, la coloration DAPI des noyaux d'*A. thaliana* permet la détection en général de 8 à 10 chromocentres bien définis (FRANSZ *et al.* 2002). En fonction de la forme et intensité fluorescente de ces derniers, les noyaux ont été classés en deux groupes : noyaux avec des chromocentres définis (très condensés ou condensés) ou avec des chromocentres non définis (peu condensés ou non visibles) (Figure 3.10A). En accord avec les publications précédentes, l'observation microscopique des noyaux de cellules de feuilles montre une décondensation des chromocentres lors de la TF (Figure 3.10). En effet, le pourcentage de noyaux présentant des chromocentres définis (HX) pas de 77% à 40% (Figure 3.10B).

Pour avoir un premier aperçu de l'impact de la TF dans l'identité des NADs, j'ai isolé par FANoS deux échantillons de 600.000 nucléoles, à partir de feuilles de plantes âgées de 29 jours, chez lesquelles la tige de la hampe florale était visible. Des données de séquençage d'ADN nucléaire de plantes de la même lignée Col-0 pEG::FIB2::YFP âgées de 21 jours et ayant poussé dans les mêmes conditions, ont été utilisées pour cette première étude pilote.

En effet, un échantillon d'ADN nucléaire et deux d'ADN nucléolaire de plantes Col-0 pEG::FIB2::YFP de 21 jours ont été séquencés (séquençage illumina® en «paired-end», 2x150 pb) par la plateforme de séquençage Bioenvironnement de l'Université de Perpignan *Via Domitia* (cf. Matériel et Méthodes). L'enrichissement nucléolaire des fenêtres de 100 kb a été calculé (Figure 3.1). Les caractéristiques de séquençage et d'alignement des différentes bibliothèques de séquences ADN se trouvent en Annexe 1. Ensuite, l'enrichissement nucléolaire des fenêtres de 100kb après transition florale a été comparé à l'enrichissement chez les plantes âgées de 21 jours.

Dans la Figure 3.11 est représenté l'enrichissement nucléolaire des fenêtres génomiques de 100 kb après la TF par rapport à l'enrichissement nucléolaire de ces dernières chez les plantes de 21 jours (FC/FC). On observe que $\log_2(\text{FC}/\text{FC})$ de ces fenêtres dans les cinq chromosomes est proche de 0 (*i.e.* : FC/FC=1), ce qui n'indique pas de changements au niveau des NADs induits par la TF.

A l'échelle génique, comme ce qui a été observé suite à un CT, l'identité des gènes associés au nucléole diffère grandement entre les deux conditions (Figure 3.12A). Par ailleurs, il y a peu de superposition entre les gènes identifiés en 2016 (PONTVIANNE *et al.* 2016b) et les gènes trouvés lors de ce nouveau séquençage, alors qu'il s'agit de plantes de la même lignée et du même âge ayant poussé dans les mêmes chambres de culture (Figure 3.12A). La même observation est faite lorsque ces données sont comparées avec

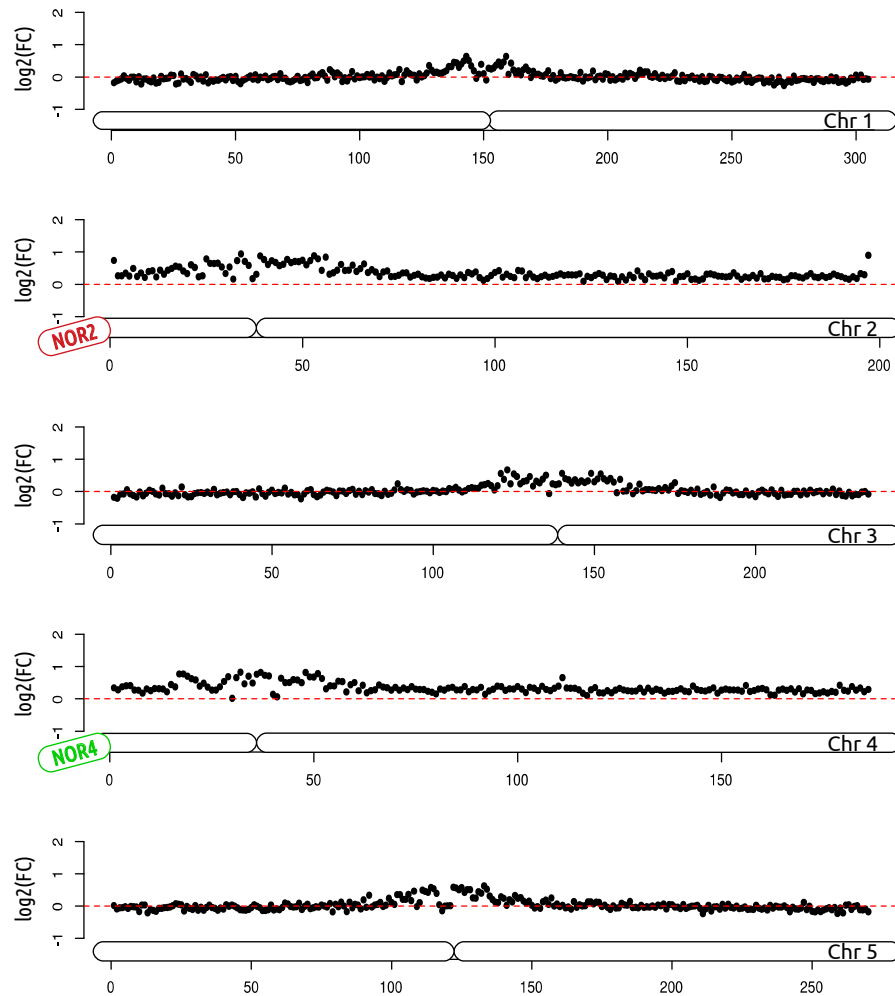


Figure 3.11 – Les NADs d'*Arabidopsis thaliana* suite à la la transition florale. Enrichissement nucléolaire des fenêtres génomiques de 100 kb après la transition florale par rapport à l'enrichissement nucléolaire de plantes de 21 jours. En ordonnée le $\log_2(\text{FC}/\text{FC})$.

les NAD-gènes des plantes contrôle pour le CT.

Toutefois, la majorité de NAD-gènes sont localisés dans les régions associées avec le nucléole lorsqu'elles sont identifiées par des fenêtres de 100 kb. C'est notamment le cas du bras court du chromosome 4 (Figure 3.12B). Ces résultats semblent indiquer que la méthode utilisée ne permet pas une précision au gène. Dans ce cas, une augmentation de la couverture de séquençage pourrait régler en partie ce problème. Cependant, l'identification des NADs est performante pour déterminer les régions génomiques de 100 kb associées au nucléole.

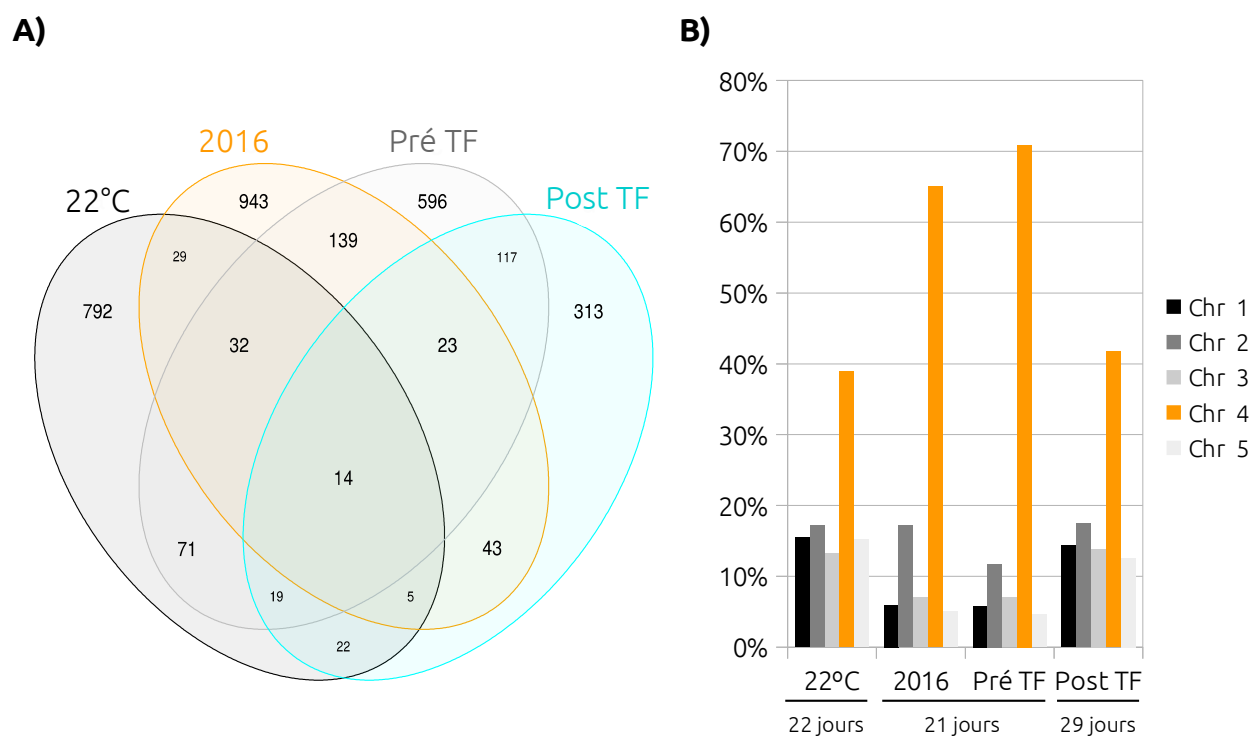


Figure 3.12 – Comparaison des NAD-gènes entre différents échantillons. A) Diagrammes de Venn représentant la superposition entre les NAD-gènes identifiés dans les différents échantillons d'ADN. B) Proportion de NAD-gènes par chromosome (en %) dans les cellules de feuilles de plantes de 22 jours qui ont servi de témoin pour le CT (22°C), de plantes de 21 jours de la première étude de 2016 (PONTVIANNE *et al.* 2016b) et générées pendant ma thèse (Pré TF) et de plantes de 29 jours (Post TF).

3.5 Rôle des gènes d'ARNr 45S dans la composition des NADs

Dans cette section est incluse l'article publié en 2019 dans la revue *Nucleus*, où se trouve la comparaison entre les NADs et les domaines associés à la périphérie du noyau des cellules d'*A. thaliana*. Cet article souligne également l'importance du positionnement

et de l'expression des gènes d'ARNr, pour l'identité des NADs et l'organisation de *hubs* inter-chromosomiques (interactions entre 3 régions chromosomiques ou plus) de gènes peu ou pas transcrits.

EXTRA VIEW

 OPEN ACCESS  Check for updates

Ribosomal RNA genes shape chromatin domains associating with the nucleolus

Ariadna Picart-Piccolo , Nathalie Picault , and Frédéric Pontvianne 

CNRS, Laboratoire Génome et Développement des Plantes (LGDP), Perpignan, France; Université de Perpignan Via Domitia, LGDP, Perpignan, France

ABSTRACT

Genomic interactions can occur in addition to those within chromosome territories and can be organized around nuclear bodies. Several studies revealed how the nucleolus anchors higher order chromatin structures of specific chromosome regions displaying heterochromatic features. In this review, we comment on advances in this emerging field, with a particular focus on a recent study published by Quinodoz *et al.*, that developed a new method to characterize simultaneous genomic interactions in the same cell. Highlighting studies conducted in animal and plant cells, we then discuss the establishment of inactive chromatin at nucleolus organizer region (NOR)-bearing chromosomes.

ARTICLE HISTORY

Received 15 November 2018
Revised 23 January 2019
Accepted 31 January 2019

KEYWORDS

Chromatin; genome organization; nucleolus; rRNA genes

Introduction

In 1928, the cytogeneticist Heitz described two major chromatin states that differ according to their condensation level in the nucleus [1,2]. He named the condensed chromatin regions heterochromatin, while more loosely packed chromatin regions were named euchromatin. The identification of genome-wide chromatin–chromatin interactions maps confirmed this bipartite organization. Today, euchromatin and heterochromatin are generally referred to as A/B compartments [3]. This type of organization is relevant to both plant and animal cells [4]. However, this bipartite organization can be subdivided into several states according to their epigenetic signatures such as, but is not restrictive to, DNA methylation, histone modifications, histone variant distribution or nucleosome positioning [5–7]. Notably, these different states strongly participate in gene transcriptional regulation by modifying the access, or the attractiveness, of any loci to the transcriptional machineries.

At the nucleus scale, cytological analyses already revealed the existence of chromosome territories that can be distributed randomly, as in *Arabidopsis thaliana* or non-randomly, as in human cells, within interphase nuclei [8–11]. In addition, recent technical advances deeply increased our knowledge of how the genome is organized within the nucleus. For

example, the development of the chromosome conformation capture (3C) approach to a genome-wide scale revealed in various organisms and cell-types how inter and intra-chromosomal DNA contacts shape this 3D organization [12]. These studies revealed the existence of topologically associating domains or (TADs) that are self-interacting genomic regions, separated by boundaries, and confined either in A or B compartments, depending on their chromatin signatures. In plants, although the presence of TADs is not obvious in *Arabidopsis thaliana*, they were described in other plant species with larger genomes [4,13].

Recent studies pushed forward our understanding of the 3D genome organization, by taking into consideration the subnuclear compartments such as the nucleolus and the nuclear speckles. In this review, we aim to combine data of recent studies that demonstrate an important role of ribosomal RNA genes distribution and expression in global genome organization.

The nucleolus, a platform for genome organization

Ribosomal RNA (rRNA) genes are organized as tandem repeats in specific genomic regions named nucleolus organizer regions (NORs) by

CONTACT Frédéric Pontvianne  frederic.pontvianne@univ-perp.fr  CNRS, Laboratoire Génome et Développement des Plantes (LGDP); Université de Perpignan Via Domitia, LGDP, UMR 5096, Perpignan 66860, France

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Barbara McClintock [14]. The nucleolus is the site of ribosome biogenesis, starting with rRNA transcription, and is present in most eukaryote cells. Its size is dependent on the ribosome biogenesis activity, but can represent a large volume of the cell, especially in actively dividing cells [15]. In human cells, several nucleoli are usually present, while in *A. thaliana*, usually one round nucleolus is present. The nucleolus is a direct consequence of ribosome biogenesis. However, this large subnuclear compartment is also implicated in other cellular processes, especially in the maturation or sequestration of several ribonucleoproteins complexes such as the spliceosome, the proteasome, the RNA telomerase and others implicated in RNA degradation [15–20].

In 2010, two studies conducted in human cells demonstrated the direct role of the nucleolus in genome organization [21,22]. DNA co-purified with nucleoli from two types of human cells were analyzed, demonstrating that all chromosomes contain, at least, one region associating with the nucleolus. These genomic regions were named NADs for nucleolus-associated chromatin domains. In parallel, a network composed by nuclear lamina also anchors genomic regions with repressive epigenetic features at the nuclear periphery [23]. These regions are named LADs for lamina-associated domains. NADs were subsequently identified in *A. thaliana* cells [24]. In both plant and animal cells, excluding rRNA genes, NADs are composed of around 4% of the genome and display repressive chromatin features [24]. Interestingly, genes are also present in NADs and tend to be lowly expressed. Majority of NADs are presumed to be located at the nucleolus periphery, but cytological analyses demonstrated that the DNA inside the nucleolus is not only composed of rRNA genes [24]. Since RNA polymerase II is absent from inside the nucleolus, it has been proposed that the nucleolus would have a negative impact on gene expression [25,26].

Among the important aspects pointed out by Nemeth *et al.* [22], is the potential redundancy between NADs and LADs. Additional studies confirmed this observation and demonstrated that a portion of LADs, identified as facultative LADs, can reshuffle stochastically either at the nuclear or the nucleolar periphery after mitosis [27,28]. Conversely, LADs remaining at the nuclear

periphery in the daughter cells are now identified as constitutive LADs [29]. In plant cells, no homolog of nuclear lamina proteins have been identified so far, but putative homologs like nuclear matrix constituent protein (NMCP), nuclear envelope-associated proteins (NEAPs) or CROWDED NUCLEI (CRWN) proteins, all localized at the nuclear periphery, are suspected to play a similar role [30–33]. Chromatin domains associating with proteins located at the nuclear periphery have not been identified so far, but Chang Liu's laboratory was able to identify the genomic regions associated with the nuclear periphery using a derived chromatin immunoprecipitation (ChIP) protocol. The nucleoporin NUP1, which is one of the subunits of the nuclear pore complex, was used as a bait [34]. As for NADs, chromosome regions identified in this study (hereafter referred to as NUP1-enriched domains) are enriched in repressive chromatin signatures, transposable elements and contain genes that tend to be lowly expressed [34]. Since at least a portion of NADs overlap with LADs in human cells, we analyzed the potential connection between NADs and NUP1-enriched domains in *A. thaliana* (Figure 1). At the chromosome scale, NADs correspond to genomic regions flanking telomeres, as well as the short arm of chromosome 4, where active rRNA genes are located [35,36]. In contrast, NUP1-enriched domains mainly correspond to peri-centromeric regions, and large portions of chromosome arms (Figure 1(a)) [24,34]. The partial overlap observed on the short arm of chromosome 4 could potentially link the nucleolus and the nuclear periphery. Interestingly, the heterochromatic knob, corresponding to the conspicuous gap located between the 3,5 and 4,5 MB region, only associate with the nucleolus.

Therefore, the overlap between NADs and NUP1-enriched domains appears minor, and genes usually either associate with the nuclear periphery or with the nucleolar periphery (Figure 1(b)).

Influence of RNA genes in nucleolus-associated chromatin domains and inactive hubs

The Hi-C approach allows the detection of direct or indirect chromatin interactions in a cell population [12]. However, Hi-C is not compatible to detect

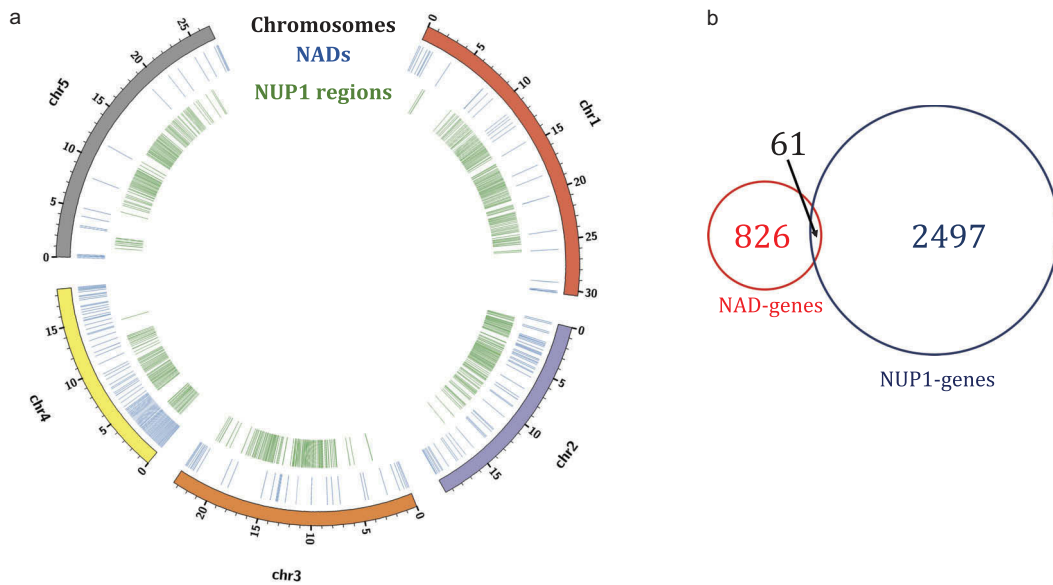


Figure 1. Overlap between NADs and NUP1-regions in *Arabidopsis thaliana*. (a) Genome-wide view of NAD-genes and the regions enriched at the nuclear periphery (NUP1-enriched regions) in the *A. thaliana* genome produced by Circos [49]. Considering the concentric circles from the outside inward, the outermost (and first) circle indicates the five chromosomes of *A. thaliana*. The next two circles are showing the location of NADs-genes (blue) and NUP-enriched regions (green) throughout the genome. (b) Venn diagram of NAD-genes and NUP1-enriched regions corresponding to protein-coding genes in TAIR10. Data from [24,34] were used to generate this figure.

simultaneous chromatin interactions and cell-to-cell variability. Single-cell Hi-C has recently been developed to counteract this problem, but sequencing depth remains an issue, leading to very low genome coverage [37]. In a recent study, Quinodoz and colleagues enable the detection of simultaneous interactions *via* the development of a new technique named SPRITE for split-pool recognition of interactions by tag extension [1]. SPRITE allows the detection of multiple chromatin interactions mediated by an individual complex within a nucleus, revealing the context of the interactions. Briefly, unique complex, from which all RNA, DNA and protein components are covalently crosslinked, are sorted in individual wells. Therefore, all DNA reads from one individual complex are identified as a SPRITE cluster [1]. In other words, a chromatin loop can be associated with its TAD and its compartment. Using the SPRITE technique, Quinodoz *et al.*, linked some A/B compartments with nuclear bodies. They identified active hubs, corresponding to A compartments, in association with nuclear speckles. In parallel, inactive hubs, corresponding to B compartments, were shown to associate with the nucleolus [1]. This corroborates previous

observations made in both animal and plant cells [25,38], regarding the important role of the nucleolus in organizing inactive chromatin regions at the nucleolar periphery.

Importantly, inactive hubs are present on rRNA-genes bearing chromosomes [1]. In other words, inactive hubs correspond to genomic regions flanking rRNA genes. This raises the question of the role of rRNA gene expression and location on chromosomes in defining the identity of inactive hubs at the nucleolar periphery. In *A. thaliana*, NADs are mainly composed of genomic regions corresponding to the NOR-bearing chromosomes 2 and 4, as well as genomic regions from the subtelomeric regions [24]. In wild-type leaves, only NOR4-derived rRNA genes associate with the nucleolus and are actively transcribed, while NOR2-derived rRNA genes are silent and excluded from the nucleolus [35,36,39–42]. Interestingly, NOR2-derived rRNA genes expression provokes a complete reorganization of the short arm of chromosome 2 that associates with the nucleolus [24]. Collectively, these data suggest that rRNA genes expression or silencing impact the global heterochromatin nuclear distribution. To test the

potential presence of inactive hubs in *A. thaliana*, we analyzed the global gene expression of nucleolus-associated genomic regions. As in mammalian cells, the short arms of NOR-bearing chromosomes tend to contain low-expressed genes compared to the entire genome (Figure 2). However, we do not see such correlation for genes of the subtelomeric regions that also associate with the nucleolus in a wild-type (Col-0) *A. thaliana* line [43].

Conclusions and perspectives

Overall, recent studies revealed the importance of nuclear bodies in the distribution of the epigenome. In these cases, nuclear body specific proteins play crucial roles in these regulations. For example, the nucleolar protein Nucleolin was demonstrated to regulate (i) telomere nucleolar clustering in *A. thaliana* [24], (ii) the two-cell embryo master regulator gene *DUX* in human embryonic stem cells [44] or (iii) centromere nucleolar clustering in drosophila [45]. A missing aspect is an overlap between inactive hubs identified using the SPRITE method and NADs identified in human cells [1,19,20]. We indeed expect NADs and inactive hubs to overlap. To further demonstrate how rRNA genes distribution in the genome defines NADs or inactive hubs, it will be important to identify these regions in cells with the variable NOR chromosomal location, or in mutant showing reduction or increase numbers of rRNA genes. Previous observations in yeast and in drosophila demonstrated that reducing the number of rRNA genes alter heterochromatin organization and genome stability [46–48]. In the light of these new

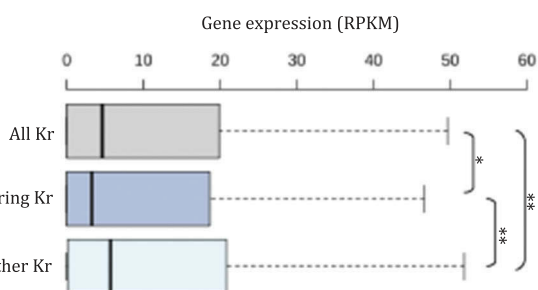


Figure 2. Gene expression of nucleolus-associated genomic regions in *A. thaliana*. Box plot representing the global expression of genes located in the short arm of NOR-bearing chromosomes 2 and 4 (NOR-bearing Kr) compare to other chromosomes 1, 3 and 5 (other Kr 1, 3 & 5) or all five chromosomes (all Kr) in *A. thaliana*. A *p* value < 0.05 *; *p* value < 0.01: ** (Wilcoxon rank sum test).

data, we expect to observe changes in global genome organization in cells with a reduced number of rRNA genes.

Acknowledgments

The authors thank Gwénaëlle Détourné, Guillaume Moissiard and anonymous reviewers for critical reading of the manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work and APC PhD fellowship are supported by the ANR NucleoReg [ANR-15-CE12-0013-01] and the COST ACTION CA16212 INDEPTH.

ORCID

Ariadna Picart-Piccolo  <http://orcid.org/0000-0001-6380-7969>

Nathalie Picault  <http://orcid.org/0000-0003-2770-9688>

Frédéric Pontvianne  <http://orcid.org/0000-0002-2913-4104>

References

- [1] Quinodoz SA, Ollikainen N, Tabak B, et al. Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus. *Cell*. 2018;174:744–757.e24.
- [2] Passarge E. Emil Heitz and the concept of heterochromatin: longitudinal chromosome differentiation was recognized fifty years ago. *Am J Hum Genet*. 1979;31:106–115.
- [3] Lieberman-Aiden E, van Berkum NL, Williams L, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 2009;326:289–293.
- [4] Dogan ES, Liu C. Three-dimensional chromatin packing and positioning of plant genomes. *Nat Plants*. 2018;4:521–529.
- [5] Fillion GJ, van Bommel JG, Braunschweig U, et al. Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell*. 2010;143:212–224.
- [6] Roudier F, Ahmed I, Berard C, et al. Integrative epigenomic mapping defines four main chromatin states in *Arabidopsis*. *Embo J*. 2011;30:1928–1938.
- [7] Sequeira-Mendes J, Araguez I, Peiro R, et al. The functional topography of the *Arabidopsis* genome is organized in a reduced number of linear motifs of chromatin states. *Plant Cell*. 2014;26:2351–2366.

- [8] Heard E, Bickmore W. The ins and outs of gene regulation and chromosome territory organisation. *Curr Opin Cell Biol.* 2007;19:311–316.
- [9] Pecinka A, Schubert V, Meister A, et al. Chromosome territory arrangement and homologous pairing in nuclei of *Arabidopsis thaliana* are predominantly random except for NOR-bearing chromosomes. *Chromosoma.* 2004;113:258–269.
- [10] Schubert V, Berr A, Meister A. Interphase chromatin organisation in *Arabidopsis* nuclei: constraints versus randomness. *Chromosoma.* 2012;121:369–387.
- [11] Franz P, De Jong JH, Lysak M, et al. Interphase chromosomes in *Arabidopsis* are organized as well defined chromocenters from which euchromatin loops emanate. *Proc Natl Acad Sci U S A.* 2002;99:14584–14589.
- [12] Dekker J, Marti-Renom MA, Mirny LA. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat Rev Genet.* 2013;14:390–403.
- [13] Dong P, Tu X, Chu P-Y, et al. 3D chromatin architecture of large plant genomes determined by local A/B compartments. *Mol Plant.* 2017;10:1497–1509.
- [14] McClintock B. The relationship of a particular chromosomal element to the development of the nucleoli in *Zea mays*. *Zeit Zellforsch Mik Anat.* 1934;21:294–328.
- [15] Pederson T. The nucleolus. *Cold Spring Harb Perspect Biol.* 2011;3:a000638.
- [16] Boulon S, Westman BJ, Hutten S, et al. The nucleolus under stress. *Mol Cell.* 2010;40:216–227.
- [17] Lange H, Zuber H, Sement FM, et al. The RNA helicases AtMTR4 and HEN2 target specific subsets of nuclear transcripts for degradation by the nuclear exosome in *Arabidopsis thaliana*. *PLoS Genet.* 2014;10:e1004564.
- [18] Pendle AF, Clark GP, Boon R, et al. Proteomic analysis of the *Arabidopsis* nucleolus suggests novel nucleolar functions. *Mol Biol Cell.* 2005;16:260–269.
- [19] Montacie C, Durut N, Opsomer A, et al. Nucleolar proteome analysis and proteasomal activity assays reveal a link between nucleolus and 26S proteasome in *A. thaliana*. *Front Plant Sci.* 2017;8:1815.
- [20] Tomlinson RL, Ziegler TD, Supakorndej T, et al. Cell cycle-regulated trafficking of human telomerase to telomeres. *Mol Biol Cell.* 2006;17:955–965.
- [21] van Koningsbruggen S, Gierlinski M, Schofield P, et al. High-resolution whole-genome sequencing reveals that specific chromatin domains from most human chromosomes associate with nucleoli. *Mol Biol Cell.* 2010;21:3735–3748.
- [22] Nemeth A, Conesa A, Santoyo-Lopez J, et al. Initial genomics of the human nucleolus. *PLoS Genet.* 2010;6:e1000889.
- [23] Pombo A, Dillon N. Three-dimensional genome architecture: players and mechanisms. *Nat Rev Mol Cell Biol.* 2015;16:245–257.
- [24] Pontvianne F, Carpentier M-C, Durut N, et al. Identification of nucleolus-associated chromatin domains reveals a role for the nucleolus in 3D organization of the *A. thaliana* genome. *Cell Rep.* 2016;16:1574–1587.
- [25] Picart C, Pontvianne F. Plant nucleolar DNA: green light shed on the role of Nucleolin in genome organization. *Nucl Austin Tex.* 2017;8:11–16.
- [26] Schubert V, Weisshart K. Abundance and distribution of RNA polymerase II in *Arabidopsis* interphase nuclei. *J Exp Bot.* 2015;66:1687–1698.
- [27] Kind J, Pagie L, de Vries SS, et al. Genome-wide maps of nuclear lamina interactions in single human cells. *Cell.* 2015;163:134–147.
- [28] Kind J, Pagie L, Ortabozkoyun H, et al. Single-cell dynamics of genome-nuclear lamina interactions. *Cell.* 2013;153:178–192.
- [29] van Steensel B, Belmont AS. Lamina-associated domains: links with chromosome architecture, heterochromatin, and gene repression. *Cell.* 2017;169:780–791.
- [30] Meier I, Richards EJ, Evans DE. Cell biology of the plant nucleus. *Annu Rev Plant Biol.* 2017;68:139–172.
- [31] Pawar V, Poulet A, Detourne G, et al. A novel family of plant nuclear envelope-associated proteins. *J Exp Bot.* 2016;67:5699–5710.
- [32] Ciska M, Masuda K, Moreno Diaz de la Espina S. Lamin-like analogues in plants: the characterization of NMCP1 in *Allium cepa*. *J Exp Bot.* 2013;64:1553–1564.
- [33] Dittmer TA, Stacey NJ, Sugimoto-Shirasu K, et al. LITTLE NUCLEI genes affecting nuclear morphology in *Arabidopsis thaliana*. *Plant Cell.* 2007;19:2793–2803.
- [34] Bi X, Cheng Y-J, Hu B, et al. Nonrandom domain organization of the *Arabidopsis* genome at the nuclear periphery. *Genome Res.* 2017;27:1162–1173.
- [35] Pontvianne F, Blevins T, Chandrasekhara C, et al. Subnuclear partitioning of rRNA genes between the nucleolus and nucleoplasm reflects alternative epiallelic states. *Genes Dev.* 2013;27:1545–1550.
- [36] Chandrasekhara C, Mohannath G, Blevins T, et al. Chromosome-specific NOR inactivation explains selective rRNA gene silencing and dosage control in *Arabidopsis*. *Genes Dev.* 2016;30:177–190.
- [37] Nagano T, Lubling Y, Yaffe E, et al. Single-cell Hi-C for genome-wide detection of chromatin interactions that occur simultaneously in a single cell. *Nat Protoc.* 2015;10:1986–2003.
- [38] Nemeth A, Langst G. Genome organization in and around the nucleolus. *Trends Genet.* 2011;27:149–156.
- [39] Pontvianne F, Blevins T, Chandrasekhara C, et al. Histone methyltransferases regulating rRNA gene dose and dosage control in *Arabidopsis*. *Genes Dev.* 2012;26:945–957.
- [40] Pontvianne F, Abou-Ellail M, Douet J, et al. Nucleolin is required for DNA methylation state and the expression of rRNA gene variants in *Arabidopsis thaliana*. *PLoS Genet.* 2010;6:e1001225.
- [41] Mohannath G, Pontvianne F, Pikaard CS. Selective nucleolus organizer inactivation in *Arabidopsis* is a chromosome position-effect phenomenon. *Proc Natl Acad Sci U S A.* 2016;113:13426–13431.

- [42] Earley KW, Pontvianne F, Wierzbicki AT, et al. Mechanisms of HDA6-mediated rRNA gene silencing: suppression of intergenic Pol II transcription and differential effects on maintenance versus siRNA-directed cytosine methylation. *Genes Dev.* **2010**;24:1119–1132.
- [43] Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature.* **2000**;408:796–815.
- [44] Percharde M, Lin C-J, Yin Y, et al. A LINE1-Nucleolin partnership regulates early development and ESC identity. *Cell.* **2018**;174:391–405.e19.
- [45] Padeken J, Mendiburo MJ, Chlamydas S, et al. The nucleoplasmin homolog NLP mediates centromere clustering and anchoring to the nucleolus. *Mol Cell.* **2013**;50:236–249.
- [46] Ide S, Miyazaki T, Maki H, et al. Abundance of ribosomal RNA gene copies maintains genome integrity. *Science.* **2010**;327:693–696.
- [47] Paredes S, Maggert KA. Ribosomal DNA contributes to global chromatin regulation. *Proc Natl Acad Sci U S A.* **2009**;106:17829–17834.
- [48] Paredes S, Branco AT, Hartl DL, et al. Ribosomal DNA deletions modulate genome-wide gene expression: “rDNA-sensitive” genes and natural variation. *PLoS Genet.* **2011**;7:e1001376.
- [49] Krzywinski M, Schein J, Birol I, et al. Circos: an information aesthetic for comparative genomics. *Genome Res.* **2009**;19:1639–1645.

Lors de l'analyse comparative entre les NADs et les domaines associés à la périphérie du noyau des cellules d'*A. thaliana* de cette étude, ce sont les domaines chromatiniens purifiés par le biais de la protéine NUP1 qui ont été utilisés (Bi *et al.* 2017). Plus tard, l'équipe de Chang Liu a identifié les domaines associés à la périphérie nucléaire, en utilisant les protéines CRWN (CROWDED NUCLEI) (Hu *et al.* 2019). Les domaines identifiés ont été appelés PLADs (Plant-like Lamina associated chromatin Domains).

J'ai utilisé ces nouvelles données pour refaire l'analyse comparative entre les NADs et les PLADs (Figure 3.13). Les résultats obtenus sont très similaires à ceux obtenus en comparant les NADs avec les domaines NUP1. En effet, il y a très peu de superposition entre les NADs et les PLADs. Les domaines qui à la fois peuvent être associés au nucléole ou à la périphérie nucléaire sont des régions qui se trouvent au niveau du bras court du chromosome 4.

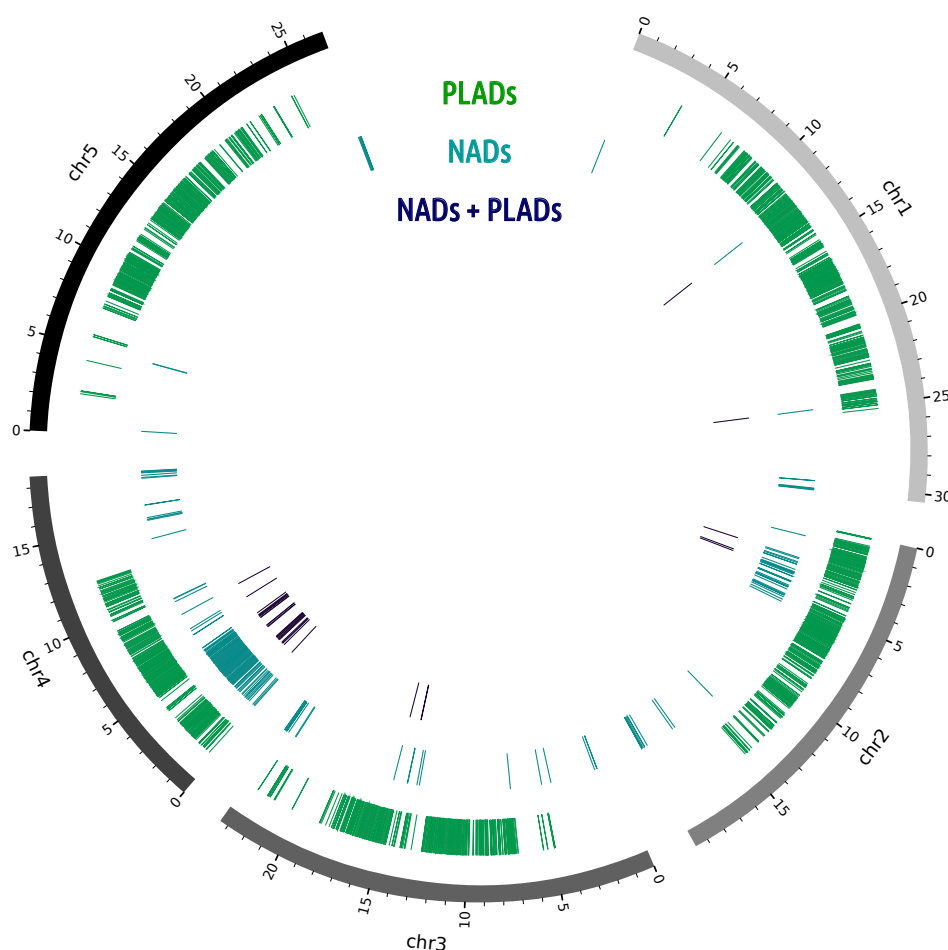


Figure 3.13 – Comparaison entre les NADs et les PLADs. Circos (KRZYWINSKI *et al.* 2009) représentant la position des PLADs et des NADs le long des 5 chromosomes d'*A. thaliana*, et les régions qui ont été à la fois identifiées comme des NADs et des PLADs.

3.6 Discussion

Au cours de ma thèse, j'ai pu analyser les NADs en condition de stress thermique et suite à la transition florale. Lors d'un CT, la décondensation des chromocentres est un processus transitoire impliqué dans la réponse au stress thermique. Lors de la transition florale, la restructuration de la chromatine est due à un changement développemental qui implique la modification de l'expression de gènes de façon plus prolongée dans le temps. Cependant, dans les deux cas, on ne constate pas de différences au niveau des domaines associés avec le nucléole.

Au contraire, ces résultats montrent une grande stabilité des NADs à l'échelle de 100 kb, qui correspondent à des domaines hétérochromatiques très spécifiques. En effet, l'identité des NADs est surtout déterminée par l'activité des NORs et le regroupement des télomères autour du nucléole. De même, chez les mammifères, les NADs sont très enrichis en séquences proches des NORs. Les *hubs* inter-chromosomiques qui sont localisés autour du nucléole se forment entre des chromosomes possédant des NORs (GUETG et SANTORO, 2012a; KONINGSBRUGGEN *et al.* 2010; NÉMETH *et al.* 2010; QUINODOZ *et al.* 2018; VERTII *et al.* 2019).

Mis à part le chromocentre associé au NOR actif, les régions péricentromériques ne font pas partie des NADs. Contrairement aux animaux, qui possèdent aussi des régions péri-centromériques de plusieurs chromosomes, parmi lesquels des chromosomes ne comportant pas de NOR, associées avec le nucléole (DILLINGER *et al.* 2016; GUETG et SANTORO, 2012a; KONINGSBRUGGEN *et al.* 2010; NÉMETH *et al.* 2010).

Ces résultats étaient attendus en raison de la conformation en rosette du génome d'*A. thaliana* (FRANSZ *et al.* 2002), avec les centromères localisés à la périphérie nucléaire et les télomères à la périphérie nucléolaire. Cette conformation explique également pourquoi on constate peu de chevauchement entre les NADs et les domaines associés à la périphérie nucléaire contrairement à la levure ou l'humain. Cependant, un nombre non négligeable de régions du bras court du chromosome 4 ont également été identifiées comme des PLADs. Une explication possible est le rapprochement observé dans de nombreux noyaux du nucléole avec la périphérie nucléaire.

De plus, mis à part les régions péricentromériques, la fraction inactive des NORs et le knob hk4S (chez Col-O), le génome d'*A. thaliana* ne comporte pas de régions intergéniques hétérochromatiques. Par conséquent, il n'y a pas d'autres régions propices à se retrouver associées avec le nucléole. Cette stabilité est également observée entre des différentes lignées cellulaires d'humain et souris, de même qu'au cours de la sénescence cellulaire chez l'humain (DILLINGER *et al.* 2016; VERTII *et al.* 2019).

Il est à noter que les NADs sont aussi composés de régions hétérochromatiques qui

ne sont pas localisées au niveau du bras court du chromosome 4 ni des régions sub-télomériques. En effet, des domaines hétérochromatiques de tous les chromosomes, chez *A. thaliana* comme chez l'humain et la souris, font partie des NADs. De plus, la proximité physique avec un NOR est importante mais ne suffit pas pour se retrouver associé avec le nucléole.

Par exemple, le NOR et le bras court du chromosome 2 sont exclus du nucléole. Chez l'humain, des régions actives entre des NADs proches d'un NOR ne se retrouvent pas associées avec le nucléole (DILLINGER *et al.* 2016). Ceci indique que si une région a une activité transcriptionnelle elle est exclue de la périphérie nucléolaire. Un autre exemple est le chromosome X inactif qui est associé avec le nucléole, mais des régions de ce dernier qui « échappent » à la répression mise en place au cours du développement ne sont pas associées (DILLINGER *et al.* 2016).

En conclusion, nos résultats ne nous permettent pas de valider l'existence d'une voie active de régulation transcriptionnelle par séquestration (ou libération) nucléolaire de certains gènes transcrits par l'ARN pol II. De mon point de vue, l'ensemble des données suggèrent plutôt que l'association de domaines hétérochromatiques avec le nucléole est un phénomène passif. Une conséquence du fait d'être à proximité d'un NOR ou transcriptionnellement inactif.

Chapitre 4

Importance de l'ADN ribosomique dans l'intégrité du génome d'*Arabidopsis thaliana*

Le deuxième axe de ma thèse avait comme but d'étudier l'importance des copies inactives des gènes d'ARNr chez *A. thaliana*. Les résultats obtenus font l'objet de l'article en préparation.

Low ribosomal RNA genes copy number provoke genomic instability and chromosomal segment duplication events that modify global gene expression and plant-pathogen response

Ariadna Picart-Piccolo^{1,2}, Stefan Grob³, Nathalie Picault^{1,2}, Michal Franek⁴, Thierry Alter⁵, Thomas Maier⁶, Christel LLauro^{1,2}, Edouard Jobet^{1,2}, Panpan Zhang^{1,2,7}, Thomas Baum⁶, Lionel Navarro⁵, Martina Dovrackova⁴, Marie Mirouze^{1,2,7}, Frederic Pontvianne^{1,2#}

¹CNRS, LGDP UMR5096, Université de Perpignan, Perpignan, France

²UPVD, LGDP UMR5096, Université de Perpignan, Perpignan, France

³Institute of Plant and Microbial Biology, University of Zurich, Zurich, Switzerland

⁴CEITEC, Masaryk University, Brno, Czech Republic

⁵ENS, IBENS, CNRS/INSERM, PSL Research University, Paris, France

⁶Department of Plant Pathology and Microbiology, Iowa State University, Ames, IA, USA

⁷IRD, UMR232 DIADE, Montpellier, France

#*corresponding author: fpontvia@univ-perp.fr*

ABSTRACT

Among the hundreds to thousands of ribosomal RNA (rRNA) genes copies in the genome present in the nucleolus organizer regions (NORs), only a portion are usually actively expressed and participate in the ribosome biogenesis process in the nucleolus. The role of these extra-copies remains elusive, but previous studies suggest their importance in genome stability and global gene expression. The nucleolus is also an important platform for the nuclear organization. Therefore, we tested the impact of having a low amount of rRNA genes

copies in the *Arabidopsis thaliana* 3D genome organization and stability, using a line only displaying 20% of them (20rDNA line). Compare to the wild-type Col-0, the 20rDNA line accumulate several signs of genomic instability like variations in the 3D genome organization, spontaneous double stranded breaks accumulation, transcriptomic changes and significantly higher DNA methylation level. Strikingly, using genomic and microscopic approaches, we identified up to 7 large tandem duplications in direct orientation (TDDO) from 60 kb to 1,44 Mb. In consequence, more than 600 genes are now duplicated, often provoking an increase of their expression level. Among them, we have found genes implicated in the plant-pathogen responses, whose up-regulation seem to have modified the ability of the 20rDNA line to resist to both bacterial and nematode infections. Finally, we show that the TDDO provoke gene fusions and/or truncations and discuss their potential implications in plant genome evolution.

INTRODUCTION

The nucleolus, site of ribosome biogenesis, plays an important role in the spatial organisation of the chromosomes in the nucleus (1). Chromatin domains repressive to transcription (inactive) are particularly enriched at the nucleolar periphery (2, 3), as well as at the nuclear periphery retained by a network of proteins named Lamina in mammal cells (4). Due to their nuclear positioning, these chromatin domains are referred as Lamina Associated chromatin Domains (LADs) and as Nucleolus-Associated chromatin Domains (NADs). Plant-LADs (pLADs) were recently described in *A. thaliana*, associating with Crowded Nuclei (CRWN) proteins displaying similar characteristics than Lamina (5, 6). In parallel, using Fluorescent assisted cell sorting, we recently isolated pure nucleoli extracts, allowing the identification of NADs (7). Both LADs-like and NADs in *A. thaliana* are enriched in repressive chromatin

signatures. As a consequence, genes present in these regions tend to be not or low expressed (6, 8).

NADs identification revealed a direct link between Nucleolus Organizer Regions (NORs) and NADs other than ribosomal RNA (rRNA) genes (7). In *Arabidopsis thaliana* Col-0, 375 of tandemly 45S rRNA gene copies are at the top of both chromosomes 2 (NOR2) and 4 (NOR4) (9). Although rRNA genes are highly similar, small variabilities allow to separate rRNA genes that belong to the NOR2 or the NOR4 (10, 11). Previous studies clearly demonstrated that in wild-type Col-0 leaf cells, NOR4 derived rRNA genes are expressed and associate with the nucleolus. Conversely, NOR2 is excluded from the nucleolus and NOR2 derived rRNA genes are silent (11, 12). As a consequence, the entire short arm of the chromosome 4 (kr4s), as well as its centromeric and pericentromeric regions, that juxtapose the active NOR4, associate with the nucleolus (7). This association was not detected for the short arm of chromosome 2, except in the *nucleolin 1* (*nuc1*) mutant, where normally silent rRNA gene copies are expressed resulting in the association of both NOR2 and 4 with the nucleolus (7, 8). These data demonstrate that rRNA genes organization and expression have direct impact on global genome organization within the nucleus (13).

As in other Eukaryotes, rRNA copy numbers is highly variable among *Arabidopsis thaliana* populations (14–17). In natural inbred lines found in Sweden, rRNA copy number heterogeneity is responsible of the 10% genome size variation (18). Worldwide, *A. thaliana* ecotypes can be found in nature with a copy number of the rRNA gene ranging from 500 to 2500 in haploid cells, knowing that the number of copies of Col-0 is about 750 (18). Therefore, 500 copies could be considered as the lowest rRNA gene copy number found *in natura* (17). In plant cells, the impact of this variability is not known, but in budding yeast

and in *Drosophila*, previous studies suggest that a minimum amount of inactive rRNA gene copies is necessary for global genome stability (19, 20). 100 to 200 rRNA gene units are usually found in budding yeast, but genome engineering allowed the creation of viable yeast lines with only 40 rRNA gene units. Nevertheless, genome stability is highly compromised in these cells (19, 21). In the same time, rRNA gene copy number shifts can affect genome-wide euchromatic patterns and alter gene expression in flies (20, 22).

FASCIATA (FAS) 1 and 2 compose the CAF complex required for proper deposition of H3 and H4 (23). Their knock-out provoke a drastic in the number of copies of rRNA genes, leading to a global reorganization of the epigenetic states and the sub-nuclear positioning of NORs (12, 24). Interestingly, crossing *fas1-4* and *fas2-4* mutants lead to the creation of two independent *A. thaliana* lines displaying wild-type *FAS* genes with only 20% of the amount of rRNA gene copies (20rDNA) from Col-0 (25). These two 20rDNA lines (6 and 9) have a wild-type phenotype and retain low amount of rRNA genes for 5 generations (25). In this study, we took advantage of the two 20rDNA lines to test the impact of the NORs size in NADs composition, 3D genome organization, genome stability and global gene expression. Our data reveal unexpected consequences of genomic instability in the 20rDNA line.

RESULTS

Low rDNA copies impact NADs composition and global 3D genome organization

As rRNA gene nuclear distribution seems to have an important impact in NADs identity both in plant and animal cells (7, 13, 26), we analyzed NADs composition and the global 3D

organization in an *A. thaliana* line containing a low amount of rRNA genes named 20rDNA line 6. The 20rDNA line6 has been obtained after the crossing of the *fas1-4* and *fas2-4* mutants line as described in (Figure 1A). We introgressed a transgene expressing ectopically the FIBRILLARIN 2 nucleolar protein fused to the Yellow Fluorescent Protein (FIB2:YFP) in the 5th generation of the 20rDNA line 6 (20rDNA L6F5). Thanks to the use of the FIB2:YFP nucleolar marker, we isolated nuclei and nucleoli from the transformant and identify NADs from the recovered nuclear and nucleolar DNA as described previously (27, 28). Compare to wild-type, where NADs are essentially distributed in the short arm of the chromosome 4 (kr4s), NADs in the 20rDNA L6F6 line contain both chromosome 2 and 4 short arms (Figure 1B). Among the 434 genes that gain nucleolar association, 144 belong to the chromosome 2 (33%). In contrast, chromosome 4 NADs remain stable since only 19 genes gain nucleolus association. These results are consistent with the rRNA genes transcriptional state since all left-over NOR2 and NOR4 derived rRNA are actively transcribed and associate with the nucleolus (25). As in wild-type, sub-telomeric regions remain associated with the nucleolus in 20rDNA L6F6 line (suppl. Figure 1A-B). In total, NADs identification in 20rDNA L6F6 line revealed that 5,6 Mb of chromatin domains mainly enriched in silent epigenetic signatures changed their subnuclear distribution which suggest a substantial reorganisation of the nuclear genome. To get a more global view the epigenome organization in 3D, we analysed all chromatin-chromatin interactions using chromosome-conformation capture sequencing (Hi-C) (Figure 1C).

We generated triplicate Hi-C samples from both wild-type and 20rDNA L6F6 14-day-old seedling populations, allowing us to perform robust statistical analyses (Suppl. Figure 2). We developed a novel method to assess difference between two given sets of Hi-C samples statistically. To date, comparisons between Hi-C data sets are often unsatisfactory, as they

rely merely on visual inspection of relative difference matrices. Making use of our triplicate Hi-C data sets, we performed student T-test on each contact frequency (pixel of the Hi-C matrix) and determined whether it significantly changed between wild-type and the 20rDNA L6F6 line (Figure 1C and Figure 1E-G). Surprisingly, we found several regions displaying a significant ($p < 0.01$) increase of contact frequencies at several chromosomal locations (Figure 1C, Sign_increase signals). The largest represents 1,44Mb on the kr4s and contains the knob, a large heterochromatic regions outside the peri-centromeres, and an euchromatic regions distal to the knob. Additional regions are present on the other chromosomes: 200kb around the 20Mb position of chromosome 1, 370kb around the 15,7Mb position of chromosome 2 and 60kb around the 12,5Mb of chromosome 5. Contact frequencies assayed by Hi-C can be used to detect chromosomal rearrangements (29). We suspected that these increased interaction frequencies are not primarily based on significant changes in 3D folding principles of the regions but that large duplications in the 20rDNA L6F6 line could have given rise to the increase in contact frequencies. A duplication would lead to a 2-fold increase in coverage of the affected region and, thus, doubling of interaction frequencies, which involve the duplicated region. Indeed, by analysing the genome-wide coverage using unpaired raw Hi-C sequencing reads, we found a significant increase in coverage in all affected regions (Figure 1D). We subsequently normalized our Hi-C matrices, which led to the removal of nearly all significantly different contact frequencies stemming from the duplicated regions. However, difference analysis in coverage normalized Hi-C data showed that short-range contact frequencies within the duplicated regions are significantly depleted. Whether this depletion of contact frequencies is biologically significant or represents an artefact of the normalization procedure is extremely difficult to determine.

To further examine potential differences in 3D folding principles between wild type Col-0 and 20rDNA L6F6, we performed PCA analysis to retrieve the Eigenvector, which is characteristic of 3D-folding pattern of a Hi-C data set (30, 31). Sign-changes in the Eigenvector delineate basic 3D folding domains, known as Closed Structural Domains (CSDs) and Loose Structural Domains (LSDs), which are analogous to animal A and B compartments (31). We could not identify significant changes in the Eigenvectors between the two lines that would go beyond the variation observed within the lines. We then calculated the contact density for each genomic bin, which is determined by the number of contact frequencies observed within a given region surrounding the genomic bin of interest. In line with our difference analysis, we could observe a decrease in contact densities in duplicated regions, however, no significant difference elsewhere in the genome.

Subsequently, we were interested whether the location of the duplications might have been biased based on a priori folding principles in the wild type. Hence, we compared duplication breakage points with the Eigenvector obtained by the PCA analysis of the wild type Hi-C data (Suppl. Figure 3). Intriguingly, we observed that in a majority of duplications at least one of the breakage points coincides with sign-changes (CSDs to LSDs) or directional changes (valleys and peaks within a structural domain) in the Eigenvector. Hence, the changes 3D-conformation may have facilitated the occurrence of duplications. For the knob *hk4s*, this was most prominent, where the more central breakpoint exactly colocalizes with the change between the CSD and the LSD, which defines the knob border in 3D and furthermore, exactly colocalizes with the ancient inversion breakpoint that gave rise to the knob (Figure 2H) (32). This suggests the existence of continuously fragile chromosomal regions and that borders between structural domains may be diagnostic for these regions.

Characterization and impact of duplication events on gene expression

To confirm and map the duplicated regions in the 20rDNA L6F6 line, we performed long-read resequencing using the Nanopore technology. We obtained 6,4 Gb of total sequences with a mid-size of 6 kb. We then analyzed the sequencing coverage using the TAIR10 *A. thaliana* Col-0 reference genome to visualize the highly covered regions (Figure 2A). We confirmed the existence of the 4 major large duplications found by HiC. We also identified 3 additional minor duplications: a 80kb long duplication around the 6,4 Mb position of chromosome 2, a 60 kb long duplication around the 10,7Mb position on chromosome 4 and an additional duplication of 100kb in the already duplicated region on the kr4s (Figure 2A). These 7 duplication events correspond to tandem duplications in direct orientation (TDDO) and were named from TDDO1 to TDDO7. Combined, they correspond to a gain of 2,31 Mb per haploid genome. As a consequence, genes and Transposable Elements (TEs) copy number variations (CNV) occurs in the 20rDNA L6F6 line. These CNVs concern 626 genes and 851 TEs. In parallel, we also confirmed a significant reduction of reads mapped against rRNA genes.

Duplication events occurrence is a sign of genomic instability. These chromosomal rearrangements could be the consequence of double stranded breaks (DSBs). Although we do not know when these rearrangements happened, we compared the occurrence of spontaneous double stranded breaks between 20rDNA L6F6 and WT Col-0 cells. Spontaneous phosphorylated γ -H2Ax foci, corresponding to DSB sites, were detected at a higher rate in 20rDNA L6F6 nuclei compare to WT nuclei (Figure 2B and Suppl. Figure 4A). Accumulation of DSB foci can potentially be associated to a DNA repair defect. This

hypothesis is supported by an increased susceptibility of the 20rDNA L6F6 line to treatment with the genotoxin bleomycin (Suppl. Figure 4B).

Changes in the organization of the 3D genome as well as variations in the number of gene copies can have an impact on gene expression and local epigenetic modifications. We therefore analysed the global gene expression pattern by polyA⁺ RNA-seq and the methylome by whole-genome bisulfite sequencing (WGBS) in WT Col-0 *versus* the 20rDNA L6F6 line. Four replicates per samples of RNA-seq analyses identified differentially accumulating transcripts: 321 UP genes and 14 UP TEs, as well as 37 DOWN genes in the 20rDNA L6F6 line (with an adjusted p-value <0.01 and log₂(FoldChange) > 1,5) (Figure 2A and Suppl. Figure 5A). For some of the genes and TEs, quantitative RT-PCR (qRT-PCR) was performed to confirm the differential accumulation of their transcripts (Suppl. Figure 5B). Differentially accumulating transcripts belong from genes spread on all five chromosomes and do not necessary concentrate on the lost or gain NADs in the 20rDNA L6F6 line (Figure 2A; Suppl. Figure 6).

Differentially methylated regions (DMRs) were identified from triplicates WGBS in WT Col-0 and in the 20rDNA L6F6 line. We did not reveal an important overlap between upregulated genes and hypo-DMRs (Suppl. Figure 6). Curiously, we observed a global increase of methylation in the 20rDNA L6F6 line. We analysed this increase for all genes, duplicated genes and up-regulated genes. We always found a significant increase in CG, CHG and CHH methylation in the 5' and 3' regulatory sequences flanking the genes (Figure 2E-G; Suppl. Figure 7). Only body methylation remains unchanged in the upregulated genes. Global methylation is affected in the body of TEs in the CHG and CHH contexts, but not in the CG context (Figure 2H-I; Suppl. Figure 8). Although significant changes can be observed at the

DNA methylation level, they seem to concern all genes and TEs, and not a specific group of genes. In other words, both duplicated or upregulated genes are concerned by the global DNA methylation increase. Only gene body methylation of the upregulated genes are not concerned by this DNA methylation increase, which is consistent with their higher expression levels (Suppl. Figure 6).

Yet, we found that duplicated genes and TEs are significantly more expressed (Figure 2C-D). If we consider the 321 upregulated genes with a foldchange enrichment of 1.5, we found that 22% of them (71) belongs to duplicated genes, while the TDDOs only represent 2% of the genome. Conversely, no overlap can be found between duplicated and down-regulated genes. In the same time, 57% (8) of the Upregulated TEs are also duplicated (Figure 2C). Box-plot analyses of all genes *versus* the duplicated genes indeed reveal their important ability to overaccumulate more transcripts in the 20rDNA L6F6 line (Figure 2J). Overall, our data suggest that duplicated genes are preferentially more expressed than in WT.

Duplication events are linked to a higher pathogen resistance

Genes categories implicated in biotic and stress responses are particularly enriched in the upregulated genes in the 20rDNA L6F6 line (Figure 3A). We performed qRT-PCR analyses and confirmed the overaccumulation of key genes required in the plant-pathogen response (Figure 3B). Among them are *Pathogen Response (PR)* 1 and 5 genes, whose higher expression level is usually correlated with better resistance against bacteria and nematodes. (33). Interestingly, some of these genes actually compose the duplicated regions of chromosome 2 and 4 (TDDO3 and TDDO4). It is therefore possible that their higher expression rate is a consequence of the duplication events (Suppl. Figure 9). It is the case of

ASYMMETRIC LEAVES 1 (AS1) that has an evolutionarily conserved role in plant-pathogen interactions (34). AS1 indeed acts as positive regulator of extracellular defenses against bacterial pathogens in a salicylic acid-independent manner (35). In addition, genes encoding four Cysteine-rich receptor-like kinases (CRKs), located in the duplicated part of the *kr4s*, also over accumulates transcripts in the 20rDNA L6F6 line (Suppl. Figure 9). Among them is found *CRK36*, whose overexpression was shown to be sufficient to enhanced pattern-triggered immunity response and bacterial pathogens resistance (36).

We therefore challenged the 20rDNA L6F6 line and wild-type plants with phytopathogenic agents to test their resistance capabilities. *A. thaliana* is susceptible to various pathogens, from prokaryotes to multicellular organisms. We first tested the ability of the sugar beet cyst nematode *Heterodera schachtii*, which also infects the model plant *Arabidopsis thaliana* to affect the WT Col-0 and the 20rDNA L6F6 line (Figure 3C). We observed that twice less females were able to infect the 20rDNA L6F6 in comparison with WT Col-0 plants (Figure 3C). However, we did not observe a change in the syncytium size, that correspond to reprogrammed plant root cells by the female to serve as a feeding site. Secondly, we tested the ability of the 20rDNA L6F6 to resist against the virulent bacteria *Pseudomonas* strain DC3000. Three days after inoculation, bacterial growth was significantly lower in the 20rDNA L6F6 line (Figure 3D). In conclusion, we were able to demonstrate that the higher accumulation of transcripts from genes implicated in the plant-pathogens response is able to maintain the 20rDNA L6F6 line in a primed state, increasing its ability to resist against at least two types of distinct pathogens.

Appearance of the duplication events in the 20rDNA line

One of the key question is when did these TDDOs come up? The 20rDNA L6F6 line was indeed created thanks to a cross between *fas1-4* and *fas2-4* mutants that both display low amount of rDNA copies (12, 24). We therefore tested by quantitative PCR (qPCR) the presence of the duplicated regions in the parent lines of origin and in the offspring. We focused on the TDDO3 and TDDO4, the two largest genomic duplication (Figure 2A). Our analyses revealed that the TDDO4 is already present in the *fas2-4* parent, but not in the *fas1-4*. It therefore segregated in F2 after their cross, staying in the line 6, but was lost in the line 9 (Figure 4A). Its absence/presence was also confirmed by DNA-Fluorescent *In situ* Hybridization (FISH) (Figure 4B). We used two probes generated from BAC clones: one recognizing a portion of the TDDO4 (hk4 – T5H22) and one recognizing an unduplicated genomic region located between the TDDO4 and the NOR4 (F5I10). Different cell-types were analysed from vegetative and reproductive tissues (Figure 4B and Suppl. Figure 10). While a similar number of signals were detected in WT Col-0 nuclei for both probes, more signals corresponding to the TDDO4 were detected in the 20rDNA L6F6 nuclei (Suppl. Figure 10). Analyses of pachytene chromosomes clearly demonstrated that the additional signal actually belong from the same chromosome, which confirm the duplication hypothesis (Figure 4B). In parallel, the analyses of nuclei from the 20rDNA L9 line have confirmed the absence of the TDDO4 at the 5th generation (F5) (Figure 4B and Suppl. Figure 10).

Interestingly, qPCR analyses revealed that the TDDO3 is absent from the *fas1-4* and *fas2-4* mutant genomes, as well as in the first generated offspring. Our data indeed suggest an appearance of the TDDO3 between the 3rd and 6th generation of Line 6 (Figure 4A). Late apparition of the TDDO3 is also confirmed by its absence from the 20rDNA L9 line.

Altogether, these analyses suggest that the TDDOs appeared independently, either in the parent line or in the offspring of the *fas1-4* and *fas2-4* cross (Figure 4C). Furthermore, the correlation between low rDNA abundance and TDDO appearance remains valid, since *fas2-4* also contained only 20% of the rDNA copies compare to WT.

Duplication events create chimeric genes

Tandem duplications in direct orientation keep genes intact and do not lead to gene loss. Truncated genes can be generated at the breakpoint junction, while keeping intact genes on the edges of duplication. However, when breakpoints are located in two different genes in the same orientation, gene fusion can be created if the reading frame is preserved. In the 20rDNA L6F6, we systematically analyzed the breakpoints junction (Suppl. Figure 11). Out of the 7 cases of TDDO identified in our study, 3 potentially create fused or truncated proteins (Figure 5). On the chromosome 1, the TDDO1 fused the first exon of the gene *AT1G55325* that encode for the N-terminal domain of the MEDIATOR 13-like with 4 of the 5 exons of *AT1G54770* that encode for the FCF2 pre-rRNA processing factor. On the chromosome 2, although genes are in the opposite orientation, the TDDO3 creates a shorter ORF of the *AT2G38460* gene that potentially produce a truncated FERROPORTIN 1 protein. Finally, on the chromosome 4, the TDDO4 fuses the *AT4G05475* gene to a TE (*AT4G02960*), leading to the potential expression of three new ORFs, including 1 which encode a protein with 2 Leucin Rich Repeats (LRR) (Figure 5). We then investigate whether these chimeric genes were transcribed. To our surprise, analysing our RNA-seq data revealed that they are all able to accumulate transcripts. However, their potential to be translated remains to be determined.

DISCUSSION

Genomic structural variations shape animal and plant genomes (37). Within a period of several millions of years, numerous rearrangements have occurred to shape the *Arabidopsis thaliana* genome and include duplications, translocations, inversions and deletions (38, 39). Recently, genome analysis of seven accessions of *Arabidopsis thaliana* revealed that they contain an average of 15 MB of rearranged sequences, generating CNVs for thousands of genes (40). In this case, deletion, gain or loss of copies are considered important sources of CNV, and have potentially occurred in tens of thousands of years of evolution (41). CNVs occurring in the context of tandem duplication events represent between 3 and 4 Mb of genomic sequences in each of the 7 accessions sequenced (40). In our case, only few generations were necessary to gain 2,3 Mb of genomic sequences by tandem duplications.

The rapid occurrence of these rearrangement is particularly intriguing. The relative sensitivity to genotoxic stress and the detection of a higher rate of spontaneous DSB in our 20rDNA line is certainly one of source of their appearance (Figure 2B and Suppl. Figure 4), but the mechanisms remain to be determined. One possibility is the implication of non-allelic homologous recombination (NAHR), usually responsible of TDDO with sometimes the implication of TEs (37, 42). In that case, segmental duplication or deletion occurs. In the 20rDNA L6F6, we only detect TDDO representing at least dozen of genes or TEs. The fact that we did not detect deletions is probably due to their important deleterious effect. TDDOs only generate gene CNVs, no *knock-out* or gene loss.

Two other particular aspects of the detected TDDOs are their large sizes and locations, ranging from 60 kb to 1.44 Mb on 4 of the 5 chromosomes. (Figure 2A). The seven TDDO sites do not share genetic features and breaking junctions are for example not enriched in

repetitive elements or genes for example (Suppl. Figure 3). However, sign-changes in the Eigenvector seem to be overrepresented at breaking junctions, suggesting a potential link between the 3D genome folding and the occurrence of TDDOs. The systematic identification and characterization of additional TDDOs would be necessary to strength this potential link.

It is also, intriguing that 5 out the 7 TDDO, which represent 89% of the total segmental duplications, are located on NOR-bearing chromosomes (Figure 2A). Due to their tandemly repeated nature, NORs are indeed subjected to an inherent instability. Therefore, the existence of a sensing system monitoring their abundance have been proposed (42), potentially *via* unequal sister chromatid exchange (43, 44). The 20rDNA lines have been created after the cross of *fas1* and *fas2* mutants, whose mutations provoked a gradual loss of rRNA genes copies (24). Interestingly, lineages 6 and 9 of 20rDNA are the only siblings where the number of rRNA genes remained stably at a low level, whereas all other lineages quickly acquired rRNA genes (25). However, our data actually show that rRNA genes copies is increasing progressively in the 20rDNA line L6F6 (Figure 4A), suggesting that the CNVs are found not only at the level of the TDDOs, but also at the level of the NORs. If a link between the rRNA genes gain and the appearance of TDDO exists and if the same mechanisms are involved remains to be elucidated. Though, in cancer cells, a loss of rRNA genes copies have been shown to be associated with genomic instability and hypersensitivity to DNA damage (45, 46).

Short term consequences of gene duplications have been studied in animal, especially in cancer cells, were multiple *de novo* tandem duplication events induce *de novo* gene CNVs (47, 48). The 20rDNA L6F6 is an unprecedent opportunity to study the transcriptional behavior of newly duplicated genes. Globally, duplicated genes tend to be more expressed

(Figure 2J). In the most recent duplication located on the chromosome 2 (TDDO2), 49 out of the 97 genes are at least upregulated more than 2 fold, while 15 of them can be considered as unexpressed genes. In the short term, the expression of tandem gene duplicates is often greater than two-fold (49). Although we cannot exclude that the detected transcripts come from only one of the duplicated genes, it is more likely that equivalent additive expression will occur for the duplicated genes. During evolution, duplicated gene expression quickly evolve to specialized, often in a tissue-specific manner, although a significant number retain correlated transcriptional profiles (50, 51). In our case, we were able to correlate this change in gene expression with the acquisition of increased resistance to pathogens (Figure 3). Analyzing the expression profile of the genes present in the TDDO2 as well as the ability of the 20rDNA L6F6 to resist against pathogens in future generations will allow us to evaluate if rapid transcriptional regulation will occur to modulate their expression.

Plant genomes are rapidly evolving and their potential to adapt to environmental changes is crucial. Like genome hybridization and TEs mobilization, TDDO is one important aspect enabling genome evolution (52, 53). Previous observation and our data demonstrate the importance of systematical detection of gene CNVs and the presence of chimeric genes formation due to duplication events. CNVs can indeed associate with adaptive traits and arise as a underexplored driven force of evolution (53, 54). In our case, we indeed found a potential link between CNV and pathogen resistance (Figure 3 and Suppl. Figure 9). CNV were already shown to be implicated in nematode resistance in soybean (55), but also in potato cultivar genome heterogeneity (56). The 20rDNA lines offer an unprecedented tool to study the occurrence of CNV and their consequences in the time frame of few generations. This last point is particularly interesting in the context of plant breeding. In addition, TDDOs have the potential to create chimeric genes (Figure 5). TDDO events have been shown to promote

cancer cells formation, *via* the activation of oncogenes (48). In that case, breaking junctions can indeed change the expression of an oncogene by modifying its regulation by enhancer for examples. In our study, the chimeric genes created are expressed and although we do not have evidence concerning their potential ability to be translated or if the protein produced would be stable, it is tempting to speculate that they can be used by the organism by exaptation. Studying the consequences of TDDOs on future generations will certainly shed light on their potential impact on genome evolution and plant adaptation.

MATERIAL ET METHODS

Plant materials

Seeds corresponding to the *fas1-4* (SAIL_662_D10) and *fas2-4* (SALK_033228) were previously reported in (57). 20rDNA lines depicted in Figure 4C were previously reported in (25). For NADs identification, wild-type Col-0 expressing the FIB2:YFP fusion protein was described in (12). The 20rDNA L6F5 line was transformed by agroinfiltration to insert a transgene expressing FIB2:YFP fusion protein as described previously (12).

NADs identification

Nuclei and nucleoli were isolated as previously described (27) using a S3 cell sorter (Biorad®). Sorted nuclei or nucleoli were treated with RNase A and proteinase K prior to purification and concentration of the DNA using a phenol/chloroforme purification, followed by two precipitation steps. DNA libraries were generated via the kit Nextera XT DNA sample preparation (Illumina®) according to manufacturer's instruction, and were then subjected to high throughput paired-end 2X150nt sequencing on a Next-seq 550 apparatus (Illumina ®) at

the Bioenvironment sequencing platform (Perpignan University, Perpignan, France). NADs identification was then performed as described in (28).

Chromatin-chromatin interactions analyses

Hi-C samples were generated according to (30), using HindIII as a restriction enzyme. For both wild-type (Columbia-0 accession) and 20rDNA, Hi-C samples were generated in triplicates using 14-day-old seedling populations grown on MS culture media (22°C; 16h light/8h dark). Single ends of paired-end sequencing reads were trimmed to 35 bp using cutadapt (58) and aligned separately using bowtie (59) using the following parameters: bowtie -a -v 0 -m 1. After sorting and indexing using samtools (60), single end alignments were further processed using HiCdat (61). Before mapping read-pairs to genomic bins (10 kb and 50 kb), read pairs were filtered to remove pairs, which paired inwards with less than 1kb and outward with less than 25 kb. The resulting Hi-C matrices were further analyzed in R using HiCdatR package (61) and other in-house R scripts. Hi-C matrices were normalized for sequencing depth (fragment counts per million). To normalize for coverage, we mapped single-end alignments to genomic bins (equal bin size than the corresponding Hi-C matrices) and created bed files indicating the coverage for each genomic bin. Subsequently, the bed files were normalized to obtain read per million values. Hi-C matrices were subsequently converted to coverage-normalized matrices using the R-base scale() function. Principal component analysis (PCA) were performed using f.principle.component.analysis() function from the HiCdatR package using either entire chromosomes or chromosome arms only, excluding pericentromeric regions (Chr1:13Mb-17Mb; Chr2:2Mb-5Mb; Chr3: 11.5Mb-15.5Mb; Chr4: 3Mb-5Mb; Chr5:10Mb;14Mb). Hi-C contact densities were assessed by calculating median interaction frequency of all contacts within 5 times the bin size of a given genomic bin. As input coverage-normalized Hi-C matrices were used. Significant differences

between Hi-C data sets were determined by performing individual Student t-tests on each entry of the triplicate Hi-C matrices. Subsequently, we determined 5 categories of changes in contact frequencies (significantly less, less, no change, more, and significantly more). Significant difference matrices were generated using these categories and plotted with an according color code. All plots were generated using R-base.

Nanopore sequencing and data analyses

Genomic DNA preparation was performed as previously described in (62). After Qubit dosage (dsDNA High Sensitivity (Thermo Fisher Scientific, USA), a second step of DNA purification was performed with the Genomic DNA Clean & Concentrator kit (Zymo Research, USA) and precipitated. A last Qubit dosage was performed before library preparation using the 1D Genomic DNA by ligation kit SQK-LSK109 (Oxford Nanopore Technologies, UK), following manufacturer's instructions. The R9.5 ONT flow-cell FLO-MIN106D (Oxford Nanopore Technologies, UK) was used.

ONT reads were mapped on the TAIR10 reference genome using minimap2 with -a -Q -map-ont options. The alignment files were converted into bed files using bedtools and the coverage per 100 kb window was calculated using coverageBed. For each 100 kb window the ratio $r = 20\%rDNA \text{ coverage} / \text{WT coverage}$ was calculated. The mean m and standard error (SE) were calculated across the entire genome. Differentially covered regions in the 20%rDNA line were defined as regions for which: $r \geq m + 2SE$ or $r \leq m - 2SE$.

Transcriptome analyses

Total RNA was extracted from four pools of 3-week-old *Arabidopsis* plant leaf tissues of wild-type Col-0 or 20rDNA L6F6 using TRIzol reagent (MRC). Sequencing was performed by the Bioenvironment sequencing platform (Perpignan University, Perpignan, France) using

a Nextseq 550 to generate 2×75 -bp-long reads. Illumina reads from non-stranded, polyA+ RNA deep sequencing libraries were aligned to the *A. thaliana* TAIR10-annotated genome reference using HISAT 2 (63). Count the number of reads aligned to each genome coding sequences was performed with HTseq-count (64) and differential expression profile analyses with DESeq2 (Bioconductor - R package) (65).

Methylome analyses

Total genomic DNA was extracted from three pools of 3-week-old *Arabidopsis* plant leaf tissues of wild-type Col-0 or 20rDNA L6F6 using the Illustra Phytopure DNA extraction kit (GE Healthcare®, UK), following manufacturer's instructions. Bisulfite treatment, libraries preparation and sequencing was performed by the Novogene company (Hong-Kong) using the Illumina® technology. DMRs were identified with Bismark bisulfite mapper (66) and Methylkit (67).

Immunostaining

Nuclei of *A. thaliana* plants were isolated and immunostained as described (72). A antibody targeting specifically the Phosphorylated γ -H2Ax was used at a dilution of 1:500 in PBS and applied onto a slide pre-coated with nuclei and incubated overnight at 4°C. The nuclei were labeled with anti-rabbit-Alexa Fluor 488 antibodies (Abcam) at a dilution of 1:1000, counterstained with DAPI (4', 6-diamidino-2-phenylindole), mounted using the Vectashield medium (Vector Laboratories), and analyzed in a Zeiss LSM 700 inverted confocal microscope.

Bleomycin assay

Sterile were sowed on Petri dishes containing medium ~ MS (Murashige and Skoog, Duchefa Biochemie M0231) 1X agar: 4.7g / L of MS medium, 5g / L of MES and 10g / L of Plant Agar (Gold Biotechnology) with a pH 5,7. For the treatment, media was supplemented with 1.5 ~ µg / mL bleomycin to obtain a 10^{-6} M concentration of bleomycin.

Nematode Assay

For the infection assay, ten day-old *A. thaliana* seedlings grown on modified Knop's medium at 24°C under 16 h-light/8 h-dark were inoculated with J2 *H. schachtii* nematodes (68). Four weeks after inoculation, adult females in each plant were counted, and the data were analyzed by a modified *t*-test using the Statistical Software Package SAS ($P < 0.05$). Root segments containing *H. schachtii*-induced syncytia and adjacent root segments without syncytia were dissected under a light microscope as described in (69). Syncytium size measurements was performed 21 days after inoculation of *A. thaliana* with *H. schachtii* (70). For each line, 3 pool of 35 single-female syncytia were randomly selected, size was measured and average size for each line determined. Statistically significant differences were determined in a modified *t*-test using the statistical software package SAS ($P < 0.05$).

DC3000 infection assay and bacteria counting

Four plants per condition were dip-inoculated using *Pseudomonas* strain DC3000 at 5×10^7 cfu/ml (OD₆₀₀ of 0.2 corresponds to 10^8 cfu/ml) supplemented with 0.02 % Silwet L-77, and immediately placed in chambers with high humidity. Bacterial growth was determined 3 days after infection. For the quantification, infected leaves were harvested, washed for one minute in 70 % (v/v) EtOH and one minute in water. Leaf discs with a diameter of 5 mm have been excised, ground and homogenized in 200 µl of 10 mM MgCl₂. Each data point consists of

four leaf discs. 10 µl of each homogenate were then plated undiluted and at different dilutions onto NYGA plates. Growth of bacteria was determined after 36 h of incubation at 28°C by colony counting.

Fluorescence in situ hybridization (FISH), slide preparation, probe labelling

For nuclei isolation for FISH (T5H22, F5I10 ratio) 14 days-old seedlings of WT Col-0 and L6 (generation F6) were used. The seedlings were first fixed in 4% paraformaldehyde for 15 min at RT, chopped with a razor blade in ice cold nucleus isolation buffer (NIB - 0.5 M sucrose; 10 mM EDTA; 2.5 mM DTT; 100 mM KCl; 1 mM spermine; 4 mM spermidine in 10 mM Tris-Cl, pH = 9.5), and filtered through a 50 µm and 30 µm pore-sized disposable filters (CellTrics, Sysmex, Germany). Afterwards, the filtrate was supplemented with 1/10 volume of 10% Triton-X in NIB and centrifuged at 2000 x g for 10 minutes at 4°C. Pellet, containing nuclei, was then resuspended in 1x PBS and nuclei were dripped onto superfrost slides (10 µl / slide). After brief drying at 4°C, cells were fixed in chilled 3:1 methanol / acetic acid mixture for 10 minutes and air dried. After a rinse in 2x SSC, slides were incubated with RNase A (Dnase free, Applichem) in a humid chamber for 45 minutes at 37°C. Slides were then rinsed in 2xSSC and immersed in 10mM HCl for 5 minutes and subsequently in 10mM HCl with 8 µg/ml pepsin for 5 minutes (both at 37°C). After two washes in 2xSSC, slides were dehydrated in ethanol series (70%, 80%, 96%), 2 minutes each, and air-dried.

For pachytene chromosomes, entire inflorescences were fixed in ethanol:glacial acetic acid (3:1) overnight and stored in 70% ethanol at -20°C until use. WT Col-0, L6 (generation F4) and L9 (generation F6) were used. Selected immature flower buds were rinsed in distilled water (2 × 5 min) and in citrate buffer (10 mM sodium citrate, pH 4.8; 2 × 5 min), then anthers were dissected and incubated in an enzyme mix 0.15% each of cellulase (Onozuka R10, Serva), cytohelicase (Sigma), and pectolyase (Duchefa) in citrate buffer (10 mM sodium

citrate buffer pH 4.5), at 4°C ON. Individual anthers were put on the microscope slide, disintegrated by the needle in a drop of citrate buffer. Then the suspension was softened by adding 15 to 30 µL of 50% acetic acid and spread by stirring with a needle on a hot plate at 50°C for 1 min. Chromosomes were fixed by adding 100 µL of ethanol:acetic acid (3:1) fixative, the slides were air dried, fixed in 4% formaldehyde dissolved in 2xSSC for 10 min, washed 3x5min in 2xSSC, and treated with RNase and pepsin (5-15min) as described above. After the pepsin treatment, slides were rinsed in 2xSSC, 3x5min and the fixation in paraformaldehyde was repeated, slides washed and dried in the ethanol series, 2 min each, air dried and hybridized with appropriate probes.

Hybridization conditions were used as follows: 50% formamide (Sigma)/10% dextran sulfate/2x SSC solution, denaturation for 3 minutes at 80 °C, overnight incubation at 37 °C. Afterwards, slides were rinsed in 2x SSC at RT, then washed 2x5 min in 50% formamide in 2x SSC and 2x5 min in 2x SSC, (both at 42°C). Slides were blocked in 5% BSA in 4T buffer (4x SSC, 0.05% Tween-20) for 30 minutes at RT. After blocking, incubation with Alexa fluor 594-conjugated streptavidin (#S32356; Life technologies) and Alexa fluor 488-conjugated anti-digoxigenin (#IC75206; R&D systems) was performed for 45 minutes at RT. Antibodies were diluted in blocking buffer (1:200 and 1:10, respectively). Subsequently, slides were washed 2x5 min in 4T buffer and 2x5 min in 2xSSC buffer at 37°C with mild shaking.

Slides were stained with DAPI (4,6-diamidino-2-phenylindole, 2µg/ml) diluted in Vectashield (Vector laboratories), images were acquired using a Zeiss Axioimager Z1 or Olympus BX61, with appropriate filter set (AHF Analysentechnik, <http://www.ahf.de/>). Images were processed in Photoshop and Image J (<https://imagej.nih.gov/ij/>) softwares.

The BAC probes (T5H22, F5I10, available at www.arabidopsis.org) were kindly provided by T.Mandáková (CEITEC MU, Brno, CR). Approximately 1 µg of isolated BAC DNA

(NucleoBond® Xtra Midi kit, Macherey-Nagel) were labeled by nick translation with biotin-dUTP or digoxigenin-dUTP (Jena Bioscience) using either commercial nick translation labelling kit (#07J00-001; Abbot - labelling was performed according to the manufacturer's instructions using 5 µl of the enzyme mix and 8h labelling time); or by home-made reaction mix (71), containing 5 µL of 10× NT buffer (0.5 M Tris-HCl, pH 7.5, 50 mM MgCl₂, and 0.05% BSA), 10mM β-mercaptoethanol, 0.2mM each dATP, dCTP, dGTP and 40 µM dTTP (Jena Bioscience), 0.2 mM labelled-dUTP, 3 µL of DNase I (#10104159001, Roche ,diluted to 8 µg/ml) and 5U of DNA polymerase I (#EP0041,Fermentas).The nick translation mixture was incubated at 15°C for ~ 4h (or longer) to obtain a fragment length of ~200 to 500 bp. The reaction was stopped by adding 1 µL of 0.5 M EDTA, pH 8.0, and incubation at 65°C for 10 min.

Quantification of foci in isolated interphase nuclei

In the first analysis we counted the number of hk4 and subNOR4 signals in interphase nuclei and calculated their ratio. Two independent counts were performed and the ratio of hk4 to subNOR4 was evaluated foci in both wild-type cells and L6 cell line. To avoid conscious bias, one of the counts was blind and integrated all the foci into the analysis (singlet, doublet, partially split etc.) The results presented in Suppl. Figure 10 are the average of the two counts. The floral bud nuclei showed were 2, 3 or 4 of hk4 signals. The number and spatial distribution of these signals were evaluated. Total counts are shown in the table in (Suppl. Figure 10B).

FUNDINGS

This work and APC PhD fellowship are supported by the ANR JCJC NucleoReg [ANR-15-CE12-0013-01] to FP. FP was supported by the French Laboratory of Excellence project TULIP (ANR-10-LABX-41 and ANR-11-IDEX-0002-02).

AUTHORS CONTRIBUTION:

FP and APC conceived and designed the analysis, APC SG MF TA TM CL EJ FP collected the data, APC SG PZ TB LN MD MM FP contributed data or analysis tools, APC SG NP MM FP performed the analysis, FP wrote the paper, APC SG MM edited the paper, FP acquired funding.

ACKNOWLEDGEMENTS

The authors are grateful to Rémy Merret and Michèle Laudie for Illumina sequencing and lab members for fruitful discussions. We also thank the flow cytometry facility, the microscopic facility and the sequencing facility of Perpignan University *Via Domitia* Bioenvironnement (Perpignan, France).

REFERENCES

1. C. Bersaglieri, R. Santoro, Genome Organization in and around the Nucleolus. *Cells* **8** (2019).
2. A. Nemeth, *et al.*, Initial genomics of the human nucleolus. *PLoS Genet.* **6**, e1000889 (2010).
3. S. van Koningsbruggen, *et al.*, High-resolution whole-genome sequencing reveals that specific chromatin domains from most human chromosomes associate with nucleoli. *Mol. Biol. Cell* **21**, 3735–3748 (2010).
4. W. A. Bickmore, B. van Steensel, Genome architecture: domain organization of interphase chromosomes. *Cell* **152**, 1270–1284 (2013).
5. T. A. Dittmer, N. J. Stacey, K. Sugimoto-Shirasu, E. J. Richards, LITTLE NUCLEI genes affecting nuclear morphology in *Arabidopsis thaliana*. *Plant Cell* **19**, 2793–2803 (2007).

6. B. Hu, *et al.*, Plant lamin-like proteins mediate chromatin tethering at the nuclear periphery. *Genome Biol.* **20**, 87 (2019).
7. F. Pontvianne, *et al.*, Identification of Nucleolus-Associated Chromatin Domains Reveals a Role for the Nucleolus in 3D Organization of the *A. thaliana* Genome. *Cell Rep.* **16**, 1574–1587 (2016).
8. C. Picart, F. Pontvianne, Plant nucleolar DNA: Green light shed on the role of Nucleolin in genome organization. *Nucl. Austin Tex* **8**, 11–16 (2017).
9. G. P. Copenhaver, C. S. Pikaard, RFLP and physical mapping with an rDNA-specific endonuclease reveals that nucleolus organizer regions of *Arabidopsis thaliana* adjoin the telomeres on chromosomes 2 and 4. *Plant J. Cell Mol. Biol.* **9**, 259–272 (1996).
10. F. Pontvianne, *et al.*, Nucleolin is required for DNA methylation state and the expression of rRNA gene variants in *Arabidopsis thaliana*. *PLoS Genet.* **6**, e1001225 (2010).
11. C. Chandrasekhara, G. Mohannath, T. Blevins, F. Pontvianne, C. S. Pikaard, Chromosome-specific NOR inactivation explains selective rRNA gene silencing and dosage control in *Arabidopsis*. *Genes Dev.* **30**, 177–190 (2016).
12. F. Pontvianne, *et al.*, Subnuclear partitioning of rRNA genes between the nucleolus and nucleoplasm reflects alternative epiallelic states. *Genes Dev.* **27**, 1545–1550 (2013).
13. A. Picart-Piccolo, N. Picault, F. Pontvianne, Ribosomal RNA genes shape chromatin domains associating with the nucleolus. *Nucl. Austin Tex*, 1–6 (2019).
14. T. Kobayashi, Regulation of ribosomal RNA gene copy number and its role in modulating genome integrity and evolutionary adaptability in yeast. *Cell. Mol. Life Sci. CMLS* **68**, 1395–1403 (2011).
15. E. B. Dopman, D. L. Hartl, A portrait of copy-number polymorphism in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 19920–19925 (2007).
16. J. G. Gibbons, A. T. Branco, S. A. Godinho, S. Yu, B. Lemos, Concerted copy number variation balances ribosomal DNA dosage in human and mouse genomes. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 2485–2490 (2015).
17. F. A. Rabanal, *et al.*, Unstable Inheritance of 45S rRNA Genes in *Arabidopsis thaliana*. *G3 Bethesda Md* **7**, 1201–1209 (2017).
18. Q. Long, *et al.*, Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. *Nat. Genet.* **45**, 884–890 (2013).
19. S. Ide, T. Miyazaki, H. Maki, T. Kobayashi, Abundance of ribosomal RNA gene copies maintains genome integrity. *Science* **327**, 693–696 (2010).
20. S. Paredes, A. T. Branco, D. L. Hartl, K. A. Maggert, B. Lemos, Ribosomal DNA deletions modulate genome-wide gene expression: “rDNA-sensitive” genes and natural variation. *PLoS Genet.* **7**, e1001376 (2011).

21. T. Kobayashi, M. Sasaki, Ribosomal DNA stability is supported by many 'buffer genes'-introduction to the Yeast rDNA Stability Database. *FEMS Yeast Res.* **17** (2017).
22. B. Lemos, L. O. Araripe, P. Fontanillas, D. L. Hartl, Dominance and the evolutionary accumulation of cis- and trans-effects on gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 14471–14476 (2008).
23. E. Ramirez-Parra, C. Gutierrez, The many faces of chromatin assembly factor 1. *Trends Plant Sci.* **12**, 570–576 (2007).
24. I. Mozgova, P. Mokros, J. Fajkus, Dysfunction of chromatin assembly factor 1 induces shortening of telomeres and loss of 45S rDNA in *Arabidopsis thaliana*. *Plant Cell* **22**, 2768–2780 (2010).
25. V. Pavlistova, *et al.*, Phenotypic reversion in *fas* mutants of *Arabidopsis thaliana* by reintroduction of FAS genes: variable recovery of telomeres with major spatial rearrangements and transcriptional reprogramming of 45S rDNA genes. *Plant J. Cell Mol. Biol.* **88**, 411–424 (2016).
26. S. A. Quinodoz, *et al.*, Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell* **174**, 744–757.e24 (2018).
27. F. Pontvianne, M. Boyer-Clavel, J. Saez-Vasquez, Fluorescence-Activated Nucleolus Sorting in *Arabidopsis*. *Methods Mol. Biol. Clifton NJ* **1455**, 203–211 (2016).
28. M.-C. Carpentier, A. Picart-Piccolo, F. Pontvianne, A Method to Identify Nucleolus-Associated Chromatin Domains (NADs). *Methods Mol. Biol. Clifton NJ* **1675**, 99–109 (2018).
29. A. Himmelbach, *et al.*, Discovery of multi-megabase polymorphic inversions by chromosome conformation capture sequencing in large-genome plant species. *Plant J. Cell Mol. Biol.* **96**, 1309–1316 (2018).
30. S. Grob, M. W. Schmid, U. Grossniklaus, Hi-C analysis in *Arabidopsis* identifies the KNOT, a structure with similarities to the flamenco locus of *Drosophila*. *Mol. Cell* **55**, 678–693 (2014).
31. E. Lieberman-Aiden, *et al.*, Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
32. L. Zapata, *et al.*, Chromosome-level assembly of *Arabidopsis thaliana* Ler reveals the extent of translocation and inversion polymorphisms. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E4052–4060 (2016).
33. M. J. E. Wubben, J. Jin, T. J. Baum, Cyst nematode parasitism of *Arabidopsis thaliana* is inhibited by salicylic acid (SA) and elicits uncoupled SA-independent pathogenesis-related gene expression in roots. *Mol. Plant-Microbe Interact. MPMI* **21**, 424–432 (2008).
34. J.-Y. Yang, *et al.*, betaC1, the pathogenicity factor of TYLCCNV, interacts with AS1 to alter leaf development and suppress selective jasmonic acid responses. *Genes Dev.* **22**, 2564–2577 (2008).

35. P. L. Nurnberg, *et al.*, The developmental selector AS1 is an evolutionarily conserved regulator of the plant immune response. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 18795–18800 (2007).
36. Y.-H. Yeh, Y.-H. Chang, P.-Y. Huang, J.-B. Huang, L. Zimmerli, Enhanced Arabidopsis pattern-triggered immunity by overexpression of cysteine-rich receptor-like kinases. *Front. Plant Sci.* **6**, 322 (2015).
37. K. V. Krasileva, The role of transposable elements and DNA damage repair mechanisms in gene duplications and gene fusions in plant genomes. *Curr. Opin. Plant Biol.* **48**, 18–25 (2019).
38. G. Blanc, A. Barakat, R. Guyot, R. Cooke, M. Delseny, Extensive duplication and reshuffling in the Arabidopsis genome. *Plant Cell* **12**, 1093–1101 (2000).
39. Y. Henry, M. Bedhomme, G. Blanc, History, protohistory and prehistory of the Arabidopsis thaliana chromosome complement. *Trends Plant Sci.* **11**, 267–273 (2006).
40. W.-B. Jiao, K. Schneeberger, Chromosome-level assemblies of multiple *Arabidopsis thaliana* accessions reveal hotspots of genomic rearrangements. *bioRxiv*, 738880 (2019).
41. A. Fulgione, A. M. Hancock, Archaic lineages broaden our view on the history of Arabidopsis thaliana. *New Phytol.* **219**, 1194–1198 (2018).
42. J. Zhang, T. Zuo, T. Peterson, Generation of tandem direct duplications by reversed-ends transposition of maize ac elements. *PLoS Genet.* **9**, e1003691 (2013).
43. K. D. Tartof, Unequal mitotic sister chromatid exchange and disproportionate replication as mechanisms regulating ribosomal RNA gene redundancy. *Cold Spring Harb. Symp. Quant. Biol.* **38**, 491–500 (1974).
44. K. D. Tartof, Unequal mitotic sister chromatid exchange as the mechanism of ribosomal RNA gene magnification. *Proc. Natl. Acad. Sci. U. S. A.* **71**, 1272–1276 (1974).
45. M. Wang, B. Lemos, Ribosomal DNA copy number amplification and loss in human cancers is linked to tumor genetic context, nucleolus activity, and proliferation. *PLoS Genet.* **13**, e1006994 (2017).
46. B. Xu, *et al.*, Ribosomal DNA copy number loss and sequence variation in cancer. *PLoS Genet.* **13**, e1006771 (2017).
47. Y. Wee, T. Wang, Y. Liu, X. Li, M. Zhao, A pan-cancer study of copy number gain and up-regulation in human oncogenes. *Life Sci.* **211**, 206–214 (2018).
48. D. A. Quigley, *et al.*, Genomic Hallmarks and Structural Variation in Metastatic Prostate Cancer. *Cell* **174**, 758-769.e9 (2018).
49. D. W. Loehlin, S. B. Carroll, Expression of tandem gene duplicates is often greater than twofold. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 5988–5992 (2016).

50. G. Blanc, K. H. Wolfe, Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. *Plant Cell* **16**, 1679–1691 (2004).
51. K. Guschanski, M. Warnefors, H. Kaessmann, The evolution of duplicate gene expression in mammalian organs. *Genome Res.* **27**, 1461–1474 (2017).
52. L. Quadrana, *et al.*, Transposition favors the generation of large effect mutations that may facilitate rapid adaption. *Nat. Commun.* **10**, 3421 (2019).
53. I. Gabur, H. S. Chawla, R. J. Snowdon, I. A. P. Parkin, Connecting genome structural variation with complex traits in crop plants. *TAG Theor. Appl. Genet. Theor. Angew. Genet.* **132**, 733–750 (2019).
54. F. A. Kondrashov, Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc. Biol. Sci.* **279**, 5048–5057 (2012).
55. D. E. Cook, *et al.*, Copy number variation of multiple genes at Rhg1 mediates nematode resistance in soybean. *Science* **338**, 1206–1209 (2012).
56. G. M. Pham, *et al.*, Extensive genome heterogeneity leads to preferential allele expression and copy number-dependent expression in cultivated potato. *Plant J. Cell Mol. Biol.* **92**, 624–637 (2017).
57. V. Exner, P. Taranto, N. Schonrock, W. Gruissem, L. Hennig, Chromatin assembly factor CAF-1 is required for cellular differentiation during plant development. *Dev. Camb. Engl.* **133**, 4163–4172 (2006).
58. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 3 (2011).
59. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
60. H. Li, *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinforma. Oxf. Engl.* **25**, 2078–2079 (2009).
61. M. W. Schmid, S. Grob, U. Grossniklaus, HiCdat: a fast and easy-to-use Hi-C data analysis tool. *BMC Bioinformatics* **16**, 277 (2015).
62. E. Debladis, C. Llauro, M.-C. Carpentier, M. Mirouze, O. Panaud, Detection of active transposable elements in Arabidopsis thaliana using Oxford Nanopore Sequencing technology. *BMC Genomics* **18**, 537 (2017).
63. D. Kim, B. Langmead, S. L. Salzberg, HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
64. S. Anders, P. T. Pyl, W. Huber, HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinforma. Oxf. Engl.* **31**, 166–169 (2015).
65. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

66. F. Krueger, S. R. Andrews, Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinforma. Oxf. Engl.* **27**, 1571–1572 (2011).
67. A. Akalin, *et al.*, methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* **13**, R87 (2012).
68. T. J. Baum, M. J. Wubben, K. A. Hardy, H. Su, S. R. Rodermel, A Screen for *Arabidopsis thaliana* Mutants with Altered Susceptibility to *Heterodera schachtii*. *J. Nematol.* **32**, 166–173 (2000).
69. D. Hermsmeier, J. K. Hart, M. Byzova, S. R. Rodermel, T. J. Baum, Changes in mRNA abundance within *Heterodera schachtii*-infected roots of *Arabidopsis thaliana*. *Mol. Plant-Microbe Interact. MPMI* **13**, 309–315 (2000).
70. T. Hewezi, *et al.*, Cellulose binding protein from the parasitic nematode *Heterodera schachtii* interacts with *Arabidopsis* pectin methylesterase: cooperative cell wall modification during parasitism. *Plant Cell* **20**, 3080–3093 (2008).
71. T. Mandakova, M. A. Lysak, Chromosomal phylogeny and karyotype evolution in x=7 crucifer species (Brassicaceae). *Plant Cell* **20**, 2559–2570 (2008).
72. N. Durut, *et al.*, A duplicated NUCLEOLIN gene with antagonistic activity is required for chromatin organization of silent 45S rDNA in *Arabidopsis*. *Plant Cell* **26**, 1330–1344 (2014).
73. M. Krzywinski, *et al.*, Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
74. H. Thorvaldsdottir, J. T. Robinson, J. P. Mesirov, Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).

FIGURE LEGENDS

Figure 1 : Affected 3D genome organization in the 20rDNA line 6

A. Schematic representation of the obtaining of the 20rDNA line 6 and its relative content in rRNA gene copies. NOR2 and NOR4 rRNA gene copies are both affected by the reduction and according DNA-FISH experiment, their nuclear distribution in the nucleus changes compare to WT, since all rRNA gene copies associate with the nucleolus (25). **B.** Chromosome plots displaying the relative enrichment of a given genomic segment with the nucleolus. The y-axis displays the fold change nucleolus enrichment between wild-type Col-0

and the 20rDNA line 6. Each dot represents a 100kb window. Nucleolus-enriched genomic regions above the threshold (red-dotted line) are colored in red, while depleted regions are colored in violet. **C.** Coverage normalized T-test difference matrix (50 kb bins). The color of each pixel of the matrix is defined by the result of a T-test using the triplicate contact frequencies from wild type and 20rDNA coverage normalized Hi-C samples, respectively. (alternative: T-test difference matrix (50 kb bins). The color of each pixel of the matrix is defined by the result of a T-test using the triplicate contact frequencies from wild type and 20rDNA Hi-C samples, respectively. The two magnified area correspond to the two regions displaying the highest level of contact frequency changes **D.** Summed coverage over triplicate Hi-C data sets in 50 kb bins. The coverage was calculated based on unpaired single end raw Hi-C reads. **E-F** Non-normalized Hi-C snap shot showing the contact frequencies on the short arm of chromosome 4 in WT Col-0 (**E**) *versus* the 20rDNA L6F6 (**F**). TEs and genes are annotated to illustrate the occurrence of euchromatin and heterochromatin, respectively. **G.** Ratio between Hi-C contact frequencies from wild-type and 20rDNA L6F6. Negative ratios: More contacts in the wild-type, Positive ratios: More contacts in the 20rDNA L6F6. **H.** Eigenvector of the Col-0 WT Hi-C data set and annotation of the duplication affecting the knob hk4s. Note the central duplication breakpoint exactly coincides with a change between LSD and a CSD.

Figure 2: Identification of large tandem duplication in direct orientation and characterization of their transcriptome and epigenome.

A. Distribution of the reads obtained by nanopore sequencing along the chromosomes in the 20rDNA line 6 compare to wilt-type Col-0 (bottom_blue). Upregulated genes (red bars) and TEs (green bars) with an adjusted p value < 0.01 and a $\log_2\text{FoldChange} > 2$ (top). The blue arrow represents the localization and the orientation of the TDDO identified in the 20rDNA

line 6. **B.** Histogram showing the percentage of nuclei displaying at least one phosphorylated γ H2Ax foci in WT Col-0 *versus* 20rDNA line 6. **C-D.** Venn diagrams representing the proportion of duplicated TEs (C) or genes (D) that accumulated a higher level of transcripts, using an adjusted p value < 0.01 and a \log_2 FoldChange > 1.5 . **E-I.** global DNA methylation analyses from genome-wide bisulfite sequencing experiments in WT col-0 *versus* the 20rDNA line 6. Global CG, CHG and CHH methylation is shown for all genes (E), duplicated genes (F), Upregulated genes with an adjusted p value < 0.01 and a \log_2 FoldChange > 1.5 (G), all TEs (H) and duplicated TEs (I). **J.** Dot-plot revealing the relative expression of all genes or duplicated (DUP) genes in 3 week-old leaves in WT Col-0 or in the 20rDNA line 6. p value = 0,0005 (***) calculated using a Wilcoxon test.

Figure 3: The 20rDNA line 6 overexpressed biotic stress genes and is more resistant to nematod and bacterial pathogenes.

A. Go term enriched in the upregulated genes pool identified in the 20rDNA line 6. **B.** Histogram displaying the relative transcript enrichment for 8 genes implicated in the biotic stress response using qRT-PCR in WT Col-0 *versus* the 20rDNA line 6. **C.** WT Col-0 and the 20rDNA line 6 were inoculated with the sugar beet cyst nematode (*Heterodera schachtii*). 4 weeks after inoculation, the number of adult females per plant was determined. Data are the average number of adult females \pm SE (n = 35 X3). Data from the 3 independent experiments were pooled and are shown (Left histogram). The relative size of the syncytium cells was measured between both lines but no significant changes were noticed (right histogram). **D.** WT Col-0 and the 20rDNA line 6 were inoculated with Pseudomonas strain DC3000 at 5 x 10⁷ cfu/ml. Relative bacterial growth was determined 3 days after infection and is shown on the plot.

Figure 4: Apparition of the 2 major TDDO events.

A. CNV of genes present or not in TDDO and of rRNA genes were determined by qPCR. Their relative enrichment was determined in WT Col-0, the parent lines *fas1-4* and *fas2-4* and in 2 of the siblings line 6 and 9 during several generations. Relation between the lines are shown in (C). RRNA genes CNVs was followed *via* probes amplifying the 18S, the *loci* AT4G05475 and AT4G16580 that are not duplicated are controls, while the *loci* AT4TE12140 / AT4G05030 and AT2G38250 / AT2G38340 allow the identification of the 2 largest duplication TDDO4 and TDDO2 respectively. **B.** DNA-FISH analyses of 2 *loci* present on kr4s, distant by 1.6 Mb: one present on the TDDO4 (BAC T5H22 – green) and one absent of the duplicated region (BAC F5I10 – red). Three Pachytene chromosomes from WT Col-0, 20rDNA Line 6 and 20rDNA Line 9 are shown. A zoom of the unlooped kr4s from the middle panel of the 20rDNA line6 is presented at the bottom of the panel. **C.** Schematic representation of the obtaining of the 20rDNA line 6 and 9 and its relative. The putative appearance of the TDDO3 and TDDO4 events are shown.

Figure 5: TDDO provoke chimeric genes formation

A-C. Schematic representation of the TDDO 1, 3 and 4 that provoked TE and/or gene fusion in the 20rDNA line 6. Genes or TEs present in the breaking junction (and their orientation are shown. Potential fused protein sequences are displayed with the putative nature and their domains. **A.** Three ORFs can be generated from the fusion of large 5' sequence of a TE (AT4G02960) with the second and last exon of gene AT4G05475. The ORF2 could potentially fused a small TE coding sequence of two Leucine Rich Repeats of AT4G05475 (LRR). **B.** The chimeric gene created between AT2G37520 and AT2G38460 potentially encode for a truncated FERROPORTIN protein. **C.** The TDDO1 provoke the fusion of the

first exon of AT1G55325 that encode for an ATPase motif of MEDIATOR 13 and the last four exons of AT1G54770 that contains an RNA processing domain.

Supplemental Figure 1:

A-B. Chromosome plots displaying the relative enrichment of a given genomic segment with the nucleolus in WT Col-0 (A) and in the 20rDNA L6F6 (B). The y-axis displays the fold change between the nuclear *versus* the nucleolar DNA. Each dot represents a 100kb window. Nucleolus-enriched genomic regions above the threshold (red-dotted line) are colored in red.

Supplemental Figure 2:

A-B. HiC matrix displaying the chromatin-chromatin interactions identified by the chromosome capture conformation technique in wild-type Col-0 (A) and the 20rDNA L6F6 (B). Enriched interactions are displayed in yellow.

Supplemental Figure 3:

Eigenvector of the Col-0 wild-type Hi-C data set and annotation of the TDDO.

Supplemental Figure 4:

A. Representative confocal images showing the immunolocalization of the P- γ H2Ax foci (green) in WT Col-0 (top panels) or in 20rDNA F6L6 (bottom panels) isolated nuclei. The Blue staining correspond to the DNA labeled with DAPI. **B.** WT Col-0 and 20rDNA L6F6 seeds were grown for 18 days on an MS plate with or without Bleomycine at 10^{-5} M. Individual plant growth was then measured using their fresh weights.

Supplemental Figure 5:

A. Representation of the chromosomal position of genes and TEs of differentially accumulated transcripts in the 20rDNA L6F6 *versus* WT Col-0 using *Circos* (73). Up-regulated TEs are in green, Up-regulated genes are in red and down-regulated genes are in blue. **B.** QPCR analyses of differentially accumulated genes. Their relative accumulation was analyzed against the *ACTIN7* (red bars) or the *HC7* (blue bars) household genes.

Supplemental Figure 6:

A. Venn diagram showing the relation between genes that lost their association with the nucleolus in the 20rDNA L6F6 (lost NAD-genes in 20rDNA), Genes that are upregulated in 20rDNA L6F6 with an adjusted p -value < 0.01 and a $\log_2\text{FoldChange} > 2$ (Genes UP in 20rDNA) and with hypomethylated genes in 20rDNA L6F6. **B.** Venn diagram showing the relation between genes that gain association with the nucleolus in the 20rDNA L6F6 (Acquired NAD-genes in 20rDNA), Genes that are down-regulated in 20rDNA L6F6 with an adjusted p -value < 0.01 and a $\log_2\text{FoldChange} > 2$ (Genes Down in 20rDNA) and with hypermethylated genes in 20rDNA L6F6. **C.** Venn diagram showing the relation between TEs that lost their association with the nucleolus in the 20rDNA L6F6 (lost NAD-TEs in 20rDNA), TEs that are upregulated in 20rDNA L6F6 with an adjusted p -value < 0.01 and a $\log_2\text{FoldChange} > 2$ (TEs UP in 20rDNA) and with hypomethylated TEs in 20rDNA L6F6.

Supplemental Figure 7:

A-F. Dot plots showing the relative methylation levels in CG, CHG and CHH contexts in WT Col-0 or in 20rDNA L6F6. The analysis was performed by genome-wide bisulfite sequencing. 2 WT Col-0 and 2 20rDNA L6F6 (labelled 20) replicates (1 and 2) are shown. The average methylation between both replicates is shown (labelled ave.). The analysis was performed with All genes (A-B), with Upregulated genes with an adjusted p -value < 0.01 and a

$\log_2\text{FoldChange} > 1.5$ (C-D) and with the Duplicated genes (E-F). For this analysis, we separated Gene body sequences (top panels A, C and E) from the flanking sequences of the genes (5' and 3' – bottom panels B, D and F). *($p < 0,00025$); **($p < 5.10^{-10}$); ***($p < 2.10^{-16}$) calculated using a Wilcoxon test.

Supplemental Figure 8:

A-D. Dot plots showing the relative methylation levels in CG, CHG and CHH contexts in WT Col-0 or in 20rDNA L6F6. The analysis was performed by genome-wide bisulfite sequencing. 2 WT Col-0 and 2 20rDNA L6F6 (labelled 20) replicates (1 and 2) are shown. The average methylation between both replicates is shown (labelled ave.). The analysis was performed with All TEs (A-B) and with the Duplicated TEs (C-D). For this analysis, we separated TE body sequences (top panels A and C) from the flanking sequences of the TEs (5' and 3' – bottom panels B and D). *($p < 0,00025$); **($p < 5.10^{-10}$); ***($p < 2.10^{-16}$) calculated using a Wilcoxon test.

Supplemental Figure 9:

A-B. Genome browser (IGV - (74)) screenshot showing examples of Up-genes in 20rDNA L6F6 in duplicated regions. DNA coverage was obtained by nanopore sequencing for WT Col-0 (green) and for 20rDNA L6F6 (blue). Reads from RNA sequencing are shown for WT Col-0 (red) and for 20rDNA L6F6 (yellow). On the bottom, genes location is represented in blue and TEs in green. Gene names are surrounded in blue (non-significant) and in red (significant) when they are differentially accumulated transcripts with an adjusted p -value < 0.01 and a $\log_2\text{FoldChange} > 1.5$. This value is noted in red. Two locations are presented: the left border of the TDDO3 (A) and portion of the TDDO4 (B).

Supplemental Figure 10:

A-B. DNA-FISH analyses of a portion of the TDDO4 in leaf nuclei. **A.** Two *loci* present on kr4s, distant by 1.6 Mb: one present on the TDDO4 (BAC T5H22 – green) and one absent of the duplicated region (BAC F5I10 – red). Three leaf nuclei from WT Col-0 and 20rDNA Line 6 are shown and the ratio of green versus red signals as well as its ratio noted in the figure. **B.** In that case, the TDDO4 locus (BAC T5H22 – green) was analyzed in floral bud nuclei and four representatives images are presented in the figure. DNA is labelled with DAPI stain (Top panel, with DAPI in grey). Between 58 and 77 nuclei per genotype (WT Col-0, rDNA20 L6 and L9) and the distribution of each nuclei found per genotype is presented in the table.

Supplemental Figure 11:

Graphical representation of the 7 TDDO along the 5 chromosomes of *A. thaliana*. The genomic context of the breaking junctions is presented for each TDDO.

Figure 1

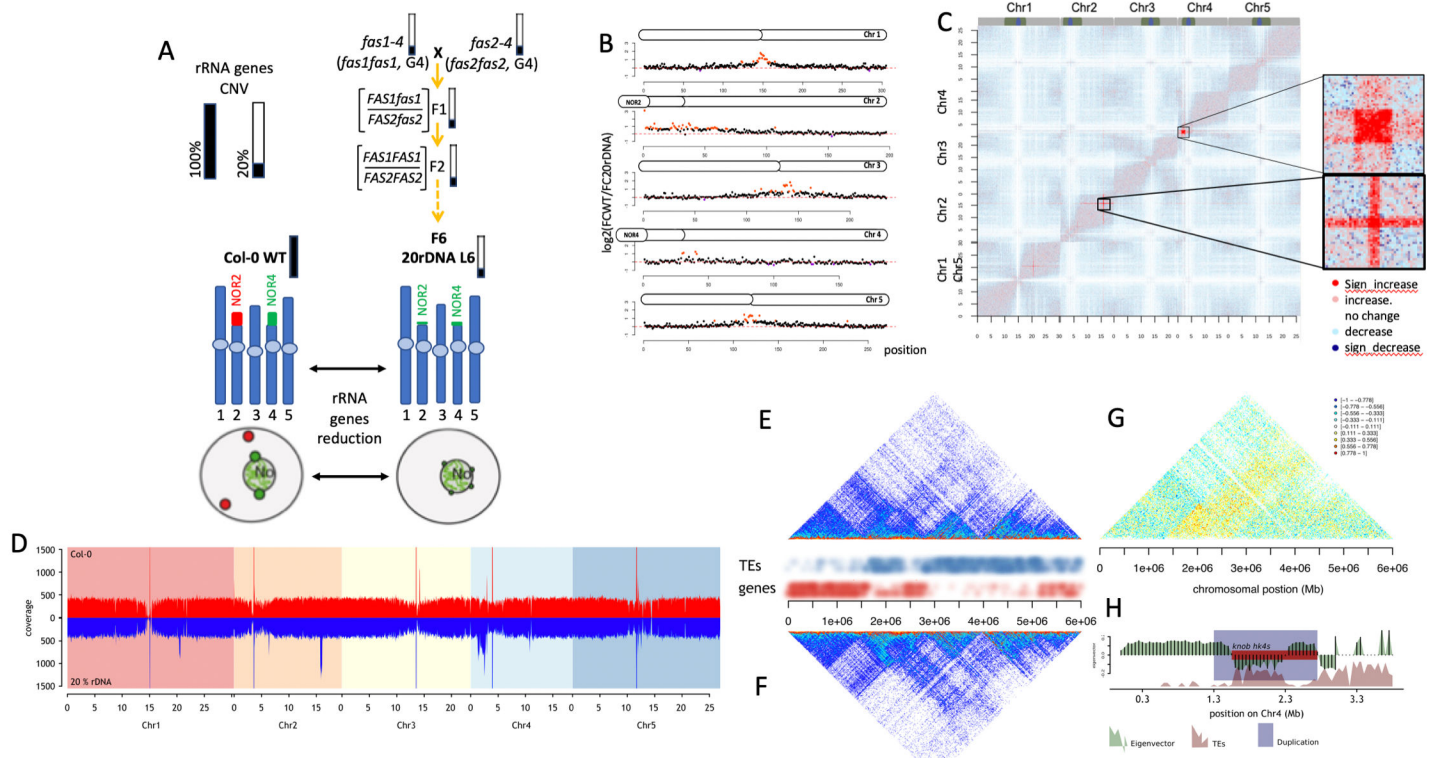


Figure 1 : Affected 3D genome organization in the 20rDNA line 6

A. Schematic representation of the obtaining of the 20rDNA line 6 and its relative content in rRNA gene copies. NOR2 and NOR4 rRNA gene copies are both affected by the reduction and according DNA-FISH experiment, their nuclear distribution in the nucleus changes compare to WT, since all rRNA gene copies associate with the nucleolus (25). **B.** Chromosome plots displaying the relative enrichment of a given genomic segment with the nucleolus. The y-axis displays the fold change nucleolus enrichment between wild-type Col-0 and the 20rDNA line 6. Each dot represents a 100kb window. Nucleolus-enriched genomic regions above the threshold (red-dotted line) are colored in red, while depleted regions are colored in violet. **C.** Coverage normalized T-test difference matrix (50 kb bins). The color of each pixel of the matrix is defined by the result of a T-test using the triplicate contact frequencies from wild type and 20rDNA coverage normalized Hi-C samples, respectively. (alternative: T-test difference matrix (50 kb bins). The color of each pixel of the matrix is defined by the result of a T-test using the triplicate contact frequencies from wild type and 20rDNA Hi-C samples, respectively. The two magnified area correspond to the two regions displaying the highest level of contact frequency changes **D.** Summed coverage over triplicate Hi-C data sets in 50 kb bins. The coverage was calculated based on unpaired single end raw Hi-C reads. **E-F** Non-normalized Hi-C snap shot showing the contact frequencies on the short arm of chromosome 4 in WT Col-0 (**E**) versus the 20rDNA L6F6 (**F**). TEs and genes are annotated to illustrate the occurrence of euchromatin and heterochromatin, respectively. **G.** Ratio between Hi-C contact frequencies from wild-type and 20rDNA L6F6. Negative ratios: More contacts in the wild-type, Positive ratios: More contacts in the 20rDNA L6F6. **H.** Eigenvector of the Col-0 WT Hi-C data set and annotation of the duplication affecting the knob hk4s. Note the central duplication breakpoint exactly coincides with a change between LSD and a CSD.

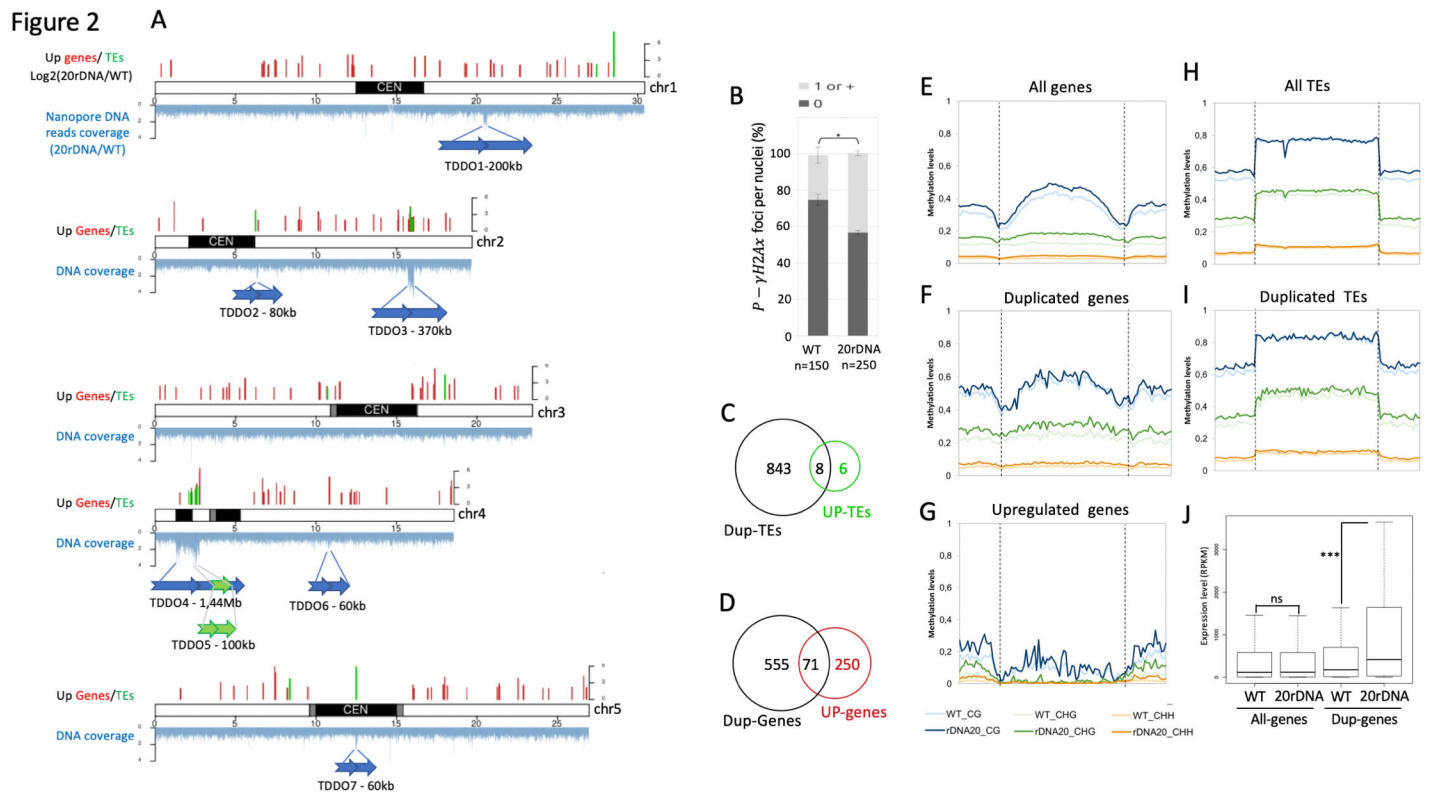


Figure 2: Identification of large tandem duplication in direct orientation and characterization of their transcriptome and epigenome.

A. Distribution of the reads obtained by nanopore sequencing along the chromosomes in the 20rDNA line 6 compare to wilt-type Col-0 (bottom_blue). Upregulated genes (red bars) and TEs (green bars) with an adjusted p value < 0.01 and a $\log_2\text{FoldChange} > 2$ (top). The blue arrow represents the localization and the orientation of the TDDO identified in the 20rDNA line 6. **B.** Histogram showing the percentage of nuclei displaying at least one phosphorylated γH2Ax foci in WT Col-0 versus 20rDNA line 6. **C-D.** Venn diagrams representing the proportion of duplicated TEs (C) or genes (D) that accumulated a higher level of transcripts, using an adjusted p value < 0.01 and a $\log_2\text{FoldChange} > 1.5$. **E-I.** global DNA methylation analyses from genome-wide bisulfite sequencing experiments in WT col-0 versus the 20rDNA line 6. Global CG, CHG and CHH methylation is shown for all genes (E), duplicated genes (F), Upregulated genes with an adjusted p value < 0.01 and a $\log_2\text{FoldChange} > 1.5$ (G), all TEs (H) and duplicated TEs (I). **J.** Dot-plot revealing the relative expression of all genes or duplicated (DUP) genes in 3 week-old leaves in WT Col-0 or in the 20rDNA line 6. p value = 0,0005 (***) calculated using a Wilcoxon test.

Figure 3

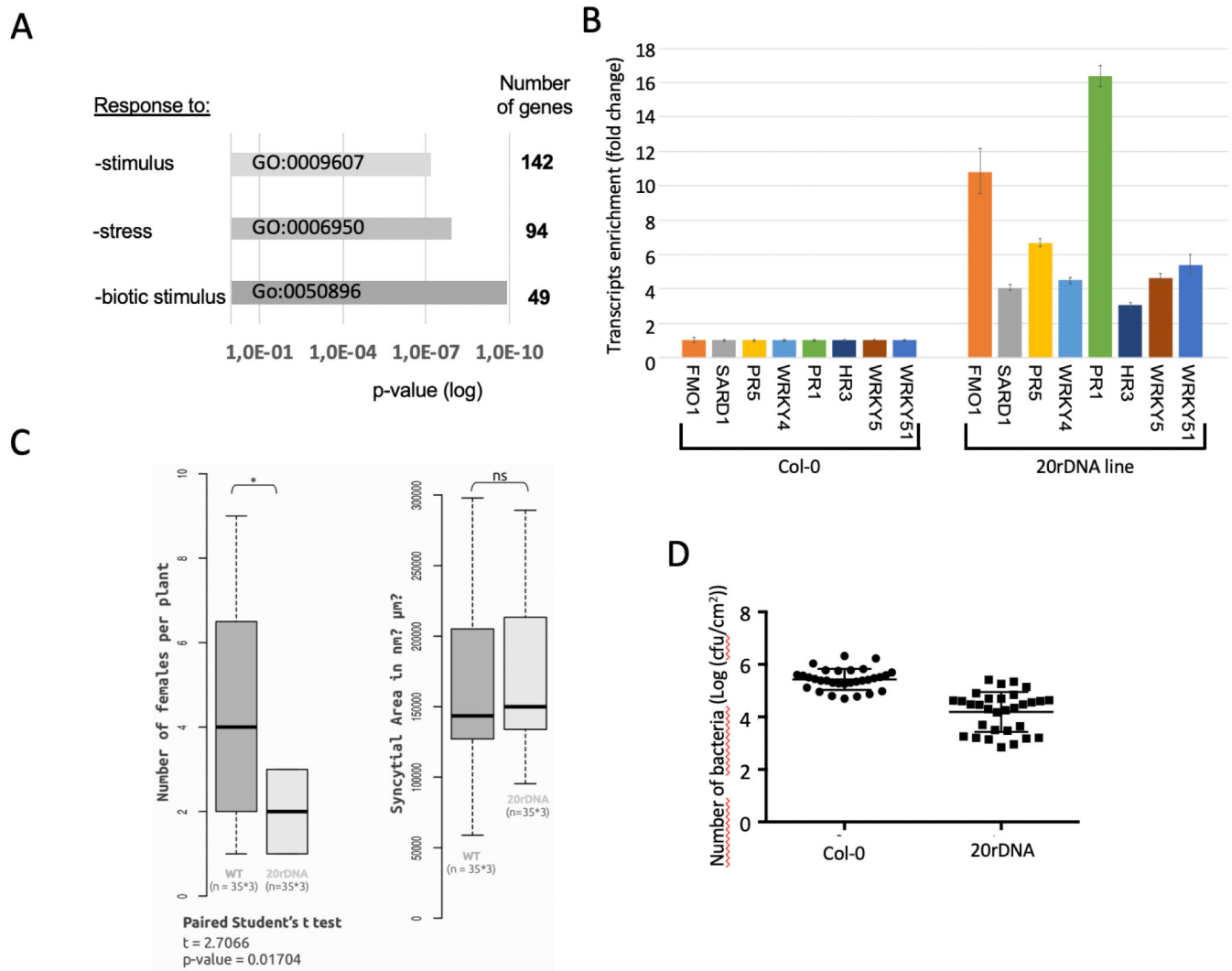


Figure 3: The 20rDNA line 6 overexpressed biotic stress genes and is more resistant to nematod and bacterial pathogens.

A. Go term enriched in the upregulated genes pool identified in the 20rDNA line 6. **B.** Histogram displaying the relative transcript enrichment for 8 genes implicated in the biotic stress response using qRT-PCR in WT Col-0 versus the 20rDNA line 6. **C.** WT Col-0 and the 20rDNA line 6 were inoculated with the sugar beet cyst nematode (*Heterodera schachtii*). 4 weeks after inoculation, the number of adult females per plant was determined. Data are the average number of adult females \pm SE (n = 35 X3). Data from the 3 independent experiments were pooled and are shown (Left histogram). The relative size of the syncytium cells was measured between both lines but no significant changes were noticed (right histogram). **D.** WT Col-0 and the 20rDNA line 6 were inoculated with *Pseudomonas* strain DC3000 at 5×10^7 cfu/ml. Relative bacterial growth was determined 3 days after infection and is shown on the plot.

Figure 4

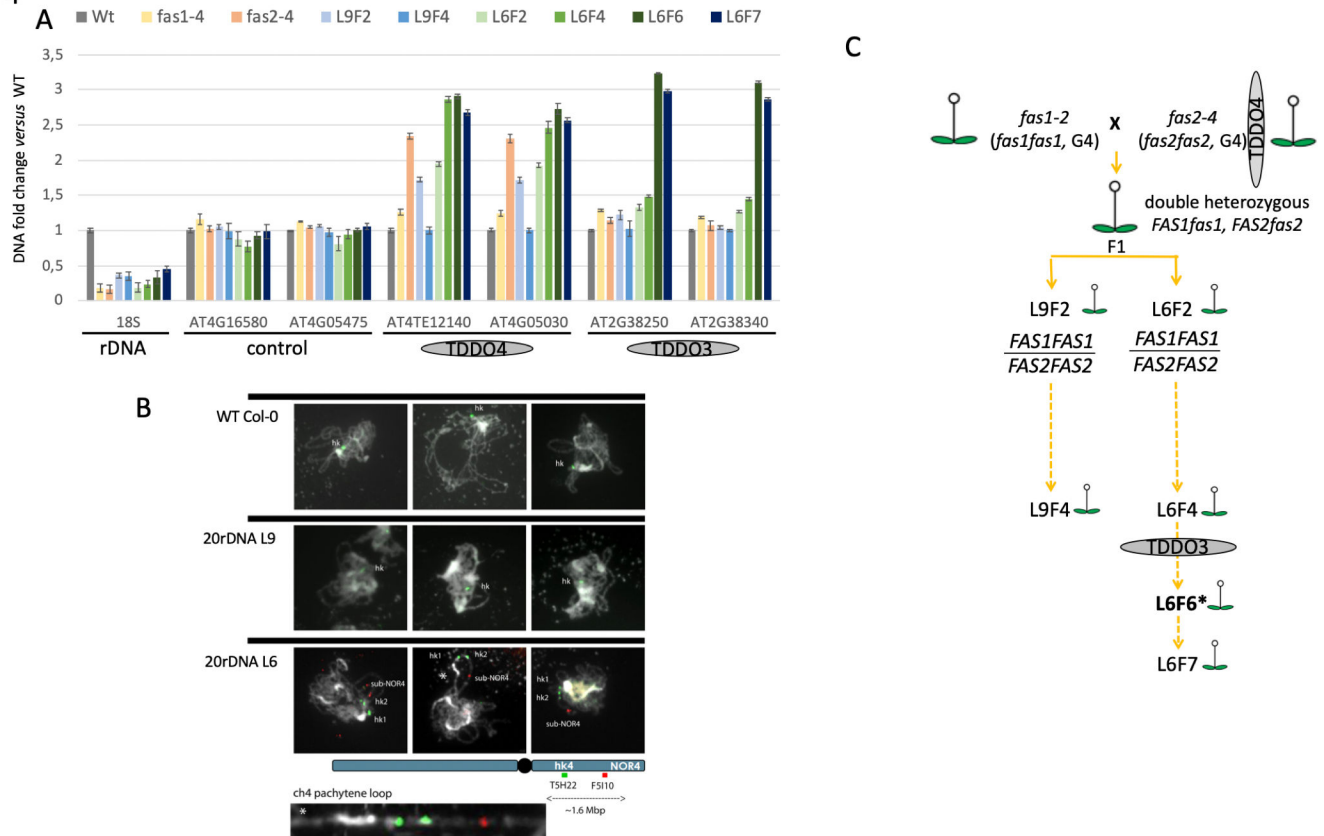


Figure 4: Apparition of the 2 major TDDO events.

A. CNV of genes present or not in TDDO and of rRNA genes were determined by qPCR. Their relative enrichment was determined in WT Col-0, the parent lines *fas1-4* and *fas2-4* and in 2 of the siblings line 6 and 9 during several generations. Relation between the lines are shown in (C). RRNA genes CNVs was followed *via* probes amplifying the 18S, the *loci* AT4G05475 and AT4G16580 that are not duplicated are controls, while the *loci* AT4TE12140 / AT4G05030 and AT2G38250 / AT2G38340 allow the identification of the 2 largest duplication TDDO4 and TDDO2 respectively. **B.** DNA-FISH analyses of 2 *loci* present on kr4s, distant by 1.6 Mb: one present on the TDDO4 (BAC T5H22 – green) and one absent of the duplicated region (BAC F5I10 – red). Three Pachytene chromosomes from WT Col-0, 20rDNA Line 6 and 20rDNA Line 9 are shown. A zoom of the unlooped kr4s from the middle panel of the 20rDNA line6 is presented at the bottom of the panel. **C.** Schematic representation of the obtaining of the 20rDNA line 6 and 9 and its relative. The putative appearance of the TDDO3 and TDDO4 events are shown.

Figure 5

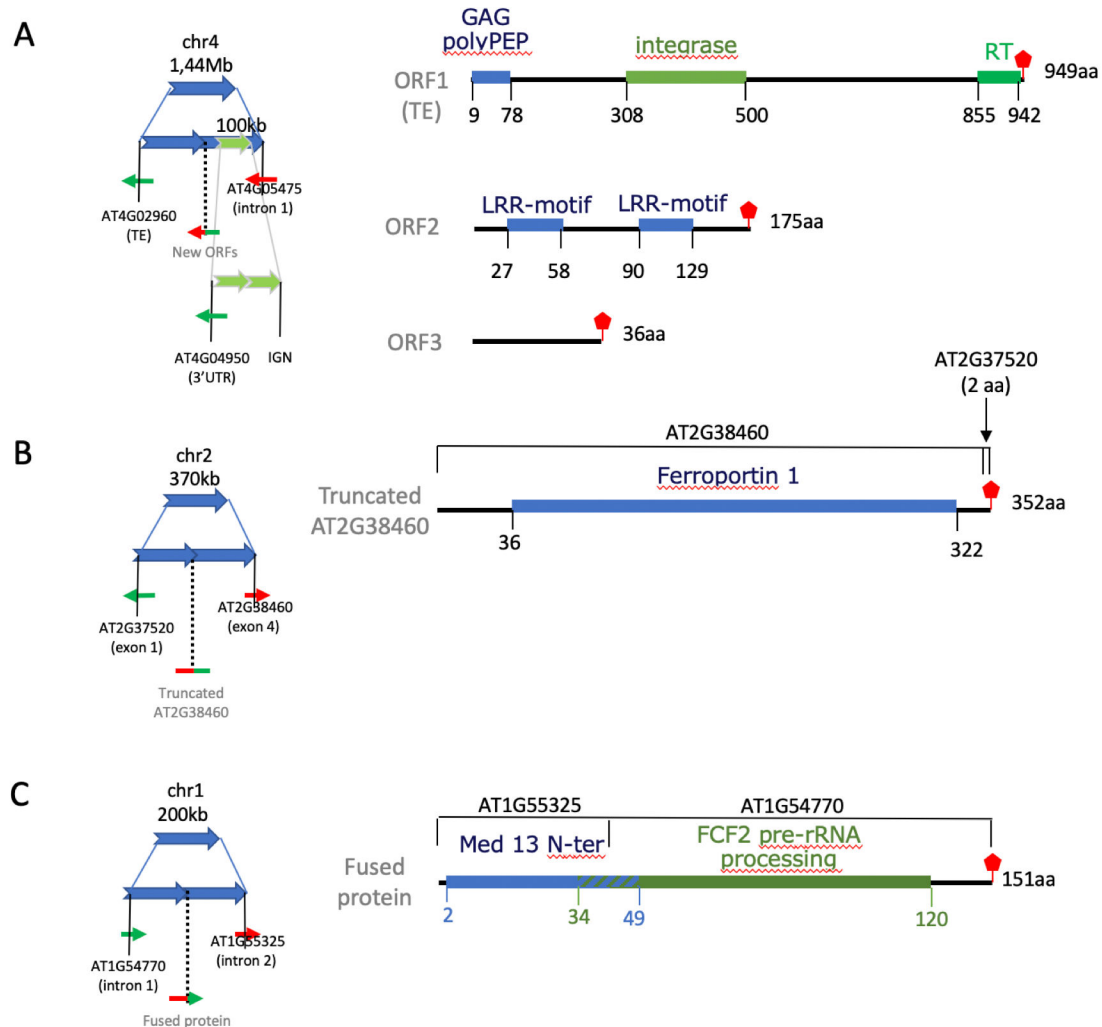
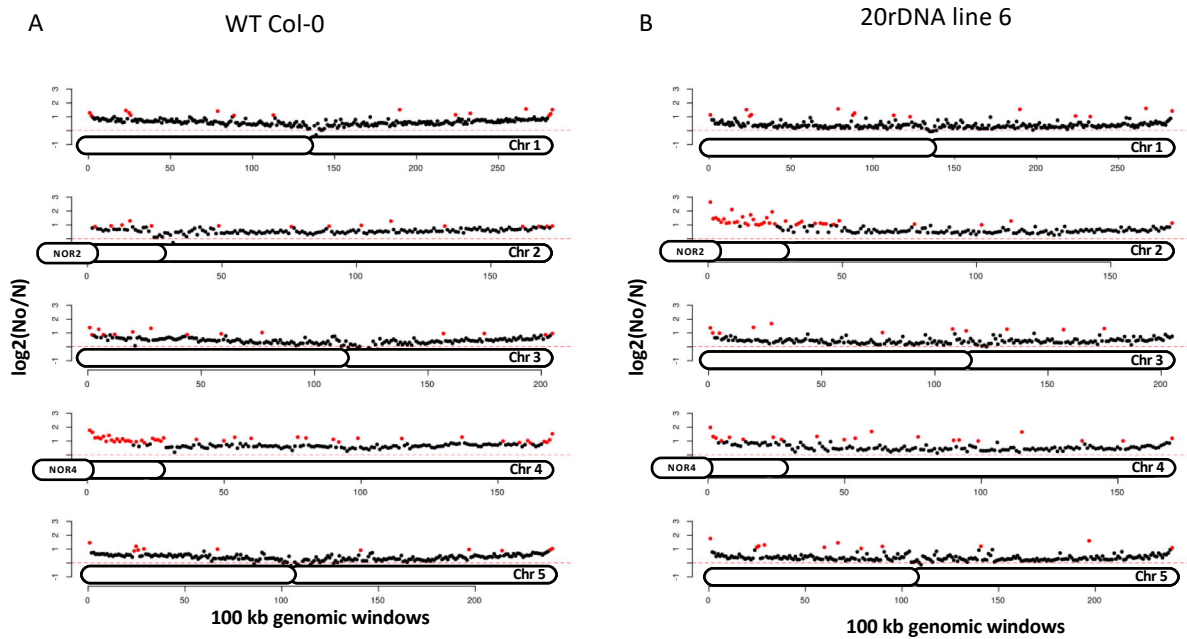


Figure 5: TDDO provoke chimeric genes formation

A-C. Schematic representation of the TDDO 1, 3 and 4 that provoked TE and/or gene fusion in the 20rDNA line 6. Genes or TEs present in the breaking junction (and their orientation are shown). Potential fused protein sequences are displayed with the putative nature and their domains. **A.** Three ORFs can be generated from the fusion of large 5' sequence of a TE (AT4G02960) with the second and last exon of gene AT4G05475. The ORF2 could potentially fused a small TE coding sequence of two Leucine Rich Repeats of AT4G05475 (LRR). **B.** The chimeric gene created between AT2G37520 and AT2G38460 potentially encode for a truncated FERROPORTIN protein. **C.** The TDDO1 provoke the fusion of the first exon of AT1G55325 that encode for an ATPase motif of MEDIATOR 13 and the last four exons of AT1G54770 that contains an RNA processing domain.

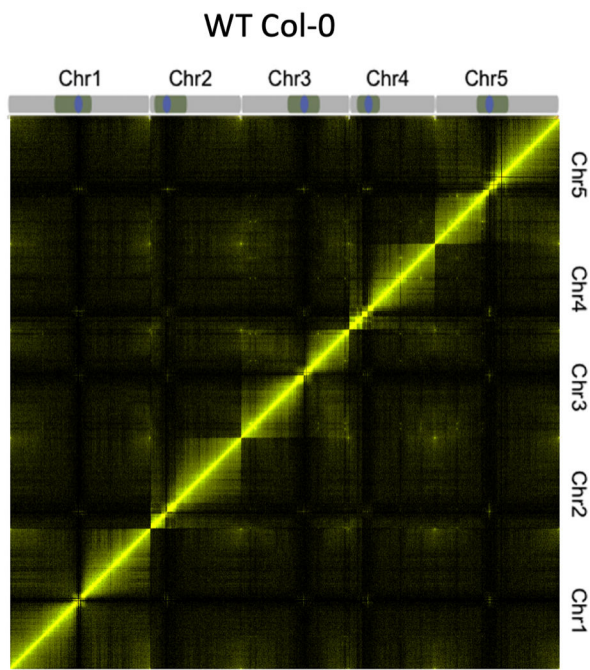
Suppl. Figure 1

**Supplemental Figure 1:**

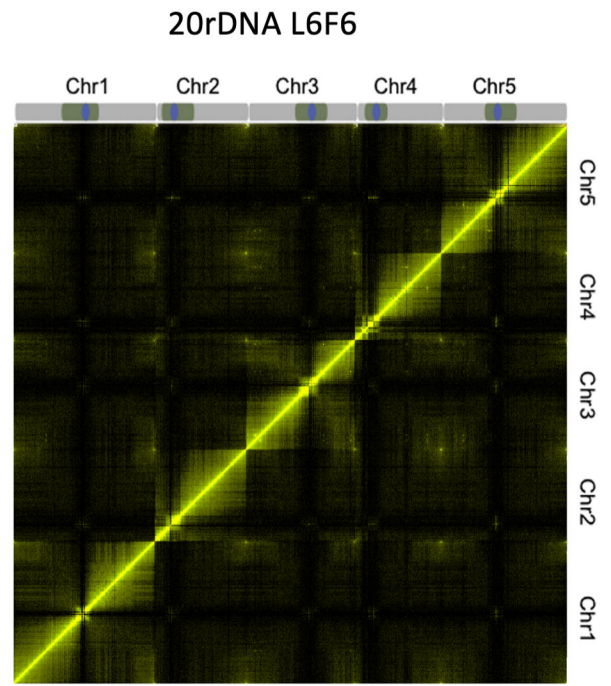
A-B. Chromosome plots displaying the relative enrichment of a given genomic segment with the nucleolus in WT Col-0 (A) and in the 20rDNA L6F6 (B). The y-axis displays the fold change between the nuclear *versus* the nucleolar DNA. Each dot represents a 100kb window. Nucleolus-enriched genomic regions above the threshold (red-dotted line) are colored in red.

Suppl. Figure 2

A

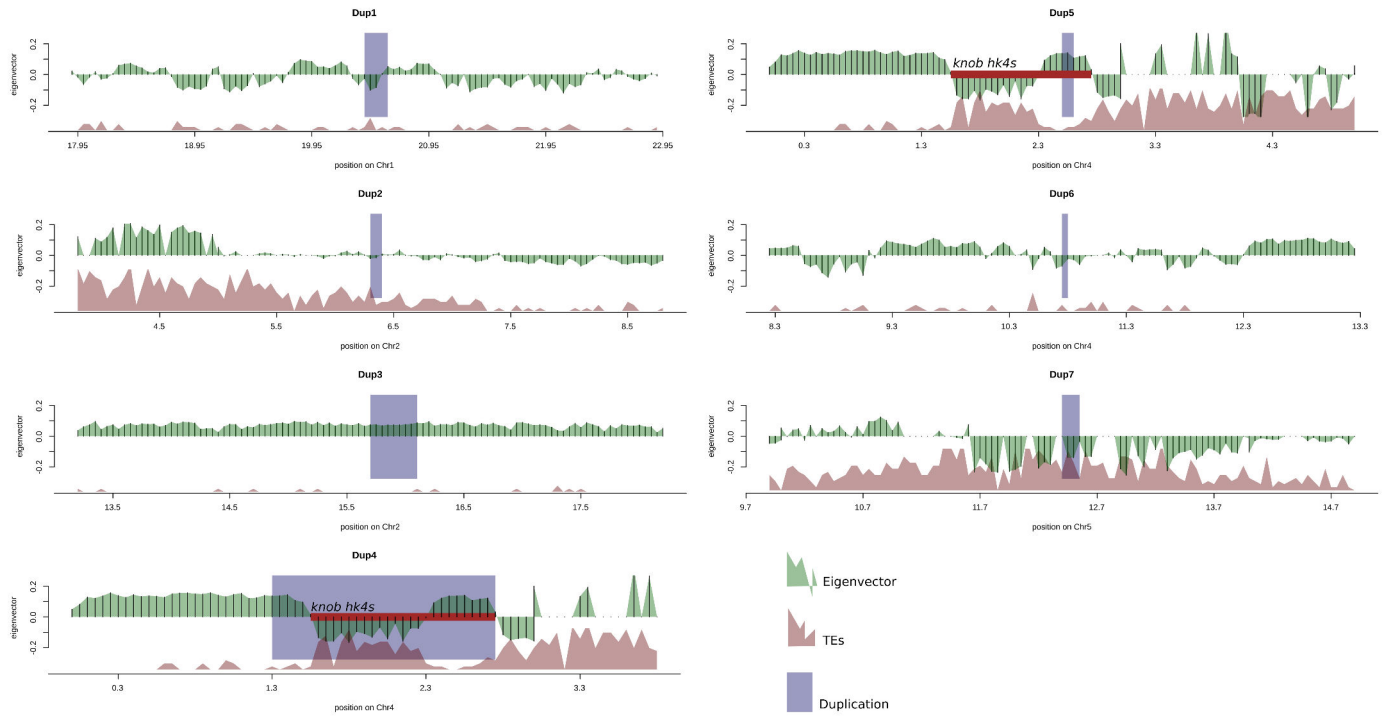


B

**Supplemental Figure 2:**

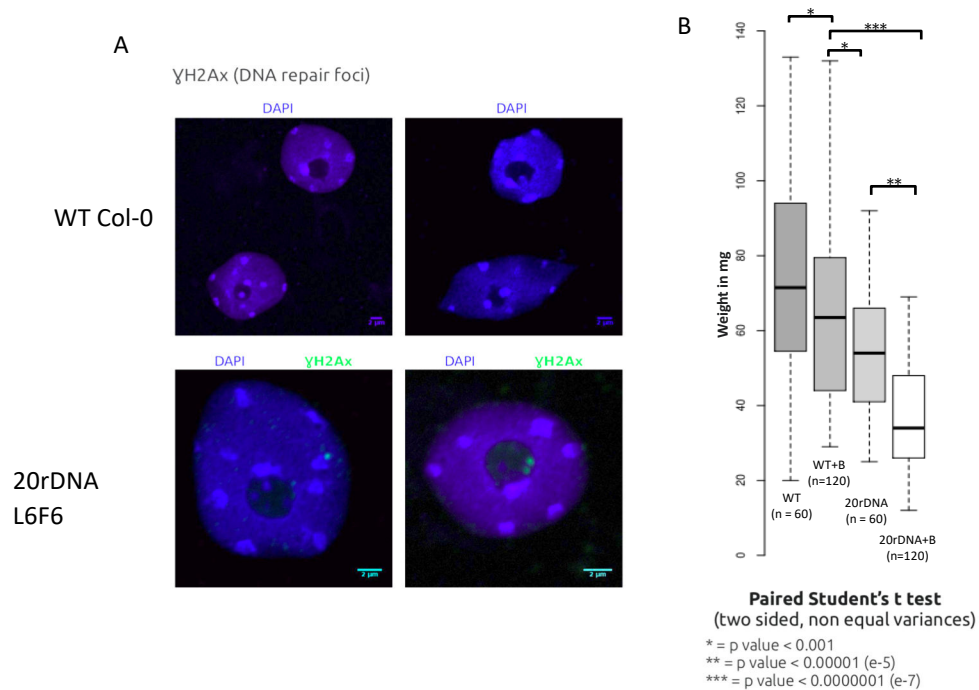
A-B. HiC matrix displaying the chromatin-chromatin interactions identified by the chromosome capture conformation technique in wild-type Col-0 (A) and the 20rDNA L6F6 (B). Enriched interactions are displayed in yellow.

Suppl. Figure 3

**Supplemental Figure 3:**

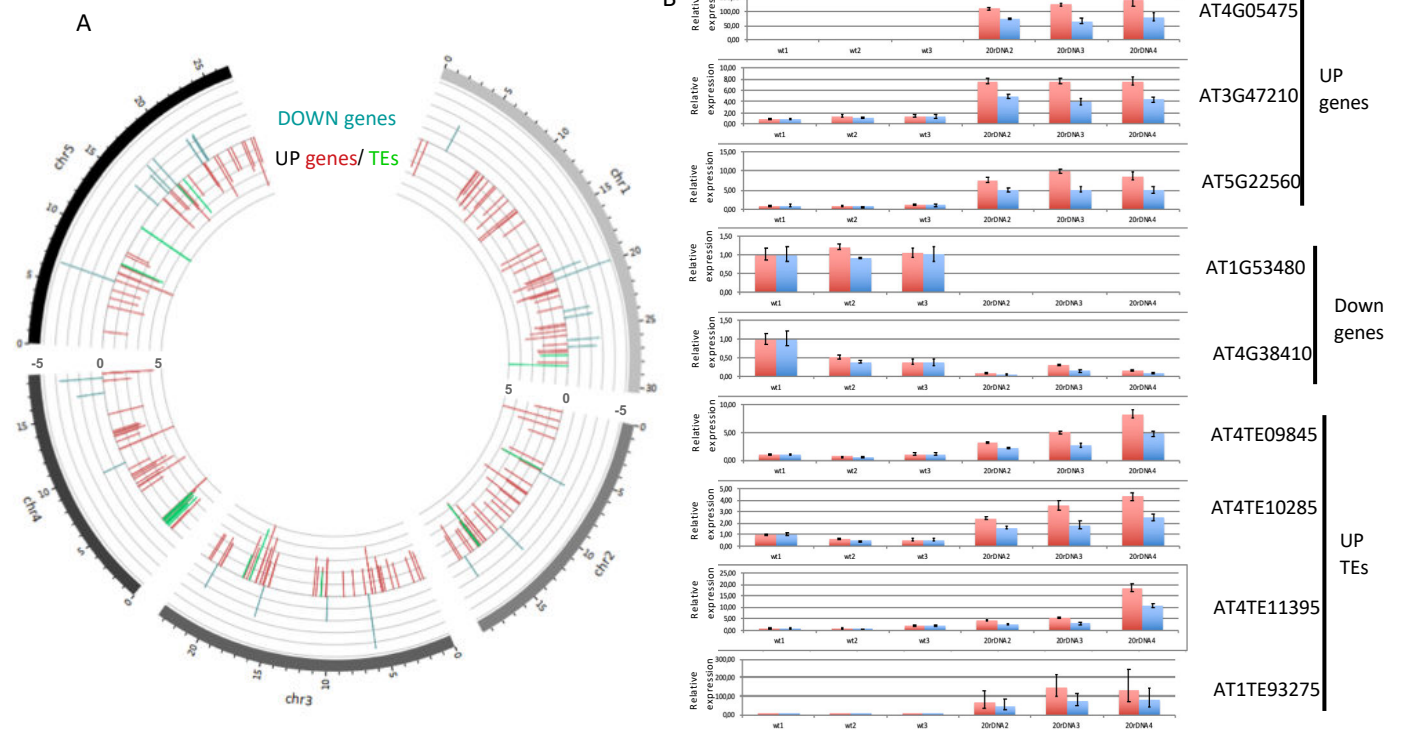
Eigenvector of the Col-0 wild-type Hi-C data set and annotation of the TDDO.

Supp. Figure 4

**Supplemental Figure 4:**

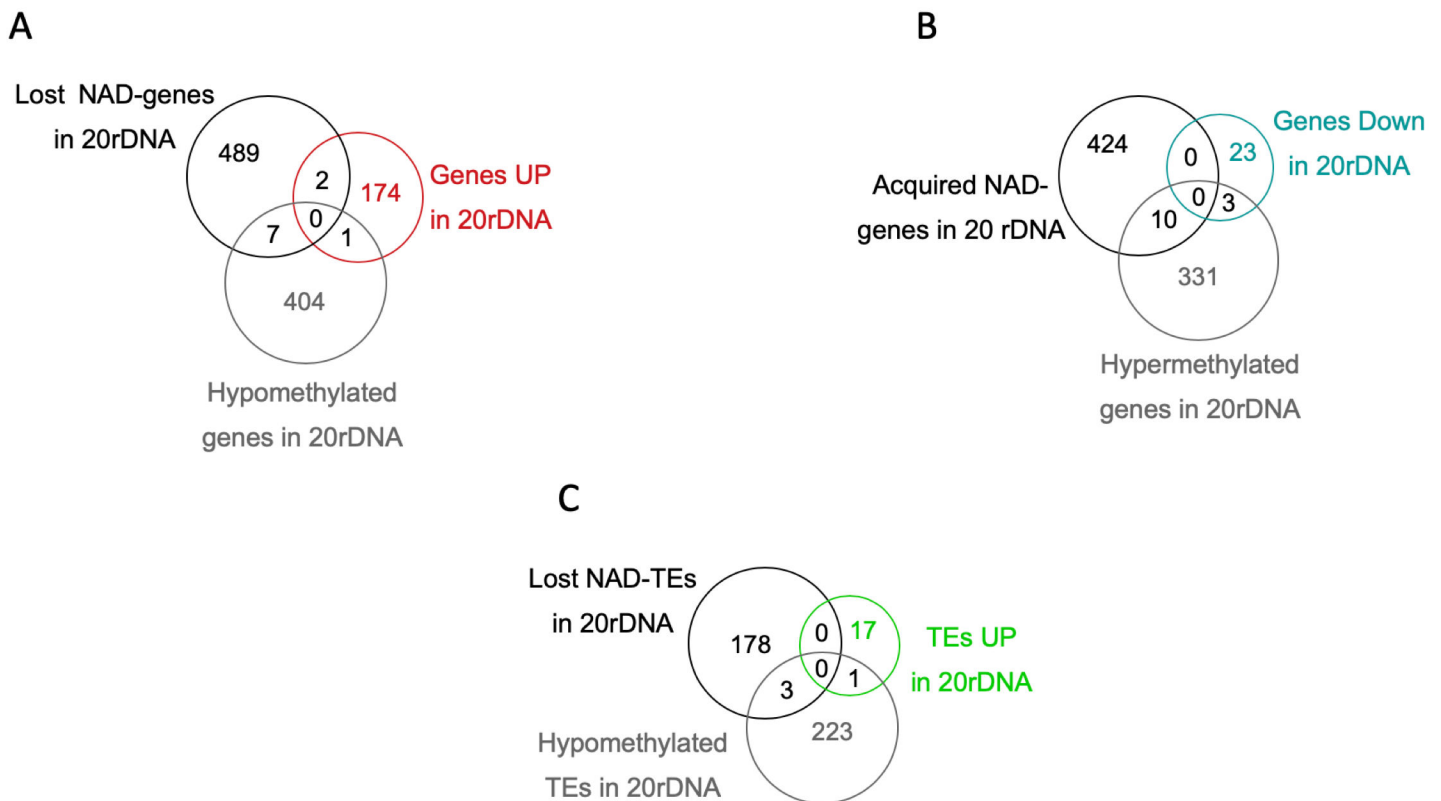
A. Representative confocal images showing the immunolocalization of the P- γ H2Ax foci (green) in WT Col-0 (top panels) or in 20rDNA F6L6 (bottom panels) isolated nuclei. The Blue staining correspond to the DNA labeled with DAPI. **B.** WT Col-0 and 20rDNA L6F6 seeds were grown for 18 days on an MS plate with or without Bleomycine at 10^{-5} M. Individual plant growth was then measured using their fresh weights.

Suppl. Figure 5

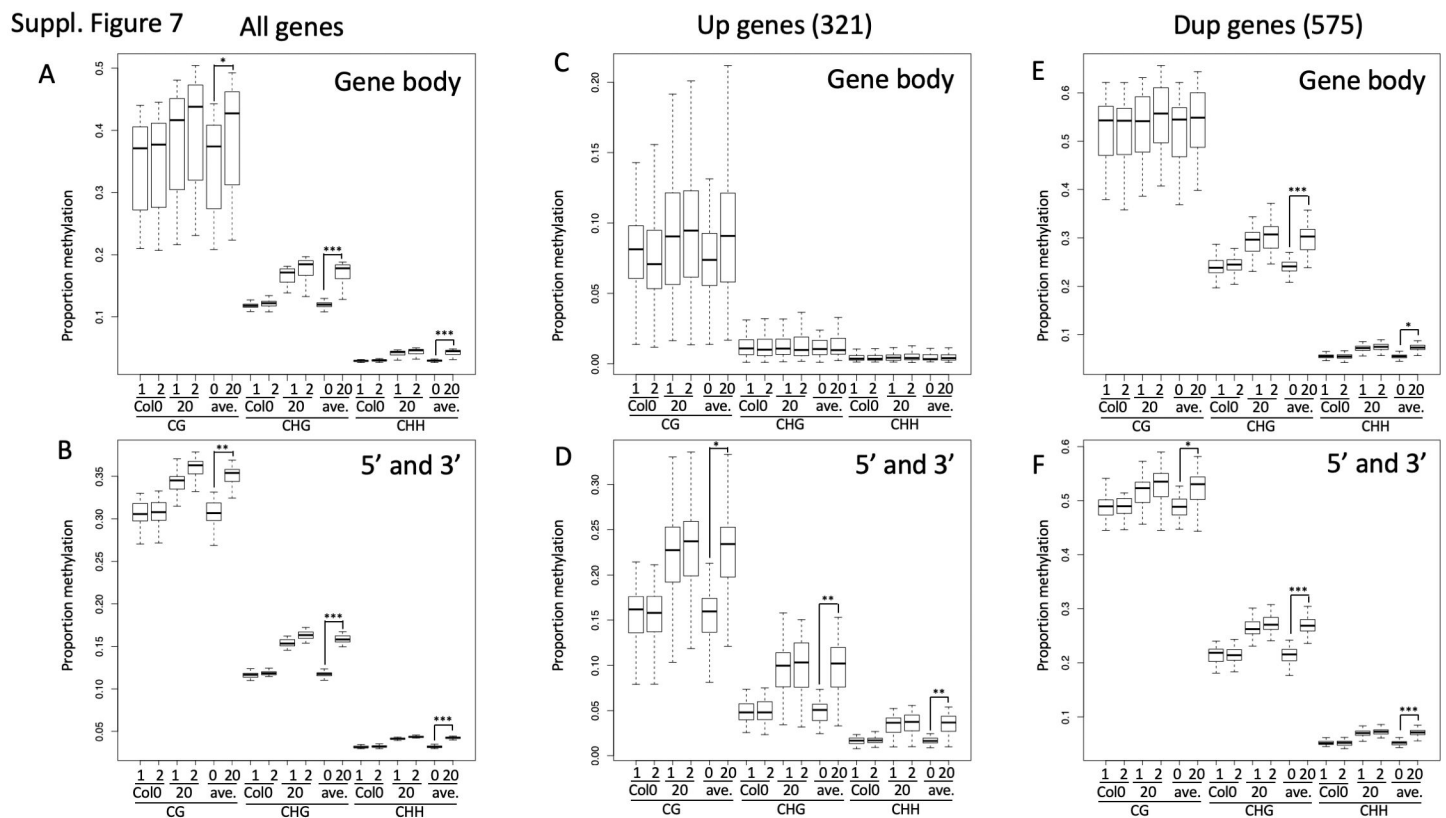
**Supplemental Figure 5:**

A. Representation of the chromosomal position of genes and TEs of differentially accumulated transcripts in the 20rDNA L6F6 *versus* WT Col-0 using *Circos* (72). Up-regulated TEs are in green, Up-regulated genes are in red and down-regulated genes are in blue. **B.** QPCR analyses of differentially accumulated genes. Their relative accumulation was analyzed against the *ACTIN7* (red bars) or the *HC7* (blue bars) household genes.

Suppl. Figure 6

**Supplemental Figure 6:**

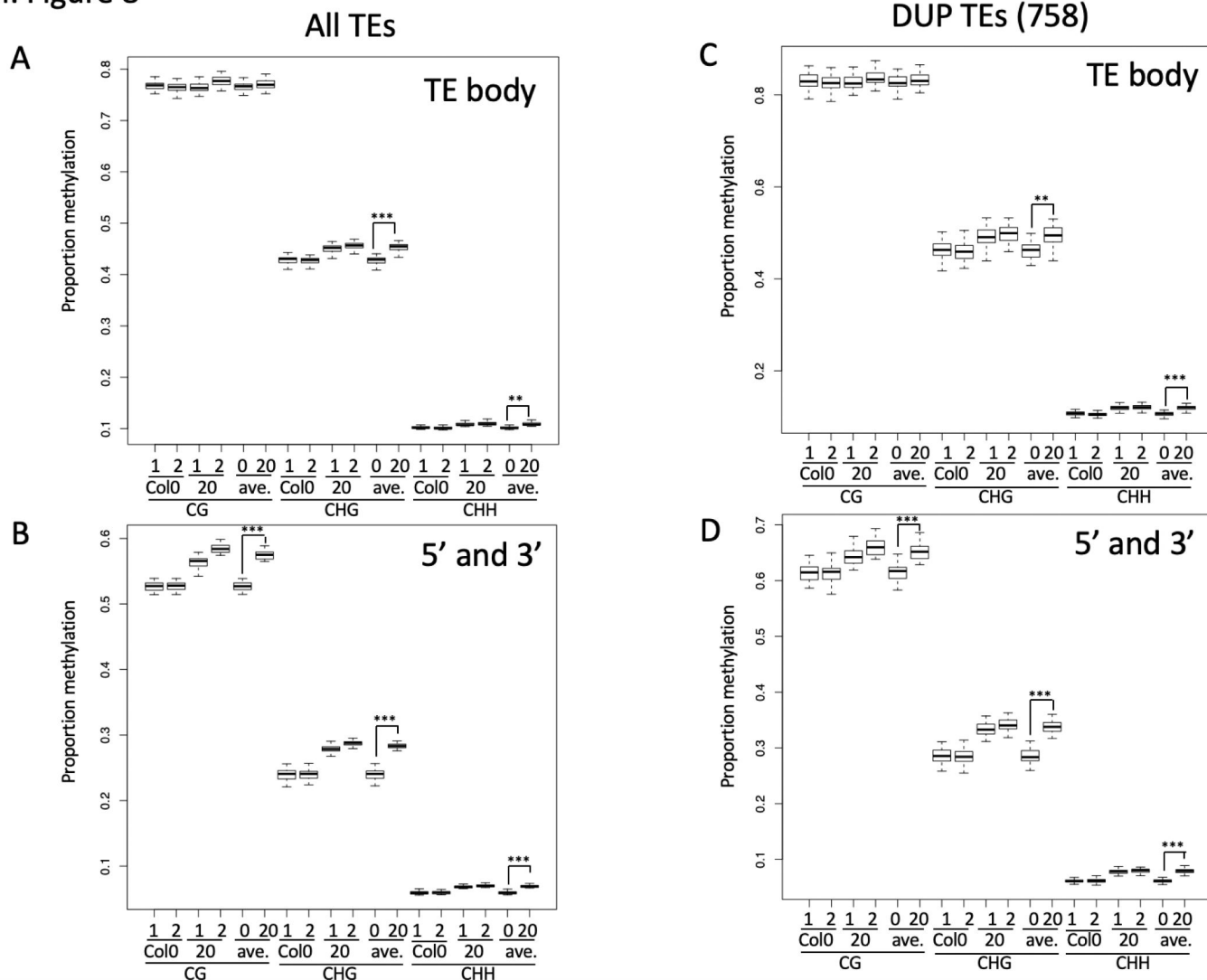
A. Venn diagram showing the relation between genes that lost their association with the nucleolus in the 20rDNA L6F6 (lost NAD-genes in 20rDNA), Genes that are upregulated in 20rDNA L6F6 with an adjusted p -value < 0.01 and a $\log_2\text{FoldChange} > 2$ (Genes UP in 20rDNA) and with hypomethylated genes in 20rDNA L6F6. **B.** Venn diagram showing the relation between genes that gain association with the nucleolus in the 20rDNA L6F6 (Acquired NAD-genes in 20rDNA), Genes that are down-regulated in 20rDNA L6F6 with an adjusted p -value < 0.01 and a $\log_2\text{FoldChange} > 2$ (Genes Down in 20rDNA) and with hypermethylated genes in 20rDNA L6F6. **C.** Venn diagram showing the relation between TEs that lost their association with the nucleolus in the 20rDNA L6F6 (lost NAD-TEs in 20rDNA), TEs that are upregulated in 20rDNA L6F6 with an adjusted p -value < 0.01 and a $\log_2\text{FoldChange} > 2$ (TEs UP in 20rDNA) and with hypomethylated TEs in 20rDNA L6F6.



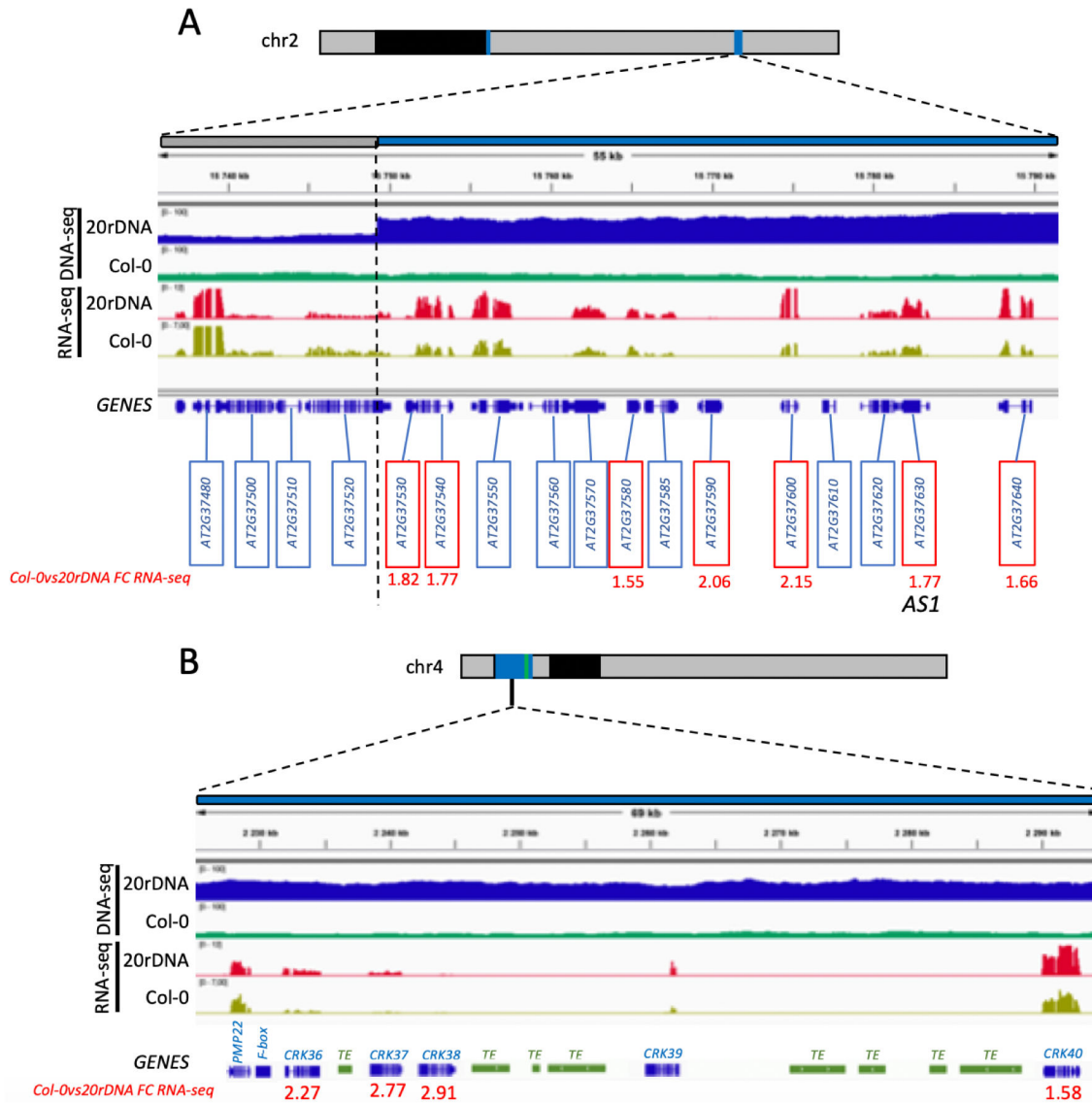
Supplemental Figure 7:

A-F. Dot plots showing the relative methylation levels in CG, CHG and CHH contexts in WT Col-0 or in 20rDNA L6F6. The analysis was performed by genome-wide bisulfite sequencing. 2 WT Col-0 and 2 20rDNA L6F6 (labelled 20) replicates (1 and 2) are shown. The average methylation between both replicates is shown (labelled ave.). The analysis was performed with All genes (A-B), with Upregulated genes with an adjusted p -value < 0.01 and a $\log_2\text{FoldChange} > 1.5$ (C-D) and with the Duplicated genes (E-F). For this analysis, we separated Gene body sequences (top panels A, C and E) from the flanking sequences of the genes (5' and 3' – bottom panels B, D and F). *($p < 0.00025$); **($p < 5.10^{-10}$); ***($p < 2.10^{-16}$) calculated using a Wilcoxon test.

Suppl. Figure 8

**Supplemental Figure 8:**

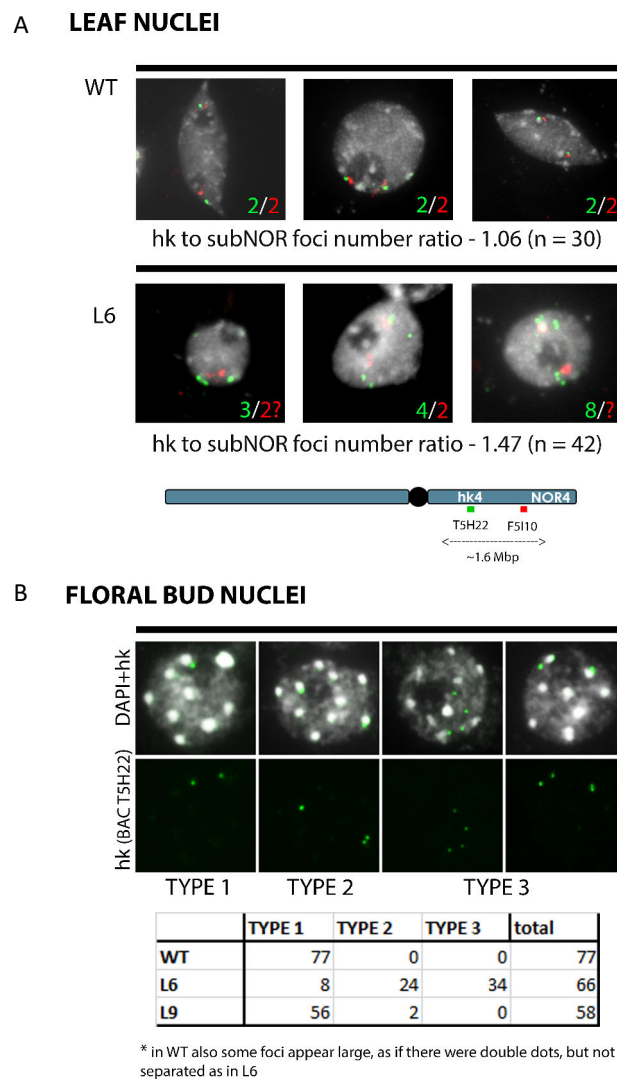
A-D. Dot plots showing the relative methylation levels in CG, CHG and CHH contexts in WT Col-0 or in 20rDNA L6F6. The analysis was performed by genome-wide bisulfite sequencing. 2 WT Col-0 and 2 20rDNA L6F6 (labelled 20) replicates (1 and 2) are shown. The average methylation between both replicates is shown (labelled ave.). The analysis was performed with All TEs (A-B) and with the Duplicated TEs (C-D). For this analysis, we separated TE body sequences (top panels A and C) from the flanking sequences of the TEs (5' and 3' – bottom panels B and D). *($p < 0.00025$); **($p < 5.10^{-10}$); ***($p < 2.10^{-16}$) calculated using a Wilcoxon test.



Supplemental Figure 9:

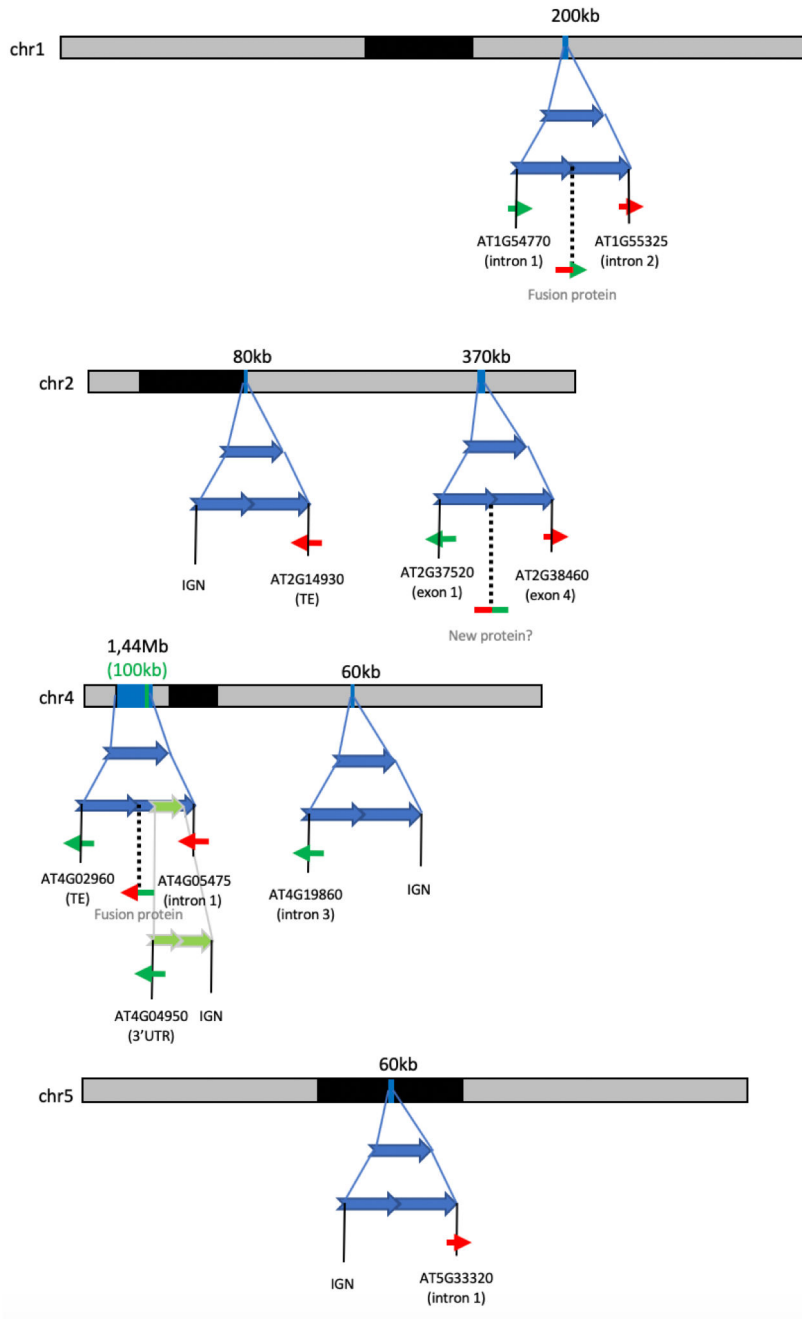
A-B. Genome browser (IGV - (73)) screenshot showing examples of Up-genes in 20rDNA L6F6 in duplicated regions. DNA coverage was obtained by nanopore sequencing for WT Col-0 (green) and for 20rDNA L6F6 (blue). Reads from RNA sequencing are shown for WT Col-0 (red) and for 20rDNA L6F6 (yellow). On the bottom, genes location is represented in blue and TEs in green. Gene names are surrounded in blue (non-significant) and in red (significant) when they are differentially accumulated transcripts with an adjusted p -value < 0.01 and a \log_2 FoldChange > 1.5 . This value is noted in red. Two locations are presented: the left border of the TDDO3 (A) and portion of the TDDO4 (B).

Suppl. Figure 10



Supplemental Figure 10:

A-B. DNA-FISH analyses of a portion of the TDDO4 in leaf nuclei. **A.** Two *loci* present on kr4s, distant by 1.6 Mb: one present on the TDDO4 (BAC T5H22 – green) and one absent of the duplicated region (BAC F5I10 – red). Three leaf nuclei from WT Col-0 and 20rDNA Line 6 are shown and the ratio of green versus red signals as well as its ratio noted in the figure. **B.** In that case, the TDDO4 locus (BAC T5H22 – green) was analyzed in floral bud nuclei and four representatives images are presented in the figure. DNA is labelled with DAPI stain (Top panel, with DAPI in grey). Between 58 and 77 nuclei per genotype (WT Col-0, rDNA20 L6 and L9) and the distribution of each nuclei found per genotype is presented in the table.

**Supplemental Figure 11:**

Graphical representation of the 7 TDDO along the 5 chromosomes of *A. thaliana*. The genomic context of the breaking junctions is presented for each TDDO.

Discussion générale

Mes travaux de thèse avaient deux objectifs majeurs : i) étudier la dynamique des domaines chromatiniens associés au nucléole d'*Arabidopsis thaliana*; ii) révéler l'importance des copies inactives des gènes d'ARNr 45S sur l'organisation et l'intégrité du génome de cette espèce.

Pour ce faire, j'ai d'abord optimisé la méthode d'identification des NADs et l'ai adapté au FACS BioRad S3TM du laboratoire. J'ai également créé des nouveaux outils pour l'étude des NADs comme la lignée transgénique 20%ADNr-pEG ::FIB2 ::YFP.

Chez *A. thaliana* Col-0, seul un des deux NORs est actif. Mais au sein de ce NOR actif, seule une portion des gènes d'ARNr 45S sont activement transcrits et forment le nucléole. Ainsi, les gènes d'ARNr 45S inactifs du NOR actif sont très condensés et localisés à la périphérie du nucléole. Ils participent à la formation d'une couche d'hétérochromatine constitutive autour du nucléole, à laquelle s'ajoutent les régions centromériques et péri-centromériques adjacentes au NOR actif. Dans le cas du NOR4, le *knob* *hk4S* est également associé.

Les régions subtélomériques des 5 chromosomes sont également associées au nucléole. Puisque la chromatine périnucléolaire est enrichie en marques de type répressives et que les gènes présents sont peu ou pas transcrits, le premier axe de ma thèse a été d'identifier les NADs dans différentes conditions. A terme, nous voulions vérifier l'existence d'une possible voie de régulation transcriptionnelle par séquestration nucléolaire.

L'identification des NADs en condition de choc thermique et suite à la transition florale a permis de démontrer que la composition des NADs était stable. Cependant, aucune voie de régulation transcriptionnelle par relocalisation de certains gènes autour du nucléole ou vice-versa, n'a pu être démontrée.

Cependant, nous avons constaté une grande hétérogénéité quant à l'identification des NAD-gènes, même entre les conditions contrôles. La profondeur de séquençage de ce genre d'approches est un élément clé qui détermine la résolution de l'analyse. Plus elle sera grande, plus la détection des NADs sera précise et fine. Avec une faible profondeur de séquençage, comme lors de ma thèse, les NADs ne peuvent être déterminés qu'avec une résolution comprise entre 50 et 100 kb. Dans PONTVIANNE *et al.* 2016b, les NAD-gènes avaient été identifiés avec une plus grande profondeur de séquençage. Ainsi, le chevauchement entre les réplicats était meilleur.

Finalement, l'ensemble des résultats obtenus concernant la dynamique des NADs indiquent que les régions associées avec le nucléole sont des régions très spécifiques du génome, surtout déterminées par la proximité avec les télomères et le NOR actif.

L'importance de la proximité avec un NOR pour l'association avec le nucléole est un constat qui est vrai chez les autres organismes étudiés. Cette observation est logique,

puisque le nucléole se forme autour des NORs. C'est pourquoi la proximité avec un NOR mais également son activité, vont déterminer si les régions adjacentes s'associent avec le nucléole.

L'identification des NADs dans la lignée 20%rDNA-pEG ::YFP ::FIB2, où l'ensemble du bras court du chromosome 2 et son chromocentre font partie des NADs, est un autre exemple confirmant cet effet de l'activité des NORs. En effet, cette observation avait déjà été faite dans le mutant *nuc1* où les deux NORs sont actifs (PONTVIANNE *et al.* 2007; PONTVIANNE *et al.* 2016b).

Lorsqu'on analyse la composition des bras courts des chromosomes acrocentriques 2 et 4, on s'aperçoit que 7,5% des gènes qu'ils contiennent sont des pseudogènes. Pourtant, sur les autres bras des chromosomes ils ne représentent que 2% des gènes. Concrètement, parmi les 915 pseudogènes annotés chez *A. thaliana*, 72 (7,9%) sont présents dans les NADs. Ceci démontre la spécificité de ces régions chromosomiques, appauvries en gènes fonctionnels.

Chez la majorité des espèces étudiées (plus de 75% des caryotypes), les NORs sont également localisés en position terminale, proches des télomères ou en position péri-centromérique (GARCIA *et al.* 2014; GARCIA *et al.* 2017; SOCHOROVÁ, 2018). La position péri-centromérique des NORs est moins récurrente, surtout si l'espèce possède un seul NOR par génome haploïde. Dans tous les cas, la matrice d'ADNr se trouve proche de régions d'hétérochromatine constitutive, enrichies en séquences répétées et appauvries en gènes (GARCIA *et al.* 2014; GARCIA *et al.* 2017; SOCHOROVÁ, 2018).

Il semblerait donc qu'au niveau évolutif, les NORs occupent des positions préférentiellement distinctes du reste des chromosomes et des gènes transcrits par l'ARN Pol II. Plusieurs hypothèses permettent d'expliquer cette convergence évolutive.

Premièrement, il est probable que le positionnement de l'ADNr en position terminale et/ou proche de régions non transcrites, soit important pour la formation des structures nucléolaires. En effet, le nucléole est une agglomération de centaines de facteurs protéiques et d'ARNs sur des gènes d'ARNr actifs. L'ensemble de ces molécules possèdent des caractéristiques physiques qui induisent la formation d'un compartiment «phase-séparé», non miscible avec le nucléoplasme (BRANGWYNNE *et al.* 2011; FERIC *et al.* 2016). Si le NOR était entouré de domaines euchromatiques, cela pourrait compliquer la formation du nucléole. Ou, à l'inverse, la formation du nucléole pourrait interférer avec l'activité transcriptionnelle des domaines euchromatiques. Ainsi, la chromatine très condensée et appauvrie en gènes entourant les NORs mais aussi le nucléole (les NADs), sert peut être de barrière physique entre le nucléole et le reste du génome. Ces structures sont d'ailleurs appelées «Nucleolar Shell» du fait de leur densité (GUETG et SANTORO, 2012b).

Deuxièmement, ceci peut être dû à la nature très instable des gènes d'ARNr. En effet, l'ADNr est une des régions les plus sensibles à des agents endommageant de l'ADN, qu'ils soient exogènes (*e.g.* : bléomycine) ou endogènes (*e.g.* : transcription). Par conséquent, à cause de sa nature répétée, des erreurs de réparation de l'ADNr sont récurrents. Ces erreurs induisent des délétions ou amplifications de copies de gènes d'ARNr au sein du NOR (*cf.* : Introduction générale). Par exemple, dans les cellules cancéreuses, caractérisées par une grande instabilité génétique, un des loci les plus affectés est l'ADNr qui présente beaucoup d'événements d'expansion ou contraction des NORs par rapport aux cellules saines du même individu (VALORI *et al.* 2019; WANG et LEMOS, 2017; XU *et al.* 2017).

Une des causes principales de l'instabilité de l'ADNr est la forte activité transcriptionnelle au sein des NORs. En effet, lorsque l'activité transcriptionnelle est augmentée de façon artificielle, on constate une instabilité accrue de l'ADNr (ALDRICH et MAGGERT, 2015; PENG et KARPEN, 2007). La transcription entraîne la formation de structures secondaires de l'ADN, un état ouvert de la chromatine et des possibles collisions avec la machinerie de réplication de l'ADN. Tous ces facteurs peuvent induire des DSBs. Les copies d'ADNr inactives ainsi que l'hétérochromatine entourant les NORs pourraient ainsi atténuer le stress transcriptionnel du bras chromosomique en question.

D'un autre côté, il est avéré dans les organismes étudiés, que les copies de gènes d'ARNr inactives et condensées participent à la stabilité des copies actives. Ainsi, chez la levure, des lignées présentant un nombre réduit de copies sont plus sensibles au méthanesulfonate de méthyle (MMS) ou à la lumière ultraviolette (UV). Par ailleurs, des cellules souches hématopoïétiques de souris qui ont un nombre réduit de copies de gènes d'ARNr, sont également plus sensibles aux agents endommageant de l'ADN tels que la bléomycine (XU *et al.* 2017).

Cette instabilité peut être expliquée par deux mécanismes chez la levure (IDE *et al.* 2010; KOBAYASHI, 2011, 2014).

Premièrement, des copies inactives sont intercalées avec les copies actives et permettent de réduire le stress transcriptionnel au niveau de la partie active du NOR. Ainsi, la chromatine est moins encombrée par la machinerie de transcription de l'ARN Pol I et présente une conformation moins «ouverte». Ceci réduit les possibilités de cassure de cette dernière.

Deuxièmement, les copies inactives facilitent la réparation des DSBs au niveau des NORs. En effet, l'ADNr est surtout réparé par recombinaison homologue (MARNEF *et al.* 2019; McSTAY, 2016; SLUIS et McSTAY, 2015; TSANG et CARR, 2008). Parmi les protéines impliquées dans ce processus, on retrouve des complexes protéiques formés par des protéines STRUCTURAL MAINTENANCE OF CHROMOSOMES (SMC) tels que la cohésine ou la condensine. Ces complexes, sont essentiels à la cohésion des chromatides pour la ré-

paration des chromosomes par recombinaison homologue (WATANABE *et al.* 2009). Les copies inactives servent de point d'ancrage de ces protéines, car les copies actives sont occupées par la machinerie de transcription (IDE *et al.* 2010).

Dans le cas de la levure, c'est la condensine qui est impliquée dans la cohésion des chromatides lors de la recombinaison homologue au niveau de l'ADNr (IDE *et al.* 2010). Dans des lignées artificielles présentant un nombre très réduit de copies, toutes deviennent actives pour maintenir le niveau de production d'ARNr. Par conséquent, la chromatine devient plus sensible aux agents endommageant de l'ADN et il y a une augmentation des DSBs. De plus, la condensine a moins d'espace pour se fixer et les DSBs sont plus difficiles à réparer (IDE *et al.* 2010).

Grâce à la lignée 20%ADNr, on a pu mettre en évidence que chez *A. thaliana*, ces extra copies sont également importantes pour la stabilité de l'ADNr. En effet, comme chez la levure, la souris ou la drosophile, la lignée 20%ADNr présente une croissance normale et on constate le même niveau de production d'ARNr. Ces copies supplémentaires ne sont donc pas nécessaires aux fonctions nucléolaires de base.

Cependant, en conditions normales de croissance, les noyaux de cette lignée présentent plus de DSBs que chez l'écotype sauvage. Et la majorité de ces signaux (plus de 60%) sont localisés au niveau du nucléole, c'est à dire au niveau des gènes d'ARN 45S actifs (données non montrées).

Comme chez la levure, il se peut que l'augmentation du nombre de DSBs dans la lignée 20%ADNr soit due à la fois à une augmentation du stress transcriptionnel et à une réduction de l'efficacité de réparation. Chez les plantes, les condensines sont aussi impliquées dans la recombinaison homologue. Plus particulièrement la condensine II, qui est impliquée dans la réparation de l'ADN durant l'interphase et est localisée au niveau du nucléole chez *A. thaliana* (SAKAMOTO *et al.* 2019). En outre, l'analyse d'expression différentielle entre l'écotype sauvage et la lignée 20%ADNr, n'a pas mis en évidence l'inhibition de gènes impliqués dans la réparation de l'ADN qui pourraient expliquer cette sensibilité.

L'instabilité accrue de l'ADNr causée par la réduction du nombre de copies, entraîne une réduction de la durée de vie des levures (IDE *et al.* 2010). Dans des cellules humaines, un lien entre l'instabilité de l'ADNr et la sénescence cellulaire a également été mis en évidence (PAREDES *et al.* 2018). Dans cette étude, les auteurs se sont intéressés au rôle de l'histone déacétylase SIRTOUINE7 (SIRT7) qui se lie aux séquences régulatrices des ADNr et participe à leur régulation, dans le vieillissement cellulaire (PAREDES *et al.* 2018). En cas d'inhibition de la protéine SIRT7, les auteurs ont observé une augmentation substantielle du nombre de cellules sénescents. De plus, ils ont pu démontrer que la sénescence observée était plus précoce et indépendante de la sénescence induite par le

raccourcissement des télomères.

Ces cellules présentaient également une instabilité accrue de l'ADNr (augmentation des niveaux de $\gamma H2AX$) ainsi qu'une réduction de jusqu'à 50% du nombre de copies d'ADNr. Afin de mettre en évidence un lien direct entre l'instabilité de l'ADNr et la sénescence cellulaire, ils ont utilisé la technologie CRISPR-Cas9 pour générer des DSBs spécifiquement au niveau des séquences d'ADNr. Grâce à cette expérience, un lien direct entre l'instabilité de l'ADNr et la sénescence précoce des cellules a pu être établi (PAREDES *et al.* 2018).

Dans notre cas, la croissance des plantes de la lignée 20%ADNr est beaucoup plus affectée par l'exposition à la bléomycine. Par contre, nous n'avons pas observé de signes de sénescence précoce dans cette lignée.

Dans les populations naturelles, l'ADNr reste très instable, même lorsqu'il contient de nombreuses copies inactives. Le grand nombre de DSBs et les NAHR fréquentes au niveau de ces *loci*, induisent des variations du nombre de copies d'ADNr mais également des changements structuraux, surtout des duplications, au niveau des séquences adjacentes (McSTAY, 2016). En outre, le fait que les NORs soient souvent sur des bras courts de chromosomes acrocentriques (GARCIA *et al.* 2014; GARCIA *et al.* 2017; SOCHOROVÁ, 2018) est peut-être une conséquence d'événements de recombinaison homologue non allélique. En effet, il se peut que ces événements aient induit la délétion de séquences adjacentes aux NORs au cours du temps. Ces délétions auraient été sélectionnées positivement pour séparer les NORs du reste du chromosome et ne pas impacter négativement les gènes adjacents.

Pour résumer, au cours de l'évolution, il est très possible que la localisation des NORs au sein de régions hétérochromatiques, ait été sélectionnée, pour servir de barrière physique entre le nucléole et le reste du génome et aider ainsi à la stabilité de l'ensemble du génome. De plus, ces copies inactives stabilisent aussi le NOR en réduisant le stress transcriptionnel et en facilitant sa réparation. Ces deux raisons, pourraient expliquer pourquoi un grand nombre de copies d'ADNr est retrouvé chez tous les eucaryotes.

Il est à noter que les gènes d'ARNr sont les plus anciens du génome et on les retrouve chez tous les organismes. Les bactéries possèdent que quelques opérons codant pour leurs trois ARNr (16S, 23S et 5S) et n'ont pas de système d'amplification génique (KOBAYASHI, 2014). Mais avec l'apparition de la pluri-cellularité, les organismes plus complexes ont eu besoin d'augmenter la production de protéines, et donc de ribosomes. Pour répondre à la demande en ARNr, il y a eu une amplification du nombre de copies des gènes d'ARNr. La grande instabilité générée par le fort taux de transcription aurait été toxique pour les cellules, en augmentant leur sensibilité aux dommages de l'ADN, comme ceux causés

par les rayons UV. Les cellules capables de maintenir un nombre élevé de copies d'ARNr par un système d'amplification génique, auraient ainsi été sélectionnées (GARCIA *et al.* 2014; GARCIA *et al.* 2017; KOBAYASHI, 2014; SOCHOROVÁ, 2018).

Ce système d'amplification génique permet de maintenir un nombre de copies entre un intervalle, spécifique à chaque espèce. Cependant, il existe une grande variabilité du nombre de copies au niveau intra- et interspécifique (*cf.* : Introduction générale). Chez l'humain (GIBBONS *et al.* 2014) et le maïs (LI *et al.* 2018), la variation du nombre de copies dans des populations naturelles a été corrélée à des différences d'expression d'une quarantaine de gènes chez le maïs et environ 200 chez l'humain. Ainsi, il semblerait que les copies inactives participent à la formation d'hétérochromatine constitutive ayant un impact sur l'ensemble du génome et l'expression des gènes (PAREDES et MAGGERT, 2009; PAREDES *et al.* 2011).

On a pu mettre en évidence cette influence également chez *A. thaliana*, où on constate une grande surexpression de gènes dans la lignée 20%ADNr comparée à l'écotype sauvage. Plus de 85% des gènes différentiellement exprimés sont surexprimés. Chez l'humain et le maïs, l'expression des gènes influencés par la variation du nombre de copies était le plus souvent corrélée positivement avec les différences de quantité d'ADNr. En d'autres mots, si l'individu présente plus de copies, le gène est plus exprimé ou vice-versa. Par contre, chez la drosophile, la réduction artificielle de copies de gènes d'ARNr induit la sur expression mais également la sous expression de gènes à une proportion à peu près égale (PAREDES *et al.* 2011).

Chez le maïs, les gènes corrélés à la variation du nombre de copies de gènes 45S sont impliqués dans la biogenèse des ribosomes (LI *et al.* 2018). Par contre, chez *A. thaliana* comme chez l'humain et la drosophile, les gènes impactés sont répartis sur l'ensemble des chromosomes et ne sont pas uniquement liés à la fonction nucléolaire (GIBBONS *et al.* 2014; PAREDES *et al.* 2011). De plus, une étude a démontré, grâce à la technique de Hi-C, que les gènes identifiés comme étant associés à la variation du CN en ADNr parmi les génotypes de populations humaines, n'étaient pas en contact avec l'ADNr au sein du noyau (YU et LEMOS, 2018).

Ceci laisse penser que les gènes d'ARNr ont un effet pléiotropique indirect sur l'expression de certains gènes transcrits par l'ARN Pol II. Les mécanismes par lesquels les gènes d'ARNr impactent l'expression d'autres gènes est inconnue. Il a été suggéré que ce mécanisme de régulation indirecte pouvait être dû à des changements au niveau de l'épigénome, induits par des changements du nombre de répétitions de l'ADNr.

Parmi les gènes dont l'expression est corrélée chez l'humain avec la variation du nombre de copies d'ARNr, se trouvent des gènes codant pour des facteurs impliqués dans la régulation de la chromatine. C'est le cas par exemple, des gènes codant pour la pro-

téine structurale HP1 β (HETEROCHROMATIN PROTEIN 1-beta) ou encore le facteur de transcription CTCF (GIBBONS *et al.* 2014). Suite à ces résultats, les auteurs ont proposé que le dosage de l'ADNr pourrait altérer les états de la chromatine, en altérant l'expression de ces gènes qui seraient «dosage-sensitive» (*i.e.* : dépendants du nombre de copies d'ADNr). Ces protéines codées par les gènes «dosage-sensitive», se fixent et régulent également la chromatine au niveau des NORs. Le dosage différentiel des NORs provoquerait des différences de concentration dans le noyau de ces protéines, et pourraient donc modifier la chromatine et l'expression de certains gènes sur l'ensemble du génome (GIBBONS *et al.* 2014; YU et LEMOS, 2018).

Cependant, chez *A. thaliana*, l'analyse transcriptomique de la lignée 20%rDNA n'a pas mis en évidence des changements d'expression de gènes impliqués dans la régulation de la chromatine. Soit cette régulation «dosage-sensitive» n'existe pas dans cette espèce, soit des possibles facteurs impliqués dans ce mécanisme sont régulés de façon post-transcriptionnelle ou sont tout simplement inconnus. De même, les différences au niveau des NADs de cette lignée, ne permettent pas d'expliquer l'expression différentielle des gènes.

En revanche, nos données de séquençage ont permis de mettre en évidence 7 duplications en tandem en orientation directe (TDDO). Ces duplications ont été validées par l'analyse de la couverture d'alignement des séquences issues du séquençage Nanopore de la lignée 20%ADNr, sur le génome de référence TAIR10, mais aussi par Hi-C et microscopie. L'ensemble de ces duplications comptent pour un gain de 2,31 Mb par génome haploïde, induisant la duplication de 626 gènes et 851 TEs. Ces duplications ont permis d'expliquer la différence d'expression de 20% des gènes sur-exprimés dans la lignée 20%rDNA (70 des 321 gènes sur-exprimés sont dupliqués). En effet, il est connu que des variations du nombre de copies de gènes induisent des changements d'expression de ces derniers (ZHANG *et al.* 2019; ŻMIENKO *et al.* 2014).

De façon intéressante, mis à part le changement d'expression de nombreux gènes, ces duplications ont également induit la formation de gènes chimériques (fusion de gènes) exprimés par la cellule. Les transcrits issus de ces gènes chimériques ont en effet été retrouvés dans les données d'ARN-seq. Pour vérifier si ces transcrits donnent lieu à des protéines stables, il serait très intéressant de créer des lignées transgéniques exprimant ces derniers. C'est à dire insérer un construit contenant un promoteur ubiquitaire, comme le 35S, et la séquence chimérique issue de la fusion de gènes et voir si la plante traduit les transcrits et a protéine synthétisée est stable.

Nos données ne nous permettent pas d'expliquer le mécanisme par lequel les duplications en tandem se sont produites. Plusieurs hypothèses sont plausibles, mais celles qui me paraissent le plus en accord avec les résultats obtenus sont les suivantes. Premiè-

rement, la recombinaison homologue non allélique (*c.f.* : Introduction Générale), surtout dans le cas de la TDDO3 et 4, car ce sont des régions enrichies en séquences répétées. Deuxièmement, la réparation par excision de base (BER) «long-patch», qui se produit également suite à un DSB, suivie d'une ligation des extrémités cassées par NHEJ. Comme la NAHR, la réparation par BER «long-patch» combinée à la NHEJ induit des translocations et duplications en tandem (pour une explication détaillée de ce mécanisme, voir : KRASILEVA, 2019). En effet, les autres duplications ne sont pas enrichies en séquences répétées, et donc la NAHR semble moins probable.

Cependant, ce qui est fortement probable, c'est que ces duplications sont les conséquences d'une erreur de réparation de l'ADN suite à un DSBs. Que ce soit la perte de copies d'ADNr ou les erreurs de réplication dus aux mutations *fas1* ou *fas2* (*c.f.* : plus bas), dans les deux cas il y a une augmentation du le nombre de DSBs dans le génome (*c.f.* : article en préparation et MUCHOVÁ *et al.* 2015). Les différentes régions dupliquées mises en évidence dans la lignée 20%ADNr, possèdent une caractéristique commune. En effet, au moins un des deux sites où s'est produit la cassure est localisé au niveau de la limite entre un domaine LSD (Loose Structural Domain) et un domaine CSD (Closed Structural Domain), les analogues compartiments A et B chez les animaux. Comme il s'agit d'une frontière entre deux types de chromatine très différents, il se peut que ce soit des sites fragiles et donc plus affectés dans le cas d'une instabilité génomique accrue.

La duplication du bras court du chromosome 4 (TDDO4) est très probablement due à la réduction de copies d'ADNr. Il est vrai que cette duplication est également retrouvée dans la génération 4 du mutant *fas2*, mais ce mutant présente déjà une réduction de 80% de l'ensemble des copies. De plus, une TDDO au niveau du bras court du chromosome 4 comprenant également le *knob* *hk4S*, a été retrouvée dans une lignée d'*A. thaliana* Col-O indépendante, où la réduction de copies d'ADNr a été réalisée par CRISPR-Cas9 (communication personnelle).

L'augmentation de l'instabilité de l'ADNr due à la réduction du nombre de copies d'ADNr et la subséquente réparation par NAHR, a dû provoquer la duplication de la région génomique adjacente aux copies actives (il n'y a plus de copies inactives qui les séparent) dans les deux lignées.

Mis à part cette TDDO, nos données nous permettent pas d'expliquer l'origine des autres duplications. En effet, la lignée 20%ADNr est issue du croisement entre les mutants *fas1* et *fas2*. Chez ces mutants, il y a également une grande augmentation des DSBs au niveau de l'ensemble du génome et la fréquence de recombinaison homologue est multipliée par 40 (MUCHOVÁ *et al.* 2015). La duplication TDDO3 semble être propre à la lignée L6F6, et donc à la perte de copies d'ADNr car elle n'a pas été retrouvée chez les parents *fas1* ou *fas2*. Dans tous les cas, il est possible qu'elle soit originaire du parent *fas1* G4. C'est pour cela, qu'un séquençage Nanopore des mutants *fas1*, *fas2*, et de la lignée 9 20%ADNr, va

être réalisé. Ce séquençage nous permettra d'affiner l'apparition de l'ensemble des TDDOs depuis les parents *fas1*, *fas2*, jusqu'aux descendants. Il permettra également d'apporter des nouvelles données pour comprendre les mécanismes par lesquels ces TDDOs se sont formées.

Il est possible que les résultats du séquençage Nanopore montrent un effet synergique entre la mutation *fasciata* et la perte de copies d'ADNr, car dans les deux cas une augmentation du nombre de DSBs est observée. De plus, il est possible qu'il y ait une autre cause à cette instabilité. En effet, la création des lignées 6 et 9 20%ADNr a été possible car parmi toute la descendance du croisement entre les mutants *fas1* et *fas2*, ces deux lignées ne récupéraient pas de copies d'ADNr, contrairement au reste. On n'a pas d'explication pour ce constat, mais peut être que le mécanisme impliqué a également eu une influence sur la création des TDDO.

Dans la majorité de cas, les mutations de l'ADN (que ce soit des changements de bases azotées ou des réarrangements chromosomiques) causées par des erreurs de réparation de l'ADN, sont délétères, provoquant souvent des maladies (FRANKE *et al.* 2016 ; PENG et KARPEN, 2008). Mais dans certains cas, cette instabilité peut avoir un effet neutre, voir positif. Il est connu que les évènements de duplication de l'ensemble du génome (polyploïdisation) ou d'un segment d'ADN (comme les TDDO), ont grandement participé à l'évolution des espèces par le biais de la duplication ou fusion de gènes. Suite à de telles duplications, les gènes peuvent évoluer et acquérir des nouvelles fonctions. À une échelle de temps plus courte, ces variations du nombre de copies induisent des variations d'expression des gènes concernés.

Cette variation de l'expression des gènes peut également participer à la fitness de l'individu et contribue de manière significative à la variation phénotypique de populations naturelles, et donc influence l'évolution des espèces. Il existe de nombreux exemples de variation du nombre de copies d'un gène et donc de son expression, liées à des différences dans la réponse à des stress, ou encore dans le moment de la transition florale (pour une *review*, voir : ŻMIENKO *et al.* 2014).

Dans le cas de la lignée 20%ADNr F6L6, on a constaté un grand enrichissement de gènes impliqués dans la réponse biotique dans les gènes différentiellement exprimés. Cette sur-expression de gènes liés à la réponse biotique, met un place un «priming» (sans contact préalable avec l'agent infectieux) chez la plante, qui lui confère une meilleure résistance en cas d'infection biotique. Il est à noter que ces expériences ont été réalisées en laboratoire, dans des conditions contrôlées. Il est possible que dans la nature, le fait d'exprimer ces gènes interfère avec la croissance où la réponse à d'autres stress, et ne s'avère pas positif.

En conclusion, il est possible que la variation naturelle du nombre de copies d'ADNr,

entraîne des changements d'expression de gènes dans les populations naturelles, participant ainsi à la variation phénotypique de ces dernières. Il se peut également, que dans certains cas beaucoup moins fréquents, ces changements aient provoqué des duplications dans d'autres parties du génome, qui aient été maintenues au cours de l'évolution. Cependant, dans notre cas, la variation de copies d'ADNr est artificielle et très forte. On ne retrouve pas d'individus avec uniquement 20% de copies d'ADNr dans les populations naturelles d'*A. thaliana*. Enfin, il est possible que les duplications soient dues à des effets synergiques entre les mutations *fas1*, *fas2* et la réduction du nombre de copies d'ADNr. Pour avoir un aperçu plus ample et ainsi mieux comprendre l'effet de la variation du nombre de copies d'ADNr, même si nous savons que la TDDO3 n'a pas été générée dans un contexte de mutation *fasciata*.

Matériel et Méthodes

6.1 Conditions de culture des plantes

Culture des plantes en chambre de culture

Les graines sont semées sur terre (contenant 1 volume de vermiculite pour 3 volumes de terreau) puis déposées 48 heures en chambre froide (6°C) pour uniformiser le moment de la germination (étape de stratification).

Ensuite, elles sont mises en culture dans des chambres de croissance, avec une photopériode de 16 heures de lumière à $100 \mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ (ampoules fluorescentes à spectre blanc de 6500K, Sylvania) et 8 heures d'obscurité, 60 à 75% d'humidité et 20°C. Ou en chambre de culture Percival® pour une croissance dans des conditions contrôlées, avec une photopériode de 16 heures de lumière à $150 \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ et 22°C pour 8 heures d'obscurité à 20°C ; 55% d'humidité constante.

Culture *in vitro* de plantes

Les graines sont stérilisées dans un bain de 12 minutes dans une solution d'éthanol 75% et 0,05% de SDS, puis rincées avec trois lavages successifs de 3 minutes à l'éthanol 95% avant d'être séchées.

Les plantes sont cultivées *in vitro* dans des boîtes de Pétri contenant du milieu MS (Murashige et Skoog, Duchefa Biochemie M0231) 1X gélosé : 4,7g/L de milieu MS, 5g/L de MES et 10g/L de Plant Agar (Gold Biotechnology). Le pH est ajusté à 5,7 avec une solution de KOH. Les boîtes sont incubées pendant 48 heures à 6°C dans le noir (étape de stratification) et placées dans une chambre de croissance à 20°C avec une photopériode de 16 heures de lumière à $130 \mu\text{E}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ (DEL à spectre blanc 4500K, Vegeled) et 8 heures d'obscurité.

Traitement Bléomycine

Les graines de l'écotype *Arabidopsis thaliana* Col-0 et de la lignée 20%ADNr L6F6 ont été stérilisées, et les plantes sont cultivées *in vitro* dans des boîtes de Pétri contenant du milieu MS 1X gélosé, additionné de $1,5 \mu\text{g}/\text{mL}$ de bléomycine. Après 18 jours de croissance, les plantes ont été pesées.

6.2 Création de lignées transgéniques FIB2 ::YFP

Clonage d'un gène dans un plasmide

Amplification d'un fragment d'ADN par Polymerase Chain Reaction (PCR)

La PCR permet d'amplifier exponentiellement un fragment d'ADN délimité par deux amorces (ou oligonucléotides). Pour la réalisation du clonage, les PCR sont réalisées avec l'enzyme Phusion™ High Fidelity (HF) DNA polymerase (Finnzymes). La réaction de PCR est réalisée dans le mélange suivant : du tampon Phusion HF à 1X, 200 pmoles de dNTPs, 10 pmoles de chaque amorce, 100 ng d'ADN matrice, 0,5 U de polymérase. Des thermocycleurs Biometra® Tprofessional ont été utilisés pour effectuer les réactions PCR. Ensuite, une électrophorèse sur gel d'agarose (1,2 %) avec l'agent intercalant RedGel® à une concentration finale de 1/10000ème est réalisée (25 minutes à 130 V). Les amorces utilisées permettent d'ajouter aux extrémités du fragment d'ADN amplifié les séquences nécessaires à son clonage (i. e. : sites de restriction ou de recombinaison homologue). Pour introduire dans le vecteur pDONRTM221®, la séquence codante de la FIBRILLARINE2, les sites de recombinaison homologue AttB1 et AttB2 ont été rajoutés grâce aux amorces : FIB2+AttB1–F (GGGGACAAGTTTGTACAAAAAAGCAGGCTTTATGAGACCTCCTCTAACTGG) et FIB2+AttB2–R (GGGGACCACTTTGTACAAGAAAGCTGGGTCTCAAGCAGCAGTAGCAGCCTTTGG).

Purification sur gel d'un fragment d'ADN

Suite à l'électrophorèse, le marqueur de taille 1Kb Plus DNA Size Ladder (à 100 ng/μl, Gibco), est utilisé pour estimer la taille des fragments d'ADN sur le gel. Le fragment d'intérêt est excisé du gel, avant d'être purifié avec le kit GENE CLEAN Turbo Protocols® (Qbiogen, États-Unis), selon les directives du fabricant.

Clonage Gateway® (Thermo Fisher Scientific, États Unis)

Le clonage Gateway® exploite la recombinaison homologue du bactériophage lambda pour transformer des plasmides (HARTLEY, 2000). L'enzyme BP Clonase™ permet d'insérer dans un plasmide dit donneur (pDONRTM221®), un fragment d'ADN possédant à ses extrémités les sites de recombinaison homologue AttB1 et AttB2 ajoutés par PCR. Le plasmide pDONRTM221® contenant le fragment d'ADN d'intérêt permet d'insérer ce dernier dans un plasmide de destination par la LR Clonase™ II. Ce clonage est réalisé selon les recommandations du fabricant (Invitrogen®).

Transformation de bactéries avec un plasmide

Les souches bactériennes *Escherichia coli* DH5 α thermo-compétentes sont transformées par le biais d'un choc thermique. D'abord, 50 μ l de cellules compétentes sont décongelées sur glace et additionnées de 10 μ l de tampon KCM 5X (0,5M de KCl; 0,15M de CaCl₂ et 0,25M de MgCl₂) à une concentration finale d'1X, ainsi que le plasmide ou le produit de ligation à transformer. Le mélange est ensuite incubé 30 minutes dans la glace. Après un choc thermique de 45 secondes à 42°C suivi de 2 min dans la glace, 1 mL de milieu liquide LB (Luria Bertani) est ajouté. Ce milieu est composé d'extraits de levure (5 g/L), de Bacto Tryptone (10 g/L), de NaCl 10 g/L, le tout ajusté à pH 7,2. Les bactéries sont ensuite incubées 30 minutes à 37°C sous agitation (320 rpm). Pour terminer, 250 μ L de culture sont étalées sur des boîtes de Pétri contenant du milieu solide LB-Agar (à 15 g/L) et de la kanamycine (Duchefa) à 50 μ g/mL (uniquement les bactéries ayant intégrées le plasmide pourront se développer), puis sont incubées à 37°C toute la nuit.

PCR (Polymerase Chain Reaction) sur colonie

Une partie de chacune des colonies à tester est ajoutée à un mélange réactionnel contenant 2U de GoTaq[®] Polymerase (Promega), 4 μ L de tampon GoTaq[®] (concentration finale de 1X, Promega), 10 pmoles de chacune des amorces et 200 pmoles de dNTPs. Les réactions PCR sont réalisées dans les thermocycleurs Biometra[®] Tprofessional : 30 cycles avec 30 secondes à 96°C, 30 secondes à 56°C et 150 secondes à 72°C. Une élongation finale de 10 minutes à 72°C termine la réaction. Les produits PCR sont ensuite analysés par électrophorèse sur gel d'agarose (1,2 %).

Amplification, purification et dosage d'un plasmide

Les colonies bactériennes possédant le plasmide sont mises en culture dans 5 mL de milieu LB liquide contenant de la kanamycine à 50 μ g/mL et sont incubées sur la nuit à 37°C (320 rpm). Ensuite, les bactéries sont culotées par centrifugation (1 min à température ambiante, 14000 rpm). Les plasmides sont alors purifiés à l'aide du kit Wizard Plus SV Minipreps DNA Purification System[®] (Promega, États-Unis), avant d'être dosés avec le spectrophotomètre NanoDrop[™]2000 (Thermo Fisher Scientific, États-Unis).

Séquençage Sanger de plasmides

Les réactions de séquençage sont réalisées à l'aide du kit Big Dye v3.1 (Applied Biosystems, États-Unis) et du séquenceur « 3130 xl Genetic Analyser » (Applied Biosystems). Les réactions sont préparées dans 8 μ L finaux avec 5 pmol d'amorces et 100 ng d'ADN pour un plasmide de 4000 paires de bases. Les séquences sont par la suite analysées via le logiciel Geneious[®] version 8.1.8. Les amorces utilisées pour vérifier la séquence codante de la FIBRILLARINE2 intégrée dans le plasmide pEG ::FIB2 ::YFP sont

les suivantes : FIB2–R1 (CACCACTGAATCCACCGTCA) et FIB2–R2 (CCTCGACCACCACTGAATCC).

Transformation de plantes

Transformation et sélection d'*Agrobacterium tumefaciens*

La souche C58C1 d'*Agrobacterium tumefaciens* (compétentes) a été utilisée pour réaliser les transformations de plantes. Cette dernière est résistante à la rifampicine et contient le plasmide pMP90, nécessaire pour l'insertion de l'ADN-T dans le génome de la plante, résistant à la gentamicine.

1 μg de plasmide est incorporé dans 100 μL d'*A. tumefaciens* compétentes, suite à un choc thermique de 5 minutes à 37°C. Puis, 2 mL de LB liquide sont ajoutés, et l'ensemble est incubé à 28°C pendant 3 heures sous agitation (320 rpm). Ensuite, 200 μL de culture sont étalées sur boîte de Pétri avec du milieu LB - Agar additionné des antibiotiques rifampicine (Duchefa, à une concentration finale de 12,5 $\mu\text{g}/\text{mL}$), gentamicine (Sigma, à 25 $\mu\text{g}/\text{mL}$) et kanamycine (à une concentration finale de 50 $\mu\text{g}/\text{mL}$). Les boîtes sont alors incubées pendant 48 heures à 28°C et les colonies analysées par PCR sur colonie.

Transformation des plantes par insertion d'AND-T

La transformation repose sur la méthode d'inoculation au Silwet® (De Sangosse, France), décrite par Clough et Bent (1998) et qui permet de transformer les plantes au stade de la floraison.

Une colonie d'agrobactéries transformées est cultivée dans 5 mL de LB liquide additionné des 3 antibiotiques aux mêmes concentrations que dans le milieu solide, sur la nuit, à 28°C (120 rpm). De cette pré-culture, 2 mL sont utilisés pour ensemercer 200 mL de culture (contenant les antibiotiques), puis incubée pendant 20 h à 28°C (120 rpm). Les bactéries sont culotées par centrifugation (1 min à 20°C, 7000 rpm) et reprises dans 200 mL de solution d'inoculation (5 % de sucrose ; 200 $\mu\text{L}/\text{L}$ de Silwet® ; 10 mM de MgCl_2). Les inflorescences sont immergées dans la suspension d'*A. tumefaciens* pendant 45 secondes puis entourées de film plastique afin de les protéger de la dessiccation. Après 48 heures en dehors de la lumière directe, les plantes sont replacées en pièce de culture.

Sélection des plantes ayant intégré le marqueur nucléolaire FIB2 ::YFP

Les transformants de la première génération sont sélectionnés par un traitement Basta®, et les F2 homozygotes sont sélectionnés au microscope inversé à fluorescence Nikon® Eclipse TS100-F®, en utilisant le filtre pour la fluorescence de la YFP (FL2 : filtre passe-bande 510-560 nm).

6.3 Identification des NADs chez *Arabidopsis thaliana*

Le protocole d'identification des NADs fait l'objet d'une section du chapitre X de ce manuscrit (L'étude des NADs chez *Arabidopsis thaliana*). Cependant c'est le FACS BioRad S3™ avec le logiciel ProSort™ (BioRad, États Unis) qui ont été utilisés pour trier des noyaux ou des nucléoles. Un graphique de l'intensité du FSC en fonction de la fluorescence de la YFP (FL1) est créée, la zone de tri est ensuite définie en entourant toutes les particules qui émettent un signal $-\log(\text{FL1}) \leq 100$ (ajuster en fonction de l'échantillon WT). La méthode d'extraction au phénol-chloroforme pour purifier l'ADN et le kit QuantiFluor® dsDNA System pour le doser ont également été utilisés. Le séquençage des échantillons d'ADN a été réalisé par la plateforme Bio-Environnement de l'UPVD : préparation des bibliothèques d'ADN avec le kit Nextera® XT (illumina, États-Unis), selon les recommandations du fabricant ; séquençage illumina (séquençage en «paired-end», 2x150 pb) par le système NextSeq^{MC}550 (illumina, États-Unis).

purification de l'ADN au phénol-chloroforme

Pour purifier l'ADN au phénol-chloroforme, un demi-volume de phénol à pH 8 et un demi-volume de chloroforme sont ajoutés à l'échantillon. Après 30 secondes de vortex, l'échantillon est centrifugé 15 minutes à vitesse maximale et à 40°C. La phase aqueuse (surnageant) est récupérée puis un volume de chloroforme y est ajouté avant de centrifuger 10 minutes à vitesse maximale et à 40°C. Le surnageant est de nouveau récupéré et un dixième de volume d'acétate de sodium (NaAc), puis deux volumes et demi d'éthanol 100% y sont additionnés avec 0,75 μL de glycogène d'azur. Les échantillons sont incubés 1 heure à - 80°C pour permettre la précipitation des ADN, puis centrifugés 20 minutes à vitesse maximale et à 40°C. Après un lavage avec de l'éthanol à 70%, le culot est repris dans 25 μL de tampon TE à pH 8 (10mM de Tris à pH 8 ; 1mM d'EDTA à pH 8).

6.4 Visualisation de noyaux et immunolocalisation de protéines nucléaires au microscope confocal

Fixation des noyaux de feuilles sur des lames en polysine

Les feuilles sont fixées dans du tampon Tris froid (10 mM de Tris- HCl pH 7,5 ; 10mM d'NaEDTA ; 100mM de NaCl) contenant 4% de formaldéhyde pendant 20 minutes, puis lavées deux fois pendant 10 minutes avec du tampon Tris froid. Les feuilles sont hachées finement à l'aide d'une lame de rasoir dans 500 μL de tampon LB01 (15 mM de Tris-HCl pH 7,5, 2mM d'NaEDTA, 0,5 mM de spermine, 80 mM de KCl, 20 mM de NaCl et 0.1% Triton X-100). Le liquide obtenu est filtré avec un filtre de 30 μm CellTrics® (Sysmex

Partec, Japon) puis additionné d'un volume égal de tampon de tri (100 mM de Tris-HCl pH 7,5; 50mM KCl; 2mM de MgCl₂; 0,05% de Tween[®]20 [Sigma-Aldrich, États Unis]; 5% de saccharose). 20 *mu*L sont ensuite déposés sur une lame en polysine, séchés à température ambiante. Ensuite l'échantillon est fixé sur la lame avec 20 *mu*L de tampon 1X PBS (137 mM de NaCl; 2,7 mM de KCl; 10 mM de Na₂HPO₄; 1,8 mM de KH₂PO₄) contenant 2% de formaldéhyde pendant 5 minutes, ensuite il est rincé trois fois au 1X PBS puis laissé sécher à température ambiante.

Coloration DAPI et visualisation des noyaux

Une fois que la lame est sèche suite aux trois lavages au 1X PBS, un lavage de 5 minutes au 1X PBS additionné de DAPI (4',6-diamidino-2-phénylindole) à une concentration finale de 1 *mu*g/mL est réalisé à température ambiante, puis un dernier lavage de 5 minutes au 1X PBS. La lame est montée avec 18 μ L de milieu Vectashield[®] (Maravai LifeSciences, États-Unis), additionné de DAPI à une concentration finale de 1 *mu*g/mL. Les observations et images sont réalisées et analysées avec un microscope confocal à balayage laser Zeiss LSM 700 (Carl Zeiss, Allemagne) et le logiciel ZEN (Carl Zeiss, Allemagne).

Immunolocalisation de protéines nucléaires

Une fois que la lame est sèche suite aux trois lavages au 1X PBS, ajouter directement sur les lames du 1X PBS additionné de 0,5% de détergeant Triton et incubé pendant 15 minutes à température ambiante dans une chambre humide. Ensuite, 3 lavages de 5 minutes avec du PBS 1X sont réalisés à température ambiante. Ensuite, l'anticorps primaire anti γ H2A_X (dilution à 1/500) contenu dans une solution de PBS 1X, 3% d'Albumine de Sérum Bovine (BSA) et 0,05% de Tween[®]20; est déposé sur la lame, puis les lames sont incubées sur la nuit à 4°C dans une chambre humide et à l'obscurité. Le lendemain, réaliser 3 lavages de 5 minutes des lames avec du 1X PBS à température ambiante. Ajouter sur la lame l'anticorps secondaire (dilution à 1/1000) anti-souris Alexa 488 (Abcam, Royaume-Uni) contenu dans une solution de PBS 1X; 3% d'Albumine de Sérum Bovine (BSA) et 0,05% de Tween[®]20. Les lames sont incubés pendant 2 à 3 heures dans une chambre humide à température ambiante et un lavage de 5 minutes avec du PBS 1X à température ambiante est réalisé. Ensuite, un lavage de 5 minutes au 1X PBS additionné de DAPI à une concentration finale de 1 *mu*g/mL est réalisé à température ambiante, puis un dernier lavage de 5 minutes au 1X PBS. La lame est montée avec 18 μ L de milieu Vectashield[®] (Maravai LifeSciences, États-Unis), additionné de DAPI à une concentration finale de 1 *mu*g/mL. Les observations et images sont réalisées et analysées avec un microscope confocal à balayage laser Zeiss LSM 700 (Carl Zeiss, Allemagne) et le logiciel ZEN (Carl Zeiss, Allemagne).

6.5 Analyse de l'expression des gènes

Purification d'ARN au TRIzol et séquençage ARN

L'ARN total de feuilles de l'écotype *Arabidopsis thaliana* Col-0 et de la lignée 20%ADNr L6F6 a été purifié par la méthode de purification au TRIzol. Les feuilles sont broyées dans de l'azote liquide pour obtenir une poudre. ~1 g de poudre est transféré dans un tube en polypropylène de 15 mL et 10 ml de réactif TRI Reagent® (Zymo Research, États-Unis) sont ajoutés. Les tubes sont ensuite vortexés et incubés 5 minutes à température ambiante. 2 mL de chloroforme froid sont ajoutés, les tubes sont vortexés pendant 20 secondes puis incubés 3 minutes à température ambiante. Après une centrifugation de 15 minutes à $8\,000 \times g$ et 4°C , la phase aqueuse est transférée dans un nouveau. 6mL d'isopropanol froid sont ajoutés (mélanger doucement par inversion) et les tubes sont incubés 30 minutes sur la glace. L'ARN est sédimenté par centrifugation pendant 30 minutes à $8000 \times g$ et 4°C , le culot est lavé avec 75% d'éthanol, puis laissé sécher. Le culot est remis en suspension dans 50 μL d'eau milliQ puis dosé avec le spectrophotomètre NanoDrop™2000.

Le séquençage des échantillons d'ARN de feuilles d'*A. thaliana* Col-0 et de la lignée 20%ADNr L6F6 a été réalisé par la plateforme Bio-Environnement de l'UPVD : séquençage illumina (séquençage en «paired-end», 2x75 pb) par le système NextSeq^{MC} 550 (illumina, États-Unis). Les ARN à queue polyadénylée (polyA) totaux de plantes stressées (22 jours à 37°C) et contrôle (22 jours à 22°C) ont été séquencés à la plateforme MGX-Montpellier GenomiX (50 paires de bases en «single-end», séquenceur HiSeq® 2500 d'illumina®). Les données de séquençage des ARN (ARN-seq) ont été récupérées à partir de fichiers au format fastq.

Analyse bioinformatique des séquences d'ARN totaux (sous linux)

Analyse de qualité et nettoyage des séquences

L'outil FastQC (version 0.11.5, Babraham bioinformatics, ©Andrews, Lindenbaum, Howard et Ewels) a été utilisé pour réaliser les analyses de qualité. L'analyse de qualité est faite avec l'outil fastQC. La détection des séquences des adaptateurs et le filtre des séquences issues du ARN-seq en fonction du score de qualité Sanger sont faits avec l'outil Trimmomatic® (version 0.36 ; BOLGER *et al.* 2014).

Alignement des séquences sur le génome de référence d'*A. thaliana* Col-0 TAIR10

Le logiciel utilisé est HISAT2 (KIM *et al.* 2015), il permet l'alignement de séquences issues de transcrits sur un génome, à l'aide de l'algorithme *Bowtie2* (version 2.3.4 ; LANGMEAD *et al.* 2009), et du fichier d'annotation du génome de référence (TAIR10), pour tenir

compte de la localisation des introns (lorsqu'il s'agit de transcrits épissés). Ensuite, les fichiers d'alignement au format BAM générés par HISAT2 ont été convertis au format SAM avec l'outil samtools *view* (version 1.3; LI, 2011).

Quantification de l'expression des gènes

Le logiciel HTseq-count (version 0.6.0; Simon Anders, European Molecular Biology Laboratory, 2010) est utilisé pour compter le nombre de séquences alignées sur chaque séquence codante (CDS) du génome de référence (TAIR10). Les tableaux de comptage générés par HTseq-count sont regroupés en un seul tableau de comptage au format .csv pouvant être lu par le logiciel R. Dans ce tableau, chaque ligne correspond à un CDS, et chaque colonne à un échantillon. Un deuxième fichier descriptif au format .csv est aussi construit pour indiquer les caractéristiques de chacun des échantillons pour qu'ils soient traités par le progiciel DESeq2 (LOVE *et al.* 2014).

Analyse différentielle de l'expression des gènes

Le progiciel de R utilisé est DESeq2 (LOVE *et al.* 2014), avec la version 3.6.1 de R (R Core Team, 2019). DESeq2 réalise, à partir du tableau de comptage et du fichier descriptif, l'analyse d'expression différentielle par une loi binomiale négative et teste la significativité des différences d'expression avec le test de Wald. La distance Euclidienne (classification hiérarchique) entre les échantillons à partir du tableau de comptage a également été calculée, pour vérifier l'accord des données avec l'origine des échantillons (stressés ou contrôles).

Analyse des voies métaboliques différentiellement exprimées

Cette analyse a été réalisée par le biais de l'outil ThaleMine (version 1.9.0; <https://apps.araport.org/thalemine>) de l'ARAbidopsis information PORTal (ARAPORT; KRISHNAKUMAR *et al.* 2015).

Validation des différences d'expression de gènes par RT-qPCR

Afin d'obtenir des ADN complémentaires, les ARN ont été traités avec le kit Invitrogen™ SuperScript™ IV Reverse Transcriptase (Thermo Fisher Scientific, États Unis), selon les recommandations du fabricant. La préparation des échantillons pour la qPCR a été réalisée avec le kit Takyon™ NoROX SYBR® Master mix dTTTP Blue (Eurogentec, Belgique), selon les recommandations du fabricant. Les qPCR ont été réalisées dans le LightCycler® 480 Instrument II. Les résultats sont analysés en utilisant la méthode comparative CT à l'aide du logiciel d'analyse LightCycler®. Les amorces qui ont été utilisées pour la vérification de l'ARNseq d'*A. thaliana* Col-0 et de la lignée 20%ADNr L6F6 se trouvent en Annexe 2.

6.6 Identification des régions différentiellement méthylées (DMRs)

Le kit illustra Nucleon® Phytocomplete Genomic DNA Extraction (Tepnel Life Sciences PLC, Royaume Uni) a été utilisé pour purifier de l'ADN de l'écotype *Arabidopsis thaliana* Col-0 et de la lignée 20%ADNr L6F6. Les échantillons d'ADN ont été dosés avec le kit QuantiFluor® dsDNA System, selon les recommandations du fabricant. Le traitement au bisulfite et le séquençage ADN a été réalisé par la plateforme de séquençage Novogene (Chine), les données de séquençage ont été récupérées au format fastq.

Analyse bioinformatique des séquences d'ADN (sous linux)

Alignement des séquences sur le génome de référence d'*A. thaliana* Col-0 TAIR10

L'analyse de qualité et le nettoyage des séquences a été réalisé comme pour les séquences d'ARN. L'alignement sur le génome de référence a été réalisé *via* le logiciel *Bismark bisulfite mapper* (version 0.22.1, [algorithme *Bowtie2* version 2.3.4; LANGMEAD *et al.* 2009], alignement en «paired-end», paramètres par défaut; KRUEGER et ANDREWS, 2011) Le logiciel *Bismark bisulfite mapper* génère un seul fichier d'alignement au format *Binary Alignment Map* (BAM). Pour obtenir les statistiques de cet alignement, l'outil *Flagstat* (version 2.0, paramètres par défaut) du progiciel *Samtools* (version 1.3 ; LI, 2011) a été utilisé. Les fichiers d'alignement BAM générés par le logiciel *Bismark bisulfite mapper* sont triés en fonction du nom du «scaffold» avec l'outil *Sort a BAM file* (version 1.0.0, paramètres par défaut) du progiciel *Samtools* (version 0.1.3 ; LI, 2011), puis convertis en format SAM avec l'outil *BAM-to-SAM* (version 1.0.4, paramètres par défaut) également du progiciel *Samtools* (version 1.3 ; LI, 2011); pour être lus par le progiciel *methylKit*.

Identification des régions différentiellement méthylées avec le progiciel *methylKit*

Le progiciel *methylKit* (AKALIN *et al.* 2012) du logiciel R (version 3.6.1, R Development Core Team, 2019), lit les fichiers d'alignement pour un seul contexte de méthylation de la cytosine (CpG, CHH ou CHG) à la fois. Ainsi, les fichiers d'alignement issus de *Bismark bisulfite mapper* sont lus trois fois, une fois par contexte de méthylation. Ensuite, les statistiques de la méthylation [*getMethylationStats()*] et de la couverture d'alignement [*getCoverageStats()*] pour chaque échantillon, et chaque contexte de méthylation de la cytosine, sont calculées. Pour la suite des analyses, les données sont filtrées en fonction de la couverture d'alignement [*filterByCoverage()*] par cytosine (dans le contexte analysé); les cytosines ayant une couverture de moins de 8X sont éliminées. Deuxièmement, les données de méthylation par cytosine (dans un des trois contextes), sont caractérisées par régions génomiques de 500 bases [*tileMethylCounts()*], puis seule-

ment les régions méthylées dans les six échantillons sont retenues pour l'analyse et réunies en un fichier [unite()]. Les données obtenues sont ensuite clustérisées selon une classification hiérarchique (métrique de distance Euclidienne entre les méthylomes des échantillons). Finalement, les régions de 500 bases ayant des différences de méthylation significatives (régression logistique, valeur $q < 0.01$) sont obtenues avec la commande `calculateDiffMeth()`; et grâce à la commande `get.methylDiff()`, les zones avec une différence de taux de méthylation significative égale ou supérieure à 20% pour les contextes CHH et CHG, ou égale ou supérieure à 50% pour le contexte de méthylation CpG sont définies comme des régions différentiellement méthylées.

6.7 Séquençage Nanopore

Extraction et purification de l'ADN

Extraction de l'ADN de feuilles de l'écotype *Arabidopsis thaliana* Col-0 et de la lignée 20%ADNr L6F6 au Bromure de cetyltriméthylammonium (CTAB) (CLARKE, 2009). Suite à cette extraction, un traitement RNase est fait, l'RNase A est ajoutée dans chaque échantillon d'ADN pour une concentration finale de 10 $\mu\text{g}/\text{mL}$, puis les échantillons sont incubés à 37°C pendant 30 minutes. Ensuite l'ADN est dosé avec le kit Qubit™ dsDNA High Sensitivity (Thermo Fisher Scientific, États Unis), selon les recommandations du fabricant. Une deuxième étape de purification est réalisée avec le kit Genomic DNA Clean & Concentrator-10 (Zymo Research, États-Unis), selon les recommandations du fabricant (l'ADN est élué dans un volume de 100 μL d'«elution buffer» préchauffé à 65°C). Finalement l'ADN est précipité à l'acétate de sodium et éthanol : ajouter 1/10 de volume d'acétate de sodium 3M et pH=5,2 puis 2 volumes d'éthanol absolu froid, ensuite les échantillons sont centrifugés à 4°C pendant 45 minutes à 13000 rpm. Le culot d'ADN est rincé avec 500 μL d'éthanol 75% froid puis les échantillons sont centrifugés à 4°C pendant 15 minutes à 13000 rpm. Le surnageant est éliminé, le culot est séché à température ambiante puis resuspendu dans 30 μL de tampon TE à pH 8 (10mM de Tris à pH 8; 1mM d'EDTA à pH 8) et incubé à 4°C sur la nuit. L'ADN est à nouveau dosé avec le kit Qubit™ dsDNA High Sensitivity (Thermo Fisher Scientific, États Unis).

Séquençage Nanopore avec le séquenceur MinION

La préparation des banques d'ADN pour le séquençage a été réalisée avec le kit 1D Genomic DNA by ligation SQK-LSK109 (Oxford Nanopore Technologies, Royaume-Uni), selon les recommandations du fabricant. La flowcell utilisée avec le séquenceur MinION a été la FLO-MIN106D version R9 (Oxford Nanopore Technologies, Royaume-Uni).

Bibliographie

- AGRAWAL, S. et A. R. D. GANLEY (2018). “The conservation landscape of the human ribosomal RNA gene repeats”. In : *PLOS ONE* 13.12. Sous la dir. d'I. B. ROGOZIN, e0207531. DOI : 10.1371/journal.pone.0207531.
- AKALIN, A. *et al.* (2012). “methylKit : a comprehensive R package for the analysis of genome-wide DNA methylation profiles”. In : *Genome Biology* 13.10, R87. DOI : 10.1186/gb-2012-13-10-r87.
- ALDRICH, J. C. et K. A. MAGGERT (2014). “Simple Quantitative PCR Approach to Reveal Naturally Occurring and Mutation-Induced Repetitive Sequence Variation on the *Drosophila* Y Chromosome”. In : *PLoS ONE* 9.10. Sous la dir. de N. SINGH, e109906. DOI : 10.1371/journal.pone.0109906.
- (2015). “Transgenerational Inheritance of Diet-Induced Genome Rearrangements in *Drosophila*”. In : *PLOS Genetics* 11.4. Sous la dir. de G. CAVALLI, e1005148. DOI : 10.1371/journal.pgen.1005148.
- ANDERSEN, J. S. *et al.* (2005). “Nucleolar proteome dynamics”. In : *Nature* 433.7021, p. 77-83. DOI : 10.1038/nature03207.
- ANDERSON, R. M., M. KWON et S. A. STROBEL (2007). “Toward Ribosomal RNA Catalytic Activity in the Absence of Protein”. In : *Journal of Molecular Evolution* 64.4, p. 472-483. DOI : 10.1007/s00239-006-0211-y.
- ARMSTRONG, S. J., F. C. FRANKLIN et G. H. JONES (2001). “Nucleolus-associated telomere clustering and pairing precede meiotic chromosome synapsis in *Arabidopsis thaliana*”. In : *Journal of Cell Science* 114 (Pt 23), p. 4207-4217.
- AUDAS, T. E., M. D. JACOB et S. LEE (2012). “Immobilization of Proteins in the Nucleolus by Ribosomal Intergenic Spacer Noncoding RNA”. In : *Molecular Cell* 45.2, p. 147-157. DOI : 10.1016/j.molcel.2011.12.012.
- BACHELLERIE, J.-P. et J. CAVAILLÉ (1997). “Guiding ribose methylation of rRNA”. In : *Trends in Biochemical Sciences* 22.7, p. 257-261. DOI : 10.1016/S0968-0004(97)01057-8.
- BI, X. *et al.* (2017). “Nonrandom domain organization of the *Arabidopsis* genome at the nuclear periphery”. In : *Genome Research*.
- BOISVERT, F.-M. *et al.* (2007). “The multifunctional nucleolus”. In : *Nature Reviews Molecular Cell Biology* 8.7, p. 574-585. DOI : 10.1038/nrm2184.
- BOLGER, A. M., M. LOHSE et B. USADEL (2014). “Trimmomatic : a flexible trimmer for Illumina sequence data”. In : *Bioinformatics* 30.15, p. 2114-2120. DOI : 10.1093/bioinformatics/btu170.
- BONEV, B. et G. CAVALLI (2016). “Organization and function of the 3D genome”. In : *Nature Reviews Genetics* 17.11, p. 661-678. DOI : 10.1038/nrg.2016.112.
- BOULON, S. *et al.* (2010). “The Nucleolus under Stress”. In : *Molecular Cell* 40.2, p. 216-227. DOI : 10.1016/j.molcel.2010.09.024.

- BRANGWYNNE, C. P., T. J. MITCHISON et A. A. HYMAN (2011). "Active liquid-like behavior of nucleoli determines their size and shape in *Xenopus laevis* oocytes". In : *Proceedings of the National Academy of Sciences* 108.11, p. 4334-4339. DOI : 10.1073/pnas.1017150108.
- BROWN, J. W. S. et P. J. SHAW (2008). "The Role of the Plant Nucleolus in Pre-mRNA Processing". In : *Nuclear pre-mRNA Processing in Plants*. Sous la dir. d'A. S. N. REDDY et M. GOLOVKIN. Réd. par R. W. COMPANS *et al.* T. 326. Berlin, Heidelberg : Springer Berlin Heidelberg, p. 291-311. DOI : 10.1007/978-3-540-76776-3_16.
- CHANDRASEKHARA, C. *et al.* (2016). "Chromosome-specific NOR inactivation explains selective rRNA gene silencing and dosage control in *Arabidopsis*". In : *Genes & Development*. DOI : 10.1101/gad.273755.115.
- CHESS, A. *et al.* (1994). "Allelic inactivation regulates olfactory receptor gene expression". In : *Cell* 78.5, p. 823-834. DOI : 10.1016/s0092-8674(94)90562-2.
- CLARKE, J. D. (2009). "Cetyltrimethyl Ammonium Bromide (CTAB) DNA Miniprep for Plant DNA Isolation". In : *Cold Spring Harbor Protocols* 2009.3, pdb.prot5177-pdb.prot5177. DOI : 10.1101/pdb.prot5177.
- CLOUGH, S. J. et A. F. BENT (1998). "Floral dip : a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*". In : *The plant journal* 16.6, p. 735-743.
- COEN, E. S. et G. A. DOVER (1982). "Multiple Pol I initiation sequences in rDNA spacers of *Drosophila melanogaster*". In : *Nucleic Acids Research* 10.21, p. 7017-7026. DOI : 10.1093/nar/10.21.7017.
- COPENHAVER, G. P. et C. S. PIKAARD (1996a). "RFLP and physical mapping with an rDNA-specific endonuclease reveals that nucleolus organizer regions of *Arabidopsis thaliana* adjoin the telomeres on chromosomes 2 and 4". In : *The Plant Journal : For Cell and Molecular Biology* 9.2, p. 259-272.
- (1996b). "Two-dimensional RFLP analyses reveal megabase-sized clusters of rRNA gene variants in *Arabidopsis thaliana*, suggesting local spreading of variants as the mode for gene homogenization during concerted evolution". In : *The Plant Journal : For Cell and Molecular Biology* 9.2, p. 273-282.
- CREMER, M. *et al.* (2001). "Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells". In : *Chromosome Research : An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology* 9.7, p. 541-567.
- DILLINGER, S., T. STRAUB et A. NEMETH (2016). *Nucleolus association of chromosomal domains is largely maintained in cellular senescence despite massive nuclear reorganization*. Cold Spring Harbor Laboratory. DOI : 10.1101/054908.
- DITTMER, T. A. *et al.* (2007). "LITTLE NUCLEI Genes Affecting Nuclear Morphology in *Arabidopsis thaliana*". In : *THE PLANT CELL ONLINE* 19.9, p. 2793-2803. DOI : 10.1105/tpc.107.053231.
- DOĞAN, E. S. et C. LIU (2018). "Three-dimensional chromatin packing and positioning of plant genomes". In : *Nature Plants* 4.8, p. 521-529. DOI : 10.1038/s41477-018-0199-5.
- DONG, P. *et al.* (2017). "3D Chromatin Architecture of Large Plant Genomes Determined by Local A/B Compartments". In : *Molecular Plant* 10.12, p. 1497-1509. DOI : 10.1016/j.molp.2017.11.005.

- DONG, P. *et al.* (2019). "Tissue-specific Hi-C analyses of rice, foxtail millet and maize suggest non-canonical function of plant chromatin domains". In : *Journal of Integrative Plant Biology*, jipb.12809. DOI : 10.1111/jipb.12809.
- EAGLE, S. H. et T. J. CREASE (2012). "Copy number variation of ribosomal DNA and Pokey transposons in natural populations of *Daphnia*". In : *Mobile DNA* 3.1, p. 4. DOI : 10.1186/1759-8753-3-4.
- EARLEY, K. (2006). "Erasure of histone acetylation by *Arabidopsis* HDA6 mediates large-scale gene silencing in nucleolar dominance". In : *Genes & Development* 20.10, p. 1283-1293. DOI : 10.1101/gad.1417706.
- EARLEY, K. W. *et al.* (2010). "Mechanisms of HDA6-mediated rRNA gene silencing : suppression of intergenic Pol II transcription and differential effects on maintenance versus siRNA-directed cytosine methylation". In : *Genes & Development* 24.11, p. 1119-1132. DOI : 10.1101/gad.1914110.
- ELLWOOD, M. et M. NOMURA (1982). "Chromosomal Locations of the Genes for rRNA in *Escherichia coli* K-12t". In : *J. BACTERIOL.* 149, p. 11.
- FENG, S. *et al.* (2014). "Genome-wide Hi-C Analyses in Wild-Type and Mutants Reveal High-Resolution Chromatin Interactions in *Arabidopsis*". In : *Molecular Cell* 55.5, p. 694-707. DOI : 10.1016/j.molcel.2014.07.008.
- FERIC, M. *et al.* (2016). "Coexisting Liquid Phases Underlie Nucleolar Subcompartments". In : *Cell* 165.7, p. 1686-1697. DOI : 10.1016/j.cell.2016.04.047.
- FILIPOWICZ, W. *et al.* (1999). "Structure and biogenesis of small nucleolar RNAs acting as guides for ribosomal RNA modification". In : *Acta Biochimica Polonica* 46.2, p. 377-389.
- FRANKE, M. *et al.* (2016). "Formation of new chromatin domains determines pathogenicity of genomic duplications". In : *Nature* 538.7624, p. 265-269. DOI : 10.1038/nature19800.
- FRANSZ, P. *et al.* (1998). "Cytogenetics for the model system *Arabidopsis thaliana*". In : *The Plant Journal* 13.6, p. 867-876. DOI : 10.1046/j.1365-313X.1998.00086.x.
- FRANSZ, P. *et al.* (2002). "Interphase chromosomes in *Arabidopsis* are organized as well defined chromocenters from which euchromatin loops emanate". In : *Proceedings of the National Academy of Sciences* 99.22, p. 14584-14589.
- FRANSZ, P. *et al.* (2016). "Molecular, genetic and evolutionary analysis of a paracentric inversion in *Arabidopsis thaliana*". In : *The Plant Journal* 88.2, p. 159-178. DOI : 10.1111/tpj.13262.
- FROTTIN, F. *et al.* (2019). "The nucleolus functions as a phase-separated protein quality control compartment". In : *Science*, eaaw9157. DOI : 10.1126/science.aaw9157.
- GABORIAUD, J. et P.-Y. WU (2019). "Insights into the Link between the Organization of DNA Replication and the Mutational Landscape". In : *Genes* 10.4, p. 252. DOI : 10.3390/genes10040252.
- GALARDI, S. *et al.* (2002). "Purified Box C/D snoRNPs Are Able To Reproduce Site-Specific 2'-O-Methylation of Target RNA In Vitro". In : *Molecular and Cellular Biology* 22.19, p. 6663-6668. DOI : 10.1128/MCB.22.19.6663-6668.2002.
- GALGANSKI, L., M. O. URBANEK et W. J. KRZYZOSIAK (2017). "Nuclear speckles : molecular organization, biological function and role in disease". In : *Nucleic Acids Research* 45.18, p. 10350-10368. DOI : 10.1093/nar/gkx759.
- GARCIA, S. *et al.* (2014). "Plant rDNA database : update and new features". In : *Database* 2014.0, bau063-bau063. DOI : 10.1093/database/bau063.

- GARCIA, S. et A. KOVAŘÍK (2013). "Dancing together and separate again : gymnosperms exhibit frequent changes of fundamental 5S and 35S rRNA gene (rDNA) organisation". In : *Heredity* 111.1, p. 23-33. DOI : 10.1038/hdy.2013.11.
- GARCIA, S. *et al.* (2010). "Repeated reunions and splits feature the highly dynamic evolution of 5S and 35S ribosomal RNA genes (rDNA) in the Asteraceae family". In : *BMC Plant Biology* 10.1, p. 176. DOI : 10.1186/1471-2229-10-176.
- GARCIA, S. *et al.* (2017). "Cytogenetic features of rRNA genes across land plants : analysis of the Plant rDNA database". In : *The Plant Journal* 89.5, p. 1020-1030. DOI : 10.1111/tpj.13442.
- GIBBONS, J. G. *et al.* (2014). "Ribosomal DNA copy number is coupled with gene expression variation and mitochondrial abundance in humans". In : *Nature Communications* 5.1, p. 4850. DOI : 10.1038/ncomms5850.
- GIBBONS, J. G. *et al.* (2015). "Concerted copy number variation balances ribosomal DNA dosage in human and mouse genomes". In : *Proceedings of the National Academy of Sciences* 112.8, p. 2485-2490. DOI : 10.1073/pnas.1416878112.
- GINISTY, H. *et al.* (1999). "Structure and functions of nucleolin". In : *Journal of Cell Science* 112 (Pt 6), p. 761-772.
- GONZALEZ, I. L. et J. E. SYLVESTER (1995). "Complete Sequence of the 43-kb Human Ribosomal DNA Repeat : Analysis of the Intergenic Spacer". In : *Genomics* 27.2, p. 320-328. DOI : 10.1006/geno.1995.1049.
- GRANNEMAN, S. et S. J. BASERGA (2005). "Crosstalk in gene expression : coupling and coregulation of rDNA transcription, pre-ribosome assembly and pre-rRNA processing". In : *Current Opinion in Cell Biology* 17.3, p. 281-286. DOI : 10.1016/j.ceb.2005.04.001.
- GROB, S., M. W. SCHMID et U. GROSSNIKLAUS (2014). "Hi-C Analysis in Arabidopsis Identifies the KNOT, a Structure with Similarities to the flamenco Locus of Drosophila". In : *Molecular Cell* 55.5, p. 678-693. DOI : 10.1016/j.molcel.2014.07.009.
- GROZDANOV, P., O. GEORGIEV et L. KARAGYOZOV (2003). "Complete sequence of the 45-kb mouse ribosomal DNA repeat : analysis of the intergenic spacer". In : *Genomics* 82.6, p. 637-643. DOI : 10.1016/S0888-7543(03)00199-X.
- GRUENDLER, P. *et al.* (1991). "rDNA intergenic region from Arabidopsis thaliana structural analysis, intraspecific variation and functional implications". In : *Journal of Molecular Biology* 221.4, p. 1209-1222. DOI : 10.1016/0022-2836(91)90929-Z.
- GUELEN, L. *et al.* (2008). "Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions". In : *Nature* 453.7197, p. 948-951. DOI : 10.1038/nature06947.
- GUETG, C. et R. SANTORO (2012a). "Formation of nuclear heterochromatin : The nucleolar point of view". In : *Epigenetics* 7.8, p. 811-814. DOI : 10.4161/epi.21072.
- (2012b). "Formation of nuclear heterochromatin : The nucleolar point of view". In : *Epigenetics* 7.8, p. 811-814. DOI : 10.4161/epi.21072.
- HACOT, S. *et al.* (2010). "Isolation of Nucleoli". In : *Current Protocols in Cell Biology* 47.1, p. 3.36.1-3.36.10. DOI : 10.1002/0471143030.cb0336s47.
- HARTLEY, J. L. (2000). "DNA Cloning Using In Vitro Site-Specific Recombination". In : *Genome Research* 10.11, p. 1788-1795. DOI : 10.1101/gr.143000.
- HASTINGS, P. J. *et al.* (2009). "Mechanisms of change in gene copy number". In : *Nature Reviews Genetics* 10.8, p. 551-564. DOI : 10.1038/nrg2593.

- HER, J. et S. F. BUNTING (2018). "How cells ensure correct repair of DNA double-strand breaks". In : *Journal of Biological Chemistry* 293.27, p. 10502-10511. DOI : 10.1074/jbc.TM118.000371.
- HERRERA, M. L. *et al.* (2009). "Strain-Dependent Variation in 18S Ribosomal DNA Copy Numbers in *Aspergillus fumigatus*". In : *Journal of Clinical Microbiology* 47.5, p. 1325-1332. DOI : 10.1128/JCM.02073-08.
- HU, B. *et al.* (2019). "Plant lamin-like proteins mediate chromatin tethering at the nuclear periphery". In : *Genome Biology* 20.1, p. 87. DOI : 10.1186/s13059-019-1694-3.
- HUANG, M. *et al.* (2012). "Plant 45S rDNA Clusters Are Fragile Sites and Their Instability Is Associated with Epigenetic Alterations". In : *PLoS ONE* 7.4. Sous la dir. de M. BRYK, e35139. DOI : 10.1371/journal.pone.0035139.
- IDE, S. *et al.* (2010). "Abundance of Ribosomal RNA Gene Copies Maintains Genome Integrity". In : *Science* 327.5966, p. 693-696. DOI : 10.1126/science.1179044.
- JACOB, M. D. *et al.* (2013). "Environmental cues induce a long noncoding RNA-dependent remodeling of the nucleolus". In : *Molecular Biology of the Cell* 24.18. Sous la dir. d'A. G. MATERA, p. 2943-2953. DOI : 10.1091/mbc.e13-04-0223.
- KALININA, N. O. *et al.* (2018). "The Multiple Functions of the Nucleolus in Plant Development, Disease and Stress Responses". In : *Frontiers in Plant Science* 9, p. 132. DOI : 10.3389/fpls.2018.00132.
- KIM, D., B. LANGMEAD et S. L. SALZBERG (2015). "HISAT : a fast spliced aligner with low memory requirements". In : *Nature Methods* 12, p. 357.
- KIM, N. et S. JINKS-ROBERTSON (2012). "Transcription as a source of genome instability". In : *Nature Reviews Genetics* 13.3, p. 204-214. DOI : 10.1038/nrg3152.
- KIND, J. *et al.* (2013). "Single-Cell Dynamics of Genome-Nuclear Lamina Interactions". In : *Cell* 153.1, p. 178-192. DOI : 10.1016/j.cell.2013.02.028.
- KLAPPENBACH, J. A. (2001). "rrndb : the Ribosomal RNA Operon Copy Number Database". In : *Nucleic Acids Research* 29.1, p. 181-184. DOI : 10.1093/nar/29.1.181.
- KOBAYASHI, T. (2011). "Regulation of ribosomal RNA gene copy number and its role in modulating genome integrity and evolutionary adaptability in yeast". In : *Cellular and Molecular Life Sciences* 68.8, p. 1395-1403. DOI : 10.1007/s00018-010-0613-2.
- (2014). "Ribosomal RNA gene repeats, their stability and cellular senescence". In : *Proceedings of the Japan Academy, Series B* 90.4, p. 119-129. DOI : 10.2183/pjab.90.119.
- KONINGSBRUGGEN, S. van *et al.* (2010). "High-resolution whole-genome sequencing reveals that specific chromatin domains from most human chromosomes associate with nucleoli". In : *Molecular biology of the cell* 21.21, p. 3735-3748.
- KOUZARIDES, T. (2007). "Chromatin Modifications and Their Function". In : *Cell* 128.4, p. 693-705. DOI : 10.1016/j.cell.2007.02.005.
- KRASILEVA, K. V. (2019). "The role of transposable elements and DNA damage repair mechanisms in gene duplications and gene fusions in plant genomes". In : *Current Opinion in Plant Biology* 48, p. 18-25. DOI : 10.1016/j.pbi.2019.01.004.
- KRISHNAKUMAR, V. *et al.* (2015). "Araport : the Arabidopsis Information Portal". In : *Nucleic Acids Research* 43 (D1), p. D1003-D1009. DOI : 10.1093/nar/gku1200.
- KRUEGER, F. et S. R. ANDREWS (2011). "Bismark : a flexible aligner and methylation caller for Bisulfite-Seq applications". In : *Bioinformatics* 27.11, p. 1571-1572. DOI : 10.1093/bioinformatics/btr167.
- KRZYWINSKI, M. *et al.* (2009). "Circos : An information aesthetic for comparative genomics". In : *Genome Research* 19.9, p. 1639-1645. DOI : 10.1101/gr.092759.109.

- KUO, L. J. et L.-X. YANG (2008). “ γ -H2AX – A Novel Biomarker for DNA Double-strand Breaks”. In : *in vivo*, p. 5.
- KURIHARA, Y. *et al.* (1994). “Chromosomal locations of Ag-NORs and clusters of ribosomal DNA in laboratory strains of mice”. In : *Mammalian Genome* 5.4, p. 225-228. DOI : 10.1007/BF00360550.
- LANGMEAD, B. *et al.* (2009). “Ultrafast and memory-efficient alignment of short DNA sequences to the human genome”. In : *Genome Biology* 10.3, R25. DOI : 10.1186/gb-2009-10-3-r25.
- LAW, J. A. et S. E. JACOBSEN (2010). “Establishing, maintaining and modifying DNA methylation patterns in plants and animals”. In : *Nature Reviews Genetics* 11.3, p. 204-220. DOI : 10.1038/nrg2719.
- LI, B. *et al.* (2018). “Coregulation of ribosomal RNA with hundreds of genes contributes to phenotypic variation”. In : *Genome Research* 28.10, p. 1555-1565. DOI : 10.1101/gr.229716.117.
- LI, H. (2011). “A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data”. In : *Bioinformatics* 27.21, p. 2987-2993. DOI : 10.1093/bioinformatics/btr509.
- LIEBERMAN-AIDEN, E. *et al.* (2009). “Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome”. In : *Science* 326.5950, p. 289-293. DOI : 10.1126/science.1181369.
- LIU, C. *et al.* (2016). “Genome-wide analysis of chromatin packing in *Arabidopsis thaliana* at single-gene resolution”. In : *Genome Research* 26.8, p. 1057-1068. DOI : 10.1101/gr.204032.116.
- LIU, C. *et al.* (2017). “Prominent topologically associated domains differentiate global chromatin packing in rice from *Arabidopsis*”. In : *Nature Plants* 3.9, p. 742-748. DOI : 10.1038/s41477-017-0005-9.
- LOFGREN, L. A. *et al.* (2019). “Genome-based estimates of fungal rDNA copy number variation across phylogenetic scales and ecological lifestyles”. In : *Molecular Ecology* 28.4, p. 721-730. DOI : 10.1111/mec.14995.
- LONG, E. O. et I. B. DAWID (1980). “Repeated Genes in Eukaryotes”. In : *Annual Review of Biochemistry* 49.1, p. 727-764. DOI : 10.1146/annurev.bi.49.070180.003455.
- LOVE, M. I., W. HUBER et S. ANDERS (2014). “Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2”. In : *Genome Biology* 15.12. DOI : 10.1186/s13059-014-0550-8.
- LUGER, K. et T. J. RICHMOND (1998). “DNA binding within the nucleosome core”. In : *Current Opinion in Structural Biology* 8.1, p. 33-40. DOI : 10.1016/S0959-440X(98)80007-9.
- LUPIÁÑEZ, D. G., M. SPIELMANN et S. MUNDLOS (2016). “Breaking TADs : How Alterations of Chromatin Domains Result in Disease”. In : *Trends in Genetics* 32.4, p. 225-237. DOI : 10.1016/j.tig.2016.01.003.
- LUPIÁÑEZ, D. G. *et al.* (2015). “Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions”. In : *Cell* 161.5, p. 1012-1025. DOI : 10.1016/j.cell.2015.04.004.
- LYCKEGAARD, E. M. et A. G. CLARK (1989). “Ribosomal DNA and Stellate gene copy number variation on the Y chromosome of *Drosophila melanogaster*.” In : *Proceedings of the National Academy of Sciences* 86.6, p. 1944-1948. DOI : 10.1073/pnas.86.6.1944.

- MAEDA, M., T. SHIMADA et A. ISHIHAMA (2015). "Strength and Regulation of Seven rRNA Promoters in *Escherichia coli*". In : *PLOS ONE* 10.12. Sous la dir. de S. SEMSEY, e0144697. DOI : 10.1371/journal.pone.0144697.
- MAO, J. *et al.* (1982). "The 5S RNA genes of *Schizosaccharomyces pombe*". In : *Nucleic Acids Research* 10.2, p. 487-500. DOI : 10.1093/nar/10.2.487.
- MARNEF, A., S. COHEN et G. LEGUBE (2017). "Transcription-Coupled DNA Double-Strand Break Repair : Active Genes Need Special Care". In : *Journal of Molecular Biology* 429.9, p. 1277-1288. DOI : 10.1016/j.jmb.2017.03.024.
- MARNEF, A. *et al.* (2019). "A cohesin/HUSH- and LINC-dependent pathway controls ribosomal DNA double-strand break repair". In : *Genes & Development* 33.17, p. 1175-1190. DOI : 10.1101/gad.324012.119.
- MCCLINTOCK, B. (1934). "The relation of a particular chromosomal element to the development of the nucleoli in *Zea mays*". In : *Zeitschrift für Zellforschung und Mikroskopische Anatomie* 21.2, p. 294-326. DOI : 10.1007/BF00374060.
- MCSTAY, B. (2016). "Nucleolar organizer regions : genomic 'dark matter' requiring illumination". In : *Genes & Development* 30.14, p. 1598-1610. DOI : 10.1101/gad.283838.116.
- MCSTAY, B. et I. GRUMMT (2008). "The Epigenetics of rRNA Genes : From Molecular to Chromosome Biology". In : *Annual Review of Cell and Developmental Biology* 24.1, p. 131-157. DOI : 10.1146/annurev.cellbio.24.110707.175259.
- MÉLÈSE, T. et Z. XUE (1995). "The nucleolus : an organelle formed by the act of building a ribosome". In : *Current Opinion in Cell Biology* 7.3, p. 319-324. DOI : 10.1016/0955-0674(95)80085-9.
- MILLER, O. L. et B. R. BEATTY (1969). "Visualization of nucleolar genes". In : *Science (New York, N.Y.)* 164.3882, p. 955-957. DOI : 10.1126/science.164.3882.955.
- MIROUZE, M. *et al.* (2009). "Selective epigenetic control of retrotransposition in *Arabidopsis*". In : *Nature* 461.7262, p. 427-430. DOI : 10.1038/nature08328.
- MITREA, D. M. et R. W. KRIWACKI (2016). "Phase separation in biology; functional organization of a higher order". In : *Cell Communication and Signaling* 14.1, p. 1. DOI : 10.1186/s12964-015-0125-7.
- MONTACIE, C. (2019). "Proteasome and iron : roles and/or regulations in *Arabidopsis thaliana* nucleolus". Theses. Université de Perpignan.
- MOZGOVÁ, I., P. MOKROŠ et J. FAJKUS (2010). "Dysfunction of Chromatin Assembly Factor 1 Induces Shortening of Telomeres and Loss of 45S rDNA in *Arabidopsis thaliana*". In : *The Plant Cell* 22.8, p. 2768-2780. DOI : 10.1105/tpc.110.076182.
- MUCHOVÁ, V. *et al.* (2015). "Homology-dependent repair is involved in 45S rDNA loss in plant CAF-1 mutants". In : *The Plant Journal* 81.2, p. 198-209. DOI : 10.1111/tpj.12718.
- NELSON, J. O. *et al.* (2019). "Mechanisms of rDNA Copy Number Maintenance". In : *Trends in Genetics* 35.10, p. 734-742. DOI : 10.1016/j.tig.2019.07.006.
- NÉMETH, A. *et al.* (2010). "Initial Genomics of the Human Nucleolus". In : *PLoS Genetics* 6.3. Sous la dir. d'A. AKHTAR, e1000889. DOI : 10.1371/journal.pgen.1000889.
- NOZAWA, R.-S. et N. GILBERT (2019). "RNA : Nuclear Glue for Folding the Genome". In : *Trends in Cell Biology* 29.3, p. 201-211. DOI : 10.1016/j.tcb.2018.12.003.
- PAREDES, S. et K. A. MAGGERT (2009). "Ribosomal DNA contributes to global chromatin regulation". In : *Proceedings of the National Academy of Sciences* 106.42, p. 17829-17834.

- PAREDES, S. *et al.* (2011). "Ribosomal DNA Deletions Modulate Genome-Wide Gene Expression : "rDNA-Sensitive" Genes and Natural Variation". In : *PLoS Genetics* 7.4. Sous la dir. d'A. G. CLARK, e1001376. DOI : 10.1371/journal.pgen.1001376.
- PAREDES, S. *et al.* (2018). "The epigenetic regulator SIRT7 guards against mammalian cellular senescence induced by ribosomal DNA instability". In : *Journal of Biological Chemistry* 293.28, p. 11242-11250. DOI : 10.1074/jbc.AC118.003325.
- PAVLIŠTOVÁ, V. *et al.* (2016). "Phenotypic reversion in *fas* mutants of *Arabidopsis thaliana* by reintroduction of *FAS* genes : variable recovery of telomeres with major spatial rearrangements and transcriptional reprogramming of 45S rDNA genes". In : *The Plant Journal* 88.3, p. 411-424. DOI : 10.1111/tpj.13257.
- PECINKA, A. *et al.* (2004). "Chromosome territory arrangement and homologous pairing in nuclei of *Arabidopsis thaliana* are predominantly random except for NOR-bearing chromosomes". In : *Chromosoma* 113.5, p. 258-269. DOI : 10.1007/s00412-004-0316-2.
- PECINKA, A. *et al.* (2010). "Epigenetic Regulation of Repetitive Elements Is Attenuated by Prolonged Heat Stress in *Arabidopsis*". In : *The Plant Cell* 22.9, p. 3118-3129. DOI : 10.1105/tpc.110.078493.
- PEDERSON, T. (1998). "The plurifunctional nucleolus". In : *Nucleic Acids Research* 26.17, p. 3871-3876. DOI : 10.1093/nar/26.17.3871.
- PEDERSON, T. et J. C. POLITZ (2000). "The Nucleolus and the Four Ribonucleoproteins of Translation". In : *The Journal of Cell Biology* 148.6, p. 1091-1096. DOI : 10.1083/jcb.148.6.1091.
- PENDLE, A. F. *et al.* (2005). "Proteomic analysis of the *Arabidopsis* nucleolus suggests novel nucleolar functions". In : *Molecular biology of the cell* 16.1, p. 260-269.
- PENG, J. C. et G. H. KARPEN (2007). "H3K9 methylation and RNA interference regulate nucleolar organization and repeated DNA stability". In : *Nature Cell Biology* 9.1, p. 25-35. DOI : 10.1038/ncb1514.
- PENG, J. C. et G. H. KARPEN (2008). "Epigenetic regulation of heterochromatic DNA stability". In : *Current Opinion in Genetics & Development* 18.2, p. 204-211. DOI : 10.1016/j.gde.2008.01.021.
- PHIZICKY, E. M. et A. K. HOPPER (2010). "tRNA biology charges to the front". In : *Genes & Development* 24.17, p. 1832-1860. DOI : 10.1101/gad.1956510.
- PICKERSGILL, H. *et al.* (2006). "Characterization of the *Drosophila melanogaster* genome at the nuclear lamina". In : *Nature Genetics* 38.9, p. 1005-1014. DOI : 10.1038/ng1852.
- POMBO, A. et N. DILLON (2015). "Three-dimensional genome architecture : players and mechanisms". In : *Nature Reviews Molecular Cell Biology* 16.4, p. 245-257. DOI : 10.1038/nrm3965.
- PONTVIANNE, F. *et al.* (2012). "Histone methyltransferases regulating rRNA gene dose and dosage control in *Arabidopsis*". In : *Genes & Development* 26.9, p. 945-957. DOI : 10.1101/gad.182865.111.
- PONTVIANNE, F. *et al.* (2013). "Subnuclear partitioning of rRNA genes between the nucleolus and nucleoplasm reflects alternative epiallelic states". In : *Genes & Development* 27.14, p. 1545-1550. DOI : 10.1101/gad.221648.113.
- PONTVIANNE, F. (2016). "Rôle du nucléole dans l'organisation nucléaire chez *Arabidopsis thaliana*". Habilitation à diriger des recherches. Université de Perpignan Via Domitia (UPVD).

- PONTVIANNE, F., T. BLEVINS et C. S. PIKAARD (2010). "Arabidopsis Histone Lysine Methyltransferases". In : *Advances in Botanical Research*. T. 53. Elsevier, p. 1-22. DOI : 10.1016/S0065-2296(10)53001-5.
- PONTVIANNE, F., M. BOYER-CLAVEL et J. SÁEZ-VÁSQUEZ (2016a). "Fluorescence-Activated Nucleolus Sorting in Arabidopsis". In : *The Nucleolus*. Sous la dir. d'A. NÉMETH. T. 1455. New York, NY : Springer New York, p. 203-211. DOI : 10.1007/978-1-4939-3792-9_15.
- PONTVIANNE, F. *et al.* (2007). "Characterization of AtNUC-L1 Reveals a Central Role of Nucleolin in Nucleolus Organization and Silencing of AtNUC-L2 Gene in Arabidopsis". In : *Molecular Biology of the Cell* 18, p. 11.
- PONTVIANNE, F. *et al.* (2016b). "Identification of Nucleolus-Associated Chromatin Domains Reveals a Role for the Nucleolus in 3D Organization of the *A. thaliana* Genome". In : *Cell Reports* 16.6, p. 1574-1587. DOI : 10.1016/j.celrep.2016.07.016.
- POULET, A. *et al.* (2017). "The LINC complex contributes to heterochromatin organisation and transcriptional gene silencing in plants". In : *Journal of Cell Science* 130.3, p. 590-601. DOI : 10.1242/jcs.194712.
- PROBST, A. V. *et al.* (2004). "Arabidopsis Histone Deacetylase *HDA6* Is Required for Maintenance of Transcriptional Gene Silencing and Determines Nuclear Organization of rDNA Repeats". In : *The Plant Cell* 16.4, p. 1021-1034. DOI : 10.1105/tpc.018754.
- PROKOPOWICH, C. D., T. R. GREGORY et T. J. CREASE (2003). "The correlation between rDNA copy number and genome size in eukaryotes". In : *Genome* 46.1, p. 48-50. DOI : 10.1139/g02-103.
- QUINODOZ, S. A. *et al.* (2018). "Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus". In : *Cell* 174.3, 744-757.e24. DOI : 10.1016/j.cell.2018.05.024.
- RABANAL, F. A. *et al.* (2017a). "Epistatic and allelic interactions control expression of ribosomal RNA gene clusters in Arabidopsis thaliana". In : *Genome Biology* 18.1, p. 75. DOI : 10.1186/s13059-017-1209-z.
- RABANAL, F. A. *et al.* (2017b). "Unstable Inheritance of 45S rRNA Genes in Arabidopsis thaliana". In : *Genes/Genomes/Genetics* 7.4, p. 1201-1209. DOI : 10.1534/g3.117.040204.
- RABL, C. (1885). "Über Zelltheilung. Morphologisches Jahrbuch. 10". In :
- RITOSSA, F. M. et S. SPIEGELMAN (1965). "Localization of DNA complementary to ribosomal RNA in the Nucleolus Organizer Region of *Drosophila melanogaster*". In : *Proceedings of the National Academy of Sciences of the United States of America* 53, p. 737-745.
- ROGERS, S. O. et A. J. BENDICH (1987). "Ribosomal RNA genes in plants : variability in copy number and in the intergenic spacer". In : *Plant Molecular Biology* 9.5, p. 509-520. DOI : 10.1007/BF00015882.
- ROSA, S. et P. SHAW (2013). "Insights into Chromatin Structure and Dynamics in Plants". In : *Biology* 2.4, p. 1378-1410. DOI : 10.3390/biology2041378.
- ROSATO, M. *et al.* (2016). "Conserved Organisation of 45S rDNA Sites and rDNA Gene Copy Number among Major Clades of Early Land Plants". In : *PLOS ONE* 11.9. Sous la dir. de X.-Q. LI, e0162544. DOI : 10.1371/journal.pone.0162544.
- SAEZ-VASQUEZ, J. et M. DELSENY (2019). "Ribosome Biogenesis in Plants : from Functional 45S Ribosomal DNA Organization to Ribosome Assembly Factors". In : *The Plant Cell*, tpc.00874.2018. DOI : 10.1105/tpc.18.00874.

- SAKAMOTO, T. *et al.* (2019). "Plant condensin II is required for the correct spatial relationship between centromeres and rDNA arrays". In : *Nucleus* 10.1, p. 116-125. DOI : 10.1080/19491034.2019.1616507.
- SALIM, D. et J. L. GERTON (2019). "Ribosomal DNA instability and genome adaptability". In : *Chromosome Research* 27.1, p. 73-87. DOI : 10.1007/s10577-018-9599-7.
- SALIM, D. *et al.* (2017). "DNA replication stress restricts ribosomal DNA copy number". In : *PLOS Genetics* 13.9. Sous la dir. d'A. K. HOPPER, e1007006. DOI : 10.1371/journal.pgen.1007006.
- SASAKI, M., J. LANGE et S. KEENEY (2010). "Genome destabilization by homologous recombination in the germ line". In : *Nature Reviews Molecular Cell Biology* 11.3, p. 182-195. DOI : 10.1038/nrm2849.
- SCHÜBELER, D. (2015). "Function and information content of DNA methylation". In : *Nature* 517.7534, p. 321-326. DOI : 10.1038/nature14192.
- SCHUBERT, I. *et al.* (2004). "DNA damage processing and aberration formation in plants". In : *Cytogenetic and Genome Research* 104.1, p. 104-108. DOI : 10.1159/000077473.
- SCHUBERT, V. et K. WEISSHART (2015). "Abundance and distribution of RNA polymerase II in Arabidopsis interphase nuclei". In : *Journal of Experimental Botany* 66.6, p. 1687-1698. DOI : 10.1093/jxb/erv091.
- SEQUEIRA-MENDES, J. *et al.* (2014). "The Functional Topography of the Arabidopsis Genome Is Organized in a Reduced Number of Linear Motifs of Chromatin States". In : *The Plant Cell* 26.6, p. 2351-2366. DOI : 10.1105/tpc.114.124578.
- SLUIS, M. van et B. McSTAY (2015). "A localized nucleolar DNA damage response facilitates recruitment of the homology-directed repair machinery independent of cell cycle stage". In : *Genes & Development* 29.11, p. 1151-1163. DOI : 10.1101/gad.260703.115.
- SOCHOROVÁ, J. (2018). "Evolutionary trends in animal ribosomal DNA loci : introduction to a new online database". In : p. 10.
- STEENSEL, B. van et E. E. M. FURLONG (2019). "The role of transcription in shaping the spatial organization of the genome". In : *Nature Reviews Molecular Cell Biology* 20.6, p. 327-337. DOI : 10.1038/s41580-019-0114-6.
- STĘPIŃSKI, D. (2014). "Functional ultrastructure of the plant nucleolus". In : *Protoplasma* 251.6, p. 1285-1306. DOI : 10.1007/s00709-014-0648-6.
- STULTS, D. M. *et al.* (2007). "Genomic architecture and inheritance of human ribosomal RNA gene clusters". In : *Genome Research* 18.1, p. 13-18. DOI : 10.1101/gr.6858507.
- SUZUKI, A. *et al.* (1996). "Structural and Functional Characterization of the Intergenic Spacer Region of the rDNA in *Daucus carota*". In : *Plant and Cell Physiology* 37.2, p. 233-238. DOI : 10.1093/oxfordjournals.pcp.a028937.
- SZCZEPIŃSKA, T., A. M. RUSEK et D. PLEWCZYŃSKI (2019). "Intermingling of chromosome territories". In : *Genes, Chromosomes and Cancer* 58.7, p. 500-506. DOI : 10.1002/gcc.22736.
- TESSADORI, F. *et al.* (2007). "Light-regulated large-scale reorganization of chromatin during the floral transition in Arabidopsis : Chromatin reorganization during floral transition". In : *The Plant Journal* 50.5, p. 848-857. DOI : 10.1111/j.1365-313X.2007.03093.x.
- THOMPSON, M. *et al.* (2003). "Nucleolar Clustering of Dispersed tRNA Genes". In : *Science* 302.5649, p. 1399-1401. DOI : 10.1126/science.1089814.

- TSANG, E. et A. CARR (2008). "Replication fork arrest, recombination and the maintenance of ribosomal DNA stability". In : *DNA Repair* 7.10, p. 1613-1623. DOI : 10.1016/j.dnarep.2008.06.010.
- TUROWSKI, T. W. et D. TOLLERVEY (2015). "Cotranscriptional events in eukaryotic ribosome synthesis : Cotranscriptional ribosome synthesis". In : *Wiley Interdisciplinary Reviews : RNA* 6.1, p. 129-139. DOI : 10.1002/wrna.1263.
- VALORI, V. et al. (2019). *Human rDNA Copy Number Is Unstable in Metastatic Breast Cancers*. preprint. *Cancer Biology*. DOI : 10.1101/623595.
- VERTII, A. et al. (2019). "Two contrasting classes of nucleolus-associated domains in mouse fibroblast heterochromatin". In : *Genome Research* 29.8, p. 1235-1249. DOI : 10.1101/gr.247072.118.
- WANG, C. et al. (2015a). "Genome-wide analysis of local chromatin packing in *Arabidopsis thaliana*". In : *Genome Research* 25.2, p. 246-256. DOI : 10.1101/gr.170332.113.
- WANG, L.-C. et al. (2013). "Arabidopsis HIT4 encodes a novel chromocentre-localized protein involved in the heat reactivation of transcriptionally silent loci and is essential for heat tolerance in plants". In : *Journal of Experimental Botany* 64.6, p. 1689-1701. DOI : 10.1093/jxb/ert030.
- WANG, L.-C. et al. (2015b). "Arabidopsis HIT4, a regulator involved in heat-triggered reorganization of chromatin and release of transcriptional gene silencing, relocates from chromocenters to the nucleolus in response to heat stress". In : *New Phytologist* 205.2, p. 544-554. DOI : 10.1111/nph.13088.
- WANG, M. et B. LEMOS (2017). "Ribosomal DNA copy number amplification and loss in human cancers is linked to tumor genetic context, nucleolus activity, and proliferation". In : *PLOS Genetics* 13.9. Sous la dir. de C. ENG, e1006994. DOI : 10.1371/journal.pgen.1006994.
- WANG, W. et al. (2019). "Remarkable variation of ribosomal DNA organization and copy number in gnetophytes, a distinct lineage of gymnosperms". In : *Annals of Botany* 123.5, p. 767-781. DOI : 10.1093/aob/mcy172.
- WATANABE, K. et al. (2009). "The STRUCTURAL MAINTENANCE OF CHROMOSOMES 5/6 Complex Promotes Sister Chromatid Alignment and Homologous Recombination after DNA Damage in *Arabidopsis thaliana*". In : *The Plant Cell* 21.9, p. 2688-2699. DOI : 10.1105/tpc.108.060525.
- WATKINS, N. J. et M. T. BOHNSACK (2012). "The box C/D and H/ACA snoRNPs : key players in the modification, processing and the dynamic folding of ribosomal RNA : Box C/D and H/ACA snoRNPs". In : *Wiley Interdisciplinary Reviews : RNA* 3.3, p. 397-414. DOI : 10.1002/wrna.117.
- WEIS, B. L. et al. (2015). "Plant-Specific Features of Ribosome Biogenesis". In : *Trends in Plant Science* 20.11, p. 729-740. DOI : 10.1016/j.tplants.2015.07.003.
- XU, B. et al. (2017). "Ribosomal DNA copy number loss and sequence variation in cancer". In : *PLOS Genetics* 13.6. Sous la dir. de C. ENG, e1006771. DOI : 10.1371/journal.pgen.1006771.
- YEUNG, J. et F. NAEF (2018). "Rhythms of the Genome : Circadian Dynamics from Chromatin Topology, Tissue-Specific Gene Expression, to Behavior". In : *Trends in Genetics* 34.12, p. 915-926. DOI : 10.1016/j.tig.2018.09.005.
- YU, S. et B. LEMOS (2018). "The long-range interaction map of ribosomal DNA arrays". In : *PLOS Genetics* 14.3. Sous la dir. de G. Bosco, e1007258. DOI : 10.1371/journal.pgen.1007258.

- ZHANG, H. *et al.* (2019). "The effects of Arabidopsis genome duplication on the chromatin organization and transcriptional regulation". In : *Nucleic Acids Research* 47.15, p. 7857-7869. DOI : 10.1093/nar/gkz511.
- ŽMIENKO, A. *et al.* (2014). "Copy number polymorphism in plant genomes". In : *Theoretical and Applied Genetics* 127.1, p. 1-18. DOI : 10.1007/s00122-013-2177-7.

Annexes

Annexe 1

Tableau 1 – Caractéristiques des bibliothèques de séquences ADN ou ARN

Echantillon	# de séquences après le filtre de qualité avec Trimmomatic	# de séquences alignées sur le génome TAIR10	Couverture d'alignement (X)
N WT	13803877	8249056	18
Nuc 2 WT	14802693	8568483	19
Nucléoles 1 WT	22622115	13062116	29
Noyaux 20%ADNr	16663143	10894583	24
Nucléoles 2 20%ADNr	23959776	13968295	31
Nucléoles 1 20%ADNr	7142317	4586191	10
Noyaux 22°C	16694003	9981574	22
Nucléoles 1 22°C	14373115	8965292	20
Nucléoles 2 22°C	14587596	6901038	15
Noyaux 37°C	10904109	7354816	16
Nucléoles 1 37°C	11198102	7064690	16
Nucléoles 2 37°C	18838247	9251233	21
Nucléoles 1 Post TF	23424785	13748195	31
Nucléoles 2 Post TF	28104170	15219094	34
ARN 22°C 1	46807612	23359468	52
ARN 22°C 2	53452659	27368512	61
ARN 37°C 1	55962086	25891433	58
ARN 37°C 2	45838702	23728651	53
ARN WT1	13120041	3094522	7
ARN WT2	10343273	2506789	6
ARN WT3	13427880	3190052	7
ARN WT4	13618679	3252209	7
ARN 20rDNA L6F6 1	14478130	3445659	8
ARN 20rDNA L6F6 2	9788466	2489506	6
ARN 20rDNA L6F6 3	17332045	4043949	9
ARN 20rDNA L6F6 4	15647767	3872857	9
WGBS WT1	13764960	8052397	18
WGBS WT2	16343090	9726943	22
WGBS WT3	17041530	10260091	23
WGBS 20rDNA L6F6 1	18361132	11578448	26
WGBS 20rDNA L6F6 2	13844885	8557714	19
WGBS 20rDNA L6F6 3	17837081	10883453	24
Hi-C Col1 R1		28689042	64
Hi-C Col1 R2		26353843	59
Hi-C Col2 R1		27256052	61
Hi-C Col2 R2		25644805	57
Hi-C Col3 R1		31592696	70
Hi-C Col3 R2		29811064	66
Hi-C 20rDNA1 R1		36976966	82

Echantillon	# de séquences après le filtre de qualité avec Trimmomatic	# de séquences alignées sur le génom TAIR10	Couverture d'alignement (X)
Hi-C 20rDNA1 R2		34756589	77
Hi-C 20rDNA2 R1		31994373	71
Hi-C 20rDNA2 R2		30307582	67
Hi-C 20rDNA3 R1		42287518	94
Hi-C 20rDNA3 R2		39023549	87

Annexe 2

Tableau 2 – Amorces pour la validation des différences d'expression entre la lignée 20%ADNr et l'écotype sauvage

Gène	Séquence de l'amorce
AT1TE93275 (forward)	CAACACAATTATGGGTCAGGT
AT1TE93275 (reverse)	TCCAAGCCTCTTTCATTCTCA
AT4G05475 (forward)	CCAATGACGAACAAAGGAGTAA
AT4G05475 (reverse)	TCACGAGAGTTTCAAGCAGT
AT3G47210 (forward)	AGCTTCTTCAACCTCTCAATG
AT3G47210 (reverse)	GAAGTTTCTGTGTTTGACCCA
AT5G22560 (forward)	AAGAATGAATCTTGATGCAAGC
AT5G22560 (reverse)	CTTGGTACGAAAGTCTTACGG
AT1G53480 (forward)	GCTCAGTGTTAAGAAGGTACAG
AT1G53480 (reverse)	TCTTCACAACCAAGCTCGAAA
AT4G38410 (forward)	GAAGTTGCTGAAGATCAATGTG
AT4G38410 (reverse)	CATCCTCATGGCCTATACCC
AT4TE09845 (forward)	CACGGTGGATGAATACATCAA
AT4TE09845 (reverse)	GAATCTACGCAAGCACTCTAA
AT4TE10285 (forward)	CGATCACATGAGTTAGCAACTT
AT4TE10285 (reverse)	CCTATTGAGGCGAGGATTATCT
AT4TE11395 (forward)	CCCAATAAGGTAAATCTTCGGG
AT4TE11395 (reverse)	CAACCAGAACCGAAACTGAC

Résumé

Le nucléole est le site de biogenèse des ribosomes, qui commence par la transcription des gènes d'ARN ribosomique (ARNr). Cependant, le nucléole est également impliqué dans d'autres processus cellulaires, comme l'organisation 3D du génome. Ainsi, des régions génomiques appelées NADs pour Nucleolus-Associated chromatin Domains, ont été identifiées dans des cellules animales et végétales. Ces régions sont surtout hétérochromatiques et les gènes associés ont tendance à être peu ou pas transcrits. Un des objectifs de ma thèse a été d'étudier l'implication du nucléole dans l'organisation de la chromatine au sein du noyau et la régulation transcriptionnelle de gènes transcrits par l'ARN Polymérase II chez *Arabidopsis thaliana*. Par ailleurs, parmi les centaines de copies de gènes d'ARNr, uniquement une fraction participe au processus de biogenèse des ribosomes. Dans un second temps, j'ai donc étudié le rôle de ces copies inactives. On a pu démontrer que l'absence des gènes d'ARNr inactifs n'engendre pas de changements majeurs dans la fonction nucléolaire. Par contre, ces copies participent à la stabilité du génome. En effet, en leur absence, des duplications génomiques allant jusqu'à plusieurs centaines de kilobases s'accumulent, entraînant des duplications de gènes et des différences du niveau d'expression de ces derniers. Finalement, les effets de ces changements structuraux sur la biologie de la plante sont discutés.

Abstract

The nucleolus is the site of ribosome biogenesis, which begins with the transcription of ribosomal RNA (rRNA) genes. However, the nucleolus is also involved in other cellular processes, such as the 3D genome organization. Thus, genomic regions called NADs for Nucleolus-Associated chromatin Domains, have been identified in animal and plant cells. These regions are mostly heterochromatic and the associated genes tend to be poorly transcribed. One of the objectives of my thesis was to study the involvement of the nucleolus in the 3D genome organization and the transcriptional regulation of genes transcribed by RNA Polymerase II in *Arabidopsis thaliana*. In addition, only a fraction of rRNA gene copies participates in the process of ribosome biogenesis. In a second time, I studied the role of the inactive rRNA gene copies. We show that in their absence, there is no major changes in the nucleolus function. However, these copies contribute to genome stability. Indeed, in their absence, up to several hundred of kilobases long duplication events accumulate, resulting in the duplication and the differential expression of hundreds of genes. Finally, the impact of these structural changes on the plant biology are discussed.